

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



CrossMark

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 intrarater 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, Consolability and Cry Scale

Categories	Scoring		
	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)

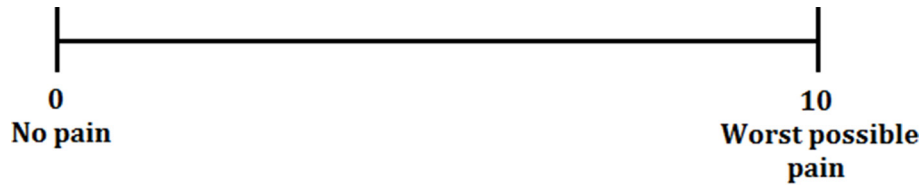
VAS is a tool designed to measure and quantify subject experiences such as pain and distress.³¹ The scale is a 10 cm line anchored at either end with labels such as 'no pain' and 'worst possible pain' or 'no distress' and 'worst possible distress'. When applied by an observer, they are asked to estimate 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	2	3	4	5
1. Provides information that is clinically useful	1 <input type="checkbox"/> Clinically not very useful	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/> Clinically very useful	
2. Is it clear and easy to understand	1 <input type="checkbox"/> Not clear and easy	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/> Clear and easy	
3. Is quick to apply	1 <input type="checkbox"/> Very slow	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/> Very quick	
4. Is easy to apply	1 <input type="checkbox"/> Very difficult	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/> Very easy	
5. Reflects the extent of procedural pain	1 <input type="checkbox"/> Does not reflect at all	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/> Reflects the extent well	
6. Discriminates children with pain from children without pain	1 <input type="checkbox"/> Does not discriminate at all	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/> Discriminates well	
7. Score is readily understood and supports decisions about pain management	1 <input type="checkbox"/> Not readily understood and doesn't support	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/> Readily understood and supports decisions decisions	
8. Reflects procedural pain-specific features	1 <input type="checkbox"/> Doesn't reflect procedural pain related features	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/> Reflects procedural pain related features	

Procedural pain: Scale evaluation (PROPose) study

You now have the opportunity to review the segment of video as many times as you like before entering your final score.

Your final score should be the score that you are confident wouldn't change no matter how many times you view the video. This score may be the same as your first score or it may be different.

MBPS

Face

0 1 2 3 NA Other...

Cry

0 1 2 3 4 NA Other...

Movement

0 2 3 NA Other...

If you are unable to score an item for the following reasons

- restraint interfering
- procedure interfering
- comfort measures

Select NA and then the reason.

If there is another reason you cant score select other and



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 O₂ 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:

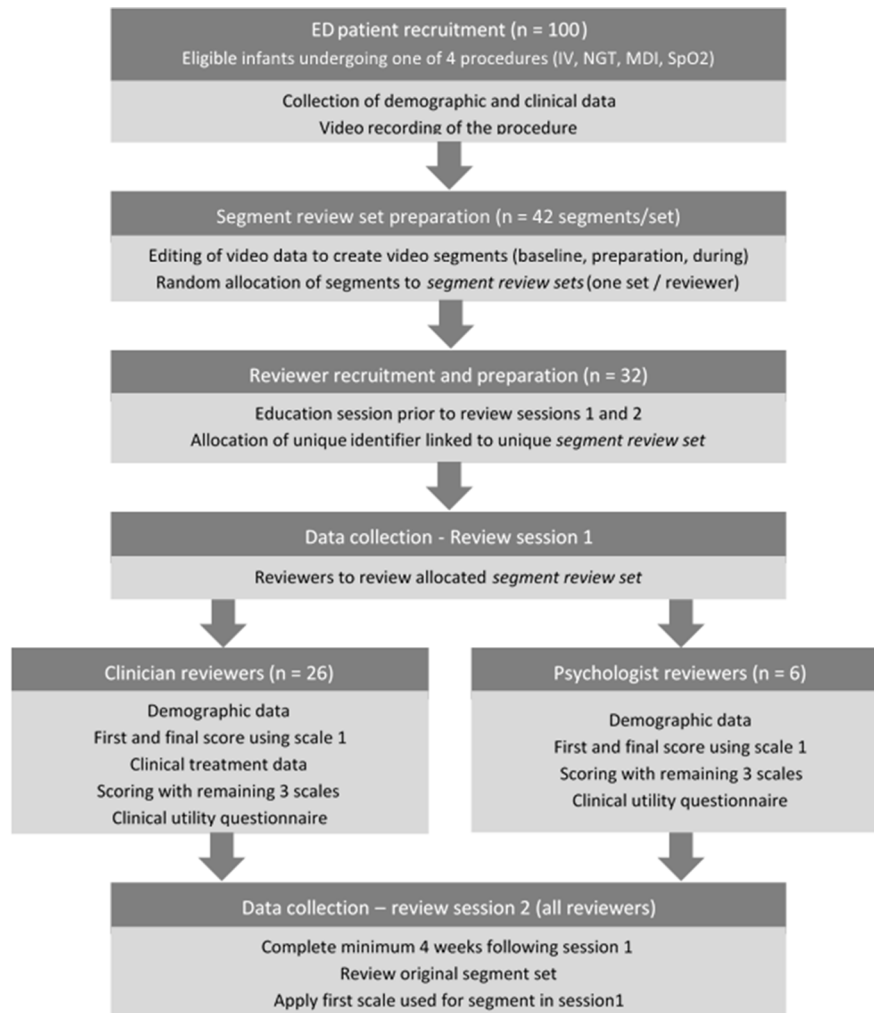


Figure 3 Study procedure. ED, emergency department; IMD, Inhaled medication delivery; IV, intravenous; NGT, nasogastric tube insertion; SpO₂, oxygen saturation measurement.

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 5s 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 15s 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 Stata³⁴ 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.

CI's 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

¹Department of Nursing, The University of Melbourne, Melbourne, Australia

²Clinical Sciences, Murdoch Children's Research Institute (MCRI), Melbourne, Australia

³Royal Children's Hospital, Melbourne, Australia

⁴Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa, Canada

⁵Clinical Epidemiology and Biostatistics Unit, MCRI, Melbourne, Australia

⁶Department 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, development 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 Biostatistics 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.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Sinatra R. Causes and consequences of inadequate management of acute pain. *Pain Med* 2010;11:1859–71.
- 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.
- 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.
- 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.
- 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.
- Fry M, Bennetts S, Huckson S. An Australian audit of ED pain management patterns. *J Emerg Nurs* 2011;37:269–74.
- 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.
- 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.
- 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: 2016/12/03).
- 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.
- 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.
- 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).
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Taddio A, Nulman I, Koren BS, et al. A revised measure of acute pain in infants. *J Pain Symptom Manage* 1995;10:456–63.
- 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.
- Crellin, Dianne J, Babl, Franz E, Santamaria, Nick, Harrison, Denise. The MBPS: a systematic review to determine its role in assessing pain in infants and children. International Symposium on Pediatric Pain; 2017; Kuala Lumpur.
- Celeste Johnston C, von Baeyer CL. A measure of pediatric pain intensity across ages and clinical conditions. *Pain* 2012;153:1545–6.
- 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.
- 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.
- Vazirani J, Knott JC. Mandatory pain scoring at triage reduces time to analgesia. *Ann Emerg Med* 2012;59:134–8.
- 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.
- 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.
- 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.
- 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.
- 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.