PROCEDURAL PAIN ASSESSMENT IN INFANTS AND YOUNG CHILDREN: IDENTIFYING A SUITABLE BEHAVIOURAL ASSESSMENT SCALE.

Dianne J Crellin

ORCID: 0000-0003-2629-0483

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

August 2018

The Department of Nursing, Melbourne School of Health Sciences

The University of Melbourne

Abstract

Infants and young children frequently experience pain as a consequence of medical procedures associated with their healthcare. Pain management is often suboptimal, and this is in part due to the difficulties associated with assessment of pain of infants and children too young to self-report pain intensity. Observable behaviours indicative of pain have long been considered a viable alternative and scales comprised of these behaviours have proliferated in the literature. However, it remains unclear which scales are best suited for procedural pain assessment and whether they are well supported by psychometric data.

The aims of this project were to: identify behavioural observation scales potentially suitable for procedural pain assessment, summarise available psychometric data and prospectively test the psychometric properties of potentially suitable scales when used to assess procedural pain in infants and young children. These aims were addressed in three phases of work: i) a thorough interrogation of the literature to identify scales considered potentially suitable for assessing procedural pain in infants and children, ii) a series of systematic reviews to summarise the evidence supporting the psychometric properties of the identified scales and iii) a prospective observational study to test the psychometric properties of these scales used to assess procedural pain in infants and young children.

Three scales, the Face, Legs, Activity, Cry and Consolability Scale (FLACC), the Modified Behavioral Pain Scale (MBPS) and the Visual Analogue Scale for observers (VASobs), met predefined criteria and were considered potentially suitable for inclusion in this project. The systematic reviews showed that available psychometric data was insufficient to recommend these scales for procedural pain assessment of infants and children. There was data to tentatively support the MBPS and to a lesser extent the VASobs for assessing immunisation related pain. The data regarding the FLACC scale was inconclusive.

The results of the prospective study confirmed that all scales were sensitive to pain. The FLACC scale and MBPS scores were reliable (intraclass correlation (ICC) 0.92 and 0.87, respectively) but VASobs scores were less reliable (ICC 0.55). The FLACC scores showed the highest sensitivity (94.9%) and specificity (72.5%) for procedure type (painful vs non-painful) at the lowest cut-off score (pain score 2, area under the curve (AUC 0.83)). Similar results were achieved at a MBPS cut-off score of 4 (sensitivity 91.5%, specificity 77.5%, AUC 0.85). The FLACC scale resulted in more incomplete scores ($p < 0.000$) and was changed more often than

other scale scores. Reviewers liked the VASobs most, considered it the quickest and easiest to apply, but judged the FLACC scale and MBPS to be more likely to be useful.

In conclusion, three behavioural observational pain scales to assess procedural pain in infants and young children were identified and included in systematic reviews. This work culminated in a prospective study, the results of which support use of the FLACC scale, but not without reservation as there are practical limitations when used to assess procedural pain. These results build on promising existing evidence that suggests that the FLACC scale may currently be a suitable scale for procedural pain assessment in infants and young children.

Declaration

This is to certify that:

1. This thesis comprises only my original work towards the Doctor of Philosophy except where indicated in the preface.
2. Due acknowledgement has been made in the text to all other material used.
3. This work has not been submitted previously, in whole or in part, to qualify for any other higher academic degree.
4. The thesis is less than 100,000 words in length, exclusive of tables, bibliographies and appendices.

Name: Dianne Crellin

Signature:  **Date:** 31 August 2018

Preface

This thesis is a compilation of original material written specifically for the thesis and publications arising from the research conducted as part of this project. The Advisory Committee have approved the inclusion of these publications in this thesis and a completed *Declaration for a Thesis with Publication Form,* signed by my principal supervisor, is provided in Appendix G for each publication.

Publications

Publications arising from the work conducted during my PhD candidature and which are included in this thesis are as follows:

Section 3, Chapter 6

1. Crellin DJ, Harrison D, Santamaria N, Babl FE. Systematic Review of the FLACC scale for assessing pain in infants and children: is it reliable, valid, and feasible for use? Pain. 2015;156(11):2132-51.

Section 3, Chapter 7

2. Crellin DJ, Babl FE, Santamaria N, Harrison D. A Systematic Review of the Psychometric Properties of the Modified Behavioral Pain Scale (MBPS). Journal of Pediatric Nursing: Nursing Care of Children and Families. 2018;40:14-26.

Section 4, Chapter 8

3. Crellin DJ, Harrison D, Hutchinson A, Schuster T, Santamaria N Babl FE. Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months. BMJ Open. 2017;7(9).

Section 4, Chapter 10

4. Crellin DJ, Harrison D, Santamaria N, Huque, H and Babl, F E. The psychometric properties of the FLACC scale used to assess procedural pain. J Pain 2018: 19(8):862-72.

Section 4, Chapter 11

5. Crellin DJ, Babl FE, Santamaria N, Harrison D. The Psychometric Properties of the MBPS Scale Used to Assess Procedural Pain. J Pain. 2018;19(6):660-70.

I was lead author for all publications stemming from this thesis and was responsible for writing and revising the manuscript in all cases. My supervisors were co-authors on all publications and contributed to the conception and design for each publication and revision of drafts. The third publication includes additional authors who collaborated in the development of the protocol. Dr Schuster, a Clinical and Epidemiology Unit (CEBU) statistician assisted with the analysis plan and Mr Adrian Hutchinson, Chief Nursing Information Officer, the Royal Children's Hospital, developed the purpose-built electronic data collection tool used in the psychometric evaluation study. Dr Huque made significant contribution to the analysis plan and the manuscript for the publication presented in Chapter 10. Declarations from my fellow authors to acknowledge that I contributed at least 50% to the manuscripts are available in the Appendices (Appendix G).

Invited key note oral presentation

1. Crellin DJ, Harrison D, Santamaria N, Babl FE. Procedural pain assessment in children: the state of the science. Paper presented at the Australian Pain Society Annual Conference, Perth, Australia. 2016

Published abstracts

1. Crellin DJ, Babl FE, Santamaria N, Harrison D. The MBPS: A systematic review to determine its role in assessing pain in infants and children. Poster presented at the International Symposium on Pediatric Pain, Kuala Lumpur. 2017.
2. Crellin DJ, Harrison D, Santamaria N, Babl FE. The FLACC scale: is it reliable and valid used to assess procedural pain experienced by infants and young children? Poster presented at the International Symposium on Pediatric Pain, Kuala Lumpur. 2017.
3. Crellin DJ, Harrison D, Santamaria N, Babl FE. 2017. The psychometric properties of the FLACC scale for procedural pain assessment. Paper presented at the 7th International Nursing Forum, Hong Kong. 2017.
4. Crellin DJ, Harrison D, Santamaria N, Babl FE. The FLACC scale for assessing pain in infants and children; a systematic review. Paper presented at the International Conference for Emergency Nursing, Brisbane, Australia. 2015

5. Crellin DJ, Harrison D, Santamaria N, Babl FE. Procedural pain assessment in children: what's the problem? Paper presented at the International Conference for Emergency Nursing, Brisbane, Australia. 2015
6. Crellin DJ, Harrison D, Santamaria N, Babl FE. The FLACC scale for assessing pain in infants and children; a systematic review. Paper presented at the ACEM Annual Scientific Meeting, Brisbane, Australia. 2015.
7. Crellin DJ, Harrison D, Santamaria N, Babl FE. Procedural pain assessment in children: what's the problem? Paper presented at the ACEM Annual Scientific Meeting, Brisbane, Australia. 2015.

Workshops and local presentations

1. Crellin DJ, Harrison D, Santamaria N, Babl FE. Procedural pain assessment: is there a good option for this? RCH Research Symposium, Melbourne Australia. 2018
2. Crellin DJ, Harrison D, Santamaria N, Babl FE. The feasibility of the MBPS for procedural pain assessment. The University of Melbourne Research Colloquium, Melbourne, Australia. 2017
3. Crellin DJ, Harrison D, Santamaria N, Babl FE. Procedural pain RCH Research Symposium, Melbourne Australia. 2015
4. Crellin DJ, Harrison D, Santamaria N, Babl FE. A systematic review of the FLACC scale for assessing pain in infants and children. The University of Melbourne Research Colloquium, Melbourne, Australia. 2017.
5. Crellin D. Acute paediatric pain management in the emergency department. Acute pain management workshop. Australian Pain Society Conference. Hobart Australia. 2014

Funding and other support

During my candidature I received a small clinical funding grant from the Clinical Sciences theme, Murdoch Children's Research Institute (MCRI). This grant was used to support data collection for the psychometric evaluation study.

I was awarded a travel grant by the Melbourne School of Health Sciences which supported attendance at the International Symposium of Pediatric Pain to present two posters.

For the duration of my candidature I was also supported as a member of Pain in Child Health (PICH), a Canadian research training initiative that unites the international paediatric pain

research community. This support included a travel grant to attend the International Forum on Pediatric Pain and a PICH workshop in Halifax Canada.

Ethics approval

Ethics approval was not required for the comprehensive literature search or the series of systematic reviews which are presented in Section 2 and Section 3 of this thesis.

Ethics approval for the psychometric evaluation study, which is reported in Section 4 of the thesis, was obtained from the Human Research and Ethics Committee of the Royal Children's Hospital, Melbourne Australia. The certificate of approval and approval for minor modifications are available in Appendix F.

Acknowledgements

There are numerous people that I would like to thank for their support, guidance and encouragement over the duration of my candidature.

My supervisors: Professor Nick Santamaria, Professor Denise Harrison and Associate Professor Babl, have provided immeasurable support, expert advice and guidance throughout this work. Their availability, commitment to assisting me to achieve my goals and never-ending encouragement has ensured completion of the research, publication of the results and submission of the thesis. I would like to extend particular thanks to Associate Professor Babl for his longstanding and unwavering confidence in my capacity as a clinical researcher. He has been a colleague, a collaborator, a mentor, an inspiration and a friend for many years.

The psychometric evaluation study could not have occurred without the assistance of a number of people. Associate Professor Susan Donath, Deputy Director Clinical Epidemiology and Biostatistics Unit (CEBU) provided invaluable support to this work, committing the resources of her department to assist with the analysis plan and analysis of the final data set. Dr Tibor Schuster and Dr Hamidul Huque also deserve considerable thanks for their help designing the analysis plan, assisting with the analysis and helping me to interpret the results. Dr Hamidul's patience explaining complicated concepts to a relative novice was tireless. Mr Adrian Hutchinson deserves considerable thanks for his work developing an electronic data management system for the prospective phase of the project. This saved countless hours of my time entering data manually. This system also provided a seamless environment for the emergency department (ED) clinicians in which to review the video segments and enter their scores. The staff who recruited and filmed some of these procedures also deserve acknowledgement for their contribution. I would also like to thank the study participants for their willingness to take part in this study. This includes the children and their families who were filmed during their ED visit and the clinicians and researchers who gave up their valuable time to review segments of video.

My colleagues and friends who encouraged and supported me over the duration of my candidature were a constant source of motivation. Ms Jayne Hughan, nurse unit manager of the emergency department, deserves specific thanks, as without her willingness to allow me to creatively and flexibly manage my work commitments throughout my candidature this work would not have been possible.

Finally, I am immensely grateful to my partner for his enthusiasm and generous support. I have received enormous encouragement and tireless practical support from him without a single word

of complaint. In many ways he deserves this higher degree as much as I do. My mother also deserves my thanks as she contributed her support on numerous occasions and has always been a source of personal and professional inspiration.

and 🐓 🐓 🐓 🐄 🐄

Table of contents

List of Figures

List of Tables

SECTION 1.

This section of the thesis is presented in two chapters: the first of which illustrates the rationale for this work, articulates the research questions and outlines the approach to the research and the layout of the thesis. The second chapter in this section provides a summary of the concepts that underpin this project: pain, the assessment of pain and the methods used to test the performance of (pain) assessment tools.

CHAPTER 1.

The assessment and management of the pain experienced by infants and children during painful procedures has gained increasing attention from clinicians and researchers over the last 20 years and yet there are still major deficits in our understanding and our practice. The focus of this research is procedural pain assessment in infants and young children. This chapter outlines the rationale for this and articulates the specific research questions. The phases of the research undertaken to answer these questions and the layout of this thesis are also outlined in this chapter.

1.1 Rationale

Infants and young children frequently experience pain as a consequence of medical procedures associated with their healthcare. Current immunisation schedules include multiple injections at regular intervals throughout infancy and early childhood (1). Hospitalisation also results in frequent exposure to painful diagnostic and therapeutic procedures. Prevalence of pain is high in neonates, infants and young children who make up the bulk of paediatric hospital admissions. (2-8). Prevalence studies focusing on neonates have reported that the average neonate in a neonatal intensive care unit (NICU) experiences an average of 5.6 painful procedures per week (2) but up to as many as 16 per day (3, 7, 9). Infants and children in other parts of the hospital fare better but nonetheless still experience a substantial number of painful procedures during their admission. In a Canadian study of children admitted to one of 32 inpatient units across 8 paediatric hospitals, 78% of children had at least one painful procedure in the 24 hours prior to data collection and the average number of procedures for those children who had a least one procedure was 6.3 (8). In a cohort of 252 children presenting to a Mexican emergency department, 369 painful procedures were performed (5).

There is strong evidence illustrating the negative consequences of exposure to pain during infancy. Over 20 years ago, Taddio and colleagues reported the impact that neonatal circumcision had on immunisation related pain in infancy (10, 11). Based on post-hoc analysis of data derived from a randomised controlled trial (RCT), they reported higher pain intensity scores during 4 – 6 month immunisation in male infants who were circumcised in the newborn period than male infants who had not been circumcised (11). In their second study designed specifically to test this association between circumcision and immunisation, scores were again higher in infants who had been circumcised as a neonate (10). In addition, repeated heel lances in the first 24 to 36 hours of

life resulted in more intense pain responses to venepuncture than was recorded in newborns who did not receive repeated heel lances (12).

Several research groups have concentrated their research efforts on developing an understanding of the impact of early exposure to pain on neurocognitive development. Valeri and colleagues, in a systematic review of 13 studies addressing the impact of neonatal pain exposure on developmental outcomes, found strong associations between the number of painful procedures and delayed post-natal growth, poorer early neurodevelopment and poorer cognitive and motor function at one year of age, and higher levels of cortical activation and changes in cortical rhythmicity and cortical thickness in 7 year old children (13). Animal models and the advent of newer imaging modalities have provided a mechanism for exploring the impact of painful procedures on neurodevelopmental physiology that may be responsible for the adverse outcomes. The premature neonate's brain continues to undergo significant maturation which involves; the programming of the hypothalamus-pituitary-adrenal axis, the formation of new synaptic connections, activity related selective apoptosis to shape the brain, proliferation and differentiation of glial cells, sub-plate neurons, elaboration of dendrites and axons and alignment, orientation and layering of cortical neurons (14). It seems likely that stress and inflammation associated with injury and pain play a key role in interrupting these activities of development. White and grey brain matter changes are seen on MRI studies of premature neonates and these changes persist into adulthood (14).

There is also a body of evidence illustrating the effects of unrelieved pain in adults, much of which concentrates on the impact of postoperative pain (15). Acute physiological effects include hyperalgesia, hyperglycaemia, protein catabolism, changes in water and electrolyte flux and increased sympathetic activity (16-20). Postoperative cognitive dysfunction has been linked to postoperative pain; the aetiology of which is not well understood but is likely the result of dysregulation of cognitive neurotransmitters (21, 22). The psychological effects of unrelieved pain include alterations to sleep patterns (23), increased anxiety (24) and, importantly, an increased risk of the development of persistent pain (25, 26).

Much of the work exploring the negative effects of procedural pain has concentrated on the short and long-term effects of unrelieved procedural pain experienced by preterm and term neonates or unrelieved postoperative pain on adults (15). However, a study of the psychological effect of serious illness on older children confirms the deleterious effects of pain on older age groups of children (27). The number of invasive painful procedures was shown to be the strongest predictor of significant psychological sequelae at 6 weeks and 6 months following children's discharge from paediatric critical care units (27).

Despite the frequency with which infants and children are exposed to painful procedures, the implications of unrelieved pain and the proliferation of evidence supporting strategies to minimise procedural pain, there is still considerable concern about procedural pain management. In Johnston and colleagues' neonatal pain prevalence study, half of the procedures were performed without analgesic interventions (2) and a quarter of the children in the Canadian survey of hospitalised children reported having received no analgesic or local anaesthetics prior to their last needle-stick procedure (4). One of the potential barriers to providing effective pain management is the capacity to accurately recognise pain and measure its intensity to determine the need for, and the effectiveness of, pain relief.

1.1.1 Procedural pain assessment

Self-report of pain is considered ideal but is not an option in patients either too unwell or cognitively or developmentally unable to self-report. In these circumstances clinicians and researchers are reliant on proxy measures as an indirect surrogate for estimating the patient's pain experience. A range of methods have been proposed, developed and/or tested to quantify pain intensity; the most commonly used and recommended of which are observational behavioural scales. Large numbers of these scales have been developed to assess pain in infants and children and many of these scales have been tested and/or are used in a range of circumstances to assess pain. However, there is no consensus as to the scale best suited to procedural pain assessment.

Two systematic reviews published in 2007 synthesised the evidence regarding the psychometric properties of available observational scales to generate recommendations for use. Von Baeyer and Spagrud recommended the Face, Legs, Activity, Cry, Consolability (FLACC) scale (28) and the Children's Hospital of Eastern Ontario Pain Scale (CHEOPS) (29) for assessment of procedural pain in children aged three to 18 years (30). They claimed excellent reliability, validity and responsiveness for these two pain assessment tools. The FLACC scale is acknowledged as less burdensome as it is based on five easily recalled items all scored on a scale of 0 to 2 (30). In contrast, the strength of the CHEOPS is the reliance on easily observed behaviours that do not rely on a judgement in the same way that the consolability item of the FLACC scale does (30). Crellin and colleagues were more cautious and claimed that these scales while showing promise were not supported by sufficient evidence to claim them as appropriate for procedural pain assessment use in infants and young children (31). Despite this, in the absence of an appropriate alternative they recommended these scales for clinical purposes.

Although these reviews were conducted approximately 10 years ago and these scales are likely to have undergone further testing, more recent reviews to re-evaluate our understanding of the performance of these scales to assess procedural pain do not appear to be available. Professional organisations responsible for recommendations regarding pain assessment and management do not make consistent recommendations without reservations for a specific scale for procedural pain assessment. The Royal College of Nursing in their clinical practice guideline aimed at advising clinicians regarding the assessment and management of pain in infants and children recommended the FLACC scale, CHEOPS and the University of Wisconsin Pain Scale (32). However, significant limitations to the studies on which recommendations were based were noted. More recently, the Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine published the 4th edition of their synthesis of the evidence for acute pain management and in this document, they acknowledged that 'there is no gold standard' available for procedural pain assessment in infants and children but that based on current data the FLACC scale and the CHEOPS are most suitable (15).

1.2 Study aims/questions

The aim of this project was to identify observational assessment scales with adequate psychometric properties to recommend the scale for assessing procedural pain in infants and young children who are unable to self-report pain.

This aim generated three research questions which are as follows:

1. Is there an observational pain assessment scale considered suitable for assessing the procedural pain experienced by infants and young children?
2. Is this scale/Are these scales supported by sufficient psychometric data to recommend the scale for use?
3. Can the selected observation scales be recommended for procedural use following psychometric testing?

1.3 Overview of study/methods

This project was conducted in a series of sequential steps aimed at answering the research questions.

The first of these steps was to systematically search the literature to identify scales that were considered *potentially suitable* for procedural pain assessment in infants and young children. Criteria were established to define considered 'suitable' and they included scale design characteristics, evidence of psychometric testing for procedural use and recommendations supporting the scales use.

The second of these steps was to establish the strength of evidence supporting the psychometric properties of the identified scales. This was achieved by systematic reviews. A robust search strategy was developed to identify relevant literature and study quality was reviewed using appropriate quality assessment tools. Studies that aimed to assess the psychometric properties of the scale were included in these reviews. In addition, high quality randomised controlled trials (RCT) that used the scale to measure a study outcome were also included as they potentially contribute validation data. Data was extracted from the studies, reviewed and a narrative synthesis completed the analysis of the psychometric properties of each scale.

The third and final stage of this research was a study designed to test the psychometric properties of the chosen scales (identified in phase 1 and systematically reviewed in phase 2) applied to infants aged 6 to 42 months who underwent a painful procedure. It was intended that the results of this phase would answer the third research question. This study was designed to assess the feasibility of application and clinical utility of these scales and the reliability, responsiveness, discrimination and convergence of scores. A convenience sample of twenty-six clinicians viewed segments of video depicting 100 children aged six to 42 months undergoing a procedure and applied the FLACC scale, the MBPS and the VASobs to score the infant or child's pain intensity. Procedures considered painful and non-painful were purposefully include in this study to determine whether the scales could discriminate between pain and non-pain related distress.

1.4 Thesis outline

This thesis is comprised of 13 chapters, organised into five sections which include a literature review, sections addressing each research question and a final discussion.

Section 1 includes 2 chapters, the first of which addresses the rationale for this work and the research questions that the project is intended to answer (Chapter 1). An overview of the phases of this project and the methods used to address each question is also provided in this chapter. The second chapter of section 1 briefly explains our current understanding of pain and the methods

used to assess pain intensity (Chapter 2). A summary of the principles of and the methods used to assess the psychometric properties of assessment tools are also provided in this chapter.

Section 2 of this thesis details the first stage of this work in a single chapter (Chapter 3). This documents the systematic search for a scale considered potentially suitable for assessing procedural pain in infants and young children unable to self-report pain. From this work three scales were identified as potentially suitable for this purpose; the FLACC Scale, the MBPS and the VASobs.

Section 3 of this thesis addresses the second research question of this thesis. Chapter 4 reports the methods used to conduct a series of systematic reviews to summarise the psychometric properties of the FLACC scale, the MBPS and the VASobs. Chapters 5 to 7 report the results of the systematic reviews for each scale. Chapter 6 is a Portable Document Format (PDF) version of the FLACC scale review published in *Pain* and Chapter 7 is a PDF version of the MBPS review published in *Journal of Pediatric Nursing*.

Section 4 of this thesis details the psychometric testing of the FLACC scale, the MBPS and the VASobs in response to the third research question. Chapter 8 is a PDF of the protocol used for the observational study, which was published in *BMJ Open*. Chapters 9 to 11 report the results of the psychometric testing of the FLACC scale, MBPS and the VASobs. The FLACC scale and the MBPS results were published in the *Journal of Pain* and are presented as PDF versions of these papers. In Chapter 12 the results of a comparison of the psychometrics of these scales are reported.

This thesis concludes with **Section 5** which is presented in a single chapter (Chapter 13). This chapter summarises the key findings of this research and discusses the implications of these findings. Recommendations for clinical practice and future research are also reported.

Ancillary documents and supplementary figures and tables of data which are not key to the thesis but are referred to are included in a series of Appendices.

CHAPTER 2.

This chapter provides an overview of pain, briefly describes the neurophysiology and the pain experience, summarises methods of pain intensity assessment and finally identifies the approach to developing and testing the psychometric performance of observational (pain) scales. It is intended to serve as general background to the thesis and not as a detailed appraisal of the literature regarding these topics or as they relate specifically to paediatric pain. More detailed reviews of relevant literature are presented prior to each phase of this research.

2.1 Pain

The International Association for the Study of Pain (IASP) provides the most widely accepted definition of pain: '*an unpleasant sensory and emotional experience associated with potential and actual tissue damage or described in terms of such damage*' (33). Traditionally, pain was thought to be experienced only in the setting of potential or actual tissue damage and to reflect the extent of the damage, i.e. the more damage to the tissue the more severe the pain and vice versa (34). Advances in our understanding of pain have resulted in acceptance that pain is a much more complex experience potentially influenced by factors not directly related to tissue damage, that this description better explains nociception and that these terms (pain and nociception) are not synonymous (35). In the following sections pain and nociception are explored more fully to better understand the IASP definition of pain which takes us beyond the traditional nociceptive definition of pain.

2.1.1 Nociception

Pain is an integral part of the body's defence mechanisms and demonstrates a stimulus-response relationship where noxious stimuli trigger a cascade of neuro-physiological actions to prevent or limit tissue damage. Nociception describes the process behind physiological pain and is defined as the 'neural encoding of noxious stimuli', where noxious stimuli are defined as an 'actual or potentially tissue damaging event' (33). Transduction, transmission, modulation and perception describe the neurophysiological sequence that characterise nociception (36).

2.1.1.1 Transduction and transmission

Nociceptors are specialised high-threshold sensory receptors of primary somatosensory neurons located as free nerve endings in the tissues (cutaneous and visceral) (37). Nociceptive processing begins with activation of the nociceptor by the noxious stimuli and *transduction*, which is encoding of the noxious stimuli as a receptor potential (38). Nociceptors are variably responsive to specific stimuli such as mechanical (stretching, cutting or pinching), thermal (extremes of temperature) and/or chemical (exogenous and endogenous chemicals) stimuli (38). It was traditionally thought that receptors responded to only one type of stimulus. However, it is increasingly recognised that most receptors are polymodal and will respond to all types of stimuli given the right circumstances, e.g. intensity of stimulus (39). A subset of receptors found in joints and viscera are considered 'silent' and only respond and transduce stimuli once they have been sensitised to inflammatory mediators (37-39). The receptor potential is then conveyed as an action potential via the peripheral nervous system to the spinal cord; *transmission* (38).

The cell bodies of the primary somatosensory neurons, which are located within the dorsal root and trigeminal ganglia, give rise to axons that are classified as either; myelinated Aδ-fibres or unmyelinated polymodal C-fibres (37). These two classes of fibres respond differently to noxious stimuli. The nociceptors of the Aδ-fibre are triggered by noxious stimuli, have a narrow receptive field, rapidly transmit the signal and only continue to signal in the presence of the noxious stimulus. These fibres are responsible for initial sharp well localised pain that occurs immediately after the stimulus (39). The receptors associated with these fibres are largely located in the skin. In contrast, nociceptors of C-fibres are recruited more slowly, continue to send electrical signals beyond the termination of the noxious stimulus and have a larger receptive field. These fibres are responsible for the dull ache that develops more slowly and is also slower to resolve (39). Activation of the nociceptor also provokes release of neuropeptides substance P calcitonin gene-related peptide (CGRP) at the peripheral terminal, which triggers neurogenic inflammation (39). Release of substances from injured and inflamed tissues such as: bradykinin, cytokines, tumour necrosis factor (TNF), prostaglandins and histamine act to either lower the activation threshold of these receptors or increase excitability, which in both cases amplifies nociceptive signalling (40).

The axons of these fibres transmit the electrical signal to second order neurons in the dorsal horn of the spinal cord predominantly in laminae I and II (40, 41). Synaptic transmission of the sensory signal across the synaptic junction is triggered by release of the nociceptors' primary neurotransmitter, glutamate (37, 41). The post-synaptic terminals are found in different types of dorsal horn neurons: interneurons, propriospinal neurons and projection neurons. Interneurons

may be either excitatory or inhibitory and participate in local processing, e.g. transmission of nociceptive signals received in laminae I and II to second-order neurons in laminae IV, V and VI (42). Propriospinal neurons extend over spinal segments and are involved in segmental reflex activity. Finally, projection neurons are responsible for transmission of signals beyond the spinal cord to the brain via the spinothalamic tract and several associated tracts which includes the spinobulbar tract (43, 44).

The second order projection neurons of the spinothalamic tract decussate in the anterior white commissure at about the level that they enter the spinal cord and ascend in the anterolateral quadrant of the cord (43). They pass through the brainstem, including the medulla oblongata, pons and midbrain, and terminate in the ventral caudal nucleus of the thalamus where they synapse with third order neurons which project to the primary somatosensory cortex. The exact site in the thalamus depends on the location of the original stimulus (e.g. cutaneous versus visceral and body versus face and head) (43). The somatosensory cortex, specifically somatosensory areas I and II in the post central gyrus and superior wall of sylvian fissure are the main sensory receptive regions of the brain responsible for the sense of touch and along with the spinothalamic tract, is somatotopically organised to make localisation of sensation, in this case pain, possible (45).

Fibres of the spinobulbar tract that project to the pons and mesencephalon enter the posterior horn and decussate at the level of entry and synapse with neurons in the anterior horn in a similar pattern to the spinothalamic tract fibres (43). In contrast, the fibres of this pathway that terminate in the medulla ascend via an ipsilateral tract. The fibres of the spinobulbar tract terminate in four regions of the brainstem: the catecholamine cell groups, the parabrachial nucleus (PB), the periaqueductal gray (PAG) and the reticular formation (43). These regions of the brain act to integrate sensory data with homeostatic functions and trigger autonomic cardiorespiratory responses. Collateral projections from the PB link the hypothalamus, amygdala and regions of the thalamus that relay signals to the anterior cingulate cortex, the insular cortex and the forebrain (43). These terminations help explain the autonomic and emotional responses associated with pain.

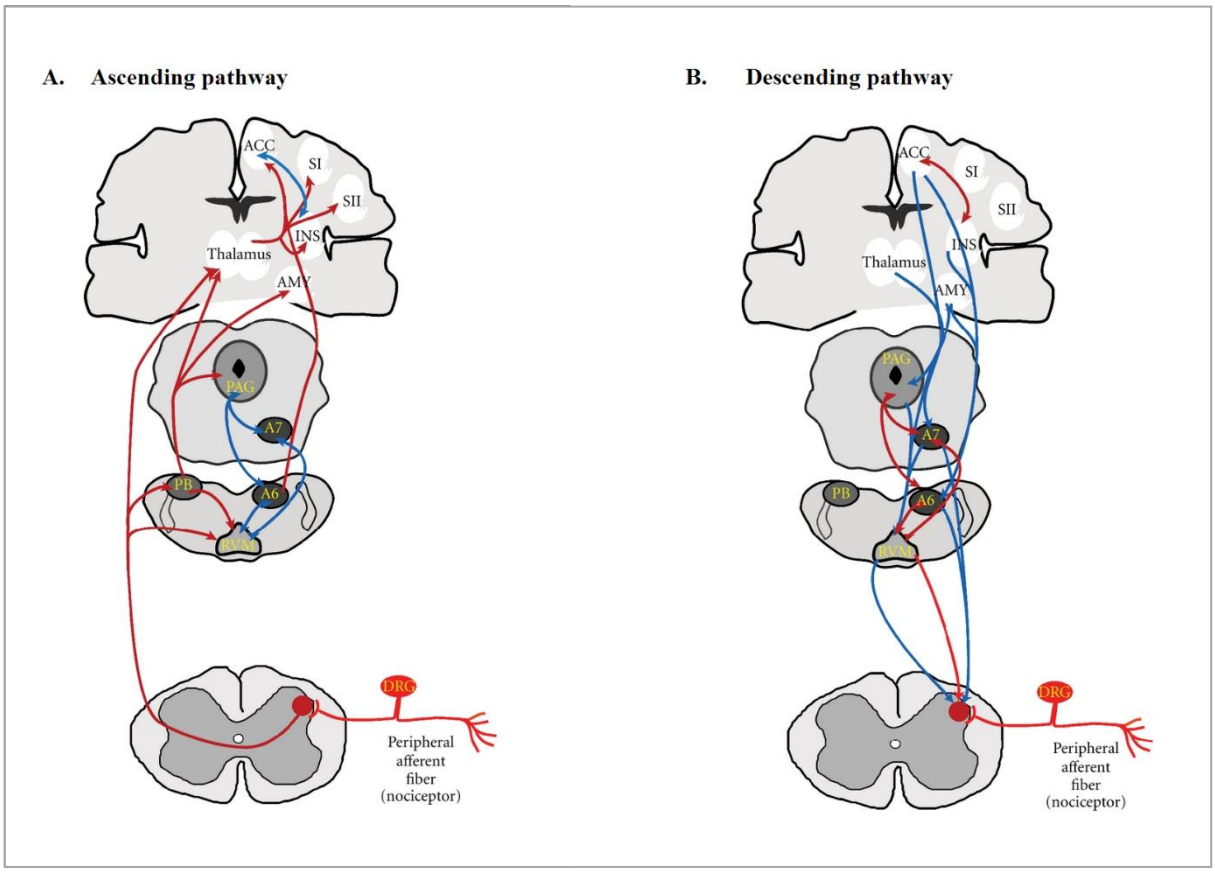


Figure 2-1 A diagram showing the major ascending and descending pain pathways

The ascending tracts in A are represented in red, and the blue 2-headed arrows indicate bilateral communications. Descending projections in B are shown in blue, and the 2-headed arrows in red indicate bilateral communications.

Abbreviations: A6 and A7: noradrenergic nuclei, ACC: anterior cingulate cortex, AMY: amygdala, DRG: dorsal root ganglion, INS: insular cortex, PAG: periaqueductal grey, PB: parabrachial nuclei, RVM: rostroventromedial medulla, SI: primary somatosensory cortex, and SII: secondary somatosensory cortex.

Taken from: Ossipov MH. The Perception and Endogenous Modulation of Pain. Scientifica. 2012;2012:25. Figure 1. The major ascending (a) and descending (b) pain modulatory systems are illustrated in this schematic representation. p.6 (40)

2.1.1.2 Pain perception

Pain can be categorised as either somatic or visceral and this effects pain *perception* (46). Somatic pain defines pain that originates from the skin, muscles or peripheral nerves (cutaneous, superficial pain) or deeper structures such as tendons and fascia (deep pain). Visceral pain defines pain that originates from the visceral organs. Somatic pain is characterised by perception of an initial sharp pain that is well localised followed by a more defuse burning or aching sensation.

The first pain reflects stimulation of Aδ-fibres and transmission of these signals via the spinothalamic tract to the sensory cortex (37, 43). Stimulation of C-fibres results in slower transmission of signals to the cortex and in a delayed perception of a burning pain. Signals from the tendons, fascia and joint receptors are also perceived as dull aching or burning pain. In addition to pain, signals from deeper tissues may also trigger autonomic responses such as sweating and changes in heart rate and blood pressure. Visceral pain arises from mechanical and chemical stimulation of nociceptors in visceral organs (37, 43). These signals are poorly localised and are usually accompanied by autonomic responses (47). In some circumstances visceral pain is referred to sites other than the original site of injury or inflammation. Predictable patterns for referred pain are recognised e.g. angina perceived as pain in the chest wall and down the left arm (42). One explanation postulated for this is that this may be the result of convergence in the spinal cord of the pathways carrying noxious stimuli for the viscera and the sites where visceral pain is referred (47).

It would be short-sighted not to recognise the affective impact of pain signalling and how this impacts on pain perception. Several regions of the brain have been shown to participate and observations made in animal and human studies have elucidated the role that they play in the pain experience, which will be addressed in more detail in section 2.1.2 (48). Almost two decades ago, Ploner and colleagues described the case of an adult, with injury to regions of the somatosensory cortex as a result of stroke, experiencing 'unpleasantness' and a desire to avoid a thermo-nociceptive stimulus, despite not recognising this stimulus as 'painful' (49). Conversely, over 50 years ago Foltz and colleagues described that ablation of the anterior cingulate cortex in patients with intractable pain resulted in immediate relief (50). These patients described a sensation of pain but without the associated unpleasantness. Pain is in most circumstances perceived as both noxious (sensory component) and unpleasant (affective component) and both contribute to the more complex constellation of sensations, emotions and responses that defines the pain experience (section 2.1.2).

2.1.1.3 Pain modulation

As has already been alluded to pain is not exclusively the product of a stimulus-response relationship. Complex mechanisms suppress and potentiate the passage of pain signals to the brain, the process referred to as *modulation*. This may in part explain the wide variation in the pain experienced by individuals in response to the same stimulus. Delineation of the neuronal pathways, the neurotransmitters and chemical mediators involved in modulation are some of the most significant advances to our understanding of the science of pain made by researches in the last few decades. Modulation involves a complex array of interconnected regions of the brain and

chemical mediators directly and indirectly involved in nociceptive neurotransmission which include; endogenous opioids, noradrenaline, adrenaline, serotonin, dopamine and inflammatory mediators (36).

The higher centres implicated in descending modulation of pain are the periaqueductal gray (PAG) in the midbrain, the nucleus raphe magnus (NRM) in the rostral ventromedial medulla (RVM), the locus coeruleus and the nucleus reticularis gigantocellularis (Rgc) (51). The PAG is located in the midbrain and is key to central modulation of pain. The PAG is rich in opioid receptors and highly responsive to exogenous and endogenous opioids. Release of opioids in the PAG provokes powerful anti-nociception (40, 51). The PAG receives ascending signals and signals from the hypothalamus, thalamus, the anterior cingulate cortex, the insula and collaterals of the spinothalamic tract. Integration of these signals provokes opioid-mediated inhibition of nociceptive signalling via enkephalanergic projections to synaptic junctions in the NRM and the locus coeruleus (51). Serotoninergic projections from the post-synaptic neurons of the NRM coupled with noradrenergic projections from the post-synaptic neurons of the locus coeruleus transmit inhibitory signals to block the dorsal horn nociceptive neurons from transmitting ascending nociceptive signals (40, 51).

In addition, stimulation of the PAG and the NRM triggers central and spinal cord release of endogenous opioids (44). At least 10 endogenous opioids (e.g. β-endorphins, enkephalins and dynorphins) and their preferred receptors (e.g. μ, δ and κ) have been identified by researchers. Opioid receptor activation results in inhibition of the presynaptic calcium and potassium channels preventing the release of pain-related neurotransmitters into the synaptic junction, hence blocking the passage of pain signals. This occurs in the central and peripheral nervous system (40).

Immediately following tissue injury an area of inflammation develops. Algogenic substances, such as: histamine and bradykinin which sensitise C-fibres and reduce their threshold, are released from damaged and inflamed tissues (52). Peripheral sensitisation of nociceptors and their neurons caused by repeated stimulation results in: decreased threshold, increased signal frequency, decreased latency to response and spontaneous firing despite the cessation of the noxious stimuli (39, 53). The results of changes in signalling in both circumstances results in increased painful sensation in response to additional noxious stimuli and this is known as hyperalgesia (39). This modulation of responses may well play a key role in the development of chronic pain (39).

2.1.2 Pain experience

Historically, pain was considered to reflect an evolutionary mechanism to reduce tissue injury and this gave rise to the view that the experience of pain is a direct function of the extent of the tissue injury (54). Furthermore, it was considered that noxious stimuli triggered a predictable response that would result in a consistent sensation (severity and nature) entirely dependent on the stimulus (55). These early descriptions of pain have repeatedly been shown to represent an overly simplistic understanding of the experience of pain. Increasing understanding of the neurophysiology of pain, the extensive regions of the brain involved in pain processing and extensive and meticulous observations made of individuals experiencing pain have paved the way for recognition of influences well beyond direct stimulation of nociceptors to explain the pain experience. It is a complex interplay between sensory perception and cognitive and emotional processing and behavioural and physical reactions to pain and this section addresses some of the key elements that contribute to the experience. Although these elements are inseparable as they each have a profound influence on the contribution that make to the pain experience, for the purposes of this discussion these elements will be grouped and discussed in the following sections.

2.1.2.1 Cognitive processing

Attention has long been considered a function of the resources allocated to processing of neural data, the intention of which is to achieve goal directed actions (behavioural and physiological) that are consistent with the stimuli (56, 57). Pain by its aversive nature demands attention which is aimed at provoking an appropriate response to prevent additional harm. The corollary to this is that when attention is focused on pain it is experienced more intensively (56, 58). Furthermore, paradoxically, attempting to suppress thoughts about pain may actually increase the pain experienced (59). Functional MRI results have shown that people distracted from the noxious stimulus have less neural activity in the regions of the brain involved in sensory and affective pain processing (somatosensory cortex, insula and anterior cingulate cortex, and the thalamus) and reported lower levels of pain than people whose attention was not refocused (60-62). Concurrently increased activation of the prefrontal cortex, anterior cingulate cortex and PAG is seen in the setting of distraction. Studies in children undergoing painful procedures repeatedly report the positive effects of music, television, video games and virtual reality on the pain experience associated with these procedures (63-67). There is strong evidence that supports an interaction between attention and the descending facilitatory or inhibitory modulation of pain. However, there is a similar body of data that highlights that this interaction is not uniform and that there may be marked individual differences in the impact of attention and that this may also be situational (68).

Pain signals provoked by sensory data originating from the periphery are interpreted consciously and/or unconsciously and based on the likelihood that they represent actual or potential tissue damage (59). Interpretation will be heavily influenced by past experience, memories of these experiences, expectations of the pain and personality factors such as the individual's perception of their ability to cope with the pain. This interpretation of the meaning of pain signals modulates the experience such that these signals may generate a more or less unpleasant and painful experience across a range of individuals who perceive the meaning of the sensory signals differently.

Experience and the memory of previous events is an opportunity to learn and hence institute appropriate behaviours to manage potentially or actually harmful circumstances. There is evidence that pain memory activates unique centres in the brain when compared with memories of a non-noxious nature or imagined pain (69). However, evidence published more than a decade ago showed that memory of previous painful experience and the efficacy of treatment is often exaggerated (59). Furthermore, experience also lays the foundation for anticipatory fear and anxiety which can have a significant effect on pain experience (1, 70).

How we interpret pain will also be influenced by our expectations for the pain, e.g. how long it might last or how intensely it might be felt. Studies have reported strong links between the level of pain expected, activation of regions of the brain responsible for pain processing and the pain intensity reported (71-73). This may also underlie the mechanism for the effectiveness of placebo medications whereby the expectation of efficacy is sufficient to cause a reduction in the pain experienced (56). Expectations are influenced by previous experience, learning, culture and social setting (74) which contribute to our schematic model for pain that allows us to assess the salience of the pain and make predictions about painful experiences. For example, more severe pain may be considered indicative of more severe illness or injury. Concerns such as this may also feed concern, fear and anxiety based on the perceived threat associated with the pain. This is particularly the case when sensory data conflicts with expectations. Two studies, both conducted more than 10 years ago, reported increased pain scores for the same intensity stimulus when participants were cued to expect a much lower intensity stimulus compared with more accurate cues regarding intensity (72, 75).

The degree of threat that pain poses is in part modified by the individual's perceived ability to control the pain (56). Perceived control over the painful stimuli can decrease the pain experience. Salomons reported reduced activity in the ACC, insula and somatosensory cortex (76) and in a second study the prefrontal cortex (77) in participants given perceived control over the stimulus. In contrast, a perceived lack of control may generate feelings of hopelessness which amplify the

perception of pain. Strongly linked to perceived control over the pain is perception of capacity to cope with pain and concern about inability to cope can give rise to negative emotions such as hopelessness, frustration and guilt all of which are potentially pain promoting (78). There is a growing body of research that reports the deleterious effect of pain catastrophizing on pain experience and contrastingly the positive effects of reinterpreting pain as benign sensory data on pain intensity (56, 79-84).

2.1.2.2 Emotional processing

Pain is an unpleasant and aversive experience linked to a range of negative emotions that have already been alluded to and they include; fear, anxiety, guilt, anger, frustration and hopelessness. The amygdala plays a role in generating emotional understanding of the significance of sensory stimuli, including pain. As has been suggested in the previous section there is significant interaction between cognitive and emotional processing to inform the affective features of the pain experience (48). This interaction is illustrated in a study examining the impact on anxiety and perceived control over the noxious stimuli (85). Controllable pain was associated with decreased anxiety and activation of the amygdala compared with uncontrollable pain. While, the emotional features of pain are what creates the unpleasantness of pain it is also clear that they play a modulatory role and facilitate or inhibit pain perception. There are clinical and experimental studies which demonstrate that emotions and mood impact both positively and negatively on the pain associated with chronic disease and that positive manipulation of emotion and mood decreases the perception of pain and that provoking negative emotions and/or mood has the opposite effect on pain perception (68).

Furthermore, the centres responsible for processing emotion are also heavily involved in arousal and attention. This means that negative emotions increase arousal and serve to focus attention on pain and hence increase the unpleasantness of the sensation. Finally, these emotions blunt function of the pre-frontal cortex which decreases higher cognitive control to institute measures to control and modify the situation. This reduction in the sense of control over the pain experience, which has been described earlier, is likely to magnify the experience of pain (58).

Research focused on the interactions between emotion, attention, memory and pain has greatly increased our understanding of the strong connection between these states but has equally underscored the complexity of these relationships.

2.1.2.3 Physiological responses

The central autonomic system spans an extensive network of interconnected neurons across a range of regions of the central nervous system (CNS), which include but are not limited to the insular and anterior cingulate cortices, the amygdala, the hypothalamus, the PAG, the parabrachial nucleus, and the medullary raphe nuclei (86). These areas of the CNS receive ascending nociceptive and visceral signals from the periphery in addition to collateral neural data from other regions of the CNS. This convergence of neuro-signalling provokes specific patterns of autonomic response which are transmitted to preganglionic parasympathetic and sympathetic neurons to generate peripheral responses. These responses are extensive and involve but are not limited to regulation of temperature, blood flow, fluid and electrolyte balance, circadian rhythm and sleep/wake cycles, stress responses, arousal and gastrointestinal activity (87). These centres are also key to descending modulation of pain and the autonomic system plays an important role in these processes (87).

The interactions and results of this network of signalling depend to a large extent on the origins of the noxious stimuli and therefore the pathway and where the signal terminates, which will influence the autonomic, endocrine and emotional response that results from the noxious stimuli (87). For example, somatic, cutaneous pain is relayed to the lateral PAG via the dorsal horn and provokes descending signals designed to provoke a 'fight' or flight' response to 'escapable stimuli' (86). This stimulates sympathetic activity to increase blood flow to the face and the legs, heart rate and blood pressure. An increase in muscle contraction and anxiety is also seen with sympathetic activity associated with pain. Many of these responses intensify pain and the unpleasantness but likely served an evolutionary purpose to prompt action to reduce the risk of further injury. In contrast, 'non-escapable' poorly localised deep somatic or visceral stimuli terminates in the venterolateral PAG which sends descending signals to the parasympathetic system to reduce heart rate, blood pressure and activity (86).

Autonomic and pain modulatory responses triggered by noxious stimuli are also controlled peripherally by triggering of preganglionic autonomic effector neurons at the same spinal segment as the afferent signal. These signals generate somatosympathetic and viscerosympathetic reflexes, which include increased innovation to the blood vessels of the head, face and neck, the lacrimal and nasal mucosa (86).

Pain signals that reach the hypothalamus are also responsible for a cascade of neuroendocrine responses. The first of which is the release of corticotropin releasing hormone (CRH), gonadal releasing hormone and thyroid releasing hormone from the hypothalamus. These hormones act

on the anterior pituitary and stimulate the release of hormones that act on the end glands of this cascade: the thyroid and adrenal glands and the gonads to release cortisol, pregnenolone, dehydroepaindrostone, testosterone, estrogen, progesterone and thyroxine (88). These hormones have a range of actions that feed pain modulation and they include; immune and anti-inflammatory activity, glucose control, modulation of CNS receptors and nerve conduction and control of the blood brain barrier (88).

2.1.2.4 Behavioural responses

Guarding and withdrawal from a source of potential or actual tissue damage and avoidance of activities that provoke pain are examples of behavioural responses aimed at reducing pain (89). Some behaviours are learned while others are a reflex response to painful stimulus and there is some recent evidence to suggest that the relationship between behavioural responses and pain perception may in fact be bidirectional where behaviours also influence perception (89). Nociceptive signals reach the dorsal horn via Aδ-fibres which have projections that synapse with spinal interneurons and motor efferent fibres in addition to the ascending projections. The motor signals are responsible for ipsilateral flexion of muscles to cause withdrawal from the painful stimulus and a contralateral response to provide postural support. There is a complex descending modulation of these reflexes and if this pathway is damaged flexion reflexes occur in response to non-noxious stimuli (90).

The inclination to rub a painful area of the body is also an example of the way in which behaviour may help to alleviate pain. Melzack and Wall's Gate theory of pain is thought to explain the physiology behind this behaviour and how it may play a part in the modulation of the pain experience (91). This is described in section 2.1.3. In some instances behaviours will increase the experience of pain and this is most obvious in the setting of chronic pain where avoidance behaviours have been shown to contribute to long-term increases in pain (58).

Behaviours are also used to communicate pain to others and this is achieved most obviously by the pain related changes that occur to facial expression. Communication of pain may result in behaviours in others that reduce the individual's pain or their suffering, e.g. the provision of sympathy, helping with activities etc. The social significance of pain gave rise to the social communication model of pain which will be described briefly in the next section (92, 93).

2.1.3 Pain theories

The complexity of pain processing and the pain experience has led researchers to propose several theoretical frameworks or theories to explain the numerous clinical and research observations made. Melzack and Wall's widely accepted Gate Control theory and the more recent pain matrix are described in this section (91). The social communication model concentrates on the social impact of pain and is used to explain our behavioural responses to pain and this is also described in this section (92).

2.1.3.1 Gate Control theory of pain

The gate control theory was proposed in 1965 to explain inhibitory pain modulation (91). Melzack and Wall proposed that complex interactions between inhibitory interneurons and the descending inhibitory tracts from higher centres occur within the dorsal horn to modulate transmission of the nociceptive signal. Simply put these authors proposed that signals from non-painful stimuli 'close the gate' to signals from painful stimuli. This theory postulates that Aβ-fibres, which are similarly capable of fast transmission of signals to Aδ-fibres but have a low activation threshold and are responsive to non-noxious mechanical stimuli such as: light touch, modulate pain transmission. It is proposed that the non-noxious stimuli transmitted by these fibres results in inhibition of the dorsal root nociceptor fibres preventing them from transmitting noxious signals via C-fibres to the central nervous system. It is thought that this mechanism explains the impulse to rub an area of pain. Furthermore, this theory proposes that signals descending from the brain also have the capacity to block transmission of ascending afferent signals (91).

2.1.3.2 Pain matrix

Observations made of people experiencing phantom limb pain following cord transection or amputation of the limb and growing awareness of the complexity of the pain experience and the poor correlation between pain severity and injury served as the basis for Melzack's conceptual model to explain pain processing: the neuromatrix theory (94). He described an extensive network of neuronal connections between many areas of the brain, which have been described briefly in previous sections and include regions such as: the primary and secondary somatosensory cortices and the insular, anterior cingulate and prefrontal cortices. These connections are genetically predetermined but remodelled by sensory input. The matrix comprises loops of neurons between these structures that allow parallel processing of all signals from the body but converge repeatedly throughout the matrix to allow for interactions between the output products of processing. This

matrix perceives the body as a whole and processing and synthesis of signals through this network occurs in a characteristic pattern that is recognised as the neurosignature.

Fundamental to this theory of pain is acceptance that the CNS is responsible for producing pain experience and not damaged tissue. Patterns of neuronal activity in the brain comprise a neurosignature for pain which is responsible for the sensory, affective and cognitive dimensions of the pain experience. These patterns may be triggered by noxious stimuli but may also be provoked under circumstances where there is no stimulus. They are also a function of genetic programming and integrated inputs from the brain which include factors other than those that stem directly from nociceptive pathways (94).

New imaging modalities such as, functional magnetic resonance imaging (fMRI) have confirmed the extensive network of structures throughout the brain that are active during a painful experience, which is consistent with the theory that it is a matrix of neurons and not a single centre responsible for pain experience. However, increasingly it is acknowledged that this activity, although sensitive to pain, may not be specific (95)

2.1.3.3 Social communication model

The need to communicate pain to solicit assistance from others has formed the basis for a model that Craig and colleagues have used to describe paediatric pain (92, 93). They assert that self-report and behaviours that can be recognised by others as indicative of pain are a means to convey the individual's experience to others and that pain is inherently a social experience. This model accounts for the sequencing of events that occur with pain experience and the social context in which pain is experienced. The social communication model (SCM) of pain describes the process for transducing the pain experience into a pain expression that can be interpreted by others to provoke a response.

The pain experience of the child is the sum of the biological, psychological and social factors that influence pain perception (93). The complex interaction between these influences and their impact on processing of pain signalling in the brain has already been described in this chapter. The second stage of this model involves the transduction of perception into expression, which is the sum of physiological and behavioural responses and the product of conscious and unconscious control. Pain expression manifests in numerous ways which is influenced by cognitive development. Infants express pain by behaviours such as: crying, changes in facial expression and other body movements. With increasing cognitive development children increasingly use language to express

pain. This model also recognised a bidirectional relationship between experience and expression such that the expression of pain will also impact on the experience of pain (93).

Stage three of this model involves the decoding of the expression by the observer (93). This is also based on a bidirectional relationship as the observer will interpret the expression based on their understanding of the expression and its relationship to pain. This will be influenced by cultural, contextual, experiential, interpersonal characteristics and the emotional state of the observer (93). In turn, their response to the expression of pain will potentially modulate the experience and/or expression of pain in the infant or child. For example, pain expression may elicit caring behaviours that result in removal of the painful stimulus or provision of pain-relieving interventions which reinforces the behaviours. Alternatively, a negative response to pain expression may prompt the child to inhibit their expression.

2.1.4 Pain and the developing nervous system

The traditionally held view that the nervous system of the neonate was insufficiently developed for young babies to experience pain was repeatedly challenged in the latter half of the last century. It is now accepted that neonates are born with functioning nociceptive pathways and do experience pain. However, evidence demonstrates that there are some differences in these pathways and the responses of preterm and term neonates when compared to those of adults. Furthermore, exposure to pain during the neonatal period is likely to have a significant effect on the neural circuits responsible for pain processing (96).

Unlike the afferent nociceptive pathways, the pathways involved in central descending modulation of afferent signals, are not fully mature at birth. Preterm neonates exhibit evidence of more substantial immaturity in these pathways and the associated circuits (97, 98). This renders these neonates more susceptible to pain associated with stimuli during the neonatal period and to the long-term negative effects that result from exposure to noxious stimuli during the neonatal period (99). Activation of immature pathways provokes unique and exaggerated patterns of activation across multiple regions of the brain (97, 100). Preterm infants also show lower thresholds and greater reflex response to stimuli than term neonates (101).

Using electroencephalography (EEG), researchers have demonstrated that neural pathways able to distinguish between noxious and tactile stimulus are present at 35-37 weeks gestation (97, 102, 103). However, infants younger than 33 weeks postconceptional age show heightened touch sensitivity and there is evidence that non-painful stimuli such as nappy changing trigger similar

stress responses to noxious stimuli such as needle procedures and that exposure to these procedures sensitises infants such that painful procedures are more painful when they follow non-invasive handling procedures (104, 105). Studies of older children show both hyper-reactivity and hypo-reactivity in adolescents exposed to noxious stimuli who were born very prematurely. The type of pain and the duration of the pain stimulus has been shown to influence the extent and direction of the effect on reactivity (96).

Noxious stimuli provoke behavioural responses in preterm and term neonates which include withdrawal and changes in facial expression (106). Observation of these responses reveals differences from those seen in adults and subtle differences between preterm and term neonates implying changes in these pathways or modulation of these responses associated with maturation. Although neonates, like adults, demonstrate flexor muscle activity and withdrawal from a stimulus the duration of the response is longer, is seen on the contralateral side, is likely to provoke a stronger physical response and may be provoked by other tactile stimuli (106). These differences become increasingly more apparent with decreasing gestational age. Data from electrophysiological studies confirm electrophysiological differences associated with these behaviours. For example; inhibitory signalling is weaker in neonates than adults which goes some way to explaining the stronger reflex seen in neonates. Additionally, with increasing gestational age, the dorsal horn circuits demonstrate increased spatial organisation and therefore capacity to localise sensory data and reduced sensitivity to stimuli and therefore greater specificity for nociceptive stimuli (106).

Pain also provokes a brain stem reflex which results in consistent changes in facial expression in neonates which have been described as the ‘primal face of pain’ (107). Although facial expression is not identical between individuals, very similar patterns of expression are seen across genders and cultures (107) and are common from infancy to adulthood (108). Evidence shows that facial reflex response develops early in the foetus and is present even in the setting of significant damage to white brain matter (106). However, with increasing age facial expression comes under increasing conscious control and is modified by signals from areas of the brain responsible for emotions (106). Children as young as eight years have been shown capable of suppressing expressions of pain or faking pain expressions using the same changes in expression associated with genuine pain (albeit the expression is often exaggerated) (109).

Haemodynamic studies have improved our understanding of the role of the cerebral cortex in pain processing in preterm and term neonates. Seminal work using near-infrared spectroscopy, functional magnetic resonance imagining (fMRI) and EEG have demonstrated activity in the neonatal cortex following noxious stimuli confirming that noxious signals reach the brain (95, 97,

103, 110-112). Goksan and colleagues' work reveals an active network in the neonate that includes 18 of the 20 regions of the adult brain excited by noxious stimuli (112). It is revealing to note that noxious signalling in the neonate does not reach the amygdala or the orbitofrontal cortex, centres that play a significant role in the affective responses to pain. Furthermore, many of these regions are activated by any salient stimuli (95). With increasing gestational age, the capacity to distinguish between tactile and noxious stimuli increases. Conversely, several regions are only active in the neonatal brain and they include: bilateral parietal lobe, pallidum and precuneus cortices (only contralateral in adults), bilateral auditory cortices, hippocampus and caudate nucleus (112). These studies also highlight more extensive bilateral activity and more widespread connections in the neonatal brain when compared with the adult brain. Electrophysiological studies confirm event related potentials in the somatosensory cortex which is age specific (106). The interpretation of the significance of the differences in brain activity and how this relates to the adult response to pain remains unclear.

Central sensitisation and the long-term effects resulting from triggering of nociceptive pathways by nociceptive stimuli in the developed nervous system is well established (96). Increasingly, laboratory and clinical studies have revealed the impact of noxious stimuli on the highly plastic neonatal nervous system and the long-term outcomes of noxious stimuli. In studies using animal models, long-term outcomes are most often defined in adulthood as changes to either; baseline sensory sensitivity or responsiveness to noxious stimuli (96). The type and intensity of the stimulus, the region of the nervous system triggered and the post-delivery age of the neonate have all been shown to influence the extent of the long-term impact of noxious stimuli (98). The work of a number of research teams has also revealed a strong connection between neonatal pain and alterations in sensitivity and response to noxious stimuli in later life. Interestingly, similar patterns have also been seen when neonates are exposed to maternal separation and stress (96).

Research has focused on several cohorts to explore the effects of noxious stimuli on human infants. A systematic review of 13 studies examining the impact of neonatal pain exposure on developmental outcomes in children born preterm reported that the number of painful procedures were associated with alterations in early and later adverse developmental outcomes including negative impact on growth, cortical processing, and behaviour (13). Neonates and very young infants who experience circumcision have been shown to demonstrate increased pain behaviours in response to infant immunisations than their uncircumcised counterparts (10, 11). The long-term outcomes of early exposure to surgery are likely to include changes in peripheral and central pain processing. Results from studies are inconsistent and predictable patterns are not apparent. However, the impact of surgery probably depends on multiple factors which include; the site and extent of the surgery, the age of the infant at the time of the surgery and the post-operative pain

management received (99). Infants suffering burns are another cohort studied to determine the effect of noxious stimuli on later pain experience and altered mechanical and pain sensitivity has been shown in adolescents who suffered a burn in infancy (113).

2.1.5 Classification of pain

The heterogeneity of the pain experience has led to multiple systems of classification for pain. The distinction between nociception and pain is at the core of one these systems and is based on the aetiology of pain: defining it as either nociceptive or neuropathic. Nociceptive pain has already been describe but in short is considered the pain associated with noxious stimulus (33). Neuropathic pain is initiated by nervous system dysfunction that causes excess and/or disproportionate stimulation of nociceptive pathways or decreases the inhibitory action of neural pathways. The result of which is an imbalance in painful and non-painful input causing the sensation of pain in the absence of a potentially or actual noxious stimuli. The IASP states that for pain to be considered neuropathic a demonstrable lesion or disease of the somatosensory nervous system is necessary (33).

Pain is also categorised based on the duration of the pain e.g. acute or chronic, although this classification often includes an aetiological component. Definitions vary but clinically, acute pain is most often considered pain of short term duration (less than 3 months), associated with recognisable pathology and resolves within an expected time frame; e.g. resolution of the pathological state (15). In contrast chronic pain is pain that lasts beyond 3 months and the expected time of healing. Furthermore, the mechanisms are likely to be different to those associated with the original disease or injury. In keeping with the link to the underlying pathology, acute pain is generally considered adaptive as it serves to change behaviours to avoid tissue injury while chronic pain is often described as maladaptive as there is no demonstrable biological value to the individual (114). The principle of this is sound but in practice the classification of acute versus chronic is not as easily applied. Healing occurs at varying rates within and between lesions and between individuals and the relationship with the symptom of pain is not always consistent. The presence or absence of pain may not be associated with the progress of healing e.g. scar formation and maturation is generally painless and in other circumstances complete healing may not be possible e.g. chronic rheumatological disease (115). This underlies the reliance on a logically but somewhat arbitrarily derived time frame for differentiating acute from chronic pain.

Cancer-related pain is the most frequently used disease-related pain classification. However, cancer pain does not describe a single pain experience. Cancer-related pain encapsulates disease

and treatment related pain, pain originating from differing aetiology and pathology, acute and chronic pain and nociceptive and neuropathic pain. In addition, it is often heavily influenced by temporal factors, most notably the potential fear and distress that stems from a diagnosis of cancer (116). More detailed discussion of this complex symptom is beyond the scope of this summary.

2.2 Assessment and measurement of pain

Clinicians and researchers strive for ways to assess and measure pain in patients to guide management or measure study outcomes. Most often the characteristic of interest is pain intensity and methods to quantify pain intensity, even though this may oversimplify the overall experience of pain. More correctly, pain assessment should include review of the location, nature, quality, functional impact and emotional features of pain. However, in this section the focus will be on the methods used to assess and quantify pain intensity.

As pain has been increasingly acknowledged as 'what-ever the patient says that it is, existing whenever they say it does' (117), self-report has in turn increasingly been viewed as the 'gold standard' for assessment. Tools that allowed the individual to provide an estimate of the intensity of their pain using numeric scales or ordinal verbal scales that imply quantity using terms such as: mild, moderate and severe are widely used for clinical and research purposes. The visual analogue scale (VAS) was developed to quantify subjective experiences and is an early example of a numeric scale used to self-report and quantify pain intensity (118). This scale uses the now widely accepted 0 to 10 metric for this purpose where '0' represents 'no pain' and '10' represents the 'most severe pain'.

In recent years conventional thinking has been challenged and experts have redefined self-report as ideal but not necessarily the gold standard for pain assessment (119). This shift reflects growing concern about the oversimplification of pain assessment based exclusively on intensity and the lack of congruence in some circumstances between self-reported pain and evidence from other sources e.g. denial of pain where pathology would suggest severe pain and conversely reports of extreme pain to affect a specific outcome such as work avoidance etc. Concerns about accepting self-report as a gold standard for assessment of pain intensity are particularly obvious in children. The cognitive development of children may restrict their understanding of numbers making their assessment using a number-based scale unreliable (120). Graphical tools have been developed as an alternative. These scales use a combination of numbers and images to represent increasing pain intensity. Several self-report scales designed for children use facial images representing expressions intended to align with the intensity of pain (121). Most commonly these scales are

comprised of six faces the first representing a score of ‘0’ and subsequent faces an increase in pain in increments of ‘2’ with a maximum score of ‘10’ (120). Use of scales of this type is not without criticism. It cannot be assumed that these scales provide interval or ratio data despite each face having a numeric label (122). This has significant implications for interpreting the scores reported by children using these scales. It has been suggested that the use of the face and inclusion of numbers on the scale serves to confuse the affective features of pain with intensity (123). Furthermore, the style and expression of the faces varies across different versions of these scales and there is evidence to show that this influences the way child, particularly young children score their pain using these scales(121). This is seen most clearly in results that show a propensity for young children to allocate higher scores when using a scale with a smiling face for no-pain. Although, children as young as 4 years are considered able to self-report pain (124), there is also evidence that question the capacity of children 3 to 5 years to provide meaningful report when using 6-point self-report scales (125). They demonstrate a propensity to select the end points of these scales and are considered likely to better manage a scale with fewer options e.g. three choices corresponding to ‘mild’, ‘moderate’ and ‘severe’.

For a range of reasons individuals are not always able to self-report pain intensity. This includes infants and young children who are developmentally unable to communicate and/or understand the necessary concepts to apply an ordinal scale. Similarly, individuals with congenital or acquired cognitive impairment may also have difficulty understanding the scale and its application to their experience of pain. Language barriers and serious illness also make a potentially significant contribution to impairing an individual’s ability to self-report. Alternative methods to self-report for quantifying pain intensity have been developed to meet the needs of these groups. The next sections address methods of assessment such as: physiological measures, imaging and electrophysiological studies and observational behavioural scales which have been used or are proposed as an alternative to self-report.

2.2.1 Imaging and electrophysiological studies

Imaging and electrophysiological studies have been the basis for great advances in our understanding of the neurophysiology of nociceptive pathways and pain signalling. They have also been proposed as a potential means to assessing pain. A range of techniques have been studied for this purpose and which will be discussed briefly in this section. They include electroencephalography (EEG), near infra-red spectroscopy (NIRS) and functional magnetic resonance imaging (fMRI).

EEG provides a non-invasive method for measuring cortical brain activity by measuring small electrical currents generated by postsynaptic potentials (48). It is proposed that as this method can be used to map the activity of areas of the brain active following painful stimuli such as, the somatosensory cortex and the anterior cingulate cortex, therefore it could be used to support pain assessment. Researchers have worked to demonstrate the capacity of this modality to localise parts of the brain involved in pain processing (126) and to show that currents correlate with the stimulus intensity and subjective experience (126, 127). Furthermore, it has been shown that the EEG waves are modulated by factors that impact on pain such as attention (128, 129) and placebos (130, 131).

Fernandez and colleagues demonstrated a difference in EEG activity in infants exposed to noxious heel stroking randomised to receive water compared with those randomised to receive glucose (132). Furthermore, those that received glucose showed no change in their EEG before and after the stimuli while those that did not receive glucose demonstrated EEG activity consistent with noxious stimuli. These results are evidence of the capacity of this modality to assess and differentiate pain. However, they are at odds with results from a later study by Norman and colleagues (133). In this study, the authors reported clear differences in pain scores (Premature Infant Pain Profile (PIPP)) based on group allocation (glucose or placebo and/or procedure type). However, no meaningful differences in EEG activity were seen across groups or before and after the stimuli (133). More recently, Slater and colleagues have reported results that are comparable to those of Fernandez and colleagues. (103). In a small cohort of infants, the EEG recordings of the infants' baseline activity and their responses to a tactile and a painful stimulus were compared. The results of these comparisons allowed the authors to conclude that nociceptive-specific activity occurs in infants and can be detected by EEG. However, in line with findings by Norman and colleagues, Slater and colleagues also demonstrated in a small subgroup of infants enrolled in a larger RCT that EEG activity did not differ between infants receiving sucrose compared to water during a heel lance. The capacity for the EEG to distinguish between painful and non-painful stimuli in a sample of neonates experiencing heel lance and touch has also been shown in Worley's work (134). Although, perhaps it is accepted that painful stimuli create a unique electrical signature that can be recorded using EEG it is unclear how this relates to the characteristics of the pain experience, i.e. the intensity of the pain. This area of study is still relatively new, and further work is required to ascertain if EEG activity is a useful measure of pain in newborns.

Near infrared spectroscopy provides a measure of haemoglobin oxygenation by passing two infrared lights of different wavelengths through the tissue (135). Light is variably absorbed which enables measurement of concentration changes in deoxygenated and oxygenated haemoglobin

and therefore calculation of blood volume. Therefore, it has been used to provide an estimate of blood flow to body tissues, including the brain (135). It was used as long ago as 1995 as an indirect measure of cortical activity in response to noxious stimuli in an RCT by Bucher and colleagues who tested the effect of glucose on pain associated with heel lancing (136). The changes in cerebral blood flow, measured by NIRS, was a secondary outcome of this study. Significant differences in heart rate and crying time between groups were reported. In contrast, no difference in cerebral blood flow between groups was reported. The obvious paradox between the outcome measures would suggest that cortical blood flow to areas of the brain responsible for pain processing and pain experience are not synonymous and therefore that this measurement may not be useful for assessing pain. In more recent years researchers have reported results that are consistent with or challenge those of Bucher's study leaving some uncertainty about the role of NIRS in assessment of pain (137, 138).

A team at University College London have made significant contributions to our understanding of cortical processing of pain in neonates and imaging techniques that may be useful for clinical assessment of pain. In their 2006 study, infants were studied using NIRS to measure cerebral oxygenation over the somatosensory cortex during a heel lance (111). A clear increase in cerebral oxygenation in the contralateral cortex was demonstrated in infants aged between 25 and 45 weeks postmenstrual age. The contralateral response was greater in awake infants and increased with age. Furthermore, latency in response also changed with age. This response to noxious stimuli was in stark contrast to what was seen associated with non-noxious tactile stimulation where there was no increase in cerebral activity in the contralateral somatosensory cortex even in infants that exhibited reflex withdrawal from the stimulus. Bartocci and colleagues reported similar results to those of the University College London team (110). Haemodynamic changes in the somatosensory cortices of 28 to 36-week gestational aged neonates subjected to venepuncture and tactile stimulus also demonstrated patterns specific to procedure type. The results of a study examining the cerebral haemodynamics and behavioural responses to painful stimuli in critically ill infants following cardiac surgery also reported a positive association between pain scores and cerebral activation demonstrated by NIRS and similar responses in both measures to analgesic administration (139).

In another study from the University College London group, Slater and colleagues compared cortical activity with observational pain scores (PIPP) in neonates aged 25 to 45 weeks postmenstrual following a heel lance (137). Correlation between the two was good and the 'facial expression' item of the behavioural scale correlated most strongly. However, cortical responses (measured by NIRS) were seen in infants that did not demonstrate a change in facial expression. The authors conclude that some infants may be in pain despite a low behavioural pain score.

However, this conclusion, based on a small subgroup of infants has been criticised by some researchers and may highlight a potential disconnect between pain experience and current mapping of neural activity associated with painful stimulus.

Two studies in infants exposed to skin breaking procedures also challenge the validity of NIRS used as a measure of pain perception. These studies reported no correlation between pain scores and cortical activation (140, 141). Bembich and colleagues explored the effects of breast feeding and glucose on the pain associated with heel lance (142). Their results demonstrated no significant changes in cortical activity associated with glucose administration but higher pain scores than were seen in breast fed infants. Furthermore, cortical activation occurred in breast fed infants during heel lance. In a third study, an RCT to determine the impact on the pain associated with heel lance of three different non-pharmacological interventions, results were also inconsistent (143). The heart rate of the preterm infants increased in response to the stimulus for all treatment groups. In contrast peripheral oxygen saturation and NIRS measurements did not change following heel lance.

Studies have also used NIRS in adult populations which provides a unique opportunity to correlate cortical activity with self-report of pain, an opportunity denied neonatal researchers. Using an experimental pain stimulus, Azar demonstrated a correlation between self-reported pain and cortical activity demonstrated by increases in cerebral oxygenation (144). Similarly, increases in cerebral oxygenation were seen during nociceptive procedures in adults undergoing cardiac surgery in the work of Gelinas and colleagues (145). However, no correlation existed between cerebral oxygenation changes, pain behaviours and self-reported pain intensity.

There are clearly still gaps in our understanding of the relationship between cortical activation in response to painful stimuli measured by NIRS and pain experience. Furthermore, as with all indicators of pain, NIRS does not fully explain the experience of pain. Research teams continue to use this modality to map cortical activity of the brain in response to painful and non-painful stimuli and explore the relationship between this mapping and other assessment modalities to better appreciate its potential role as an assessment tool (97, 134, 146, 147).

Functional magnetic resonance imaging (fMRI) is another imaging technique that has been increasingly used in the quest to better understand supraspinal activity associated with pain. By measuring changes in blood flow, fMRI indirectly measures neural activity in the brain (48). To achieve this, fMRI uses one of two approaches, detection of the difference in magnetic susceptibility between oxygenated blood and deoxygenated blood (blood oxygen level-dependent (BOLD) imaging) or tracing of magnetically labelled endogenous water through the cerebral

circulation (48, 148). The choice of approach depends largely on temporal factors. The first approach is used to image acute short-term experiences of pain and is the focus of the following discussion. Although, the literature includes studies that confirm a predictable and consistent pattern of activation of centres in the brain in association with painful stimulus (112, 149-153), infants are the focus of only two of these studies (112, 153).

Williams and colleagues used experimental pain in a recent study of full-term infants to determine the feasibility of using fMRI to measure brain activity in response to a range of mechanical stimuli (153). Specific patterns of activation resulted from different stimuli and activation increased in response to the intensity of the stimulus. The authors concluded that fMRI shows promise as a potential measure of neural processing of mechanical stimuli. The regions of the brain activated by pain in neonates were compared with those of adults by Goksan and colleagues and there was considerable overlap. This was described in Section 2.1.4 of this chapter (112). These results are considered evidence that infant perception of pain is also likely to overlap that of adults.

Neuroimaging and haemodynamic studies have allowed scientists to make great leaps in our understanding of brain activity associated with pain. Accepting that these modalities provide a valid measure of a pain is predicated on accepting that the patterns of activity consistently associated with pain represent pain experience. Observations have been made that cast doubt on this premise. Salamons and colleagues reported the same patterns of neural activity to painful stimulus in individuals with congenital insensitivity to pain as control individuals (154). Similarly, the conflicting results reported in studies using NIRS raise questions about what neurological activation tells us about pain experience. Finally, these modalities require specialised equipment and significant patient preparation rendering them impractical for use in circumstances other than research designed specifically to explore neuronal activity.

2.2.2 Physiological measures

Use of physiological parameters such as: heart rate, respiratory rate, blood pressure and oxygen saturations as a measure of pain intensity is based on an assumption that these measures change predictably with changes in pain intensity. As has been described, noxious stimuli and pain signalling trigger autonomic responses which result in changes in these physiological parameters. A systematic review conducted in 2011 to determine how effective physiological parameters are as indicators of neonatal pain included seven RCTs where these parameters were used as a measure of pain (155). Heart rate was the most frequently used of these and results were inconsistent; in some studies heart rate changes in response to pain were corroborated by other

measures of pain intensity while in others this was not the case. Although not included in this systematic review, results from other studies show similar inconsistencies in the way these indices respond to painful stimuli (156-171). Raeside and colleagues concluded that these physiological indices do not show sufficient sensitivity or specificity to be used as an independent indicator of pain (155). They also reiterate the conclusion drawn several years earlier by Stevens and colleague that although these parameters may change in response to pain potentially making them useful indices to detect pain, they are unlikely to be specific for pain or able to quantify pain intensity (172).

Studies have also evaluated these parameters when paired with other observable measures such as behaviours to assess the capacity of a composite of indices to estimate pain intensity. Results are mixed as to their additive value and in some circumstances the internal validity of the scale is not improved by the addition of physiological measures. The COMFORT scale designed by van Dijk and colleagues (173) is an example where heart rate and blood pressure items were removed from a subsequent iteration of this scale following internal reliability testing to determine the contribution of physiological and behavioural items to the scale (COMFORT-B scale) (174).

Biochemical markers have also been explored as a potential surrogate for self-report of pain. The biochemical markers implicated include endogenous neuropeptides, such as: beta-endorphin, epinephrine, norepinephrine and inflammatory mediators such as: serotonin, prostaglandins, histamines, cytokines and chemokines (175). Increased levels of these biochemical markers have been linked to pain in animal and human models. However, it remains unclear what role these markers may play in the detection of pain and it seems unlikely that they can be used as an estimate of pain intensity. Furthermore, many of these markers are located in the dorsal root ganglia or the spinal cord making sampling procedurally difficult and associated with significant risk of harm (175).

Cortisol is another biochemical marker linked to the cascade of neuroendocrine responses to painful stimuli. It is released by the adrenal glands in response to stress (88). Salivary cortisol levels correlate well with serum cortisol levels providing a simple, non-invasive way of sampling this marker (176). This has led to research interest in its potential role in pain assessment. Increases in cortisol levels associated with immunisation (177-179) and surgical procedures (180, 181) have been reported in multiple studies. Zhao and colleagues reported rises in salivary cortisol in children with cerebral palsy undergoing painful therapies which were corroborated by pain scores (182). A study designed to assess the effect of pet therapy (the presence of a therapy dog) on the pain associated with venepuncture reported a reduction in self-reported distress and salivary cortisol levels but not in self-reported pain associated with the presence of the dog (183).

In contrast, Ha and Kim reported lower pain and affective responses to pain in the experimental group of their RCT but no difference in salivary cortisol levels (184). A positive correlation between salivary cortisol levels and postoperative morphine consumption was shown in an RCT of children randomised to receive varying levels of information prior to surgery (181). Morelius and colleagues reported lower increases in infants receiving glucose and a pacifier for the immunisation (179). At odds with these results are those of a study examining premature infant responses to routine heel lances which reported no rise in salivary cortisol in response to the heel lance (185). Hansen also reported no significant difference in cortisol levels, despite significant differences in pain scores between groups randomised to either sequential or simultaneous immunisations (178). Finally, data from a 2005 study by Grunau and colleagues also demonstrated lower cortisol levels following cumulative exposure to painful procedures (186) and in preterm neonate compared with term neonates (187). The results across studies are not consistent and salivary cortisol levels vary widely making it impossible to accept cortisol levels as sensitive or specific measures of pain.

Skin conductance is another physiological measure of stress that has been explored as a potential measure of pain. It is accepted that changes in electrical conductance in the skin occur in response to changes in activity of the sweat glands which is mediated by changes in sympathetic activity and that therefore skin conductance is an indirect measure of neuro-physical arousal (188). This has served as the impetus of several research groups to evaluate skin conduction as a potential means to detect pain (188-201). Storm has led a body of work across several centres to test the capacity of this modality to measure stress and more specifically pain. Early work in adults demonstrated increased fluctuations in skin conductance associated with perioperative stress (202, 203). This was followed by two studies confirming correlations between patient-reported postoperative pain scores and the number of fluctuations in skin conductance (204, 205). Similarly, early data in infants showed promise (189, 206-209). However, researchers are not without reservations about these findings. Harrison and colleagues reported significant increases in skin conductance during a heel lance procedure (189). However, in the same study there was no significant difference in skin conductance changes between infants undergoing a heel lance and those experiencing routine nursing care. Erikson and colleagues tested this modality on premature infants exposed to painful and non-painful stimuli and although they reported the capacity for skin conductance to differentiate between stimuli, they echoed Harrison's call for more data before this technology is accepted as a valid measure of pain (206). Gunther and colleagues reported rises in skin conductance associated with painful stimulus. However, similar rises in motor activity associated with this stimulus led them to suggest that fluctuations in skin conductance may be better for detecting emotional distress than pain (209).

These physiological measures all suffer the same limitation as to their specificity for pain. These parameters are responsive to physiological stress and pain is only one trigger for stress. A physiological stress response may also be triggered by illness, emotional distress and other aversive emotions etc. Therefore, these experiences are likely to result in similar physiological changes as those seen in pain. This may in part go some way to explaining the inconsistencies in the results of studies, particularly those that concentrate on cortisol levels. The participants in many studies are experiencing a range of stress-related triggers in addition to pain. Studies exploring the effect of early painful procedures on hypothalamic pituitary adrenal (HLA) axis function in later infancy illustrate the complex relationship between pain and stress (210, 211). Physiological and biochemical markers are not sufficiently specific to pain to make them a useful unidimensional measure of the presence and intensity of pain.

2.2.3 Behavioural pain scales

Behavioural pain scales are one of the most commonly used and most practical methods for pain assessment. The premise behind these scales is that observable behavioural responses to pain are consistent and predictable rendering them suitable surrogate measures of pain (212). This has given rise to the development of nearly 50 unique scales for pain measurement in infants and children (213). Specific scales considered likely to be suitable for assessment of procedural pain in infants and children will be described in the body of the thesis. This section will describe scales in terms of overall type and design and highlight those that have been recommended specifically for procedural pain assessment in infants and children.

2.2.3.1 Measure type

Behavioural pain scales are most commonly either checklists, rating scales or global rating scales (30). Checklists are comprised of a set of behaviours, which are assessed dichotomously as either present or absent. The pain intensity score is based on the number of pain-related behaviours present during the observation period. These scales are easily applied and relatively unambiguous and lend themselves well to clinical use. However, these scales lack the capacity to acknowledge the intensity of behaviours or to give some behaviours greater weighting (30). Rating scales include behavioural indices that are individually rated based on the intensity, frequency or duration of the observed behaviour. The sum of the item scores is the pain intensity score. These scales allow for assigning variable weighting to behaviours and are potentially more sensitive to subtle changes in pain intensity (30). Well-designed rating scales provide clear descriptors for behaviours of increasing intensity and simple metrics for scoring individual items, making them

only marginally more difficult to use than checklists but better suited to more precise measurement.

Global rating scales, such as the visual analogue scale (VAS), although not usually designed as an observational pain scale, have been increasingly used to assess pain in those unable to self-report their pain intensity. An observer provides a global rating of pain intensity based on their perception of the intensity of pain experienced by the infant or child. The anchors used for these scales quantify the intensity in words or numbers but do not describe behaviours. For example, the VAS is a graphical scale that uses a line with anchors at each end defined as 'no pain' and 'worst pain' while other global scales include otherwise undefined increments of 1 or 2 between the 0 and 10 anchors at either end (30). Global rating scales are very quickly and easily applied. However, they are criticised for the impact on scoring of observer bias (30). Observers' perception of pain intensity will potentially be variably influenced by specific behaviours, experience, circumstantial factors such as the administration of analgesics and familiarity with the infant or child.

2.2.3.2 Design

The specificity and sensitivity of behavioural responses can be influenced by contextual factors (214), which has been the driver for development of measures designed to assess pain in either specific populations and/or under specific circumstances. Available scales can be grouped according to their design purpose, some of the most common of which are scales for acute, chronic and disease specific pain, neonates and children with cognitive impairment

There are recognised differences between acute and chronic pain that are likely to be reflected in pain-related behaviours of infants and children experiencing acute pain or chronic pain (30). With this in mind, measures are designed to assess one or the other; acute or chronic pain. Many of those developed for acute pain use were specifically developed for post-operative pain assessment, e.g. Children's Behaviour Coding System PACU (CBCS-P) (215) Children's and Infants' Postoperative Pain Scale (216), the Children's Hospital Eastern Ontario Pain Scale (CHEOPS) (29) and the Face, Legs, Cry, Activity and Consolability (FLACC) scale. The FLACC scale was the scale of choice for assessment of postoperative pain based on recommendations made in two systematic reviews published in 2007 (30, 31). There has not been the same emphasis on developing scales for alternative acute circumstances such as procedural pain. Many of the scales developed for postoperative pain have been used in alternative circumstances and in the same systematic reviews, the FLACC scale and the CHEOPS were recommended for procedural

pain assessment (30, 31). This is despite limited psychometric testing of these scales for this purpose.

Disease-related tools also appear in the assessment literature, although most are designed as a measure for self-report of quality of life indicators including pain. Cancer related-pain receives the most attention. Infants and children experience both acute and chronic pain in association with the disease and its treatment. Hence, measures for acute procedural pain and chronic pain are used in this cohort. There is also acknowledgement of the significant affective responses to pain under these circumstances which has resulted in the development of scales such as the Observational Scale for Behavioural Distress (OSBD) for procedural pain. In addition, the impact of prolonged pain has informed scales such as; Douleur Enfant Gustave Roussy, (DEGRR) Scale for prolonged pain in cancer (217).

It has been previously acknowledged that pain processing, specifically modulation, matures from approximately 25 weeks gestational age and that neuro-physical responses to pain vary with maturation. For this reason, scales that capture the nuances in behavioural responses to pain in preterm and term neonates e.g. the PIPP (218) have been developed. However, even this population is not homogenous and there are likely to be significant differences in the responses exhibited by a 26-week gestation neonate in intensive care compared with those of a healthy term neonate (219). Furthermore, most of the tools developed were validated for assessing acute pain. Despite our understanding of their differences and increased interest in assessing neonatal pain, widely accepted recommendations regarding the most appropriate scale choice are not available. The PIPP scale has been studied most extensively and a systematic review by the scale developers in 2010 summarised the psychometric data available (220). The authors concluded that it is a valid and reliable option but more data to address the feasibility and clinical utility is needed.

Assessment of pain in children with cognitive impairment has also been recognised as uniquely challenging as many of their behaviours are inconsistent with those of their normally developing contemporaries and it is likely that this includes their pain-related behaviours. Effort has been made to develop or modify existing scales to measure pain intensity in these children. Only a few scales have been purposefully designed for individuals with cognitive impairment and they include the Non-Communicating Children's Pain Checklist (NCCPC-PV) (221) and the revised version (222), the Individualised Numeric Rating Scale (INRS) (223) and the Pediatric Pain Profile (PPP) (224). The FLACC scale is an example of a scale designed originally for typically developing infants and children which has been adapted for use with children with cognitive impairment (225). Crosta and colleagues reviewed available scales to make recommendations for

the scale best suited to assessing pain in hospitalised children with CI and they cautiously concluded that the revised FLACC was probably the most suitable (226).

Pain-related behaviours in critically unwell infants and children may also be significantly modified by the circumstances where they are either too unwell or too heavily sedated to demonstrate the usual patterns of behaviour seen in infants and children in pain. Scales have been developed that recognise these limitations and include the COMFORT scale (227).

The purpose of pain assessment is also likely to influence scale design. Scales developed for clinicians to support pain management decisions need to be easily and quickly understood and applied without extensive training. In contrast, scales used for research purposes may sacrifice ease of application in return for improved performance. There are several scales, such as the Child Facial Coding System (CFSC) (228, 229), that require considerable expertise to apply and are more reliably applied to video recordings where repeated viewing is possible. Therefore, this scale and others like it may be better suited to research use.

Item selection

Item selection is also key to the design of the scale and scale developers acknowledge several strategies to identify and rationalise the behavioural indices included in the measure. Many describe interrogating the literature or surveying experts as a means to identifying items considered representative of pain. Psychometric testing (discussed in detail in the next section) of the scale applied to the population and/or circumstances for which it was designed is used to validate their selection. In some studies methods to specifically assess the internal reliability of each item were employed to better understand the contribution that they make to the performance of the scale. Although the measures identified in the previous section all vary slightly there is considerable overlap and some behaviours appear more frequently than others.

Facial expression is one of the most common behavioural items to feature in observational scales. It is also the most extensively and independently researched of all behaviours thought to be indicative of pain. Early attempts to define facial expressions associated with emotion were undertaken by Charles Darwin (230) and pursued again in the 1970s by Ekman and Friesen who provided a framework for identifying expressions (231). They objectively coded individual muscle movements that occurred with a range of emotions into ‘action units’ and identified constellations of consistently occurring action units. Similar facial actions were seen across early studies assessing adult responses to experimental pain supporting the notion of a ‘pain expression’ (232-234). Furthermore, facial action units were not influenced by culture or the type of stimuli

(235). In the pursuit of measures to accurately detect and rate pain intensity in infants and children too young to self-report, early paediatric pain researchers recognised the potential for similar pain related expressions to be seen in infants and children. Grunau and Craig undertook pioneering work to detail specific facial actions seen in neonates experiencing pain associated with a heel lance (236). The results of this study were used to inform their Neonatal Facial Coding System (NFCS). Similar efforts went in to developing a coding system for the facial expressions of preschool aged children which is aptly known as the Child Facial Coding System (CFCS) (229).

Close inspection of the action units implicated in pain expression for adults, children and neonates reveals some obvious similarities supporting the contention that the face of pain is hardwired, universal and can be recognised across the life-span (108). It has been suggested that this hardwired expression is a method for communicating infant's needs to carers and this is the basis for the Social Communication Model of Infant Pain, which was described briefly in section 2.1.3 of this chapter (92). Schiavenato and colleagues coined the term the 'primal face of pain' to capture the innate biologically determined capacity to communicate pain via facial expression (107).

Many research teams have since added to our understanding of the sensitivity and specificity of facial expression for pain and the factors that influence expression associated with pain. Preterm and term neonatal facial expressions in response to pain are particularly well studied since Grunau and Craig's original work (107, 237-244). In many cases, the original 9-point scale has been reduced to a 3 or 4-point scale using only the most commonly occurring facial expressions of pain (218, 236). A 2009 study aimed at testing the capacity of the NFCS to differentiate between painful and distressing procedures questions the specificity of the pain expression or at least the specificity of the NFCS (245). Chang and colleagues focused their attention on the face item of six observational scales and the associated descriptors for these items (246). Their aim was to determine the impact on observer judgement of different descriptors. Their results supported their hypothesis that descriptors that were inconsistent with empirically derived descriptions of facial expression were associated with poor reliability. Despite evidence that consistent facial expressions are associated with pain in infants and children, this has not translated into consistent use of evidence based facial descriptors across observational measures. Furthermore, available data exploring facial expression is derived from studies using an acute painful stimulus for pain. There are no studies that attempt to identify whether the 'pain expression' is the same with other types of pain e.g. postoperative and chronic pain.

Cry is another behaviour that has traditionally been considered indicative of pain in infants and children and hence it features frequently in observational measures of pain intensity. Like facial

expression, cry has received research attention to determine whether it can be used to assess pain. Various characteristics of cry such as; latency to cry following exposure to a painful stimulus and the duration of cry following a painful stimulus, have been explored. Grunau and colleagues reported significant differences in latency to cry and cry duration between infants having a painful or non-painful procedure (238). In a second study by this team, temporal patterns were demonstrated: they reported significantly longer latency to cry for sleeping infants than awake neonates following heel lance (236). Trials assessing the effectiveness of treatments to relieve pain associated with immunisations and needle-related procedures frequently use crying time as an outcome measure and results confirm that pain relief is likely to reduce crying time (247-250). However, predictable patterns that can be used for assessment of pain based on latency to cry or crying time are not available.

The acoustic qualities of the cry have also been tested to examine a potential association with pain. Bellieni and colleagues demonstrated that neonates with higher pain scores during blood sampling also had a higher fundamental frequency (251) and Facchini reported more noise patterns in the sound spectrograms in infants with higher pain scores (252). Decoupling between fundamental frequency and intensity contours during crying is considered to represent cortical control and therefore another potential way to indirectly measure pain. This relationship was explored in a study of neonates during heel lance and authors reported a trend towards a relationship between the extents of decoupling and pain scores. This infers a potential role for this analysis in pain assessment (253). Lehr and colleagues reported that the power and velocity of the cry (amplitude and rapidity) were stronger predictors of pain scores than features of arousal (turbulence and intensity) and the pitch of crying (frequency characteristics) (254). Positive correlations between cough frequency associated with cry and pain scores and negative correlations between fundamental frequencies and pain scores were reported in term neonates during venepuncture (255). In a study to examine patterns in crying associated with anger, fear and pain, objective analysis did not identify a consistent pattern associated with each emotion and carers were unable to correctly identify the emotion that was the cause of the crying (256). Furthermore, several research teams have reported variation within subjects and demonstrated that cry is also influenced by age (257) and age and gender (258). Finally, the results of other studies do not provide a convincing case to suggest that the acoustic characteristics of cry could be used to detect pain or assess intensity (238, 259-262)

Like the studies exploring the association between facial expression and pain, studies focused on the characteristics of cry use a procedural pain model and concentrate on neonates. However, it is less likely that there is a consistent, predictable and universal cry response to pain. Analysis of results to date have not included efforts to establish cut off values that might assist in

differentiating painful from non-painful stimulus, e.g. the duration of cry (seconds) following a stimulus which would differentiate a tactile stimulus from a heel lance. This makes interpretation of a single infant's cry response almost impossible. It is likely, given the extent of the variability within infants and between infants, that cut offs with high levels of sensitivity and specificity do not exist.

2.2.4 Pain mimics

The challenge for clinicians and researchers is identifying measures that are not only sensitive to pain but are also specific. It has long been recognised that there is potential overlap between the responses seen with other negative emotions and physiological arousal states and the responses associated with pain. A systematic review from 2007 highlights the paucity of data available that addresses the capacity of measures to discriminate between pain and other negative emotions (30). Since this review, Ahola Kohut and colleagues reported the results of their study that aimed to determine the capacity of the NFCS to distinguish between neonates experiencing a tactile stimulus and those experiencing a painful stimulus (245). They concluded that the NFCS was unable to make this distinction. As has been described there is limited evidence to confirm the capacity of any of the assessment techniques explored (imaging, physiological or behavioural) to differentiate between pain and non-pain related distress.

Von Bayer and Spagrud point out that scales are variably referred to as either a measure of pain, distress or pain and distress, despite similarities in the designs of these instruments and the similarities in their responsiveness to pain when tested (30). They stop short of recommending that all instruments are referred to as a measure of distress and not pain or pain and distress. However, Blount did make this recommendation and suggested that scales are uniformly considered distress scales (263). This differentiation between pain responses and other emotions can be particularly relevant for procedural pain where anticipatory fear and anxiety play a significant role in the experience. It has been suggested that pre-painful procedure distress is accepted as evidence of fear and anxiety and the procedural phase response as evidence of pain. Blount challenges this view based on data which demonstrate positive correlations between pre-procedural anxiety and procedural pain. His contention was that the behaviours seen in anticipation of and during the procedure are a chain of behaviours representing different intensities of pain-related distress rather than different emotions and may be explained by classic conditioning. Cues in the pre-procedural phase provoke a conditioned response which approximates the response associated with the original conditioning (e.g. painful procedure).

Blount likened this to ‘Pavlov’s dog’ and suggested that anticipatory distress is analogous to procedural pain.

Finally, children under the age of 8 to 9 years struggled to differentiate between the sensory experience of pain and their affective response to painful sensations in Goodenough and colleagues’ study (264). This adds weight to a contention that differentiation between pain and other negative emotions associated with procedures may well be fruitless as infants and children may not experience or recognise a meaningful difference between these experiences.

2.3 Psychometric properties and pain scale testing

Numerous indices of health cannot be measured directly and as a result instruments designed to approximate these indices have proliferated. This is particularly the case where the aim is to measure a subjective experience, such as pain. These instruments are frequently comprised of multiple items and may take the form of self-administered questionnaires and scales or observational tools applied by clinicians and researchers. In this section the concepts that relate to the development of these instruments and assessment of their performance are described, particularly as they relate to observational pain scales.

2.3.1 Instrument scales

Medical science conventionally used categorical scales to measure the presence or absence of an attribute, typically a disease. The social sciences have traditionally used an approach felt to be better suited for attributes that are considered to exist on a continuum and instruments are designed to grade the magnitude or severity of the attribute and place individuals along this continuum. These scales use ordinal, interval and ratio scales. Unquestionably, pain intensity cannot be dichotomised and is best described on a scale that allows for multiple levels of intensity.

Ideally the number of levels on a scale should reflect the capacity of the individual to discriminate between these levels so as not to lose valuable information about the construct or to exceed the precision with which the individual perceives the attribute. There is evidence to suggest that across a range of constructs individuals have difficulty perceiving more than seven (plus or minus two) levels and in the pain literature there is evidence that young children are unable to use traditional 6-point scales (125). Interestingly, the reliability is lower for instruments with fewer categories and this becomes most obvious where there are fewer than seven categories (265).

Respondents also show a preference for scales that include 5 to 9 categories and show a strong dislike for dichotomous scales, considering them less ‘accurate’ and ‘reliable’ and more ‘ambiguous’ and ‘restrictive’ (265).

2.3.2 Performance testing

Determining the diagnostic accuracy testing for measures where the results are either positive or negative uses methods where the results of the test of interest (‘positive’ and ‘negative’) are compared with the results of an existing test considered the ‘gold standard. The results are then defined as representative of either a ‘true’ or ‘false’ result depending on whether they agree with the ‘gold standard’ and summarised in terms of sensitivity and specificity (266). This paradigm is only possible where an accepted gold standard is available and furthermore where the results of the application of the test can be dichotomised. As has been noted, pain assessment for infants and young children experiencing a painful procedure does not meet either of these criteria.

For these measurement instruments to be useful for research and clinical practice, it is still essential to understand how well they perform. To achieve this, alternative methods to establish the measurement (psychometric) properties are employed. Studies are designed to assess reliability, validity and/or the feasibility of using the instrument, in this case an observational pain scale, to measure the construct of interest (pain).

It is tempting to consider that reliability and validity are fixed and that results of psychometric testing describe the reliability or validity of the instrument. However, these properties are a function of the instrument’s performance under the circumstances under which it was tested. This means that reliability and validity are more rightly considered properties of the measurements and not the instrument. This is an important distinction as it underpins the need for assessment of the measurement properties in a range of circumstances to replicate the circumstances of intended use. These properties (reliability, validity and feasibility) and the ways in which they are tested, to establish the psychometrics of pain scales, are discussed in the next sections of this chapter.

2.3.3 Reliability

Reliability refers to the reproducibility of a measure when either the instrument is applied by different observers, on different occasions (where the true value is unchanged) or using similar or parallel tests. Measurements can be considered reliable if these results are comparable. Reliability describes the proportion of the variability in scores which is due to true differences between

individuals and as such it is reported as a value between 0 and 1, where 0 is 'no reliability' and 1 is 'perfect reliability'(265). Reliability can also be described as a measure of the amount of random error in the measurements (267).

There are different ways in which reliability can be assessed and they include: internal reliability, test-retest reliability, intra-rater reliability and inter-rater reliability testing. Not all methods are relevant to all measurement instruments. For example, testing internal consistency is not possible for a single item instrument and inter-rater reliability is not applicable for self-administered instruments. In this section these approaches to reliability testing will be briefly described.

Internal consistency measures the average correlation between items for multi-item instruments (265, 267). The premise for this approach is that items measuring the same construct would be expected to correlate strongly. Conversely, if the correlation results are low it can be assumed that the items are either measuring different things or that respondents are inconsistent across different items. Internal consistency is an important step in the development of new instruments with multiple items and will guide inclusion and exclusion of items. Poorly correlated items are removed and less obviously those achieving near perfect correlations may also be removed as they do not add a great deal to the instrument. Removal of the physiological items from the original COMFORT scale (227) was in response to low internal validity for these items (268) and is an example of how internal consistency is used in scale design. Internal consistency is measured by testing a single application of the instrument by several observers and Cronbach's alpha is commonly used to analyse the result. However, there are limitations to this approach particularly where the instrument includes large numbers of items, e.g. in excess of 15 (267). Furthermore, this test does not account for variations in time or observers and therefore results may be falsely elevated and overestimate the extent of the reliability (267).

Test-retest reliability refers to the capacity of a test to provide similar results over time when the construct of interest is considered unlikely to have changed (267). The capacity to measure test-retest reliability and interpretation of the results are highly dependent on the stability of the construct of interest and the interval between the tests. For example, personality traits are unlikely to change over time, so repeat application of an instrument designed to measure personality traits weeks later is unlikely to result in different outcomes unless the tool is unreliable. The interval between tests may also be influenced by recall of previous responses. To overcome this, repeat applications are delayed until it is reasonable to consider that respondents will be unable to recall their response on the first test. The conventional interval is between 10 to 14 days (267). In contrast, pain is dynamic, and intensity may change frequently making interpretation of test-retest reliability regardless of the interval impossible.

Inter-rater reliability measures the consistency achieved between observers when the instrument is not self-administered. Many health-related instruments are observational scales and as such will be applied by many different observers (researchers and clinicians). To have confidence in the results, it must be shown that when the instrument is applied by different observers, they make similar observations. Inter-rater reliability is tested by correlating the independent responses of different observers measuring the same patient using the same instrument. Ideally the timing of these observations is also the same to eliminate the possibility that the condition of the individual has changed (267). Alternatively, intra-rater reliability describes the reliability or agreement between repeat measures by the same observer on separate occasions.

Test-retest, intra-rater and inter-rater reliability are calculated by correlating the results of the tests. The magnitude of the correlation coefficient is interpreted to determine reliability. There is considerable debate about the moșt appropriate correlation test and the interpretation of the coefficients. Pearson's r has been used to calculate a correlation coefficient but there is increased acceptance that the intraclass correlation coefficient (ICC) using a repeated measures analysis of variance assessing for absolute agreement is better suited as this accounts for the bias between or among observers and multiple assessments of the same individual over time (267).

The interpretation of reliability is further complicated by the lack of an agreed standard for the magnitude of the correlation coefficient that defines acceptable reliability. This is in part a measure of the range of factors that impact on the value of the coefficient and therefore the interpretation of reliability. As reliability is calculated based on a ratio of variability between subjects to total variability the magnitude of reliability coefficient for a heterogeneous sample will be much higher than for a more homogeneous sample with much lower between subject variability (265). Interpretation of the coefficient will also depend on the intended application of the instrument and whether this is similar to the circumstances under which it was tested and the importance of absolute agreement e.g. high stakes diagnoses such as cancer versus low stakes diagnoses such as minor injury (269).

There is growing interest in the Bland-Altman method for examining levels of agreement. Bland and Altman introduced this method about 20 years ago as an alternative to correlation coefficients which they considered inappropriate for comparing the agreement between two instruments (270). They argued that high correlations reflect strong relationships between variables but not necessarily agreement. Perfect agreement is only achieved if we get the same results using the two different instruments. Perfect correlation occurs when there is consistent difference in scores using the two different instruments even if the scores are not the same e.g. the scores from application of one instrument are always half the other instrument's score. Mathematically,

agreement can be expressed as X = Y whereas correlation can be summarised as X = aY + b (where 'X' represents the scores from instrument ' and 'Y' the scores from instrument Y, 'a' is the slope of the line and 'b' is the point at which X cuts the Y axis) (265).

The aim of the Bland-Altman approach is to determine the extent of the difference between the scores from two instruments to establish whether this is clinically significant (270). The first step in this approach is to plot the difference between the results from the two instruments against the mean of the results to visually demonstrate the between methods differences. The bias is estimated by the mean difference and the standard deviation of the differences. It would be expected that all differences would be within 2 standard deviations of the mean difference. These values are considered the upper and lower limits of agreement. They represent the extent to which the results of one instrument might vary from the other and the clinical significance of these values are used to determine whether the instruments can be considered interchangeable. The standard error and confidence intervals are calculated to report the precision of the estimated limits of agreement.

2.3.4 Validity

Validity defines the extent to which the scores represent a true measure of the construct under investigation and whether scores accurately distinguish between individuals with higher or lower levels of the construct (265). In other words, whether the instrument accurately measures what it purports to measure and appropriately discriminates between varying degrees of the construct. Using a pain intensity scale as an example, to consider scores valid individuals who score 'zero' should have no pain and in reverse individuals with no pain should score 'zero'. The same should be seen for high scores; they should reflect an experience of high levels of pain.

The reliability of scores also impacts on the assessment of validity as the maximum value for validity is limited by the reliability. This is expressed in the following equation;

$$\text{Validity}_{max} = \sqrt{(\text{Reliability}_{\text{new test}})\,(\text{Reliability}_{\text{criterion}})} \quad (265)$$

This means that if reliability for the new test is low (e.g. 0.4), even if the reliability of the criterion to which it is compared is high (e.g. 0.9), the maximum value for validity will be limited by this and cannot exceed 0.6. Intuitively this makes sense as if something is not measured consistently by observers then it is likely that the construct is also not measured accurately, in other words if the answers are not the same, they cannot both be right.

Traditional definitions of validity have included different types of validity and including: content, criterion and construct. More recently, the agreed understanding of validity acknowledges these are methods to test validity rather than different types of validity and that validity is demonstrated cumulatively and circumstantially using a range of methods (269).

2.3.4.1 Face and content validation

Face and content validation are judgements of whether or not the instrument is reasonable (265). Both are determined by experts and are rarely supported by empirical data. Face validation is the first judgement made about a scale and describes whether on the 'face' of it the instrument measures what is intended. Content validation extends this to include an assessment of whether the instrument captures all domains of the construct. Furthermore, judgements of the relevancy of included items are made to ensure that items that do not measure the construct of interest are removed. This means that a scale designed to measure pain based on observed behaviours should include all behaviours that can be used to discriminate the individual experiencing pain from one who is not and to establish the intensity of the pain. Construct validation will also ensure that the scale does not include items that measure experiences other than pain, for example fatigue. The content of the instrument is determined by the instrument developer based on other instruments, the literature and the views of experts. Although content validation does not often involve statistical tests, it can be seen to be inversely related to internal consistency (267). Therefore, decisions designed to improve the internal consistency of the items on an instrument may result in the removal of items with low correlation to other instrument items. Paradoxically, this may reduce the scope of the construct assessed by the instrument.

2.3.4.2 Criterion validation

Criterion validation is used when there is at least one other existing measure of the phenomenon of interest to which the performance of the experimental instrument can be compared (265). Criterion validation methods include concurrent validation where the experimental instrument and the 'gold' standard are applied at the same time and predictive validation where the 'gold' standard is applied later and the capacity of the results of the new instrument to predict a later outcome is determined (265). In both cases to accept the instrument as valid relies on a strong correlation between the scores from the test of interest with those derived from application of the 'gold' standard.

The assumption behind criterion validation is that the existing measure is the 'gold' standard and therefore provides scores that most closely replicate a 'true' measure of the construct (265). There

are several flaws in the logic that underpin this approach to validation. The premise for developing a new instrument is often the development of a new, more practical, safer or less expensive alternative to the existing instrument or an instrument that provides a timelier result than the criterion. However, on some occasions a new instrument is developed as a more accurate alternative to the existing instrument. In this circumstance it is counterintuitive to measure the performance of the new and improved instrument against the existing potentially less accurate one. This creates confusion when interpreting the correlation coefficients. (265)

Another challenge to the logic and interpretation of criterion validation is that for many instruments used in health care there is not always an objective criterion available and the developers of the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) Checklist were in agreement that no gold standards exist for patient reported health measures (271). In these circumstances the new or experimental instrument is compared with the performance of an existing instrument that is also not a direct measure of the construct of interest. This requires that the existing tool is supported by validation data. This becomes an endless cascade of validation studies as each instrument requires validation against another criterion which is likely another indirect measure of the construct (272). Without complete confidence in the accuracy of the existing instrument it is impossible to interpret the performance of the experimental instrument based on the performance of the existing one. For example, if the scores are highly correlated it can be concluded that the instruments are measuring the same thing but not necessarily what they are measuring. Conversely if they are poorly correlated it can be concluded that the scales are not measuring the same thing. However, in this situation it cannot be concluded that the original instrument is accurate and that the experimental instrument is not valid.

Messick has offered more descriptive terms to better describe the rationale for the development of a new tool and therefore the validation methods; diagnostic utility and predictive utility (273). Diagnostic utility describes the methods of validation used to confirm that the new, less expensive, safer or more practical, instrument performs at least as well as the existing criterion. The new instrument and the existing criterion would be applied at the same time and the outcomes compared. The FLACC scale was developed as a more practical (easily remembered) scale than existing pain scales and could be validated in this way (28). Diagnostic utility would see the performance of this scale assessed against an existing scale that was considered the criterion or the ‘gold’ standard. Predictive utility defines the capacity of an instrument to predict an outcome that has as yet not occurred. There does not seem an obvious role for prediction in pain assessment. Hence, predictive validation methods are unlikely to be useful for validation of pain assessment scales.

Finally, criterion contamination exists when the outcome of the criterion ('gold' standard) is predicated on the results based on the new instrument (265). This may be the case where the criterion instrument includes data from the new instrument; for example; patient's report of pain using a visual analogue scale (VAS) is the criterion and is applied after the patient report of their pain using a simpler four-point verbal rating scale. The score given using the VAS is informed by the score allocated using the new instrument and so high levels of correlation between the scores could be reasonably expected. There are scenarios where the results of the criterion are likely to be predicted by the new instrument and that this is acceptable, for example; where the new instrument is a modification e.g. a shorter version of an existing instrument or one that uses similar items to the criterion but is more easily remembered. In these situations, providing the application of the scales does not bias the outcomes, high correlation is both predictable and desirable (265).

2.3.4.3 Construct validation

Construct validation includes methods to indirectly establish validation of a new instrument where there are no existing instruments or criteria available to measure the construct (265), for example the first observational pain scale developed to assess pain in children too young to self-report. Additionally, construct validation is used where a new instrument has been developed as the existing instruments were considered a poor measure of the construct and therefore should not be used in criterion validation studies. Construct validation involves testing predictions about the construct that would confirm the capacity of the instrument to measure it accurately (265). Construct validation relies on an understanding of the construct and its relationship with other constructs, which can be used to generate these predictions or hypotheses (265, 267). Construct validation could be said to provide circumstantial evidence of validity and therefore a single study supporting a hypothesis will not be sufficient to assert validity. Evidence must be built from multiple tests of a range of hypotheses. Interestingly, it has been said that it only takes negative results from one well-designed study to question the entire construct (274). Furthermore, it is unclear where the problem lies: with the experimental instrument, the hypothesis or with both. Commonly used methods which are examples of construct validation include extreme groups and convergent and discriminant validation (267).

'Extreme groups' tests the performance of the instrument used to measure the construct in two groups; one group known to have the trait and the other which does not (265, 271). It is expected that the instrument scores would accurately distinguish between these two groups. This may also be described as discriminative validation. The key problem with this method is the identification of these two groups, particularly where the instrument has been developed in the absence of an

appropriate method for measurement of the construct as is the case for pain experienced by infants and children unable to self-report. This method is also blunt in as much as it demonstrates the capacity of the instrument to measure the extremes but does not test the capacity to recognise more equivocal conditions, e.g. mild to moderate pain rather than no pain versus severe pain (265).

Comparing the performance of a new instrument against the performance of other instruments measuring the same construct of measures or similar related variables is defined as convergent validation and is another construct validation method (265). Studies aiming to demonstrate the validity of a pain scale frequently correlate the scores from the experimental scale with scores from another pain scale. Less commonly, correlations between pain scores from the experimental scales are correlated with other variables based on an assumption that they are related in some way e.g. pain and fear. The corollary to convergence is discriminant validation. Not only should the instrument scores correlate strongly with scores using another instrument measuring the same construct but there should be no correlation with scores for dissimilar and unrelated constructs, e.g. pain and breathlessness. Once again, unpredictable results may reflect a flaw in the scale, the hypothesis or both (265).

Over the last 30 years there has been a trend to consider all validation methods construct validation (269). The premise for this is that criterion validation is based on the hypothesis that there will be a positive correlation between the scores generated with a new instrument and the criterion (273). Content validation is also based on the hypothesis that all items are measuring a facet of the same construct and that a significant feature of the construct is not overlooked. Therefore, content validation is considered by many as another type of construct validation (265).

Responsiveness and sensitivity to change are methods for validation that have been increasingly used to explore validity. These terms are often used interchangeably by authors; however Liang makes a distinction between them which is based on the relevance of the result. Sensitivity to change is the ability of the instrument to detect significant change even if it is not meaningful while responsiveness is the ability of the instrument to measure meaningful change (275). This is well illustrated using pain scale scores where *sensitivity to change* is shown if there is a statistical difference between pain scores, however the difference is clinically meaningless e.g. a difference of 0.5 between scores that is reported as statistically significant. *Responsiveness* is shown if there is a statistically significant difference between scores that is also clinically significant e.g. a difference of 2 between pain scores (clinical significance for many scales is acknowledged as ‘2’ (276)). Both approaches to validation are also a construct validation method based on a hypothesis

that scores will change over time or in response to treatment (e.g. analgesic) or a stimulus (e.g. painful).

Practically, the language used to define validation is less important than the need to use a range of validation methods to demonstrate the validity of the scores and to repeat this testing under similar and different circumstances. Repeated demonstration of the validity of the scores provides a platform for extrapolating from these scores to the potential for scores associated with the next application of the scale under these circumstances to be valid.

2.3.5 Feasibility and clinical utility

Clinical utility describes the extent to which the results influence clinical decision-making and clinical outcomes (277). It is now considered insufficient to introduce an instrument to clinical practice based solely on validation data. It must also be shown that the instrument produces clinically useful results. This concept has entered the diagnostic accuracy literature, but evaluation of this property has been variable. A common approach has been to assess how useful clinicians consider the information provided by the instrument and whether it is likely to influence their treatment choices. This has been criticised as too subjective and not necessarily a reflection of clinician behaviour as there is evidence that clinicians reported intentions do not always align with their practice. Ideally, assessment of the clinical utility of a measure will include methods to determine the extent to which use of the measure results in changes in clinical decisions and more importantly improved clinical outcomes (277).

Instruments used for clinical or research purposed must accurately measure the construct in question. However, they must also be feasible for the circumstances. Instruments that are overly complicated, rely on extensive training or specialised equipment to use them, take too long to apply or are expensive are unlikely to gain favour regardless of how well they perform. The threshold for what is consider feasible will differ depending on whether they are proposed for research or clinical use (277). The assessment of feasibility may also involve methods designed to measure the ease of application, efficiency and cost analysis.

2.4 Conclusion

Pain has attracted a great deal of interest over many decades and recent advances in technology correspond with exponential growth in our understanding of the complex neurophysiology and the associated experience of pain. It is well beyond the scope of this chapter and this thesis to

explore the evidence that details current understanding. This chapter provides a simple summary of these concepts to help place the research reported in this thesis in context.

SECTION 2.

Section 2 reports the first phase of this project which was a detailed interrogation of the literature to identify behavioural pain scales that are potentially suitable to assess procedural pain in infants and young children. This work is reported in one chapter (Chapter 3).

CHAPTER 3.

The aim of this phase of the project, the approach to the search, the search results, the criteria used to define 'potentially suitable' pain scales and the scales identified in the literature that met these predefined criteria, are reported in this chapter. It concludes with a description of the eligible scales and current understanding of their role in assessment of procedural pain in infants and young children.

3.1 Background and aims

During their healthcare, infants and children are often exposed to painful and uncomfortable procedures. It is increasingly recognised that unrelieved or poorly managed pain can have significant sequelae (including long-term sequelae), especially for preterm infants and children. A growing body of evidence demonstrates alterations in neural pathways (278, 279), physical development (280) and pain-related behaviours (11, 281) as a result of neonatal exposure to pain (13). The impact of painful experiences during infancy and childhood has not been as well studied but is likely to provoke significant fear and anxiety associated with healthcare and future avoidance of healthcare (282, 283). In light of this, procedural pain management must be a priority for clinicians and researchers involved in treating infants and children and to identify efficacious pain management strategies for use in diverse settings where painful procedures take place. However, good management is contingent on recognition of pain and accurate assessment to guide treatment choices.

Self-report of pain is generally accepted as the ideal way in which to assess pain, determine the need for strategies to relieve pain and assess the success of interventions (119, 284, 285). However, the cognitive and language development of infants and young children prevents them from self-reporting their pain experience. Clinicians and researchers rely on proxy measures to estimate the infant and young child's pain experience. Observational behavioural assessment scales are one of the most commonly used and most practical methods for pain assessment. These scales are comprised of behaviours and in some cases physiological features considered indicative of pain in infants and children. Over a decade ago, Duhn and colleagues reported the existence of over 40 published scales in the literature designed to assess pain in this age group (286) making it challenging for clinicians and researchers to select appropriately.

Systematic reviews, clinical practice guidelines and experts in the field have provided clues as to the best scales for assessing pain in this age group (213). However, consensus eludes us and there remains inconsistency across recommendations and in clinical and research application (30, 213). The aim of this phase of the project was to identify observational behavioural scales that were designed for and considered suitable or used for assessment of procedural pain experienced by infants and young children unable to self-report pain that may be used for clinical or research purposes.

3.2 Methods

A systematic review of the literature was conducted to identify observational behavioural scales that could be considered *potentially suitable* to assess and quantify procedural pain intensity experienced by infants and young children. A search strategy and the criteria used to define scales as *potentially suitable* were defined prior to conducting this review of the literature. The results of the search were then reviewed against these criteria to identify scales considered *potentially suitable*.

3.2.1 Inclusion/exclusion criteria

Scales were considered *potentially suitable* to quantify procedural pain intensity in infants and children if they met a series of criteria, which are described here and summarised in Table 3-1. The criteria were carefully considered and agreed to by the authors with the aim of identifying a scale that could be used for clinical and research purposes to assess procedural pain in infants and children unable to self-report. Several scale attributes formed the basis for considering a scale *potentially suitable*. The scale must assess pain using behavioural features of pain that can be identified by an observer without the need for extensive training or specialist equipment. Scales that included physiological parameters such as heart rate in addition to behavioural indices were not excluded providing these parameters could be collected at the same time to support immediate scoring of pain. Use of a 0 to 10 scale to quantify pain intensity has gained increasing favour as the accepted metric to support consistency and comparability in measurement and communication (287). For this reason, only scales quantifying pain intensity using a 0 to 10 metric were accepted as *potentially suitable*. In addition to pain scales, scales referred to as a measure of ‘pain-related distress’ or ‘pain and distress’ were included to reflect the diversity of language often use to describe scales with similar assessment purpose.

To narrow the focus, the criteria for scales considered *potentially suitable* were further defined as scales considered *potentially appropriate* for procedural use and scales considered to have been *accepted* for this purpose. The agreed principle underpinning the criteria defining *potentially appropriate* was that well-designed scales and those with data to support reliability and validity when applied to assess procedural pain would be appropriate for this purpose. Therefore, scales were identified as *potentially appropriate* if they were recommended in a systematic review for assessing procedural pain in infants and/or children or had undergone psychometric testing to evaluate its reliability and validity when used for this purpose. Furthermore, it was assumed that scales purposefully designed for procedural pain assessment would be designed to capture the unique circumstances of a medical procedure and therefore should also be considered *potentially appropriate*.

Table 3-1 Criteria for identifying scales *potentially suitable* for assessing procedural pain in infants and young children.

Scale inclusion criteria	Scale exclusion criteria
Scale attributes: - Scale quantifies 'pain', 'pain-related distress' or 'pain & distress' AND - Observational pain scale based on behavioural indices of pain AND - Scores intensity using a 0 to 10 metric Potentially appropriate: - Scales recommended by systematic review of the psychometrics of pain scales OR - Scales designed to assess procedural pain &/OR - Undergone psychometric testing for procedural pain assessment use Accepted: Used in a minimum of 5 RCTs from unique research groups OR Recommended in an expert consensus statement or CPG Available in English	Designed exclusively as a measure of distress (no reference to pain assessment) Designed &/or tested exclusively for neonates Designed &/or tested exclusively for cognitively impaired infants & children Designed &/or tested exclusively for adults Scale does not rely on assessment of observable behaviours or uses indices of pain that are not immediately measurable e.g. physiological parameters such as: salivary cortisol Scores intensity using a metric other than 0 to 10. Scale &/or publications not available in English

Abbreviations: CPG – clinical practice guideline, RCT – randomised controlled trial

The foundation for including *accepted* scales was the view that acceptance is indirectly evidence of expert opinion acknowledging the likely fitness for purpose of the scale. Therefore, *accepted* scales were defined as those scales that were either recommended in published expert consensus statements or clinical practice guidelines or were used in at least five randomised controlled trials (RCT) published by unique research groups. Expert consensus statements and guidelines were defined as statements and guidelines published by or endorsed by specialty organisations, associations, societies and colleges. These groups were further defined as national and international professional bodies representing a medical and/or allied health professional membership with a pain, emergency care, anaesthetic, pain or paediatric focus. National and international organisations established with the explicit purpose of producing evidence-based consensus statements e.g. Joanna Briggs Institute, were also included as experts.

Scales designed solely as a measure of distress which did not acknowledge a role in assessing pain were excluded. Scales that were designed and assessed for use in one language and culture cannot be assumed to be appropriate for use in another without adequate translation and testing. For this reason, scales not available in English were also excluded. Additionally, scales exclusively designed for and tested to assess neonatal (term and preterm) pain or pain experienced by cognitively impaired children and/or children with altered conscious states were excluded. Finally, scales where psychometric testing was based on application of the scale to adults and children and the data for analysis was pooled were also excluded.

3.2.2 Search strategy

Identification of scales that met the pre-defined criteria was achieved by a series of searches to target the specific criteria. Each search is described separately, and the search terms used for these searches are listed in Appendix A (Box 1 – 4). In addition, the reference lists of all relevant publications were reviewed to identify other potentially relevant publications and data sources. See Table 3-2 for a summary of these search strategies.

3.2.2.1 Search: Potentially appropriate scales

Two searches were conducted to identify *potentially appropriate* scales; the first to identify systematic reviews of studies examining the psychometric properties of pain assessment scales used to assess procedural pain and the second to retrieve individual studies assessing the psychometric properties of scales used to assess procedural pain. Both searches used search terms to identify publications focused on the population of interest e.g. ‘infant’ or ‘child’ and the topic of interest e.g. ‘pain assessment’ or ‘pain measurement’. The search to identify systematic reviews

limited the search by publication type e.g. ‘systematic review’ and the search to identify psychometric evaluation studies included additional search terms e.g. ‘validation’, ‘validity’, or ‘reliability’ etc.

The electronic databases; MEDLINE, Embase and PsycINFO using Ovid, the Cochrane Database of Systematic Reviews and Cochrane Controlled Trials and Cumulative Index Nursing and Allied Health Literature (CINAHL) using Ebsco were searched. The search occurred in week 4 of June 2014. All searches were limited to publications in English.

3.2.2.2 Search: Accepted scales

Identification of *accepted* scales involved two search strategies: the first to identify RCTs using an observational behavioural scale to assess procedural pain in infants and children and the second to locate expert consensus statements and clinical practice guideline recommending scales for this purpose.

To retrieve as many relevant RCTs as possible the search was broadened compared to the earlier searches which focused on ‘pain assessment/measurement’ and instead the terms ‘pain’ and ‘pain management’ were used. However, to restrict the results to trials focussing on procedural pain a range of terms such as; ‘procedure’, ‘blood sampling’, ‘immunisation’ and ‘wound care’ were added to the search strategy. Finally, the search was also limited by publication type to ‘clinical trial’ OR ‘RCT’.

A search to identify expert consensus statements and clinical practice guidelines also used terms to define the population of interest; ‘infant’ OR ‘child’ and the topic of interest which was once again kept broad; e.g. ‘pain’. To restrict the results to expert consensus statements and clinical practice guidelines, terms such as; ‘consensus’ OR ‘guideline’ OR ‘practice guideline’ were included.

MEDLINE, Embase and PsycINFO using the Ovid platform, the Cochrane Database of Systematic Reviews and Cochrane Controlled Trials and CINAHL using the Ebsco platform were searched separately in the same week as the search for *potentially appropriate* scales (week 4, June 2014) to retrieve relevant publications.

Additionally, a search of the web using the Google search engine to identify the websites of specialty organisations, associations, societies and colleges was completed using combinations of

the following terms; ‘organisation’ OR ‘college’ OR ‘society’ OR ‘professional body’ ‘pain’ OR ‘emergency’ OR anaesthetics’ OR ‘paediatrics’ and ‘medical’ OR ‘nursing’ OR ‘psychology’.

Table 3-2 Criteria to identify appropriate publication sources for potentially suitable scales.

Criteria	Inclusion criteria	Exclusion criteria
Potentially appropriate		
Identified in a systematic review	Described in title or abstract as: ‘systematic’ or ‘integrative’ review. Review methods documented, specifically: search, inclusion & exclusion criteria AND Includes analysis of data addressing observer assessment of procedural pain assessment in infants &/or children	Publication not identified as a systematic or integrative review Results for psychometric properties of scales for procedural pain assessment use in infants & children not reported independently
OR		
Psychometric data available	Study provides data assessing scale score; reliability, test-retest, validation, clinical utility or feasibility when used to assess procedural pain in infants &/or children using behavioural observation scales	Study does not analyse procedural pain assessment data for infants &/or children separately
Accepted scale		
Use in RCT	Scale used by observer to assess procedural pain in infants & children as a study outcome	Non-randomised trials
OR		
Consensus statements & CPGs	Statements published, commissioned or endorsed by national & international specialty organisations (associations, colleges & societies) Organisation official language English AND Membership base is; medicine, nursing or psychology & specialty focus is: paediatrics/child health, emergency care, anaesthetics or pain Statements published by independent agencies established to publish evidence-based guidelines & statements	Statements published or commissioned by: government agencies, professional regulatory authorities, panels established for commercial purposes e.g. pharmaceutical company panels

Abbreviations: CPG – Clinical Practice Guideline, RCT - Randomised Controlled Trial

3.2.3 Study and scale selection

The publications retrieved from all searches, following removal of duplicates, were reviewed and the scales reported in the retrieved publications assessed for eligibility. The behavioural observation scales that met the criteria for *potentially appropriate* and *accepted* were reviewed in detail to confirm that the scale attributes also met the inclusion criteria, e.g. observer applied scales based on behavioural indices considered indicative of pain and designed to quantify pain intensity using a 0 to 10 metric. Scales that were designed and tested exclusively to assess pain in neonates or cognitively impaired children, to exclusively assess distress or were not available in English were excluded.

3.3 Results

The following sections report the results of each search (Table 3-3) and the review of the reported scales against the eligibility criteria. In summary, 36 scales were identified from a total of 2669 citations retrieved from the database searches and the web search. Several scales were identified in more than one search strategy. Only two of the identified scales were originally designed to assess procedural in infants and young children; the MBPS (288) and the EVENDOL (289). Following review of these scales and supporting literature, three eligible scales were identified for inclusion; the Face, Legs Activity, Cry and Consolability (FLACC) scale, the Modified Behavioural Pain Scale (MBPS) and the Visual Analogue Scale applied by an observer (VASobs).

Table 3-3 Search summary (Search date: 30 June 2014).

Database searched	Systematic reviews	Psychometric evaluation studies	RCTs	Consensus statements & CPGs
Medline	115	459	185	18
Embase	3	9	513	186
CINAHL	62	395	298	174
PsychInfo	5	32	10	0
Cochrane [a]	0	0	556	-
Web search	-	-	-	8
Secondary search [b]	0	1	0	2
Total reviewed [c]	179	782	1378	330
Relevant publications	2	18	171	5

[a] Cochrane Database of Systematic Reviews and Cochrane Controlled Trials

[b] Reference list of other relevant publications

[c] Abstracts retrieved minus duplicate citations

Abbreviations: CPG – Clinical Practice Guidelines, RCT – Randomised Controlled Trial

3.3.1 Potentially appropriate scales

3.3.1.1 Systematic review

Only two systematic reviews were identified that made recommendations for the most suitable behavioural observation scale for procedural pain assessment in infants and/or young children (30, 31). The FLACC scale was recommended in both systematic reviews, while the Children's Hospital of Eastern Ontario Pain Scale (CHEOPS) was only recommended in von Baeyer's systematic review (30). A review by van Dijk and colleagues (290) was excluded as although a review of available evidence to summarise the reliability and validity of the VASobs, it was not reported as a systematic or integrative review and did not meet the methodological criteria for eligibility.

3.3.1.2 Psychometric evaluation

A total of 18 papers were identified that addressed psychometric evaluation of 13 observational behavioural scales used to assess procedural pain experienced by infants and/or children. These scales and the studies are listed in Table 3-4.

The psychometric properties of the Children's and Infant's Postoperative Pain Scale (ChIPPS) used to assess procedural pain was examined in a cross validation to test the scales performance following translation into Brazilian Portuguese. There are no studies available testing the scale's psychometric properties when used to assess procedural pain using the English version of this scale. For this reason, this scale did not meet the criteria for *potentially appropriate* and was excluded from further review. Similarly, the EVENDOL was developed by a French team and to date psychometric analysis of the scale has only been completed for the French version of this scale hence, EVENDOL was also excluded from further review.

3.3.2 Accepted scales

3.3.2.1 RCT

Thirty-two scales (Table 3-5) were identified in 171 RCTs where at least one observational behavioural pain scale was used to assess procedural pain in infants and children. Twenty-two studies reported use of more than one scale. Four scales; FLACC, MBPS, VASobs and CHEOPS, account for those used in over three quarters of the trials assessed (n = 133, 77%). The VASobs was the most frequently used of all scales and appeared in 66 trials, while MBPS and FLACC were used in similar numbers of studies (n = 26 and n = 23, respectively). More than a third of the scales (n = 14, 43%) were only used once and the remaining scales (n = 15, 45%) on fewer than five occasions each. Therefore, for this reason these scales were not considered as *accepted* for procedural pain assessment and were excluded from further review.

Table 3-4 Scales identified with data addressing the psychometric properties of the scale used to assess procedural pain in infants and children and the studies reporting this data.

Scale	Studies
Children's Hospital Eastern Ontario Pain Scale (CHEOPS) (29)	Stein PR. Indices of pain intensity: construct validity among preschoolers. Pediatric Nursing. 1995;21(2):119-23.
Children's & Infant Postoperative Pain Scale (ChIPPS)[a] (167)	Alves MM, Carvalho PR, Wagner MB, Castoldi A, Becker MM, Silva CC. Cross-validation of the Children's and Infants' Postoperative Pain Scale in Brazilian Children. Pain Practice. 2008;8(3):171-6.
COMFORT (227)	de Jong AEE, Bremer M, van Komen R, Middelkoop E, Tuinebreijer W, Baartmans M, et al. Reliability, validity and practicality of the Pain Observation Scale for Young Children, the COMFORT Scale and the Visual Analogue Scale in young children with burns. Burns. 2009;35:S6.
EVENDOL[b] (289)	Fournier-Charrière E, Tourniaire B, Carbajal R, Cimerman P, Lassauge F, Ricard C, et al. EVENDOL, a new behavioral pain scale for children ages 0 to 7years in the emergency department: Design and validation. Pain. 2012;153(8):1573-82.
Face, Legs, Activity, Cry, Consolability (FLACC) scale (28)	Gomez RJ, Barrowman N, Elia S, Manias E, Royle J, Harrison D. Reliability of the FLACC scale for evaluating pain in toddlers during immunization. Pain Research and Management. 2012;17 (3):210-1. Ranger M, Celeste Johnston C, Rennick JE, Limperopoulos C, Heldt T, du Plessis AJ. A Multidimensional Approach to Pain Assessment in Critically Ill Infants During a Painful Procedure. Clinical Journal of Pain. 2013. Taddio A, Hogan ME, Moyer P, Girgis A, Gerges S, Wang L, et al. Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. Vaccine. 2011;29(7):1390-4. Nilsson S, Finnstrom B, Kokinsky E. The FLACC behavioral scale for procedural pain assessment in children aged 5-16 years. Paediatric Anaesthesia. 2008;18(8):767-74.

Modified Behavioural Pain Scale (MBPS) (288)	Pillai Riddell R, Flora DB, Stevens SA, Stevens B, Cohen LL, Greenberg S, et al. Variability in infant acute pain responding meaningfully obscured by averaging pain responses. Pain. 2013;154(5):714-21. Taddio A, Hogan ME, Moyer P, Girgis A, Gerges S, Wang L, et al. Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. Vaccine. 2011;29(7):1390-4. Taddio A, Nulman I, Koren BS, Stevens B, Koren G. A revised measure of acute pain in infants. J Pain Symptom Manage. 1995;10(6):456-63.
Modified Paediatric Pain Scale (mPEPPS) (291)	Schultz AA, Strout TD, Jordan P, Worthing B. Safety, tolerability, and efficacy of iontophoresis with lidocaine for dermal anesthesia in ED pediatric patients. Journal of Emergency Nursing. 2002;28(4):289-96.
Neonatal Facial Coding System (NFCS) (236)	Ahola Kohut S, Pillai Riddell R. Does the Neonatal Facial Coding System differentiate between infants experiencing pain-related and non-pain-related distress? Journal of Pain. 2009;10(2):214-20.
Neonatal & Infant Pain Scale (NIPS) (292)	Taddio A, Hogan ME, Moyer P, Girgis A, Gerges S, Wang L, et al. Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. Vaccine. 2011;29(7):1390-4.
Numeric Rating Scale observer (NRSobs)[c] (293)	Eyelade OR, Oladokun RE, Fatiregun AA. Convergent validity of pain measuring tools among Nigerian children. African Journal of Medicine & Medical Sciences. 2009;38(4):333-6.
Pain Observation Scale for Young Children (POCIS) (294)	de Jong AE, Bremer M, Schouten M, Tuinebreijer WE, Faber AW. Reliability and validity of the pain observation scale for young children and the visual analogue scale in children with burns. Burns. 2005;31(2):198-204. de Jong AEE, Bremer M, van Komen R, Middelkoop E, Tuinebreijer W, Baartmans M, et al. Reliability, validity and practicality of the Pain Observation Scale for Young Children, the COMFORT Scale and the Visual Analogue Scale in young children with burns. Burns. 2009;35:S6.
University Wisconsin Children's Hospital Pain Scale (UWCHPS) (295)	Soetenga D, Frank J, Pellino TA. Assessment of the validity and reliability of the University of Wisconsin Children's Hospital Pain scale for Preverbal and Nonverbal Children. Pediatric Nursing. 1999;25(6):670-6.

Visual Analogue Scale observer (VASobs)[c] (118)	de Jong AE, Bremer M, Schouten M, Tuinebreijer WE, Faber AW. Reliability and validity of the pain observation scale for young children and the visual analogue scale in children with burns. Burns. 2005;31(2):198-204. de Jong AEE, Bremer M, van Komen R, Middelkoop E, Tuinebreijer W, Baartmans M, et al. Reliability, validity and practicality of the Pain Observation Scale for Young Children, the COMFORT Scale and the Visual Analogue Scale in young children with burns. Burns. 2009;35:S6. McClellan CB, Schatz JC, Mark TR, McKelvy A, Puffer E, Roberts CW, et al. Criterion and convergent validity for 4 measures of pain in a pediatric sickle cell disease population. The Clinical Journal of Pain. 2009;25(2):146-52. Eyelade OR, Oladokun RE, Fatiregun AA. Convergent validity of pain measuring tools among Nigerian children. African Journal of Medicine & Medical Sciences. 2009;38(4):333-6. Taddio A, O'Brien L, Ipp M, Stephens D, Goldbach M, Koren G. Reliability and validity of observer ratings of pain using the visual analog scale (VAS) in infants undergoing immunization injections. Pain. 2009;147(1-3):141-6. Schultz AA, Strout TD, Jordan P, Worthing B. Safety, tolerability, and efficacy of iontophoresis with lidocaine for dermal anesthesia in ED pediatric patients. Journal of Emergency Nursing. 2002;28(4):289-96.

[a] Study tested scale translated into Brazilian Portuguese

[b] Scale developed and tested in French only

[c] Scale developed for self-report but applied by an observer

References relate to the original scale author/publication

Table 3-5 Scales used in an RCT to measure procedural pain in infants and children.

Scale	No of RCTs
Visual Analogue Scale observer[a] (VASobs) (118)	66
Modified Behavioural Pain Scale (MBPS) (288)	26
Face, Legs, Activity, Cry, Consolability (FLACC) scale (28)	23
Children's Hospital Eastern Ontario Pain Scale (CHEOPS) (29)	15
5-point scale author defined (296-299)	5
Face Pain Scale-Revised observer[a] (FPS-Robs) (300)	4
3-point scale author defined scale (301-304)	4
4-point scale author defined scale (305-308)	4
University Wisconsin Children's Hospital Pain Scale (UWCHPS) (295)	4
Modified Frankl Rating Scale (mFRANKL) (309)	3
Neonatal & Infant Pain Scale (NIPS) (292)	3
Neonatal Facial Coding System (NFCS) (236)	3
Numeric Rating Scale observer[a] (NRSobs)	3
Verbal Rating Scale observer[a] (VRSobs)	3
Children's & Infant Postoperative Pain Scale (ChIPPS) (167)	2
Faces Pain Scale observer[a] (FPSobs) (310)	2
Facial Affective Scale observer[a] (FASobs) (311)	2
Observational Pain Scale (OPS) (312)	2
OUCHER observer[a] (OUCHERobs) (313)	2
0 - 10 scale (314)	1
2-point author defined scale (315)	1
CASobs[a] (311)	1
Children's Facial Coding System (CFCS) (229)	1
Douleur Aiguë du Nouveau-né (DAN) (316)	1
Faces scale observer[a] (317)	1
Facial Grimace Scale observation (318)	1
Maximally Discriminative Facial Movement Coding System (MAX) (319)	1
Modified Children's Hospital Eastern Ontario Pain Scale (mCHEOPS) (320)	1
Modified Neonatal Facial Coding System (mNFCS) (321)	1
Modified Observational Pain Scale (mOPS) (322)	1
Modified Riley Pain Scale (mRILEY PS) (323)	1
Premature Infant Pain Profile (PIPP) (218)	1

[a] Observer denotes a scale designed for self-report but applied by an observer

References relate to the original scale author/publication

3.3.2.2 Expert statements

A total of 71 association, society, academy, collaboration, organisation and network sites located in Australia, New Zealand, United States, Canada and the British Isles were reviewed. In addition, several databases and repositories that maintain lists of evidence-based guidelines were also searched to identify relevant publications (Listed in Appendix A, Box 5*).* Recommendations for suitable observational scales for procedural pain assessment were available from, or endorsed by, eight professional groups via clinical guidelines and consensus statements (Table 3-6). Two organisations endorsed a third-party statement (the Royal College of Nursing (RCN) statement) rather than authoring their own. This resulted in six unique guidelines and consensus statement in which scale recommendations could be found. Statements and guidelines which had not been found via the web search of association, society, academy, collaboration, organisation and network sites were not located as a result of searching databases such as Medline etc.

The FLACC scale was recommended in six documents (15, 32, 285, 324-326), the CHEOPS in three (15, 32, 285) and in addition to these scales, the COMFORT and the University of Wisconsin Children's Hospital (UWCH) pain scale were also recommended in the Royal College of Nursing Guideline (32). The Pediatric Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (PedIMMPACT) Consensus Statement endorsed by the International Association for the Study of Pain (IASP) recommendation is specific for assessment scales used in clinical trials and the FLACC scale and CHEOPS were both recommended for procedural use in this statement (285).

Table 3-6 Expert statements and clinical practice guidelines (CPG).

Association	Scale
American Academy of Pediatrics (326)	FLACC scale
Association of Paediatric Anaesthetists of Great Britain and Ireland (324)	FLACC scale
Australian & New Zealand College of Anaesthetists and Faculty of Pain Medicine (15)	CHEOPS FLACC scale
European Society for Paediatric Anaesthesiology *	Endorsed RCN (UK) guideline
International Association for the Study of Pain (IASP) (285)	FLACC scale CHEOPS
Royal Australian College of Physicians (325)	FLACC scale
Royal College of Nursing (UK) (32)	CHEOPS COMFORT scale FLACC scale UWCH pain scale
Royal College of Paediatrics and Child Health *	Endorsed RCN (UK) guideline

* College or society acknowledges endorsement of alternative guideline

3.3.3 Eligible scales

Review of the scales and their supporting literature resulted in three eligible scales and 33 ineligible scales. The FLACC scale was recommended for use in two systematic reviews of the psychometric properties of pain assessment scales, has been evaluated in four psychometric evaluations studies and used in 23 RCTs and has been recommended in four expert consensus statements. The MBPS has also been evaluated in three psychometric evaluation studies and has been used in 26 RCTs while the VASobs has been used in 66 RCTs and the scale's psychometric performance has been assessed in five studies. The MPBS and the VASobs have not been recommended in a systematic review of pain scale psychometrics or in a consensus guideline providing recommendations for assessment.

Ineligible scales were excluded for either reasons related to the attributes of the scale; most commonly that they use an ordinal rating scale or do not quantify pain using a 0 to 10 metric or a lack of evidence to suggest that the scale is potentially appropriate for procedural pain assessment in infants and children e.g. no psychometric data to support the scale. Additionally, scales were excluded if there was no/insufficient evidence that the scale has been accepted for this purpose,

i.e. used in too few RCTs or they were not recommended in a consensus statement or CPG. See Table 3-7 for a list of eligible scales and Table 3-8 for scales assessed as ineligible and the reasons for their exclusion.

Table 3-7 Eligible scales.

Scale	Design purpose	Age range	Description	Metric	Eligibility criteria - summary
Face, Legs, Activity, Cry & Consolability (FLACC) scale (28)	Postoperative pain	2mth – 7yr	Behavioural observational scale comprised of 5 items: 'facial expression', 'legs', 'activity', 'cry' & 'consolability'	0 - 10	Scale: no exclusion criteria Appropriate: systematic review & psychometric data Accepted: RCT use >5 & consensus statement recommendations
Modified Behavior Pain Scale (MBPS) (288)	Procedural pain	Infants	Behavioural observational scale comprised of 3 items: 'facial expression', 'cry', 'movement'	0 - 10	Scale: no exclusion criteria Appropriate: psychometric data Accepted: RCT use >5
Visual Analogue Scale applied by an observer (VASobs) (118)	Pain	NA	10cm line with anchors at each end 'no pain' &'worst possible pain' – observer places a mark on line for extent of pain demonstrated	0 - 10	Scale: no exclusion criteria Appropriate: psychometric data Accepted: RCT use >5

Abbreviations: RCT – Randomised Controlled Trial

Table 3-8 Scales assessed as ineligible

Scale	Design purpose*	Intended age range*	Scale description	Scale metric	Eligibility criteria - summary
0 - 10 scale (314)	NA	NA	Author defined – no details provided	0 - 10	Scale: no exclusion criteria Appropriate: criteria not met Accepted: criteria not met
2-point scale (315)	NA	NA	Author defined	2 ordinal categories	Scale: metric Appropriate: criteria not met Accepted: criteria not met
3-point scale (301-304)	NA	NA	Author defined	3 ordinal categories	Scale: metric Appropriate: criteria not met Accepted: criteria not met
4-point scale (305-308)	NA	NA	Author defined	4 ordinal categories	Scale: metric Appropriate: criteria not met Accepted: criteria not met
5-point scale (296-299)	NA	NA	Author defined	5 ordinal categories	Scale: metric Appropriate: criteria not met Accepted: criteria not met
7-point scale	NA	NA	Author defined	7 ordinal categories	Scale: metric Appropriate: criteria not met Accepted: criteria not met
Colour Analogue Scale (CASobs) (311) [a]	Pain	5 – 17 years	A vertically orientated analogue scale that uses changing colour to reflect pain intensity	0 - 10	Scale: no exclusion criteria Appropriate: criteria not met Accepted: criteria not met

Child Face Coding System (CFCS) (229)	Postoperative	1 – 5 years	Behavioural scale comprised of 13 facial actions	Intensity & frequency	Scale: metric Appropriate: criteria not met Accepted: criteria not met
Children's Hospital Eastern Ontario Pain Scale (CHEOPS) (29)	Postoperative	1 – 5 years	Behavioural scale comprised of 6 items: comprised of items: facial, cry, child verbal, torso, touch, legs	4 - 13	Scale: metric Appropriate: meets criteria Acceptance: meets criteria
Children & Infants Postoperative Pain Scale (ChIPPS) (167)	Postoperative	0 – 5 years	Behavioural scale comprised of 5 items: facial expression, crying, trunk posture, leg posture, motor restlessness	0 - 10	Scale: no exclusion criteria Appropriate: no English testing Accepted: criteria not met
COMFORT (227)	Critical illness	0 – 18 years	Behavioural scale comprised of 8 items: alertness, calmness, respiratory response, movement, mean arterial pressure, heart rate, muscle tone, facial expression	0 - 30	Scale: metric Appropriate: meets criteria Accepted: criteria not met
Douleur Aigue¨ du Nouveau-ne´ (DAN) (316)	Procedural	Premature and term neonates	Behavioural scale comprised of 3 items: facial expression, limb movements, vocal expressions	0 - 10	Scale: neonatal scale Appropriate: criteria not met Accepted: criteria not met
EVENDOL(289)	Acute pain including procedural pain	0 – 7 years	Behavioural scale comprised of 5 items: complaint, grimace, movements, posture, interaction with surroundings	0 - 15	Scale: metric, Non-English Appropriate: no English testing Accepted: criteria not met

Wong and Baker faces scale applied by an observer (FACESobs) (317) [a]	Postoperative & procedural	3 – 18 years	Analogue scale comprised of 6 faces representing pain expressions	0 - 10	Scale: no exclusion criteria Appropriate: criteria not met Accepted: criteria not met
Facial Grimacing score applied by an observer (318)	Procedural	Neonates	Behavioural scale comprised of 3 facial actions from NFCS: brow bulges, nasolabial furrowing, eye squeeze	Presence & frequency	Scale: metric, neonate Appropriate: criteria not met Accepted: criteria not met
Facial Affective Scale (FASobs) (311) [a]	Illness related	5 – 16 years	Analogue scale comprised of 9 faces representing pain expressions		Scale: no exclusion criteria Appropriate: criteria not met Accepted: criteria not met
Face Pain Scale applied by an observer (FPSobs) (310) [a]	Pain	6 – 9 years	Analogue scale comprised of 7 faces representing pain expressions	0 – 6	Scale: no exclusion criteria Appropriate: criteria not met Accepted: criteria not met
Face pain scale – revised applied by an observer (FPS-Robs) (300) [a]	Procedural	5- 12 years	Analogue scale comprised of 6 faces representing pain expressions	0 – 10	Scale: no exclusion criteria Appropriate: criteria not met Accepted: criteria met
Maximally Discriminant Facial Movement Coding System (MAX) (319)	Pain	Infant	Behavioural scale comprised of facial actions: brow lowering, forehead furrowing, nasal root bulging, eye/nose/cheek, mouth	Presence & frequency	Scale: metric Appropriate: criteria not met Accepted: criteria not met
Modified CHEOPS (mCHEOPS) (320)	Postoperative	1 – 13 years	Behavioural scale comprised of 5 items: cry, facial	0 - 10	Scale: verbal Appropriate: criteria not met Accepted: criteria not met

			expression, verbal response, torso, legs		
mFRANKL (309)	Dental/procedural	18 months & 5 years	Behavioural scale defining response to procedure attempt	5 ordinal categories	Scale: metric Appropriate: criteria not met Accepted: criteria not met
Modified neonatal facial coding system (mNFCS) (321)	Procedural	Neonates	Behavioural scale comprised of 5 items: cry, brow bulge, eye squeeze, nasolabial fold, and open mouth	0 - 5	Scale: metric Appropriate: criteria not met Accepted: criteria not met
Modified observational pain scale (mOPS) (322)	Dental	7 – 12 years	Behavioural scale comprised of 5 items: crying, movement, agitation, positive complaints of pain (inc. localising)	0 - 10	Scale: verbal response required Appropriate: criteria not met Accepted: criteria not met
Modified Riley pain scale (mRIPS) (323)	Postoperative	0 – 3 years inc: CI	Behavioural scale comprised of 3 items: facial expression, body movement and vocal/verbal	0 - 9	Scale: metric Appropriate: criteria not met Accepted: criteria not met
Neonatal facial coding system NFCS (236)	Procedural	Neonates	Behavioural scale comprised of 9 facial action units: brow bulge, eye squeeze, nasolabial furrow, open lip, stretch mouth (horizontal & vertical), taut tongue, chin quiver, lip purse	0 - 9	Scale: metric Appropriate: criteria met Accepted: criteria not met
Neonatal infant pain scales (NIPS) (292)	Procedural	Neonates	Behavioural scale comprised of 6 items: facial expression,	0 - 7	Scale: metric Appropriate: criteria met

			cry, breathing patterns, arms, legs, state of arousal		Accepted: criteria not met
Numeric rating scale (NRSobs) [a] (327)	NA	NA	Author defined	0 – 10	Scale: no exclusion criteria Appropriate: criteria not met Accepted: criteria not met
Observational pain scale (OPS) (312)	Postoperative	1.5 – 12 years	Behavioural scale comprised of 5 items: crying, movement, agitation, positive complaints of pain (inc. localising), blood pressure	0 - 12	Scale: metric Appropriate: criteria not met Accepted: criteria not met
OUCHER applied by an observer (OUCHERobs) (313) [a]	Pain	3 – 7 years	Analogue scale comprised of vertical scale 0 – 100 and 6 photographs of children in pain	0 - 10	Scale: no exclusion criteria Appropriate: criteria not met Accepted: criteria not met
Pain observation scale for young children (POCIS) (294)	Postoperative	1 – 4 years	Behavioural scale comprised of items: facial expression, crying, breathing, torso movements, movement of the arms and legs and restlessness	0 – 7	Scale: metric Appropriate: criteria met Accepted: criteria not met
Premature infant pain scale (PIPP) (218)	Acute	Neonates	Behavioural scale comprised of 7 items: gestational age, behavioural state, HR, SpO2, brow bulge, eye squeeze, nasolabial furrow	0 - 21	Scale: metric Appropriate: criteria not met Accepted: criteria not met

University Wisconsin Children's Hospital Pain Scale (UWCHPS) (295)	Acute pain: operative and procedural	0 – 3 years	Behavioural scale comprised of 5 items: vocal/cry, facial, behavioural, body movement/posture and sleep	0 – 5	Scale: metric Appropriate: criteria met Accepted: criteria not met
Verbal rating scale (VRSobs) (328) [a]	NA	NA	Author defined – 9 faces ranging from happy face to distressed	0 - 1	Scale: metric Appropriate: criteria not met Accepted: criteria not met

[a] Observer application of scale designed for self-report of pain – original scale described in table

Metric – use of a scale other than numeric 0 to 10

Design purpose – designed and used exclusively for neonates or cognitively impaired children

3.4 Discussion

Despite the large numbers of behavioural observation pain scales reported in the literature and the frequency with which infants and children experience procedural pain, we identified very few scales that met our criteria for scales that may be *potentially suitable* to assess procedural pain. Only three scales met our criteria: the FLACC scale, MBPS and VASobs. The remaining scales were excluded: many as the attributes of the scale rendered them unsuitable while others were excluded as there were no studies available that reported psychometric data, or they were not *accepted* as suitable by experts and researchers.

Only one of the three eligible scales, the MBPS, was designed specifically to assess procedural pain (288). Two of the eligible scales, the FLACC scale and MBPS, were comprised of specific items considered indices of pain. Each item is scored based on demonstration of behaviours described for that score level. In contrast, the third scale, the VAS applied by the observer, is a single item scale and quantifies pain intensity based on the observer's overall impression, which may be based on a composite of the behaviours or a single behaviour that the observer considers indicative of pain.

3.4.1 Face, Legs, Activity, Cry, Consolability (FLACC) scale

The FLACC scale was designed to assess postoperative pain in infants aged two months to children aged seven years (28). It is comprised of five behavioural items; 'face', 'legs', 'activity', 'cry' and 'consolability each scored on a scale from 0 to 2 to provide a total score from 0 to 10. This scale has been used extensively and was the most widely recommended of all the scales identified in this review.

Two systematic reviews recommended the FLACC scale for procedural pain (30, 31). Von Baeyer and Spagrud were commissioned by the Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (PedIMMPACT) to complete a review of studies evaluating the psychometric performance of pain scales used to assess pain in children aged three to 18 years (30). This review was published in 2007 and the authors concluded that the FLACC scale and the CHEOPS had been both extensively used and were supported by sufficient evidence of reliability, validity and responsiveness to recommend both scales for use. In their conclusions they pointed out one of the relative advantages of the FLACC scale; a scoring system based on a standard 0 to 10-metric. A second review published the same year repeated this recommendation but with more

reserve. It stated that in the absence of a more suitable scale, one supported by more substantial evidence of the scales' psychometric properties, these scales may be cautiously recommended as the most suitable for procedural pain assessment (31). Methods to evaluate the quality of the studies were not included limiting our capacity to weigh the strength of the data supporting the recommendations made in both reviews.

The FLACC scale was also recommended for procedural pain assessment use in four expert consensus statements (15, 32, 285, 324). The RCN (UK) commissioned an evidence-based review of their existing guideline to provide recommendations regarding the recognition and assessment of acute pain in children (32). In addition, the European Society for Pediatric Anaesthesiology and the Royal College of Paediatrics and Child Health endorsed this document for use by their membership. This document was not described as a systematic review. However, the methods for the search and analysis of the studies identified also approximate those described in the Preferred Reporting Items for Systematic Review (PRISMA) Statement. The RCN went one step further than the two systematic reviews described and provided narrative assessment of the methodological 'shortcomings' of included studies. However, details of the analysis of the data and therefore the basis for their conclusions were limited making it difficult to confirm the strength of their assertions.

One of the most robust of the expert statements, which although also not cited as a systematic review, approximated the rigor associated with the methods described by the (PRISMA) Statement (329). The 4th edition of Acute Pain Management: Scientific Evidence authored by a working group of the Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine made recent recommendations to support use of the FLACC scale for procedural pain assessment (15). The CHEOPS, which we excluded on the basis that it does not use a 0 to 10 metric to score pain intensity, was also recommended in this synthesis of the scientific evidence. The Acute Pain Management: Scientific Evidence is a widely respected and accepted summary of available evidence and the authors note that the 3rd edition was endorsed by the International Association for the Study of Pain (IASP) and by Colleges, Societies and Associations from the United Kingdom, Ireland, Hong Kong, Singapore and Malaysia and recommended by the American Academy of Pain Medicine (15). The specific recommendations for procedural pain assessment reference data summarised in McGrath and colleagues' consensus statement (285) and von Baeyer's systematic review (30) and make no further attempt to describe the psychometric data supporting their recommendation for the FLACC or CHEOPS scale.

The Pediatric Initiative on Methods Measurement and Pain Assessment in Clinical Trials (PedIMMPACT) Statement, commissioned and endorsed by the IASP is the only clear guidance

available for selection of appropriate scales to assess procedural pain in clinical trials. This statement refers us to the recommendations of the 2007 systematic review completed by von Baeyer and Spagrud. However, as has been pointed out this review was limited by the absence of a critical review of the quality of the psychometric studies included in this study (330). Furthermore, these reviews are now ten years old and our search for psychometric evaluation studies found four studies published since publication of this review. For this reason, a contemporary systematic review of the evidence addressing the performance of the FLACC scale to assess procedural pain which includes methods to review the quality of included studies is warranted

3.4.2 Modified Behavioural Pain Scale (MBPS)

The MBPS first appeared in the literature in 1995 is comprised of three behaviours; face, cry and movement and each is scored on a subscale based on the demonstration of described behaviours which are added to provide a final score out of ten (288). Our search of the literature identified three psychometric evaluation studies addressing the performance of the MBPS used to assess procedural pain and 26 RCTs using the scale to measure procedural pain in infants and children and therefore meets our criteria for scales potentially suitable to assess procedural pain in infants and children.

The scale was uniquely designed to assess procedural pain in young infants and was initially to assess immunisation related pain (288). However, it has escaped the attention of both systematic reviews published in 2007 and all published consensus statements. The focus of most of the work since the original testing has been limited to immunisation related pain, which may explain why this scale has not attracted the level of attention that a scale designed for procedural pain might. To better understand the psychometric performance of the MBPS a systematic review should be completed to identify, assess and summarise the psychometric studies evaluating the performance of this scale, specifically for procedural pain assessment. This will serve as a platform for recommendations regarding MBPS clinical and research use.

3.4.3 Visual Analogue Scale observer

The VAS applied by an observer met our criteria for scales considered *potentially suitable*. Five psychometric evaluation studies to establish the measurement properties of the VASobs or to compare performance with another scale were identified by our search. Furthermore, the VASobs was the most frequently used scale to measure procedural pain in the RCTs located for this review.

A total of 64 RCTs used the scale to assess procedural pain and it is likely that this reflects the practical advantages of the VASobs which is quick and simple to apply. Observers are asked to make a mark on an unmarked 10cm line between the anchors at each end 'no pain' and the 'worst pain' to indicate their impression of the infant or child's pain. Unlike many other scales the observers require little if any training and can apply the scale and record a score within seconds. These practical advantages may also encourage its use in clinical practice where pain assessment is needed, although it is not clear whether this is the case and to what extent it is used in clinical practice.

The VASobs was not recommended in either systematic reviews or expert consensus statements or clinical practice guidelines which belies its extensive use in RCTs to measure procedural pain. It is likely that the VASobs has been overlooked in these reviews for similar reasons to those presented by van Dijk and colleagues in their 2002 review of available data addressing the reliability and validity of the VAS used as an observational pain scale (290). The authors concluded that the results supporting the psychometric properties of the VASobs were limited and that further testing was required before the scale could be accepted as suitable for assessing paediatric pain. Only one study cited in this review examined the scale's performance used to assess procedural pain (331). The authors called for further testing in circumstances where the child is unable to self-report, using study designs able to test the scales sensitivity to pain. Considering the favour shown to the VASobs and the concerns of van Dijk in 2002, it is critical that clinicians and researchers are provided with a clear understanding of the evidence supporting this choice of scale for clinical and research purposes.

3.4.4 Limitations

There are several limitations to this review that increase the likelihood that scales that may be reliable and valid when used to assess procedural pain in infants and children were not identified in this review. The criteria used to select eligible scales were carefully considered and were based on an assumption that scales that were designed for procedural pain assessment, had undergone psychometric testing and were accepted as appropriate for procedural pain are likely to be suitable for procedural pain assessment. Criteria also limited scales to ones using a 0 to 10 metric as it is widely accepted that use of a standard metric will improve interpretation of scores and treatment decisions. However, as all scales can be converted mathematically to a 0 to 10 metric, it could be argued that this was an unnecessary restriction. Particularly, as it is conceivable that scales that may be potentially reliable and valid for procedural pain assessment may have been excluded by these criteria. Scales were also limited to those available and tested in English. As has been said,

scales assessed as valid in one language cannot be assumed to be valid in another. However, exclusion based on language may also have resulted in the exclusion of a *potentially suitable* scale. The design of the search was aimed at identifying all relevant publications. However, as this review did not apply the methodological rigour of a systematic review it is possible that the search did not identify all studies and therefore scales that may have met inclusion criteria.

3.4.5 Recommendations and future directions

This review aimed to identify pain scales that we may consider *potentially suitable*. However, it has not identified scales that can be conclusively considered valid for use to assess procedural pain in infants or children for clinical or research purposes. The three scales identified show promise as viable options for this purpose. To draw firm conclusions about each of these scales and make recommendations regarding their use, the psychometric (reliability and validity) and practical (feasibility) properties and clinical utility of each should be carefully assessed in systematic reviews of all available psychometric data. The review design should employ search strategies which minimise the likelihood that eligible data is overlooked, is based on the principles of the PRISMA statement and includes methods to assess the quality of the included studies. This is the focus of the next phase of this project.

3.5 Addendum: New literature

The searches for this phase were completed in June 2014 and included studies and guidelines published before this date. We recently repeated these searches to identify relevant literature published more recently that might alter the conclusions drawn from this search. This search was conducted in the first week of May 2018 and includes studies and guidelines published between June 2014 and May 2018. We used the search terms and inclusion and exclusion criteria used in our original review to identify scales that might now meet our criteria for scales considered *potentially suitable* for procedural pain assessment use.

The recent search did not identify new scales designed for procedural pain assessment or more recent systematic reviews summarising the psychometric properties of existing scales, other than those that arose from the next phase of this project (330, 332). Only two studies were completed to evaluate the psychometric properties of a scale used to assess procedural pain in infants or children, the focus of both was the FLACC scale (333, 334). Furthermore, there were no new consensus statements or changes to existing recommendations from professional associations, academies, societies or organisations regarding the most suitable observational scale for procedural pain assessment use.

Fifty-seven RCTs that measured procedural pain using an observation pain scale were identified and 24 (40%) of them used the FLACC scale, seven used the MBPS and six used the VASobs to measure a study outcome. The Neonatal Infant Pain Scale (NIPS) was used in eight trials. However, the NIPS is scored on a scale of 0 to 7 and was therefore excluded. The remaining RCTs used nine different scales, some of which were not identified in the original search, such as, the Analgesia Nociceptive Index (ANI) and the Sound, Eye and Motor (SEM) scale. The Faces Pain Scale-Revised applied by an observer (FPS-Robs) was used in one RCT since the original search, taking the total number of RCTs to five.

3.5.1 Implications

These results do not alter our assessment of the eligibility of the FLACC scale, the MBPS or the VASobs. However, notably there was a reduction in the frequency of the use of the VASobs in RCTs compared with the FLACC scale. The Faces Pain Scale-Revised applied by an observer (FPS-Robs) would now meet the pre-defined criteria to be considered *potentially suitable* to assess procedural pain in infants and children.

SECTION 3.

This section includes four chapters: the first of which reports the methods used and the following three chapters report the results of three systematic reviews conducted to summarise the existing evidence addressing the psychometric properties of the Face, Legs, Activity, Cry and Consolability (FLACC) scale, the Modified Behavioural Pain Scale (MBPS) and the Visual Analogue Scale applied by an observer (VASobs). Chapters 6 and 7 are presented as published versions of the FLACC scale and MBPS reviews.

CHAPTER 4.

This chapter reports the methods used for the three systematic reviews conducted to summarise the evidence addressing the psychometric properties of the Face, Legs, Activity, Cry and Consolability (FLACC) scale, the Modified Behavioural Pain Scale (MBPS) and the Visual Analogue Scale applied by an observer (VASobs). These scales were selected from the many (over 40 scales) that appear in the literature (212, 286) as they met the criteria developed to identify scales that may be *potentially suitable* for procedural pain assessment.

4.1 Methods

Systematic reviews were conducted to identify and appraise the evidence for the psychometric properties, the clinical utility and the feasibility for use of the FLACC, MBPS and VASobs. A protocol, which was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA statement) (329) was developed by me with the support of my supervisors for this purpose. The protocol was registered separately for each review with the International Prospective Register of Systematic Reviews (FLACC: CRD42014014296, MBPS: CRD42016041722 and VASobs: CRD42016041724). They are available in full text on the PROSPERO Web Site (335-337).

4.1.1 Inclusion/exclusion criteria

Studies reporting reliability, validity, clinical utility or feasibility data for the FLACC scale, the MBPS and the VASobs applied to infants and children were included in these reviews. This included studies where the aim of the study was to examine the psychometrics of the scale, compare the scale with alternative scales or assessment tools or establish the psychometrics of an alternative scale or assessment tool using one of these scales as a reference. The review also included RCTs using the FLACC scale, MBPS or VASobs as an outcome measure as trial methods are a method for construct validation (i.e.: the capacity of the scale to detect a difference between known or extreme groups). Infants and children were defined as participants aged from birth to 18 years.

Studies that did not report or analyse FLACC, MBPS or VASobs scores separately, did not include infants or children or report their data separately, were only published as an abstract or

were not available in English were excluded from this review. Finally, studies using or evaluating a modified version of any of these scales (including translated versions) were also excluded. Following quality assessment, RCTs with Jadad scores (described in section 4.1.5) less than three were excluded as they were considered at risk of significant bias and therefore unlikely to contribute significant evidence to this review (338).

4.1.2 Search strategy

Relevant search terms were used to search the following databases: MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews and Cochrane Controlled Trials, Cumulative Index Nursing and Allied Health Literature (CINAHL) and PsycINFO using the Ovid, PubMed and EBSCOhost platforms. Google Scholar was also searched using the same search terms and the reference lists of the included studies and identified reviews were also searched. The search was limited to texts available in English. The terms used and the date ranges used for searches are reported in Table 4-1.

Table 4-1 Search terms and search dates used for the FLACC scale, MBS and VASobs systematic reviews.

Review	Search terms	Search date
Face, Legs, Activity, Cry Consolability (FLACC)	FLACC scale OR FLACC OR Face Legs Activity Cry Consolability AND infant OR child	31st August 2014 *
Modified Behavioral Pain Scale (MBPS)	MBPS OR Modified Behavioural Pain Scale AND infant OR child	31st July 2016 *
Visual Analogue Scale applied by an observer (VASobs)	VAS OR VAS observer OR Visual Analogue Scale AND infant OR child	31st July 2016

* Searches were re-run prior to submission of manuscripts for publication: FLACC – 21 May 2015, MBPS – 16 August 2017.

4.1.3 Study selection

Following removal of duplicates, abstracts were reviewed by two independent reviewers (DC and one of NS, FB, or DH). Where eligibility was ambiguous the full text article was reviewed. A third reviewer was used to reach consensus where study eligibility remained unclear.

4.1.4 Data extraction

Data extraction was completed by one reviewer (DC) and recorded on one of two extraction tools: for the psychometric evaluation studies, a modification of the QAREL data extraction form (339) designed for appraisals of diagnostic reliability studies was used and for the RCTs, a modification of the Cochrane Collaboration data collection tool designed for intervention studies (340) was used. The modifications of these forms included the deletion of irrelevant fields and the addition of fields to capture relevant methods and results not included in the original form.

The data extracted included; participant details (e.g. numbers, demographics), setting and circumstances of the pain being measured (e.g. associated with disease, operative or procedural), scale description (e.g. items and modifications), study methods (design, psychometric properties evaluated, and statistical methods), sources of bias and study results.

Data was extracted by one reviewer (DC) and checked and confirmed by a second reviewer (FB, DH or NS). A third reviewer completed data extraction independently to resolve any inconsistencies between the first two reviewers.

4.1.5 Quality assessment

One of two quality assessment tools were applied independently by two reviewers (DC and one of NS, FB, or DH) and a third where agreement was not achieved by the first two reviewers. The methodological quality of the psychometric evaluation studies was appraised using the COSMIN checklist (341) and the RCTs were appraised using the Jadad score (338). The COSMIN checklist was also used to assess the methods of an RCT where other psychometric properties, such as reliability and responsiveness, were assessed in the trial.

The COSMIN checklist was developed to assess the quality of studies focused on health related patient-reported outcome measures and provides standards for study design, statistical methods and acceptable outcome values (342). The checklist is also considered suitable for clinical rating

scales that measure constructs not self-reported but not directly measurable and has been previously used in a systematic review examining the psychometric properties of an observer applied fatigue assessment scale for children (343).

The checklist is comprised of four steps:

1. Identification of the measurement properties (see Table 4-3 for measurement properties),
2. Assessment of the item response theory methods applied,
3. Evaluation of the quality of the methods used to assess the measurement properties identified in step one and
4. Assessment of the generalisability of the results.

The measurement properties: internal consistency, reliability, measurement error, content validity, construct validity (structural validity, hypothesis testing, cross-cultural validity), criterion validity and responsiveness are assessed against a series of criteria. Each criterion is scored on a 4-point scale (‘poor’, ‘fair’, ‘good’ or ‘excellent’) depending on the standard met. The lowest item rating forms the final assessment for the methods used to assess that property. A study may receive different assessments for the methods addressing different psychometric properties. The COSMIN taxonomy and the terms commonly used to describe the methods in pain scale evaluations studies are defined in Table 4-2. The checklist is included in full in Appendix B and a full description of the COSMIN checklist and application of scoring system can be found on the COSMIN website (http://www.cosmin.nl/the_cosmin_checklist.html).

The Jadad Scale (Table 4-3) for assessing the quality of RCTs focuses on randomisation, blinding and participant follow-up and has been used in a previous pain scale reviews to assess the quality of RCTs (220). Each item contributes to a total score out of five, where five is a perfect score. High scoring studies will provide higher levels of evidence of the scales capacity to distinguish between known groups than low scoring studies. For these reviews we made a minor modification to the definition for participant follow-up and scored this as acceptable if, in the absence of an explicit statement 'there were no withdrawals from this study', there was sufficient detail in the results to account for all study participants.

The intention was to report the feasibility and the clinical utility of scale application in clinical practice in these reviews. A tool to support the assessment of the quality of the methods used to evaluate the practicality or feasibility or the clinical utility of the scale was not identified. For this reason, a pragmatic and descriptive approach to assessing these studies was adopted.

Table 4-2 Pain scale validation strategies and COSMIN taxonomy (342).

COSMIN measurement property	Pain scale measurement property	Pain scale evaluation study method
Internal consistency	Internal consistency	Correlations between items on the scale
Reliability	Inter-rater reliability	Correlation between pain scores provided simultaneously by independent reviewers
	Intra-rater reliability	Correlation between scores allocated by single reviewer to the same episode of pain on separate occasions (achieved using video-recorded footage)
Measurement error		Rarely tested in pain scale evaluation studies
Content validity (including face)	Content validity	Relevance & comprehensiveness of the items assessed by experts
Structural validity		Principal component analysis (used for new scale development)
Hypothesis testing	Convergent validity	Correlation with assessments using other pain assessment tools/scales - observational scale
	Discriminant validity	Correlation with other unrelated constructs (e.g.: pain & hunger)
	Construct validity	Extreme or known groups comparison - correlation between different procedure/treatment groups
Cross cultural validity	Cross cultural validity	Translation – backwards & forwards, content review for cultural appropriateness
Criterion validity	Concurrent validity	Correlation with assessments using the gold standard (other valid tools/scales & self-report)
Responsiveness	Responsiveness	Change over time where change expected e.g.: before & after analgesic or pain producing procedure
Interpretability	Clinical utility	Impact of score on clinical decision making

Table 4-3 The Jadad scale (338)

Question	Response Yes / no
1. Was the study described as randomized (this includes the use of words such as randomly, random and randomisation	
2. Was the study described as double blind?	
3. Was there a description of withdrawals and dropouts?	

Scoring the items:

Either give a score of 1 point for each ”yes” or 0 points for each “no.”

Give 1 additional point each if:

For question 1, the method to generate the sequence of randomization was described **and** it was appropriate (table of random numbers, computer generated, etc.)

If for question 2 the method of double blinding was described, and it was appropriate (identical placebo, active placebo, dummy, etc.)

Deduct 1 point each if:

For question 1, the method to generate the sequence of randomization was described and it was inappropriate (patients allocated alternately, or according to date of birth, hospital number, etc.)

For question 2, the study was described as double blind, but the method of blinding was inappropriate (e.g. comparison of tablet vs. injection with no double dummy)

4.1.6 Data synthesis

The results of the search and study selection were described using the PRISMA flow chart (329). Studies using different designs were included in this review, therefore pooling of data for meta-analysis was not considered possible. A narrative synthesis of the evidence provided by each study was therefore used to address each of the study outcomes. It was also anticipated that eligible studies would apply the scale to different populations (e.g. age groups) and under different circumstances (e.g. postoperative, procedural and illness/injury related) to those for which the scale was developed and originally tested. These studies were reviewed separately to the studies concentrating on the original population and circumstances and the results summarised for these cohorts (in most cases defined by age and the circumstances of pain). As it was not possible to identify a population for which the VASobs was originally designed, studies eligible for this review were grouped based on age ranges selected to reflect potential differences in behavioural responses to pain which may in turn influence the observer’s impression of the pain experienced

by the infant or child and therefore their application of the VASobs. The ranges were: neonates, infants and toddlers aged 1 month to 3 years, young children aged 3 to 12 years and older children (over 12 years). Subgroups were also created based on the circumstances of the pain (procedural, postoperative, illness/injury related) as it is accepted that these circumstances may impact on the infant of child's experience and the behaviours demonstrated e.g. procedures are likely to be associated with significant fear related distress, the behaviours of which may mimic pain-related behaviours (263).

The weight of evidence derived from each study was estimated based on the strength of the results and the quality of the methods used in the study. Comparison between groups of other pain related variables measured in the RCTs also contributed to this assessment. Where these results were consistent with the results of the between groups comparison of the FLACC, MBPS or VASobs scores they helped to support the strength of the contribution of these results to our understanding of the validity of the scale.

A narrative synthesis of the results from the studies included in this review was completed as the heterogeneity of the eligible studies meant that pooling of data for meta-analysis was not possible. The results were summarised for the various subgroups created based on age and circumstances. The evaluation criteria for IMMPACT reviews (344) (Table 4-4) provided a framework to define the strength of the evidence base supporting the psychometrics of a scale or measure based on the number, results and independence of the evaluation studies. Assessment of study quality is not included in the criteria which is a significant limitation to this framework. However, the principles used to underpin these assessment criteria are used here to guide our synthesis.

Table 4-4 IMMPACT evaluation criteria for the level of evidence supporting the psychometrics properties of a scale. (Taken from Cohen and colleagues (344))

Standard	Criteria
Well established	The measure must have been presented in at least two peer-reviewed articles by different investigators or investigatory teams. Sufficient detail about the measure to allow critical evaluation & replication (e.g., measure & manual provided or available upon request). Detailed (e.g., statistics presented) information indicating good validity & reliability in at least one peer-reviewed article.
Approaching well established	The measure must have been presented in at least two peer-reviewed articles, which might be by the same investigator or investigatory team. Sufficient detail about the measure to allow critical evaluation & replication (e.g., measure & manual provided or available upon request). Validity & reliability information presented in either vague terms (e.g., no statistics presented) or moderate values.
Promising assessment	The measure must have been presented in at least one peer-reviewed article. Sufficient detail about the measure to allow critical evaluation & replication (e.g., measure & manual provided or available upon request). Validity & reliability information presented in either vague terms (e.g., no statistics presented) or moderate values.

4.2 Addendum

The systematic reviews conducted for this phase of the project were conducted between 2014 and 2016. The searches and analysis of the results for each scale was repeated in May 2018 to identify data published since the original review, specifically data that may impact on the original conclusions drawn. The same search terms and inclusion and exclusion criteria were used to identify eligible studies and these studies were reviewed to determine whether they were likely to alter the conclusions drawn based on the results of the original review. To make a broader contribution to our understanding of the psychometric properties of the scales the original reviews were not restricted to procedural pain assessment. However, the recent searches and reviews were restricted to studies (psychometric evaluation studies and RCTs) that focused on the scale used to assess procedural pain in infants and children.

CHAPTER 5.

A systematic review of the psychometric properties of the observer applied Visual Analogue Scale (VASobs).

The results of an unpublished systematic review to summarise the data that describes the psychometric properties of the Visual Analogue Scale applied by an observer (VASobs) to assess pain in infants and children are reported in this chapter. As this work is currently unpublished it is presented here formatted for the thesis. The chapter concludes with a summary of relevant data published since the original review and their impact on the assessment of the psychometric properties of the VASobs.

Abstract

Introduction: The visual analogue scale has widely been used by observers to assess pain in infants and children unable to self-report pain (VASobs). However, there are concerns about the reliability and validity of the scale when applied by observers. The aim of this systematic review was to provide a current summary of the psychometric properties of the VASobs.

Methods: Databases searched were MEDLINE, CINAHL, EMbase, PsycINFO, The Cochrane Database of Systematic Reviews and Cochrane Controlled Trials and Google Scholar. Studies examining the psychometric properties of the VASobs and RCTs using the VASobs as an outcome measure for participants aged from birth to 18 years were eligible for inclusion. Quality assessment of the included studies was completed using the COSMIN Checklist and the Jadad Scale.

Results: Thirty-two psychometric evaluation studies and 65 RCTs were included. The study population, circumstances and the study quality varied greatly. There was promising but not conclusive data to support the psychometrics of the scale for assessing immunisation related pain, less convincing data supporting the psychometrics for assessing other procedures and based on contradictory findings and limited data, very limited evidence to support the psychometrics of the scale used to assess postoperative pain.

Discussion: The conclusions drawn from this review are unchanged from those of a similar review published 15 years ago. The VASobs although widely used cannot be unequivocally recommended for assessing the pain experienced by infants and children.

5.1 Introduction

Graphical tools, such as the Visual Analogue Scale (VAS), were designed to quantify the extent of a range of subjective experiences where individuals had difficulty capturing the experience in words, specifically to articulate symptom intensity precisely (345). Use of the VAS to support self-assessment of pain intensity is supported by early work reporting the validity of the scale when used for this purpose (118, 345-348). The ease with which it could be applied and the strength of the evidence supporting validity of self-reported VAS scores served as the platform for use of this scale applied by observers to assess the intensity of pain experienced by individuals unable to use the scale to self-report, such as infants and young children (290).

The VAS is a line measuring 10 centimetres (cm) with verbal anchors at each end; most commonly ‘no pain’ and ‘worst possible pain’ which correspond to a score of ‘0’ and ‘10’ respectively (118, 345). The line is otherwise unmarked and patients or a suitable observer are asked to place a mark on the line to indicate the pain they (or the patient) are currently experiencing. The distance from the zero mark is measured and this is considered their pain score.

In 2002, van Dijk and colleagues published a review of the evidence to establish the usefulness of the VAS applied by an observer for paediatric pain assessment (290). They drew our attention to significant gaps in the evidence supporting observer applied VAS in this population. The results for inter-rater reliability were most often reported as Pearson correlation coefficients which gave us no understanding of the level of agreement between scores and only the extent to which they correlate. It is therefore theoretically possible for scores to differ substantially but correlate strongly if the observer scores change in the same direction from one observation to the next, i.e. both observers scores increase (or decrease) from one observation to the next. Intra-rater reliability had at the time of their review not been tested. Validity testing was limited to demonstrating the extent to which scores correlated with scores derived from application of an alternative pain scale or with self-reported scores and results were mixed. In summary, the authors concluded that there was insufficient evidence to support the reliability and validity of the VAS used by observers to quantify pain intensity in infants and children and that further psychometric testing was needed.

In light of van Dijk and colleagues conclusions 15 years ago (290) and continued use of the VASobs to measure study outcomes, it is reasonable to undertake a current review to determine whether there is now sufficient high-quality data to support use of this scale to assess pain in infants and children.

5.1.1 Review aim

The aim of this systematic review was to provide a current summary of the psychometric and practical properties of the VASobs to assess pain in infants and children. The specific objectives were to 1) identify and describe the paediatric populations and circumstances for which psychometric data is available, 2) systematically review the quality of this data, 3) analyse and summarise the strength of evidence that support the psychometric properties, clinical utility and feasibility of the VAS and provide recommendations to guide clinical and research use of this scale to assess pain in infants and children.

5.2 Methods

The methods for this study are described in the protocol which is provided in Chapter 4.

5.3 Results

The search for this review was completed in Aug 2016 and a total of 97 studies (32 psychometric evaluation studies and 65 RCTs) were eligible for this review. The results of the search and eligibility screening are shown in Figure 5-1.

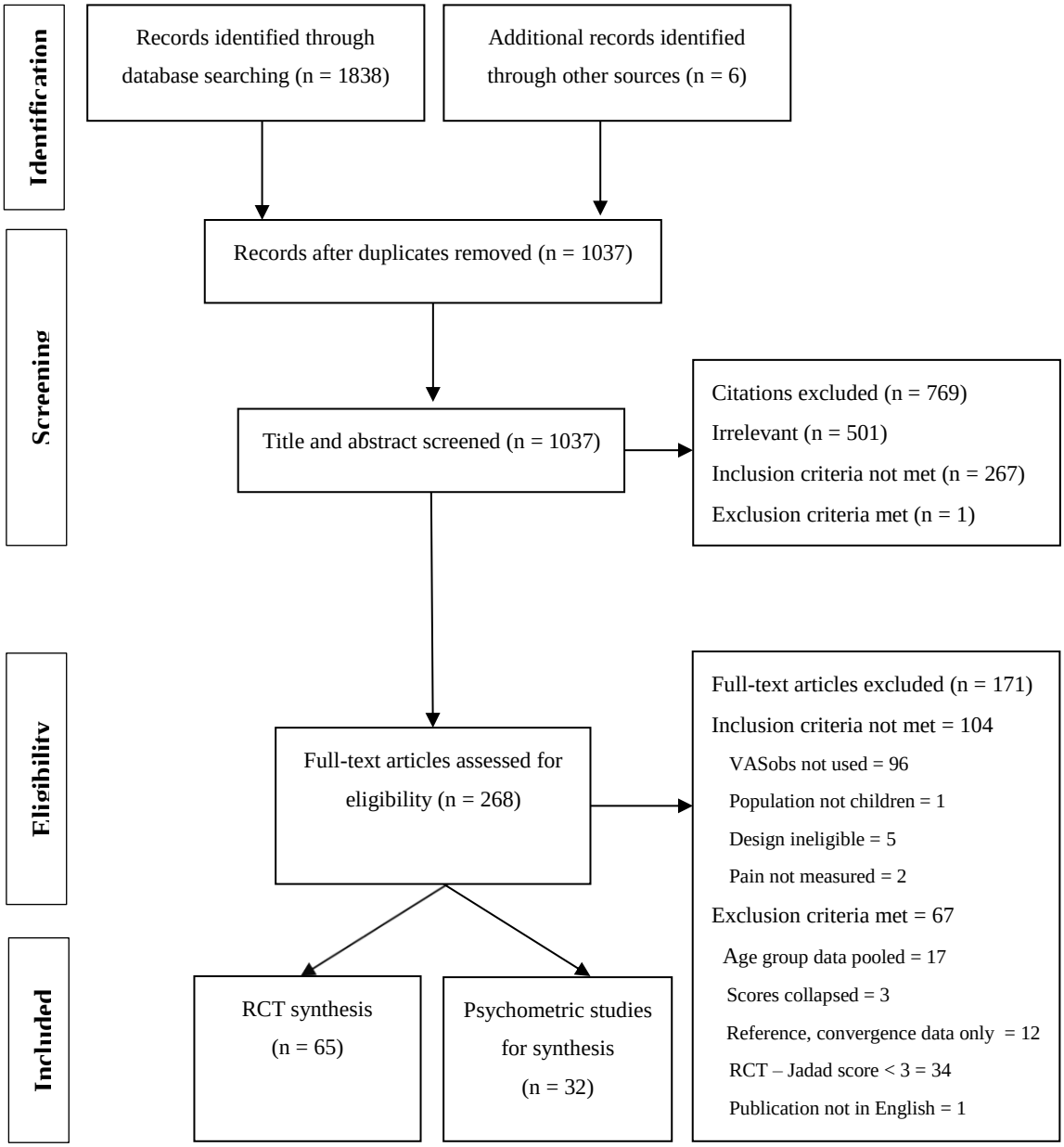


Figure 5-1 Flow chart detailing the search and study screening results

Abbrev. VAS – Visual Analogue Scale, RCT – Randomised Controlled Trial

5.3.1 Study and patient characteristics

5.3.1.1 Psychometric evaluation studies

Thirty-two studies, summarised by psychometric properties in Tables 5-3 to 5-6, included data addressing the psychometric properties of the VASobs. However, almost half (n = 13, 40.6%) used the VASobs as a reference scale in a study designed to assess the psychometric properties of another scale. These studies were eligible as they either included analysis of the performance of the VASobs other than correlation with a new scale or the index scale had undergone previous psychometric testing and was reported as valid. The results from correlations between the index scale and the VASobs for two eligible studies was excluded as the index scale was a newly developed scale which had not been tested and therefore could not be assumed to be valid (292, 349). The scale was applied by parents of infants and/or children (n = 22), nurses (n = 11), physicians (n = 5), researchers (n = 3) and other observers (n=3) for infants and children experiencing postoperative (n = 12), procedural (n = 11) and disease or injury related (n = 9) pain. The age of participants in these studies ranged from newborn to 18 years old. A more comprehensive summary of each study can be found in Appendix C, Table 1.

5.3.1.2 Randomised Controlled Trials

Table 5-1 lists the 65 RCTs included in this review. The study populations ranged in age from newborn to 18 years old and included infants and children experiencing procedural (n = 50), postoperative (n = 11) and disease/injury related (n = 4) pain. The VASobs was applied most often by parents (n = 49), then nurses (n = 27), physicians (n = 9), investigators (n = 11) and finally by other observers (n = 6). These studies are summarised in more detail in Appendix C, Table 2.

Table 5-1 Summary of the RCTs using VASobs to measure a study outcome.

Study	Subject	Circumstances / Setting	Quality score	Psychometric properties: results – strength of evidence	Evidence strength
Procedural					
Abuelkheir et al, 2014 (350)	2 months to 6 years	Procedural (immunisation) Well baby paediatric clinic (Saudi Arabia)	5	Hypothesis testing: difference in VASobs scores (nurse) consistent with independent related variables (MBPS scores, crying time) & dependent related variable (cry). Responsiveness: scores increase, no significance testing.	Moderate
Babl et al, 2009 (351)	1 – 5 years	Procedural (nasogastric tube insertion) ED (Australia)	5	Hypothesis testing: difference in VASobs scores (nurse, parent, observer), inconsistent with independent related variable (FLACC score).	Very low
Balan et al, 2009 (352)	5 – 12 years	Procedural (venepuncture) Inpatient department tertiary care centre (India)	3	Hypothesis testing: consistent differences in VASobs scores (parent, nurse, & investigator).	Very low
Barkan et al, 2014 (353)	1 – 10 years	Procedural (laceration repair) Paediatric ED (Israel)	5	Hypothesis testing: no difference VASobs scores (parent, investigator), difference in independent unrelated variable (sedation score). Study agent sedative may be evidence of discrimination	Nil
Bhatnagar et al, 2008 (354)	1 – 10 years	Procedural (lumbar puncture) Cancer hospital (India)	3	Hypothesis testing: no difference VASobs (investigator), difference in independent unrelated variable (sedation score). Study agent sedative may be evidence of discrimination	Nil
Bishai et al, 1999 (355)	5 – 16 years	Procedural (IV port access) Setting unstated	5	Hypothesis testing: no difference in VASobs scores (parents, nurse), consistent with independent related variable (self-report)	Low

Chapman et al, 2011 (356)	0 – 17 years	Procedural (IV catheter insertion) Paediatric ED (US)	3	Hypothesis testing: No difference in VASobs scores (parents, nurse), consistent with independent related variable (self-report) in 8 – 17y	Very low
Cignacco et al; 2008 (357)	neonates 24 – 37 weeks PMA	Procedural (endotracheal suctioning) NICU (Switzerland)	3	Hypothesis testing: No difference in VASobs scores, consistent with dependent related variables (PIPP, Bernese Pain Scale scores*)	Nil
Cohen et al, 2006 (358)	1 – 24 months	Procedural (immunization) University medical centre & private practice Office. (US)	3	Hypothesis testing: No difference in VASobs scores (parents, nurse), inconsistent with independent related variable	Nil
Cohen et al, 2009 (359)	4 – 6 years	Procedural (immunization) University outpatient primary care clinic (US)	3	Hypothesis testing: No difference in VASobs scores (caregiver, nurse), consistent with independent related variable (self-report), Responsiveness shown (COSMIN poor)	Very low Very low
Di Liddo et al, 2006 (360)	2 – 18 years	Procedural (fracture reduction) ED/orthopaedic clinic (Canada)	5	Hypothesis testing: no difference in VASobs score, no independent/dependent variables to confirm result	Nil
Dulai et al, 2016 (361)	3 – 16 years	Procedural (percutaneous pin removal) Orthopaedic dept (Canada)	5	Hypothesis testing: no difference in VASobs scores consistent with independent related variable (self-report) Responsiveness demonstrated (COSMIN poor)	Low Very low
Fatovich et al, 1999 (362)	1 – 10 years	Procedural (LA infiltration) ED (Australia)	5	Hypothesis testing: No difference in VASobs scores (parent), consistent with independent related variables (PPAT scores, facial expression, restraint use, self-report in adult subset)	Low
Ha et al, 2013 (184)	3 – 10 years	Procedural (laceration repair) ED (Korea)	3	Hypothesis testing: Difference in VASobs scores (parent), consistent with independent related variable (PBCL scores) & inconsistent with independent related variable (self-report)	Nil

Harrison et al, 2014	12 – 18 months 3 – 5 years	Procedural (immunisation) Immunisation centre (Australia)	3	Hypothesis testing: No difference in VASobs scores (parent) or independent related variable (FLACC)	Low
Heden et al, 2009 (363)	1 – 18 years	Procedural (IV port access) Paed oncology / haematology setting (Sweden)	5	Hypothesis testing: no difference in VASobs scores (parent, nurse), consistent with independent related variable (self-report), inconsistent with independent related variable (CHEOPS score) Convergence with fear & distress shown (COSMIN - poor)	Very low Very low
Heden et al, 2009 (364)	2 - 7 years	Procedural (IV port access) Paed oncology/ haematology setting (Sweden)	3	Hypothesis testing: no differences in VASobs scores (parent, nurse), difference in parent VASobs distress & VASobs fear scores	Nil
Heden et al, 2011 (365)	1 – 18 years	Procedural (IV port access) Paed oncology/haematology setting (Sweden)	5	Hypothesis testing: no difference in VASobs scores (parent, nurse), consistent with independent related variables (self-report, CHEOPS, PBCL scores)	Low
Heden et al, 2014 (366)	1 – 18 years	Procedural (IV port access) Paed oncology/haematology setting (Sweden)	5	Hypothesis testing: No difference in VASobs scores (parent, nurse), consistent with independent related variables (self-report, CHEOPS scores, cortisol levels), difference in self-reported distress	Low
Hogan et al, 2014 (367)	4 to 6 months	Procedural (immunisation) Primary care practice (Canada)	3	Hypothesis testing: no difference in VASobs scores, (parent) consistent with independent related variable (MBPS scores)	Nil
Hopkins et al, 1988 (368)	1 – 5 years	Procedural (IV catheter insertion) Day surgery unit (England)	3	Hypothesis testing: difference in VASobs scores (OR assistant), consistent with dependent related variable (VRS)	Very low
Horn et al, 1999 (369)	4 – 6 years	Procedural (immunisation) Private paed office (US)	3	Hypothesis testing: no difference in VASobs scores (parents), consistent with independent related variables (self-report distress, OSBD-R scores)	Very low

Hua et al, 2015(370)	4 – 16 years	Procedural (dressing change) Paed centre of tertiary hospital (China)	3	Hypothesis testing: difference in VASobs scores (caregivers), consistent with independent related variables (self-report, FLACC)	Low
Ipp et al, 2004 (371)	12 months	Procedural (immunisation) Community paediatricians clinic (Canada)	5	Hypothesis testing: difference in VASobs scores (parent & paed), consistent with independent related variable (MBPS), Responsiveness shown (COSMIN poor)	Low Very low
Ipp et al, 2006 (372)	4 – 6 years	Procedural (immunisation) Urban primary care paed practice (Canada)	5	Hypothesis testing: no difference in VASobs scores (parent, physician), inconsistent with independent related variables (self-report, cry related)	Nil (very lownegative)
Ipp et al, 2007 (373)	4 to 6 months	Procedural (immunisation) Primary care Practice (Canada)	3	Hypothesis testing: Difference in VASobs scores (parent & paediatrician) consistent with independent related variables (MBPS scores, cry related)	Low
Ipp et al, 2009 (374)	2 to 6months	Procedural (immunisation) Paed community practice (Canada)	5	Hypothesis testing: Difference in VASobs scores (parent, paediatrician), consistent with independent related variable (MBPS scores)	Low
Knutsson et al 2006 (375)	18 – 24 months	Procedural (immunisation) Child health centre (Sweden)	5	Hypothesis testing: Difference in VASobs scores (parent), consistent with independent related variable (CHEOPS score)	Low
Kozer et al, 2006 (376)	0 – 2 months	Procedural (urine collection) University hospital (Israel)	3	Hypothesis testing: Difference in VASobs scores (parent, nurse), consistent with independent related variable (DAN scores)	Very low
Lee-Jayaram et al, 2010 (377)	5 – 17 years	Procedural (fracture manipulation) ED (US)	3	Hypothesis testing: Difference in VASobs scores (parent), consistent with independent related variable (OSBD-r score)	Very low

Lindh et al, 2003 (378)	3 months	Procedural (immunisation) Paed Outpatient (Sweden)	5	Hypothesis testing: Difference in VASobs scores (parent, nurse), consistent with independent related variables (MBPS scores, HR changes, cry related variables).	Moderate
Luhmann et al, 2004 (379)	4 – 17 years	Procedural (IV catheter insertion) ED (US)	3	Hypothesis testing: No difference in VASobs scores (parent, nurse), consistent with independent related variable (self-report)	Very low
Luhmann et al, 2006 (380)	5 – 17 years	Procedure (fracture reduction) ED (US)	3	Hypothesis testing: Difference in VASobs scores (parent), consistent with independent related variable (self-report, PBC)	Very low
Marec-Berard et al, 2009 (381)	2 – 18 years	Procedure (lumbar puncture) Several oncology centres France)	3	Hypothesis testing: No difference in VASobs scores (parent), consistent with independent related variable (self-report, procedure success)	Very low
McErlean et al, 2003 (382)	9month – 6years	Procedural (IV catheter insertion) ED (US)	3	Hypothesis testing: Difference in VASobs (parent), consistent with difference in independent related variable but not significant (VASobs (observer)).	Very low
McGowen et al, 2013 (383)	2 – 6 months	Procedural (immunisation) Immunisation clinic (Wales)	3	Hypothesis testing: difference (not significant) in VASobs (parent), independent related variable significant difference	Nil to very low
Miller et al, 2011 (384)	3 – 10 years	Procedural (burns dressing procedure) Burns outpatient clinic (Australia)	3	Hypothesis testing: Difference in VASobs (parent) scores. Consistent with independent related variables (self-report & FLACC scores)	Moderate
Newbury et al, 2009 (385)	3 months	Procedural (IV catheter insertion) ED (New Zealand)	3	Hypothesis testing: No difference in VASobs scores (observer), consistent with independent related variable (FLACC scores)	Nil – very low
Ravikiran et al, 2011 (386)	neonates	Procedural (immunisation) Paed outpatient dept (India)	3	Hypothesis testing: Difference in VASobs scores (observer), consistent with independent related variable (NIPS scores)	Very low

Rubinstein et al, 2016 (387)	1 – 10 years	Procedural (laceration repair) ED (Israel)	5	Hypothesis testing: No difference in VASobs scores (parent, investigator), consistent with independent related variable (self-report)	Low
Shah et al, 2008 (388)	neonates (≥ 37 weeks GA age)	Procedural (IM injection) Neonatal unit (Canada)	5	Hypothesis testing: No difference in VASobs scores (nurse, parent), inconsistent with independent related variables (cry, latency to cry)	Nil
Shaikh et al, 2011 (389)	6 – 36 months	Procedural (tympanocentesis) Outpatient general pediatric clinic (US)	3	Hypothesis testing: No difference in VASobs scores (parent, nurse, physician), consistent with independent related variables (cry, cry duration), inconsistent with independent related variable (HR). Inter-rater reliability (COSMIN poor)	Nil - very low Very low
Shavit et al, 2009 (390)	12 – 16 years	Procedure (venepuncture) ED (Israel)	3	Hypothesis testing: no difference in VASobs scores (nurse), consistent with independent related variable (self-report distress)	Nil to very low
Simons et al, 2003 (391)	newborns	Procedure: procedures/ stressful events Neonatal ICU (Netherlands)	5	Hypothesis testing: no difference in VASobs scores (nurse), consistent with independent related variables (PIPP & NIPS scores)	Low
Sinha et al, 2006 (392)	6 – 18 years	Procedural (laceration repair) Tertiary paed ED (US)	3	Hypothesis testing: difference in VASobs scores (parent), consistent with independent related variable (self-report >10y), inconsistent with independent related variable (self-report <10y)	Very low & very low negative * age related
Skarbek-Borowska et al, 2006 (393)	8 – 18 years	Procedural (IV catheter insertion) ED (US)	5	Hypothesis testing: no difference in VASobs scores (nurses), inconsistent with independent related variables (parent VASobs & self-report)	Nil

Young et al, 1996 (328)	6mth – 18 years	Procedural (venepuncture) Paed outpatient (US)	3	Hypothesis testing: Difference in VASobs (parent) scores. Consistent with independent (self-report, effectiveness assessment (observer)) & dependent (VRSobs (parent), effectiveness assessment (parent) related variables.	Moderate
Zempsky et al, 1997 (394)	5 – 18 years	Procedural (suturing) ED (US)	3	Hypothesis testing: VASobs scores did not differ, consistent with independent (self-report) & dependent (supplemental anaesthesia requirement) variable.	Low
Zempsky et al, 2008(395)	3 – 18 years	Procedural (venepuncture, IV catheter insertion) Children's Hospitals (US)	5	Hypothesis testing: difference in VASobs scores (parent), consistent with independent related variable (self-report)	Moderate
Zempsky et al, 2008 (396)	3 – 7 years	Procedural (venepuncture) Children's Medical Centre (US)	3	Hypothesis testing: No difference in VASobs scores (parent), consistent with independent related variable (self-report)	Very low
Postoperative					
Bouwmeester et al 2001 (397)	0 – 3 years	Postoperative (non-cardiothoracic & abdominal) Paed Surgical ICU (Netherlands)	3	Hypothesis testing: no difference in VASobs scores (nurse), conflicts with independent related variable (COMFORT score) at one time point. Evidence	Nil
Hamers et al, 1999 (398)	3 - 12 years	Postoperative (tonsil & adenoid surgery Setting not stated	3	Hypothesis testing: no difference in VASobs scores (parent, researcher) or independent related variables (FLACC, CHEOPS, self-report, activity)	Low
Kjeldgaard Pedersen et al, 2016 (399)	3 – 13 years with CP	Postoperative (osteotomy) Paed orthopaedic department (Denmark)	3	Hypothesis testing: Difference in VASobs scores (parent), consistent with independent related variable (r-FLACC)	Low
Knutsson et al, 2006 (375)	3 – 10 years	Postoperative (adenoidectomy)	3	Hypothesis testing: No difference in VASobs scores (nurse), consistent with independent related variable (self-report)	Very low

		Otorhinolaryngology department (Sweden)			
Muthusamy et al, 2010 (400)	3 – 18 years	Post-operative (outpatient surgical procedures) Children's Hospital (US)	3	Hypothesis testing: Difference in VASobs scores (parent), consistent with dependent related variable (TQPM questionnaire)	Low
Oztekin et al, 2002 (401)	5 – 14 years	Postoperative (tonsillectomy) PACU & ward (US)	3	Hypothesis testing: Difference in VASobs scores (blinded investigator), consistent with independent related variable (analgesic consumption)	Low
Prins et al, 2008 (402)	6months – 2 years	Postoperative (craniofacial surgery) Paed ICU (Netherlands)	3	Hypothesis testing: No difference in VASobs scores (observer), inconsistent with independent related variable (COMFORT-B scores, drug plasma levels)	Nil – very low negative
Splinter et al, 1995 (403)	1 – 13 years	Postoperative (hernia repair) Tertiary paed hospital (Canada)	3	Hypothesis testing: no difference in VASobs scores, consistent with independent related variable (mCHEOPS)	Very low
Splinter et al, 1997 (404)	2 – 6 years	Postoperative (hernia repair) Tertiary paed hospital (Canada)	5	Hypothesis testing: difference in VASobs scores	Very low - low
van der Marel et al, 2001 (405)	3 months to 3 years	Postoperative (craniofacial surgery) Surgical referral centre (Netherlands)	5	Hypothesis testing: difference in VASobs scores (nurse), consistent with independent related variable (COMFORT scores)	Low
van der Marel et al, 2007 (406)	0 – 1yr ≥36 weeks PCA	Postoperative (thoracic or abdominal) Paed ICU (Netherlands)	5	Hypothesis testing: no difference in VASobs scores (nurse, investigator) consistent with independent (COMFORT scores) & dependent related variable (morphine consumption)	Low
Disease/injury related					

Bolt et al, 2008 (407)	3 – 17 years	Acute pain (otalgia) Paed ED (Australia)	3	Hypothesis testing: No difference in VASobs scores (physician), conflicts with independent related variable (self-report)	Nil – very low negative
Coda et al, 2014 (408)	5 – 18 years	Chronic pain (disease specific pain – JIA) Paed rheumatology dept (Scotland)	3	Hypothesis testing: Difference in VASobs scores.	Nil – very low
Koller et al, 2007 (409)	6 – 18 years	Acute pain (secondary to injury) ED (US)	3	Hypothesis testing: no difference in VASobs scores, consistent with independent related variable (self-report) Interrater reliability: (COSMIN poor), Responsiveness shown (COSMIN poor), consistent with independent related variables (BP & SpO2 measurements)	Very low Very low
Miner et al, 2007 (410)	6months – 17 years	Acute pain ED (US)	3	Hypothesis testing: Difference in VASobs scores (physician), consistent with independent related variable (self-report, CHEOPS scores), inconsistent with independent related variable (physician assessed adequacy of analgesia)	Very low

Note: independent/dependent refers to whether measurement of variable is likely to influence assessment of pain using the VASobs and related/unrelated refers to whether variables are considered to contribute to an assessment of pain.

* age-related – the strength of the evidence was different for the age cohorts included in the study

Abbreviations: BP - blood pressure, CHEOPS – Children's Hospital of Easter Ontario Pain Scale, CP – cerebral palsy, DAN - Douleur Aiguë du Nouveau-né, dept – department, ED – emergency department, FLACC - Face, Legs, Activity, Cry Consolability scale, GA – gestational age, HR – heart rate, ICU – intensive care unit, IV – intravenous, JIA – Juvenile Idiopathic Arthritis, MBPS – Modified Behavioral Pain Scale, NICU – neonatal intensive care unit, NIPS – Neonates and Infant Pain Scale, OR – operating room, OSBD – observational behavioural distress scale, paed – paediatric, PBCL – procedure behaviors check list, PCA – Post conception age, PIPP – Premature Infant Pain Profile, PPAT – Pediatric Pain Assessment Tool, TQPM – Total Quality Pain Management, US – United States, VASobs – Visual Analogue Scale observer, VRS – verbal rating scale.

5.3.2 Psychometric properties and study quality

The studies eligible for this review evaluated the following psychometric properties: intra-rater reliability (2 psychometric evaluation studies), inter-rater reliability (19 studies: 17 psychometric evaluation studies and 2 RCTs), criterion validity (17 psychometric evaluation studies), responsiveness (10 studies: 6 psychometric evaluation studies and 4 RCTs) and hypothesis testing (74 studies), specifically; convergent validity (15 psychometric evaluation studies and 1 RCT), discriminant validity (1 psychometric evaluation study) and between known groups (67 studies: 2 psychometric evaluation studies and 65 RCTs). Measurement error, content validity, structural validity and cross-cultural validity were not reported in any of the studies eligible for this review.

The quality of the methods used to evaluate the psychometric properties varied across studies and within studies: ranging from ‘poor’ to ‘excellent’, with only one study scoring ‘excellent’ for methods assessing criterion validity and one for inter-rater reliability (see Table 5-2). Almost two thirds (67.3%) of the methods used to assess the psychometrics of the VASobs scored either ‘fair’ or ‘poor’ and methods used to assess scale responsiveness were not scored above ‘poor’ in any of the studies in this review. The RCT methods scored ‘3’ (n = 35) and ‘5’ (n = 23) using the Jadad Scale.

Table 5-2 COSMIN Checklist (quality) scores for psychometric parameters (342).

Study	Reliability	Hypothesis testing	Cross cultural validity	Criterion validity	Responsiveness
Psychometric evaluation studies					
Abu-Saad et al, 1995 (411)	Fair			Fair	
Bai et al, 2012 (412)				Poor	
Berntson et al, 2001 (413)				Poor	
Breau et al, 2010 (414)		Poor			
Breau et al, 2009 (415)	Poor				
Breau et al, 2001 (416)	Good			Exc	
de Jong et al, 2010 (417)	Fair				
de Jong et al, 2005 (331)	Fair				
Eyelade, et al, 2009 (418)	Fair	Fair			Poor
Filocamo et al, 2010 (419)		Good			
Garcia-Munits et al, 2006 (420)	Fair	Poor		Fair	
Hirschfeld et al, 2013 (421)		Fair			
Jylli et al, 1995 (422)	Poor			Good	
Kelly et al, 2002 (423)				Good	
Knutsson et al, 2006 (375)	Good			Good	Poor
Lawrence et al, 1993 (292)	Poor				
Liaw et al, 2012 (424)	Exc			Poor	Fair
McClellan et al, 2009 (425)		Good			Poor
McNair et al, 2004 (426)		Good			Poor
Miller et al, 1996 (427)	Fair			Poor	
Ramelet et al, 2007 (428)		Poor			
Romsing et al, 1996 (429)	Good			Fair	Poor
Schulz et al, 2002 (291)		Good		Fair	
Singer et al, 2002 (430)	Good			Good	
Spence et al, 2005 (431)		Good			
Stein et al, 1995 (432)		Good		Good	
Suominen et al, 2004 (349)	Fair				
Taddio et al, 2009 (433)	Poor	Good		Poor	

Study	Reliability	Hypothesis testing	Cross cultural validity	Criterion validity	Responsiveness
Tarbell et al, 1992 (434)		Fair			
Terstegen et al, 2003 (435)	Fair				
Valitalo et al, 2016 (436)		Good			
van Dijk et al, 2000 (173)		Fair			
Varni et al, 1987 (437)	Poor	Poor		Poor	
Voepel-Lewis et al, 2002 (438)		Poor			
Wilson et al, 1996 (439)	Good				
Randomised controlled trials					
Cohen et al, 2009 (359)					Poor
Dulai et al, 2016 (361)					Poor
Heden et al, 2009 (363)		Poor			
Ipp et al, 2004 (371)					Poor
Koller et al, 2007 (409)	Poor				Poor
Shaik et al, 2011 (389)	Poor				

The most common and clinically significant factors which impacted on the quality of the studies were: small sample sizes, failure to report missing data and inappropriate analysis methods. Specific factors impacting on method quality included: poorly described methods for reliability, poor reporting of the hypothesis, use of an inappropriate reference scale for criterion validity, observers applying both tools for convergence assessment and no blinding of observers to circumstance for responsiveness testing.

5.3.3 Data synthesis

The evidence from the psychometric evaluation studies and the RCTs were integrated to draw conclusions about the psychometric properties (reliability, validity, feasibility and clinical utility) of the VASobs and the results are grouped by psychometric property (e.g. reliability) and method of evaluation (e.g. responsiveness) and the circumstance of pain (e.g. postoperative) and summarised in Tables 5-3 to 5-6. Studies were also grouped and results summarised by age groups based on potential variability in behavioural responses to pain: neonates (preterm and term), older infants and children aged up to 3 years, children aged 3 to 12 years and over 12 years. De Jong

and colleagues reported results separately for children with injury (burn) and procedure related pain in their two studies (331, 417). Harrison reported results separately for two age cohorts (12 to 18 months and 3 to 5 years) undergoing immunisation (440) and Chapman undertook sub analysis of the results for infants aged 0 to 2 years (356).

5.3.3.1 Reliability

The two studies evaluating intra-rater reliability examine VASobs performance in circumstances and age groups that are broadly similar (infants undergoing immunisation and infants and children up to 4 years of age undergoing burns dressing). Both studies reported inconsistent reliability where kappa scores (0.69 – 0.91) (433) and intra-class correlations (ICC) (0.52 – 0.82) (331) vary considerably. Furthermore, the methods used were assessed as 'poor' and 'fair' respectively. Consequently, there is currently insufficient data of high enough quality to draw conclusions about the intra-rater reliability of the VASobs used to assess procedural pain.

Inter-rater reliability of the VASobs was evaluated when used to assess procedural pain (n = 9), postoperative pain (n = 5) and disease or injury related pain (n = 7). The age range for studies focusing on procedural and postoperative pain was from birth to 16 years, while those focusing on disease/injury related pain included children up to 18 years.

Procedural pain

Two studies addressed procedural pain assessment in term and pre-term neonates and it was possible to conclude that the VASobs showed promise as a reliable tool to assess procedural pain in neonates based on convincing results from Liaw and colleagues' study (ICC range 0.8 to 0.89 and 'excellent' methods) (424). The results of the second study were variable (r = 0.42 to 0.91), generated using methods assessed as 'poor' and although less convincing they imply potential reliability (292). Similarly, only two studies used a sample of infants and toddlers aged from 1 month to three years and only one makes a contribution to the assessment of reliability. Taddio and colleagues reported ICC results ranging from 0.55 to 0.97 (433). Considering the use of methods rated as 'poor' it is difficult to interpret these results. There is insufficient high-quality data to draw conclusions about the VASobs used to assess infants and toddlers experiencing procedural pain.

A single high-quality study (COSMIN – 'good') reported good correlation between the VASobs scores of observers (r = 0.73) used to assess procedural pain in a cohort of children aged four to five (416). No other studies provided evidence to support the reliability of the scale for young

children aged 3 to 12 years and none were found specifically addressing reliability in children older than 12 years. However, four studies used samples where age groups included infants and children across age ranges between 0 to 16 years (331, 417, 418, 422). Two psychometric evaluation studies from the same research group reported ICCs ranging from 0.56 to 0.64 using methods rated in both studies as 'fair' for cohorts aged 0 to 4 years and 0 to 5 years (331, 417). These results contribute low levels of evidence to support an assertion that VASobs is reliable when used to assess procedural pain in young children. A study which recruited children aged 0 to 16 years was excluded as this study also used non-conventional analysis to demonstrate reliability (difference between scores using an independent t-test) and reported results that confirmed a difference (422). Finally, Eyelade and colleagues reported an ICC of 0.73 for VASobs scores for children aged 6 months to 12 years undergoing venepuncture using methods rated as 'fair' (418). This study is also not sufficient to conclude that the VASobs is reliable when used to assess procedural pain in older children, but it does suggest the potential for reliable scores in this age group.

The sum of these results implies that the VASobs used to assess procedural pain in infants and young children is likely to result in scores with fair to good levels of reliability but is not conclusive in the absence of more high-quality evidence.

Postoperative pain

There are no studies addressing reliability of the VASobs in neonates or infants experiencing postoperative pain. Five studies recruited children aged between 0 and 16 years. Wilson and colleagues reported a correlation of 0.69 for children aged 2 to 11 years (439) and Knutsson and colleagues (441) correlations of 0.66 & 0.67 for children aged 3 to 10 years. Correlations between scores for children aged 3 to 15 years in Romsing and colleagues' study were slightly lower 0.52 to 0.60 but not conflicting (429). All three studies used methods rated as 'good' and contribute evidence suggesting 'good' levels of reliability for VASobs scores in this age group (442). In contrast, Miller and colleagues reported poor to fair correlations ($r = 0.36$ & 0.47) for children aged 7 to 11 years but used methods rated as 'fair' (427). Suominen and colleagues used alternative analysis methods and reported Lin's concordance correlation (0.61) and bias was 3.1 mm and the 95% limits of agreement were – 30.2 to 36.3 indicating considerable variation in scores for a cohort of children aged 0 to 16 years (349). From these studies it is possible to conclude that the reliability of VASobs scores for children aged 2 to 16 years experiencing postoperative pain is likely to be fair to good.

Illness/injury related pain

Three studies addressed the use of the VASobs to assess pain associated with juvenile arthritis (JA) in children aged from three to 18 years. The methods were assessed as ‘fair’ and ‘poor’ and correlations ranged from 0.47 to 0.94. Widely variable results and lower quality methods make it impossible to draw conclusions about VASobs use in children with JA from these studies. The remaining four studies addressed acute illness/injury related pain. In two separate studies using methods rated as ‘fair’, De Jong and colleagues reported ICCs for children with burn related pain aged between 0 to 5 years which ranged from 0.52 to 0.59 (331, 417). Using more robust methods (COSMIN = ‘good’), Singer and colleagues reported weak correlations (r = 0.4) between the VASobs scores of observers in a cohort aged 4 to 7 years (430). It should be noted that their study reported pooled data and therefore reliability results include results for children assessed during a procedure (22% of the sample). Finally, Koller and colleagues reported results for a broader age range (6 to 18 years). Their methods were rated as ‘poor’ and their results did not provide convincing evidence of VASobs reliability. The sum of the results of these studies is unclear and no conclusion about the capacity of the VASobs to generate reliable scores when used to assess acute illness/injury related pain in young or older children can be drawn in the absence of higher quality studies.

Table 5-3 Reliability results.

Study	Sample	Circumstances	Observer	Reliability	Quality
Procedural					
Breau et al, 2001 (416)	123 children aged 4-5 years	Procedural (immunisations)	Parents, technicians	Inter-rater: correlation r = 0.73, p < 0.001; n = 116	COSMIN – Good
de Jong et al, 2010 (417)	154 children aged 0 – 5 years 102 nurses rated	Procedural & background pain (burn care)	Nurse	Inter-rater: ICC for procedural pain VASobs = 0.60 (CI 0.55 – 0.65) & background pain VASobs 0.55 (CI 0.51 – 0.59)	COMSIN – Fair VAS scored last – potentially affecting scores
de Jong et al, 2005 (331)	24 children aged 0-4 years 73 nurses (grouped by hospital A & B)	Procedural & background pain (burn care)	Nurse	Inter-rater: ICC for Group A procedural 0.56 (CI 0.38 – 0.79) & background 0.52 (CI 0.20 – 0.98) Group B procedural 0.64 (CI 0.43 – 0.87) & background 0.59 (CI 0.27 – 0.98) Intra-rater: ICC for Group A procedural 0.52 (CI 0.41 – 0.61) & background 0.70 (CI 0.56 – 0.80) Group B procedural 0.82 (CI 0.76 – 0.86) & background 0.75 (CI 0.62 – 0.84)	COSMIN – Fair Sample size = 48
Eyelade et al, 2009 (418)	179 children aged 6 months to 12 years	Procedural (venepuncture)	Researcher	Inter-rater: VASobs ICC = 0.727	COMSIN – Fair Independence of observers unclear
Jylli et al, 1995 (422)	129 infants and children aged less than 16 years	Procedural pain	Parent, nurse	Inter-rater: Proportion considered to be in pain higher for parents than nurses (60% vs 77%, p < p <0.005). 27% of assessments differed	COSMIN - Poor
Lawrence et al, 1993 (292)	38 preterm and term neonates	Procedural (capillary, venous or arterial puncture)	Nurse	Inter-rater: Pearson correlations ranged from 0.42 to 0.91	COSMIN - Poor Sample size = 38 Independence of observers unclear.

Study	Sample	Circumstances	Observer	Reliability	Quality
				Paired t-tests (n = 6) used to assess inter-rater differences; one significant & two approached statistical significance	
Liaw et al, 2012 (424)	60 preterm newborns	Procedural (heel stick)	Nurses	Inter-rater: ICC ranged from 0.80 – 0.89	COSMIN - Excellent
Shaikh et al, 2011 (389)	58 children aged 6 – 36 months	Procedural (tympanocentesis)	Physician, nurse	Inter-rater: parent scores higher (62 vs 41 vs 37; P < .001)	COSMIN – Poor Analysis not suitable
Taddio et al, 2009 (433)	120 infants aged 1 year	Procedural (immunisation)	Physician, nurse, graduate student	Inter-rater: ICC ranged from 0.55 (95% CI 0.27 – 0.74) to 0.97 (95% CI 0.84 – 0.99) Intra-rater: kappa ranged from 0.69 to 0.91, where cut-off ≥ 30mm kappa ranged from 0.35 to 0.91 4.5 to 14.29% of rater scores varied more than 20mm	COSMIN – Poor Testing circumstances not consistent for observers
Post-operative					
Knutsson et al, 2006 (443)	100 children aged 3 – 10 years	Post-operative	Nurse, parent	Inter-rater: Correlation 0.66 and 0.672 (p = 0.01) Parent scores higher than nurse scores (49.02 vs 35.45, p< 0.001) and (40.79 vs 27.95, p < 0.001)	COSMIN - Good
Miller et al, 1996 (427)	20 children aged 7 – 11 years	Postoperative	Mother, nurse	Inter-rater reliability: Pearson's correlation coefficients; 0.36, p = 0.12, 0.55, p = 0.01 & 0.47, p = 0.07	COSMIN – Fair Sample size n = 20
Romsing et al, 1996 (429)	100 children aged 3 – 15 years	Postoperative	Nurse	Interrater: correlations between 2 nurse reviewers r = 0.52 – 0.60, p < 0.001	COSMIN - Good
Suominen et al, 2004 (349)	32 children aged 0 – 16 years	Postoperative (cardiac surgery)	Nurses*	Interrater: Lin's concordance correlation – 0.61 (95% CI 0.38 – 0.83). Bland Altman limits of agreement for 95% of scores -30.2 to 36.3	COSMIN - Fair

Study	Sample	Circumstances	Observer	Reliability	Quality
		& post non-painful stimulus			* bedside & other - may potentially increase agreement
Wilson et al, 1996 (439)	40 children aged 2 – 11 years	Postoperative (general & ENT)	Parent & medical observer	Interrater: correlation 0.69, $p < 0.01$ and 0.73, $p < 0.01$. 95% CI for the limits for the difference btw observer scores -7 to -15mm	COSMIN - Good
Injury and disease related					
Abu-Saad et al, 1995 (411)	33 children aged 7 – 16 years	Disease related pain (Juvenile arthritis)	Parent, physician	Inter-rater: Spearman rank order correlation for parent and physician VASobs score not significant ($r = 0.10$)	COSMIN – Fair Missing data not described
Garcia-Munitis et al, 2006 (420)	94 children aged 5 – 18 years	Disease related pain (juvenile arthritis)	Mother, father, physicians	Inter-rater: correlation ranged between 0.47 and 0.94	COSMIN – Fair Missing data not described
Koller et al, 2007 (409)	66 children aged 6 – 18 years	Acute pain (injury)	Parent, nurse, investigator	Inter-rater: Differences btw scores significant ($p < 0.001$)	COSMIN – Poor Analysis not suitable for reliability
Singer et al, 2002 (430)	57 children aged 4 – 7 years	Acute & procedural pain	Parent & clinician	Interrater: correlation $r = 0.04$ ($p = 0.001$)	COSMIN – Good Data not analysed separately
Varni et al, 1987 (437)	25 children aged 4 – 16 years	Disease related pain (Juvenile arthritis)	Parent, physician	Inter-rater: correlation $r = 0.85$ ($p < 0.001$)	COSMIN – Poor Sample size = 25

Abbreviations: CI – confidence intervals, ICC – intraclass correlation, VAS – visual analogue scale

5.3.3.2 Validity

Procedural pain

A total of 58 studies (8 psychometric evaluation studies and 50 RCT) used or evaluated use of the VASobs for procedural pain assessment. Following closer assessment, the results of comparison of the VASobs between groups in 14 RCTs were not mirrored by the results for other variables in the study e.g. other measures of pain. Therefore, these studies did not provide evidence to support VASobs validity and are not reported in the tables. This left 44 studies, which are summarised in the following sections addressing validation. Participants recruited for studies that provided validation data included neonates to 18-year-old children experiencing a range of procedures the most common of which was immunisation (n = 18). Other needle related procedures e.g. intravenous (IV) catheter insertion and intramuscular (IM) injection, accounted for the procedure in another 22 studies.

Four studies contributed evidence to support VASobs validation in a neonatal sample experiencing immunisation, heel prick and endotracheal suction related pain (386, 391, 424, 436). The RCTs reported results that contributed low and very low levels of evidence that supported the capacity of the VASobs to distinguish between known groups (386, 391). Liaw and colleagues demonstrated scale responsiveness (baseline 2.32, SD 1.94 vs procedure 7.59, SD 2.82; $p < 0.001$) but as the quality of the methods were rated as 'fair' their contribution to VASobs validation is weakened (424). Although the quality of the methods used by Valitalo and colleagues were rated as 'good', widely varying correlations between VASobs score and other related constructs from 0.02 (physiological items) to 0.8 (pain model) make interpretation of their results and their contribution to VASobs validity in neonates difficult (436). Liaw and colleagues also reported strong correlations between VASobs and PIPP scores (ICC 0.75 to 0.82) as evidence of excellent criterion validity (424). However, as PIPP scores cannot be considered 'gold standard' the COSMIN rating was 'poor'. These results are more appropriately viewed as evidence of convergence, in which case they provide moderate levels of evidence to support VASobs validity. Finally, Kozer and colleagues conducted an RCT using the VASobs to measure the difference between groups of infants aged from birth to 2 months (376). Although not limited to neonates, the age range was sufficiently narrow to consider that the evidence (low) provided by their study added t the weight of evidence addressing the capacity of the scale to differentiate between known groups of neonates. Nonetheless, there was insufficient data to confidently conclude that the VASobs is valid when used to assess procedural pain in neonates but sufficient to consider it possible.

Nine studies recruited infants and toddlers aged between 1 month and 3 years or reported results for this age group separately (356, 371, 373-375, 378, 383, 433, 440) and eight of these studies concentrated on immunisation related pain (371, 373-375, 378, 383, 433, 440). Ipp's research group are responsible for three of the seven RCTs which demonstrate the capacity of the VASobs to differentiate between known groups, each contributing low levels of evidence for VASobs validity used to assess immunisation related pain (371, 373, 374). Four independent RCTs also provided very low (n = 2), low and moderate levels of evidence respectively, supporting the VASobs capacity to distinguish between known groups of infants and young children (356, 375, 378, 440). In contrast, Taddio and colleagues' study designed to assess the validity of the VASobs used to assess immunisation related pain in infants aged 12 months old provided no evidence that the VASobs can differentiate between known groups (433). Excellent correlations between VASobs and MBPS scores are reported as evidence of strong criterion validity. However, as the MBPS cannot be accepted as 'gold standard' this would also be better described as convergence validation and as such the quality of the methods could be considered 'good' with results providing strong evidence to support the validity of the scale. The sum of the results from these studies was sufficient to cautiously suggest that the VASobs is valid when used to assess immunisation related pain in infants and toddlers. However, there was very limited data (356) addressing the scales performance used to assess alternative procedures in this age group.

Eight studies reported relevant data for populations between 3 and 12 years (352, 359, 369, 384, 396, 416, 432, 440). Five studies used a cohort undergoing immunisation (359, 369, 416, 432, 440) while the other three concentrated on children undergoing venepuncture (n = 2) (352, 396) and burns care (384). Five RCTs provided very low levels of evidence (352, 359, 369, 396, 440) and another moderate levels of evidence (384) that the VASobs can distinguish between known groups of children in this age range. Responsiveness was only tested in one study by Cohen and colleagues and scores were higher during immunisation than at baseline (ANOVA $F(1, 49) = 71.15$, $P<0.001$) (359). The methods used were rated as 'poor' and for this reason only contributed very low levels of evidence. Correlations with self-reported pain scores for children undergoing immunisation (parent: $r = 0.59$ & technician: $r = 0.60$) were reported by Breau and colleagues as evidence of the criterion validity of the scale (416). This study used methods rated as 'good' and provided moderate evidence to support the assertion that the VASobs is valid for assessing pain associated with immunisation. The sum of these results was not strong but is sufficient to suggest that the VASobs may also be valid when used to assess children aged between 3 and 12 years experiencing immunisation and needle-related pain.

No studies were found that contributed results addressing VASobs validity in children older than 12 years of age. However, many studies recruited children with wide-ranging ages (n = 24) of

which 16 included but were not limited to children aged over 12 years, nine included but were not limited to infants and toddlers and all included but were not limited to children aged between three to 13 years. None of these studies analysed the data for different age cohorts separately making it impossible to credit positive results to more specifically aged infants and children. Where age ranges are similar between studies they have been grouped and described together in the following section.

Five studies (four RCTs and one psychometric evaluation study) recruited children aged from two months old to seven years and provided evidence for this age range. The RCT results contributed very low (n = 3) (351, 368, 440) and moderate (350) levels of evidence that the VASobs can differentiate between known groups undergoing intravenous catheter (n = 3) and nasogastric tube insertion and immunisation. Schultz and colleagues reported correlations between VASobs scores and Modified Pre-verbal Early-verbal Pediatric Pain Scale (M-PEPPS) scores that provided limited evidence to support convergent validity (r = 0.37 and 0.47), particularly as the M-PEPPS had not had extensive previous psychometric testing (291). In this study correlations with self-reported scores in children old enough to self-report were reported as relatively strong (r = 0.63 & 0.92). Criterion validation in this study can only be accepted for those children that provided scores and as the study age range was 12 to 84 months and the age of the children who provided self-reported scores was not reported we can only speculate about the age group for which criterion validation was demonstrated. The results of these studies are similar to those reported in studies that included infants and toddlers and children aged 3 to 12 years and were not strong enough to substantially strengthen the conclusions drawn earlier.

In the remaining 19 studies with broad age ranges, all of which were RCTs, seven studies included infants and toddlers and the remainder included children aged over three years. The procedures included in these studies fell into two broad groups; needle related e.g. IV catheter insertion, venepuncture and port access (n = 12) and wound or fracture related procedures e.g. dressings, suturing and fracture reduction (n = 7). The consistency of the results comparing self-reported pain scores between groups and the VASobs scores between groups was used in 14 RCTs (328, 355, 356, 363, 365, 366, 370, 379-381, 387, 392, 394, 444) to establish whether the VASobs was measuring a true relationship between groups or not. In 10 studies children providing self-reported pains scores was a subset of the total study population and they ranged in age from children over 4, 6, 7 (n= 3), 8 and 10 years. However, in the remaining three studies, the authors reported that self-reported scores were provided ‘where possible’ and the age range was not defined (328, 381, 387). In the absence of robust alternative variables assessing the relationship between the known groups, these three RCTs were considered to make no contribution to the evidence of VASobs validity. Therefore, these trials contributed very low (n= 6), low (n = 7) and moderate (n = 1)

levels of evidence of the scales capacity to differentiate between known groups for children aged 0 to 18 years.

Responsiveness and convergence were tested in three studies (363, 418, 445, 446). The quality of McClellan's methods for convergence analysis was rated as 'good' and correlations between VASobs scores and the modified Observation Scale for Behavioural Distress (mOSBD) scores was (r = 0.33) and self-reported pain scores was (r = 0.43). Correlations between self-reported pain scores and VASobs scores could be considered evidence of criterion validity but the evidence supporting overall validity was only 'fair'. The other two studies contribute low and varied levels of evidence of convergence and therefore low evidence of the validity of the VASobs (363, 418). The responsiveness of the scale was shown in all three studies, however, the contribution that they make to demonstrating the scale's validity was weakened by methods 'rated as 'poor' (418, 445, 446). The results from studies using wider age ranges did not add persuasive evidence to alter the conclusions drawn based on evidence from studies with narrower age ranges.

Postoperative pain

There were 21 studies identified which contributed to understanding the psychometric performance of the VASobs when used to assess postoperative pain. Following assessment of the quality of the studies and the results, we concluded that two trials did not make significant contribution to validation of the VASobs when used to assess postoperative pain in children. The remaining 19 studies included samples ranging in age from neonates to 18 years.

Only two studies focused specifically on a neonatal population. McNair and colleagues correlated VASobs scores with PIPP and CRIES (Crying, Requires O2 for Oxygen Saturation) scores over 72 hours postoperatively with varied results ranging from 0.07 to 0.88, using methods rated as 'good' (426). The second psychometric evaluation study concentrating on neonates was designed to validate the largely untested PAT scale and reported correlations between VASobs scores and CRIES pain scale scores and PAT scale scores (r = 0.47 and r = 0.38, respectively) using methods described as only 'fair' (431). Furthermore, the psychometric properties of the PAT scale had not been adequately demonstrated to consider this scale sufficiently validated to support validation of the VASobs. These results did not provide convincing evidence of the validity of the VASobs used to assess postoperative pain in neonates.

Table 5-4 Validity (hypothesis) results.

Study	Sample	Circumstances	Observer	Results	Quality
Procedural					
Eyelade et al, 2009 (418)	179 children aged 6 months to 12 years	Procedural (venepuncture)	Researchers	Convergent – Correlation between VASobs with Oucher before r = 0.87 (p , 0.0001) & during r = 0.63, p < 0.0001) procedure	COSMIN – Fair Observer applied all scales – inc. potential for correlation
Heden et al, 2009 (363)	50 children aged 1 – 18 years	Procedural (intravenous port access)	Parent, nurse	Convergent: correlations between parent's fear & pain r = 0.60, & distress & pain r = 0.71; for nurses scores were 0.69, & 0.79, respectively	COSMIN - Poor Hypothesis not identified
McClellan et al, 2009 (425)	48 children aged 2 – 17 years (sickle cell disease)	Procedural (venepuncture)	Parents	Convergent: VASobs reactivity scores correlated with child report (0.43, p < 0.050), mOSBD (0.33, p < 0.05) & HR (0.38, p < 0.05)	COSMIN - Good
Schultz et al, 2002 (291)	38 children aged 12 – 84 months	Procedural (venous access)	Parents	Convergent: correlation between VASobs & M-PEPPS during iontophoresis r = 0.37 (p = 0.024) & during venous access r = 0.47 (p = 0.003)	COSMIN - Good
Stein et al, 1995 (432)	149 children aged 4 – 5 years	Procedural (immunisation)	Parent	Convergent – correlation btw VASobs & CHR = 0.27, p < 0.01, CHEOPS = 0.38 - 0.4, p < 0.01, GMS = 0.26, p < 0.01 CHR – change in heart rate	COSMIN - Good
Taddio et al, 2009 (433)	120 infants aged 1 year	Procedural (immunisation)	Physician, nurse, graduate student	Known groups: mean scores for non-physician raters lower in the amethocaine group (15.1, SD 19.8 vs 19.5 SD 19.2, p = 0.025). No difference in mean scores for physician raters	COSMIN - Good

Study	Sample	Circumstances	Observer	Results	Quality
Valitalo et al, 2016 (436)	118 newborns (516 scores)	Procedural (endotracheal suction)	Investigator	Convergent: correlation between VASobs & behavioural COMFORT items = 0.57 to 0.85 & behavioural PIPP items = 0.31 to 0.48 & physiological COMFORT items = 0.05 – 0.08 & physiological PIPP items = 0.02 to 0.06. Correlation between latent variable (pain estimated by graded response model) & VAS scores (r = 0.80)	COSMIN - Good
Post-operative					
McNair et al, 2004 (426)	51 neonates	Postoperative	Expert nurse	Convergent: correlations between nurse, PIPP & CRIES scores varied (ICC ranged from 0.07 – 0.88 immediately to 72hrs postop)	COSMIN - Good
Ramlet et al, 2007 (428)	19 critically unwell children aged 0 – 31 months	Postoperative & critically ill	Parents	Convergent: comparison between parents & MAPS scores using Bland-Altman method = mean of the differences of -0.29; limits of agreement 1.78 to -2.37	COSMIN – Poor Sample size
Spence et al, 2003 (431)	144* preterm & term infants	Postoperative (critically ill)	Parent	Convergent: correlation between parent & CRIES r = 0.47 ($p < 0.001$) & PAT r = 0.38 ($p < 0.01$)	COSMIN – Good
Tarbell et al, 1992 (434)	74* children aged 12 & 64 months	Postoperative (inguinal hernia or hydrocele repair)	Nurse, parent	Discriminant – VASobs scores of nurses at interval 1 [$F (93, 49) = 0.07$], VASobs score of nurses at interval 2 [$F(3, 34) = 0.53$] & VASobs score of parents [$F (3, 44) = 0.88$] did not differentiate patients based on intra-operative analgesics	COSMIN – Fair Sample size = 34

Study	Sample	Circumstances	Observer	Results	Quality
van Dijk et al, 2000 (173)	158 children aged 0 – 3 years	Postoperative (abdominal & thoracic)	Nurse	Convergent – correlations between VASobs & COMFORT behaviour = 0.89 – 0.96	COSMIN – Fair Observers used both scales – VASobs 2nd
Voepel-Lewis, Merkel et al, 2002 (438)	79 cognitively impaired children aged 4 – 18 140 observations.	Post-operative (orthopaedic or general surgery)	Parent	Convergent: correlation btw parent & FLACC scores of bedside nurse r = 0.651 & both blinded nurses r = 0.609 & 0.519 (p< 0.001). VASobs scores higher than FLACC scores bedside (bias 0.59, precision ±2.3) & blinded nurses (0.51 ±2.4 & 0.65 ±2.6)	COSMIN – Poor FLACC not validated for this population
Injury & disease related					
Filocamo et al, 2010 (419)	397 children*	Disease related pain (juvenile arthritis)	Parent	Convergent – VASobs correlated with MD global score (r = 0.61), parent global (r = 0.82), functional scale (r = 0.58), CHAQ (r = 0.54), CHQ (r = -0.75 & -0.24, physical & psychosocial respectively)	COSMIN – Good
Garcia-Munitis et al, 2006 (420)	94 children aged 5 – 18 years	Disease related pain (juvenile arthritis)	Mother, father, physicians	Convergent correlation between pain & overall well-being for observers rating scales on the same form & those rating them on separate forms were 0.93 & 0.79 (p = 0.005) respectively, for the mothers & 0.89 & 0.73 (p = 0.02) respectively, for the fathers	COSMIN – Poor Correlations inc by rating on same form
Hirschfeld et al, 2013 (421)	2276 children aged 3 – 10 years*	Nationwide survey (Germany)	Parents	Convergent = correlations between pain & disability for children (r = 0.42; 95% CI 0.38–0.45; P < 0.001)	COSMIN – Fair Survey data for prev 3mths Hypothesis not articulated

Study	Sample	Circumstances	Observer	Results	Quality
Varni et al, 1987 (437)	25 children aged 4 – 16 years	Disease related pain (Juvenile arthritis)	Parent, physician	Convergent - Physician rated disease activity index increases corresponded with increase in child, parent & physician pain scores (no significance testing)	COSMIN – Poor Sample size = 25

Abbreviations: CHEOPS – Children's Hospital Eastern Ontario pain scale, CHAQ - Childhood Health Assessment Questionnaire, CHQ – child health questionnaire, CI – confidence interval, CRIES - Crying, Requires O2 for oxygen saturation, FLACC – Face, legs, activity, cry consolability scale, HR – heart rate, ICC – intraclass correlation, MAPS - Multidimensional Assessment Pain Scale, MPEPPS – modified Pre-Verbal, Early verbal, Pediatric Pain Scale, mOSBD – modified Observation Behavioral Distress Scale, PIPP – Premature Infant Pain Profile, SD – standard deviation, VASobs – visual analogue scale applied by observer.

Children aged from 1 month to 3 years was the age range of the participants in five studies (3 psychometric evaluation studies and 2 RCTs). The two RCTS, both from the same research group (405, 406), contributed evidence that the VASobs can distinguish between known groups. However, the results were not well supported and the contribution that they made to the evidence of validity was considered low. Bai and colleagues reported excellent correlations between VASobs and FLACC (r = 0.86) but poor correlations with COMFORT-B scores (r = 0.31) both described as evidence of criterion validity (412). As the FLACC and COMFORT-B scales cannot be considered gold standard the methods were assessed as 'poor'. However, these results could be considered stronger evidence if assessed as convergence validation. Two other studies provided results for convergent validity; van Dijk and colleagues reported correlations between COMFORT-B ranging from 0.64 to 0.83 and Ramlet and colleagues reported the results of Bland-Altman analysis (bias = 0.29, limits of agreement 1.78 to -2.37) (173, 428) comparing VASobs scores with Multidimensional Assessment Pain Scale (MAPS) scores. However, the quality of the methods used to assess convergent validity were assessed as 'fair' and 'poor' and the MAPS had had limited psychometric testing, reducing the strength of the evidence for VASobs validity. The results drawn from these studies provided insufficient high-quality evidence to claim that the VAS is valid for assessing postoperative pain in infants and toddlers.

Four studies included children ranging in age from 3 to 12 years. One psychometric evaluation study correlated VASobs scores with self-reported scores and reported widely varying results compared to the child's mother (r = 0.43 to 0.83) and nurse (r = 0.23 to 0.54) and when coupled with methods assessed as 'poor' can only be considered to contribute low levels of support for VASobs validity (427). However, the quality of the second study examining criterion validity was rated as 'good' and the results were similar. The correlations between parent and nurse VASobs scores and self-reported pain scores were poor (r = 0.27) and fair (r = 0.59), respectively. Considering the strength of these methods and consistency with the results of the other study, these results should be considered evidence that criterion validity is at best fair under these circumstances. Two RCTs provided very low and low levels of evidence demonstrating the capacity of the VASobs to differentiate between known groups (398, 443). Analysis of the responsiveness of the VASobs in a fourth study, also a psychometric evaluation study, offered no support for the validation of the VASobs for this age group (443). The sum of the results from these four studies did not provide evidence to consider VASobs valid for assessing postoperative pain in children aged 3 to 12 years.

Table 5-5 Validity (responsiveness) results.

Study	Sample	Circumstances	Observer	Results	Quality
Procedural					
Cohen et al, 2009 (359)	57 children aged 4 – 6 years	Procedural (routine immunizations)	Care-giver, nurses	Scores higher during the injection compared with baseline (caregiver, F (1, 54) = 89.10, P<0.001, & nurse reports, F (1, 53) = 25.21, P<0.00)	COSMIN – Poor Observers not blinded to circumstances
Dulai et al, 2016 (361)	281 children aged 3 – 16 years	Procedural (Percutaneous pin (PP) removal)	Parents & orthopaedic technician	Pre vs post PP removal: parents reported a 2.10 (SD=2.72) increase in pain (p<0.001) & orthopaedic technicians reported a 1.76 (SD=2.10) increase in pain	COSMIN – Poor Observers not blinded to circumstances Results consistent with self-report
Eyelade et al, 2009 (418)	179 children aged 6 months to 12 years	Procedural (venepuncture)	Researchers	Increase in scores across phases median [range] baseline 2 [0 – 10], procedural 4 [0 – 10]	COSMIN – Poor Descriptive only
Ipp et al, 2004 (371)	49 infants aged 12 months	Procedural (immunisation)	Parent, physician	Median difference in pain scores (after minus before) for Priorix vs M-M-R II were as follows: paediatrician 15 vs 53 (p=.003) & parent, 22 vs 47 (p=.008)	COSMIN – Poor Descriptive only
Liaw et al, 2012 (424)	60 preterm newborns	Procedural (heel stick)	Nurses	Mean scores increased from baseline (2.32, SD 1.94) to heel-stick (7.59, SD 2.82) & then reduced following from 4.80, SD 3.36 to 2.06 SD 2.45 (p < 0.001)	COSMIN – Fair A priori hypothesis - unclear
McClellan et al, 2009 (425)	48 children aged 2 – 17 years (sickle cell disease)	Procedural (venepuncture)	Parent	Mean scores increased post venepuncture (3.2, SD 6.6 vs 29.5 SD 28.7, t(44) = 6.25, p < 0.001)	COSMIN – Poor Pattern consistent with self-report although scores higher

Study	Sample	Circumstances	Observer	Results	Quality
Post-operative					
Knutsson et al, 2006	100 children aged 3 – 10 years	Post-operative	Nurse, parent	Mean score 10 min post-operatively higher than 30min post for parents (49.02 vs 40.79, $p < 0.001$)	COSMIN – Poor Same for nurses but not significant
McNair et al, 2004 (426)	51 neonates	Postoperative	Expert nurse	Scores (n = 45) decreased over 1st 12hrs postop, remained low until 48hrs & then rose slightly	COSMIN – Poor Analysis descriptive
Romsing et al, 1996 (429)	100 children aged 3 – 15 years	Postoperative	Nurse	Scores (before & after analgesic) more pronounced for nurse scores 53 – 58% than child's PCT scores 17% ($p < 0.001$)	COSMIN – Poor Not blinded to circumstances
Injury/disease					
Koller et al, 2007 (409)	66 children aged 6 – 18 years	Acute pain (injury)	Parent, nurse, investigator	Differences in scores over time were significant ($p < 0.001$)	COSMIN – Poor Not blinded to circumstances

Abbreviations: MMR – measles, mumps & rubella, PCT – Poker chip tool, SD – standard deviation

There were no studies that specifically contributed to an understanding of the validation of the VASobs used to assess postoperative pain in children over the age of 12 years. Validity in this age group can only be inferred from the studies providing evidence addressing validity in samples that include but are not limited to this age group.

The final seven studies (two psychometric evaluation studies and five RCTs) that contributed evidence addressing validity for postoperative pain assessment across much wider age ranges (from 0 to 18 years) included two studies where the sample included but was not limited to infants and toddlers. Both studies were RCTs and were conducted by the same research team (403, 404). They provided very low levels of evidence to suggest that the VASobs can be used to differentiate between known groups in samples aged 1 – 18 and 1 to 7 years. Therefore, these studies offered little more to the assessment of the validity of the VASobs for infants and toddlers experiencing postoperative pain.

All seven studies included children aged over three years and all 5 RCTs (399-401, 403, 404) contributed low level evidence of the validity of the VASobs used to assess postoperative pain in this age group. One of the two psychometric evaluation studies reported limited evidence to support the responsiveness of the VASobs and correlations with self-report that ranged from 0.35 to 0.43 (429). The methods used for criterion validity were assessed as only 'fair' making it difficult to determine what these results tell us about VASobs validation for these circumstances. The second psychometric study reported results demonstrating the capacity of the newly developed index scale to differentiate between known groups but with no difference shown for VASobs scores between groups (434). This study was not an RCT and the quality of the methods were assessed as 'fair'. The results added nothing to our understanding of VASobs validity.

Illness/Injury related pain

Finally, the psychometric performance of the VASobs used to assess pain related to injury or illness across a range of age groups has been tested in eight studies (410, 411, 413, 419-421, 423, 430, 437). The validity of the VASobs used to assess pain in children with juvenile arthritis (JA) was assessed in five studies (411, 413, 419, 420, 437). Correlation with self-reported scores was used to explore validity in four studies and results ranged from 0.24 to 0.45 (413) to 0.32 to 0.77 (411). The quality of the methods was no higher than 'fair' for any of these studies. Three studies also assessed the convergence of VASobs scores with scores derived from other measures. The absence of analysis and significance testing prevented one study from contributing to the evidence of VASobs validity (437). Filocamo and colleagues' study reported correlation coefficients ranging from 0.54 to 0.82 for several measures (419). However, it should be noted that all

measures for correlation were rated by the same observers. Garcia-Munitis and colleagues demonstrated that correlations increased if the rating instruments were on the same page of the data collection tool rather than presented separately (mothers' ratings 0.93 vs 0.79 and fathers' ratings 0.89 versus 0.73) (420). Considering this it can be suggested that the act of rating two constructs at the same time is likely to increase convergence independent of the convergence in scores that occurs when valid tools for assessment are used to measure related constructs. This reduced the contribution that the results of these two studies made to our understanding of VASobs validity. Based on limited evidence we can suggest that the VASobs is only likely to achieve fair levels of validity when used to assess the pain experienced by children with juvenile arthritis (JA).

The remaining four studies used samples of children aged from infancy to 18 years experiencing more acute illness and injury related pain. Singer reported correlations between self-reported pain scores and VASobs scores for children with acute pain presenting to the ED (430). The quality of the methods was 'good' but the results for different observers were markedly different (parent r = 0.47 and clinician r = 0.008). Kelly and colleagues in a very similar study reported higher levels of correlation (r = 0.63) between parent and child scores. However, they examined absolute agreement using a Bland-Altman plot and reported substantial variation between scores (bias of 5mm with 95% limits of agreement of -38 to 47 mm). Furthermore, the RCT results provided no evidence of the capacity of the VASobs to detect a difference between groups. In fact, the results were the reverse of those achieved for CHEOPS scores between groups. There was insufficient evidence to draw conclusions about the validity of the VASobs when used to assess acute pain.

Table 5-6 Validity (criterion) results.

Study	Sample	Circumstances	Observer		Quality
Procedural					
Breau et al, 2001 (221)	123 children aged 4 – 5 years	Procedural (immunisations)	Parents, technicians	Correlations between child & parents scores r = 0.59 ($p < 0.001$; n = 118) & technician's r = 0.60 ($p < 0.001$; n = 120)	COSMIN – Excellent
Jylli et al, 1995 (422)	129 infants & children aged less than 16 years	Procedural pain	Parent, nurse	Correlation between child & parents scores r = 0.33 (n = 96) Children rated pain higher than nurses 34 vs 10 ($p < 0.001$) & parents 34 vs 26 ($p < 0.01$)	COSMIN - Good
Liaw et al, 2012 (424)	60 preterm newborns	Procedural (heel stick)	Nurses	Correlations with PIPP scores ranged from 0.75 – 0.82 across phases	COSMIN - Poor PIPP not 'gold standard'
Schultz et al, 2002 (291)	38 children aged 12 – 84 months	Procedural (venous access)	Parents	Correlations during iontophoresis between child & parent scores r = 0.63 ($p = 0.016$) & during venous access r = 0.92 ($p = 0.000$)	COSMIN – Fair Both higher than the M-PEPPS results
Stein et al, 1995 (432)	149 children aged 4 – 5 years	Procedural (immunisation)	Parent	Correlation between child & parent scores r = 0.3 ($p < 0.01$)	COSMIN - Good
Taddio et al, 2009 (433)	120 infants aged 1 year	Procedural (immunisation)	Physician, nurse, graduate student	Correlations with MBPS scores range from 0.81 – 0.94 using Pearson's rho	COSMIN – Poor MBPS not 'gold standard'
Post-operative					
Bai, Hsu et al, 2012 (412)	174 children aged 0 – 7 months	Post-operative (cardiac surgery)	Investigator	Correlation with FLACC r =0.86 ($p = .0001$) & COMFORT-$B_{Chinese}$ scores r=0.31 ($p=.0001$)	COSMIN – Poor FLACC not 'gold standard'

Study	Sample	Circumstances	Observer		Quality
Knutsson et al, 2006 (443)	100 children aged 3 – 10 years	Postoperative	Nurse, parent	Correlation at 30min between child & parent scores r = 0.27 (p = 0.03) & nurse r = 0.595 (p = 0.01)	COSMIN - Good
Miller et al, 1996 (427)	20 children aged 7 – 11 years	Postoperative	Mother, nurse	Correlations between child & mother scores for 3 occasions: r = 0.71 (p = 0.0005), 0.83 (p = 0.0001) & 0.46 (p = 0.07) & nurse r = 0.50 (p = 0.02), 0.54 (p = 0.01) & 0.23 (p = 0.39)	COSMIN – Poor Sample size = 20
Romsing et al, 1996 (429)	100 children aged 3 – 15 years	Postoperative	Nurse	Correlations between child & nurse scores r = 0.35 – 0.43 (p < 0.001)	COSMIN – Fair Bedside nurses therefore inc. potential correlation between scores
Injury & disease related					
Abu-Saad et al, 1995 (411)	33 children aged 7 – 16 years	Disease related pain (Juvenile arthritis)	Parent, physician	Correlations between child & parents for present pain r = 0.53 (p < 0.001)) & worst pain r = 0.77 (p < 0.05)) & between child & doctor r = 0.32 (p < 0.05)	COSMIN – Fair Missing data not described
Berntson et al, 2001 (413)	26 children aged 2 – 18 years	Disease related pain (Juvenile arthritis)	Parent	Children rated pain higher than parents & 27% of all pairs of observations were disordered (D = 0.27, MA = 0.46)	COSMIN – Poor Results reversed using the VDS-4 scale
Garcia-Munitis et al, 2006 (420)	94 children aged 5 – 18 years	Disease related pain (juvenile arthritis)	Mother, father, physicians	Correlations between child & mothers scores r = 0.45 & father r = 0.31 & physician r = 0.26 & independent physician r = 0.24	COSMIN – Fair Missing data not described

Study	Sample	Circumstances	Observer		Quality
Kelly et al, 2002 (423)	78 children aged 8 – 15 years	Acute illness/injury related pain	Parents	Correlation between child & parents' scores r= 0.63. Bias = 5mm, 95% CI -38mm to 47mm	COSMIN -
Singer et al, 2002 (430)	57 children aged 4 – 7 years	Acute & procedural pain	Parent & clinician	Correlation between child & parent scores 0.47 ($p < 0.001$) & clinician 0.008, ($p = 0.54$)	COSMIN – Good
Varni et al, 1987 (437)	25 children aged 4 – 16 years	Disease related pain (Juvenile arthritis)	Parent, physician	Correlations between child & parents scores for current pain $r = 0.72$ ($p < 0.001$) & worst pain $r = 0.54$ ($p < 0.013$) & physician scores for present pain $r = 0.65$ ($p < 0.001$). Paired t-tests showed no difference between scores	COSMIN – Poor Sample size = 25

Abbreviations: FLACC – face, legs, activity, cry & consolability, MBPS – modified behavioural pain scale, PEPPS – pre-verbal, early-verbal pediatric pain scale

5.4 Discussion

This is the most extensive review conducted to examine the psychometric properties of the VASobs in recent years and the evidence to support a contention that the scale is reliable and valid is not strong. Although the results of several studies have become available since publication of van Dijk and colleague's review of the psychometric properties of the VAS applied by an observer, in which they called for more psychometric testing, the findings of the current review were similar to those reported by these authors (290). This is particularly concerning given the frequency of the scale's use in research. In this review alone, where studies were restricted to higher quality studies RCTs (Jadad score of at least '3'(338)), 66 studies used the VASobs to measure a study outcome. It is likely that there are many more studies in the published literature using the VASobs to measure pain intensity despite insufficient data to recommend it for this purpose. In some cases, the data gave rise to concern that the scale may be unsuitable.

Traditionally, authors use terms such as; content and face, construct, criterion, concurrent, predictive and discriminant to describe different types of validity. However, these terms would be better used to describe methods of assessment that are used to explore validity rather than to imply that there are different types of validity (265). Validity is more appropriately considered a single quality and refers to whether the scale measures what it is designed to measure (265). This means that the results from these different methods of assessment cumulatively contribute to an estimate of validity. This is particularly true where there is no 'gold standard' against which the scores can be compared to determine the true accuracy of the scores. These assessment methods are therefore indirect measures of validity, making it critical that the validity of a scale is tested using multiple methods (447). Furthermore, it is increasingly accepted that testing can only demonstrate the validity of the scores rather than the scale as validity is not an immutable quality of the scale but rather will vary depending on the circumstances under which it is applied and should therefore more rightly be attributed to the scores rather than the scale (265). And finally, it is simplistic to consider that the results from studies assessing the psychometrics of a scale (or the scores generated by application of the scale under specific circumstances) can be interpreted as binary (e.g. evidence that reliability or validity is present or absent). This is well illustrated by Knutsson's study which reported ICC for interrater reliability of 0.66 using methods rated as 'good' (443). It would be inaccurate to conclude from these results that the VASobs is unreliable, but these results also do not support an assertion that the scale is highly reliable. It would be fairer to suggest that these results indicate that the VASobs can be applied by observers with moderate levels of reliability. Results are interpreted considering the strength of the results, the quality of

the methods, the range of methods used, the population and circumstances to which it was applied and the extent to which the absolute accuracy of the scores is clinically important (330).

The quality of the methods used in psychometric evaluation studies is often not high enough to place great confidence in the results and therefore the contribution that they make to our understanding of the psychometrics of a scale and the studies in this review were no exception. The methods of the studies included in this review were most often rated as 'poor' and 'fair' with only two studies using methods rated as 'excellent'; (reliability (424) and criterion validity (416)) and as a result conclusion drawn by the authors of studies must be viewed cautiously. Furthermore, the authors of these studies reported with considerable reserve, recognising the limits of the evidence that address VASobs psychometrics.

For the purposes of this review we grouped studies and analysed psychometric properties for different arbitrarily generated age ranges that we considered might potentially demonstrate differences in pain related behaviours (448-450). It was thought that this may in turn influence VASobs scores, which are an observer's global impression of the pain experienced by the infant or child, and impact on the psychometric performance of the scale for different ages. It was surprising to find that a substantial number of the psychometric evaluation studies included samples of children spanning very wide age ranges and did not report sub-analysis for narrower age groups. The results of the studies in this review can only contribute to an understanding of the psychometrics of the scale for the circumstances and age group for which it was tested, and they were often not strong. Although speculative it is possible that the results for different age groups in these studies may have differed making it more or less suitable for some cohorts of children.

Encouragingly several methods have been used to assess VASobs validity which include; hypothesis testing, convergence, criterion validation and responsiveness. In 2002, Van Dijk specifically called for studies to assess the responsiveness of the VASobs and although a total of 10 studies reported in this review assessed VASobs responsiveness, all but one of which was published since van Dijk's 2002 review, the quality of the methods used in nine of these studies were rated as 'poor'. The reviewers were not blinded to the circumstances in four studies, which may have influenced the observers' scores and created a false difference in scores over time. Finally, another four studies reported results that were either not statistically significant or that did not include significance testing. The most convincing results address the responsiveness of the VASobs to procedural pain and yet it remains insufficient to accept that the scale is responsive even for procedural pain across a range of ages.

Criterion testing was used in 15 studies assessing the performance of the VASobs under a range of circumstances and the quality of the methods used was 'good' (n = 4) or 'excellent' (n = 1) in a third of these studies, increasing our confidence in the results from these studies. However, the results of the studies using criterion testing varied widely with correlations between observer applied VAS scores and self-reported pain scores ranging from 0.008 (430) to 0.92 (291). Correlations of almost zero suggest that the scale does not measure the same construct as self-report of pain and should be abandoned for measuring pain intensity in infants and children unable to self-report. At odds with this are the results of studies that report near perfect correlations which would support accepting the VASobs as an ideal proxy measure for pain in infants and children unable to self-report. Van Dijk and colleagues also reported that the observer scores for acute pain were often lower than those self-reported and we found nothing in the studies in this review or in the adult literature (451, 452) to persuade us that this is not a consistent pattern which places patients at risk of poor pain management if decisions about treatment of acute pain are based on observer reported VAS scores.

In most circumstances the results for RCTs eligible for this review comparing VASobs scores between groups were not well supported by the results of comparisons between related but independent variables (i.e. self-reported pain scores). For this reason, these studies infrequently contributed more than low levels of evidence to suggest that the VASobs can be used to distinguish between known groups. The validity of the VASobs was also tested using correlations between VASobs scores and scores obtained using other pain scales or scores for constructs considered likely to correlate strongly with pain such as; distress, disease severity etc. Interpreting these results relies on having an excellent understanding of the data that supports the validity of scores derived from the alternative measure. So often this becomes circular as correlations between scores for different combinations of scales with limited evidence of validity are used from one study to the next to support the validity of these scales. Taddio's work to establish the validity of the VASobs is an excellent e xample of the challenge faced by researchers using an alternative measure to support the validity of the scale of interest. In this study the Modified Behavioral Pain Scale (MBPS) was used for this purpose (433). However, in earlier work intended to demonstrate the validity of the MBPS, MBPS scores were correlated with VASobs scores (288). This approach may only be accepted as a method to determine the extent to which these two scales measure the same thing. Interpreting the net result of the convergence data in this review is further complicated by inconsistent comparators and variable results.

Researchers have been criticised for using correlation coefficients such as Pearson's *r*, that measure the existence of an association between variables but do not address absolute agreement (290). There is a growing trend towards the use of analysis techniques better suited to address the

questions researchers and clinicians have about the performance of measurement scales such as the VASobs. The intraclass correlation coefficient (ICC) is a measure of reliability that provides a model for absolute agreement. Furthermore, it is better suited to circumstances where different observers rate the same phenomenon across several observations rather than where the aim is to determine the relationship between two constructs rated for the same observation (453). Although van Dijk called for this in 2002 to provide greater clarity about whether the VASobs can be reliably used to assess pain, only four studies included in this review used the intraclass correlation coefficient (ICC) to report reliability(331, 417, 424, 433). The results from these studies vary (ICCs ranged from 0.52 to 0.97) and when coupled with the results of the remaining studies which were primarily of low quality we were unable to make confident recommendations about the reliability of the VASobs for all age groups and circumstances. However, there is sufficiently promising data for several age groups and circumstances, particularly procedural pain and postoperative pain in children, to echo van Dijk's call for more high-quality studies designed to address VASobs reliability.

Intra-rater reliability of the VASobs has received scant attention and only two studies identified for this review, both published since van Dijk's review, reported results of an assessment of intra-rater reliability. In both cases the VASobs was used to assess procedural pain (331, 433). Again, the results were variable (ICC ranging from 0.52 to 0.82) and the methods used were rated as 'poor' and 'fair' and were not sufficiently robust to accept that these results contribute to our understanding of the intra-rater reliability of VASobs.

5.4.1 Limitations

There are several limitations to this review. The review was restricted to studies published in English and available in full text. It is unclear what role the results of validation of the VASobs used by non-English speakers may play in understanding the scale validity used with English speakers. Unpublished studies or work not available in full were also excluded for pragmatic reasons. As there is a significant publication bias towards studies with positive results it is possible that results that demonstrate that the VASobs is not reliable and not valid remain unpublished. The search strategy was designed to maximise the number of eligible studies found. However, it is possible that relevant studies were not identified and that they may have contributed valuable results.

The VASobs was applied by different observers in these studies which included; nurses, physicians and parents. It can be argued that the VASobs applied by different reviewers could be

considered a different scale. We did not examine data addressing the psychometrics of the scale used by different observers separately which is a limitation of this review.

5.4.2 Conclusion and recommendations

Although there are large numbers of studies reporting on various psychometric measures of the VASobs when used to measure pain there are insufficient numbers of robust studies contributing strong positive results supporting the reliability and validity of the VASobs to recommend it for infants and children for clinical and research purposes. However, there is some data to suggest that the VASobs can be applied with a fair level of reliability and that it may show reasonable validity when used to assess immunisation related pain. Based on some promising results, widespread acceptance of this scale, and the scales practical advantages we would recommend that studies of high quality evaluate the psychometric properties of this scale. Studies using larger sample sizes, a range of validation methods that include; responsiveness, capacity to distinguish between known groups and discriminate between pain and non-pain-related distress and appropriate data analysis techniques which include analysis of data based on narrower age ranges and use of appropriate statistical tests are urgently needed to support conclusive recommendations about the scale's continued use.

5.5 Addendum: New literature

An updated search was completed in the last week of May 2018 to identify studies published since the original search was completed in Aug 2016. More recent published data comes only from RCTs and of these, five studies were eligible for consideration (i.e. procedural pain assessment, Jadad score ≥3 etc.) and confirm the capacity of the VASobs to distinguish between groups based on differences in pain (361, 387, 454-456). These studies assessed interventions across a range of procedures which included: intraosseous pin removal, burns dressing, immunisation, intravenous cannula insertion and laceration repair. They also included similar and wide age ranges (infancy to adolescence) except for Cohen and colleagues' study examining the role of distraction in reducing immunisation related pain in pre-schoolers.

5.5.1 Implications

These results add to the weight of promising evidence supporting the potential for VASobs to provide a valid option for assessing procedural pain in infants and children. However, they are insufficient to alter the recommendations of this systematic review.

CHAPTER 6.

A systematic review of the FLACC scale for assessing pain in infants and children.

This chapter reports the results of a systematic review to summarise the data that describes the psychometric properties of the Face, Legs, Activity, Cry and Consolability Scale (FLACC) scale. This work has been published and the PDF of this publication is reproduced in this chapter. The chapter also includes a summary of the studies published since this review and the contribution that they make to our understanding of the psychometric properties of the FLACC scale used to assess procedural pain.

Publication:

Crellin DJ, Harrison D, Santamaria N, Babl FE. Systematic Review of the FLACC scale for assessing pain in infants and children: is it reliable, valid, and feasible for use? Pain. 2015;156(11):2132-51.

Comprehensive Review

PAIN®

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?

Dianne J. Crellin[a,b,c,*], Denise Harrison[a,b,d], Nick Santamaria[a], Franz E. Babl[b,c,e]

Abstract

The Face, Legs, Activity, Cry and Consolability (FLACC) scale is one of the most widely used behavioural observation pain scales. However, the psychometrics of the scale have not been adequately summarised and evaluated to provide clear recommendations regarding its use. The aim of this study was to rigorously evaluate the reliability, validity, feasibility, and utility of the scale for clinical and research purposes and provide recommendations regarding appropriate use of the scale. Databases searched were MEDLINE, CINAHL, Embase, PsycINFO (using the Ovid, PubMed, and Ebscohost platforms), The Cochrane Database of Systematic reviews and Cochrane Controlled Trials, and Google Scholar. Psychometric evaluation studies reporting feasibility, reliability, validity, or utility data for the FLACC scale applied to children (birth to 18 years) and randomised controlled trials (RCT) using the FLACC scale to measure a study outcome in infants and children. Data extraction included study design, population demographics, and psychometric data. Analysis involved in this study are quality assessment of the psychometric evaluation studies and the RCTs using the COSMIN checklist and the Jadad scale, respectively, and narrative synthesis of all results. Twenty-five psychometric evaluations studies and 52 RCTs were included. The study population, circumstances, and quality of the studies varied greatly. Sufficient data addressing postoperative pain assessment in infants and children exist. Some positive data support the psychometrics of the scale used to assess postoperative pain in children with cognitive impairment. Limited and conflicting data addressing procedural pain assessment exist. Content validity and scale feasibility have had limited psychometric evaluation. There are insufficient data to support the FLACC scale for use in all circumstances and populations to which is currently applied.

Keywords: Pain assessment, FLACC scale, Infant, Child, Psychometric properties

1. Background

Pain assessment is widely considered to be integral to effective pain management, and patient self-report, when available, is commonly identified as the primary source of information for assessment of pain. However, many cannot self-report pain, including infants, young children, the cognitively impaired, and the critically unwell. To facilitate pain assessment for infants and children unable to self-report, in excess of 40 multidimensional observational scales have been developed over the last 2 decades to assess and quantify pain intensity.[38] Many of these have since been adapted and used in other populations unable to self-report. These scales are a composite of a number of parameters considered indicative of pain that can be detected and graded by an observer. Commonly, these parameters are a combination of behaviours (facial expressions, body movements, and cry) and physiological parameters (heart rate, oxygen saturation, and blood pressure). The Face, Legs, Activity, Cry and Consolability (FLACC) scale, designed to assess postoperative pain in young children, is one of the most commonly used scales.[74] The FLACC scale scores pain intensity by rating 5 behaviours on a 0 to 2 scale; face, legs, activity, consolability, and cry resulting in a maximum score of 10 (**Table 1**).

The FLACC scale was published in 1997, developed as a more practical alternative to existing pain measurement scales.[74] The authors considered existing scales impractical for clinical use as they were too long or difficult to remember or score. They drew heavily from existing scales and clinical experts to identify appropriate pain-related behaviours and the descriptors to grade these behaviours. The new scale was comprised exclusively behavioural items and was originally designed and validated for use in infants and children aged 2 months to 7 years to measure postoperative pain.[74] Two observers blinded to each other's scores independently and simultaneously applied the scale to 30 children on 3 occasions each. Results showed almost perfect interrater reliability (r = 0.97). The kappa scores for the 5 items showed moderate to substantial agreement, ranging from

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

[a] *Department of Nursing, The University of Melbourne, Melbourne, Australia,* [b] *Murdoch Children's Research Institute, Melbourne, Australia,* [c] *Emergency Department, Royal Children's Hospital, Melbourne, Australia,* [d] *School of Nursing, Faculty of Health Sciences, University of Ottawa,* [e] *Department of Paediatrics, The University of Melbourne, Melbourne, Australia*

**Corresponding author. Address: Nurse Practitioner/Lecturer, Royal Children's Hospital & The University of Melbourne. c/o Emergency Department, Royal Children's Hospital, Flemington Rd, Parkville 3052, Australia. Tel.: +61 39345 5331; fax:+61 39345 5938. E-mail address: dianne.crellin@rch.org.au (D. J. Crellin).*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.painjournalonline.com).

PAIN 156 (2015) 2132–2151

http://dx.doi.org/10.1097/j.pain.0000000000000305

November 2015 • Volume 156 • Number 11 www.painjournalonline.com 2133

Table 1

Face, Legs, Activity, Cry and Consolability (FLACC) scale.[74]

Categories	Scoring		
	0	1	2
Face	No particular expression or smile	Occasional grimace or frown, withdrawn, disinterested	Frequent to constant quivering chin, clenched jaw
Legs	Normal position or relaxed	Uneasy, restless, tense	Kicking or legs drawn up
Activity	Lying quietly, normal position, moves easily	Squirming, shifting back and forth, tense	Arched, rigid, or jerking
Cry	No cry (awake or asleep)	Moans or whimpers, occasional complaint	Crying steadily, screams, or sobs, frequent complaints
Consolability	Content, relaxed	Reassured by occasional touching, hugging, or being talked to, distractible	Difficult to console or comfort

0.52 (face) to 0.82 (cry). Responsiveness to analgesics was shown (FLACC scores decreased postanalgesic from 7.0 ± 2.9 to 1.7 ± 2.2 at 10 minutes, 1.0 ± 1.9 at 30 minutes, and 0.02 ± 0.05 at 60 minutes [$P < 0.001$ at each interval]) in a second group of children (n = 30). However, the study was at risk of bias as observers were not blinded to the use of analgesics. Substantial agreement between FLACC and Objective Pain Scale scores, an existing scale designed to assess postoperative pain in infants,[15] ($r = 0.80$, $P < 0.001$) was used to demonstrate convergent validity in a third cohort of children (n = 29).

The results of the original study, although promising, were insufficient to confirm the reliability and validity of the FLACC scale when used to assess postoperative pain in infants and children aged 2 months to 7 years. Studies have attempted to provide confirmatory evidence, with many focusing on application of the scale to alternate populations (eg, older children, children with cognitive impairment, and adults) and under alternate circumstances (eg, procedural pain and critical illness).[14,71,74,96–98,105–108,112] The results of studies published before 2007 have been summarised in 2 separate systematic reviews.[27,110] They each recommend the FLACC scale to assess pain in children unable to self-report based on the strength of the available evidence. However, the conclusions of both systematic reviews suggest that there remain too many limitations to claim the scale as reliable and valid for use in all circumstances associated with its previous use.

The objectives of this systematic review were to determine the suitability of the FLACC scale to assess pain in infants and children by rigorously evaluating the psychometric properties of the scale. Specifically, to (1) identify and describe studies providing psychometric data and the populations and circumstances to which FLACC has been applied in these studies, (2) systematically review the quality of these studies using appropriate assessment tools, (3) analyse and synthesise the evidence for the psychometric and practical properties (feasibility and utility) of the FLACC scale, and (4) provide contemporary recommendations for the scale's role in pain assessment.

2. Methods

A systematic review was conducted to identify and appraise the evidence for the psychometric properties of the FLACC scale using a protocol developed by the authors and based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA statement).[77] The protocol was registered with the International Prospective Register of Systematic Reviews (CRD42014014296) and is available in full text on the PROSPERO Web Site.[26]

Primary outcomes were the reliability, validity, feasibility, and utility of the FLACC scale for assessing pain in infants and children.

2.1. Inclusion/exclusion criteria

Studies were included in this review if they reported on the reliability, validity, feasibility, or utility of the FLACC scale used to assess pain in infants and children. This included studies where the aim was evaluation of the FLACC scale, comparison between scales, including FLACC, or evaluation of an alternate scale where the FLACC scale was used as the reference. As construct validity can be demonstrated by the difference between known or extreme groups, randomised controlled trials (RCT) using the FLACC scale to measure a study outcome in infants and children were also included in the review. Infants were defined as participants aged from birth to 1 year of age, and children were defined as participants aged from 1 to 18 years.

Studies were excluded from this review if FLACC scores were not reported or analysed separately, if the sample did not include children or their results were not analysed separately, or if the study was not available in full or in English. Additionally, RCTs considered of low quality and at high risk of bias (Jadad scores <3)[57] were also excluded from analysis as they were considered unlikely to contribute evidence to the review.

2.2. Search strategy

Electronic databases searched were MEDLINE (1996—week 4, August 2014), EMBASE (1996—week 35, August 2014), the Cochrane Database of Systematic Reviews and Cochrane Controlled Trials (1996—Issue 8, August 2014), Cumulative Index Nursing and Allied Health Literature (CINAHL), and PsycINFO (1996—August 2014) using the Ovid, PubMed, and EBSCOhost platforms. Google Scholar and the reference lists of the included studies and identified reviews were also searched.

The search terms used were combinations of "FLACC" or "Face Legs Activity Cry Consolability" and "infant" or "child." Because the original publication describing the FLACC scale was published in 1997, the search range included 1 year before this date to account for publication delays and earlier developmental works. No other limits were applied.

2.3. Study selection

Duplicates were removed, and relevant abstracts were reviewed by 2 independent reviewers (D.J.C. and one of N.S., F.E.B., or D.H.). Full-text articles were reviewed where eligibility could not be determined from the abstract. A third reviewer was used to reach consensus where study eligibility remained unclear.

2.4. Data extraction and analysis

Data extraction was completed by 1 reviewer (D.J.C.) and recorded on one of 2 extraction tools: for the psychometric

evaluation studies, a modification of the QAREL data extraction form[70] designed for appraisal of diagnostic reliability studies was used and for the RCTs, a modification of the Cochrane Collaboration data collection tool designed for intervention studies[22] was used. Modifications of these forms included the deletion of irrelevant fields and the addition of fields to capture relevant methods and results not included in the original form.

Data extracted included participant details (eg, numbers, demographics), setting and circumstances of the pain being measured (associated with disease, operative, or procedural), scale description and application (eg, modifications and translation), study methods (design, psychometric properties evaluated, and statistical methods), sources of bias, and study results.

A second reviewer (F.E.B., D.H., or N.S.) checked and confirmed details of the data extraction and where there was disagreement a third assessor extracted data independently to reach consensus.

2.4.1. Quality assessment

Quality of the studies included in this review was assessed using one of 2 tools. The COSMIN checklist and a 4-point rating scale was used to assess the methodological quality of the studies focusing on psychometric evaluation,[100] and the Jadad score was used to assess the quality of RCTs.[57] Both tools were applied independently by 2 reviewers and a third if agreement was not achieved by the first 2 reviewers.

The COSMIN checklist and 4-point rating scale was developed to assess the quality of studies focused on health-related patient-reported outcome measures and provides standards for study design, statistical methods, and acceptable outcome values.[100] The checklist is also considered suitable for other clinical rating scales that measure constructs not directly measurable.

The checklist is comprised of 12 boxes, which focus on measurement properties (9 boxes), interpretability, item response theory methods, and generalizability. The measurement properties addressed by the COSMIN checklist are internal consistency, reliability (test–retest, interrater and intrarater), measurement error, content validity, construct validity (structural validity, hypothesis testing, cross-cultural validity), criterion validity, and responsiveness. Each item within these boxes is scored on a 4-point scale ("poor," "fair," "good" or "excellent") depending on the standard met by the study. The lowest item rating forms the final assessment for that property. The COSMIN taxonomy and the terms commonly used in pain scale evaluations studies are defined in **Table 2**.

The Jadad scale for assessing the quality of RCTs focuses on randomisation, blinding, and participant follow-up and results in a total score of 5, where 5 is a perfect score.[57] A minor adjustment from the original scale was made to the definition for participant follow-up. In this review, we scored follow-up as acceptable if, in the absence of the explicit statement that "there were no withdrawals from the study," all participants could be accounted for in the results.

For RCTs where reliability and/or responsiveness was assessed, these methods were evaluated using the relevant items from the COSMIN tool.

2.5. Data synthesis

The results of the search and study selection were described using the PRISMA flowchart.[77] Studies using different designs were included in this review; therefore, pooling of data for meta-analysis was not considered possible. A narrative synthesis of the evidence provided by each study was therefore used to address each of the study outcomes. It was also anticipated that eligible studies would apply the scale to different populations and under different circumstances to those for which the scale was developed and originally tested. These studies were reviewed separately to the studies concentrating on the original population and circumstances. Subgroups were created for studies focusing on similar populations based on parameters such as age (older and younger than the original age range), circumstances of the pain (procedural), and modifications to the scale (translation to another language).

Two approaches to assessing the weight of evidence have been used for similar purposes,[28,110] Method Guidelines for systematic reviews in the Cochrane Collaboration Back Review Group[103] and evaluation criteria for Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT) reviews;[23] however, each has limitations. The Cochrane Back Review Group for systematic reviews defines the weight of evidence required of RCTs and controlled clinical trials evaluating treatment effects. It is reasonable to consider that a greater

Table 2

Pain scale validation strategies and COSMIN taxonomy.

COSMIN measurement property	Pain scale measurement property	Pain scale evaluation study method
Reliability	Interrater reliability	Correlation between pain scores provided simultaneously but independently by more than 1 reviewer
	Intrarater reliability	Correlation between scores allocated by a single reviewer to the same episode of pain on separate occasions (achieved using video-taped segments)
Internal consistency	Internal consistency	Correlations between items on the scale
Measurement error	Measurement error	Calculation of the SEM, SDC, or LoA from 2 independent measures under the same conditions
		Rarely calculated in pain scale evaluation studies
Criterion validity	Concurrent validity	Correlation with assessments using the gold standard (other valid tools/scales and self-report)
Structural validity	Content validity	Principal component analysis
Cross-cultural validity	Cross-cultural validity	Translation—backwards and forwards, content review for cultural appropriateness
Hypothesis testing	Convergent validity	Correlation with assessments using other pain assessment tools/scales and/or observational scale
	Discriminant validity	Correlation with other unrelated constructs (eg, pain and hunger)
	Construct validity	Extreme or known groups comparison eg, correlation between groups undergoing different procedures or treatments
Responsiveness	Responsiveness	Change over time where change expected, eg, before and after analgesic or pain producing procedure

weight of evidence would be required to confirm the psychometric properties of a scale than would be required where the outcome is demonstrated in RCTs and controlled clinical trials. The IMMPACT review standards make no reference to the quality of the studies providing data to the review, which is a significant limitation of this approach.

3. Results

A total of 78 full-text articles were included in this review (26 psychometric evaluation studies[1,5,7,14,21,31,32,43,47,52,59,71,72,74,83,86–88,96–98,105–108,112] and 52 RCTs[2–4,6,9,10,12,16–18,20,29,30,36,37,40–42,44,45,48–50,53–56,58,62–69,75,76,80–82,84,85,90–92,95,99,101,104,109,114]) (**Fig. 1**).

3.1. Study and patient characteristics

3.1.1. Psychometric evaluation studies

Twenty-six studies evaluated psychometric properties of the FLACC scale, which included: the original evaluation of FLACC,[74] 2 additional studies evaluating the psychometric properties of FLACC when applied to the same population under the same circumstances,[14,112] and 18 studies evaluating the psychometric properties of FLACC applied to alternate populations or in alternate circumstances than those for which the scale was originally designed.[1,5,7,31,32,47,59,71,72,83,88,96–98,105–108] In 4 of these studies, item descriptors were also modified to better suit the new population (eg, to describe pain behaviours of children with cognitive impairment).[59,71,105,106] Seven studies used the FLACC scale translated into another language (Chinese, Thai, Swedish, and Brazilian Portuguese).[7,31,32,59,83,96,97] Finally, 5 studies evaluated measurement properties of another pain assessment scale and used FLACC as a reference scale.[21,43,52,86,87]

These studies are summarised in **Table 3**, and a more detailed summary of these studies is available in Appendix A (available online as Supplemental Digital Content at http://links.lww.com/PAIN/A134).

3.1.2. Randomised controlled trials

A total of 86 eligible trials were identified, which potentially met the inclusion criteria. However, 34 RCTs were excluded on retrieval of the full-text article, as they did not report FLACC scores separately (13 trials) or the Jadad quality score was less than 3 (21 trials). The population, setting and circumstances, and quality scores for the 52 included studies are summarised in **Table 4**. More details about these studies are available in Appendix B (available online as Supplemental Digital Content at http://links.lww.com/PAIN/A135).

In 26 RCTs, the FLACC scale was applied to the original population (infant and children aged 2 months to 7 years) and

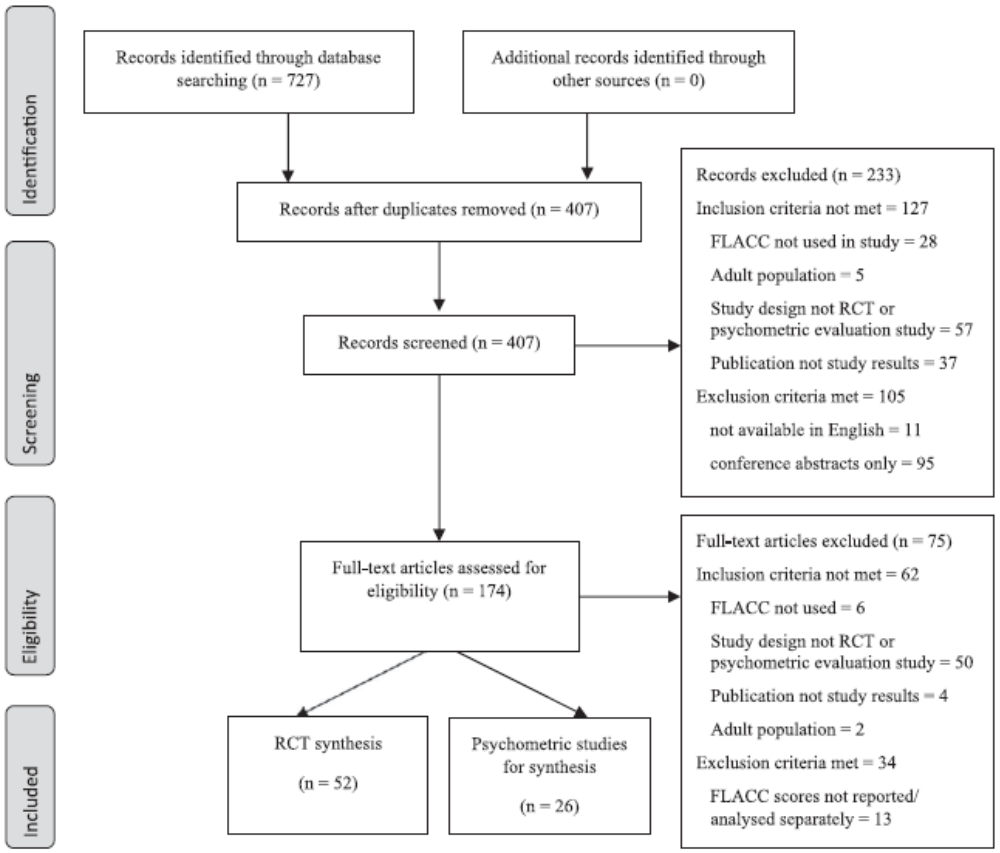


Figure 1. PRISMA flow chart detailing the search and study screening results.

Table 3

Summary of population, circumstances, and the methods used for psychometric testing for the psychometric evaluation studies.

Study	Subjects	Circumstances (Pain/setting)	Scale	Psychometric properties			Evidence
				Reliability	Validity	Feasibility/utility	
Original study							
Merkel et al.[74]	2 mo-7 y	Postoperative pain, PACU	Original FLACC	Interrater: between individuals and individual items	Hypothesis testing (convergent—OPS), responsiveness (change over time)	Not assessed	Reliability—moderate Validity—low Correlations with reference scales predictable as their items used to develop FLACC No blinding to circumstances weakening results for construct validity
FLACC repeat validation studies							
Bringuier et al.[14]	1-7 y	Postoperative pain, inpatient surgical centre, France	Original FLACC	Interrater, internal consistency	Content Structural (construct factor analysis). Hypothesis testing (convergent—CHEOPS, CHIPPS, OPS: discriminant). Criterion: (concurrent—self-report). Responsiveness (change over time)	Utility: sensitivity and specificity, identification of factors resulting in false positives and negatives	Reliability—moderate Validity—low Evidence of ability to discriminate btw anxiety and pain low Low evidence of correlation with self-report—correlations with other scales strong Highly specific but lower sensitivity (eg: false negatives) Evidence for utility—moderate
Willis et al.[112]	3-7 y	Postoperative pain, inpatient units	Original FLACC	Interrater	Criterion; (concurrent—faces)	Not assessed	Reliability—very low Small numbers—particularly reliability 6 pairs Validity—low to moderate Correlation with FACES in 5-9 y but not <5 y
FLACC validation for alternate sample, circumstances, or with scale modifications							
Anh and Jun[1]	Mean GA 32.4 at 1 wk	Procedural, NICU, Korea	Original FLACC	Tested before study	Hypothesis testing (convergent—CRIES, PIPP, known groups)	Not assessed	Validity—low Weakened as observers not blinded to circumstances and single observer applied all scales

(*continued on next page*)

Table 3 (*continued*)

Study	Subjects	Circumstances (Pain/setting)	Scale	Psychometric properties			Evidence
				Reliability	Validity	Feasibility/utility	
Babl et al.[5]	6-42 mo	Procedural, emergency department	Original FLACC	Not assessed	Hypothesis testing (known groups), responsiveness (painful procedure)	Not assessed	Validity—very low Attempt to reduce observer awareness of circumstances Responsiveness and difference between known groups—no *P* values reported
Bai et al.[7]	0-7 y	Postoperative, (cardiac) CICU China	Original FLACC translated (Chinese)	Tested before study	Criterion (concurrent—COMFORT-B)* hypothesis testing (discriminant)	Utility: sensitivity and specificity for VAS <4 and VAS ≥4. Association with independent variables	Timing, frequency, and impact of sedation and muscle relaxant on pain behaviours unclear Validity—low Weakened as observers likely not blinded to circumstances and single observer applied all scales
da Silva et al.[31]	7-17 y	Secondary to disease	Original FLACC translated (Brazilian Portuguese)	Internal consistency	Hypothesis testing (convergent—FPS-R)	Not assessed	Validity (cross-cultural)—low Some changes to descriptors required Strength of evidence weakened by incomplete cross validation and low COSMIN score
da Silva and Thuler[32]	Health professionals	Outpatient and inpatient ward, Brazil	Original FLACC translated (Brazilian Portuguese)	Not assessed	Cross-cultural validity	Not assessed	Brazilian translation* Face validity—moderate Cross-cultural validity—poor
Gomez et al.[47]	12-18 mo	Procedural pain, immunisation centre	Original FLACC	Intrarater, interrater	Not assessed	Not assessed	Reliability—high
Johansson and Kokinsky[59]	0-9 y; critically ill	Postoperative pain, PICU, Sweden	Translated (Swedish) and modified for intubated patients	Interrater	Criterion (concurrent—COMFORT-B)* responsiveness (analgesic) (change over time)	Utility: comparison of median scores for VAS ≤3 and VAS >3	Reliability—low Validity Weakened by lack of blinding (circumstances and applied both scales)
Malviya et al.[71]	4-19 y; cognitively impaired			Interrater, intrarater	Criterion (concurrent—self-report), hypothesis testing (convergent—NAPI, parent VAS), responsiveness (analgesic)	Utility: reliability with clinical significant pain severity categories	Reliability—moderate Validity—moderate Feasibility/utility —moderate

(*continued on next page*)

Table 3 (*continued*)

Study	Subjects	Circumstances	Scale	Psychometric properties			Evidence
		(Pain/setting)		Reliability	Validity	Feasibility/utility	
Manworren and Hynan[72]	1 d-34 mo	Acute and postoperative, PACU, PICU, surgical trauma and haematology/oncology units	Original FLACC	Not assessed	Responsiveness (analgesics)	Utility: treatment choice and FLACC scores	Validity—low Results markedly weakened by lack of blinding (circumstances and observer also carer) Feasibility/utility —low Not blinded to circumstances which are basis for utility assessment
Nilsson et al.[83]	5-16 y	Procedural pain, surgical and oncology units: Sweden	Original FLACC translated (Swedish)	Interrater	Content, criterion (FAS, CAS), responsiveness (painful procedure)	Not assessed	Reliability—moderate Validity—low Weakened as observers not blinded to circumstances
Ranger et al.[88]	<12 mo; critically ill	Procedural pain, cardiac ICU	Original FLACC	Interrater, intrarater	Hypothesis testing (associations between variables—univariate linear regression), responsiveness (painful procedure and analgesic)	Not assessed	Reliability—moderate Validity—low Contribution of results to validity conflicting
Suraseranivongse et al.[96]	Parents (age 1-12 y)	Postoperative (herniorrhaphy or hydrocelectomy), recovery room, Thailand	Original FLACC translated (Thai)	Interrater	Not assessed	Utility: ease of parental use—VAS score	Reliability—moderate Feasibility—low
Suraseranivongse et al.[97]	1-5.5 y	Postoperative, PACU and surgical ward, Thailand	Original FLACC translated (Thai)	Interrater, intrarater	Content, hypothesis testing: (convergent—CHEOPS, OPS TPPPS), responsiveness (change over time)	Feasibility: time and staff assessment Utility: agreement with decision to treat	Reliability—moderate Validity—moderate Weakened slightly by single observer applying all scales to video
Taddio et al.[98]	2-6 mo	Procedural (vaccination), private outpatient practice	Original FLACC	Interrater, intrarater	Criterion (concurrent—MBPS, NIPS)*, hypothesis testing (extreme groups), responsiveness (painful procedure)	Feasibility: reliability between first real time and repeated, time for assessment, user preference	Utility/feasibility—moderate Similar to OPS and CHEOPS Reliability—moderate Validity—low Weakened as observers not blinded to circumstances and single observer applied all scales Similar result to other scales in study Utility/feasibility—low Evidence supports MBPS
Voepel-Lewis et al.[108]	5-6 yr; critically ill	Acute (preanalgesia) and procedural, ICU	Original FLACC	Interrater, internal consistency	Criterion: (concurrent—COMFORT-B)* responsiveness (analgesic and painful procedure)	Not assessed	Reliability—moderate Validity—low to moderate Weakened as observers not blinded to circumstances

(*continued on next page*)

Table 3 (*continued*)

Study	Subjects	Circumstances (Pain/setting)	Scale	Psychometric properties			Evidence
				Reliability	Validity	Feasibility/utility	
Voepel-Lewis et al.[106]	4-19 y; CI	Postoperative	FLACCr; modified descriptors for children with CI	Interrater	Not assessed	Utility: clinical utility attribute score; feasibility: time taken to apply score	Reliability—low Weakened as comparison between 2 data sets Feasibility/utility—low Similar results for NAPI
Voepel-Lewis et al.[105]	4-19 y; CI	Postoperative	FLACCr; modified descriptors for children with CI	Interrater	Hypothesis testing (convergent—NRSobs), criterion: (nurse/parent FLACC and child rating), responsiveness (analgesic)	Not assessed	Reliability (parent applied FLACC)—moderate Validity—low Weakened as observers not blinded to circumstances
Voepel-Lewis et al.[107]	4-18 y; CI	Post-operative pain, bedside; modified descriptors	FLACCr; modified descriptors for children with CI	Interrater, intrarater	Hypothesis testing: (convergent—VASobs), responsiveness (analgesic)	Utility: association between level of pain and analgesia given	Reliability—moderate Validity—moderate Feasibility/utility—low
Alternate index scale (FLACC used as reference scale)							
Chorney et al.[21]	2-11 y	Postoperative pain, PACU	Original FLACC	Not assessed for FLACC	Hypothesis testing: (convergent—verbal and nonverbal distress composite score), responsiveness: (analgesic)	Not assessed	Validity—very low Reference for untested scale Scale measuring distress not pain
Fournier-Charrière et al.[43]	0-83 mo	Acute and procedural pain, ED France	Original FLACC	Not assessed for FLACC	Hypothesis testing: (convergent—EVENDOL)	Not assessed for FLACC	Not analysed separately
Hartrick and Kovan[52]	1.1-5.1 y	Postoperative and nonpainful radiological procedures, operative and radiology suites	Original FLACC	Not assessed for FLACC	Responsiveness (analgesic)	Not assessed for FLACC	Validity—low Weakened as observers not blinded to circumstances
Ramlet et al.[86]	0-31 mo; critically unwell	Postoperative, PICU—Australia and Switzerland	Original FLACC	Not assessed for FLACC	Hypothesis testing; (convergent—MAPS and FLACC)	Not assessed	Validity—very low Reference for untested scale
Ramlet et al.[87]	0-31 mo; critically ill	Postoperative, PICU	Original FLACC	Not assessed for FLACC	Convergent (MAPS and FLACC): Bland and Altman method	Not assessed for FLACC	Validity—low Reference for scale only tested once previously

* Described by author as criterion validity; however, reference scale is not an adequate gold standard.

CAS, Colour Analogue Scale; CHEOPS, Children's Hospital Eastern Ontario Pain Scale; CHIPPS, Children's and Infants' Postoperative Pain Scale; CI, cognitive impairment; CICU, cardiac intensive care unit; FAS, Facial Affective Scale; FLACC, Face, Legs, Activity, Cry, Consolability; FPS-R, Face Pain Scale-Revised; GA, gestational age; ICU, intensive care unit; MAPS, Multidimensional Assessment of Pain Scale; MBPS, Modified Behavioural Pain Scale; NAPI, Nursing Assessment Pain Intensity; NIPS, Neonatal Infant Pain Scale; NRS, Numeric Rating Scale; OPS, Objective Pain Scale; PACU, postoperative acute care unit; PICU, paediatric ICU; PIPP, Premature Infant Pain Profile; TPPPS, Toddler Preschooler Postoperative Pain Scale; VAS, Visual Analogue Scale.

Table 4
Summary of randomised controlled trials.

Study	Subject	Circumstances (specifics) and setting	Quality score	Psychometric property: results—strength of evidence
Sample and circumstances consistent with original FLACC design				
Amin[2]	4-6 y	Postoperative (adenotonsillectomy), not stated	3	Hypothesis testing: difference in FLACC scores and dependent variables (analgesic requirement)—very low evidence
Anand et al.[3]	6 mo-6 y	Postoperative (urogenital), not stated	3	Hypothesis testing: difference in FLACC scores, independent (sedation) and dependent (analgesic requirements) variables—low evidence
Ahsrey and Bosat[4]	1-7 y	Postoperative (penile surgery), recovery	3	Hypothesis testing: difference in FLACC scores not supported by independent (physiological) variables—no evidence
Batra et al.[9]	1-12 mo	Postoperative (lower abdominal and urogenital), PACU	3	Hypothesis testing: difference in FLACC scores and dependent variables—low evidence
Chandler et al.[18]	2-6 y	Postoperative (strabismus repair), PACU	3	Hypothesis testing; difference in FLACC scores and independent variable (incidence emergence delirium) and relationship between FLACC scores and related variable—low evidence
Cho et al.[20]	5-84 mo	Postoperative, not stated	5	Hypothesis testing: difference in FLACC scores and dependent variables under some circumstances—low evidence
Dewhirst et al.[36]	1-7.7 y	Postoperative (BMT), PACU	5	Hypothesis testing: difference in FLACC and dependent variables (OPS scores)—low evidence
Elshammaa et al.[41]	2-7 y	Postoperative (tonsillectomy), PACU	3	Hypothesis testing: difference in FLACC scores and independent variables (surgical time)—very low evidence
Ghai et al.[45]	6 mo-6 y	Postoperative (cataract surgery), PACU	5	Hypothesis testing: difference in FLACC scores and dependent and independent (sedation scores) variables—moderate evidence
Hippard et al.[53]	6-10 mo	Postoperative (BMT), PACU	5	Hypothesis testing: no difference in FLACC scores and correlation with independent variables (PAED score, time discharge, parental satisfaction with pain management)—low evidence
Hong et al.[54]	1-5 y	Postoperative (orchiopexy), PACU	5	Hypothesis testing: difference in FLACC and dependent (CHEOPS scores) variables—low evidence
Hong et al.[55]	2-5 y	Postoperative (hernioplasty), not stated	3	Hypothesis testing: difference in FLACC scores and no difference in independent variables (sedation score, time to oral intake and discharge)—no evidence
Hughes et al.[56]	5-10 mo	Postoperative (cleft palate repair), ward	3	Hypothesis testing: no difference in FLACC scores and dependent variables—no evidence
Jindal et al.[58]	6-48 mo	Postoperative (cleft lip repair), not stated	5	Hypothesis testing: no difference in FLACC score, difference in dependent variable—no evidence
Jonnavithula et al.[62]	5-60 mo	Postoperative (cleft lip repair), not stated	3	Hypothesis testing: no difference in FLACC scores or independent variable (sedation)—no evidence
Jonnavithula et al.[63]	8-62 mo	Postoperative (palatoplasty), postoperative recovery	3	Hypothesis testing: difference in FLACC scores and independent variable (parental satisfaction with pain relief)—moderate evidence
Kil et al.[64]	1-5 y	Postoperative (orchiopexy), day surgery unit	5	Hypothesis testing: difference in FLACC scores and dependent (CHEOPS, rescue analgesic requirements) and independent (anxiety scores, sedation) variables—low evidence
Kim et al.[65]	1-5 y	Postoperative (ambulatory surgery), PACU	5	Hypothesis testing: difference in FLACC scores and independent (sedation score, physiologic and emergence agitation) and dependent (CHEOPS) variables—moderate evidence
Kundu et al.[67]	9 mo-4 y	Postoperative (dental work), PACU	5	Hypothesis testing: no difference in FLACC or dependent variables—no evidence

(continued on next page)

Table 4 (*continued*)

Study	Subject	Circumstances (specifics) and setting	Quality score	Psychometric property: results—strength of evidence
Loetwiriyakul et al.[68]	6 mo-7 y	Postoperative (intraabdominal surgery), OR and recovery room	5	Hypothesis testing: difference n FLACC scores and dependent variables, no difference in independent variables (physiologic)—low evidence
Lorenzo et al.[69]	0-6 y	Postoperative (pyeloplasty), tertiary centre	5	Hypothesis testing: difference in FLACC scores and dependent variables—very low evidence
Sethi et al.[91]	2-6 y	Postoperative, PACU	5	Hypothesis testing: no difference in FLACC scores or independent variables (PAED scores, mYPAS scores, LOS PACU, anaesthetic duration)—low evidence
Stuth et al.[95]	2-37 mo	Postoperative (cardiac surgery), CICU	3	Hypothesis testing: no difference in FLACC scores or independent variables (NIPS scores, analgesic requirements)—very low evidence
Takmaz et al.[99]	<2 y	Postoperative (cleft lip repair), recovery and ward	4	Hypothesis testing: difference in FLACC scores, dependent and independent (parent satisfaction) variables—very low evidence
Townsend et al.[101]	3-5.5 y	Postoperative (dental), PACU	5	Hypothesis testing: no difference in FLACC scores or FACES scores (collected later)—no evidence
Voepel-Lewis et al.[109]	<28 mo	Postoperative (minor noninvasive), PACU	3	Hypothesis testing: difference in FLACC scores and dependent variables—low–moderate evidence.
Alternate sample and circumstances to original FLACC design (n = 26)				
Babl et al.[6]	1-3.8 y	Procedural (NGT insertion), ED	5	Hypothesis testing: no difference in FLACC. Responsiveness: demonstrated* (not blinded to circumstance)—low evidence
Bharti et al.[10]	1-8 y	Postoperative (lower abdominal surgery), day surgery unit	5	Hypothesis testing: difference in FLACC scores, independent (sedation) and dependent (analgesic) variables—mod evidence
Boots and Edmundson[12]	2 mo-8 y	Procedural (urinary catherisation), radiology department	3	Hypothesis testing: no difference in FLACC or independent variables (parent perception)—no evidence
Brown et al.[16]	4-13 y	Procedural (dressing change), burn centre	3	Hypothesis testing: no difference in FLACC or independent (self-report, physiologic, anxiety scores) or dependent variables—no evidence
Chadha et al.[17]	3-12 y	Procedural (nasendoscopy), otolaryngology ambulatory clinic	5	Hypothesis testing: no difference in FLACC scores or independent (self-report) variables—low evidence
Curry et al.[29]	1-7 mo	Procedural (immunisation), ambulatory paediatric clinic (2 hospitals)	3	Hypothesis testing: no difference in FLACC or independent variables (crying time)—very low evidence
Curtis et al.[30]	0-6 mo	Procedural (venipuncture), ED	3	Hypothesis testing: no difference in FLACC or independent variables (crying time)—very low evidence
Diao et al.[37]	<13 y	Postoperative (choleductal cyst excision), not stated	3	Hypothesis testing: difference in FLACC scores and independent variable (activity), responsiveness: demonstrated (not blinded to circumstance)—moderate evidence
El-Sharkawi et al.[40]	5-7 y	Procedural (dental procedure), dental clinic	3	Reliability: intrarater reliability—item kappa = 0.89-1.0. Hypothesis testing: difference in FLACC and independent variables (self-report)—low evidence
Fernandes et al.[42]	1-10 y	Postoperative (infraumbilical urological and genital procedures), not stated	5	Hypothesis testing: difference in FLACC scores and some dependent variables at some time points—low evidence
Frawley et al.[44]	1 mo-10 y	Postoperative (lower abdominal surgery), OR and recovery	5	Hypothesis testing: no difference in FLACC scores and independent (physiologic) variables—limited evidence
Grove et al.[48]	6-48 mo	Procedural (tape removal), research clinic	3	Hypothesis testing: difference in FLACC scores and independent variables—low evidence
Hall et al.[49]	11-108 d	Postoperative (pyloromyotomy), not stated	5	Hypothesis testing: no difference in FLACC between groups, difference in dependent variables, difference over time—very low evidence

(*continued on next page*)

Table 4 (*continued*)

Study	Subject	Circumstances (specifics) and setting	Quality score	Psychometric property: results—strength of evidence
Hamers et al.[50]	3-12 y	Postoperative (tonsillectomy and adenoidectomy), PACU and ward	3	No difference in FLACC and independent (time to oral intake, VASobs, and Oucher) and dependent (CHEOPS) variables—low evidence
Kim et al.[65]	4-15 y	Procedural (venipuncture), OR	5	Hypothesis testing: difference in FLACC scores between groups—low evidence
Miller et al.[75]	3-10 y	Procedural (burn care), burns outpatient centre	3	Hypothesis testing: difference in FLACC scores, dependent (FACES scale, VASobs), and independent (days to healing, dressing time)—low evidence
Miller et al.[76]	3-10 y	Procedural (burns dressing), burns clinic	3	Hypothesis testing: difference in FLACC scores and dependent scores (FACES, VASobs)—very low evidence
Natarajan Surendar et al.[80]	4-14 y	Procedural (dental), not stated	3	Hypothesis testing: difference in FLACC scores, no difference in independent variables—no evidence
Newbury et al.[81]	3 mo-15 y	Procedural (IV cannulation), ED	3	Reliability: interrater reliability—ICC = 0.86. Hypothesis testing: no difference in FLACC or dependent variables (VASobserver)—no evidence
Nilsson et al.[82]	5-12 y	Procedural (wound dressing), day care unit	3	Hypothesis testing: difference in FLACC scores and independent variables (anxiety, distress), no difference in other independent variables (self-report), responsiveness shown (not blinded to circumstances)—low evidence
Nilsson et al.[84]	7-16 y	Postoperative (day surgical procedure), PACU	3	Hypothesis testing: no difference in FLACC and independent variables (CAS, FAS and anxiety scores)—very low evidence
Nord and Belew[85]	1-21 y	Postoperative, PACU	3	Hypothesis testing: no difference in FLACC scores or independent variables (parental satisfaction with pain mx)—low evidence
Saha et al.[90]	4-12 y	Postoperative (appendicectomy), Surgical department	3	Hypothesis testing: difference in FLACC scores and dependent (analgesic requirements) and independent (complications) variables—very low evidence
Singh et al.[92]	1-10 y	Postoperative, OR and observation ward	3	Hypothesis testing: difference in FLACC scores and dependent variables (analgesia), no difference in independent variables (sedation scores, physiologic)—no evidence
Vaughan et al.[104]	<2 y	Procedural (urinary catheterisation), ED	4	Reliability: interrater reliability, ICC = 0.92-0.99. Hypothesis testing: no difference in FLACC scores, responsiveness shown (not blinded to circumstance)—low evidence
Zier et al.[114]	1-16 y	Procedural (injections—sedation) Outpatient clinic sedation area	5	Hypothesis testing: difference in FLACC scores and independent variables (VASobs nurse and parent)—moderate evidence

BMT, myringotomy and tympanostomy tube placement; CHEOPS, Children's Hospital Eastern Ontario Pain Scale; CICU, cardiac intensive care unit; ED, emergency department; FLACC, Face, Legs, Activity, Consolability, Cry; ICC, intraclass coefficient; OPS, Objective Pain Scale; OR, operating room; PACU, postoperative acute care unit; PAED, Paediatric Assessment of Emergence Delirium; VAS, Visual Analogue Scale; VASobs, VAS observer.

under similar conditions (postoperative pain) as the original scale.[2–4,9,18,20,36,41,45,53–56,58,62–64,66–69,91,95,99,101,109] In 10 of the remaining 26 trials, the sample included infants and children aged older and younger than the age range for which FLACC was originally intended,[10,37,42,44,49,50,84,85,90,92] and in the other 16 trials, the FLACC scale was used to measure procedural pain,[6,12,16,17,29,30,40,48,65,75,76,80–82,104,114] of which, 10 included older children[12,16,17,65,75,76,80–82,114] and one younger children.[30]

3.2. Psychometric properties and study quality

The 26 psychometric evaluation studies included evaluated the measurement properties: reliability (15 studies[14,47,59,71,74,83,88,96–98,105–108,112]), internal consistency (4 studies[14,31,47,98] plus one excluded as results not reported separately for children[108]), content validity (2 studies[14,97]), structural validity (one study[14]), hypothesis testing (14 studies[1,5,7,14,21,43,71,74,86–88,97,98,107,112]), cross-cultural validity (one study[32]), criterion validity (7 studies[7,31,59,83,98,105,108]), and responsiveness (13 studies[5,7,14,52,59,72,74,83,88,97,98,105,107] plus one excluded as paediatric data not analysed separately[108]). Measurement error was not reported in any of the studies. The RCTs all compared known groups using the FLACC scale to measure a study outcome (n = 52). In addition, 3 trials evaluated reliability,[40,81,104] and 2 trials assessed the scale's responsiveness.[6,104]

The quality of the methods used to measure the psychometric properties of the FLACC scale was variable, ranging from "poor" to "good," with only 9 studies scoring "good" for at least 1 property. The COSMIN checklist scores for the psychometric properties for each study are shown in **Table 5**. There were a number of common design limitations that impacted on the

November 2015 • Volume 156 • Number 11 www.painjournalonline.com 2143

Table 5

COSMIN checklist (quality) scores for psychometric parameters.

Study	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	Cross-cultural validity	Criterion validity	Responsiveness
Psychometric evaluation studies									
Merkel et al.[74]		Fair				Poor			Poor
Bringuier et al.[14]	Good	Good		Poor	Excellent	Fair			Fair
Willis et al.[112]		Poor						Fair	
Ahn and Jun[1]						Fair			
Babl et al.[5]						Fair			Fair
Bai et al.[7]						Fair		Poor	
da Silva et al.[31]	Fair							Good	
da Silva and Thuler[32]							Poor		
Gomez et al.[47]	Poor	Good							
Johansson and Kokinsky[59]		Good						Poor	Poor
Malviya et al.[71]		Good				Good		Poor	
Manworren and Hynan[72]									Poor
Nilsson et al.[83]		Fair						Fair	Fair
Ranger et al.[88]		Poor				Poor			Poor
Suraseranivongse et al.[96]		Fair							
Suraseranivongse et al.[97]		Good		Poor		Fair			Good
Taddio et al.[98]	Poor	Good				Fair		Poor	Good
Voepel-Lewis et al.[108]	Fair	Good			Fair			Poor	Fair
Voepel-Lewis et al.[106]		Poor							
Voepel-Lewis et al.[105]		Good				Fair		Poor	Fair
Voepel-Lewis et al.[107]		Fair				Poor			Fair
Chorney et al.[21]						Fair			
Fournier-Charriere et al.[43]						Poor			
Hattrick and Kovan[52]									Poor
Ramlet et al.[86]									Poor
Ramlet et al.[87]									Poor
Randomised Controlled Trails									
Babl et al.[6]									Poor
Diao et al.[37]									Poor
El Sharkawi et al.[40]		Poor							
Hall et al.[49]									Poor
Newbury and Herd[81]		Poor							
Nilsson et al.[83]									Poor
Vaughan et al.[104]		Poor							Poor

quality of these methods, the most noteworthy of which are noted in **Table 3**.

The design of the psychometric evaluation studies and the RCTs allowed unbiased blinded assessment of reliability. Despite this, 8 of the psychometric evaluation studies scored "fair" or "poor" on the COSMIN checklist[71,74,83,96,105,106,108,112] and the RCTs scored similarly. Reliability assessments were most commonly limited by small sample sizes and statistical analysis techniques. Sample size also had a substantial impact on the quality of the methods used to assess internal consistency.

The methods used for hypothesis testing in the psychometric studies (correlation with alternate scales and comparison between known groups) were limited by small sample sizes, failure to report missing values, or provide adequate descriptions of how they were managed and insufficient detail describing the comparator scale used for convergence testing and its measurement properties. As low-quality RCTs were excluded, the quality of the included RCTs was relatively higher than for the psychometric evaluation studies.

Correlation of the FLACC scale with another scale was described as criterion validity testing in 6 studies.[7,31,59,83,98,108] However, an alternate behavioural scale, which cannot be considered a gold standard for pain assessment was used in 4 of these studies,[7,59,98,105,108] resulting in low COSMIN scores. Hence, these evaluations should be more correctly considered convergence and not concurrent (criterion) testing.

Responsiveness to analgesics and procedural pain was commonly used to demonstrate the validity of the FLACC scale.[7,14,52,59,72,74,83,88,97,98,105–107] However, the quality of most methods used to assess this was poor as the observers, although blinded to group allocation, were not blinded to the patient's circumstances (eg, administration of analgesics or pain producing procedures) potentially biasing their assessments.[52,59,72,74,83,88,105–107]

3.3. Data synthesis

The level of evidence of the psychometric evaluation studies and the RCTs are provided in **Table 3 and 4**, respectively. The following sections provide a synthesis of this evidence in attempt to draw conclusions about the feasibility, reliability, validity, and clinical utility of the FLACC scale.

3.3.1. Reliability

Of the 18 studies addressing the reliability of the FLACC scale, only 3 studies evaluated reliability in the population and circumstances for which FLACC was originally intended (infants and children aged 2 months to 7 years; for postoperative pain).[14,74,112]

Merkel et al. (1997), the authors of the original FLACC scale, reported near perfect levels of agreement between observers (r = 0.97). Agreement for each category ranged from a kappa value of 0.52 (face) to 0.82 (cry). The strength of these results was limited by the small sample size and the use of a nonweighted kappa to describe agreement for an ordinal scale.[74] Both factors contributed to a COSMIN score of "fair." Bringuier et al.[14] reported near perfect interrater agreement (intraclass coefficients [ICC] = 0.9) for FLACC scores in a study rated "good" on the COSMIN checklist. The study by Willis reporting percentage agreement between 6 pairs of observations scored "poor" on the COSMIN checklist and therefore adds limited evidence to support interrater reliability.[112]

Video-taping of infants and children for later scoring by observers was only a feature of one of the studies focusing on postoperative pain in infants and children aged between 2 months and 7 years.[14] However, observers did not score these segments of video a second time to enable calculation of intrarater reliability. Hence, there has been no evaluation of intrarater reliability of the FLACC scale when applied to infants and children from the original age group experiencing postoperative pain.

Of the remaining 15 studies, 7 studies focused on children undergoing a painful procedure,[40,47,81,83,88,98,104] 4 of which included either infants younger than 2 months or children older than 7 years.[81,83,88,104] Four studies concentrated on older children with cognitive impairment.[71,105–107] Three of these also applied the scale with modified descriptors,[71,105,106] one of which also assessed application by the parents.[105] Finally, 4 studies evaluated the scale after translation to another language (Thai = 2[96,97] and Swedish = 2[59,83]) all to different populations or under unique circumstances. All but 1 study[40] evaluated the interrater reliability and 7 assessed the intrarater reliability of the FLACC scale.[47,71,88,97,98,107]

The reliability for observers applying the FLACC scale to children experiencing procedural pain was reported as almost perfect agreement with intraclass correlations ranging from 0.85 to 0.99.[40,47,83,88,98,104] Three of these 6 studies used rigorous methods and included sufficiently similar populations to more confidently draw conclusions about application of the FLACC scale to infants and young children undergoing a painful procedure.[47,88,98] Two studies also reported interrater reliability separately for each phase of the procedure (eg, baseline and during the procedure) with contrasting results. Vaughan et al.[104] and Gomez et al.[47] reported similar interrater reliability during the painful phase of the procedure (95% confidence interval [CI], [0.92-0.99] and ICC = 0.95, respectively).[47,104] However, ICCs for the baseline phases differ markedly (95% CI, [0.93-0.99] and ICC = 0.4, respectively). These results provide evidence to support the reliability of the scale for assessing pain during the painful phase of the procedure but raise questions about the reliability during nonpainful phases of a procedure. It is possible that the lower correlation demonstrated in the study by Gomez et al. is a function of lower FLACC scores with little variance. However, it should be noted that the intrarater correlation was 0.88 for the same phase of the procedure. These factors make it difficult to interpret the low interrater reliability results in their study.

Intrarater reliability of the FLACC scale has been assessed in 4 studies applying the scale to children experiencing procedural pain[47,88,98] and in 3 studies, ICC ranged from 0.88 to 0.98.[47,88,98] The age ranges of children included in each of these studies varied with no 2 studies including similar age ranges. This makes it difficult to draw conclusions about the intrarater reliability of the FLACC scale when used to measure acute procedural pain.

Application of the FLACC scale to children with cognitive impairment has been evaluated in 4 psychometric evaluation studies,[71,105–107] all conducted by the same research group. The age of study participants included children older than the original age range, and in 3 of the 4 studies, the item descriptors had been modified to better suit assessment of pain in children with cognitive impairment. Reliability coefficients for interrater reliability ranged from 0.507 to 0.92 for total scores[71,105–107] and 0.339 to 0.826 for scale items.[71,107] The quality of these studies was generally "fair" to "good," and these studies provide promising evidence for the interrater reliability of the FLACC scale applied to children with cognitive impairment, both with or without modification to the descriptors. The intrarater reliability of the FLACC scale modified to be applied to children with cognitive impairment aged between 4 and 19 years was reported in 1 "good" quality study (ICC = 0.8-0.883).[107]

Reliability testing has been reported for the Thai and Swedish versions of the scale.[59,83,96,97] The reliability of the Thai version applied to assess children experiencing postoperative pain from the original age group was tested on 2 occasions by the same authors and on 1 occasion applied by the parent.[96,97] After minor modifications to ensure the same meaning in Thai as was intended in the English version, interrater agreement (ICC = 0.949 and 0.948) and intrarater agreement (ICC = 0.095-0.99) were similarly high. It may be possible to cautiously accept the reliability of the Thai version of the FLACC scale applied to infants and children aged 2 months to 7 years experiencing postoperative pain.

The 2 studies testing the Swedish version of the FLACC scale reported application to specific and different populations to each other, and in 1 study, the cry item had been modified.[59,83] Conclusions about the reliability of the Swedish version of the FLACC scale cannot be drawn at this time.

3.3.2. Internal consistency

Four studies evaluated internal consistency of the FLACC scale.[14,31,47,98] Two studies used immunisation pain in children aged 2 to 18 months,[47,98] one used postoperative pain in children

aged 1 to 7 years[14] and one used pain in children aged 7 to 17 years with cancer after translation of the scale to Brazilian Portuguese.[31] Taddio et al.[98] included 120 children and reported an alpha correlation coefficient of 0.88 at both baseline and during the immunisation. Gomez et al.[47] reported that different items of the scale make more or less significant contributions to the overall score depending on the circumstance, eg, consolability made the largest contribution during needle insertion (0.903), whereas 15 seconds after needle insertion, the cry item made the larger contribution (0.957). In contrast to Taddio's results, all items scored low at baseline. This study was limited by a small sample size (n = 29), which was considered insufficient to meet the requirements for a high-quality examination of internal consistency. However, the findings raise important questions about the performance of the scale during different phases of a painful procedure. The study details for the other 2 studies are shown in **Table 3** and the results in Appendix A (available online as Supplemental Digital Content at http://links.lww.com/PAIN/A134).[14,31]

3.3.3. Validation

The first attempt to establish content validity was made by Bringuier et al.[14] in a study comparing the psychometric properties of 4 behavioural pain assessment scales. This assessment did not determine whether the scale was sufficiently comprehensive to assess pain and whether each item was an appropriate measure of pain and relevant to all groups for whom it had been designed. As these factors are key to content validation, this study was assigned a low score on the COSMIN checklist. Only 1 other attempt to establish the content validity of the scale has been made and that occurred after translation into Thai. In light of these results, it cannot be said that content validation for the FLACC scale has been achieved.

Three attempts to validate the FLACC scale, applied to the same population and under the same conditions as those for which it was designed, were identified.[14,74,112] Validity of the FLACC scale was inferred in the original validation study using correlation with the objective pain scale (r = 0.8) and by responsiveness of the score to treatment with analgesics (7 ± 2.9-1 ± 1.9, 30 minutes after analgesic). However, limitations of the methods used in this study significantly reduce the strength of their findings (**Table 3**).

Willis et al.[112] sought to validate the FLACC scale by correlating FLACC scores with self-reported pain scores. They demonstrated almost perfect correlation (r = 0.83, $P = 0.0001$) between the FLACC scores and self-report in older children (aged 5 to 7 years), which is much higher than the summary concordance reported in a 2008 meta-analysis (child and parent, r = 0.64) and (child and nurse, r = 0.58).[113] The correlation between FLACC scores and self-report in younger children (aged 3 to 5 years) in Willis' study is not similarly high (r = 0.254, $P = 0.381$). This result may be a function of the increasing evidence that young children have difficulty using self-report pain scales.[34,35,93,102] In light of only fair study quality, the results should be cautiously accepted as contributing evidence of FLACC validation for children aged 5 to 7 years old.

Bringuier et al.[14] provided equivocal evidence of the FLACC scale's validity applied to the original population and circumstances. Responsiveness was shown by change over time in postoperative pain scores using repeated-measures analysis of variance. However, only the P value was reported ($P < 0.001$). Correlations between FLACC and 3 other behavioural rating scales (0.88-0.94) and facial action summary score (FASS) (0.71-0.78) were high. Conversely, correlations with self-report of pain at the same times were only fair to moderate (0.31-0.51). Furthermore, FLACC scores were also moderately correlated with self-reported anxiety at 2 time points (0.63 in postoperative acute care unit and 0.63 one day postoperatively) suggesting limited capacity for the scale to discriminate between pain and anxiety.

A study aimed at assessing the psychometric properties of the Toddler, Preschooler Postoperative Pain Scale (TPPPS) compared the validity of this scale to other observational measures of pain, including the FLACC scale.[52] Results showed that FLACC scores were responsive to the administration of analgesics to children aged 1 to 5 years experiencing postoperative pain (FLACC = 5 (interquartile range 4.25-7.75) to FLACC = 0 (IQR = 0-0) $P < 0.0002$). However, the risk of bias was high as observers were not blinded to the treatment.

Nineteen RCTs provided evidence (3 contributing moderate levels[45,63,66]) supporting the validity of the scale to measure pain in children aged 2 months to 7 years experiencing postoperative pain.[2,3,9,18,20,36,41,45,53,54,63,64,66,68,69,91,95,99,109]

Finally, another 3 psychometric evaluation studies used the FLACC scale as a reference to evaluate a newly developed scale and comprised samples which included older children[21] and younger infants[86,87] than the original studies. The FLACC scale was only used as the reference scale to which the newly developed scale was compared to demonstrate concurrent validity of the new scale. Agreement with a scale not previously validated offers no evidence of FLACC validity. Where these studies used other methods such as scale responsiveness to demonstrate validity, only the new scale was assessed. Nine RCTs[10,37,42,44,49,50,84,85,90] contribute evidence (2 contributing moderate levels[10,37]) towards the validation of the FLACC scale to measure postoperative pain for older children than was originally intended (7 years). No RCTs included younger infants.

The remaining psychometric evaluation studies (n = 16)[1,5,7,31,43,59,71,72,83,88,97,98,105,107,108] and RCTs (n = 16)[6,12,16,17,29,30,40,48,65,75,76,80–82,104,114] applied the FLACC scale under different circumstances, to different populations (other than age related differences), and/or after modification to the scale descriptors or translation into languages other than English. The focus of 5 psychometric evaluation studies was the assessment of the validity of the FLACC scale applied to children undergoing a painful procedure.[1,5,82,88,98] Two other studies, focusing on assessment of acute and procedural pain, did not report results independently for procedural pain and cannot add evidence to the validity of the FLACC scale for procedural pain assessment.[43] Sixteen trials included assessment of procedural pain. Study participants in these 20 studies ranged from neonates to adolescents, and on 1 occasion, the scale was translated into another language.[82]

A study by Taddio et al.[98] applied FLACC in its original form to infants aged 2 to 6 months undergoing immunisation. The results of this study, rated "fair" on the COSMIN checklist, demonstrated that the FLACC scale was responsive to immunisation pain (FLACC scores preimmunisation mean = 0.6 [SD = 1.6] and postimmunisation mean = 6.5 (SD = 3.0), $P < 0.001$), strongly correlated with other measures (eg, Neonatal Infant Pain Scale, r = 0.92, $P < 0.001$ and Modified Behavioural Pain Scale (MBPS), r = 0.84, $P < 0.001$), and can discriminate between known groups (receipt of different vaccines, mean = 5.3 [SD = 3.30] vs mean = 7.8 [SD = 1.9], $P < 0.001$). Babl et al.[5] aimed to determine whether the FLACC scale can distinguish between pain and distress in infants and children using discrimination between known groups (painful and nonpainful procedures) and

responsiveness (before during and after a painful procedure). However, the quality of their methods rated "fair" and they did not report *P* values limiting the capacity to interpret the results. Low levels of evidence derived from 4 RCTs support the validity of FLACC to measure procedural pain in samples aged within the original limits (2 months to 7 years).[6,29,40,48]

Studies by Anh and Ranger included infants less than 2 months in their studies using FLACC to assess procedural pain.[1,88] Anh and Jun demonstrated the capacity of the FLACC scale to differentiate between neonates undergoing painful (mean FLACC score = 4.58 [SD = 2.42]) and nonpainful stimulus (routine care mean FLACC score = 1.41 [SD = 1.86] and auditory stimulus mean FLACC score = 0.69 [SD = 1.38]), $P < 0.001$. Their results also show a strong correlation between FLACC scores and CRIES scores across each category of procedure (r = 0.826, 0.843, and 0.824; $P < 0.01$ in all). Ranger et al.[88] also demonstrated responsiveness with changes in scores across phases of the procedure (baseline 0.25 [SD = 0.12], 95% CI, [0.01-0.51]; tactile 3.25 [SD = 0.56], 95% CI, [2.08-4.23]; and noxious 6.7 [SD = 0.66], 95% CI, [5.32-8.08], $P < 0.001$). However, they were unable to demonstrate similar responsiveness to the administration of analgesics and could show no correlation between FLACC scores and near-infrared spectroscopy results.

An additional 12 RCTs included older children (10 trials[12,16,17,65,75,76,80–82,114]) and younger infants (2 trials[30,104]) in trials using the FLACC scale to assess procedural pain, 7 of which contribute low levels of evidence and 1 moderate level of evidence towards validation of the FLACC scale for assessing procedural pain. Three trials also reported scale responsiveness.[6,82,104] However, observers were not blinded to the circumstances (procedure or analgesics), and the quality of these methods scored "poor." Therefore, these results contribute little to the evidence.

The use of the FLACC scale as a valid measure of pain in children with cognitive impairment aged between 4 and 18 to 21 years has been examined in 3 studies.[71,105,107] Malviya et al.[71] reported correlations between FLACC scores and Nursing Assessment of Pain Intensity scores ranging from 0.78 to 0.87, ($P < 0.01$), FLACC and parent applied Visual Analogue Scale scores ranging from 0.65 to 0.82 ($P < 0.01$), and FLACC scores and child report 0.67, ($P = 0.051$) to 0.86, ($P < 0.001$). They also demonstrated scale responsiveness with lower scores after analgesics assessed by video observers (6.1 ± 2.6 vs 1.9 ± 2.7; $P < 0.001$) and bedside observers (6.1 ± 2.5 vs 2.2 ± 2.4; $P < 0.001$). Voepel-Lewis et al. (2005) reported agreement between observer FLACC scores and Numeric Rating Scale, (ICC = 0.81 [CI, 0.70-0.89]), and child rating, (kappa = 0.65). Additionally, they demonstrated FLACC scale responsiveness to analgesics in their 2002 study[107] (blinded nurses' scores: 5.1 ± 2.9 vs 2.2 ± 3.0, $P = 0.001$) and again in their 2005 study[105] (FLACC 6.6 ± 2.4 vs 2.6 ± 2; $P = 0.003$). It should be noted that in 2 of these studies, the descriptors for the scale were modified to include pain behaviours of the included children and that these studies were all conducted by the same research group. There are no RCTs of sufficient quality to support these results.

The FLACC scale was translated into 4 languages (Brazilian Portuguese, Chinese, Swedish, and Thai), and validity was assessed in 6 psychometric evaluation studies.[7,31,32,83,96,97] Only 2 studies by the same authors focus on a similar population and circumstances (7 to 17-year olds with oncological disease) after translation into the same language (Brazilian Portuguese).[31,32] These studies provide insufficient data to support the use of FLACC after translation to these 4 languages and do not contribute to validation of the English version of the scale.

The remaining studies (n = 3) each applied the FLACC scale to different populations and circumstances and cannot be grouped.[59,72,108] Single studies do not provide sufficient validation data to draw conclusions about the validity of the scale applied to that population. These results are presented in **Table 4**.

3.3.4. Feasibility and utility

The utility of the FLACC scale has been evaluated on 9 occasions[7,14,59,71,72,96,97,106,107] and the feasibility on 3 occasions[97,98,106] in a range of populations and circumstances. Due to heterogeneity of studies, populations of children, and circumstances, it is not possible to confidently draw broad conclusions about scale feasibility or utility. Taddio et al.[98] have made the most objective attempt to determine the feasibility of several behavioural scales used to assess procedural pain, including the FLACC scale. Three observers recorded a pain score after one viewing of video-taped segments of 120 infants undergoing an immunisation procedure, then viewed the segment as frequently as necessary to reach a final score. The correlations between these scores were almost perfect for all scales (0.97-0.99, FLACC = 0.98), and there was no difference in the proportion of final pain scores achieved after the first viewing across scales (50%-66%, FLACC = 50%, $P = 0.06$). Application of the FLACC scale took the longest time in total (5 hours 25 minutes to 6 hours 50 minutes, FLACC = 6 hours 55 minutes), and only 20% preferred the FLACC scale (80% preferred the MBPS).

Nine studies evaluated the clinical utility of the FLACC scale, across heterogeneous populations, circumstances, and/or after modification or translation.[7,14,59,71,72,96,97,106,107] Only 1 cohort, children aged 4 to 19 years with cognitive impairment, was studied on more than 1 occasion (3 studies).[71,106,107] Each of these studies was conducted by the same research group and after modification to individualise the descriptors for the children included in the study. These studies used similar approaches to each other to show good levels of agreement between observers' scores coded to clinically meaningful categories: "mild," "moderate," and "severe."[71,106,107] These data provide a reasonable foundation to accept the clinical utility of a modified FLACC scale applied to children aged 4 to 19 years with cognitive impairment.

4. Discussion

This systematic review is the first comprehensive and robust review of the psychometric properties of the FLACC scale to be undertaken since it was developed. Previously published reviews assessing the psychometrics of the FLACC scale offer limited insight into the quality of the studies contributing evidence and make limited attempts to quantify the weight of evidence required to support their recommendations.[27,110] This review attempts to address these limitations and provides a unique platform for making recommendations about the application of the FLACC scale in practice and identifying directions for future research and development.

The FLACC scale was developed as existing behavioural pain assessment scales were considered too long and difficult to score and remember and impractical to apply clinically.[74] For example, the Children's Hospital Eastern Ontario Pain Scale comprised 6 items, to contribute to a total score ranging from 4 to 13.[73] Similar to existing scales from which it was in part derived,[8,15] the FLACC behaviours are scored on a consistent scale and the total score ranges from 0 to 10. The most obvious advance on existing scales is the potential ease with which

clinicians might remember the behaviours as the first letter of each item has been used to name the scale (**Table 1**). Whilst feasibility was a primary reason for the development of the FLACC scale, it was not tested by the authors of the scale and has yet to be examined in the original population, circumstances, and setting. Taddio et al.[98] completed a robust assessment of the feasibility of application of the FLACC scale to assess pain in infants undergoing immunisation. However, their conclusions suggest a clinician preference for the MBPS rather than the FLACC scale.

Procedural pain assessment presents unique challenges for a scale designed to assess postoperative pain in children. Although there is increasing concern about the use of physical restraint during procedures, in clinical, practice restraint continues to be used.[25] No attempts have been made to determine the impact of restraint on the feasibility of using the FLACC scale where restraint is likely to directly interfere with the behaviour of the child, the capacity of carers to console, or the capacity of the assessor to assess the behaviour.

There is widespread acceptance that fear and anxiety generally accompany pain during a procedure and that the behaviours associated with these emotions may significantly modify or mimic the behaviours of children experiencing procedural pain. Data demonstrating the scale's responsiveness to pain and analgesics and the capacity of the scale to differentiate between known groups undergoing painful and nonpainful procedures are cited as evidence of validity. However, questions about the capacity of the FLACC scale to discriminate pain from fear can be raised from this same data. Babl et al.[5] demonstrated the FLACC scale's responsiveness across the phases of a procedure. However, they note that FLACC scores were still high for nonpainful procedures and during the preparation phases of all procedures. These data were confirmed by other studies using responsiveness to support validity.[88]

Despite the data providing support for the validity of the FLACC scale, when examined closely, a number of concerns about the validity of the FLACC scale present themselves.[47] For example, the results of 2 separate studies, one using the Facial Action Coding System[39,46] and the other the Child Facial Coding System, demonstrate that infants and children rarely showed "jaw clenching" or "chin quivering" as an indication of pain, both of which are descriptors for the FLACC facial expression item.[14,19] This is echoed in work completed by Breau et al.[13] and more recently by Chang et al.,[19] where observers coded the postoperative facial expressions of 44 infants and children aged 13 to 74 months using the facial items of the Child Facial Coding System and those found in common behavioural scales, including the FLACC. Results confirmed concerns about reliability, face, and convergent validity, and the authors concluded that where behavioural descriptors are inconsistent with what is observed empirically the scales are likely to perform more poorly.

Furthermore, several of the descriptors of the FLACC are open to interpretation as demonstrated by Harrison et al.,[51] who showed in their recently published study that clinicians reinterpret the facial expression descriptors to include behaviours more commonly seen in infants and children experiencing pain and score accordingly. It is also unclear from the original scale description to what lengths efforts to console the child should be made before the consolability item is scored. This item is particularly problematic for procedural assessment where conduct of the procedure may impede attempts to console the child. Despite these concerns, the consolability item has been shown to make the largest contribution during needle insertion.[47] Rigorous attempts to examine the descriptors for the items included in the scale and clarify how these items should be scored have not been attempted, so doubt remains about the accuracy of the descriptors and in turn this arguably forms the basis for concerns about the validity of the FLACC scale.

Many behavioural scales are derivatives of others, and the FLACC scale is no exception. Studies aiming to validate these scales frequently use correlation between the newly developed scale and other behavioural scales to claim validity. The logic of using behavioural scales all derived from each other or a similar foundation to validate each other seems circular and likely to confirm only that they all test the same construct, but not that this construct is necessarily pain. This is a problem often faced by researchers attempting to validate tools assessing a construct where there is no clear gold standard.[89,94] To accept these correlations as evidence of the validity of the FLACC scale, the validity of the comparator scale must be established for the population and circumstances to which they are applied for comparison purposes. The authors of the studies included in this review using this approach claim validity for the comparator scales but frequently cite unconvincing evidence.

In the absence of a gold standard for comparison, researchers use multiple approaches to validation. Responsiveness is an example of a technique considered well suited to assessing the validity of pain scores. Unfortunately, many of the studies where the responsiveness of the FLACC scale to anticipated changes in pain is demonstrated did not blind observers to these circumstances. In their study measuring the reliability of the FLACC scale applied to children receiving an immunisation, the data from 2 studies confirm the notion that clinicians alter their scores to account for the circumstances at the time.[33] Furthermore, as pain behaviours may also be the behaviours of a frightened child, responsiveness may in fact be to changes in the child's levels of fear and anxiety.

Psychometric properties are not intrinsic to the scale but rather an interaction between the scale, the population to which it is applied and the circumstances under which it is applied.[94] This review has drawn attention to the diversity of populations and circumstances to which the FLACC scale has been tested and applied. Previous reviews have largely focused on the distinction between assessment of pain in children experiencing postoperative and procedural pain and have made recommendations for use of the FLACC scale for these 2 groups. It is tempting to consider that the data supporting the psychometrics of the scale applied to one cohort can be unreservedly contributed to the evidence for the psychometrics of another. However, it is widely proposed and supported by a growing body of evidence that the pain behaviours of children vary significantly with age. A number of studies show that the behaviours of premature neonates are blunted when compared with those of their term contemporaries.[24,60,61,78,79,111] The results of 2 studies from this review also challenge the notion that pain behaviours are consistent across age groups.[83,112] Correlations between FLACC scores and self-report varied across the age groups included in these studies, 3 to 7 years old[112] and 5 to 16 years old.[83] It is not clear where the boundaries exist to distinguish one age related cohort demonstrating unique pain behaviours from another. However, scales have been developed to assess neonates and preterm infants on the assumption that they are likely to exhibit unique response to pain compared with older infants. Until data are available to demonstrate that infants less than 2 months of age and children older than 7 years behave similarly to infants and children aged 2 months to 7 years, studies including infants and children outside the age limits for which FLACC was originally developed and tested must be considered to include a new population and will need reliability and validity data to address

application in this population. Similarly, studies addressing application of the scale where there are other population or circumstantial differences are needed to provide psychometric data.

Unfortunately, the small numbers of studies evaluating the measurement properties of the FLACC scale in discrete and definable groups, the limitations in the methods of many of these studies, and the concerns about the validity of the item descriptors make it difficult to confidently make recommendations about the psychometrics of the FLACC scale. However, clinicians and researchers continue to seek guidance about the role of the FLACC scale in assessing pain in different populations of children and circumstances so despite these significant limitations, recommendations based on the strength of available evidence are provided.

4.1. Recommendations

The weight of evidence for the reliability and validity of the FLACC scale applied to infants and children aged 2 months to 7 years experiencing postoperative pain is sufficient to recommend the scales' use under these circumstances. However, in the absence of feasibility data, it is not possible to recommend the scale as feasible for practice. Similarly, as the evidence is limited to the results of 1 study, it is only possible to suggest that the scale may have clinical utility when applied to this population and setting.

4.1.1. Age

There are insufficient data supporting the measurement properties of the scale used to assess postoperative pain in infants younger than 2 months of age, in particular, in neonates, to recommend the FLACC for this age group. The body of evidence addressing application of the scale to older children is larger and is sufficient to cautiously recommend the FLACC scale as valid for assessing postoperative pain in older children. However, no recommendations about the scale's feasibility or clinical utility under these circumstances can be made.

4.1.2. Cognitive impairment

There are also sufficient data to cautiously recommend the FLACC scale, particularly after modification to better suit the child, as valid for assessing pain in children aged 4 to 19 years with cognitive impairment. Based on limited data, application in this cohort is probably feasible and likely to be clinically useful, but it is insufficient to make a stronger assertion.

4.1.3. Procedural pain

For reasons elucidated earlier, accepting that available evidence supports the psychometrics of the FLACC scale used to measure procedural pain is problematic. Since publication of the 2007 reviews, data have been published that contribute some evidence to reliability, add to the validity, and suggest a clinician preference for an alternative to measure procedural pain but continue to leave some key questions unanswered. It is no longer possible to recommend the FLACC scale for procedural pain assessment despite the absence of an acceptable alternative.

4.1.4. Language and other modifications

The FLACC scale has been evaluated after modification, most often translation to another language. However, most are single studies of variable quality testing the measurement properties of the modified scale. Many of these studies report positive results but in the absence of a greater weight of evidence, no recommendation to support the scale's use in these circumstances can be made.

4.2. Future directions

There are a number of concerns regarding the FLACC scale that need to be addressed before recommendations for its use in clinical and research practice can be confirmed. Review of the appropriateness of the descriptors for the items of the scale, specifically the faces item, is urgently needed. This may culminate in modification of the scale, provoking the need for evaluation of the measurement properties of the newly modified scale. Before attempts to validate this or a modified version of the scale take place, the feasibility of using the scale in various populations and in a range of circumstances and clinical settings should also be explored. Adaptation of the scale or the descriptors to account for the circumstances of a procedure, eg, restraint, may also be needed to improve the feasibility and validity of the scale for these circumstances. Furthermore, modifications to the scale should be informed by a better understanding of the measurement properties of individual scale items and their relationship with the other scale items and the total score. To date, there are insufficient data for the various populations and circumstances to which the scale is applied in these studies to support the measurement properties of the scale items.

Additionally, the capacity for scales to discriminate between pain and distress is a cause for concern. Blount contends that behavioural indices are unlikely to be specific to pain or distress, which would make this discrimination unachievable.[11] However, to continue using the FLACC scale in circumstances where the aim is to measure pain independently of distress, for example, in studies evaluating the efficacy of pain relief, data demonstrating the capacity of the FLACC scale to discriminate between pain and distress are required. Finally, more compelling evidence demonstrating reliability, validity, and clinical utility for the range of populations, circumstances, and settings to which the scale is applied is also needed.

To provide this psychometric data, careful study design using robust methods for validation is required to reduce the bias that detracts from the results of many existing studies. The tools used in this review such as the COSMIN checklist, although not designed for this purpose, could be used to guide the development of the methods for future psychometric evaluation studies.

Finally, there are some promising data to support future research efforts focusing on the FLACC scale. However, review of alternate behavioural scales and pain assessment modalities to identify a method with better psychometric and practical properties than the FLACC scale for the assessment of pain in children is also warranted.

4.3. Limitations

There are a number of limitations to this review. A positive publication bias may well mean that unpublished data are available which conflicts with the bulk of the published data supporting the reliability, validity, and clinical utility of the FLACC scale. Similarly, excluding studies not published in English may have denied a source of data that support translation of FLACC to another language. Restriction of the focus of this study to patients from birth to 18 years means that data examining the

measurement properties of FLACC in adults have not been included. The role of FLACC to assess pain in nonverbal adults is being increasingly explored but readers should be cautioned against generalising the results of this review to this population.

Only a small number of studies provided data addressing detailed item analysis and none applying the FLACC under consistent circumstances or to a consistent population. Hence, item analysis has not been provided in this review.

There were no appropriate objective criteria available to quantify the weight of evidence considered sufficient to demonstrate the measurement properties of an assessment tool. In this review, the weight of evidence considered adequate was derived from a subjective assessment by the authors of this review.

5. Conclusions

The results of this review challenge the long-held view that the strong psychometric properties of the FLACC scale supports its use to assess pain in children from infancy to adolescents in a range of circumstances. The data used to support the psychometric properties of the scale, in particular validity, are either absent or limited and frequently derived from studies with methodological flaws. Continued application of a scale designed for postoperative pain assessment to procedural pain assessment is unsupported. It is clear that further work is required to provide a foundation from which confident recommendations about the future of the FLACC scale in paediatric pain assessment can be made.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Acknowledgements

D. J. Crellin is a trainee member of *Pain in Child Health*, a research training initiative of the Canadian Institutes of Health Research.

Appendices. Supplemental Digital Content

Supplemental Digital Content associated with this article can be found online at http://links.lww.com/PAIN/A134, and http://links.lww.com/PAIN/A135.

Article history:
Received 1 June 2015
Received in revised form 14 July 2015
Accepted 15 July 2015
Available online 23 July 2015

References

[1] Ahn Y, Jun Y. Measurement of pain-like response to various NICU stimulants for high-risk infants. Early Hum Dev 2007;83:255–62.
[2] Amin SM. Evaluation of gabapentin and dexamethasone alone or in combination for pain control after adenotonsillectomy in children. Saudi J Anaesth 2014;8:317.
[3] Anand VG, Kannan M, Thavamani A, Bridgit MJ. Effects of dexmedetomidine added to caudal ropivacaine in paediatric lower abdominal surgeries. Indian J Anaesth 2011;55:340–6.
[4] Ashrey E, Bosat B. Single-injection penile block versus caudal block in penile pediatric surgery. Ain-Shams J Anaesthesiol 2014;7:428.
[5] Babl FE, Crellin D, Cheng J, Sullivan TP, O'Sullivan R, Hutchinson A. The use of the faces, legs, activity, cry and consolability scale to assess procedural pain and distress in young children. Pediatr Emerg Care 2012;28:1281–96.
[6] Babl FE, Goldfinch C, Mandrawa C, Crellin D, O'Sullivan R, Donath S. Does nebulized lidocaine reduce the pain and distress of nasogastric tube insertion in young children? A randomized, double-blind, placebo-controlled trial. Pediatrics 2009;123:1548–55.
[7] Bai J, Hsu L, Tang Y, van Dijk M. Validation of the COMFORT Behavior scale and the FLACC scale for pain assessment in Chinese children after cardiac surgery. Pain Manag Nurs 2012;13:18–26.
[8] Barrier G, Attia J, Mayer MN, Amiel-Tison C, Shnider SM. Measurement of post-operative pain and narcotic administration in infants using a new clinical scoring system. Intensive Care Med 1989;15(suppl 1):S37–39.
[9] Batra YK, Rajeev S, Panda NB, Lokesh VC, Rao KLN. Intrathecal neostigmine with bupivacaine for infants undergoing lower abdominal and urogenital procedures: dose response. Acta Anaesthesiol Scand 2009;53:470–5.
[10] Bharti N, Praveen R, Bala I. A dose-response study of caudal dexmedetomidine with ropivacaine in pediatric day care patients undergoing lower abdominal and perineal surgeries: a randomized controlled trial. Paediatr Anaesth 2014;24:1158–63.
[11] Blount RL, Loiselle KA. Behavioural assessment of pediatric pain. Pain Res Manag 2009;14:47–52.
[12] Boots BK, Edmundson EE. A controlled, randomised trial comparing single to multiple application lidocaine analgesia in paediatric patients undergoing urethral catheterisation procedures. J Clin Nurs 2010;19:744–8.
[13] Breau LM, McGrath PJ, Craig KD, Santor D, Cassidy KL, Reid GJ. Facial expression of children receiving immunizations: a principal components analysis of the child facial coding system. Clin J Pain 2001;17:178–86.
[14] Bringuier S, Picot MC, Dadure C, Rochette A, Raux O, Boulhais M, Capdevila X. A prospective comparison of post-surgical behavioral pain scales in preschoolers highlighting the risk of false evaluations. PAIN 2009;145:60–8.
[15] Broadman L, Rice L, Hannallah RS. Testing the validity of an objective pain scale for infants and children. Anesthesiology 1988;69:A770.
[16] Brown NJ, Kimble RM, Rodger S, Ware RS, Cuttle L. Play and heal: randomized controlled trial of Ditto intervention efficacy on improving re-epithelialization in pediatric burns. Burns 2014;40:204–13.
[17] Chadha NK, Lam GOA, Ludemann JP, Kozak FK. Intranasal topical local anesthetic and decongestant for flexible nasendoscopy in children: a randomized, double-blind, placebo-controlled trial. JAMA Otolaryngol Head Neck Surg 2013;139:1301–5.
[18] Chandler JR, Myers D, Mehta D, Whyte E, Groberman MK, Montgomery CJ, Ansermino JM. Emergence delirium in children: a randomized trial to compare total intravenous anesthesia with propofol and remifentanil to inhalational sevoflurane anesthesia. Paediatr Anaesth 2013;23:309–15.
[19] Chang J, Versloot J, Fashler SR, McCrystal KN, Craig KD. Pain assessment in children: validity of facial expression items in observational pain scales. Clin J Pain 2015;31:189–97.
[20] Cho JE, Kim JY, Hong JY, Kil HK. The addition of fentanyl to 1.5 mg/ml ropivacaine has no advantage for paediatric epidural analgesia. Acta Anaesthesiol Scand 2009;53:1084–7.
[21] Chorney JM, Tan ET, Martin SR, Fortier MA, Kain ZN. Children's behavior in the postanesthesia care unit: the development of the child behavior coding system-PACU (CBCS-P). J Pediatr Psychol 2011;37:338–47.
[22] Cochrane Collaboration. Data collection form for intervention reviews: RCTs and non-RCTs. The Cochrane Collaboration, 2014. http://www.cochrane.org/search/site/Data%20collection%20form%20Intervention%20review%20%E2%80%93%20RCTs%20and%20non-RCTs. Accessed 8 September 2014.
[23] Cohen LL, La Greca AM, Blount RL, Kazak AE, Holmbeck GN, Lemanek KL. Introduction to special issue: evidence-based assessment in pediatric psychology. J Pediatr Psychol 2008;33:911–15.
[24] Craig KD, Whitfield MF, Grunau RV, Linton J, Hadjistavropoulos HD. Pain in the preterm neonate: behavioural and physiological indices. PAIN 1993;52:287–99.
[25] Crellin D, Babl FE, Sullivan TP, Cheng J, O'Sullivan R, Hutchinson A. Procedural restraint use in preverbal and early-verbal children. Pediatr Emerg Care 2011;27:622–7.
[26] Crellin D, Santamaria N, Babl F, Harrison D. A systematic review of the Faces Legs Activity Cry Consolability (FLACC) pain scale for assessment of pain in children. 2014. PROSPERO. International Prospective Register of Systematic Reviews. Available at: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014014296. Accessed 17 October 2014.
[27] Crellin D, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. Paediatr Anaesth 2007;17:720–33.

[28] Crichton A, Knight S, Oakley E, Babl FE, Anderson V. Fatigue in child chronic health conditions: a systematic review of assessment instruments. Pediatrics 2015;135:e1015–31.
[29] Curry DM, Brown C, Wrona S. Effectiveness of oral sucrose for pain management in infants during immunizations. Pain Manag Nurs 2012; 13:139–49.
[30] Curtis SJ, Jou H, Ali S, Vandermeer B, Klassen T. A randomized controlled trial of sucrose and/or pacifier as analgesia for infants receiving venipuncture in a pediatric emergency department. BMC Pediatr 2007;7:27.
[31] da Silva FC, Santos Thuler LC, de Leon-Casasola OA. Validity and reliability of two pain assessment tools in Brazilian children and adolescents. J Clin Nurs 2011;20:1842–8.
[32] da Silva FC, Thuler LCS. Cross-cultural adaptation and translation of two pain assessment tools in children and adolescents. J Pediatr 2008;84: 344–9.
[33] De Ruddere L, Goubert L, Stevens MA, Deveugele M, Craig KD, Crombez G. Health care professionals' reactions to patient pain: impact of knowledge about medical evidence and psychosocial influences. J Pain 2014;15:262–70.
[34] de Tovar C, von Baeyer CL, Wood C, Alibeu JP, Houfani M, Arvieux C. Postoperative self-report of pain in children: interscale agreement, response to analgesic, and preference for a faces scale and a visual analogue scale. Pain Res Manag 2010;15:163–8.
[35] Decruynaere C, Thonnard JL, Plaghki L. How many response levels do children distinguish on faces scales for pain assessment? Eur J Pain 2009;13:641–8.
[36] Dewhirst E, Fedel G, Raman V, Rice J, Barry ND, Jatana KR, Elmaraghy C, Merz M, Tobias JD. Pain management following myringotomy and tube placement: intranasal dexmedetomidine versus intranasal fentanyl. Int J Pediatr Otorhinolaryngol 2014;78:1090–4.
[37] Diao M, Li L, Cheng W. To drain or not to drain in Roux-en-Y hepatojejunostomy for children with choledochal cysts in the laparoscopic era: a prospective randomized study. J Pediatr Surg 2012;47:1485–9.
[38] Duhn LJ, Medves JM. A systematic integrative review of infant pain assessment tools. Adv Neonatal Care 2004;4:126–40.
[39] Ekman P, Friesen WV. Facial action coding system: A technique for the measurement of facial movement. Palo Alto: Consulting Psychologists Press, 1978.
[40] El-Sharkawi HFA, El-Housseiny AA, Aly AM. Effectiveness of new distraction technique on pain associated with injection of local anesthesia for children. Pediatr Dent 2012;34:e35–38.
[41] Elshammaa N, Chidambaran V, Housny W, Thomas J, Zhang X, Michael R. Ketamine as an adjunct to fentanyl improves postoperative analgesia and hastens discharge in children following tonsillectomy—a prospective, double-blinded, randomized study. Paediatr Anaesth 2011;21:1009–14.
[42] Fernandes ML, Pires KCC, Tiburcio MA, Gomez RS. Caudal bupivacaine supplemented with morphine or clonidine, or supplemented with morphine plus clonidine in children undergoing infra-umbilical urological and genital procedures: a prospective, randomized and double-blind study. J Anesth 2012;26:213–18.
[43] Fournier-Charrière E, Tourniaire B, Carbajal R, Cimerman P, Lassauge F, Ricard C, Reiter F, Turquin P, Lombart B, Letierce A. EVENDOL, a new behavioral pain scale for children ages 0 to 7 years in the emergency department: design and validation. PAIN 2012;153: 1573–82.
[44] Frawley GP, Downie S, Huang GH. Levobupivacaine caudal anesthesia in children: a randomized double-blind comparison with bupivacaine. Paediatr Anaesth 2006;16:754–60.
[45] Ghai B, Ram J, Makkar JK, Wig J, Kaushik S. Subtenon block compared to intravenous fentanyl for perioperative analgesia in pediatric cataract surgery. Anesth Analg 2009;108:1132–8.
[46] Gilbert CA, Lilley CM, Craig KD, McGrath PJ, Court CA, Bennett SM, Montgomery CJ. Postoperative pain expression in preschool children: validation of the child facial coding system. Clin J Pain 1999;15: 192–200.
[47] Gomez RJ, Barrowman N, Elia S, Manias E, Royle J, Harrison D. Establishing intra- and inter-rater agreement of the face, legs, activity, cry, consolability scale for evaluating pain in toddlers during immunization. Pain Res Manag 2013;18:e124–8.
[48] Grove GL, Zerweck CR, Ekholm BP, Smith GE, Koski NI. Randomized comparison of a silicone tape and a paper tape for gentleness in healthy children. J Wound Ostomy Continence Nurs 2014;41:40–8.
[49] Hall NJ, Pacilli M, Eaton S, Reblock K, Gaines BA, Pastor A, Langer JC, Koivusalo AI, Pakarinen MP, Stroedter L, Beyerlein S, Haddad M, Clarke S, Ford H, Pierro A. Recovery after open versus laparoscopic pyloromyotomy for pyloric stenosis: a double-blind multicentre randomised controlled trial. Lancet 2009;373:390–8.
[50] Hamers JP, Huijer Abu-Saad H, Geisler FE, van den Hout MA, Schouten HJ, Halfens RJ, van Suijlekom HA. The effect of paracetamol, fentanyl, and systematic assessments on children's pain after tonsillectomy and adenoidectomy. J Perianesth Nurs 1999;14:357–66.
[51] Harrison D, Sampson M, Reszel J, Abdulla K, Barrowman N, Cumber J, Fuller A, Li C, Nicholls S, Pound CM. Too many crying babies: a systematic review of pain management practices during immunizations on YouTube. BMC Pediatr 2014;14:134.
[52] Hartrick CT, Kovan JP. Pain assessment following general anesthesia using the Toddler Preschooler Postoperative Pain Scale: a comparative study. J Clin Anesth 2002;14:411–15.
[53] Hippard HK, Govindan K, Friedman EM, Sulek M, Giannoni C, Larrier D, Minard CG, Watcha MF. Postoperative analgesic and behavioral effects of intranasal fentanyl, intravenous morphine, and intramuscular morphine in pediatric patients undergoing bilateral myringotomy and placement of ventilating tubes. Anesth Analg 2012;115:356–63.
[54] Hong JY, Han S, Kim W, Kim E, Kil H. Effect of dexamethasone in combination with caudal analgesia on postoperative pain control in day-case paediatric orchiopexy. Br J Anaesth 2010;105:506–10.
[55] Hong JY, Lee I, Shin S, Park E, Ban S, Cho J, Kil H. Caudal midazolam does not affect sevoflurane requirements and recovery in pediatric day-case hernioplasty. Acta Anaesthesiol Scand 2008;52:1411–14.
[56] Hughes J, Lindup M, Wright S, Naik M, Dhesi R, Howard R, Sommerlad B, Kangesu L, Sury M. Does nasogastric feeding reduce distress after cleft palate repair in infants? Nurs Child Young People 2013;25:26–30.
[57] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1–12.
[58] Jindal P, Khurana G, Dvivedi S, Sharma JP. Intra and postoperative outcome of adding clonidine to bupivacaine in infraorbital nerve block for young children undergoing cleft lip surgery. Saudi J Anaesth 2011;5: 289–94.
[59] Johansson M, Kokinsky E. The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. Nurs Crit Care 2009; 14:122–30.
[60] 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;52:201–8.
[61] Johnston CC, Stevens B, Yang F, Horton L. Developmental changes in response to heelstick in preterm infants: a prospective cohort study. Dev Med Child Neurol 1996;38:438–45.
[62] Jonnavithula N, Durga P, Kulkarni DK, Ramachandran G. Bilateral intra-oral, infra-orbital nerve block for postoperative analgesia following cleft lip repair in paediatric patients: comparison of bupivacaine vs bupivacaine-pethidine combination. Anaesthesia 2007;62:581–5.
[63] Jonnavithula N, Durga P, Madduri V, Ramachandran G, Nuvvula R, Srikanth R, Damalcheruvu MR. Efficacy of palatal block for analgesia following palatoplasty in children with cleft palate. Paediatr Anaesth 2010;20:727–33.
[64] Kil HK, Kim WO, Han SW, Kwon Y, Lee A, Hong JY. Psychological and behavioral effects of chloral hydrate in day-case pediatric surgery: a randomized, observer-blinded study. J Pediatr Surg 2012;47:1592–9.
[65] Kim CH, Yoon JU, Lee HJ, Shin SW, Yoon JY, Byeon GJ. Availability of a 5% lidocaine patch used prophylactically for venipuncture- or injection-related pain in children. J Anesth 2012;26:552–5.
[66] Kim NY, Kim SY, Yoon HJ, Kil HK. Effect of dexmedetomidine on sevoflurane requirements and emergence agitation in children undergoing ambulatory surgery. Yonsei Med J 2014;55:209–15.
[67] Kundu A, Lin Y, Oron AP, Doorenbos AZ. Reiki therapy for postoperative oral pain in pediatric patients: pilot data from a double-blind, randomized clinical trial. Complement Ther Clin Pract 2014;20:21–5.
[68] Loetwiriyakul W, Asampinwat T. Caudal block with 3 mg/Kg Bupivacaine for intraabdominal surgery in pediatric patients: a randomized study. Asian Biomed 2011;5:93–9.
[69] Lorenzo AJ, Lynch J, Matava C, El-Beheiry H, Hayes J. Ultrasound guided transversus abdominis plane vs surgeon administered intraoperative regional field infiltration with bupivacaine for early postoperative pain control in children undergoing open pyeloplasty. J Urol 2014;192:207–13.
[70] Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol 2010;63:854–61.
[71] Malviya S, Voepel-Lewis T, Burke C, Merkel S, Tait AR. The revised FLACC observational pain tool: improved reliability and validity for pain assessment in children with cognitive impairment. Paediatr Anaesth 2006;16:258–65.

[72] Manworren RCB, Hynan LS. Clinical validation of FLACC: preverbal patient pain scale. Pediatr Nurs 2003;29:140–6.
[73] McGrath PJ, Johnson G, Goodman JT, Schillinger J, Dunn J, Chapman J. CHEOPS: a behavioural scale for rating postoperative pain in children. In: Fields HL, Dubner R, Cervero F, editors. Advances in Pain Research and Therapy, Volume 9. New York: Raven Press, 1985. p. 395–402.
[74] Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S. The FLACC: a behavioral scale for scoring postoperative pain in young children. Pediatr Nurs 1997;23:293–7.
[75] Miller K, Rodger S, Bucolo S, Greer R, Kimble RM. Multi-modal distraction. Using technology to combat pain in young children with burn injuries. Burns 2010;36:647–58.
[76] Miller K, Rodger S, Kipping B, Kimble R. A novel technology approach to pain management in children with burns: a prospective randomized controlled trial. Burns 2011;37:395–405.
[77] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.
[78] Morison SJ, Grunau RE, Oberlander TF, Whitfield MF. Relations between behavioral and cardiac autonomic reactivity to acute pain in preterm neonates. Clin J Pain 2001;17:350–8.
[79] Morison SJ, Holsti L, Grunau RE, Whitfield MF, Oberlander TF, Chan HW, Williams L. Are there developmentally distinct motor indicators of pain in preterm infants? Early Hum Dev 2003;72:131–46.
[80] Natarajan Surendar M, Kumar Pandey R, Kumar Saksena A, Kumar R, Chandra G. A comparative evaluation of intranasal dexmedetomidine, midazolam and ketamine for their sedative and analgesic properties: a triple blind randomized study. J Clin Pediatr Dent 2014;38:255–61.
[81] Newbury C, Herd DW. Amethocaine versus EMLA for successful intravenous cannulation in a children's emergency department: a randomised controlled study. Emerg Med J 2009;26:487–91.
[82] Nilsson S, Enskär K, Hallqvist C, Kokinsky E. Active and passive distraction in children undergoing wound dressings. J Pediatr Nurs 2013;28:158–66.
[83] Nilsson S, Finnstrom B, Kokinsky E. The FLACC behavioral scale for procedural pain assessment in children aged 5-16 years. Paediatr Anaesth 2008;18:767–74.
[84] Nilsson S, Kokinsky E, Nilsson U, Sidenvall B, Enskar K. School-aged children's experiences of postoperative music medicine on pain, distress, and anxiety. Paediatr Anaesth 2009;19:1184–90.
[85] Nord D, Belew J. Effectiveness of the essential oils lavender and ginger in promoting children's comfort in a perianesthesia setting. J Perianesthesia Nurs 2009;24:307–12.
[86] Ramelet AS, Rees N, McDonald S, Bulsara M, Abu-Saad HH. Development and preliminary psychometric testing of the Multidimensional Assessment of Pain Scale: MAPS. Paediatr Anaesth 2007;17:333–40.
[87] Ramelet AS, Rees NW, McDonald S, Bulsara MK, Huijer Abu-Saad H. Clinical validation of the multidimensional assessment of pain scale. Paediatr Anaesth 2007;17:1156–65.
[88] Ranger M, Celeste Johnston C, Rennick JE, Limperopoulos C, Heldt T, du Plessis AJ. A multidimensional approach to pain assessment in critically Ill infants during a painful procedure. Clin J Pain 2013;29:613–20.
[89] Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol Assess 2007;11:iii, ix–51.
[90] Saha N, Saha DK, Rahman MA, Islam MK, Aziz MA. Comparison of post operative morbidity between laparoscopic and open appendectomy in children. Mymensingh Med J 2010;19:348–52.
[91] Sethi S, Ghai B, Ram J, Wig J. Postoperative emergence delirium in pediatric patients undergoing cataract surgery—a comparison of desflurane and sevoflurane. Paediatr Anaesth 2013;23:1131–7.
[92] Singh R, Kharbanda M, Sood N, Mahajan V, Chatterji C. Comparative evaluation of incidence of emergence agitation and post-operative recovery profile in paediatric patients after isoflurane, sevoflurane and desflurane anaesthesia. Indian J Anaesth 2012;56:156–61.
[93] Stanford EA, Chambers CT, Craig KD. The role of developmental factors in predicting young children's use of a self-report scale for pain. PAIN 2006;120(1-2):16–23.
[94] Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford University Press: Oxford, United Kingdom, 2008.
[95] Stuth EA, Berens RJ, Staudt S, Robertson FA, Scott JP, Stucke AG, Hoffman GM, Troshynski TJ, Tweddell JS, Zuperku EJ. The effect of caudal vs intravenous morphine on early extubation and postoperative analgesic requirements for stage 2 and 3 single-ventricle palliation: a double blind randomized trial. Pediatr Anesth 2011;21:441–53.
[96] Suraseranivongse S, Kraiprasit K, Petcharatana S. Postoperative pain assessment in ambulatory pediatric patients by parents. J Med Assoc Thai 2002;85(suppl 3):S917–922.
[97] Suraseranivongse S, Santawat U, Kraiprasit K, Petcharatana S, Prakkamodom S, Muntraporn N. Cross-validation of a composite pain scale for preschool children within 24 hours of surgery. Br J Anaesth 2001;87:400–5.
[98] Taddio A, Hogan ME, Moyer P, Girgis A, Gerges S, Wang L, Ipp M. Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. Vaccine 2011;29:1390–4.
[99] Takmaz SA, Uysal HY, Uysal A, Kocer U, Dikmen B, Baltaci B. Bilateral extraoral, infraorbital nerve block for postoperative pain relief after cleft lip repair in pediatric patients: a randomized, double-blind controlled study. Ann Plast Surg 2009;63:59–62.
[100] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7.
[101] Townsend JA, Ganzberg S, Thikkurissy S. The effect of local anesthetic on quality of recovery characteristics following dental rehabilitation under general anesthesia in children. Anesth Prog 2009;56:115–22.
[102] Tsze DS, von Baeyer CL, Bulloch B, Dayan PS. Validation of Self-Report Pain Scales in Children. Pediatrics 2013;132:e971–9.
[103] van Tulder MP, Furlan AMD, Bombardier CMDF, Bouter LP; The Editorial Board of the Cochrane Collaboration Back Review G. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine 2003;28:1290–9.
[104] Vaughan M, Paton EA, Bush A, Pershad J. Does lidocaine gel alleviate the pain of bladder catheterization in young children? A randomized, controlled trial. Pediatrics 2005;116:917–20.
[105] Voepel-Lewis T, Malviya S, Tait AR. Validity of parent ratings as proxy measures of pain in children with cognitive impairment. Pain Manag Nurs 2005;6:168–74.
[106] Voepel-Lewis T, Malviya S, Tait AR, Merkel S, Foster R, Krane EJ, Davis PJ. A comparison of the clinical utility of pain assessment tools for children with cognitive impairment. Anesth Analg 2008;106:72–8, table of contents.
[107] Voepel-Lewis T, Merkel S, Tait AR, Trzcinka A, Malviya S. The reliability and validity of the Face, Legs, Activity, Cry, Consolability behavioral tool as a measure of pain in children with cognitive impairment. Anesth Analg 2002;95:1224–9.
[108] Voepel-Lewis T, Zanotti J, Dammeyer JA, Merkel S. Reliability and validity of the face, legs, activity, cry, consolability behavioral tool in assessing acute pain in critically ill patients. Am J Crit Care 2010;19:55–61; quiz 62.
[109] Voepel-Lewis TD, Malviya S, Burke C, D'Agostino R, Hadden SM, Siewert M, Tait AR. Evaluation of simethicone for the treatment of postoperative abdominal discomfort in infants. J Clin Anesth 1998;10:91–4.
[110] von Baeyer CL, Spagrud LJ. Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. PAIN 2007;127:150.
[111] Williams AL, Khattak AZ, Garza CN, Lasky RE. The behavioral pain response to heelstick in preterm neonates studied longitudinally: description, development, determinants, and components. Early Hum Dev 2009;85:369–74.
[112] Willis MHW, Merkel SI, Voepel-Lewis T, Malviya S. FLACC behavioural pain assessment scale: a comparison with a child's self-report. Pediatr Nurs 2003;29:195–8.
[113] Zhou H, Roberts P, Horgan L. Association between self-report pain ratings of child and parent, child and nurse and parent and nurse dyads: meta-analysis. J Adv Nurs 2008;63:334–42.
[114] Zier JL, Rivard PF, Krach LE, Wendorf HR. Effectiveness of sedation using nitrous oxide compared with enteral midazolam for botulinum toxin A injections in children. Dev Med Child Neurol 2008;50:854–8.

6.1 Additional material

The supplemental tables online that are referred to in this paper are reproduced in Appendix D of the thesis.

6.2 Addendum: New literature

An updated search was completed in the last week of May 2018 to identify studies published since the original search was completed in Aug 2014. Two psychometric studies have been conducted since the 2015 systematic review of the FLACC; one of which involved testing a translated version of the FLACC scale (333) and would have been deemed ineligible for inclusion and the other tested reliability when used to assess pain associated with burns dressings (334). Shen and colleagues' study increased our understanding of the factors that impact on the accuracy of FLACC scores when used to assess procedural pain. However, this study did not contribute to an assessment of the psychometric properties of the scale used for procedural pain assessment. The third psychometric evaluation study published since this systematic review was completed in an emergency department but did not assess procedural pain and therefore made no contribution to an assessment of the FLACC scale to assess procedural pain (457).

Nineteen RCTs and one systematic review with meta-analysis used the FLACC scale to assess procedural pain have also been published since this systematic review. These studies include participants aged from 1 month to 17 years and most focus on treatments aimed at alleviating the pain associated with immunisation, venepuncture or IV cannula insertion and dental procedures. A published meta-analysis was excluded as scores derived using the FLACC scale were pooled with scores generated using other scales for analysis. The quality of the RCTs was examined and nine met the quality requirement for inclusion. Of these studies, four reported results that did not add evidence to support the capacity of the FLACC scale to differentiate between known groups. The results of the remaining five were conflicting: three studies provided moderate support for the scales capacity to differentiate between groups (458-460) while two studies provided negative evidence for the scales capacity to differentiate between groups (461, 462).

Two RCTs included methods to assess the inter-rater reliability of the scores allocated in the study. In both cases the methods used were assessed as 'fair' using the COSMIN Checklist and the results were expressed as kappa scores (0.79 (463) and 0.89 (462))

6.2.1 Implications

The studies that have been published since the systematic review do not add greatly to the data that was summarised in this systematic review. There are two studies that confirm the assessment of the reliability of scores when the FLACC scale is used to assess procedural pain. However, the sum of the RCTs do not support the validity of the scale for procedural pain assessment. The recommendations of the systematic review would not alter with the inclusion of more recently published studies.

CHAPTER 7.

A systematic review of the Modified Behavioural Pain Scale for assessing pain in infants and children

This chapter reports the results of a systematic review to summarise the data that describes the psychometric properties of the Modified Behavioral Pain Scale (MBPS). This work has been published and the PDF of this publication is reproduced in this chapter. A summary of the studies identified in a more recent search (May 2018) and their contribution to our understanding of the psychometric properties of the MBPS used to assess procedural pain concludes this chapter.

Publication:

Crellin DJ, Babl FE, Santamaria N, Harrison D. A Systematic Review of the Psychometric Properties of the Modified Behavioral Pain Scale (MBPS). Journal of Pediatric Nursing: Nursing Care of Children and Families. 2018;40:14-26.

Published abstract

Crellin DJ, Babl FE, Santamaria N, Harrison D. The MBPS: A systematic review to determine its role in assessing pain in infants and children. Paper presented at the International Symposium on Pediatric Pain, Kuala Lumpur. 2017.

Journal of Pediatric Nursing 40 (2018) 14–26

Contents lists available at ScienceDirect

Journal of Pediatric Nursing

A Systematic Review of the Psychometric Properties of the Modified Behavioral Pain Scale (MBPS)[1]

Dianne J. Crellin, MN [a,b,c,*], Franz E. Babl, MD [b,c,d], Nick Santamaria, PhD [a], Denise Harrison, PhD [a,b,e]

[a] Department of Nursing, The University of Melbourne, Melbourne, Australia
[b] Murdoch Children's Research Institute, Melbourne, Australia
[c] Royal Children's Hospital, Melbourne, Australia
[d] Department of Paediatrics, The University of Melbourne, Melbourne, Australia
[e] Children's Hospital of Eastern Ontario, University of Ottawa, Canada

ARTICLE INFO

Article history:
Received 30 August 2017
Revised 14 January 2018
Accepted 5 February 2018
Available online xxxx

Keywords:
MBPS/Modified Behavioral Pain Scale
Pain assessment
Infant
Child
Psychometric properties
Validation
Reliability

ABSTRACT

Problem: Clinicians and researchers require a valid way to assess procedural pain experienced by infants and children. The Modified Behavioral Pain Scale (MBPS) has been used to assess immunisation pain. However, it is unknown whether it is valid for this purpose and whether use can be extended to other procedures. The aim of this study was to rigorously assess evidence addressing the psychometric properties of this scale and to provide recommendations for its use.
Eligibility criteria: Psychometric evaluation studies reporting feasibility, reliability, validity, or utility data for the MBPS applied to children (birth to 18 years) and randomised controlled trials (RCT) using the MBPS were included.
Sample: Twenty-eight studies (8 psychometric and 20 RCTs) were included.
Results: Studies were of varying quality. Sufficient data was available to cautiously accept the MBPS as valid for assessing immunisation related pain in infants aged 2 to 22 months. There was insufficient data to support the psychometrics in other age groups or in circumstances other than immunisation. There is no data addressing the clinical utility of the MBPS.
Conclusions: It is not possible at this time to confidently accept the MBPS as suitable for assessing all procedural pain in young children.
Implications: Studies to evaluate the capacity of the MBPS to assess pain in a range of procedures and to distinguish between pain and non-pain related distress are needed if it is to be recommended.

Introduction

Multidimensional observational scales have been developed as a means to assess and quantify pain intensity in infants and children too young to provide self-report and there are now over 40 reported in the literature (Lee & Stevens, 2014). The Modified Behavioral Pain Scale (MBPS) was developed over 20 years ago at a time when few scales were available and those that were available were not considered sufficiently able to capture the variability in responses in young infants. For this reason, Taddio and her colleagues modified an existing scale developed for measuring postoperative pain in children one to five years of age, the Children's Hospital Eastern Ontario Pain Scale (CHEOPS) (McGrath et al., 1985). The aim was to develop a clinically applicable pain intensity measure for young infants and test the new scale's performance when used to measure pain intensity in infants aged four to six months receiving a routine immunisation (Taddio, Nulman, Koren, Stevens, & Koren, 1995).

The MBPS comprises a composite of three behaviours considered indicative of pain in infants; facial expression, cry and body movements. The role of facial expression for pain assessment is well supported in the literature and is included as an assessment item in most behavioral observation scales (Chang, Versloot, Fashler, McCrystal, & Craig, 2015; Prkachin, 2009). Similarly, cry and various cry characteristics, such as latency to cry and cry duration are considered manifestations of pain in infants and young children and also feature in many observation scales (Chorney MacLaren & McMurtry, 2014). Finally, specific patterns in body movement have been shown to exist in infants and children exposed to painful stimuli (Craig, Whitfield, Grunau, Linton, & Hadjistavropoulos, 1993). Each of the behaviours included in the

1 Registration details: International Prospective Register of Systematic Reviews (CRD42016041722).

* Corresponding author at: Royal Children's Hospital, The University of Melbourne, c/o Emergency Department, Royal Children's Hospital, Flemington Road, Parkville, 3052, Australia.
E-mail address: dianne.crellin@rch.org.au (D.J. Crellin).

https://doi.org/10.1016/j.pedn.2018.02.005
0882-5963/

MBPS is assessed, scored and the item scores added to generate a pain intensity score from zero to ten. The items, their descriptors and the associated scores for each item descriptor for the MPBS are shown in Table 1.

The reliability (internal consistency, inter-rater and test retest reliability) and concurrent and construct validity of the new scale were tested in a randomised controlled trial (RCT) evaluating the efficacy of topical Eutectic Mixture of Local Anaesthetic (EMLA) cream to reduce immunisation pain in infants aged four to six months (Taddio et al., 1995). Observers independently scored participants five seconds before the immunisation and 15 s after the injection. The authors concluded that the MBPS showed high levels of internal consistency and near perfect inter- and intra-rater reliability. The validity of the scale was supported by positive correlations between MBPS scores and observer Visual Analogue Scale (VAS) scores, the capacity of the scale to detect increases in pain during the vaccination and the difference in the amount of pain experienced by infants who received their immunisation with or without the application of EMLA.

The importance of assessing and managing procedural pain has gained increasing attention and the potential that infants and children may demonstrate different behaviours, complicated by fear and non-pain related distress that often accompanies procedures, to those exhibited when experiencing post-operative pain has also been recognised (Blount, Piira, Cohen, & Cheng, 2006; von Baeyer & Spagrud, 2007). The MBPS has been cited as a valid procedural pain assessment tool on the basis of the results of Taddio's original study (Taddio et al., 1995) and a subsequent study testing the measurement properties of the scale (Taddio et al., 2011). In their 2011 study, Taddio and her colleagues evaluated the psychometric properties of three scales, including the MBPS, used to assess immunisation related pain in 120 infants aged 2 to 6 months. Participating infants were randomised to two different vaccines and pain responses during vaccination were evaluated. Reviewers applied the scales to video segments. Inter-rater reliability was high (intraclass correlations 0.90 and 0.94), scores were higher during the painful phase ($p < 0.001$) and differed between groups ($p < 0.001$). However as data from all studies addressing the psychometric properties of the scale have yet to be synthesised and evaluated, the weight of evidence supporting this contention is unclear and the circumstances and population to which the scale can be reasonably applied is similarly unclear. Addressing this at a time when the validity of the scale most often recommended for this purpose, the Face Legs, Activity, Cry and Consolability (FLACC) scale (Crellin, Sullivan, Babl, O'Sullivan, & Hutchinson, 2007; von Baeyer & Spagrud, 2007), has recently been challenged (Crellin, Harrison, Santamaria, & Babl, 2015) is an important step towards providing clinicians and researchers with suitable means by which to assess procedural pain experienced by infants and young children.

Table 1
Modified Behavioral Pain Scale (MBPS) (Taddio et al., 1995).

Item	Descriptor	Score
Facial expression	Definite positive expression (smiling)	0
	Neutral expression	1
	Slightly negative expression (grimace)	2
	Definite negative expression (furrowed brow eyes closed tightly)	3
Cry	Laughing or giggling	0
	Not crying	1
	Moaning quiet vocalizing gentle or whimpering cry	2
	Full lunged cry or sobbing	3
	Full lunged cry more than baseline cry (scored only if child crying at baseline)	4
Movements	Usual movements and activity	0
	Resting and relaxed	0
	Partial movement (squirming arching limb tensing clenching)	2
	Attempt to avoid pain by withdrawing the limb where puncture is done	2
	Agitation with complex/generalized movements involving the head torso or other limbs	3
	Rigidity	3

Note: Item scores are summed to generate a total score out of 10.

Review Aim

The aim of this systematic review was to provide the first attempt to summarise the psychometric and practical properties of the MBPS used to assess pain in children aged from birth to 18 years. The specific objectives were to 1) identify and describe the populations and circumstances for which psychometric data are available, 2) systematically review the quality of these data, 3) analyse and summarise the strength of evidence that support the psychometric and practical properties of the MBPS and provide recommendations to guide clinical and research use of this scale.

Materials and Methods

A systematic review was conducted to identify and appraise the evidence for the psychometric and practical (feasibility) properties of the MBPS using a protocol developed by the authors and based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA statement). The protocol was registered with the International Prospective Register of Systematic Reviews (CRD42016041722) and is available in full text on the PROSPERO Web Site.

Inclusion/Exclusion Criteria

Studies reporting reliability, validity, feasibility or clinical utility data for the MBPS applied to infants and children aged two months to 18 years were included in this review. This included studies where the aim of the study was to examine the psychometrics of the MBPS, compare the MBPS with alternative scales or assessment tools or establish the psychometrics of an alternative scale or assessment tool using the MBPS as a reference. The review also included RCTs using MBPS as an outcome measure as a potential source of construct validity (i.e. the capacity of the scale to detect a difference between known or extreme groups). The authors of the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) Checklist, which is described in a following section, recommend excluding RCTs from a systematic review. They were included in this review for several reasons; it is a common validation methodology used by pain researchers and was used by Taddio and colleagues in their original and follow-up studies (Taddio et al., 2011; Taddio, Nulman, Goldbach, Ipp, & Koren, 1994) and it has also been used in two previous systematic reviews of the psychometrics of pain scales (Crellin et al., 2015; Stevens, Johnston, Taddio, Gibbins, & Yamada, 2010). Furthermore, this is a robust method that may potentially contribute validation evidence providing the results are interpreted appropriately.

Studies that did not report or analyse MBPS scores separately, did not include children or report their data separately, were only published as an abstract or were not available in English were excluded from this review. Finally, studies using or evaluating a modified version of the MBPS (including translated versions) were also excluded. Following quality assessment, RCTs with Jadad scores less than three were excluded as they were considered to be at risk of significant bias and therefore unlikely to contribute significant evidence to this review (Jadad et al., 1996).

Search Strategy

The search terms 'MBPS' and 'Modified Behavioral Pain Scale' were used to search the following databases; MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews and Cochrane Controlled Trials, Cumulative Index Nursing and Allied Health Literature (CINAHL)

and PsycINFO using the Ovid, PubMed and EBSCOhost platforms. Google Scholar was searched using all search terms and the operator 'AND' and the reference lists of the included studies and identified reviews were also searched. The date range used for all searches was 1990 to 31st July 2016 to ensure that studies using the MBPS and published prior to the original psychometric evaluation study were not overlooked. The search was limited to texts available in English. The search strategy was informed and reviewed by a librarian and was conducted by the first author.

Study Selection

Following removal of duplicates, abstracts were reviewed by two independent reviewers (DC and one of NS, FB, or DH). Where eligibility was ambiguous the full text article was reviewed. A third reviewer was used to reach consensus where study eligibility remained unclear.

Data Extraction

Two data extraction tools were modified for use; the 'Quality Appraisal of Diagnostic Reliability (QAREL) data extraction form' (used for the psychometric studies) and the 'Cochrane Collaboration data collection tool for intervention studies' (used for the RCTs) (Cochrane Collaboration, 2014; Lucas, Macaskill, Irwig, & Bogduk, 2010). The QUAREL data extraction form was developed to support data extraction from studies assessing the reliability of diagnostic tests (Lucas et al., 2010). It is comprised of 11 items that address seven principles which include; demographics of subjects and the reviewers, reviewer blinding, the timing of assessments etc. The modifications to both forms included the deletion of irrelevant fields and the addition of fields relevant to this review.

The data extracted included; participant details (e.g. numbers, demographics), setting and circumstances of the pain being measured (associated with disease, operative or procedural), scale description and application (e.g. modifications), study methods (design, psychometric properties evaluated, and statistical methods), sources of bias and study results.

Data were extracted by one reviewer (DC) and checked and confirmed by a second reviewer (FB, DH or NS). A third reviewer completed data extraction independently to resolve any inconsistencies between the first two reviewers.

Quality Assessment

Quality assessment tools were applied independently by two reviewers (DC and one of NS, FB, or DH) and a third where agreement was not achieved by the first two reviewers. The methodological quality of the psychometric evaluation studies were appraised using the COSMIN checklist (Terwee et al., 2007), and the RCTs were appraised using the Jadad score (Jadad et al., 1996). The COSMIN checklist was also used to assess the methods of an RCT where other psychometric properties, such as reliability and responsiveness, were also assessed.

The COSMIN checklist, has been used to assess the quality of studies in a recently published review assessing the psychometric properties of the Face, Legs, Activity, Cry and Consolability scale (Crellin et al., 2015). The checklist provides standards for study design, statistical methods and acceptable outcome values and rates items on a 4-point scale as 'poor', 'fair', 'good' or 'excellent'. The checklist is comprised of four steps: 1. Identification of the measurement properties (see Table 2 for measurement properties), 2. Assessment of the item response theory methods applied, 3. Evaluation of the quality of the methods used to assess the measurement properties identified in step 1 and 4. Assessment of the generalisability of the results. The measurement properties are assessed against criteria and the lowest rating forms the final score for that measurement property. A study may receive different assessments for the methods addressing different psychometric properties. The COSMIN taxonomy and the terms commonly used in pain scale evaluation studies are defined in Table 2 (Crellin et al., 2015). A full description of the COSMIN checklist and application of scoring system can be found on the COSMIN website (http://www.cosmin.nl/the_cosmin_checklist.html).

The Jadad Scale requires assessment and scoring of the quality of blinding, randomisation and participant follow-up. The items are scored to contribute to a total score ranging from 0 to 5, where 5 is a perfect score. This scale has been used in two previous pain scale reviews to assess the quality of identified RCTs. Each item contributes to a total score out of five, where five is a perfect score. High scoring studies will provide higher levels of evidence of the scales capacity to distinguish between known groups than low scoring studies. In this review we made a minor modification to the definition for participant follow-up and scored this as acceptable if, in the absence of an explicit statement 'there were no withdrawals from this study' there was sufficient detail in the results to account for all study participants.

It was intended that data reporting the feasibility and the clinical utility of scale application in clinical practice would also be presented in this review. A tool to support the assessment of the quality of the methods used to evaluate the practicality or feasibility or the clinical utility of the scale has not been identified. For this reason a pragmatic and descriptive approach to assessing these studies will be adopted.

Data Synthesis

The PRISMA flow chart was used to present the results of the search and study selection (Moher, Liberati, Tetzlaff, & Altman, 2009).

Table 2
Pain scale validation strategies and COSMIN taxonomy (Terwee et al., 2007).

COSMIN measurement property	Pain scale measurement property	Pain scale evaluation study method
Internal consistency	Internal consistency	Correlations between items on the scale
Reliability	Interrater reliability	Correlation between pain scores provided simultaneously by independent reviewers
	Intrarater reliability	Correlation between scores allocated by a single reviewer to the same episode of pain on separate occasions (achieved using video taped segments for repeat review)
Measurement error		Rarely tested in pain scale evaluation studies
Content validity	Content validity	Relevance and comprehensiveness of the items tested with experts
Structural validity		Principal component analysis, exploratory factor analysis (used for new scale development) and confirmatory factor analysis (validation of scale)
Hypothesis testing	Convergent validity	Correlation with assessments using other pain assessment tools/scales - observational scale
	Discriminant validity	Correlation with other unrelated constructs (e.g.: pain and hunger)
	Construct vali1dity	Extreme or known groups comparison - correlation between groups undergoing different procedures or treatments
Cross cultural validity	Cross cultural validity	Translation – backwards and forwards, content review for cultural appropriateness
Criterion validity	Concurrent validity	Correlation with assessments using the gold standard (other valid tools/scales and self-report)
Responsiveness	Responsiveness	Change over time where change expected e.g.: before and after analgesic or pain producing procedure
Interpretability	Clinical utility	Methods to determine value of score for clinical decision making

Accepted standards for interpreting the strength of results and clinical judgement to establish consensus agreed thresholds were used to classify the strength of the result and the weight of evidence (Shrout & Fleiss, 1979; Streiner & Norman, 2008; Tomlinson, von Baeyer, Stinson, & Sung, 2010). The weight of evidence derived from each study was estimated based on the strength of the results and the quality of the methods used in the study i.e. methods rated as either 'poor' or 'fair' weakened confidence in the results and therefore the strength of the evidence that they could be considered to provide. 'Nil/none', 'low', 'moderate' and 'high' were used to describe the level of evidence. In some circumstances, results were considered negative evidence, e.g. correlation <0.4; in other words evidence that the MBPS is not reliable or valid. The results of other pain related variables e.g. other pain scores, analgesic consumption etc. measured in the RCTs also contributed to the assessment of the weight of evidence for validation provided by the RCT results. Where these results were consistent with those generated using the MBPS they helped to support the validity of the scale and where they were inconsistent, the RCT results were not considered evidence of MBPS score validation.

A narrative synthesis of the results from the studies included in this review was completed as the heterogeneity of the eligible studies meant that pooling of data for meta-analysis was not possible. Studies were grouped and analysed according to the age of the participants (e.g.: 4 to 6 months, the original cohort, or older than 6 months or younger than 4 months), the circumstances of the pain (e.g.: immunisation or non-immunisation related pain) and scale modifications.

The evaluation criteria for Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) reviews (Cohen et al., 2008) provide a framework to define the strength of the evidence base supporting the psychometrics of a scale or measure on the basis of the number, results and independence of the evaluation studies. Assessment of study quality is not included in the criteria which is a significant limitation to this framework. In our review it was used to assess the strength of the results of each study as described. The principles used to underpin the IMMPACT assessment criteria are used here to guide our synthesis. Our modifications to these criteria are shown in Table 3. The most significant of these is the need for multiple validation methods to confirm validity based on our increasing understanding that validity is indirectly assessed and best achieved using multiple approaches or validation methods (Streiner & Kottner, 2014).

Results

Twenty-eight papers (8 psychometric evaluation studies and 20 RCTs) were assessed as eligible for inclusion in this systematic review (Abuelkheir et al., 2014; Anninger, Forbes, Quinn, & Schreiner, 2007; Carbajal et al., 2008; Cohen, Bernard, McClelland, & MacLaren, 2005; Dyer, Collins, Baghurst, Saxon, & Meachan, 2004; Fallah, Gholami, Ferdosian, & Binesh, 2016; Girish & Ravi, 2014; Hillgrove-Stuart, Riddell, Horton, & Greenberg, 2013; Hogan et al., 2014; Ipp, Cohen, Goldbach, & Macarthur, 2004; Ipp, Parkin, Lear, Goldbach, & Taddio, 2009; Ipp, Taddio, Sam, Gladbach, & Parkin, 2007; Kassab, Sheehy, King, Fowler, & Foureur, 2012; Lindh, Wiklund, Blomquist, & Hakansson, 2003; McClellan, Cohen, & Joseph, 2003; McGowan, Cottrell, Roberts, & Lankshear, 2013; Mijovic et al., 2010; Mularoni et al., 2009; O'Brien, Taddio, Ipp, Goldbach, & Koren, 2004; Silva et al., 2010; Taddio et al., 1994; Taddio et al., 1995; Taddio et al., 2009; Taddio et al., 2011; Taddio et al., 2014; Taddio et al., 2015; Verriotis et al., 2015). As shown in Fig. 1, 247 citations identified in the search were excluded, of which 194 were deemed irrelevant, 32 did not meet inclusion criteria and the remaining 21 met exclusion criteria. Another 30 citations were excluded following full text review; 20 as they did not meet inclusion criteria and 12 as they met exclusion criteria. The details of the search results are documented in Fig. 1. The original evaluation study was an RCT designed to determine the efficacy of topical anaesthetic for reducing immunisation related pain. As this study also included methods designed to evaluate the psychometric properties of the scale, it was counted and assessed as a psychometric evaluation study and the quality of the RCT methods was assessed using the Jadad tool.

Study and Patient Characteristics

Psychometric Evaluation Studies

Eight psychometric evaluation studies were included in this review (Cohen et al., 2005; McClellan et al., 2003; Mijovic et al., 2010; Silva et al., 2010; Taddio et al., 1995; Taddio et al., 2009; Taddio et al., 2011; Verriotis et al., 2015). Three studies (including the original) focused on immunisation pain in infants and/or young children (McClellan et al., 2003; Taddio et al., 1995; Taddio et al., 2011) and five studies used the MBPS as the reference scale to evaluate the psychometric properties of another scale or pain assessment tool in infants of varying ages (Cohen et al., 2005; Mijovic et al., 2010; Silva et al., 2010; Taddio et al., 2009; Verriotis et al., 2015). Immunisation-related pain was the focus in all but two studies, which both focused on venous blood sampling (Mijovic et al., 2010; Silva et al., 2010). Tables 4 to 7 provide a summary of these studies and more details can be found in Appendix A.

Randomised Controlled Trials

Twenty RCTs were eligible for inclusion in this review (Abuelkheir et al., 2014; Anninger et al., 2007; Carbajal et al., 2008; Dyer et al., 2004; Fallah et al., 2016; Girish & Ravi, 2014; Hillgrove-Stuart et al., 2013;

Table 3
(Modified) IMMPACT criteria for establishing the weight of evidence supporting reliability and validity.

Assessment	Criteria
I. A well-established assessment	High level evidence from: a. Minimum 2 peer review publications from different investigators/investigatory teams AND b. Minimum 2 different validation methods used e.g. concurrent, convergent, discriminant or responsiveness validation AND Sufficient detail about the measure to allow critical evaluation and replication.
II. Approaching well-established assessment	High/moderate level evidence from: c. Minimum 2 peer review publications from different investigators/investigatory teams AND d. Minimum 2 different validation methods used e.g. concurrent, convergent, discriminant or responsiveness validation AND Sufficient detail about the measure to allow critical evaluation and replication.
III. Promising assessment	High/moderate level evidence from: e. Minimum 1 peer review publications from different investigators/investigatory teams AND f. Minimum 2 different validation methods used e.g. concurrent, convergent, discriminant or responsiveness validation AND Sufficient detail about the measure to allow critical evaluation and replication.

Note: Modified from the IMMPACT criteria (Cohen et al., 2008).

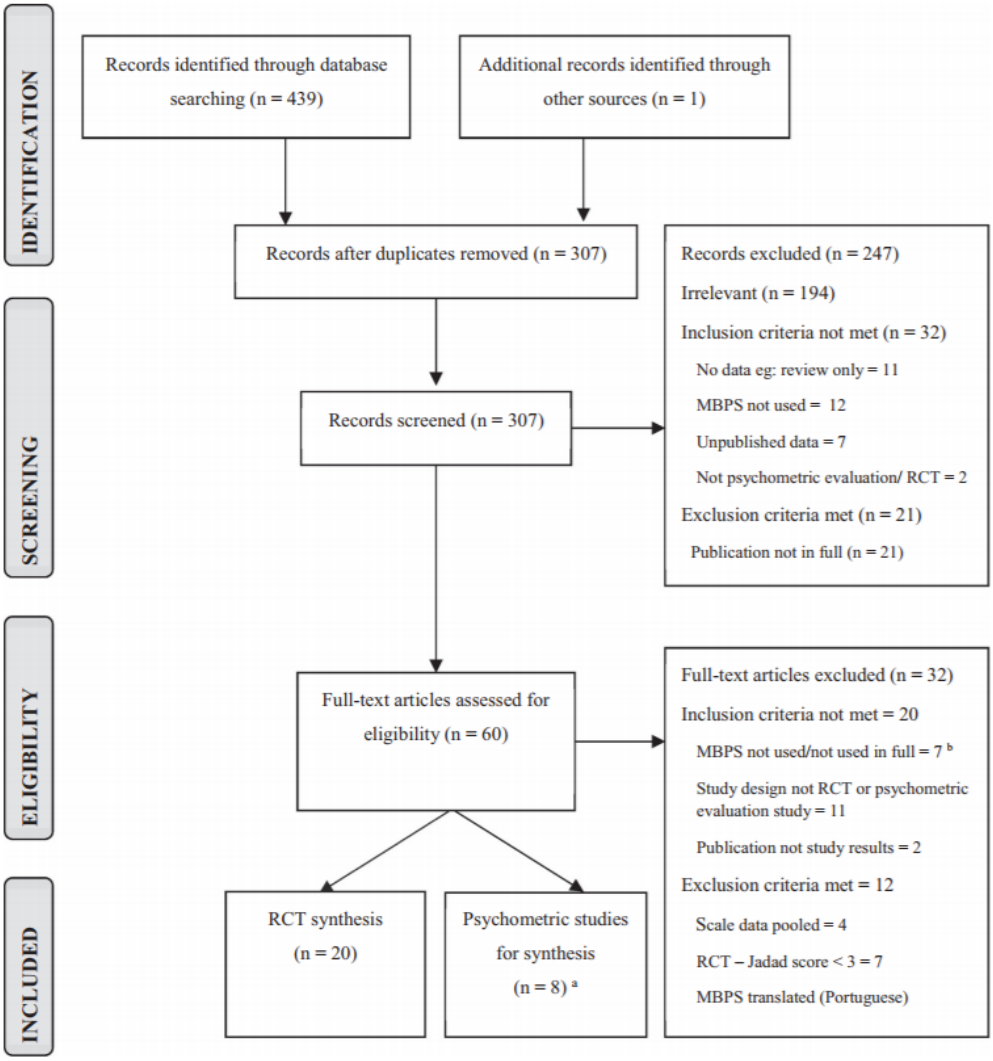


Fig. 1. Flow chart detailing the search and study screening results. [a] Original psychometric evaluation of the newly developed MBPS was conducted in an RCT which included other psychometric evaluation methods. Counted as a psychometric evaluation study. [b] MBPS not described and applied as defined by original authors. Abbrev. MBPS – Modified Behavioral Pain Scale, RCT – Randomised Controlled Trial.

Hogan et al., 2014; Ipp et al., 2004; Ipp et al., 2007; Ipp et al., 2009; Kassab et al., 2012; Lindh et al., 2003; McGowan et al., 2013; Mularoni et al., 2009; O'Brien et al., 2004; Taddio et al., 1994; Taddio et al., 2014; Taddio et al., 2015). The MBPS was used in fifteen RCTs to assess immunisation related pain in infants aged from 1 to 20 months (Fallah et al., 2016; Girish & Ravi, 2014; Hillgrove-Stuart et al., 2013; Hogan et al., 2014; Ipp et al., 2004; Ipp et al., 2007; Ipp et al., 2009; Kassab et al., 2012; Lindh et al., 2003; McGowan et al., 2013; O'Brien et al., 2004; Taddio et al., 1994; Taddio et al., 2014; Taddio et al., 2015; Taddio et al., 2015). An additional trial included children up to 6 years of age undergoing immunisation (Abuelkheir et al., 2014). Three trials also studied interventions aimed at alleviating pain in alternative procedures; intramuscular injection (Carbajal et al., 2008; Dyer et al., 2004) and urethral catheter insertion (Mularoni et al., 2009) and a fourth assessed the efficacy of peri-operative pain management techniques (Anninger et al., 2007). These studies, their quality score and strength of evidence are summarised in Tables 4 to 7 and more details can be found in Appendix B.

Psychometric Properties and Study Quality

Measurement error and content, structural and cross cultural validity for the MBPS have not been reported.

Data Synthesis

The evidence from the psychometric evaluation studies and the RCTs were integrated to draw conclusions about the psychometric properties (reliability, validity, feasibility and clinical utility) of the MBPS.

Table 4
Summary of results and level of evidence supporting MBPS score reliability.

Study	Sample	Circumstances	Reliability results	Quality	Level of evidence
(Cohen et al., 2005)	0.13–1.86 years (n = 18)	Procedural (immunisation) Rural health facilities (USA)	Inter-rater: kappa – face 0.61, cry 0.77, movement 0.67	COSMIN: poor (sample size)	Low
(Lindh et al., 2003)	3 months (n = 45)	Procedural (immunisation) Outpatient pediatric practice (Sweden)	Inter-rater: kappa - baseline 1.0, injection 0.5	COSMIN: fair	Low
(McClellan et al., 2003)	2 and 22 months (n = 18)	Procedural (immunisation) Rural health facility (USA)	Inter-rater: kappa – face 0.61, cry 0.77, movement 0.67	COSMIN: poor (sample size)	Low
(Mularoni et al., 2009)	2–24 months (n = 18)	Procedural (urethral catherisation)/Emergency department (USA)	Inter-rater: kappa 0.79, cry 0.82	COSMIN: poor (sample size)	Low to moderate
(Taddio et al., 1995)	4–6 months (n = 96)	Procedural (immunisation) Outpatient clinic (Canada)	Internal consistency: item correlation – face 0.66, movement 0.55, cry 0.60 ($p < 0.001$) Inter-rater: ICC 0.95 Intra-rater: ICC 0.95	COSMIN: fair COSMIN: fair COSMIN: fair	Low to mod Moderate Moderate
(Taddio et al., 2011)	2–6 months (n = 120)	Procedural (immunisation) Outpatient clinic (Canada)	Internal consistency: Cronbach's alpha baseline 0.94, injection 0.83 Inter-rater: ICC 0.90–0.94 Intra-rater: ICC0.96	COSMIN: poor COSMIN: good COSMIN: good	Low High High
(Verriotis et al., 2015)	1 to 2 & 12 months (n = 15)	Procedural (immunisation) Outpatient clinic (England)	Inter-rater: ICC baseline 0.81, injection 0.89	COSMIN: poor (sample size)	Low

NOTE: Abbreviations: ED- emergency department, ICC – intraclass correlation.

Reliability

The reliability of the MBPS was assessed in seven studies (Cohen et al., 2005 49; McClellan et al., 2003; Mularoni et al., 2009; Taddio et al., 2011; Taddio et al., 1995; Verriotis et al., 2015; Lindh et al., 2003), the results of which are shown in Table 4, All studies examined inter-rater reliability and two assessed intra-rater reliability and internal consistency (both by the original scale authors) (Taddio et al., 1995; Taddio et al., 2011).

Excellent levels of agreement between test and retest scores (intra-rater reliability) were reported in Taddio and colleagues original study of infants aged 4 to 6 months experiencing immunisation related pain (ICC 0.95) (Taddio et al., 1995) and in their subsequent study of 2 months also undergoing immunisation (ICC 0.96) (Taddio et al., 2011). The quality of the earlier study reduced the contribution to the evidence for intra-reliability. Hence there is only sufficient evidence to suggest that the intra-rater reliability of the MBPS is promising.

The inter-rater reliability of the MBPS used to assess immunisation pain was evaluated in six studies; including Taddio and colleagues' original study (5 psychometric evaluation studies (Cohen et al., 2005; McClellan et al., 2003;Taddio et al., 2011; Verriotis et al., 2015) and 1 RCT (Lindh et al., 2003)) The age range across these studies varied but included children aged from 1 to 22 months old and all experienced immunisation-related pain. Two studies reported excellent reliability with intraclass correlations ranging from 0.81 to 0.94 and contributed 'high' and 'low' levels of evidence of score reliability (Taddio et al., 2011; Verriotis et al., 2015). Lindh et al. (2003) reported variable reliability results (kappa ranging from 0.5 to 1.0) using methods rated 'fair', hence contributed only 'low' levels of evidence of reliability. Kappa values for the scale items reported in two studies provided low to moderate levels of evidence for the reliability of the scale items (face 0.61, movement 0.67 and cry 0.77) (Cohen et al., 2005; McClellan et al., 2003). Remaining studies used 'low' quality methods but reported results consistent with those described here. It can be cautiously concluded based on promising evidence, that the MBPS is reliable when used to assess immunisation related pain in infants and children up to the age of almost two years.

Finally, the reliability of the MBPS was also reported as 'good' (kappa 0.79) when used to assess children up to two years having a urethral catheter inserted (Mularoni et al., 2009). Similar to other studies, the quality of the methods was rated 'poor' using the COSMIN checklist. Evidence based on the results from a single study assessing the capacity of the MBPS to reliably assess pain associated with procedures other than immunisation is 'limited' and therefore insufficient to draw conclusions.

Internal consistency was examined by Taddio and her colleagues in the original psychometric evaluation study and again in their 2011 study (Taddio et al., 1995; Taddio et al., 2011). Item correlations ranged from 0.48–0.67, $p < 0.001$ (Taddio et al., 1995) and Cronbach alpha at baseline and during the procedure was >0.83 (Taddio et al., 2011). However, limitations in the methods of both studies resulted in a quality rating of 'poor' and reduced the contribution that these studies make to the evidence confirming the internal consistency of the MBPS, which is at best 'limited'.

Validity

Hypothesis testing (comparison between known groups (n = 21) and correlation with an alternative pain measure (n = 8)) (Table 5), criterion validation (n = 4) (Table 6) and responsiveness testing (n = 7) (Table 7) were the methods used in 27 studies potentially contributing evidence of MBPS validity. Most studies were conducted in infants and children under the age of 2 years while only four apply the MBPS to alternative circumstances and/or age groups.

Immunisation-related Pain

Taddio and colleague's original validation study used three validation methods to test the performance of the MBPS; the capacity of the scale to distinguish between known groups, convergent validation and responsiveness testing. Participants were randomised to two groups and MBPS scores were lower in the treatment group when compared with the control group (6.8, SD 1.9 versus 8, SD 1.5, p, 0.001) (Taddio et al., 1995). The Visual Analogue Scale (VAS) was applied by an observer and the correlations between VASobs and MBPS scores were good (observer $r = 0.68$, pediatrician $r = 0.74$, $p < 0.001$). Scale responsiveness was demonstrated by a significant increase in MBPS scores from baseline to after the immunisation (1.9, SD 0.8 to 7.3 SD 1.8, $p < 0.001$). Demonstration of the capacity of the scale to distinguish between known groups in this age group is also demonstrated in Ipp and colleagues' RCT evaluating two immunisation techniques which contributes moderate levels of evidence to support the validity of the MBPS for immunisation pain measurement (Ipp et al., 2007).

The bulk of the data contributing to our understanding of the validity of the MBPS is derived from studies that have included infants older

Table 5
Summary of results and level of evidence supporting MBPS score validity using hypothesis testing.

Study	Sample	Circumstances	Method design	Results	Quality	Level of evidence
(Abuelkheir et al., 2014)	2 months–6 years (n = 216)	Procedural (Immunisation)/Well baby pediatric clinic (Saudi Arabia)	RCT	Btw groups: difference in MBPS scores (2.56 ± 1.96 vs 3.95 ± 2.20, p < 0.001) & related independent variables (VAS scores) & dependent variables (cry & crying time)	Jadad 5	High
(Anninger et al., 2007)	1–12 years (n = 88)	Postoperative (strabismus surgery)/PACU (USA)	RCT	Btw groups: difference in MBPS scores (p < 0.033) and related independent variables (emergence scores)	Jadad 4	Moderate
(Carbajal et al., 2008)	<24 months (n = 55)	Procedural (injection)/pulmonary outpatient department (2 hospitals in France)	RCT	Btw groups: difference in MBPS scores (8.2 ± 1.8 v's 9.3 ± 1.0 v's 8.8 ± 1.2, p < 0.001) & related independent variables (VAS scores)	Jadad 5	Moderate
(Dyer et al., 2004)	1 month–18 years (n = 20)	Procedural (G-CSF injection administration) Haematology/Oncology department (Australia)	RCT	Btw groups: no difference in MBPS scores or related independent variables (FAS scores).	Jadad 3 (Sample size not large enough to generate significant results)	Nil
(Fallah et al., 2016)	18 month (n = 70)	Procedural (immunisation)/Primary Health Care Clinic (Iran)	RCT	Btw groups: difference in total MBPS scores (15.61 ± 2.6 vs 14.23 ± 1.35 p = 0.04), no difference in individual vaccine MBPS scores no difference in dependent variables (cry duration	Jadad 3	Nil
(Girish & Ravi, 2014)	6 weeks–1.5 years (n = 200)	Procedural (immunisation)/Hospital (India)	RCT	Btw groups: difference in MBPS scores (8.4 ± 0.75 vs7.8 ± 1.17, p = 0.00) and related dependent variable (crying).	Jadad 3	Nil to low
(Hillgrove-Stuart et al., 2013)	12–20 months (n = 99)	Procedural (immunisation)/Pediatrician's clinic (Canada)	RCT	Btw groups: no difference in MBPS scores * highlights potential inability of scale to differentiate between pain and distress	Jadad 3	Nil
(Hogan et al., 2014)	4–6 months (n = 120)	Procedural (immunisation)/Primary care practice (Canada)	RCT	Btw groups: no difference in MBPS scores or related independent variables (VAS scores).	Jadad 3	Nil
(Ipp et al., 2004)	12 months (n = 49)	Procedural (immunisation)/Community pediatricians clinic (Canada)	RCT	Btw groups: difference in MBPS scores (6 vs 8, p = 0.02), no difference in related dependent variables.	Jadad 5	Nil to low
(Ipp et al., 2007)	4–6 months (n = 113)	Procedural (immunisation)/Primary Care Practice (Canada)	RCT	Btw groups; difference in MBPS scores, (5.6, 5 to 6.3 vs 3.3, 2.6 to 3.9) related independent variables (parent & pediatrician VAS, duration of injection) and related dependent variables (cry and cry duration).	Jadad 3	Moderate
(Ipp et al., 2009)	2–6 months (n = 120)	Procedural (immunisation)/Pediatric community practice (Canada)	RCT	Btw groups: difference in MBPS scores (7.6 ± 1.5 vs 8.2 ± 1.5, p = 0.037) and related independent variables (parent VAS).	Jadad 5	Moderate
(Kassab et al., 2012)	2 months (n = 120)	Procedural (immunisation)/Maternity and child health care centre (Jordan)	RCT	Btw groups: difference in MBPS scores during (median 8, IQR = 1 versus 9, IQR = 1) & post-immunisation (median 4, IQR – 1 and 6, IQR = 3) (p < 0.001) & related dependent variables (crying time).	Jadad 5	Low to moderate
(Lindh et al., 2003)	3 months (n = 90)	Procedural (immunisation)/outpatient pediatric practice (Sweden)	RCT	Btw groups: difference in MBPS scores at 0–10 s (5.5 ± 2.0 v 7.7 ± 1.7) & 11–20 s (5.4 ± 2.4 v 6.8 ± 2.2), related independent variables (parent and nurse VAS, HR changes)	Jadad 5	High
(McClellan et al., 2003)	2 and 22 months (n = 18)	Procedural (immunisation) Rural health facility (USA)	Obs	Convergent – no correlation between MBPS scores and VASobs pain or VASobs distress scores.	COSMIN: poor (sample size)	Low
(McGowan et al., 2013)	2–6 months (n = 36)	Procedural (immunisation)/Immunisation clinic (Wales)	RCT	Btw groups: difference in MBPS scores - median change (pre-post) less in simultaneous group at 15 s (p = 0.05), greater in simultaneous group at 30 s (p < 0.05), 45 s (p = 0.01) and 120 s (p = 0.02).	Jadad 3	Nil to low
(Mijovic et al., 2010)	Term neonates (n = 24)	Procedural (blood sampling) Postnatal unit, University Hospital (Belgium)	Obs	Convergent – correlations between MBPS scores and intensity contours of cry bouts & fundamental frequencies r = 0.39 to 0.83)	COSMIN: poor (untested index measure)	Nil
(Mularoni et al., 2009)	2–24 months (n = 45)	Procedural (urethral catherisation)/Emergency department (USA)	RCT	Btw groups: no difference in MBPS scores.	Jadad 5	Nil
(O'Brien et al., 2004)	1 year (n = 120)	Procedural (immunisation)/Pediatric outpatient clinic (Canada)	RCT	Btw groups: difference in MBPS scores (1.51 vs 2.29, p = 0.029).	Jadad 5	Low
(Silva et al., 2010)	Term infants (n = 47)	Procedural (blood sampling) Postnatal unit (Belgium)	Obs	Convergent – no correlation between MBPS scores and no of cries & fundamental frequency of cries	COSMIN: fair (no significant correlations, untested index measure)	Nil
(Taddio et al., 1994)	4–6 months (n = 96)	Procedural (immunisation)/Pediatric outpatient clinic (Canada)	RCT	Btw groups: difference in MBPS scores (post-immunisation (7 v 8, p = 0.001) & difference between pre & post-scores (5 v 6, p = 0.001) & related independent variable (VASobs scores) & related dependent variables (cry related variables).	Jadad 3	Moderate
(Taddio et al., 1995)	4–6 months (n = 96)	Procedural (immunisation)/Pediatric outpatient clinic (Canada)	RCT	Btw groups: difference in MBPS scores (6.8, ±–1.9 vs 8, ±–1.5, p < 0.001) and related independent variables (VASobs scores)	Jadad 1	Low to moderate
(Taddio et al.,	2–6	Procedural (immunisation)	RCT*	Btw groups: difference in MBPS scores (6.3 ± 2.8 and 8.2 ± 1.4,	COSMIN: fair	Low to

Table 5 (*continued*)

Study	Sample	Circumstances	Method design	Results	Quality	Level of evidence
2011)	months (n = 120)	Outpatient clinic (Canada)		p < 0.001) and related variables.		moderate
(Taddio et al., 2014)	1–12 months (n = 121)	Procedural (immunisation)/Private pediatric clinic (Canada)	RCT	Btw groups: no difference in MBPS scores or related dependent variables (cry related variables), difference in related independent variable (parent and pediatrician NRS scores).	Jadad 3	Nil
(Taddio et al., 2015)	<24 weeks (n = 160)	Procedural (immunisation)/Outpatient pediatric clinic (Canada)	RCT	Btw groups: Stage 1 – no difference in MBPS scores or related independent (Parent NRS) and dependent variables (cry related), Stage 2 – difference in MBPS scores ((7.8 v 8.3, p = 0.002), no difference in related independent (parent NRS) and dependent variables (cry related).	Jadad 3	Nil
(Taddio et al., 2015)	2–4 months (n = 120)	Procedural (immunisation)/Outpatient pediatric clinic (Canada)	RCT	Btw groups: no difference in MBPS score or related independent (parent NRS and clinician NRS) and dependent (cry duration) variables.	Jadad 5	Low
(Verriotis et al., 2015)	1–2 months and 12 months (n = 15)	Procedural (immunisation) Outpatient clinic (England)	Obs	Convergent: 1- to 2-month-olds (Spearman rank order correlation coefficient; waveform 1: r = −0.15, p = 0.62 and waveform 2: r = −0.20, p = 0.52; n = 13)	COSMIN: poor (sample size)	Nil

NOTES: *Data derived from an RCT but study not reported as an RCT. Abbreviations: ED- emergency department, FAS – Facial Action Scale, HR – heart rate, NRS – Numeric Rating Scale, Obs – Observational study, PACU - Post Anesthetic Care unit, VAS – Visual Analogue Scale, VASobs - Visual Analogue Scale observer.

and/or younger than the original age cohort (4–6 months). Data from 14 studies using hypothesis testing contributed evidence of MBPS validity in this age group (Abuelkheir et al., 2014; Girish & Ravi, 2014; Ipp et al., 2004; Ipp et al., 2007; Ipp et al., 2009; Kassab et al., 2012; Lindh et al., 2003; McClellan et al., 2003; McGowan et al., 2013; O'Brien et al., 2004; Taddio et al., 1994; Taddio et al., 1995; Taddio et al., 2011; Taddio et al., 2015). The most convincing evidence derived from hypothesis testing was provided by two RCTs, both testing the efficacy of EMLA cream applied prior to immunisation to infants and children up to 6 years of age (Abuelkheir et al., 2014; Lindh et al., 2003). In both studies the difference in MBPS scores between groups was substantiated by differences between groups in other independent pain-related variables such as; heart rate and VASobs scores. Results from additional work from the scale authors contribute moderate and low to moderate levels of evidence of scale validity in two studies demonstrating the capacity of the scale to differentiate between groups (Taddio et al., 1994; Taddio et al., 2011). It should be noted that the later study used data derived from an RCT but the methods for this testing was not described as an RCT. Hence the quality of the methods was assessed using the COSMIN Checklist.

Using a similar design to their original validation study, Taddio and colleague's second psychometric evaluation of the MBPS reported very similar results for infants aged two to six months receiving one of two vaccines (Taddio et al., 2011). Significant differences between the two trial groups were reported (6.3 SD 2.8 and 8.2 SD 1.4, p < 0.001). Correlation coefficients between MBPS and other scales used to measure pain in this study were high (FLACC r = 0.84; Neonatal Infant Pain Scale (NIPS) r = 0.87). The responsiveness of the scale to immunisation pain was also demonstrated by a marked increase in MBPS score from baseline to post the immunisation (2.3 versus 7.7).

Responsiveness of MBPS scores to immunisation-related pain was reported in seven studies and demonstrated statistically in 5 studies (Abuelkheir et al., 2014; Ipp et al., 2004; McClellan et al., 2003; Mularoni et al., 2009; Taddio et al., 1995; Taddio et al., 2011; Verriotis et al., 2015). McClellan and her colleagues expressed and compared as a scale item average across phases of the procedure and a significant rise from baseline and pre-immunisation to during and post-immunisation phases was reported (baseline 0.84 SD 0.48, pre- 0.96 SD 0.64, during 2.26 SD 0.43 and post 2.05 SD 0.61, p < 0.001) (McClellan et al., 2003). Uniquely, these results are substantiated by the demonstration of similar patterns in other variables collected across phases, e.g. heart rate, nurse pain ratings and nurse and parent distress ratings. Unfortunately, the quality of the methods used in these studies were all rated 'fair' and contributed low level evidence to the validation of the MBPS used to assess immunisation-related pain. The methods used to test responsiveness were limited in Taddio's three studies and McClellan's study by the inability to blind the raters to the circumstances. Scores changed substantially in the remaining two studies e.g. baseline 2.0 to 9.0 following immunisation in Verriotis' sample of 2 month old infants (Abuelkheir et al., 2014; Verriotis et al., 2015). However, in the absence of statistical testing it is difficult to accept their results as evidence of responsiveness.

In four studies, the correlations between MBPS and an alternative scale were described as concurrent (criterion) validation of the MBPS but in the absence of convincing validation data, the VASobs, NIPS and

Table 6
Summary of results and level of evidence supporting MBPS score validity using criterion validation.

Study	Sample	Circumstances	Results	Quality	Level of evidence
(Cohen et al., 2005)	0.13–1.86 years (n = 18)	Procedural (immunisation) Rural health facilities (USA)	Correlation with MSAID scores (r = 0.44, p < 0.001).	COSMIN: poor (untested index scale, not 'gold standard')	Nil
(Taddio et al., 1995)	4–6 months (n = 96)	Procedural (immunisation) Outpatient clinic (Canada)	Correlation with VASobs scores (observer r = 0.68 and pediatrician r = 0.74, p < 0.001).	COSMIN: poor (index scale not 'gold standard')	Low
(Taddio et al., 2009)	1-year old (n = 120)	Procedural (immunisation) Pediatric outpatient clinic (Canada)	Correlation with VASobs scores (r = 0.81–0.94)	COSMIN: poor (index scale not 'gold standard')	Low
(Taddio et al., 2011)	2–6 months (n = 120)	Procedural (immunisation) Outpatient clinic (Canada)	Correlation with FLACC scores (r = 0.84) & NIPS (r = 0.87)	COSMIN: poor (index scale not 'gold standard')	Low

NOTE: Abbreviations: ED- emergency department, FLACC – Face, Legs, Activity, Cry Consolability, MSAID - Measure of Adult and Infant Soothing & Distress, NIPS – Neonatal and Infant Pain Scale, VAS – Visual Analogue Scale, VASobs - Visual Analogue Scale observer.

Table 7
Summary of results and level of evidence supporting MBPS score validity using responsiveness testing.

Study	Sample	Circumstances	Results	Quality	Level of evidence
(Abuelkheir et al., 2014)	2 months–6 years (n = 216)	Procedural (Immunisation) Well baby pediatric clinic (Saudi Arabia)	Demonstrated: Difference in pre–post-scores both groups (2.56 ± 1.96 versus 3.95 ± 2.20)	COSMIN: poor (no significance testing)	Nil to low
(Ipp et al., 2004)	12 months (n = 49)	Procedural (immunisation) Community pediatric clinic (Canada)	Demonstrated median difference in pre and post-scores for both groups (3 and 5)	COSMIN: fair	Low
(McClellan et al., 2003)	2 and 22 months (n = 18)	Procedural (immunisation) Rural health facility (USA)	Demonstrated: baseline 0.84 ± 0.48, pre- 0.96 ± 0.64, during 2.26 ± 0.43 and post 2.05 ± 0.61, p < 0.001	COSMIN: fair (raters not blinded to circumstances)	Low
(Mularoni et al., 2009)	2–24 months (n = 45)	Procedural (urethral catherisation) ED (USA)	Demonstrated: baseline scores lower than instillation (t44 = 3.53, p = 0.001) or catheterization scores (t42 = 3.14, p = 0.003)	COSMIN: fair	Low
(Taddio et al., 1995)	4–6 months (n = 96)	Procedural (immunisation) OP clinic (Canada)	Demonstrated: before (1.9, ±0.8) to after immunisation (7.3, ±1.8) p < 0.01.	COSMIN: fair (raters not blinded to circumstances)	Low
(Taddio et al., 2011)	2–6 months (n = 120)	Procedural (immunisation) OP clinic (Canada)	Demonstrated: scores; baseline 2.3 versus post-immunisation 7.7, p < 0.001	COSMIN: fair (raters not blinded to circumstances)	Low
(Verriotis et al., 2015)	1–2 months and 12 months (n = 15)	Procedural (immunisation) Outpatient clinic (England)	Demonstrated: baseline scores 2.0 [2.0–2.0] for 1- to 2-mnth-olds & 2.0 [1.0–2.0] for 12-month-olds) during immunisation (8.0 [7.5–8.0] & 9.0 [9.0–9.0], respectively)	COSMIN: poor (no significance testing)	Nil to low

NOTE: ED- emergency department, OP – outpatients, USA – United States of America.

the FLACC scales cannot be considered as 'gold standard' rendering the methods of these studies 'poor'. Only three of these studies, all from the same research team, contributed evidence ('low') to the validation of the MBPS (Taddio et al., 1995; Taddio et al., 2009; Taddio et al., 2011).

On the basis of the data supporting the responsiveness of the MBPS to changes in pain and its capacity to differentiate between known groups, it can be cautiously concluded that the MBPS is a valid measure of immunisation pain in infants from up to 24 months old. Abuelhkeir and colleagues provide the only data reported in the review for children older than 24 months receiving routine immunisation (Abuelkheir et al., 2014). Four years-olds receiving routine immunisations accounted for approximately 6.5% of the sample, while all remaining children were aged <24 months. This is insufficient to extend this recommendation to older children.

Other Procedures

Four RCTs have used the MBPS to assess pain related to circumstances other than immunisations (Anninger et al., 2007; Carbajal et al., 2008; Dyer et al., 2004; Mularoni et al., 2009). No difference in MBPS scores between treatment groups for infants aged two to 24 months during urethral catheter insertion (Mularoni et al., 2009) or infants aged one month to 18 years undergoing two different two injection techniques (Dyer et al., 2004) were observed. Dyer's results were supported by independent pain related variables suggesting that the MBPS scores may be a true measure of the existence of no difference between groups. In contrast, in a study evaluating MBPS during intramuscular injection in infants/children up to 24 months and receiving topical anaesthetic and/or inhaled nitrous oxide, there were significant differences in scores between the groups (topical anaesthetic and inhaled nitrous oxide (8.2 =/– 1.8) v's topical anaesthetic (9.3 ± 1.0) v's inhaled nitrous oxide (8.8 ± 1.2), $p < 0.001$) (Carbajal et al., 2008). The disparate age ranges in these studies raises some concerns about pooling these results to draw conclusions about the validity of the MBPS used to assess pain associated with procedures other than immunisation.

Postoperative Pain

Finally, the MBPS was used to assess pain intensity in children aged up to 12 years emerging from anaesthetic after eye surgery. Differences in MBPS scores ($p < 0.033$) and emergence scores ($p < 0.019$) between groups (A - placebo before and after surgery, B - placebo before and anaesthetic eye drops after surgery, C - anaesthetic eye drops before and after surgery) were reported (Anninger et al., 2007). This data was not sufficient to make strong recommendations but should encourage researchers not to discard this scale as an option without further assessment.

MBPS Used as a Reference Scale

The MBPS has been used on four occasions as a reference scale in studies examining the performance of another assessment scales (Cohen et al., 2005; Mijovic et al., 2010; Silva et al., 2010; Verriotis et al., 2015). Results of the correlation between alternative scale or tool and the MBPS varied between conflicting, moderate, good and excellent. However, methods were assessed as 'poor' on the COSMIN Checklist. Each of these studies is based on an assumption that the MBPS is valid and therefore can be used to test newly developed methods for assessing infant pain or scales validated for alternative circumstances (e.g.: post-operative pain) but as yet untested for immunisation related pain. Regardless of the results of these studies, they cannot be accepted as evidence of MBPS validity as convergence has been demonstrated with a scales with unknown psychometric properties (Cohen et al., 2005; Mijovic et al., 2010; Silva et al., 2010; Verriotis et al., 2015).

Feasibility and Clinical Utility

Taddio and colleagues, using a variety of approaches in a well-designed study, make the only attempt to date to assess the feasibility

and clinical utility of the MBPS (Taddio et al., 2011). A single study can only form a basis for encouraging further work but is insufficient to claim that the MBPS is either feasible or clinically useful.

Study Quality

The COSMIN scores ranged from 'poor' to 'good', although the methods for reliability, hypothesis testing and responsiveness testing were only assessed as 'good' for two (7.5%) studies (Taddio et al., 2009; Taddio et al., 2011). The RCT methods of the original validation study by Taddio scored '1' on the Jadad Scale, however was included as it also meet the requirements for inclusion as a psychometric evaluation study. The significant factors which impacted on the quality of the psychometric evaluation methods are highlighted in Tables 4 to 7 and include; small sample sizes, use of non-recommended analysis techniques, failure to report missing data, use of non-gold standard reference scales for concurrent validation and use of non-blinded raters for scoring for responsiveness testing.

For reliability or validity to be considered 'well established' the IMMPACT criteria recommend evidence derived from different researchers or research teams (Cohen et al., 2008). The original author responsible for the MBPS scale development and testing was also an investigator and co-author in three of the psychometric evaluation studies and eight of the RCTs.

Discussion

The evidence to establish the psychometric properties of the MBPS had not been previously summarised, despite extensive use of the scale to assess immunisation pain intensity and a recognised need to accurately assess procedural pain for clinical and research purposes. This review attempts to addresses this research gap and provides the first opportunity to make recommendations about the application of the MBPS in clinical practice and research and to identify future research directions. This review focused attention on concerns regarding the premise for the development of scale, the circumstances under which it has been tested and their impact on the scale's performance, the specificity of the scale for pain assessment, the feasibility of using it for procedural assessment, and the quality of the methods used to assess the psychometric properties.

Scale Development

The MBPS, a modification of the CHEOPS, was developed to capture the variability of responses to pain likely to be demonstrated by young infants. The MBPS reduces the number of scale items in the CHEOPS from six items to three and modifies descriptors to describe behaviours more consistent with those of young infants. Despite the premise for the development of the scale, there are no published studies examining content validation, contrasting the performance of the MBPS with its precursor the CHEOPS, or comparing performance in varying age ranges. The psychometric data for MBPS was provided largely in studies including children up to the age of 24 months, although the scale was originally developed for younger infants aged 4–6 months. It is unclear from these studies or literature whether pain related behaviours are likely to vary between infants aged four to six months and older infants to make a unique scale necessary. As an aside, these factors make it surprising that it was chosen as the scale of choice in the three RCTs using the scale to assess pain in samples that included children aged 6 years (Abuelkheir et al., 2014), up to 12 years (Anninger et al., 2007) and up to 18 years (Dyer et al., 2004) and they do not contribute a great deal to the evidence for the scales validity in this age group. Data from studies included in this review support the contention that the MBPS can be reliably used to assess immunisation related pain across a broader age range than was first proposed and furthermore that the scale is sensitive to immunisation related pain in this age cohort. However, when the performance of the MBPS is compared with the performance of other scales it cannot be convincingly claimed that the MBPS is better suited to assessment of pain in these age groups or that it is suitable for procedures other than immunisation.

Procedural Versus Immunisation Related Pain

The psychometric properties of a scale are contingent not only on the population to which the scale is applied but also on the circumstances in which the scale is applied. Many observational pain scales were initially either designed or tested for assessing postoperative pain. However, post-operative pain experience may not be the same as a procedural pain experience and therefore associated pain related behaviours may also differ. For this reason, scales have either been developed purposively, for assessing procedural pain, or attempts have been made to establish the psychometrics of existing scales when used to assess procedural pain. The MBPS was introduced as a measure for young infants experiencing procedural pain (Taddio et al., 1995). However, evaluation of the MBPS performance has mostly concentrated on the scale's capacity to assess pain intensity during immunisation. To accept the data from these studies as evidence of the psychometrics of the MBPS for other procedures it would be necessary to accept that immunisation is a representative model for procedural pain. However, common medical procedures such as venepuncture, nasogastric tube, intravenous cannula insertion and urethral catheterisation often take more time and occur under different circumstances to routine immunisation. Contextual factors such as the associated symptoms of the illness, the clinical environment in which these procedures are usually performed and parental anxiety are likely to impact on the pain experience. Although the MBPS shows promise, there are currently insufficient data to confidently claim that the MBPS is a valid measure of all procedural pain.

Specificity for Procedural Pain

In addition, fear and anxiety are likely to manifest in distress behaviours that mimic those of pain, making independent assessment of pain difficult if not impossible (Babl et al., 2012). Ahola Kohut and colleagues elegantly illustrate this in their study demonstrating that the Neonatal Facial Coding System is unable to differentiate between infants experiencing pain-related and non-pain-related distress (Ahola Kohut & Pillai Riddell, 2009). Validation of a measure aimed at assessing procedural pain, where fear or anxiety levels can be high, must establish not only the scale's sensitivity to pain (responsiveness and capacity to distinguish between groups) but also include methods to demonstrate the measure's specificity; in other words the capacity of the measure to distinguish between pain related behaviours and those associated with other negative emotions that may be experienced during a medical procedure. Blount and Loiselle acknowledge behavioral assessment scales are measures of distress and not exclusively pain (Blount & Loiselle, 2009). It may be unrealistic to expect that any scale can achieve high levels of specificity for pain. However, while these scales continue to be used where specificity is necessary, e.g.: trials measuring the efficacy of therapies targeting pain, such as analgesics, or distress, such as sedation, it is essential that we understand their limitations. The specificity of the MBPS is yet to be explored. Indeed this is a line of enquiry largely missing from attempts to validate most pain assessment measures.

Feasibility

Procedural pain poses unique challenges to designers of assessment measures, namely the potential impact of the procedure and/or physical restraint on the psychometrics of the scale or the feasibility of application of the scale. Despite this and the frequent use of restraint (Crellin et al., 2011), modifications to the CHEOPS to produce the MBPS resulted

in the removal of descriptors to assist the observer to score the 'movement' item for restrained infants. It has also been recognised that the impact of the procedure and restraint on behaviours and therefore the score is largely unknown (Crellin et al., 2015). For example; young children may be likely to struggle more when restrained than they might have otherwise. The feasibility of the MBPS has only been assessed once in a study by Taddio and colleagues. Although not explicitly stated, it is implied from the results that immunisation and associated restraint did not interfere with scoring (Taddio et al., 2011). Further to this and perhaps surprisingly, Carbajal and colleagues state that the use of a face mask during the procedure did not hamper assessor's ability to score the 'face' in their RCT (Carbajal et al., 2008).

Quality

According to Cohen and colleagues, to claim that the psychometrics have been 'well-established', the measure must have been presented in two peer-reviewed papers by different investigatory teams (Cohen et al., 2008). There is no doubt that this has been achieved for MBPS. However, it is worth noting that 9 of the 29 studies in this review included one of the original research team responsible for development of the MBPS. In addition, another four studies were conducted in Canada by members of a broader and strongly collaborative network of pediatric pain researchers (Pain in Child Health http://paininchildhealth.dal.ca/), which includes the original scale authors. The effect of study authorship is not known but potential bias resulting in more positive results can't be easily dismissed.

This IMMPACT framework (Cohen et al., 2008) provides guidance about the number of studies that are necessary to make claims about the strength of the evidence for the psychometrics of a measure. However, strength of evidence is a function of both the volume of data from multiple independent studies and the quality of the studies and this framework does not advise about criteria and standards to establish study quality. The COSMIN checklist helps to fill this gap providing criteria to establish the quality of the methods used to assess the psychometrics of the measure. Conversely, the Checklist provides no guidance about the number of studies required to claim validity. The factors influencing an assessment of validity may be too complicated to distil to a scale defining validation strength based on the number of studies and the quality of the study methods.

The methods used to establish the validity of a new measure such as the MBPS, where an existing gold standard does not exist, rely on indirect validation methods such as; responsiveness to changes in condition, capacity to distinguish between known groups and correlations with similar measures. However, these methods all have potential limitations, for example, methods used to assess responsiveness often do not or are unable to blind the reviewers to the circumstances surrounding the procedure/analgesic administration (Crellin et al., 2015). The studies in this review were no different: quality was variable but most frequently assessed as 'fair' or 'poor'.

Convergence and concurrence are widely accepted methods used to contribute to the validation of an assessment measure and were used in many of the studies included in this review. Data derived from a study using MBPS to validate a new scale, where the MBPS validity is assumed, cannot then be confidently used to support the validity of the MBPS regardless of the strength of the results. Similar circular logic underpins studies using alternative observational scales to validate new scales. This is made most clear by Taddio's use of the VAS to validate the MBPS and then the MBPS to validate the VAS for procedural pain assessment (Taddio et al., 1995; Taddio et al., 2009). One measure must have been ideally demonstrated as valid for the purpose in question to provide support for the validity of the second measure. In the absence of a good understanding of the psychometrics of a measure, achieved through robust systematic review of the evidence, it could be argued that using this scale as a reference to establish the psychometrics of another measure is problematic.

Limitations

This systematic review had several limitations which may impact on the final outcome of the review and the associated recommendations. It is possible that the search did not identify all studies providing evidence of the psychometric properties of the MBPS and this is particularly the case for our search for RCTs using the MBPS to measure a study outcome. The review was restricted to studies published in English and available in full. Non-English publications likely documented validation of a translated version of the scale. This was unlikely to contribute useful data to an evaluation of the English version of the MBPS. In contrast, excluded unpublished data may have provided data to support or conflict with the data from eligible studies. The search identified a number of studies which used the MBPS to explore relationships between pain and other variables. These studies were not validation studies and were based on an assumption of MBPS validity and were excluded as their aims where not consistent with the review inclusion criteria (Pillai Riddell, Flora, Stevens, Greenberg, & Garfield, 2014; Pillai Riddell et al., 2013; Pillai Riddell et al., 2011; Pillai Riddell, Stevens, Cohen, Flora, & Greenberg, 2007; Racine, Riddell, Flora, Garfield, & Greenberg, 2012; Stevens et al., 2013). On closer review of these studies, it was unlikely that these studies could contribute significant data to a review of the measurement properties of the MBPS. Finally, heterogeneous study designs made meta-analysis impossible.

Conclusion and Recommendations

Systematic reviews summarising available pain scale psychometric data are a vital means to providing a platform on which to make recommendations about their use. To our knowledge two systematic reviews published in 2007 are the only substantial reviews of the literature available to guide researchers and clinicians seeking an appropriate tool to assess procedural pain assessment experienced by infants (Crellin et al., 2007; von Baeyer & Spagrud, 2007). Both reviews and published clinical guidelines (Fein, Zempsky, Cravero, & The Comittee on Pediatric Emergency Medicine Section on Anesthesiology Pain Medicine, 2012; Howard et al., 2008; Royal Australasian College of Physicians & Paediatric and Child Health Division, 2005; Royal College of Nursing, 2009; Schug et al., 2015), recommend the FLACC scale for procedural pain assessment, the validity of which has been more recently challenged in a systematic review of the scales psychometric properties (Crellin et al., 2015). The MBPS was developed to capture the nuances and variability in pain response of young infants during acute procedural pain. However, it is not yet possible to claim that this objective has been met in full or that this scale outperforms the FLACC scale. The results of the current review support recommending the MBPS to assess immunisation pain in infants between 2 and 24 months but provide no evidence to support its use for older infants and children or used to assess other procedures.

In the absence of these data and viable alternatives, well-designed studies conducted by independent research teams studying infants experiencing a range of procedures and aimed at testing the scales capacity to discriminate between pain and non-pain related behaviours are urgently needed. In the interim, the MBPS should be considered potentially suitable to assess procedural pain in children too young to self-report.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pedn.2018.02.005.

Funding and Other Material Support

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

All authors declare no conflicts of interest.

Acknowledgements

Dianne Crellin is a trainee member of *Pain in Child Health*, a research training initiative of the Canadian Institutes of Health Research.

References

Abuelkheir, M., Alsourani, D., Al-Eyadhy, A., Temsah, M. H., Meo, S. A., & Alzamil, F. (2014). EMLA(R) cream: A pain-relieving strategy for childhood vaccination. *Journal of International Medical Research*, *42*(2), 329–336https://doi.org/10.1177/0300060513509473.

Ahola Kohut, S., & Pillai Riddell, R. (2009). Does the Neonatal Facial Coding System differentiate between infants experiencing pain-related and non-pain-related distress? *The Journal of Pain*, *10*(2), 214–220.

Anninger, W., Forbes, B., Quinn, G., & Schreiner, M. S. (2007). The effect of topical tetracaine eye drops on emergence behavior and pain relief after strabismus surgery. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *11*(3), 273–276.

Babl, F. E., Crellin, D., Cheng, J., Sullivan, T. P., O'Sullivan, R., & Hutchinson, A. (2012). The use of the faces, legs, activity, cry and consolability scale to assess procedural pain and distress in young children. *Pediatric Emergency Care*, *28*(12), 1281–1296https://doi.org/10.1097/PEC.0b013e3182767d66.

von Baeyer, C. L., & Spagrud, L. J. (2007). Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. *Pain*, *127*(1–2), 150.

Blount, R. L., & Loiselle, K. A. (2009). Behavioral assessment of paediatric pain. *Pain Research & Management*, *14*(1), 47–52.

Blount, R. L., Piira, T., Cohen, L. L., & Cheng, P. S. (2006). Pediatric procedural pain. *Behavior Modification*, *30*(1), 24–49.

Carbajal, R., Biran, V., Lenclen, R., Epaud, R., Cimerman, P., Thibault, P., ... Fauroux, B. (2008). EMLA cream and nitrous oxide to alleviate pain induced by palivizumab (Synagis) intramuscular injections in infants and young children. *Pediatrics*, *121*(6), e1591–1598.

Chang, J., Versloot, J., Fashler, S. R., McCrystal, K. N., & Craig, K. D. (2015). Pain assessment in children: Validity of facial expression items in observational pain scales. *Clinical Journal of Pain*, *31*(3), 189–197. https://doi.org/10.1097/AJP.0000000000000103.

Chorney MacLaren, J., & McMurtry, M. C. (2014). Behavioural measures of pain. In P. J. McGrath, B. J. Stevens, S. M. Walker, & W. T. Zempsky (Eds.), *Oxford textbook of paediatric pain* (1 ed.). Oxford: Oxford University Press.

Cochrane Collaboration. (2014). *Data collection form for intervention reviews: RCTs and non-RCTs*. http://training.cochrane.org/resource/data-collection-forms-intervention-reviews, The Cochrane Collaboration.

Cohen, L. L., Bernard, R. S., McClelland, C. B., & MacLaren, J. E. (2005). Assessing medical room behavior during infants' painful procedures: The measure of adult and infant soothing and distress (MAISD). *Children's Health Care*, *34*(2), 81–94.

Cohen, L. L., La Greca, A. M., Blount, R. L., Kazak, A. E., Holmbeck, G. N., & Lemanek, K. L. (2008). Introduction to special issue: Evidence-based assessment in pediatric psychology. *Journal of Pediatric Psychology*, *33*(9), 911–915. https://doi.org/10.1093/jpepsy/jsj115.

Craig, K. D., Whitfield, M. F., Grunau, R. V., Linton, J., & Hadjistavropoulos, H. D. (1993). Pain in the preterm neonate: Behavioural and physiological indices. *Pain*, *52*(3), 287–299.

Crellin, D. J., Babl, F. E., Sullivan, T. P., Cheng, J., O'Sullivan, R., & Hutchinson, A. (2011). Procedural restraint use in preverbal and early-verbal children. *Pediatric Emergency Care*, *27*(7), 622–627.

Crellin, D. J., Harrison, D., Santamaria, N., & Babl, F. E. (2015). Systematic Review of the FLACC scale for assessing pain in infants and children: Is it reliable, valid, and feasible for use? *Pain*, *156*(11), 2132–2151. https://doi.org/10.1097/j.pain.0000000000000305.

Crellin, D. J., Sullivan, T. P., Babl, F. E., O'Sullivan, R., & Hutchinson, A. (2007). Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. *Paediatric Anaesthesia*, *17*(8), 720–733.

Dyer, S. L., Collins, C. T., Baghurst, P., Saxon, B., & Meachan, B. (2004). Insuflon versus subcutaneous injection for cytokine administration in children and adolescents: A randomized crossover study. *Journal of Pediatric Oncology Nursing*, *21*(2), 79–86.

Fallah, R., Gholami, H., Ferdosian, F., & Binesh, F. (2016). Evaluation of vaccines injection order on pain score of intramuscular injection of diphtheria, whole cell pertussis and tetanus vaccine. *Indian Journal of Pediatrics*. https://doi.org/10.1007/s12098-016-2186-z.

Fein, J. A., Zempsky, W. T., Cravero, J. P., & The Comittee on Pediatric Emergency Medicine Section on Anesthesiology Pain Medicine (2012). Relief of pain and anxiety in pediatric patients in emergency medical systems. *Pediatrics*, *130*(5), e1391–e1405. https://doi.org/10.1542/peds.2012-2536.

Girish, G. N., & Ravi, M. D. (2014). Vaccination related pain: Comparison of two injection techniques. *Indian Journal of Pediatrics*, *81*(12), 1327–1331https://doi.org/10.1007/s12098-014-1347-1.

Hillgrove-Stuart, J., Riddell, R. P., Horton, R., & Greenberg, S. (2013). Toy-mediated distraction: Clarifying the role of distraction agent and preneedle distress in toddlers. *Pain Research & Management: The Journal of the Canadian Pain Society*, *18*(4), 197.

Hogan, M. E., Probst, J., Wong, K., Riddell, R. P., Katz, J., & Taddio, A. (2014). A randomized-controlled trial of parent-led tactile stimulation to reduce pain during infant immunization injections. *Clinical Journal of Pain*, *30*(3), 259–265https://doi.org/10.1097/AJP.0b013e318296079e.

Howard, R., Carter, B., Curry, J., Morton, N., Rivett, K., Rose, M., ... Ireland (2008). Pain assessment. *Paediatric Anaesthesia*, *18*(Suppl. 1), 14–18.

Ipp, M., Cohen, E., Goldbach, M., & Macarthur, C. (2004). Effect of choice of measles-mumps-rubella vaccine on immediate vaccination pain in infants. *Archives of Pediatrics & Adolescent Medicine*, *158*(4), 323–326.

Ipp, M., Parkin, P. C., Lear, N., Goldbach, M., & Taddio, A. (2009). Order of vaccine injection and infant pain response. *Archives of Pediatrics & Adolescent Medicine*, *163*(5), 469–472https://doi.org/10.1001/archpediatrics.2009.35.

Ipp, M., Taddio, A., Sam, J., Gladbach, M., & Parkin, P. C. (2007). Vaccine-related pain: Randomised controlled trial of two injection techniques. *Archives of Disease in Childhood*, *92*(12), 1105–1108. https://doi.org/10.1136/adc.2007.118695.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, *17*(1), 1–12.

Kassab, M., Sheehy, A., King, M., Fowler, C., & Foureur, M. (2012). A double-blind randomised controlled trial of 25% oral glucose for pain relief in 2-month old infants undergoing immunisation. *International Journal of Nursing Studies*, *49*(3), 249–256https://doi.org/10.1016/j.ijnurstu.2011.09.013.

Lee, G. Y., & Stevens, B. (2014). Neonatal and infant pain assessment. In P. J. McGrath, B. J. Stevens, S. M. Walker, & W. T. Zempsky (Eds.), *Oxford textbook of paediatric pain* (1 ed.). Oxford: Oxford University Press.

Lindh, V., Wiklund, U., Blomquist, H. K., & Hakansson, S. (2003). EMLA cream and oral glucose for immunization pain in 3-month-old infants. *Pain*, *104*(1–2), 381–388.

Lucas, N. P., Macaskill, P., Irwig, L., & Bogduk, N. (2010). The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal of Clinical Epidemiology*, *63*(8), 854–861https://doi.org/10.1016/j.jclinepi.2009.10.002.

McClellan, C. B., Cohen, L. L., & Joseph, K. E. (2003). Infant distress during immunization: A multimethod assessment. *Journal of Clinical Psychology in Medical Settings*, *10*(4), 231–238.

McGowan, A., Cottrell, S., Roberts, R., & Lankshear, A. (2013). Minimising pain response during routine infant immunisation. *Community Practitioner*, *86*(6), 24–28.

McGrath, P. J., Johnson, G., Goodman, J. T., Schillinger, J., Dunn, J., & Chapman, J. (1985). CHEOPS: A behavioural scale for rating postoperative pain in children. In H. L Fields, R. Dubner, & F. Cervero (Eds.), *Advances in pain research and therapy*. 9. (pp. 395–402). New York: Raven Press.

Mijovic, B., Silva, M., Van den Bergh, B., Allegaert, K., Aerts, J. -M., Berckmans, D., & Van Huffel, S. (2010). Assessment of pain expression in infant cry signals using Empirical Mode Decomposition. *Methods of Information in Medicine*, *49*(5), 448–452.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Mularoni, P. P., Cohen, L. L., DeGuzman, M., Mennuti-Washburn, J., Greenwald, M., & Simon, H. K. (2009). A randomized clinical trial of lidocaine gel for reducing infant distress during urethral catheterization. *Pediatric Emergency Care*, *25*(7), 439–443.

O'Brien, L., Taddio, A., Ipp, M., Goldbach, M., & Koren, G. (2004). Topical 4% amethocaine gel reduces the pain of subcutaneous measles-mumps-rubella vaccination. *Pediatrics*, *114*(6), e720–724.

Pillai Riddell, R., Flora, D. B., Stevens, S., Greenberg, S., & Garfield, H. (2014). The role of infant pain behaviour in predicting parent pain ratings. *Pain Research & Management: The Journal of the Canadian Pain Society*, *19*(5), e124.

Pillai Riddell, R., Flora, D. B., Stevens, S. A., Stevens, B., Cohen, L. L., Greenberg, S., & Garfield, H. (2013). Variability in infant acute pain responding meaningfully obscured by averaging pain responses. *Pain*, *154*(5), 714–721.

Pillai Riddell, R., Racine, N. M., Turcotte, K., Uman, L. S., Horton, R. E., Din Osmun, L., ... Gerwitz-Stern, A. (2011). Non-pharmacological management of infant and young child procedural pain. *Cochrane Database of Systematic Reviews*, *10*. https://doi.org/10.1002/14651858.CD006275.pub2 http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006275.pub2/abstract (Retrieved from).

Pillai Riddell, R., Stevens, B. J., Cohen, L. L., Flora, D. B., & Greenberg, S. (2007). Predicting maternal and behavioral measures of infant pain: The relative contribution of maternal factors. *Pain*, *133*(1), 138–149.

Prkachin, K. M. (2009). Assessing pain by facial expression: Facial expression as nexus. *Pain Research & Management*, *14*(1), 53–58.

Racine, N. M., Pillai Riddell, R. R., Flora, D., Garfield, H., & Greenberg, S. (2012). A longitudinal examination of verbal reassurance during infant immunization: occurrence and examination of emotional availability as a potential moderator. *Journal of Pediatric Psychology*, *37*(8), 935–944. https://doi.org/10.1093/jpepsy/jss066.

Royal Australasian College of Physicians, & Paediatric and Child Health Division (2005). Guideline statement: Management of procedure-related pain in children and adolescents. Retrieved from Sydney http://www.racp.edu.au/hpu/paed/pain/index.htm.

Royal College of Nursing (2009). *The recognition and assessment of acute pain in children. Update of full guideline*. London: Royal College of Nursing.

Schug, S., Palmer, G., Scott, D., Halliwell, R., Trinca, J., Working Group of the Australian, & New Zealand College of Anaesthetists and Faculty of Pain Medicine (2015). *Acute pain management: Scientific evidence* (4th ed.). Melbourne: ANZCA & FPM.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Silva, M., Mijovic, B., Van den Bergh, B. R., Allegaert, K., Aerts, J. M., Van Huffel, S., & Berckmans, D. (2010). Decoupling between fundamental frequency and energy envelope of neonate cries. *Early Human Development*, *86*(1), 35–40. https://doi.org/10.1016/j.earlhumdev.2009.12.006.

Stevens, B., Johnston, C., Taddio, A., Gibbins, S., & Yamada, J. (2010). The premature infant pain profile: Evaluation 13 years after development. *Clinical Journal of Pain, 26*(9), 813–830.
Stevens, S. A., Racine, N., Riddell, R. P., Horton, R. E., Garfield, H., & Greenberg, S. (2013). Infant pain regulation as an early indicator of childhood temperament. *Pain Research & Management: The Journal of the Canadian Pain Society, 18*(6), 313.
Streiner, D. L., & Kottner, J. (2014). Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing, 70*(9), 1970–1979. https://doi.org/10.1111/jan.12402.
Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use*. Oxford University Press.
Taddio, A., Flanders, D., Weinberg, E., Lamba, S., Vyas, C., Ilersich, A. F., ... McNair, C. (2015). A randomized trial of rotavirus vaccine versus sucrose solution for vaccine injection pain. *Vaccine, 33*(25), 2939–2943https://doi.org/10.1016/j.vaccine.2015.04.057.
Taddio, A., Ho, T., Vyas, C., Thivakaran, S., Jamal, A., Ilersich, A. F., ... Shah, V. (2014). A randomized controlled trial of clinician-led tactile stimulation to reduce pain during vaccination in infants. *Clinical Pediatrics, 53*(7), 639–644https://doi.org/10.1177/0009922814526976.
Taddio, A., Hogan, M. E., Moyer, P., Girgis, A., Gerges, S., Wang, L., & Ipp, M. (2011). Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. *Vaccine, 29*(7), 1390–1394https://doi.org/10.1016/j.vaccine.2010.12.051.
Taddio, A., Nulman, I., Goldbach, M., Ipp, M., & Koren, G. (1994). Use of lidocaine-prilocaine cream for vaccination pain in infants. *Journal of Pediatrics, 124*(4), 643–648.
Taddio, A., Nulman, I., Koren, B. S., Stevens, B., & Koren, G. (1995). A revised measure of acute pain in infants. *Journal of Pain and Symptom Management, 10*(6), 456–463.
Taddio, A., O'Brien, L., Ipp, M., Stephens, D., Goldbach, M., & Koren, G. (2009). Reliability and validity of observer ratings of pain using the visual analog scale (VAS) in infants undergoing immunization injections. *Pain, 147*(1–3), 141–146.
Taddio, A., Parikh, C., Yoon, E. W., Sgro, M., Singh, H., Habtom, E., ... Shah, V. (2015). Impact of parent-directed education on parental use of pain treatments during routine infant vaccinations: A cluster randomized trial. *Pain, 156*(1), 185–191.
Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., ... de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42.
Tomlinson, D., von Baeyer, C. L., Stinson, J. N., & Sung, L. (2010). A systematic review of faces scales for the self-report of pain intensity in children. *Pediatrics, 126*(5), e1168–1198. https://doi.org/10.1542/peds.2010-1609.
Verriotis, M., Fabrizi, L., Lee, A., Ledwidge, S., Meek, J., & Fitzgerald, M. (2015). Cortical activity evoked by inoculation needle prick in infants up to one-year old. *Pain, 156*(2), 222–230.

7.1 Additional material

The supplemental tables found online that are referred to in this paper are reproduced in Appendix E.

7.2 Addendum: New literature

An updated search was completed in the last week of May 2018 to identify studies published since the original search was completed in July 2016. Since publication of the systematic review summarising MBPS psychometric data (332) and our prospective study (464) no psychometric evaluation studies have been published to provide additional evidence and there are no new recommendations regarding pain assessment. Seven RCTs using the MBPS to measure a study outcome have been published and all but one was focused on interventions to alleviate immunisation related pain. This study evaluated the effect of lidocaine on the pain associated with urinary catheter insertion and the results do not add evidence to support the capacity of the MBPS to measure pain (465). Of the six RCTs concentrating on immunisation related pain, only two studies provided additional data to support the capacity of the MBPS to assess immunisation related pain (466, 467).

7.2.1 Implications

There is no data to alter the conclusions drawn from the original review regarding MBPS use to assess pain for procedures other than immunisations. Furthermore, the studies that have been published since the review also focused on immunisation-related pain and do not change the weight of evidence sufficiently to alter the conclusions or recommendations of the published systematic review regarding use of the MBPS for immunisation-related pain assessment.

SECTION 4.

Section 4 is comprised of five chapters that describe the methods used to conduct a prospective evaluation of the psychometric properties of the Face, Legs Activity, Cry and Consolability (FLACC) scale, the Modified Behavioural Pain Scale (MBPS) and the Visual Analogue Scale applied by an observer (VASobs) used to assess procedural pain in infants and children aged 6 to 42 months. A PDF of the publication of the methods is provided in Chapter 8 and the results of this study are reported in Chapters 9 to 12: the psychometrics of each scale and a comparison of the psychometrics of the scales. The results for the FLACC scale and the MBPS are also presented as PDFs of the published versions of these results (Chapters 10 and 11).

CHAPTER 8.

Study protocol for an evaluation of the psychometric properties of the FLACC scale, MBPS and the VASobs

This chapter reports the study design for the prospective evaluation of the psychometric properties of the Face, Legs, Activity, cry and Consolability (FLACC) scale, the Modified Behavioural Pain Scale (MBPS) and the Visual Analogue Scale applied by an observer (VASobs). This has been published in BMJ Open and the PDF of this publication is presented in this chapter.

Publication:

Crellin DJ, Harrison D, Hutchinson A, Schuster T, Santamaria N, Babl FE. Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months. BMJ Open. 2017;7(9).

Open Access Protocol

BMJ Open

Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months

Dianne J Crellin,[1,2,3] Denise Harrison,[1,2,4] Adrian Hutchinson,[3] Tibor Schuster,[5] Nick Santamaria,[1] Franz E Babl[2,3,6]

To cite: Crellin DJ, Harrison D, Hutchinson A, *et al*. Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months. *BMJ Open* 2017;**7**:e016225. doi:10.1136/bmjopen-2017-016225

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-016225).

Received 1 February 2017
Revised 24 May 2017
Accepted 29 June 2017

For numbered affiliations see end of article.

Correspondence to
Dianne J Crellin;
dianne.crellin@rch.org.au

ABSTRACT

Introduction Infants and children are frequently exposed to painful medical procedures such as immunisation, blood sampling and intravenous access. Over 40 scales for pain assessment are available, many designed for neonatal or postoperative pain. What is not well understood is how well these scales perform when used to assess procedural pain in infants and children.

Aim The aim of this study was to test the psychometric and practical properties of the Face, Legs, Activity, Cry and Consolability (FLACC) scale, the Modified Behavioural Pain Scale (MBPS) and the Visual Analogue Scale (VAS) observer pain scale to quantify procedural pain intensity in infants and children aged from 6–42 months to determine their suitability for clinical and research purposes.

Methods and analysis A prospective observational non-interventional study conducted at a single centre. The psychometric and practical performance of the FLACC scale, MBPS and the VAS observer pain scale and VAS observer distress scale used to assess children experiencing procedural pain will be assessed. Infants and young children aged 6–42 months undergoing one of four painful and/or distressing procedures were recruited and the procedure digitally video recorded. Clinicians and psychologists will be recruited to independently apply the scales to these video recordings to establish intrarater and inter-rater reliability, convergent validity responsiveness and specificity. Pain score distributions will be presented descriptively; reliability will be assessed using the intraclass correlation coefficient and Bland-Altman plots. Spearman correlations will be used to assess convergence and linear mixed modelling to explore the responsiveness of the scales to pain and their capacity to distinguish between pain and distress.

Ethics and dissemination Ethical approval was provided by the Royal Children's Hospital Human Research Ethics Committee, approval number 35220B. The findings of this study will be disseminated via peer-reviewed journals and presented at international conferences.

Strengths and limitations of this study

- Publication providing details of the methods used for a psychometric evaluation study examining pain scales.
- Multiple strategies used for validation where no 'gold standard' is available.
- Methods used to reduce bias for evaluation of responsiveness.
- Methods used to reduce bias resulting from the application of multiple scales.
- Large sample sizes of reviewers and participants.
- Single-centre study.
- Reviewers not blinded to circumstances of procedure, which may bias their application of scales.

INTRODUCTION

Pain is a common feature of illness and injury in infants and children and assessment and treatment frequently involve pain-inducing procedures such as blood sampling and intravenous access. Despite the increasing weight of evidence that infants' and children's experience of pain has a negative impact on short-term and long-term outcomes,[1–4] pain continues to be poorly managed, particularly in infants and children presenting to emergency departments (EDs).[5–12] Reasons cited for the suboptimal treatment of pain and distress in EDs include among other factors, poor recognition of significant pain by medical providers.[13] The generally accepted standard for pain assessment is self-report. However, infants and children less than 3 years of age are unable to self-report pain and there are some doubts about the capacity of children aged 3–5 years to self-report

pain using traditional scales designed for young children.[14–16] Behavioural observation scales are one of the most commonly used alternatives to self-report. Over 40 tools have been identified in the literature, many of which were designed for either neonates or infants and children experiencing postoperative pain.[17] Each of these scales has been subjected to varying levels of psychometric testing, and many have been used in a variety of circumstances other than for which they were originally intended. What is not well understood is how well these scales perform when used to assess procedural pain in infants and children.

Current evidence suggests that available scales may not be practical or psychometrically suitable for procedural pain assessment and that they may have difficulty in differentiating pain from other distress-related behaviours.[18 19] Furthermore, many of these scales do not use the commonly accepted 0–10 metric. Only two scales have been designed specifically for or to include assessment of procedural pain in infants and/or children: the Modified Behavioural Pain Scale (MBPS)[20] and EVENDOL.[21] However, they are not supported by sufficient feasibility or psychometric data to unreservedly support their use. MBPS has had some testing to establish scale performance when used to assess immunisation pain in infants, and the results are promising.[22] The MBPS has not been widely tested to establish its measurement properties when used to assess pain associated with other procedures such as blood sampling, intravenous cannula insertion and other diagnostic and therapeutic procedures. EVENDOL was specifically developed to measure pain experienced by infants and children in the ED and was tested on children presenting with acute pain and those experiencing procedural pain. This scale was considered unsuitable for several reasons.[23] The scale items are scored based on 'duration' and 'intensity', making scoring ambiguous. For example, reviewers are likely to be confused about how to score a brief (low score) but intense reaction (high score) or a frequent but low-intensity response. Furthermore, the maximum EVENDOL score is 15. The accepted metric for pain assessment is 0–10 to standardise scoring to improve the clinical usefulness of the scores.

In the absence of a purposefully designed scale, scales designed for alternative populations and circumstances have been repeatedly used clinically and in research to assess procedural pain. Only a small number of these scales have been subjected to psychometric evaluation for this purpose and are recommended following systematic review or in well-supported clinical practice guidelines. The Face, Legs, Activity, Cry and Consolability (FLACC) scale and the Visual Analogue Scale (VAS) applied by an observer (VAS observer) are two such scales.

The FLACC scale, designed to assess postoperative pain in young children, is one of the most well-known and most commonly used scale.[24] It has been used extensively as an outcome measure in studies examining procedural pain and procedural pain management strategies. However, a recent systematic review of the psychometrics of the FLACC scale raises a number of questions about the validity and feasibility of the scale and concludes that there is currently insufficient data to accept the scale as reliable and valid for procedural pain assessment.[24]

The VAS observer is commonly used in clinical trials and other studies to measure pain intensity for children unable to self-report. Data addressing the psychometric properties of this scale used to assess procedural pain in infants and young children are limited. The authors of a review in 2002 conclude that insufficient data are available to confidently support the psychometric properties of the VAS observer and that studies addressing reliability, responsiveness and cut-offs are needed.[25] Our recent and, as yet, unpublished review of the evidence supporting the psychometrics does little to change this conclusion.

Integral to effective pain management is accurate assessment of pain, and it has been shown that mandating the use of pain assessment improves analgesic administration in the ED.[26] Furthermore, clinical trials testing the efficacy of pain management strategies depend on the availability of instruments to measure trial outcomes with a tool likely to provide valid results. It is therefore essential that appropriate and validated means to assess pain are identified. The aim of this study is to test the psychometric and practical properties of three scales for clinical and research purposes that have been either designed for procedural pain assessment (MBPS) and/or are used and recommended for this purpose (FLACC scale and VAS observer pain scale) to quantify procedural pain intensity in infants and children aged from 6–42 months to determine their suitability .

Study objectives

Primary objectives

The primary objectives of this study are to test the (1) feasibility, (2) reliability, (3) validity and (4) clinical utility of the FLACC scale, the MBPS and VAS observer pain scale for assessing procedural pain intensity in infants and children aged 6–42 months.

Secondary objectives

The study aims to meet several secondary objectives: (1) to determine whether there is a difference in the inter-rater and intra-rater reliability of the scale when applied by clinicians compared with application by *clinically naive* (italicised terms are defined in the accompanying 'Definition list') reviewers (psychologist researchers) and (2) to establish whether reviewing the phases of a procedure in sequence influences the scores allocated to each phase.

METHODS

Study overview

This study will use a prospective observational non-interventional design and will be conducted in the ED of a tertiary paediatric hospital in Melbourne, Australia. The

Royal Children's Hospital (RCH) ED, Melbourne has an annual census of approximately 90 000 children.

We will assess and compare the psychometric and practical performance of the FLACC scale, MBPS and the VAS observer pain scale and VAS observer distress scale when used to assess infants and young children experiencing procedural pain. Infants and young children aged 6–42 months undergoing one of four painful and/or distressing procedures in the ED were recruited, and the procedure was digitally video recorded to create a dataset for review. Demographic and clinical data will be collected during the ED presentation. Video recordings of each procedure will be independently reviewed by the recruited reviewers and assessed at three different time points using a behavioural pain scale. A sample of clinicians and psychologists will be recruited to complete these reviews.

The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) Checklist was used to support the development of the design for this study. This checklist was developed to provide standards for evaluating the methodological quality of studies addressing the psychometric properties of health measurement instruments but can be used to guide design and reporting of a study.[27]

Sample

Two samples are required for this study: patients (infants and children experiencing a medical procedure) and reviewers (ED clinicians and clinically naive reviewers).

Patient participants

Demographic and video-recorded data for participants recruited prospectively will be used in this study. This is a convenience sample of infants and children aged between 6 and 42 months presenting to the ED who were filmed while experiencing one of four nominated painful and/or distressing procedures. The inclusion and exclusion criteria are provided in table 1. Data have been collected for 132 children.

Reviewer participants

The reviewers will be recruited prospectively from two cohorts for this study: clinicians from the RCH ED and research psychologists (clinically naive reviewers) affiliated with or appointed to RCH or the campus research partner, the Murdoch Children's Research Institute (MCRI).

Eligible clinicians include qualified doctors and nurses of any level of experience practising in the ED and clinically naive reviewers are psychologists who have completed at least their basic training and are therefore recognised as at least a provisional psychologist .

Studies addressing the psychometric properties of pain scales frequently use non-clinical research assistants to apply the scale to generate a pain score. The results from these studies are used to claim validity of the scale for clinical use. However, as there is some evidence that clinicians apply clinical judgement when applying assessment scales,[28 29] using the two intended cohorts will provide an opportunity to test the assumption that the scale will perform similarly when applied by clinical and non-clinical raters.

Instruments

Scales that met the following criteria were selected for psychometric evaluation: observational scales using a 0–10 metric that were designed for procedural pain assessment, scales with psychometric data to support their capacity to generate valid procedural pain scores or scales recommended by systematic review or consensus clinical practice guidelines. The following scales, which are presented in tables 2 and 3 and figure 1, will be applied to each phase of each procedure: the FLACC scale, MBPS, VAS observer pain and VAS observer distress. Once the reviews are completed, reviewers will complete a clinical utility questionnaire. An electronic data management (EDM) tool was developed for this study to present the videos and collect reviewer data.

FLACC scale

The FLACC scale (table 2) was developed as a more practical alternative to existing pain scales and first published in 1997.[30] It is a composite of five behaviours considered indicative of pain that can be detected and graded by an observer and easily remembered using the acronym 'FLACC' ('face', 'legs', 'activity', 'cry' and 'consolability'). Each item is scored on a 0–2 scale resulting in a maximum score of 10. The FLACC scale was originally designed and validated for use in infants and children aged 2 months to 7 years to measure postoperative pain. The original instructions for use recommended observing the child

Table 1 Inclusion and exclusion criteria for infants recruited

Inclusion criteria	Exclusion criteria
Infant/child aged 6–42 months Infant/child undergoing one of the following painful or non-painful procedures: 1. Intravenous cannula insertion 2. Nasogastric tube insertion 3. Inhaled medication delivery via mask and spacer device 4. Measurement of oxygen saturations (SpO_2)	Infant/child requires immediate treatment History of cognitive delay or altered conscious state, significant comorbid disease Parent/guardian non-English speaking Recordings of the procedure incomplete Inadequate video quality

Table 2 Faces, Legs, Activity, Consolabilty and Cry Scale

	Scoring		
Categories	0	1	2
Faces	No particular expression or smile	Occasional grimace or frown, withdrawn, disinterested	Frequent to constant quivering chin, clenched jaw
Legs	Normal position or relaxed	Uneasy, restless, tense	Kicking or legs drawn up
Activity	Lying quietly, normal position, moves easily	Squirming, shifting back and forth, tense	Arched, rigid or jerking
Cry	No cry (awake or asleep)	Moans or whimpers, occasional complaint	Crying steadily, screams or sobs, frequent complaints
Consolability	Content, relaxed	Reassured by occasional touching, hugging or being talked to, distractible	Difficult to console or comfort

Item scores are summed to generate a total score of 10.

for 1–5 min and matching the observed behaviours to those described in the scale for each item.

MBPS

MBPS (table 3) is a modification of an earlier paediatric pain scale (the Children's Hospital of Eastern Ontario Pain Scale) and was designed to better capture the variability of young infant responses to pain.[20] Furthermore, the scale was specifically aimed at assessing procedural pain and much of the validation data are derived from studies including infants undergoing routine immunisation. MBPS is a behavioural scale composed of three behaviours: facial expression, cry and body movements. Each of the behaviours included is assessed and scored and the scores added to generate a pain intensity score from 0 to 10. In the original validation studies, observers watched 5 s of video footage of the infant prior to the procedure and 15 s of the infant during the procedure and instructed to score the maximum reaction that occurred during the time observed.

VAS observer (pain and distress)

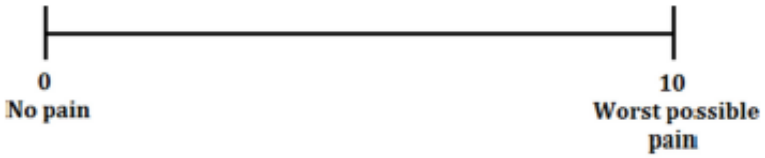
VAS is a tool designed to measure and quantify subject experiences such as pain and distress.[31] The scale is a 10 cm line anchored at either end with labels such as 'no pain' and 'worst possible pain' or no distress' and 'worst possible distress'. When applied by an observer, they are asked to estimate the intensity of the pain or distress observed by placing a mark on the line. The distance from the zero point on the line is measured, and this represents the pain score.

The VAS observer scale is included to gain an estimate of the level of distress that reviewers perceived the infant to be experiencing during the phases of the procedure. These scores will be used to explore the

Table 3 Modified Behavioural Pain Scale

Item	Descriptor	Score
Facial expression	Definite positive expression (smiling)	0
	Neutral expression	1
	Slightly negative expression (grimace)	2
	Definite negative expression (furrowed brow eyes closed tightly)	3
Cry	Laughing or giggling	0
	Not crying	1
	Moaning quiet vocalising gentle or whimpering cry	2
	Full lunged cry or sobbing	3
	Full lunged cry more than baseline cry (scored only if child crying at baseline)	4
Movements	Usual movements and activity	0
	Resting and relaxed	0
	Partial movement (squirming arching limb tensing clenching)	2
	Attempt to avoid pain by withdrawing the limb where puncture is done	2
	Agitation with complex/generalised movements involving the head torso or other limbs	3
	Rigidity	3

Item scores are summed to generate a total score of 10.



NOTE: A mark is placed on the line at the point that represents the level of pain observed. This is measured in millimeter from the left anchor 'no pain' to generate a pain score. The word 'distress' replaces 'pain' to create a distress scale

Figure 1 The Visual Analogue Scale.

capacity of the pain scales to discriminate between pain and distress.

Feasibility and utility questionnaire

A feasibility and clinical utility questionnaire was used to capture the reviewers' assessments of how easy the scale is to use and how well it performs (see table 4). The utility scale was developed by de Jong and colleagues,[32] based on utility criteria defined by Harris and Warren,[33] and includes nine statements that are rated using a 5-point Likert Scale to assess the extent to which the reviewer agrees with the statement.

Data collection tool

An EDM system has been specifically designed and developed for this study. The EDM system allowed reviewers to watch and review the footage of the procedures while simultaneously entering data into the database. The system will also record time stamps to allow for time-related variables. The interface used by the reviewer is shown in figure 2. The video data collection tool was developed specifically for this study by one of the researchers using FileMaker 7.0 (FileMaker, Santa Clara, California).

Study procedure

Recruitment and consent

Patient participants

Infants and children meeting the inclusion criteria for the study were identified by a member of the clinical staff or a research team member and then approached to participate. Recruitment occurred when a member of the research team was available to complete data collection. Parents/guardians were provided with the study information sheet, an opportunity to answer questions and an assurance that participation was voluntary and that their decision would not impact on their child's care.

Written consent to participate in the study was provided by legal guardians of the infants and children presenting to the ED and verbal consent from the staff present during the procedure.

Reviewer participants

Clinicians and psychologists appointed/affiliated to the ED and MCRI, respectively, will be recruited using similar strategies. Hard-copy notices in the two departments and

Table 4 Feasibility and clinical utility questionnaire

Statement	Scale				
1. Provides information that is clinically useful	1 ▫ Clinically not very useful	2 ▫	3 ▫	4 ▫	5▫ Clinically very useful
2. Is it clear and easy to understand	1 ▫ Not clear and easy	2 ▫	3 ▫	4 ▫	5▫ Clear and easy
3. Is quick to apply	1 ▫ Very slow	2 ▫	3 ▫	4 ▫	5 ▫ Very quick
4. Is easy to apply	1 ▫ Very difficult	2 ▫	3 ▫	4 ▫	5 ▫ Very easy
5. Reflects the extent of procedural pain	1 ▫ Does not reflect at all	2 ▫	3 ▫	4 ▫	5 ▫ Reflects the extent well
6. Discriminates children with pain from children without pain	1 ▫ Does not discriminate at all	2 ▫	3 ▫	4 ▫	5 ▫ Discriminates well
7. Score is readily understood and supports decisions about pain management	1 ▫ Not readily understood and doesn't support	2 ▫	3 ▫	4 ▫	5 ▫ Readily understood and supports decisions decisions
8. Reflects procedural pain-specific features	1 ▫ Doesn't reflect procedural pain related features	2 ▫	3 ▫	4 ▫	5 ▫ Reflects procedural pain related features

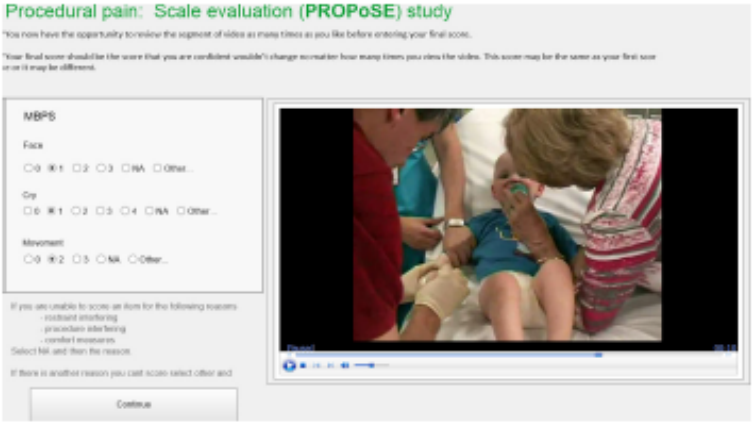


Figure 2 Screenshot of the electronic data management system that will be used to capture reviewer data. MBPS, Modified Behavioural Pain Scale.

electronic notices via existing closed department social media forums and other electronic communication systems will be posted to advertise the study. An email distribution list in the ED will be used to circulate the study information sheet and a generic invitation to participate. Psychologists will be identified by MCRI research theme heads who will forward the study information sheet and the invitation to participate.

The invitation will advise those interested in participating to return a signed copy of the consent form directly to the principal investigator.

Data collection

Demographics and clinical data were collected during the ED visit and recorded on a piloted case report form. The procedure was video recorded to create a video dataset for later review by the recruited clinician and psychologist reviewers. Clinicians and parents scored the child's pain and distress during the procedure using VAS.

The reviewers will use the FLACC scale, MBPS and VAS to assess the pain and VAS to assess the distress experienced by the infant/child shown in the video. Clinicians will also be asked to identify the pain and distress management strategies that they would use to manage the experience of the infant/child. For each new segment of video, reviewers will allocate a 'first' score and a 'final' score to establish clinical utility. The time taken for reviewers to allocate the first score will also be recorded. Reviewers will also be asked a series of questions to establish their judgement of the feasibility and utility of the scale.

Data collection will involve a series of steps that are shown in sequence in figure 3 and described in detail in the following sections.

ED presentation data collection

A member of the research team was responsible for collecting demographic and clinical data for consented infants and video recording the procedure. A hand-held video recorder was used, and researchers aimed to focus on the infant to capture their face and body. Recordings were commenced from the time the infant and their parent/caregiver were moved to the procedure area but before any contact to prepare the infant for the procedure occurred. The recording ended once the procedure had been completed.

All clinical decisions were made by the treating clinicians and based on department and hospital guidelines.

This may have resulted in the application of topical anaesthesia prior to intravenous cannula insertion, lubrication of the nasogastric tube (NGT) prior to insertion and comfort and/or distraction during the procedure. Infants and children having an intravenous cannula inserted lay flat or in a semirecumbent position on the trolley and those having an NGT inserted lay flat. Restraint was used as required, and this was provided in most circumstances by a member of the nursing staff who restrained either the involved limb or the torso and limb or torso (intravenous insertion) and head (NGT insertion). Finally, infants and children requiring inhaled medication or had their O_2 saturation measured either sat on their parents lap or independently on the trolley or a chair. Where restraint was needed, this involved either stabilising the mask on the face and keeping the child on the parent's lap or restraining the limb. No further effort was made to standardise the procedures. Parents were present for the procedure.

Video data preparation

The video recordings of the procedures will be reviewed to select recordings of sufficient quality to allow reviewers to apply the scales. This requires that the face and body movements of the infant are visible to make application of the scales possible. In the event that a larger number of recordings are eligible, participants and their recordings will be randomly selected.

The video recordings will be divided into segments to demonstrate different phases of the procedure, which are defined as follows:

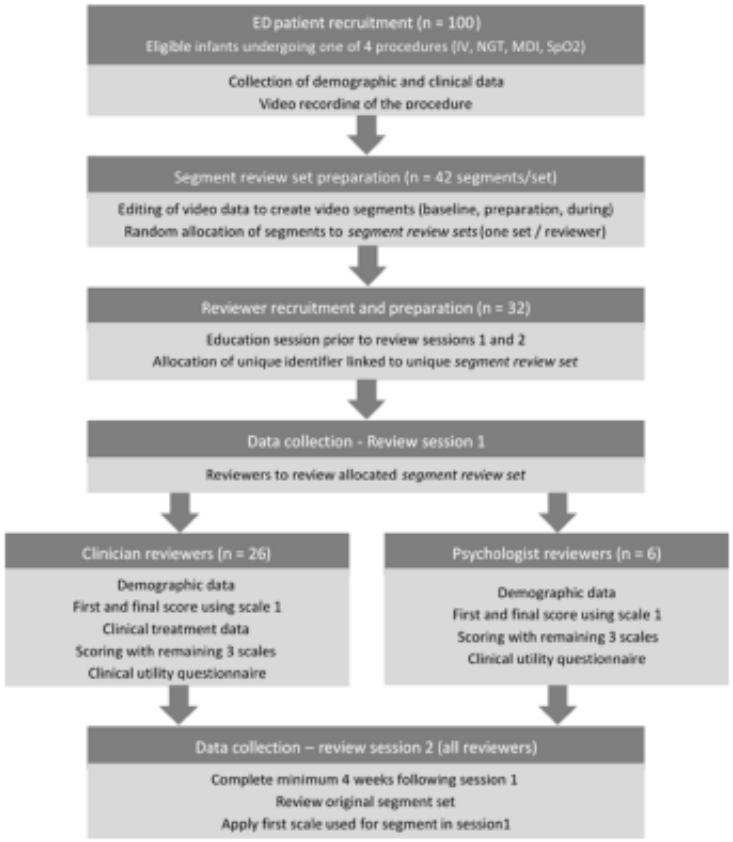


Figure 3 Study procedure. ED, emergency department; IMD, Inhaled medication delivery; IV, intravenous; NGT, nasogastric tube insertion; SpO2, oxygen saturation measurement.

1. Baseline (B): before any attempt to prepare the infant/child for the procedure is made (eg, while still in parents arms).
2. Preparation (P): preparation phase of the procedure (eg, while restrained but prior to painful stimulus).
3. During (D): during the painful/distressing part of the procedure (eg, within 5 s of needle insertion).

The procedures presumed painful will be divided into all three segments and the procedures presumed distressing but not painful will only be divided into two segments (B and D) as these procedures cannot be separated into a non-painful contact phase to prepare the infant and a painful procedural phase. This will result in a total of 260 segments of video for review.

Each video segment will be 15 s long and show the infant/child's face and body. The segments will be grouped by procedure and allocated to review sets (one per reviewer) to ensure that the following criteria are met:

1. Each review set has similar numbers of segments from each procedure and each procedural phase.
2. Review sets include different combinations of segments.
3. Each segment is included in the same number of review sets.
4. A review set does not contain more than one segment from the same procedure (infant/child).

These criteria are designed to ensure that all segments are reviewed by the same number of reviewers, that reviewers provide assessments of a range of procedures and phases but never for the same child and that different combinations of reviewers review each segments. Allocation will be automated using a Stata[34] script to prevent bias occurring with manual allocation of segments to review sets.

The four scales will be used in varying order to assess each video segment. The sequencing of the scales will be randomly allocated with only one stipulation: that each scale is applied first on equal numbers of occasions. This sequencing will be generated using a random sequence generator (https://www.random.org/sequences/).

The reviewer will access the system with their unique study identification number, which will ensure that they

assess the video segments allocated to their review set and that the data that they enter are recorded in the database against their unique study number.

Reviewer preparation

The reviewers will attend an education session before they commence the data collection stage of the study. They will be familiarised with the EDM system and each of the scales that they will be using to assess pain and distress. The reviewers will have an opportunity to trial the data collection system and apply each of the scales during this session. Reviewers will be allocated a unique study number. These training sessions are designed to replicate the training used to prepare clinicians to use an assessment tool in practice. No attempt will be made to improve inter-rater or intrarater reliability before data collection as we are interested to evaluate reliability among clinicians, replicating as close as possible the clinical circumstances in which these scales are used.

Reviewer data collection

Each reviewer will complete two review sessions, a minimum of 4 weeks apart. On each occasion, the principal investigator will set up the EDM system and provide the reviewer with headphones and laminated copies of the FLACC scale and MBPS and a ruler to use for the VAS observer pain and VAS observer distress. The reviewer will log in using their unique study number to access their review set.

Review session 1

On the first occasion, they will complete the demographic data section of the EDM system and provide an assessment of each segment in their review set using the scales.

The video segments will be loaded for the reviewer to watch, and a randomly selected scale will be presented beside the video viewing window. The reviewer will not be able to stop the video segment from playing or rewind and review the video until they have entered their first score. This is intended to as closely as possible replicate real-time clinical pain assessment using this scale. Once the score has been entered, the clinician reviewer will be asked several questions about the treatment that they considered necessary for the infant/child in the video segment. This will include checkboxes and an option for free text. Then they will be able to review the video as many times as they like before entering their final pain score using this scale. The final score is the one that they consider unlikely to change regardless of how many more times they watch the video. The database will time stamp the start of the video for the first viewing of the video and entry of their first score. Once these scores have been entered, the reviewer will score the segment again using the remaining scales presented in the preallocated sequence. They will have the option of watching the video segment again as many times as required to support application of the other scales.

Once all reviews are completed, reviewers will be asked to complete the feasibility and utility questionnaire.

Review session 2

Reviewers will be asked to provide assessments for the same review set (eg, same video segments) as was used in review session 1. However, they will only be asked to use one scale (the first scale used for the segment in the first review) and they will not be asked any questions about treatment or to respond to the feasibility/clinical utility statements. For half of the video segments reviewed at review session 2, the EDM will present all segments (phases) of the procedure in sequence to be watched before the reviewer views the segment for assessment and applies the nominated scale.

Data management

The patient participant data, which include video and demographic data, are identified by their study number and stored in password-protected folders on a secure network drive. The consent forms will be stored separately in secure storage.

Reviewer participants will also be identified by a unique study number, and all data collected will be identified by this study number. This data will also be stored in files stored in password-protected folders on a secure network drive. A password-protected file stored separately will match the reviewer name with their unique study number to ensure that data from their two review sessions can be matched for data analysis. Signed consent forms will be secured in a locked cupboard.

Access to all data and video files and consent forms is restricted to members of the research team. Database access (EDM system) for the purposes of data collection will only be possible from a private office computer via password. Individual reviewer dataset access from the EDM system will be further restricted to their review sets by their unique study number. The data will be kept until all participants have reached 25 years of age.

Sample size

Sample size estimations for reliability testing using measures of agreement rely on an estimate of the true variation in the sample. There are data in other circumstances (eg, postoperative pain) but very limited data to establish the likely variability in scores associated with the medical procedures included in this study. The senior biostatistician informing protocol development advised that limited data and the inclusion of several procedures and multiple raters made estimating variation and therefore sample size difficult and unreliable. The advice was to base the sample size on current recommendations and the sample sizes used in similar psychometric evaluation studies. Therefore, the number of observations made by each observer in this study is based on the recommendations of the COSMIN Checklist.[27] These standards rate a sample of 50–99 as 'good' and over 100 as 'excellent'. Therefore, a sample of 100 children will be sought for this study.

The use of small numbers of raters in previously published psychometric evaluation studies assumes that the raters are representative of a larger pool of raters. Our decision to recruit a larger number of raters is based on our unwillingness to accept a largely untested assumption about the representativeness of the raters applying pain scales. The larger number of reviewers does not completely overcome this assumption but seeks to acknowledge the potential for variability between reviewers.

As studies addressing the psychometric properties of pain scales frequently use small numbers of raters (2–5) and observations (<50), logic suggests that larger numbers should increase our confidence in the results. However, we acknowledge that the reductions in error conferred by larger sample sizes are not linear. These substantial increases in sample sizes may only confer modest improvements in the margin of error in the results.

Patient participants

One hundred procedures will be included in this study: 60 presumed painful procedures (30 intravenous cannula insertions and 30 NGT insertions) and 40 distressing but presumed non-painful procedures (20 inhaled medication administrations via mask and 20 oxygen saturation measurements). A total of 260 segments of video will be created by dividing the segments into the phases described in a following section. All procedures will be reviewed by the reviewer clinicians, and a subset of 40 procedures (14 intravenous insertions, 14 NGT insertions, 7 inhaled medication administrations and 7 SpO2 measurements), to generate 112 segments, will be reviewed by the clinically naive reviewers.

Reviewer participants

Clinician reviewers

A sample of 25 clinicians will be recruited to the study. This number was chosen to ensure that each segment of video was reviewed by at least four clinicians and that the review sets contained the same number of segments in each and where not prohibitively large (42 segments each).

Psychologist reviewer

The aim is for at least two psychologists to review each segment of a subset of 40 procedures (106 segments in total). This will require a sample of six psychologists.

Statistical analysis

Statistical analyses will be conducted using the statistical software package 'R' (R Core Team (2016).[35]

Demographic data and pain scores

The demographic data collected from the reviewers and the demographics of the infants and children involved in the procedures that were collected at the time that the procedure was filmed will be summarised using descriptive statistics.

Reliability

Intraclass correlations will be calculated to establish the inter-rater and intrarater reliability of the scales. Coefficients will be calculated separately for clinician and psychologist reviewers. Reliability will be considered excellent for coefficients greater than 0.75. Bland-Altman plots will also be used to assess agreement.

Validity

Comparison between pain scale scores will be used to examine convergent validity. This will be achieved by calculating the Spearman correlation coefficient, and strong positive correlation between FLACC, MBPS and VASobs scores (r>0.75) will be considered to support our hypothesis that these scales measure the same construct. In contrast, a weak positive correlation between the pain scales and VASobs distress is expected, as pain-related and non-pain-related distress, although often linked, are different constructs. The responsiveness of the scale to changes in pain experience will be determined by analysis of the change in scores over the phases of the procedure using linear mixed models to estimate fixed effects of time (phase on procedure) and procedure type (painful vs non-painful). Children and reviewers will be considered as random effects. Random effects for children will be allowed to vary across sequences (nested random effects). CIs for fixed effects will be computed using bootstrap samples as implemented in the *confint* function in R (R package: stats). It is hypothesised that scores will be low during the baseline and preparation phases and will rise significantly during the procedure phase of a painful procedure. As the change in pain score is dependent on the real change in pain, it is not possible to establish an accepted standard for the extent to which pain scores must rise to accept the scale as responsive. As a change in scores of 2 is generally accepted as evidence of a clinically significant change (36), we will consider responsiveness demonstrated if the change in scores exceeds 2. This change in scores should not be seen for non-painful procedures.

Finally, the specificity of the scale for pain will be evaluated by grouping infants and children into those with pain scores greater than 3 during the baseline and/or preparation phase and those with scores less than 3 and comparing the change in score during the procedure phase of the procedure. We hypothesise that the extent of the change in score across phases will be similar for children in these groups, reflecting the capacity of the scale to distinguish between pain-related and non-pain-related distress.

Feasibility and clinical utility

The percentage of valid scores allocated for each scale will be described and compared. Where a valid score is not allocated, the reason for this will be summarised to establish the potential limitations to the scale.

The average time taken by reviewers to allocate a pain score using each of the pain scales following the first viewing

of the video will be compared using a Student's t-test. Furthermore, the first and final scores allocated by the reviewer using the same scale will be compared to identify the percentage of scores that change following multiple viewings of the segment of video.

Feasibility and clinical utility will be tested using a series of self-report statements, first used by de Jong *et al*[32] and also by Taddio and colleagues in their study addressing the psychometric properties of MBPS.[36] The results will be summarised descriptively and compared between scales.

Finally, correlations between treatment choice and pain scores will be calculated to contribute to an assessment of clinical utility.

CIs and p values (set for significance at 0.5) will be used to establish statistical significance where appropriate.

Risks

Study participation had no impact on patient care. The most significant risk to these children and their families is inappropriate access to video footage and loss of confidentiality. Stringent measures to avoid this have been put in place.

There are no other additional risks to the original patient cohort and their families from participation in the current study and no risks to the reviewers likely from participation in this study.

Limitations

There are a number of limitations to methods used to evaluate the psychometric properties of scales and tools where a gold standard does not exist. Assessment is therefore dependent on the results from a range of indirect measures of validity, all of which have limitations. It is not possible to blind the reviewers to the circumstances surrounding the infant or child, therefore potentially biasing reviewer application of the scale. To help overcome this potential bias, unique reviewers were used to score each phase of the procedure. Reviewers were also broadly aware of the purpose of the study, and although specific details and hypotheses were not revealed, this may have influenced their application of the scales. Finally, establishing the validity of one measure based on correlation with another can be considered to rely on circular logic, hence the use of multiple methods to establish scale validity.

It is not possible to establish the most appropriate sample size to measure reliability as the true variation in the population is not known. However, using a larger than previously used sample of raters, both clinicians and naive raters, and a large dataset of video segments will address shortcomings related to small sample sizes in previous studies evaluating psychometric properties of pain scales in infants and children.

Current status

Recruitment of patients for this study has been completed, and recruitment of reviewers has commenced. It is anticipated that the reviewer data collection will be completed by July 2017.

ETHICS AND DISSEMINATION

Research ethics

This study has been approved by the Human Research Ethics Committee of the RCH, Melbourne (reference number: RCH/EHRC 35220B).

Particular attention will be paid to ensure the appropriate storage and use of the video data used in this study. Patient and reviewer confidentiality will be maintained and no identifying features will be published.

Dissemination

There are currently very limited data to assist clinicians or researchers in their choice of the most appropriate scale for procedural pain assessment. This study will provide psychometric data addressing the performance of the FLACC scale, MBPS and VAS observer pain and VAS observer distress when used to assess procedural pain in infants and young children aged 6–42 months. This has the potential to identify the most reliable and valid scale for clinical and research purposes.

Results from this study will be disseminated to clinicians and researchers through peer-reviewed publications and conferences and in a higher-degree thesis.

Definitions

Clinician reviewer: emergency doctor or nurse of any level of experience recruited from the ED who has consented to participate in the study.

Distressing procedure: a procedure that is anticipated to cause distress but that is not considered to be painful.

Final score: score allocated with the first scale presented following review of the video segment as many times as needed until the reviewer is confident that their score will not change.

First score: score allocated with the first scale presented following a single uninterrupted view of the video segment.

Clinically naive: a healthcare professional with no clinical experience where they may have been responsible for assessing and/or treating pain.

Painful procedure: a procedure that is considered to be painful, for example, skin-breaking procedures.

Procedure phases: the procedure has been divided into sections to represent stages (phases) of the procedure; baseline, preparation and during.

Baseline phase: the phase (stage) of the procedure before any attempt to prepare or complete the procedure is made.

Preparation phase: the phase (stage) of the procedure during which contact is made by the clinician with the infant to prepare them for the procedure. This phase does not include stimulus presumed to be painful:

During phase: the phase (stage) of the procedure during which the procedural stimulus is applied.

Psychologist reviewer: a researcher affiliated/appointed to MCRI who has completed their basic training as a psychologist and who is, therefore, recognised as a psychologist or provisional psychologist.

Review session: data collection during which reviewer completes data collection using their allocated *segment review set*. Each reviewer will complete two review sessions using the same *segment review set*.

Segment review set: a unique set of video segments that will be allocated to a reviewer.

Video data: a digital video recording of the infant's clinical procedure.

Video segment: a 15s section of the video-recorded data that shows a procedure phase.

Author affiliations
[1]Department of Nursing, The University of Melbourne, Melbourne, Australia
[2]Clinical Sciences, Murdoch Children's Research Institute (MCRI), Melbourne, Australia
[3]Royal Children's Hospital, Melbourne, Australia
[4]Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa, Canada
[5]Clinical Epidemiology and Biostatistics Unit, MCRI, Melbourne, Australia
[6]Department of Paediatrics, The University of Melbourne, Melbourne, Australia

Contributors DJC: conception and design of the study, drafting of the protocol, design of the electronic data management system, development of the analysis plan, drafting of the manuscript. DH: design of the study, revision of the protocol, revision of the manuscript. AH: design of the study, development of the electronic data management system. TS: design of the study, design of the analysis plan. NS: revision of the protocol, revision of the manuscript. FEE: conception and design of the study, revision of the protocol, revision of the manuscript. All authors have approved the manuscript.

Funding This work is supported by a grant provided by the Murdoch Children's Research Institute (MCRI). Furthermore, MCRI is providing assistance with recruitment of psychologist reviewers. Advice regarding the study design and analysis plan was provided by the Clinical Epidemiology and Biostatics Unit. Support for recruitment during the original study was provided by staff of the emergency department. This assistance will also be provided for recruitment and data collection during the current study.

Competing interests None declared.

Patient consent Obtained.

Ethics approval Human Research Ethics Committee of the Royal Children's Hospital, Melbourne.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Sinatra R. Causes and consequences of inadequate management of acute pain. *Pain Med* 2010;11:1859–71.
2. Victoria NC, Murphy AZ. The long-term impact of early life pain on adult responses to anxiety and stress: historical perspectives and empirical evidence. *Exp Neurol* 2016;275 Pt 2(Pt 2):261–73.
3. Grunau RE, Tu MT, Whitfield MF, *et al*. Cortisol, behavior, and heart rate reactivity to immunization pain at 4 months corrected age in infants born very preterm. *Clin J Pain* 2010;26:1–704.
4. de Graaf J, van Lingen RA, Valkenburg AJ, *et al*. Does neonatal morphine use affect neuropsychological outcomes at 8 to 9 years of age? *Pain* 2013;154:449–58.
5. Dong L, Donaldson A, Metzger R, *et al*. Analgesic administration in the emergency department for children requiring hospitalization for long-bone fracture. *Pediatr Emerg Care* 2012;28:1–14.
6. Fry M, Bennetts S, Huckson S. An Australian audit of ED pain management patterns. *J Emerg Nurs* 2011;37:269–74.
7. Guéant S, Taleb A, Borel-Kühner J, *et al*. Quality of pain management in the emergency department: results of a multicentre prospective study. *Eur J Anaesthesiol* 2011;28:97–105.
8. Herd DW, Babl FE, Gilhotra Y, *et al*. Pain management practices in paediatric emergency departments in Australia and New Zealand: a clinical and organizational audit by National Health and Medical Research Council's National Institute of Clinical Studies and Paediatric Research in Emergency Departments International Collaborative. *Emerg Med Australas* 2009;21:210–21.
9. Robb AL, Ali S, Poonai N, *et al*. Pain management of acute appendicitis in Canadian pediatric emergency departments. *CJEM* 2016:1–7 (published Online First: 2016/12/09).
10. Shenoi R, Ma L, Syblik D, *et al*. Emergency department crowding and analgesic delay in pediatric sickle cell pain crises. *Pediatr Emerg Care* 2011;27:911–7.
11. Sills MR, Fairclough DL, Ranade D, *et al*. Emergency department crowding is associated with decreased quality of analgesia delivery for children with pain related to acute, isolated, long-bone fractures. *Acad Emerg Med* 2011;18:1330–8.
12. Yackey KJ, Rominger AH. Are we adequately treating pain in children who present to US emergency departments? Factors that contribute to pain treatment in pediatric patients. *Pediatr Emerg Care* 2016 (published Online First: 2016/06/03).
13. Stalnikowicz R, Mahamid R, Kaspi S, *et al*. Undertreatment of acute pain in the emergency department: a challenge. *Int J Qual Health Care* 2005;17:173–6.
14. Chan JY, von Baeyer CL. Cognitive developmental influences on the ability of preschool-aged children to self-report their pain intensity. *Pain* 2016;157:997–1001.
15. Emmott AS, West N, Zhou G, *et al*. Validity of simplified versus standard self-report measures of pain intensity in preschool-aged children undergoing venipuncture. *J Pain* 2017;18:564–73.
16. von Baeyer CL, Chambers CT, Forsyth SJ, *et al*. Developmental data supporting simplification of self-report pain scales for preschool-age children. *J Pain* 2013;14:1116–21.
17. Lee GY, Stevens B , *et al*. Neonatal and Infant Pain Assessment. In: McGrath PJ, Stevens BJ, Walker SM, eds. *Oxford Textbook of Paediatric Pain. 1 ed*. Oxford: Oxford University Press, 2014.
18. Ahola Kohut S, Pillai Riddell R. Does the Neonatal Facial Coding System differentiate between infants experiencing pain-related and non-pain-related distress? *J Pain* 2009;10:214–20.
19. Babl FE, Crellin D, Cheng J, *et al*. The use of the faces, legs, activity, cry and consolability scale to assess procedural pain and distress in young children. *Pediatr Emerg Care* 2012;28:1281–96.
20. Taddio A, Nulman I, Koren BS, *et al*. A revised measure of acute pain in infants. *J Pain Symptom Manage* 1995;10:456–63.
21. Fournier-Charrière E, Tourniaire B, Carbajal R, *et al*. EVENDOL, a new behavioral pain scale for children ages 0 to 7 years in the emergency department: design and validation. *Pain* 2012;153:1573–82.
22. Crellin, Dianne J, Babl, Franz E, Santamaria, Nick, Harrison, Denise. The MBPS: a systematic reivew to determine its role in assessing pain in infants and children. International Symposium on Pediatric Pain; 2017; Kuala Lumpur.
23. Celeste Johnston C, von Baeyer CL. A measure of pediatric pain intensity across ages and clinical conditions. *Pain* 2012;153:1545–6.
24. Crellin DJ, Harrison D, Santamaria N, *et al*. 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;156:2132–51.
25. van Dijk M, Koot HM, Saad HH, *et al*. Observational visual analog scale in pediatric pain assessment: useful tool or good riddance? *Clin J Pain* 2002;18:310–6.
26. Vazirani J, Knott JC. Mandatory pain scoring at triage reduces time to analgesia. *Ann Emerg Med* 2012;59:134–8.
27. Mokkink LB, Terwee CB, Patrick DL, *et al*. 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;19:539–49.
28. De Ruddere L, Goubert L, Stevens M, *et al*. Discounting pain in the absence of medical evidence is explained by negative evaluation of the patient. *Pain* 2013;154:669–76.
29. Gomez RJ, Barrowman N, Elia S, *et al*. Establishing intra- and inter-rater agreement of the Face, Legs, Activity, Cry, Consolability scale for evaluating pain in toddlers during immunization. *Pain Res Manag* 2013;18:e124–e128.
30. Merkel SI, Voepel-Lewis T, Shayevitz JR, *et al*. The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatr Nurs* 1997;23:293–7.
31. McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. *Psychol Med* 1988;18:1007–19.

32. de Jong A, Baartmans M, Bremer M, *et al*. Reliability, validity and clinical utility of three types of pain behavioural observation scales for young children with burns aged 0-5 years. *Pain* 2010;150:561–7.
33. Harris MR, Warren JJ. Patient outcomes: assessment issues for the CNS. *Clin Nurse Spec* 1995;9:82–6.
34. Stata Statistical Software [program]. 14 version College Station Texas:Stata Corporation. 2015.
35. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/. 2016.
36. Taddio A, Hogan ME, Moyer P, *et al*. Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. *Vaccine* 2011;29:1390–4.

8.1 Additional material

A copy of the Human Research and Ethics Committee letter of approval for this study and approval for minor modifications are provided in Appendix F.

8.2 Protocol amendments

The protocol was prepared and submitted for publication prior to completing the prospective study. Several amendments to the published protocol provided in this chapter were made and they are described in the following sections.

8.2.1 Sample

The protocol described two samples: patient and reviewers. The patients included in this study were as described in the protocol: 100 infants and young children (aged 6 – 42 months) undergoing one of the four nominated painful or non-painful procedures. A sample of 25 clinicians was sought for the study and 26 were recruited. The protocol also reported the intention to recruit a sample of psychologists and six psychologists participated in this study. However, their data has not been reported in this thesis.

8.2.2 Instruments

The instruments used in the study were as described in the published protocol. No changes were made to these instruments.

8.2.3 Procedure

The study was completed using the procedure as described in the protocol. All data described was collected during the study and the video segments were also prepared and presented to reviewers as described in the published protocol.

8.3 Statistical analysis

Results from this study are reported in the following chapters (chapters 9 to 12). The data were analysed as described in most cases e.g. intraclass correlations and Bland-Altman plots for

reliability testing. However, there are several variations to what was described in the protocol. The responsiveness of the scales was tested using linear mixed modelling as intended but an interaction between procedure type and procedure phase was added to the original model to account for the relationship between these two factors. In addition, the mean difference between baseline scores and procedure scores for painful and non-painful procedure were compared using the Student's t-test, which was not described in the protocol. Comparisons using Student's t-tests were also made between the scores for the procedural phase of painful and non-painful procedures. Finally, to determine the capacity of the scales to differentiate between procedures at different cut-off scores, sensitivity and specificity and area under the curve (AUC) were calculated using receiver operating characteristics (ROC). This analysis was also not described in the published protocol.

The protocol also describes collection of data that has not been presented in these chapters and this includes: time-based variables, clinicians' assessment of appropriate management for the pain &/or distress exhibited by the patient and data resulting from review of the video segments by the six psychologists recruited. This data, along with the unpublished data reported in the thesis, e.g. VASobs related data, will be analysed and reported via publication following submission of this thesis.

CHAPTER 9.

The psychometric properties of the Visual Analogue Scale applied by an observer to assess procedural pain in infants and young children.

The results of the psychometric properties of the Visual Analogue Scale applied by an observer (VASobs) used to assess procedural pain in infants and children are reported in this chapter. This work is unpublished and is presented here formatted for the thesis.

Abstract

Introduction: The VASobs is one of the most practical ways to assess pain in infants and children unable to self-report their pain. Despite widespread use, there is insufficient data to recommend it for assessment of procedural pain. Therefore, the aim of this study was to test the psychometric and practical properties of the VASobs to quantify procedural pain in infants and young children.

Methods: Twenty-six clinicians independently applied the VASobs to segments of video collected from 100 children aged six to 42 months undergoing a procedure. Video segments were scored by four reviewers.

Results: Reliability (intra- and inter-rater) was poor to fair (ICC ranged from 0.35 to 0.55). At a cut-off score of 3, sensitivity and specificity were 84.7% and 95.0%, respectively. Linear mixed modelling confirmed scale responsiveness to pain. Pain scores increased across phases (baseline to procedure) for painful procedures (regression slope 4.95) and more modestly for non-painful procedures (regression slope = 0.41). The correlation between FLACC and Visual Analogue Scale observer (VASobs) pain and FLACC scores was good (r = 0.74). VASobs was easily applied and preferred by clinicians in this study.

Discussion: Despite evidence of sensitivity and responsiveness to pain, incongruously the reliability results were sufficiently poor to raise concerns about the VASobs for assessing procedural pain in infants and young children.

9.1 Introduction

Infants and children are frequently exposed to painful and distressing procedures during their health care and management of the pain associated with these procedures has gained increasing attention in the clinical and research literature. Key to improvement of the procedural experience is accurate assessment of procedural pain. Ideally this is achieved by self-report. However, infants and young children do not have the verbal or cognitive skills to report and quantify their pain intensity. Therefore, clinicians and researchers are reliant on proxy measures to estimate pain intensity.

Multidimensional observational pain scales have repeatedly been proposed as a viable option to support pain assessment in infants and children. However, scales can be complex and impractical for use, particularly for clinical application. In contrast, the perceived ease with which the Visual Analogue Scale (VAS) can be applied and the validity of self-reported pain scores served as the impetus to use this scale, applied by an observer (VASobs), to assess the pain experienced by younger children and infants unable to self-report. Use of the VASobs is seen most frequently in the procedural pain literature. Our recent, as yet unpublished, review of the research literature showed that the VASobs was used three times more often to measure procedural pain in RCTs than other behavioural pain scales (chapter 3).

The VAS is a line measuring 10cm with verbal anchors at each end, most commonly ‘no pain’ and ‘worst possible pain’, which correspond to a score of ‘0’ and ‘10’ respectively. It is widely used to assess adults and older children able to self-report (118, 345-348) and patients are asked to place a mark on the otherwise unmarked line to indicate the pain they are currently experiencing. The distance from the zero mark is measured and this is considered the pain score (118, 345). Application of this scale by an observer is similar; the observer places a mark on the line to indicate their perception of the intensity of the patient’s (e.g. infant or child) pain.

Despite widespread application of this scale by an observer, the psychometric evidence to support use of the VASobs to assess pain in infants and young children is limited (Chapter 5). In 2002, Van Dijk and colleagues summarised the results of studies contributing to our understanding of the psychometrics of the VASobs used to quantify pain intensity in infants and children (290). The authors concluded that interrater reliability (correlation coefficients ranging from 0.36 to 0.91), correlation with self-report (0.23 to 0.83) and correlations with other pain measures (0.42 to 0.86) were variable, making it difficult to draw conclusions about its potential performance in even similar circumstances. They also concluded that the absence of adequate data addressing reliability, scale responsiveness and optimal cut-offs prevented drawing conclusions about the

role that this scale might have for pain intensity assessment. A recent review, completed in 2016, does little to change our understanding of how well the VASobs performs when used to assess pain in infants and children (Chapter 2). In light of such widespread use establishing the psychometrics of the scale is critical.

The aim of this study was therefore to fill this gap in the literature and to test the psychometric and practical properties (feasibility, reliability, validity and clinical utility) of the VASobs to quantify procedural pain intensity in infants and children aged from six to 42 months to determine its suitability for clinical and research purposes.

9.2 Methods

The methods for this study are described in the published protocol which was presented in Chapter 8 (468).

9.3 Results

Twenty-six ED clinicians were recruited for this study and made 1088 observations of 100 children at the first review and 358 at the second review (Figure 9-1). There were no missing observations as reviewers could not advance unless each data field was completed. The clinicians included; 19 nurses of varying levels of experience (range 1 to 20 years, mean 10.1), 12 with postgraduate specialty training in paediatrics and/or emergency care and seven doctors of whom three were considered senior (defined as having completed their specialty training). The mean age of the children was 22.5 (±10.3) months, 58% (n = 58) were boys and 38% were diagnosed with respiratory disease, 29% with dehydration and gastroenteritis, while the remaining 36% spanned a range of diagnoses.

The mean, median and distribution of VASobs scores across the phases of each procedure from review session one are presented in boxplots in Figure 9-2A-D. Mean VASobs (pain) scores during the procedure phase were highest for nasogastric tube insertion (6.4 ± 2.0) and lowest for saturation measurement (0.7 ± 0.86). The VASobs (distress) scores followed similar patterns but were universally higher than pain scores; the mean nasogastric tube insertion VASobs (distress) score was 8.1 (± 1.8) and saturation measurement score was 2.7 ± (5.2).

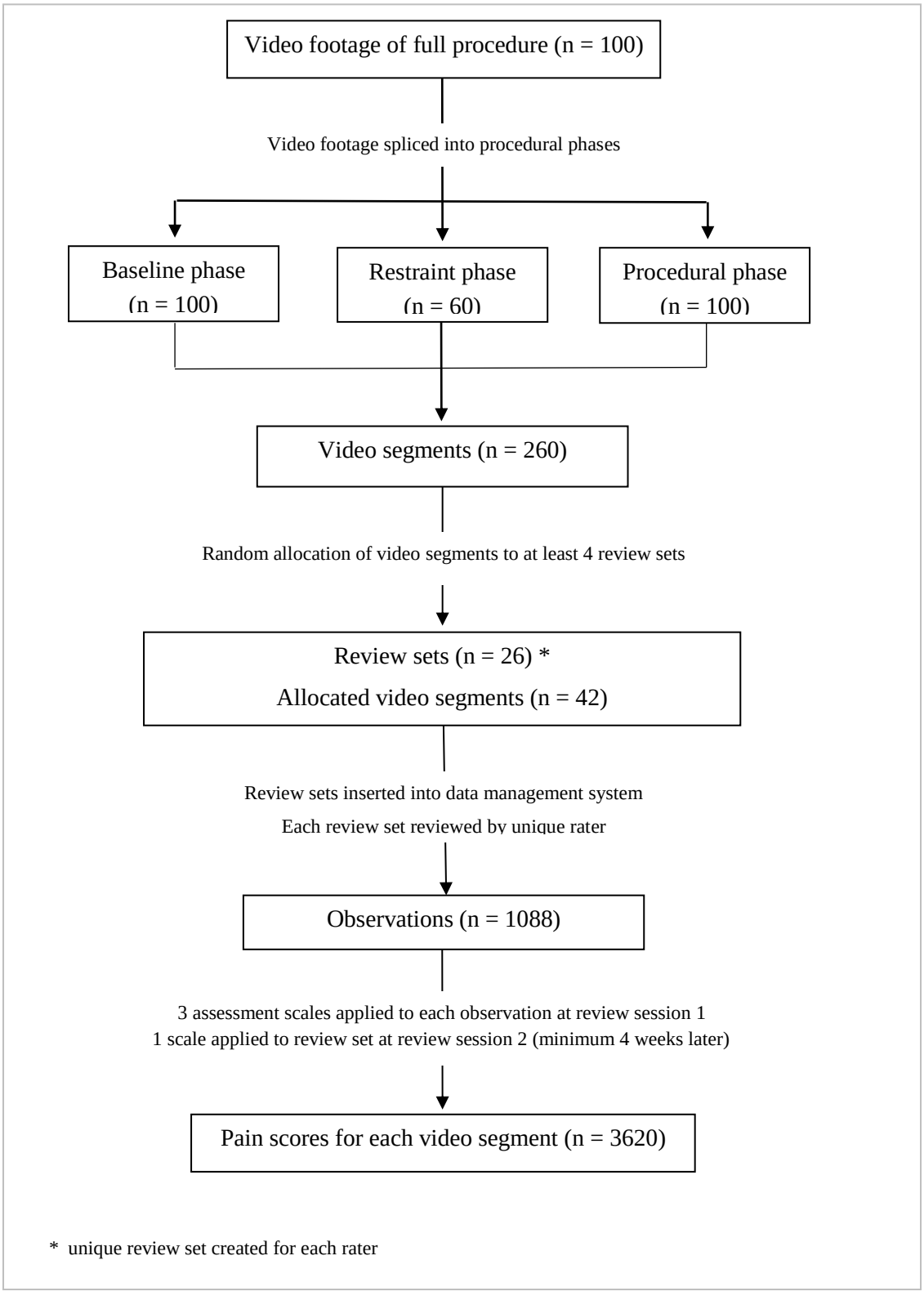


Figure 9-1 Overview from creation of video segments to final data set.

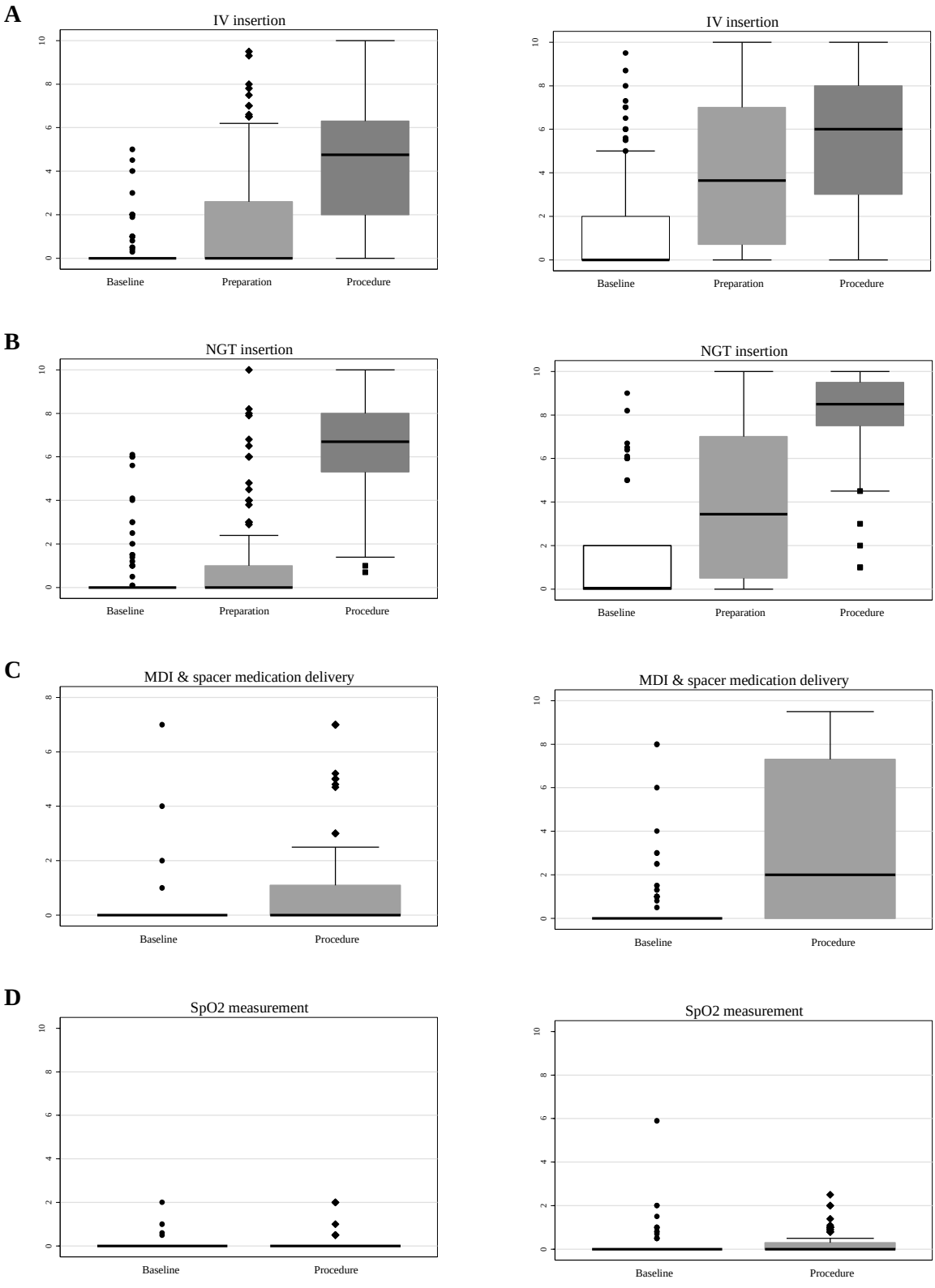


Figure 9-2 Boxplots for observer Visual Analogue Scale pain (left) and Visual Analogue Scale distress (right) scores for each phase of each procedure (A - IV insertion, B – NGT insertion, C – metered dose inhaler (MDI) medication administration & D – SpO2 measurement).

9.3.1 Psychometric evaluation

9.3.1.1 Feasibility and clinical utility

Reviewers were able to allocate a VASobs (pain) score on all but 10 scoring occasions (0.9%) and a distress score on all but two scoring occasions (0.2%). The correlation between the first score allocated after one uninterrupted view of the video segment and the final score allocated was near perfect (r = 0.97). As scores were not normally distributed, Wilcoxon signed rank test was used to demonstrate that the there was no statistically significant difference between first and final scores for either group (Table 9-1).

Table 9-1 Comparisons between first score and final score.

Scale	First score	Final score	scores changed by 1	Correlation*	P value**
VASobs (pain)	1.6 (2.7)	1.6 (2.7)	8.8%	0.94	0.63
VASobs (distress)	3.7 (3.7)	3.6 (3.7)	9.9%	0.92	0.58

Values are mean (standard deviation) / median [interquartile range]

* Spearman correlation coefficient

** Wilcoxon signed rank test with continuity correction

Reviewers used a Likert scale to indicate the extent to which they agreed (5) or disagreed (1) with a series of utility statements. Reviewers agreed (score of 4 or 5) that the VASobs (pain) was; ‘easy to understand’ (74.1%) and ‘quick’ (88.9%) and ‘easy’ (81.5%) to apply. However, the extent to which they agreed that the VASobs ‘reflects procedural pain-specific features’ was lower (53.8%) and lower still when they rated the extent to which they agreed that the scale is ‘readily understood and supports decisions about pain management’ (26.9%) and is ‘clinically useful’(26.9%). Finally, few reviewers agreed that the VASobs ‘reflects the **extent** of procedural pain’ (14.8%) or ‘discriminates children with pain from those without’ (18.5%).

9.3.1.2 Reliability

The overall ICC for inter-rater reliability for the VASobs (pain) scores was 0.55 (Table 9-2). Reliability varied across phases and procedure types; ranged from 0.27 (baseline, non-painful procedures) to 0.48 (procedure, painful procedure). Overall reliability for VASobs (distress) was

higher ICC = 0.78 and ranged between 0.60 and 0.89 across procedures and phases. The extent of the variation between reviewer scores for VASobs (pain) and VASobs (distress) are demonstrated in Figures 9-3 and 9-4 where standard deviations of the difference in scores are plotted against the reviewers' mean scores. There is a subtle trend towards higher levels of inconsistency for scores allocated from the middle of both scales (2 to 7) when compared to the variation in scores allocated from the extremes of either scale (0 to 1 and 8 to 10).

Table 9-2 The inter-rater reliability of VASobs (pain) and VASobs (distress) overall and for each procedural phase of painful and non-painful procedures.

Scale		Painful			Non-painful	
	overall	baseline	preparation	procedure	baseline	procedure
VASobs (pain)	0.559	0.37	0.35	0.48	0.27	0.35
VASobs (distress)	0.78	0.7	0.78	0.65	0.6	0.89

Values are intra-class correlation coefficients (ICC) using one-way random effects model

Inter-rater reliability was confirmed by linear mixed modelling, the results of which are reported in full in section 9.3.1.3, where variability in VASobs pain scores attributed to the effect of the reviewer was close to zero (variance = 0.35 ± SD 0.59) and for VASobs distress scores was almost zero (variance = 0.146 ± SD 0.382).

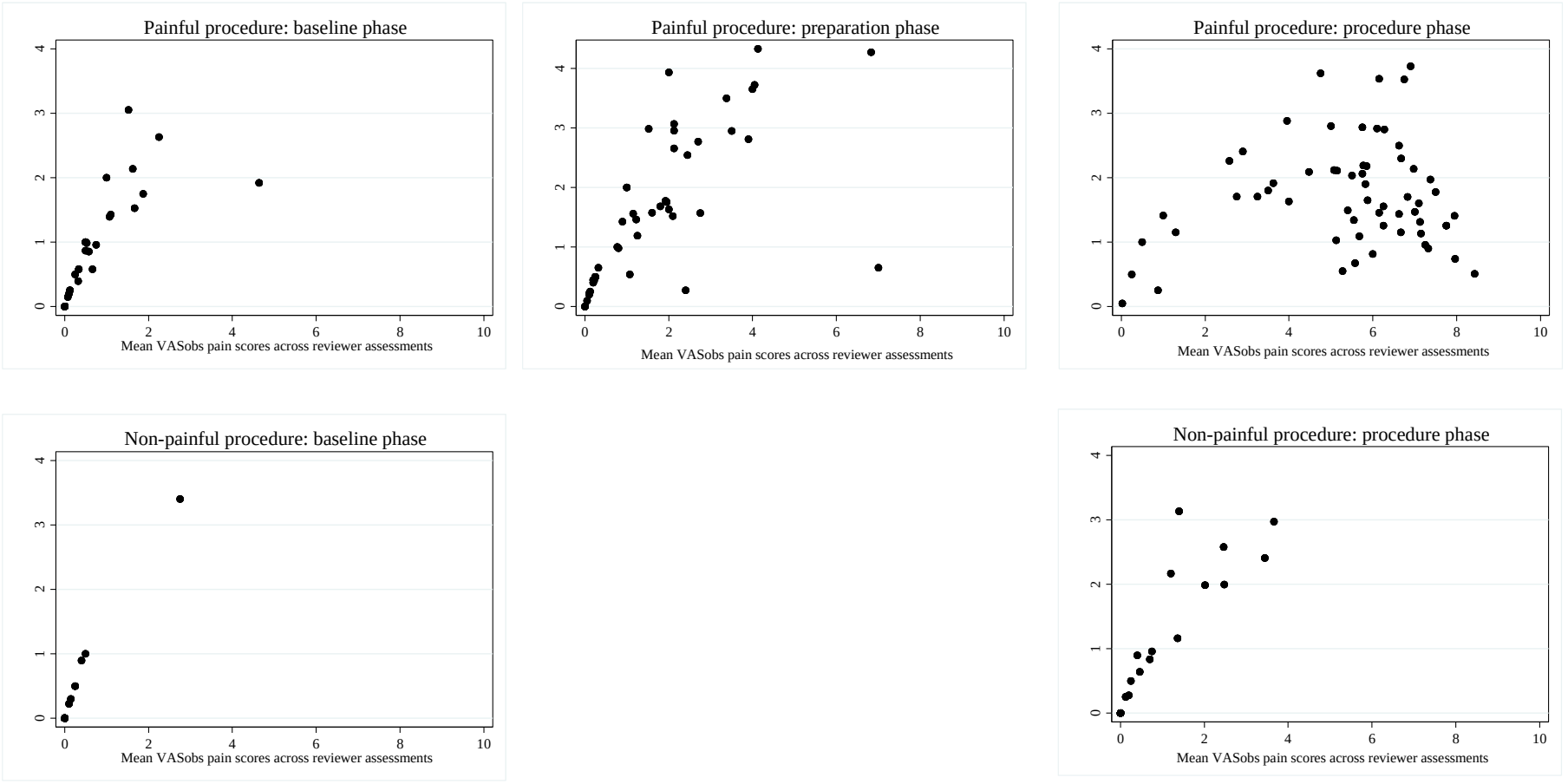


Figure 9-3 Inter-rater reliability for VASobs (pain) scores: variation of reviewer assessments within child (standard deviation displayed on y-axis) versus average rating over all assessments (mean displayed on x-axis).

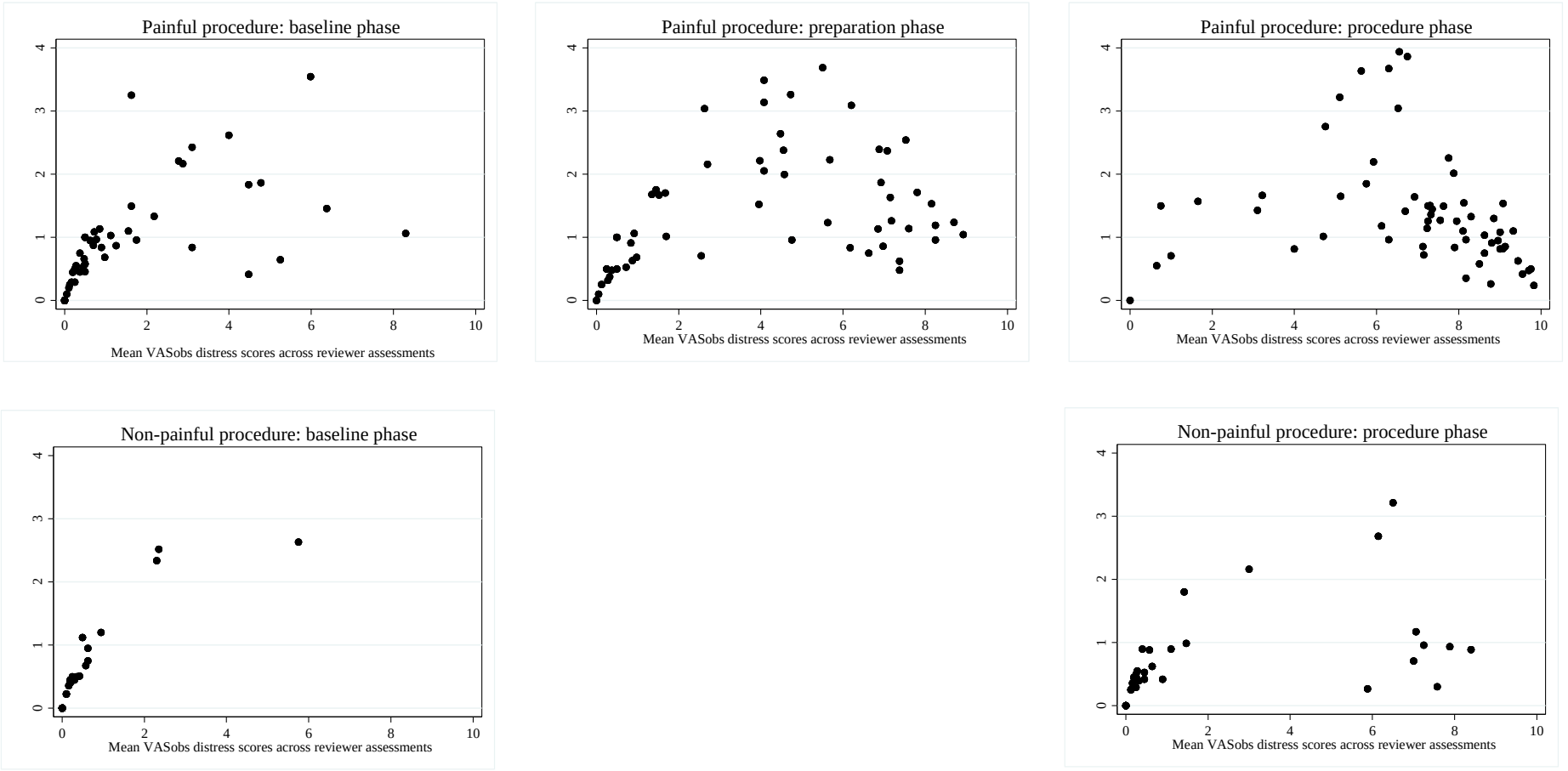


Figure 9-4 Inter-rater reliability for VASobs (distress) scores: variation of reviewer assessments within child (standard deviation displayed on y-axis) versus average rating over all assessments (mean displayed on x-axis).

The ICC for intra-rater reliability was 0.77 for VASobs (pain) and slightly higher for VASobs (distress) (ICC = 0.81). The Bland and Altman plots shown in Figure 9-5 and 9-6 demonstrated that the variability between scores and the mean difference for pain scores was -0.64 (SD ±1.93) and for distress scores was -0.09 (± SD 2.27) with a tendency for the second score to be lower than the first. There was also a predictable trend towards higher levels of inconsistency for scores allocated from the middle of the scale (2 to 7) when compared to the variation in scores allocated at the extremes of the scale (0 to 1 and 8 to 10).

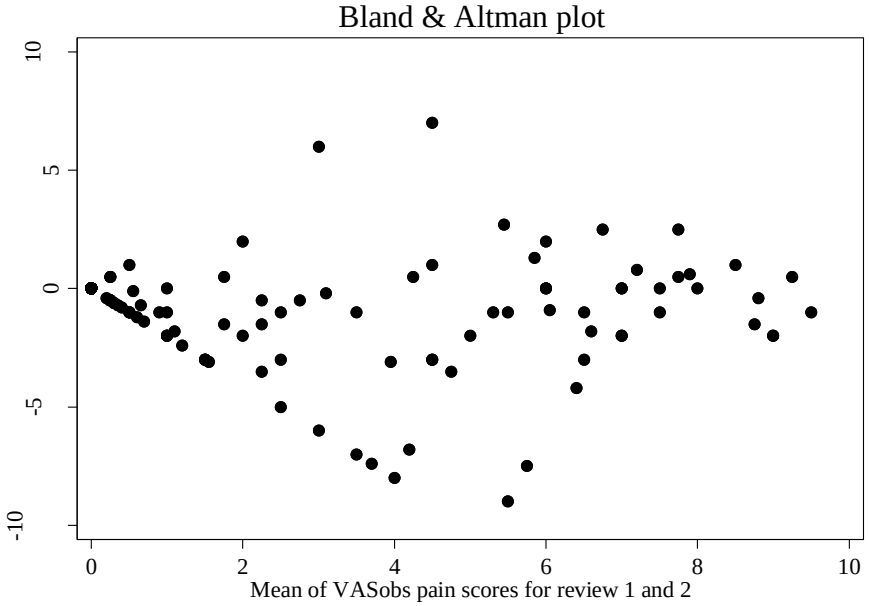


Figure 9-5 Difference between VASobs pain scores plotted against the mean score for review session 1 and 2. Mean difference is -0.64 (SD 1.93), 95% limits of agreement are -4.50 and 3.22.

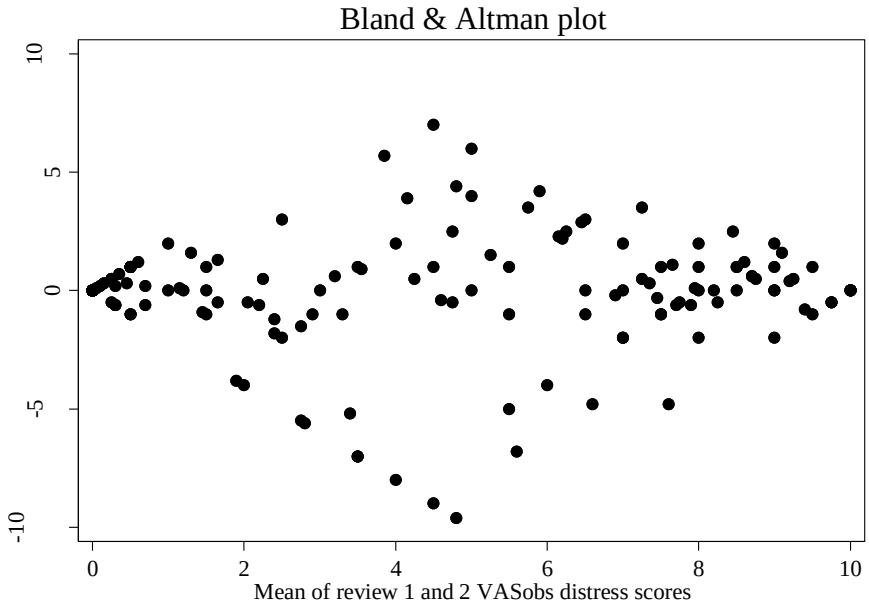


Figure 9-6 Difference between VASobs distress scores plotted against the mean score for review session 1 and 2. Mean difference is -0.09 (SD 2.27), 95% limits of agreement are -4.45 and 4.63.

9.3.1.3 Construct validation: known groups, responsiveness and discrimination

Mean VASobs pain scores and mean VASobs distress scores plotted across phases of the procedure for non-painful and painful procedures (Figures 9-7 and 9-8) illustrate the relationships between scores across the phases of these procedures.

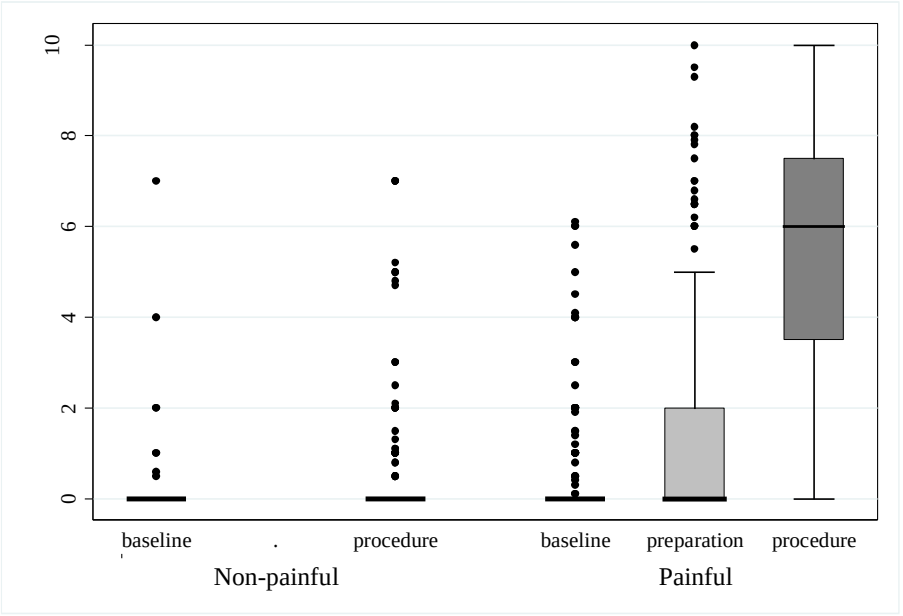


Figure 9-7 Boxplots/ representing change of VASobs pain scores over procedure phases (baseline, preparation and procedure) in the two procedure cohorts (painful and non-painful procedures).

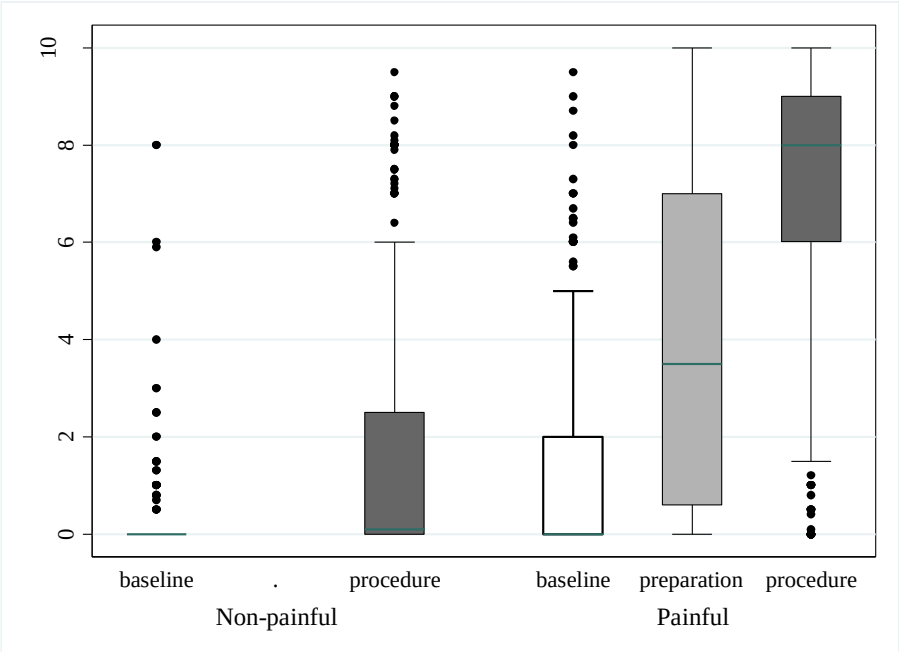


Figure 9-8 Boxplots/ representing change of VASobs distress scores over procedure phases (baseline, preparation and procedure) in the two procedure cohorts (painful and non-painful procedures).

A significant difference between the mean VAS pain procedural phase scores for painful (5.4 ± 2.7) and non-painful procedures (0.6 ± 1.4) was shown $t(418) = -22.11$, $p < 0.000$. Similarly, a significant difference between VASobs distress scores during the procedural phase of painful (6.9 ± 2.8) and non-painful (2.0 ± 3.0) procedures was shown $t(422) = -17.13$, $p < 0.000$. This demonstrates a difference between to two procedure groups. To confirm the capacity of the VASobs to differentiate between painful and non-painful procedures at different cut-off scores for VASobs pain and VASobs distress, we also calculated sensitivity, specificity and AUC using receiver operating characteristics (ROC). The results for various cut-offs are reported in Tables 9-5 and 9-6. It can be seen from these results that a VASobs pain score of '3' provides the best sensitivity (84.7%), specificity (95.0%) and AUC (0.90) and that a VASobs distress score of 3 also provides the best sensitivity (91.5%), specificity (77.5%) and AUC (0.84) for VASobs distress.

Table 9-3 Sensitivity, specificity and area under the curve values (AUC) calculated for different cut-off VASobs distress scores to differentiate procedure type (painful and non-painful).

VASobs distress score cut-off	Sensitivity (%)	Specificity (%)	Area under the curve (AUC)
Score > 0	98.3	17.5	0.57
Score > 1	93.2	67.5	0.80
Score > 2	91.5	75.0	0.83
Score > 3	91.5	77.5	0.84
Score > 4	86.4	77.5	0.82
Score > 5	85.0	77.5	0.80
Score > 6	74.6	80.0	0.77
Score > 7	61.0	87.5	0.74

Table 9-4 Sensitivity, specificity and area under the curve values (AUC) calculated for different cut-off scores for VASobs pain to differentiate procedure type (painful and non-painful).

VASobs pain score cut-off	Sensitivity (%)	Specificity (%)	Area under the curve (AUC)
Score > 0	100	60	0.80
Score > 1	91.5	80.0	0.86
Score > 2	89.8	87.5	0.89
Score > 3	84.7	95.0	0.90
Score > 4	76.3	100	0.88
Score > 5	71.2	100	0.86
Score > 6	44.1	100	0.72
Score > 7	18.6	100	0.59

There was a difference between the mean difference in VASobs pain scores between baseline and procedural phase for children experiencing painful (4.96 ± 2.16) versus non-painful procedures (0.42 ± 0.99), $t(97) = -12.45$, $p < 0.000$. This was repeated for VASobs distress scores where the mean difference in distress scores from baseline to procedure for children experiencing a painful procedure (5.51 ± 2.94) was significantly different to the mean difference in distress scores between phases for children experiencing a non-painful procedure (1.53 ± 2.67), $t(97) = - 6.86$, $p < 0.000$.

Responsiveness was also tested by linear mixed modelling to determine the impact of procedure and phase on VASobs pain and distress scores (Table 9-5 and 9-6). Both variables were considered fixed effects and reviewer and child were added to the model as random effects. From the model we can see that pain scores at baseline were a little over zero (intercept = 0.13) as were distress scores (intercept = 0.429). Pain scores increased across phases (baseline to procedure) for painful procedures (regression slope 4.95). The difference across phases for non-painful procedures is substantially more modest (regression slope = 0.41) Distress scores across phases (baseline to procedure) increased for painful and non-painful procedures. However, this change was much more pronounced for painful compared with non-painful procedures (regression slope 5.52 versus 1.52, respectively).

Table 9-5 The variances and estimates for random and fixed effects for the model used to demonstrate the responsiveness of VASobs pain scores to the procedure type (painful vs non-painful) and phase.

Random effects		Variance	Standard deviation	
Patient	(Intercept)	0.556	0.746	
Reviewer	(Intercept)	0.351	0.592	
Residual		2.481	1.575	
Fixed effects		**Estimate**	**Standard error**	**t value**
Intercept		0.127	0.203	0.623
Painful procedure		0.248	0.220	0.293
Preparation phase		0.953	0.150	6.363
Procedure phase		0.410	0.169	2.430
Procedure phase: Painful procedure		4.535	0.226	20.114

Model = (1|Reviewer) + (1|Child) + procedure_type + procedure_phase + procedure_type*procedure_phase

Table 9-6 The variances and estimates for random and fixed effects for the model used to demonstrate the responsiveness of VASobs distress scores to the procedure type (painful vs non-painful) and phase.

Random effects		Variance	Standard deviation	
Patient	(Intercept)	2.734	1.653	
Reviewer	(Intercept)	0.146	0.382	
Residual		4.173	2.043	
Fixed effects		**Estimate**	**Standard error**	**t value**
Intercept		0.429	0.312	1.374
Painful procedure		0.969	0.394	2.458
Preparation phase		2.444	0.191	12.769
Procedure phase		1.521	0.218	6.991
Procedure phase: Painful procedure		3.996	0.289	13.805

Model = (1|Reviewer) + (1|Child) + procedure_type + procedure_phase + procedure_type*procedure_phase

Independent t-test were also calculated to compare VASobs pain and VASobs distress scores overall, by procedure type (painful versus non-painful) and procedural phase (baseline, preparation and procedural). Mean pain scores were significantly lower than distress scores for each comparison; overall (1.9 vs 3.09, t-test(2126) = -10.36), painful procedures (2.38 vs 4.03, t-test(714) = - 18.95), non-painful procedures (0.33 vs 1.19, t-test(361) = -8.69), baseline phase (0.27 vs 0.90, t-test(412) = -8.58), preparation phase (1.32 vs 3.83, t-test(243) = -13.98) and procedural phase painful procedures (5.38 vs 6.86, t-test(238) = - 11.91 and procedural phase non-painful procedures (0.56 vs 2.01, t-test(180) = - 8.17).

9.3.1.4 Convergent validation

Correlations between VASobs pain with other behavioural scales Face, Legs, Activity, Cry, Consolability (FLACC) scale and the Modified Behaviour Pain Scale (MBPS) were both 0.74. The correlations between distress scores (VASobs distress) and FLACC scores (r = 0.89) and MBPS scores (r = 0.87) were higher than when correlated with VASobs pain scores. The Spearman correlation coefficient for pairwise correlation between VASobs scores for pain and distress was = 0.77. Scatter plots of the comparisons between scores are shown in Figure 9-9 and this shows an obvious trend towards lower VASobs pain scores compared with VAS distress scores and to some extent compared with FLACC and MBPS scores.

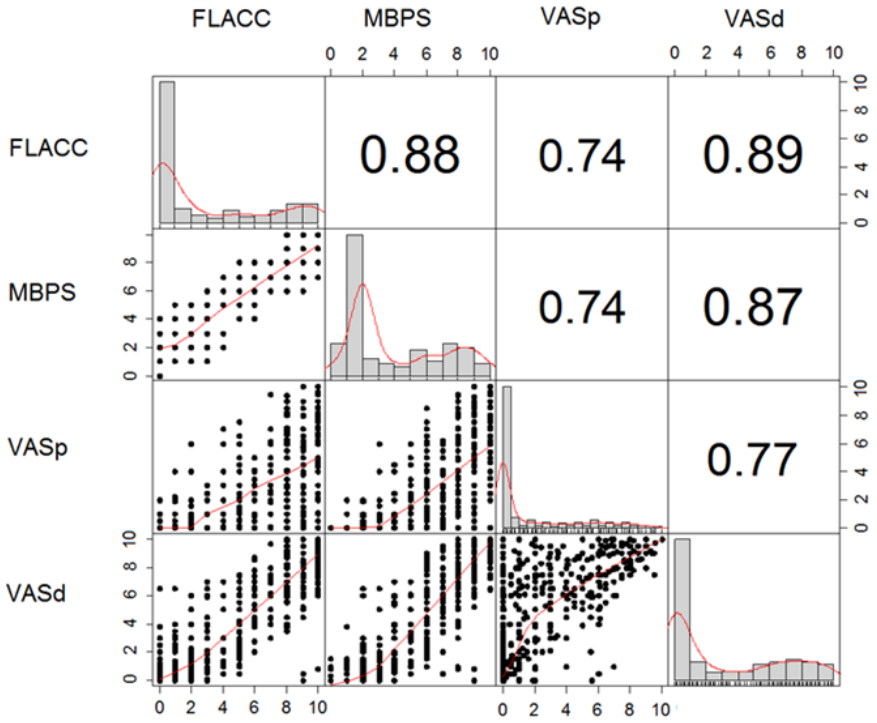


Figure 9-9 Distributions for the scores of each scale on the X axis, correlations between the scores for the scales on the X and Y axis and plots of scores on the X axis against the scores for the scale on Y axis are shown.

Note: correlations were calculated using Spearman correlation coefficient.

9.4 Discussion

The frequency of use of the VASobs to assess pain in infants and children underscores the critical need to better understand the psychometrics of the scale when used for this purpose. The aim of this study was to address this deficit in the literature, specifically by assessing the feasibility and clinical utility, reliability and validity of the scale used to assess procedural pain. Despite some results from this study suggesting that the VASobs pain is valid for assessing procedural pain, the results of this study raise some interesting questions about the role of the VASobs in paediatric practice.

The validity of the VASobs to assess pain in infants and young children experiencing procedural pain was tested using several methods and the results of each confirm the capacity of the scale to

measure the pain associated with intravenous catheter and nasogastric tube insertion. The responsiveness of the scale was also demonstrated with a marked increase in pain across phases particularly for these painful procedures (although it is worth noting that scores also increased for non-painful procedures but to a much lesser extent). Observer VAS pain scores correlated well with FLACC scale and MBPS scores (r = 0.74), which have both been shown sensitive to pain (464, 469).

As pain and distress often go hand in hand it was not surprising to find that the correlation between VASobs pain and VASobs distress was high (r = 0.77). However, for a scale to be considered specific for pain it should also be able to differentiate between pain and distress. Baseline phases and the inclusion of presumed non-painful procedures were an opportunity to determine whether VASobs pain scores were specific for pain or were in part a measure of non-pain related distress. Comparison of the VASobs pain scores with VASobs distress scores showed that pain scores were significantly lower than distress scores for non-painful circumstances. Furthermore, the capacity of the VASobs pain and VASobs distress to distinguish between painful and non-painful procedures was confirmed. However, this was at a cut-off of 3, which has been described as the upper limit of 'mild' pain (470). It should also be noted that mean pain scores for phases and procedures presumed not to be painful were not zero. As reviewers were asked to apply both scales, this may have prompted a conscious attempt to differentiate between these constructs that may not occur if VASobs pain is applied on its own.

The VASobs pain and VASobs distress proved practical and reviewers largely agreed that the VASobs for pain assessment was easy to understand and 'quick' and 'easy to apply'. In addition, restraint, the procedure or other factors rarely impinged on reviewers' capacity to score the infant or child's pain (n = 10) or distress (n = 2). However, reviewers were less positive about the scales capacity to measure procedural pain. Only a quarter of reviewers agreeing that the scale was 'readily understood and supportive of decisions about pain management' or 'clinically useful' and even smaller numbers agreed that the scale 'discriminates children in pain from those without' (18.5) and 'reflects the extent of procedural pain' (14.8%). An absence of clinician confidence in the validity of the scale has substantial ramifications for how clinicians interpret and respond to VASobs scores. Research findings and clinical assessment results that do not accord with the clinician's perception may be easily dismissed based on a fundamental assumption that the scale is not clinically useful or capable of measuring procedural pain.

Inter-rater reliability results for VASobs pain were not strong (r = 0.55) and poor across phases for both painful and non-painful procedures (0.27 to 0.48). Results were better for VASobs distress, the overall ICC was 0.78 and ranged across phases for painful and non-painful

procedures from 0.60 to 0.89. This may reflect a willingness on the part of the reviewers to accept the behaviours as evidence of distress but in the absence of a visible painful stimulus a reluctance to accept these behaviours as evidence of pain, rather than a reflection of a clear capacity to differentiate distress-related behaviours from pain-related behaviours. Intra-rater reliability was markedly higher VASobs pain scores (r = 0.77) but almost identical for VASobs distress scores (r = 0.81). The scatter plots highlight a predictable trend for scores more polarised towards the anchors to show less variation than for scores from two to seven. Reviewers are likely to differ in their perception of the 'intensity' of the behaviours seen and therefore the intensity of the score where the behaviours do not conform to the extremes, but they may also differ in how they interpret these behaviours, i.e. as evidence of pain or as evidence of distress.

It has been traditionally accepted that if something is not reliable it cannot be considered valid and the theoretical rationale for this is sound. For a scale to accurately measure a construct (validity) it must be able to do so consistently when applied by different reviewers repeatedly (inter-rater reliability) and repeatedly by the same reviewer where no change in condition has occurred (intra-rater reliability). However, in this study scores were not highly reliable, but the scores were shown to be sensitive and responsive to pain. These results suggest that a theory such as this oversimplifies our understanding of the relationship between these psychometric properties. Reliability is shown in degrees (as has been demonstrated here) rather than as an absolute (present or absent) and therefore the effect of reliability on validity will strengthen or weaken validity in degrees. In this study the reliability of VASobs pain scores was only fair. Therefore, regardless of the strength of the validity results, based on the reliability results the validation results can only be considered at best fair.

9.4.1 Strengths and Limitations

This was a single centred study and reviewers could not be blinded to the circumstances e.g. needle insertion. Furthermore, establishing the validity of one measure based on correlation with another can be considered circular logic. Until recently, there was insufficient data to accept the FLACC scale and the MBPS as well validated for procedural pain assessment (330, 471). Therefore, use of these scales for convergence testing was a limitation of this study. The use of video recordings may have altered the experience of the reviewer compared with real time assessment impacting on their scores, particularly those that relate to feasibility.

Several strategies were employed to overcome the impact of these limitations. Unique reviewers were used to assess each phase of a single procedure to prevent reviewers from creating logical

patterns in the scores across phases of a procedure. Application of the scale after a single uninterrupted view of the video segment to generate the first score was used to simulate as closely as possible the clinical application of the scale. In addition, multiple methods were used to establish validity to overcome the limitations of each validation method. Studies addressing the psychometric properties of pain scales frequently use non-clinical research assistants and their results used to claim validity of the scale for clinical use. However, as there is some evidence that clinicians apply clinical judgement when applying assessment scales (472-474), using clinicians provided an opportunity to test the assumption that the scale will perform adequately when applied by clinicians. These methodological limitations should be considered when interpreting the results of this study.

9.4.2 Conclusions and future directions

Extensive review of the literature confirms that the VASobs is frequently used for procedural pain assessment in infants and young children in clinical trials in the absence of robust evidence to support this practice. The purpose of our study was to evaluate the psychometric performance of the VASobs pain when used to assess procedural pain in infants and children and to make recommendations about its use for clinical and research purposes. There remains insufficient data to support its psychometric properties and despite frequent use, too little clinician confidence in its capacity to be useful to recommend it for this purpose.

CHAPTER 10.

The psychometric properties of the FLACC scale used to assess procedural pain

This chapter reports results of the psychometric properties of the Face, Legs, Activity, Cry Consolability (FLACC) Scale used to assess procedural pain in infants and children. This work has been published and the PDF of this publication is reproduced in this chapter.

Publication

Crellin DJ, Harrison D, Santamaria N, Huque H, Babl FE. The Psychometric Properties of the FLACC Scale Used to Assess Procedural Pain. J Pain. 2018. 19(8):862-872.

Published abstract

Crellin DJ, Harrison D, Santamaria N, Babl FE. The FLACC scale: is it reliable and valid used to assess procedural pain experienced by infants and young children? International Symposium on Pediatric Pain; Kuala Lumpur: International Association for the Study of Pain; 2017.

American Pain Society
RESEARCH
EDUCATION
TREATMENT
ADVOCACY
PUBLISHED BY ELSEVIER

The Journal of Pain, Vol 19, No 8 (August), 2018: pp 862-872
Available online at www.jpain.org and www.sciencedirect.com

The Psychometric Properties of the FLACC Scale Used to Assess Procedural Pain

Dianne J. Crellin,*,†,‡ Denise Harrison,*,‡,§ Nick Santamaria,* Hamidul Huque,‡ and Franz E. Babl†,‡,¶

*Department of Nursing, The University of Melbourne, Melbourne, Australia.
†Department of Emergency Medicine, Royal Children's Hospital, Melbourne, Australia.
‡Clinical Sciences, Murdoch Children's Research Institute, Melbourne, Australia.
§Department of Nursing, Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa, Ontario, Canada.
¶Department of Paediatrics, The University of Melbourne, Melbourne, Australia.

Abstract: **The Face, Legs, Activity, Cry, and Consolability (FLACC) scale is one of the most commonly and widely used behavioral observation pain scales. The aim of this study was to test the psychometric and practical properties of the FLACC scale to quantify procedural pain in infants and young children. Twenty-six clinicians independently applied the FLACC scale to segments of video collected from 100 children aged 6 to 42 months undergoing a procedure. Video segments were scored by 4 reviewers. Inter- and intrarater reliability coefficients were high (.92 and .87, respectively). Linear mixed modeling confirmed scale responsiveness (differences in difference between FLACC scores across phases for painful versus nonpainful procedures was 4.2, 95% confidence interval = 3.67-4.81). Sensitivity and specificity were 94.9% and 73.5%, respectively, at a cutoff of 2. However, the mean difference across phases for children with baseline scores >3 was much lower than for children with scores <3, $P = .0001$. Correlations between FLACC and Visual Analog Scale observer pain and distress were good ($r = .74$ and $r = .89$, respectively). This study supports the reliability and sensitivity of the FLACC scale for procedural pain assessment. However, the circumstances of procedures interfered with application of the scale and the findings question the capacity of the scale to differentiate between pain- and nonpain-related distress.**

Perspective: ***This article provides evidence that the FLACC scale is reliable and sensitive to pain for procedural pain assessment. Concerns remain about specificity and scale design. Identification of a scale valid for this purpose is needed to provide a platform for improved procedural pain management in infants and young children.***

Key words: *Face, Legs, Activity, Cry, and Consolability scale, pain assessment, psychometric evaluation, validation, reliability, FLACC scale.*

Infants and children receiving medical assessment and treatment in a range of settings frequently experience pain-inducing procedures such as injections, blood sampling, and intravenous access. Despite the increasing weight of evidence that infants' experience of pain, especially for preterm infants, can have a negative effect on short- and long-term outcomes,[31] pain continues to be poorly managed, particularly in infants and children presenting to emergency departments (EDs), where many of these procedures occur.[17,19,30] Reasons cited for the suboptimal treatment of pain and distress in EDs include, among other factors, poor recognition of significant pain by clinicians.[23] It is therefore essential that appropriate

Received September 7, 2017; Revised December 31, 2017; Accepted February 22, 2018.
Funding and other material support provided by a Murdoch Children's Research Institute clinical research grant.
The authors have no conflicts of interest to declare.
Supplementary data accompanying this article are available online at www.jpain.org and www.sciencedirect.com.
Address reprint requests to Dianne Crellin, RN, BN, MN, Royal Children's Hospital & The University of Melbourne, c/o Emergency Department, Royal Children's Hospital, Flemington Road Parkville, 3052, Australia. E-mail: dianne.crellin@rch.org.au
1526-5900/$36.00

https://doi.org/10.1016/j.jpain.2018.02.013

and validated means to assess procedural pain are available. Equally, researchers focusing on procedural pain require a valid tool to measure this in young study participants.

The generally accepted standard for pain assessment is self-report. However, infants and children younger than 3 years of age are unable to self-report pain and behavioral observation scales are one of the most commonly used alternatives. More than a decade ago, >60 pain assessment tools were identified in the literature; many of which were designed for either neonates or infants and children experiencing postoperative pain.[4] Each of these scales has been subjected to varying levels of psychometric testing and many have been used in a variety of circumstances other than for which they were originally intended. There is currently insufficient evidence to confidently recommend an existing scale for procedural pain assessment for infants and children.[11] Furthermore, some data suggest that available scales may be practically or psychometrically unsuitable for procedural pain assessment and that they may have difficulty differentiating pain from other distress-related behaviors.[1,27]

Measuring procedural pain in young children is challenging, yet identifying a tool that works is a high priority for clinicians and researchers. The Face, Legs, Activity, Cry, and Consolability (FLACC) scale (Table 1) was first published in 1997.[21] The FLACC scale scores pain intensity by rating 5 behaviors (face, legs, activity, consolability, and cry) to derive a score out of 10. The descriptors for each item are considered indicative of behaviors exhibited by children in pain and the descriptors associated with each score level to represent an escalation consistent with increasing pain intensity. It was originally designed to assess postoperative pain in infants and children, aged 2 months to 7 years and is now one of the most commonly used pain assessment scales including, as an outcome measure in studies examining procedural pain and procedural pain management strategies.[11] It has also been recommended for procedural use in 2 early systematic reviews[9,35] and is frequently referenced in guidelines as a suitable choice for procedural pain assessment on the basis of available data.[3,18,28] Most notably, Gomez and colleagues reported high levels of agreement (intraclass correlation [ICC] = .80-.97) between reviewers assessing toddlers aged 12 to 18 months receiving immunization[16] although data to support validity was limited by the quality of the study. However, a study also focusing on immunization-related pain using an independent t-test showed the potential of the FLACC scale by showing responsiveness (increase in scores from baseline to procedure: .6 vs 6.5, $P < .001$) and capacity to differentiate between groups (5.3 ± 3.3 vs 7.8 ± 1.9, $P < .001$).[29] However, a recent systematic review concluded that there was insufficient evidence to support the FLACC scale for procedural pain assessment although results from several of the included studies was promising.[11]

The FLACC scale is potentially a suitable tool for assessment of procedural pain in young children but in the absence of high-quality validity and reliability data, cannot be recommended as such. Furthermore, extensive use of this scale make establishment of the psychometrics of this scale for procedural pain assessment critical.[11] We therefore set out to conduct a rigorous validation study to assess the psychometric performance of this scale used to assess several commonly performed procedures in the ED setting.

Methods

We evaluated the psychometric and practical performance of the FLACC scale applied by 26 clinicians to assess 100 infants and young children experiencing procedural pain using data obtained from a more extensive study designed to assess the validity, reliability, utility, and feasibility of 3 scales used to assess procedural pain. The protocol, summarized in this report but presented in detail elsewhere[10] was approved by Royal Children's Hospital ethics committee (reference number: 35220B).

The Consensus-based Standards for the Selection of Health Measurement Instruments Checklist was used to support development of the design for this study. This checklist was developed to provide standards for evaluating the methodological quality of studies addressing the psychometric properties of health measurement instruments but can also be used to guide design and reporting of a study.[22]

Table 1. FLACC Scale

	Scoring		
Categories	0	1	2
Face	No particular expression or smile	Occasional grimace or frown, withdrawn, disinterested	Frequent to constant quivering chin, clenched jaw
Legs	Normal position or relaxed	Uneasy, restless, tense	Kicking or legs drawn up
Activity	Lying quietly, normal position, moves easily	Squirming, shifting back and forth, tense	Arched, rigid, or jerking
Cry	No cry (awake or asleep)	Moans or whimpers, occasional complaint	Crying steadily, screams or sobs, frequent complaints
Consolability	Content, relaxed	Reassured by occasional touching, hugging or being talked to, distractible	Difficult to console or comfort

NOTE. Item scores are summed to generate a total score out of 10.

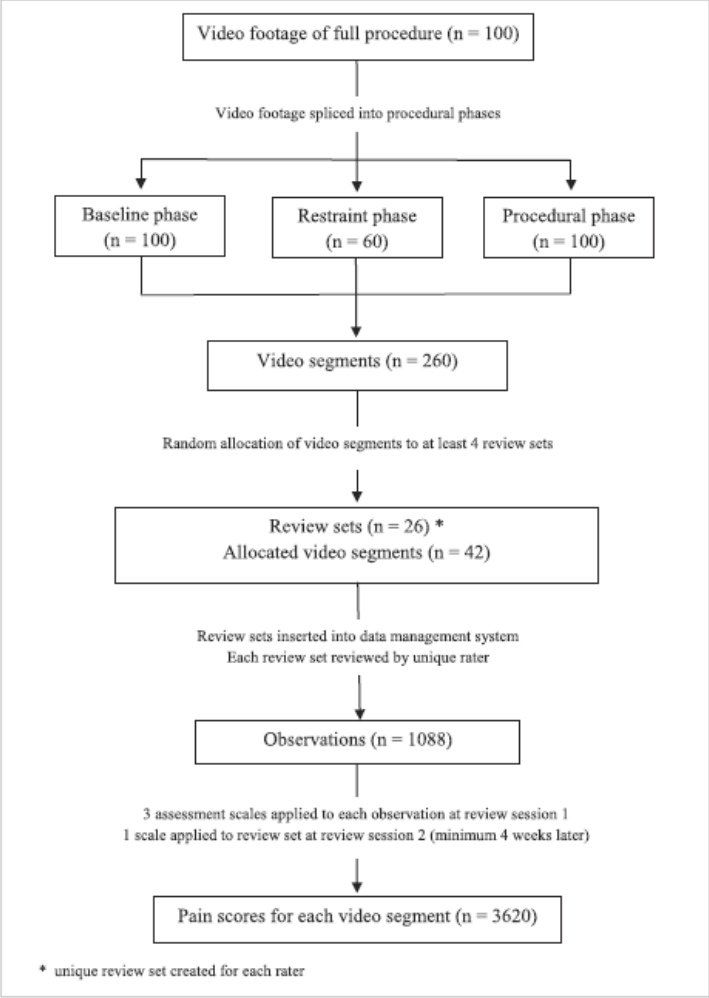


Figure 1. Overview from creation of video segments to final data set.

Sample and Sample Size

A convenience sample of 132 infants and young children aged 6 to 42 months undergoing 1 of 4 nominated presumed painful and nonpainful procedures in the ED at tertiary pediatric hospital in Melbourne, Australia was recruited. Children were excluded if they required immediate treatment, had a history of cognitive delay, altered conscious state or significant comorbid disease, the parent/caregiver did not speak English, or the video-recording of the procedure was incomplete or did not show the infant or child adequately.

The final sample of 100 video-recorded procedures for review included; 30 infants and children each for the painful procedure (intravenous catheter and nasogastric tube insertion), and 20 each for the nonpainful procedures (inhaled medication administration and cutaneous oxygen saturation [SpO2] measurement; Fig 1). Painful and nonpainful procedures were included to assess the capacity of the FLACC scale to distinguish between known groups. It was not possible to establish the most appropriate sample size to measure reliability because the true variation in the population was not known. The number of procedures chosen for this study was on the basis of the recommendations of the Consensus-based Standards for the Selection of Health Measurement Instruments Checklist, which rates a sample of at least 100 as "excellent."[22]

Twenty-six qualified doctors and nurses of any level of experience practicing in the ED were recruited to review the videos. A minimum of 25 reviewers was required to ensure that review sets were not prohibitively large (42 segments each) and that each review set was reviewed by at least 4 clinicians. Both samples are considerably larger than have been used in many similar studies.[16,24,33,36]

Procedure

Demographic and clinical data were collected during the ED presentation and a hand-held video recorder was used to digitally video-record the procedure. Researchers aimed to capture the infant's or child's face and body

and recorded them from before any contact with a clinician and ended when the procedure had been completed. All clinical decisions were made by the treating clinicians and no attempt was made to standardize the procedures.

The video recordings were divided into 15-second segments to demonstrate different phases of the procedure; baseline (before handling), preparation (tactile but nonpainful stimulus), and procedure (painful/nonpainful procedural stimulus). The baseline segment was the 15 seconds that immediately preceded contact with a clinician. The preparation phase commenced with the clinician touching the child and the segment for this phase was the 15 seconds immediately before the procedural phase. Finally, the procedural phase commenced when the procedural stimulus commenced (eg, needle piercing skin or application of mask). The procedural segment was the 15 seconds from the beginning of the procedure. The procedures presumed nonpainful were only divided into 2 segments (baseline and procedure) because a nonpainful preparatory phase could not be separated from the procedural phase. This resulted in a total of 260 segments of video for review. All video segments were organized into 15 seconds in length for evaluation, although the actual phases of the procedures in real time all exceeded 15 seconds.

The segments were allocated to 26 unique review sets to ensure that: all segments were reviewed by at least 4 reviewers, reviewers provided assessments of a range of procedures and phases but never for the same infant or child, and that different combinations of reviewers reviewed each segment. Allocation was automated using a Stata Statistical Software script (Stata Corp, College Station, TX). Similarly, the scales were applied to each video by the reviewer in varying order, the sequence of which was randomly allocated using a random sequence generator (https://www.random.org/sequences/) with only 1 stipulation: that each scale was applied first to a segment on equal numbers of occasions.

A purpose-built electronic data management system was developed to present the videos to the reviewer and allow for them to independently enter their scores. Reviewers allocated their first score after 1 uninterrupted viewing of the video segment. This was intended to as closely as possible replicate the circumstances of real-time clinical pain assessment. When the score was entered the reviewer was able to watch the video as many times as needed before entering their final pain score using the same scale. The final score was the score they considered unlikely to change with additional viewings of the video. Where reviewers were unable to score an item they were asked to indicate their reason; the effect of restraint, the procedure, or attempts to comfort the infant or child. There was also a free text option to capture other reasons for omitting an item score. To finish, reviewers completed a feasibility and utility questionnaire. A minimum of 4 weeks later reviewers repeated pain assessments using the first scale for the same video review set.

Instruments

The FLACC scale, the Visual Analog Scale (VAS) observer pain, the VAS observer (VASobs) distress and the Modified Behavioural Pain Scale were used in this study. Results for the Modified Behavioural Pain Scale and VASobs are reported elsewhere.

FLACC Scale

The FLACC scale (Table 1) is a composite of 5 behaviors ("face," "legs," "activity," "cry," and "consolability") considered indicative of pain that can be detected and graded by an observer. Each item is scored on a 0 to 2 scale resulting in a pain intensity score ranging from 0 to 10. The original instructions for use recommended observing the child for 1 to 5 minutes and matching the observed behaviors to those described in the scale for each item.

VASobs (Pain and Distress)

The VAS is a tool designed to measure and quantify subject experiences such as pain and distress.[2] The scale is a 10-cm line anchored at either end with labels such as "no pain" and "worst possible pain" or "no distress" and "worst possible distress." When applied by an observer they are asked to estimate the intensity of the pain or distress observed by placing a mark on the line. The distance from the 0 point on the line is measured to provide a pain intensity score from 0 to 10.

The VASobs scale was included to gain a separate estimate of the intensity of pain and the level of distress that reviewers perceived the infant to be experiencing during the procedure. These scores were used to explore convergence validity and the capacity of the FLACC scale to discriminate between pain and distress.

Feasibility and Utility Questionnaire

A feasibility and clinical utility questionnaire was used to capture the reviewers' assessments of how easy the scale is to use and how well it performs.[12] It includes 9 statements that are rated using a 5-point Likert scale to assess the extent to which the reviewer agrees with the statement (1 = strongly disagree and 5 = strongly agree). The questionnaire is available in Supplementary Table 1.

Statistical Analysis

Statistical analyses were conducted using the statistical software package R: A language and environment for statistical computing.[26] Descriptive statistics were used to report the demographic characteristics, pain scores, the percentage of complete pain scores, and the reason for unscored items. Incomplete scores were treated as missing data for the remaining analyses. Wilcoxon signed rank test was used to compare the first and final FLACC scores to identify whether multiple viewings of the video segment significantly affect the final scores. All subsequent analyses were on the basis of the observers' first score. Reliability was examined using the ICCs (1-way random effects) and Bland–Altman statistics were also used to assess agreement. Kappa scores were calculated for the 5 scale items using methods designed for multiple raters and categories and the presence of missing values. Convergent validity was tested by comparing

FLACC scale scores with VASobs scores using the Spearman correlation coefficient. We completed receiver operating characteristic analysis and area under the curve (AUC); sensitivity and specificity at various FLACC score cutoffs were calculated to assess the capacity of the scale to detect the difference between known groups.

Using lme4[5] in R,[26] the responsiveness of the scale to changes in pain was determined by analysis of the change in scores over the phases of the procedure using linear mixed models to estimate the fixed effects of time (phase of the procedure) and procedure type (painful vs nonpainful). The model also accounted for an interaction between the phase and the procedure type. Infants and children and reviewers were considered as random effects. This method of analysis allows for correlated data (ie, repeated measures from the same child), the effect of multiple independent variables, and missing data. Confidence intervals (CIs) for fixed effects were computed using bootstrap samples as implemented in the confint function in R.[26]

To determine the effect of preprocedural distress, children were grouped according to their scores in the baseline phase (I<3 vs ≥3) and the mean difference between their scores across phases and the procedure were compared using the t-test.

Ninety-five percent CIs and *P* values set for significance at .05 were used to establish statistical significance where appropriate.

Results

One hundred infants and children filmed having had 1 of 4 procedures; intravenous cannula insertion (n = 30), nasogastric tube insertion (n = 30), inhaled medication administration (n = 20), or oxygen saturation measurement (n = 20) contributed 260 video segments (Fig 1). The mean age of the children was 22.5 (±10.3) months and 58% (n = 58) were boys. Thirty-eight percent were diagnosed with a respiratory disease, 29% with dehydration and gastroenteritis, whereas the remaining 36% spanned a range of disparate diagnoses such as; fever and/or viral illness (n = 3), pathology testing only (n = 2), rash (n = 2), convulsion (n = 1), collapse (n = 1), hypoglycemia (n = 1), and mouth ulcers (n = 1).

Twenty-six ED clinicians (19 nurses and 7 physicians) reported experience ranging from <1 year to 20 years (mean = 10.1) Twelve (63%) nurses had completed postgraduate specialty training in pediatrics and/or emergency care and 3 of the 7 physicians had completed specialty training. The remaining physicians were part way through their pediatric or emergency specialty training and all nurses had had hospital-based education to prepare them for caring for infants and children in pain at the time they participated in this study. Importantly, all participants had regular (weekly) experience of the included procedures in children before participating in the study.

At the first review session a total of 1,088 observations were made of the 260 segments of video using the FLACC and VASobs pain and VASobs distress. Forty-eight segments of video were reviewed by 5 reviewers whereas the remaining 212 were reviewed by 4 reviewers. All data were included in the analysis. Another 356 observations using the FLACC scale were made at the second review. The mean, median, and distribution of FLACC scores across the phases of each procedure from review session 1 are presented in box plots in Fig 2. Mean scores during the procedure phase were highest for nasogastric tube insertion (9.5, SD ± .8), followed by intravenous catheter insertion (6.4, SD ± 3.1), and lowest for SpO2 measurement (.5, SD ± .9). Mean scores for infants and children at baseline were not 0 for any procedure and ranged from .3 (±SD .6) for SpO2 measurement to 1.6 (SD ± 2.3) and 1.7 (SD ± 2.7) for nasogastric tube and intravenous catheter insertion, respectively. Scores for all procedures were not normally distributed.

Psychometric Evaluation

Feasibility and Clinical Utility

Reviewers were unable to score a total of 261 (4.8%) FLACC scale items resulting in 159 (14.6%) incomplete scores (Table 2). The most frequent impediment to allocation of an item was restraint—precluding assessment of body movements ("leg" and "activity"). Reviewers reported that the absence of attempts to console the infant or child prevented the allocation of a score for the "consolability" item on 30 occasions.

Using the 5-point Likert scale to indicate the extent to which they agreed (5) or disagreed (1) with statements, reviewers agreed (score "4" and "5") that the FLACC scale as easy to understand (50%) and quick (34.5%) and easy (46.2%) to apply. Their ratings were lower for the extent to which they agreed that the FLACC scale is clinically useful (26.9%), reflects the extent of procedural pain (26.9%), discriminates infants and children with pain from those without (19.3%), is readily understood and supports decisions about pain management (7.7%), and reflects procedural pain-specific features (23.1%).

As another measure of clinical utility, the first scores allocated by reviewers after 1 uninterrupted view of the video segment were compared with their final scores and although the correlation between scores was very strong (r = .91), a third of scores (33.0%) changed by "1" and a Wilcoxon signed rank test showed a difference between the first and final score (Z = –2.136, *P* = .033).

Reliability

ICCs were calculated to establish interrater reliability and ranged from .79 to .94 depending on the procedure type and phase of the procedure. Fig 3 shows the extent of the variability of scores according to procedure type and phase on the basis of the mean score and no obvious patterns emerged for painful procedures. The data suggest a tendency for greater variability for scores across phases for nonpainful procedures. Linear mixed modeling included reviewers as a fixed effect and results reveal that the effect of the reviewer on FLACC scale scores is very low (variance = .004 ± .063). Finally, κ scores for scale items were; .61 (consolability), .72 (face, legs, and activity), and .82 (cry).

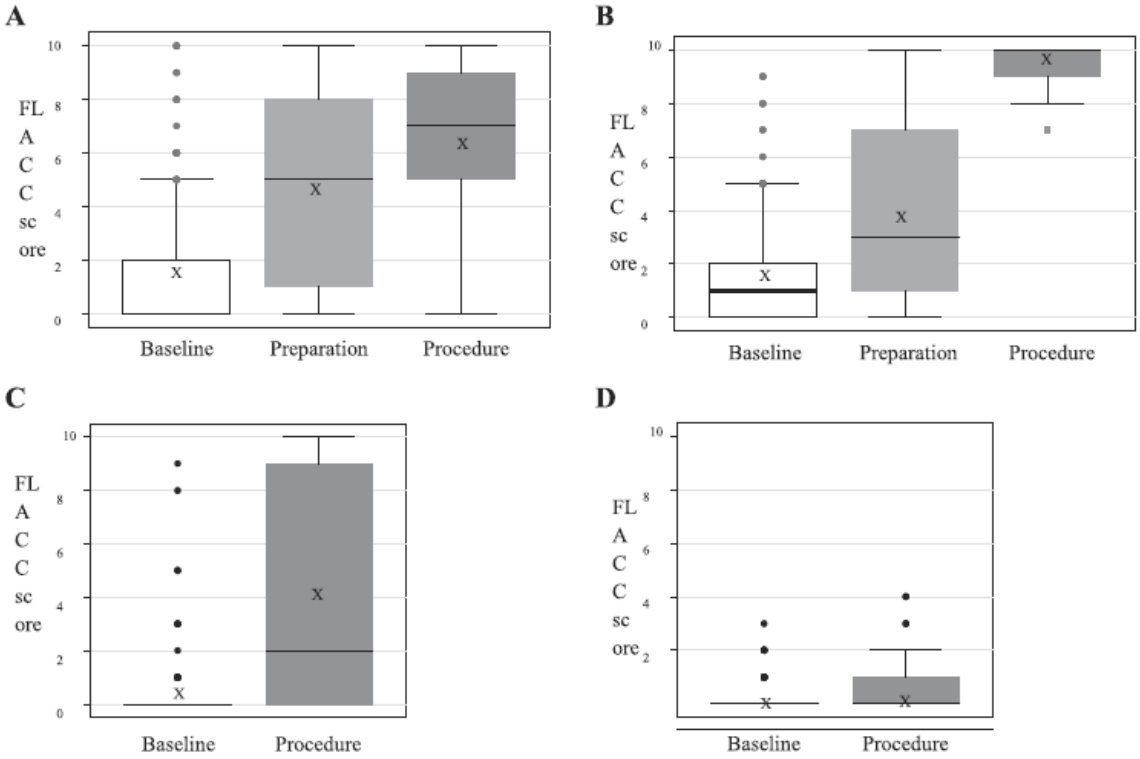


Figure 2. Boxplots for FLACC scores for each phase of each procedure. **(A)** Intravenous (IV) catheter insertion, **(B)** nasogastric tube (NGT) insertion, **(C)** inhaled medication administration, and **(D)** SpO2 measurement. Mean values are highlighted by the X.

The ICC for intrarater reliability was similarly high (ICC = .87). Intrarater agreement was also assessed using a Bland–Altman plot (Fig 4) and the mean difference between the scores was −.12 (±1.49), showing a trend toward lower scores at the second review.

Construct Validity: Known Groups, Responsiveness, and Discrimination

A box plot (Fig 5) is used to visually represent the relationships between the phases of the procedures and the types of procedures. The mean scores at baseline for painful procedures and nonpainful procedures were 1.63 ± 2.48 and .40 ± 1.18, respectively, and during the procedure were 7.47 ± 3.0 and 2.08 ± 3.3, respectively. The difference in means across phases (baseline to procedural) for painful and nonpainful procedures were 6.09 ± 3.36 and 1.99 ± 3.34, respectively.

We calculated the sensitivity, specificity, and AUC to analyze the capacity of the scale to differentiate between painful and nonpainful procedures using varying cutoff values for the FLACC score. The results are presented in

Table 2. Capacity for Reviewer to Score FLACC Items

	Item					
	Face	Legs	Activity	Cry	Consolability	Total Score
Valid item score	1,066 (98.0)	987 (90.7)	1,000 (92.0)	1,082 (99.5)	1,044 (96.0)	929 (85.39)
Not applicable	22 (2.0)	101 (9.3)	88 (8.0)	6 (.5)	44 (4.0)	159 (14.6)
Restraint interference	–	94	85	–	4	120
Procedural interference	20	4	1	6	8	32
Comfort interference	2	2	1	–	2	7
Other	–	1	1	–	30	29

NOTE. Values are frequency (%) unless otherwise indicated.

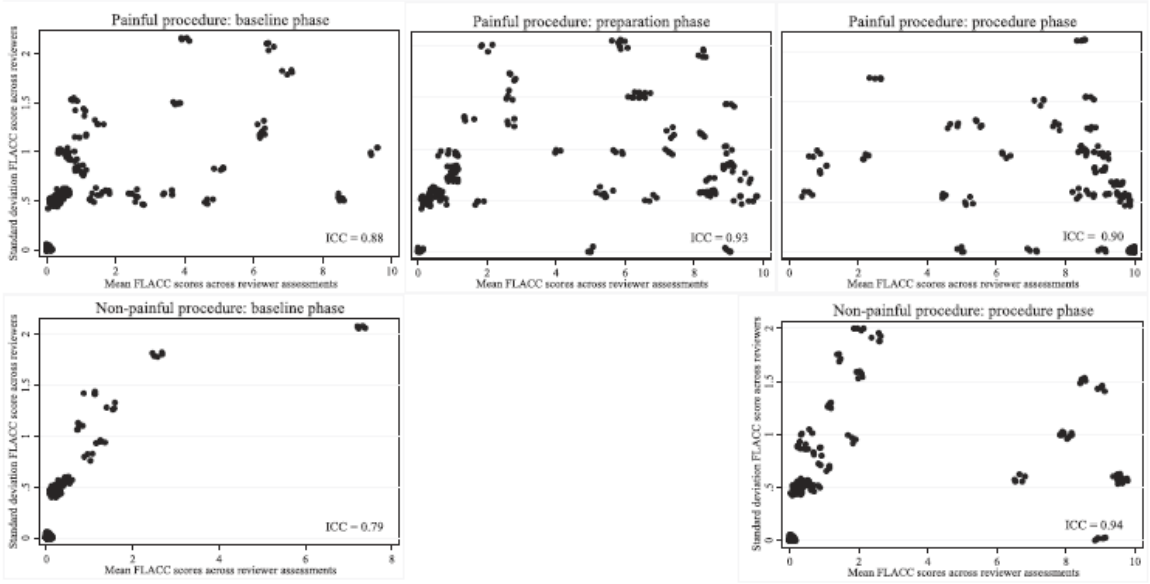


Figure 3. Inter-rater reliability: variation of reviewer assessments within child (standard deviation displayed on y-axis) versus average rating over all assessments (mean displayed on x-axis). Overall intraclass correlation coefficient (ICC): 0.92 (using one-way random effects model).

Table 3 and show that a cutoff >2 results in optimum sensitivity (94.9%), specificity (72.5%), and AUC value (.83) to distinguish between groups.

To adjust for the underlying variation in the child and the baseline FLACC scores when comparing FLACC scores for the nonpainful versus painful procedures, we fitted a linear mixed model with infant and child- as well as reviewer-specific random effects and an indicator variable defining whether the procedure was a painful procedure (Table 4). The model indicates that scores at baseline were not 0 for either group (intercept = .4708), but that these scores were slightly higher for the painful than for nonpainful procedures (mean difference in score was 1.19, 95% CI = .35-1.95). The model also confirmed that the scores during the procedural phase for children undergoing a painful procedure were markedly higher than during a nonpainful procedure (mean difference in score was 5.35 with 95% CI = 3.95-6.62). The model also confirmed scale responsiveness by including interaction between the procedure types (painful vs nonpainful procedure) and the phase of the procedure (baseline vs during procedure) while adjusting for the child- and reviewer-specific random effect. The resultant differences-in-difference between FLACC score for painful versus nonpainful procedure from baseline to the procedure phase was 4.2 (95% CI = 3.67-4.81). However,

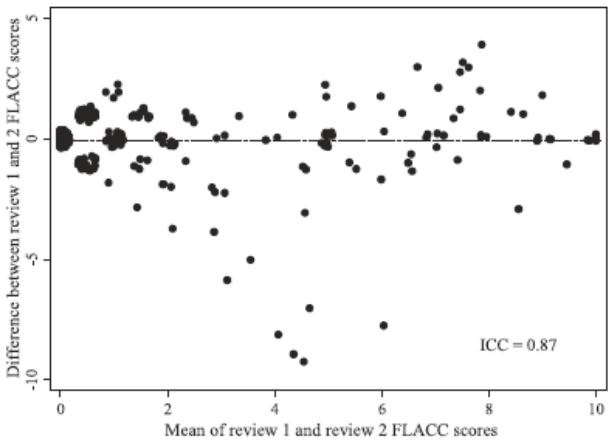


Figure 4. Intra-rater reliability: Bland–Altman plot showing variation between review scores (difference between review 1 and 2 FLACC scores are displayed on the y-axis) versus average FLACC score across review sessions (mean of paired scores for review session 1 and 2 are displayed on the x-axis). Mean difference is –.12 (SD = 1.49) and 95% limits of agreement are –3.10 and 2.86; ICC using 1-way random effects model. Zero line is highlighted with a dotted line.

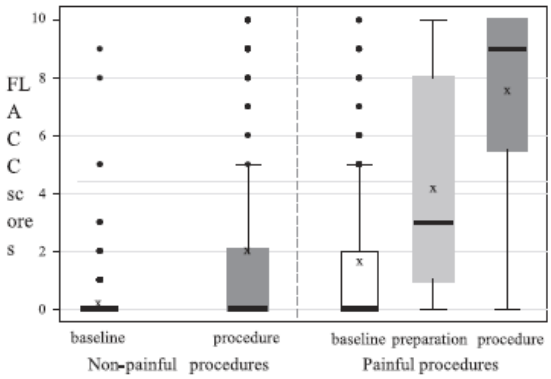


Figure 5. Box plots representing change of values over procedure phases (baseline, preparation, and procedure) in the 2 procedure cohorts (painful and nonpainful procedures). Mean values are highlighted by the X.

Table 3. Sensitivity, Specificity, and AUC Calculated for Different Cutoff FLACC Scores to Differentiate Procedure Type

FLACC Score Cutoff	*Sensitivity, %*	*Specificity–*	*AUC*
Score > 0	100.0	20.0	.60
Score > 1	94.9	60.0	.77
Score > 2	94.9	72.5	.83
Score > 3	91.5	75.0	.83
Score > 4	91.5	75.0	.83
Score > 5	84.8	75.0	.80
Score > 6	81.4	75.0	.78
Score > 7	76.3	77.5	.77

Table 4. The Variances and Estimates for Random and Fixed Effects for the Model Used to Show the Responsiveness of FLACC Scores to the Procedural Phase and Procedure Type (Painful vs Nonpainful)

Random Effects	Variance	SD	
Patient (intercept)	3.951	1.988	
Reviewer (intercept)	.004	.063	
Residual	4.151	2.037	
Fixed Effects	*Estimate*	*Standard Error*	*t Value*
Intercept	.471	.349	1.348
Painful procedure	1.248	.454	2.750
Preparation phase	2.499	.198	12.594
Procedure phase	1.782	.225	7.913
Procedure phase: painful procedure	4.122	.314	13.104

NOTE. Model = (1|reviewer) + (1|child) + procedure_type + procedure_phase + procedure_type × procedure_phase.

there was a difference in baseline to procedure scores for non-painful procedures (mean difference in score was 1.78).

Independent t-tests were also run to determine if there were differences in the mean change in scores across phases of the procedure for infants and children who had higher scores before the procedure than those with lower scores before the procedure. Although the mean difference in FLACC scores from baseline to the procedure of children with a baseline score of <3 (4.57 ± 3.99) differed from those of children with a baseline FLACC score of ≥3 (2.42 ± 2.77), this difference was not statistically significant (t_{87} = 1.80, P = .076). However, comparison between the mean difference in FLACC scores from baseline to the procedure for infants and children undergoing a painful procedure showed a significant difference for those with a baseline score of <3 (7.01 ± 2.90) compared with children with a score of ≥3 (2.70 ± 2.72, t_{48} = 4.40, P = .0001).

Convergent Validity

Spearman correlation coefficients for correlations between FLACC and VASobs pain and VASobs distress were higher for distress (r = .89) and pain (r = .74). Correlation of VASobs pain scores with VASobs distress scores was .77.

Discussion

We believe that results of this study make a significant contribution to addressing gaps in our understanding of the psychometrics of the FLACC scale used to assess procedural pain in infants and young children unable to self-report pain. These results also confirm, to some extent the FLACC scale as an option for assessment of procedural pain. However, some important questions and concerns about the scale used to assess procedural pain remain. The reliability of FLACC scores was good and the data confirm the sensitivity of the scale to procedural pain. In contrast, the feasibility of scoring pain during a procedure was impaired by the circumstances of the procedure and has implications for the specificity of the scale for pain associated with procedures.

This study convincingly confirms the capacity of the scale to detect pain in the children in this study undergoing a painful procedure evidenced by an increase in scores from the baseline phase to the procedure phase for children experiencing a painful procedure and the capacity of the scale to differentiate between children experiencing a painful and a nonpainful procedure, evidenced by very high sensitivity and high specificity. Strong correlation between FLACC scores and VASobs pain scores, suggesting that these scales measure a similar construct, adds indirect weight to this contention. However, use of this approach can be challenged because it is contingent on accepting that the VAS applied by an observer to assess pain is a valid measure of pain. This scale is widely used in other studies to establish convergence validity in the absence of a "gold standard" or a suitable alternative that is reliable and valid in these circumstances.

However, a 2002 review reports that there was at that time insufficient evidence to assert that VASobs was valid for procedural pain assessment.[32] Furthermore, if the VASobs measures the same construct logic supports use of a less complex scale for assessment of procedural pain, particularly because of the limitations of the FLACC.

The decision to use clinicians and a larger sample of reviewers than is generally used in studies examining the psychometrics of a scale recognizes the contribution that the rater may have on the performance of an assessment tool. Scales like the FLACC scale are often evaluated in studies with small numbers of specifically trained raters designed to ensure higher levels of inter-rater reliability.[16,20] This approach, although reasonable to assess scale performance in research conditions, has done little to add to our understanding of the reliability of scores across a range of clinicians. The techniques used to determine reliability (ICC, Bland–Altman statistics, and linear mixed modeling) all consistently show high levels of intra- and inter-rater reliability. Furthermore, the plots from Figs 3 and 4 support a contention that reliability is consistent across scores from 0 to 10. Our data show that FLACC scores allocated by clinicians (doctors and nurses) of varying experience are likely to be reliable. However, the lower κ values for certain scale items cannot be entirely ignored.

Despite high sensitivity for pain, question remains about the capacity of the FLACC scale to distinguish between distress behaviors associated with pain and those associated with other experiences such as fear and anxiety, which are frequently linked with procedures. Infants and children in this study did not consistently score "0" during the nonpainful phases of painful procedures or during nonpainful procedures. Additionally, although the specificity of the FLACC scale to differentiate between the painful and nonpainful groups was high (72.5%) this was on the basis of a cutoff score of >2, which was chosen in preference to 3 and 4 because it provides for the highest sensitivity. There are no studies that examine cutoff scores that define pain intensity or a minimally clinically significant difference for FLACC scores. However, intuitively a score of 2 does not suggest that the child is pain-free and empirically this is supported by the results of studies evaluating cutoffs and minimally clinically significant differences for the VAS and the numeric rating scale in acute and postoperative pain, which show that children report a difference in pain with changes of 10 mm[25] and a score of "1."[34] The demonstration of responsiveness was used to provide evidence of validity of a pain scale. However, for children already exhibiting distress before the painful stimulus, FLACC scores did not increase as much as they did for children with lower scores, which is likely the result of range restriction of the scale, where the scores of distressed children cannot extend beyond the maximum value of the scale. Results similar to these are described by Ahola Kohut and Pillai Riddell in their study examining the Neonatal Facial Coding System specificity for pain.[1] The risk for children distressed by the circumstances as a result of the potential blunting of score responsiveness in contrast to those children who score "0" at baseline, is that their pain is not recognized or minimized resulting in inadequate management. Finally, the clinicians in this study also raised their concerns about the extent to which the FLACC scale "reflects procedural pain specific features" and the "extent of procedural pain" and the capacity of the scale to "discriminate children in pain from those without pain." Data from the study cannot confirm the specificity of the FLACC scale for assessing procedural pain in infants and young children and support a long-held view that pain scales are measures of distress, a composite of pain and nonpain related factors, and not a measure exclusively of pain.[6]

Chang and her colleagues have previously expressed concerns about the descriptors for the face item, stating that they describe behaviors not often seen in infants in pain. They state that this may explain the relatively poor levels of agreement for the "face" item in their study.[7] These results are not supported in the current study, where reliability for the face item (κ = .72) was consistent with the level of agreement for other scale items (κ scores ranged from .61 to .82). Furthermore, clinicians did not indicate that inconsistency in the descriptors for the "face" item and the behaviors exhibited by the child prevented them from scoring this item. These results invite an interesting line for enquiry regarding the application of the scale by clinicians and the extent to which their clinical judgement influences their scores. There is some existing evidence to suggest that clinicians do apply judgement when assessing pain[13,14] and that this occurs with application of the FLACC.[16]

However, the results from this study indicate that this cohort of clinicians have other practical concerns about the FLACC scale used to assess procedural pain. The circumstances of the procedure, most commonly restraint use and the absence of attempts to console, had a significant effect on their ability to apply items of the scale, resulting in incomplete scores for almost 1 in 8 observations and only "fair" reliability for the "consolability" item indicated by a κ score of .61. Although there has been significant efforts to reduce the use of restraint during procedures, its practice as seen in this study has not been completely eliminated.[8] For a scale to be practical and valid for procedural use it must account for the effects of restraint. Furthermore, because procedural sedation to manage procedural distress has gained increasing favor, the effect of sedation should also be included when considering the capacity of a behavioral tool to sensitively capture pain and differentiate this from other nonpain-related distress.

Interpretation of the results of this study introduces several lines of debate relating to psychophysics, some of which have already been addressed in the discussion, but also include the attribution of the psychometric properties to the scale when they should be more rightly attributed to the scores. The performance of a measurement scale is contingent on the circumstances and population to which it was applied and who applied the scale. We largely draw our conclusions about the scores and use these to make tentative comments about the performance more broadly of the scale. There is also a risk that the result will be interpreted on the basis of an

assumption that the FLACC scale is a ratio scale; this is yet to be established.

Strengths and Limitations

This was a single-center study and raters could not be blinded to the circumstances (eg, needle insertion). Because the data to accept the VASobs as well validated for procedural pain assessment is limited,[32] use of this scale for convergence testing was a limitation of this study. The use of video recordings although a limitation because it differs from clinical practice, is also a strength because it allows for multiple reviews to score the same situation. However, strategies were used to overcome the effects of these limitations; recruitment of a larger than usual sample size, the use of unique reviewers to assess each phase of a child's procedure, and application of the scale after a single uninterrupted view of the video segment to generate the first score. In addition, multiple methods were used to establish validity to overcome the limitations of each validation method. However, methodological limitations such as those identified should be considered when interpreting the results. Studies addressing the psychometric properties of pain scales frequently use nonclinical research assistants and their results used to claim validity of the scale for clinical use. However, because there is some evidence that clinicians apply clinical judgement when applying assessment scales,[13,14,16] using clinicians provided an opportunity to test the assumption that the scale will perform adequately when applied by clinicians.

Conclusions

The FLACC scale has been widely used to assess procedural pain in infants and young children unable to self-report despite a lack of evidence to support the psychometrics when used for this purpose. This study makes a contribution to our understanding of the performance of the FLACC scale in these circumstances. Although data addressing the sensitivity and specificity of the FLACC scale applied to a broader range of procedures are needed, as are efforts to determine cutoffs to define intensity and minimal clinically significant differences, it is possible to conclude from our data that the FLACC scale is sensitive to procedural pain. However, on the basis of non-zero scores during presumed nonpainful phases and procedures and a cutoff to distinguish painful from nonpainful procedures of 2, doubt remains about the capacity of the scale to differentiate between pain and nonpain-related distress. Indeed it is likely that distinguishing between pain-related and nonpain-related distress cannot be achieved using observable behaviors. This has significant implications for pain assessment during medical procedures, where fear and anxiety are likely to be high. Current clinical and research practice must recognize that FLACC scores quantify procedural distress and should not be relied upon where discrimination between pain and nonpain-related distress is required. Furthermore, we have shown that the circumstances of the procedure has the potential to interfere with application of the scale. Revision of the FLACC scale may overcome the problems associated with restraint and other procedure-related concerns. The FLACC scale may profit from additional work to better understand how restraint, the absence of attempts to console, the appropriateness of their attempts to console, and the procedure effect on application. This may assist us to make appropriate adjustments in a revision of the scale for procedural use. On the basis of the deficiencies, particularly in the "face" and "consolability" items and the effect of restraint on the "legs" and "activity" items, revision may require changes to each of these items. The descriptors for the "face" item require amendment to better reflect common pain-related facial expressions (eg, the included "chin quiver" is rarely seen in infants experiencing pain). "legs" and "activity" scoring may be more feasible if an option is provided that acknowledges the use of varying levels of restraint. Similarly, "consolability" scoring options may require inclusion of scoring options for which interventions aimed at consoling the child are not offered or are inappropriate. These modifications will require similar testing to determine the revised version's role in procedural pain assessment.

Acknowledgments

Dianne Crellin is a trainee member of Pain in Child Health, a research training initiative of the Canadian Institutes of Health Research.

Supplementary Data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jpain.2018.02.013.

References

1. Ahola Kohut S, Pillai Riddell R: Does the Neonatal Facial Coding System differentiate between infants experiencing pain-related and non-pain-related distress? J Pain 10:214-220, 2009

2. Aitken RC: Measurement of feelings using visual analogue scales. Proc R Soc Med 62:989-993, 1969

3. American Academy of Pediatrics Committee on Psychosocial Aspects of Child Family Health, American Pain Society Task Force on Pain in Infants Children Adolescents: The assessment and management of acute pain in infants, children, and adolescents. Pediatrics 108:793-797, 2001

4. Andersen RD, Langius-Eklof A, Nakstad B, Bernklev T, Jylli L: The measurement properties of pediatric observational pain scales: A systematic review of reviews. Int J Nurs Stud 73:93-101, 2017

5. Bates D, Machler M, Bolker B, Walker S: Fitting linear mixed-effects models using {lme4}. J Stat Softw 67:1-48, 2015

6. Blount RL, Loiselle KA: Behavioural assessment of pediatric pain. Pain Res Manag 14:47-52, 2009

7. Chang J, Versloot J, Fashler SR, McCrystal KN, Craig KD: Pain assessment in children: Validity of facial expression items in observational pain scales. Clin J Pain 31:189-197, 2015

8. Crellin D, Babl FE, Sullivan TP, Cheng J, O'Sullivan R, Hutchinson A: Procedural restraint use in preverbal and early-verbal children. Pediatr Emerg Care 27:622-627, 2011

9. Crellin D, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A: Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. Paediatr Anaesth 17:720-733, 2007

10. Crellin DJ, Harrison D, Hutchinson A, Schuster T, Santamaria N, Babl FE: Procedural Pain Scale Evaluation (PROPoSE) study: Protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months. BMJ Open 7:2017

11. Crellin DJ, Harrison D, Santamaria N, Babl FE: Systematic review of the FLACC scale for assessing pain in infants and children: Is it reliable, valid, and feasible for use? Pain 156:2132-2151, 2015

12. de Jong A, Baartmans M, Bremer M, van Komen R, Middelkoop E, Tuinebreijer W, van Loey N: Reliability, validity and clinical utility of three types of pain behavioural observation scales for young children with burns aged 0-5 years. Pain 150:561-567, 2010

13. De Ruddere L, Goubert L, Stevens M, de Williams C, Amanda C, Crombez G: Discounting pain in the absence of medical evidence is explained by negative evaluation of the patient. Pain 154:669-676, 2013

14. De Ruddere L, Goubert L, Stevens MA, Deveugele M, Craig KD, Crombez G: Health care professionals' reactions to patient pain: Impact of knowledge about medical evidence and psychosocial influences. J Pain 15:262-270, 2014

16. Gomez RJ, Barrowman N, Elia S, Manias E, Royle J, Harrison D: Establishing intra- and inter-rater agreement of the face, legs, activity, cry, consolability scale for evaluating pain in toddlers during immunization. Pain Res Manag 18:e124-e128, 2013

17. Herd DW, Babl FE, Gilhotra Y, Huckson S, PREDICT group: Pain management practices in paediatric emergency departments in Australia and New Zealand: A clinical and organizational audit by National Health and Medical Research Council's National Institute of Clinical Studies and Paediatric Research in Emergency Departments International Collaborative. Emerg Med Australas 21:210-221, 2009

18. Howard R, Carter B, Curry J, Morton N, Rivett K, Rose M, Tyrrell J, Walker S, Williams G: Pain Assessment. Pediatr Anaesth 18:14-18, 2008

19. MacLean S, Obispo J, Young KD: The gap between pediatric emergency department procedural pain management treatments available and actual practice. Pediatr Emerg Care 23:87-93, 2007

20. Manworren RC, Hynan LS: Clinical validation of FLACC: Preverbal patient pain scale. Pediatr Nurs 29:140-146, 2003

21. Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S: The FLACC: A behavioral scale for scoring postoperative pain in young children. Pediatr Nurs 23:293-297, 1997

22. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, 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 19: 539-549, 2010

23. Motov SM, Khan AN: Problems and barriers of pain management in the emergency department: Are we ever going to get better? J Pain Res 2:5-11, 2009

24. Nilsson S, Finnstrom B, Kokinsky E: The FLACC behavioral scale for procedural pain assessment in children aged 5–16 years. Paediatr Anaesth 18:767-774, 2008

25. Powell CV, Kelly AM, Williams A: Determining the minimum clinically significant difference in visual analog pain score for children. Ann Emerg Med 37:28-31, 2001

26. R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria; 2016

27. Ranger M, Celeste Johnston C, Rennick JE, Limperopoulos C, Heldt T, du Plessis AJ: A multidimensional approach to pain assessment in critically ill infants during a painful procedure. Clin J Pain 29:613-620, 2013

28. Schug SA, Palmer GM, Scott DA, Halliwell R, Trinca J, APM: SE Working Group of the Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine: Acute Pain Management: Scientific Evidence, 4th ed. Melbourne, ANZCA & FPM, 2015

29. Taddio A, Hogan ME, Moyer P, Girgis A, Gerges S, Wang L, Ipp M: Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. Vaccine 29:1390-1394, 2011

30. Todd KH, Ducharme J, Choiniere M, Crandall CS, Fosnocht DE, Homel P, Tanabe P: PEMI Study Group. Pain in the emergency department: Results of the pain and emergency medicine initiative (PEMI) multicenter study. J Pain 8:460-466, 2007

31. Valeri BO, Holsti L, Linhares MB: Neonatal pain and developmental outcomes in children born preterm: A systematic review. Clin J Pain 31:355-362, 2015

32. van Dijk M, Koot HM, Saad HH, Tibboel D, Passchier J: Observational visual analog scale in pediatric pain assessment: Useful tool or good riddance? Clin J Pain 18:310-316, 2002

33. Voepel-Lewis T, Zanotti J, Dammeyer JA, Merkel S: Reliability and validity of the face, legs, activity, cry, consolability behavioral tool in assessing acute pain in critically ill patients. Am J Crit Care 19:55-61, quiz 62, 2010

34. Voepel-Lewis TM, Burke CN, Jeffreys N, Malviya S, Tait AR: Do 0–10 numeric rating scores translate into clinically meaningful pain measures for children? Anesth Analg 112: 415-421, 2011

35. von Baeyer CL, Spagrud LJ: Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. Pain 127:150, 2007

36. Willis MH, Merkel SI, Voepel-Lewis T, Malviya S: FLACC behavioural pain assessment scale: A comparison with a child's self-report. Pediatr Nurs 29:195-198, 2003

CHAPTER 11.

The psychometric properties of the MBPS used to assess procedural pain.

The psychometric properties of the Modified Behavioral Pain Scale (MBPS) used to assess procedural pain in infants and children are reported in this chapter. This work has been published and the PDF of this publication is reproduced in this chapter.

Publication

Crellin DJ, Babl FE, Santamaria N, Harrison D. The Psychometric Properties of the MBPS Scale Used to Assess Procedural Pain. J Pain. 2018;19(6):660-70.

American Pain Society
RESEARCH
EDUCATION
TREATMENT
ADVOCACY
PUBLISHED BY ELSEVIER

The Journal of Pain, Vol 19, No 6 (June), 2018: pp 660-669
Available online at www.jpain.org and www.sciencedirect.com

The Psychometric Properties of the MBPS Scale Used to Assess Procedural Pain

Dianne J. Crellin,*,†,‡ Franz E. Babl,†,‡,§ Nick Santamaria,* and Denise Harrison*,†,¶

*Department of Nursing, The University of Melbourne, Melbourne, Australia.
†Clinical Sciences, Murdoch Children's Research Institute, Melbourne, Australia.
‡Royal Children's Hospital, Melbourne, Australia.
§Department of Paediatrics, The University of Melbourne, Melbourne, Australia.
¶School of Nursing Faculty of Health Sciences, Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa, Ontario, Canada.

Abstract: **The Modified Behavioral Pain Scale (MBPS) was designed to assess procedural pain in infants and is considered valid for assessing immunization pain. The aim of this study was to assess the practical and psychometric properties of the MBPS when applied to other commonly performed procedures. Twenty-six clinicians independently applied the MBPS scale to segments of video collected from 100 infants and children aged 6 to 42 months undergoing 1 of 4 procedures in the emergency department. Positive correlation between MBPS and Visual Analogue Scale observer applied (VASobs) pain (r = .74) was shown and inter- and intrarater reliability coefficients were high (.87 and .89, respectively). Construct validity was shown by scale responsiveness to painful stimuli (4.6 times increase in scores across phases) and the capacity of the scale to distinguish between painful versus nonpainful procedures ($P < .001$). However, mean baseline scores for procedures were not 0 (likely a function of item descriptors for a "0" score) and the mean difference increased across phases for children with baseline scores >3, which was much lower than for children with scores <3 ($P = .0001$). Finally, 28% of scores changed after the second viewing of a video segment. The MBPS appears reliable and sensitive to procedural pain when applied by clinicians. Results question the capacity of the scale to differentiate between pain- and nonpain-related distress, the feasibility of this scale, and the appropriateness of item descriptors for medical procedures.**

Perspective: ***This article presents the psychometric and practical properties of the MBPS applied to assess procedural pain. Identification of a suitable scale for this purpose will support improved pain management in infants and young children who undergo painful procedures.***

Key words: *Modified Behavioral Pain Scale, pain assessment, reliability, validation/validity, psychometric evaluation, pediatrics, infants.*

Received August 17, 2017; Revised January 12, 2018; Accepted January 30, 2018.
This work was supported by a Murdoch Children's Research Institute clinical research grant.
The authors have no conflicts of interest to declare.
Supplementary data accompanying this article are available online at www.jpain.org and www.sciencedirect.com.
Address reprint requests to: Dianne Crellin, RN, BN, MN, Royal Children's Hospital and The University of Melbourne, c/o Emergency Department, Royal Children's Hospital, Flemington Road Parkville, 3052, Melbourne, Australia. E-mail: dianne.crellin@rch.org.au
1526-5900/$36.00

https://doi.org/10.1016/j.jpain.2018.01.012

The importance of assessing and managing procedural pain has gained increasing attention in recent years particularly as infants and children are frequently exposed to painful and distressing procedures as part of their health care. Although self-report is the accepted standard for pain assessment, infants and young children are developmentally unable to provide self-report. For this reason clinicians and researchers are reliant on proxy measures of pain intensity. Multidimensional observational pain scales have repeatedly been proposed as a viable option to support pain assessment in

Table 1. MBPS

Item	Descriptor	Score
Facial expression	Definite positive expression (smiling)	0
	Neutral expression	1
	Slightly negative expression (grimace)	2
	Definite negative expression (furrowed brow eyes closed tightly)	3
Cry	Laughing or giggling	0
	Not crying	1
	Moaning quiet vocalizing gentle or whimpering cry	2
	Full lunged cry or sobbing	3
	Full lunged cry more than baseline cry (scored only if child crying at baseline)	4
Movements	Usual movements and activity	0
	Resting and relaxed	0
	Partial movement (squirming, arching, limb-tensing, clenching)	2
	Attempt to avoid pain by withdrawing the limb where puncture is done	2
	Agitation with complex/generalized movements involving the head, torso, or other limbs	3
	Rigidity	3

NOTE. Item scores are summed to generate a total score of 10.

infants and children and >60 behavioral scales appear in the literature.[3] However, most scales are unsuitable for procedural pain assessment because of insufficient psychometric data and/or unsuitable design and this includes the most commonly recommended of all behavioral scales; the Face, Legs, Activity, Cry, and Consolability (FLACC) scale.[19]

The Modified Behavioral Pain Scale (MBPS) was developed >20 years ago at a time when fewer scales were available. The available scales were not considered sufficiently able to capture the variability in responses that may be shown by young infants to procedural pain and the FLACC scale[19] did not appear in the literature until 2 years later. To develop a clinically applicable pain intensity measure for this purpose, Taddio and her colleagues modified an existing scale, The Children's Hospital Eastern Ontario Pain Scale (CHEOPS),[18] which was developed for measuring postoperative pain in children 1 to 5 years of age. Several items of the CHEOPS were combined to create a single item in the MBPS and the "wound evaluation" item was dropped altogether. The newly created MBPS comprised 3 behavioral items; face, movement, and cry.[26] The role of facial expression for pain assessment is well supported in the literature and is included as an assessment item in most behavioral observation scales.[6,22] Similarly, cry and various cry characteristics, such as latency to cry and cry duration are considered manifestations of pain in infants and young children and also feature in many observation scales.[7] Finally, patterns in body movement have been shown to be exhibited in infants and children exposed to painful stimuli.[8] Each of the behaviors included is assessed, scored, and the scores added to generate a pain intensity score from 0 to 10. The items, their descriptors, and the associated scores are shown in Table 1.

The validity of the MBPS for procedural pain assessment is claimed on the basis of the results of Taddio's original study[26] and a subsequent study testing the measurement properties of the scale.[25] However, a recent systematic review concluded that psychometric data addressing the MBPS is almost exclusively focused on the scale's performance when used to assess the pain associated with immunization in infants and young children and to accept the scale for use in other contexts it must first be tested in in these circumstances.[10] Because of the frequency of which infants and young children undergo painful procedures such as injections, intravenous (i.v.) access, and other skin-breaking and painful procedures during medical treatment and the absence of a suitable alternative scale (eg, the FLACC scale) for procedural pain assessment, it is critical to extend our understanding of the psychometric properties of the MBPS to include procedures other than immunization.

The aim of this study was therefore to fill this gap in the literature and to test the psychometric and practical properties (feasibility, reliability, validity, and clinical utility) of the MBPS to quantify procedural pain intensity in infants and children aged from 6 to 42 months to determine its suitability for clinical and research purposes.

Methods

Study Design and Setting

This study was conducted in the emergency department (ED) of a tertiary hospital in Melbourne, Australia using a prospective observational noninterventional design. The Royal Children's Hospital, Melbourne ED has an annual census of approximately 90,000 children. The Consensus-based Standards for the Selection of Health Measurement Instruments Checklist was used to support the development of the design for this study.[20] This checklist was developed to provide standards for evaluating the methodological quality of studies addressing the psychometric properties of health measurement instruments. However, the authors of this checklist also suggest that it has a role to play in guiding study design and reporting.

Sample and Sample Size

Infants and children experiencing a medical procedure and ED clinicians (physicians and nurses) to assess their pain were recruited for this study.

Patient Participants

This was a convenience sample of infants and children aged between 6 and 42 months presenting to the ED who were filmed while experiencing 1 of 4 nominated procedures anticipated to be painful (nasogastric tube [NGT] or i.v. catheter insertion) and distressing but not painful (inhaled medication administration via mask and spacer device and oxygen saturation measurement). The families of 159 infants and children were approached to participate and 27 declined to participate. Data for 132 infants and children were collected and 100 selected on the basis of the quality of the video and to ensure the intended sample was included, painful and nonpainful procedures were included to assess the capacity of the MBPS to distinguish between known groups. Infants and children were excluded if they required immediate treatment, had a history of cognitive delay, altered conscious state, or significant comorbid disease, or the parent/caregiver did not speak English. Participants were also excluded if the video recording of the procedure was incomplete, of poor quality, or did not show the child adequately to allow for application of the scale.

The aim was to provide a final sample of 100 video-recorded procedures for review featuring 30 infants or children each for the painful procedures (i.v. and NGT insertion) and 20 each for the nonpainful procedures (inhaled medication delivery and oxygen saturation measurement). The sample size chosen for this study was on the basis of the recommendations of the Consensus-based Standards for the Selection of Health Measurement Instruments Checklist, which rate a sample of 50 to 99 as "good" and a sample of at least 100 as "excellent."[21]

Reviewer Participants

Qualified doctors and nurses of any level of experience practicing in the ED, from a pool of 45 doctors and 115 nurses were invited to participate. A minimum of 25 reviewers was required to ensure that review sets were not prohibitively large (41-42 segments each) and that each review set was reviewed by at least 4 clinicians.

Instruments

The MBPS (Table 1), a visual analog scale (VAS) observer pain and VAS observer distress were applied to segments of video by the recruited reviewers. When reviews were complete, reviewers also completed a clinical utility questionnaire. An electronic data management (EDM) tool was purposefully designed to present the videos to the reviewers and record the data.

MBPS

The MBPS (Table 1) is a composite of 3 behaviors ("face," "movement," and "cry") considered indicative of pain in young infants experiencing immunization. Data support the validity of the MBPS used to assess immunization pain. Because the pain experienced during immunization is likely to change rapidly, the observation periods are brief; baseline scores are generated in the 5 to 15 seconds before the procedure, and postimmunization scores within 15 seconds of the immunization. Baseline scores are often described as the reaction observed for most of the time. In contrast, immunization scores are frequently defined as the maximum reaction observed.[26] The sum of the scores for each item generate a pain intensity score ranging from 0 to 10. There are insufficient data to determine the validity of scores when used to assess pain associated with other procedures. The MBPS was used in its original format for this study.

VAS Observer (Pain and Distress)

The VAS is a 10-cm line anchored at either end with labels such as "no pain" and "worst possible pain" or no distress" and "worst possible distress." It was originally proposed for use as a scale for self-report of a subjective experience, including pain.[2] However, it has also been frequently used for individuals unable to self-report their experience and applied by an observer. The observer is asked to estimate the intensity of the pain or distress observed by placing a mark on the line. The distance from the 0 point on the line is measured to provide a pain intensity score from 0 to 10.

The VAS observer scale is included to gain a separate estimate of the intensity of pain and the level of distress that reviewers perceived the infant to be experiencing during the procedure. These scores will be used to explore convergence validity and the capacity of the MBPS scale to discriminate between pain and distress.

Feasibility and Utility Questionnaire

Reviewers provided an assessment of how easy the scale was to use and how well it performed using a clinical feasibility and utility scale developed by de Jong and colleagues,[12] which was on the basis of utility criteria defined by Harris and Warren.[16] Reviewers were asked to respond to 9 statements using a 5-point Likert scale to assess the extent to which the reviewer agrees with the statement. The questionnaire is available online in Supplementary Table 1.

Data Collection Tool

The EDM system was designed to present reviewers with a predetermined set of video segments and the scales in a predetermined sequence to assess the experience of the child in the video segment. It allowed reviewers to watch and review the video segments while simultaneously entering data into the database. The system was also designed to prevent reviewers from progressing until all fields were complete to avoid missing data. The system was developed specifically for this study by one of the researchers using FileMaker 7.0 (FileMaker, Inc, Santa Clara, CA).

Study Procedure

The study was conducted in a series of steps, which are described in detail in the following sections.

ED Presentation Data Collection

A research team member was responsible for approaching families of eligible infants and children, obtaining informed consent, collecting demographic and clinical data, and video recording the procedure. A hand-held video recorder was used and researchers aimed to capture the infant's face and body from the time the infant and their parent/caregiver were moved to the procedure area but before any contact to prepare the infant for the procedure occurred until the procedure was completed.

No effort was made to standardize the procedures and all clinical decisions were made by the treating clinicians and were on the basis of department and hospital guidelines. This resulted in the application of topical anaesthesia before i.v. and lubrication of the NGT before insertion and in many cases comfort and/or distraction during the procedure. Parents were present for the procedures and could offer whatever comfort or distraction they considered appropriate.

Video Data Preparation

Recordings were made of 132 infants and children during the data collection period. Twenty-three videos were considered unsuitable and excluded because they did not provide a clear view of the face and body movements of the infant/child. From the remaining 109 video recordings; 30 each of the painful procedures and 20 each of the nonpainful procedures were randomly selected for this study (100 in total).

The video recordings were divided into 15-second segments to show different phases of the procedure, which were defined as follows: 1) baseline—before any attempt to prepare the infant/child for the procedure was made (eg, while still in parents arms but in the procedure space); (2) preparation phase of the procedure (eg, immediately before the procedure commenced while restrained and/or being physically prepared); 3) during the first attempt of the painful/distressing part of the procedure (eg, at the time of needle/tube insertion or application of face mask or oxygen measurement probe).

The procedures presumed painful were divided into these 3 segments. The procedures presumed nonpainful were only divided into 2 segments (1 and 3) because a nonpainful preparatory phase could not be separated from the procedural phase. This resulted in a total of 260 segments of video for review.

The segments were allocated to review sets (1 unique review set per reviewer) to ensure that: all segments were reviewed by at least 4 reviewers, reviewers provided assessments of a range of procedures and phases but never for the same child, and that different combinations of reviewers reviewed each segment. Allocation was automated using a Stata[24] script to prevent bias occurring with manual allocation of segments to review sets.

The scales were applied to each video segment in varying order. The sequencing of the scales was randomly allocated with only 1 stipulation: that each scale was applied first on equal numbers of occasions. This sequencing was generated using a random sequence generator (https://www.random.org/sequences/).

Pain Scoring

After an education session to familiarize the reviewer with scales and the EDM system, they were provided with an opportunity to trial the data collection system and apply the scales. The instructions to the reviewers regarding application of the MBPS were similar to those described by Taddio et al.[26]

Two review sessions for each reviewer where conducted a minimum of 4 weeks apart. On each occasion they were provided with headphones and laminated copies of the MBPS and a pain scale ruler (marked on one side and blank on the other) with a slider to use for VAS observer pain and VAS observer distress scoring.

Reviewers provided demographic information at the first review session. The EDM presented the videos in the viewing window in the predetermined random sequence. They were not able to stop the video segment from playing or rewind and review the video until they entered their first score using the predetermined randomly allocated scale. This was intended to as closely as possible replicate the circumstances of real-time clinical pain assessment using this scale. When the score was entered the clinician reviewer was able to review the video as many times as they liked before entering their final pain score using this scale. The final score was the one that they considered unlikely to change regardless of how many more times they watched the video. The reviewer then applied the remaining scales to the video segment. Reviewers were not able to advance until all items were completed. When all video segments were reviewed the reviewer completed the feasibility and utility questionnaire.

A minimum of 4 weeks later, reviewers repeated pain assessments for the same review set using the scale used first at the original review session.

Statistical Analysis

Statistical analyses was conducted using the statistical software package R: A language and environment for statistical computing (https://www.R-project.org/).[23]

Descriptive statistics were used to report the demographic and clinical information and the responses to the feasibility and utility questionnaire. Distributions of the pain scores were presented in box plots. The percentage of complete pain scores are also described. Where items were not allocated the reason for this has also been summarized to establish the potential limitations to the MBPS. This prevented calculation of a total of MBPS scores for some observations, which were then treated as missing data for remaining analysis. The first and final MBPS scores allocated by the reviewer scale were compared descriptively and using the Wilcoxon signed rank test to identify the percentage of scores that changed after multiple viewings of the video segment.

Reliability was examined using the intraclass correlations (1-way random effects) and Bland–Altman plots were also used to assess agreement. Kappa scores were calculated for the 5 scale items using methods designed for any number of raters, any number of categories, and in the presence of missing values. Convergent validity was

Table 2. Capacity for Reviewer to Score MBPS Items

	Item			
	Face	Movement	Cry	Total score
Valid item score	1,071 (98.5)	1,006 (92.5)	1,085 (99.7)	995 (91.5)
Not applicable	16 (1.5)	82 (7.5)	3	92 (8.5)
Restraint interference	1	78	-	78
Procedural interference	14	1	2	16
Comfort interference	1	2	-	3
Other	-	-	1	1

NOTE. Values are frequency (%) unless otherwise indicated.

tested by comparing MBPS scores with VASobs scores, which was achieved by calculating the Spearman correlation coefficient. The independent t-test was used to compare mean scores for the procedural phases of painful and nonpainful procedures to document the capacity of the scale to detect the difference between known groups.

Using "lme4"[4] in R,[23] the responsiveness of the scale to changes in pain was determined by analysis of the change in scores over the phases of the procedure using linear mixed models to estimate the fixed effects of time (phase of the procedure) and procedure type (painful vs nonpainful). The model also accounted for an interaction between the phase and the procedure type. Children and reviewers were considered as random effects. This method of analysis accounts for correlated data (ie, MBPS scores across procedural phases from the same child), the effect of multiple independent variables, and missing data. Confidence intervals for fixed effects were computed using bootstrap samples as implemented in the "confint" function in R (R package: stats).

To determine the level of distress experienced by infants and children before the painful or intrusive segment of the procedure participants were grouped according to their scores in the baseline and preparation phases (<3 vs ≥3) and the mean difference between their scores in these phases and the procedure were compared using the t-test.

Ninety-five percent confidence intervals and *P* values set for significance at .05 was used to establish statistical significance where appropriate.

Ethics

The study protocol was approved by the Royal Children's Hospital Human Research Ethics Committee (reference number: 35220B). Written consent was provided by the parents of the children and reviewers before participation.

Results

Twenty-six ED clinicians were recruited and made 1,088 observations of 100 infants and children for this study at the first review and 358 at the second review (Fig 1). There were no missing observations as reviewers could not advance unless each data field was completed. The clinicians included 19 nurses of varying levels of experience (range = 1-20, mean = 10.1), 12 with postgraduate specialty training in pediatrics and/or emergency care and 7 doctors of whom 3 were considered senior (defined as having completed their specialty training). The mean age of the infants and children was 22.5 (±10.3) months, 58% (n = 58) were boys and 38% were diagnosed with respiratory disease, 29% with dehydration and gastroenteritis, whereas the remaining 36% spanned a range of diagnoses.

The mean, median, and distribution of MBPS scores across the phases of each procedure from review session 1 are presented in box plots (Fig 2). Mean scores during the procedure phase were highest for NGT insertion (8.78 ± 1.09) and lowest for saturation measurement (1.99 ± .86). Scores were not normally distributed.

Psychometric Evaluation

Feasibility and Clinical Utility

Reviewers were unable to score 101 (3.1%) MBPS items resulting in 92 (8.5%) incomplete scores (Table 2). Restraint was the most common source of interference to scoring an item. The correlation between the first score allocated after 1 uninterrupted view of the video segment and the final score allocated was near perfect (r = .97). Despite this, one-quarter (28.2%) of scores changed by at least a score of 1 (Table 3). The range of the differences between scores was –3 to +4

Table 3. Correlation Between First MBPS Score and Final MBPS Score

Scale	First Score*	Final Score*	Scores Changed by 1†	Correlation Coefficient‡	P§
MBPS	4.7 (3.1) 4 [6]	4.7 (3.1) 4 [6]	28.2%	.97	.96

*Values are mean (SD) and median [interquartile range], except where otherwise noted.
†Percentage.
‡Spearman correlation coefficient.
§Wilcoxon signed rank test with continuity correction.

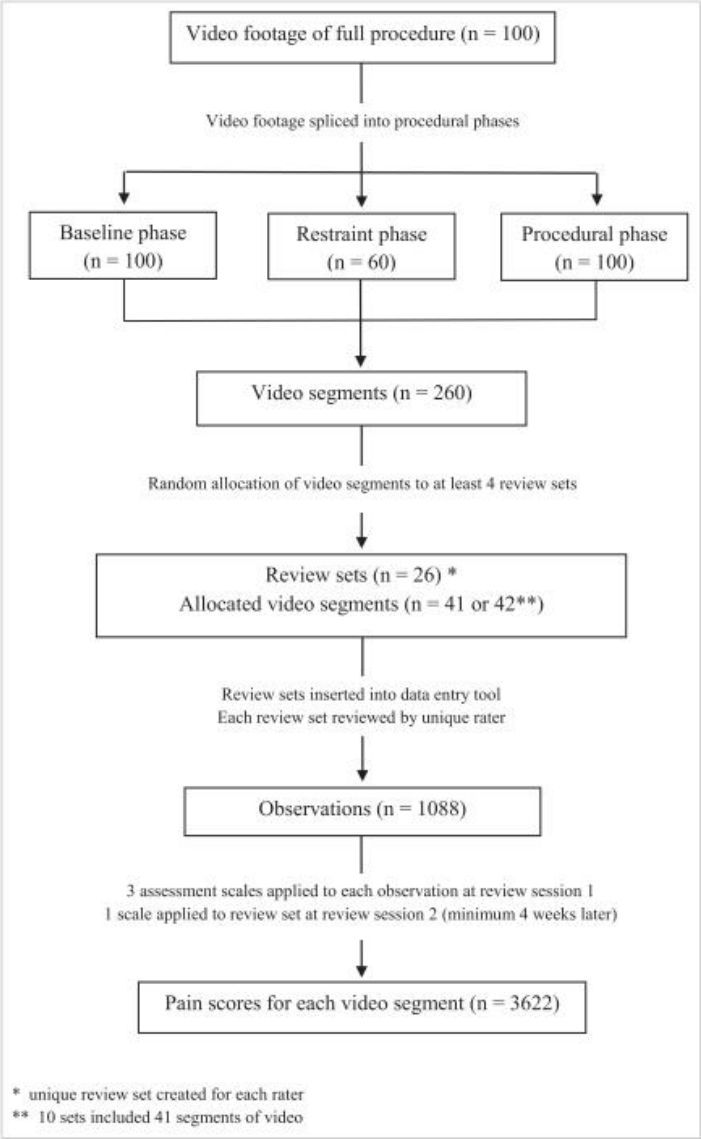


Figure 1. Overview from creation of video segments to final data set.

and the average difference between scores was .02 (SD = .71).

Using a Likert scale to indicate the extent to which they agreed (5) or disagreed (1) with statements, reviewers agreed (score of "4" or "5" on the Likert scale) that the MBPS was: "easy to understand" (53.8%) and "quick" (53.8%) and "easy" (61.5%) to apply. However, their rating for the extent to which they agree that the MBPS is "clinically useful" was lower (26.9%) and lower still when they rated the extent to which they agreed that the scale "reflects the extent of procedural pain" (11.5%), "discriminates children with pain from those without" (15.5%), is "readily understood and support decisions about pain management" (19.2%) and "reflects procedural pain-specific features" (15.4%).

Reliability

The overall intraclass correlation coefficient (ICC) for inter-rater reliability was .87, ranging from .69 (nonpainful, baseline) to .89 (painful, preparation), depending on the procedure phase and type. The ICCs for the remaining phases were as follows: nonpainful, procedure .87; painful, baseline .85; and painful, procedure .82. Strong inter-rater reliability was also demonstrated by linear mixed modeling, where variability attributed to the effect of the reviewer was close to 0 (variance = .016 ± .125). The ICC for intrarater reliability was similarly high (ICC = .89) Bland and Altman analysis showed a mean difference between scores of –.27 ± 1.49 and 95% limits of agreement of –3.27 to 2.71.

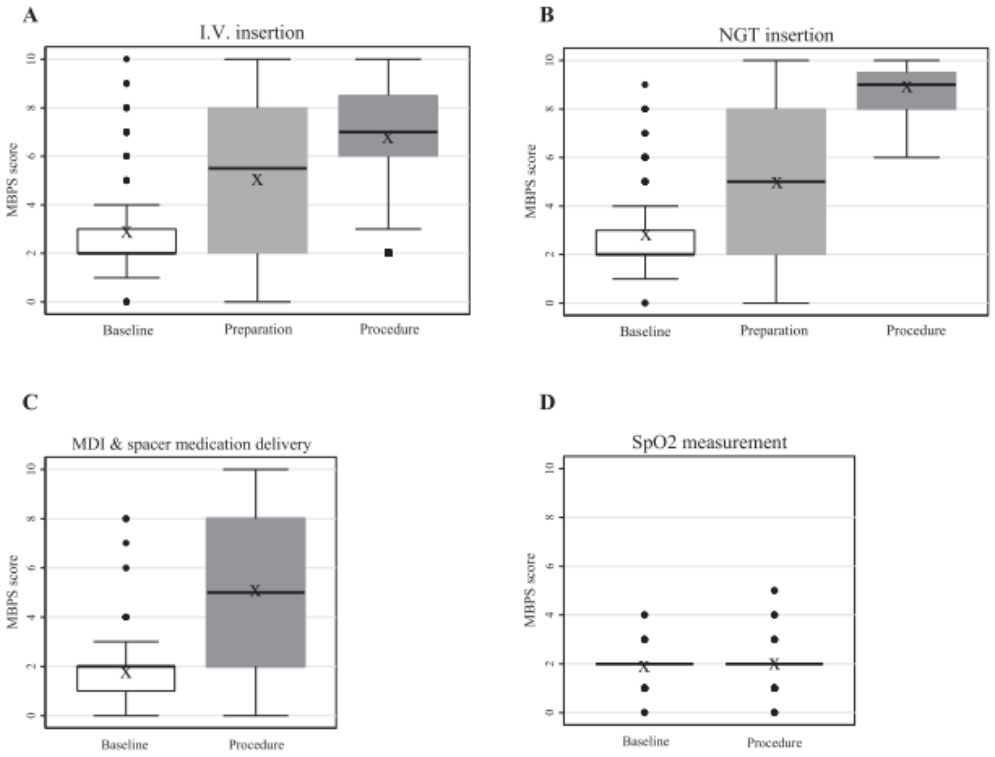


Figure 2. Box plots for MBPS scores for each phase of each procedure: (**A**) i.v. insertion, (**B**) NGT insertion, (**C**) inhaled medication administration, and (**D**) oxygen saturation measurement. Mean values are highlighted by the crosses (X).

Construct Validity: Known Groups, Responsiveness, and Discrimination

A significant difference between the mean scores for painful (7.5 ± 2.3) and nonpainful procedures (3.4 ± 2.7) show the capacity of the scale to differentiate between known groups (t_{994} = −12.70, P = .000). Furthermore, the results of an independent t-test comparing the difference in means from phase 1 to 3 for infants and children experiencing a painful procedure (4.73 ± 2.63) compared with those experiencing a nonpainful procedure (1.50 ± 2.73) show a significant difference between these cohorts (t_{95} = −5.86, P = .000).

Mean MBPS scores plotted across phases of the procedure for nonpainful and painful procedures (Fig 3) illustrate scale responsiveness. This was statistically tested using linear mixed modeling to determine the effect of procedure and phase on MBPS scores, which were considered fixed effects. Reviewer and child were added to the model as random effects. From the model we can see that scores at baseline were not 0 for either group (intercept = 1.92) and that scores increased across phases (baseline to procedure) for painful as well as nonpainful procedures (Table 4). This increase was more pronounced for painful compared with nonpain procedures (average increase in scores of 4.6 vs 1.5, respectively).

Analysis also considered the capacity of the scale to differentiate between related constructs (pain and nonpain related distress). This was achieved by running independent t-tests to determine if there were differences in the mean change in scores across phases of the procedure for infants and children who had higher scores before the procedure than those with lower scores before the procedure. There was a significant difference in the mean change in MBPS score from baseline to the procedure for infants and children with a baseline score of <3 (mean difference = 3.72 ± 3.13) compared with those

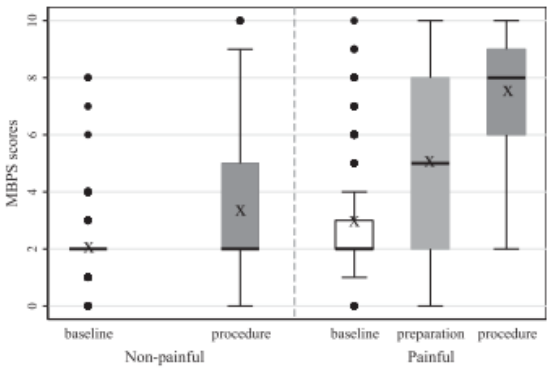


Figure 3. Box plots representing change of values over procedure phases (baseline, preparation, and procedure) in the 2 procedure cohorts (painful and nonpainful procedures). Mean values are highlighted by the crosses (X).

Table 4. The Variances and Estimates for Random and Fixed Effects for the Model Used to Demonstrate the Responsiveness of MBPS Scores to the Procedural Phase and Procedure Type (Painful vs Nonpainful)

Random Effects	Variance	SD	
Patient (intercept)	2.287	1.512	
Reviewer (intercept)	.016	.125	
Residual	2.844	1.687	
Fixed effects	**Estimate**	**Standard Error**	**t**
Intercept	1.917	.271	7.068
Painful procedure	1.044	.351	2.971
Preparation phase	2.056	.160	12.862
Procedure phase	1.482	.182	8.135
Procedure phase: painful procedure	3.140	.249	12.631

NOTE. Model = (1|Reviewer) + (1|Child) + procedure_type + procedure_phase + procedure_type × procedure_phase.

with a score of ≥3 (mean difference = 1.74 ± 2.38, t_{95} = 2.40, P = .019). Comparisons between the mean change in scores were also made for infants and children undergoing a painful procedure. Similarly, there was significant difference in the mean change in MBPS scores from baseline to preparation in participants with a baseline score of <3 (5.71 ± 1.94) compared with children with a score of ≥3 (1.72 ± 2.16, t_{55} = 6.49, P = .0000).

Convergent Validity

The MBPS were strongly correlated with VASobs scores for pain (r = .74) and VASobs scores for distress (r = .87).

Discussion

The results of this study support the conclusion that the MBPS is sensitive to the pain experienced by infants and children during venous access and NGT insertion. The mean scores during these painful and commonly performed procedures were high (7.5 ± 2.3). In contrast mean scores were significantly lower during the procedure phase for procedures presumed nonpainful (3.4 ± 2.7, t_{994} = –12.70, P = .000), confirming the capacity of the scale to differentiate between known groups. The responsiveness of the scale to painful stimuli was also demonstrated by a significant rise in scores across the phases of a painful procedure (mean difference = 4.7 ± 2.6) compared with a nonpainful procedure (1.5 ± 2.7, t_{95} = –5.86, P = .000), and linear mixed modeling confirmed these results independent of the effects of the reviewer and the child. And finally, strong correlation with the VASobs pain scores (r = .74) indirectly supports the conclusion that MBPS is sensitive to procedural pain.

Contemporary understanding of reliability and validity attribute these properties to the scores generated by the scale and not to the scale and as such much is made of the need to test application of a scale to each population and circumstances for which the scale is intended. The MBPS was designed to assess procedural pain in infants,[26] however, our understanding and claims about validity are to date limited to its performance when used to assess immunization-related pain.[10] These results help fill this gap and provide evidence that confirms the MBPS as an option to assess pain associated with procedures other than immunization.

In contrast to population and circumstance, the potential effect of the reviewer on the psychometric properties receives relatively scant attention. Until now, our understanding of MBPS reliability was on the basis of data from studies using no more than 5 nonclinical trained observers to apply the scale.[17,25,26] Our study shows that the MBPS can be reliably applied by clinicians to assess procedural pain and distress. ICCs were very high for inter- (ICC = .87) as well as intrarater reliability (ICC = .89) and these results were supported by the linear mixed modeling results, which confirmed that raters contributed very little to the variability in scores. Additionally, the Bland–Altman plots and statistics also confirmed a high level of intrarater reliability.

The aim of the study did not include an assessment of the content validity of the scale. However, the results point to the need to examine the items and descriptors of the scale more carefully. Reviewers rated the extent to which the scale "reflects procedural pain" and "reflects procedural pain-specific features" poorly. Furthermore, despite three-quarters of the observations involving infants and children during nonpainful phases of the procedure or nonpainful procedures, reviewers recorded "0" for only 33 observations (3.6%). The descriptors for the "face" and "cry" items may provide a simple explanation for these results. A score of "0" for these items requires the infant to demonstrate "definite positive expression (smiling)" and "laughing or giggling," respectively. Infants and children in unfamiliar surroundings are arguably unlikely to demonstrate these behaviors regardless of whether or not they are experiencing pain. Adjustment to the descriptors for these items may also address the concerns of reviewers about the extent to which this scale is suited to procedural pain assessment. It is also interesting to note that the descriptors for a maximum score on several items from CHEOPS (activity of torso, touching, and response of lower limbs) include the need for restraint. This descriptor seems particularly relevant for procedural pain and yet these items were merged to create the summary MBPS item "movement" and the references to restraint use were removed from the descriptors for this item.

Furthermore, this study highlights practical concerns regarding the MBPS when used to score procedural pain. Restraint and the procedure interfered with reviewers' capacity to score scale items resulting in incomplete scores in these instances. In addition, although first and final scores allocated by reviewers correlated strongly, almost one-quarter of these scores differed from each other. Although this may in part be attributed to the effect of video review, with numbers of incomplete scores reaching 8.5% and 28% of scores changing, the practical limitations of the MBPS for procedural pain assessment cannot be ignored.

It has been widely suggested that observational pain scales should be considered measures of distress and not

specifically of pain because the behaviors demonstrated by children experiencing other distress-causing emotions such as fear mimic those demonstrated by children in pain. This study provides evidence for this long-held view.[5] Although there were marked differences in the procedure scores for painful versus nonpainful procedures and responsiveness was greater in children experiencing a painful procedure, very few observations, including those made at baseline, resulted in a score of "0." As previously noted, this may relate to the appropriateness of the descriptors but may also reflect limited capacity of the MBPS to differentiate pain from nonpain-related distress. That responsiveness for nonpainful procedures was demonstrated at all reinforces this potential limitation of the scale. Finally, MBPS scores in children experiencing higher baseline scores did not rise during the procedural phase (mean difference = 1.74) as much as those in children with lower scores at baseline (mean difference = 3.72, $P = .019$). This means that a more limited response to painful stimuli is seen in children already exhibiting distress behaviors. This may signal a complex relationship between nonpain- and pain-related distress, which results in a blunted behavioral response to pain in distressed children. However, it may also provide additional evidence that the MBPS struggles to differentiate pain- from nonpain-related distress behaviors. This is not the first time that the effect of baseline distress on procedural scores has been seen. Similar observations were made by Ahola Kohut and Pillai Riddell in their study addressing the specificity of the Neonatal Facial Coding Scale for pain.[1] These authors concluded that their results highlighted poor specificity and conclude that the Neonatal Facial Coding Scale is sensitive but not specific for pain in neonates. Finally, the clinicians in this study also raised their concerns about the capacity of the scale to discriminate children in pain from those without pain.

Strengths and Limitations

Multiple methods for validation were used to overcome the absence of a "gold standard" to which the MBPS could be compared. Furthermore, establishing the validity of one measure on the basis of the correlation with another can be considered circular logic. Unique reviewers were used to assess each phase of a single procedure because reviewers could not be blinded to the circumstances of the phases of the procedure. It was intended that this would prevent reviewers from creating logical patterns in the scores across phases of a procedure. The use of video recordings was necessary to make multiple assessments by a range of reviewers possible. However, application of the scale to video recordings may alter the experience of the rater compared with real-time assessment affecting the results, particularly those that relate to feasibility. The aim of recruiting a larger sample of clinicians was on the basis of our unwillingness to accept that results from studies addressing the psychometrics of scales using small samples of nonclinical research assistants trained to improve reliability can be used to draw conclusions about the performance of the scale when applied by a range of clinicians. In fact there is some evidence that clinicians apply clinical judgement when applying assessment scales.[13-15] Finally, this study was conducted in a single center, which may limit its generalizability.

Conclusions

Extensive review of the literature confirms that there are few alternatives to the MBPS for procedural pain assessment in infants and young children. The MBPS is one of only several scales purposefully developed to assess procedural pain or that has been psychometrically tested for this purpose. Alternative scales, such as the FLACC, although tested for this purpose are unsupported by sufficient data to accept them as better suited for procedural pain assessment than the MBPS.[9,11,28] Furthermore, unlike many of the scales recommended for procedural pain assessment, such as CHEOPS, the MBPS uses a 0 to 10 metric, which is the commonly accepted standard for pain assessment scales.[27] The evidence for reliability and sensitivity of the MBPS from the current study is compelling. However, it is also clear from these results that there are still some important areas of concern and some practical limitations to using the MBPS to assess procedural pain, preventing us from unreservedly supporting a recommendation for MBPS as an appropriate choice of scale for assessment of procedural pain.

Work to better understand the effect of the circumstances of the procedure and restraint on scores and application of the scale is needed. Furthermore, the differences in behavioral responses across this age range (6–42 months), sex, and other patient-related factors and in reviewer experience and education and their possible effect on application of the scale and the scores needs exploration. Studies addressing the capacity to differentiate pain- and nonpain-related distress are also crucial to expand our understanding of procedural distress, pain assessment, and management.

The results of this study also point to a need for closer scrutiny of the scale items and their descriptors before we consider assessing performance of the scale in additional studies. Specifically, clinicians may have greater confidence if descriptors for a score of 0 did not require the child to show a positive expression and laughter or giggling, which are at odds with what might be reasonably expected by clinicians from a child being prepared for a procedure in a health care environment. Steps to improve the content validity of the scale to address this and clinicians' concerns that the scale does not differentiate between infants and children with and without pain or guide treatment decisions are needed. Identification empirically of the behaviors of anxious infants and children in a health care environment who are preparing for a procedure may help to inform revisions in the descriptors for the items of this scale. Engaging end users—clinicians and researchers—in the redesign phase may also improve their confidence in the performance of the scale.

Although, the MBPS may be suitable, in light of the limitations of the scale clinicians and researchers should

use MBPS to assess procedural pain clinically and to measure study outcomes with caution and are probably well advised to regard the scale as a measure of procedure-related distress (including pain) and not specifically pain alone.

Acknowledgments

Dianne Crellin is a trainee member of Pain in Child Health, a research training initiative of the Canadian Institutes of Health Research.

References

1. Ahola Kohut S, Pillai Riddell R: Does the Neonatal Facial Coding System differentiate between infants experiencing pain-related and non-pain-related distress? J Pain 10:214-220, 2009

2. Aitken RC: Measurement of feelings using visual analogue scales. Proc R Soc Med 62:989-993, 1969

3. Andersen RD, Langius-Eklof A, Nakstad B, Bernklev T, Jylli L: The measurement properties of pediatric observational pain scales: A systematic review of reviews. Int J Nurs Stud 73:93-101, 2017

4. Bates D, Machler M, Bolker B, Walker S: Fitting linear mixed-effects models using {lme4}. J Stat Softw 67:1-48, 2015

5. Blount RL, Loiselle KA: Behavioural assessment of pediatric pain. Pain Res Manag 14:47-52, 2009

6. Chang J, Versloot J, Fashler SR, McCrystal KN, Craig KD: Pain assessment in children: Validity of facial expression items in observational pain scales. Clin J Pain 31:189-197, 2015

7. Chorney MacLaren J, McMurtry MC: Behavioural measures of pain, in McGrath PJ, Stevens BJ, Walker SM, Zempsky WT (eds): Oxford Textbook of Paediatric Pain. Oxford, Oxford University Press, 2014, pp 379-390

8. Craig KD, Whitfield MF, Grunau RV, Linton J, Hadjistavropoulos HD: Pain in the preterm neonate: Behavioural and physiological indices. Pain 52:287-299, 1993

9. Crellin D, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A: Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. Paediatr Anaesth 17:720-733, 2007

10. Crellin DJ, Babl FE, Santamaria N, Harrison D: The MBPS: A systematic review to determine its role in assessing pain in infants and children. Presented at the International Symposium on Pediatric Pain. Kuala Lumpur, International Association for the Study of Pain, July 6-9, 2017

11. Crellin DJ, Harrison D, Santamaria N, Babl FE: Systematic review of the FLACC scale for assessing pain in infants and children: Is it reliable, valid, and feasible for use? Pain 156:2132-2151, 2015

12. de Jong A, Baartmans M, Bremer M, van Komen R, Middelkoop E, Tuinebreijer W, van Loey N: Reliability, validity and clinical utility of three types of pain behavioural observation scales for young children with burns aged 0–5 years. Pain 150:561-567, 2010

13. De Ruddere L, Goubert L, Stevens M, de C Williams AC, Crombez G: Discounting pain in the absence of medical evidence is explained by negative evaluation of the patient. Pain 154:669-676, 2013

14. De Ruddere L, Goubert L, Stevens MA, Deveugele M, Craig KD, Crombez G: Health care professionals' reactions to patient pain: Impact of knowledge about medical evidence and psychosocial influences. J Pain 15:262-270, 2014

15. Gomez RJ, Barrowman N, Elia S, Manias E, Royle J, Harrison D: Establishing intra- and inter-rater agreement of the face, legs, activity, cry, consolability scale for evaluating pain in toddlers during immunization. Pain Res Manag 18:e124-e128, 2013

16. Harris M, Warren JJ: Patient outcomes: Assessment issues for the CNS. Clin Nurse Spec 9:82-86, 1995

17. McClellan CB, Cohen LL, Joseph KE: Infant distress during immunization: A multimethod assessment. J Clin Psychol Med Settings 10:231-238, 2003

18. McGrath PJ, Johnson G, Goodman JT, Schillinger J, Dunn J, Chapman J: CHEOPS: A behavioural scale for rating postoperative pain in children, in Fields HL, Dubner R, Cervero F (eds): Advances in Pain Research and Therapy. New York, Raven Press, 1985, pp 395-402

19. Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S: The FLACC: A behavioral scale for scoring postoperative pain in young children. Pediatr Nurs 23:293-297, 1997

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

21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, 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 19: 539-549, 2010

22. Prkachin KM: Assessing pain by facial expression: Facial expression as nexus. Pain Res Manag 14:53-58, 2009

23. R Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing, 2016

24. StataCorp: Stata Statistical Software: Release 14. College Station, TX, Stata Corp LP, 2015

25. Taddio A, Hogan ME, Moyer P, Girgis A, Gerges S, Wang L, Ipp M: Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. Vaccine 29:1390-1394, 2011

26. Taddio A, Nulman I, Koren BS, Stevens B, Koren G: A revised measure of acute pain in infants. J Pain Symptom Manage 10:456-463, 1995

27. von Baeyer C, Hicks C: Support for a common metric for pediatric pain intensity scales. Pain Res Manag 5:157-160, 2000

28. von Baeyer CL, Spagrud LJ: Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. Pain 127:150, 2007

11.1 Additional data analysis

Sensitivity, specificity and area under the curve (AUC) at various MBPS cut-off scores were calculated using receiver operating characteristics analysis. The results of this analysis are presented in Table 11-1 and demonstrated that the best levels of sensitivity (91.5%), specificity (77.5%) and AUC 0.85 are achieved at an MBPS cut-off of '4'.

Table 11-1 Sensitivity, specificity and area under the curve values (AUC) calculated for different cut-off MBPS scores to differentiate procedure type (painful and non-painful).

MBPS score cut-off	Sensitivity (%)	Specificity (%)	Area under the curve (AUC)
Score > 0	na*	na*	na*
Score > 1	100	2.5	0.51
Score > 2	96.6	57.5	0.77
Score > 3	91.5	70.0	0.81
Score > 4	91.5	77.5	0.85
Score > 5	91.5	77.5	0.85
Score > 6	84.7	80.0	0.82
Score > 7	71.2	82.5	0.77

na – not applicable. Too few '0' observations to perform analysis based on a cut-off of '0'

11.1.1 Implications

This analysis highlights the limitations of the MBPS. Although, sensitivity was very high, and specificity was good, this was achieved at a cut-off score of 4. A scale where scores differentiate painful from non-painful at a score of 4 makes interpretation very difficult as a score of 4 has also been defined as the lower end of moderate pain (470). This result strengthens our concern about using the MBPS to assess procedural pain and the recommendation to regard scores as an indicator of distress and not exclusively pain. The full implications of this analysis is given more attention in the discussion (Chapter 13).

CHAPTER 12.

Comparing the psychometric properties of the FLACC scale, the MBPS and the observer applied Visual Analogue Scale used to assess procedural pain.

This chapter reports a comparison of the psychometric properties of the Face, Legs, Activity, cry and Consolability (FLACC) scale, the Modified behavioural Pain Scale (MBPS) and the Visual Analogue Scale applied by an observer (VASobs) used to assess procedural pain in infants and children. This work is currently unpublished and is presented here formatted for the thesis.

Abstract

Introduction: The Face Legs, Activity, Cry and Consolability scale (FLACC) and the Modified Behavioural Pain Scale (MBPS) are observational pain scales considered valid for procedural pain assessment. The Visual Analogue Scale applied by observers is a widely used alternative. However, it is not clear whether one performs better for this purpose. The aim of this study was to compare the psychometric and practical properties (reliability, validity, feasibility and utility) to quantify procedural pain in infants and young children.

Methods: A convenience sample of twenty-six clinicians used the FLACC scale, the MBPS and the VASobs to segments of video from 100 children aged six to 42 months undergoing a procedure.

Results: The FLACC scale resulted in more incomplete scores ($p < 0.000$) than the other scales. Reviewers liked the VASobs (pain) most, considered it quickest and easiest to apply but all scales were considered of limited use for procedural pain assessment. Observers changed their MBPS scores more often than they changed FLACC or VASobs scores, but the degree of change was greater for FLACC scores ($p = 0.033$). Inter-rater reliability was poorest for VASobs pain scores (ICC – 0.55). VASobs pain scores were lower than FLACC and MBS scores during the procedure but MBPS scores were higher during non-painful phases ($p < 0.001$). The FLACC scale provided the best sensitivity (94.9%) and specificity (72.5%) for the lowest cut-off score (pain score 2). Correlations between FLACC and VASobs (distress) were strongest ($r = 0.89$).

Discussion: This study supported the reliability and sensitivity of the FLACC and MBPS. There were practical concerns for application of the FLACC scale and the MBPS and doubt about the capacity of both scales to differentiate between pain- and non-pain related distress exists. The VASobs, although practical, was less reliable than either the FLACC scale or the MBPS. The results of this study demonstrated that the FLACC scale may be best suited for procedural pain assessment.

12.1 Introduction

Pain assessment informs decisions about treatment and is a frequent outcome measure in trials evaluating methods designed to reduced pain. The generally accepted standard for pain assessment is self-report, however for those unable to self-report the most practiced and recognised alternative is the observational pain scale. Over 40 tools have been identified in the literature (212). Despite the frequency with which infants and children experience painful medical procedures, the appropriate choice of assessment strategy to assess procedural pain in infants and young children for clinical and research purposes is far from obvious.

There is widespread recognition that procedural pain, particularly in infants and young children, may manifest differently to other types of pain, e.g. postoperative and chronic pain. However, few scales are recommended for this purpose and fewer still have been purposefully designed for procedural pain assessment. Although many scales are used in circumstances other than for which they were originally designed most are not suitable for procedural pain assessment. They are often not supported by sufficient psychometric data and/or the design makes them unsuitable e.g. non-standard metrics to accept them as fit for purpose. Recent systematic reviews of the properties of three observational pain scales, the Face, Legs, Activity, Cry Consolability (FLACC) scale (330), the Modified Behavioural Pain Scale (MBPS) (471) and the Visual Analogue Scale applied by an observer (VASobs) [Section 3, Chapter 2], prompted work to assess the performance of these scales when used to assess procedural pain in infants and young children.

The MBPS was developed at a time when few scales were available and were not considered sufficiently able to capture the variability in responses in young infants to procedural pain (288). The validity of the MBPS for procedural pain assessment until recently is largely supported by the results of studies addressing the scale's performance when used to assess the pain associated with immunisation in infants and young children (471). In contrast, the FLACC scale was originally designed to assess postoperative pain in infants and children aged two months to seven years (28) and is now one of the most well-known and widely recommended pain assessment scales (15, 30, 475, 476). Despite this, available data, summarised in a 2015 systematic review, was insufficient data to recommend FLACC for procedural pain assessment (330). Finally, the VASobs, a unidimensional scale based on the VAS used for self-report of pain, has been extensively used to assess procedural pain in infants and young children, particularly for research purposes (Chapter 3). This is despite two systematic reviews concluding that there was inadequate data to support the psychometric properties of this scale for this purpose (290)(Chapter 5).

A recent psychometric evaluation study assessed the performance of these three scales used to assess pain associated with procedures frequently experienced by infants and young children aged six to 42 months of age during an emergency department presentation. The results for each scale have been previously reported/published in detail in previous chapters and associated publications (464, 469). In summary, the FLACC scale and the MBPS were reliably applied by clinicians. However, results reaffirm long held concerns about the reliability of the VASobs when used to assess pain (290). The sensitivity of all scales to procedural pain was demonstrated but none could be shown to be highly specific for procedural pain. There were circumstantial factors, such as restraint and the steps of the procedure that interfered with application of the FLACC and the MBPS. In contrast, the VASobs was confirmed as easily applied and rarely affected by circumstantial factors that interfered with application of the other scales. Data also gave rise to concerns about the design of the MBPS. Infants and children in this study infrequently scored '0' even for procedures and phases of procedures not considered painful.

The aim of this study was to compare the psychometric and practical properties of the FLACC scale, the MBPS and the VASobs used to assess procedural pain in infants and young children to provide decisive recommendations for clinicians and researchers regarding scale selection for clinicians and researchers.

12.2 Methods

The methods for this study are described in the published protocol which is presented in Chapter 8 (468).

12.3 Results

The mean age of the children was 22.5 (±10.3) months and 58% were boys. Thirty-eight percent were diagnosed with a respiratory disease, 29% with dehydration and gastroenteritis, while the remaining 36% spanned a range of diagnoses. Twenty-six ED clinicians participated,19 nurses and seven physicians. The nurses reported varying levels of experience ranging from less than one year to twenty years (mean = 10.1 years) and 12 (63%) had postgraduate specialty training. Three of the seven physicians had completed specialty training.

Clinicians allocated scores that ranged across procedures and phases from zero to 9.5 and mean and median scores were highest for nasogastric tube (NGT) insertion (FLACC 9.5 ± 0.8, [10 IQR

9 - 10], MBPS 8.8 ±1.1 IQR 5.3 - 8] 8.5 IQR 7.3 - 9.5]) and lowest for oxygen saturation measurement (SpO2) (FLACC 0.5 ± 0.9, [0 IQR 0 – 1], MBPS 2.0 ±0.9, [2 IQR 2 – 2],). VASobs pain scores were lowest for all phases of all procedures except for the procedural phase of SpO2 measurement. MBPS scores were highest for all baseline and preparation phases and the procedural phase of intravenous cannula insertion. Mean and median scores are reported in Table 12.1.

Table 12-1 Pain and distress scores for each scale for each phase of the four procedures.

Procedure	Scale			
Phase	FLACC	MBPS	VAS pain	VAS distress
IV cannula insertion				
Baseline	1.7 (2.7) 0 [0 - 2]	2.9 (2.2) 2 [2 - 3]	0.3 (0.9) 0 [0 - 0]	1.3 (2.2) 0 [0 - 2]
Preparation	4.4 (3.7) 5 [1 - 8]	5.0 (2.8) 5.5 [2 - 8]	1.6 (2.4) 0 [0 - 2.6]	3.9 (3.4) 3.6 [0.7 - 7]
Procedural	6.4 (3.1) 7 [5 - 9]	6.6 (2.4) 7 [6 - 8.5]	4.4 (2.8) 4.7 [2.0 - 6.4]	5.6 (3.2) 6 [3 - 8]
NGT insertion				
Baseline	1.6 (2.3) 1 [0 - 2]	2.9 (1.8) 2 [2 - 3]	0.5 (1.3) 0 [0 - 0]	0.9 (1.6) 0 [0 - 2]
Preparation	3.8 (3.6) 3 [1 - 7]	4.9 (2.7) 5 [2 - 8]	1.1 (2.1) 0 [0 - 1]	3.8 (3.3) 3.4 [0.5 - 7]
Procedural	9.5 (0.8) 10 [9 - 10]	8.8 (1.1) 9 [8 - 9.5]	6.4 (2.0) 6.7 [5.3 - 8]	8.1 (1.8) 8.5 [7.3 - 9.5]
MDI administration				
Baseline	0.5 (1.5) 0 [0 - 0]	1.8 (1.5) 2 [1 - 2]	0.2 (0.9) 0 [0 - 0]	0.6 (1.6) 0 [0 - 0]
Procedural	4.2 (4.1) 2 [0 - 9]	5.1 (3.2) 5 [2 - 8]	1.1 (1.9) 0 [0 - 1.1]	3.7 (3.5) 2 [0 - 7.3]
SpO2 measurement				
Baseline	0.3 (0.6) 0 [0 - 0]	1.9 (0.7) 2 [2 - 2]	0.0 (0.2) 0 [0 - 0]	0.2 (0.7) 0 [0 - 0]
Procedural	0.5 (0.9) 0 [0 - 1]	2.0 (0.9) 2 [2 - 2]	0.7 (3.2) 0 [0 - 0]	2.7 (5.2) 0 [0 - 0.3]

Values are mean (standard deviation) and medians [interquartile ranges]

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, IV – intravenous, MBPS – Modified Behavioral Pain Scale, MDI – metered dose inhaler, NGT – nasogastric tube, VASobs – Visual Analogue Scale observation

Data were pooled by procedure type and analysed using Wilcoxon signed rank sum, a non-parametric test to compare scale scores for painful and non-painful procedures. The results demonstrated a significant difference between scores for painful and non-painful procedures at baseline and during the procedure. The aggregated means and medians and the results of the comparison are reported in Table 12.2.

Table 12-2 Comparison of FLACC, MBPS scores, VASobs pain and VASobs distress for baseline and procedural phases of painful and non-painful procedures.

Procedure	Scores				Comparisons					
	FLACC	MBPS	VASobs (pain)	VASobs (distress)	FLACC: MBPS	FLACC: VASobs (pain)	FLACC: VASobs (distress)	MBPS: VASobs (pain)	MBPS: VASobs (distress)	VASobs (pain): VASobs (distress)
Painful procedures										
Baseline	0 [0 – 2] 1.6 (2.5)	2 [2 – 3] 2.9 (2.0)	0 [0 – 0] 0.4 (1.1)	0 [0 – 2] 1.4 (2.2)	-11.89*	9.26*	4.17*	13.42*	12.26*	-8.74*
Procedure	9 [5.3 – 10] 7.5 (3.0)	8 [6 – 8] 7.5 (2.3)	6 [3 – 8] 5.4 (2.7)	8 [6 – 9] 6.9 (2.8)	0.99	9.44*	6.83*	10.89*	5.91*	-10.66*
Non-painful procedures										
Baseline	0 [0] 0.4 (1.2)	2 [1.75 – 2] 1.9 (1.2)	0 [0 – 0 0.1 (0.6)	0 [0 – 0] 0.4 (1.2)	-11.51*	5.16*	1.30	12.04*	11.40*	-5.14*
Procedure	0 [0 – 2] 2.1 (3.3)	2 [2 – 5] 3.4 (2.7)	0 [0 – 0.9] 0.6 (1.4)	0.1 [0 – 2.6] 2.0 (3.0)	-8.40*	8.47*	5.84*	11.53*	10.65*	-8.64*

Values are median [interquartile ranges] and mean (standard deviation)

* p value < 0.05 based on Wilcoxon signed rank sum

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, MBPS – Modified Behavioral Pain Scale, VASobs – Visual Analogue Scale observation

12.3.1 Psychometric evaluation

12.3.1.1 Feasibility and clinical utility

Application of the FLACC scale resulted in a larger number of incomplete scores than for the MBPS (14.6% vs 8.5%, $\chi^2 = 473.7$, $p < 0.000$). VASobs pain and VASobs distress scores were relatively infrequently omitted by reviewers (0.9% and 0.2% of scoring occasions, respectively) so they were not considered further for this comparison. The most frequent impediment to allocation of items for the FLACC scale and MBPS was restraint. Uniquely, the absence of attempts to console the child prevented the allocation of a score for the FLACC item 'consolability' item on 30 occasions.

Using a Likert scale reviewers indicated the extent to which they agreed (5) or disagreed (1) with statements (Table 12.3). Reviewers rated the utility of the scale similarly for a number of items: clinically useful, able to discriminate children with and without pain, reflective of procedural pain specific features. Comparison between FLACC and MBPS ratings using Wilcoxon signed rank sum revealed no statistically significant differences in ratings between the FLACC scale and MBPS. However, larger numbers of reviewers agreed that the VASobs was 'quick' and 'easy' to apply when compared with their response for the FLACC scale ($z = 4.15$, $p < 0.000$ and $z = 2.081$, $p = 0.037$, respectively) and the MBPS ($z = 3.023$, $p = 0.003$ and $z = 2.043$, $p = 0.041$, respectively). When asked to rank the scales in order of preference, reviewers liked the VASobs the most (n = 14) and the MBPS the least (n = 11) (Table 12.4).

Table 12-3 Clinical Utility Questionnaire responses (responding 'agree' or 'strongly agree') and comparison across scales (FLACC scale, MBPS and VASobs (pain)).

Utility statement	Frequency n (%)			Comparison z score		
	FLACC	MBPS	VASobs	FLACC: MBPS	FLACC: VASobs	MBPS: VASobs
Provides information that is clinically useful	7 (26.9)	7 (26.9)	7 (26.9)	0.15	-0.34	0.07
Is quick to apply	9 (34.6)	14 (53.8)	23 (88.5)	-1.48	4.15*	3.02*
Is easy to apply	12 (46.1)	16 (61.5)	22 (84.6)	-0.99	2.08*	2.04*
It is clear & easy to understand	13 (50.0)	14 (53.8)	20 (76.9)	0.57	1.74	1.54
Reflects the extent of procedural pain	7 (26.9)	3 (11.5)	4 (15.4)	0.872	-0.63	0.04
Discriminates children with pain from children without pain	5 (19.2)	4 (15.4)	5 (19.2)	0.90	0.92	1.70
Score is readily understood & supports decisions about pain management	2 (7.7)	4 (15.4)	7 (26.9)	0.68	0.67	0.73
Reflects procedural pain-specific features	6 (23.1)	4 (15.4)	4 (15.4)	0.68	0.52	1.05

Responses on a 5-point Likert scale: strongly disagree (1), disagree (2), neutral (3), agree (4), strongly agree (5).

* p value < 0.05 based on Wilcoxon signed rank sum

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, MBPS – Modified Behavioral Pain Scale, VASobs – Visual Analogue Scale observation

Table 12-4 Reviewer rankings of their preference for the scales (n = 26).

Ranking	Scale		
	MBPS	FLACC	VAS pain
1st	5	7	14
2nd	10	10	4
3rd	11	9	8

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, MBPS – Modified Behavioral Pain Scale, VASobs – Visual Analogue Scale observation

As another measure of clinical utility, the first scores allocated by reviewers following one uninterrupted view of the video segment were compared with their final scores (Table 12.5). Reviewers' MBPS scores changed more often than FLACC and VASobs pain scores (28.2% vs 23.0% vs 8.8%, respectively). Wilcoxon signed rank sum test results show that there was a statistically significant difference between the first and final median FLACC scores (0 vs 2, p = 0.033) but not for the first and final scores for MBPS and VASobs scores. Correlations between the first and final scores were similarly high for all scales and all coefficients exceeded 0.90.

Table 12-5 Comparison between first score and final score.

Scale	First score	Final score	% scores changed	Correlation coefficient [a]	P value [b]
FLACC	1.6 (2.7) 0 [2]	1.9 (2.9) 1 [2]	23.0	0.91	0.033
MBPS	4.7 (3.1) 4 [6]	4.7 (3.1) 4 [6]	28.2	0.97	0.96
VASobs (pain)	1.6 (2.7)	1.6 (2.7)	8.8	0.94	0.63
VASobs (distress)	3.7 (3.7)	3.6 (3.7)	9.9	0.92	0.58

Values are median [interquartile range] / mean (standard deviation)

[a] Spearman correlation coefficient

[b] Wilcoxon signed-rank test with continuity correction

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, MBPS – Modified Behavioral Pain Scale, VASobs – Visual Analogue Scale observation

12.3.1.2 Reliability

Intra-class correlations were calculated to establish inter- and intra-rater reliability for each scale. Correlations for the FLACC and MBPS were very high and ranged from 0.69 to 0.94 (Table 12.6). Correlations for VASobs pain and distress were lower, ranging from 0.27 to 0.77 and 0.60 to 0.89, respectively. The results of linear mixed modelling confirm that the effect of the reviewer on FLACC scores (variance = 0.004, SD ± 0.063) and MBPS scores (variance = 0.016, SD ± 0.125) was very low and only slightly higher for VASobs distress (variance = 0.146, SD ± 0.382) and VASobs pain (variance = 0.35, SD ± 0.592).

Table 12-6 The reliability of the FLACC scale, MBPS, VASobs pain and VASobs distress - inter-rater overall and for each procedural phase of painful and non-painful procedures and intra-rater overall.

Measure	FLACC	MBPS	VASobs pain	VASobs distress
Inter-rater – overall	0.92	0.87	0.55	0.78
Painful – baseline	0.88	0.85	0.37	0.7
Painful – preparation	0.93	0.89	0.35	0.78
Painful – procedure	0.90	0.82	0.48	0.65
Non-painful – baseline	0.79	0.69	0.27	0.6
Non-painful – procedure	0.94	0.88	0.35	0.89
Intra-rater – overall	0.87	0.88	0.77	0.81

Values are intra-class correlation coefficients (ICC)

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, MBPS – Modified Behavioral Pain Scale, VASobs – Visual Analogue Scale observation

12.3.1.3 Validity

Sensitivity, specificity and AUC using receiver operating characteristics (ROC) at various cut-offs were calculated for each scale, the results of which are reported in Table 12.7. All scales demonstrate the capacity to distinguish between known groups (painful versus non-painful procedures). However, the FLACC scale provided the best sensitivity (94.9%), specificity

(72.5%) and AUC (0.83) for the lowest cut-off score (FLACC score 2). The MBPS was most sensitive (91.4%) and specific (77.5%) at the highest cut-off score (MBPS score 4).

The scores for each scale across phases for painful versus non-painful procedures can be seen in Figure 12.1 and visually demonstrates the responsiveness of all scales to pain. Independent t-tests were used to determine whether the mean difference in scores across phases differed for painful versus non-painful procedures. The responsiveness of the FLACC scale, MBPS and the VASobs to pain (mean difference from baseline to procedure) differed significantly for painful versus non-painful procedures (FLACC 6.09 ± 3.36 vs 1.99 ± 3.34, $p < 0.000$, MBPS 4.73 ± 2.63 vs 1.50 ± 2.73, $p < 0.000$ and VASobs pain 4.96 ± 2.16 vs 0.42 ± 0.99, $p < 0.000$).

These observations were also tested by linear mixed modelling to determine the impact of procedure and phase (fixed effects) on scores where the children and reviewers were considered random effects. The model indicates that there was an average increase of 5.9 for FLACC scores, 4.2 for MBPS scores and 5.5 for VASobs pain scores across phases for children undergoing a painful procedure. Responsiveness was more modest for non-painful procedures with an average increase of 1.8 for FLACC scores and 1.5 for MBPS scores and markedly so for VASobs pain scores (average increase = 0.4). VASobs distress scores showed a similar pattern with an average increase in scores of 5.5 for painful procedures and 1.5 for non-painful procedures.

Independent t-tests were run to determine if there were differences in the mean change in scores across phases of the procedure for infants and children who had higher scores prior to the procedure than those with lower scores prior to the procedure. For children undergoing a painful procedure, the mean difference in scores across phases for both scales was significantly different for children with baseline scores less than three (FLACC 7.01 ±2.90, MBPS 5.71 ±1.94) compared with children with baseline scores of three or more (FLACC 2.70 ± 2.72, MBPS 1.72 ± 2.16), $p = 0.0001$ and $p = 0.0000$, respectively. As there was only one child with a mean VASobs pain score at baseline over '3', this analysis was not completed for VASobs pain.

The correlation between FLACC and VASobs pain ($r = 0.74$) and MBPS and VASobs pain ($r = 0.74$) were identical and very similar for VASobs pain and VASobs distress scores ($r = 0.77$). The correlations between FLACC and MBPS and VASobs distress scores were stronger than for VASobs pain scores ($r = 0.89$ and $r = 0.87$, respectively). The relationships between scores for the different scales are shown in the scatterplots provided in Figure 12.2, where scores allocated by the clinicians for one scale are plotted against scores allocated with an alternative. With the exception of VASobs pain and VASobs distress, where distress scores show a tendency to be higher, there are no obvious patterns in the relationships between scores for different scales.

Table 12-7 Sensitivity, specificity and area under the curve values (AUC) calculated for different cut-off for FLACC, MBPS, VASobs pain and VASobs distress scores to differentiate procedure type.

Cut-off scores	FLACC			MBPS			VASobs pain			VASobs distress		
	Sensitivity (%)	Specificity (%)	AUC	Sensitivity (%)	Specificity (%)	AUC	Sensitivity (%)	Specificity (%)	AUC	Sensitivity (%)	Specificity (%)	AUC
> 0	100.0	20.0	0.60	na*	na*	na*	100	60	0.80	98.3	17.5	0.57
> 1	94.9	60.0	0.77	100	2.5	0.51	91.5	80.0	0.86	93.2	67.5	0.80
> 2	**94.9**	**72.5**	**0.83**	96.6	57.5	0.77	89.8	87.5	0.89	91.5	75.0	0.83
> 3	91.5	75.0	0.83	91.5	70.0	0.81	**84.7**	**95.0**	**0.90**	**91.5**	**77.5**	**0.84**
> 4	91.5	75.0	0.83	**91.5**	**77.5**	**0.85**	76.3	100	0.88	86.4	77.5	0.82
> 5	84.8	75.0	0.80	91.5	77.5	0.85	71.2	100	0.86	85.0	77.5	0.80
> 6	81.4	75.0	0.78	84.7	80.0	0.82	44.1	100	0.72	74.6	80.0	0.77
> 7	76.3	77.5	0.77	71.2	82.5	0.77	18.6	100	0.59	61.0	87.5	0.74

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, MBPS – Modified Behavioral Pain Scale, VASobs – Visual Analogue Scale observation. The values at which the scores most accurately discriminate between painful and non-painful procedures are highlighted in bold.

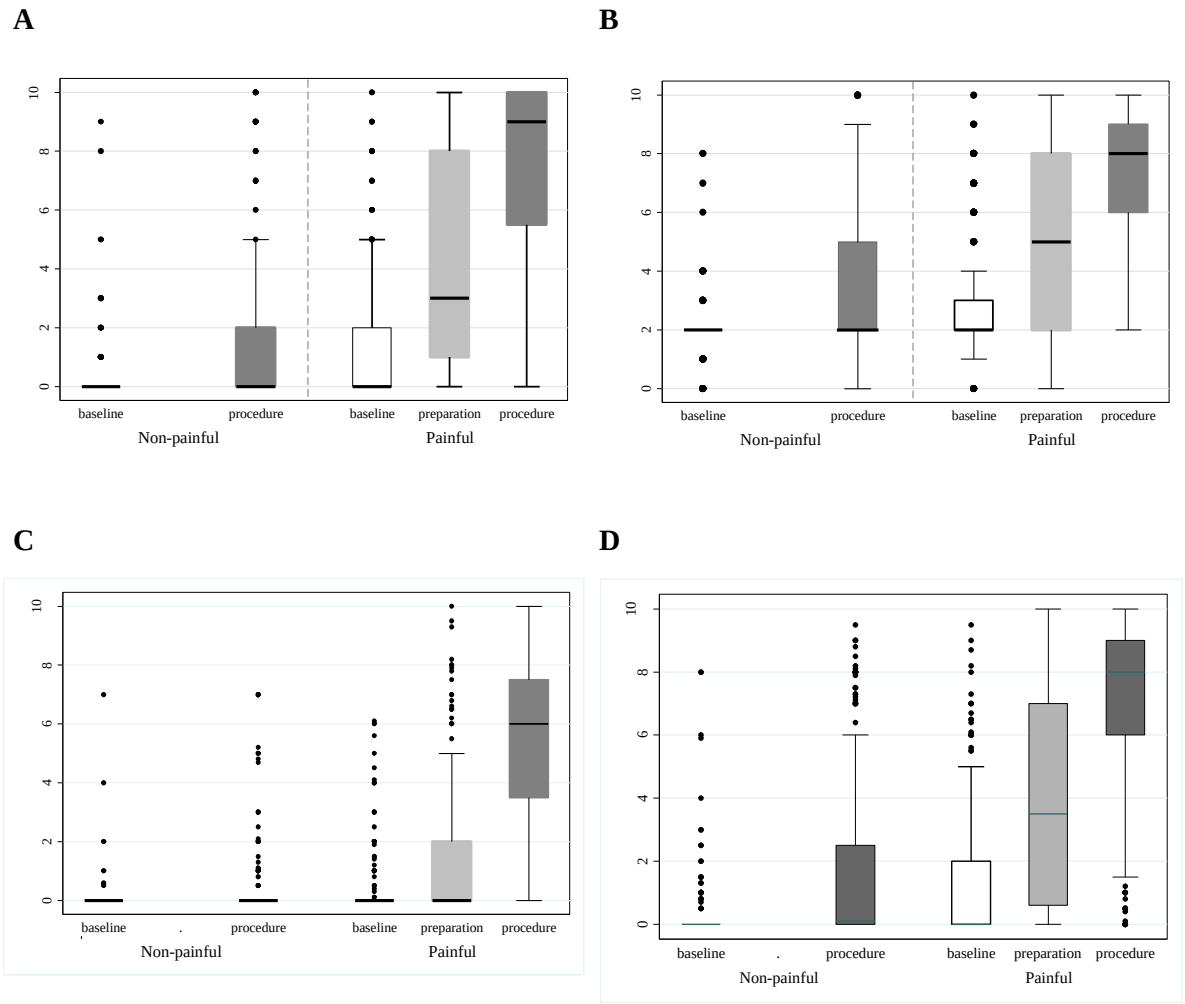


Figure 12-1 Boxplots representing change of values over time (procedural phases) in the two procedure cohorts (painful and non-painful procedures).

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, MBPS – Modified Behavioral Pain Scale, VASobs – Visual Analogue Scale observation

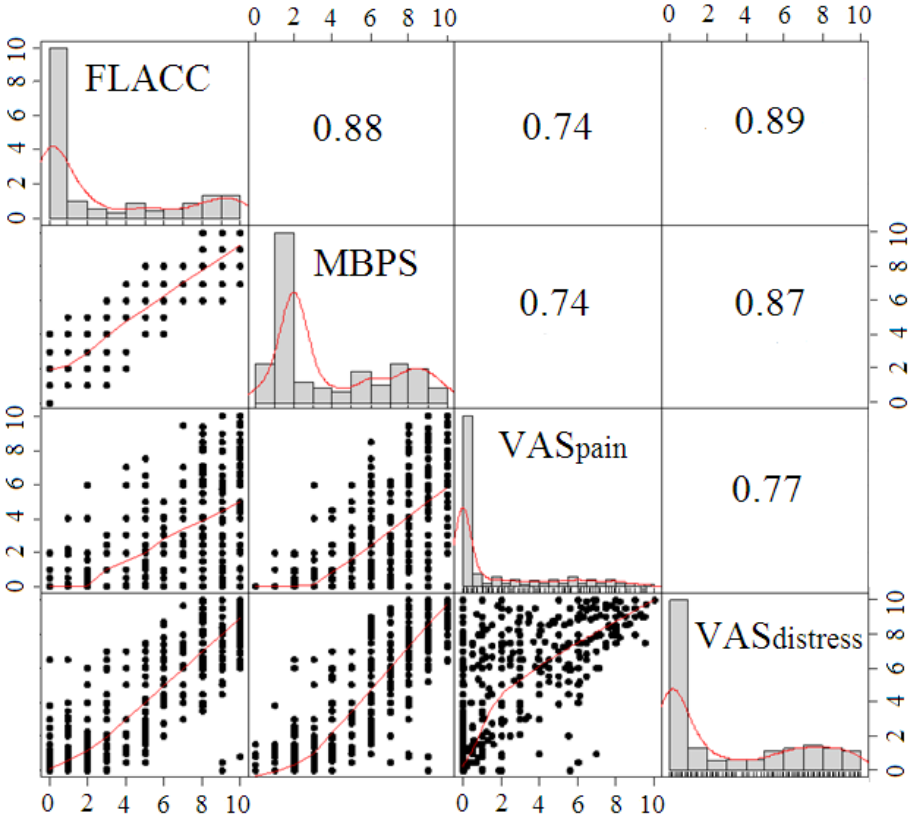


Figure 12-2 Scatter plots demonstrating relationships between scores.

Abbreviations: FLACC – Face, Legs, Activity, Cry, Consolability, MBPS – Modified Behavioral Pain Scale, VASobs – Visual Analogue Scale observation

12.4 Discussion

We have previously shown that the FLACC scale, the MBPS and VASobs pain are sensitive to pain but that they each have limitations to their capacity to differentiate pain related and non-pain related distress (Chapters 9 - 11). Our intention in this study was to determine which scale, FLACC scale, MBPS or VASobs pain, is better suited to assessing procedural pain experienced by infants and young children.

A comparison of the scores across scales reveals an obvious trend for VASobs pain scores to be lower than the score allocated using the other scales. It has been shown that VASobs scores for acute pain are generally lower than self-reported scores (416, 423, 477, 478). Although it is not

possible to replicate these results in a sample of children too young to self-report it is possible that observers using the VASobs would also underestimate pain in this age group. If we accept this, we could assume that the FLACC scale and MBPS scores, which were higher than VASobs scores, more closely represent self-reported scores. Clinically underestimated and undertreated pain is a greater concern for infants and children receiving healthcare than is overestimation or overtreatment (479) supporting our concerns about a scale that consistently scores lower than others. In contrast, MBPS scores for eight of the ten phases were higher than the scores allocated using other scales and, in both cases, these were procedural phases (nasogastric tube insertion and oxygen saturation measurement). Very few observations resulted in an MBPS score of 'zero' and averaged almost 'two' even at baseline. This might be best explained by a possible flaw in the design of the scale. The descriptors for 'face' and 'cry' items require the infant to be 'smiling' and 'laughing or giggling', respectively, to achieve a score of 'zero'. The absence of these behaviours does not necessarily equate to the presence of pain. This is particularly relevant for children undergoing procedures other than immunisation where the context, the more extensive preparation and the concern of their parents is likely to heighten their pre-procedural fear and anxiety making smiling and giggling unlikely.

All scales were shown to be responsive to pain, demonstrated by marked increase in scores across phases (baseline to procedural). Responsiveness of scores to the pain associated with painful procedures was highest for the FLACC scale with an average increase of 5.9, closely followed by the VASobs pain with average increases of 5.5. The responsiveness of MBPS scores was slightly lower with average increases of 4.2. To establish the capacity of these scales to differentiate between pain and non-pain related distress the responsiveness to pain for infants and children demonstrating distress in the baseline phase (score ≥ 3) was compared with those who were more settled (score < 3). The responsiveness of the FLACC scale and the MBPS scores was blunted for infants and children distressed at baseline. Although this may reflect the natural tendency for the difference between two unrelated variables, the values of which are randomly selected from within discrete limits, to be smaller if the differences are compared for two groups based on higher and lower values of one variable, it may reflect limited capacity for these scales to differentiate between non-pain related and pain-related distress. This analysis could not be completed for the VASobs pain scores as there were so few observations scoring at least three in the baseline phase. This is likely to reflect the use of separate VASobs for pain and distress which allowed reviewers to distinguish between pain and non-pain related distress in a way that the FLACC scale and MBPS did not. Finally, analysis of the scales' responsiveness for non-pain related procedures was intended to support our understanding of the scales' capacity to differentiate between pain and non-pain related distress. A highly specific pain scale should ideally show no response or increase in scores in circumstances where there is no increase in pain, i.e. a non-panful procedure. The

average increase in scores across phases for non-painful procedures was more modest, although not zero, for the FLACC scale (1.8) and the MBPS (1.5) but almost zero for VASobs pain (0.4). This implies greater specificity for pain for the VASobs pain than the FLACC scale and the MBPS, which may be the result of the capacity for the observer to make allowances for the context when making their assessment of pain using the VASobs.

These concerns about the capacity of the scales to differentiate between pain and non-pain related distress was shared by the reviewers who reported in the utility survey that they did not consider either scale well suited to procedural pain assessment, or capable of differentiating children with pain from those without or supporting clinical decisions about pain management. However, our greatest concerns about capacity for discrimination lie with the MBPS. Scores for segments of video featuring infants and children presumed not to be in pain (baseline, preparation and non-painful procedures) were significantly higher for MBPS than the FLACC scale and even at baseline averaged almost two. This impacted on the degree of responsiveness demonstrated by the MBPS compared with the FLACC scale and the VASobs pain.

VASobs distress scores followed similar patterns to the FLACC scale and MBPS scores and an average increase of 5.5 was seen for painful procedures and 1.5 for non-painful procedures. This similarity in the way the FLACC scale, the MBPS and the VASobs responded to pain was reinforced by strong correlations between the VASobs distress scores and FLACC scale scores ($r = 0.89$) and the MBPS scores ($r = 0.87$). The correlation between the FLACC scale and MBPS scores was similarly high ($r = 0.88$), while all correlations between VASobs pain scores and the other scales were slightly lower and ranged between 0.74 and 0.77.

The FLACC scale and MBPS were reliable when applied by clinicians in this study and one scale cannot be considered more reliable than the other. This is perhaps not surprising given the similarities between the scales and that each reviewer applied both scales. However, despite the similarities between these scales significant differences existed between the scores allocated using these two scales for most phases. Furthermore, the practical performance of the two scales was not consistent. Reviewers more often changed their score when given an opportunity to review the video segment when applying MBPS, but scores changed more significantly when applying the FLACC scale. Furthermore, reviewers were unable to score FLACC items more often than they were unable to score MBPS items resulting in fewer complete FLACC scores. These results suggest that the FLACC scale may have more practical limitations than the MBPS. In contrast are the results for VASobs pain and VASobs distress. Reliability for both scales was markedly lower than for FLACC and MBPS while the practical performance of the VASobs exceeded that of the FLACC scale and the MBPS; scores were more often complete and changed less frequently.

Furthermore, reviewers rated it as much easier and quicker to apply and indicated a preference for the VASobs pain for pain assessment over the FLACC scale and the MBPS.

12.4.1 Strengths and limitations

There were strengths and limitations to this study. This was a single centre study and reviewers could not be blinded to the circumstances e.g. needle insertion. However, several strategies were employed to overcome the impact of these limitations. Including a larger than usual sample size and the use of unique reviewers to assess each phase of a child's procedure. Multiple methods were used to establish validity to overcome the limitations of each validation method. Finally, it was not possible to statistically compare all psychometric properties of the scales and our results and conclusions are to some extent based on a pragmatic comparison of the performance of the FLACC scale, the MBPS and the VASobs pain and VASobs distress scales.

12.4.2 Conclusions and future directions

The reliability of the VASobs challenges results that suggest that it may be the most valid scale for assessing pain and differentiating pain and non-pain related distress. Furthermore, scores were consistently lower than FLACC scale and MBPS scores raising concerns about the potential for VASobs scores to underestimate pain. For this reason, the VASobs cannot be recommended. Although the MBPS was designed to assess procedural pain, the evidence suggests that it does not perform as well as the FLACC scale to differentiate pain and non-pain related distress. Questions about the design and performance of both the FLACC scale and the MBPS remain which prevent unreserved support for the use of either scale to assess procedural pain. Based on these results we cautiously recommend the FLACC scale for procedural use but advocate for accepting scores as a measure of procedure-related distress and not as a measure of procedure-related pain. We also recommend review and potential revision of the scale to improve practical application and align item descriptors with empirical data.

SECTION 5.

The final section of this thesis is presented in a single chapter and summarises the key findings arising from this work and discusses the implications of these findings. Recommendations for clinical and research practice and based on this work are made. Finally, the remaining gaps in the literature are highlighted and suggestions are made as to how best to address these gaps in our understanding via future research initiatives.

CHAPTER 13.

This thesis was underpinned by three research questions each addressed in separate phases of work. The first was a detailed interrogation of the literature to identify observational behavioural assessment scales that could be considered potentially suitable for assessment of procedural pain. The second phase included three systematic reviews to summarise existing psychometric data to determine the extent to which the three scales' (FLACC scale, MBPS and VASobs) psychometric properties when used to assess procedural pain are supported. The final stage was prospective and tested the psychometric properties of three observational scales (FLACC scale, MBPS and VASobs) used to assess procedural pain in infants and young children aged 6 to 42 months undergoing one of four painful and/or distressing procedures. The results of this work have addressed these questions and bring us closer to achieve our aim which was to identify an observational pain scale that could be used clinically and for research purposes to assess procedural pain in infants and young children.

The aim of this chapter is to summarise the key findings of this work and place them in context of the existing literature and understanding of these scales and draw conclusions about their use. Table 13.1 lists the research questions and directs the reader to the chapters that report the results and provide more comprehensive discussion of the implications of these results of each phase of work. In this chapter, the results of each phase will be integrated to provide an overall assessment of the implications of this project to support recommendations for clinical practice and research.

Table 13-1 Research questions and the chapters that report the results and discussion.

Research question	Chapter
1. Is there an observational pain assessment scale considered suitable for assessing the procedural pain experienced by infants and young children?	Chapter 3
2. Is this scale/Are these scales supported by sufficient psychometric data to recommend the scale for use	Chapters 5 to 7
3. Can the selected observation scales be recommended for procedural use following psychometric testing?	Chapters 9 to 12

13.1 Key findings

The first two phases of this project provided a comprehensive summary of our understanding of the use of behavioural pain scales to assess procedural pain in infants and young children. The increasing focus on the assessment of paediatric pain has resulted in the proliferation of observational behavioural pain scales for this purpose. Over 10 years ago Duhn and colleagues reported the existence of at least 40 scales (286) and then in 2017 Anderson and colleagues in a systematic search of the literature, identified 65 unique scales for assessing pain in neonates, infants and children (480). Despite these numbers there are few clear recommendations regarding the most suitable scale for procedural pain assessment. Two systematic reviews published in 2007 made recommendations based on limited evidence and the absence of suitable alternatives (30, 31). These reviews concluded that for procedural pain assessment in infants and children, the FLACC scale and the CHEOPS showed the most promise but that the evidence supporting these scales was still limited. A working group of the Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine has recently made similar recommendations to those of the two systematic reviews and with similar reservations following extensive review of the literature (15).

We established criteria to help identify an existing scale that **might** be suitable for assessing procedural pain in infants and young children too young to self-report. The absence of strong support for the recommendations made to date was apparent in the results of our attempts to identify accepted scales (based on our criteria for acceptance), which are reported in Chapter 3. A total of 32 different scales were used to assess procedural pain in the identified RCTs, which suggests a lack of consensus about the most suitable scale for use. Furthermore, very few expert associations, societies, academies and organisations made recommendations regarding pain assessment (8/71) and those that did recommended four different scales (most commonly the FLACC scale). In contrast the VASobs was not recommended in any of these statements or in the two systematic reviews mentioned earlier. Despite this, the VASobs was used in the identified RCTs more than twice as often as other commonly used scales (MBPS and FLACC).

Our search also confirmed that despite the significant numbers of scales and the increasing attention directed towards procedural pain assessment and management, only two scales were designed for this purpose (MBPS (288) and EVENDOL (289)). The EVENDOL was also designed to assess acute illness/injury related pain in addition to procedural pain. This scale was excluded as it uses a 0 to 15 metric for scoring, did not meet the criteria for acceptance and has not been tested in English. Eighteen studies testing the psychometrics of scales used to assess procedural pain concentrated on 13 scales and only three scales (MBPS, FLACC and VASobs)

were evaluated in more than two studies (Chapter 3). Three scales were considered potentially suitable for use based on the criteria set a priori. However, this was insufficient to unequivocally support use of these scales for procedural pain assessment.

The inconsistencies exposed in this phase of work likely reflect either insufficient data to confirm the scales' suitability or inconsistent interpretation of available data. These results supported the need for closer scrutiny of the psychometric data available for the eligible scales: the FLACC scale, the MBPS and the VASobs.

13.1.1 Psychometric data

Phase two and three of this thesis focused on comprehensively summarising available data describing the psychometric properties of the FLACC scale, the MBPS and the VASobs used to assess procedural pain and testing these properties in a prospective study. Although the systematic reviews summarised all psychometric data for the scales used to assess pain in infants and children, only the findings as they relate to procedural pain assessment will be referred to in this discussion. A brief integrated summary of the key findings from these phases is presented for each scale and a comparison between scales.

13.1.1.1 FLACC

The FLACC scale is the most frequently recommended and the most extensively evaluated observational behavioural scale identified. However, based on a systematic review of available data we could not conclude that there was sufficient support for the psychometric properties of the FLACC scale to recommend it for procedural pain assessment (330). This conclusion is more cautious than the one drawn 10 years earlier by von Baeyer and Spagrud who concluded that there was sufficient data to provisionally recommend the FLACC scale for procedural use (30). This may reflect the results of robust assessment of the quality of the studies included in the current review. This level of rigour has not applied to review of studies addressing scale performance and only a small number of studies were assessed as using methods rated as 'good' or 'excellent'. This reduced our confidence in these results and lessened their contribution to an assessment of FLACC performance to assess procedural pain.

To combat the lack of high-quality studies evaluating the psychometric properties of the FLACC scale used to assess procedural pain in infants and children, a prospective study to provide this data was conducted in the third phase of this project. The results from this study confirmed that procedural pain can be reliably and sensitively measured by the FLACC scale (469). Inter-

observer agreement was high (intra-class correlation coefficients (ICC) ranging from 0.79 to 0.94 depending on the type and phase of the procedure and intra-rater reliability was similarly high (overall ICC 0.87). The mean difference between the scores was small (-0.12), although the limits of agreement were wide (-3.10 to 2.86). The responsiveness of FLACC scores to painful stimuli also supported the capacity of the scale to assess pain. Scores rose substantially from baseline to the procedural phase of painful procedures (mean difference = 5.35). Sensitivity to pain was shown by the capacity of the FLACC score to differentiate between painful and non-painful procedures at a cut-off score of 2 (sensitivity 94.2%, AUC 0.83). However, specificity results, although promising, were not as convincing. The specificity at a cut-off score of 2 was 72.5% and did not improve much as scores rose. A comparison of the responsiveness of the scale to painful and non-painful procedures confirmed that the scores rose much more from baseline to the procedure during painful procedures when compared to non-painful procedures. However, scores rose from baseline to the procedure during non-painful procedures (mean difference = 1.78). Although the minimal clinical difference in pain scores for the FLACC scale has not been confirmed, based on the results from work to establish this for other scales (481, 482) it is likely that a rise of 1.78 reflects a clinically significant rise in pain score. Furthermore, the mean difference in scores across phases for children with higher scores (FLACC ≥ 3) during the baseline phase was significantly lower than for children with lower baseline scores (FLACC scores < 3). Although may be influenced by the natural tendency for the difference between two unrelated variables, the values of which are random numbers selected from within defined limits, to be smaller if the variables are compared based on higher or lower values for one variable, these results raise questions about the capacity of the FLACC scale to differentiate between pain and non-pain related responses. This will be discussed in more detail in the following section addressing the implications of the results of this work.

Clinicians also expressed concerns about the capacity of the scores to make a meaningful contribution to procedural pain assessment, specifically indicating a lack of confidence in the capacity to differentiate between procedural pain and non-pain related distress or influence treatment decisions. Furthermore, a substantial number of scores (14.5%) could not be completed as items could not be scored due to the impact of restraint or the procedure itself. This is the first time that the feasibility of using this scale for procedural pain assessment has been explored and the results give rise to concerns about how well the scale performs in this context.

13.1.1.2 MBPS

Studies have largely concentrated on evaluating the capacity of the MBPS to assess immunisation related pain in infants and young children and this was reflected in the recently published results

of the systematic review reported in chapter 6 of this thesis (332). The conclusion of this review was that the MBPS was supported by sufficient data to recommend the scale to assess immunisation related pain. However, although it was designed for acute pain and not specifically immunisation related pain, there was insufficient data available to make any assessment of the psychometrics of the scale used to assess pain associated with other procedures. The psychometric evaluation study conducted in the third phase of this work was intended to fill this evidence gap to provide a platform from which to make recommendations regarding use of the MBPS to assess pain associated with procedures other than immunisation.

The results of the prospective study confirmed that the MBPS can be used reliably by clinicians (ICC 0.87 across all procedures and phases) (332) and that the scale is sensitive to procedural pain. Scores rose substantially from baseline to the procedural phase of painful procedures and linear mixed modelling confirmed scale responsiveness to pain (mean difference 4.6). Furthermore, the MBPS was able to differentiate between painful and non-painful procedures with a sensitivity of 91.5% and specificity of 77.5% (AUC 0.85). However, these results were achieved at a cut-off score of 4, which has been defined as the lower end of moderate pain (470). Furthermore, reviewers only recorded a score of '0' for 3.6% of observations despite three quarters of the segments scored depicting infants and children during phases and procedures not considered painful. Based on these results, there is reason for considerable concern about the specificity of the MBPS for pain and the likely reasons for this and the implications will be explored later in this chapter. Finally, reviewers' ratings of their perception of the scale reflected their concern and they rated the scale unlikely to '*reflect procedural pain*' and unlikely to '*contribute to treatment decisions*'.

13.1.1.3 VASobs

The VASobs is the observational scale used most frequently to measure procedural pain in RCTs and yet our systematic review could not confirm the adequacy of the psychometrics of the scale used for this purpose. Over 15 years ago van Dijk and colleagues completed a systematic review to summarise the reliability, validity and cut-offs points for the VASobs to highlight the strengths and limitations of this scale (290). They concluded that the scale required additional testing, specifically to address intra-observer reliability, responsiveness and optimal cut-off points. The systematic review completed in phase two of this project (chapter 5) included 20 studies published since the 2002 review, of which 8 were focused on procedural pain assessment. In addition, this review also included RCTs to provide validation data and 50 studies concentrating on procedural pain were analysed. Despite additional data, the conclusions from this review were similar to those of van Dijk's review 15 years earlier.

Concern was raised in both reviews regarding the reliability of VASobs scores when used to assess procedural pain experienced by infants and children (290). This concern has not been allayed by data from studies published since our 2016 review reported in Chapter 5, which did not test reliability, or the results of the prospective psychometric study completed in phase three of this project. Our data confirmed an overall intra-class correlation of 0.55 for pain scores allocated by clinicians in this study and they were lower for procedures and phases of procedures considered not painful. When asked to apply the scale as a measure of distress reliability improved markedly (ICC 0.78). Intra-rater reliability was better for pain (ICC 0.77, mean difference - 0.64, ± 1.93) and distress (ICC 0.81, mean difference - 0.09 ± 2.27) scores.

Review of available data in the systematic review reported in Chapter 5 was sufficient to suggest that the scale may be suitable for assessment of neonatal procedural pain and could be cautiously recommended for assessing immunisation related pain in infants, toddlers and potentially older children. However, in the absence of data from studies focusing on alternative procedures no recommendations could be made beyond the one made for immunisation-related pain. The psychometric evaluation study conducted in phase 3 of this project provided prospective data regarding VASobs used to assess procedural pain. Pain scores increased predictably in response to painful stimuli (mean difference 4.95) and only a modest increase in scores was seen across phases of non-painful procedures (mean difference 1.52), confirming responsiveness of the scale to the painful procedures in this study. The scale was able to differentiate between painful and non-painful procedures with a sensitivity of 84.7% and a specificity of 95.0% (AUC 0.90) at a cut-off score of 3, although this has been defined as the upper end of mild pain (470). The VASobs distress scores also differentiated between painful and non-painful procedures most accurately at a cut-off score of 3 but sensitivity was slightly higher (91.5%) and specificity was lower (77.5%). Predictably the correlation between pain and distress scores was good (r = 0.77) and correlations with FLACC scale and MBPS scores were similar (r = 0.74). Finally, although clinicians did not show great confidence in the performance of the scale there were very few occasions where they were unable to allocate a pain or a distress score using the VASobs.

13.1.1.4 Scale comparison

Comparison of the performance of the scales highlighted differences in their psychometric properties. There were significant differences between the scores allocated by each scale to the segments of video and MBPS scores were consistently higher during procedures and phases considered non-painful. This pattern did not extend to the procedural phases of painful procedures and FLACC scores were higher for nasogastric tube insertion than MBPS scores. In most

circumstances (phases and procedures) the VASobs pain scores were lower than the other scale scores. The distress scores, while higher than VASobs pain scores, were also lower than the observation scales scores. All scales investigated were shown to be sensitive and responsive to pain. However, the FLACC scale results were more convincing than the results of other scales. Most notably, the FLACC scale, although similarly able to differentiate between painful and non-painful procedures with similar sensitivity and specificity as the other scales did so at a lower cut-off score (FLACC cut-off 2, MBPS cut-off 3 and VASobs pain and distress cut-off 3). However, this was still at a score that is likely to have significant clinical impact as it has been defined as 'mild pain' (470).

The FLACC scale and the MBPS were similarly reliable when applied by clinicians to the range of procedures included in this study (ICC FLACC scale 0.92 and MBPS 0.87). Distress scores using the VASobs were also reliable (ICC 0.78). However, the reliability of VASobs pain scores was considerably lower (ICC 0.55).

Reviewers had the most difficulty applying the FLACC scale and were unable to allocate a score more often than when applying the other scales. Reviewers altered their MBPS scores most often following subsequent review of the video segment compared with the other scales. However, the difference between first and final FLACC scores was significant but was not significant for the other scales. Despite the problems that reviewers had with application of the FLACC scale, the MBPS was liked the least by reviewers and thought to perform worse than the FLACC on most measures of clinical utility other than those relating to ease and speed of application. The VASobs pain scale was ranked highest by over half of the participants (54%) and 88.5% and 84.6% respectively, agreed or strongly agreed that the scale was quick and easy to apply. It is of concern that reviewers were not convinced that any of the scales measured procedural pain well enough to support clinical decisions. This may reflect an intuitive recognition of some of the problems highlighted by the results of this study, which are explored in the following section.

13.2 Implications of findings

The results of this psychometric evaluation study are unique as there are very limited data addressing the psychometrics of observational pain scales applied to procedures other than immunisation. Furthermore, significant attention has been focused on neonatal pain and distress and less on older infants and young children who are too young to self-report pain. These results serve to add to our understanding of the role that the FLACC scale, the MBPS and the VASobs may play in procedural pain assessment for procedures other than immunisation and in age groups

other than neonates. In addition, they highlight several issues that relate to the design of these scales and the impact that this has on their psychometric and practical properties. The impact of infant related and observer related influences on pain assessment are attracting increasing attention and although this was not the focus of the project, the prospective data alludes to the potential importance of these factors. The limited capacity of these scales to discriminate between pain related and non-pain related distress reignites the debate about whether behavioural scales should be considered pain, pain and distress or distress assessment scales. The unique circumstances of medical procedures and the impact that they might have on the pain experience and responses have been illuminated in the prospective study of phase 3. The work undertaken in this project has also unearthed several methodological concerns for researchers to address before additional efforts to assess the psychometric properties of pain assessment measures are made.

13.2.1 Scale related factors

The authors of the FLACC scale designed this scale to provide clinicians with a systematic approach to assessment of post-operative pain that was more easily recalled than some of the other options available (28). They used items from other scales and consulted with clinicians to inform the design before testing the reliability and validity of the scale to assess post-operative pain in infants and children aged 2 months to 7 years. The results of the prospective phase of this study may be the result of attempting to use a scale designed for an alternative circumstance to assess procedural pain. Reviewers had difficulty applying items of the FLACC scale, in particular legs, activity and consolability. They identified restraint as the factor most often impeding their ability to make an assessment of legs and activity. It is unlikely that restraint and the impact of an invasive procedure were considerations when the FLACC scale was developed for assessment of postoperative pain and the current descriptors do not make provision for scoring these items under these circumstances. The consolability item was also difficult to assess on occasions as reviewers reported no attempt to console the infant or child. Furthermore, use of the term 'consolability' is out of step with current evidence regarding the use of non-pharmacological strategies for minimising pain. The Oxford English Dictionary defines *to console* as '*to make someone….feel better by giving them comfort or sympathy*' (483). Evidence points towards a negative effect for some traditional comforting behaviours and that efforts to distract are more likely to have a positive effect on pain experience (484, 485). In addition to the use of appropriate distraction techniques, the aim for management of procedural pain and distress should include reduction of the use of restraint. Nonetheless, revision of the items and/or their descriptors may be warranted to improve the feasibility of using the FLACC scale for procedural pain and to align it with current evidence.

In contrast to the FLACC scale, the MBPS was designed to assess 'acute iatrogenic' pain in young infants. The authors drew heavily on existing literature to design the MBPS and the scale includes items to assess facial expression, movement and cry (288). Facial expression was described as the most specific response to pain but the coding systems for assessment available were acknowledged by Taddio and colleagues as burdensome. Taddio and colleagues used a simplified assessment of facial expression in the MBPS. Cry characteristics, although commonly used and supported by some data, were acknowledged by Taddio and colleagues as inconsistent. Nonetheless the authors included this item in the MBPS. Finally, they referred to the potential for coding systems for body movements based on evidence of behaviours associated with noxious stimuli similar to facial expression coding systems. Their aim was to provide a clinically applicable pain measure and they tested the psychometrics of this newly developed scale to assess immunisation pain in infants aged 4 to 6 months with positive results.

Despite the intended purpose of the MBPS scale and the robust approach to its design, the prospective study did highlight concerns about the design of the MBPS, in particular the face and cry items. As has been noted, reviewers rarely applied a score of zero even to segments of video depicting children during non-painful procedures or procedural phases. Closer examination of the scale provides a potential explanation for this finding. The descriptor for a score of zero on the face item requires the infant to be smiling and the cry item requires the infant to be giggling or cooing to receive a score of zero. The absence of these behaviours does not necessarily confirm the presence of pain. Both behaviours are unlikely for infants and children in unfamiliar circumstances surrounded by strangers. However, in the scales current form, even neutral emotions could be interpreted as pain. Consistent scores of 2 during non-painful procedures no doubt contributed to a cut-off score of 4 for MBPS scores to most accurately discriminate between painful and non-painful procedures. Inclusion of descriptors for scores of zero that provide for neutral facial expressions and the absence of vocalisation may resolve this issue.

Facial expression is widely recognised as one of the most consistent and universal behavioural responses to pain and hence it is included as an item in most observational pain sales, including the two that have been studied here. However, there is increasing concern about how this item performs across scales and this may be a function of the design of the item. Chang and colleagues undertook a study specifically designed to examine the face item of six well-known scales, which included the FLACC scale but not the MBPS. They reported poor reliability for the FLACC scale face item (246). The authors concluded that inconsistency of the descriptors compared with expressions recorded empirically in response to pain was likely the reason for variation in scores based on facial expression. This concern was not reflected in the current prospective study where

overall reliability for FLACC scores was very high (ICC 0.92), the kappa scores for the item scores ranged from 0.61 to 0.82 and the face item kappa score was 0.72. This was traditionally considered substantial agreement, although increasingly it is argued that this might be too generous an interpretation of agreement for health related measures (486). Nonetheless, these results are not consistent to those seen in Chang's study. These results potentially challenge Chang's concerns held about the descriptors of face item of the FLACC scale or alternatively, and perhaps more likely, lend support to the view that clinician judgement is exercised when applying pain scales, including the FLACC scale. This is discussed in more detail in a later sub-section of this chapter.

The absence of strong empirical work to identify sensitive and specific behaviours indicative of procedural pain may be at the heart of the problems previously described. Although scale authors went to some effort to ensure that the FLACC scale and the MBPS were based on behaviours indicative of pain, there is not a strong body of evidence to support that the included behaviours and descriptors of the behaviours are consistently observed in infants and young children experiencing procedural pain.

Finally, the VASobs was originally designed to support self-assessment of symptom intensity. It had long been used for self-report of pain intensity before it was used as an observational tool to assess pain in patients unable to self-report pain intensity. Unlike the MBPS and the FLACC scale, observers are not directed to specific behaviours when applying the VASobs. They are instructed to allocate a score based on their perception of the intensity of the pain experienced. The scores are therefore heavily influenced by the judgements of the observer and their understanding of the behaviours most likely to indicate pain. The impact of observer related factors on application of the scales and the scores allocated are discussed in more detail in a later section. However, it is can be suggested that the design of the VASobs allows for greater observer judgement to impact on the final score than the other two scales.

13.2.2 Patient and observer related factors

The Social Communication Model was developed by Craig and colleagues to explain paediatric pain (92, 487). They considered an infant or child's response to pain a complex social interaction between the infant or child and their carer. In this model the infant or child experiences pain, generates an observable response and the carer interprets this response as indicative of pain and potentially initiates an action aimed at alleviating the pain or removing them from the source of the pain (harm). It is easy to see from this model that factors that influence either the infant/child

or the observer are likely to impact on the pain responses exhibited or the judgements made as to how these responses should be interpreted. Several studies provide persuasive evidence to this effect (334, 473, 474, 488, 489).

The complexity of these relationships and the challenge for researchers in their efforts to understand them is illustrated by studies reporting the impact of pain intensity on the psychometrics of scores. The reliability and accuracy of pain scores has been shown to improve with higher scores (334, 473, 474, 488). The findings from our prospective study where reliability for pain scores (VASobs pain, MBPS and FLACC) was higher in phases where scores were higher is consistent with these studies. These results may represent increasing consistency in the behaviours demonstrated by infants and children at higher pain intensity, more consistent interpretation of a range of behaviours when these behaviours are demonstrated at higher intensity or a combination of factors.

In the following sections the relationship between clinician judgement and pain scoring and the relationships between pain experience, pain behaviours and individual infant/children related factors are used to reflect on the results of this study. Although no attempt was made in the prospective study to examine these relationships, the assumption that observer and patient related factors were likely to affect the psychometric properties of the scores was responsible for the decision to use multiple observers selected from a pool of clinicians and a larger than usual sample of infants and children.

13.2.2.1 Observer judgement

De Ruddere and colleagues have explored the impact on observers' responses to pain of the presence or absence of potentially painful pathology and other psychosocial influences (472, 473). Observers reported lower pain scores, sympathy and inclination to help when there was no medical evidence for pain when responding to adult patient vignettes (472). Furthermore, observers were less likely to be influenced by the patient's self-reported pain when psychosocial influences on pain such as: anxiety and emotional distress, were included in the vignette. In a follow-up study using clinicians as observers rather than lay people, De Ruddere reported lower pain scores, less sympathy and greater suspicion of deception for patients where there is no medical evidence for pain (473). The perception of deception and negative impressions of the patient were increased by the presence of psychosocial factors considered to influence pain. Recognition of a potential link between clinical data e.g. diagnosis and observer assessments of pain is not new.

Similar effects to those seen in De Ruddere's studies were reported in medical students over 20 years ago (489). Medical students' judgements of pain were higher for patients who had medical evidence of a painful condition compared with those with no medical evidence of a painful condition. Furthermore, students did not consistently underrate pain and for low intensity pain, students augmented the score slightly regardless of the medical evidence. Although the clinicians in our study were not given any clinical or demographic information about the infant or child in the video segment, it was not possible to blind them to the infant or child's circumstances. Hence, it is possible that circumstances may have had similar effects on judgement as those seen in the studies described. A significant difference between VASobs pain and VASobs distress scores (1.9 vs 3.09, t-test(2126) = -10.36) was shown, which may be the result of one of two possibilities; clinicians recognised behaviours that distinguished pain related distress from non-pain related distress or they interpreted the behaviours as indicative of non-pain related distress and not pain based on their perception of the level of pain or distress associated with the circumstances (e.g. absence of a painful stimulus during baseline and preparation phases and non-painful procedures).

It is likely that the clinicians in this study used their judgement when applying the face item of the FLACC scale and very likely when applying the VASobs. Reviewers were fairly consistent in their application of the 'face' item of the FLACC scale. However, as noted earlier Chang and colleagues work reported considerable variability in scores for the face item which they attributed to the inconsistencies in the descriptors (246). Even though clinicians in the current study were given the opportunity not to score an item, they scored the face item in most cases and on those occasions where they were unable to do so this was not because of inconsistent descriptors. This implies that reviewers scored the face item based on behaviours they saw and assessed as most severe regardless of the descriptors provided. Furthermore, these results suggest that the clinicians in this study made similar judgements to other clinicians in this study regarding the facial expressions of the infants and children seen in this study. Gomez and colleagues make similar observations which they attributed to observers reinterpreting the descriptors for the face item to achieve consistent scoring (474). It seems likely that clinicians apply personal judgement if the scale item and/or descriptors do not accord with the behaviours they see and that this will reflect their understanding of the behaviours representative of pain.

In a series of studies, Pillai Riddell and colleagues have provided evidence that the application of judgement also affects the assessments made by parents (490-493). They have repeatedly shown that the variance in parental pain ratings was not explained exclusively by infant behaviours. Furthermore, they have demonstrated relationships between parental factors such as parental age, number of children and infant related factors such as the age and sex of the infant and parent reported pain scores.

Shen and colleagues explored the impact of assessor and patient related factors on the accuracy of FLACC scale scores for children having burn dressing changes (334). They reported that nurses with more than 11 years nursing experience agreed with the expert panel less often than nurses with less than 5 years experience (334). They also found that nurses in their study were unable to distinguish between moderate and high pain (defined as FLACC scale score 4 and 6). Comparison between observer assessments and self-report of pain scores has repeatedly demonstrated that observers underestimate pain (423) and that this is more pronounced for clinicians (430, 494, 495) and with increasing experience (496). It was not the aim of this study to specifically examine the relationship between the demographics of the clinicians and their scores. Clinicians in this study were experienced (average of 10.1 years of experience, ± 4.5 years) and only two nurses had less than 5 years experience. Therefore, sub-group analysis to compare the reliability of experienced and inexperienced clinicians was not possible.

Researchers have exposed a concerning potential relationship between repeated exposure to the pain experience of others and observer judgements of the pain of others (497-500). Gregoire and colleagues reported reduced estimates of the pain intensity experienced by others following repeated exposure to images of facial expressions of pain. The interest in the relationship between exposure and observer assessments has been taken up by researchers using imaging to map brain activity. There is evidence that there is overlap between the regions of the brain that become active during a painful experience and those that become active when the pain of others is witnessed (501). In related work, Gregoire's research group demonstrated that in addition to judging pain as less intense, there are changes in brain activity in observers following repeated exposure to facial expressions of pain (502). In several studies the changes seen confirmed diminished brain activity in observers following repeated exposure to images of facial pain expressions (502, 503). These observational and imaging studies may give us insight into our results, which showed lower VASobs scores compared with FLACC scale and MBPS scores. In light of the absence of specific descriptors to guide assessment using the VASobs logic suggests that application of the VASobs leaves more room for observer judgement than application of the two behavioural scales. Although not confirmed empirically, it may be surmised that clinician habituation to pain in part explains these observations.

These studies and some of the results from the current prospective study challenge us to consider the impact of observer related factors on the application of pain scales and how observations are interpreted and scored and support the need for studies that aim to expand our knowledge of the factors that affect the way that observers assess pain.

13.2.2.2 Patient pain responses

Attention has been increasingly devoted to establishing the role that patient related factors such as infant temperament, experience and the capacity to retain pain-related memories, might have on pain experience and pain-related responses. Experience and memory and discussed in the next section as they are particularly pertinent to a discussion focused on procedural pain.

In a systematic review aimed at identifying the factors that increase anticipatory distress, the authors drew a link between temperament and anticipatory distress (504). Other authors have attempted to understand the relationship between temperament and pain behaviours in response to immunisation (505). Stevens and colleagues demonstrated that by 2 months of age pain behaviours predicted parent assessment of temperament. Racine and colleagues point us to the developmental literature suggesting that children with internalising or externalising problems also have problems regulating their affect which may in turn impact on the behaviours that these infants and children exhibit in response to pain. A longitudinal study mapping the responses to their 2, 4, 6 and 12 month immunisations highlighted the considerable variation in response to a similar stimulus of 750 infants (506). Pillai-Riddell and colleagues recognised stable groups of infants at each age that demonstrated different patterns of response which became increasingly apparent with age. These results serve as the basis for their concern regarding the validity of using overall group means as they deny the potential 'trait-like differences in affect regulation, distress or pain responding'.

Temperament and its impact on pain-responses were not examined in the prospective study reported in Section 4 of this thesis. However, infants and children in this study were grouped according to their level of distress to determine whether pre-existing distress had an impact on scores during the procedural phase of the procedures. Negative emotions such as fear and anxiety are known to modify pain experiences by increasing pain perception (92, 507). Counterintuitively, FLACC scale and MBPS scores in this study did not rise as much during the painful procedure in infants and children who were distressed at baseline when compared with infants and children who were not distressed at baseline. This may be the product of the natural tendency for the difference between unrelated variables to be lower where the value of one variable is higher if the values of the variables are random numbers selected from within discrete limits or a real tendency for pre-existing distress to impact on a child's experience of pain or importantly it may reflect poor capacity for the scale to measure pain and distress cumulatively. Similar results were found in Humphrey and colleagues' study to determine the level of distress associated with venepuncture (508). They reported strong correlations between preparation phase distress and procedural distress in children and adolescents. More significantly, they reported that for the

majority of participants the distress during the preparation phase was the same as during the procedure phase. There are several ways in which the prospective results of this project can be interpreted, but in light of the work of Pillai-Riddell and others these results may reflect the existence of infant/child specific factors that alter their experience of or their behaviours in response to a stimulus. Alternatively, this may be linked to the phenomenon seen in the work of Rhudy and colleagues where pain responses in adults who were anxious increased while pain responses for those who were fearful actually decreased (509). These studies underscore the complexity of the pain experience which has significant implications for interpreting pain scores based on the assessment of behaviours thought to represent pain experience.

13.2.3 Context -related factors

Procedures create a unique set of circumstances likely to impact on the capacity to generate reliable and valid pain scores. There are practical challenges to application of pain assessment scales during medical procedures. Furthermore, children and infants are known to experience anticipatory fear and distress prior to and during medical procedures, which are likely to influence their pain experience (12, 510-512). The capacity to differentiate between the infant or child's responses to anticipatory fear and other aversive emotions from those associated with pain is vital for a scale intended to assess procedural pain.

Anticipatory distress is a function of the perception of threat and this will be fed by memories of previous experiences of pain associated with an event (513). Researchers have increased our understanding of the capacity of infants and young children to retain pain-related memories (12, 514, 515). The impact of memories on future pre-procedural distress was demonstrated in Taddio and colleagues' study of infants born to diabetic mothers. Infants who underwent repeated heel lances to check blood glucose levels following birth showed higher levels of distress in response to non-painful procedural cues such as skin cleansing when compared to neonates who had not experienced heel pricks for blood glucose testing (12). It is worth noting that the same tool for assessment was used to measure pre-procedural distress and procedural pain, so this study not only provides evidence that neonates remember painful procedures and that it affects anticipatory distress, it is also potentially evidence of the limited capacity of assessment tools to differentiate between pain and distress. The results of the current prospective study offer no insight into the role of memory or past procedural experience, but this literature may assist us to understand why some infants and children in our study exhibited behaviours suggestive of higher levels of pre-procedural distress than others. Our study was designed to assess the psychometric properties of the chosen scales and not to explore factors that might impact on their experience. However, as

exposure to previous painful procedures was recorded and approximately half of the infants and children in our study were procedurally naïve it would be possible to compare the scores of these two groups.

Our results confirm long held concern that observational scales are not specific for pain-related distress (263). In 2007, von Baeyer and Spagrud highlight how few studies reported results addressing the capacity of the scale to differentiate between pain-related and non-pain-related distress. Over ten years later, this gap remains in the research literature. The studies reviewed in our three systematic reviews infrequently used discriminant validation methods and did not provide persuasive data to support the specificity of the FLACC scale, the MBPS or the VASobs for assessing procedural pain (330, 332).

Procedures not considered painful (inhalational medication administration and oxygen saturation measurement) were purposefully included in our prospective study to assess the capacity of the scales to discriminate between painful and non-painful procedures. Receiver operating characteristic (ROC) analysis demonstrated that all scales could make this distinction, the FLACC and MBPS with very high levels of sensitivity (FLACC 94.9%, MBPS 91.5%, VASobs pain 84.7% and VASobs distress 77.5%). Specificity although high was slightly lower for the FLACC scale (72.5%), MBPS (77.5%) and VASobs distress (77.5%), meaning that these scales, although likely to correctly identify most painful procedures as painful they were also likely to identify some non-painful procedures as painful. This seems an acceptable compromise as over-assessment of pain is in most circumstances more acceptable than under-assessment of pain and the likely impact that this has on treatment. The VASobs scores performed similarly to the observational scale with the exception of VASobs specificity which was notably higher than for other scales (95%). This may be explained by the inclusion of separate pain and distress scales to allow clinicians to make the distinction between pain and distress. As the clinicians were not blinded to the circumstances, they were able to decide whether the infant or child was likely to be experiencing pain and therefore whether the behaviours represented pain and/or distress.

The most concerning result of the ROC analysis was the cut-off score at which these results were achieved. The MBPS was most accurate (AUC 0.85) at a cut-off score of MBPS = 4, while FLACC was most accurate (AUC 0.83) at a cut-off of FLACC = 2. Pain severity based on scores for the FLACC scale, MBPS and the VASobs have not been defined empirically. However, in older children able to self-report, moderate pain has been defined as a Face Pain Scale-Revised (FPS-R) score from 4 to 6 and mild pain as an FPS-R score from 1 to 3 (470). Therefore, it is reasonable to conclude that the MBPS did not meaningfully distinguish between pain-related and non-pain related distress in this study. The FLACC scale fared much better but nonetheless, a cut-

off score of ‘2’ makes scores of ‘2’ difficult to interpret i.e. establishing whether they represent pain or non-pain related distress. Closer inspection of the data reinforces these results; scores (particularly MBPS scores) during phases and procedures not considered painful were not universally zero. As has been described this is likely linked to the design of the MBPS. However, it probably suffers the same challenges as other observational scales related to differentiation of pain and non-pain related behaviours.

Iannetti and colleagues provided a very cogent argument to explain a significant logic flaw that they consider underlies the commonly accepted interpretation of data from brain imaging studies (516). They argued that researchers have applied reverse inference to make claims based on study data that demonstrate the regions of the brain that become active in response to painful stimulus. These results are used to support the reverse inference that activity in these regions of the brain can be assumed to mean that pain is experienced. However, Iannetti and colleagues provided considerable evidence to show that these regions of the brain are not specific to pain and may be activated under other circumstances. Parallels can be seen between this and the literature devoted to observational pain-related behaviours. Behaviours such as specific facial expressions, cry and movement have been consistently observed in response to painful stimuli. Logic has been used to suggest that therefore observation of these behaviours must be associated with the experience of pain and that they can therefore be used to measure pain experience. The results of this prospective study and the absence of strong evidence from other studies does not support this inference and suggests that this may well represent similarly flawed logic to the inferences made based on imaging studies. These behaviours are likely to be indicative of pain but are also likely to be provoked by other non-pain related behaviours as has been discussed.

The contrasting view is that to suggest a distinction between pain and non-pain related distress can be made is a false dichotomy. Blount referred us to operant conditioning to explain the possibility that the distress associated with medical procedures is akin to pain (510). He reasoned that through previous exposure to a painful procedure infants and children are conditioned to the same experience in response to similar triggers even when the stimuli is not noxious. Based on this conditioning the distress experience during the preparation phase of a procedure is felt by the infant or child as pain. Taddio and colleagues applied this reasoning to the results of their study demonstrating the impact of repeated procedures and memory on pre-procedural distress (12). They proposed that repeated exposure conditioned infants to exhibit pain behaviours in response to non-painful stimuli and that these infants may even be experiencing pain.

To our knowledge, this is the first study to assess the feasibility of using the FLACC scale, the MBPS or the VASobs for procedural pain assessment and as has been described earlier the results

have raised questions about the feasibility of using the FLACC scale and to a lesser extent the MBPS for procedural pain assessment. Restraint was the most commonly cited reason for clinicians not to score an item on the FLACC scale and the MBPS. Notwithstanding the design related challenge of restraint on assessment, the impact of restraint on the infants and child's experience and their behaviours is unclear. It is conceivable that the act of restraining an infant or child provokes a response to the restraint which is independent of their response to other sensations that they may be concurrently experiencing. Restraint removes a sense of control and there is evidence that supports a relationship between perceived loss of control and anticipatory fear and perceived pain (504). Lacey and colleagues reported the impact of positioning for immunisation on children aged 4 – 6 years (517). Children who were placed in the sitting position prior to and for the immunisation were less frightened than children who were placed in the supine position. These findings suggest that restraint plays a much bigger part in procedural pain assessment than just interference with the capacity of the observer to score the behaviour.

Sedation is widely recommended and has become common practice for managing the distress experienced by infants and children associated with diagnostic and therapeutic procedures (15, 518-520). The impact of sedation on pain perception and behavioural responses to pain have not been studied. Infants and children in this study were not sedated for these procedures and these questions are beyond the scope of this project. However, as the use of sedation for procedures increases, this is a question that should be addressed to gain a comprehensive understanding of the experience of the infant or child and how best to assess this experience.

Procedural interference was the second most often cited reason for being unable to score an item. Logically, procedures that involve the face e.g. nasogastric tube insertion and inhaled medication administration via face mask and spacer could be expected to interfere with scoring the facial expression. Equipment may obscure the face but direct contact with the face may also alter the expression e.g. wriggling the nose to avoid tube insertion or an attempt at local withdrawal in response to direct stimulation, such that they are no longer consistent with the facial actions described by the item of the scale. So perhaps it is surprising how often observers were able to allocate a score for this item of the FLACC scale and the MBPS for these procedures. The direct impact of different types of procedures on behaviours is unknown but perhaps warrants attention when considering the assessment of procedural pain

13.2.4 Study design and analysis related factors

Establishing the psychometric properties of measures such as pain scales is methodologically complex, particularly in the absence of a 'gold standard' against which the performance of the measure of interest can be compared. There has been an evolution in our understanding of what it means to validate a measure and the design and analysis methods needed to achieve this rigorously, which in turn impacts on interpretation of results based on methodological factors

In the absence of a gold standard to which the scale of interest can be compared, studies should employ multiple validation methods to indirectly build a case for validity (447). Psychometric evaluation studies have traditionally used several methods to support the validity of the scale and they include correlation between the scores of the scale of interest and another measure (see Table 4-3). These results should be interpreted cautiously. In most circumstances where this method is used the reference scale is not supported by sufficient data to confidently claim that the scale is valid. This method is based on circular logic which was clearly illustrated in Taddio and colleagues' studies to establish the validity of the MBPS and the VASobs (288, 488). This was described in more detail in *Chapter 5*, but in short both scales were used as a reference scale to support the validity of the other. The net result of these studies is evidence to confirm that these scales measure the same thing but not what they are measuring.

Responsiveness is a commonly used validation method to test the psychometric properties of pain scales. It is based on the hypothesis that as pain is associated with many procedures e.g. needle related procedures, pain scores will increase in response to these stimuli. Conversely, if the procedure is not painful scores would not be expected to rise. In studies where responsiveness has been tested the analysis has often involved comparison of the mean scores for the different phases of the procedure. However, this is a simplistic approach which does not consider the effects of the child, the reviewer or repeated measures. Mixed linear modelling was used in this study to account for these factors in the analysis of the change in scores across the phases of the procedures.

Validation of scale performance includes methods to demonstrate the capacity of the scale to differentiate between pain and non-pain related distress. Traditionally, t-tests to establish a statistical difference between groups has been used. Taddio and colleagues used this approach analysis to examine the capacity of the MBPS to differentiate between two groups of infants receiving different immunisations (488). The t-test was also used to analyse the MBPS data from our study for publication and a statistically significant difference between the scores of infants and children undergoing painful versus non-painful procedures was shown. (464). The t-test

measures whether a significant difference between the means exists. The aim of validation is to determine whether a specific score can predict the group (painful versus non-pain procedure) that the infant or child belongs on the basis of their score. This is appropriately assessed using receiver operating characteristic (ROC) analysis to report sensitivity, specificity and area under the curve (265). The ROC analysis for the MBPS data in our study provided much more meaningful results than those generated using the t-test. This analysis confirmed that the MBPS discriminated between painful and non-painful procedures most accurately (sensitivity 94% and specificity 77%, AUC 0.84) at a MBPS cut-off score of 4. However, ROC analysis highlighted a substantial limitation in performance of the MBPS which was not made apparent by testing the difference in the means.

Analysis of reliability data should use techniques designed to provide an estimate of absolute agreement (290). The traditional approach to reliability data has been to a calculate correlation coefficient, e.g. Pearson's *r* and many of the studies examining reliability identified in the systematic reviews used this method of analysis. This method of analysis reports the extent of the association between observers' scores but does not provide a means to determining the level of absolute agreement between responses. In the prospective study, reported in Section 4, intra-class correlation coefficients were used to calculate intra- and inter-rater reliability. Furthermore, Bland-Altman plots and statistics were used to assess intra-rater reliability. Both analysis techniques assess absolute agreement and are described in more detail in Chapter 2. They are recommended for analysis of reliability in preference to traditional correlations such as Spearman's correlation (342, 521, 522).

We have previously acknowledged the trap in interpreting reliability as binary, e.g. as either 'reliable' or 'unreliable'. This is a simplistic interpretation and reliability is more correctly interpreted by degrees e.g. low, moderate and high levels of reliability. Therefore, it can be concluded that the retrospective and prospective reliability data for the VASobs confirms at best moderate levels of reliability for VASobs scores when the scale is used to assess procedural pain. However, when used to assess distress much higher levels of reliability are achieved.

This data infers that the VASobs pain score could be used to assess procedural pain. However, the conventional view of psychometrics places reliability and validity on a hierarchy where for scores to be considered valid they must be reliable. Furthermore, if they are shown to be valid they are accepted as reliable. Quantitatively, validity is a function of the reliability of the scores using the test of interest and the reliability of the criterion against which it is compared. The relationship between these two properties is described mathematically by the equation reported in chapter 2. If this holds true, the concerns held about the reliability of the VASobs make it

difficult to accept the VASobs as valid for procedural pain assessment. However, it is undeniable that the results of validation testing for VASobs pain and VASobs distress were promising and perhaps should not be dismissed so readily. The complexity of the construct of interest (pain) and the many factors influencing its assessment may make a framework such as the one described to explain the relationship between reliability and validity too rigid for interpreting these results. Pain is a qualitative rather than a quantitative experience and pain behaviours have been described as a system of communication designed to elicit a response (92). In this paradigm pain assessment is a function of the patient/clinician dyad and providing the patient elicits the response needed, the specifics of the assessment may be less important.

13.3 Strengths and limitations

Procedural pain management has attracted considerable attention in the last two decades, but this is the first focused project to use a series of steps to identify an observational pain scale adequately supported by psychometric data to recommend it for assessing pain experienced by infants and young children undergoing a diagnostic or therapeutic procedure. Previously published systematic reviews have included reviews of the psychometric properties of existing scales. However, there were limitations to these reviews that were largely overcome by the design rigour of this project. A detailed interrogation of the literature preceded robust systematic reviews to summarise the psychometric properties of the scales identified as potentially suitable. These reviews uniquely included a quality assessment of the included studies to provide a strong foundation for the recommendations made based on the review results. In each case these recommendations included the need for new studies to assess the psychometric properties of the identified scales when used to assess procedures, which was the final stage of this project.

While a strength of this work, the search strategy and the criteria designed to assist in the identification of potentially suitable scales may also have been a weakness. The criteria used for this purpose were developed for this thesis and were based on several assumptions: that a scale designed explicitly for assessment of procedural pain was likely to perform well, that a scale psychometrically tested for this purpose may be supported by data to this effect and that a scale that has been accepted by the clinical and research community is also likely to perform well. In addition, the criteria used to define acceptance were based on logic rather than an established convention, in particular the number of RCTs needed to confirm acceptance. It was thought that these criteria would narrow a field of 65 scales to those best suited for purpose. However, it is also possible that these criteria resulted in the exclusion of scales that although less well known,

less widely used or unsupported by data were better suited for procedural pain than those that were eligible based on the criteria set.

The systematic reviews conducted in phase two and reported in Section 3 of this thesis used a method based on the principles of the PRISMA Statement to ensure rigour (329). Furthermore, the methods used to test the psychometric properties of the scales were assessed using quality assessment tools designed for this purpose (COSMIN Checklist (342) and the Jadad Scale (338)). Anderson and colleagues. in a recent systematic review highlighted how infrequent quality assessment of the eligible studies is included in the methods for pain scale related reviews. However, there were limitations, which have been reported in Chapters 5 to 7, and they included the effects of a potential publication bias based on the inclusion of only published studies. Studies published in languages other than English were excluded as the contribution that evaluation of a translated version of the scale might make on the English version is unclear. Similarly, data from studies focused on adults may potentially make a contribution to our understanding of the application of pain scales. However, as psychometric properties are associated with the scores (e.g. the circumstances to which the scale is applied) these sources of data were excluded. Despite these limitations, these reviews made a unique contribution to the literature regarding the FLACC scale, the MBPS and the VASobs and provided a platform for recommendations regarding their role in procedural pain assessment.

The prospective study in this project was designed to overcome many of the limitations of the methods used in studies included in the systematic reviews by using multiple validation methods, large sample sizes, unique reviewers across phases of the procedure for the same child and appropriate analysis methods. The COSMIN checklist was originally developed to assess methods quality (342). However, the authors have acknowledged a role for the Checklist when designing a psychometric evaluation study and it was used to inform the protocol for the prospective study, which was reproduced in Chapter 8 (468). Furthermore, statisticians were consulted during planning to assist the development of an appropriate analysis plan. Nonetheless there were several limitations to the prospective psychometric evaluation study. The use of video recordings allowed multiple assessments of the same infant or child under controlled circumstances where reviewers were uninterrupted. This is a common strategy used by pain researchers aiming to test reliability and validity (246, 334, 448, 474, 488, 523-529). However, as real time, real world application may impact on a clinician's application of the scale, recommendations made based on video-based data collection should be applied cautiously to clinical practice. Furthermore, reviewers scored a large number of videos on a single occasion providing opportunity to practice and compare application across different segments. This may have resulted in relatively higher

reliability estimates than those that might have been achieved if scoring was completed in real-time clinical practice.

A large cohort of clinicians responsible for paediatric procedural pain assessment was recruited to ensure that the results reflected the performance of the scale when used by clinicians and not by researchers trained to use the scale to improve the reliability of scores. However, this cohort was recruited from a single centre, where clinicians may be socialised to approach pain and pain assessment in a particular way. This also makes it important that this is considered before generalising to other clinicians. Clinician reviewers could not be blinded to the circumstances experienced by the infant or child, which may have impacted on their judgement in ways that have been described earlier. Attempts to limit the impact of this were made by preventing reviewers from assessing more than one segment of the procedure from the same child. This was to prevent reviewers from creating patterns in their assessments to reflect a predictable pattern in scores across a procedure. However, this does not reflect the circumstances under which these scales are used clinically where clinicians assessing pain are not blinded to the procedure or denied the opportunity to see patterns in behaviours across phases of the procedure. Given the complexity of the pain experience and the potential for variability in individual responses it is possible that the change in behaviour rather than the raw score provides the best measure of the infant or child's pain experience. This is similar to the approach that has been taken by experts exploring the best ways to assess pain in infants and children with cognitive impairment (222, 223, 225). Scales, such as the FLACC scale, are tailored to include individualised responses that allow for the patterns of behaviour seen in individuals which are thought to be their unique pain related behaviours.

The use of clinician reports of the clinical utility of the FLACC scale, MBPS and VASobs is a limitation of the prospective study. Testing of clinical utility is best achieved by assessing the impact of scale use on the clinical decisions made as clinician reported behaviours may not accurately reflect the decisions and behaviours demonstrated clinically. Therefore, the results can only be considered an estimate of utility (277).

13.4 Recommendations and future directions

Recommendations regarding FLACC scale, MBPS and VASobs use for clinical and research purposes can be made based on the results of this project. Furthermore, the results highlighted remaining gaps in our understanding of the performance of these scales used to assess procedural pain experienced by infants and young children. These gaps serve as the basis for

recommendations for future research. Finally, knowledge translation and implementation of these results and the ensuing recommendations is briefly addressed.

13.4.1 Clinical and research use

The FLACC scale is supported by the most convincing evidence to support recommending it for clinical and research use. However, the practical shortcomings of the FLACC scale used to assess procedural pain need to be acknowledged. Restraint, the absence of attempts to console and the procedure itself all have the potential to interfere with the allocation of a FLACC score. The relative ease with which the VASobs is applied may in part explain its frequent use. However, the reliability of scores in this study prevent recommending this scale for use. The MBPS also cannot be recommended for procedural pain as the results call into question the capacity of the scale to adequately differentiate pain from non-pain related distress and to differentiate pain and non-pain related distress and from a neutral state.

Although, the FLACC scale is reliable and sensitive to pain, there is data to suggest that it is not as specific. Based on these results it must be assumed that the scale measures pain AND to some extent non-pain related distress. Scores should be considered indicative of a composite of experiences and clinical decisions regarding treatment should reflect a multidimensional experience. This shift has largely occurred in clinical practice where approaches to management have acknowledged the need for distress management to address pain, fear and anxiety. This includes non-pharmacological strategies to alleviate distress (484) and in the last two decades recommendations for sedation to assist with management of fear and anxiety (15, 519, 530-532).

A similar shift may be needed for study designs as many of the RCTs identified in the reviews were designed to test the efficacy of an analgesic or local anaesthetic. This approach to the question and research design is reliant on use of a scale with high levels of sensitivity and specificity for procedural pain. In the absence of a scale that can make this distinction it is possible that the efficacy of the analgesic or anaesthetic agent might be masked by overwhelming non-pain related distress associated with the diagnostic and therapeutic procedure. A trial conducted to test the efficacy of nebulised lignocaine to reduce the pain associated with nasogastric tube insertion in infants and young children may be an example of a study where the results were compromised by the capacity to distinguish between pain and non-pain related distress (351). Despite evidence of the effectiveness of this strategy for minimising the pain associated with this procedure in adults (533-535), Babl and colleagues could not demonstrate a difference between the lignocaine and placebo group based on their FLACC scores. There are potentially equally

plausible alternative explanations for these results. However, the results of this RCT in light of adult data, give reason for concern about the capacity of this scale used to distinguish between pain and non-pain related distress.

One approach to this challenge is to consider the distinction between pain and non-pain related distress unwarranted and trials designed to assess the efficacy of an analgesic or anaesthetic, unnecessary. This would require a paradigm shift in approach to research design that acknowledges the duality of the procedural experience (pain and non-pain related distress) and would result in trials designed to test the efficacy of management regimens aimed at treating procedural distress concurrently rather than regimens designed to solely treat pain. This approach aligns with clinical practice recommendations that acknowledges pain as well as non-pain related distress and advocate for a composite of strategies such as analgesics and local anaesthetics and sedation and other non-pharmacological measures aimed at alleviating distress.

13.4.2 Future research

Gaps remain in our understanding of the assessment of the experience of infants and children; some reflect areas that were not explored by this project and some that have been revealed as a result of this work. These gaps underlie the recommendations described in this section. Furthermore, limitations to the methods used also contribute to recommendations for additional research in this area.

Before commencing new studies to address these gaps, the capacity to answer other questions based on the existing dataset should be explored. Relationships between variables and direct comparison between reviewer responses not analysed as part of this thesis may increase our understanding of the potential influences on the psychometrics properties of the studied scales. For example; comparing responses where the clinician indicated that they were unable to score an item with those of the other reviewers to determine how often reviewers disagreed about the capacity to score items may point to the influence of clinician judgement when allocating scores. Additional analysis could also include analysis intended to provide us with increased understanding of the methods used to test the psychometrics properties e.g. to examine order effects based on which scale was applied by the reviewers first.

Confirmation of the clinical utility of a tool is key to accepting it for use and in the prospective study this was a clinician reported assessment of utility. Future research should aim to test application of the scale in the clinical environment to provide opportunity to test the impact of the

pain scores on clinical decisions to provide an estimate of clinical utility (277). It is widely believed that self-reported behaviours are not a reliable proxy measure of clinical behaviour, although a systematic review testing the relationship between self-reported behaviour and clinical behaviour concluded that the evidence was inconclusive as to its reliability as proxy measure (536). Nonetheless, sufficient criticism exists to warrant testing of clinical behaviours rather than the reported intensions or impressions of clinicians (277). Furthermore, this approach to testing the application of the scale will more closely approximate its performance in clinical practice than application of the scales to video recordings of the procedures. In the clinical environment procedures and clinician assessments may be potentially affected by parental involvement, clinical information, other clinicians, available resources, clinician relationships with the infant or child and their parents, competing priories and other incidental factors such as clinical interruptions (537). Furthermore, the exploration of clinical utility of pain scores will provide an opportunity to better understand the clinical significance of the differences in scores across individuals and the change in scores over time within individuals.

Psychometric testing of a scale provides data to support the reliability and validity of the scores rather than the scale per se as the scores are a function of the performance of the scale applied in the circumstances under which it was tested (269). As has been articulated earlier these circumstances include patient related factors, e.g. their age, gender, temperament, past experience, culture: clinician related factors, e.g. their discipline, specialty, experience, education, personal demographics such as age, gender and culture and procedure related factors, although these are less well understood. As this was completed in a single centre in the emergency department similar studies should be repeated in other departments, other organisations and using a range of procedures to explore further the potential impact of the reviewer on application of the scales and therefore the psychometric properties of the scores. Increasing evidence of the reliability and validity of scores in a range of circumstances add to our confidence that the scale can be reliably and validly used to measure procedural pain in circumstances that approximate those in which it has been tested.

Confirming the capacity of the scales capacity to differentiate between pain and non-pain related distress is crucial to understanding the ways in which the scale can be used and the impact on clinical and research application of these scales based on current data has already been described. However, although this study has made a significant contribution to this question it is insufficient to be considered conclusive. The conduct of similar studies to this one to address the need for data generated under different circumstances is also an opportunity to include methods aimed at testing specificity.

Several concerns regarding the design of the FLACC scale and the MBPS were revealed by the assessment of the feasibility of application of these scales and closer inspection of the scores. Although it is tempting to suggest that all efforts should concentrate on revision of the FLACC scale given the results of the psychometric analysis, the potential value of revision of the MBPS should not be ignored. The ways in which both scales might be revised and the rationale for these potential changes has been discussed more extensively earlier but in summary the MBPS scale descriptors for face and cry items might benefit from inclusion of more neutral expressions and revision of the FLACC scale face item descriptors and review of the legs, activity and consolability items to improve this scale for procedural use should be considered. Revisions must be made based on empirical data and the internal consistency of the items of revised versions tested before the scale is subjected to closer scrutiny.

A revised scale would require psychometric and feasibility testing before it could be recommended for clinical or research use. Careful attention to the design of studies and data analysis strategies is required for any studies examining the psychometric properties of the existing scales or revised versions to ensure methodological rigour. For future studies we recommend using the COSMIN Checklist as a template for the design. The quality of methods used in many studies was variable with many studies using methods rated on the COSMIN checklist as ‘poor’ and ‘fair’ (342).

Finally, careful consideration should be given to undertaking work to identify behaviours consistently demonstrated by infants and young children in response to procedural pain to expand this evidence base. The results of this work may be used to support the revision of existing scales or to inform the development of new empirically derived procedural pain scale.

13.4.3 Dissemination and knowledge translation

Critical to research is consideration for how the findings are disseminated to ensure that they impact on practice. The results of the search for a potentially suitable pain scale highlight a conflict between available evidence and practice (538). The psychometric properties of the VASobs were summarised in a review in 2002 which concluded that there was insufficient data to recommend the scale for use. In 2007 von Baeyer and Spagrud’s review recommended the FLACC scale and CHEOPS for procedural pain assessment and there was no mention of the VASobs as a potential option (30). Despite these publications, the VASobs was used almost three times as often as the next most frequently used scale (FLACC scale) amongst the RCTs identified. This trend seems to have reversed in the years since this search was completed.

The decision to complete this thesis with publication was aimed at ensuring dissemination of the results of this work to the pain community. The results of the systematic reviews and the psychometric evaluation studies have been published in leading journals for the discipline and specialty: *Pain, The Journal of Pain* and *Journal of Pediatric Nursing* (332, 464, 469, 539). Furthermore, the protocol was published to provide a potential template for other researchers (468). Additionally, opportunities to present these results at international, national and local conferences and seminars have been sought. It should also be noted that there are plans to publish unpublished data from the thesis and any additional analysis that is completed. These are the traditional approaches to dissemination but in recent years it has been shown that there are considerable delays to achieving practice change in response to evidence and it has been suggested that traditional methods are insufficient to drive practice change (540). This may be more likely to come from changes to pain related guidelines and recommendations made on the basis of new evidence (541). Use of the evidence to inform education programs aimed at clinicians responsible for assessing and managing procedural pain may also generate change in practice (542).

There is a growing body of paediatric pain related work based on the principles of knowledge translation and implementation science and ChildKind International (543) leads much of this work (538). A systematic review of knowledge translation studies identified 60 studies that tested a knowledge translation strategy aimed at paediatric pain (538). Three quarters of the studies tested practice-level changes, which included changes to pain assessment procedures. Although many of these initiatives were geared towards increasing the frequency of assessments or documentation of pain scores, some focused on implementing a new assessment protocol. All studies demonstrated a positive change as a result of initiatives that included in-service and workshop training and resource intensive activities such as auditing and reporting of assessment practices, visual reminders in clinical areas and changes in forms and charts. On completion of this project and dissemination of the results via traditional channels, strategies to ensure that these results inform knowledge translation initiatives is vital. This may include local level guideline changes which is likely to have both local, national and international impact as the study hospital produces and published clinical practice guideline that are internationally accessed. Furthermore, the study hospital (The Royal Children's Hospital, Melbourne) is a member of a an Australian and New Zealand paediatric emergency research network (PREDICT (544)) which provides another vehicle for dissemination and collaboration, nationally and internationally.

13.5 Conclusion

This project was a robust attempt to identify an observational scale that could be recommended as suitable to assess procedural pain experienced by infants and young children. Of the many unique paediatric observational pain scales identified in the literature (480) very few were either designed for procedural pain assessment use, supported by psychometric evaluation data or universally accepted for this purpose. Three scales met pre-established criteria and were considered potentially suitable: the FLACC scale, the MBPS and the VASobs. Following systematic reviews of the available psychometric data for these scales it was concluded that these scales could not be recommended on the basis of the existing data. The results of the prospective study to examine the psychometric properties of the FLACC scale, the MBPS and the VASobs pain and the VASobs distress, confirmed the FLACC scale as the scale supported by the most convincing data. However, significant concerns regarding this scale's performance remained: namely the feasibility of applying this scale for medical procedures and the specificity of the scores for pain. The circumstances of a procedure create a potentially unique experience for infants and children that impact on their behavioural responses. It seems likely that observational scales, including the FLACC scale, measure a composite of pain and non-pain related distressed and should be considered distress rather than pain scales and that this should be considered when using these scales for clinical or research purposes and for interpreting scores.

There are also still considerable gaps in our understanding of pain assessment, specifically for infants and children undergoing diagnostic and therapeutic procedures. If we are to base treatment decisions and interpret research on assessments generated using these scales it is critical that these gaps are explored. It is vital that clinicians and researchers understand what the scale measures and the factors that influence these measurements. Scale revisions may make these scales better suited for procedural pain assessment. However, revised versions would again require extensive psychometric testing before they could be recommended for use.

REFERENCES

1. McMurtry CM, Pillai Riddell R, Taddio A, Racine N, Asmundson GJ, Noel M, et al. Far From "Just a Poke": Common Painful Needle Procedures and the Development of Needle Fear. Clinical Journal of Pain. 2015;31(10 Suppl):S3-11.
2. Johnston C, Barrington KJ, Taddio A, Carbajal R, Filion F. Pain in Canadian NICUs: have we improved over the past 12 years? Clinical Journal of Pain. 2011;27(3):225-32.
3. Simons SH, van Dijk M, Anand KS, Roofthooft D, van Lingen RA, Tibboel D. Do we still hurt newborn babies? A prospective study of procedural pain and analgesia in neonates. Archives of Pediatric and Adolescent Medicine. 2003;157(11):1058-64.
4. Harrison D, Joly C, Chretien C, Cochrane S, Ellis J, Lamontagne C, et al. Pain prevalence in a pediatric hospital: raising awareness during Pain Awareness Week. Pain Research and Management. 2014;19(1):e24-30.
5. Ortiz MI, López-Zarco M, Arreola-Bautista EJ. Procedural pain and anxiety in paediatric patients in a Mexican emergency department. Journal of Advanced Nursing. 2012;68(12):2700-9.
6. Stevens B, McGrath P, Gibbins S, Beyene J, Breau L, Camfield C, et al. Procedural pain in newborns at risk for neurologic impairment. Pain. 2003;105(1-2):27-35.
7. Carbajal R, Rousset A, Danan C, Coquery S, Nolent P, Ducrocq S, et al. Epidemiology and treatment of painful procedures in neonates in intensive care units. JAMA. 2008;300(1):60-70.
8. Stevens B, Abbott LK, Yamada J, Harrison D, Stinson J, Taddio A, et al. Epidemiology and management of painful procedures in children in Canadian hospitals. Canadian Medical Association Journal. 2011;183(7):E403-10.
9. Johnston CC, Collinge JM, Henderson SJ, Anand KJ. A cross-sectional survey of pain and pharmacological analgesia in Canadian neonatal intensive care units. Clinical Journal of Pain. 1997;13(4):308-12.
10. Taddio A, Katz J, Ilersich AL, Koren G. Effect of neonatal circumcision on pain response during subsequent routine vaccination. Lancet. 1997;349(9052):599-603.
11. Taddio A, Goldbach M, Ipp M, Stevens B, Koren G. Effect of neonatal circumcision on pain responses during vaccination in boys. Lancet. 1995;345(8945):291-2.
12. Taddio A, Shah V, Gilbert-MacLeod C, Katz J. Conditioning and hyperalgesia in newborns exposed to repeated heel lances. JAMA. 2002;288(7):857-61.
13. Valeri BO, Holsti L, Linhares MB. Neonatal pain and developmental outcomes in children born preterm: a systematic review. Clinical Journal of Pain. 2015;31(4):355-62.

14. Grunau RE. Neonatal pain in very preterm infants: long-term effects on brain, neurodevelopment and pain reactivity. Rambam Maimonides Medical Journal. 2013;4(4):e0025.

15. Schug S, Palmer G, Scott D, Halliwell R, Trinca J, Working Group of the Australian and New Zealand College of Anaesthetists (ANZCA) and Faculty of Pain Medicine (FPM). Acute Pain Management: Scientific Evidence. 4th ed. Melbourne: ANZCA & FPM; 2015.

16. Liu S, CL W. Neural blockade: impact on outcome. In: Cousins MJ, Carr DB, Horlocker TT, Bridenbaugh PO, editors. Neural Blockade in Clinical Anesthesia and Pain Medicine. 113. 4 ed. Philadelphia: Lippincott, Williams & Wilkins; 2008. p. 133–43.

17. Cardinale F, Chinellato I, Caimmi S, Peroni DG, Franceschini F, Miraglia Del Giudice M, et al. Perioperative period: immunological modifications. International Journal of Immunopathology and Pharmacology. 2011;24(3 Suppl):S3-12.

18. Blackburn GL. Metabolic considerations in management of surgical patients. Surgical Clinics of North America. 2011;91(3):467-80.

19. Prabhakar A, Mancuso KF, Owen CP, Lissauer J, Merritt CK, Urman RD, et al. Perioperative analgesia outcomes and strategies. Best Practice in Research and Clinical Anaesthesiology. 2014;28(2):105-15.

20. Carli F, Schricker T. Modification of Metabolic Response to Surgery by Neural Blockade. In: Cousins MJ, Carr DB, Horlocker TT, Bridenbaugh PO, editors. Neural Blockade in Clinical Anesthesia and Pain Medicine. 113. 4 ed. Philadelphia: Lippincott, Williams & Wilkins; 2008.

21. Newman S, Stygall J, Hirani S, Shaefi S, Maze M. Postoperative cognitive dysfunction after noncardiac surgery: a systematic review. Anesthesiology. 2007;106(3):572-90.

22. Monk TG, Weldon BC, Garvan CW, Dede DE, van der Aa MT, Heilman KM, et al. Predictors of cognitive dysfunction after major noncardiac surgery. Anesthesiology. 2008;108(1):18-30.

23. Cousins MJ, Brennan F, Carr DB. Pain relief: a universal human right. Pain. 2004;112(1-2):1-4.

24. Young Casey C, Greenberg MA, Nicassio PM, Harpin RE, Hubbard D. Transition from acute to chronic pain and disability: a model including cognitive, affective, and trauma factors. Pain. 2008;134(1-2):69-79.

25. Jenewein J, Moergeli H, Wittmann L, Buchi S, Kraemer B, Schnyder U. Development of chronic pain following severe accidental injury. Results of a 3-year follow-up study. Journal of psychosomatic research. 2009;66(2):119-26.

26. Williamson OD, Epi GD, Gabbe BJ, Physio B, Cameron PA, Edwards ER, et al. Predictors of moderate or severe pain 6 months after orthopaedic injury: a prospective cohort study. Journal of Orthopedic Trauma. 2009;23(2):139-44.

27. Rennick JE, Morin I, Kim D, Johnston CC, Dougherty G, Platt R. Identifying children at high risk for psychological sequelae after pediatric intensive care unit hospitalization. Pediatric Critical Care Medicine. 2004;5(4):358-63.

28. Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S. The FLACC: a behavioral scale for scoring postoperative pain in young children. Pediatric Nursing. 1997;23(3):293-7.

29. McGrath PJ, Johnson G, Goodman JT, Schillinger J, Dunn J, Chapman J. CHEOPS: A Behavioural Scale for Rating Postoperative Pain in Children. In: Fields HL, Dubner R, Cervero F, editors. Advances in Pain Research and Therapy. 9. New York: Raven Press; 1985. p. 395-402.

30. von Baeyer CL, Spagrud LJ. Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. Pain. 2007;127(1-2):150.

31. Crellin DJ, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. Paediatric Anaesthesia. 2007;17(8):720-33.

32. Royal College of Nursing. The recognition and assessment of acute pain in children. Update of full guideline. London: Royal College of Nursing; 2009.

33. International Association for the Study of Pain. IASP Pain Taxonomy 2017 [updated 14 December 2017. Available from: http://www.iasp-pain.org/AM/Template.cfm?Section=General_Resource_Links&Template=/CM/HTMLDisplay.cfm&ContentID=3058#Pain.

34. Melzack R, Katz J. Pain. Wiley interdisciplinary reviews Cognitive science. 2013;4(1):1-15.

35. Loeser JD, Treede R-D. The Kyoto protocol of IASP Basic Pain Terminology. Pain. 2008;137(3):473-7.

36. Walker SM, Baccei ML. Nociceptive signalling in the periphery and spinal cord. In: McGrath PJ, Stevens BJ, Walker SM, Zempsky WT, editors. Oxford Textbook of Paediatric Pain. Oxford: Oxford University Press; 2015. p. 53-64.

37. Dubin AE, Patapoutian A. Nociceptors: the sensors of the pain pathway. The Journal of Clinical Investigation. 2010;120(11):3760-72.

38. Gold MS. Molecular Biology of Sensory Transduction. In: McMahon SB, Koltzenburg M, Tracey IMA, Turk DC, editors. Wall & Melzack's Textbook of Pain. Philadelphia: Saunders; 2013. p. 31-47.

39. Ringkamp M, Raja SN, Campbell JN, Meyer RA. Peripheral Mechanisms of Cutaneous Nociception. In: McMahon SB, Koltzenburg M, Tracey IMA, Turk DC, editors. Wall & Melzack's Textbook of Pain. Philadelphia: Saunders; 2013. p. 1-30.

40. Ossipov MH. The Perception and Endogenous Modulation of Pain. Scientifica. 2012;2012:25.

41. Todd AJ, Koerber HR. Neuroanatomical Substrates of Spinal Nociception. In: McMahon SB, Koltzenburg M, Tracey IMA, Turk DC, editors. Wall & Melzack's Textbook of Pain. Philadelphia: Saunders; 2013. p. 77-93.

42. Purves D, Augustine G, Fitzpatrick D, et al., editors. Central Pain Pathways: The Spinothalamic Tract. 2001. In: Neuroscience [Internet]. Sunderland (MA): Sinauer Associates. Available from: https://www.ncbi.nlm.nih.gov/books/NBK10967/.

43. Dostrovsky JO, Craig AD. Ascending Projection Systems. In: McMahon SB, Koltzenburg M, Tracey IMA, Turk DC, editors. Wall & Melzack's Textbook of Pain. Philadelphia: Saunders; 2013. p. 182-97.

44. Ossipov MH, Dussor GO, Porreca F. Central modulation of pain. Journal of Clinical Investigation. 2010;120(11):3779-87.

45. Apkarian AV, Bushnell MC, Schweinhardt P. Representation of Pain in the Brain. In: McMahon SB, Koltzenburg M, Tracey IMA, Turk DC, editors. Wall & Melzack's Textbook of Pain. Philadelphia: Saunders; 2013. p. 111-28.

46. Steeds CE. The anatomy and physiology of pain. Surgery. 2009;27(12):507-11.

47. Arendt-Nielsen L, Svensson P. Referred Muscle Pain: Basic and Clinical Findings. Clinical Journal of Pain. 2001;17(1):11-9.

48. Morton DL, Sandhu JS, Jones AK. Brain imaging of pain: state of the art. Journal of Pain Research. 2016;9:613-24.

49. Ploner M, Freund HJ, Schnitzler A. Pain affect without pain sensation in a patient with a postcentral lesion. Pain. 1999;81(1-2):211-4.

50. Foltz EL, White LE, Jr. Pain "relief" by frontal cingulumotomy. Journal of Neurosurgery. 1962;19:89-100.

51. Heinricher MM, Fields HL. Central Nervous System Mechanisms of Pain Modulation. In: McMahon SB, Koltzenburg M, Tracey IMA, Turk DC, editors. Wall & Melzack's Textbook of Pain. Philadelphia: Saunders; 2013. p. 129-42.

52. Dawes JM, Andersson DA, Bennett DLH, Bevan S, McMahon SB. Inflammatory Mediators and Modulators of Pain. In: McMahon SB, Koltzenburg M, Tracey IMA, Turk DC, editors. Wall & Melzack's Textbook of Pain. Philadelphia: Saunders; 2013. p. 48-67.

53. Gold MS, Gebhart GF. Nociceptor sensitization in pain pathogenesis. Nature Medicine. 2010;16(11):1248-57.

54. Melzack R. The Perception of Pain. Scientific American. 1961;204(2):41-9.

55. McGrath PA. Psychological aspects of pain perception. Archives of Oral Biology. 1994;39 Suppl:55S-62S.

56. Wiech K, Ploner M, Tracey I. Neurocognitive aspects of pain perception. Trends in Cognitive Sciences. 2008;12(8):306-13.

57. Posner MI, Petersen SE. The Attention System of the Human Brain. Annual Review of Neuroscience. 1990;13(1):25-42.

58. Garland EL. Pain Processing in the Human Nervous System: A Selective Review of Nociceptive and Biobehavioral Pathways. Primary Care. 2012;39(3):561-71.

59. Hansen GR, Streltzer J. The psychology of pain. Emergency Medicine Clinics of North America. 2005;23(2):339-48.

60. Bantick SJ, Wise RG, Ploghaus A, Clare S, Smith SM, Tracey I. Imaging how attention modulates pain in humans using functional MRI. Brain: A Journal of Neurology. 2002;125(2):310-9.

61. Tracey I, Ploghaus A, Gati JS, Clare S, Smith S, Menon RS, et al. Imaging attentional modulation of pain in the periaqueductal gray in humans. Journal of Neuroscience. 2002;22(7):2748-52.

62. Dunckley P, Aziz Q, Wise RG, Brooks J, Tracey I, Chang L. Attentional modulation of visceral and somatic pain. Neurogastroenterology and Motility. 2007;19(7):569-77.

63. Piira T, Hayes B, Goodenough B, von Baeyer CL. Effects of attentional direction, age, and coping style on cold-pressor pain in children. Behaviour Research and Therapy. 2006;44(6):835-48.

64. Dahlquist LM, Weiss KE, Law EF, Sil S, Herbert LJ, Horn SB, et al. Effects of videogame distraction and a virtual reality type head-mounted display helmet on cold pressor pain in young elementary school-aged children. Journal of Pediatric Psychology. 2010;35(6):617-25.

65. Ploner M, Lee MC, Wiech K, Bingel U, Tracey I. Flexible cerebral connectivity patterns subserve contextual modulations of pain. Cerebral Cortex. 2011;21(3):719-26.

66. Bellieni CV, Tei M, Coccina F, Buonocore G. Sensorial saturation for infants' pain. Journal of Maternal-Fetal and Neonatal Medicine. 2012;25 Suppl 1:79-81.

67. Ozdemir FK, Tufekci FG. The effect of using musical mobiles on reducing pain in infants during vaccination. Journal of Research in Medical Sciences. 2012;17(7):662-7.

68. Villemure C, Bushnell CM. Cognitive modulation of pain: how do attention and emotion influence pain processing? Pain. 2002;95(3):195-9.

69. Upadhyay J, Granitzka J, Bauermann T, Baumgartner U, Breimhorst M, Treede RD, et al. Detection of central circuits implicated in the formation of novel pain memories. Journal of Pain Research. 2016;9:671-81.

70. Noel M, Chambers CT, Petter M, McGrath PJ, Klein RM, Stewart SH. Pain is not over when the needle ends: a review and preliminary model of acute pain memory development in childhood. Pain Management. 2012;2(5):487-97.

71. Fairhurst M, Wiech K, Dunckley P, Tracey I. Anticipatory brainstem activity predicts neural processing of pain in humans. Pain. 2007;128(1-2):101-10.

72. Keltner JR, Furst A, Fan C, Redfern R, Inglis B, Fields HL. Isolating the modulatory effect of expectation on pain transmission: a functional magnetic resonance imaging study. Journal of Neuroscience. 2006;26(16):4437-43.

73. Porro CA, Baraldi P, Pagnoni G, Serafini M, Facchin P, Maieron M, et al. Does anticipation of pain affect cortical nociceptive systems? Journal of Neuroscience. 2002;22(8):3206-14.

74. Linton SJ, Shaw WS. Impact of Psychological Factors in the Experience of Pain. Physical Therapy. 2011;91(5):700-11.

75. Koyama T, McHaffie JG, Laurienti PJ, Coghill RC. The subjective experience of pain: where expectations become reality. Proceedings of the National Academy of Sciences of the United States of America 2005;102(36):12950-5.

76. Salomons TV, Johnstone T, Backonja MM, Davidson RJ. Perceived controllability modulates the neural response to pain. Journal of Neuroscience. 2004;24(32):7199-203.

77. Salomons TV, Johnstone T, Backonja MM, Shackman AJ, Davidson RJ. Individual differences in the effects of perceived controllability on pain perception: critical role of the prefrontal cortex. Journal of Cognitive Neuroscience. 2007;19(6):993-1003.

78. Wiech K, Tracey I. Pain, decisions, and actions: a motivational perspective. Frontal Neuroscience. 2013;7:46.

79. Piira T, Taplin JE, Goodenough B, von Baeyer CL. Cognitive-behavioural predictors of children's tolerance of laboratory-induced pain: implications for clinical assessment and future directions. Behaviour Research and Therapy. 2002;40(5):571-84.

80. Salomons TV, Moayedi M, Erpelding N, Davis KD. A brief cognitive-behavioural intervention for pain reduces secondary hyperalgesia. Pain. 2014;155(8):1446-52.

81. Verhoeven K, Goubert L, Jaaniste T, Van Ryckeghem D, Crombez G. Pain catastrophizing influences the use and the effectiveness of distraction in school children. European Journal of Pain. 2012;16(2):256-67.

82. Williams SE, Blount RL, Walker LS. Children's pain threat appraisal and catastrophizing moderate the impact of parent verbal behavior on children's symptom complaints. Journal of Pediatric Psychology. 2011;36(1):55-63.

83. Vervoort T, Eccleston C, Goubert L, Buysse A, Crombez G. Children's catastrophic thinking about their pain predicts pain and disability 6 months later. European Journal of Pain. 2010;14(1):90-6.

84. Lu Q, Tsao JC, Myers CD, Kim SC, Zeltzer LK. Coping predictors of children's laboratory-induced pain tolerance, intensity, and unpleasantness. Journal of Pain. 2007;8(9):708-17.

85. Salomons TV, Nusslock R, Detloff A, Johnstone T, Davidson RJ. Neural emotion regulation circuitry underlying anxiolytic effects of perceived control over pain. Journal of Cognitive Neuroscience. 2015;27(2):222-33.

86. Benarroch EE. Pain-autonomic interactions. Neurological Sciences. 2006;27(2):s130-s3.

87. Jänig W, Levine JD. Autonomic, Endocrine, and Immune Interactions in Acute and Chronic Pain. In: McMahon SBF, Koltzenburg MMD, Tracey IMA, Turk DC, editors. Wall & Melzack's Textbook of Pain. Philadelphia: Saunders; 2013. p. 198-210.

88. Tennant F. The Physiologic Effects of Pain on the Endocrine System. Pain and Therapy. 2013;2(2):75-86.

89. May ES, Tiemann L, Schmidt P, Nickel MM, Wiedemann N, Dresel C, et al. Behavioral responses to noxious stimuli shape the perception of pain. Scientific Reports. 2017;7:44083.

90. Purves D, Augustine G, Fitzpatrick D, et al, [editors]. Flexion Reflex Pathways. 2001. In: Neuroscience [Internet]. Sunderland (MA): Sinauer Associates. 2nd. Available from: https://www.ncbi.nlm.nih.gov/books/NBK11091/.

91. Melzack R, Wall PD. Pain mechanisms: a new theory. Science. 1965;150(3699):971-9.

92. Hadjistavropoulos T, Craig KD. A theoretical framework for understanding self-report and observational measures of pain: a communications model. Behaviour Research and Therapy. 2002;40(5):551-70.

93. Hadjistavropoulos T, Craig KD, Duck S, Cano A, Goubert L, Jackson PL, et al. A biopsychosocial formulation of pain communication. Psychological Bulletin. 2011;137(6):910-39.

94. Melzack R. Pain and the neuromatrix in the brain. Journal of Dental Education. 2001;65(12):1378-82.

95. Iannetti GD, Mouraux A. From the neuromatrix to the pain matrix (and back). Experimental Brain Research. 2010;205(1):1-12.

96. Walker SM. Long-term effects of early pain and injury: animal models. In: McGrath PJ, Stevens BJ, Walker SM, Zempsky WT, editors. Oxford Textbook of Paediatric Pain. Oxford: Oxford University Press; 2015. p. 20-9.

97. Fabrizi L, Slater R, Worley A, Meek J, Boyd S, Olhede S, et al. A shift in sensory processing that enables the developing human brain to discriminate touch from pain. Current Biology. 2011;21(18):1552-8.

98. Fitzgerald M, Walker SM. Infant pain management: a developmental neurobiological approach. Nature Clinical Practice Neurology. 2009;5(1):35-50.

99. Grunau RE. Long-term effects of pain in children. In: McGrath PJ, Stevens BJ, Walker SM, Zempsky WT, editors. Oxford Textbook of Paediatric Pain. Oxford: Oxford University Press; 2015. p. 30-8.

100. Hohmeister J, Kroll A, Wollgarten-Hadamek I, Zohsel K, Demirakca S, Flor H, et al. Cerebral processing of pain in school-aged children with neonatal nociceptive input: an exploratory fMRI study. Pain. 2010;150(2):257-67.

101. Fitzgerald M, Beggs S. The neurobiology of pain: developmental aspects. Neuroscientist. 2001;7(3):246-57.

102. Fabrizi L, Worley A, Patten D, Holdridge S, Cornelissen L, Meek J, et al. Electrophysiological measurements and analysis of nociception in human infants. Journal of Visualized Experiments. 2011;58.

103. Slater R, Worley A, Fabrizi L, Roberts S, Meek J, Boyd S, et al. Evoked potentials generated by noxious stimulation in the human infant brain. European Journal of Pain. 2010;14(3):321-6.

104. Holsti L, Grunau RE, Oberlander TF, Whitfield MF. Prior pain induces heightened motor responses during clustered care in preterm infants in the NICU. Early Human Development. 2005;81(3):293-302.

105. Holsti L, Grunau RE, Whifield MF, Oberlander TF, Lindh V. Behavioral responses to pain are heightened after clustered care in preterm infants born between 30 and 32 weeks gestational age. Clinical Journal of Pain. 2006;22(9):757-64.

106. Fitzgerald M. What do we really know about newborn infant pain? Experimental Physiology. 2015;100(12):1451-7.

107. Schiavenato M, Byers JF, Scovanner P, McMahon JM, Xia Y, Lu N, et al. Neonatal pain facial expression: evaluating the primal face of pain. Pain. 2008;138(2):460-71.

108. Prkachin KM. Assessing pain by facial expression: facial expression as nexus. Pain Research and Management. 2009;14(1):53-8.

109. Larochette A-C, Chambers CT, Craig KD. Genuine, suppressed and faked facial expressions of pain in children. Pain. 2006;126(1-3):64-71.

110. Bartocci M, Bergqvist LL, Lagercrantz H, Anand KJ. Pain activates cortical areas in the preterm newborn brain. Pain. 2006;122(1-2):109-17.

111. Slater R, Cantarella A, Gallella S, Worley A, Boyd S, Meek J, et al. Cortical pain responses in human infants. Journal of Neuroscience. 2006;26(14):3662-6.

112. Goksan S, Hartley C, Emery F, Cockrill N, Poorun R, Moultrie F, et al. fMRI reveals neural activity overlap between adult and infant pain. Elife. 2015;4.

113. Wollgarten-Hadamek I, Hohmeister J, Zohsel K, Flor H, Hermann C. Do school-aged children with burn injuries during infancy show stress-induced activation of pain inhibitory mechanisms? European Journal of Pain. 2011;15(4):423.e1-10.

114. Bonica JJ. Introduction to symposium on pain. Archives of Surgery. 1977;112(6):749.

115. IASP Taxonomy Working Group. Classification of Chronic Pain, Second Edition (Revised) 2011. Available from: https://www.iasp-pain.org/PublicationsNews/Content.aspx?ItemNumber=1673.

116. Hjermstad MJ, Fainsinger R, Kaasa S. Assessment and classification of cancer pain. Current opinion in supportive and palliative care. 2009;3(1):24-30.

117. Melzack R. Pain measurement and assessment. Michigan: Raven Press; 1983.

118. Huskisson EC. Measurement of Pain. Lancet. 1974;304(7889):1127-31.

119. Twycross A, Voepel-Lewis T, Vincent C, Franck LS, von Baeyer CL. A Debate on the Proposition that Self-report is the Gold Standard in Assessment of Pediatric Pain Intensity. Clinical Journal of Pain. 2015;31(8):707-12.

120. Cohen LL, Lemanek K, Blount RL, Dahlquist LM, Lim CS, Palermo TM, et al. Evidence-based assessment of pediatric pain. Journal of Pediatric Psychology. 2008;33(9):939-55; discussion 56-7.

121. Tomlinson D, von Baeyer CL, Stinson JN, Sung L. A systematic review of faces scales for the self-report of pain intensity in children. Pediatrics. 2010;126(5):e1168-98.

122. McGrath PJ, Cunningham SJ, Goodman JT, Unruh A. The Clinical Measurement of Pain in Children: A Review. Clinical Journal of Pain. 1985;1(4):221-8.

123. McGrath PA. An assessment of children's pain: a review of behavioral, physiological and direct scaling techniques. Pain. 1987;31(2):147-76.

124. Stinson JN, Kavanagh T, Yamada J, Gill N, Stevens B. Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. Pain. 2006;125(1-2):143-57.

125. von Baeyer CL, Chambers CT, Forsyth SJ, Eisen S, Parker JA. Developmental data supporting simplification of self-report pain scales for preschool-age children. Journal of Pain. 2013;14(10):1116-21.

126. Kakigi R, Inui K, Tamura Y. Electrophysiological studies on human pain perception. Clinical Neurophysiology. 2005;116(4):743-63.

127. Lorenz J, Garcia-Larrea L. Contribution of attentional and cognitive factors to laser evoked brain potentials. Neurophysiologie Clinique. 2003;33(6):293-301.

128. Franz M, Nickel MM, Ritter A, Miltner WHR, Weiss T. Somatosensory spatial attention modulates amplitudes, latencies, and latency jitter of laser-evoked brain potentials. Journal of Neurophysiology. 2015;113(7):2760-8.

129. Legrain V, Guerit JM, Bruyer R, Plaghki L. Attentional modulation of the nociceptive processing into the human brain: selective spatial attention, probability of stimulus occurrence, and target detection effects on laser evoked potentials. Pain. 2002;99(1-2):21-39.

130. Tiemann L, May ES, Postorino M, Schulz E, Nickel MM, Bingel U, et al. Differential neurophysiological correlates of bottom-up and top-down modulations of pain. Pain. 2015;156(2):289-96.

131. Wager TD, Matre D, Casey KL. Placebo effects in laser-evoked pain potentials. Brain, Behaviour and Immunity. 2006;20(3):219-30.

132. Fernandez M, Blass EM, Hernandez-Reif M, Field T, Diego M, Sanders C. Sucrose attenuates a negative electroencephalographic response to an aversive stimulus for newborns. Journal of developmental and behavioral pediatrics : JDBP. 2003;24(4):261-6.

133. Norman E, Rosen I, Vanhatalo S, Stjernqvist K, Okland O, Fellman V, et al. Electroencephalographic response to procedural pain in healthy term newborn infants. Pediatric Research. 2008;64(4):429-34.

134. Worley A, Fabrizi L, Boyd S, Slater R. Multi-modal pain measurements in infants. Journal of Neuroscience Methods. 2012;205(2):252-7.

135. Villringer A, Planck J, Hock C, Schleinkofer L, Dirnagl U. Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neuroscience Letters. 1993;154(1-2):101-4.

136. Bucher HU, Moser T, von Siebenthal K, Keel M, Wolf M, Duc G. Sucrose reduces pain reaction to heel lancing in preterm infants: a placebo-controlled, randomized and masked study. Pediatric Research. 1995;38(3):332-5.

137. Slater R, Cantarella A, Franck L, Meek J, Fitzgerald M. How well do clinical pain assessment tools reflect pain in infants? PLoS Medicine 2008;5(6):e129.

138. Stevens B, Yamada J, Ohlsson A, Haliburton S, Shorkey A. Sucrose for analgesia in newborn infants undergoing painful procedures. Cochrane Database of Systematic Reviews. 2016;7:Cd001069.

139. Ranger M, Celeste Johnston C, Rennick JE, Limperopoulos C, Heldt T, du Plessis AJ. A Multidimensional Approach to Pain Assessment in Critically Ill Infants During a Painful Procedure. Clinical Journal of Pain. 2013;29(7):613-20.

140. Bembich S, Brovedani P, Cont G, Travan L, Grassi V, Demarini S. Pain activates a defined area of the somatosensory and motor cortex in newborn infants. Acta Paediatrica. 2015;104(11):e530-3.

141. Ozawa M, Kanda K, Hirata M, Kusakawa I, Suzuki C. Influence of repeated painful procedures on prefrontal cortical pain responses in newborns. Acta Paediatrica. 2011;100(2):198-203.

142. Bembich S, Davanzo R, Brovedani P, Clarici A, Massaccesi S, Demarini S. Functional neuroimaging of breastfeeding analgesia by multichannel near-infrared spectroscopy. Neonatology. 2013;104(4):255-9.

143. Gerull R, Cignacco E, Stoffel L, Sellam G, Nelle M. Physiological parameters after nonpharmacological analgesia in preterm infants: a randomized trial. Acta Paediatrica. 2013;102(8):e368-73.

144. Azar AA. Monitoring blood oxygenation changes due to acute pain stimuli using functional near-infrared spectroscopy (fNIRS). Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2009;2009:1375-9.

145. Gelinas C, Choiniere M, Ranger M, Denault A, Deschamps A, Johnston C. Toward a new approach for the detection of pain in adult patients undergoing cardiac surgery: near-infrared spectroscopy--a pilot study. Heart and Lung. 2010;39(6):485-93.

146. Hartley C, Poorun R, Goksan S, Worley A, Boyd S, Rogers R, et al. Noxious stimulation in children receiving general anaesthesia evokes an increase in delta frequency brain activity. Pain. 2014;155(11):2368-76.

147. Hartley C, Slater R. Neurophysiological measures of nociceptive brain activity in the newborn infant--the next steps. Acta Paediatrica. 2014;103(3):238-42.

148. Benoit B, Martin-Misener R, Newman A, Latimer M, Campbell-Yeo M. Neurophysiological assessment of acute pain in infants: a scoping review of research methods. Acta Paediatrica. 2017;106(7):1053-66.

149. Favilla S, Huber A, Pagnoni G, Lui F, Facchin P, Cocchi M, et al. Ranking brain areas encoding the perceived level of pain from fMRI data. Neuroimage. 2014;90:153-62.

150. Johnstone T, Salomons TV, Backonja MM, Davidson RJ. Turning on the alarm: the neural mechanisms of the transition from innocuous to painful sensation. Neuroimage. 2012;59(2):1594-601.

151. Kunz M, Chen JI, Lautenbacher S, Vachon-Presseau E, Rainville P. Cerebral regulation of facial expressions of pain. Journal of Neuroscience. 2011;31(24):8730-8.

152. Liang M, Mouraux A, Iannetti GD. Parallel processing of nociceptive and non-nociceptive somatosensory information in the human primary and secondary somatosensory cortices: evidence from dynamic causal modeling of functional magnetic resonance imaging data. Journal of Neuroscience. 2011;31(24):8976-85.

153. Williams G, Fabrizi L, Meek J, Jackson D, Tracey I, Robertson N, et al. Functional magnetic resonance imaging can be used to explore tactile and nociceptive processing in the infant brain. Acta Paediatrica. 2015;104(2):158-66.

154. Salomons TV, Iannetti GD, Liang M, Wood JN. The "Pain Matrix" in Pain-Free Individuals. JAMA Neurology. 2016;73(6):755-6.

155. Raeside L. Physiological measures of assessing infant pain: a literature review. British Journal of Nursing. 2011;20(21):1370-6.

156. Gibbins S, Stevens B, Beyene J, Chan PC, Bagg M, Asztalos E. Pain behaviours in Extremely Low Gestational Age infants. Early Human Development. 2008;84(7):451-8.

157. Guinsburg R, Kopelman BI, Anand KJ, de Almeida MF, Peres CdA, Miyoshi MH. Physiological, hormonal, and behavioral responses to a single fentanyl dose in intubated and ventilated preterm neonates. Journal of Pediatrics. 1998;132(6):954-9.

158. Johnston CC, Stevens B, Yang F, Horton L. Developmental changes in response to heelstick in preterm infants: a prospective cohort study. Developmental Medicine and Child Neurology. 1996;38(5):438-45.

159. Lindh V, Wiklund U, Sandman PO, Hakansson S. Assessment of acute pain in preterm infants by evaluation of facial expression and frequency domain analysis of heart rate variability. Early Human Development. 1997;48(1-2):131-42.

160. Morison SJ, Holsti L, Grunau RE, Whitfield MF, Oberlander TF, Chan HW, et al. Are there developmentally distinct motor indicators of pain in preterm infants? Early Human Development. 2003;72(2):131-46.

161. Ranger M. Near-Infrared Spectroscopy (NIRS) to measure nociception following noxious stimulation in critically ill infants: McGill University (Canada); 2012.

162. Taksande AM, Vilhekar KY, Jain M, Chitre D. Pain response of neonates to venipuncture. Indian Journal of Pediatrics. 2005;72(9):751-3.

163. Van Cleve L, Johnson L, Pothier P. Pain responses of hospitalized infants and children to venipuncture and intravenous cannulation. Journal of Pediatric Nursing. 1996;11(3):161-8.

164. van Dijk M, de Boer JB, Koot HM, Duivenvoorden HJ, Passchier J, Bouwmeester N, et al. The association between physiological and behavioral pain measures in 0- to 3-year-old infants after major surgery. Journal of Pain and Symptom Management. 2001;22(1):600-9.

165. Bauer K, Ketteler J, Hellwig M, Laurenz M, Versmold H. Oral glucose before venepuncture relieves neonates of pain, but stress is still evidenced by increase in oxygen consumption, energy expenditure, and heart rate. Pediatric Research. 2004;55(4):695-700.

166. Bossart P, Fosnocht D, Swanson E. Changes in heart rate do not correlate with changes in pain intensity in emergency department patients. Journal of Emergency Medicine. 2007;32(1):19-22.

167. Buttner W, Finke W. Analysis of behavioural and physiological parameters for the assessment of postoperative analgesic demand in newborns, infants and young children: a comprehensive report on seven consecutive studies. Paediatric Anaesthesia. 2000;10(3):303-18.

168. Gaspardo CM, Chimello JT, Cugler TS, Martinez FE, Linhares MBM. Pain and tactile stimuli during arterial puncture in preterm neonates. Pain. 2008;140(1):58-64.

169. Harrison D, Evans C, Johnston L, Loughnan P. Bedside assessment of heel lance pain in the hospitalized infant. Journal of Obstetric, Gynecologic and Neonatal Nursing. 2002;31(5):551-7.

170. Pokela ML. Pain relief can reduce hypoxemia in distressed neonates during routine treatment procedures. Pediatrics. 1994;93(3):379-83.

171. Tousignant-Laflamme Y, Rainville P, Marchand S. Establishing a link between heart rate and pain in healthy subjects: a gender effect. Journal of Pain. 2005;6(6):341-7.

172. Stevens BJ, Pillai Riddell RR, Oberlander TE, Gibbins S. Assessment of Pain in Neonates and Infants. In: Anand KJ, J. SB, McGrath PJ, editors. Pain in Neonates and Infants. Toronto: Elsevier; 2007. p. 67-90.

173. van Dijk M, de Boer JB, Koot HM, Tibboel D, Passchier J, Duivenvoorden HJ. The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. Pain. 2000;84(2-3):367-77.

174. Ista E, van Dijk M, Tibboel D, de Hoog M. Assessment of sedation levels in pediatric intensive care patients can be improved by using the COMFORT "behavior" scale. Pediatric Critical Care Medicine. 2005;6(1):58-63.

175. Arif-Rahu M, Fisher D, Matsuda Y. Biobehavioral Measures for Pain in the Pediatric Patient. Pain Management Nursing. 2012;13(3):157-68.

176. Hadjian A, Chedin M, Cochet C, E. C. Cortisol binding to proteins in plasma in the human neonate and infant. Pediatric Research. 1975;9:40-5.

177. Grunau RE, Tu MT, Whitfield MF, Oberlander TF, Weinberg J, Yu W, et al. Cortisol, behavior, and heart rate reactivity to immunization pain at 4 months corrected age in infants born very preterm. Clinical Journal of Pain. 2010;26(8):698-704.

178. Hanson D, Hall W, Mills LL, Au S, Bhagat R, Hernandez M, et al. Comparison of distress and pain in infants randomized to groups receiving standard versus multiple immunizations. Infant Behavior and Development. 2010;33(3):289-96.

179. Morelius E, Theodorsson E, Nelson N. Stress at three-month immunization: parents' and infants' salivary cortisol response in relation to the use of pacifier and oral glucose. European Journal of Pain. 2009;13(2):202-8.

180. Franck LS, Ridout D, Howard R, Peters J, Honour JW. A comparison of pain measures in newborn infants after cardiac surgery. Pain. 2011;152(8):1758-65.

181. Wennstrom B, Tornhage C-J, Nasic S, Hedelin H, Bergh I. The perioperative dialogue reduces postoperative stress in children undergoing day surgery as confirmed by salivary cortisol. Paediatric Anaesthesia. 2011;21(10):1058-65.

182. Zhao X, Chen M, Du S, Li H, Li X. Evaluation of stress and pain in young children with cerebral palsy during early developmental intervention programs: a descriptive study. American Journal of Physical Medicine and Rehabilitation. 2015;94(3):169-75.

183. Vagnoli L, Caprilli S, Vernucci C, Zagni S, Mugnai F, Messeri A. Can Presence of a Dog Reduce Pain and Distress in Children during Venipuncture? Pain Management Nursing. 2015;16(2):89-95.

184. Ha YO, Kim HS. The effects of audiovisual distraction on children's pain during laceration repair. International Journal of Nursing Practice. 2013;19 Suppl 3:20-7.

185. Cignacco E, Denhaerynck K, Nelle M, Bührer C, Engberg S. Variability in pain response to a non-pharmacological intervention across repeated routine pain exposure in preterm infants: a feasibility study. Acta Paediatrica. 2009;98(5):842-6.

186. Grunau RE, Holsti L, Haley DW, Oberlander T, Weinberg J, Solimano A, et al. Neonatal procedural pain exposure predicts lower cortisol and behavioral reactivity in preterm infants in the NICU. Pain. 2005;113(3):293-300.

187. Grunau RE, Haley DW, Whitfield MF, Weinberg J, Yu W, Thiessen P. Altered basal cortisol levels at 3, 6, 8 and 18 months in infants born at extremely low gestational age. Journal of Pediatrics. 2007;150(2):151-6.

188. Storm H, Fremming A, Odegaard S, Martinsen OG, Morkrid L. The development of a software program for analyzing spontaneous and externally elicited skin conductance changes in infants and adults. Clinical Neurophysiology. 2000;111(10):1889-98.

189. Harrison D, Boyce S, Loughnan P, Dargaville P, Storm H, Johnston L. Skin conductance as a measure of pain and stress in hospitalised infants. Early Human Development. 2006;82(9):603-8.

190. Dalal PG, Doheny KK, Klick L, Britcher S, Rebstock S, Bezinover D, et al. Analysis of acute pain scores and skin conductance measurements in infants. Early Human Development. 2013;89(3):153-8.

191. Ahmed M, Sobithadevi D, Mostafa S, Pore P, Maiden J, Boswell S, et al. Pain evaluation in preterm infants using skin conductance algesimeter. Archives of Disease in Childhood. 2012;97:A457.

192. Tristao RM, Garcia NV, de Jesus JA, Tomaz C. COMFORT behaviour scale and skin conductance activity: what are they really measuring? Acta Paediatrica. 2013;102(9):e402-6.

193. Breimhorst M, Sandrock S, Fechir M, Hausenblas N, Geber C, Birklein F. Do intensity ratings and skin conductance responses reliably discriminate between different stimulus intensities in experimentally induced pain? Journal of Pain. 2011;12(1):61-70.

194. Choo EK, Magruder W, Montgomery CJ, Lim J, Brant R, Ansermino JM. Skin conductance fluctuations correlate poorly with postoperative self-report pain measures in school-aged children. Anesthesiology. 2010;113(1):175-82.

195. Gjerstad AC, Wagner K, Henrichsen T, Storm H. Skin conductance versus the modified COMFORT sedation score as a measure of discomfort in artificially ventilated children. Pediatrics. 2008;122(4):e848-53.

196. Hullett B, Chambers N, Preuss J, Zamudio I, Lange J, Pascoe E, et al. Monitoring electrical skin conductance: a tool for the assessment of postoperative pain in children? Anesthesiology. 2009;111(3):513-7.

197. Ledowski T, Pascoe E, Ang B, Schmarbeck T, Clarke MW, Fuller C, et al. Monitoring of intra-operative nociception: skin conductance and surgical stress index versus stress hormone plasma levels. Anaesthesia. 2010;65(10):1001-6.

198. Loggia ML, Juneau M, Bushnell MC. Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. Pain. 2011;152(3):592-8.

199. Munsters J, Wallstrom L, Agren J, Norsted T, Sindelar R. Skin conductance measurements as pain assessment in newborn infants born at 22-27 weeks gestational age at different postnatal age. Early Human Development. 2012;88(1):21-6.

200. Storm H. Changes in skin conductance as a tool to monitor nociceptive stimulation and pain. Current Opinion in Anaesthesiology. 2008;21(6):796-804.

201. Valkenburg AJ, Niehof SP, van Dijk M, Verhaar EJM, Tibboel D. Skin conductance peaks could result from changes in vital parameters unrelated to pain. Pediatric Research. 2012;71(4 Pt 1):375-9.

202. Storm H, Myre K, Rostrup M, Stokland O, Lien MD, Raeder JC. Skin conductance correlates with perioperative stress. Acta Anaesthesiologica Scandinavica. 2002;46(7):887-95.

203. Storm H, Shafiei M, Myre K, Raeder J. Palmar skin conductance compared to a developed stress score and to noxious and awakening stimuli on patients in anaesthesia. Acta Anaesthesiologica Scandinavica. 2005;49(6):798-803.

204. Ledowski T, Bromilow J, Paech MJ, Storm H, Hacking R, Schug SA. Monitoring of skin conductance to assess postoperative pain intensity. British Journal of Anaesthesia. 2006;97(6):862-5.

205. Ledowski T, Bromilow J, Wu J, Paech MJ, Storm H, Schug SA. The assessment of postoperative pain by monitoring skin conductance: results of a prospective study. Anaesthesia. 2007;62(10):989-93.

206. Eriksson M, Storm H, Fremming A, Schollin J. Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants. Acta Paediatrica. 2008;97(1):27-30.

207. de Jesus JA, Tristao RM, Storm H, da Rocha AF, Campos D, Jr. Heart rate, oxygen saturation, and skin conductance: a comparison study of acute pain in Brazilian newborns. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Confer. 2011;2011:1875-9.

208. Pereira-da-Silva L, Virella D, Monteiro I, Gomes S, Rodrigues P, Serelha M, et al. Skin conductance indices discriminate nociceptive responses to acute stimuli from different heel prick procedures in infants. Journal of Maternal-Fetal and Neonatal Medicine. 2012;25(6):796-801.

209. Gunther AC, Bottai M, Schandl AR, Storm H, Rossi P, Sackey PV. Palmar skin conductance variability and the relation to stimulation, pain and the motor activity assessment scale in intensive care unit patients. Critical Care. 2013;17(2):R51.

210. Brummelte S, Chau CM, Cepeda IL, Degenhardt A, Weinberg J, Synnes AR, et al. Cortisol levels in former preterm children at school age are predicted by neonatal procedural pain-related stress. Psychoneuroendocrinology. 2015;51:151-63.

211. Grunau RE, Cepeda IL, Chau CM, Brummelte S, Weinberg J, Lavoie PM, et al. Neonatal pain-related stress and NFKBIA genotype are associated with altered cortisol levels in preterm boys at school age. PLoS One. 2013;8(9):e73926.

212. Chorney MacLaren J, McMurtry MC. Behavioural measures of pain. In: McGrath PJ, Stevens BJ, Walker SM, Zempsky WT, editors. Oxford Textbook of Paediatric Pain. 1 ed. Oxford: Oxford University Press; 2014.

213. Lee GY, Stevens B. Neonatal and Infant Pain Assessment. In: McGrath PJ, Stevens BJ, Walker SM, Zempsky WT, editors. Oxford Textbook of Paediatric Pain. 1 ed. Oxford: Oxford University Press; 2014.

214. Sellam G, Cignacco EL, Craig KD, Engberg S. Contextual factors influencing pain response to heelstick procedures in preterm infants: what do we know? A systematic review. European Journal of Pain. 2011;15(7):661.e1-15.

215. Chorney JM, Tan ET, Martin SR, Fortier MA, Kain ZN. Children's Behavior in the Postanesthesia Care Unit: The Development of the Child Behavior Coding System-PACU (CBCS-P). Journal of Pediatric Psychology. 2011;37(3):338-47.

216. Buttner W, Finke W, Hilleke M, Reckert S, Vsianska L, Brambrink A. [Development of an observational scale for assessment of postoperative pain in infants]. Anasthesiologie, Intensivmedizin, Notfallmedizin, Schmerztherapie. 1998;33(6):353-61.

217. Gauvain-Piquard A, Rodary C, Rezvani A, Serbouti S. The development of the DEGR(R): A scale to assess pain in young children with cancer. European Journal of Pain. 1999;3(2):165-76.

218. Stevens B, Johnston C, Petryshen P, Taddio A. Premature Infant Pain Profile: development and initial validation. Clinical Journal of Pain. 1996;12(1):13-22.

219. van Dijk M, Tibboel D. Update on pain assessment in sick neonates and infants. Pediatric Clinics of North America. 2012;59(5):1167-81.

220. Stevens, Johnston C, Taddio A, Gibbins S, Yamada J. The premature infant pain profile: evaluation 13 years after development. Clinical Journal of Pain. 2010;26(9):813-30.

221. Breau LM, Camfield C, McGrath PJ, Rosmus C, Finley GA. Measuring pain accurately in children with cognitive impairments: refinement of a caregiver scale. Journal of Pediatrics. 2001;138(5):721-7.

222. Breau LM, McGrath PJ, Camfield CS, Finley GA. Psychometric properties of the non-communicating children's pain checklist-revised. Pain. 2002;99(1-2):349-57.

223. Solodiuk J, Curley MAQ. Pain assessment in nonverbal children with severe cognitive impairments: the Individualized Numeric Rating Scale (INRS). Journal of Pediatric Nursing. 2003;18(4):295-9.

224. Hunt A, Goldman A, Seers K, Crichton N, Mastroyannopoulou K, Moffat V, et al. Clinical validation of the paediatric pain profile. Developmental Medicine and Child Neurology. 2004;46(1):9-18.

225. Malviya S, Voepel-Lewis T, Burke C, Merkel S, Tait AR. The revised FLACC observational pain tool: improved reliability and validity for pain assessment in children with cognitive impairment. Paediatric Anaesthesia. 2006;16(3):258-65.

226. Crosta QR, Ward TM, Walker AJ, Peters LM. A review of pain measures for hospitalized children with cognitive impairment. Journal for Specialists in Pediatric Nursing. 2014;19(2):109-18.

227. Ambuel B, Hamlett KW, Marx CM, Blumer JL. Assessing distress in pediatric intensive care environments: the COMFORT scale. Journal of Pediatric Psychology. 1992;17(1):95-109.

228. Gilbert CA, Lilley CM, Craig KD, McGrath PJ, Court CA, Bennett SM, et al. Postoperative pain expression in preschool children: validation of the child facial coding system. Clinical Journal of Pain. 1999;15(3):192-200.

229. Chambers CT, Cassidy KL, McGrath PJ, C. A. Gilbert CA, Craig KD. Child Facial Coding System Revised Manual. Vancouver, Canada: Dalhousie University, Halifax, Canada, University of British Columbia; 1996.

230. Darwin C. Special Expressions of Man: Suffering and Weeping. The Expression of the Emotions in Man and Animals. New York: D. Appleton & Company.; 1872. p. 146-75.

231. Ekman P, Friesen WV. Facial action coding system: A technique for the measurement of facial movement. Palo Alto, California: CA: Consulting Psychologists Press; 1978.

232. LeResche L. Facial expression in pain: A survey of candid photographs. Journal of Nonverbal Behaviour. 1982;7:45-56.

233. Craig KD, Patrick CJ. Facial expression during induced pain. Journal of Personality and Social Psychology. 1985;48(4):1080-91.

234. Patrick CJ, Craig KD, Prkachin KM. Observer judgments of acute pain: facial action determinants. Journal of Personality and Social Psychology. 1986;50(6):1291-8.

235. Prkachin KM. The consistency of facial expressions of pain: a comparison across modalities. Pain. 1992;51(3):297-306.

236. Grunau RV, Craig KD. Pain expression in neonates: facial action and cry. Pain. 1987;28(3):395-410.

237. Craig KD, Hadjistavropoulos HD, Grunau RV, Whitfield MF. A comparison of two measures of facial activity during pain in the newborn child. Journal of Pediatric Psychology. 1994;19(3):305-18.

238. Grunau RV, Johnston CC, Craig KD. Neonatal facial and cry responses to invasive and non-invasive procedures. Pain. 1990;42(3):295-305.

239. Schiavenato M, Butler-O'Hara M, Scovanner P. Exploring the association between pain intensity and facial display in term newborns. Pain Research and Management. 2011;16(1):10-2.

240. Brahnam S, Chuang C-F, Shih FY, Slack MR. Machine recognition and representation of neonatal facial displays of acute pain. Artificial Intelligence in Medicine. 2006;36(3):211-22.

241. Chimello JT, Gaspardo CM, Cugler TS, Martinez FE, Linhares MBM. Pain reactivity and recovery in preterm neonates: latency, magnitude, and duration of behavioral responses. Early Human Development. 2009;85(5):313-8.

242. Gholami B, Haddad WM, Tannenbaum AR. Relevance vector machine learning for neonate pain intensity assessment using digital imaging. IEEE Transactions on Biomedical Engineering. 2010;57(6):1457-66.

243. Grunau RE, Oberlander T, Holsti L, Whitfield MF. Bedside application of the Neonatal Facial Coding System in pain assessment of premature neonates. Pain. 1998;76(3):277-86.

244. Guinsburg R, de Araujo Peres C, Branco de Almeida MF, de Cassia Xavier Balda R, Cassia Berenguel R, Tonelotto J, et al. Differences in pain expression between male and female newborn infants. Pain. 2000;85(1-2):127-33.

245. Ahola Kohut S, Pillai Riddell R. Does the Neonatal Facial Coding System differentiate between infants experiencing pain-related and non-pain-related distress? Journal of Pain. 2009;10(2):214-20.

246. Chang J, Versloot J, Fashler SR, McCrystal KN, Craig KD. Pain assessment in children: validity of facial expression items in observational pain scales. Clinical Journal of Pain. 2015;31(3):189-97.

247. Bueno M, Yamada J, Harrison D, Khan S, Ohlsson A, Adams-Webber T, et al. A systematic review and meta-analyses of nonsucrose sweet solutions for pain relief in neonates. Pain Research and Management. 2013;18(3):153.

248. Harrison D, Stevens B, Bueno M, Yamada J, Adams-Webber T, Beyene J, et al. Efficacy of sweet solutions for analgesia in infants between 1 and 12 months of age: a systematic review. Archives of Disease in Childhood. 2010;95(6):406-13.

249. Harrison D, Yamada J, Adams-Webber T, Ohlsson A, Beyene J, Stevens B. Sweet tasting solutions for reduction of needle-related procedural pain in children aged one to 16 years. Cochrane Database of Systematic Reviews. 2015(5):CD008408.

250. Kassab M, Foster Jann P, Foureur M, Fowler C. Sweet-tasting solutions for needle-related procedural pain in infants one month to one year of age. Cochrane Database of Systematic

Reviews [Internet]. 2012; (12). Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008411.pub2/abstract http://onlinelibrary.wiley.com/store/10.1002/14651858.CD008411.pub2/asset/CD008411.pdf?v=1&t=i8trtyio&s=41c6eeb9b177a87c62306a846be38288cf565d6a.

251. Bellieni CV, Sisto R, Cordelli DM, Buonocore G. Cry features reflect pain intensity in term newborns: an alarm threshold. Pediatric Research. 2004;55(1):142-6.

252. Facchini A, Bellieni C, Marchettini N, Puselli F, Tiezzi E. Relating pain intensity of newborns to onst of non-linear phenomena in cry recordings. Physician's Letter. 2005;338:332-7.

253. Mijović B, Silva M, Van den Bergh BRH, Allegaert K, Aerts JM, Berckmans D, et al. Assessment of pain expression in infant cry signals using empirical mode decomposition. Methods of Information in Medicine. 2010;49(5):448-52.

254. Lehr VT, Zeskind PS, Ofenstein JP, Cepeda E, Warrier I, Aranda JV. Neonatal facial coding system scores and spectral characteristics of infant crying during newborn circumcision. Clinical Journal of Pain. 2007;23(5):417-24.

255. Branco A, Fekete SMW, Rugolo LMSS, Rehder MI. The newborn pain cry: descriptive acoustic spectrographic analysis. International Journal of Pediatric Otorhinolaryngology. 2007;71(4):539-46.

256. Choliz M, Fernandez-Abascal EG, Martinez-Sanchez F. Infant crying: pattern of weeping, recognition of emotion and affective reactions in observers. Spanish Journal of Psychology. 2012;15(3):978-88.

257. Runefors P, Arnbjonsson E. A sound spectrogram analysis of children's crying after painful stimuli during the first year of life. Folia Phoniatrica et Logopedica. 2005;57(2):90-5.

258. Fuller BF. Infant gender differences regarding acute established pain. Clinical Nursing Research. 2002;11(2):190-203.

259. Gibbins S, Stevens B, McGrath P, Dupuis A, Yamada J, Beyene J, et al. Changes in physiological and behavioural pain indicators over time in preterm and term infants at risk for neurologic impairment. Early Human Development. 2008;84(11):731-8.

260. Hadjistavropoulos HD, Craig KD, Grunau RV, Johnston CC. Judging pain in newborns: facial and cry determinants. Journal of Pediatric Psychology. 1994;19(4):485-91.

261. Johnston CC, Sherrard A, Stevens B, Franck L, Stremler R, Jack A. Do cry features reflect pain intensity in preterm neonates? A preliminary study. Biology of the Neonate. 1999;76(2):120-4.

262. Silva M, Mijovic B, Van den Bergh BRH, Allegaert K, Aerts J-M, Van Huffel S, et al. Decoupling between fundamental frequency and energy envelope of neonate cries. Early Human Development. 2010;86(1):35-40.

263. Blount RL, Loiselle KA. Behavioural assessment of pediatric pain. Pain Research and Management. 2009;14(1):47-52.

264. Goodenough B, van Dongen K, Brouwer N, Abu-Saad HH, Champion G. A comparison of the Faces Pain Scale and the Facial Affective Scale for children's estimates of the intensity and unpleasantness of needle pain during blood sampling. European Journal of Pain. 1999;3(4):301-15.

265. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use.: Oxford University Press; 2008.

266. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.

267. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. Journal of psychosomatic research. 2010;68(4):319-23.

268. Carnevale FA, Razack S. An item analysis of the COMFORT scale in a pediatric intensive care unit. Pediatric Critical Care Medicine. 2002;3(2):177-80.

269. Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. Journal of Advanced Nursing. 2014;70(9):1970-9.

270. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.

271. Mokkink L, Terwee C, Patrick D, Alonso-Coello P, Stratton SJ, Knol D, et al. COSMIN Checklist Manual: EMGO Institute for Health and Care Research; 2010. Available from: www.cosmin.nl (no longer available).

272. Kane MT. Current Concerns in Validity Theory. Journal of Educational Measurement. 2001;38(4):319-42.

273. Messick S. Test validity and the ethics of assessment. American Psychologist. 1980;35(11):1012-27.

274. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychological Bulletin. 1955;52(4):281-302.

275. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Medical Care. 2000;38(9 Suppl):Ii84-90.

276. Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. Medical Care. 2000;38(9 Suppl):Ii14-25.

277. Bossuyt PMM, Reitsma JB, Linnet K, Moons KGM. Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. Clinical Chemistry. 2012;58(12):1636-43.

278. Zwicker JG, Grunau RE, Adams E, Chau V, Brant R, Poskitt KJ, et al. Score for neonatal acute physiology-II and neonatal pain predict corticospinal tract development in premature newborns. Pediatric Neurology. 2013;48(2):123-9.e1.

279. Walker SM. Translational studies identify long-term impact of prior neonatal pain experience. Pain. 2017;158 Suppl 1:S29-s42.

280. Grunau RE, Whitfield MF, Petrie-Thomas J, Synnes AR, Cepeda IL, Keidar A, et al. Neonatal pain, parenting stress and interaction, in relation to cognitive and motor development at 8 and 18 months in preterm infants. Pain. 2009;143(1-2):138-46.

281. Grunau RV, Whitfield MF, Petrie JH, Fryer EL. Early pain experience, child and family factors, as precursors of somatization: a prospective study of extremely premature and fullterm children. Pain. 1994;56(3):353-9.

282. Kortesluoma RL, Nikkonen M. 'The most disgusting ever': children's pain descriptions and views of the purpose of pain. Journal of Child Health Care. 2006;10(3):213-27.

283. Kortesluoma RL, Nikkonen M. 'I had this horrible pain': the sources and causes of pain experiences in 4- to 11-year-old hospitalized children. Journal of Child Health Care. 2004;8(3):210-31.

284. von Baeyer CL. Children's self-report of pain intensity: what we know, where we are headed. Pain Research and Management. 2009;14(1):39-45.

285. McGrath PJ, Walco GA, Turk DC, Dworkin RH, Brown MT, Davidson K, et al. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. Journal of Pain. 2008;9(9):771-83.

286. Duhn LJ, Medves JM. A systematic integrative review of infant pain assessment tools. Advances in Neonatal Care. 2004;4(3):126-40.

287. von Baeyer C, Hicks C. Support for a common metric for pediatric pain intensity scales. Pain Research and Management. 2000;5(2):157-60

288. Taddio A, Nulman I, Koren BS, Stevens B, Koren G. A revised measure of acute pain in infants. Journal of Pain and Symptom Management. 1995;10(6):456-63.

289. Fournier-Charrière E, Tourniaire B, Carbajal R, Cimerman P, Lassauge F, Ricard C, et al. EVENDOL, a new behavioral pain scale for children ages 0 to 7 years in the emergency department: Design and validation. Pain. 2012;153(8):1573-82.

290. van Dijk M, Koot HM, Saad HH, Tibboel D, Passchier J. Observational visual analog scale in pediatric pain assessment: useful tool or good riddance? Clinical Journal of Pain. 2002;18(5):310-6.

291. Schultz AA, Strout TD, Jordan P, Worthing B. Safety, tolerability, and efficacy of iontophoresis with lidocaine for dermal anesthesia in ED pediatric patients. Journal of Emergency Nursing. 2002;28(4):289-96.

292. Lawrence J, Alcock D, McGrath P, Kay J, MacMurray SB, Dulberg C. The development of a tool to assess neonatal pain. Neonat Netw 12(6):59-66. 1993;12(6):59-66.

293. Taddio A, Parikh C, Yoon EW, Sgro M, Singh H, Habtom E, et al. Impact of parent-directed education on parental use of pain treatments during routine infant vaccinations: a cluster randomized trial. Pain. 2015;156(1):185-91.

294. Boelen-van der Loo WJ, Scheffer E, de Haan RJ, de Groot CJ. Clinimetric evaluation of the pain observation scale for young children in children aged between 1 and 4 years after ear, nose, and throat surgery. Journal of Developmental and Behavioral Pediatrics. 1999;20(4):222-7.

295. Soetenga D, Frank J, Pellino TA. Assessment of the validity and reliability of the University of Wisconsin Children's Hospital Pain scale for Preverbal and Nonverbal Children. Pediatric Nursing. 1999;25(6):670-6.

296. Anthony CL, Tonkin IL, Marin-Garcia J, Paul RN, Bell EDJ, Brooks E, et al. A double-blind randomized clinical study of the safety, tolerability and efficacy of Hexabrix in pediatric angiocardiography. Invest Radiol. 1984;19(6 Suppl):S335-43.

297. Baxter AL, Cohen LL, McElvery HL, Lawson ML, von Baeyer CL. An integration of vibration and cold relieves venipuncture pain in a pediatric emergency department. Pediatric Emergency Care. 2011;27(12):1151-6.

298. Harrison A. Preparing children for venous blood sampling. Pain. 1991;45(3):299-306.

299. Kuttner L, Bowman M, Teasdale M. Psychological treatment of distress, pain, and anxiety for young children with cancer. Journal of Developmental and Behavioral Pediatrics. 1988;9(6):374-81.

300. Hicks CL, von Baeyer CL, Spafford PA, van Korlaar I, Goodenough B. The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. Pain. 2001;93(2):173-83.

301. Bauchner H, Vinci R, Bak S, Pearson C, Corwin MJ. Parents and procedures: a randomized controlled trial. Pediatrics. 1996;98(5):861-7.

302. Chang PC, Goresky GV, O'Connor G, Pyesmany DA, Rogers PC, Steward DJ, et al. A multicentre randomized study of single-unit dose package of EMLA patch vs EMLA 5% cream for venepuncture in children. Canadian Journal of Anaesthesia. 1994;41(1):59-63.

303. Kurien L, Kollberg H, Uppfeldt A. Venepuncture pain can be reduced. J Trop Med Hyg. 1985;88(6):397-9.

304. Nilsson A, Boman I, Wallin B, Rotstein A. The EMLA patch-a new type of local anaesthetic application for dermal analgesia in children. Anaesthesia. 1994;49(1):70-2.

305. Capeding MRZ, Jica C, Macura-Biegun A, Rauscher M, Alberto E. Interchangeability of Quinvaxem during primary vaccination schedules: Results from a phase IV, single-blind, randomized, controlled, single-center, non-inferiority study. Vaccine. 2014;32(7):888-94.

306. Sethna NF, Verghese ST, Hannallah RS, Solodiuk JC, Zurakowski D, Berde CB. A randomized controlled trial to evaluate S-Caine patch for reducing pain associated with vascular access in children. Anesthesiology. 2005;102(2):403-8.

307. Sims C. Thickly and thinly applied lignocaine-prilocaine cream prior to venepuncture in children. Anaesthesia and Intensive Care. 1991;19(3):343-5.

308. Wolf AR, Stoddart PA, Murphy PJ, Sasada M. Rapid skin anaesthesia using high velocity lignocaine particles: a prospective placebo controlled trial. Archives of Disease in Childhood. 2002;86(4):309-12.

309. Shaw EG, Routh DK. Effect of Mother Presence on Children's Reaction to Aversive Procedures1. Journal of Pediatric Psychology. 1982;7(1):33-42.

310. Bieri D, Reeve RA, Champion GD, Addicoat L, Ziegler JB. The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties. Pain. 1990;41(2):139-50.

311. McGrath PA, Seifert CE, Speechley KN, Booth JC, Stitt L, Gibson MC. A new analogue scale for assessing children's pain: an initial validation study. Pain. 1996;64(3):435-43.

312. Hannallah RS, Broadman LM, Belman AB, Abramowitz MD, Epstein BS. Comparison of caudal and ilioinguinal/iliohypogastric nerve blocks for control of post-orchiopexy pain in pediatric ambulatory surgery. Anesthesiology. 1987;66(6):832-4.

313. Beyer JE, Aradine CR. Content validity of an instrument to measure young children's perceptions of the intensity of their pain. Journal of Pediatric Nursing. 1986;1(6):386-95.

314. Newby BD, Faschoway GD, Soukoroff CI. Tetracaine (Ametop) compared to placebo for reducing pain associated with intramuscular injection of palivizumab (Synagis). Journal of Pediatric Nursing. 2009;24(6):529-33.

315. Capeding MR, Cadorna-Carlos J, Book-Montellano M, Ortiz E. Immunogenicity and safety of a DTaP-IPV//PRP~T combination vaccine given with hepatitis B vaccine: A randomized open-label trial. Bulletin of the World Health Organization. 2008;86(6):443-51.

316. Carbajal R, Paupe A, Hoenn E, Lenclen R, Olivier-Martin M. APN: evaluation behavioral scale of acute pain in newborn infants. Archives of Pediatrics. 1997;4(7):623-8.

317. Wong DL, Baker CM. Pain in children: comparison of assessment scales. Pediatric Nursing. 1988;14(1):9-17.

318. Ipp M, Taddio A, MG G, Ben David S, Stevens B, Koren G. Effects of age, gender and holding on pain response during infant immunization. Canadian Journal of Clinical Pharmacology. 2004;11:e2-e7.

319. Izard CE. The Maximally Discriminative Facial Movement Coding System (MAX) Rev. ed ed. Newark: University of Delaware, Instructional Resources Center; 1995.

320. Splinter W, Semelhago L, Chou S. The reliability and validity of a modified CHEOPS pain score. Anesthesia and analgesia. 1994;78:S413.

321. Rushforth JA, Levene MI. Behavioural response to pain in healthy neonates. Archives of Disease in Childhood, Fetal and Neonatal Edition. 1994;70(3):F174-F6.

322. Huet A, Lucas-Polomeni M-M, Robert J-C, Sixou J-L, Wodey E. Hypnosis and dental anesthesia in children: A prospective controlled study. International Journal of Clinical and Experimental Hypnosis. 2011;59(4):424-40.

323. Schade JG, Joyce BA, Gerkensmeyer J, Keck JF. Comparison of three preverbal scales for postoperative pain assessment in a diverse pediatric sample. Journal of Pain and Symptom Management. 1996;12(6):348-59.

324. Howard R, Carter B, Curry J, Morton N, Rivett K, Rose M, et al. Pain assessment. Paediatric Anaesthesia. 2008;18 Suppl 1:14-8.

325. Royal Australasian College of Physicians, Paediatric and Child Health Division. Guideline Statement: Management of Procedure-related Pain in Children and Adolescents. Sydney: Royal Australasian College of Physicians; 2005.

326. Fein JA, Zempsky WT, Cravero JP, The Comittee on Pediatric Emergency Medicine Section on Anesthesiology Pain Medicine. Relief of Pain and Anxiety in Pediatric Patients in Emergency Medical Systems. Pediatrics. 2012;130(5):e1391-e405.

327. Taddio A, Smart S, Sheedy M, Yoon EW, Vyas C, Parikh C, et al. Impact of prenatal education on maternal utilization of analgesic interventions at future infant vaccinations: a cluster randomized trial. Pain. 2014;155(7):1288-92.

328. Young SS, Schwartz R, Sheridan MJ. EMLA cream as a topical anesthetic before office phlebotomy in children. Southern Medical Journal. 1996;89(12):1184-7.

329. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine. 2009;6(7):e1000097.

330. Crellin DJ, Harrison D, Santamaria N, Babl FE. Systematic Review of the FLACC scale for assessing pain in infants and children: is it reliable, valid, and feasible for use? Pain. 2015;156(11):2132-51.

331. de Jong AE, Bremer M, Schouten M, Tuinebreijer WE, Faber AW. Reliability and validity of the pain observation scale for young children and the visual analogue scale in children with burns. Burns. 2005;31(2):198-204.

332. Crellin DJ, Babl FE, Santamaria N, Harrison D. A Systematic Review of the Psychometric Properties of the Modified Behavioral Pain Scale (MBPS). Journal of Pediatric Nursing. 2018;40:14-26.

333. Matsuishi Y, Hoshino H, Shimojo N, Enomoto Y, Kido T, Hoshino T, et al. Verifying the validity and reliability of the Japanese version of the Face, Legs, Activity, Cry, Consolability (FLACC) Behavioral Scale. PLoS One. 2018;13(3):e0194094.

334. Shen J, Giles SA, Kurtovic K, Fabia R, Besner GE, Wheeler KK, et al. Evaluation of nurse accuracy in rating procedural pain among pediatric burn patients using the Face, Legs, Activity, Cry, Consolability (FLACC) Scale. Burns. 2017;43(1):114-20.

335. A systematic review of the Face Legs Activity Cry Consolability (FLACC) pain scale for assessment of pain in children [Internet]. International Prospective Register of Systematic Reviews. 2014. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014014296

336. A systematic review of the Modified Behavioral Pain Scale (MBPS) for assessment of pain in children [Internet]. International Prospective Register of Systematic Reviews. 2016. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014014296

337. A systematic review of the Visual Analogue Scale applied by an observer (VASobs) for assessment of pain in children [Internet]. International Prospective Register of Systematic Reviews. 2016. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014014296

338. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clinical Trials. 1996;17(1):1-12.

339. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). Journal of Clinical Epidemiology. 2010;63(8):854-61.

340. Cochrane Collaboration. Data collection form for intervention reviews: RCTs and non-RCTs. http://community.cochrane.org/sites/default/files/uploads/forums/u389/ERC%20data%20collection%20form%20for%20intervention%20reviews%20for%20RCTs%20and%20non-RCTs.doc: The Cochrane Collaboration; 2014.

341. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology. 2007;60(1):34-42.

342. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research. 2012;21(4):651-7.

343. Crichton A, Knight S, Oakley E, Babl FE, Anderson V. Fatigue in Child Chronic Health Conditions: A Systematic Review of Assessment Instruments. Pediatrics. 2015;135(4):e1015-e31.

344. Cohen LL, La Greca AM, Blount RL, Kazak AE, Holmbeck GN, Lemanek KL. Introduction to Special Issue: Evidence-based Assessment in Pediatric Psychology. Journal of Pediatric Psychology. 2008;33(9):911-5.

345. Aitken RC. Measurement of feelings using visual analogue scales. Proceedings of the Royal Society of Medicine. 1969;62(10):989-93.

346. Ohnhaus EE, Adler R. Methodological problems in the measurement of pain: A comparison between the verbal rating scale and the visual analogue scale. Pain. 1975;1(4):379-84.

347. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain. 1983;17(1):45-56.

348. Scott J, Huskisson EC. Graphic representation of pain. Pain. 1976;2(2):175-84.

349. Suominen P, Caffin C, Linton S, McKinley D, Ragg P, Davie G, et al. The cardiac analgesic assessment scale (CAAS): a pain assessment tool for intubated and ventilated children after cardiac surgery. Paediatric Anaesthesia. 2004;14(4):336-43.

350. Abuelkheir M, Alsourani D, Al-Eyadhy A, Temsah MH, Meo SA, Alzamil F. EMLA(R) cream: a pain-relieving strategy for childhood vaccination. Journal of International Medical Research. 2014;42(2):329-36.

351. Babl FE, Goldfinch C, Mandrawa C, Crellin D, O'Sullivan R, Donath S. Does nebulized lidocaine reduce the pain and distress of nasogastric tube insertion in young children? A randomized, double-blind, placebo-controlled trial. Pediatrics. 2009;123(6):1548-55.

352. Balan R, Bavdekar SB, Jadhav S. Can Indian classical instrumental music reduce pain felt during venepuncture? Indian Journal of Pediatrics. 2009;76(5):469-73.

353. Barkan S, Breitbart R, Brenner-Zada G, Feldon M, Assa A, Toledano M, et al. A double-blind, randomised, placebo-controlled trial of oral midazolam plus oral ketamine for sedation of children during laceration repair. Emergency Medicine Journal. 2014;31(8):649-53.

354. Bhatnagar S, Mishra S, Gupta M, Srikanti M, Mondol A, Diwedi A. Efficacy and safety of a mixture of ketamine, midazolam and atropine for procedural sedation in paediatric oncology: a randomised study of oral versus intramuscular route. Journal of Paediatrics and Child Health. 2008;44(4):201-4.

355. Bishai R, Taddio A, Bar-Oz B, Freedman MH, Koren G. Relative efficacy of amethocaine gel and lidocaine-prilocaine cream for Port-a-Cath puncture in children. Pediatrics. 1999;104(3):e31.

356. Chapman LL, Sullivan B, Pacheco AL, Draleau CP, Becker BM. VeinViewer-assisted Intravenous catheter placement in a pediatric emergency department. Academic Emergency Medicine. 2011;18(9):966-71.

357. Cignacco E, Hamers JP, van Lingen RA, Zimmermann LJ, Muller R, Gessler P, et al. Pain relief in ventilated preterms during endotracheal suctioning: a randomized controlled trial. Swiss Medical Weekly. 2008;138(43-44):635-45.

358. Cohen LL, MacLaren JE, Fortson BL, Friedman A, DeMore M, Lim CS, et al. Randomized clinical trial of distraction for infant immunization pain. Pain. 2006;125(1-2):165-71.

359. Cohen LL, MacLaren JE, DeMore M, Fortson B, Friedman A, Lim CS, et al. A randomized controlled trial of vapocoolant for pediatric immunization distress relief. Clinical Journal of Pain. 2009;25(6):490-4.

360. Di Liddo L, D'Angelo A, Nguyen B, Bailey B, Amre D, Stanciu C. Etomidate versus midazolam for procedural sedation in pediatric outpatients: a randomized controlled trial. Annals of Emergency Medicine. 2006;48(4):433-40, 40.e1.

361. Dulai SK, Firth K, Al-Mansoori K, Cave D, Kemp KA, Silveira A, et al. Does Topical Anesthetic Reduce Pain During Intraosseous Pin Removal in Children? A Randomized Controlled Trial. Journal of Pediatric Orthopedics. 2016;36(2):126-31.

362. Fatovich DM, Jacobs IG. A randomized controlled trial of buffered lidocaine for local anesthetic infiltration in children and adults with simple lacerations. Journal of Emergency Medicine. 1999;17(2):223-8.

363. Hedén L, von Essen L, Frykholm P, Ljungman G. Low-dose oral midazolam reduces fear and distress during needle procedures in children with cancer. Pediatric Blood and Cancer. 2009;53(7):1200-4.

364. Hedén L, Von Essen L, Ljungman G. Randomized interventions for needle procedures in children with cancer. European Journal of Cancer Care. 2009;18(4):358-63.

365. Heden LE, von Essen L, Ljungman G. Effect of morphine in needle procedures in children with cancer. European Journal of Pain. 2011;15(10):1056-60.

366. Heden L, von Essen L, Ljungman G. Effect of high-dose paracetamol on needle procedures in children with cancer-A RCT. Acta Paediatrica. 2014;103(3):314-9.

367. Hogan ME, Probst J, Wong K, Riddell RP, Katz J, Taddio A. A randomized-controlled trial of parent-led tactile stimulation to reduce pain during infant immunization injections. Clinical Journal of Pain. 2014;30(3):259-65.

368. Hopkins CS, Buckley CJ, Bush GH. Pain-free injection in infants. Use of a lignocaine-prilocaine cream to prevent pain at intravenous induction of general anaesthesia in 1-5-year-old children. Anaesthesia. 1988;43(3):198-201.

369. Horn MI, McCarthy AM. Children's responses to sequential versus simultaneous immunization injections. Journal of Pediatric Healthcare. 1999;13(1):18-23.

370. Hua Y, Qiu R, Yao W-y, Zhang Q, Chen X-l. The effect of virtual reality distraction on pain relief during dressing changes in children with chronic wounds on lower limbs. Pain Management Nursing. 2015;16(5):685-91.

371. Ipp M, Cohen E, Goldbach M, Macarthur C. Effect of choice of measles-mumps-rubella vaccine on immediate vaccination pain in infants. Archives of Pediatrics and Adolescent Medicine. 2004;158(4):323-6.

372. Ipp M, Cohen E, Goldbach M, Macarthur C. Pain response to M-M-R vaccination in 4-6 year old children. Canadian Journal of Clinical Pharmacology. 2006;13(3):e296-9.

373. Ipp M, Taddio A, Sam J, Gladbach M, Parkin PC. Vaccine-related pain: randomised controlled trial of two injection techniques. Archives of Disease in Childhood. 2007;92(12):1105-8.

374. Ipp M, Parkin PC, Lear N, Goldbach M, Taddio A. Order of vaccine injection and infant pain response. Archives of Pediatrics and Adolescent Medicine. 2009;163(5):469-72.

375. Knutsson N, Jansson UB, Alm B. Immediate injection pain in infants aged 18 months during vaccination against measles, mumps and rubella with either Priorix or MMR-II. Vaccine. 2006;24(31-32):5800-5.

376. Kozer E, Rosenbloom E, Goldman D, Lavy G, Rosenfeld N, Goldman M. Pain in infants who are younger than 2 months during suprapubic aspiration and transurethral bladder catheterization: a randomized, controlled study. Pediatrics. 2006;118(1):e51-6.

377. Lee-Jayaram JJ, Green A, Siembieda J, Gracely EJ, Mull CC, Quintana E, et al. Ketamine/midazolam versus etomidate/fentanyl: procedural sedation for pediatric orthopedic reductions. Pediatric Emergency Care. 2010;26(6):408-12.

378. Lindh V, Wiklund U, Blomquist HK, Hakansson S. EMLA cream and oral glucose for immunization pain in 3-month-old infants. Pain. 2003;104(1-2):381-8.

379. Luhmann J, Hurt S, Shootman M, Kennedy R. A comparison of buffered lidocaine versus ELA-Max before peripheral intravenous catheter insertions in children. Pediatrics. 2004;113(3 Pt 1):e217-20.

380. Luhmann JD, Schootman M, Luhmann SJ, Kennedy RM. A randomized comparison of nitrous oxide plus hematoma block versus ketamine plus midazolam for emergency department forearm fracture reduction in children. Pediatrics. 2006;118(4):e1078-86.

381. Marec-Berard P, Bissery A, Kebaili K, Schell M, Aubert F, Gaillard S, et al. A positioning pillow to improve lumbar puncture success rate in paediatric haematology-oncology patients: a randomized controlled trial. BMC Cancer. 2009;9:21.

382. McErlean M, Bartfield JM, Karunakar TA, Whitman MC, Turley DM. Midazolam syrup as a premedication to reduce the discomfort associated with pediatric intravenous catheter insertion. Journal of Pediatrics. 2003;142(4):429-30.

383. McGowan A, Cottrell S, Roberts R, Lankshear A. Minimising pain response during routine infant immunisation. Community Practitioner. 2013;86(6):24-8.

384. Miller K, Rodger S, Kipping B, Kimble RM. A novel technology approach to pain management in children with burns: A prospective randomized controlled trial. Burns. 2011;37(3):395-405 11p.

385. Newbury C, Herd DW. Amethocaine versus EMLA for successful intravenous cannulation in a children's emergency department: a randomised controlled study. Emergency Medicine Journal. 2009;26(7):487-91.

386. Ravikiran SR, Kumar PMJ, Meundi AD. Pain response in newborns to the order of injecting BCG and hepatitis-B vaccines: A randomized trial. Indian Journal of Pediatrics. 2011;78(6):693-7.

387. Rubinstein O, Barkan S, Breitbart R, Berkovitch S, Toledano M, Weiser G, et al. Efficacy of oral ketamine compared to midazolam for sedation of children undergoing laceration repair: A double-blind, randomized, controlled trial. Medicine. 2016;95(26):e3984.

388. Shah VS, Taddio A, Hancock R, Shah P, Ohlsson A. Topical amethocaine gel 4% for intramuscular injection in term neonates: a double-blind, placebo-controlled, randomized trial. Clinical Therapeutics. 2008;30(1):166-74.

389. Shaikh N, Hoberman A, Kurs-Lasky M, Rockette HE, Chandra A, Colborn DK, et al. Pain management in young children undergoing diagnostic tympanocentesis. Clinical Pediatrics. 2011;50(3):231-6.

390. Shavit I, Hadash A, Knaani-Levinz H, Shachor-Meyouhas Y, Kassis I. Lidocaine-based topical anesthetic with disinfectant (LidoDin) versus EMLA for venipuncture: A randomized controlled trial. Clinical Journal of Pain. 2009;25(8):711-4.

391. Simons SH, van Dijk M, van Lingen RA, Roofthooft D, Duivenvoorden HJ, Jongeneel N, et al. Routine morphine infusion in preterm newborns who received ventilatory support: a randomized controlled trial. JAMA. 2003;290(18):2419-27.

392. Sinha M, Christopher NC, Fenn R, Reeves L. Evaluation of nonpharmacologic methods of pain and anxiety management for laceration repair in the pediatric emergency department. Pediatrics. 2006;117(4):1162-8.

393. Skarbek-Borowska S, Becker BM, Lovgren K, Bates A, Minugh PA. Brief focal ultrasound with topical anesthetic decreases the pain of intravenous placement in children. Pediatric Emergency Care. 2006;22(5):339-45.

394. Zempsky WT, Karasic RB. EMLA Versus TAC for Topical Anesthesia of Extremity Wounds in Children. Annals of Emergency Medicine. 1997;30(2):163-6.

395. Zempsky WT, Robbins B, Richards PT, Leong MS, Schechter NL. A novel needle-free powder lidocaine delivery system for rapid local analgesia. Journal of Pediatrics. 2008;152(3):405-11.

396. Zempsky WT, Robbins B, McKay K. Reduction of topical anesthetic onset time using ultrasound: a randomized controlled trial prior to venipuncture in young children. Pain Medicine. 2008;9(7):795-802.

397. Bouwmeester NJ, Anand KJ, van Dijk M, Hop WC, Boomsma F, Tibboel D. Hormonal and metabolic stress responses after major surgery in children aged 0-3 years: a double-blind, randomized trial comparing the effects of continuous versus intermittent morphine. British Journal of Anaesthesia. 2001;87(3):390-9.

398. Hamers JP, Huijer Abu-Saad H, Geisler FE, van den Hout MA, Schouten HJ, Halfens RJ, et al. The effect of paracetamol, fentanyl, and systematic assessments on children's pain after tonsillectomy and adenoidectomy. Journal of Perianesthesia Nursing. 1999;14(6):357-66.

399. Kjeldgaard Pedersen L, Nikolajsen L, Rahbek O, Uldall Duch B, Moller-Madsen B. Epidural analgesia is superior to local infiltration analgesia in children with cerebral palsy undergoing unilateral hip reconstruction. Acta Orthopedica. 2016;87(2):176-82.

400. Muthusamy K, Recktenwall SM, Friesen RM, Zuk J, Gralla J, Miller NH, et al. Effectiveness of an anesthetic continuous-infusion device in children with cerebral palsy undergoing orthopaedic surgery. Journal of Pediatric Orthopedics. 2010;30(8):840-5.

401. Oztekin S, Hepaguslar H, Kar AA, Ozzeybek D, Artikaslan O, Elar Z. Preemptive diclofenac reduces morphine use after remifentanil-based anaesthesia for tonsillectomy. Paediatric Anaesthesia. 2002;12(8):694-9.

402. Prins SA, Van Dijk M, Van Leeuwen P, Searle S, Anderson BJ, Tibboel D, et al. Pharmacokinetics and analgesic effects of intravenous propacetamol vs rectal paracetamol in children after major craniofacial surgery. Paediatric Anaesthesia. 2008;18(7):582-92.

403. Splinter WM, Bass J, Komocar L. Regional anaesthesia for hernia repair in children: local vs caudal anaesthesia. Canadian Journal Of Anaesthesia. 1995;42(3):197-200.

404. Splinter WM, Reid CW, Roberts DJ, Bass J. Reducing pain after inguinal hernia repair in children: caudal anesthesia versus ketorolac tromethamine. Anesthesiology. 1997;87(3):542-6.

405. van der Marel CD, van Lingen RA, Pluim MA, Scoones G, van Dijk M, Vaandrager JM, et al. Analgesic efficacy of rectal versus oral acetaminophen in children after major craniofacial surgery. Clinical pharmacology and therapeutics. 2001;70(1):82-90.

406. van der Marel CD, Peters JW, Bouwmeester NJ, Jacqz-Aigrain E, van den Anker JN, Tibboel D. Rectal acetaminophen does not reduce morphine consumption after major surgery in young infants. British Journal of Anaesthesia. 2007;98(3):372-9.

407. Bolt P, Barnett P, Babl FE, Sharwood LN. Topical lignocaine for pain relief in acute otitis media: results of a double-blind placebo-controlled randomised trial. Archives of Disease in Childhood. 2008;93(1):40-4.

408. Coda A, Fowlie PW, Davidson JE, Walsh J, Carline T, Santos D. Foot orthoses in children with juvenile idiopathic arthritis: a randomised controlled trial. Archives of Disease in Childhood. 2014;99(7):649-51.

409. Koller DM, Myers AB, Lorenz D, Godambe SA. Effectiveness of oxycodone, ibuprofen, or the combination in the initial management of orthopedic injury-related pain in children. Pediatric Emergency Care. 2007;23(9):627-33.

410. Miner JR, Kletti C, Herold M, Hubbard D, Biros MH. Randomized clinical trial of nebulized fentanyl citrate versus i.v. fentanyl citrate in children presenting to the emergency department with acute pain. Academic Emergency Medicine. 2007;14(10):895-8.

411. Abu-Saad HH, Uiterwijk M. Pain in Children with Juvenile Rheumatoid Arthritis: A Descriptive Study. Pediatr Res. 1995;38(2):194-7.

412. Bai J, Hsu L, Tang Y, van Dijk M. Validation of the COMFORT Behavior scale and the FLACC scale for pain assessment in Chinese children after cardiac surgery. Pain Management Nursing. 2012;13(1):18-26.

413. Berntson L, Svensson E. Pain assessment in children with juvenile chronic arthritis: a matter of scaling and rater. Acta Paediatrica. 2001;90(10):1131-6.

414. Breau L, Camfield C, Camfield P. Development and initial validation of the Batten's Observational Pain Scale: A preliminary study. Journal of Pain Management. 2010;3(3):283-92.

415. Breau LM, Finley G, Camfield CS, McGrath PJ. Facial expression of pain in children with intellectual disabilities following surgery. Journal of Pain Management. 2009;2(1):51-62.

416. Breau LM, McGrath PJ, Craig KD, Santor D, Cassidy KL, Reid GJ. Facial expression of children receiving immunizations: a principal components analysis of the child facial coding system. Clinical Journal of Pain. 2001;17(2):178-86.

417. de Jong A, Baartmans M, Bremer M, van Komen R, Middelkoop E, Tuinebreijer W, et al. Reliability, validity and clinical utility of three types of pain behavioural observation scales for young children with burns aged 0-5 years. Pain. 2010;150(3):561-7.

418. Eyelade OR, Oladokun RE, Fatiregun AA. Convergent validity of pain measuring tools among Nigerian children. African Journal of Medicine and Medical Sciences. 2009;38(4):333-6.

419. Filocamo G, Davì S, Pistorio A, Bertamino M, Ruperto N, Lattanzi B, et al. Evaluation of 21-numbered circle and 10-centimeter horizontal line visual analog scales for physician and parent subjective ratings in juvenile idiopathic arthritis. Journal of Rheumatology. 2010;37(7):1534-41.

420. Garcia-Munitis P, Bandeira M, Pistorio A, Magni-Manzoni S, Ruperto N, Schivo A, et al. Level of agreement between children, parents, and physicians in rating pain intensity in juvenile idiopathic arthritis. Arthritis and Rheumatology. 2006;55(2):177-83.

421. Hirschfeld G, Zernikow B. Cut points for mild, moderate, and severe pain on the VAS for children and adolescents: what can be learned from 10 million ANOVAs? Pain. 2013;154(12):2626-32.

422. Jylli L, Olsson GL. Procedural pain in a paediatric surgical emergency unit. Acta Paediatrica. 1995;84(12):1403-8.

423. Kelly A-M, Powell CV, Williams A. Parent visual analogue scale ratings of children's pain do not reliably reflect pain reported by child. Pediatric Emergency Care. 2002;18(3):159-62.

424. Liaw J-J, Yang L, Chou H-L, Yin T, Chao S-C, Lee T-Y. Psychometric analysis of a Taiwan-version pain assessment scale for preterm infants. Journal of Clinical Nursing. 2012;21(1/2):89-100.

425. McClellan CB, Schatz JC, Mark TR, McKelvy A, Puffer E, Roberts CW, et al. Criterion and convergent validity for 4 measures of pain in a pediatric sickle cell disease population. Clinical Journal of Pain. 2009;25(2):146-52.

426. McNair C, Ballantyne M, Dionne K, Stephens D, Stevens B. Postoperative pain assessment in the neonatal intensive care unit. Archives of Disease in Childhood Fetal and Neonatal Edition. 2004;89(6):F537-41.

427. Miller D. Practice applications of research. Comparisons of pain ratings from postoperative children, their mothers, and their nurses. Pediatric Nursing. 1996;22(2):145-9.

428. Ramelet AS, Rees NW, McDonald S, Bulsara MK, Huijer Abu-Saad H. Clinical validation of the Multidimensional Assessment of Pain Scale. Paediatric Anaesthesia. 2007;17(12):1156-65.

429. Romsing J, Moller-Sonnergaard J, Hertel S, Rasmussen M. Postoperative pain in children: comparison between ratings of children and nurses. Journal of Pain and Symptom Management. 1996;11(1):42-6.

430. Singer AJ, Gulla J, Thode HC, Jr. Parents and practitioners are poor judges of young children's pain severity. Academic Emergency Medicine. 2002;9(6):609-12.

431. Spence K, Gillies D, Harrison D, Johnston L, Nagy S. A reliable pain assessment tool for clinical assessment in the neonatal intensive care unit. Journal of Obstetric, Gynecologic and Neonatal Nursing. 2005;34(1):80-6.

432. Stein PR. Indices of pain intensity: construct validity among preschoolers. Pediatric Nursing. 1995;21(2):119-23.

433. Taddio A, O'Brien L, Ipp M, Stephens D, Goldbach M, Koren G. Reliability and validity of observer ratings of pain using the visual analog scale (VAS) in infants undergoing immunization injections. Pain. 2009;147(1-3):141-6.

434. Tarbell SE, Cohen IT, Marsh JL. The Toddler-Preschooler Postoperative Pain Scale: an observational scale for measuring postoperative pain in children aged 1-5. Preliminary report. Pain. 1992;50(3):273-80.

435. Terstegen C, Koot HM, de Boer JB, Tibboel D. Measuring pain in children with cognitive impairment: pain response to surgical procedures. Pain. 2003;103(1-2):187-98.

436. Välitalo PAJ, van Dijk M, Krekels EHJ, Gibbins S, Simons SHP, Tibboel D, et al. Pain and distress caused by endotracheal suctioning in neonates is better quantified by behavioural than physiological items: a comparison based on item response theory modelling. Pain. 2016;157(8):1611-7.

437. Varni JW, Thompson KL, Hanson V. The Varni/Thompson Pediatrie Pain Questionnaire. I. Chronic musculoskeletal pain in juvenile rheumatoid arthritis. Pain. 1987;28(1):27-38.

438. Voepel-Lewis T, Merkel S, Tait AR, Trzcinka A, Malviya S. The reliability and validity of the Face, Legs, Activity, Cry, Consolability observational tool as a measure of pain in children with cognitive impairment. Anesthesia and analgesia. 2002;95(5):1224-9.

439. Wilson GA, Doyle E. Validation of three paediatric pain scores for use by parents. Anaesthesia. 1996;51(11):1005-7.

440. Harrison D, Elia S, Manias E, Royle J. Sucrose and lollypops to reduce immunisation pain in toddlers and young children: Two pilot randomised controlled trials. Neonatal, Paediatric and Child Health Nursing. 2014;17(1):19.

441. Knutsson J, Tibbelin A, Von Unge M. Postoperative pain after paediatric adenoidectomy and differences between the pain scores made by the recovery room staff, the parent and the child. Acta Oto-Laryngologica. 2006;126(10):1079-83.

442. Cicchetti DV. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. Psychological Assessment. 1994;6(4):284-90.

443. Knutsson J, Tibbelin A, Von Unge M. Adjuvant local anaesthetics in the epipharyngeal space in day-case adenoidectomy: a prospective, randomized, double-blind, placebo-controlled trial. Acta Oto-Laryngologica. 2006;126(1):51-5.

444. Zempsky WT, Bean-Lijewski J, Kauffman RE, Koh JL, Malviya SV, Rose JB, et al. Needle-free powder lidocaine delivery system provides rapid effective analgesia for venipuncture or cannulation pain in children: randomized, double-blind Comparison of Venipuncture and Venous Cannulation Pain After Fast-Onset Needle-Free Powder Lidocaine or Placebo Treatment trial. Pediatrics. 2008;121(5):979-87.

445. McClellan CB, Cohen LL, Joseph KE. Infant distress during immunization: A multimethod assessment. Journal of Clinical Psychology in Medical Settings. 2003;10(4):231-8.

446. Dulai SK, Firth K, Al-Mansoori K, Cave D, Kemp KAR, Silveira A, et al. Does Topical Anesthetic Reduce Pain During Intraosseous Pin Removal in Children? A Randomized Controlled Trial. Journal of Pediatric Orthopedics. 2016;36(2):126-31.

447. Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technology Assessment. 2007;11(50):iii, ix-51.

448. Ahola Kohut S, Pillai Riddell R, Flora DB, Oster H. A longitudinal analysis of the development of infant facial expressions in response to acute pain: Immediate and regulatory expressions. Pain. 2012;153(12):2458-65.

449. Lilley CM, Craig KD, Grunau RE. The expression of pain in infants and toddlers: developmental changes in facial action. Pain. 1997;72(1-2):161-70.

450. Slater R, Cantarella A, Yoxen J, Patten D, Potts H, Meek J, et al. Latency to facial expression change following noxious stimulation in infants is dependent on postmenstrual age. Pain. 2009;146(1-2):177-82.

451. Ahlers SJ, van Gulik L, van der Veen AM, van Dongen HP, Bruins P, Belitser SV, et al. Comparison of different pain scoring systems in critically ill patients in a general ICU. Critical Care. 2008;12(1):R15.

452. Maguire K, Morrell K, Westhoff C, Davis A. Accuracy of providers' assessment of pain during intrauterine device insertion. Contraception. 2014;89(1):22-4.

453. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979;86(2):420-8.

454. Miller K, Tan X, Hobson AD, Khan A, Ziviani J, O'Brien E, et al. A Prospective Randomized Controlled Trial of Nonpharmacological Pain Management During Intravenous Cannulation in a Pediatric Emergency Department. Pediatric Emergency Care. 2016;32(7):444-51.

455. Gee Kee EL, Kimble RM, Cuttle L, Khan A, Stockton KA. Randomized controlled trial of three burns dressings for partial thickness burns in children. Burns. 2015;41(5):946-55.

456. Cohen LL, Rodrigues NP, Lim CS, Bearden DJ, Welkom JS, Joffe NE, et al. Automated parent-training for preschooler immunization pain relief: a randomized controlled trial. Journal of Pediatric Psychology. 2015;40(5):526-34.

457. Kochman A, Howell J, Sheridan M, Kou M, Shelton Ryan EE, Lee S, et al. Reliability of the Faces, Legs, Activity, Cry, and Consolability Scale in Assessing Acute Pain in the Pediatric Emergency Department. Pediatric Emergency Care. 2017;33(1):14-7.

458. Deepak V, Challa RR, Kamatham R, Nuvvula S. Comparison of a New Auto-controlled Injection System with Traditional Syringe for Mandibular Infiltrations in Children: A Randomized Clinical Trial. Anesth Essays Res. 2017;11(2):431-8.

459. Gouin S, Gaucher N, Lebel D, Desjardins MP. A Randomized Double-Blind Trial Comparing the Effect on Pain of an Oral Sucrose Solution vs. Placebo in Children 1 to 3 Months Old Undergoing Simple Venipuncture. Journal of Emergency Medicine. 2018;54(1):33-9.

460. Raslan N, Masri R. A randomized clinical trial to compare pain levels during three types of oral anesthetic injections and the effect of Dentalvibe on injection pain in children. International Journal of Paediatric Dentistry. 2018;28(1):102-10.

461. Franck LS, Berberich FR, Taddio A. Parent participation in a childhood immunization pain reduction method. Clinical Pediatrics. 2015;54(3):228-35.

462. Weiss JE, Haines KA, Chalom EC, Li SC, Walco GA, Nyirenda TL, et al. A randomized study of local anesthesia for pain control during intra-articular corticosteroid injection in children with arthritis. Pediatric Rheumatology Online Journal. 2015;13:36.

463. Elbay US, Elbay M, Kaya E, Yildirim S. Effects of Two Different Anesthetic Solutions on Injection Pain, Efficacy, and Duration of Soft-Tissue Anesthesia with Inferior Alveolar Nerve Block for Primary Molars. Journal of Clinical Pediatric Dentistry. 2016;40(6):456-63.

464. Crellin DJ, Babl FE, Santamaria N, Harrison D. The Psychometric Properties of the MBPS Scale Used to Assess Procedural Pain. Journal of Pain. 2018;19(6):660-70.

465. Uspal NG, Strelitz B, Gritton J, Follmer K, Bradford MC, Colton TL, et al. Randomized Clinical Trial of Lidocaine Analgesia for Transurethral Bladder Catheterization Delivered via Blunt Tipped Applicator in Young Children. Pediatric Emergency Care. 2018;34(4):273-9.

466. Taddio A, Riddell RP, Ipp M, Moss S, Baker S, Tolkin J, et al. A Longitudinal Randomized Trial of the Effect of Consistent Pain Management for Infant Vaccinations on Future Vaccination Distress. Journal of Pain. 2017;18(9):1060-6.

467. Taddio A, Flanders D, Weinberg E, Lamba S, Vyas C, Ilersich AF, et al. A randomized trial of rotavirus vaccine versus sucrose solution for vaccine injection pain. Vaccine. 2015;33(25):2939-43.

468. Crellin DJ, Harrison D, Hutchinson A, Schuster T, Santamaria N, Babl FE. Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months. BMJ Open. 2017;7(9).

469. Crellin DJ, Harrison D, Santamaria N, Huque H, Babl FE. The Psychometric Properties of the FLACC Scale Used to Assess Procedural Pain. Journal of Pain. 2018;19(8):862-72.

470. Tsze DS, Hirschfeld G, Dayan PS, Bulloch B, von Baeyer CL. Defining No Pain, Mild, Moderate, and Severe Pain Based on the Faces Pain Scale-Revised and Color Analog Scale in Children With Acute Pain. Pediatric Emergency Care. 2016.

471. Crellin DJ, Babl FE, Santamaria N, Harrison D. The MBPS: A systematic review to determine its role in assessing pain in infants and young children. 11th International Symposium on Pediatric Pain; Kuala Lumpur: Special Interest Group on Pain in Childhood; 2017.

472. De Ruddere L, Goubert L, Stevens M, de C. Williams AC, Crombez G. Discounting pain in the absence of medical evidence is explained by negative evaluation of the patient. Pain. 2013;154(5):669-76.

473. De Ruddere L, Goubert L, Stevens MA, Deveugele M, Craig KD, Crombez G. Health care professionals' reactions to patient pain: impact of knowledge about medical evidence and psychosocial influences. Journal of Pain. 2014;15(3):262-70.

474. Gomez RJ, Barrowman N, Elia S, Manias E, Royle J, Harrison D. Establishing intra- and inter-rater agreement of the face, legs, activity, cry, consolability scale for evaluating pain in toddlers during immunization. Pain Research and Management. 2013;18(6):e124-e8.

475. American Academy of Pediatrics. Committee on Psychosocial Aspects of Child and Family Health. Task Force on Pain in Infants Children and Adolescents. The assessment and management of acute pain in infants, children, and adolescents. Pediatrics. 2001;108(3):793-7.

476. Howard R, Carter B, Curry J, Morton N, Rivett K, Rose M, et al. Good practice in postoperative and procedural pain management. Background. Paediatric Anaesthesia. 2008;18 Suppl 1:1-3.

477. Schultz AA, Murphy E, Morton J, Stempel A, Messenger-Rioux C, Bennett K. Preverbal, Early Verbal Pediatric Pain Scale (PEPPS): development and early psychometric testing. Journal of Pediatric Nursing. 1999;14(1):19-27.

478. Singer AJ, Richman PB, Kowalska A, Thode HC, Jr. Comparison of patient and practitioner assessments of pain from commonly performed emergency department procedures. Annals of Emergency Medicine. 1999;33(6):652-8.

479. Krauss BS, Calligaris L, Green SM, Barbi E. Current concepts in management of pain in children in the emergency department. Lancet. 2016;387(10013):83-92.

480. Andersen RD, Langius-Eklof A, Nakstad B, Bernklev T, Jylli L. The measurement properties of pediatric observational pain scales: A systematic review of reviews. International Journal of Nursing Studies. 2017;73:93-101.

481. Stevens B, Gibbins S. Clinical utility and clinical significance in the assessment and management of pain in vulnerable infants. Clinics in Perinatology. 2002;29(3):459-68.

482. Birnie KA, McGrath PJ, Chambers CT. When does pain matter? Acknowledging the subjectivity of clinical significance. Pain. 2012;153(12):2311-4.

483. Oxford Uo. Oxford English Dictionary Online: Oxford University Press; 2018 [Available from: https://en.oxforddictionaries.com/.

484. Pillai Riddell R, Racine NM, Turcotte K, Uman L, S., Horton R, E., Din Osmun L, et al. Non-pharmacological management of infant and young child procedural pain. Cochrane Database of Systematic Reviews [Internet]. 2011; (10). Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006275.pub2/abstract http://onlinelibrary.wiley.com/store/10.1002/14651858.CD006275.pub2/asset/CD006275.pdf?v=1&t=i8ts09ht&s=e4b7f494d8c1216d32838d8504d82e2f1ff277d7.

485. Birnie KA, Chambers CT, Taddio A, McMurtry CM, Noel M, Pillai Riddell R, et al. Psychological Interventions for Vaccine Injections in Children and Adolescents: Systematic Review of Randomized and Quasi-Randomized Controlled Trials. Clinical Journal of Pain. 2015;31:S72-S89.

486. McHugh ML. Interrater reliability: the kappa statistic. Biochemia Medica. 2012;22(3):276-82.

487. Craig KD. The Social Communication Model of Pain. Canadian Psychology. 2009;50(1):22-32.

488. Taddio A, Hogan ME, Moyer P, Girgis A, Gerges S, Wang L, et al. Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. Vaccine. 2011;29(7):1390-4.

489. Chibnall JT, Tait RC, Ross LR. The Effects of Medical Evidence and Pain Intensity on Medical Student Judgments of Chronic Pain Patients. Journal of Behavioral Medicine. 1997;20(3):257-71.

490. Pillai Riddell RR, Craig KD. Judgments of infant pain: the impact of caregiver identity and infant age. Journal of Pediatric Psychology. 2007;32(5):501-11.

491. Pillai Riddell RR, Horton RE, Hillgrove J, Craig KD. Understanding caregiver judgments of infant pain: contrasts of parents, nurses and pediatricians. Pain Research and Management. 2008;13(6):489-96.

492. Pillai Riddell RR, Badali MA, Craig KD. Parental judgements of infant pain: importance of perceived cognitive abilities, behavioural cues and contextual cues. Pain Research and Management. 2004;9(2):73-80.

493. Pillai Riddell R, Flora DB, Stevens S, Greenberg S, Garfield H. The role of infant pain behaviour in predicting parent pain ratings. Pain Research and Management. 2014;19(5):e124.

494. Guru V, Dubinsky I. The patient vs. caregiver perception of acute pain in the emergency department. Journal of Emergency Medicine. 2000;18(1):7-12.

495. Rajasagaram U, Taylor DM, Braitberg G, Pearsell JP, Capp BA. Paediatric pain assessment: differences between triage nurse, child and parent. Journal of Paediatrics and Child Health. 2009;45(4):199-203.

496. Prkachin KM, Solomon PE, Ross J. Underestimation of pain by health-care providers: towards a model of the process of inferring pain in others. Canadian Journal of Nursing Research. 2007;39(2):88-106.

497. Gregoire M, Coll MP, Tremblay MP, Prkachin KM, Jackson PL. Repeated exposure to others' pain reduces vicarious pain intensity estimation. Europena Journal ofJ Pain. 2016;20(10):1644-52.

498. Prkachin KM, Berzins S, Mercer SR. Encoding and decoding of pain expressions: a judgement study. Pain. 1994;58(2):253-9.

499. Prkachin KM, Mass H, Mercer SR. Effects of exposure on perception of pain expression. Pain. 2004;111(1-2):8-12.

500. Prkachin KM, Rocha EM. High levels of vicarious exposure bias pain judgments. Journal of Pain. 2010;11(9):904-9.

501. Lamm C, Decety J, Singer T. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. NeuroImage. 2011;54(3):2492-502.

502. Coll MP, Gregoire M, Prkachin KM, Jackson PL. Repeated exposure to vicarious pain alters electrocortical processing of pain expressions. Experimental Brain Research. 2016;234(9):2677-86.

503. Preis MA, Kröner-Herwig B, Schmidt-Samoa C, Dechent P, Barke A. Neural Correlates of Empathy with Pain Show Habituation Effects. An fMRI Study. PLoS One. 2015;10(8):e0137056.

504. Racine NM, Riddell RR, Khan M, Calic M, Taddio A, Tablon P. Systematic Review: Predisposing, Precipitating, Perpetuating, and Present Factors Predicting Anticipatory Distress to Painful Medical Procedures in Children. Journal of Pediatric Psychology. 2016;41(2):159-81.

505. Stevens SA, Racine N, Riddell RP, Horton RE, Garfield H, Greenberg S. Infant pain regulation as an early indicator of childhood temperament. Pain Research and Management. 2013;18(6):313.

506. Pillai Riddell R, Flora DB, Stevens SA, Stevens B, Cohen LL, Greenberg S, et al. Variability in infant acute pain responding meaningfully obscured by averaging pain responses. Pain. 2013;154(5):714-21.

507. Wiech K, Tracey I. The influence of negative emotions on pain: behavioral effects and neural mechanisms. Neuroimage. 2009;47(3):987-94.

508. Humphrey GB, Boon CMJ, Van den Heuvell GFECVL, Van de Wiel HBM. The occurrence of high levels of acute behavioral distress in children and adolescents undergoing routine venipunctures. Pediatrics. 1992;90(1 PART 1):87-91.

509. Rhudy JL, Meagher MW. Fear and anxiety: divergent effects on human pain thresholds. Pain. 2000;84(1):65-75.

510. Blount RL, Sturges JW, Powers SW. Analysis of child and adult behavioral variations by phase of medical procedure. Behavioral Therapy. 1990;21(1):33-48.

511. Jacobson RM, Swan A, Adegbenro A, Ludington SL, Wollan PC, Poland GA, et al. Making vaccines more acceptable--methods to prevent and minimize pain and other common adverse events associated with vaccines. Vaccine. 2001;19(17-19):2418-27.

512. Taddio A, Ipp M, Thivakaran S, Jamal A, Parikh C, Smart S, et al. Survey of the prevalence of immunization non-compliance due to needle fears in children and adults. Vaccine. 2012;30(32):4807-12.

513. Scholtz S, editor Threat: concept analysis. Nursing Forum; 2000: Wiley Online Library.

514. Zonneveld LN, McGrath PJ, Reid GJ, Sorbi MJ. Accuracy of children's pain memories. Pain. 1997;71(3):297-302.

515. Noel M, Chambers CT, McGrath PJ, Klein RM, Stewart SH. The influence of children's pain memories on subsequent pain experience. Pain. 2012;153(8):1563-72.

516. Iannetti GD, Salomons TV, Moayedi M, Mouraux A, Davis KD. Beyond metaphor: contrasting mechanisms of social and physical pain. Trends in Cognitive Sciences. 2013;17(8):371-8.

517. Lacey CM, Finkelstein M, Thygeson MV. The impact of positioning on fear during immunizations: supine versus sitting up. Journal of Pediatric Nursing. 2008;23(3):195-200.

518. South West Anaesthetic Research Matrix (SWARM). Sedation practice in six acute hospitals - a snapshot survey. Anaesthesia. 2015;70(4):407-15.

519. Czarnecki ML, Turner HN, Collins PM, Doellman D, Wrona S, Reynolds J. Procedural pain management: a position statement with clinical practice recommendations. Pain Management Nursing. 2011;12(2):95-111.

520. Borland M, Esson A, Babl F, Krieser D. Procedural sedation in children in the emergency department: a PREDICT study. Emergency Medicine Australasia. 2009;21(1):71-9.

521. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. Journal of Clinical Epidemiology. 2006;59(10):1033-9.

522. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. Statistics in Medicine. 2002;21(22):3431-46.

523. Bringuier S, Picot M-C, Dadure C, Rochette A, Raux O, Boulhais M, et al. A prospective comparison of post-surgical behavioral pain scales in preschoolers highlighting the risk of false evaluations. Pain. 2009;145(1-2):60-8.

524. Holsti L, Grunau RE, Oberlander TF, Osiovich H. Is it painful or not? Discriminant validity of the Behavioral Indicators of Infant Pain (BIIP) scale. Clinical Journal of Pain. 2008;24(1):83-8.

525. Hunseler C, Merkt V, Gerloff M, Eifinger F, Kribs A, Roth B. Assessing pain in ventilated newborns and infants: validation of the Hartwig score. European Journal of Pediatrics. 2011;170(7):837-43.

526. Hunt A, Wisbeach A, Seers K, Goldman A, Crichton N, Perry L, et al. Development of the paediatric pain profile: role of video analysis and saliva cortisol in validating a tool to assess pain in children with severe neurological disability. Journal of Pain and Symptom Management. 2007;33(3):276-89.

527. Milesi C, Cambonie G, Jacquot A, Barbotte E, Mesnage R, Masson F, et al. Validation of a neonatal pain scale adapted to the new practices in caring for preterm newborns. Archives of Disease in Childhood Fetal and Neonatal Edition. 2010;95(4):F263-6.

528. Uyan ZS, Bilgen H, Topuzoglu A, Akman I, Ozek E. Comparison of three neonatal pain scales during minor painful procedures. Journal of Maternal-Fetal and Neonatal Medicine. 2008;21(5):305-8.

529. Voepel-Lewis T, Malviya S, Tait AR, Merkel S, Foster R, Krane EJ, et al. A comparison of the clinical utility of pain assessment tools for children with cognitive impairment. Anesthesia and analgesia. 2008;106(1):72-8, table of contents.

530. American Academy of Pediatrics, American Academy of Pediatric Dentistry, Cote CJ, Wilson S, Work Group on Sedation. Guidelines for monitoring and management of pediatric patients during and after sedation for diagnostic and therapeutic procedures: an update. Pediatrics. 2006;118(6):2587-602.

531. Godwin SA, Caro DA, Wolf SJ, Jagoda AS, Charles R, Marett BE, et al. Clinical policy: procedural sedation and analgesia in the emergency department. Annals of Emergency Medicine. 2005;45(2):177-96.

532. Lago P, Garetti E, Merazzi D, Pieragostini L, Ancora G, Pirelli A, et al. Guidelines for procedural pain in the newborn. Acta Paediatrica. 2009;98(6):932-9.

533. Wolfe TR, Fosnocht DE, Linscott MS. Atomized lidocaine as topical anesthesia for nasogastric tube placement: A randomized, double-blind, placebo-controlled trial. Annals of Emergency Medicine. 2000;35(5):421-5.

534. Cullen L, Taylor D, Taylor S, Chu K. Nebulized lidocaine decreases the discomfort of nasogastric tube insertion: a randomized, double-blind trial. Annals of Emergency Medicine. 2004;44(2):131-7.

535. Ducharme J, Matheson K. What is the best topical anesthetic for nasogastric insertion? A comparison of lidocaine gel, lidocaine spray, and atomized cocaine. Journal of Emergency Nursing. 2003;29(5):427-30.

536. Hrisos S, Eccles MP, Francis JJ, Dickinson HO, Kaner EF, Beyer F, et al. Are there valid proxy measures of clinical behaviour? a systematic review. Implementation Science. 2009;4(1):37.

537. Croskerry P. Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias. Academic Emergency Medicine. 2002;9(11):1184-204.

538. Gagnon MM, Hadjistavropoulos T, Hampton AJ, Stinson J. A Systematic Review of Knowledge Translation (KT) in Pediatric Pain: Focus on Health Care Providers. Clinical Journal of Pain. 2016;32(11):972-90.

539. Crellin DJ, Harrison D, Santamaria N, Babl FE. The FLACC scale for assessing pain in infants and children; a systematic review. ACEM Annual Scientific Meeting; Brisbane, Australia2015.

540. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. Journal of the Royal Society of Medicine. 2011;104(12):510-20.

541. Chan S, Pielak K, McIntyre C, Deeter B, Taddio A. Implementation of a new clinical practice guideline regarding pain management during childhood vaccine injections. Paediatrics and Child Health. 2013;18(7):367-72.

542. Watt-Watson J, McGillion M, Hunter J, Choiniere M, Clark AJ, Dewar A, et al. A survey of prelicensure pain curricula in health science faculties in Canadian universities. Pain Research and Management. 2009;14(6):439-44.

543. ChildKind International. ChildKind 2018 [Available from: http://childkindinternational.org.

544. PREDICT. Paediatric Research in Emergency Departments International Collaborative 2018 [Available from: http://www.predict.org.au/.

545. Willis MHW, Merkel SI, Voepel-Lewis T, Malviya S. FLACC behavioural pain assessment scale: A comparison with a child's self-report. Pediatric Nursing. 2003;29(3):195-8.

546. Ahn Y, Jun Y. Measurement of pain-like response to various NICU stimulants for high-risk infants. Early Human Development. 2007;83(4):255-62.

547. da Silva FC, Thuler LCS. Cross-cultural adaptation and translation of two pain assessment tools in children and adolescents. Jornal de Pediatria. 2008;84(4):344-9.

548. da Silva FC, Santos Thuler LC, de Leon-Casasola OA. Validity and reliability of two pain assessment tools in Brazilian children and adolescents. Journal of Clinical Nursing. 2011;20(13/14):1842-8.

549. Johansson M, Kokinsky E. The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. Nursing in Critical Care. 2009;14(3):122-30.

550. Manworren RCB, Hynan LS. Clinical validation of FLACC: preverbal patient pain scale. Pediatric Nursing. 2003;29(2):140-6.

551. Nilsson S, Finnstrom B, Kokinsky E. The FLACC behavioral scale for procedural pain assessment in children aged 5-16 years. Paediatric Anaesthesia. 2008;18(8):767-74.

552. Suraseranivongse S, Kraiprasit K, Petcharatana S. Postoperative pain assessment in ambulatory pediatric patients by parents. Journal of the Medical Association of Thailand. 2002;85 Suppl 3:S917-22.

553. Voepel-Lewis T, Zanotti J, Dammeyer JA, Merkel S. 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. 2010;19(1):55-61; quiz 2.

554. Voepel-Lewis T, Malviya S, Tait AR. Validity of parent ratings as proxy measures of pain in children with cognitive impairment. Pain Management Nursing. 2005;6(4):168-74.

555. Amin SM. Evaluation of gabapentin and dexamethasone alone or in combination for pain control after adenotonsillectomy in children. Saudi Journal of Anaesthesia. 2014;8(3):317.

556. Anand VG, Kannan M, Thavamani A, Bridgit MJ. Effects of dexmedetomidine added to caudal ropivacaine in paediatric lower abdominal surgeries. Indian journal of anaesthesia. 2011;55(4):340-6.

557. Ashrey E, Bosat B. Single-injection penile block versus caudal block in penile pediatric surgery. Ain-Shams Journal of Anaesthesiology. 2014;7(3):428.

558. Batra YK, Rajeev S, Panda NB, Lokesh VC, Rao KLN. Intrathecal neostigmine with bupivacaine for infants undergoing lower abdominal and urogenital procedures: Dose response. Acta Anaesthesiologica Scandinavica. 2009;53(4):470-5.

559. Bharti N, Praveen R, Bala I. A dose-response study of caudal dexmedetomidine with ropivacaine in pediatric day care patients undergoing lower abdominal and perineal surgeries: a randomized controlled trial. Paediatric Anaesthesia. 2014;11(24):1158 - 63.

560. Boots BK, Edmundson EE. A controlled, randomised trial comparing single to multiple application lidocaine analgesia in paediatric patients undergoing urethral catheterisation procedures. Journal of Clinical Nursing. 2010;19(5-6):744-8.

561. Brown NJ, Kimble RM, Rodger S, Ware RS, Cuttle L. Play and heal: Randomized controlled trial of Ditto intervention efficacy on improving re-epithelialization in pediatric burns. Burns. 2014;40(2):204-13.

562. Chadha NK, Lam GOA, Ludemann JP, Kozak FK. Intranasal topical local anesthetic and decongestant for flexible nasendoscopy in children: a randomized, double-blind, placebo-controlled trial. JAMA Otolaryngology, Head and Neck Surgery. 2013;139(12):1301-5.

563. Chandler JR, Myers D, Mehta D, Whyte E, Groberman MK, Montgomery CJ, et al. Emergence delirium in children: a randomized trial to compare total intravenous anesthesia with propofol and remifentanil to inhalational sevoflurane anesthesia. Paediatric Anaesthesia. 2013;23(4):309-15.

564. Cho JE, Kim JY, Hong JY, Kil HK. The addition of fentanyl to 1.5 mg/ml ropivacaine has no advantage for paediatric epidural analgesia. Acta Anaesthesiologica Scandinavica. 2009;53(8):1084-7.

565. Curry DM, Brown C, Wrona S. Effectiveness of oral sucrose for pain management in infants during immunizations. Pain Management Nursing. 2012;13(3):139-49.

566. Curtis SJ, Jou H, Ali S, Vandermeer B, Klassen T. A randomized controlled trial of sucrose and/or pacifier as analgesia for infants receiving venipuncture in a pediatric emergency department. BMC Pediatrics. 2007;7:27.

567. Dewhirst E, Fedel G, Raman V, Rice J, Barry ND, Jatana KR, et al. Pain management following myringotomy and tube placement: Intranasal dexmedetomidine versus intranasal fentanyl. International Journal of Pediatric Otorhinolaryngology. 2014;78(7):1090-4.

568. Diao M, Li L, Cheng W. To drain or not to drain in Roux-en-Y hepatojejunostomy for children with choledochal cysts in the laparoscopic era: a prospective randomized study. Journal of Pediatric Surgery. 2012;47(8):1485-9.

569. Elshammaa N, Chidambaran V, Housny W, Thomas J, Zhang X, Michael R. Ketamine as an adjunct to fentanyl improves postoperative analgesia and hastens discharge in children following tonsillectomy - a prospective, double-blinded, randomized study. Paediatric Anaesthesia. 2011;21(10):1009-14.

570. El-Sharkawi HFA, El-Housseiny AA, Aly AM. Effectiveness of new distraction technique on pain associated with injection of local anesthesia for children. Pediatric Dentistry. 2012;34(2):e35-8.

571. Fernandes ML, Pires KCC, Tiburcio MA, Gomez RS. Caudal bupivacaine supplemented with morphine or clonidine, or supplemented with morphine plus clonidine in children undergoing infra-umbilical urological and genital procedures: a prospective, randomized and double-blind study. Journal of Anesthesia. 2012;26(2):213-8.

572. Frawley GP, Downie S, Huang GH. Levobupivacaine caudal anesthesia in children: a randomized double-blind comparison with bupivacaine. Paediatric Anaesthesia. 2006;16(7):754-60.

573. Ghai B, Ram J, Makkar JK, Wig J, Kaushik S. Subtenon block compared to intravenous fentanyl for perioperative analgesia in pediatric cataract surgery. Anesthesia and analgesia. 2009;108(4):1132-8.

574. Grove GL, Zerweck CR, Ekholm BP, Smith GE, Koski NI. Randomized Comparison of a Silicone Tape and a Paper Tape for Gentleness in Healthy Children. Journal of Wound Ostomy and Continence Nursing. 2014;41(1):40-8.

575. Hall NJ, Pacilli M, Eaton S, Reblock K, Gaines BA, Pastor A, et al. Recovery after open versus laparoscopic pyloromyotomy for pyloric stenosis: a double-blind multicentre randomised controlled trial. Lancet. 2009;373(9661):390-8.

576. Hippard HK, Govindan K, Friedman EM, Sulek M, Giannoni C, Larrier D, et al. Postoperative analgesic and behavioral effects of intranasal fentanyl, intravenous morphine, and intramuscular morphine in pediatric patients undergoing bilateral myringotomy and placement of ventilating tubes. Anesthesia and analgesia. 2012;115(2):356-63.

577. Hong JY, Lee I, Shin S, Park E, Ban S, Cho J, et al. Caudal midazolam does not affect sevoflurane requirements and recovery in pediatric day-case hernioplasty. Acta Anaesthesiologica Scandinavica. 2008;52(10):1411-4.

578. Hong J-Y, Han S, Kim W, Kim E, Kil H. Effect of dexamethasone in combination with caudal analgesia on postoperative pain control in day-case paediatric orchiopexy. British Journal of Anaesthesia. 2010;105(4):506-10.

579. Hughes J, Lindup M, Wright S, Naik M, Dhesi R, Howard R, et al. Does nasogastric feeding reduce distress after cleft palate repair in infants? . Nursing Children and Young People. 2013;25(9):26-30.

580. Jindal P, Khurana G, Dvivedi S, Sharma JP. Intra and postoperative outcome of adding clonidine to bupivacaine in infraorbital nerve block for young children undergoing cleft lip surgery. Saudi Journal of Anaesthesia. 2011;5(3):289-94.

581. Jonnavithula N, Durga P, Kulkarni DK, Ramachandran G. Bilateral intra-oral, infra-orbital nerve block for postoperative analgesia following cleft lip repair in paediatric patients: comparison of bupivacaine vs bupivacaine-pethidine combination. Anaesthesia. 2007;62(6):581-5.

582. Jonnavithula N, Durga P, Madduri V, Ramachandran G, Nuvvula R, Srikanth R, et al. Efficacy of palatal block for analgesia following palatoplasty in children with cleft palate. Paediatric Anaesthesia. 2010;20(8):727-33.

583. Kil HK, Kim WO, Han SW, Kwon Y, Lee A, Hong J-Y. Psychological and behavioral effects of chloral hydrate in day-case pediatric surgery: a randomized, observer-blinded study. Journal of Pediatric Surgery. 2012;47(8):1592-9.

584. Kim NY, Kim SY, Yoon HJ, Kil HK. Effect of Dexmedetomidine on Sevoflurane Requirements and Emergence Agitation in Children Undergoing Ambulatory Surgery. Yonsei Medical Journal. 2014;55(1):209-15.

585. Kim C-H, Yoon J-U, Lee H-J, Shin S-W, Yoon J-Y, Byeon G-J. Availability of a 5% lidocaine patch used prophylactically for venipuncture- or injection-related pain in children. Journal of Anesthesia. 2012;26(4):552-5.

586. Kundu A, Lin Y, Oron AP, Doorenbos AZ. Reiki therapy for postoperative oral pain in pediatric patients: Pilot data from a double-blind, randomized clinical trial. Complementary Therapies in Clinical Practice. 2014;20(1):21-5.

587. Loetwiriyakul W, Asampinwat T. Caudal block with 3 mg/Kg Bupivacaine for intraabdominal surgery in pediatric patients: a randomized study. Asian Biomedicine. 2011;5(1):93-9.

588. Lorenzo AJ, Lynch J, Matava C, El-Beheiry H, Hayes J. Ultrasound Guided Transversus Abdominis Plane vs Surgeon Administered Intraoperative Regional Field Infiltration with Bupivacaine for Early Postoperative Pain Control in Children Undergoing Open Pyeloplasty. Journal of Urology. 2014;192(1):192.

589. Miller K, Rodger S, Bucolo S, Greer R, Kimble RM. Multi-modal distraction. Using technology to combat pain in young children with burn injuries. Burns. 2010;36(5):647-58.

590. Natarajan Surendar M, Kumar Pandey R, Kumar Saksena A, Kumar R, Chandra G. A Comparative evaluation of Intranasal Dexmedetomidine, Midazolam and Ketamine for their sedative and analgesic properties: A Triple Blind Randomized Study. Journal of Clinical Pediatric Dentistry. 2014;38(3):255-61.

591. Nilsson S, Enskär K, Hallqvist C, Kokinsky E. Active and passive distraction in children undergoing wound dressings. Journal of Pediatric Nursing. 2013;28(2):158-66.

592. Nilsson S, Kokinsky E, Nilsson U, Sidenvall B, Enskar K. School-aged children's experiences of postoperative music medicine on pain, distress, and anxiety. Paediatric Anaesthesia. 2009;19(12):1184-90.

593. Nord D, Belew J. Effectiveness of the essential oils lavender and ginger in promoting children's comfort in a perianesthesia setting. Journal of Perianesthesia Nursing. 2009;24(5):307-12.

594. Saha N, Saha DK, Rahman MA, Islam MK, Aziz MA. Comparison of post operative morbidity between laparoscopic and open appendectomy in children. Mymensingh Medical Journal. 2010;19(3):348-52.

595. Sethi S, Ghai B, Ram J, Wig J. Postoperative emergence delirium in pediatric patients undergoing cataract surgery - A comparison of desflurane and sevoflurane. Paediatric Anaesthesia. 2013;23(12):1131-7.

596. Singh R, Kharbanda M, Sood N, Mahajan V, Chatterji C. Comparative evaluation of incidence of emergence agitation and post-operative recovery profile in paediatric patients after isoflurane, sevoflurane and desflurane anaesthesia. Indian journal of anaesthesia. 2012;56(2):156-61.

597. Stuth EA, Berens RJ, Staudt S, Robertson FA, Scott JP, Stucke AG, et al. The effect of caudal vs intravenous morphine on early extubation and postoperative analgesic requirements for stage 2 and 3 single-ventricle palliation: a double blind randomized trial. Pediatric Anesthesia. 2011;21(4):441-53.

598. Takmaz SA, Uysal HY, Uysal A, Kocer U, Dikmen B, Baltaci B. Bilateral extraoral, infraorbital nerve block for postoperative pain relief after cleft lip repair in pediatric patients: a randomized, double-blind controlled study. Annals of Plastic Surgery. 2009;63(1):59-62.

599. Townsend JA, Ganzberg S, Thikkurissy S. The effect of local anesthetic on quality of recovery characteristics following dental rehabilitation under general anesthesia in children. Anesthesia Progress. 2009;56(4):115-22.

600. Vaughan M, Paton EA, Bush A, Pershad J. Does lidocaine gel alleviate the pain of bladder catheterization in young children? A randomized, controlled trial. Pediatrics. 2005;116(4):917-20.

601. Voepel-Lewis TD, Malviya S, Burke C, D'Agostino R, Hadden SM, Siewert M, et al. Evaluation of simethicone for the treatment of postoperative abdominal discomfort in infants. Journal of Clinical Anesthesia. 1998;10(2):91-4.

602. Zier JL, Rivard PF, Krach LE, Wendorf HR. Effectiveness of sedation using nitrous oxide compared with enteral midazolam for botulinum toxin A injections in children. Developmental Medicine and Child Neurology. 2008;50(11):854-8.

603. Taddio A, Nulman I, Goldbach M, Ipp M, Koren G. Use of lidocaine-prilocaine cream for vaccination pain in infants. Journal of Pediatrics. 1994;124(4):643-8.

604. Cohen LL, Bernard RS, McClelland CB, MacLaren JE. Assessing medical room behavior during infants' painful procedures: The measure of adult and infant soothing and distress (MAISD). Children's Health Care. 2005;34(2):81-94.

605. Verriotis M, Fabrizi L, Lee A, Ledwidge S, Meek J, Fitzgerald M. Cortical activity evoked by inoculation needle prick in infants up to one-year old. Pain. 2015;156(2):222-30.

606. Anninger W, Forbes B, Quinn G, Schreiner MS. The effect of topical tetracaine eye drops on emergence behavior and pain relief after strabismus surgery. Journal of American Association for Pediatric Ophthalmology and Strabismus. 2007;11(3):273-6.

607. Carbajal R, Biran V, Lenclen R, Epaud R, Cimerman P, Thibault P, et al. EMLA cream and nitrous oxide to alleviate pain induced by palivizumab (Synagis) intramuscular injections in infants and young children. Pediatrics. 2008;121(6):e1591-8.

608. Cohen LL. Reducing infant immunization distress through distraction. Health Psychology. 2002;21(2):207.

609. Cramer-Berness LJ, Friedman AG. Behavioral interventions for infant immunizations. Children's Health Care. 2005;34(2):95-111.

610. Dyer SL, Collins CT, Baghurst P, Saxon B, Meachan B. Insuflon versus subcutaneous injection for cytokine administration in children and adolescents: a randomized crossover study. Journal of Pediatric Oncology Nursing. 2004;21(2):79-86.

611. Fallah R, Gholami H, Ferdosian F, Binesh F. Evaluation of Vaccines Injection Order on Pain Score of Intramuscular Injection of Diphtheria, Whole Cell Pertussis and Tetanus Vaccine. Indian Journal of Pediatrics. 2016;83(12-13):1405-9.

612. Girish GN, Ravi MD. Vaccination Related Pain: Comparison of Two Injection Techniques. Indian Journal of Pediatrics. 2014;81(12):1327-31.

613. Hillgrove-Stuart J, Riddell RP, Horton R, Greenberg S. Toy-mediated distraction: Clarifying the role of distraction agent and preneedle distress in toddlers. Pain Research and Management. 2013;18(4):197.

614. Kass FC, Holman JR. Oral glucose solution for analgesia in infant circumcision. The Journal of Family Practice. 2001;50(9):785-8.

615. Kassab M, Sheehy A, King M, Fowler C, Foureur M. A double-blind randomised controlled trial of 25% oral glucose for pain relief in 2-month old infants undergoing immunisation. International Journal of Nursing Studies. 2012;49(3):249-56.

616. Kaur L, Kaur S, Kalia R, Bharti B. Analgesic effect of breast feeding in infants during immunization injections: Randomized control trial. Nursing and Midwifery Research. 2009;5(4).

617. Mularoni PP, Cohen LL, DeGuzman M, Mennuti-Washburn J, Greenwald M, Simon HK. A randomized clinical trial of lidocaine gel for reducing infant distress during urethral catheterization. Pediatric Emergency Care. 2009;25(7):439-43.

618. O'Brien L, Taddio A, Ipp M, Goldbach M, Koren G. Topical 4% amethocaine gel reduces the pain of subcutaneous measles-mumps-rubella vaccination. Pediatrics. 2004;114(6):e720-4.

619. Pathak P, Kalia R, Bharti B. Effect of needle gauge on perception of pain intensity among infants receiving DPT vaccination. Nursing and Midwifery Research. 2007;3(4).

620. Ram D, Amir E. Comparison of articaine 4% and lidocaine 2% in paediatric dental patients. International Journal of Paediatric Dentistry. 2006;16(4):252-6.

621. Sundar S, Ramesh B, Dixit PB, Venkatesh S, Das P, Gunasekaran D. Live Music Therapy as an Active Focus of Attention for Pain and Behavioral Symptoms of Distress During Pediatric Immunization. Clinical Pediatrics. 2016;55(8):745-8.

622. Taddio A, Ho T, Vyas C, Thivakaran S, Jamal A, Ilersich AF, et al. A randomized controlled trial of clinician-led tactile stimulation to reduce pain during vaccination in infants. Clinical Pediatrics. 2014;53(7):639-44.

APPENDICES

APPENDIX A:

Search terms

Appropriate scales

Box 1. Search terms used to identify systematic reviews providing pain assessment scale recommendations

Medline using Ovid	
1	Infant/
2	Child/
3	1 OR 2
4	Pain/
5	Limit 3 to pre 1986
6	Pain measurement/
7	Pain assessment.mp
8	Pain scale.mp
9	Pain tool.mp
10	5 OR 6 OR 7 OR 8 OR 9
11	3 AND 10
12	Limit 7 to (meta-analysis or systematic reviews)
13	Limit 12 to English
Embase using Ovid	
1	Infant/
2	Child/
3	1 OR 2
4	Pain measurement/
5	Pain assessment/
6	Pain scale.mp
7	Pain tool.mp
8	4 OR 5 OR 6 OR 7
9	3 AND 8
10	Limit 8 to (meta-analysis or “systematic review”)
11	Limit to English
PsychINFO using Ovid	
1	Infant.mp

2	Child.mp
3	Pediatrics/
4	1 OR 2 OR 3
5	Pain measurement/
6	Pain assessment.mp
7	Pain scale.mp
8	Pain tool.mp
9	4 OR 5 OR 6 OR 7 OR 8
10	4 AND 9
11	Limit to (meta-analysis or systematic review)
12	Limit to English
CINAHL using Ebsco	
1	infant
2	child
3	1 OR 2
4	Pain measurement
5	Pain assessment
6	Pain scale
7	Pain tool
8	4 OR 5 OR 6 OR 7
9	3 AND 8
10	Limiters: Publication Type: meta-analysis, systematic review
11	Limiters: Language: English

Box 2. Search terms used to identify studies assessing the psychometric properties of scales used to assess procedural pain in infants and children.

Medline using Ovid	
1	Infant/
2	Child/
3	1 OR 2
4	Pain/
5	Limit 3 to pre 1986
6	Pain measurement/
7	Pain assessment.mp
8	Pain scale.mp
9	Pain tool.mp
10	5 OR 6 OR 7 OR 8 OR 9

11	Psychometrics/
12	Validation studies/ OR validation studies/
13	“Reproducibility of results”/
14	Evaluation Studies/
15	“Sensitivity and specificity”/
16	feasibility studies/
17	validity.mp
18	reliability.mp
19	clinical utility.mp
20	feasibility.mp
21	11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20
22	3 AND 10 AND 21
23	Limit to English
Embase using Ovid	
1	Infant/
2	Child/
3	1 OR 2
4	Pain measurement/
5	Pain assessment/
6	Pain scale.mp
7	Pain tool.mp
8	4 OR 5 OR 6 OR 7
9	validation process/ or validation study/ or instrument validation/
10	intrarater reliability/ or interrater reliability/ or test retest reliability/ or reliability/
11	feasibility study/
12	“sensitivity and specificity”/
13	feasibility.mp
14	reliability.mp.
15	validation.mp
16	clinical utility.mp
17	9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16
18	3 AND 8 AND 17
19	Limit to English
PsychINFO using Ovid	
1	Infant.mp
2	Child.mp
3	Pediatrics/
4	1 OR 2 OR 3
5	Pain measurement/

6	Pain assessment.mp
7	Pain scale.mp
8	Pain tool.mp
9	4 OR 5 OR 6 OR 7 OR 8
10	Psychometrics
11	Test validity
12	Test reliability
13	Interrater reliability
14	10 OR 11 OR 12 or 13
15	4 AND 9 AND 14
16	Limit to English
CINAHL using Ebsco	
1	infant
2	child
3	1 OR 2
4	Pain measurement
5	Pain assessment
6	Pain scale
7	Pain tool
8	4 OR 5 OR 6 OR 7
9	Valid&
10	Reliab$
11	sensitivity
12	specificity
13	9 OR 10 OR 11 OR 12
14	3 AND 8 AND 13
14	Limiters: Language: English

Accepted scales

Box 3. Search terms used to identify randomised controlled trials using an observational behavioural pain scale to measure procedural pain in infants and/or children.

Medline using Ovid	
1	Infant/
2	Child/
3	1 OR 2
4	Painful procedure.mp
5	Medical procedure.mp
6	Therapeutic procedure.mp
7	diagnostic procedure.mp
8	intravenous catheter insertion.mp
9	intravenous cannula insertion.mp
10	venepuncture mp
11	venipuncture mp
12	venepuncture mp
13	venipuncture mp
14	EXP Blood specimen collection/
15	Phlebotomy/
16	EXP catheterisation/
17	Intubation, gastrointestinal/
18	Nasogastric tube insertion.mp
19	Manipulation orthopaedic/
20	Suture techniques/
21	EXP foreign bodies/
22	EXP immunisation/
23	EXP vaccination/
24	4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23
25	Pain/
26	Pain management/
27	25 or 26
28	3 AND 24 AND 27
29	Limit by publication (Clinical trial, all RCT)
30	Limit to English
Embase using Ovid	
1	infant/
2	child/

3	1 OR 2
4	procedures
5	painful procedure(s).mp
6	medical procedure.mp
7	therapeutic procedures
8	diagnostic procedure
9	intravenous catheter insertion.mp
10	intravenous cannula insertion.mp.
11	vein puncture/
12	venepuncture
13	venipuncture
14	nasogastric tube insertion.mp
15	blood sampling/
16	Phlebotomy/
17	catheterization/
18	digestive tract intubation/
19	orthopedic manipulation
20	fracture reduction/
	wound care/
21	foreign body/
22	immunization/
23	vaccination/
24	4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23
25	EXP pain/
26	analgesia
27	local anesthesia/ or regional anesthesia/
28	25 OR 26 OR 27
29	3 AND 24 AND 28
30	Limit by publication (Clinical trial, RCT)
31	Limit to English
PsychINFO via Ovid	
1	Infant.mp
2	Child.mp
3	Pediatrics/
4	1 OR 2 OR 3
5	therapeutic processes/ or medical treatment (general)/
6	drug therapy
7	physical treatment methods/

8	immunisation
9	5 OR 6 OR 7 OR 8 OR 9
10	Pain
11	pain management
12	10 OR 11
13	4 AND 9 AND 12
14	limit by publication (1200 meta-analysis or "2000 treatment outcome/clinical trial")
15	limit to English

CINAHL via Ebsco

1	Infant
2	Child, preschool
3	procedure
4	painful procedure
5	medical procedure
6	therapeutic procedure
7	diagnostic procedure
8	intravenous cannula insertion
9	intravenous catheter insertion
10	venepuncture
11	venipuncture
12	nasogastric tube insertion
13	blood specimen collection
14	phlebotomy
15	catheterisation
16	wound care
17	wound closure
18	suturing
19	wound drainage
20	foreign body removal
21	fracture reduction
22	immunisation
23	vaccination
24	pain
25	pain management
26	clinical trials
27	randomis(z)ed controlled trial
28	meta-analysis
29	1 or 2
30	3 or……23

31	24 or 25
32	26 or 27 or 28
33	29 and 30 and 31 and 32
34	limit to English
35	29 and 30 and 31
36	limit by publication
37	limit to English
38	34 or 38

Box 4. Search terms used to identify expert consensus guidelines and clinical practice guidelines making recommendations for observational behavioural pain scales to measure procedural pain in infants and/or children.

Medline using Ovid	
1	Infant/
2	Child/
3	Pain/
4	Pain measurement/
5	Consensus/
6	Clinical guideline/
7	1 OR 2
8	3 OR 4
9	5 OR 6
10	7 AND 8 AND 9
Embase using Ovid	
1	infant/
2	child/
3	pain assessment/
4	pain tool.mp
5	Consensus/
6	Clinical guideline/
7	1 OR 2
8	3 OR 4
9	5 OR 6
10	7 AND 8 AND 9
PsychINFO via Ovid	
1	Infant
2	Child

3	Pain assessment
4	Pain tool
5	Pain scale
6	Pain measurement
7	Consensus/
8	Clinical guideline/
9	1 OR 2
10	3 OR 4 OR 5 OR 6
11	7 OR 8
12	9 AND 10 AND 11
CINAHL via Ebsco	
1	Infant
2	Child, preschool
3	Pain assessment
4	Pain tool
5	Pain scale
6	pain measurement
7	Consensus/
8	Clinical guideline/
9	1 OR 2
10	3 OR 4 OR 5 OR 6
11	7 OR 8
12	9 AND 10 AND 11

Box 5. Association, society, academy, collaboration, organisation and network sites searched to identify expert consensus statements and clinical practice guidelines

Academic Paediatrics Association
American Academy of Emergency Medicine
American Academy of Emergency Nurse Practitioners
American Academy of Pediatrics *
American College of Emergency Physicians
American College of Pediatrics
American Pain Society *
American Society for Anesthesiologists
American Society for Pain Management Nursing
American Society of Regional Anesthesia and Pain Medicine
Association for British Paediatric Nurses
Association for the Wellbeing of Children in Healthcare
Association of Paediatric Anaesthetists of Great Britain and Ireland *
Australasian College for Emergency Medicine
Australia College of Children and Young Peoples Nurses
Australian and New Zealand College of Anaesthetists*
Australian College of Emergency Nursing
Australian College of Paediatric and Child Health Nurses
Australian Paediatric Society
Australian Pain Society
Canadian Anesthesiologists' Society
Canadian Association of Emergency Physicians
Canadian Paediatric Association
Canadian Pain Coalition
Canadian Pain Society
Canadian Pediatric Anesthetist Society
Centre for Evidence-Based Pharmacotherapy
Centre for Pediatric Pain Research
Centres for Reviews and Dissemination (York University)
Childkind International
CIHR Team in Children's Pain
Cochrane Collaboration
College of Emergency Nurses, Australasia
College of Nursing (Australia)
Emergency Nurses Association
European Academy of Paediatrics
European Observatory on Health Care for Chronic Conditions
European Society for Anaesthesiology
European Society for Emergency Medicine
European Society for Emergency Nurses
European Society for Paediatric Anaesthesiology *
Faculty of Emergency Nursing
Faculty of Paediatrics (Royal College of Physicians of Ireland)
Guidelines International Network
Indian Academy of Paediatrics

Institute for Clinical Evaluative Sciences
Institute for Clinical Systems Improvement:
Institute of Medicine Report on Pain (USA)
International Association for Hospice and Palliative Care
International Association for the Study of Pain (IASP) *
International Pediatric Association
Joanna Briggs Institute for Evidence Based Nursing & Midwifery
National Emergency Nurses Association (Canada)
National Guidelines Clearinghouse (US)
National Institute for Health and Clinical Excellence (NICE)
New Zealand Guidelines Group
Paediatric Society of New Zealand
Pain Assessment and Management Initiative
Royal Australian College of Physicians *
Royal College of Anaesthetists (UK)
Royal College of Emergency Medicine
Royal College of Nursing (UK) *
Royal College of Paediatrics and Child Health *
Scottish Intercollegiate Guidelines Network
Society for Academic Emergency Medicine
Society of Academic Anesthesia Associations
Society of Pediatric Nurses (US)
The University of Toronto Centre for the Study of Pain
United States Preventive Service Task Force
University of Wisconsin Comprehensive Cancer Centre - Pain & Policy Studies Group
World Health Organization

* Organisations that author/endorse statements/guidelines making scale recommendations for procedural pain assessment.

APPENDIX B

COSMIN checklist

Step 1	Step 2	Step 3	Step 4
Mark properties assessed Internal consistency ☐ Reliability ☐ Measurement error ☐ Content validity ☐ Construct validity ☐ Structural validity ☐ Hypothesis testing ☐ Cross-cultural validity ☐ Criterion validity ☐ Responsiveness ☐ Interpretability ☐	**Have IRT methods been used** Yes ☐ Complete IRT box No ☐	**Complete associated box for each property marked in Step 1**	**Complete Generalizability box for each property marked in Step 1**

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

Box General requirements for studies that applied Item Response Theory (IRT) models					
		Exc	Good	Fair	Poor
1	Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	☐	☐		
2	Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	☐	☐		
3	Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	☐	☐		
4	Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	☐	☐	☐	

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency					
		yes	no	?	
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?	☐	☐	☐	
Design requirements		Exc	Good	Fair	Poor
2	Was the percentage of missing items given?	☐	☐		
3	Was there a description of how missing items were handled?	☐	☐	☐	
4	Was the sample size included in the internal consistency analysis adequate?	☐	☐	☐	☐
5	Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	☐	☐	☐	☐
6	Was the sample size included in the unidimensionality analysis adequate?	☐	☐	☐	☐
7	Was an internal consistency statistic calculated for each (unidimensional) (sub) scale separately?	☐			☐
8	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					
9	For Classical Test Theory (CTT), continuous scores: Was Cronbach's alpha calculated?	☐		☐	☐
10	For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	☐		☐	☐
11	For IRT: Was a goodness of fit statistic at a global level calculated? e.g. $\chi 2$, reliability coefficient of estimated latent trait value (index of (subject or item) separation)	☐			☐
	FINAL ASSESSMENT	☐	☐	☐	☐

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)					
Design requirements		Exc	Good	Fair	Poor
1	Was the percentage of missing items given?	☐	☐		
2	Was there a description of how missing items were handled?	☐	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐	☐
4	Were at least two measurements available?	☐			☐
5	Were the administrations independent?	☐	☐	☐	☐
6	Was the time interval stated?	☐		☐	
7	Were patients stable in the interim period on the construct to be measured?	☐	☐	☐	☐
8	Was the time interval appropriate?	☐		☐	☐
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	☐	☐	☐	☐

10	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					
11	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	☐	☐	☐	☐
12	For dichotomous/nominal/ordinal scores: Was kappa calculated?	☐			☐
13	For ordinal scores: Was a weighted kappa calculated?	☐		☐	☐
14	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	☐	☐		
	FINAL ASSESSMENT	☐	☐	☐	☐

Box C. Measurement error: absolute measures					
Design requirements		Exc	Good	Fair	Poor
1	Was the percentage of missing items given?	☐	☐		
2	Was there a description of how missing items were handled?	☐	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐	☐
4	Were at least two measurements available?	☐			☐
5	Were the administrations independent?	☐	☐	☐	☐
6	Was the time interval stated?	☐		☐	
7	Were patients stable in the interim period on the construct to be measured?	☐	☐	☐	☐
8	Was the time interval appropriate?	☐		☐	
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	☐	☐	☐	☐
10	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					
11	For CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	☐	☐		☐
	FINAL ASSESSMENT	☐	☐	☐	☐

Box D. Content validity (including face validity)					
General requirements		Exc	Good	Fair	Poor
1	Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	☐		☐	☐
2	Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	☐	☐	☐	☐
3	Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	☐	☐	☐	

4	Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	☐		☐	☐
5	Were there any important flaws in the design or methods of the study?	☐		☐	☐
	FINAL ASSESSMENT	☐	☐	☐	☐

Box E. Structural validity					
		Yes	No	?	
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?	☐	☐	☐	
Design requirements		Exc	Good	Fair	Poor
2	Was the percentage of missing items given?	☐	☐		
3	Was there a description of how missing items were handled?	☐	☐	☐	
4	Was the sample size included in the analysis adequate?	☐	☐	☐	☐
5	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					
6	For CTT: Was exploratory or confirmatory factor analysis performed?	☐	☐		☐
7	For IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?	☐			☐
	FINAL ASSESSMENT	☐	☐	☐	☐

Box F. Hypotheses testing					
Design requirements		Exc	Good	Fair	Poor
1	Was the percentage of missing items given?	☐	☐		
2	Was there a description of how missing items were handled?	☐	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐	☐
4	Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	☐	☐	☐	☐
5	Was the expected *direction* of correlations or mean differences included in the hypotheses?	☐	☐		
6	Was the expected absolute or relative *magnitude* of correlations or mean differences included in the hypotheses?	☐	☐		
7	For convergent validity: Was an adequate description provided of the comparator instrument(s)?	☐	☐	☐	☐
8	For convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	☐	☐	☐	☐
9	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					

10	Were design and statistical methods adequate for the hypotheses to be tested?	☐	☐	☐	☐
	FINAL ASSESSMENT	☐	☐	☐	☐

Box G. Cross-cultural validity					
Design requirements		Exc	Good	Fair	Poor
1	Was the percentage of missing items given?	☐	☐		
2	Was there a description of how missing items were handled?	☐	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐	☐
4	Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	☐			☐
5	Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	☐	☐	☐	
6	Did the translators work independently from each other?	☐	☐	☐	☐
7	Were items translated forward and backward?	☐	☐	☐	☐
8	Was there an adequate description of how differences between the original and translated versions were resolved?	☐	☐		
9	Was the translation reviewed by a committee (e.g. original developers)?	☐	☐		
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	☐	☐	☐	☐
11	Was the sample used in the pre-test adequately described?	☐		☐	
12	Were the samples similar for all characteristics except language and/or cultural background?	☐	☐	☐	☐
13	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					
14	For CTT: Was confirmatory factor analysis performed?	☐			☐
15	For IRT: Was differential item function (DIF) between language groups assessed?	☐			☐
	FINAL ASSESSMENT	☐	☐	☐	☐

Box H. Criterion validity					
Design requirements		Exc	Good	Fair	Poor
1	Was the percentage of missing items given?	☐	☐		
2	Was there a description of how missing items were handled?	☐	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐	☐

4	Can the criterion used or employed be considered as a reasonable 'gold standard'?	☐	☐	☐	☐
5	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					
6	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?	☐			☐
7	For dichotomous scores: Were sensitivity and specificity determined?	☐			☐
	FINAL ASSESSMENT	☐	☐	☐	☐

Box I. Responsiveness					
Design requirements		Exc	Good	Fair	Poor
1	Was the percentage of missing items given?	☐	☐		
2	Was there a description of how missing items were handled?	☐	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐	☐
4	Was a longitudinal design with at least two measurement used?	☐			☐
5	Was the time interval stated?	☐			☐
6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	☐	☐	☐	
7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	☐	☐	☐	☐
Design requirements for hypotheses testing where gold standard not available					
8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	☐		☐	☐
9	Was the expected *direction* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	☐	☐		
10	Were the expected absolute or relative *magnitude* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	☐	☐		
11	Was an adequate description provided of the comparator instrument(s)?	☐		☐	☐
12	Were the measurement properties of the comparator instrument(s) adequately described?	☐	☐	☐	☐
13	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					
14	Were design and statistical methods adequate for the hypotheses to be tested?	☐		☐	☐
Design requirement for comparison to a gold standard: where gold standard is available					

15	Can the criterion for change be considered as a reasonable gold standard?	☐	☐	☐	☐
16	Were there any important flaws in the design or methods of the study?	☐		☐	☐
Statistical methods					
17	For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	☐			☐
18	For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	☐			☐
	FINAL ASSESSMENT	☐	☐	☐	☐

Box J. Interpretability		
1	Was the percentage of missing items given?	
2	Was there a description of how missing items were handled?	
3	Was the sample size included in the analysis adequate?	
4	Was the distribution of the (total) scores in the study sample described?	
5	Was the percentage of the respondents who had the lowest possible (total) score described?	
6	Was the percentage of the respondents who had the highest possible (total) score described?	
7	Were scores and change scores (i.e. means and SD) presented for relevant (sub) groups? e.g. for normative groups, subgroups of patients, or the general population	
8	Was the minimal important change (MIC) or the minimal important difference (MID) determined?	
9	Were there any important flaws in the design or methods of the study?	

Step 4: Determining the Generalisability of the results

Box Generalisability		
Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of:		
1	Median or mean age (with standard deviation or range)?	
2	Distribution of sex?	
3	Important disease characteristics (e.g. Severity, status, duration) and description of treatment?	
4	Setting(s) in which the study was conducted? E.g. General population, primary care or hospital/rehabilitation care	
5	Countries in which the study was conducted?	
6	Language in which the HR-PRO instrument was evaluated?	
7	Was the method used to select patients adequately described? E.g. Convenience, consecutive, or random	
8	Was the percentage of missing responses (response rate) acceptable?	

APPENDIX C

Table 1. VAS psychometric evaluation study details.

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
Validation as index scale						
Berntson et al, 2001 (413)	Observational study. To evaluate the concordance between pain assessments made on a VAS & on a 4-point verbal descriptor scale (VDS-4). Parents & children aged over 10 years independently rated the pain experienced in the last week using pain scales. Scales applied in random order.	26 children aged 2 – 18 years. Pain: disease related pain (Juvenile arthritis). Scales: VASobs (parent) VDS-4obs (parent), VAS (child) VDS-4 (child) GRS. Setting: paediatric ward (Sweden).	Not tested	Criterion: Children rated pain higher than parents using the VAS & 27% of all possible pairs of observations were disordered (D = 0.27, MA = 0.46). Children rated pain lower than parents using the VDS-4 scale (D = 0.05, MA = 0.91). COSMIN – Poor.	Not tested	Only 12 children provided pain scores.
de Jong et al, 2010 (417)	Observational study. To investigate whether the POCIS, the COMFORT-B & the nurse observational VASobs are reliable, valid & clinically useful instruments to measure pain in children with burns aged 0–5 years. Two nurses independently & simultaneously applied scales at baseline &	154 children aged 0 – 5 years. 102 nurses rated. Pain: procedural & background pain (burn care). Index: VASobs (nurse), POCIS, COMFORT-B. Reference: NA.	Inter-rater: ICC for procedural pain VASobs = 0.60 (CI 0.55 – 0.65), POCIS = 0.81 (CI 0.78 – 0.84) & COMFORT-B = 0.82 (CI 0.80 – 0.85) & background pain VASobs 0.55 (CI 0.51 – 0.59), POCIS = 0.75 (CI 0.72 – 0.77) & COMFORT-B = 0.83 (CI 0.82 – 0.85). COMSIN – Fair.	Not tested*	Not tested	Sequencing of POCIS & COMFORT-B altered, VAS always last – therefore potentially influenced by results from application of other scales. Validity of the scale for other populations assumed.

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	associated with procedure (peak pain & overall pain).	Setting: burn centre (Netherlands).				*Validity testing not completed due to poor reliability results.
de Jong et al, 2005 (331)	Observational study. To assess if the pain observation scale for young children (POCIS) & the visual analogue scale (VAS) are reliable & valid instruments to measure procedural & background pain in burned children aged 0–4 years. Sample of nurses rated pain using two scales for 24 segments of video of children undergoing a procedure or at rest. The VAS was applied first followed by the POCIS. 2 months later sample completed re-test, applying scales to video segments a second time.	24 children aged 0-4 years. 73 nurses (grouped by hospital A & B) rated 24 video segments. Pain: procedural & background pain (burn care). Index: VASobs (nurse), POCIS. Reference: NA. Setting: burn centre (Netherlands).	Inter-rater: ICC for Group A procedural VASobs = 0.56 (CI 0.38 – 0.79) & POCIS = 0.40 (CI 0.22 – 0.72) & background VASobs = 0.52 (CI 0.20 – 0.98) & POCIS = 0.97 (CI 0.89 – 0.99). ICC for Group B procedural VASobs = 0.64 (CI 0.43 – 0.87) & POCIS = 0.79 (CI 0.63 – 0.93) & background VASobs = 0.59 (CI 0.27 – 0.98) & POCIS = 0.65 (CI 0.32 – 0.99). COSMIN – Fair. Intra-rater: ICC for Group A procedural VASobs = 0.52 (CI 0.41 – 0.61) & POCIS = 0.53 (CI 0.43 – 0.62) & background VASobs = 0.70 (CI 0.56 – 0.80) & POCIS = 0.97 (CI 0.95 – 0.98). ICC for Group B procedural VASobs = 0.82 (CI 0.76 – 0.86) & POCIS = 0.88 (CI 0.84 – 0.91) & background VASobs= 0.75 (CI 0.62 – 0.84) &	Not tested*	Not tested	*Validity testing not completed due to poor reliability results. VAS assessed first

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
			POCIS = 0.74 (CI 0.60 – 0.83). COMSIN – Fair.			
Eyelade et al, 2009 (418)	Observational study. To determine the convergent validity of the Oucher, OPS, VAS & the NRS among Nigerian children aged 6months to 12 years who required venepuncture/ phlebotomy. Four pain scales applied to each child by 3 researchers at baseline, during & after procedure to assess pain.	179 children aged 6 months to 12 years. Exc: developmental delay, altered sensorium, clinical instability. Pain: procedural (venepuncture/ phlebotomy. Index: VASobs (researchers), NRS, Oucher, OPS. Reference: NA. Setting: children's outpatient department (Nigeria).	Inter-rater: VASobs ICC = 0.727, COMSIN – Fair.	Hypothesis: Convergent – VASobs correlation with Oucher before r = 0.87 (p , 0.0001) & during r = 0.63, p < 0.0001) COSMIN - Fair Responsiveness: Increase in scores for all scales – no significance testing reported) COSMIN - Poor	Not measured	Authors state their premise that if scales agree they must all be measuring pain – flawed assumption all derived from similar criteria etc. Limited description of scale application procedure – unclear whether VAS was self-report or observer, all observers applied each scale to each phase of procedure for all children (therefore responsiveness biased), independence not described. Language of scales unclear.
Filocamo et al, 2010 (419)	Observational study. To evaluate the measurement properties of 21-numbered circle VAS & traditional 10-cm horizontal line VAS for physician & parent subjective ratings in children with juvenile idiopathic arthritis. During clinic appointment, children & parents	397 children*. Pain: disease related pain (juvenile arthritis). Index scale: VASobs (parent). Reference: Parent Global Wellbeing score, Physician Global Disease Activity score, Childhood	Not tested	Hypothesis: Convergent – VASobs correlated with MD global score (r = 0.61), parent global (r = 0.82), functional scale (r = 0.58), CHAQ (r = 0.54), CHQ (r = -0.75 & -0.24, physical & psychosocial respectively), COSMIN – Good.	Not measured	Disease specific Ratings by parents (global & pain) likely to correlate strongly. * age (mean or range) not reported. Unclear if tool translated into Italian.

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	completed summary pain assessments describing pain over the last week. Physicians completed disease activity score.	Health Assessment Questionnaire (CHAQ), Child Health Questionnaire (CHQ) Setting: study units (Italy).				
Garcia-Munitis et al, 2006 (420)	Observational study. To investigate the level of agreement between patients, mothers, fathers, & physicians in rating pain intensity in juvenile idiopathic arthritis & to identify factors explaining discrepancies between raters. Prior to clinic appointment, child & parents provided independent ratings of current pain & pain experienced in the last week. Following the physical examination, the physician & an observer (independent physician) rated the child's current pain.	94 children aged 5 - 18 years with juvenile arthritis. Pain: disease related pain (juvenile arthritis). Index scale: VASobs (mother, father, physician, independent physician), VAS self-report. Setting: outpatient clinic (Italy).	Inter-rater: correlation between btw mother & father r = 0.73, mother & attending doctor r = 0.51, father & attending r = 0.47 & attending & independent physician r = 0.94, COSMIN – Fair.	Hypothesis: convergent correlation between pain & overall well-being for observers rating scales on the same form & those rating them on separate forms were 0.93 & 0.79 (p = 0.005) respectively, for the mothers & 0.89 & 0.73 (p = 0.02) respectively, for the fathers. COSMIN – Poor. Criterion: correlations btw child & mother r = 0.45, child & father r = 0.31, child & attending doctor r = 0.26 & child & independent physician r = 0.24, COSMIN – Fair.	Not measured	Disease specific Ratings by parents (global wellbeing & pain) likely to correlate strongly. Evidence that scoring similar constructs likely to influence the scores for each reported Power calculation completed to determine reliability sample size. Unclear if tool translated into Italian.
Hirschfeld et al, 2013 (421)	Observational study. To use a method similar to that developed by Serlin & colleagues to establish optimal cut-offs for mild, moderate, & severe pain	2276 children aged 3 – 10 years*. Pain: acute and chronic. Index: VASobs (parent).	Not tested	Hypothesis: convergent = correlations between pain & disability for children (r = 0.42; 95% CI 0.38–0.45; P < 0.001), COSMIN – Fair.	Optimal cut-off points identified for VASobs mild = 25 & moderate = 60. Optimal cut-off points for self-reported VAS mild = 43, mod = 73.	Verbal anchor point for '0' was 'hardly noticeable pain'. Self-reported VAS data collected from adolescents & results

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score) Reliability	Validity	Feasibility & clinical utility	Comments
	using the VAS to be used in population-based paediatric research. Study analyses data collected in a larger study. Random community sample provided self-administered surveys to capture health-related data, including pain related data. Subset who reported pain in the last 3 months include in this analysis. Parental report of pain used in children less than 11 years.	Setting: nationwide survey (Germany). * Study reports self-report data for 11 – 17 years.				consistent with those where pain scored by parents. Unclear if tool translated into German.
Huijer Abu-Saad et al, 1995 (411)	Observational study. To describe, using the PPAT, the pain experienced by children with juvenile arthritis. Following physical examination, physician rated pain & then administered PPAT, which includes the VAS to the child. PPAT was administered to parents independently.	33 children aged 7 – 16 years. Pain: disease related pain (Juvenile arthritis). Index: VASobs (parent, physician). Setting: outpatient department – two academic teaching hospitals (Netherlands).	Inter-rater: Spearman rank order correlation for parent & physician VASobs score not significant (r = 0.10). COSMIN – Fair.	Criterion: Correlations between child & parents were significant for present pain ($r = 0.53$, $p < 0.001$)) & worst pain ($r = 0.77$, $p < 0.05$)) & correlations between child & doctor were significant ($r = 0.32$, $p < 0.05$), COSMIN – Fair.	Not tested	Physician rated pain & then administered PPAT which includes the VAS to the child – therefore potentially biasing their result.
Jylli et al, 1995 (422)	Observational study. To assess procedural pain as perceived by the child, but also by the parents & nurses,	129 infants & children aged less than 16 years. Pain: procedural pain.	Inter-rater: Proportion considered to be in pain higher for parents than nurses (60% vs 77%, p < p <0.005). On 27% of	Criterion (n = 96): Median pain scores higher in children compared with scores from nurses (34 vs 10, $p < 0.001$) & parents	Not tested	Not blinded to circumstances & independence of raters scores & timing of scores not made explicit

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	& thus to identify situations & procedures associated with an unacceptable level of pain. Nurses, parents & children rated pain on admission to ED, administration of analgesic & during the procedure.	Scales: VASobs (parent, nurse), VAS (child aged > 10y) Smiley Five Faces Scale (3 – 9yr). Setting: ED (Sweden).	occasions nurse & parents differed in their assessment of the child's pain, COSMIN – Poor.	(34 vs 26, p < 0.01). Correlation between parents & child scores (r = 0.33). COSMIN – Good.		Reliability analysis did not use appropriate measures of agreement. Unclear if tool translated into Swedish.
Kelly et al, 2002 (423)	Prospective, two-group, repeated measures, blinded study. To determine whether parent & child VAS scores for the pain of acute conditions in the child agreed sufficiently for these methods of measurement to be used interchangeably in pain & analgesia research.	78 children aged 8 – 15 years. Exc: conscious or cognitive impairment. Pain: acute illness/injury related pain. Scale: VASobs (parent), VAS (self-report). Setting: ED (Australia).		Criterion: (concurrent) Correlation between child & parent VAS pain scores 6 = 0.63, (95% CI, 0.56–0.70). Bias plot analysis: bias = 5mm. 95% limits of agreement from -38 to 47 mm.	Not tested	
Knutsson et al, 2006 (441)	Observational study. To evaluate the correlation between the parents & the health care professionals regarding how postoperative pain is estimated & to identify age & gender differences regarding the pain after adenoidectomy. Parents & nurse independently scored pain	100 children aged 3 – 10 years. Exc: mental retardation, current pain medication. Pain: post-operative. Scales: VASobs (nurse, parent) Wong-Baker FACES.	Inter-rater: Correlation between parents & nurses 0.66 at 10min & 0.672 at 30min (p = 0.01) Parent scores higher than nurse scores at 10min postoperatively (49.02 vs 35.45, p< 0.001) & 30min postoperatively (40.79 vs 27.95, p < 0.001). COSMIN – Good.	Criterion: (Concurrent) Correlation at 30min between parent & child r = 0.27, p = 0.03 & nurse r = 0.595, p = 0.01. COSMIN – Good. Responsiveness: Mean score 10 min post-operatively higher than 30min post for nurses (35.45 vs 27.95, p = 0.051) & parents (49.02	Not tested	2 nurses used to reduce inter-observer differences in nurse VAS scoring. Parents & nurse aware of the circumstances, may impact on scores & therefore responsiveness results. Swedish speaking.

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	10min & 30 min after discharge from the operating room. Child scored their pain at 30min unless asleep.	Setting: recovery room (Sweden).		vs 40.79, $p < 0.001$) COSMIN – Poor..		
McClellan et al, 2009 (425)	Prospective pre-post study. To evaluate the psychometric properties of 4 measures of acute pain in youth with sickle cell disease (SCD) during a medical procedure. Children & parents scored pain pre & post venepuncture while in the clinic. Heart rates were recorded at the same time by the phlebotomist. The procedure was videotape recorded & two undergraduate coders applied the m-OSBD to the two time points	48 children aged 2 – 17 years (sickle cell disease). Pain: procedural (venepuncture). Index: VASobs (parent), Wong-Baker Faced Scale, m-OSBD, HR. Setting: haematology clinic (US).	Not tested	Hypothesis: Convergent – VASobs reactivity scores correlated with child report (0.43, $p < 0.050$), mOSBD (0.33, $p < 0.05$) & HR (0.38, $p < 0.05$). COSMIN – Good. Responsiveness: mean VASobs scores inc post venepuncture (3.2, SD 6.6 vs 29.5 SD 28.7, $t(44) = 6.25$, $p < 0.001$). Self-report scores (n = 32) inc post venepuncture (0.7 SD 1.5 vs 3.2 SD 2.8, $t(35) = 5.41$, $p < 0.001$) COSMIN – Poor.	Not tested	Children & observers aware of circumstances.
McNair et al, 2004 (426)	Prospective, repeated measures correlational study. To compare the convergent validity of two measures of pain PIPP & CRIES in real life postoperative pain assessment in infants. Three observers rated pain using one each of the tools	51 neonates. Pain: postoperative. Index: PIPPS, CRIES. Reference: VASobs (expert nurse).	Not tested	Hypothesis: convergent – correlations between VASobs, PIPP & CRIES scores varied (ICC ranged from 0.07 – 0.88 at times from immediately – to 72hrs postop) COSMIN – Good. Responsiveness: VASobs scores (n = 45) decreased	Not tested	Raters aware of circumstance & one observer (PIPP or CRIES) was patient's bedside nurse. No significance testing for responsiveness. Timing of daily VAS varied but scheduled with PIPP & CRIES scoring.

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	(researcher used VAS for all observations) at regular intervals; immediately after surgery, PIPP & CRIES every 4 hours for 24 hours & then every 8 hours & before & after analgesic administration, VAS once every 24 hours.	Setting: NICU university affiliated hospital (Canada).		over 1st 12hrs postop, remained low until 48hrs & then rose slightly. Slope of change for each scale correlated COSMIN – Poor.		
Miller et al, 1996 (427)	Descriptive correlation study. To identify whether nurses & mothers of paediatric patients accurately assess the child's pain intensity as determined by the child. Nurse, parent & child rated pain independently rated pain using the VAS at three times between 8 & 28 hours post-operatively.	20 children aged 7 – 11 years. Pain: postoperative. Index: VASobs (mother). Reference: VAS (child). Setting: (US).	Inter-rater reliability: Pearson's correlation coefficients for mother/ nurse scores across 3 occasions - 0.36, p = 0.12, 0.55, p = 0.01 & 0.47, p = 0.07, COSMIN – Fair.	Criterion: Correlations for mother/child across 3 occasions – 0.71, p = 0.0005, 0.83, p = 0.0001 & 0.46 p = 0.07 & nurse/child 0.50, p = 0.02, 0.54, p = 0.01 & 0.23, p = 0.39, COSMIN – Poor.	Not tested	Repeat scores may result in reduced independence of child/mother scores – possible for them to confer about results of previous score. Observers aware.
Romsing et al, 1996 (429)	Observational study. To examine the relationship between children's ratings of their pain & the nurses' rating of the children's pain in a Danish Hospital. Children & 2 nurses independently & simultaneously scored pain the day after surgery just	100 children aged 3-15 years. Pain: postoperative (tonsillectomy). Index: VASobs (nurse). Reference: Poker Chip Tool (PCT).	Interrater: correlations between 2 nurse reviewers r = 0.52 – 0.60, p < 0.001. COSMIN – Good.	Criterion: correlations between child & nurse scores significant (r = 0.35 – 0.43, p < 0.001). COSMIN – Fair. Responsiveness: The change in pain scores (before & after analgesic) more pronounced for nurse VASobs scores 53 – 58% than child's PCT	Not tested	Reviewers not blinded to circumstances which may impact on responsiveness data.

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	before analgesics were given & 2 hours after administration.	Setting: hospital (Denmark).		scores 17% ($p < 0.001$). COSMIN – Poor.		
Singer et al, 2002	Prospective, descriptive study. To determine whether assessments of pain severity by children aged 4–7 years correlate with similar assessments made by their parents & health care practitioners. Children, parents & nurses provided a pain rating at one of two time points; for children having a painful procedure score provided immediately following procedure & for children with acute painful condition score provided following practitioern assessment.	57 children aged 4-7 years. Pain: acute & procedural pain. Index: VASobs (parent & clinician). Reference: Smiley Analogue Scale (SAS). Setting: ED (US).	Interrater: correlation between parent & practitioner 0.04 ($p = 0.001$) COSMIN – Good.	Criterion: correlation between child & parent VASobs scores 0.47 ($p < 0.001$) & child & clinician 0.008, ($p = 0.54$), COSMIN – Good.	Not tested	
Taddio et al, 2009 (433)	RCT – double blinded placebo controlled trial. To test the reliability & validity of observer-rated pain in infants undergoing immunization using the visual analogue scale (VAS).	120 infants aged 1-year old Convenience sample from an RCT. Pain: procedural (immunisation).	Inter-rater: ICC values range from 0.55 (95% CI 0.27 – 0.74) to 0.97 (95% CI 0.84 – 0.99). Intra-rater: kappa ranged 0.69 to 0.91, where cut-off ≥ 30mm kappa ranged from 0.35 to 0.91.	Hypothesis: known groups – mean scores for non-physician raters lower in the amethocaine group (15.1, SD 19.8 vs 19.5 SD 19.2, $p = 0.025$). No difference in mean scores for physician raters between groups. COSMIN - Good	Percentage of scores that differed by more than 20mm for real time versus video review assessments across raters ranged from 4.5 – 14.3%.	Intra-rater reliability – circumstances of 1st & 2nd review differ (real-time versus video). Criterion: more appropriately defined as convergent validity Responsiveness testing: raters not blinded to

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	Participants randomised to 2 groups: 4% amethocaine topically Placebo topically Immunisation video-taped, VAS scores allocated real time & from video. MBPS scores allocated to video. Raters repeated all assessments for same child on a 2nd occasion.	Index: VASobs (physician n=2, nurse, graduate student). Reference: MBPS*. Setting: paediatric outpatient clinic (Canada).	4.5% to 14.29% of rater scores varied more than 20mm. COSMIN – Poor.	Criterion: VAS correlations with MBPS range from 0.81 – 0.94 using Pearson's rho. COSMIN – Poor.		circumstances resulting in potential bias * VASobs used in original study to support validation of MBPS – circular logic to criterion validation. ** methods for RCT described.
Varni et al, 1987 (437)	Observational study. To utilise the PPQ-Child Form to assess chronic musculoskeletal pain in children with juvenile rheumatoid arthritis. Children were selected from clinic database & data collected at a clinic appointment. Researcher independently administered PPQ to parent & child. Physician also independently completed a pain score following the physical examination.	25 children aged 4 – 16 years. Pain: disease related pain (Juvenile arthritis). Index scale: PPQ-Child form inc. VASobs (parent, physician). Setting: rheumatology clinic at children's hospital (US).	Inter-rater reliability: significant correlation between parent & physician VASobs scores ($r = 0.85$, $p < 0.001$), COSMIN – Poor.	Hypothesis: Convergent - Physician rated disease activity index increases corresponded with increase in child, parent & physician pain scores (no significance testing) COSMIN – Poor. Criterion: significant correlations between child pain scores & parents VASobs scores for current pain ($r = 0.72$, $p < 0.001$) & worst pain ($r = 0.54$, $p < 0.013$) & between child pain scores & physician VASobs scores for present pain ($r = 0.65$, $p < 0.001$). Paired t-tests showed no difference between scores, COSMIN – Poor.		

Study	Study aim / design	Sample/Circumstance/ Scale (modification	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
Wilson et al, 1996 (439)	Observational study. To assess the suitability for use by parents of three pain assessment scales commonly used in paediatric anaesthesia Children simultaneously assessed by parent & doctor using three scales on two occasions (on awakening & 1 hour later).	40 children aged 2 – 11 years. Pain: postoperative (general & ENT). Scales: OPS, 4-point numerical score VASobs (Parent & single medical observer). Setting: recovery (Scotland).	Interrater reliability: correlation between VASobs parent & medical observer scores at time 1 = 0.69, $p < 0.01$ (OPS = 0.77) & time 2 = 0.73, $p < 0.01$ (OPS = 0.81). 95% CI for the limits for the difference btw parent & medical VASobs scores -7 to -15mm, COSMIN – Good.	Not tested	Not tested	Sequencing of scores for either observer unclear.

Abbreviations: CRIES - crying, requires oxygen, increased vital signs, expression, & sleepless, ED – emergency department, ENT – Ears, nose & Throat, FLACC – Face, Legs Activity Cry and Consolability, HR – Heart Rate, ICC – intraclass correlation, ICU – intensive care department, IVH – intraventricular haemorrhage, MBPS – Modified behavioural Scale, mOSBD – modified Observational Scale for Behavioural Distress, NICU – neonatal intensive care department, NRS – Numeric Rating Scale, OPS – Observational Pain Scale, PIPP- Premature Infant Pain Profile, POCIS - Pain Observation Scale for Young Children, PPAT - Pediatric Pain Assessment Tool, PPQ – Pediatric Pain Questionnaire, RCT – Randomised Controlled Trial, SD – Standard Deviation, US – United States, VAS - Visual Analogue Scale, VASobs – VAS observer.

Table 2. VAS RCT details

Study	Design/Aim	Subjects/Circumstances/ Setting	Intervention / Pain measures	Results	Quality score/Comments
Abuelkheir et al, 2014 (350)	Randomized, double-blind, placebo-controlled study. To evaluate the effectiveness of topical eutectic mixture of local anaesthetics (EMLA) cream in reducing the pain associated with vaccination injections.	216 children aged 2 months to 6 years. Exc: analgesic or sedative in last 12hrs. Pain: procedural (Immunisation). Setting: well baby paediatric clinic (Saudi Arabia).	2 groups – participants randomised to: Treatment group: EMLA cream. Control group: placebo cream. Pain scoring: VASobs (nurse) MBPS, crying.	VASobs scores at needle prick & after injection lower in the EMLA group (1.60 ± 1.67 versus 3.24 ± 2.01; 3.29 ± 2.27 versus 4.86 ± 2.20; respectively). Pre & post MBPS scores lower in the EMLA group (2.56 ± 1.96 versus 3.95 ± 2.20, respectively). Numbers who cried after the vaccination were lower in the EMLA group: 77.64% versus 92.7% (p=0.002). Total crying time shorter in the EMLA group (24.8 ± 20.6 s versus 43.3 ± 20.5 s, p<0.001).	Jadad score = 5.
Babl et al, 2009 (351)	Randomised, double blind placebo controlled trial. To investigate the role of nebulized lidocaine in reducing pain & distress of nasogastric tube insertion in young children.	36* children aged 1 – 5 years. Exc: chronic disease, epilepsy, neurological disease, cognitive impairment. Pain; procedural (nasogastric tube insertion). Setting: ED (Australia).	2 groups – participants randomised to: Treatment group - nebulized 2% lidocaine at 4 mg/kg Placebo group – equivalent volume of normal saline placebo. Administered via nebuliser 10minutes prior to NGT insertion. Pain scoring: FLACC, secondary - VASobs pain & VASobs distress (nurse, parent & video observer).	Researcher VASobs pain (p = 0.01) & distress (p = 0.02) lower in the treatment group during the post-insertion phase. Nurse VASobs scores lower in the treatment group immediately after the procedure (median: 22 mm; [IQR: 12–37 mm]; placebo: 47 mm [IQR: 25–56 mm]; P<0.05). No difference in parent VASobs scores. No difference in FLACC scores between groups at any time.	Jadad score = 5. * trial concluded early due to concerns re distress associated with administration of trial nebuliser. Therapy effective in adults.
Balan et al, 2009 (352)	Randomised controlled prospective study.	150 children aged 5 – 12 years. Exc: hypersensitivity to local anaesthetics, severe hepatic	3 groups – participants randomised to: LA group – EMLA cream & earphones with no music	All category of observers' median VASobs scores higher in placebo group than other groups at all time points.	Jadad score = 3. All observers, parents & patients aware of

	To determine the comparative efficacy & safety of local anaesthetic cream, Indian classical instrumental music & placebo in reducing pain due to venepuncture in children.	disease, G6PD, altered sensorium, hearing impairment, requirement for urgent treatment. Pain: procedural (venepuncture). Setting: inpatient department tertiary care centre (India).	Music therapy group – Placebo gel & earphones with music. Placebo group – Placebo gel & earphones no music. Pain scoring: VASobs (parent, nurse, investigator & independent observer).	Investigator VASobs scores lower in the LA group than the music group at all time intervals ($p<0.05$) & parent VASobs scores at 1min ($p = 0.033$) & independent observer VASobs scores at 5min ($p=0.38$).	purpose of study & unblinded to group allocation.
Barkan et al, 2014 (353)	Randomised double-blind placebo controlled trial. To compare the efficacy of oral midazolam alone with a combination of oral midazolam & ketamine in children requiring laceration repair.	80 children aged 1 – 10 years. Exc: contraindications or hypersensitivity to the drug, neurological impairment. Pain; procedural (laceration repair). Setting: paediatric ED (Israel).	2 groups – participants randomised to: Treatment group: ketamine Placebo group: normal saline. Pain scoring: VASobs (parent, investigator), Sedation Activity Score.	No difference in VASobs between the groups score at the time of local anaesthetic injection; parent (4.40 ±2.78 cm & 4.50 ±3.33 cm) & investigator (4.49 ±3.16 cm & 4.05 ±2.87 cm). Sedation score lower in the treatment group: mean 2.36 ±0.89 & 3.5 ±0.67, respectively (mean difference 1.14, 95% CI 0.67 to 1.6, $p = 0.001$).	Jadad score = 5. Evidence that VAS used as a measure of pain only made clear in the discussion.
Bhatnagar et al, 2008 (354)	Prospective randomised trial. To evaluate the efficacy & safety of a mixture of ketamine, midazolam & atropine given orally by comparing the same mixture given through the well-established intramuscular route.	60 children with cancer aged 1 – 10 years. Exc: history of recent head injury, active respiratory disease, increased intracranial pressure, neurological dysfunction, drug allergies. Pain: procedural (lumbar puncture). Setting: cancer hospital (India).	2 groups – participants randomised to: Group 1 – IM ketamine (6 mg/kg), midazolam (0.05 mg/kg) & atropine (0.02 mg/kg) Group 2 – oral ketamine (10 mg/kg), midazolam (0.2 mg/kg) & atropine (0.05 mg/kg). Pain scoring: VASobs (investigator).	Mean (±SD) VASobs scores during the procedure were 8.33 (±15.99) & 9.33 (±16.39) in Group 1 & Group 2, respectively ($P = 0.892$). Median sedation score at the start of the procedure was 1 in Group 1 (range = 0–2) & 2 in Group 2 (range = 0–3) ($P < 0.001$).	Jadad score = 3.
Bishai et al, 1999 (355)	Randomized, blinded, crossover design.	39 children aged 5 – 16 years. Exc: unable to rate pain.	2 groups – participants treated using both regimens (acted as their own controls):	No differences in parent VASobs scores parents (2.6 ± 2.2 v 2.4 ± 2.0, $p = 0.54$) or	Jadad score = 5.

	To compare the relative efficacy & safety of amethocaine gel & lidocaine–prilocaine cream in children with cancer undergoing Port-a-Cath puncture & to determine which patient factors influence judgments about pain.	Pain: procedural (Port-a-Cath access). Setting: unstated.	Group 1 - 1g amethocaine gel for 30 mins, preceded by a placebo gel for 30 minutes Group 2 - 1g of EMLA cream for 60 minutes. Pain scoring: Faces scale (child) VASobs (Parents & nurse).	nurse VASobs scores (2.0 ± 1.9 v 1.9 ± 2.1, p = 0.78) between groups. No difference in self-rated scores between groups (2.0 ± 1.4 v 1.5 ± 1.5, p = 0.09). Pain scores assigned by children & their parents were not influenced (p > 0.05) by age, gender, duration of diagnosis, or treatment group in the child.	
Bolt et al, 2008 (407)	Double-blind, randomised, placebo-controlled trial. To determine the efficacy of topical aqueous 2% lignocaine eardrops compared with a placebo (saline) for pain relief of acute otitis media in children.	63 children aged 3 – 17 years Exc: TM perforation, allergy to lignocaine, ventilation tube insitu, hepatic, renal or cardiac disease. Pain: acute pain (otalgia). Setting: tertiary paediatric ED (Australia).	2 groups – participants randomised to: Treatment group – 2% lignocaine Placebo group – normal saline. Pain scoring: Primary - Bieri faces scale, Secondary - VASobs (physician).	VASobs scores were not significantly different between groups at any time. Self-reported pain scores different between groups at T10 & T30 for reduction by 50% & at all time points for reduction by 25% from baseline. >2-point reduction from baseline in self-reported pain was significant only at T10.	Jadad score = 3.
Bouwmeester et al 2001 (397)	Double-blinded randomised controlled trial. To test the hypothesis that postoperative analgesia with continuous morphine infusion would provide improved analgesia with lower stress responses compared with intermittent doses.	204 children aged 0 – 3 years Exc: opioid therapy less than 6 h prior to surgery, neurological, renal, or hepatic dysfunction. Pain: postoperative (non-cardio-thoracic & abdominal surgery). Setting: paediatric surgical ICU (Netherlands).	2 groups (age cohorts) – randomised to: CM – Morphine infusion 10mcg/kg/h & 3hrly placebo (saline) boluses IM – Placebo infusion & 3hrly morphine 30mcg/kg boluses. Pain scorings: VASobs (nurse), COMFORT.	No difference in VASobs scores between groups, scores differed across age groups. Scores reduced over time following surgery. COMFORT scores differed for age group 4 across treatment groups (1.8 v 20.8, p=0.02). Overall no plasma concentration differences between treatment groups. Glucose interaction between age & treatment group (p=0.04).	Jadad score = 3.
Chapman et al, 2011 (356)	Prospective randomised controlled clinical study. To assess the effectiveness of the VeinViewer for peripheral	323 children aged 0 – 17 years. Exc: need for urgent cannulation/treatment.	2 groups – participants randomised to: VeinViewer group – PIV placement using VeinViewer	No difference in VASobs & VAS scores reported by the parent, nurse or patient (children 8 – 16 years). In 0 – 2 year-old children, parents report no difference, nurses report less pain in	Jadad score = 3.

	intravenous catheter insertion in children.	Pain: procedural (peripheral IV catheter insertion). Setting: paediatric ED (US).	Control group – PIV placement without VeinViewer. Pain scoring; VAS (children> 8 years), VASobs (parents & nurses).	VeinViewer group (median VAS = 34 vs 46; p = 0.01).	
Cignacco et al; 2008 (357)	Randomised controlled trial. To assess the impact of 0.1mg/kg morphine IV prior to endotracheal suctioning & the impact on recovery of multisensorial stimulation (particularly those receiving placebo).	30 preterm neonates 24 – 37 weeks postmenstrual age. Exc: IVH grade III or IV, condition involving partial or total loss of sensitivity, morphine in last 10 hours, APGAR <3 after 5min, cord blood pH < 7.0, maternal drug addiction, ventilation postop. Pain: procedural (endotracheal suctioning). Setting: NICU (2 units in Switzerland).	4 groups – participants randomised first to: Morphine group Placebo group Then randomised to: MSS group Standard recovery method group. Pain scoring: Bernese neonatal Pain Scale, PIPP, VASobs (nurses).	No difference in VASobs scores, PIPP, & Bernese Pain Scale scores between groups. No difference between groups (all scales) between MSS & standard recovery group.	Jadad score = 3.
Coda et al, 2014 (408)	Single-blinded randomised controlled trial. To test the hypothesis that in children with JIA, fitted foot orthoses (FO)s are more effective at reducing pain & improving quality of life than control orthoses.	60 children aged 5 – 18 years. Exc; inability to walk barefoot, other MSK, central/peripheral CNS disease, previous foot surgery, currently using FOs. Pain: disease specific pain (juvenile arthritis). Setting: paediatric rheumatology department (Scotland).	2 groups – participants randomised to: Treatment group – fitted foot orthoses Control group – non-fitted FO. Pain scoring; VASobs (parents).	Difference in VASobs scores (baseline - 6 months) greater in fitted FO group (p=0.029). 8 mm reduction in pain in the fitted FOs group was clinically important.	Jadad score = 3.

Cohen et al, 2009 (359)	Randomised controlled trial. To evaluate the effectiveness of vapocoolant alone for paediatric immunization pain.	57 children aged 4 – 6 years. Exc: none stated. Pain: procedural (routine immunizations). Setting: university-based outpatient primary care clinic (US).	2 groups – participants randomised to: Treatment group – topical vapocoolant alone Control – standard treatment. Pain scoring: VASobs (caregiver & nurses), Faces Pain Scale-Revised, Distress behaviours.	VASobs scores did not differ between groups but higher during the injection compared with baseline (caregiver, F (1, 54) = 89.10, P<0.001, & nurse reports, F (1, 53)=25.21, P<0.00). No difference in self-report scores between groups but higher scores during injection than baseline, F (1, 49)=71.15, P<0.001.	Jadad score = 3.
Cohen et al, 2006 (358)	Randomised clinical trial. To examine the effectiveness of an easy-to-use & practical distraction intervention for reducing infants' immunization distress.	136 infants aged 1 – 24 months. Exc: non stated. Pain: procedural (routine immunizations). Setting: university-affiliated medical centre & a private practice office. (US).	2 groups – participants randomised to: Treatment group – program of distraction (inc: nurse & parent training) Control group – typical care. Pain scoring: VASobs (parents & nurses), MAISD.	No differences in VASobs scores existed between groups. MAISD scores favoured Treatment Group, $F(1,108) = 9.91$, $p < .05$ using ANOVA. Using paired samples t-tests; infants in Treatment group less distressed during the anticipatory, $t(109) = 2.02$, $p < .05$, & recovery phases, $t(123) = 2.35$, $p < .05$. Non-significant difference in infant distress between the Distraction & Typical Care groups during the procedure phase, $t(124) = 1.49$, ns	Jadad score = 3.
Di Liddo et al, 2006 (360)	Randomised, double-blinded trial. To compare the efficacy of etomidate & midazolam for achieving procedural sedation & analgesia in children.	128 children aged 2 – 18 years. Pain: procedural (fracture reduction) Setting: ED & orthopaedic clinic (Canada).	2 groups – participants randomised to: Midazolam group: 0.1mg/kg Etomidate: 0.2mg/kg. Pain scoring: VASobs (investigator).	No difference in mean VASobs scores (diff -2.9mm; 95% CI -11.0 to 5.1mm). Patients attaining adequate sedation higher in etomidate group: 46 of 50 (92%) versus 18 of 50 (36%) (diff 56%; 95% confidence interval [CI] 38% to 69%).	Jadad score = 5.
Dulai et al, 2016 (361)	Randomised triple-blind placebo controlled trial. To evaluate the efficacy of a fast-acting topical preparation of liposomal lidocaine in	281 children aged 3 – 16 years. Exc: contraindications to lignocaine.	2 groups – participants randomised to: Lignocaine group – liposomal lignocaine applied to pin site Placebo group – placebo cream applied to pin site.	No differences in post-procedure parent & technician VASobs scores or self-report scores. Pre vs post PP removal: children reported a 2.18 (SD=2.92) increase in pain (P<0.001), parents reported a 2.10 (SD=2.72) increase	Jadad score = 5.

	reducing the pain experienced by paediatric patients during intraosseous percutaneous pins (PP) extraction.	Pain: procedural (PP removal). Setting: orthopaedic department (Canada).	Pain scoring: Oucher scale, VASobs (parents & orthopaedic technician).	in pain (P<0.001) & orthopaedic technicians reported a 1.76 (SD=2.10) increase in pain (less than either the parent or children groups (P<0.001).	
Fatovich et al, 1999 (362)	Randomised double-blind, placebo controlled trial. To compare plain lignocaine with buffered (sodium bicarbonate) lignocaine for the repair of simple lacerations.	136 children aged 1 – 10 years. Exc: physical, visual or cognitive impairment, allergic to local anaesthetic, complex wound, requirement for LA other than 1% lignocaine infiltration. Pain: procedural (lignocaine infiltration). Setting: ED (Australia).	2 groups – participants randomised to: Buffered lignocaine – mixed with 8.4% sodium bicarbonate Plain lignocaine – mixed with normal saline. Pain scoring; VASobs (parent) Nurse-Rated Pediatric Pain Assessment Tool.	No difference in VASobs between groups (mean = 4.5 in both groups). Median nurse-rated pain scores also 4.5 in each group. No differences in facial expression, sounds the child made or the degree of restraint used. Cohort of adults – self-report VAS results mean VAS for plain lignocaine was 2.8 (SD 2.5) & buffered lignocaine 2.4 (SD 2.1) – a difference of 0.4 (95% CI -1.83 to 2.63).	Jadad score = 5.
Ha et al, 2013 (184)	Randomised controlled trial. To determine the effects of audio-visual distraction on pain in children during laceration repair in emergency room settings.	84 children aged 3 – 10 years. Exc: chronic disease, complicated lacerations, associated bone fractures or multiple injuries, previous treatment with analgesics or sedatives. Pain: procedural (laceration repair). Setting: ED (Korea).	2 groups – participants randomised to: Treatment group – audio-visual distraction Control – standard treatment. Pain scoring; Faces Pain Rating Scale (child), VASobs (primary caregiver), PBCL, salivary cortisol.	Post-procedure sensory VASobs scores higher in the control group than in the experimental group ($t = -3.768$, $P < 0.001$). The post procedure caregiver-reported affective VAS pain scores higher in the experimental group ($t = 4.607$, $P < 0.001$). Difference in PBCL between groups ($t = 4.070$, $P < 0.001$). No difference between groups in post-procedure FPRS scores ($t = -1.322$, $P < 0.001$).	Jadad score = 3. Affective/sensory pain not adequately defined. No presentation of pain & anxiety scores.
Hamers et al, 1999 (398)	Double-blind, randomized, placebo controlled (2 x 2) factorial design. 1. To evaluate the effectiveness of 2 pain	83 children aged 3 - 12 years. Exc: children unable to use self-report scales. Pain: postoperative (tonsil & adenoid surgery.	4 groups – participants randomised to: Group 1: 30-50mg/kg paracetamol suppository & 0.9% saline IM	No difference between groups in VASobs, FLACC, CHEOPS, Faces or self-reported Oucher scores or whether child had drunk at 1, 2, 3 hours post procedure.	Jadad score = 3.

	protocols used interchangeably to manage early postoperative T&A pain. 2. To investigate whether nurses' systematic pain assessments (SPA) improve pain management.	Setting: not stated.	Group 2: 30-50mg/kg paracetamol suppository, 0.9% saline IM & SPA Group 3: 30 – 50mg paracetamol suppository & 1microgram/kg fentanyl intramuscularly Group 4: 30 – 50mg paracetamol suppository & 1microgram/kg fentanyl intramuscularly & SPA. Pain scores: VASobs (parent & researcher), FLACC & CHEOPS (not blinded), Faces Pain Scale & Oucher (independent).		
Harrison et al, 2014 (440)	Randomised placebo-controlled trial. To evaluate the feasibility, acceptability & preliminary efficacy of sweet taste in reducing pain in toddlers & pre-school children during immunisation & to use the results to inform a sample size estimation for future full-scale trials.	60 children aged 12 -18mths & 3 – 5 years. Pain: procedural (immunisation). Setting: immunisation centre (Australia).	2 x 2 groups: 12 – 18 month infants randomised to: Group – 33% sucrose syrup Group – placebo (water) 3 – 5 year olds randomised to: Lollypop group – lollypop Standard care group – pinwheel or bubble blowing for distraction. Pain scores: VASobs (parent), FLACC, cry duration.	No difference in pain scores (VASobs or FLACC) in standard care versus intervention group for either pilot RCT.	Jadad score = 3*. * Allocation for 12 – 18 month children double blinded.
Heden et al, 2009 (363)	Randomised triple-blind, placebo-controlled study. To investigate whether children experience less fear, distress, &/or pain according to parents, nurses, & children	50 children with cancer aged 1 – 18 years. Exc: pain > 50 from causes other than needle insertion, needle phobia.	2 groups – participants randomised to: Midazolam – 0.3mg/kg orally Placebo – equal volume Bitrex (to match bitter taste) orally.	No difference between groups in VASobs scores or self-report scores or discomfort associated with oral intake. CHEOPS scores in midazolam group lower (6.2 vs 8.3, p = 0.012). Correlations between parent's fear & distress scores were 0.85, fear & pain 0.60,	Jadad score = 5.

	>7 years of age when they receive oral midazolam versus placebo before a needle is inserted in a subcutaneously implanted intravenous port.	Pain: procedural (needle insertion into IV port). Setting: paediatric oncology & haematology setting (Sweden).	Pain scores: VASobs (parent & nurse), VAS (child > 7y), CHEOPS.	& distress & pain 0.71; for nurses scores were 0.89, 0.69, & 0.79, & for children 0.54, 0.44, & 0.80, respectively.	
Heden et al, 2009 (364)	Randomised controlled trial with parallel groups in two steps. To examine whether children experience less fear, distress & pain connected to a routine needle insertion in an intravenous port when subjected to (1) an intervention: blowing soap bubbles or having a heated pillow in addition to standard care vs. standard care only; & when subjected to (2) blowing soap bubbles vs. heated pillow.	28 children with cancer aged 2 - 7 years. Exc: severe pain of other cause, needle phobia requiring sedation. Pain: procedural (needle insertion into IV port). Setting: paediatric oncology & haematology setting (Sweden).	2 groups – participants randomised to: Group 1 – 1st needle insertion = standard care, 2nd needle insertion = standard care + blowing soap bubbles Group 2 – 1st needle insertion = standard care & 2nd needle insertion = standard care + heated pillow. Pain scores: VASobs (parent & nurse).	No difference between nurse or parent VASobs pain, fear or distress scores between groups. Parents reported lower VASobs distress ($t = 2.28$, d.f. = 13, $P < 0.05$) & fear ($t = 2.36$, d.f. = 13, $P < 0.05$) scores in children receiving either bubble blowing + standard care versus standard care. Parents also reported lower VASobs fear scores ($t = 3.400$, d.f. = 13, $P < 0.05$). No difference in nurses VASobs fear or distress for intervention + standard care versus standard care.	Jadad score = 3. Intervention vs standard care comparison not based on randomised allocation to group.
Heden et al, 2011 (365)	Randomized, triple-blind, placebo-controlled study. To investigate whether children experience less fear, distress, &/or pain when they receive oral morphine vs. placebo before a needle is inserted in a sub-cutaneously implanted intravenous port when combined with topical anaesthesia.	50 children with cancer aged 1 – 18 years. Exc: severe pain of other cause, needle phobia, concurrent treatment with medication known to interact with morphine. Pain: procedural (needle insertion into IV port).	2 groups – participants randomised to: Morphine group – 0.05ml/kg of morphine (5mg/ml) orally Placebo group – 0.05ml/kg placebo (strongly flavoured to preserve blinding) orally. Pain scores: VASobs (parent & nurse), VAS (child > 7y), CHEOPS, PBCL.	No difference in VASobs scores (nurse or parent) or self-report VAS scores for pain, distress or fear between groups. No difference in CHEOPS or PBCL scores across groups.	Jadad score = 5.

		Setting: paediatric oncology & haematology setting (Sweden).			
Heden et al, 2014	Double-blind randomised placebo controlled trial. To investigate whether children experience less pain, fear &/or distress when they receive high-dose paracetamol compared with placebo, using a needle insertion in a subcutaneously implanted intravenous port as a model.	51 children with cancer aged 1 – 18 years Exc: severe pain of other cause, needle phobia, concurrent treatment with paracetamol or medication known to interact with paracetamol, liver disease. Pain: procedural (needle insertion into IV port). Setting: paediatric oncology & haematology setting (Sweden).	2 groups – participants randomised to: Paracetamol group – 40mg/kg paracetamol orally Placebo group – same volume of placebo solution. Pain scores: VASobs (parent & nurse), VAS (child > 7y), CHEOPS, cortisol levels.	No difference in VASobs (nurse & parent) scores for pain, fear or distress between paracetamol & placebo groups. Self-report distress scores lower following paracetamol (mean 6 ± 7 versus 20 ± 19) but no difference in self-report pain or fear scores. No difference in CHEOPS scores or cortisol levels.	Jadad score = 5.
Hogan et al, 2014 (366)	Randomised partially-blinded parallel 2 group trial. To determine the effectiveness of parent-led tactile stimulation for pain reduction when added to a combination of evidence-based pain-reducing interventions in infants undergoing immunization injections.	120 infants aged 4 to 6 months. Exc: impaired neurological development, previous seizure, local anaesthetic at site, use of sedatives or opioids in previous 24 hours. Pain: procedural (immunisation). Setting: primary care practice (Canada).	2 groups: participants randomised to; Usual care group: Tactile stimulation group: Pain scoring: MBPS, VASobs (parent).	No difference in mean VASobs (60 [20] vs. 53 [22] mm; P=0.10) or MBPS (8.2 [1.1] vs. 8.0 [1.3]; P=0.57) scores between groups.	Jadad score = 3.
Hopkins et al, 1988 (368)	Randomised double-blind placebo-controlled trial. To examine the efficacy of EMLA with respect to	120 children aged 1 – 5 years Exc: none reported. Pain: procedural (IV catheter insertion).	2 groups – participants randomised to: Intervention group: EMLA gel applied topically to IV insertion site	Lower VASobs scores ($w = 387.5$, $p < 0.0005$) & VRS scores ($x' = 20.96$, d.f. = 3, $p < 0.001$) were recorded in the children treated with EMLA cream compared with placebo.	Jadad score = 3. Questionable whether groups are randomised.

	alleviation of venepuncture pain at intravenous induction of general anaesthesia in children aged 1-5 years, to identify the optimal application time & to evaluate possible adverse reactions to the preparation.	Setting: day surgery unit (England).	Placebo group: Placebo gel applied topically to IV insertion site. Pain scoring: Verbal Rating Scale (VRS) & VASobs (operating department assistant).	* VRS scores recorded by same clinicians as VASobs.	80 in treatment & 40 in placebo group.
Horn et al, 1999 (369)	Comparative randomised study. To compare distress behaviours & perceptions of distress of 4 – 6 year old children who received 2 immunisations simultaneously with those of children of the same age receiving them sequentially.	46 children aged 4 – 6 years. Exc: Physical or mental conditions, hospitalised or receipt of injection in last 6 months. Pain: procedural (immunisation). Setting: private paediatric office (US).	2 groups – participants randomised to: Simultaneous group – immunisations administered at the same time Sequential group – immunisations given one after the other. Pain scoring: VASobs 'upset' (parents), Wong & Baker Faces (self-report), OSBD-R.	No difference in VASobs scores between groups. No difference in self-report of distress between the two groups & no difference in OSBD-R scores between the two groups. The VASobs scores differed before & after the injections.	Jadad score = 3.
Hua et al, 2015 (370)	Prospective randomised study. To investigate the effect of virtual reality distraction on alleviating pain during dressing changes in children with chronic wounds on their lower limbs.	65 children aged 4 – 16 years. Exc: non-Chinese speaking, sensory disability, other diagnosed illness, require sedative medication. Pain: procedural (dressing change). Setting: paediatric centre of tertiary hospital (China).	2 groups – participants randomised to: Virtual reality (VR) distraction Standard distraction. Pain scoring: VASobs (care-givers), FLACC (nurses), Wong & Baker Faces scale (self-report).	VAS observer scores lower in VR distraction group; before (1.87 ± 2.14 vs 0.99 ± 0.68, p = 0.028), during (6.25 ± 2.84 vs 4.35 ± 2.64, p = 0.007) & after (5.94 ± 1.59 vs 2.67 ± 1.89, p < 0 .001). Self-reported pain scores lower in VR distraction group; before (1.63 ± 1.39 vs 0.85 ± 1.12 , p = 0.016), during (4.19 ± 2.12 vs 2.42 ± 1.85, p = 0.001) & after (3.38 ± 1.48 vs 2.48 ± 1.8, p = 0.034). FLACC scores lower in VR group during (7.36 ± 3.47 vs 4.18 ± 2.97, p < 0.001) & after (5.79 ± 3.84 vs 3.68 ± 2.73, p = 0.013).	Jadad score = 3.
Ipp et al, 2004 (371)	Randomised double blinded clinical trial.	49 infants aged 12 months.	2 groups: participants randomised to:	Median pain scores lower in Priorix group: paediatrician VASobs, 15 vs 58 (P=.001);	Jadad score = 5.

	To compare acute pain response to 2 measles-mumps-rubella (MMR) vaccines.	Exc: chronic illness, anaphylaxis to egg, fever or acute illness. Pain: procedural (immunisation). Setting: community pediatrician's clinic (Canada).	Priorix group: MMR-II group: Pain scoring: VASobs (parent & paediatrician), MBPS latency to cry & cry duration.	parent VASobs, 22 vs 53 (P=.007); & MBPS, 6 vs 8 (P=.02). The median latency to first cry longer in Priorix group (1.5 seconds vs 1 sec, p= 0.26). Median difference in pain scores (after minus before) for Priorix vs M-M-R II were as follows: paediatrician VASobs, 15 vs 53 (p=.003); parent VASobs, 22 vs 47 (P=.008); & MBPS, 3 vs 5 (P=.03).	
Ipp et al, 2006 (372)	Randomised double-blind study. To determine if the acute & immediate pain response to two licensed M-M-R vaccine products (using a self-report measure) in children 4-6 years of age was similar to that demonstrated in younger infants.	60 children aged 4 – 6 years. Exc: none stated. Pain: procedural (immunisation). Setting: urban primary care paediatric practice (Canada).	2 groups – participants randomised to: Priorix MMR-II. Pain scoring: OUCHER pain scale (self-report), VASobs (physician & parent), cry duration.	VASobs scores did not differ between groups (p = 0.235). Children in the M-M-R II group had significantly higher OUCHER scores (median 20, IQR 0 - 60 vs median 0.0, IQR 0 – 20, p = 0.047) & were more likely to cry post-vaccination (17 vs 8, p = 0.018). Cry duration was longer for the MMR-II group (median 6 IQR 0 – 40 vs median 0, IQR 0 – 0, p = 0.02).	Jadad score = 5.
Ipp et al, 2007 (373)	Randomised controlled trial. To compare acute pain response during immunisation in infants using a slow standard of care injection technique versus a rapid pragmatic technique.	113 infants aged 4 to 6 months. Exc: chronic illness, anaphylaxis to egg, fever or acute illness, treatment with local anaesthetic. Pain: procedural (immunisation). Setting: primary care practice (Canada).	2 groups – participants randomised to: Standard group: slow aspiration, injection & withdrawal Pragmatic group: no aspiration, rapid injection & withdrawal. Pain scoring: MBPS, VASobs (parent & paediatrician), & cry duration.	The median (IQR) VAS scores by parents 1.9 (0.1–3.1) vs 3.5 (1.6–5.5) & paediatricians vs 1.4 (0.2–2.4) vs 2.8 (2.0–5.1) were lower for pragmatic group MBPS scores also lower for pragmatic group (3.3 95% CI 2.6 to 3.9 vs 5.6, 95% CI 5 to 6.3, p<0.001). The pragmatic group less likely to cry, 24/56 (43%) vs 47/57 (82%), to cry less, median 0 sec (IQR 0–11.30) vs 14.7 sec (8.7–35.6) & to take less time to have vaccine injected, median 0.9 s (IQR 0.8–1.1) vs 8.8 s (7.9–10.3), for all comparisons p<0.001.	Jadad score = 3.

Ipp et al, 2009 (374)	Single centre, double blinded randomised clinical trial. To determine if acute pain response after administration of the diphtheria, polio, & tetanus toxoids & acellular pertussis & Haemophilus influenzae type b (DPTaP-Hib) vaccine & the pneumococcal conjugate vaccine (PCV) is affected by the order in which they are given.	120 infants aged 2 to 6months. Exc: chronic illness, anaphylaxis to egg, fever or acute illness, treatment with local anaesthetic. Pain: procedural (immunisation). Setting: paediatric community practice (Canada).	2 groups: participants randomised to: DPTa-HiB group: received DTPa-HiB first followed by PCV PCV group: received PCV first followed by DPTa-HiB . Pain scoring: MBPS, VASob (parent & paediatrician).	VASobs scores post injection were lower when DPTaP-Hib was administered first (4.2 SD 2.3 vs 5.6 SD 2.6, p=0.003) MBPS score were also lower when DTaP-Hib was given first (7.6, SD 1.5 vs 8.2, SD 1.5, p=0.037).	Jadad score = 5.
Kjeldgaard Pedersen et al, 2016 (399)	Randomized double-blind trial. To test the efficacy y of epidural analgesia & LIA for the management of early post-operative pain in children with CP.	12 children aged 3 – 13 years with CP. Exc: previous surgical interventions in the same anatomical region, multiple-level surgery, allergy to or intolerance of study drugs, or implanted intrathecal baclofen pump. Pain: postoperative (osteotomy). Setting: Paediatric Orthopaedic department (Denmark).	2 groups – participants randomised to: LIA group: ropivacaine (2 mg/kg) & epinephrine (5 µg/mL) as infiltration & ropivacaine (0.5 mg/kg) as bolus. Placebo group: Identical volumes of isotonic saline were used in the placebo group. Pain scoring: VASobs (parent), FLACC-R.	VAS scores were lower in the epidural group (0.6 than the LIA group (5.2, p= 0.02) & the placebo group (6.5, p < 0.001).The r-FLACC scores were similarly lower in the epidural group (0.7) than in the LIA group (4.8, p = 0.02) & the placebo group (5.2, p = 0.01).	Jadad score = 3.
Knutsson et al, 2006 (443)	Prospective, randomized, double-blind, placebo-controlled trial. To evaluate the effect of peri-operative application of	98 children aged 3 – 10 years. Exc: simultaneous surgery during the same anaesthesia (except for insertion of transmyringeal ventilation tubes); mental retardation,	2 groups – participants randomised to: Local anaesthetic group: gauze soaked in 2ml mepivacaine 10mg/ml	No difference in VASobs scores or self-reported pain scores were seen between groups.	Jadad score = 3.

	adjuvant local anaesthesia given to reduce postoperative pain after adenoidectomy in children not undergoing simultaneous tonsillectomy.	treatment with pain medication; allergy to local anaesthetics; or refusal to take oral medication. Pain: postoperative (adenoidectomy). Setting: otorhinolaryngology department (Sweden).	Placebo group: gauze soaked in 2ml normal saline. Pain scoring: VASobs (nurse), Wong & Baker Faces (self-report).		
Knutsson et al 2006 (375)	Randomised double-blind controlled trial. To investigate whether, the vaccine Priorix® causes less immediate injection pain than MMR-II® in vaccination of infants aged 18–24 months.	295 infants aged 18 – 24 months. Exc: none stated. Pain: procedural (immunisation). Setting: child health centre (Sweden).	2 groups – participants randomised to: Priorix: administered Priorix preparation of the measles-mumps-rubella vaccine MMR-II: administered MMR-II preparation of the measles-mumps-rubella vaccine. Pain scoring: VASobs (parent), CHEOPS.	Mean VAS score were 2.3 vs 5.2 for Priorix® & MMR-II®, respectively ($p<0.001$). Mean CHEOPS scores were 1.9 vs 6.1 for Priorix® & MMR-II®, respectively ($p<0.001$).	Jadad score = 5.
Koller et al, 2007 (409)	Prospective, randomized, double-blinded, clinical trial. To investigate the effectiveness of oxycodone, ibuprofen, or their combination for the management of orthopaedic injury–related pain in children.	66 children aged 6 – 18 years Exc: FPS score <4, inability to self-report, altered mental state, allergy to agents, analgesic treatment in last 12 hours, bony deformity, open fracture, multiple trauma. Pain: acute pain (secondary to injury). Setting: ED (US).	3 groups – participants randomised to: Group 1: oxycodone [0.1 mg/kg (max 10 mg) + placebo], Group 2: ibuprofen [10 mg/kg (max 800 mg) + placebo] Group 3: oxycodone/ibuprofen [0.1 mg/kg + 10 mg/kg (max 10 mg + max 800 mg).	No difference in scores between groups. Differences in VAS scores over time were significant ($P < 0.001$). Decreases in systolic BP & SpO2 over the 120-minute observation period were significant ($P < 0.001$). FPS scores for all groups dropped over time, with a mean change of 4.2 over the 120-minute study period. Differences in VAS scores between the 3 evaluators were significant ($p < 0.001$).	Jaded score = 3.

			Pain scoring: VASobs (parent, nurse & investigator), Faces Pain Scale (self-report).		
Kozer et al, 2006 (376)	Prospective, single-blind, randomized, controlled study. To compare the pain that is experienced during suprapubic aspiration & transurethral when performed in young infants.	58 infants aged 0 – 2 months. Exc: premature, previous sepsis workup or painful procedures. Pain: procedural (urine collection). Setting: university affiliated hospital (Israel).	2 groups – participants randomised to: SPA group: EMLA applied 1hr prior to suprapubic aspirate TUC group: transurethral catheter lubricated with 2% lignocaine jelly. Pain scoring: VASobs (nurse & parent), DAN.	Mean VASobs scores higher in SPA group compared with TUC group (parent: 63 ± 27 mm vs 46 ± 26, respectively, mean difference 16.8 95% CI 1.8 – 31.8 & nurse: 3 ± 18 mm vs 43 ± 25 mm, respectively, mean difference 19.6 95% CI 7.4 – 31.8). Mean DAN scores higher in infants in SPA group compared with TUC group (7 vs 4.5, respectively, mean difference 2.5 95% CI 1.4 – 3.7).	Jadad score = 3.
Lee-Jayaram et al, 2010 (377)	Prospective, partially blinded, randomized controlled trial. to compare procedural distress during sedation with etomidate/ fentanyl (E/F) to that with ketamine/midazolam (K/M).	23 children aged 5 – 17 years Exc: developmental delay, multiple injuries, history psychosis. Pain: procedural (fracture manipulation). Setting: ED (US).	2 groups – participants randomised to: Experimental (E/F) group: initial dose of 0.2 mg/kg IV etomidate & 1 kg/kg IV fentanyl. Standard (K/M) group: 1 mg/kg IV ketamine & 0.05 mg/kg (max dose 2 mg) IV midazolam. Pain scoring: VASobs (parents), OSBD-r.	VAS scores lower with K/M than with E/F (13.7 vs 50.5, P = 0.003). Parents more satisfied with K/M on a 5-point satisfaction scale (P = 0.004). OSBD-r scores lower with K/M than E/F (0.08 vs 0.89, P = 0.001).	Jadad score = 3.
Lindh et al, 2003 (378)	Randomised double-blinded. To determine whether use of lidocaine–prilocaine 5% cream (EMLA) & oral glucose decreases pain associated with diphtheria–pertussis–tetanus	90 Infants aged 3 months. Exc: none stated. Pain: procedural (immunisation). Setting: outpatient paediatric practice (Sweden).	2 groups – participants randomised to: Treatment group: EMLA patch & glucose solution Placebo group: placebo patch & water.	VAS scores lower in treatment group (parent: 3 ± 2 vs 5 ± 2, $p < 0.05$ & nurse: 2 ± 2 vs 5 ± 2, $p < 0.05$). MBPS scores lower in treatment group at 0 – 10 sec (5.5 ± 2.0 v 7.7 ± 1.7, $p< 0.05$) & 11 – 20 sec (5.4 ± 2.4 v 6.8 ± 2.2, $p < 0.05$). Fewer infants cried in the treatment group (32 v 44, p = 0.001), latency of first cry was	Jadad score = 5.

	(DPT) immunization in 3-month-old infants.		Pain scoring: MBPS, VASobs (parent, nurse), latency to cry & total crying time, HR.	longer in the treatment group (6.4 ± 3.2 v 3.8 ± 2.3, p < 0.001). A biphasic transient heart rate response with a marked deceleration followed by an acceleration seen more often in the placebo group (p = 0.03).	
Luhmann et al, 2004 (379)	Randomised trial. To compare the reduction of pain & anxiety during PIV insertion provided by subcutaneous buffered 1% lidocaine or topical ELA-Max in children.	69 children aged 4 – 17 years Exc: analgesic administration prior to PIV insertion, allergy to anaesthetic. Pain: procedural (peripheral IV catheter insertion). Setting: ED (US).	2 groups – participants randomised to: Buffered lignocaine group: local infiltration ELA-Max: 4% lignocaine gel topical. Pain scoring: VASobs (parent, nurse, observer), VAS (child).	No difference in VAS scores (pain & anxiety) between groups. No difference in self-reported pain or anxiety between groups.	Jadad score = 3.
Luhmann et al, 2006 (380)	Prospective randomised trial. To compare the efficacy & adverse effects of intravenous K/M versus N2O/HB for analgesia & anxiolysis during forearm fracture reduction in children, including differences in recovery times.	103 children aged 5 – 17 years. Exc: adverse reactions to medication used in study, psychiatric illness, unfasted. Pain: procedure (fracture reduction). Setting: ED (US).	2 groups – participants randomised to: K/M group: IV midazolam 0.1 mg/kg & glycopyrrolate 5μg/kg N2O/HB group: inhaled 50% N2O/50% O2. Pain scoring: PBCL, VASobs (parent), VAS (patient).	VAS scores lower with N2O/HB (2.5 v 4.1, diff 1.6, 95% CI 0.6 – 2.6). Self-reported scores also lower memory of pain during reduction with N2O/HB (1.8 v 2.9, diff – 1.1, 95% CI 0.0 – 2.1). Mean change in Procedure Behavioural Checklist was less for nitrous oxide/hematoma block (p = 0.02). Parents & patients reported less anxiety with N2O/HB.	Jadad score = 3.
Marec-Berard et al, 2009 (381)	Randomised controlled trial. To estimate the success rate of LPs using the LP pillow compared to the usual procedure.	124 children aged 2 – 18 years. Exc: medical condition contraindicating use of the pillow, previous pillow use. Pain: procedure (lumbar puncture).	2 groups – participants randomised to: Pillow group: Sitting position with LP pillow for support Standard group: Sitting position, no LP pillow. Pain scoring: VASobs (parent), VAS (child).	No difference in VASobs pain (median VAS 17.5 vs. 10 mm, p = 0.16) or anxiety (median score 4 vs. 3.5, p = 0.28) between groups. No difference in children's report of pain in either group (median VAS 15 vs. 25mm, p = 0.39). No difference in procedural success in either group (primary outcome).	Jadad score = 3.

		Setting: several oncology centres (France).			
McErlean et al, 2003(382)	Randomised double-blind placebo-controlled trial. To compare the use of oral midazolam versus placebo (PLA) in an ED setting.	51 infants & children aged 9months to 6 years. Pain: procedural (IV catheter insertion) Setting: ED (US).	2 groups - participants randomised to: MID group: 0.5mg/kg oral midazolam syrup PLA group: equal volume of look-alike, taste-alike placebo. Pain scoring: VASobs (parent, clinician observer).	Parent MID group VASobs scores lower (median, 25; IQR 10-63 mm) than PLA VASobs scores (median, 72; IQR, 34-98 mm) (p = .002). Observers' MID group VASobs scores lower (median, 38; IQR, 15-55 mm) than PLA group VASobs scores (median, 49; IQR, 31-69 mm) not significant (p = 0.16).	Jadad score = 3.
McGowan et al, 2013 (383)	Randomised controlled trial. To compare pain response during routine immunisation of infants using simultaneous versus sequential administration techniques.	36 infants aged 2 to 6 months. Exc: known physical or psychological conditions, needle phobic parents. Pain: procedural (immunisation). Setting: immunisation clinic (Wales).	2 groups – participants randomised to: Intervention group: simultaneous immunisations Control group: sequential immunisations. Pain scoring: MBPS, VAS (parent).	Median change in VASobs scores greater in sequential group (5.6cm) than in simultaneous group (4.7cm) – not significant (p = 0.06). Median change in MBPS scores greater in simultaneous group at 15s (p = 0.05) & in sequential group at 30s (p < 0.05), 45s (p = 0.01) & 120s (p = 0.02).	Jadad score = 3.
Miller et al, 2011 (384)	Randomised (controlled) trial. To determine if a combined MMD protocol (preparation & distraction) will reduce the pain & distress of 3–10 year olds undergoing burn care procedures when compared with children provided SD (current typical treatment).	40 children aged 3 – 10 years. Exc: cognitive impairment, sedation & anxiolytics. Pain: procedural - burn dressing procedure. Setting: burns outpatient centre (Australia).	2 groups – participants randomised to: Group SD: standard distraction Group MMD: Multimodal distraction. Pain scores: FLACC, Wong & Baker Faces, VASobs (parents) not blinded).	VASobs scores lower in MMD group at all 4 points e.g. post dressing removal (3.23 ± 2.38 vs 6.15 ± 2.91, p = 0.01) & post dressing application (2.55 ± 1.73 vs 6.05 ± 0.76, p < 0.001). Self-reported scores & FLACC scores in MMD group lower than SD group (p < 0.001) for all points except pre-dressing removal.	Jadad score 3.
Miner et al, 2007 (410)	Randomised clinical trial. To determine if nebulized fentanyl citrate safely provides pain relief that is	41 children aged 6months – 17 years. Exc: allergy to fentanyl, hepatic or renal disease, acute	2 groups – participants randomised to: IV fentanyl: 1.5 µg/kg IV bolus fentanyl citrate	Mean decrease in VASobs was 55.1mm (95% CI = 40.3 to 70.0) for the IV group & 77.8 mm. (95% CI = 67.4 to 88.4) for the nebulized group. The mean VAS scores reported by	Jadad score = 3. Parents (n = 4) allowed to change treatment

	similar to that of IV fentanyl citrate.	intoxication, significant respiratory disease/distress. Pain: acute pain. Setting: ED (US).	Nebulised fentanyl: 3.0 µg/kg nebulised fentanyl citrate diluted with normal saline to 5ml via breath actuated nebuliser. Pain scoring: VASobs (physician), VAS (child), CHEOPS.	the children were 27.7mm (95% CI 12.6 – 42.8) for IV fentanyl & 15.6 (95% CI 6.3 – 24.9). Mean CHEOPS scores were 6.2 (95% CI 3.6 – 8.8) for IV fentanyl & 5.5 (95% CI 4.7 – 6.2) for nebulised fentanyl. No difference in physician assessed adequacy of pain treatment (8/14 vs 20/27, p = 0.42).	group (analysis based on intention to treat). Subjects (n = 5) removed from study due to inadequate pain control.
Muthusamy et al, 2010 (400)	Prospective, randomised study. To determine if the use of a pain pump in conjunction with oral analgesics is an effective post-operative pain management strategy for children with CP undergoing lower extremity outpatient procedures.	37 children with CP aged 3 – 18 years. Exc: allergy to bupivacaine, kidney or liver disease, concurrent orthopaedic surgery. Pain: post-operative (outpatient surgical procedures). Setting: children's hospital (US).	2 groups – participants randomised to: Pain pump: pain pump placed at surgical site to deliver 0.2 mg/kg/h bupivacaine to site for 48 hours & oral analgesic PRN Control group: oral analgesics PRN only. Pain scoring: VASobs (parent) NCCPCPV, TQPM.	VAS scores higher in the control group overall (P<0.0001) & at days 0, 1, & 2 (3.79 ±2.86 vs 6.7 ±3.21, p<0.007, 2.78 ±1.99 vs 5.17 ±2.31, p=0.002, 1.79 ±1.44 vs 3.43 ±2.06, p=0.01, respectively), but not on the third day (1.74 ±2.10 vs 2.10 ±1.68, p=0.59). VAS scores with movement differed between groups (5.47 ±2.72 vs 7.58 ±2.12, p = 0.029). Difference in medication given on days 1 & 2 (p=0.04 & p=0.03, respectively), but overall not significant (P=0.29) Majority of the responses to the TQPM questionnaire did not differ between groups.	Jadad score = 3.
Newbury et al, 2009 (385)	Parallel, randomised, blinded, controlled study. To determine if amethocaine increases first attempt success rates compared with EMLA, & whether amethocaine provides superior analgesia to EMLA in a children's emergency environment.	679 children aged over 3 months (65 subjects for secondary outcomes – VAS scores). Exc: none stated. Pain: procedural (IV catheter insertion). Setting: ED (New Zealand).	2 groups – participants randomised to: EMLA group: application to 2 sites for 90 minutes, Amethocaine group: application to sites for 45minutes. Pain scoring: VASobs (observer), FLACC.	No difference in the VAS scores or the FLACC scale (p.0.05 using both a parametric t test & a non-parametric Wilcoxon test). Primary outcome: First attempt cannulation success rate for amethocaine was 75.8% compared with 73.9% for EMLA (p=0.56).	Jadad score = 3.

Oztekin et al, 2002 (401)	Randomised double-blind trial. To investigate the effect of pre-emptive diclofenac given rectally on postoperative pain scores & morphine requirements of children undergoing tonsillectomy with remifentanil–propofol anaesthesia.	40 children aged 5 – 14 years. Exc: FHx coagulation disorder, history of adverse reactions to opioids, upper GI bleeding, renal disease, asthma, acute tonsillitis, hypersensitivity to diclofenac. Pain: postoperative (tonsillectomy). Setting: PACU & ward (US).	2 groups – participants randomised to: Diclofenac group: receipt of diclofenac suppository (approx. 1 mg/kg) towards the end of anaesthesia Control group: no suppository. Pain scoring: VASobs (blinded investigator).	VAS scores lower in diclofenac group on arrival to the PACU (2.85 ± 0.77 & 7.60 ± 0.83, respectively, p<0.01). Total morphine consumption lower in diclofenac group in the PACU & the ward (p = 0.012, p = 0.021, respectively).	Jadad score = 3.
Prins et al, 2008 (402)	Double-blind placebo controlled study. To compare the effectiveness of intravenous propacetamol & rectal paracetamol in young children between 6 months & 2 years of age undergoing major craniofacial surgery.	26 children aged 6months – 2 years. Exc: cognitive impairment, contraindication to study medications. Pain: postoperative (craniofacial surgery). Setting: paediatric surgical intensive care unit (Netherlands).	2 groups – participants randomised to: IV proparacetamol: 15 min infusion of 40 mg/kg propacetamol IV every 6 hours Rectal paracetamol: 20 mg kg)1 paracetamol rectally every 6 hours. Pain scoring: VASobs (nurses, medical students), COMFORT-B.	No difference in median AUC of the VAS score: 5.2 cm/h)1 (range 0–20 cm/h)1) for the IV treatment group & 8.2 cm/h)1 (range 0–28 cm/h)1) for the rectal treatment group, respectively (P = 0.68). Fewer patients in the IV group received midazolam for COMFORT-B scores exceeding 17 (3 v 9, p < 0.05). IV group plasma concentrations were higher in the 2-5hr time period (P < 0.05). During the 24-h period, paracetamol concentrations in the IV & rectal treatment group were 68% (46–95%) & 84% (39–97%) of the time above the 10 mg/l) threshold level, respectively (P = 0.26).	Jadad score = 3.
Ravikiran et al, 2011 (386)	Randomised parallel group trial. To determine if acute pain response after administration of the BCG vaccine & the Hepatitis-B vaccine is affected by the order in which they are given.	76 healthy term neonates. Pain: procedural (immunisation). Setting: paediatric outpatient department (India).	2 groups – participants randomised to: BCG first followed by Hep B vaccine Hep B vaccine followed by BCG. Pain scoring: NIPS, VASobs (nurse).	Overall VAS scores were lower when BCG vaccine given 1st (6.25 [0.80] v 6.58 [0.54], p=0.04). Similarly overall NIPS scores for BCG vaccine given 1st lower (5.55 [0.54] vs. 5.84 [0.29], p = 0.005). Scores (VASobs & NIPS) during BCG were lower regardless of sequencing (p<0.001).	Jadad score = 3.

Rubinstein et al, 2016 (387)	Double-blind randomised controlled trial. To assess the efficacy of oral ketamine versus oral midazolam for sedation during laceration repair at a paediatric ED.	68 children aged 1 – 10 years. Exc: extensive trauma, neurologic impairment, hypersensitivity or contraindications to midazolam or ketamine, psychiatric disease. Pain: procedural (laceration repair). Setting: ED (Israel).	2 groups – participants randomised to: Ketamine group: oral ketamine 5mg/kg (max 70mg) Midazolam group: oral midazolam 0.7mg/kg (max 20mg). Pain scoring: VAS (parent & investigator), VAS (child).	No difference in VASobs scores between groups & no difference in self-reported VAS scores. No difference in sedation scores between groups: average UMSS of 1.6±0.84 vs 1.7± 0.65, respectively (MD –0.1, 95% CI: –0.47 to 0.27). Failure to achieve adequate sedation was more common among children treated with ketamine.	Jadad score = 5.
Shah et al, 2008 (388)	Double-blind, placebo-controlled, randomized trial. To determine the effectiveness & tolerability of amethocaine gel 4% in full-term neonates undergoing IM injection of vitamin K.	110 full-term neonates (≥ 37 weeks gestational age). Exc: major congenital/ neurological abnormalities, evaluation for sepsis, requiring NICU. Pain: procedural (IM injection). Setting: neonatal unit (Canada).	2 groups – participants randomised to: Amethocaine group: 1 g of topical amethocaine gel 4% Placebo group: 1g of placebo gel. Pain scoring: VASobs (Parent & nurses).	No difference in VASobs scores between groups. The mean (SD) latency to cry was significantly longer in the amethocaine group compared with the placebo group (4.7 [4.5] vs 2.7 [2.3] seconds; p = 0.01). No difference in cry duration mean (SD): 55% (34%) vs 62% (38%), respectively (p = 0.34).	Jadad score = 5.
Shaikh et al, 2011 (389)	Randomised trial. To describe the pain & distress associated with diagnostic tympanocentesis in children with AOM aged 6 to 36 months.	58 children aged 6 – 36 months Exc: sensitivity to medications, craniofacial anomalies. Pain: procedural (tympanocentesis). Setting: outpatient general paediatric clinic (US).	3 groups – participants randomised to: Group 1: paracetamol 15mg/kg orally 45min prior to procedure Group 2: paracetamol 15mg/kg plus codeine 1mg/kg orally 45min prior to procedure Group 3: ibuprofen 10mg/kg plus midazolam 0.7mg/kg 30 min prior to procedure.	No differences in VASobs scores, total cry duration or cry percentage between treatment groups. Heart rate higher during restraint phase in group 1 compared with group 2 (158 vs 137, p = 0.02) & group 3 (158 vs 139; p = 0.02). During needle aspiration, heart rate lower in group 2 compared with group 1 (162 vs 185, p<0.001) & group 3 (162 vs 186; p < .001).	Jadad score = 3.

			Pain scoring: VASobs (physician, nurse, parent), cry time, Global Mood Scale (GMS), HR.	GMS scores were higher during restraint for group 1 than group 2 (5.7 vs 4.6; p<0.001) or group 3 (5.7 vs 3.7, p<0.001). Parents assigned higher pain levels than did nurs0es or physicians (62 vs 41 vs 37; P < .001)	
Shavit et al, 2009 (390)	Single-blind, randomized, controlled trial. To examine the efficacy & safety of a new topical anaesthetic containing a disinfection ingredient (LidoDin cream) in reducing the pain associated with venepuncture by comparing it with the proven eutectic mixture of lidocaine 2.5% & prilocaine 2.5% (EMLA cream).	20 children aged 12 – 16 years. Exc: allergy to anaesthetic agents, chronic disease, requirement for urgent treatment, local skin disease, scars or tattoos at sight. Pain: procedure (venepuncture). Setting: ED (Israel).	2 groups – participants randomised to: LidoDine group: 1-2g of LidoDin gel to antecubital fossa EMLA group: 1-2g EMLA gel to antecubital fossa. Pain scoring: VASobs (nurse), VAS (child).	No difference in VASobs or VAS self-report scores between groups.	Jadad score = 3.
Simons et al, 2003 (391)	Randomised double blind placebo-controlled trial. To evaluate the effects of continuous intravenous morphine infusion on pain responses, incidence of intraventricular haemorrhage (IVH) & poor neurological outcome.	150 ventilated newborns. Exc: severe asphyxia, severe IVH, major congenital malformations & neuromuscular blocker administration. Pain: procedure: repeated procedures & stressful events. Setting: NICU (Netherlands).	2 groups – participants randomised to: Treatment: loading dose (100µg/kg) & infusion (10 µ/kg/hr) morphine Control: loading does & infusion of placebo. Pain scoring: VASobs (nurse, investigators) NIPS, PIPP.	No difference in VASobs, PIPP or NIPS scores. Morphine reduced the incidence of IVH but did not alter neurological outcomes.	Jadad score = 5.
Sinha et al, 2006 (392)	Randomised clinical trial. To evaluate the effect of using distraction as an adjunct on the sensory & affective	240 children aged 6 – 18 years. Exc: complex or multiple lacerations, other injuries.	2 groups – participants randomised to: Intervention group: offered choice of age appropriate distracters	Children < 10y: VASobs scores lower in the intervention group (0.25, 95% CI 0.11 – 0.39 vs 1.19 95% CI 0.71 – 1.67, p < 0.01). Change in FPS between groups not significant.	Jadad score = 3.

	components of pain during laceration repair among paediatric patients in the ED.	Pain: procedural (laceration repair). Setting: tertiary paediatric ED (US).	Control group: distracters not offered. Pain scoring: VASobs distress (parents) FPS (child), STAIC (child).	Children > 10y: No difference in change in VASobs scores between groups (VASobs scores 082, 95% CI 0.53 - 1.12 vs 0.086, 95%CI 0.59 – 1.12) or FPS scores (FPS scores 0.37 95%CI 0.14 – 0.59 vs 0.49 95%CI 0.29 – 0.69). Change in anxiety scores in older children lower in intervention group (STAI scores 26.72, 95%CI 25.51 – 27.93 vs 30.41, 95 %CI 29.04 – 31.78, p <0.01).	
Skarbek-Borowska et al, 2006 (393)	Randomized, double-blind, placebo-controlled trial. To determine whether brief, focal pre-treatment of children's skin with low-frequency ultrasound followed by a 5minute application of a 4% lidocaine topical anaesthetic decreases the pain of intravenous (IV) catheter placement.	77 children aged 8 – 18 years. Exc: rash, inflammation etc at site, pacemaker, allergy to lignocaine cream. Pain: procedural (IV catheter placement). Setting: ED (US).	2 groups – participants randomised to: Treatment group: Application 4% lidocaine cream following SonoPrep for 90 seconds Control group: Application of placebo cream following SonoPrep for 90 seconds. Pain scoring: VASobs (parent, nurse), VAS (child).	Self-report VAS scores lower in treatment group (mean VAS scores, 2.29 vs. 3.23; p = 0.023). VASobs scores (parent) also lower in treatment group (mean 2.47 vs. 3.39; P = 0.038). No significant difference between the groups for the nurses' VAS scores (p = 0.103).	Jadad score = 5.
Splinter et al, 1995 (403)	Randomised, single-blind investigation. To compare the effect of local anaesthesia (LA) with that of caudal anaesthesia (CA) on post-operative care of children under-going inguinal hernia repair.	202 children aged 1 – 13 year. Exc: none stated. Pain: postoperative (inguinal hernia repair). Setting: tertiary paediatric hospital (Canada).	2 groups – participants randomised to: CA group: 1ml/kg (max 20ml) 0.2% bupivacaine with 5µg adrenaline LA group: Ilio-inguinal & iliohypo-gastric nerve block plus local infiltration of 0.3mg/kg bupivocaine with 5µg adrenaline. Pain scoring: VASobs (parent), mCHEOPS.	No difference in post-operative VASobs or mCHEOPS scores between groups.	Jadad score = 3.

Splinter et al, 1997 (404)	Randomised double-blind trial. To compare the analgesic efficacy, adverse events & the costs associated with supplementation of local anaesthesia (infiltration of the wound) with either intravenous ketorolac or caudal analgesia in children having inguinal hernia repair.	164 children aged 2 – 6 years. Exc: none stated. Pain: postoperative (inguinal hernia repair). Setting: tertiary paediatric hospital (Canada).	2 groups – participants randomised to: Caudal analgesia: 1ml/kg 0.2% bupivacaine with 1/200,000 adrenaline IV analgesics: 1mg/kg ketorolac IV. Pain scoring: VASobs (parents), mCHEOPS.	VASobs (parent) scores at home were significantly lower in the ketorolac group [median 10 (0-80) vs 20 (0-80)]. In hospital mCHEOPS scores were not significantly different between groups.	Jadad score = 5.
van der Marel et al, 2001 (405)	Randomised. To determine the differences in acetaminophen plasma concentrations & effects between children receiving either multiple doses of acetaminophen rectally or equal doses of oral acetaminophen after an initial rectal loading dose.	40 children aged 3 months to 3 years. Exc: craniotomy for tumours, hydrocephalus, or trauma, liver or kidney disorders, mental retardation, GCS < 8, postoperative ventilation. Pain: postoperative (craniofacial surgery). Setting: surgical referral centre (Netherlands).	2 groups – participants randomised to: Group: 20 mg/kg acetaminophen orally Group: 20 mg/kg acetaminophen rectally After receipt of 40mg/kg rectally intraoperatively. Pain scoring: VASobs (nurse), COMFORT.	The AUC of the VAS scores (mean AUC rectal group: 16.1 cm · h; mean AUC oral group: 22.5 cm · h) & the AUC of the COMFORT scores (mean AUC rectal group: 265.4; mean AUC oral group: 286.2) were higher in patients receiving oral acetaminophen (p = 0.04 & p = 0.02, respectively). No relation between acetaminophen plasma concentrations & pain scores.	Jadad score = 5.
Van der Marel et al, 2007 (406)	Randomised controlled trial. To test the hypothesis that the addition of acetaminophen decreased morphine consumption in this age group after major thoracic (non-cardiac) or abdominal surgery.	54 infants aged 0 – 1yr ≥36 weeks PCA. Exc: current analgesics, sedatives or neuromuscular blockade, hepatic or renal disease, CNS anomalies, severe spasticity or hypotonia. Pain: postoperative (thoracic or abdominal surgery).	2 groups – participants randomised to: Group A - rectal acetaminophen (loading dose = 30 mg/kg for children <4 kg & 40 mg/kg for children ≥4 kg, followed by 20 mg/kg 6 hourly) Group B – Placebo.	VAS [median (25–75th percentile) acetaminophen 0.0 (0.0–0.2) & placebo 0.0 (0.0–0.3)] scores & COMFORT [median (25–75th percentile) acetaminophen 10 (9–12) & placebo 11 (9–13)] did not differ between groups (p = 0.73 & p = 0.06, respectively). No difference in total morphine consumption, respectively, 7.91 (6.59–14.02) & 7.19 (5.45–12.06) μg/kg/h for the acetaminophen & placebo group (p = 0.60).	Jadad score = 5. * post op morphine dosing dependent on VAS scores.

		Setting: paediatric surgical ICU (Netherlands).	Pain scoring: VASobs (nurse & investigator, COMFORT).		
Young et al. 1996 (328)	Randomised double-blinded placebo controlled study. To compare the effectiveness of EMLA cream with that of a placebo in reducing distress associated with venepuncture in paediatric outpatients.	60 children aged 6 months to 18 years. Exc: Hx congenital methemoglobinaemia, analgesics within 6 hours or taking nitrite/nitrate containing drugs. Pain: procedural (venepuncture). Setting: paediatric outpatient (US).	2 groups – participants randomised to: EMLA group – 2-3ml EMLA cream over venepuncture site Placebo group – 2-3ml placebo cream over venepuncture site. Pain scoring: VASobs (parent), VRS faces scale (parent, child). GDS.	VASobs scores lower in the EMLA group (11.7, SEM = 4.4 vs 41.3, SEM = 6.6, p = 0.0005). More parents & nurse/physicians considered EMLA effective 28/30 v16/29, p = 0.0007) & 29/30 vs 14/29, p = 0.0002) Self-reported VRS lower in EMLA group 0.31, SEM = 0.04 vs 0.60, SEM = 0.06, p = 0.001) VRSobs (parent) scores lower in EMLA group (0.41, SEM = 0.04 vs 0.68, SEM = 0.05, p = 0.0001). Change in HR was lower for EMLA group (+1.5/min vs 3.7/min). More children in EMLA group remained calm (GDS score) (24/30, 80% vs 20/29, 69%, p= 0.02).	Jadad score = 3.
Zempsky et al, 1997 (394)	Randomised trial. To compare the anaesthetic efficacy of EMLA cream with that of TAC solution for suturing uncomplicated extremity wounds.	32 children aged 5 – 18 years. Pain: procedural (wound suturing). Setting: ED (US).	2 groups – participants randomised to EMLA – 0.15g/kg (max 5.0g) TAC – 0.1mL/kg (max 3.0mL). Pain scoring: VASobs (parent, physician), VAS (self).	VASobs scores were not significantly different for physician or parent scores & VAS self-reported scores also did not differ. EMLA-group did not require supplemental anaesthesia as often as TAC group: 13 of 16 (85%} versus 7 of 16 (45%, P=.03).	Jadad score = 3.
Zempsky et al, 2008 (444)	Randomized, double-blind, sham placebo-controlled, single-dose, phase 3 study. To investigate whether a needle-free powder lidocaine delivery system (a sterile, prefilled, disposable system that delivers lidocaine powder into the epidermis) produces	579 children aged 3 – 18 years. Exc: recent venous access, local infection, tattoos, surgical scars, implantable devices, insufficient cognitive skill to participate, allergy to agents.	2 groups – participants randomised to: Intervention group: needle-free powder lidocaine delivery system Placebo group: sham placebo delivery system.	Mean VASobs scores were lower in the intervention group (21.35 ± 1.43 mm vs 28.67 ± 1.66 mm; p < 0.001). Self-report scores were lower in the intervention group (1.77 ± 0.09 vs 2.10 ± 0.09, p = 0.011). The mean self-report VAS scores for the 8- to 18-year-olds were lower in the intervention group (22.62 ± 1.80 mm vs 31.97 ± 1.82 mm, p < 0.001).	Jadad score = 5.

	effective local analgesia within 1 - 3 mins for venepuncture & peripheral venous cannulation procedures in children.	Pain: procedural (venepuncture, venous cannulation). Setting: children's hospitals (US).	Pain scoring: VASobs (parent), VAS (child) Wong-Baker faces pain scale.	No difference in the success rate of procedures between groups (P = 0.2886).	
Zempsky et al, 2008 (396)	Prospective controlled trial. To evaluate a low-frequency ultrasound device to facilitate absorption of topical anaesthetic in young children who require venepuncture.	70 children aged 3 – 7 years. Exc: need for emergent access, allergy to agents, local infection, skin disease. Pain: procedural (venepuncture). Setting: children's medical centre (US).	2 groups – participants randomised to: Intervention group: Ultrasound application followed by liposomal lidocaine applied for 5min to site Control group: liposomal lidocaine cream applied for 30min to site. Pain scoring: VASobs (parent), VAS (child) Wong-Baker faces pain scale.	No difference in mean VASobs scores between groups: (19.1 [95% CI; 10.3, 27.8] vs 23.2 [95% CI; 14.7, 31.7], p = 0.87). No difference in mean self-report scores between groups: (4.78 [95% CI; 3.06, 6.52] vs 4.32 [95% CI; 2.82, 5.82], p = 0.72).	Jadad score = 3.

Abbreviations: AUC – area under the curve, CHEOPS – Children's Hospital Eastern Ontario Pain Ccale, DAN - Douleur Aigue¨ du Nouveau-ne´, FPS – Facial Pain Scale, ED – emergency department, ENT – Ears, nose & Throat, FLACC – Face, Legs Activity Cry and Consolability, HR – Heart Rate, ICC – intraclass correlation, ICU – intensive care department, IVH – intraventricular haemorrhage, MAISD - Measure of Adult and Infant Soothing and Distress, MBPS – Modified behavioural Scale, OSBD - Observational Scale for Behavioural Distress, OSBD- R, OSBD revised, NICU – neonatal intensive care department, NRS – Numeric Rating Scale, OPS – Observational Pain Scale, PIPP- Premature Infant Pain Profile, POCIS - pain observation scale for young children (POCIS), PPAT - Pediatric Pain Assessment Tool, PPQ – Pediatric Pain Questionnaire, RCT – Randomised Controlled Trial, SD – Standard Deviation, US – United States, VAS - Visual Analogue Scale, VASobs – VAS observer.

APPENDIX D

Table 1. FLACC psychometric evaluation study details *

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
Original study						
Merkel et al, 1997 (28)	To evaluate reliability & validity of FLACC Tool. Descriptive repeated measures study. Phase 1 – 2 independent assessors score times at 5min intervals using FLACC. Bedside nurse score at final time point using global scale. Phase 2 - assessment before & after analgesic Phase 3 - 2 assessors blinded to each other's scores apply FLACC & OPS.	89 children postop aged 2mth – 7 years (mean = 3.0 ±2.0). Phase 1 – 30 children Phase 2 – 29 children Phase 3 – 30 children. Pain: Post-operative. Index: FLACC. Reference: OPS. Setting: PACU (US).	Inter-rater: Correlation between observers r[87]=0.94, p<.001) Kappa values for items range from 0.52 (face) – 0.82 (cry) COSMIN – poor.	Content: behaviours selected that had been described & validated in other tools (eg: CHEOPS, OPS TPPPS & Buttner/Finke). Piloted & revisions made. Hypothesis (convergent): Positive correlation between FLACC & OPS (r=.80, p<.001) COSMIN – poor. Responsiveness: FLACC scores decreased post-analgesic from pre=7.0 ± 2.9 to 10min=1.7 ±2.2, 30min=1.0 ±1.9, 60min=.02 ±.05 (p<0.001 at each interval). COSMIN – poor.	Not assessed	Designed to offer more feasible scale (shorter, more easily remembered). FLACC not obviously shorter (5 v 3-6 items) & feasibility not tested. FLACC comprised of items from existing scales (OPS, Buttner/Finke< CHEOPS etc) – validation included correlating with existing scales (positive results predictable).
FLACC repeat validation studies (n = 3)						
Bringuier et al, 1999 (523)	To compare the psychometric properties, sensitivity & specificity of CHEOPS, CHIPPS, FLACC & OPS (collectively BRS).	148 children generating 511 videos for children mean age 2.9 years (range 1 – 7 years). Pain: Post-operative.	Inter-rater: ICC observers was >0.86. COSMIN - good Internal consistency – Cronbach's alphas ranged from 0.81 to 0.93.	Content (face): FLACC & CHIPPS accepted by experts. Scoring out of 10 with cut-off of 3 preferred. COSMIN – poor.	Utility: discrimination (pain versus no pain) Specificity - FASS as reference: = 96% & FPS-R as reference = 89%. Sensitivity –	Scoring of multiple tools may impact on convergence. Only 32% of children provided self-report –

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	Comparative longitudinal prospective study. Children videoed for scoring at 4 time points T1 – day before surgery, T2 – pre-induction, T3 – PACU 20min post extubation, T4 – morning post surgery 4 raters scored each video using each scale in random order. Children (> 4years old) or parents rated pain on FPS-R & anxiety on VAS-anxiety scale. Group of nurses assessed face validity of each scale. FASS used to establish criterion index to evaluate validity of scales.	Index: CHEOPS, CHIPPS & OPS. Reference: na. Setting: inpatient surgical centre (France).	Cronbach's alpha for CHEOPS (0.81) higher without 2 items – complaint & touch (0.83; 0.82). COSMIN – good. FLACC results not described separately.	Structural (construct): principle component analysis showed that FLACC, CHIPS & OPS were homogeneous. All item correlations >0.4, the two lowest items from CHEOPS (r=.48 complaints & touching wound). COSMIN – excellent. Hypothesis (convergent): correlations between the 4 scales were 0.88–0.94. Correlation between the 4 scales & self-reports of pain only significant at T3 & T4. (OPS at T4, p >0.5). Correlation between BRS & FASS 0.71 – 0.78 (p<0.5). Correlation between FLACC & FACES scores (r(30) = 0.584, p=0.001). FLACC did not correlate with scores for children aged < 5years (r(14) = 0.254, p = 0.381). For children aged >5y (r[16] = 0.830; p=0.0001). Hypothesis (discriminant): – correlations with anxiety	FASS = 77% & FPS-R = 62). Risk factor for false negatives - silence (OR adjusted = 4.47, 95%CI: 1.71 – 11.55) & for false positives - level of parental-reported anxiety (p=.04).	numbers not increased in older aged children. BRS did not rate pain pre-op as 0 – authors conclude restlessness contributes to false positives. High correlation with anxiety but did not increase number of false positives. Only 11 children able to report anxiety in PACU. FLACC high sensitivity & highest specificity of the 4 scales. However, more likely to result in false negative than false positive. Pain under-reported – silence likely confounder - contributing to false negatives. Potential that all scale items cannot be adequately assessed from video footage.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
				only significant at T2 when anxiety assessed by parents (0.23–0.34) at T3 & T4 when anxiety assessed by child (T 3 : 0.63–0.77; T4: 0.54–0.78) & parents (T3: 0.22–0.25; T4: 0.27–0.37). Correlation coefficients higher using self-reports (t3: 0.63–0.77; t4: 0.54–0.78) than proxy reports of anxiety (T3: 0.22–0.25; T4: 0.27–0.37). COSMIN – fair. Responsiveness: All scales changed over time (p<.001). CHEOPS item – 'touched the wound' rarely seen. COSMIN – fair.		
Willis et al, 2003 (545)	To further test the validity of the FLACC Scale. Descriptive observational study Pain was scored post-operatively by nurse researcher using FLACC. Children independently self-reported pain using the FACES scale. 2nd nurse simultaneously &	30 children aged 3 – 7 years (5.01 ± 1.04). Pain: Post-operative. Index: FLACC. Reference: Self-report - Faces Scale. Setting: inpatient units (US).	Inter-rater agreement = 100% for 6 paired observations (17% of observations) COSMIN – poor.	Criterion (concurrent): Correlation between FLACC & FACES scores (r(30) = 0.584, p=0.001). FLACC did not correlate with scores for children aged < 5years (r(14) = 0.254, p = 0.381). For children aged >5y 9r[16] = 0.830; p=0.0001) COSMIN – fair.	Not assessed	Children 3-5y unable to adequately use faces scale most likely explanation Research team includes members of development & original research team.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	independently scored using FLACC.					
FLACC validation for alternate circumstances (age, pain, language) (n = 15)						
Ahn et al, 2007 (546)	To examine pain-like responses to frequent stimulants in the neonatal intensive care unit (NICU) using CRIES, FLACC & PIPP, & determine the clinical feasibility & validity of these tools. Exploratory correlational study Observations of baseline prior to & 8 different stimuli categorised as: A - invasive B - routine care C - auditory stimulants made by researcher using all three scales. Multiple observations from for each infant possible.	Sample: 110 consecutively enrolled infants mean age GA 32.43 weeks at birth – testing at 1 week of age. * Sedated infants & those with congenital & neurological anomalies excluded. 274 observations made across Groups A, B & C. Pain: Procedural. Index: FLACC, CRIES & PIPP. Reference: na. Setting: NICU (Korea)	Inter-rater: assessed using 10 cases BEFORE data collection – results not reported.	Hypothesis (known groups): Significant hierarchy for mean scores of the 3 groups for CRIES ($F(2,271) = 125.285$, $p<.001$), FLACC ($F(2, 271)=88.257$, $p<.001$) & PIPP ($F(2,271) = 56.504$, $p<.001$). Group A highest mean pain scores for all three tools ($p <.01$). Hypothesis (convergent): Strong correlation between CRIES & FLACC in each category ($r = .826, .843, \& .824$ for A, B & C, respectively; $p<.01$ in all). Low correlation between PIPP & CRIES & FLACC, although all 3 measures were significantly related ($.292<r<.521$, $p<.01$). Pain scores higher in full-term infants than in premature infants using CRIES (2.78 v 1.95;	Not tested	Scales applied randomly by single assesor except PIPP (last as required 30 sec delay to apply correctly – may have impacted on lower correlations between PIPP & CRIES & FLACCs. Scales all differentiated between the different levels of care. However, routine care associated with elevated scores – therefore painful or scales measuring another construct. Age related differences imply inadequacy of FLACC & CRIES for preterm infants. Superiority of PIPP claimed on the basis of higher scores for preterm experiencing auditory stimulus – however, auditory stimulus not painful.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
				p<.001) & FLACC (2.52 v 1.72; p<.01). Mean PIPP score from group C was lower in full-term infants than in premature infants (3.10 v 4.28; p<.01). COSMIN – fair.		
Bai et al, 2012 (412)	To identify 1) concurrent validity of the FLACC & COMFORT-B scales for pain assessment in Chinese children after cardiac surgery; 2) to evaluate the sensitivity, specificity, & the optimal FLACC & COMFORT-B scale cut-off scores; & 3) to explore factors that predict COMFORT-B and FLACC scores. Repeated observation study. VASobs, FLACC and COMFORT-B measures taken 2hrly during the day on day 0, 1 and 2 post-op – total of 18 measures. FLACC and COMFORT-B translated into Chinese. Content validity of	174 children aged 0 – 7 (median 8 months). (4 excluded – data for 170). Pain: Post-operative (cardiac surgery). Index: FLACC, COMFORT B scale (Chinese). Reference: VASobs. Setting: CICU, (China)*.	Inter-rater: testing results from assessment PRIOR to data collection reported – 4 assessments undertaken by two researchers – intra-class correlation FLACC = 0.84, COMFORT-B = 0.98.	Criterion (concurrent): VASobs high correlation with FLACC (r =0.86; p= .0001) & low correlation with COMFORT-$B_{Chinese}$ (r=0.31; p=.0001). COMFORT-$B_{Chinese}$ score moderately correlated with FLACC (r=0.51; p=.0001). COSMIN – poor. Hypothesis (convergent) No correlation btw scores and physiological markers (HR, ArtBP) p>.05.Multiple regression: FLACC higher scores assoc with younger age (p<.001) & relaxants (p=.021). Higher COMFORT-B scores assoc with decreased duration ventilation (p<.001) & lower age (p=.028),	Utility: COMFORT-B and FLACC scores for children in pain (VASobs≥4) were significantly higher than scores for children not in pain [VASobs<4] (p<0.0001). Used to establish cut-off – FLACC ≥2 sensitivity 98% and specificity 88% COMFORT-BChinese cut-off ≥13 sensitivity = 86% and specificity = 83%.	VASobs used as reference scale. However, Van Dijk, 2002 – cites correlation with self-report - 0.23 - .83, therefore questionable choice for reference scale. Correlations reported as criterion validity - observational scale used as reference – not a gold standard. Impact of medications not addressed – observations made following muscles relaxants or sedation in many cases - may impact on behaviour and therefore scores. Aim for haemodynamic stability, children receiving haemodynamically active medications (not reported) therefore unable to determine impact on physiological markers.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	COMFORT-B (chinese) tested using 3 experts. Testing at various cut-offs for FLACC and COMFORT-B to determine sensitivity and specificity for detecting pain / no pain (defined by expert applied VASobs (<4 = not in pain). Multiple regression analysis to determine predictors.			Lower scores assoc with analgesics (p=.008) & relaxants (p=.025). COSMIN – fair.		Lower cut-off for pain (≤2) than shown previously for FLACC (may reflect the population - sedated).
Da Silva et al, 2008 (547)	To translate, back-translate and cross-culturally adapt the content of the FLACC Scale-Revised (FLACC-R) scales for the evaluation of pain in Brazilian young students and adolescents. Three stage design. Translation and back translation from English to Brazilian Portuguese. Survey of 12 expert health professionals to assess cross cultural adaptation and content. Pre-test: FLACC – survey of clinicians to assess ability to understand & apply the scale. FPS-R – survey of	20 oncology patients aged 7 – 17 years. Index: FLACC (Brazilian Portuguese). Setting: outpatients and in patient ward in 22 health professionals (Brazil).	Not assessed	(Cross cultural) Face and content – Changes were made to the Brazilian translations from the literal translation to one where the intention was better expressed. Mean score (scale 0 - 10) for comprehension of the FLACC scale was 9.6 (±1.0). COSMIN – poor.	Not assessed	Full breadth of cross cultural validity assessment not completed. Assessor comments acknowledged some ambiguity in the descriptors for scoring - amendments made to scale to suit Brazilian application.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	children about their ability to use scale.					
da Silva et al, 2011 (548)	The aim of this research is to examine the validity and reliability of the Brazilian version of the Revised FLACC scale. Prospective observational validation study. Children with cancer diagnosis rated pain using FPS-R and simultaneously physician applied FLACC-R.	90 children aged 7 – 17 years. Pain: cancer-related. Index: FLACC, Revised Faces Pain Scale (Brazilian). Setting: inpatient and outpatient (Brazil).	Internal consistency: Cronbach's α – 0.76, correlations between items ranged from 0.12 – 0.65. COSMIN – fair.	Criterion (concurrent): Spearman's correlation between FLACC and FPS-R = 0.74. Mean FPS-R score 1.74 (SD 2.43), mean FLACC score = 0.78 (SD 1.44). COSMIN – good.	Not assessed	
Gomez et al, 2013 (474)	To establish inter-rater and intra-rater agreement of the FLACC scale in toddlers during immunization. Observational validation study. Children videotaped during immunisation procedure (Two raters scored video segments in random order and one set of raters rescored video segments 3 weeks later). FLACC scored at 4 time points, prior to immunisation, during insertion of needle and 15	30 children aged 12 – 18 months. Pain: Procedural (immunisation). Index: FLACC. Reference: not applicable. Setting: Immunisation drop in service (Australia).	Intra-rater: ICC were 0.88 at baseline, 0.97 at insertion of first needle, and 0.80 & 0.81 at 15 s and 30 s following the final injection, respectively. Inter-rater: ICC were 0.40 at baseline, 0.95 at insertion of first needle, and 0.81 and 0.78 at 15 s and 30 s following the final injection, respectively. COSMIN – good.	Not assessed	Not assessed	Raters blinded to each other and time delay and random order of presentation of video segments designed to reduce memory of segments for second application of FLACC for intra-rater reliability. Able to view video segment multiple times before scoring - may alter reliability results impacting on capacity to generalise to practice.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	and 30 seconds following completion of immunisation.					
Johansson et al, 2009 (549)	To evaluate the concurrent validity and reliability of Swedish versions of the behavioural COMFORT and a modified version of the FLACC scale for assessment of pain and sedation in intubated and ventilated children and to evaluate the construct validity of the FLACC scale for assessment of pain. Prospective observational study. 6 nurses trained to use scales, piloted to establish acceptable agreement. 40 children - 2 out of the 6 nurses applied both scales in random order at random times of day and 2 bedside nurses assessed using VASobs & NIS score. Another 20 children – 1 nurse assessed FLACC	40 children aged 0 – 108 months (median 4 months) resulting in 119 paired observations. 20 additional children aged 1 – 13 months (median 4months). Pain: Postoperative (cardiac). Index: FLACC (modified item - cry, Swedish), COMFORT scale (Swedish version). Reference: VASobs, Nurse interpretation of Sedation (NIS). Setting: PICU (Sweden).	Inter-rater: weighted kappa scores for FLACC scores 0.63 (95% CI 0.53–0.72) and COMFORT-B scores 0.71 (95% CI = CI 0.65–0.77). Weighted kappa for individual items for FLACC varied from 0.51 (activity) – 0.61 (face). COSMIN – good.	Criterion (concurrent) – Correlations between FLACC and VASobs 0.50 ($p < 0.05$), FLACC and NIS 0.50 ($p < 0.05$), COMFORT-B and VASobs,= 0.49 ($p < 0.05$) and COMFORT-B and NIS 0.57 ($p < 0.05$). Correlation between COMFORT-B and FLACC = 0.76 ($p < 0.05$). COSMIN – poor. Responsiveness – median FLACC score decreased from 5 to 0–2 ($p < 0.001$, Wilcoxon signed rank test) following morphine. COSMIN – poor.	Utility: median FLACC score for VASobs <3 = 0.5 (0 – 10) and VASobs>3 = 3.5 (0-8) and median COMFORT-B scores VASobs <3 = 12 (6 - 21) and VASobs>3 = 17 (11-23) (Kruskal-Wallis, $p < 0.01$). FLACC scores for three levels of sedation were 0 (0–3) = 'over-sedated', 0 (0–8) = 'adequately sedated' and 4 (0–8) = insufficiently sedated' (Kruskal-Wallis, $p<0.01$). COMFORT-B scores for the 3 levels of sedation were 9 (6–15), 12 (6–21) and 16 (7–23) respectively (Kruskal-Wallis, $p < 0.001$).	VASobs used as reference scale. However, Van Dijk, 2002 – cites correlation with self-report - 0.23 - .83. Therefore questionable choice for reference scale Correlations reported as criterion validity - observational scale used as reference – not a gold standard. Scale modified for use in intubated critically ill children – therefore cry altered to 'cry face or moaning' – no content validation attempted. Reliability for FLACC slightly less than shown in other studies – may be result of modifications (reliability not lowest for 'cry'). Only 7 patients with VASobs>3 therefore data only supports reliability &

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	scores before and after analgesic. Scales translated into Swedish using forwards and backwards method.					validity in lower pain states.
Malviya et al, 2006 (225)	To revise the FLACC tool to include behaviours more specific to children with cognitive impairment (CI) and evaluate the reliability and validity of the revised FLACC (modified descriptors) for assessment of pain in children with CI. Observational repeated measures comparison study. Scale revision using behaviours common to children with CI (literature) & those seen in children with CI videoed following surgery. Content validated by experts. Parents individualised scale. FLACC (2 nurses), parental (VASobs) and child's self-reported pain scores recorded independently post-op before & after analgesic. Randomly ordered videotaped segments scored	52 cognitively impaired children aged 4 – 19 years provided 80 observations. Pain: Post-operative. Index: FLACCr (modified descriptors). Reference: VASobs. Setting: recovery and ward (US).	Inter-rater: ICC = 0.75 (activity) – 0.87 (cry) and total score - 0.9 (CI: 0.87 - 0.92) p< 0.001 and kappa scores 0.44 (legs) – 0.57 (face) total score 0.5. Intra-rater: ICC = 0.97 (CI: 0.92 – 0.99). COSMIN – good.	Content (Face) – confirmed by expert physicians and advanced practice nurses. Hypothesis (convergent): Correlations between FLACC (nurse, bedside nurse and video observer) and NAPI (video observer) = 0.78 – 0.87 p<0.01, FLACC and parent VASobs = 0.65 – 0.82 p<0.01, FLACC and child report – 0.67, p = 0.051 (video observer) – 0.86, p<0.01 (bedside observer). COSMIN – good. Responsiveness: FLACC scores decreased following analgesic assessed by both video (6.1 ± 2.6 vs 1.9 ± 2.7; p < 0.001) and bedside observers (6.1 ± 2.5 vs 2.2 ± 2.4; p < 0.001) using Wilcoxon signed rank test.	Utility: FLACC scores were coded as mild (0–3), moderate (4–6) and severe (7–10) - previously defined. Reliability for clinically relevant categories. ICC = 0.83 (CI = 0.78 – 0.86).	Methodology has overcome most study flaws likely to bias results. Potential that all scale items cannot be adequately assessed from video footage. Author a member of original scale development and validation study team.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	independently by 4 nurses blinded to treatment using FLACC & NAPI. 2 nurses assigned scores to 20 randomly selected segments 3-4 weeks later.			COSMIN – poor.		
Manworren et al, 2003 (550)	To validate the FLACC Pain Assessment Tool as a clinical tool for assessing pain and evaluating pain management interventions in preverbal children. Descriptive repeated measures comparison study. Nurses assigned FLACC score when child assessed as in need of analgesic and then at regular intervals post administration of analgesic (10min, 30min and 60min).	147 children aged 1 day – 34 months (mean 1 year 40 days). Pain: nurse's impression (Rx decision based on impression). Index: FLACC. Reference: na. Setting: PACU, PICU, surgical trauma unit and haem/onc unit (US).	Established prior to commencement of study 25 nurse participants assigned FLACC to 10 videos – average weighted kappa >0.54. 19 demonstrated average weighted kappa > 0.6 and included as data collectors (14 collected data).	Responsiveness: FLACC scores before [7.03 (6.66 – 7.41)] and after analgesic [30min = 2.05 (1.68 – 2.43) and 60min = 0.74 (0.48 – 1.0)] significantly different ($p<0.001$). COSMIN – poor.	Utility: Pre-analgesic FLACC scores significantly higher for opioid group than other analgesic groups ($F[2,144]=5.55$, $p<0.005$). No significant difference in FLACC scores post analgesic based on analgesic group.	5 nurses did not collect data – didn't identify patients. Authors conclude that scale is feasible 'easy to use' based on speed (10min) with which nurses trained to achieve >0.54 interrater reliability. Efforts made to include all pain levels.
Nilsson et al, 2008 (551)	To evaluate the concurrent and construct validity and the interrater reliability of the Face, Legs, Activity, Cry and Consolability (FLACC) scale during procedural pain in children aged 5–16 years. Repeated measures study.	80 children aged 5 – 16 (mean age 10.5). Pain: Procedural pain (Cannula and percutaneous puncture to access subcutaneous access device). Index: FLACC (Swedish) Reference: CAS and FAS.	Inter-rater: weighted kappa coefficient for total FLACC score measured during the procedure = 0.85 ($p < 0.001$). COSMIN - fair	Criterion (concurrent): Spearman correlation between FLACC scores and self-report CAS ($r = 0.59$, $P < 0.05$, 5–10y = 0.59 and 11-16y = 0.5) and FAS ($r = 0.35$, $P < 0.05$). COSMIN – fair. Responsiveness: median FLACC scores before increased from 0 to 1	Not assessed	Patient group with chronic illness therefore increased pain experience likely, potentially altering pain experience and behavioural expression. Very low scores in study therefore does not contribute to assessment of validity of scale to

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	2 observers assessed child before, during and after procedure for half sample. One assessor only for second half of sample. Children self-rated pain using CAS and distress using the FAS Scales translated into Swedish and back translated.	Setting: Surgical and oncology units. (Sweden)*.		during ($p<0.0001$) and decreased to 0 after the procedure ($P < 0.001$) Wilcoxon rank sum test. CAS Scores before during and after were 0, 0.75 and 0 respectively and FAS scores were 0.37, 0.47 and 0.37 respectively. COSMIN – fair.		measure moderate and severe pain. Pain scores very low (during procedure median FLACC = 1). Polarised scores may increase the strength of correlations between scores. Although statistically significant a change in pain score from 0 to 1 of questionable clinical significance. FLACC correlated better with CAS scores (pain assessment) than FAS scores (distress assessment) suggesting FLACC assesses pain better than distress in this age group. Correlation between FLACC and self-report slightly higher in younger children. May suggest that older children demonstrate different or suppressed pain behaviours.
Ranger et al, 2013 (139)	To determine whether noxious stimuli is associated with regional cerebral hemodynamic changes and whether these	20 critically unwell infants less than 12 months. Pain: postoperative, procedural.	Inter-rater: For FLACC scores ICC = 0.86 (0.71 – 0.960). Intra-rater: For FLACC scores ICC = 0.9.	Hypothesis (convergent): No association between FLACC scores or its 5 items and NIRS or physiological signal	Not assessed	Increase in FLACC scores and across epochs blunted by administration of sedation but not analgesic (high FLACC scores

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	changes correlate with physiologic and behavioural measures in critically ill infants with congenital heart defects during chest-drain removal after cardiac surgery. Repeated measures study. NIRS monitoring during painful procedure which was also video-taped for later assessment by pain nurse specialist using the FLACC scale. NIRS and FLACC scores analysed for 3x 30sec epochs addressing: – Baseline – Tactile stimulation – Removal of drain. Assessments repeated 10 weeks later and a second assessor applied FLACC scores to half of the videos.	Index: NIRS, FLACC. Reference: na. Setting: Cardiac ICU (US).	COSMIN - poor	changes across the different time periods. Significant change in MAP (F1.38, 16.52 – 19.18, $p < 0.001$) and HR (F1.28, 24.27 = 6.87, p=0.01) across epochs. COSMIN – poor. Responsiveness: Mean FLACC pain scores at the 3 epochs were: baseline 0.25 (SD 0.12), 95% CI [0.01, 0.51]; tactile 3.25 (SD 0.56), 95% CI [2.08, 4.23]; and noxious 6.7 (SD 0.66), 95% CI [5.32, 8.08]. Overall, FLACC scores differed significantly between epochs (F1.19=102.64; p<0.001) ANOVA. Analgesic administration associated with reduced change in HbO2 (p=0.005), HbH (p=0.002) and HR (p=0.02) in response to noxious stimuli but no significant impact on FLACC score. Sedation associated with less change in HbO2 (p = 0.017), bilateral HbH (p =		during despite analgesic) and no association between FLACC and NIRS - suggests FLACC may be measuring constructs in addition to or instead of pain/ insufficiently sensitive to detect pain in critically unwell neonates. Potential that all scale items cannot be adequately assessed from video footage.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
				0.015 & 0.008), HR (p = 0.005) and SpO2 (p = 0.012) and FLACC score (p = 0.005). COSMIN – poor.		
Suraseranivongse et al. 2002 (552)	To assess: 1) agreement & correlation of nurses & parent scores, 2) difference between CHEOPS & FLACC scores as rated by nurses and parents, 3) effects of parental education on scores and 4) ease with which parents could score CHEOPS & FLACC. Descriptive comparison study. Parents trained to use CHEOPS and FLACC. One of 2 nurses [tested for inter-rater reliability (ICC > 0.9)] and parent (blinded) scored pain of child in the recovery room at 15minutely intervals using CHEOPS and FLACC.	Parents of 69 children, age 1-12 years (median 3.7, IQR 2-6.85 years). Pain: Post-operative (herniorrhaphy or hydrocelectomy). Index: parent FLACC and CHEOPS (Thai). Reference: nurse FLACC & CHEOPS (Thai). Setting: Recovery room (Thailand).	Inter-rater: High correlation between nurses & parents using FLACC (rho = 0.938, p < 0.001) & CHEOPS (rho = 0.945, p < 0.001). Agreement ICC = 0.949 & 0.977 respectively (p < 0.001). No difference between parent scores and nurse scores for FLACC (p = 0.166) or CHEOPS (0.544). COSMIN – fair.	Not assessed	Feasibility: Ease of use scores (VAS) for FLACC (3.38 ± 1.70) and CHEOPS (3.43 ± 1.75) were not different (p = 0.815). * Authors claim ease of use demonstrated by reliable application by parents (regardless of level of education).	Sequencing of application of scales not defined but applied at the same time - hence may impact on scores for scales. Pain scores very low (median FLACC = 0, IQ = 0 - 3). Polarised scores may increase the strength of correlations between scores.
(Suraseranivongse, Santawat et al. 2001)	To: 1. cross-validate scales (FLACC, CHEOPS, OPS and TPPPS in Thai children, measuring validity, reliability and practicality	167 children aged 1 – 5.5yr. Pain: Post-operative.	Inter-rater: CHEOPS = 0.9184, OPS = 0.9198 TPPPS = 0.9657, FLACC = 0.9488.	Content (face): FLACC accepted unchanged. 2 behaviours in CHEOPS and 1 in TPPPS opposed	Feasibility - Time taken to apply scales FLACC – 45.5s, CHEOPS – 59s, OPS – 44s & TPPPS- 40.1s,	Clinical intent to treat used to establish scale cut offs – pain diagnosed by nurse but analgesic decision made by

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	and 2. assess the discriminative ability of the scales. Cross validation study. Scale translated and tested for content validity by experts Behaviours before &after surgery, before & after analgesics videotaped, sequence randomly arranged then 4 observers scored using 4 scales 2 weeks later observers rescored 30 behaviours.	Index: FLACC, CHEOPS, OPS and TPPPS (Thai). Reference: na. Setting: PACU and surgical ward 2 centres, (Thailand).	Intra-rater: CHEOPS = 0.99 – 0.87, OPS = 0.95 – 0.99, TPPPS = 0.92 – 0.99, FLACC = 0.95 – 0.99. COSMIN – good.	as not seen in Thai children. COSMIN – poor. Hypothesis (convergent): correlations btw scales 0.62 – 0.77 ($p < 0.0001$) CHEOPS/OPS highest, FLACC/TPPPS lowest. COSMIN – fair. Responsiveness: significant change in scores before and after surgery before analgesic for all scales $p < 0.001$. COSMIN – good.	Scales ranked by 76.7 – 90% (FLACC – 90%) as 'feasible for clinical use', 73.3% (FLACC, OPS & TPPPS) & 80% (CHEOPS) as 'easy to use', 80 – 100% (FLACC = 86.7%) as helpful for assessment & 26.6 – 66.7 % (FLACC = 66.7%) as generally satisfied. Utility: cut off & decision to treat pain: CHEOPS highest cut-off = 6 and strongest agreement with intention to Rx by clinician, $\kappa = 0.83$, FLACC cut-off = 2, $\kappa = 0.659$).	researcher 'blinded to score' but aware that nurse diagnosed 'pain'. Those not diagnoses by nurse not referred to researcher for decision to treat therefore biasing cohort for treatment. Potential that all scale items cannot be adequately assessed from video footage. Scoring of multiple tools may impact on convergence .
Taddio et al, 2011 (488)	To investigate the reliability, validity and practicality of 3 observational measures of acute pain for the assessment of pain in infants undergoing vaccine injections. Convenience sample from an RCT. Infants having 1st vaccination in clinical trial	120 infants aged 2 – 6mth. Pain: Immunisation. Index: FLACC, MBPS and NIPS. Reference: na. Setting: private outpatient practice, (Canada).	Inter-rater: ICC > 0.85 for pre and post vaccination for all scales, FLACC – 0.85 and 0.94, NIPS – 0.9 and 0.92, and MBPS – 0.94 and 0.9 respectively. Intra-rater: FLACC (ICC, 0.98: 95% CI, 0.97–0.99), MBPS (ICC, 0.96: 95% CI, 0.94–0.97) & NIPS (ICC, 0.98:95%CI, 0.97–0.98).	Criterion (concurrent)*: Pearson correlation btw scales MBPS & FLACC (r=0.84), FLACC & NIPS (r=0.92) ($p<0.001$) &MBPS & NIPS (r=0.87) ($p<0.001$). COSMIN – poor. Hypothesis (discriminant): Scores lower for all scales ($p<0.001$) for infants	Feasibility: Agreement (ICC) for first score and final score, high for: FLACC (ICC, 0.98: 95% CI, 0.97–0.99), MBPS (ICC, 0.96: 95% CI, 0.94–0.97) & NIPS (ICC, 0.98:95%CI, 0.97–0.98). Percentage of pain assessments recorded after one viewing did	Correlations reported as criterion validity - observational scale used as reference – not a gold standard. Potential that all scale items cannot be adequately assessed from video footage.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	comparing pain associated with two vaccines (DPTaP-Hib or PCV). Videotaped & pain scored at baseline and 15sec after vaccination from video. Phase 1: single raters scored all infants using all 3 scales. 2nd rater scored 30 randomly selected infants. Phase 2: 3 different raters applied scale after one view of video. Scored again after watching video as often as required to score confidently. All raters surveyed about utility of scales.		COSMIN – good. Internal consistency: Cronbach's alpha > 0.83 for all scales at baseline & following vaccination. COSMIN – poor.	receiving DPTaP-Hib to infants receiving PCV. FLACC (5.3 versus. 7.8, p< 0.001), MBPS (6.8 versus 8.5, p<0.001) & NIPS (4.4 versus 6.2, p<0.001). COSMIN – fair. Responsiveness: significant increase scores for all scales pre and post vaccination (p<0.001), FLACC (0.6 versus 6.5), MBPS (2.3 versus 7.7) & NIPS (0.3 versus 5.3). COSMIN – good.	not differ significantly (p=0.06) among groups: MBPS (56%), NIPS (66%), FLACC (50%). Total time taken to assess pain lowest for MBPS (5h 25min), followed by the NIPS (5h 58min) and FLACC (6h 50min). User preference highest for MBPS (80%).	
Voepel-Lewis et al, 2010 (553)	To evaluate the reliability and validity of the FLACC Scale in assessing pain in critically ill adults and children unable to self-report pain. Observational design. 3 nurses independently, observed and scored pain behaviours (2 using FLACC and 1 Checklist of Nonverbal Pain. COMFORT scale (for children) before analgesic or	37 critically ill patients (8 children aged 5.6 years) to (results pooled for adults and children). Pain: acute pain pre-analgesic or procedural. Index: FLACC Reference: COMFORT scale, NVPI. Setting: ICU and PICU (US).	Inter-rater: (children) - exact agreement (58% - 83%), κ statistics (0.33 – 0.71) and ICC (0.43 - 0.92) for items and total FLACC = 0.85 (CI: 0.52 – 0.96). COSMIN – Good. Internal consistency: Cronbach α = 0.882 when all items included. Improved (00.934) with removal of cry. COSMIN – fair.	Criterion (concurrent): FLACC scores correlated highly with NVPI and COMFORT scores (rho = 0.963 and 0.849, respectively). COSMIN – poor. Responsiveness: decreases in FLACC scores after analgesic (or from painful to non-painful) (mean=5.27; SD, 2.3 versus mean= 0.52; SD, 1.1; p<.001). COSMIN – fair.	Not assessed	Correlations reported as criterion validity - observational scale used as reference – not a gold standard. Author a member of original scale development and validation study team.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	during a painful procedure, and 15 to 30 minutes after analgesic or the procedure concluded.					
Voepel_Lewis et al, 2008 (529)	To evaluate pragmatic attributes or clinical utility properties of 3 recently developed pain assessment tools for children with CI. Observational design. Clinicians scored 15 videotaped observations recorded during previous study over first 3 postoperative days Applied scale to 5 video segments each. Completion of Clinical Utility Attributes Questionnaire (CUAQ).	20 clinicians (5 physicians and 15 nurses) 15 cognitively impaired children. Pain: Post-operative. Index: r-FLACC (modified descriptors), NAPI & NCCPC-PV. Reference: na. Setting: not documented.	Interrater: ICC between participant scores and *originally* assigned scores range from -0.06 (NCCPC-PV to NAPI) to 0.92 (r-FLACC to r-FLACC) and kappa scores > 0.71 for all scores (r-FLACC to r-FLACC highest at 0.96). COSMIN – poor.	Not assessed	Utility: CUAQ scores higher for r-FLACC (49.6 ± 4.6) and NAPI (43.7 ± 6.7) compared to NCCPC-PV (24.9 ± 8.1.). Feasibility: time taken to score the r-FLACC and NAPI (2.9 ± 1.7 and 2.8 ± 1.5) shorter than for NCCPC-PV (5.1 ± 2.2 min) $p < 0.001$).	Participants' familiar with several assessment tools but unfamiliar with NAPI and NCCPC-PV which may influence assessment of ease in application. Potential that all scale items cannot be adequately assessed from video footage. Author a member of original scale development and validation study team.
Voepel_Lewis et al, 2005 (554)	To evaluate the validity of parental pain scores of children with CI. Observational design Parents individualised FLACC descriptors before application. Post-operatively children scored independently by 2	52 children aged 4 – 19 years (mean = 11.3 ±4.7 years) with cognitive impairment. Pain: Post-operative. Index: parent applied FLACCr (modified descriptors)	Not reported (2 nurses applied FLACC independently but data not reported). Agreement between parents & nurses FLACC scores (ICC = 0.78; CI = 0.63–0.87) & parents' NRS rating & nurse FLACC scores (ICC = 0.73; CI = 0.59–0.83).	Criterion (concurrent): Agreement between child's rating & parents' FLACC score (κ=0.43), COSMIN – poor. Hypothesis (convergent): correlation between parents' NRS score & FLACC scores - (ICC = 0.81; CI = 0.70–0.89). NRS scores higher than	Not assessed	Small numbers of children able to self-report n = 12. FLACC reaches acceptable levels but lower than nurses or parents NRS. Author a member of original scale development & validation study team.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	nurses & parent (who also scored using NRS) using FLACC. Observations repeated following analgesics for those who received analgesics.	Reference: simplified FACES Scale, observer NRS. Setting: not documented.	Parent scores tended to be slightly higher than the nurse ratings (FLACC bias = 0.98± 2.2; NRS bias = 1.45 ± 2.2). COSMIN – good.	their FLACC ratings (bias = 0.56 ± 1.82), nurses FLACC score (κ=0.65) & parents NRS score (κ=0.39). COSMIN – fair. Responsiveness: Parent assessment of pain after analgesics decreased (FLACC 6.6 ± 2.4 vs. 2.6±2; p = .003; NRS: 6.4 ± 2.5 vs. 3.1 ± 2.3; p = .004). COSMIN – fair.		
Voepel-Lewis et al, 2002 (438)	To evaluate the validity & reliability of FLACC tool for assessing pain in children with CI. Observational design. The child's nurse observed & scored pain with the FLACC tool before &after analgesics. Parents also scored pain using a VASobs, & where possible children self-reported pain. Observations videotaped for FLACC scoring by 2 nurses blinded to analgesics & real time pain scores. Reassessed 50 randomly selected observations 2-3 months later.	79 cognitively impaired children aged 4 – 18 (mean 10.11 ± 4.3 yr) resulting in 140 observations. Pain: Post-operative (orthopaedic or general surgery). Index: FLACC Reference: VASobs (parents), simplified FACES or NRS for self-report. Setting: Bedside (US).	Inter-rater: Moderate agreement between all observers for total scores (r = 0.507-0.778 p≤0.0001) & for each FLACC item (0.339 – 0826, p≤0.0001). Measures of exact agreement highest between blinded observers for all categories (κ scores, face = 0.346, legs = 0.477, activity = 0.405, cry = 0.652, consolability = 0.555). Intra-rater: correlations for test-retest FLACC scores for the blinded observers (r = 0.8–0.883; p< 0.001).	Hypothesis (convergent): Correlation between FLACC scores of bedside nurse (r113 = 0.651) & blinded nurses (r94 = 0.609 & 0.519) with parent scores (p< 0.001), Parent scores higher than bedside (bias 0.59 precision ±2.3) & blinded nurses (0.51 ±2.4 & 0.65 ±2.6) FLACC scores Bedside nurses scores higher than blinded nurses (0.2 ±1.6 & 0.09 ±2.4). COSMIN – poor. Responsiveness: Decrease in FLACC scores after analgesics (5.3 ± 2.8 vs 2.0 ± 2.4 for the bedside nurses' scores, p < 0.001;	Utility: excellent agreement for mild & severe pain categories (& good agreement for moderate pain. Children with mild pain most often received no analgesic (64%) or non-opioids (18%), those with moderate to severe pain most often received morphine (60%) or diazepam (6%) for muscle spasms.	Potential that all scale items cannot be adequately assessed from video footage. Video observers blinded – reducing the bias supports the influence of clinical information as scores consistently lower than both bedside nurses & parents scores. Author a member of original scale development & validation study team.

Study	Aim/Design & method	Subjects/circumstances/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	Pain scores were coded as mild (0–3), moderate (4–6), & severe (7–10).		COSMIN – fair.	5.1 ± 2.9 vs 2.2 ± 3.0 for blinded nurses' scores, p = 0.001). COSMIN – fair.		

* Reproduced from the version available online

Abbreviations: CAS – Colour Analogue Scale, CHEOPS – Children's Hospital Eastern Ontario Pain Ccale,, CHIPS – Children's and Infants' Postoperative Pain Scale, CRIES - crying, requires oxygen, increased vital signs, expression, & sleepless, FAS – Faces Analogue Scale, FPS – Facial Pain Scale, ED – emergency department, FLACC – Face, Legs Activity Cry and Consolability, HR – Heart Rate, ICC – intraclass correlation, ICU – intensive care unit, MBPS – Modified behavioural Scale, OSBD - Observational Scale for Behavioural Distress, OSBD- R, OSBD revised, NAPI - Nursing Assessment of Pain Intensity, NCCPC-PV - Non-Communicating Children's Pain Checklist-Postoperative Version, NICU – neonatal intensive care department, NIPS – Neonatal Infant Pain Scale, NIRS – Near Infra-red Spectroscopy, NRS – Numeric Rating Scale, OPS – Observational Pain Scale, PACU – Postoperative Care Unit, PICU – paediatric ICU, PIPP- Premature Infant Pain Profile, POCIS - pain observation scale for young children (POCIS), RCT – Randomised Controlled Trial, SD – Standard Deviation, TPPS – Toddler-Preschooler Postoperative Pain Scale, US – United States, VAS - Visual Analogue Scale, VASobs – VAS observer.

Table 2. FLACC RCT details *

Study	Design/Aim	Subjects/circumstances/ Setting	Intervention/Pain measure	Results	Quality score
Amin et al, 2014 (555)	Double-blind randomized (placebo-controlled) study. To evaluate the role of gabapentin premedication in the management of post-operative pain following adenotonsillectomy in children.	120 children aged 4 – 6 years. Exc: chronic illness, epilepsy. Pain: postoperative (adenotonsillectomy). Setting: Not stated.	3 groups – participants randomised to: Group G: - Oral gabapentin 10 mg/kg 2hrs preoperatively Group D: placebo pre-operatively & dexamethasone 0.15 mg/kg intravenously preoperatively after induction. Group C: Oral gabapentin 10 mg/kg 2hrs preoperatively & dexamethasone 0.15 mg/kg intravenously preoperatively after induction. Pain scoring: FLACC. * Analgesics determined by pain score.	FLACC score in Group C and Group G less at 4 h, 6 h and 8 h post-operatively than in Group D ($P < 0.05$). At 12h pain score in Group C less than Group G and Group D ($P < 0.05$). No difference in FLACC score at 18h post-operatively ($p > 0.05$). Time to first analgesic longer in Group C than Group G and Group D & time to first analgesic longer in Group G than in Group D ($P < 0.05$). Total pethidine dose less in Group C & Group G than in Group D ($p < 0.05$).	3
Anand et al, 2011 (556)	Randomised double blind parallel group (controlled) trial. To compare the effects of caudal dexmedetomidine combined with ropivacaine to provide postoperative analgesia in children and also to establish its safety in the paediatric population.	60 children aged 6mth – 6 years. Exc: developmental delay. Pain: postoperative (urogenital procedures). Setting: Not stated.	2 groups – participants randomised to: Group RD - 0.25% ropivacaine 1 ml/kg with dexmedetomidine 2 mμg/kg, Group R - 0.25% ropivacaine 1 ml/kg + 0.5 ml normal saline. Administer via caudal block following induction of anaesthesia. Pain scoring: FLACC. * Analgesics determined by pain score.	Group RD duration of analgesia longer than Group R ($p < 0.001$). Group RD had significantly* lower FLACC score compared with Group R (0/30 versus 20/30 scored 4 at 6th hour). Group RD more sedated than Group R ($P<0.001$) and the emergence behaviour score lower in Group RD ($p < 0.001$). * [p values not reported].	3
Ashrey et al, 2014(557)	Randomised trial. To evaluate the effect of penile block versus caudal block using bupivacaine on the quality of analgesia, and the surgeon's and parents'	80 children aged 1 – 7 years. Exc: neurological disease. Pain: Postoperative (penile surgeries).	2 groups – participants randomised to: Group P: penile block, 0.25% bupivacaine, 0.5 mg/kg Group C: caudal block, 0.25% bupivacaine, 0.5 mg/kg. Pain scoring: FLACC.	FLACC pain scores lower in group P compared with group C ($P < 0.05$).No decrease in HR & MAP compared with the baseline in Group P. A decrease in HR and MAP in Group C ($p < 0.05$). Time to first analgesic lower in group P compared with group C ($p < 0.05$). Total	3

	satisfaction after penile paediatric surgery.	Setting: Recovery and ward (Egypt).	* Analgesics determined by pain score.	analgesic requirement lower ($p < 0.05$) in group P than in group C.	
Babl et al, 2009 (351)	Randomised, double blind placebo controlled trial. To investigate the role of nebulized lidocaine in reducing pain and distress of nasogastric tube insertion in young children.	36* children aged 1 – 5 years. Exc: chronic disease, neurological disease, cognitive impairment. Pain: procedural (nasogastric tube insertion). Setting: ED (Australia). * trial stopped early due to concerns re distress associated with administration of trial medication.	2 groups – participants randomised to: Treatment group - nebulized 2% lidocaine at 4 mg/kg Placebo group – equivalent volume of normal saline placebo. Administered via nebuliser 10minutes prior to NGT insertion. Pain scoring: FLACC.	Nebulization highly distressing (lidocaine median FLACC: 6.3 [IQR: 3.0–8.0]; placebo median: 6.0 [IQR: 1.5–8.0]). FLACC scores during NGT insertion very high in both groups (lidocaine median: 9.8 [IQR: 9.0–10.0]; placebo median: 9.5 [IQR: 9.0–10.0]). Trend in post-NGT insertion period toward lower FLACC scores in the lidocaine group (lidocaine median: 3.5 [IQR: 1.5–6.0]; placebo median: 5.5 [IQR: 3.5–7.0]).	5
Batra et al, 2009 (558)	Randomised controlled trial. To assess spinal anaesthesia (SA) duration provided by four doses of spinal neostigmine added to bupivacaine for lower abdominal and urogenital procedures in infants.	73 infants aged 1 – 12 months. Exc: neurologic, neuromuscular, psychiatric, seizure. Pain: postoperative (lower abdominal and urogenital procedures). Setting: PACU (India).	5 groups – participants randomised to: Group B – control group, bupivacaine only Treatment groups bupivacaine and Group BN.25 – 0.25 mug/kg neostigmine Group BN.50 – 0.5 mug/kg neostigmine Group BN.75 – 0.75mug/kg neostigmine Group BN1.0 - 1 mug/kg neostigmine. Administered intrathecally. Pain scoring: FLACC. * Analgesics determined by pain score.	Groups BN.75 and BN1.0 had significantly reduced pain scores ($p < 0.001$). Analgesic requirement lower in neostigmine groups (paracetamol $p < 0.01$ & fentanyl $p < 0.001$). Linear increase in SA duration with IT neostigmine to 65.2 (4.3) min with 0.5 mug/kg ($P<0.01$), 88.2 (5.1) with 0.75 mug/kg ($P<0.001$) and 92 (4.3) with 1 mug/kg ($P<0.001$) from 52.4 (4.3) min with bupivacaine alone.	3
Bharti et al, 2014 (559)	Randomized double-blind controlled study.	78 children aged 1 – 8 years Exc: developmental delay or mental retardation.	4 groups – participants randomised to: Group 1: 0.2% ropivacaine 0.75 ml/kg Groups 2,3 & 4 receieved 0.2% plain ropivacaine 0.75 ml/kg and:	Recovery periods longer in Group 3 and 4 but no delayed emergence in any group.	5

	To evaluate the analgesic efficacy and safety of addition of three different doses of dexmedetomidine in caudal ropivacaine compared with plain ropivacaine for postoperative analgesia in paediatric day care patients.	Pain: postoperative. Setting: Day surgery unit (India).	Group 2: 0.5 µg/kg dexmedetomidine Group 3: 1.0 dexmedetomidine Group 4: 1.5 µg/kg dexmedetomidine Administered as a caudal block. Pain scoring: FLACC. * Analgesics determined by pain score.	Dexmedetomidine 1.5 g/kg were more sedated compared to the other groups (P < 0.01). Postoperative analgesia prolonged in all dexmedetomidine groups compared to plain ropivacaine group (P < 0.001) and lower pain scores (p < 0.01). All patients in the plain ropivacaine group required rescue analgesia within first 6 hours, none in Groups 2,3 and 4. HR lower in Groups 3 and 4 (p < 0.05), no difference in BP.	
Boots et al, 2010 (560)	Randomised single blind ~~controlled~~ (comparison) trial. To evaluate if discomfort levels are statistically significant when two different topical and intraurethral precatheterisation analgesia strategies are used.	200 children aged 2mth – 8 years. Exc: presentation that altered levels of pain perceptions (i.e. spina bifida, prior urethral surgery or trauma). Pain: procedural (urethral catheterisation). Setting: Radiology department (US).	2 groups – participants randomised to: Intervention group - one application of lidocaine five minutes prior to catheterisation. Control group - two applications, spaced five minutes apart prior to catheterisation. Pain score: FLACC score.	No significant differences (p = 0.779) in the mean FLACC pain score at the time of the catheterisation between the intervention group (mean = 3.30) and the control group (mean = 3.39). No comparison made between pre and during procedure FLACC scores No difference in parental perception of child's discomfort.	3
Brown et al, 2014 (561)	Parallel-group, superiority, randomized controlled trial. To investigate the association between DittoTM use and speed of burn wound re-epithelialization.	73 children aged 4 – 13 years. Exc: Cognitive, visual & auditory impairment, autistic. Pain: procedural (dressing change). Setting: burn centre (Australia).	2 groups – participants randomised to: Standard group - standard preparation and standard distraction Intervention group - DittoTM procedural preparation and DittoTM distraction. Pain scoring: Faces PS revised & FLACC (independent). * Analgesics determined by pain score.	No difference in FLACC and self-report pain scores, anxiety scores, physiological parameters or salivary cortisol. No comparison made between pre and during procedure FLACC scores. Wounds in the Ditto intervention group re-epithelialized faster than the standard practice group (-2.12 days (CI: -4.26 to 0.03), p-value = 0.046) adjusted for depth.	3

Chadha et al, 2013 (562)	Parallel randomised double blind placebo controlled superiority trial. To compare the degree of pain experienced by children undergoing flexible nasendoscopy after 1 of 3 intranasal sprays: placebo, decongestant with topical local anaesthetic (TLA), or decongestant without TLA.	23 children aged 3 – 12 years. Exc: previous nasendoscopy. Pain: procedural (nasendoscopy). Setting: otolaryngology ambulatory clinic (Canada).	3 groups – participants randomised to: Group A (control) – 0.9% sodium chloride Group B - xylometazoline hydrochloride, 0.05% Group C - lidocaine hydrochloride, 1%, with xylometazoline hydrochloride, 0.05%. 0.5ml solution sprayed in nostrils 10 min before procedure. Pain scoring: Wong Baker Faces & FLACC (independent).	Mean child-rated WBFP scale scores were 2.4, 1.8, and 2.2 for the placebo, decongestant, and TLA with decongestant groups, respectively (P=.45). Statistically non-significant - decongestant had the lowest observer-rated FLACC scale score. No comparison made between pre and during procedure FLACC scores.	5
Chandler et al, 2013 (563)	Randomised, double-blinded, controlled trial. To conduct a randomized-controlled trial comparing the incidence of ED in children following sevoflurane (SEVO) anaesthesia and propofol-remifentanil total intravenous anaesthesia (TIVA).	112 children aged 2 – 6 years Exc: developmental delay, neurological injury, psychiatric diagnosis. Pain: postoperative (strabismus repair). Setting: PACU (Canada).	2 groups – participants randomised to: TIVA group - intravenous induction and maintenance of anaesthesia with propofol and remifentanil SEVO group - inhalational induction and maintenance of anaesthesia with sevoflurane. Pain scoring: FLACC.	Incidence of ED was higher with SEVO (38.3% vs 14.9%, P = 0.018). Higher FLACC scores seen with SEVO (median 3 vs 1, P = 0.033). Subjects experiencing ED had higher FLACC scores vs those unaffected by ED (median 7 vs 1, P < 0.0001).	3
Cho et al 2009 (564)	Randomised (controlled) trial. To investigated the efficacy of epidural fentanyl to 1.25 or 1.5 mg/ml ropivacaine for post-operative epidural analgesia in children.	108 children aged 5 – 84 months. Exc: neurological diseases, seizures. Pain: postoperative (hypospadius repair). Setting: not stated.	4 groups – participants randomised to: R1.25 group - 1.25 mg/ml ropivacaine R1.25F group - 1.25 mg/ml ropivacaine with 0.2 mcg/kg/h of fentanyl R1.5 group - 1.5 mg/ml ropivacaine R1.5F group - 1.5 mg/ml ropivacaine with 0.2 mcg/kg/h of fentanyl. Pain scoring: FLACC. * Analgesics determined by pain score.	Need for rescue analgesia (FLACC >4) was higher in the R1.25 group compared with other three groups (all P<0.05). No difference between the groups in the median of the highest FLACC score (p> 0.5). The FLACC score was higher during 0–6h compared with the other three periods in all groups (p < 0.5), except the R1.25 group, difference was seen only during 6–24 h.	5

Curry et al, 2012 (565)	Randomised, double blind placebo controlled trial. To assess the effectiveness of oral sucrose to control infants' pain during routine immunizations at 2, 4, & 6 months of age.	109 infants, aged 1 – 7 months. Exc: acute or chronic disease. Pain: procedural (immunisation). Setting: ambulatory paediatric clinics of two hospitals (US).	3 groups – participants randomised to: Placebo group – sterile water Sucrose group 1 – 50% sucrose Sucrose group 2 – 75% sucrose. 2ml given orally prior to immunisation. Pain scoring: FLACC.	No difference in FLACC scores between treatment groups post injection (p = 0.646; F =.439; df = 2). No difference in crying time by treatment group (p = 0.24; F = 1.43; df = 2). No difference in crying time (p = 0..35) or FLACC score (p = 0 .697) by age group. No comparison made between pre and during procedure FLACC scores.	3
Curtis et al, 2007 (566)	Randomised, double blinded (sucrose), single blinded (dummy) placebo (sucrose) controlled trial. To determine the effect of sucrose, pacifier or the combination thereof for the procedural pain of venepuncture in infants in the paediatric ED population.	84 infants aged 0 – 6mths. Exc: critically unwell, local anaesthetic at venepuncture site. Pain: procedural (venepuncture). Setting: ED (Canada).	4 groups – participants randomised to: Group 1: sucrose Group 2: sucrose + dummy Group 3: water Group 4: water + dummy. 2ml given prior to venepuncture with or without dummy. Pain scoring: FLACC.	No significant difference in FLACC scores for sucrose groups (p = 0.66). No difference in crying time between groups (p = 0.16). FLACC and crying increased form baseline but no p value reported. FLACC scores lower with dummy use but not significant (no dummy = 4.3 +/- 4.5 dummy = 2.5 +/- 3.7, p = 0.06).	3
Dewhirst et al, 2014 (567)	Double-blinded, randomized (controlled) clinical trial. To compare the efficacy of intranasal (IN) dexmedetomidine with IN fentanyl for children undergoing BMT.	100 children aged 1 – 7.7 years. Exc: nil relevant. Pain: postoperative (myringotomy and tympanostomy tube placement). Setting: PACU (US).	4 groups – participants randomised to: Group 1 MD: midazolam premedication 0.5 mg/kg & IN dexmedetomidine 1 µg/kg Group 3 D: IN dexmedetomidine 1 µg/kg Group 2 MF: Midazolam premedication 0.5 mg/kg & IN fentanyl 2 µg/kg Group 4 F: IN fentanyl 2 µg/kg. Pain scores: FLACC & OPS (not blinded).	No difference in FLACC scores between Groups: 2, 3 and 4, higher in group MD (p < 0.05) than Group D & F. OPS scores higher in Group MD than Group D and higher for Group MF than Group D. No difference time PACU or time to hospital discharge between 4 groups. The heart rate (HR) lower in group D compared to other groups. No clinically significant difference was noted in blood pressure.	5
Diao et al, 2012 (568)	Randomized trial.	100 children aged < 13 years. Exc: nil relevant.	2 groups – participants randomised to: Drainage group	Time to resume normal activity shorter in non-drainage group (1.04 ± 0.19 vs	3

	To assess the need for routine drainage after choledochal cyst excision and Roux-en-Y hepatojejunostomy.	Pain: postoperative (choledochal cyst excision). Setting: not stated.	Non-drainage group. Pain scores: FLACC.	4.45 ± 2.51 days and 3.04 ± 0.19 vs 6.14 ± 2.61 days, respectively; P < .001). FLACC scores decreased in both groups from day 1 to 2 and 3 (p < 0.001). FLACC scores in drainage group higher than non-drainage group (day 1, 4.10 ± 0.73 vs 3.74 ± 0.44, P < 0.01; day 2, 3.10 ± 1.09 vs 1.60 ± 0.72, P < 0.001; day 3, 2.70 ± 1.21 vs 0.62 ± 0.49, P < 0.001). Day 2 & day 3, 7 (14%) and 19 (38%) of non-drainage group pain free vs none of drainage group (P < 0.01 & P <0 .001, respectively).	
Elshammaa, 2011 (569)	Double blinded randomised (controlled) trial. To evaluate the effect of ketamine, as an adjunct to fentanyl, on postoperative analgesia and duration of Postoperative Care Unit (PACU) stay, in children undergoing tonsillectomy.	60 children aged 2 – 7 years. Exc: chronic pain. Pain: postoperative (tonsillectomy). Setting: PACU (US).	4 groups – participants randomised to: F1 group: fentanyl 1 mcg/kg F2 group: fentanyl 2 mcg/kg K group: ketamine 0.5 mg/kg) FK group: fentanyl 1 mcg/kg & ketamine 0.5 mg/kg. Pain scores: FLACC. * Analgesics determined by pain score.	FLACC scores lower for Groups K (p = 0.02) and FK (p = 0.0048) than F1. Pain scores increased with surgical time (no p value reported). Group comparison (adjusted for surgical time) - difference between F1 and K (P = 0.02), and F1 and FK (P = 0.0048) groups. No difference in additional analgesia required between groups. F2 and FK group had a shorter PACU stay than F1 (P = 0.05 and 0.04 respectively).	3
El-Sharkawi et al, 2012 (570)	Randomised controlled trial. To evaluate the effect of a distraction technique using audio-visual (A/V) glasses on pain perception during administration of local anaesthesia for children.	84 children aged 5 – 7 years. Exc: history of unpleasant experiences in medical settings, experience with local anaesthesia injection, and any mental, visual, or auditory impairment. Pain: procedural (dental).	2 groups – participants randomised to: Group 1: distraction with AV glasses Group 2: no distraction. Investigator scored FLACC from video immediately after procedure. 20 recordings rescored 1 week later. Pain scores: FPS and FLACC (independent).	Lower FLACC scores (p = 0.02) and self-report scores (p < 0.001) in distraction group. No comparison made between pre and during procedure FLACC scores Intra-examiner reliability – kappa = Faces – 1.0, Legs – 0.90, Activity – 1.00, Cry – 0.91, Consolability – 0.89.	3

		Setting: dentistry clinic (Egypt).			
Fernandes et al, 2012 (571)	Randomised, double-blinded (controlled) trial. To evaluate postoperative analgesia of morphine, or clonidine, or morphine plus clonidine, added to caudal bupivacaine in children undergoing infra-umbilical urological and genital procedures.	80 children aged 1 – 10 years. Exc: neurological disability, history of epilepsy or taking CNS medication. Pain: postoperative (infra-umbilical urological & genital procedures). Setting: PACU (Brazil).	4 groups – participants randomised to: Group B – 1.0ml/kg bupivocaine 0.166% with epinephrine 1:600,000 Group BM - 1.0ml/kg bupivocaine 0.166% with epinephrine 1:600,000 + morphine 20mcg/kg Group BC - 1.0ml/kg bupivocaine 0.166% with epinephrine 1:600,000 + clonidine 1.0mcg/kg Group BMC - 1.0ml/kg bupivocaine 0.166% with epinephrine 1:600,000 + morphine 20mcg/kg + clonidine 1.0mcg/kg. Pain scores: FLACC. * Analgesics determined by pain score.	FLACC scores higher in Groups B and BC than Groups BM and BMC (p = 0.001) from 6 – 24hours post-surgery. No significant difference between groups <6 hours post-surgery (p > 0.5). No difference in time to 1st analgesia. Number requiring rescue analgesia higher in Group B & BC than BM & BMC (p = 0.018).	5
Frawley et al, 2006 (572)	Randomised double-blinded comparison trial. To determine if there are significant differences in the clinical effectiveness of levobupivacaine compared with racemic bupivacaine for caudal anaesthesia in children having lower abdominal surgery.	310 children aged 1mth to 10 year. Exc: chronic disease. Pain: postoperative (lower abdominal surgery). Setting: operating theatre and postoperative recovery room (Australia).	2 groups – participants randomised to: Group 1: bupivacaine 0.25% (2.5 mg/kg) Group 2: levobupivacaine 0.25% (2.5 mg/kg). Pain scores: FLACC. * Analgesics determined by pain score.	No significant difference in FLACC scores between groups at 30, 60, 90 and 120minutes post caudal block. No significant difference between groups in those experiencing satisfactory analgesia (FLACC <6). No difference in haemodynamic parameters intra-operatively between groups & no difference in motor blockade (extent or duration) between groups.	5
Ghai et al, 2009 (573)	Randomised double-blinded controlled trial. To compare the efficacy and safety of subtenon block (SB) versus IV fentanyl for	114 children aged 6 months - 6 years. Exc: nil relevant. Pain: postoperative (cataract surgery).	2 groups – participants randomised to: Group SB: SB with 0.06–0.08 mL/kg of 2% lidocaine and 0.5% bupivacaine (50:50) mixture and 0.2 mL/kg normal saline IV	Fewer in Group SB (n=17/58, 29.3%) required rescue analgesia than Group F (n=39/56, 69.6%, P < 0.001). FLACC scores lower in Group SB. Median time to first analgesic requirement longer in Group SB (16 [2–13] vs 4 [0.5–8.5] h in Group F) (P < 0.001).	5

	perioperative analgesia in paediatric cataract surgery.	Setting: PACU (India).	Group F: 1 mg/kg (0.2 mL/kg) of fentanyl IV and subtenon injection with normal saline (0.06–0.08 mL/kg). Pain scores: FLACC. * Analgesics determined by pain score.	Sedation scores at 1⁄2h were comparable, afterwards more in Group F anxious or crying than in Group SB (P < 0.05).	
Grove et al, 2014 (574)	A randomized, grader-blinded, comparative study. To compare the relative gentleness of a silicone tape to a paper tape in healthy infants and children.	24 infants aged 9 – 47 months. Exc: developmental delay. Pain: procedural (tape removal). Setting: dermatology research facility (US).	2 groups – participants randomised to: Left group: Silicone tape on the left and paper tape on the right side of the back Right group: Paper tape on the left and silicone tape on the right side of the back. Pain scores: FLACC (assessor blinded to treatment group).	FLACC scores lower for the silicone tape (mean difference from baseline 0.5 vs 3.3, p = .0002). Lower mean ± SEM erythema response for the silicone tape (0.93 ± 0.14 vs 1.35 ± 0.11, P = .0129). No measurable epidermal stripping occurred with the silicone tape compared to a mean ± SEM response of 0.29 ± 0.11 for the paper tape (p = 0.0039). Keratin removal was significantly less with the silicone tape (8.7 ± 0.5 μg/mL vs 15.2 ± 1.3 μg/mL, P < .0001). Few hairs were removed with either tape. No differences in parent tape preferences.	3
Hall et al, 2009 (575)	Double-blinded randomised controlled trial. To compare outcomes after open or laparoscopic pyloromyotomy for the treatment of pyloric stenosis.	180 infants aged 11 – 108 days. Exc: nil relevant. Pain: postoperative (pyloromyotomy). Setting: not stated.	2 groups – participants randomised to: Controlled: Open pyloromyotomy Treatment group: Laparoscopic pyloromyotomy. Pain scores: FLACC.	FLACC scores decreased significantly (no p value reported) over time but no difference between groups (p=0.28). Time to achieve full enteral feeding in the open pyloromyotomy group was (median [IQR]) 23·9 h (16·0–41·0) versus 18·5 h (12·3–24·0; p=0·002) in the laparoscopic group; postoperative length of stay was 43·8 h (25·3–55·6) versus 33·6 h (22·9–48·1; p=0·027).	5
Hamers et al, 1999 (398)	Double-blind, randomized, placebo controlled (2 x 2) factorial design. 1. To evaluate the effectiveness of 2 pain	83 children aged 3 - 12 years. Exc: nil relevant. Pain: postoperative (tonsil & adenoid surgery.	4 groups – participants randomised to: Group 1: 30-50mg/kg paracetamol suppository & 0.9% saline IM Group 2: 30-50mg/kg paracetamol suppository, 0.9% saline IM & SPA	No difference in FLACC, CHEOPS, VAS, Faces or Oucher scores or whether child had drunk between Groups at 1, 2, 3 hours post procedure.	3

	protocols used interchangeably to manage early postoperative T&A pain. 2. To investigate whether nurses' systematic pain assessments (SPA) improve pain management.	Setting: PACU and ward (Netherlands).	Group 3: 30 – 50mg paracetamol suppository & 1microgram/kg fentanyl intramuscularly Group 4: 30 – 50mg paracetamol suppository & 1microgram/kg fentanyl intramuscularly & SPA. Pain scores: FLACC & CHEOPS (not blinded), VASobs (parent & researcher), Faces Pain Scale & Oucher (independent).		
Hippard et al, 2012 (576)	Randomised double blinded placebo controlled trial. To compare the immediate postoperative analgesic and behavioural effects of 3 frequently used intra-operative techniques of postoperative pain control for patients undergoing BMT under general anaesthesia.	156 children aged 6 months – 10 years. Exc: nil relevant. Pain: postoperative (myringotomy & placement of ventilating tubes). Setting: PACU (US).	3 groups – participants randomised to: Group 1—intranasal fentanyl 2 g/kg (0.04 mL/kg), IV placebo (0.01 mL/kg), IM placebo (0.01 mL/kg); Group 2—IV morphine 0.1 mg/kg (0.01 mL/kg), intranasal placebo (0.04 mL/kg), IM placebo (0.01 mL/kg); Group 3—IM morphine 0.1 mg/kg (0.01 mL/kg), intranasal placebo (0.04 mL/kg), IV placebo (0.01 mL/kg). Normal saline was used for placebo. Pain scores: FLACC.	No significant difference in peak FLACC scores among the 3 groups (mean [95% CI] IN fentanyl - 2.0 [1.2–2.8], IV morphine - 2.7 [1.7–3.6] IM morphine - 2.9 [2.1–3.7] or FLACC scores at specific time points. Maximum FLACC scores correlated with other outcomes eg PAED score ($p = 0.76$), time to discharge ($p = 0.32$) and parental satisfaction with pain Mx ($p = 0.35$) ($P < 0.001$).	5
Hong et al, 2008 (577)	Randomized, (controlled), double-blind study. To determine whether caudal midazolam combined with ropivacaine affects anaesthetic requirements, recovery profiles, and post-operative analgesia compared with ropivacaine alone in paediatric day-case hernioplasty.	60 boys aged 2–5 years old. Exc: pre-existing neurological disease. Pain: postoperative – hernioplasty. Setting: not explicitly stated.	2 groups – participants randomised to: RM group: 0.2% ropivacaine 1ml/kg and epinephrine 1 : 200,000 with 50 mg/kg midazolam. R group: 0.2% ropivacaine 1ml/kg and epinephrine 1: 200,000. Given via caudal injection. Pain scores: FLACC.	Pain scores lower in the R group lower than the RM group ($p = 0.011$). No difference between groups in effect on MAP and HR. No difference between groups in ET-sevo prior to or following surgical stimuli. No difference between groups in time to extubation, emergence, drinking or discharge. No difference in sedation scores 1hr post-surgery.	3

Hong et al, 2010 (578)	Randomised double-blinded (controlled) study. To examine the effects of a single I.V. dose of dexamethasone in combination with caudal block on postoperative analgesia in children.	77 children aged 1 – 5 years. Exc: pre-existing neurological disease. Pain: postoperative – orchiopexy. Setting: PACU (Korea).	2 groups – participants randomised to: Treatment group: dexamethasone 0.5 mg/kg (max 10 mg). Control group: same volume of saline. Administered intravenously. Pain scores: FLACC, CHEOPS, VASobs (not blinded). * Fentanyl determined by FLACC/CHEOPS. Acetaminophen determined by VASobs.	FLACC & CHEOPS scores significantly lower in the treatment group (no p value reported). Fewer in the treatment group required fentanyl (7.9% vs 38.5%, $p < 0.01$) in PACU or acetaminophen (23.7% vs 64.1%, $p < 0.001$) after discharge. Time to first acetaminophen longer in the treatment group (646 vs 430 min, $p = 0.012$).	5
Hughes et al, 2013 (579)	Pilot study (randomised trial). To determine the effect of nasogastric (NG) feeding compared with oral feeding on morphine requirements after primary cleft palate repair, and secondarily on enteral intake.	50 children aged 5 – 10 months. Exc: nil relevant. Pain: postoperative - cleft palate repair. Setting: ward (UK).	2 groups – participants randomised to: O group: oral postoperative feeding NG group: NGT postoperative feeding. Pain scores: FLACC.	No difference in morphine consumption or painful episodes (FLACC ≥ 4) between groups. NG group received three times more feed over 24 hours than O group (Diff of means = -0.88, CI -114.9 to -61.3).	3
Jindal et al, 2011 (580)	Prospective randomised double blind controlled trial. To evaluate the efficacy of adding clonidine to bupivacaine in bilateral infraorbital blocks.	50 children aged less than 24 months. Exc: systemic disease that compromises neurological function. Pain: postoperative - cleft lip repair. Setting: not stated.	2 groups – participants randomised to: Group A: 1 ml solution of clonidine 1microgram/kg & 0.25% bupivacaine Group B: 1ml 0.25% bupivacaine. Administered as an infraorbital block. Pain scores: FLACC. * Analgesics determined by pain scores.	FLACC scores in group A slightly lower than in Group B (no p value reported). Time to rescue analgesia longer for Group A compared with Group B (p,0.05).	5
Jonnavithula et al, 2007 (581)	Randomised double blinded (controlled) study. To compared the efficacy of pethidine as an adjuvant to	40 children aged 5 – 60 months. Exc: major systemic illness.	2 groups – participants randomised to: Group B - 1 ml of 0.25% bupivacaine Group P - 1 ml of 0.25% bupivacaine + 0.25 mg.kg)1 body weight pethidine.	No difference in the highest FLACC scores achieved between the two groups $p = 0.15$, ($\chi^2 = 2.66$, df = 1).	3

	bupivacaine with the efficacy of bupivacaine alone for infra-orbital nerve block in alleviating postoperative pain in children undergoing cleft lip repair.	Pain: postoperative - cleft lip repair. Setting: not stated.	Pain scores: FLACC.	No difference in UMSS scores between the two groups p = 0.274 (χ^2 = 2.59, df = 2).	
Jonnavithula et al, 2010 (582)	Randomised controlled trial. To evaluate the efficacy of palatal block in children with cleft palate undergoing palatoplasty by evaluating its effects on intraoperative anaesthetic requirement, postoperative analgesia and parental satisfaction.	45 children aged 8 – 62 months. Exc: major illness, associated congenital anomalies. Pain: postoperative - cleft lip repair. Setting: postoperative recovery room (India).	3 groups – participants randomised to: Group NB - no block for control, Group S - 0.5 ml of normal saline Group B - 0.5 ml of 0.25% bupivacaine. Pain scores: FLACC. * Analgesics determined by pain scores.	The mean FLACC scores in group NB were higher than those in groups S and B. The Area Under Curve (AUC) of FLACC scores of group NB were greater than group B and S but no difference between group B and group S (p ~0.000). Time to rescue analgesic was less and the number of doses greater in the NB group (p ~0.000). Parental satisfaction with pain relief lowest in NB group (p ~0.000).	3
Kil et al, 2012 (583)	Prospective, randomized, observer-blinded (placebo controlled) study. To evaluate the effects of oral chloral hydrate on perioperative psychological and behavioural phenomena in children.	100 children aged 1 – 5 years. Exc: Behavioural disorders and use of psychiatric medications. Pain: postoperative: orchiopexy. Setting: day surgery unit (Korea).	2 groups – participants randomised to: CH group: 40mg/kg chloral hydrate Placebo group: placebo in appropriate volume. Pain scores: FLACC, CHEOPS (not blinded). * Analgesics determined by pain scores.	FLACC and CHEOPS scores lower in the CH group (p < 0.05). Fewer participants in CH group required rescue analgesic (p = 0.01). Anxiety scores lower in the CH group (45.7 vs 28.8, p < 0.001). Induction compliance of CH group better than control group (3.2 vs 4.8). Postoperative sedation was more frequent (62.7% vs 20.4%) in CH group. Postoperative emergence delirium and maladaptive behaviour changes similar between groups.	5
Kim et al, 2014 (584)	Randomised double-blind (placebo controlled) study. To assess the effect of dexme-	40 children aged 1 – 5 years. Exc: mental retardation, develop-mental delay,	2 groups – participants randomised to: D group: dexmedetomidine 1 μg/kg, followed by 0.1 μg/kg/h until the end of surgery	ET-sevo reduced in Group D (23.8-67%, p < 0.05). The incidence of emergence agitation lower in Group D than in Group S (5% vs. 55%, p=0.001).	5

	detomidine infusion on sevoflurane requirements, recovery profiles, and emergence agitation in children undergoing ambulatory surgery.	neurological or psychiatric illnesses. Pain: postoperative - ambulatory surgery. Setting: PACU (Korea).	S group: volume matched saline. Pain scores: FLACC, CHEOPS. * Analgesics determined by pain scores.	Sedation scores higher at 0min and 30min in Group D ($p < 0.05$). No difference in pain scores except at 30min CHEOPS and FLACC lower in D group ($p < 0.05$). No difference in discharge time between groups. Mean arterial pressure & HR lower in Group D during surgery ($p < 0.05$).	
Kim et al 2012 (585)	Randomised double blinded placebo controlled trial. To determine the availability of a 5% lidocaine patch used prophylactically for venepuncture or injection-related pain during induction of anaesthesia.	72 children aged 4 – 15 years. Excuse of prescription strength analgesic in previous 24 hours. Pain: procedural – venepuncture. Setting: operating room (Korea).	2 groups – participants randomised to: Group A – 5% lidocaine patch (Lidoderm) Group B – pre-treatment with a placebo patch. Pain scores: FLACC.	FLACC score during venepuncture was significantly lower for treatment group (median – 0) than placebo group (median = 4) $p<0.001$.	5
Kundu et al, 2014 (586)	Randomised double-blinded controlled stud. To examine the effects of Reiki as an adjuvant therapy to opioid therapy for postoperative pain control in paediatric patients.	38 children aged 9 months – 4 years. Exc: regional blocks. Pain: postoperative - dental work. Setting: PACU (US).	2 groups – participants randomised to: Treatment group: Reiki therapy Control group: 'sham' Reiki therapy. Pain scores: FLACC. * Unclear how analgesia requirement determined.	No difference in FLACC scores between groups and no difference in opioid requirements between groups.	5
Loetwiriyakul et al, 2011 (587)	Randomised, double-blinded (controlled) clinical trial. To compare the effectiveness of 3 mg/Kg bupivacaine administered as 1.2 mL/Kg 0.25% bupivacaine and 1.5 mL/Kg 0.2% bupivacaine for caudal block in paediatric	74 children aged 6 months – 7 years. Exc: neurological disease. Pain: postoperative intra-abdominal surgery. Setting: theatre and recovery room (Thailand).	2 groups – participants randomised to: Group A: 1.2 mL/Kg 0.25% bupivacaine Group B: 1.5 mL/Kg 0.2% bupivacaine with morphine 50 µg/Kg. Administered as a caudal block. Pain scores: FLACC.	Intra-operatively, no difference in numbers requiring rescue analgesic (group A = 67% & group B = 63%). No difference in numbers requiring muscle relaxant (group A = 49% & group B = 57%). Time to extubation shorter in Group B (9.5±1.1 min) than group A (14.3±0.9 minutes), $p < 0.01$. Time to first	5

	patients undergoing intra-abdominal surgery.		* Postop analgesics determined by FLACC score. Intra-operatively anaesthetists' judgement.	analgesic required in recovery longer in Group B (202±45 minutes) than in group A (149±27 minutes), $p < 0.05$. Time to first analgesic required in ward longer in Group B (10.4±3.1 hours) than in group A (8.2±2.0 hours) $p < 0.05$. No difference in fentanyl requirements between groups, Group A = 52.5±2.0 µg & Group B = 49.5±3.0 µg. FLACC scores lower in Group B at 8 (2 v 3) and 12 hours (2 v 3) $p < 0.05$. No difference in HR or MAP between groups.	
Lorenzo et al, 2014 (588)	Parallel group, randomized, ~~controlled~~ (comparison) trial. To evaluate ultrasound guided transversus abdominis plane block superiority over surgeon delivered regional field infiltration for children undergoing open pyeloplasty at a tertiary referral centre.	32 children aged 0 – 6 years. Exc: history chronic pain. Pain: postoperative – pyeloplasty. Setting: tertiary referral centre (Canada).	2 groups – participants randomised to: TAP Group: ultrasound guided TAP block FRI Group: wound infiltration with 0.4 ml/kg bupivacaine 0.25% with 1:200,000 epinephrine before incision. Pain scores: FLACC. * Analgesics determined by pain scores.	Mean FLACC scores lower in the RFI group (5, SD +/- 5 vs 2, SD +/- 3, $p = 0.043$) in the recovery room. Fewer in RFI group required rescue morphine administration (13 of 16 receiving transversus abdominis plane block and 6 of 16 receiving regional field infiltration, $p = 0.011$). Mean +/- SD morphine consumption lower in RFI group (0.066 +/- 0.051 vs 0.028 +/- 0.040 mg/kg, $p = 0.021$). No local anaesthetic specific adverse events.	5
Miller et al, 2011 (384)	Randomised (controlled) trial. To determine if a combined MMD protocol (preparation and distraction) will reduce the pain and distress of 3–10 year olds undergoing burn care procedures as outpatients when compared with children provided with Standard	40 children aged 3 – 10 years. Exc: cognitive impairment, sedation and anxiolytics. Pain: procedural - burn care procedure. Setting: burns outpatient centre (Australia).	2 groups – participants randomised to: Group SD: standard distraction Group MMD: Multimodal distraction. Pain scores: FLACC, Wong and Baker Faces, VASobs (not blinded).	Pain scores ($p < 0.001$) and distress scores ($p < 0.001$) lower in MMD group when compared to SD (except FLACC pre removal of dressing). HR lower in MMD group ($p = 0.04$). Length of treatment ($p < 0.05$), days to healing and the number of pain adverse events were also reduced ($p < 0.05$) with the use of the MMD protocol.	3

	Distraction (SD) (current typical treatment).				
Miller et al, 2010 (589)	Randomised controlled trial. To investigate if either MMD procedural preparation (MMD-PP) or distraction (MMD-D) has a greater impact on child pain reduction compared to standard distraction (SD) or hand held video game distraction (VG), (2) to understand the impact of MMD-PP and MMD-D on clinic efficiency by measuring length of treatment across groups, and lastly, (3) to assess the efficacy of distraction techniques over three dressing change procedures.	80 children aged mean 6.2 years (SD ± 2.3). Exc: cognitive impairment, sedation and anxiolytics. Pain: procedural (burn care procedure). Setting: burns outpatient centre (Australia).	4 groups – participants randomised to: SD group: standard distraction VG group: video game distraction MMD-PP group: MMD procedural preparation MMD group: MMD distraction. Pain scores: FLACC, Wong and Baker Faces Scale, VAS observer.	MMD groups had consistent reductions in pain levels over the three procedures compared to the SD and VG groups for child reported pain ($p < 0.001$), parent observed VAS ($p < 0.001$) and FLACC scores ($p < 0.01$). No difference between MMD-PP and MMD groups for child report, parent VAS or FLACC. No difference in physiological measures.	3
Natarajan Surendar et al, 2014 (590)	Randomised triple blind comparative study. To evaluate & compare the efficacy & safety of intranasal (IN) dexmedetomidine, midazolam & ketamine in producing moderate sedation among uncooperative pediatric dental patients.	84 children aged 4 – 14 years. Exc: nil relevant. Pain: procedural (dental). Setting: not stated.	4 groups – participants randomised to: D1 group: dexmedetomidine 1µg/kg D2 group: dexmedetomidine 1.5µg/kg M group: Midazolam 0.2mg/kg K group: Ketamine 5mg/kg (K1). Pain scores: FLACC.	Intra & post-operative FLACC scores differed between D1 (3.81 ± 0.81 & 1.29 ± 0.90), D2 (3.67 ± 0.91 & 1.14 ± 0.65) and K1 (3.52 ± 0.68 & 1.10 ± 0.89) compared to M (5.62 ± 1.12 & 2.81 ± 0.60). Procedural success rate and sedation level not statistically different No significant difference in HR, RR, BP and SpO2 between groups.	3
Newbury et al, 2009 (385)	Parallel randomised double-blind ~~controlled~~ (comparison) study.	65 children aged 3months – 15 years. Exc: nil.	2 groups – participants randomised to: Group A: amethocaine Group E: EMLA. Cream applied to two vein sites.	No difference between success rates for Groups A or E. No difference in FLACC or VAS (observer) scores between groups A and E.	3

	To determine if amethocaine improves the success of cannulation compared with EMLA and whether it is a more effective topical anaesthetic.	Pain: procedural (intravenous cannula insertion). Setting: ED (New Zealand).	Pain scores: FLACC & VASobs (not blinded).	Inter-rater reliability for FLACC – 0.86 ($p<0.0001$).	
Nilsson et al, 2013 (591)	Non-blinded randomised (controlled) clinical trial. To test if serious gaming and lollipops influence pain, distress and anxiety in conjunction with a wound care session.	62 children aged 5 – 12 years. Exc: cognitive impairment & non-Swedish speaking. Pain: procedural (wound care). Setting: day care unit (Sweden).	3 groups – participants randomised to: Serious gaming group Lollipop group Control group. Pain scores: FLACC, self-report (CAS).	FLACC scores lower in serious gaming group than in other groups - effect size (d) for serious gaming was 0.72 (95% CI, 0.07–1.35) compared with lollipops and 0.80 (95% CI, 0.14–1.42) compared with the control group. Self-reported pain (CAS), did not differ between groups. Distress (FAS) lower in serious gaming group than in lollipop group but not compared to control group. The effect size (d) for serious gaming was 0.72 (95% CI, 0.06–1.34) compared with lollipops and 0.29 (95% CI, −0.34 to 0.91) compared with the control group. Serious gaming & lollipop groups reported lower anxiety (short STAI) scores after they underwent the wound dressing than control group. The effect size (d) for changes of the short STAI before and after serious gaming was 0.44 (95% CI, −0.2 to 1.06) compared with lollipops and 0.26 (95% CI, −0.37 to 0.88) compared with control group. Individual pain intensity (CAS & FLACC) increased significantly from before to during the procedure.	3
Nilsson et al, 2009 (592)	Randomised controlled trial.	80 children aged 7 – 16 years Exc: cognitive impairment, non-Swedish speaking.	2 groups – participants randomised to: Intervention group – music for 45min from arrival in post anaesthetic care unit	No significant difference in FLACC, FAS, CAS or anxiety scores between groups.	3

	To test whether postoperative music listening reduces morphine consumption and influence pain, distress, and anxiety after day surgery and to describe the experience of postoperative music listening.	Pain: postoperative (minor procedure). Setting: PACU (Sweden).	Control – no music. Pain score: FLACC, FAS & CAS (? Not blinded). * Analgesics determined by pain score.	Morphine consumption lower in music group ($p < 0.05$). No p value reported.	
Nord et al, 2009 (593)	Randomized, controlled, single-blinded study. To examine the effectiveness of an aromatherapy intervention on the reduction of children's distress in a peri-anaesthesia setting.	94 children aged 1 – 21 years. Inc: with/without cognitive impairment. Pain: postoperative Setting: PACU (US).	2 groups – participants randomised to: Group LG – Lavender and ginger oil Group J – Jojoba oil. Applied topically and inhaled. Pain scores: parent applied FLACC.	No difference in mean FLACC score ($p = 0.55$) between groups. No difference in parental satisfaction with aromatherapy between groups.	3
Saha et al, 2010 (594)	Prospective comparative study. To evaluate a short comparison between laparoscopic and open appendicectomy in children in regards to postoperative morbidity.	60 children aged 4 - 12 years. Exc: nil relevant. Pain: postoperative (appendicectomy). Setting: department of surgery (Bangladesh).	2 groups – participants randomised to: Group A: Laparoscopic appendicectomy Group B: Open appendicectomy. Pain scores: FLACC. * Analgesics determined by pain score.	FLACC score lower in group A at 6, 24 and 48hours ($p < 0.001$). Group A analgesic requirements were lower ($p = 0.0001$). Complication rates were higher in Group B ($p < 0.05$).	3
Sethi et al, 2013 (595)	Randomised double blinded study. To compare the use of desflurane and sevoflurane to determine the postoperative emergence delirium in children undergoing cataract surgery.	88 children aged 2 – 6 years. Exc: cognitive impairment. Pain: postoperative (cataract surgery). Setting: PACU (India).	2 groups – participants randomised to: Group S: desflurane Group D: sevoflurane. Both administered with 50% nitrous oxide to maintain anaesthesia. Pain score: FLACC.	Emergence from anaesthesia faster in desflurane group ($p=0.001$). PAED scores FLACC scores, m-YPAS anxiety scores, length of PACU stay and anaesthetic duration did not differ between groups.	5
Singh et al, 2012 (596)	Randomised controlled trial.	90 children aged 1 – 10 years. Exc: active CNS disorders.	3 groups – participants randomised to:	FLACC scores lower in the Group RK ($p < 0.05$).	1

	To compare the analgesic quality and duration of ropivacaine 0.2% with the addition of fentanyl with that of ropivacaine 0.2% and the addition of ketamine.	Pain: postoperative (sub-umbilical procedures). Setting: not stated.	Group R: 0.75ml/kg ropivacaine 0.2% in normal saline Group: RK: 0.75ml/kg ropivacaine 0.2% & 0.5mg/kg ketamine Group RF: 0.75ml/kg ropivacaine 0.2% & 1microgram/kg fentanyl. Pain score: FLACC. * Analgesics determined by FLACC score or patient complaint of pain.	Mean duration of analgesia longer in Group RK ($p < 0.05$). No difference in physiological parameters.	
Stuth et al, 2011 (597)	Randomised double-blinded trial. To determine whether single-shot caudal epidural with high-dose morphine (100 µg/kg) diluted in 0.25% bupivacaine with 1: 200 000 epinephrine) after induction would lead to a higher rate of successful extubation in the operating room (OR) and to delayed and lower postoperative analgesic requirements than IV morphine given after CPB but before the end of surgery.	63 children aged 75 – 1167 days (2 – 37 months). Exc: severe preoperative neurological impairment. Pain: postoperative (stage 2 & 3 cardiac palliation procedures). Setting: CICU (US).	2 groups – participants randomised to: Group C: pre-incisional caudal morphine–bupivacaine (100 µg/kg morphine with 0.25% bupivacaine with 1 : 200 000 epinephrine, total 1 ml/kg) and post cardiopulmonary bypass (CPB) intravenous (IV) droperidol (75 µg/kg) Group IV: pre-incisional caudal saline (1 ml/kg) and post-CPB IV morphine (150 µg/kg) with droperidol (75 µg/kg). Pain scores: FLACC or NIPS. * Unclear how analgesics determined.	No difference in pain scores between groups. Group IV required earlier rescue morphine in stage 3 patients ($P = 0.02$) but not in stage 2 patients ($P = 0.189$). No difference at 12h in morphine consumption ($P = 0.085$). Morphine requirements higher for stage 2 compared with stage 3 patients ($P < 0.001$).	3
Takmaz et al, 2009 (598)	Randomised double blind ~~controlled~~ (comparison) trial. To evaluate the effectiveness of bilateral extraoral infraorbital nerve block with 0.25% bupivacaine administered at the end of	40 children aged < 2 years. Exc: neurologic, or neuromuscular disease. Pain: postoperative (cleft lip repair). Setting: recovery and ward (Turkey).	2 groups – participants randomised to: Group I - 1.5 mL 0.25% bupivacaine Group II - 1.5 mL saline and 20 mg/kg rectal paracetamol. Pain score: FLACC. * Analgesics determined by FLACC score.	Recovery room FLACC scores in group I (2.0 ± 0.6) lower than group II (8.1 ± 0.9) ($p <0.001$). FLACC scores in the first 4 hours lower in group I than group II ($p < 0.001$). No difference in physiological parameters. Time to paracetamol longer & amount less in Grp 1 ($p=0.001$). Tramadol	4

	surgery in postoperative pain relief after cleft lip repair.			requirement Grp 1 (0/20 pts) versus 20/20 in Grp II (p=0.001). Parent satisfaction scores higher in Grp 1 (p=0.001).	
Townsend et al, 2009 (599)	Randomised, prospective, double blind study. To evaluate the effects of the combination of local anaesthetics and an intravenous nonsteroidal anti-inflammatory drug (NSAID) vs NSAID alone on quality of recovery following dental rehabilitation under general anaesthesia (GA).	27 children aged 3 – 5.5 years. Exc: not stated Pain: postoperative (dental rehabilitation). Setting: PACU (US).	2 groups – participants randomised to: Control group - 1 mg/kg ketorolac intravenously within 15 minutes of case completion. Experimental group - 1 mg/kg ketorolac within 15 minutes of case completion as well as local anaesthetic infiltration. Pain score: FLACC.	No difference in mean FLACC score between the experimental or control groups (L, 2.47 ± 2.69 vs C, 2.58 ± 2.54; P < 0.88) at PACU discharge. No difference between groups for highest FLACC score (P < 0.84). FACES scores at home similar between groups (L, 0.30 ± 0.21 vs C, 0.60 ± 1.35; P < 0.92). No difference between groups (L, 2 of 11 vs C, 4 of 12; P < 0.70) in analgesic use at home.	5
Vaughan et al, 2005 (600)	Randomised double blind placebo controlled trial. To evaluate the use of 2% lignocaine gel to alleviate the pain associated with BC in young children (<2 years) in the ED.	115 children aged < 2 years. Exc: altered mental status. Pain: procedural (urinary catheterisation). Setting: ED (US).	2 groups – participants randomised to: Experimental group – 1- 2ml 2% lignocaine lubricant gel Control group – 1 – 2 ml non-anaesthetic lubricant gel. Applied to genital mucosa 2 - 3 min before catherisation and used to lubricate catheter. Pain score: FLACC.	Mean FLACC scores between the control (7.55 +/- 2.56) & study groups (7.37 +/- 2.87) during catherization did not differ. Increase in FLACC scores from pre-procedure to during procedure (p < 0.01) (Not blinded to circumstances) Pre-study - Interrater reliability, ICC (95% CI: 0.93–0.99 during time 1, 0.95–0.99 during time 2, and 0.92–0.99 at time 3).	5
Voepel-Lewis et al, 1998 (601)	Randomised double blind placebo controlled trial. To evaluate the effectiveness of simethicone in treating this discomfort.	175 children aged < 28 months. Exc: mental impairment. Pain: postoperative (minor non-invasive procedure under inhalational anaesthetic). Setting: PACU (US).	2 groups – participants randomised to: Experimental group - 0.3 ml of simethicone Control group - 0.3ml placebo. Pain score: FLACC. * analgesic determined by clinician – unclear whether aware of/measuring FLACC scores.	Both groups improved over time but simethicone group had significantly less discomfort at 20 & 30 min post treatment (p < 0.05) than that control group. Rescue analgesia given 2 (12%) simethicone grp and 9 (47%) control group.	3

Zier et al, 2008 (602)	Randomised double blind placebo controlled trial. To compare the efficacy of inhaled nitrous oxide (N2O) with enteral midazolam for sedation of children with cerebral palsy (CP) undergoing botulinum toxin A (BoNT-A) injections.	50 children aged from 1 – 16 years. Exc: nil relevant. Pain: procedural (botulinum toxin A injections). Setting: outpatient clinic sedation area (US).	2 groups – participants randomised to: Midazolam group - 0.35 to 0.5mg⁄kg to a max of 10mg (orally or rectally) and 100% O2 via mask N2O group – 70% N2O via mask, titrated by clinician and equivalent volume of saline (orally or rectally). Pain scores: FLACC & VASobs (blinded).	FLACC scores were lower for the N2O grp (4, 0 – 10) than midazolam grp (6, 0 – 10) (p=0.010). VASobs nurse and parent lower for N2O grp (p = 0.007 and p = 0.009 respectively). No difference in maximum sedation (UMSS) score between groups (0.661), sedation higher at discharge in midazolam grp (p < 0.001) No difference in parent satisfaction between groups.	5

* Reproduced from the version available online

Abbreviations: BMT – myringotomy and tympanostomy tube placement, CHEOPS – Children's Hospital Eastern Ontario Pain Scale, CICU – cardiac intensive care unit, ED – emergency department, FLACC – Face, Legs, Activity, Consolability, Cry, ICC – intraclass coefficient, OPS – Objective Pain Scale, OR – operating room, PACU – postoperative acute care unit, PAED – Paediatric Assessment of Emergence Delirium, VAS – Visual Analogue Scale, VASobs – VAS observer

APPENDIX E

Table 1. MBPS psychometric evaluation study details *

Study	Study aim / design	Subjects/Circumstance/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
Original study						
Taddio et al, 1995 (11) NOTE: study appears in tables for both methods as RCT design used to test psychometric properties.	To adapt a behavioural pain measure (CHEOPS) for use in infants and to establish the reliability and validity of the measure used to measure pain secondary to immunisations. Randomised double-blind controlled trial. 2 groups: randomised to: Control: placebo cream topical Treatment; EMLA cream topical.	96 infants aged 4 – 6 months. Pain: Procedural (immunisation). Index: MBPS. Reference: VASobs. Setting: Outpatient clinic (Canada).	Internal consistency: btw items: facial – cry, r=0.67, p<0.001, body-cry, r=0.48 p<0.001, body-facial, r=0.54 p<0.001. Item total correlation: cry - r=0.6, face -r=0.66, body - r=0.5, p<0.001, COSMIN – poor. Inter-rater: 5 raters scoring 10 infants, ICC = 0.95, P<0.001, cry – 0.96, body – 0.89, body – 0.83 p<0.001. COSMIN - fair Intra-rater: 1 rater re-scored 12 months later, ICC = 0.95, COSMIN – fair.	Hypothesis: (between groups) EMLA group had lower mean MBPS scores (6.8, SD – 1.9) than placebo group (8, SD – 1.5), p<0.001 Jadad = 1. Criterion: correlation with VASobs, observer r = 0.68 and paediatrician r = 0.74, p<0.001. COSMIN – poor. Responsiveness: increase in mean scores before (1.9, SD - 0.8) to after immunisation (7.3, SD – 1.8) p<0.01. COSMIN – fair.	Not assessed	Reliability – factor analysis not included in assessment of internal consistency, poor sample size for reliability assessment. Time frame for test re-test impressively long. Hypothesis (between groups) Jadad Quality score = 1 reducing strength of results. Part of larger RCT, between groups data also reported in publication of RCT (603). Criterion: more appropriately defined as convergent validity. Responsiveness testing: raters not blinded to circumstances resulting in potential bias.
Validation for alternate circumstances (age, pain, language)						
McClellan et al, 2003) (445)	To comprehensively describe infant procedural distress and pain across assessment modalities and to	37 infants aged 2 and 22 months Pain: Procedural (immunisation)	Inter-rater: 2 observers, 18 observations: kappa scores facial exp = 0.61, cry = 0.77, body movement = 0.67	Responsiveness: mean MBPS item avg scores increased across phases; baseline 0.84, SD-0.48, pre 0.96, SD-0.64, during	Not assessed	Average item score used for MBPS and not total score. Raters: Undergraduate researchers may not be

Study	Study aim / design	Subjects/Circumstance/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	compare similarities and differences across measures. Observational study. Heart rate measured during procedure and parents and nurses scored distress using VASobs following immunisation and Immunisations were video-taped and MBPS applied by undergraduate research assistants.	Index: MBPS*, VASobs distress, VASobs pain. *item average used rather than total score. Reference: na. Setting: Rural health facility (US).	Parents rated infants' pain and distress significantly higher than nurses rated infants' pain and distress, t = 3.91, p <.001; t = 4.88, p <.001, respectively. COSMIN – poor.	2.26 SD-0.43 (p<0.001) and post (2.05 SD-0.61) higher than baseline & pre (p<0.001). Difference in MBPS avg scores across phases (F = 97.12, p <.001). Paired samples t-tests: injection distress higher than baseline and pre-injection distress, t =−15.14, p <.001, t =−11.31, p <.001, respectively. Heart rate changes increased across phases: baseline = 115.91, SD = 24.66, pre-injection = 130.33, SD = 19.44, injection = 135.37, SD = 19.12, and recovery =124.90, SD = 42.88, (p<0.005). COSMIN – fair. Convergent: Nurse distress scores, mean = 50.08 (SD = 21.30). Nurse pain scores, mean = 46.11 (SD = 22.92). Correlated r = .90, p <.001, but distress higher t = 2.36, p <.05. Parent distress scores, mean = 75.77 (SD =		transferable to alternative raters eg: clinicians. Reliability: small sample size. Responsiveness testing: raters not blinded to circumstances resulting in potential bias, small sample size. Parents and nurses reported scores for pain and distress independently and both observers rated distress higher than pain.

Study	Study aim / design	Subjects/Circumstance/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
				24.95). Parent pain scores, mean = 64.81 (SD = 23.74). Correlated r = .36, p <.05 but distress higher t = 2.42, p <.05. COSMIN – fair.		
Taddio et al, 2011 (488)	To investigate the reliability, validity and practicality of 3 observational measures of acute pain for the assessment of pain in infants undergoing vaccine injections. Descriptive study Infants having 1st vaccination in clinical trial comparing pain associated with two vaccines (DPTaP-Hib or PCV). Videotaped & pain scored at baseline and 15sec after vaccination from video Phase 1: single raters scored all infants using all 3 scales. 2nd rater scored 30 randomly selected infants. Phase 2: 3 different raters applied scale after one view of video. Scored again after watching video as often as required to score confidently.	120 infants aged 2 – 6mth. Convenience sample from an RCT. Pain: Procedural (Immunisation). Index: FLACC, MBPS & NIPS. Reference: na. Setting: private outpatient practice (Canada).	Internal consistency: Cronbach's alpha > 0.83 for all scales at baseline & following vaccination COSMIN – Poor. Inter-rater: ICC > 0.85 for pre and post vaccination for all scales, FLACC – 0.85 and 0.94, NIPS – 0.9 and 0.92, and MBPS – 0.94 and 0.9 respectively. COSMIN – Good. Intra-rater: FLACC (ICC, 0.98: 95% CI, 0.97–0.99), MBPS (ICC, 0.96: 95% CI, 0.94–0.97) & NIPS (ICC, 0.98:95%CI, 0.97–0.98). COSMIN – Good.	Hypothesis (between groups): Scores lower for all scales (p<0.001) for infants receiving DPTaP-Hib to infants receiving PCV. FLACC (5.3 versus. 7.8, p< 0.001), MBPS (6.8 versus 8.5, p<0.001) & NIPS (4.4 versus 6.2, p<0.001), COSMIN – Fair. Criterion: Pearson correlation btw scales MBPS & FLACC (r=0.84), FLACC & NIPS (r=0.92) (p<0.001) &MBPS & NIPS (r=0.87) (p<0.001). COSMIN – poor. Responsiveness: significant increase scores for all scales pre & post vaccination (p<0.001), FLACC (0.6 vs 6.5), MBPS (2.3 vs 7.7) & NIPS (0.3 vs 5.3). COSMIN – Good.	Feasibility: Agreement (ICC) for first score & final score, high for: FLACC (ICC, 0.98: 95% CI, 0.97–0.99), MBPS (ICC, 0.96: 95% CI, 0.94–0.97) & NIPS (ICC, 0.98:95%CI, 0.97–0.98). Percentage of pain assessments recorded after one viewing did not differ significantly (p=0.06) among groups: MBPS (56%), NIPS (66%), FLACC (50%). Total time taken to assess pain lowest for MBPS (5h 25min), followed by the NIPS (5h 58min) & FLACC (6h 50min). User preference highest for MBPS (80%).	Principle investigator also scale designer & PI in original validation study. Raters: 4 undergraduate students & 1 graduate student (discipline unknown) - may not be transferable to alternative raters eg: clinicians. Hypothesis: part of larger RCT, between groups data also reported in publication of RCT (374). Criterion: more appropriately defined as convergent validity. Responsiveness testing: raters not blinded to circumstances resulting in potential bias. Feasibility: Only 5 raters assessed feasibility of scale & rated preference - study PI original scale developer therefore biased.

Study	Study aim / design	Subjects/Circumstance/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	All raters surveyed about utility of scales.					
Alternate index scale (Scale as reference scale - concurrent validity testing)						
Cohen et al, 2005 (604)	To develop & validate a scale to examine the unique behaviours exhibited by infants, their caregivers & the nursing staff during painful procedures. Observational study. Parents & nurses assessed pain following procedure & immunisation video-recorded for review by unreported number of reviewers.	62 infants aged between 0.13 – 1.86 years & parents. Exc: chronic illness, treatment with medications likely to alter pain response. Pain: procedural (immunisation). Index: Measure of Adult & Infant Soothing & Distress (MAISD). Reference: MBPS (item average rather than total) . Setting: Rural health care facilities (USA).	Inter-rater: 18 randomly selected participant kappa scores - Facial expression = 0.61, Cry = 0.77 & Movements = 0.67. COSMIN – poor.	Hypothesis testing: (Convergent) MBPSavg correlated with MAISD scores ($r = 0.44$, $p<0.001$). COSMIN – fair.	Not assessed	Assumption of MBPS validity, index measure is untested therefore data cannot be considered evidence of convergent validity. An averaged MBPS total score ranging from 0 – 3.33 was used. No distribution of scores provided. Reliability: small sample size. Responsiveness testing: raters not blinded to circumstances resulting in potential bias.
Mijovic et al, 2010 (253)	To evaluate whether Empirical Mode Decompression (EMD) is a suitable technique for analysing infant cries & to assess the existence & the extent of decoupling in term neonates & whether as association between decoupling & clinical pain	24 term neonates Exc: GA < 37 weeks, APGARs > 7 at 1 & 5 minutes, wt > 2.5kg. Pain: Procedural (blood sampling). Index: EMD analysed cry Reference: MBPS.	Not assessed for MBPS	Hypothesis: (Convergent) correlation between MBPS & the mean correlations between the fundamental frequencies & intensity contours over all cry bouts for each subject ($r = 0.55$, $p = 0.006$), COSMIN – poor.	Not assessed	Assumption of MBPS validity, index measure is untested therefore data cannot be considered evidence of convergent validity. No methods for collection of MBPS data documented – therefore unable to assess quality etc.

Study	Study aim / design	Subjects/Circumstance/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	expression could be unveiled. Observational study. Vocalisations recorded & cry bouts analysed.	Setting: postnatal unit, University Hospital (Belgium).				Selection of cry bouts for analysis unclear eg: all or a subset & if a subset selection unclear. Hypothesis: sample size small.
Silva et al, 2010 (262)	To assess the existence & extent of decoupling in term neonates (neurodevelopmental relevance) & whether an association between decoupling & clinical pain expression could be unveiled (clinical relevance). Observation study. Blood taking procedure video-taped & cry recorded for analysis.	47 healthy term infants Exc: GA < 37 weeks, APGARs > 7 at 1 & 5 minutes, wt > 2.5kg. Pain: procedural (blood sampling). Index: cry decoupling Reference: MBPS. Setting: postnatal unit, University Hospital (Belgium).	Not assessed	Hypothesis: (Convergent) weak, non-significant positive correlation (r2=0.01) between MBPS score and number of cries. No relationship between MBPS scores and fundamental frequencies (r2=-0.01) or the SD of the fundamental frequencies (r2=0.03) COSMIN – Fair.	Not assessed	Assumption of MBPS validity, index measure is untested therefore data cannot be considered evidence of convergent validity.
Taddio et al, 2009 (433)	To test the reliability and validity of observer-rated pain in infants undergoing immunization using the VASobs. RCT – double blinded placebo controlled trial. Participants randomised to 2 groups:	120 infants aged 1-year old. Convenience sample from an RCT. Pain: procedural (immunisation). Index: VAS*. Reference: MBPS.	Inter-rater and intra-rater: not reported for MBPS.	Hypothesis: (convergent) correlations between VAS and MBPS scores ranged from 0.81–0.94. COSMIN – good. Criterion: VAS correlations with MBPS range from 0.81 – 0.94 using Pearson's rho. COSMIN – poor.	Not assessed	Hypothesis: Part of larger RCT, no between groups data reported for this study. Criterion: more appropriately defined as convergent validity. Responsiveness testing: raters not blinded to circumstances resulting in potential bias.

Study	Study aim / design	Subjects/Circumstance/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
	– 4% amethocaine topically – Placebo topically. Immunisation video-taped, VAS scores allocated real time and from video. MBPS scores allocated to video. Raters repeated all assessments for same child on a 2nd occasion.	Setting: Paediatric Outpatient clinic (Canada).		Responsiveness: not analysed for MBPS.		* VASobs used in original study to support validation of MBPS.
Veriotis et al, 2015 (605)	To describe the event related activities in the brain during immunisation using EEG. Observational study. Infants EEG monitored during procedure and video-recorded using a high-speed camera synchronised with the EEG machine. 2 independent observers identified point of needle contact with skin for analysis 2 independent raters scored the procedure epochs from the video using MBPS	15 infants aged 1 – 2 months and 12 months (18 inoculations) Exc: asphyxiated at birth, currently taking medication. Pain: procedural (immunisation). Index: EEG. Reference: MBPS. Setting: outpatient clinic (England).	Inter-rater (2 independent observers): ICC baseline = 0.81, procedural = 0.89. COSMIN – poor.	Hypothesis: (Convergent) No relationship between peak to peak amplitudes of the EEG and the MBPS scores 1- to 2-month-olds (Spearman rank order correlation coefficient; waveform 1: r = 0.15, P = 0.62 and waveform 2: r = 0.20 P = 0.52; n = 13) Not explored in12mth due to identical scores. (Between groups) MBPS immunisation scores higher in 12mth olds (9.0, [9.0 – 9.0]) than 1 - 2 month old infants (8.0 [7.5 – 8.0]) Wilcoxon–Mann–Whitney test; mean rank 15.0. vs 7.4 in 12-month vs 1-to 2-month-olds, respectively;	Not assessed	Reliability: small sample size. Hypothesis testing: (convergent) small sample size and correlated with an as yet untested index measure. Responsiveness only reported descriptively - likely to be a significant increase from baseline to procedure. Small sample size.

Study	Study aim / design	Subjects/Circumstance/ Setting/Pain measures	Results (inc quality score)			Comments
			Reliability	Validity	Feasibility & clinical utility	
				Z52.93, P = 0.002. COSMIN – poor. Responsiveness: MBPS scores increased from baseline 2.0 [2.0 – 2.0]) to 8.0 [7.5 – 8.0] in 12mth olds and 9.0 [9.0 – 9.0] in 1- 2mth olds. No significance value reported. COSMIN – poor.		

* Reproduced from the version available online

Abbreviations: CHEOPS – Children's Hospital Eastern Ontario Pain Scale, ED – emergency department, EEG electroencephalography, FLACC – Face, Legs, Activity, Consolability, Cry, ICC – intraclass coefficient, MAISD - Measure of Adult and Infant Soothing and Distress, MBPS – Modified Behavioral Pain Scale, NIPS – Neonatal Infant Pain Scale, OPS – Objective Pain Scale, OR – operating room, PACU – postoperative acute care unit, PAED – Paediatric Assessment of Emergence Delirium, RCT – randomised controlled trial, SD - Standard Deviation, US – United States, VAS – Visual Analogue Scale, VASobs – VAS observer

Table 2. MBPS RCT details *

Study	Study aim/Design	Subjects/Circumstances/ Setting	Intervention / Pain measures	Results	Quality score / Comments
Abuelkheir et al, 2014 (350)	Randomized, double-blind, placebo-controlled study. To evaluate the effectiveness of topical eutectic mixture of local anaesthetics (EMLA) cream in reducing the pain associated with vaccination injections.	216 children aged 2 months to 6 years. Exc: analgesic or sedative in last 12hrs. Pain: procedural (Immunisation). Setting: well baby paediatric clinic (Saudi Arabia).	2 groups – participants randomised to: Treatment group: EMLA cream Control group: placebo cream. Pain scoring: MBPS, VASobs, crying.	The difference between pre- & post-vaccination MBPS scores was lower in the EMLA group than in the placebo group (2.56 ± 1.96 versus 3.95 ± 2.20, respectively). The VAS scores at time of needle prick & after injection were lower in the EMLA group than the placebo group (1.60 ± 1.67 vs 3.24 ± 2.01; 3.29 ± 2.27 vs 4.86 ± 2.20; respectively). Fewer infants & children cried after the injection in the EMLA group than in the placebo group: 22.4% of children (n = 24) in the EMLA group did not cry compared with 7.3% (n = 8) in the placebo group (P=0.002). Total crying time was shorter in the EMLA group than the placebo group (24.8 ± 20.6 s vs 43.3 ± 20.5 s, respectively; P<0.001).	Jadad score = 5.
Anninger et al, 2007 (606)	Double –masked randomised control trial. To determine if the emergence agitation seen after eye muscle surgery was in part related to pain and whether topical anaesthetic would decrease postoperative pain and thereby diminish the incidence of emergence agitation.	88 children aged 1 to 12 years. Pain: postoperative (strabismus surgery). Setting: PACU (USA).	3 groups – participants randomised to: Group A: normal saline drops before and after surgery Group B: saline drops before and tetracaine 1% after surgery Group C: tetracaine 1% drops before and after surgery. Pain scoring: MBPS, crying. * Analgesics determined by verbal complaints of pain, crying and pain score.	Group C had lower maximum MBPS score than Groups A and B (p < 0.033). At 5min post PACU arrival: more Grp A patients had MBPS scores greater than 5 than Grp B or C patients (p < 0.013) & more patients in Grp A had emergence scores > 2 than in Grps B & C (p < 0.019). No difference between groups in total PACU time, PACU vomiting, PACU morphine use, or pain at home.	Jadad score = 4.

Carbajal et al, 2008 (607)	Randomised double blinded multicentre study. To compare the efficacy of EMLA with premixed 50% nitrous oxide/oxygen (N2O/O2), used alone or combined with EMLA, for pain alleviation during palivizumab injections.	55 children aged less than 24 months. Exc: analgesic or sedative drug during the preceding 12 hours. Pain: procedural (injection). Setting: pulmonary outpatient department (2 hospitals in France).	3 groups – participants randomised to: Group EMLA: EMLA plus air inhalation Group nitrous: inhalation of 50/50 N2O/O2 plus application of a placebo cream. Groups nitrous plus EMLA: inhalation of 50/50 N2O/O2 plus application of EMLA.	MBPS scores during the injection were lower in the nitrous and EMLA group (8.2 ± 1.8) than in the other 2 groups EMLA (9.3 ± 1.0) and nitrous (8.8 ± 1.2) and during the recovery phase MBPS scores were lower in the nitrous and EMLA group (6.9 ± 2.4) than in the EMLA group (7.8 ± 1.7) and the nitrous group (7.4 ± 1.9) ($p < 0.001$). VAS scores were EMLA 45.9 (22.1), nitrous 40.4 (22.6), and nitrous & EMLA 37.4 (23.4) for EMLA, N2O/O2, and N2O/O2 plus EMLA. A within-subjects factor analysis showed a treatment effect ($P = 0.019$).	Jadad score = 5. Authors note that face mask did not impact on scoring MBPS – unclear how this was established other than assumption based on scorers providing a 'face' item score.
Cohen et al, 2002 (608)	Randomised study. To examine nurse-directed distraction for reducing infant immunization distress.	90 infants aged 2months to 3 years. Pain: procedural (immunisation). Setting: rural health department (US).	2 groups – participants randomised to Group 1: Typical care Group 2: Distraction. Pain scoring: MBPS.	MBPS scores lower for distraction group ($M = 1.24$, $SD = 0.27$) than the control group ($M = 1.50$, $SD = 0.38$), $F(1, 88) = 12.75$, $p < 0.001$. MBPS scores increased across phases from baseline ($M = 0.73$, $SD = 0.37$) to anticipatory ($M = 0.82$, $SD = 0.48$), $t(89) = -2.82$, $p<0.001$; to injection ($M = 2.12$, $SD = 0.51$), $t(89) = -22.37$, $p<0.001$; and decreased in recovery phase ($M = 1.76$, $SD = 0.65$), $t(89) = -14.14$, $p<0.001$.	Jadad score = 0.
Cramer-Berness et al, 2005 (609)	Randomised study. To compare the benefits of parent guided distraction and parent comforting with standard care during infant immunisation.	123 infants aged 2 months to 2 years. Pain: procedural (immunisation). Setting: health care clinic (US).	3 groups – participants randomised to Distraction care: parents trained to provide distraction Supportive care: parents encouraged to use coping promoting strategies Standard care: parents not encouraged/trained to offer specific care.	MBPS scores not different during preparation phase or during immunisation. During recovery phase, pairwise comparisons showed infants in the typical care group more distressed than infants in the supportive care group ($p = .025$, $d = 0.57$). No other significant comparisons.	Jadad score = 1.

			Pain (distress) scoring: MBPS		
Dyer et al, 2004 (610)	Randomised controlled trial. To quantify and describe patient experience by a randomized crossover trial of G-CSF administration.	20 children* aged 1mth to 18 years. Pain: procedural (G-CSF injection administration). Setting: haematology/ oncology department (Australia).	2 groups: participants randomised to: Group 1: G-CSF via subcutaneous injection Group 2: G-CSF via Insuflon. Pain scoring: MBPS (n=7, age < 7 years), CAS and FAS (n=13).	No difference in MBPS or FAS scores btw groups. Trend for CAS scores to be higher in subcutaneous group (p = 0.11). Six out of seven children <7 years preferred using Insuflon for subcutaneous drug administration.	Jaded score = 3. * terminated early due to recruiting difficulties. Sample size not large enough to generate significant results.
Fallah et al, 2016 (611)	Randomised controlled non-blinded trial. To evaluate the effect of the order of injection (DwPT and MMR Or MMR and DwPT) on pain of intramuscular DwPT vaccine at 18 months of age.	70 infants aged 18 months old. Pain: procedural (immunisation). Setting: primary health care centre (Iran).	2 groups: participants randomised to: Group I: IM DwPT vaccine first and SC MMR vaccine second Group II: SC MMR vaccine first and IM DwPT vaccine second. Pain scoring: MBPS, crying time.	Pain scores did not differ between groups for each vaccination did not differ significantly. Total overall pain score was lower in Group I (15.61 ± 2.6) compared with Group II (14.23 ± 1.35) p = 0.04. Cry duration did not differ significantly between groups. Pain scores increased from pre-vaccination to during both vaccinations in both groups (eg: pre Group 1 2.26 ± 0.44 DPT vaccine 8.83 ± 1.59). No statistical comparison made.	Jadad score = 3.
Girish et al, 2014 (612)	Randomised controlled trial. To compare the acute pain response during immunization in infants using a slow" standard" injection technique vs. "pragmatic" technique.	200 children aged between 6 weeks and 1 and ½ years. Pain: procedural (immunisation). Setting: hospital (India).	2 groups: participants randomised to: Standard group: standard slow technique Pragmatic group: rapid technique. Pain scoring: MBPS, crying time.	Mean post-vaccination MBPS in standard group was 8.4 (SD – 0.75) and in pragmatic group was 7.8 (SD – 1.17) (p = 0.00). Mean crying duration in pragmatic group was less (32.1 s) than standard group (37.37 s).	Jadad score = 3.
Hillgrove et al, 2013 (613)	Randomised trial. To examine whether the agent of distraction (ie, the specific	99 children aged 12 to 20 months.	2 groups: participants randomised to: Typical care group:	Post-needle pain did not significantly differ among groups. Children who were distressed pre-needle displayed	Jadad score = 3. Very high scores pre procedure – evidence

	person conducting the distraction) & pre-needle distress behaviours impact the efficacy of distraction when toddlers were held by parents.	Exc: cognitive impairment, children born before 36 weeks. Pain: procedural (immunisation). Setting: paediatrician's clinic (Canada).	RA directed distraction group: Pain scoring: MBPS.	significantly more pain post-needle, regardless of the treatment group. 40.8% of children exhibited pre-needle pain scores 3 – 9.	that measuring another construct eg: anticipatory distress. Authors describe the tool.
Hogan et al, 2014 (367)	Randomised partially-blinded parallel , 2 group trial. To determine the effectiveness of parent-led tactile stimulation for pain reduction when added to a combination of evidence-based pain-reducing interventions in infants undergoing immunization injections.	120 infants aged 4 to 6 months. Exc: impaired neurological development, previous seizures, local anaesthetic at site, use of sedatives or opioids in last 24 hours. Pain: procedural (immunisation). Setting: primary care practice (Canada).	2 groups: participants randomised to; Usual care group: Tactile stimulation group. Pain scoring: MBPS, VASobs (parent).	No difference in mean MBPS scores & parent VAS scores between groups (8.2 [1.1] vs. 8.0 [1.3]; P=0.57) and (60 [20] vs. 53 [22] mm; P=0.10), respectively.	Jadad score = 3.
Ipp et al, 2004 (371)	Randomised double blinded clinical trial. To compare acute pain response to 2 measles-mumps-rubella (MMR) vaccines.	49 infants aged 12 months. Pain: procedural (immunisation). Setting: community paediatricians clinic (Canada).	2 groups: participants randomised to: Priorix group MMR-II group. Pain scoring: MBPS, VASob (parent and paediatrician), latency to cry and cry duration.	Median pain scores after vaccination (Priorix vs M-M-R II) were as follows: pediatrician VAS, 15 vs 58 (P=.001); parent VAS, 22 vs 53 (P=.007); and MBPS, 6 vs 8 (P=.02). The median latency to first cry was 1.5 seconds in the Priorix group and with 1 second in the M-M-R II group (p=.26). Median difference in pain scores (after minus before) for Priorix vs M-M-R II were as follows: paediatrician VAS, 15 vs 53 (p=.003); parent VAS, 22 vs 47 (P=.008); and MBPS, 3 vs 5 (P=.03).	Jadad score = 5. COSMIN – responsiveness 'fair'.
Ipp et al, 2009 (374)	Single centre, double blinded randomised clinical trial.	120 infants aged 2 to 6months.	2 groups: participants randomised to:	MBPS & parent VAS scores lower when DPTaP-Hib was administered first than	Jadad score = 5.

	To determine if acute pain response after administration of the diphtheria, polio, and tetanus toxoids and acellular pertussis and Haemophilus influenzae type b (DPTaP-Hib) vaccine and the pneumococcal conjugate vaccine (PCV) is affected by the order in which they are given.	Pain: procedural (immunisation). Setting: paediatric community practice (Canada).	DPTa-HiB group: received DTPa-HiB first followed by PCV PCV group: received PCV first followed by DPTa-HiB. Pain scoring: MBPS, VASobs (parent and paediatrician).	when PCV was administered first (MBPS score, 7.6 [1.5] vs 8.2 [1.5], P=.037; parent VAS score, 4.2 [2.3] vs 5.6 [2.6], P=.003).	
Ipp et al, 2007 (373)	Randomised controlled trial. To compare acute pain response during immunisation in infants using a slow standard of care injection technique versus a rapid pragmatic technique.	113 infants aged 4 to 6 months. Pain: procedural (immunisation). Setting: primary care practice (Canada).	2 groups – participants randomised to: Standard group: slow aspiration, injection and withdrawal Pragmatic group: no aspiration, rapid injection and withdrawal. Pain scoring: MBPS, VASob (parent and paediatrician), and cry duration.	MBPS scores lower for pragmatic group (3.3 95% CI 2.6 to 3.9 vs 5.6, 95% CI 5 to 6.3, p<0.001). VAS scores by parents 1.9 (0.1–3.1) vs 3.5 (1.6–5.5) & paediatricians vs 1.4 (0.2–2.4) vs 2.8 (2.0–5.1) were lower for pragmatic group. The pragmatic group less likely to cry, 24/56 (43%) vs 47/57 (82%), to cry less, median 0 sec (IQR 0–11.30) vs 14.7 sec (8.7–35.6) & to take less time to have vaccine injected, median 0.9 s (IQR 0.8–1.1) vs 8.8 s (7.9–10.3), for all comparisons p<0.001.	Jadad score = 3.
Kass et al, 2001 (614)	Randomized placebo-controlled blinded clinical trial. To determine if a 50% dextrose solution would reduce the percentage of circumcision procedure time a neonate spent crying by 50%, compared with water, and whether it would be similar to a dorsal penile nerve block (DPNB).	71 (term) newborn infants. Pain: procedural (circumcision). Setting: inpatient nursery of a military community hospital (US).	3 groups – participants randomised to: Group D50: 2ml 50% glucose orally Group water: 2ml sterile water orally Group DPNB: dorsal penile nerve block. Pain scoring: crying time, change in heart rate from baseline, MBPS.	Mean heart rate lowest in DPNB group (133) compared with water (171) and D50 (180) groups (p = 0.005). Percentage increase in heart rate also lowest in DPNB group (p = 0.005). Percentage crying time lower in DPNB group (33%) compared with water (82.4) and D50 (82.3) groups (p = 0.001). MBPS scores lower in DPNB group compared with water and D50 groups (MPBS scores not reported, p <.001).	Jadad score = 1. Concern about ethics approval for the treatment arms.

Kassab et al, 2012 (615)	Randomised double blind controlled trial. To determine the effectiveness of 25% oral glucose solution in reducing immunisation pain in 2-month old infants.	120 infants aged 2 months Exc: history convulsion or progressive or unstable neurological disorder. Pain: procedural (immunisation). Setting: maternity and child health care centre (Jordan).	2 groups – participants randomised to Treatment group: 2ml 25% oral glucose Control group: 2ml sterile water orally. Pain scoring: MBPS, cry duration.	Infants in the intervention group showed lower pain scores than control group during (median 8, IQR = 1 versus 9, IQR = 1) and following (median 4, IQR – 1 and 6, IQR = 3) the procedure ($p < 0.001$), and spent less time crying up to 2 min after the procedure (mean difference 38 vs. 77.9 s).	Jadad score = 5.
Kaur et al, 2009 (616)	Randomised study. To assess the effect of feeding the infant on breast during injecting vaccine on perception of pain intensity among infants.	216 infants aged 2 to 4 months. Pain: procedural (immunisation). Setting: immunisation room of Advanced Pediatric Centre (India).	2 groups – participants randomised to: Experimental group: vaccination with breast feeding Control group: vaccination without breast feeding. Pain scoring: MBPS, crying.	Net mean pain scores (post – pre scores) lower for experimental group compared to control group (4.6 v 6.8, $p<0.01$). 54.5% experimental group had moderate pain (MBPS 4 – 6) and 55.7% control group had severe pain post immunisation. No p value reported. Experimental group cried less than the control group (49.3 v 87.4 s, $p<0.02$).	Jadad score = 0.
Lindh et al, 2003 (378)	Randomised double-blinded. To determine whether use of lidocaine–prilocaine 5% cream (EMLA) and oral glucose decreases pain associated with diphtheria–pertussis–tetanus (DPT) immunization in 3-month-old infants.	90 Infants aged 3 months. Pain: procedural (immunisation). Setting: outpatient paediatric practice (Sweden).	2 groups – participants randomised to: Treatment group: EMLA patch and glucose solution Placebo group: placebo patch and water. Pain scoring: MBPS, VAS, latency to cry and total crying time, heart rate (HR).	MBPS scores lower in treatment group compared with placebo group at 0 – 10 sec (5.5 ± 2.0 v 7.7 ± 1.7) and 11 – 20 sec (5.4 ± 2.4 v 6.8 ± 2.2). Difference in MBPS scores pre- and post-injection also lower in the treatment group compared with the placebo group ($p < 0.001$). No significance testing to compare changes in scores across phases. Parent and nurse VAS scores lower in treatment group ($p < 0.05$). More infants cried in the placebo group compared with the treatment group (44 v 32, $p = 0.001$), latency of first cry was shorter in the placebo group (3.8 ± 2.3 v 6.4 ± 3.2, $p < 0.001$). A biphasic transient heart rate response with a marked deceleration followed by an acceleration was seen more	Jadad score = 5. COSMIN – reliability ‘fair’.

				often in the placebo group than the treatment group (p = 0.03). Agreement btw raters using MBPS – baseline kappa = 1.0, post kappa = 0.5.	
McGowan et al, 2013 (383)	Randomised controlled trial. To compare pain response during routine immunisation of infants using simultaneous versus sequential administration techniques.	36 infants aged 2 to 6 months. Exc: known physical or psychological conditions, needle phobic parents. Pain: procedural (immunisation). Setting: ummunisation clinic (Wales).	2 groups – participants randomised to: Intervention group: simultaneous immunisations Control group: sequential immunisations. Pain scoring: MBPS, VASobs (parent).	Median change in MBPS scores (pre- post) less in simultaneous group at 15s (p = 0.05), greater in simultaneous group at 30s (p < 0.05), 45s (p = 0.01) and 120s (p = 0.02). Median change in VAS greater in sequential group (5.6cm) than in simultaneous group (4.7cm) – not significant (p = 0.06).	Jadad score = 3.
Mularoni et al, 2009 (617)	Randomised double blind 3 armed clinical trial. To determine whether a lidocaine-enhanced lubricant used topically & instilled into the urethra decreased infants' distress associated with catheterization.	45 infants aged 2 to 24 months. Exc: altered mental status. Pain: procedural (urethral catherisation). Setting: ED (US).	3 groups – participants randomised to: Controlled group: no urethral instillation Placebo group: lubricant jelly instillation Treatment group: lidocaine-enhanced lubricant jelly instillation Instilled 2min prior to procedure. Pain scoring: MBPS, MBPScry.	MBPS scores lower but not significant in lidocaine group at the time of catheterization (phase 3; P = 0.065) MBPScry lower during the catheterization (phase 3; P = 0.036) than infants who did not have a lubricant instilled into the urethra. MBPS scores were lower for all groups in the baseline phase than either the instillation (t44 = 3.53, P = 0.001) or catheterization phases (t42 = 3.14, P = .003).	Jadad score = 5. COSMIN – reliability 'poor' responsiveness 'fair'.
O'Brien et al, 2004 (618)	Double-blind, randomized, placebo-controlled trial. To assess the efficacy and safety of 4% amethocaine in reducing the pain of	120 infants aged 1 year. Pain: procedural (immunisation).	2 groups – participants randomised to: Control group – placebo topically Treatment group – amethocaine gel topically.	Mean difference in pre and post-injection MBPS scores was lower in the amethocaine group compared to the placebo group (1.51 vs 2.29, respectively; P = 0.029). Amethocaine group experienced more skin reactions than the placebo (P < 0.001).	Jadad score = 5.

	subcutaneous measles-mumps-rubella vaccination in 1-year-old infants.	Setting: paediatric outpatient clinic (Canada).	Pain scoring: MBPS.		
Pathak et al, 2007 (619)	Randomised controlled trial To study the effect of needle gauge on pain perception of pain intensity among infants receiving D.P.T vaccination	320 infants aged up to 24 weeks Pain: procedural (immunisation) Setting: tertiary hospital child care centre (India)	2 groups – participants randomised to Group 1: 25g needle Group 2: 23g needle Pain scoring: MBPS	Mean net MBPS scores higher in group 1 compared to groups 2 (6.6 ± 1.5 v 5.9 ± 1.3, t = 4.25, df=318, (p < 0.01).	Jadad score = 1
Ram et al, 2006 (620)	Random cross over design. To evaluate and compare the reaction of children who received local anaesthesia with lidocaine 2% with 1 : 100 000 epinephrine and articaine 4% with 1 : 200 000 epinephrine and to assess the time of the onset, efficacy, duration of numbness of the soft tissues, children's sensation after treatment to both anaesthetic solutions, as well as the occurrence of adverse events.	62 children aged 5 to 13 years. Pain: procedural (dental procedure). Setting: dental clinic (Israel).	2 groups – participants randomised to one of these groups first before crossing over: Group: lidocaine 2% with 1:100,000 epinephrine Group: articaine 4% with 1:200,000 epinephrine. Pain scoring: MBPS, Wong & Baker Faces Pain Scale, anaesthetic time.	Duration of numbness longer for articaine (3·43 ± 0·7 h) than for lidocaine (3·0 ± 0·8 h) (P = 0·003). No difference in MBPS scores or self-report during injection, between sessions or block techniques.	Jadad score = 0.
Sundar et al, 2016 (621)	Randomised controlled trial. To assess effect of live music therapy intervention on pain, distress, and physiological parameters of the parent holding the child during painful immunization procedures in children.	100 infants aged less than 18 months. Pain: procedural (immunisation). Setting: paediatric outpatient department (India).	2 groups – block randomisation to: Experimental group: music therapy and visual aids Control group: no intervention.	Improvement in all MBPS items in experimental group (p < 0.001). Non-significant improvement in pain and distress scores in experimental group. Mean (±SD) duration of crying was 25.02 (±13.98) seconds in the experiment group and 41.66 (±17.29) seconds in the control group P < .05 (t test 5.2923, P = .000000738).	Jaded score = 0.

			Pain scoring: MBPS & NRS pain and NRS distress (parent applied), cry duration		
Taddio et al, 1994 (603)	Randomised double-blind controlled trial. To determine whether use of lidocaine-prilocaine 5% cream (EMLA) decreases pain associated with diphtheria-pertussis-tetanus (DPT) vaccination in infants.	96 infants aged 4 to 6 months. Exc: analgesic use within 4 hours of immunisation. Pain: procedural (immunisation). Setting: paediatric outpatient clinic (Canada).	2 groups – participants randomised to: Treatment group – EMLA patch Control group – placebo patch. Pain scoring: MBPS, 100mm VASobs, crying.	Post vaccination scores (7 v 8, p=0.001) and the difference between the pre and post vaccination scores (5 v 6, p=0.001) were lower in the EMLA group than the placebo group. VASobs scores were lower in the EMLA group (26 v 48mm, p=0.002). Latency to cry longer in EMLA group (3.4 v 2.5s, p=0.0004) and cry duration shorter in EMLA group (33.2 v 35.4s, p=0.027). Correlation between MBPS and VASobs was 0.608 (p<0.001).	Jadad score = 3. COSMIN – hypothesis 'fair'.
Taddio et al, 1995 (11) NOTE: study appears in tables for both methods as RCT design used to test psychometric properties	Randomised double-blind controlled trial. To adapt a behavioural pain measure (CHEOPS) for use in infants and to establish the reliability and validity of the measure used to measure pain secondary to immunisations.	96 infants aged 4 – 6 months. Pain: procedural (immunisation). Setting: outpatient clinic (Canada).	2 groups: randomised to: Control: placebo cream topical Treatment; EMLA cream topical. Pain scoring: MBPS, VASobs.	EMLA group had lower mean MBPS scores (6.8, SD – 1.9) than placebo group (8, SD – 1.5), p<0.001.	Jadad score = 1.
Taddio et al, 2014 (622)	Partially blinded randomised controlled trial. To evaluate the analgesic effectiveness of clinician-led tactile stimulation in infants undergoing vaccination.	121 infants aged 1 to 12 months. Exc: neurological conditions, infants receiving analgesics or sedation. Pain: procedural (immunisation). Setting: private paediatric clinic (Canada).	2 groups – participants randomised to: Tactile stimulation group No tactile stimulation group Prior, during and after immunisation. Pain scoring: MBPS, NRS parent and paediatrician, cry time.	No difference between groups in post injection MBPS scores (7.2 ± 2.4 vs 7.6 ± 1.9, p = 0.245), cry duration for the first 30 s (12 ± 11 vs 16 s ± 12, p = 0.109), and cry duration for the first 120 s (30s ± 32 vs 35s ± 33, p = 0.397). Parent & paediatrician NRS scores lower in the tactile stimulation group post immunisation (paediatrician 3.9 ± 2.5 vs 5.1 ± 2.3, p = 0.004 and parent 4.3 ± 2.8 vs 5.9 ± 2.8, p = 0.003).	Jadad score = 3.

Taddio et al, 2015 (293)	Partially blinded longitudinal cluster randomized trial. To determine the impact of educating parents about pain in outpatient paediatric clinics on their use of pain treatments during routine infant vaccinations.	160 parent infant dyads, infants aged up to 24 weeks. Exc: infants born < 30 weeks gestation, with congenital anomalies or neurological conditions. Pain: procedural (immunisation). Setting: outpatient paediatric clinic (Canada).	Clinics randomised to: Intervention clinics – parental pain management education program Control clinics – traditional approach. Pain scoring: MBPS, 11 point NRS, crying.	At first injection: non-blinded observer NRS was lower for the intervention group (5.0 v 5.7, p<0.001). MBPS, parent NRS and cry did not differ significantly between groups. At second injection 2months later MBPS (7.8 v 8.3, p=0.002) and observer NRS scores (4.7 v 5.5, p=0.002) were lower in intervention group. No difference in parent NRS or cry.	Jadad score = 3.
Taddio et al, 2015 (467)	Randomised double blind controlled trial. To compare the analgesic effectiveness of rotavirus vaccine to sucrose solution for reducing pain from vaccine injections in infants.	120 infants aged 2 – 4 months. Exc: infants with impaired neurological development, history of seizures or opioids in last 24hours. Pain: procedural (immunisation). Setting: outpatient paediatric clinic (Canada).	2 groups – participants randomised to: Grp 1 - rotavirus vaccine 1.5 mL orally 2 min prior to vaccine injections, then sucrose 24% solution 2 mL orally 1 min following vaccine injections Grp 2 - sucrose orally 2 min prior to vaccine injections then rotavirus vaccine 1 min following vaccine injections. Pain scoring: MBPS, parent and clinician NRS, cry duration.	No difference in MBPS scores, NRS or cry duration between groups at baseline or during injections. Baseline MBPS scores 3.0 and 2.7, NRS parent 0.2 and 0.1 and NRS clinician 0.3 and 0.1.	Jadad score = 5.

Abbreviations: CAS – Colour Analogue Scale, ED – emergency department, FAS - Faces Analogue Scale, FLACC – Face, Legs, Activity, Consolability, Cry, MBPS – Modified Behavioral Pain Scale, NRS – numeric rating scale, PACU – postoperative acute care unit, RCT – randomised controlled trial, SD - Standard Deviation, US – United States, VAS – Visual Analogue Scale, VASobs – VAS observer

APPENDIX F

ETHICS APPROVAL & GOVERNANCE AUTHORISATION

20 October 2015

Ms Dianne Crellin
Emergency
The Royal Children's Hospital

Dear Ms Crellin

Project Title: Procedural pain: Scale evaluation (PROPoSE) study - An evaluation of the psychometric properties of behavioural pain scales for assessment of procedural pain in infants and children aged six to 42 months

---

RCH HREC Reference Number: 35220 A

I am pleased to advise that the above project has received ethical approval from The Royal Children's Hospital Melbourne Human Research Ethics Committee (HREC). The project has also received governance authorisation at the Melbourne Children's Campus (incorporating The Royal Children's Hospital, Murdoch Children's Research Institute and the University of Melbourne Department of Paediatrics).

The Royal Children's Hospital Melbourne HREC is organised and operates in accordance with the National Health and Medical Research Council's (NHRMC) National Statement on Ethical Conduct in Research Involving Humans (2007) and all subsequent updates, and in accordance with the Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95), the Health Privacy Principles described in the Health Records Act 2001 (Vic) and Section 95A of the Privacy Act 1988 (and subsequent Guidelines).

HREC Approval Date: 20 October 2015
Please note the HREC are no longer issuing pre-determined approval periods. Ethical approval is now ongoing, subject to the submission of an annual report on the anniversary of approval.

Participating Sites:
Ethical approval for this project applies at the following sites:

Site Name
• Melbourne Children's Campus (incorporating The Royal Children's Hospital, Murdoch Children's Research Institute and the University of Melbourne Department of Paediatrics)

Approved Documents:
The following documents have been reviewed and approved:

Document	Version	Date
Protocol	1.1	12 Oct 2015
PIS & CF Psychologist	1.1	12 Oct 2015
PIS & CF Clinician	1.1	12 Oct 2015
Invite Psychologist	1.1	12 Oct 2015
Invite Clinician	1.1	12 Oct 2015
Advertising	1	5 Oct 2015

Conditions of Ethics Approval:
- You are required to submit to the HREC:

    - An Annual Progress Report (that covers all sites listed on approval) for the duration of the project. This report is due on the anniversary of HREC approval. Continuation of ethics approval is contingent on submission of an annual report, due within one month of the approval anniversary. Failure to comply with this requirement may result in suspension of the project by the HREC.
    - A comprehensive Final Report upon completion of the project.
- Submit to the reviewing HREC for approval any proposed amendments to the project including any proposed changes to the Protocol, Participant Information and Consent Form/s and the Investigator Brochure.
- Notify the reviewing HREC of any adverse events that have a material impact on the conduct of the research in accordance with the NHMRC Position Statement: *Monitoring and reporting of safety for clinical trials involving therapeutic products May 2009*.
- Notify the reviewing HREC of your inability to continue as Coordinating Principal Investigator.
- Notify the reviewing HREC of the failure to commence the study within 12 months of the HREC approval date or if a decision is taken to end the study at any of the sites prior to the expected date of completion.
- Notify the reviewing HREC of any matters which may impact the conduct of the project.
- If your project involves radiation, you are legally obliged to conduct your research in accordance with the Australian Radiation Protection and Nuclear Safety Agency Code of Practice 'Exposure of Humans to Ionizing Radiation for Research Purposes' Radiation Protection series Publication No.8 (May 2005)(ARPANSA Code).
- The HREC, authorising institution and/or their delegate/s may conduct an audit of the project at any time.

Yours sincerely

Erica Plummer
Senior Ethics Officer
The Royal Children's Hospital Melbourne
Phone : (03) 9345 5044
Email : rch.ethics@rch.org.au
Web : www.rch.org.au

ETHICS APPROVAL & GOVERNANCE AUTHORISATION

17 November 2015

Ms Dianne Crellin
Emergency
The Royal Children's Hospital

Dear Ms Crellin

Project Title: Procedural pain: Scale evaluation (PROPoSE) study - An evaluation of the psychometric properties of behavioural pain scales for assessment of procedural pain in infants and children aged six to 42 months

---

RCH HREC Reference Number: 35220B

I am pleased to advise that the below modification has received ethical approval from The Royal Children's Hospital Melbourne Human Research Ethics Committee (HREC). The modification has also received governance authorisation at the Melbourne Children's Campus (incorporating The Royal Children's Hospital, Murdoch Children's Research Institute and the University of Melbourne Department of Paediatrics).

The Royal Children's Hospital Melbourne HREC is organised and operates in accordance with the National Health and Medical Research Council's (NHRMC) National Statement on Ethical Conduct in Research Involving Humans (2007) and all subsequent updates, and in accordance with the Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95), the Health Privacy Principles described in the Health Records Act 2001 (Vic) and Section 95A of the Privacy Act 1988 (and subsequent Guidelines).

HREC Approval Date: 17 November 2015

Participating Sites:
Ethical approval for this project applies at the following sites:

Site Name
• Melbourne Children's Campus (incorporating The Royal Children's Hospital, Murdoch Children's Research Institute and the University of Melbourne Department of Paediatrics)

Approved Documents:
The following documents have been reviewed and approved:

Document	Version	Date
Protocol	2	06 Nov 2015
PIS & CF Psychologist	2	06 Nov 2015
PIS & CF Clinician	2	06 Nov 2015

Conditions of Ethics Approval:

- You are required to submit to the HREC:
  - An Annual Progress Report (that covers all sites listed on approval) for the duration of the project. This report is due on the anniversary of HREC approval. Continuation of ethics approval is contingent on submission of an annual report, due within one month of the

approval anniversary. Failure to comply with this requirement may result in suspension of the project by the HREC.
    - A comprehensive Final Report upon completion of the project.
- Submit to the reviewing HREC for approval any proposed amendments to the project including any proposed changes to the Protocol, Participant Information and Consent Form/s and the Investigator Brochure.
- Notify the reviewing HREC of any adverse events that have a material impact on the conduct of the research in accordance with the NHMRC Position Statement: *Monitoring and reporting of safety for clinical trials involving therapeutic products May 2009*.
- Notify the reviewing HREC of your inability to continue as Coordinating Principal Investigator.
- Notify the reviewing HREC of the failure to commence the study within 12 months of the HREC approval date or if a decision is taken to end the study at any of the sites prior to the expected date of completion.
- Notify the reviewing HREC of any matters which may impact the conduct of the project.
- If your project involves radiation, you are legally obliged to conduct your research in accordance with the Australian Radiation Protection and Nuclear Safety Agency Code of Practice 'Exposure of Humans to Ionizing Radiation for Research Purposes' Radiation Protection series Publication No.8 (May 2005)(ARPANSA Code).
- The HREC, authorising institution and/or their delegate/s may conduct an audit of the project at any time.

Yours sincerely

Erica Plummer
Senior Ethics Officer
Research Ethics and Governance
Phone : (03) 9345 5044
Email : rch.ethics@rch.org.au
Web : www.rch.org.au

APPENDIX G

Declaration for a thesis with publication

Professor Nick Santamaria

Declaration for a thesis with publication

THE UNIVERSITY OF MELBOURNE

PhD and MPhil students may include a primary research publication in their thesis in lieu of a chapter if:

- The student contributed greater than 50% of the content in the publication and is the "primary author", ie. the student was responsible primarily for the planning, execution and preparation of the work for publication
- The student has approval to include the publication in their thesis from their Advisory Committee
- It is a primary publication that reports on original research conducted by the student during their enrolment
- The initial draft of the work was written by the student and any subsequent editing in response to co-authors and editors reviews was performed by the student
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Students must submit this form, along with *Co-author authorisation forms* completed by each co-author, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*. If you are including multiple publications in your thesis you will need to complete a separate form for each publication. Further information on this policy is available at: *gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication*

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	Systematic Review of the FLACC scale for assessing pain in infants and children: is it reliable, valid, and feasible for use?	
Authors	Crellin DJ, Harrison D, Santamaria N, Babl FE.	
Student's contribution (%)	90%.	
Journal or book name	Pain	
Volume/page numbers	156(11):2132-51	
Status	☐ Accepted and In press ☒ Published ☐ In progress	Date accepted/ published 2015

B. STUDENT'S DECLARATION

I declare that the publication above meets the requirements to be included in the thesis

Student's name	Student's signature	Date *(dd/mm/yy)*
Dianne Crellin		19 June 2018

C. PRINCIPAL SUPERVISOR'S DECLARATION

I declare that:

- the information above is accurate
- The advisory committee has met and agreed to the inclusion of this publication in the student's thesis
- All of the co-authors of the publication have reviewed the above information and have agreed to its veracity
- 'Co-Author Authorisation' forms for each co-author are attached.

Supervisor's name	Supervisor's signature	Date *(dd/mm/yy)*
Professor Nick Santamaria		23/6/18.

The University of Melbourne
CRICOS Provider Number: 00116K

Last Updated 19 October 2017

Declaration for a thesis with publication

THE UNIVERSITY OF MELBOURNE

PhD and MPhil students may include a primary research publication in their thesis in lieu of a chapter if:

- The student contributed greater than 50% of the content in the publication and is the "primary author", ie. the student was responsible primarily for the planning, execution and preparation of the work for publication
- The student has approval to include the publication in their thesis from their Advisory Committee
- It is a primary publication that reports on original research conducted by the student during their enrolment
- The initial draft of the work was written by the student and any subsequent editing in response to co-authors and editors reviews was performed by the student
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Students must submit this form, along with *Co-author authorisation forms* completed by each co-author, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*. If you are including multiple publications in your thesis you will need to complete a separate form for each publication. Further information on this policy is available at: *gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication*

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	A Systematic Review of the Psychometric Properties of the Modified Behavioral Pain Scale (MBPS).	
Authors	Crellin DJ, Babl FE, Santamaria N, Harrison D.	
Student's contribution (%)	90%	
Journal or book name	Journal of Pediatric Nursing	
Volume/page numbers	40:14-26.	
Status	☐ Accepted and In press ☒ Published ☐ In progress	Date accepted/ published 2018

B. STUDENT'S DECLARATION

I declare that the publication above meets the requirements to be included in the thesis

Student's name	Student's signature	Date *(dd/mm/yy)*
Dianne Crellin		19 June 2018

C. PRINCIPAL SUPERVISOR'S DECLARATION

I declare that:

- the information above is accurate
- The advisory committee has met and agreed to the inclusion of this publication in the student's thesis
- All of the co-authors of the publication have reviewed the above information and have agreed to its veracity
- 'Co-Author Authorisation' forms for each co-author are attached.

Supervisor's name	Supervisor's signature	Date *(dd/mm/yy)*
Professor Nick Santamaria		23/6/18.

Declaration for a thesis with publication

THE UNIVERSITY OF MELBOURNE

PhD and MPhil students may include a primary research publication in their thesis in lieu of a chapter if:

- The student contributed greater than 50% of the content in the publication and is the "primary author", ie. the student was responsible primarily for the planning, execution and preparation of the work for publication
- The student has approval to include the publication in their thesis from their Advisory Committee
- It is a primary publication that reports on original research conducted by the student during their enrolment
- The initial draft of the work was written by the student and any subsequent editing in response to co-authors and editors reviews was performed by the student
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Students must submit this form, along with *Co-author authorisation forms* completed by each co-author, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*. If you are including multiple publications in your thesis you will need to complete a separate form for each publication. Further information on this policy is available at: *gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication*

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children.	
Authors	Crellin DJ, Harrison D, Hutchinson A, Schuster T, Santamaria N, Babl FE.	
Student's contribution (%)	90%	
Journal or book name	BMJ Open	
Volume/page numbers	7(9).	
Status	☐ Accepted and In press ☒ Published ☐ In progress	Date accepted/ published 2017

B. STUDENT'S DECLARATION

I declare that the publication above meets the requirements to be included in the thesis

Student's name	Student's signature	Date *(dd/mm/yy)*
Dianne Crellin		19 June 2018

C. PRINCIPAL SUPERVISOR'S DECLARATION

I declare that:

- the information above is accurate
- The advisory committee has met and agreed to the inclusion of this publication in the student's thesis
- All of the co-authors of the publication have reviewed the above information and have agreed to its veracity
- 'Co-Author Authorisation' forms for each co-author are attached.

Supervisor's name	Supervisor's signature	Date *(dd/mm/yy)*
Professor Nick Santamaria		28/6/18.

The University of Melbourne
CRICOS Provider Number: 00116K

Last Updated 19 October 2017

Declaration for a thesis with publication

THE UNIVERSITY OF MELBOURNE

PhD and MPhil students may include a primary research publication in their thesis in lieu of a chapter if:

- The student contributed greater than 50% of the content in the publication and is the "primary author", ie. the student was responsible primarily for the planning, execution and preparation of the work for publication
- The student has approval to include the publication in their thesis from their Advisory Committee
- It is a primary publication that reports on original research conducted by the student during their enrolment
- The initial draft of the work was written by the student and any subsequent editing in response to co-authors and editors reviews was performed by the student
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Students must submit this form, along with *Co-author authorisation forms* completed by each co-author, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*. If you are including multiple publications in your thesis you will need to complete a separate form for each publication. Further information on this policy is available at: *gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication*

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	The psychometric properties of the FLACC scale used to assess procedural pain.	
Authors	Crellin D, Harrison D, Santamaria N, Huque, H and Babl, F E.	
Student's contribution (%)	90%	
Journal or book name	Journal of Pain	
Volume/page numbers	in press	
Status	[x] Accepted and In press [ ] Published [ ] In progress	Date accepted/ published: January 2018

B. STUDENT'S DECLARATION

I declare that the publication above meets the requirements to be included in the thesis

Student's name	Student's signature	Date *(dd/mm/yy)*
Dianne Crellin		19 June 2018

C. PRINCIPAL SUPERVISOR'S DECLARATION

I declare that:

- the information above is accurate
- The advisory committee has met and agreed to the inclusion of this publication in the student's thesis
- All of the co-authors of the publication have reviewed the above information and have agreed to its veracity
- 'Co-Author Authorisation' forms for each co-author are attached.

Supervisor's name	Supervisor's signature	Date *(dd/mm/yy)*
Professor Nick Santamaria		23/6/18

The University of Melbourne
CRICOS Provider Number: 00116K

Last Updated 19 October 2017

THE UNIVERSITY OF MELBOURNE

Declaration for a thesis with publication

PhD and MPhil students may include a primary research publication in their thesis in lieu of a chapter if:

- The student contributed greater than 50% of the content in the publication and is the "primary author", ie. the student was responsible primarily for the planning, execution and preparation of the work for publication
- The student has approval to include the publication in their thesis from their Advisory Committee
- It is a primary publication that reports on original research conducted by the student during their enrolment
- The initial draft of the work was written by the student and any subsequent editing in response to co-authors and editors reviews was performed by the student
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Students must submit this form, along with *Co-author authorisation forms* completed by each co-author, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*. If you are including multiple publications in your thesis you will need to complete a separate form for each publication. Further information on this policy is available at: *gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication*

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	The Psychometric Properties of the MBPS Scale Used to Assess Procedural Pain.	
Authors	Crellin DJ, Babl FE, Santamaria N, Harrison D.	
Student's contribution (%)	90 %.	
Journal or book name	Journal of Pain	
Volume/page numbers	19(6):660-70	
Status	☐ Accepted and In press ☒ Published ☐ In progress	Date accepted/ published: Published June 2018

B. STUDENT'S DECLARATION

I declare that the publication above meets the requirements to be included in the thesis

Student's name	Student's signature	Date *(dd/mm/yy)*
Dianne Crellin		19 June 2018

C. PRINCIPAL SUPERVISOR'S DECLARATION

I declare that:

- the information above is accurate
- The advisory committee has met and agreed to the inclusion of this publication in the student's thesis
- All of the co-authors of the publication have reviewed the above information and have agreed to its veracity
- 'Co-Author Authorisation' forms for each co-author are attached.

Supervisor's name	Supervisor's signature	Date *(dd/mm/yy)*
Professor Nick Santamaria		23/6/18.

The University of Melbourne
CRICOS Provider Number: 00116K

Last Updated 19 October 2017

Co-author authorisation form

Associate Professor Franz Babl

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	Systematic Review of the FLACC scale for assessing pain in infants and children: is it reliable, valid, and feasible for use?	
Authors	Crellin D, Harrison D, Santamaria and Babl, F E.	
Student's contribution (%)	90%.	
Journal or book name	Pain	
Volume/page numbers	156(11):2132-51	
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published 2015

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Associate Professor Franz Babl		20 June 2018

The University of Melbourne
CRICOS Provider Number: 00116K

Last Updated 19 October 2017

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	A Systematic Review of the Psychometric Properties of the Modified Behavioral Pain Scale (MBPS).	
Authors	Crellin DJ, Babl FE, Santamaria N, Harrison D.	
Student's contribution (%)	90 /.	
Journal or book name	Journal of Pediatric Nursing	
Volume/page numbers	40:14-26	
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published January 2018

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Associate Professor Franz Babl		

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months.		
Authors	Crellin DJ, Harrison D, Hutchinson A, Schuster T, Santamaria N, Babl FE.		
Student's contribution (%)	90 %.		
Journal or book name	BMJ Open		
Volume/page numbers	7(9).		
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published	2017

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Associate Professor Franz Babl		20 June 2018

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	The Psychometric Properties of the FLACC Scale Used to Assess Procedural Pain.	
Authors	Crellin D, Harrison D, Santamaria N, Huque, H and Babl, F E.	
Student's contribution (%)	90 %	
Journal or book name	Journal of Pain	
Volume/page numbers	in press	
Status	[x] Accepted and In-press [ ] Published [ ] In progress	Date accepted/published January 2018

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Associate Professor Franz Babl		20 June 2018

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	The Psychometric Properties of the MBPS Scale Used to Assess Procedural Pain.	
Authors	Crellin DJ, Babl FE, Santamaria N, Harrison D.	
Student's contribution (%)	90 %.	
Journal or book name	Journal of Pain	
Volume/page numbers	19(5):660-70	
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published: June 2018

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Associate Professor Franz Babl		20 June 2018

The University of Melbourne
CRICOS Provider Number: 00116K

Last Updated 19 October 2017

Professor Denise Harrison

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS ***(to be completed by the student)***

Full title	Systematic Review of the FLACC scale for assessing pain in infants and children: is it reliable, valid, and feasible for use?	
Authors	Crellin D, Harrison D, Santamaria and Babl, F E.	
Student's contribution (%)	90 %	
Journal or book name	Pain	
Volume/page numbers	156(11):2132-51	
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published 2015

B. CO-AUTHOR'S DECLARATION ***(to be completed by the collaborator)***

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Professor Denise Harrison	Denise M Harrison	20 June 2018

The University of Melbourne
CRICOS Provider Number: 00116K

Last Updated 19 October 2017

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	A Systematic Review of the Psychometric Properties of the Modified Behavioral Pain Scale (MBPS).	
Authors	Crellin DJ, Babl FE, Santamaria N, Harrison D.	
Student's contribution (%)	90 %	
Journal or book name	Journal of Pediatric Nursing	
Volume/page numbers	40:14-26	
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published January 2018

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Professor Denise Harrison	Denise M Harrison	

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months.	
Authors	Crellin DJ, Harrison D, Hutchinson A, Schuster T, Santamaria N, Babl FE.	
Student's contribution (%)	85%	
Journal or book name	BMJ Open	
Volume/page numbers	7(9).	
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published 2017

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Professor Denise Harrison	Denise M Harrison	20 June 2018

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS ***(to be completed by the student)***

Full title	The Psychometric Properties of the FLACC Scale Used to Assess Procedural Pain.	
Authors	Crellin D, Harrison D, Santamaria N, Huque, H and Babl, F E.	
Student's contribution (%)	90 %	
Journal or book name	Journal of Pain	
Volume/page numbers	in press	
Status	[X] Accepted and In-press [ ] Published [ ] In progress	Date accepted/published January 2018

B. CO-AUTHOR'S DECLARATION ***(to be completed by the collaborator)***

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Professor Denise Harrison	Denise M Harrison	20 June 2018

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS ***(to be completed by the student)***

Full title	The Psychometric Properties of the MBPS Scale Used to Assess Procedural Pain.	
Authors	Crellin DJ, Babl FE, Santamaria N, Harrison D.	
Student's contribution (%)	90 %	
Journal or book name	Journal of Pain	
Volume/page numbers	19(6):660-70	
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published June 2018

B. CO-AUTHOR'S DECLARATION ***(to be completed by the collaborator)***

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Professor Denise Harrison	Denise M Harrison	20 June 2018

Mr Adrian Hutchinson

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	Procedural Pain Scale Evaluation (PROPoSE) study: protocol for an evaluation of the psychometric properties of behavioural pain scales for the assessment of procedural pain in infants and children aged 6–42 months.	
Authors	Crellin DJ, Harrison D, Hutchinson A, Schuster T, Santamaria N, Babl FE.	
Student's contribution (%)	90%.	
Journal or book name	BMJ Open	
Volume/page numbers	7(9).	
Status	☐ Accepted and In-press ☒ Published ☐ In progress	Date accepted/published 2017

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Mr Adrian Hutchinson		20 June 2018

The University of Melbourne
CRICOS Provider Number: 00116K

Last Updated 19 October 2017

Dr Hamidul Huque

Co-author authorisation form

THE UNIVERSITY OF MELBOURNE

All co-authors must complete this form. By signing below co-authors agree to the listed publication being included in the student's thesis and that the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

In cases where all members of a large consortium are listed as authors of a publication, only those that actively collaborated with the student on material contained within the thesis should complete this form. This form is to be used in conjunction with the *Declaration for a thesis with publication form*.

Students must submit this form, along with the *Declaration for thesis with publication form*, when the thesis is submitted to the Thesis Examination System: *https://tes.app.unimelb.edu.au/*

Further information on this policy and the requirements is available at:
gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

A. PUBLICATION DETAILS *(to be completed by the student)*

Full title	The Psychometric Properties of the FLACC Scale Used to Assess Procedural Pain.	
Authors	Crellin D, Harrison D, Santamaria N, Huque, H and Babl, F E.	
Student's contribution (%)	90%	
Journal or book name	Journal of Pain	
Volume/page numbers	in press	
Status	[x] Accepted and In-press [ ] Published [ ] In progress	Date accepted/published January 2018

B. CO-AUTHOR'S DECLARATION *(to be completed by the collaborator)*

I authorise the inclusion of this publication in the student's thesis and certify that:

- the declaration made by the student on the *Declaration for a thesis with publication form* correctly reflects the extent of the student's contribution to this work;
- the student contributed greater than 50% of the content of the publication and is the "primary author" ie. the student was responsible primarily for the planning, execution and preparation of the work for publication.

Co-author's name	Co-author's signature	Date *(dd/mm/yy)*
Dr Hamidul Huque	Hamidul Huque	20 June 2018