Pain 125 (2006) 143–157

PAIN

www.elsevier.com/locate/pain

Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents ☆

Jennifer N. Stinson [a,b,*], Tricia Kavanagh [a], Janet Yamada [a,c], Navreet Gill [a], Bonnie Stevens [a,c]

[a] *Faculty of Nursing, University of Toronto, Hospital for Sick Children, Toronto, Ont., Canada*
[b] *Department of Anaesthesia, Hospital for Sick Children, Toronto, Ont., Canada*
[c] *Department of Nursing, Hospital for Sick Children, Toronto, Ont., Canada*

Received 12 September 2005; received in revised form 19 April 2006; accepted 3 May 2006

Abstract

The aim of this study was to systematically review the psychometric properties, interpretability and feasibility of self-report pain intensity measures for children and adolescents for use in clinical trials evaluating pain treatments. Databases were searched for self-report measures of single-item ratings of pain intensity for children aged 3–18 years. A total of 34 single-item self-report measures were found. The measures' psychometric properties, interpretability and feasibility, were evaluated independently by two investigators according to a set of psychometric criteria. Six single-item measures met the a priori criteria and were included in the final analysis. While these six scales were determined as psychometrically sound and show evidence of responsivity, they had varying degrees of interpretability and feasibility. No single scale was found to be optimal for use with all types of pain or across the developmental age span. Specific recommendations regarding the most psychometrically sound and feasible measures based on age/developmental level and type of pain are discussed. Future research is needed to strengthen the measurement of pain in clinical trials with children.

Keywords: Systematic review; Pain intensity; Self-report; Children; Adolescents

1. Introduction

Clinical trials establishing the efficacy and effectiveness of pain treatment are essential for determining pain relieving interventions in children and adolescents. While multiple pediatric pain measures exist, they are inconsistently used across trials making comparison of results difficult. To address this problem, the Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (Ped-IMMPACT) has recommended core outcome domains which should be considered when designing pain clinical trials for acute and recurrent/chronic pain. The six core outcome domains for acute pain are: (1) pain, (2) global judgment of satisfaction with treatment, (3) symptoms and adverse events, (4) physical recovery, (5)

☆ Jennifer Stinson's work is supported by a Canadian Nurses Foundation/Hospital for Sick Children/Canadian Institutes of Health Research (CIHR) Doctoral Fellowship, Hospital for Sick Children Clinician Scientist Training Fellowship and the Premier's Research Excellence Award. Tricia Kavanagh's work is supported by a CIHR Doctoral Fellowship and a University of Toronto Centre for the Study of Pain Clinician Scientist Training Fellowship. Navreet Gill's work was supported by a CIHR Summer Studentship in Musculoskeletal Research. Dr. Stevens holds the Signy Hildur Eaton Chair in Paediatric Nursing Research at the Hospital for Sick Children and her research is supported through CIHR.

* Corresponding author. Tel.: +1 416 813 6243; fax: +1 416 813 7543.
E-mail address: jennifer.stinson@sickkids.ca (J.N. Stinson).

0304-3959/$32.00
doi:10.1016/j.pain.2006.05.006

emotional recovery, (6) and economic factors. The eight core outcomes for recurrent/chronic pain include: (1) pain, (2) physical functioning, (3) emotional functioning, (4) role functioning, (5) symptoms and adverse events, (6) global judgment of satisfaction, (7) sleep, (8) and economic factors (McGrath et al., 2005).

In March 2005, the Ped-IMMPACT group commissioned two systematic reviews of pain measures for children between the ages of 3 and 18 years for use in clinical pain trials. The first was a review of observational behavioral pain measures conducted by von Baeyer and Spagrud (2005). The second was a systematic review of self-report pain measures which will be the focus of this paper. Self-report and observational pain measures used in such trials must have well-established reliability, validity, responsivity, and interpretability as well as being feasible in terms of ease of use and scoring.

Three approaches to measuring pain in children include: self-report, observational or behavioral and physiological. The ideal would be a composite measure including self-report and one or more of these other approaches (Champion et al., 1998). However, this approach in clinical trials is not always practical or feasible. This approach would also not be applicable for preverbal children and nonverbal and cognitively impaired children for whom behavioral observation should be the primary source for pain measurement (von Baeyer and Spagrud, 2005). Furthermore, despite the recognition of the multidimensional nature of pain, self-reported pain intensity is the most commonly used approach in pediatric clinical trials. Self-report measures can generally be used with children who are old enough to understand and use self-report scales, are not overly distressed, who do not have impaired cognitive or communicative abilities, and whose self-report ratings are not considered exaggerated or minimized due to cognitive, emotional or situational factors (von Baeyer and Spagrud, 2005).

The objective of this study was to systematically review the research literature to evaluate the sensory intensity component of pain (pain intensity) in terms of the psychometric properties, interpretability and feasibility of self-report measures for children and adolescents. This evidence was used to make recommendations on pain intensity measures to be used in the assessment of pediatric acute, recurrent and chronic pain treatment as a core outcome in clinical trials as well as for determining areas for future research.

2. Methods

2.1. Criteria for considering studies for this review

The inclusion and exclusion criteria are outlined below in terms of types of studies, and pain measures that were included in this review. In addition, measures had to meet an a priori set of minimal criterion of psychometric evidence (i.e., reliability, validity, and responsiveness) to be classified as a well-established pain intensity measure in order to be included in this review (see Table 1). These criterion and their operational definitions are outlined in Table 2.

2.1.1. Types of studies

All published peer-reviewed English language research studies examining the psychometric properties, interpretability and feasibility of self-reported measures of pain intensity in children and adolescents were considered for inclusion in the review. Unpublished manuscripts, reviews, guidelines, commentaries and other descriptive articles were excluded. Published abstracts were also not included as the information provided in abstracts is limited and frequently non-peer reviewed. Studies published in languages other than English were also excluded due to time and financial constraints (translation costs). The exclusion of non-English language research studies was minimal (four articles including one new measure) (Jakobs and Rister, 1997; Vihunen and Sihvonen, 1998; Rossato and Angelo, 1999; Suraseranivongse et al., 2005).

2.1.2. Types of pain measures

Single-item ratings of pain intensity are the most commonly used measures in pain research and clinical settings in children 3 years of age and older who have the verbal capabilities to self-report. Behavioral observational pain measures were not included as part of this review as von Baeyer and Spagrud (2005) have recently completed this review (see http://www.immpact.org/). Composite measures that combined pain intensity, location and quality of pain were included in the review but will not be reported on here (for details on multivariate measures see http://www.immpact.org/). Furthermore, dimensions of pain relief, temporal aspects of pain and pain interference with aspects of emotional, social and physical function were not covered in this review as they were addressed by the larger Ped-IMMPACT group (McGrath et al., 2005). Measures of pain intensity and unpleasantness were also reviewed but were not included as they did not meet the minimal psychometric criteria for what constituted a good measure for use in clinical trials. Finally, measures not encompassed in this review include those developed for specific pain diagnoses (headaches, recurrent abdominal pain) or illness conditions (e.g., sickle cell disease) because these measures have generally not been validated for use across pain conditions. Given that these pain conditions are quite prevalent, these specific measures should undergo a similar review to determine the best measures to use in clinical trials.

2.2. Search strategy for identification of studies

Included studies were accessed primarily through a search of Medline (1966 to June week1, 2005), CINAHL (1982 to June week 1, 2005), EMBASE (1980 to June week 1, 2005), and Pubmed (1950's to June week 1, 2005). The search terms included MeSH headings, subjects, text words, wild cards and/or keywords relevant to the following terms: 'pain measurement', 'child', 'children' and 'pediatric' as well as specific names of pain measures and categories of measures (e.g., verbal rating scales, numerical rating scales). Lastly, only human and English language studies were selected. The date

Table 1
Categories for classification of instruments

	Criteria for categories
I. A well-established assessment	a.The measure must have been presented in at least 2 peer-reviewed articles by different investigators or investigatory teams b. Sufficient detail about the measure to allow critical evaluation and replication c. Detailed information indicating good validity and reliability in at least 1 peer-reviewed article
II. Approaching well-established assessment	a. The measure must have been presented in at least 2 peer-reviewed articles, which might be by the same investigator or investigatory team b. Sufficient detail about the measure to allow critical evaluation and replication c. Validity and reliability information either presented in vague terms (e.g., no statistics presented) or only moderate values presented
III. Promising assessment	a. The measure must have been presented in at least 1 peer-reviewed article b. Sufficient detail about the measure to allow critical evaluation and replication c. Validity and reliability information either presented in vague terms or moderate values presented

Cohen et al. (in press).

of the last search attempt was June 2005. Reference lists from all identified appropriate papers and review papers were examined followed by a hand search for other identified studies. No attempt was made to locate unpublished material or to contact authors of unpublished studies. Two reviewers (N.G. and T.K.) independently conducted the literature search.

2.3. Methods of the review

One reviewer (J.S.) screened all identified titles and abstracts for relevance. Two reviewers (J.S. and J.Y.) assessed potentially relevant studies for inclusion independently. Any disagreements were resolved by discussion including the third reviewer (B.S.). A systematic approach to data extraction was used to produce a descriptive summary of the psychometric findings, interpretability and feasibility for each self-reported measure. A second reviewer checked the data extraction. Outcome data extraction focused on the measure's age range, type of pain, reliability, validity, responsivity, interpretability, and feasibility (see Table 2 for operational definitions of these terms). A systematic approach was used by two independent reviewers (J.S. and J.Y) to rate the quality or overall level of evidence supporting each measure (see Table 1). There was 100% agreement on the rating of evidence by the two reviewers. To be included in this review the measure had to meet the a priori criteria of being a well-established measure (i.e., sound evidence of reliability, validity, responsivity, interpretability and feasibility). A total of 6 single-item pain intensity measures met the inclusion criteria for the review. Reasons for excluding the remaining 28 measures are summarized in Table 3.

3. Results

A brief description of each pain intensity measure is included in the review, as well the advantages and disadvantages of each measure and reliability and validity data are reported. An overall summary of the measures is provided in Table 4. With the exception of the Pieces of Hurt tool (Hester, 1979), which has been tested on children with acute pain only, all of the reviewed measures have psychometric evidence for both acute and chronic pain.

3.1. Pieces of hurt tool

3.1.1. Description

The Pieces of Hurt tool (Hester, 1979) is a concrete ordinal graphic rating scale (Hester, 1979). This tool consists of four red plastic poker chips, representing 'a little hurt' to 'the most hurt you could ever have'. The child is asked to select the chip that represents his/her pain intensity and the tool is scored from 0 to 4.

3.1.2. Reliability and validity data

The Pieces of Hurt tool has undergone extensive psychometric testing by various teams of investigators. This tool has evidence of test–retest reliability over 1-day (Wong and Baker, 1988) and 8-h (Gharaibeh and Abu-Saad, 2002) periods. The Pieces of Hurt tool has strong evidence of convergent validity ($r = .74$–$.98$) with other well-established pain intensity measures, including the Oucher (Beyer and Aradine, 1987, 1988), visual analogue scale (Hester et al., 1990), Wong–Baker FACES Pain Scale (Gharaibeh and Abu-Saad, 2002), Faces Pain Scale (Goodenough et al., 2005), and verbal responses and vocal behaviors from an observational behavioral measure (Hester, 1979). Conversely, two studies found that the Pieces of Hurt tool produced higher average pain scores compared to the Faces Pain Scale and a behavioral checklist (scoring observable facial, motor, verbal and vocal reactions) (Goodenough et al., 1997, 2005). Also, Goodenough et al. (2005) found that the Pieces of Hurt tool demonstrated greater upper-end bias in response distribution compared to the Faces Pain Scale and several other less well-validated pain measures especially in younger children aged 4–7 years. This tool

Table 2
Operational definition of terms

Term	Operational definition
Reliability	The reproducibility of a measure over different occasions and is concerned with minimizing sources of random error so that measures are reproducible (Streiner and Norman, 2005). In general, acceptable reliability coefficients for research and clinical purposes are $\geqslant .7$ and $\geqslant .9$ respectively (Portney and Watkins, 2000; Streiner and Norman, 2005).
a. Inter-rater (inter-observer) reliability	The agreement between different raters/observers of an observational measure of pain (Streiner and Norman, 2005).
b. Test–retest reliability	The agreement between observations with the same individuals on at least two occasions (Streiner and Norman, 2005).
c. Internal consistency	A type of reliability that includes the average of the correlation of scores from a measure with the scores of all of the items in the measure (Streiner and Norman, 2005).
Validity	Used to assess whether that the scale is measuring what it is intending to measure (Streiner and Norman, 2005).
a. Face validity	Whether the pain scale includes appropriate items that appear to measure what it is proposing to measure (Streiner and Norman, 2005).
b. Content validity	The assessment of whether the items in the pain measure include the appropriate information and content (Streiner and Norman, 2005).
c. Criterion	Includes concurrent validity and predictive validity. In concurrent validity, a new pain measure is correlated with a gold standard measure which is administered at the same time. In general, correlations between the new measure and the gold standard should be at least $r \geqslant .3$–.5. The magnitude of the coefficients are hypothesis dependent but should not be too high as to make the new measure redundant. In predictive validity, the correlation of the measure to the criterion variable is determined at a later time (Streiner and Norman, 2005).
d. Construct	Determines the validity of abstract variables that cannot be directly observed, such as pain. These constructs are assessed by their relationships with other variables (Fitzpatrick et al., 1998; Streiner and Norman, 2005).
i. Convergent validity	Evaluates how well items on a pain scale correlate with other measures of the same construct or related variables. In general, correlations between the measure and another measure of the same construct should be $r \geqslant .3$–.5; however, the magnitude of the coefficients are hypothesis dependent (Streiner and Norman, 2005).
ii. Discriminant validity	Evaluates how items on a pain scale correlate with other measures that are unrelated. In general, correlations between the measure and another unrelated measure should be $r < .3$; however, the magnitude of the coefficients are hypothesis dependent (Streiner and Norman, 2005).
Responsivity	Measures whether the measure is able to identify changes in pain over time that is clinically important to patients. An acceptable effect size should be $\geqslant .5$; however, the effect size is hypothesis dependent (Guyatt et al., 1989; Liang, 2000).
Interpretability	The meaningfulness of the scores obtained from a pain measure (Fitzpatrick et al., 1998).
Feasibility	How easily a pain measure can be scored and interpreted (Stevens and Gibbins, 2002).

Table 3
Pain intensity measures excluded from the review with rationale for exclusion

Acronym	Name of Scale	Author	Rationale for exclusion
ACCS	Analogue Chromatic Continuous Scale	Grossi et al. (1983)	Rated approaching well established assessment measure (level II evidence)
	Body outline	O'Donnell and Curley (1985)	Minimal evidence on psychometric testing
CAPS	Children's Anxiety and Pain Scales	Kuttner and LePage (1989)	Minimal evidence on psychometric testing
CGRS	Children's Global Rating Scale	Carpenter (1990)	Minimal evidence on psychometric testing
CAS	Colour Analogue Scale	McGrath et al. (1996)	Rated as approaching well established assessment measure (level II evidence)
	Eland Colour tool	Eland (1981)	Projective test – minimal psychometric testing
FACES	Facial Expression Scale	Suraseranivongse et al. (2005)	Non-English journal; minimal evidence on psychometric testing
	Faces Scale	LeBaron and Zeltzer (1984)	Rated as approaching well established assessment measure (level II evidence); not pure measure of pain intensity
	Faces Scale	Douthit (1990)	No research on psychometric properties
	Faces Scale	Maunuksela et al. (1987)	Minimal evidence on psychometric testing
	Faces Scale	Tree-Trakarn et al. (1987)	Minimal evidence on psychometric testing
	Finger Span	Franzen and Ahlquist (1989)	No research on psychometric properties
	Graphic Numerical Rating Scale	Whaley and Wong (1987)	Rated as promising assessment (level III)
	Glasses Rating Scale	Whaley and Wong (1987)	Rated as promising assessment (level III)
LAPS	Linear analogue pain scale	Broadmann et al. (1988)	Abstract only
MSPCT	Multiple Size Poker Chip Tool	St-Laurent-Gagnon et al. (1999)	Minimal evidence on psychometric testing
	Numerical Rating Scale Thermometer	Szyfelbein et al. (1985)	Minimal evidence on psychometric testing
	Pain Ladder	Hester et al. (1990)	Rated as promising assessment (level III)
	Scheffield Children's Hospital Pain Tool	Goddard and Pickup (1996)	Minimal evidence on psychometric testing
SAS	Smiley Analogue Scale	Pothmann (1990)	Rated as promising assessment (level III)
SAFE	Sydney Animated Facial Expression Scale	Goodenough et al. (1997) (abstract); Hicks et al. (2001)	No research on psychometric properties
4-VDS	4-point Verbal Descriptor Scale	Bernston and Svensson (2001)	Rated as promising assessment (level III)
4-VDS	4-point Verbal Descriptor Scale	Goodenough et al. (1997, 2005)	Rated as promising assessment (level III)
VNRS	Verbal Numerical Rating Scale	Vetter (1992)	No research on psychometric properties
VASOF	Visual Analogue Scale of Faces	Barretto et al. (2004)	Minimal evidence on psychometric testing
VAT	Visual Analogue Toy	White and Stow (1985)	Rated as approaching well established assessment measure (level II evidence)
5-WGRS	5-point Word Graphic Rating Scale	Tesler et al. (1991)	Rated as promising assessment (level III)
6-WGRS	6-point Word Graphic Rating Scale	Whaley and Wong (1987)	Rated as promising assessment (level III)

Table 4
Summary of psychometric properties, interpretability and feasibility of self-report pain intensity measures

Name of Scale (Acronym) Author (year)	Age range	Type of Pain	Reliability	Validity	Responsivity	Interpretability	Feasibility
Pieces of Hurt tool; Hester (1979)	I: 4–7 years S: 3–18 years	Acute procedural, hospital-based	Test–retest (+) Inter-rater reliability (+)	Construct (+++)	Yes (+)	No	Moderate
Faces Pain Scale (FPS); Bieri et al. (1990)	I: 4+ years	Acute procedural, post-op, disease-related	Test–retest (+)	Content (+) Construct (+++)	Yes (++)	Yes	High
Faces Pain Scale-Revised (FPS-R); Hicks et al. (2001)	S: 4–12 years	Acute procedural, post-op, disease related	Test–retest (+)	Content (+) Construct (+)	Yes(+)	No	High
Oucher–Photographic; Beyer and Aradine (1986)	I: 3–7 years S: 3–18 years	Acute procedural, post-op, disease-related	Test–retest (+)	Content (+) Construct (+++)	Yes (++)	Yes	Moderate
Oucher-NRS; Beyer and Aradine (1986)	I: 3–12 years S: 3–18 years	Acute procedural, post-op, disease-related	Test–retest (+)	Content (+) Construct (++)	Yes (++)	Yes	Moderate
Wong–Baker FACES Pain Scale; Wong and Baker (1988)	I: 3–18 years S: 9 months–18 years	Acute procedural, post-op, disease-related	Test–retest (++)	Content (+) Construct (+++)	Yes (++)	Yes	High
Visual Analogue Scale; Scott et al. (1977)	I: 2–17 years S: 3–20 years	Acute, procedural, disease-related, recurrent/chronic	Test–retest (+)	Construct (+++)	Yes (++)	Yes	Moderate

Note: + = 1–3 studies; ++ = 3–6 studies; +++ = >6 studies; I = Intended, S = studied.

has some evidence of discriminant validity ($\gamma = -.004 - .039$) in that it has demonstrated low correlations with measures of fear (Beyer and Aradine, 1988) and low to moderate correlation with pain affect (St-Laurent-Gagnon et al., 1999). The Pieces of Hurt tool has evidence of responsivity post-surgery (Beyer and Aradine, 1987) and post-analgesic administration (Beyer and Aradine, 1988). There are no data regarding the interpretability of this tool.

3.1.3. Advantages and disadvantages

Concrete ordinal rating scales are arguably appealing for use with children because concrete representations (poker chips) enhance their ability to understand the concept of levels of hurt or pain. Hester (1979) originally developed the measure for children 4–7 years old. However, this measure has been used in children as young as 3 years old to adolescents 18 years old. The majority of children studied have described the Pieces of Hurt tool as easy to use and understand. Children have indicated preference for the Pieces of Hurt tool compared to the Wong–Baker FACES Pain Scale and word descriptor scales (Gharaibeh and Abu-Saad, 2002). Conversely, some studies have indicated that children preferred the Faces Pain Scale to the Pieces of Hurt tool (West et al., 1994). Children ranked the Pieces of Hurt tool third (Wong and Baker, 1988) out of a set of five measures (i.e., a simple descriptive scale, a numerical rating scale, the Glasses Rating Scale and a color analogue scale) and fourth out of a different set of five measures (Color Analogue Scale, Faces Pain Scale, Adjective Scale, Finger Span Measure and the Sydney Animated Facial Expressions scale) (Goodenough et al., 2005) in scale preference. The Pieces of Hurt tool has been used to measure procedure-related (e.g., immunizations, intramuscular injections and venipunctures) and post-operative pain, as well as pain in children in an oncology Pediatric Intensive Care Unit. Also, the Pieces of Hurt tool has been validated for use in Jordanian (Gharaibeh and Abu-Saad, 2002) and Thai children (Suraseranivongse et al., 2005) and the instructions have been translated into Spanish. However, there are no psychometric data available for the Spanish translation.

Disadvantages of the Pieces of Hurt tool include: cleaning the chips between patient use, the potential for losing chips and the limited number of response options. Also, there have been no studies using the Pieces of Hurt tool to assess recurrent or chronic pain in children. Lastly, there is only modest evidence of reliability and validity in young preschool children between 3 and 4 years old. Younger children must pass a size-sorting task before they are deemed able to use the Pieces of Hurt tool. However, the responsivity and specificity of this size-sorting task have not been evaluated.

3.2. Faces Pain Scale and Faces Pain Scale-Revised

3.2.1. Description

The Faces Pain Scale consists of seven gender-neutral faces depicting 'no pain' (neutral face) to 'most pain possible' expressions, placed at equal intervals horizontally (Bieri et al., 1990). The child is instructed to point to the face that shows how much pain he/she feels. Ordered faces are scored from 0 to 6. In the revised version, six faces are used, rather than seven, and scoring ranges from 0 to 5 (Hicks et al., 2001). The Faces Pain Scale-Revised was developed to enhance compatibility in scoring with other self-rating and observational scales that use a common metric (0–5 or 0–10).

3.2.2. Reliability and validity data

Both the Faces Pain Scale and the Faces Pain Scale-Revised have evidence of test–retest reliability. The Faces Pain Scale demonstrated adequate stability at a two-week interval ($r = .79$) in healthy children (Bieri et al., 1990), and at one and two days post-surgery in hospitalized children (Perrott et al., 2004). The Faces Pain Scale-Revised indicated adequate stability at one month following a surgical or non-surgical painful condition ($r = .63$) (Miro and Huguet, 2004). Both measures also have established content and construct validity. Strong positive correlations ($r = .59$–$.90$) have been found with the Faces Pain Scale and other well-established self-report pain intensity measures (e.g. Pieces of Hurt tool, Oucher, Wong–Baker FACES Pain Scale) (Goodenough et al., 1997, 2005; Jacobson et al., 1997; Chambers et al., 1999, 2005). Similarly, the Faces Pain Scale-Revised has demonstrated strong positive correlations ($r = .84$ –$.92$) with visual analogue scales (Hicks et al., 2001; Migdal et al., 2005). Moderate to strong positive correlations ($r = .49$–$.90$) have been shown between the Faces Pain Scale and behavioral scales, such as the Children's Hospital of Eastern Ontario Pain Scale (Jacobson et al., 1997; Cassidy et al., 2002) and the Child Facial Coding System (Cassidy et al., 2002). The Faces Pain Scale has demonstrated responsivity following procedural pain (Goodenough et al., 1997; Wolf et al., 2002) and the Faces Pain Scale-Revised has demonstrated responsivity following administration of lidocaine during venipuncture (Migdal et al., 2005; Taddio et al., 2005). Finally, a rating of three or more on the Faces Pain Scale has been found to represent clinically significant pain in children (Gauthier et al., 1998), which confers interpretability to the scale.

3.2.3. Advantages and disadvantages

The Faces Pain Scale and Faces Pain Scale-Revised are psychometrically sound and feasible. These measures are simple, quick to use and require minimal instruction in children 4–18 years old. The Faces Pain Scale has also demonstrated a lack of upper-end bias in response

distribution of pain scores in children undergoing acute procedure-related (surgical or medical) pain compared to five other self-report measures (e.g., Pieces of Hurt tool, Color Analogue Scale, Adjective Scale, Finger Span Measure and the Sydney Animated Facial Expressions scale) (Goodenough et al., 2005). However, ratings on the Faces Pain Scale tend to result in pain intensity ratings skewed towards the 'no pain' end of the scale (Goodenough et al., 1997; Chambers et al., 1999; Perrott et al., 2004; Chambers et al., 2005). Because the majority of these studies were done with short procedure-related pain (e.g., blood draws and immunizations); these findings may represent the nature of the pain experience rather than being an artifact of the measure. An advantage of the Faces Pain Scale-Revised is that the instructions have been translated into 25 languages, which can be obtained free of charge for use in clinical practice and research from http://painsourcebook.ca/docs/pps92.html. However, there is limited evidence regarding the psychometric properties of these translated versions (von Baeyer and Piira, 2004), with testing limited to French (Wood et al., 2004), Thai (Newman et al., 2005) and Catalan (Miro and Huguet, 2004) versions in children and adolescents.

Disadvantages of the Faces Pain Scale and the Faces Pain Scale-Revised include the limited evidence regarding interpretability and mixed evidence regarding the acceptability of the scale with children. Some researchers have found low acceptability with school-aged children, wherein children have ranked the Faces Pain Scale as fourth among five presented pain scales (i.e., Color Analogue Scale, Faces Pain Scale, Wong–Baker FACES Pain Scale, and several other less well-validated faces scales) (Chambers et al., 1999; Chambers et al., 2005). However, the Faces Pain Scale has been reported as being well accepted by children aged 4–17 years (Jacobson et al., 1997; Goodenough et al., 2005), and children as young as 3 years old have used the scale with adequate comprehension (Bieri et al., 1990).

3.3. Wong–Baker FACES Pain Scale

3.3.1. Description

The Wong–Baker FACES Pain Scale consists of six hand-drawn faces that range from smiling to crying (Wong and Baker, 1988). The faces represent 'no hurt' to 'hurts worst' and the scale is scored from 0 to 5.

3.3.2. Reliability and validity data

Test–retest reliability evidence indicates a relatively high stability over 15 min ($r = .90$, Cronbach's $\alpha = .93$; Keck et al., 1996) and 8 h ($r = .84$; Gharaibeh and Abu-Saad, 2002) immediately post-procedure in children 3–18 years old. There is some evidence of discriminant validity ($r = .01$) (Wong and Baker, 1988; Stein, 1995; Keck et al., 1996). In terms of concurrent validity, strong positive correlations have been demonstrated between the Wong–Baker FACES Pain Scale and other well-established self-report measures ($r = .74$–$.78$) (e.g., Pieces of Hurt tool, Faces Pain Scale, and a visual analogue scale) (Robertson, 1993; West et al., 1994; Stein, 1995; Gharaibeh and Abu-Saad, 2002). The Wong–Baker FACES Pain Scale has indicated responsivity in terms of detecting changes in children's pain intensity following procedural (Stein, 1995; Keck et al., 1996; Kendall et al., 2001; Gharaibeh and Abu-Saad, 2002; Robert et al., 2003) and post-operative pain (Robertson, 1993).

3.3.3. Advantages and disadvantages

The Wong–Baker FACES Pain Scale is psychometrically sound and widely used in clinical practice. This scale has been translated into ten different languages (i.e., Spanish, French, Italian, Portuguese, Romanian, Bosnian, Vietnamese, Japanese, Chinese and German). However, limited psychometric evidence exists for these translations. Advantages of the Wong–Baker FACES Pain Scale include ease of administration (i.e., quick and simple to use, requiring minimal instruction) (McRae et al., 1997) and cost-effectiveness (i.e., can be obtained free of charge from http://www3.us.elsevierhealth.com/WOW/, is easily reproduced by photocopying and wearable pins can be purchased for $5 U.S.) (Keck et al., 1996). Overall, children have indicated preference for the Wong–Baker FACES Pain Scale relative to other measures, including a simple verbal rating scale, numerical rating scales, a graphic rating scale, Pieces of Hurt tool (Wong and Baker, 1988; West et al., 1994; Keck et al., 1996) and the Color Analogue Scale (Chambers et al., 2005).

A disadvantage of this scale is that young children (e.g., 4–5 years old) have demonstrated a tendency to select faces at the extremes of the scale during procedure-related pain (Stein, 1995). More research is needed to determine if this is related to the nature of the pain or if it is an artifact of the scale. Also, further testing is required to determine whether the scale has interval-quality measurement properties. Moreover, children's pain ratings appear to be influenced by the smiling 'no pain' anchor, tending to be higher relative to faces scales with neutral 'no pain' anchors (Chambers et al., 1999, 2005). Also, the tears on the face in the upper anchor may lead to an under-estimation of pain by some children (e.g., those who do not want to admit to crying). Chambers et al. (1999) suggest that scales with smiling faces may be more appropriate as measures of pain affect, rather than pain intensity. Neutral starting points for 'no pain' are generally recommended.

3.4. Oucher

3.4.1. Description

The Oucher is a combination of two separate scales: the photographic faces scale and a 0–100 mm vertical

numerical rating scale. The photographic faces scale consists of six photographs of culturally sensitive faces (Caucasian, Afro-American and Hispanic) that are scored from 0 to 5. The adjacent vertical numerical rating scale is scored from 0 to 100.

3.4.2. Reliability and validity data

The Oucher demonstrated higher test–retest reliability than the Wong–Baker FACES Pain Scale and visual analogue scale (Luffy and Grove, 2003). However, Belter et al. (1988) found that the Oucher had poor test–retest reliability for individual item ratings, despite having acceptable reliability for subscales of items for different levels of pain. The Oucher also has evidence of content (Beyer and Aradine, 1986; Villarruel and Denyes, 1991) and construct validity (i.e., known groups, convergent and discriminant) (Beyer and Aradine, 1987; Aradine et al., 1988; Beyer and Aradine, 1988; Beyer et al., 1990; Beyer et al., 1992; Ramritu, 2000; Luffy and Grove, 2003). This measure has demonstrated strong positive correlations ($r = .59$–$.95$) with other well-established pain intensity measures (e.g., Pieces of Hurt tool, Faces Pain Scale, and a visual analogue scale) (Beyer and Aradine, 1988; Jacobson et al., 1997; Luffy and Grove, 2003) and discriminant validity ($r = .21$–$.33$) with two measures of fear (Beyer and Aradine, 1988; Beyer and Knott, 1998). The Oucher has strong evidence of responsivity in terms of detecting change in children's pain intensity pre and post-operatively and pre and post-analgesic administration (Beyer and Aradine, 1987; Aradine et al., 1988; Beyer and Knott, 1998; Ramritu, 2000). Some information exists regarding the interpretability of this measure in terms of responsivity to analgesic administration. Children that stated they had 'some to good relief' had one or more scores of 0 mm after analgesic administration, while those who reported 'little or no relief' had scores greater than 40 mm (Aradine et al., 1988).

3.4.3. Advantages and disadvantages

While the Oucher photographic scale has evidence of reliability and validity, its use in younger children, especially 3–4 year olds, requires further testing (Belter et al., 1988). Moreover, the photographic scale has lower feasibility and clinical utility relative to other faces scales and there is limited evidence regarding interpretability. To overcome the cultural limitations associated with the original version, culturally sensitive photographic scales have been developed (i.e., Hispanic and African-American) (Villarruel and Denyes, 1991). However, the photos are of real children and are neither gender nor ethnically neutral. Furthermore, facial expressions in the photographs are of acute rather than chronic pain, thus limiting the range of clinical contexts in which this tool can be used.

The Oucher numerical rating scale has been extensively studied in terms of validity and responsivity, and there is beginning evidence of reliability and interpretability. The numerical rating scale appears to be generally well accepted by children and can be used in those greater than 6 years old. Beyer and Aradine (1988) found that five young children (3–7 years of age) were not able to complete the numerical rating scale. Children need to be screened in order to determine their ability to count to 100 by ones or tens, as well as to determine their ability to identify which of two numbers is larger or complete a seriation test (Beyer and Aradine, 1988). Limited testing has been done with the numerical scale in adolescents greater than 12 years old and there is a lack of clarity on whether resultant data are interval or ratio level.

Practical issues with the Oucher include that it is costly (approximately $3 U.S.) and that the laminated scale must be disinfected between patients. Furthermore, evidence regarding patient preference of the Oucher is mixed. Some researchers have found that children and health care professionals preferred the Oucher to a verbal rating scale (Ramritu, 2000). However, others have reported that the Wong–Baker FACES Pain Scale was preferred to the Oucher (Luffy and Grove, 2003). The primary limitation of the Oucher is that the numerical rating scale is combined with a photographic scale that requires further testing in young children. Future research is needed to determine if a verbal numerical rating scale or a separate graphic numerical rating scale would improve the feasibility of the Oucher for use in clinical pain trials.

3.5. Visual analogue scales

3.5.1. Description

Visual analogue scales consist of a pre-measured vertical or horizontal line, where the ends of the line represent the extreme limits of pain intensity. The child is asked to select a point or make a mark along the line to indicate the intensity of his/her pain. There were many versions of visual analogue scales found in the literature for use with children. Differences included: the anchor terminology, the presence or absence of divisions along the line, the units of measurement (e.g., cm or mm), the length of the scale (i.e., 10, 15 or 16 cm) and whether the scale was presented as a vertical or horizontal line.

3.5.2. Reliability and validity data

Test–retest reliability evidence for visual analogue scales indicated a moderate to strong positive median correlation ($r = .70$) between pain intensity ratings reported by 5–6 year olds over a two-week interval (McGrath et al., 1985). The strength of the median between-session correlation increased in children aged

13–15 years ($r = .99$). Noteworthy, however, is that Bernston and Svensson (2001) found the visual analogue scale to be less reliable compared to a 4-point verbal descriptor scale in a small sample of children aged 10–17 years. It was posited that children might be unable to reliably transform an interpretation of their pain into an assessment on a visual analogue scale. However, verbal descriptor scales have minimal evidence of reliability and validity.

In terms of construct validity, visual analogue scales have demonstrated moderate to strong correlations ($r = .63$–$.90$) with several other pain measures (e.g., Faces Pain Scale and Oucher) (Beyer & Aradine, 1988; Tyler et al., 1993; Goodenough et al., 1997; Migdal et al., 2005). Visual analogue scales have also shown responsivity to change following surgery (Tyler et al., 1993), administration of analgesics (Abu-Saad and Holzemer, 1981; Aradine et al., 1988; Tyler et al., 1993) and following application of a local anesthetic (Migdal et al., 2005). Mixed results were found regarding the effect of age on visual analogue pain intensity ratings. In children post-surgery (Beyer and Aradine, 1987) and those with juvenile rheumatoid arthritis (Beales et al., 1983) it was determined that the younger the child, the lower the pain rating. Conversely, in children undergoing venipuncture, it was found that the younger the child, the higher the pain rating (Goodenough et al., 1997). These contrasting findings may be due to the different types of pain that were measured, rather than being attributable to a scale property. Overall, visual analogue scales are thought to be less reliable for children younger than 8 years old (Beyer and Aradine, 1988; Shields et al., 2003a,b). Similar to other types of pain intensity scales (e.g., concrete ordinal, faces pain scales), young children are more likely to select the endpoints of visual analogue scales. Such age-related differences in the self-report of pain likely reflect fundamental differences in cognitive processing (Champion et al., 1998).

There is some evidence of interpretability with the minimal clinically significant difference in visual analogue pain scores for acute pain in children being 10 mm (95% confidence interval 7–12 mm) (Powell et al., 2001). Powell and colleagues (2001) asked 73 children between the ages of 8 and 15 to rate their pain on a 100 mm non-hatched visual analogue scale on admission to the emergency department. At twenty minute intervals, participants were subsequently asked to provide a global impression of change score using a 5-point categorical rating scale (“heaps better”, “a little better”, “much the same”, “a bit worse”, “heaps worse”) and rate their pain on the visual analogue scale again. The minimally clinically significant difference (MCSD) in the visual analogue pain score was defined by the authors as the mean difference between the current and preceding scores when the participant reported the pain to be “a bit worse” or “a little better”. While these findings are similar to several studies conducted in adults with acute pain (MCSD = 9 mm, Kelly, 1998; MSCD = 13 mm, Todd et al., 1996), others have found much larger differences ($\geqslant$ 2 points on a 0–10 numeric rating scale, Farrar et al., 2003; 30 mm on a visual analogue scale, Lee et al., 2003). Furthermore, these findings cannot be generalized to children with chronic pain. Finally, while this difference might be considered a noticeable difference by most children and adults, it is questionable whether a 10 mm difference represents a clinically meaningful reduction in pain.

3.5.3. Advantages and disadvantages

Visual analogue scales have been extensively researched and show good acceptability, responsivity and validity for most children aged 8 years and older (Champion et al., 1998). An advantage of visual analogue scales is increased responsivity in pain intensity scoring due to the interval or ratio measurement continuum (Champion et al., 1998). However, pain assessments derived from visual analogue scales have been determined as difficult to interpret, rather than necessarily having equal interval properties (Bernston and Svensson, 2001). Some argue that visual analogue scale assessments in younger children may produce only ordinal level data (Shields et al., 2003a,b). Creative strategies have been employed to try and improve the reliability, validity and responsiveness of visual analogue scales for use in children by using graphic (Colour Analogue Scale; McGrath et al., 1996) or other methods (i.e., Visual Analogue Toy; White and Stow, 1985) to enhance the child’s understanding of the measure. However, these measures require further psychometric testing.

Another suggested advantage of visual analogue scales is their ease and quickness of use. However, to ensure comprehension and proper use by children of varying ages and cognitive levels, visual analogue scales require careful explanation (Beales et al., 1983). Although seriation testing has been identified as the single best predictor of children’s ability to use visual analogue scales (Shields et al., 2003a,b), it is an impractical technique in busy clinical settings and increases child and researcher burden. Visual analogue scales are versatile because they can be used to rate different dimensions of pain on the same scale (e.g., pain intensity and pain affect/unpleasantness) in children greater than 8 years old (Goodenough et al., 1997). While visual analogue scales are readily reproducible, photocopying may alter the scale by increasing or decreasing the length of the line. Also, paper-based visual analogue scales require an extra step in measuring the line; measuring where the mark increases the burden on the researcher and the likelihood of error.

Finally, the use of visual analogue scales is hampered by a lack of standardization in terms of the verbal anchors, length and orientation (vertical or horizontal)

of the line. While 100 mm visual analogue scales are most commonly used, further research needs to be conducted on the most appropriate upper verbal anchors (e.g., worst pain you can imagine versus worst pain possible). For example, several visual analogue scales included a hypothetical worst imaginable pain as the upper anchor. This verbal anchor may be difficult for young children and depends on the child's experience and knowledge of other people's pain.

4. Discussion

4.1. Recommendations for measuring self-reported pain intensity in clinical trials

There are currently more than 30 pediatric self-report pain intensity measures; however, only six of these have well-established evidence of reliability and validity. These six measures have varying degrees of responsivity and modest evidence of interpretability. Of the six measures reviewed, no single scale was found to be reliable and valid across age groups or pain types, with the majority of scales lacking reliability and validity in pre-school children. Moreover, there are important differences in failure rates and children's preferences across measures. Based on the evidence of this review we have developed approximate age-based recommendations for the use of self-report measures in clinical trials which are summarized in Table 5. However, it is important to emphasize that age is a proxy for developmental level and different children will have different pain assessment needs. Furthermore, psychometric testing of measurement tools is a dynamic and ongoing process. As the body of evidence supporting psychometric properties of pain intensity measures grows, these recommendations will undoubtedly change. Finally, research recommendations on ways to strengthen pain measurement for particular pain types and developmental age groups are discussed.

4.1.1. Recommendations for measuring self-reported pain intensity in preschool children

Communication barriers pose challenges to establishing reliability and validity of pain intensity self-report measures in young children (Champion et al., 1998). Young preschool children may lack the requisite comprehension level to use measures and tend to favor the extreme ends of scales (Chambers and Johnston, 2002). None of the six measures reviewed is ideal for measurement of pain in preschool children. The Pieces of Hurt Tool (Hester, 1979) currently has the best-established reliability and validity for acute procedure-related and post-operative pain in pre-school children. This tool has not been validated in preschool children with chronic pain, which likely reflects the low prevalence rates of chronic pain conditions in this age group. The Pieces of Hurt measure was preferred to the Oucher photographic scale because the latter tool is costly, requires cleaning between children and the faces depict acute pain only and are not gender neutral.

Given the modest reliability and validity of the Pieces of Hurt Tool in preschool children, it would be prudent to include a well-established observational behavioral measure such as the FLACC (Face, Legs, Activity, Cry, Consolability; Merkel et al., 2002) to supplement self-reports of acute pain in this age group (von Baeyer and Spagrud, 2005). Using observational measures to complement self-report measures of pain intensity is crucial as preschool children may find it difficult to understand and use a self-report scale and their self-report ratings are likely to be affected by cognitive, emotional or situational factors (Chambers and Johnston, 2002).

4.1.2. Recommendations for measuring self-reported pain intensity in school-aged children (4–12 years old)

Of the six reviewed measures, the Faces Pain Scale-Revised appears to be the most psychometrically sound measure in school-aged children. Furthermore, it has been validated for use in both acute and chronic disease-related pain. The Faces Pain Scale-Revised has lower failure rates and higher patient preference ratings compared to the Oucher numerical rating scale or visual analogue scales, making it more reliable and valid in young school aged children. The abstractness of numerical rating and visual analogue scales is problematic for young school-aged children who indicate preference for faces scales. While faces scales,

Table 5
Summary of recommendations of self-report pain intensity measures for clinical trials in children 3–18 years of age

- In most clinical trials, a single-item self-report measure of pain intensity is the appropriate primary outcome dimension for children 3 years of age and older.
- Pieces of Hurt tool is recommended for acute procedure-related and post-operative pain in young children between 3 and 4 years of age.
- Given the wide variability in young children's ability to use self-report measures especially between the ages of 3 and 7 years of age, it would be prudent to consider using a behavioral observational measure as a secondary outcome in this age group.
- Faces Pain Scale-Revised is recommend for acute procedure-related, post-operative, and disease-related pain in children between 4 and 12 years of age.
- A 100 mm visual analogue scale is recommended for acute procedure-related, post-operative, and disease-related pain in children over the age of 8 years of age and adolescents.
- For children between the ages of 8 and 12 years it might be useful to use the Faces Pain Scale-Revised as a secondary outcome measure with the visual analogue scale.

such as the Faces Pain Scale-Revised and Wong–Baker FACES Pain Scales, appear superior for measuring pain in school-aged children, some children less than seven years still have difficulty using these measures. Although the Wong–Baker FACES Pain Scale appears preferable to other faces scales for clinical practice, its use in clinical trials is hampered by the anchor effect (may be more appropriately conceived as measure of pain affect) and because a painful facial expression does not appear below a score of six out of ten.

4.1.3. Recommendations for measuring self-reported pain intensity in older school-aged children and adolescents (⩾8 years old)

The Faces Pain Scale-Revised, Wong–Baker FACES Pain Scale, and Oucher numerical rating and visual analogue scales have well-established psychometric properties in older school-aged children and adolescents; however, the visual analogue scale appears to be the best measure for clinical trials. Adolescents indicate preference for visual analogue and numerical rating scales compared to faces scales. However, there is currently insufficient evidence to determine adolescents' relative preference for visual analogue or numerical rating scales. Numerical rating and visual analogue scales have a theoretical advantage in that scores can be treated as interval or ratio level data. However, it is unclear from this review which level of data scores actually provide for these two measures. Other advantages of the visual analogue scale are its ease of administration and reproduction, as well as its application as a measure of pain intensity and affect in older children and adolescents (Goodenough et al., 1999).

Disadvantages of the visual analogue scale include a lack of standardization in upper-anchor wording, line length and orientation, increased scoring burden and greater failure rates in young children compared to the Oucher-numerical rating scale. The Color Analogue Scale (McGrath et al., 1996) is a promising alternative to the standard visual analogue scale and could overcome many of its disadvantages. However, the Color Analogue Scale requires further psychometric testing and has practical drawbacks related to its cost, availability, and cleaning.

Although the 11-point numerical rating scale has been recommended for use in clinical pain trials with adults (Dworkin et al., 2005), there are disadvantages to using the Oucher-numerical rating scale with children. The appeal of this measure is that it targets a wide age group; however, the numerical rating scale is combined with a photographic faces scale that lacks adequate testing in large samples of young children. Current evidence suggests that visual analogue scales are superior for measuring pain intensity in clinical trials with adolescents. However, in light of the issues associated with visual analogue scales, supplementation with the Faces Pain Scale-Revised (especially in children 8–12 years of age) might reduce failure rates and allow for the comparison of findings across studies that use Faces Pain Scale-Revised in younger school-aged children.

4.2. Recommendations for future research on self-report pain intensity measures of pain in clinical trials

This review highlighted several important areas for further research on self-report pain intensity measures. Given the challenges of measuring pain in young children, further research should be directed towards the development and testing of other concrete measures for this population. While the Pieces of Hurt Tool has the best psychometric evidence to date, other measures such as the Multiple Size Poker Chip Tool (St-Laurent-Gagnon et al., 1999), which uses four poker chips of increasing size, might assist young children to comprehend and more accurately rate their pain. While this measure is promising, it requires further psychometric testing. Given the high failure rates and tendency to use extremes of scales in young children, more research is also needed to establish screening methods to determine which young children can and cannot provide meaningful self-reports. Lastly, the effects of standardized instructions and other methods that ensure children can understand and use self-report measures (e.g., seriation or practice with hypothetical pain vignettes) should be established.

Another key area for future research is the standardization and validation of a verbally administered 11-point numerical rating scale given its wide use in clinical practice. This measure should be studied using experimental and clinical pain stimuli over a wide age range and types of pain. It is also important to determine the responsivity of this measure to analgesic administration for acute procedural and post-operative pain. Finally, the relative superiority of verbal or graphic numerical rating scales needs to be addressed.

Lastly, more research needs to be conducted to determine the MCSD in pain at different ages, using different scales, with different clinical sources of pain. Only one study has been conducted in children to determine the MCSD in acute pain using a visual analogue scale (Powell et al., 2001). However, it is questionable whether a difference of 10 mm represents a meaningful reduction in pain. Research needs to be done on the best way to operationalize a meaningful improvement in pain. It has been suggested that for clinical trials the most cautious approach would be to define "meaningful improvement" as being both a rating of "much improved" and a 30% reduction in the rating of pain intensity (Rowbothman, 2001).

5. Conclusion

This systematic review provides evidence of the psychometric properties and feasibility of commonly used

self-report pain intensity measures in children and adolescents. Findings highlight the critical developmental issues that affect pediatric pain measurement. Recommendations were made based on the best available psychometric evidence to date. No single pain intensity measure is appropriate across ages or types of pain. Psychometrically sound measures are essential for determining the effect of pain relieving treatments and combining the results of studies in meta-analyses to facilitate evidence-based decision-making in clinical practice.

Acknowledgements

The authors thank Drs. Chambers and McGrath as well as the anonymous reviewers for their thoughtful feedback on earlier versions of this manuscript and members of the Ped-IMMPACT working group.

References

Abu-Saad H, Holzemer W. Measuring children's self-assessment of pain. Issues Compr Pediatr Nurs 1981;5:337–49.

Aradine C, Beyer J, Tompkins J. Children's pain perception before and after analgesia: a study of instrument construct validity and related issues. J Pediatr Nurs 1988;3(1):11–23.

Barretto EP, Ferreira E, Pordeus IA. Evaluation of toothache severity in children using a visual analogue scale of faces. Pediatr Dent 2004;26(6):485–91.

Beales G, Keen J, Lennox Holt P. The child's perception of the disease and the experience of pain in juvenile chronic arthritis. J Rheumatol 1983;10(1):61–5.

Belter R, McIntosh J, Finch Jr A, Saylor C. Preschoolers' ability to differentiate levels of pain: relative efficacy of three self-report measures. J Clin Child Psychol 1988;17(4):329–35.

Bernston L, Svensson E. Pain assessment in children with juvenile chronis arthritis: a matter of scaling and rater. Acta Paediatr 2001;90:1131–6.

Beyer J, Aradine C. Content validity of an instrument to measure young children's perceptions of the intensity of their pain. J Pediatr Nurs 1986;1(6):386–95.

Beyer J, Aradine C. Patterns of pediatric pain intensity: a methodological investigation of a self-report scale. Clin J Pain 1987;3:130–41.

Beyer J, Aradine C. Convergent and discriminant validity of a self-report measure of pain intensity for children. Child Health Care 1988;16(4):274–81.

Beyer J, Denyes M, Villarruel A. The creation, validation, and continuing development of the oucher: a measure of pain intensity in children. J Pediatr Nurs 1992;7(5):335–46.

Beyer J, Knott C. Construct validity estimation for the African-American and hispanic versions of the oucher scale. J Pediatr Nurs 1998;13(1):20–31.

Beyer J, McGrath P, Berde C. Discordance between self-report and behavioural pain measures in children aged 3–7 years after surgery. J Pain Symptom Manage 1990;5(6):350–6.

Bieri D, Reeve R, Champion G, Addicoat L, Ziegler J. 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;41:139–50.

Broadmann L, Rice L, Hannallah R. Evaluation of an objective pain scale for infants and children. Reg Anesth 1988;13:45.

Carpenter P. New method for measuring young children's self-report of fear and pain. J Pain Symptom Manage 1990;5(4):233–40.

Cassidy KL, Reid GJ, McGrath PJ, Finley GA, Smith DJ, Morley C, et al. Watch needle, watch TV: an audiovisual distraction in preschool immunization. Pain Med 2002;3(2):108–18.

Chambers C, Giesbrecht K, Craig K, Bennett S, Huntsman E. A comparison of faces scales for the measurement of pediatric pain: children's and parents' ratings. Pain 1999;83:25–35.

Chambers C, Hardial J, Craig K, Court C, Montgomery C. Faces scales for the measurement of postoperative pain intensity in children following minor surgery. Clin J Pain 2005;21(3):277–85.

Chambers CT, Johnston C. Developmental differences in children's use of rating scales. J Pediatr Psychol 2002;27:27–36.

Champion GD, Goodenough B, von Baeyer CL, Thomas W. Measurement of Pain by Self-Report. In: Measurement of Pain in Infants and Children. In: Finley GA, McGrath PJ, editors. Progress in Pain Research and Management, vol. 10. Seattle: IASP Press; 1998. p. 123–60.

Cohen LL, La Greca AM, Blount RL, Kazak AE, Holmbeck GN, Lemanek KL. Introduction: evidence-based assessment in pediatric psychology. J Pediatr Psychol, 2006; in press.

Douthit J. Patient care guidelines: psychological assessment and management of pediatric pain. J Emerg Nurs 1990;16(3):168–70.

Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.

Eland J. Minimizing pain associated with prekindergarten intramuscular injections. Issues Compr Pediatr Nurs 1981;5:361–72.

Farrar JT, Berlin JA, Strom BL. Clinically important changes in acute pain outcome measures: a validation study. J Pain Symptom Manage 2003;25(5):406–11.

Fitzpatrick R, Davey C, Buxton M, Jones D. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2(14):1–74.

Franzen OG, Ahlquist ML. The intensive aspect of information processing in the intradental A-delta system in man—a psychophysiological analysis of sharp dental pain. Behav Brain Res 1989;33(1):1–11.

Gharaibeh M, Abu-Saad H. Cultural validation of pediatric pain assessment tools: Jordanian perspective. J Transcult Nurs 2002;13(1):12–8.

Gauthier JC, Finley GA, McGrath PJ. Children's self-report of postoperative pain intensity and treatment threshold: determining the adequacy of medication. Clin J Pain 1998;14:116–20.

Goddard J, Pickup S. Postoperative pain in children. Anaesthesia 1996;51:588–90.

Goodenough B, Addicoat L, Champion G, McInerney M, Young B, Juniper K, et al. Pain in 4 to 6 year-old children receiving intramuscular injections: a comparison of the faces pain scale with oucher self-report and behavioral measures. Clin J Pain 1997;13(1):60–73.

Goodenough B, Thomas W, Champion G, Perrot D, Taplin J, von Baeyer C, et al. Unravelling age effects and sex differences in needle pain: ratings of sensory intensity and unpleasantness of venipuncture pain by children and their parents. Pain 1999;80:179–90.

Goodenough B, Piira T, von Baeyer CL, et al. Comparing six self-report measures of pain intensity in children. The Suffering Child 2005;8:1–25.

Grossi E, Borghi C, Cerchiara E, Della Puppa T, Francucci B. Analogue continuous chromatic scale (ACCS): a new method for pain assessment. Clin Exp Rheumatol 1983;1:337–40.

Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol 1989;42(5):403–8.

Hester N. The preoperational child's reaction to immunization. Nurs Res 1979;28(4):250–5.

Hester N, Foster R, Kristensen K. Measurement of pain in children: generalizability and validity of the pain ladder and Pieces of Hurt tool. Adv Pain Res Ther 1990;15:79–84.

Hicks C, von Baeyer C, Spafford P, van Korlaar I, Goodenough B. The faces pain scale-revised: toward a common metric in pediatric pain measurement. Pain 2001;93:173–83.

Jacobson SJ, Kopecky EA, Joshi P, Babul N. Randomized trial of oral morphine for painful episodes of sickle-cell disease in children. Lancet North Am Ed 1997;350:1358–61.

Jakobs H, Rister M. Team assessment of pain in children. Klin Padiatr 1997;209(6):384–8.

Keck J, Gerkensmeyer J, Joyce B, Schade J. Reliability and validity of the faces and word descriptor scales to measure procedural pain. J Pediatr Nurs 1996;11(6):368–74.

Kelly AM. Does the clinically significant difference in visual analogue scale pain score differ with gender, age or cause of pain? Acad Emerg Med 1998;5:1086–90.

Kendall JM, Reeves VC, Latter VS. Multicentre randomized controlled trial of nasal diamorphine for analgesia in children and teenagers with clinical fractures. Br Med J 2001;322:261–5.

Kuttner L, LePage T. Faces scales for the assessment of pediatric pain: a critical review. Can J Behav Sci 1989;21(2):199–209.

Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Med Care 2000;38(9):84–90. [Suppl. II].

LeBaron L, Zeltzer L. Assessment of acute pain and anxiety in children and adolescents by self-reports, observer reports, and a behavior checklist. J Consult Clin Psychol 1984;52(5):729–38.

Lee JS, Hobden E, Stiell IG, Wells GA. Clinically important change in visual analog scale after adequate pain control. Acad Emerg Med 2003;10(10):1128–30.

Luffy R, Grove S. Examining the validity, reliability and preference of three pediatric pain measurement tools in African-American children. Pediatr Nurs 2003;29(1):54–9.

Maunuksela EL, Olkkola KT, Korpela R. Measurement of pain in children with self-reporting and behavioral assessment. Clin Pharmacol Ther 1987;42(2):137–41.

McGrath P, de Veber L, Hearn M. Multidimensional pain assessment in children. Paper presented at the Advances in pain research and therapy: Proceedings from the 4th world congress on pain, Seattle, WA, 1985.

McGrath P, Seifert C, Speechley K, Booth J, Stitt L, Gibson M. A new analogue scale for assessing children's pain: an initial validation study. Pain 1996;64:435–43.

McGrath PJ, Turk DC, Dworkin RH, Brown MT, Davidson K, Eccleston C, et al. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: the PedIMMPACT recommendations, 2005, in submission.

McRae M, Rourke D, Imperial-Perez F, Eisenring C, Ueda J. Development of a research-based standard for assessment, intervention, and evaluation of pain after neonatal and pediatric cardiac surgery. Pediatr Nurs 1997;23(3):263–71.

Merkel S, Voepel-Lewis T, Malviya S. Pain assessment in infants and young children: the FLACC scale. Am J Nurs 2002;102(10):55–8.

Migdal M, Chudzynska-Pomianowska E, Vause E, Henry E, Lazar J. Rapid, needle-free delivery of lidocaine for reducing the pain of venipuncture among pediatric patients. Pediatrics 2005;115(4):e393–8.

Miro J, Huguet A. Evaluation of reliability, validity, and preference for a pediatric pain intensity scale: the Catalan version of the faces pain scale-revised. Pain 2004;111:59–64.

Newman CJ, Lolekha R, Limkittikul K, Luangxay K, Chotpitayasunondh T, Chanthavanich P. A comparison of pain scales in Thai children. Arch Dis Child 2005;90(3):269–70.

O'Donnell PJ, Curley H. Validation of a nonverbal instrument for pain location descriptions in children. Percept Mot Skills 1985;60(3):1010.

Perrott DA, Goodenough B, Champion GD. Children's ratings of the intensity and unpleasantness of post-operative pain using facial expression scales. Eur J Pain 2004;8:119–27.

Pothmann R. Comparison of the visual analog scale (VAS) and a smiley analog scale (SAS) for the evaluation of pain in children. Adv Pain Res Ther 1990;15:95–9.

Portney LG, Watkins MP. Statistical measures of reliability. In: Portney LG, Watkins MP, editors. Foundation of clinical research: applications to practice. 2nd ed. Connecticut: Appleton & Lange; 2000. p. 57–586.

Powell CV, Kelly AM, Williams A. Determining the minimum clinically significant difference in visual analog pain score for children. Ann Emerg Med 2001;37(1):28–31.

Ramritu P. Use of the oucher numeric and the word graphic scale in children 9–14 years with post-operative pain. J Clin Nurs 2000;9:763–73.

Robert R, Brack A, Blakeney P, Villarreal C, Rosenberg L, Thomas C, et al. A double-blind study of the analgesic efficacy of oral transmucosal fentanyl citrate and oral morphine in pediatric patients undergoing burn dressing change and tubbing. J Burn Care Rehabil 2003;24:351–5.

Robertson J. Pediatric pain assessment: validation of a multidimensional tool. Pediatr Nurs 1993;19(3):209–13.

Rossato LM, Angelo M. Using tools for pain perception assessment in hospitalized pre-school children undergoing painful procedures. Rev Esc Enferm USP 1999;33(3):236–49.

Rowbothman MC. What is a 'clinically meaningful' reduction in pain? Pain 2001;94:131–2.

Scott P, Ansell B, Huskisson E. Measurement of pain in juvenile chronic polyarthritis. Ann Rheum Dis 1977;36:186–7.

Shields B, Cohen D, Harbeck-Weber C, Powers J, Smith G. Pediatric pain measurement using a visual analogue scale: a comparison of two teaching methods. Clin Pediatr (Phila) 2003a;42(3):227–34.

Shields B, Palermo T, Powers J, Grewe S, Smith G. Predictors of a child's ability to use a visual analogue scale. Child Care Health Dev 2003b;29(4):281–90.

Stein P. Indices of pain intensity: construct validity among preschoolers. Pediatr Nurs 1995;21(2):119–23.

Stevens B, Gibbins S. Clinical utility and clinical significance in the assessment and management of pain in vulnerable infants. Clin Perinatol 2002;29(3):459–68.

St-Laurent-Gagnon T, Bernard-Bonnin AC, Villeneuve E. Pain evaluation in preschool children and by their parents. Acta Paediatr 1999;88:422–7.

Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 3rd ed. Oxford: Oxford University Press; 2005.

Suraseranivongse S, Montapaneewat T, Monon J, Chainhop P, Petcharatana S, Kraiprasit K. J Med Assoc Thai 2005;88(3):412–8.

Szyfelbein S, Osgood P, Carr D. The assessment of pain and plasma B-Endorphin immunoactivity in burned children. Pain 1985;22:173–82.

Taddio A, Soin HK, Schuh S, Koren G, Scolnik D. Liposomal lidocaine to improve procedural success rates and reduce procedural pain among children: a randomized controlled trial. Can Med Assoc J 2005;172(13):1691–5.

Tesler MD, Savedra MC, Holzemer WL, Wilkie DJ, Ward JA, Paul SM. The word graphic rating scale as a measure of children's and adolescents' pain intensity. Res Nurs Health 1991;14(5):361–71.

Todd KH, Funk KG, Funk JP, et al. Clinical significance of reported changes in pain severity. Ann Emerg Med 1996;27:485–9.

Tree-Trakarn T, Pirayavaraporn S, Lertakyamanee J. Topical analgesia for relief of post-circumcision. Anesthesiology 1987;67:395–9.

Tyler D, Tu A, Douthit J, Chapman R. Toward validation of pain measurement tools for children: a pilot study. Pain 1993;52:301–9.

Vetter T. Pediatric patient-controlled analgesia with morphine versus meperidine. J Pain Symptom Manage 1992;7(4):204–8.

Vihunen R, Sihvonen M. Use of faces scale in children's self-assessment of pain. Hoitotiede 1998;10(3):123–33.

Villarruel A, Denyes M. Pain assessment in children: theoretical and empirical validity. ANS Adv Nurs Sci 1991;14(2):32–41.

von Baeyer CL, Spagrud LJ. Review of observational or behavioral measures of pain for children and adolescents aged 3 to 18 years, University of Saskatchewan, Saskatchewan, Canada, 2005, Unpublished Manuscript.

von Baeyer C, Piira T. The faces pain scale-revised around the world: translation and adaptation for use in many cultures, 2004.

West N, Oakes L, Hinds P, Sanders L, Holden R, Williams S, et al. Measuring pain in pediatric oncology ICU patients. J Pediatr Oncol Nurs 1994;11(2):64–8.

Whaley L, Wong D. Nursing care of infants and children. 3rd ed. St. Louis: The C.V. Mosby Company; 1987. pp. 1070.

White J, Stow P. Rationale and experience with visual analogue toys. Anaesthesia 1985;40:601–3.

Wolf AR, Stoddart PA, Murphy PJ, Sasada M. Rapid skin anaesthesia using high velocity lignocaine particles: a prospective placebo controlled trial. Arch Dis Child 2002;86(4):309–12.

Wong D, Baker C. Pain in children: comparison of assessment scales. Pediatr Nurs 1988;14(1):9–17.

Wood C, von Baeyer CL, Bourrillon A, Dejos-Conant V, Clyti N, Abitbol V. Self-assessment of immediate post-vaccination pain after 2 different MMR vaccines administered as a second dose in 4- to 6-year-old children. Vaccine 2004;23(2):127–31.