Pain 127 (2007) 140–150

PAIN

www.elsevier.com/locate/pain

Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years

Carl L. von Baeyer [a,b,*], Lara J. Spagrud [a]

[a] Department of Psychology, University of Saskatchewan, Saskatoon, Canada S7N 5A5
[b] Department of Pediatrics, University of Saskatchewan, Saskatoon, Canada S7N 5A5

Received 26 April 2006; received in revised form 15 July 2006; accepted 11 August 2006

Abstract

Observational (behavioral) scales of pain for children aged 3 to 18 years were systematically reviewed to identify those recommended as outcome measures in clinical trials. This review was commissioned by the Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (www.immpact.org). In an extensive literature search, 20 observational pain scales were identified for review including behavior checklists, behavior rating scales, and global rating scales. These scales varied in their reliance on time sampling and inclusion of physiological items, facial and postural items, as well as their inclusion of multiple dimensions of assessment (e.g., pain and distress). Each measure was evaluated based on its reported psychometric properties and clinical utility. Scales were judged to be indicated for use in specific acute pain contexts rather than for general use. Two scales were recommended for assessing pain intensity associated with medical procedures and other brief painful events. Two scales were recommended for post-operative pain assessment, one for use in hospital and the other at home. Another scale was recommended for use in critical care. Finally, two scales were recommended for assessing pain-related distress or fear. No observational measures were recommended for assessing chronic or recurrent pain because the overt behavioral signs of chronic pain tend to habituate or dissipate as time passes, making them difficult to observe reliably. In conclusion, no single observational measure is broadly recommended for pain assessment across all contexts. Directions for further research and scale development are offered.

Keywords: Systematic review; Pain intensity; Observation; Behavior; Children; Adolescents

1. Introduction

Randomized clinical trials (RCTs) are important to investigate the efficacy and effectiveness of pain treatment for children and adolescents. The Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (Ped-IMMPACT; www.immpact.org) makes recommendations for core outcome domains and measures that should be considered by investigators conducting clinical trials for acute and chronic pediatric pain. In 2005, the Ped-IMMPACT group commissioned the present review of observational and behavioral pain measures for children between the ages of 3 and 18 years in clinical pain trials. A separate review was commissioned for self-report scales (Stinson et al., 2006). Pain measures must have well-established reliability and validity and must have been used to assess pain outcomes.

Three approaches to measuring pain have been identified in the literature (Walco et al., 2005): self-report; observational or behavioral; and physiological. Estimates of pain intensity based on each of these approaches are usually correlated only weakly to moderately with estimates based on the other approaches (Walco et al., 2005), suggesting that they may be measuring different constructs (pain experi-

* Corresponding author. Fax: +1 306 966 6630.
E-mail address: carl.vonbaeyer@usask.ca (C.L. von Baeyer).
URL: www.usask.ca/~vonbaeye/ (C.L. von Baeyer).

0304-3959/$32.00
doi:10.1016/j.pain.2006.08.014

ence, behavioral distress, and physiological arousal). It is considered desirable to obtain and rely most on self-report measures when these are available, since there is broad consensus that pain is primarily understood to be a subjective experience, as outlined in the definition of pain adopted by the International Association for the Study of Pain. However, the claim that self-report measures represent a 'gold standard' is overstated given the complexity and biases that may characterize self-reports of pain (Williams et al., 2000; Hodgins, 2002). Physiological changes in variables such as respiration rate and heart rate are only loosely correlated with painful events and may occur in response to many other states such as exertion or fever. Thus, observational measures of pain are needed for use with children who are (a) too young, e.g., below 4 years of age, to understand and use a self-report scale; (b) too distressed to use a self-report scale; (c) impaired in their cognitive or communicative abilities; (d) very restricted by bandages, surgical tape, mechanical ventilation, or paralyzing drugs; (e) whose self-report ratings are considered to be exaggerated, minimized, or unrealistic due to cognitive, emotional or situational factors.

The purpose of the present study was to systematically review published observational measures for pain in children aged 3 to 18. Reviewers sought to identify those measures that have well-established psychometric properties, as well as good clinical and research utility. Part of the evaluation of clinical utility was to examine the appropriateness of each measure within specific clinical contexts (e.g., post-operative pain). Emphasis was placed on reviewing and recommending measures of pain intensity, but behavioral measures of distress or fear were also examined as in some cases these overlapped with pain intensity scales. (Separate reviews are planned through Ped-IMMPACT of pain measures for infants and for children with neurological or developmental disabilities, as the types of tools and measurement issues are somewhat different for these populations.)

2. Methods

2.1. Methodological issues in reviewing observational measures of pain

Before outlining the way articles and measures were selected for this review, it is important to explain the methodology employed in the observational measures that are currently available for measurement of pain intensity, as well as methodological issues such as use of time sampling and inclusion of physiological indices. Each type of measure has advantages and disadvantages which were taken into account in assessing validity and recommending best measures for various contexts.

2.1.1. Types of observational or behavioral measures of pain

2.1.1.1. Behavior checklists. A behavior checklist provides a list of behaviors that are marked as either present (usually scored 1) or absent (usually scored 0) (Katz et al., 1980; Tarbell et al., 1992; Boelen-van der Loo et al., 1999; Chambers et al., 2003). The pain intensity score is defined as the number of items checked. The most common behavioral indices of pain in these scales include vocal, verbal, facial, postural, and motor behaviors. The instrument may or may not require observation for a specific period of time (see Section 2.1.3). No judgment of intensity or frequency of the behavior is made; each item is either absent or present. Pain intensity is assumed to be greater if a greater number of overt displays of pain is noted by the observer. One advantage of such measures is speed and ease of use: the observer simply makes a series of dichotomous judgments. However, in such scales all items are generally weighted equally which may not be justified: for example, facial expression may be a more reliable or more salient index of pain than leg movement yet both may be weighted equally in the checklist.

2.1.1.2. Behavior rating scales. Behavior rating scales incorporate a rating of the intensity, frequency, or duration of each behavior (McGrath et al., 1985; Ambuel et al., 1992; Gauvain-Piquard et al., 1999). The most frequently used rating for individual behaviors is 0 (absent) to 2 (intense or frequent), but many other metrics have been used. In some such instruments, the metric chosen for each behavior may deliberately reflect the weight placed on that behavior as an index of pain; in other instruments, all items are arbitrarily weighted equally. Similarly, the number of items reflecting a particular domain of behavior may be chosen either based on evidence-based weighting, or more commonly arbitrarily or based on the investigator's opinion. As with behavior checklists, the intensity of the pain is assumed to be reflected in the total pain intensity score defined as the sum of the individual ratings. The strength of this approach, compared with dichotomous checklists, is that it allows for gradations in intensity or frequency of expressions of pain without greatly adding to the burden of using such a tool.

2.1.1.3. Global rating scales. A global rating scale provides a rating of the observer's global impression of a patient's pain. Any metric and any tool can be used; for example, numerical rating scales (NRS), visual analog scales (VAS), and faces scales have all been used as the basis for global observational rating scales (Krane et al., 1987; Broome and Endsley, 1989; Carpenter, 1990; Humphrey et al., 1992; Schneider and LoBiondo-Wood, 1992; Tyler et al., 1993; Arts et al., 1994; Joyce et al., 1994; Calamandrei et al., 1996; Goodenough et al., 1997, 1998; Chambers et al., 1999, 2005; Gilbert et al., 1999; van Dijk et al., 2002; Voepel-Lewis et al., 2002; Bosenberg et al., 2003).

Care should be taken to distinguish these observational applications of pain scales from self-report implementation of the same scales. Global pain ratings can be provided by nurses, parents, researchers, and other observers. Some global rating scales have only the bottom (no pain) and top (maximum pain) anchors defined; others have intermediate anchors such as mild, moderate and severe pain. However,

if the instrument lists specific behaviors expected at each level of pain, it is called a *behaviorally anchored* rating scale (see next section). Global rating scales for pediatric pain intensity have been subjected to very little research in their own right; users seem to rely purely on face validity. Yet global rating scales are very commonly used as the criterion for validation of more complex pain measurement scales. This raises a question, discussed further below, as to whether such validation works both ways; i.e., whether the global rating scales are themselves validated by their correlation with more complex and more established rating scales.

A major difficulty with global rating scales is that they may be subject to many observer biases, in the absence of any specific or objective criteria for pain ratings. For example, if a child is seen to be given a painful stimulus such as a needle, then a global rating of pain may be higher than if the stimulus is not observed. Also, if the same observer is using both the global and the "molecular" (specific, detailed) observation scales, impressions gleaned from use of the "molecular" tool will be reflected in the global rating, possibly producing an inflated correlation between the global and molecular scales. Global scales may also be based on different responses in different children or in different situations. For example, for one child a specific rating could be due to crying whereas in another child the same rating could occur in the absence of crying and be the rater's response to a facial grimace. The relationship of the observer to the patient may also affect the accuracy of the assessment. An observer who is familiar with the child's normal behaviors may be better able to identify aberrant or idiosyncratic pain-related behaviors than a clinician less familiar with the child; on the other hand, children are often more expressive in the presence of parents than strangers (von Baeyer and Spagrud, 2003).

A recent review of global observational visual analog scales (van Dijk et al., 2002) concluded that there is not yet sufficient support for their use and that further research is needed to choose between them and more specific behavioral pain measures. No global observational scales were recommended as outcome measures in the present review.

A conundrum is regularly encountered in reviewing non-global behavioral pain scales in that many of these scales are validated primarily by showing moderate to high correlations with global rating scales (by nurses, parents, or other observers; in many studies the same individuals provide both scores). There is no basis for treating those global rating scales as an independent gold standard. In view of the biases affecting unanchored global observational rating scales, less reliance is placed on correlations with global scales than on evidence of responsiveness to pain-producing and pain-relieving interventions.

2.1.1.4. Global behaviorally anchored rating scales. A global behaviorally anchored rating scale lists levels of pain, e.g., 0 to 5, and for some or all of these levels it provides examples of behaviors that are likely (but are not required) to be seen at that level. It is distinguished from behavior checklists and behavior rating scales in that the behaviors listed are examples, each of which is not required to justify a rating at that level. For example, a child in severe pain may be thrashing or may be very still; both of these behaviors may be listed as examples without requiring either one. These scales rely on the observer's judgment to select and weight the behaviors observed and to translate these observations into a rating. Two such scales have been published for pediatric pain measurement for normally developing children 3 years of age and older (Humphrey et al., 1992; Peden et al., 2003). An advantage of this approach is that it allows for behaviors which are particularly salient in a particular child to be weighted highly, though they may occur rarely in other individuals or even be absent on pain behavior checklists. For example, a child who is normally active and sociable but becomes still when in severe pain could be identified on such a scale, whereas most behavior checklists would not detect significant pain in such a case. The cost may be reduced reliability as each observer may have a different basis for making a rating. There is not yet enough research on these methods either to recommend or discourage their use.

2.1.2. Role of facial expression

Considerable evidence points to a special role for facial expression in measurement of pain (Craig, 1992). Most behavioral checklists and rating scales include items referring to the face. Some of these offer vague descriptors apparently based on a lay understanding of facial expression of pain, while others are more specific and detailed. For example, the FLACC (Voepel-Lewis et al., 2002) includes a Face item ranging from 0 = "No particular expression or smile" to 2 = "Frequent to constant quivering chin, clenched jaw." The Child Facial Coding System (Gilbert et al., 1999) has been developed for more detailed measurement of facial expressions; it requires technical training using videotaped assessments to ensure standardized measurement. The required training increases the burden of administration of this measure compared with simpler instruments. In the case of children aged 3 to 18 years, there is little evidence as yet that such detailed measurement provides gains in information commensurate to the increased burden; detailed facial action scores tend to correlate well with global observational measures of pain intensity (e.g., Gilbert et al., 1999).

2.1.3. Time sampling

The instructions for administration of some scales call for specific periods of observation, e.g., observe for 5 s, record for 25 s, repeat the cycle three more times, compute mean score across the four samples (Beyer et al., 1990). This procedure increases reliability and reduces the chance of biasing a pain rating by waiting until a pain behavior occurs and then using that isolated event as the basis of the pain rating. However, literature reporting the use of such scales often does not mention whether the original time sampling procedure was followed. For example, the CHEOPS, originally designed to be used with a specific time sampling schedule, has been applied in both very brief single observations and in lengthier ongoing periods of observation.

2.1.4. Inclusion of physiological items

Some scales include heart rate, blood pressure, respiration rate, vomiting, and other items which may or may not reflect pain. These might require special equipment to observe. Correlations of physiological items with other observed behavior and with self-report are, in general, not well established (McGrath, 1998; Büttner and Finke, 2000; Walco et al., 2005). An elevated heart rate, for example, could indicate pain

but it could also be associated with exercise, medications, anxiety, excitement, or fever. There is also evidence to suggest that physiological indicators are less sensitive to changes in distress following an intervention as compared to subjective and behavioral measures (Walco et al., 2005).

2.1.5. Unidimensional versus multidimensional scales

A unidimensional scale is one that is demonstrated to measure only one construct, e.g., pain intensity. A multidimensional scale is one that simultaneously measures different constructs, whether or not it actually contains separate scales for each of these. For example, a scale may be promoted or applied as a unidimensional measure of pain intensity, but if it includes items more closely correlated with anxiety, depression, nausea, or other variables, then it would be low in internal consistency and the claim for its being unidimensional would be questionable. Some scale authors have simply assumed that their scales were internally consistent measures of a single construct, pain intensity; other scale authors have tested and reported the extent to which the scale is uni- versus multidimensional. Yet other authors (e.g., Gauvain-Piquard et al., 1999) have carried out factor analyses showing a multifactorial structure but advocate the use of the single full-scale score. For the present review, and for most clinical trials, the ideal pain scale would be one which is *designed and demonstrated* to be a unidimensional measure of pain intensity. Other non-pain outcomes of pain trials are addressed in separate discussions as part of the IMMPACT process.

2.1.6. Discrimination of pain intensity from distress, unpleasantness and fear

Most observational pain scales include items that could be interpreted as indicators of pain; or of other negative emotions such as fear, anxiety, anger, or frustration; or of efforts to cope with fear and pain. Crying, whining, physical tension, clinging, restlessness, and seeking or avoiding touch may all be observed both in pain and in non-pain distress states.

Some authors, recognizing the difficulty of discriminating pain intensity from pain unpleasantness and from other emotions such as fear, have adopted less specific terms such as "distress" or even "quality of life" in place of "pain" in the title of their scale. Nevertheless such scales may be treated by other researchers as predominantly or purely pain scales, and such "non-pain" scales are generally neither more nor less responsive to pain-producing or pain-relieving interventions than are scales explicitly labeled as measures of pain. Few researchers have presented discriminant validity data showing that their observational scales can differentiate pain intensity from its affective aspect or from other negative emotional states and reactions. Efforts made in this direction for self-report scales show that most children under eight or nine years of age have difficulty in introspectively discriminating between the sensory experience of pain and the affective response (i.e., distress or fear) to painful sensations (e.g., Goodenough et al., 1999).

While it is difficult to select unidimensional pain measures on purely empirical grounds, examination of item content may lead to favoring some measures over others on face and content validity grounds. Scales that include affect or fear laden items such as "hostility", "panic", "need for restraint" and "stall" may be less desirable as pain outcome measures than scales that contain only behaviors selected as indices of pain intensity, namely vocalizations, verbal statements, facial expressions, posture, and physical movements.

2.1.7. Single-purpose versus generic scales

Some researchers have designed instruments for specific contexts, e.g., post-circumcision or post-tonsillectomy pain in children of a specific age. These instruments have then been used (with or without validation) with other age groups and in other clinical situations. On the other hand, some instruments are presented as generic, i.e., applying to pain from different sources in children of different ages and physical conditions.

Certain scales are designed to accommodate children who are restricted in their movement or vocal expression by sedating or paralyzing drugs, use of a ventilator, bandages, and other restraints, as would be common in critical care settings (Ambuel et al., 1992). This represents a distinct and important clinical context for pain measurement that is reflected in the recommendations.

2.2. Criteria for selection of measures

The process of recommending pain measurement scales is multidimensional, taking into account the following aspects of each tool as well as the context in which pain scores are to be obtained (e.g., critical care, ward care, procedure room, home). These criteria were adapted from previous work in the IMMPACT project (Dworkin et al., 2005).

2.2.1. Evidence base for responsiveness, reliability, and validity

Responsiveness was defined as a scale's ability to detect a significant change in pain scores in the expected direction in response to pain-relieving events (e.g., administration of analgesia) and/or pain-producing events (e.g., painful procedures). Reliability was assessed by internal consistency in multi-item scales; rarely, in single-item scales, it was assessed by retest over short time intervals. Content and construct validity were assessed by appropriateness of content (see next Section). Criterion validity was assessed via the relationship with scores provided by different observers and by self-report, using the same or different instruments. A set of evaluation criteria for the assessment of *quality of evidence* for IMMPACT reviews has been established (Cohen et al., 2006). These criteria and their operational definitions are presented in Table 1.

2.2.2. Predicted non-differences

Clinical trials showing differences in pain between different analgesic conditions (responsiveness) often contribute powerful evidence for the predictive validity of pain outcome measures. On the other hand, many clinical trials entail comparison of a new treatment with an older, established treatment. To be successful, the new treatment should have some advantage over the existing therapy. It should, for example, provide a more rapid onset or longer duration of effect, or have milder side effects, or cost less than the comparator. Non-inferiority studies to demonstrate cost benefits are only valid if the assays used are sufficiently sensitive to detect differences and the sample size is adequate to detect a meaningful difference between treatments. Studies showing no significant difference between analgesic conditions provide no evidence

Table 1
Evaluation criteria for IMMPACT reviews (Cohen et al., 2006)

	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 1peer-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.

for responsiveness. For some scales, unfortunately, the majority of published studies are of this type and therefore, although they were reviewed, they did not contribute one way or another to the evidence base evaluated in the current review.

2.2.3. Aspects of pain addressed by the scale and appropriateness of content

As the primary aim of this project was to identify and recommend measures of *pain intensity*, this aspect of pain was emphasized. Observational instruments may also address duration of pain, time of pain onset, time of relief onset, frequency of pain of a certain severity, pain affect or distress, pain-related fear, functional interference, pain-related quality of life, and a variety of other variables. The recommended instruments focused on pain intensity.

2.2.4. Burden

Burden was assessed by the time taken to complete the instrument, need for equipment, and need for extensive training to administer. On this basis, for example, a scale requiring videotape of the child, a lengthy scoring process, and a high level of training for the observers was ruled out.

2.2.5. Availability of norms

A desirable feature of a scale is that data and conventions are available to interpret scores, e.g., as low, moderate, or high in severity, or severe enough to require intervention. For most observational scales, no such data are available. The scores cannot be assumed without evidence to be interpretable using the usual categories of severity.

2.2.6. Age range

The range of ages for which the measure is appropriate was taken into account. In some cases, scales were designed for a certain age but then applied without evaluation to other ages. The criterion for inclusion in this review was the availability of evidence for valid use with children aged 3 to 18.

2.3. Search strategy for identification of studies

All peer-reviewed publications that referred to the use of documented observational or behavioral pain scales with children aged 3 to 18 years were identified for consideration. Studies were primarily identified through online searches using *MEDLINE* (1966 – March Week 3, 2006), *PsycINFO* (1967–March week 4 2006) and *Web of Science* (1900-March 2006) databases. The date of the last search attempt was March, 2006. While no dates were specified in the literature searches, the articles reviewed ranged from 1979 to 2006. The search terms included subjects, text words, and keywords relevant to the following terms: 'pain measurement,' 'assessment,' 'pain,' 'observation,' 'pediatric,' as well as the specific names and acronyms of source measures and their primary authors. Limits were set to include only human populations, English language articles, peer-reviewed articles, and preschool, childhood, and adolescent populations (i.e., ages 3 to 18). Reference lists of previous reviews were also examined (McGrath, 1996; Merkel and Malviya, 2000; McGrath and Gillespie, 2001; Royal College of Nursing, 2001); Gaffney et al., 2002).

Abstracts, unpublished manuscripts, reviews, guidelines, commentaries and other descriptive articles were excluded. Many studies were carried out with children crossing the age boundaries, e.g., birth to 12 years, without reporting how many children were in the target range of 3 years and up; these studies were kept for review. Studies that incorporated modified versions of original observational measures were not included in the review because they contain idiosyncratic features that make it difficult to compare them to original published measures.

2.4. Strategy for selection of articles for review

Following examination of the scale characteristics discussed above, promising scales were kept for review (listed in Tables 2 and 3) and a search was conducted for articles reporting data on the scales. The goal that guided selection of articles for this review was to obtain enough citing articles to be able to reliably evaluate the level of evidence for each observational scale. When there were fewer than 50 citing articles for a particular scale, all of the citing articles were selected for review. When the number of citing articles exceeded 50, the goal was to obtain at least 50% of the citing articles in order to achieve a reliable evaluation of the level of evidence for each measure. A total of 129 articles were retained for review. All of the citing articles were found and reviewed for 13 of the 20 source measures. For the other seven measures at least 50% of the citing

Table 2
Scales recommended by intended context of measurement, with source, age of child for which each tool is intended, metric, rationale, and level of evidence

Recommended context of measurement	Acronym Name of tool	First author (year)	Age range[a]	Metric	Comments	Level of evidence
Procedural pain; brief painful events	FLACC Face, Legs, Arms, Cry, Consolability	Merkel et al. (1997)	I: 4–18 years S: 0–18 years	0–10: 5 items scored 0 to 2	Uses items similar to well-established CHEOPS but with a readily understood 0–10 metric. Low burden. Excellent inter-rater reliability. Moderate concurrent validity with FACES and good with VAS. Inconsistent responsiveness data. Has been used in studies of post-operative pain, minor non-invasive procedures, ear-nose–throat operations	I
	CHEOPS Children's Hospital of Eastern Ontario Pain Scale	McGrath et al. (1985)	I:1–7 years S: 4 months–17 years	4–13: 6 items scored 0 to 3	Well-established reliability and validity in many studies. Scores range from 4 to 13, with scores 4–6 indicating no pain Good indications of inter-rater and test–retest reliability. Good evidence for construct and concurrent validity, and responsiveness. Has been used in studies of general surgery; myringotomy and ear tube insertion; bladder nerve stimulation; closed fracture reduction; intravenous cannulation; sickle cell episodes; circumcision, and immunizations	I
Post-operative pain in hospital	FLACC Face, Legs, Arms, Cry, Consolability	Merkel et al. (1997)	I: 4–18 years S: 0–18 years	0–10: 5 items scored 0 to 2	See above	I
Post-operative pain at home (parent assessment)	PPPM Parents' Post-operative Pain Measure	Chambers et al. (1996)	I: 2–12 years S:1–12 years	0–15: 15 items scored 0 or 1	Well-established assessment. High inter-rater reliability and internal consistency. Good construct validity with the FPS, sensitivity, specificity, content validity. Good responsiveness data. Has been used in studies of post-operative pain (many kinds) and hernia repair	I
On ventilator or in critical care	COMFORT COMFORT Scale	Ambuel et al. (1992)	I: Newborn–17 years S: Newborn–17 years	8–40: 8 items scored 1 to 5	Only validated instrument available for this purpose. Good inter-rater reliability and internal consistency. Inconsistent responsiveness data. Has been used in studies of heart surgery; switching position to improve oxygenation; medical ventilation	II
Distress; pain-related fear or anxiety (not necessarily pain intensity; may be observed before as well as after a painful procedure)	PBCL Procedure Behavior Check List	LeBaron and Zeltzer (1984)	I: 6–17 years S: 0.1 year–19 years	Original 8–40: 8 items scored 1 to 5. Various revisions.	Good inter-rater reliability. Good construct validity and responsiveness data. Has been used in studies of bone marrow aspirations, lumbar punctures, radiation therapy, and immunization. Contains 1 unusual item	II+ (as measure of pain)
	PBRS-R Procedure Behavioral Rating Scale – Revised	Katz et al. (1980)	I: 8 months–17 years S: 3 years–10 years	0–11: 11 items scored 0 or 1	Good inter-rater, inter-item reliability. More investigation of validity and responsiveness is needed. Has been used in studies of bone marrow aspirations, immunizations and venipuncture	III+ (as measure of pain)

For level of evidence, see Table 1 and Sections 2.6 and 2.7.
[a] I = intended age range when the scale was first published; S = age range studies in subsequent research.

Table 3
Scales reviewed but not recommended at this time as outcome measures for clinical trials, listed in alphabetic order by acronym

Acronym Tool	Author (year)	Level of evidence
AHTPS Alder Hey Triage Pain Score	Stewart et al. (2004)	III
BOT Behavioral Observation Tool	Hester (1979)	–
CHIPPS Children's and Infants' Post-operative Pain Scale	Büttner and Finke (2000a)	II or III
DEGRr Douleur Enfant Gustave Roussy (and Revised version)	Gauvain-Piquard et al. (1987)	III
DPC Derbyshire Children's Hospital Paediatric Pain Chart	Peden et al. (2003)	II
GDS Groningen Distress Scale	Humphrey et al. (1992)	II
NAPI Nursing Assessment of Pain Intensity	Stevens (1990)	III
OCDS Observed Child Distress Scale	Bournaki (1997)	III
OPS Objective Pain Scale	Hannallah et al. (1987)	II+
OSBD Observational Scale of Behavioral Distress	Jay et al. (1983) Revised: Elliott et al. (1987)	II+
PMH-PAT Princess Margaret Hospital Pain Assessment Tool	Robertson (1993)	–
POCIS Pain Observation Scale for Young Children	Boelen-van der Loo et al. (1999)	–
PQL Pediatric Observational Quality of Life Measure	Myatt and Myatt (1998)	III
TPPPS Toddler Preschooler Post-operative Pain Scale	Tarbell et al. (1992)	II

For level of evidence, see Table 1.

articles were reviewed. Of the latter seven measures, three were evaluated at the highest level of evidence based on strong psychometric and clinical utility data. The remaining four measures were evaluated as having either psychometric or clinical utility problems that prohibited their being recommended at this time.

2.5. Review procedure

A team of five reviewers was assembled, comprising the authors and three research assistants. Each reviewer was responsible for a set of observational measures and for reviewing all of the selected articles associated with each one. A systematic approach to data extraction was accomplished by implementing standard review sheets to summarize each of the selected articles. After all of the selected articles for a given scale were reviewed, the data on its corresponding review sheets were aggregated and recorded on a summary sheet. Using these summary sheets, the authors independently evaluated each of the measures using the criteria (Cohen et al., 2006; see Table 1). There was 100% agreement on the independent ratings of level of evidence by the two authors.

3. Results

3.1. Scales excluded from review

The following observational scales were excluded from review: six scales specifically designed for children with developmental disabilities (for which a separate Ped-IMMPACT review is planned); one scale designed for use only with arthritis patients that requires videotaping of a prescribed sequence of actions and therefore was considered to lack clinical utility for general use in clinical trials at present; one scale designed for use with everyday pain rather than clinical pain; one research instrument involving detailed facial coding that requires use of videotape and extensive training and therefore was considered to lack clinical utility at present; and one measure of adult–child interaction in which pain was coded only incidentally. In addition, numerous scales designed and tested exclusively with neonates, infants and toddlers younger than 3 years were excluded.

3.2. Scales included in review

Tables 2 and 3 list scales included in the review, with the name of the first author and year of original publication and the level of evidence supporting each scale. Scales recommended for use are identified in Table 2 and are discussed below. Other scales, reviewed but not recommended at this time, are listed in Table 3.

3.3. Recommendations for selection of scales

3.3.1. Pain associated with medical procedures and other brief painful events

Use of either the FLACC (Face, Legs, Activity, Cry, Consolability) or the CHEOPS (Children's Hospital of Eastern Ontario Pain Scale) is recommended. Both have been very extensively used and have excellent evidence of reliability, validity, and responsiveness. The FLACC comprises five items scored 0–2 as identified in the name of the scale. It is lower in burden than the CHEOPS, and the 0–10 FLACC scores are more readily interpretable. (Of course it cannot be assumed that the 0–10 scale of the FLACC is psychometrically equivalent to an ideal or self-report 0–10 scale; the scores representing mild, moderate and severe pain, the minimum clinically significant difference, and the equality of intervals between scores have not been well established.) On the CHEOPS, behaviors are scored 0 to 1, 2 or 3, allowing for differential weighting of pain behaviors. Scores from 4 to 6 all represent no pain, and the maximum score is 13. There is a risk of scores on the CHEOPS being erroneously interpreted as if they were on the more common 0–10 metric. An advantage of the CHEOPS is that it is based

only on the child's directly observable behavior, without requiring appraisal of efforts to 'console' the child (after the procedure) as seen in the FLACC 'consolability' rating. In any particular procedural pain context, the choice between the two instruments will depend on how important it is to use an instrument with low burden and a commonly understood metric (in which the FLACC has the advantage) or to avoid inferences about 'consolability' (in which the CHEOPS has the advantage).

3.3.2. Post-operative pain

The FLACC is recommended as the first choice for post-operative pain *in hospital*, as it was designed and validated in this clinical context over a broader age range than the CHEOPS. Moreover, the variable of 'consolability,' or response to supportive contact and distraction, may be generally more feasible and important to assess in post-operative care (which lasts for a longer period of time) than it is in brief procedural pain.

For post-operative pain *at home* following discharge from hospital, the PPPM (Parents' Post-operative Pain Measure) is recommended. The unique feature of the PPPM is that it was designed specifically for use by parents in the post-operative care of their children. The PPPM is a well-established assessment tool with high inter-rater reliability and internal consistency, and good indices of construct validity, sensitivity, specificity, and responsiveness. With 15 items scored 0 or 1 it is brief and low in burden for parents. The PPPM may be useful for other longer-term types of pain but that has not yet been established.

3.3.3. Pain in critical care

The COMFORT scale is recommended for pain in children in critical care as it is the only well-studied instrument that makes explicit accommodation for constraints placed on the behavioral expression of pain by mechanical ventilation and physical restraint. The inter-rater reliability and internal consistency of this scale are strong. It requires scoring 8 items from 1 to 5, for a total score from 8 to 40; it includes a requirement for comparison of blood pressure with baseline levels which may be problematic when no baseline is available. To score at the maximum, a child would have to be engaged in "vigorous movement, including torso and head," which might not be characteristic of a child whose response to pain is mainly guarding and rigidity. These limitations might require discussion in studies employing the COMFORT scale.

3.3.4. Pain-related distress, fear, or anxiety

There is a need in many clinical trials for a measure of distress, fear, or anxiety as well as pain intensity *per se*. This need arises because it is necessary to distinguish pain (which may be responsive to analgesics) from other kinds of behavioral distress (which would not be expected to respond to analgesics). Several relevant instruments have been available since 1980. These instruments could be used before as well as after a painful procedure, as they include indices of fear and anxiety as well as pain. The scales that have the best balance of evidence, burden, and content validity are the PBCL (Procedure Behavior Check List) and the PBRS-R (Procedure Behavioral Rating Scale–Revised). Both have good indices of reliability and responsiveness to anxiety-relieving interventions. The burden for the two scales is similar. More investigation into the psychometrics of these two scales is needed to achieve a higher level of evidence.

3.3.5. Chronic or recurrent pain

No observational instruments are recommended on the basis of presently available evidence for chronic or recurrent pain lasting weeks, months or years. This is primarily because the overt behavioral signs of pain in such cases tend to habituate or dissipate as time passes, despite continued self-reported pain. Early work is in progress on eliciting and measuring brief, movement-related exacerbations of pain behavior in longer-term pain (Jaworski et al., 1995; Dworkin et al., 2005) but this approach is not yet suitable for broad clinical application in a pediatric population. It is hoped that the criteria used for evaluation of scales in the present review will inform the development and validation of scales for recurrent and chronic pain. Limitations on measurement in this context are discussed below in Section 4.4 on 'Pain behavior in relation to duration of pain.'

4. Discussion

4.1. Recommendations for outcome measures in clinical trials

Observational pain scales for use as outcome measures in clinical trials were recommended according to context (Table 2) with separate measures for (a) procedural and other brief acute pain, (b) post-operative pain in hospital, (c) post-operative pain at home, and (d) critical care. In addition, (e) two measures of pain-related distress were identified that can be used before as well as after a painful event in order to capture distress as well as pain intensity.

A number of interesting and complex issues emerged from the review and these are discussed in the following sections.

4.2. Contextual and cultural factors

Some knowledge of the clinical context is needed to make sense of scores on observational scales. As McGrath (1998) points out, "... if a child has an elevated

temperature, flushed face, or rapid breathing, one needs to know the context in which the behavior arises before one determines its meaning. If the child has been lying in bed and feeling sick and has a stiff neck, one would draw different conclusions than if the child has just run up five flights of stairs'' (p. 94).

In order to control for variations in activity that might reduce or enhance pain expression, some investigators put patients through a series of standardized, timed activities (e.g., sitting, standing, walking) and measure pain behaviors such as guarding, bracing, rubbing, and flinching (Jaworski et al., 1995). Although promising for research use, these procedures are unlikely to be broadly applicable to measurement of post-operative, procedural, and disease-related pain, so they are not reviewed here.

Contextual influences on pain expression also occur in hospital, in interactions with health care providers. For example, a study in an orthopedic ward (Byrne et al., 2001) showed that some nurses actively discouraged children from displaying their pain: they frequently ''construed pain as unreal, unwarranted or not deserving help'' (p. 69). Children who did not complain of pain or ask for analgesics were described by many nurses as ''very good'' or ''great.'' Children who displayed a lot of pain were appraised as malign, unmotivated, not coping, or even as engaging in dramatic acting. The impact of these reactions on patients may be to alter their expressions of distress. Faced with a nurse's denial of his pain, a child may intensify his expression of pain in order to be believed, or to obtain analgesics, or to try to avoid a feared procedure. Conversely, an older child may stifle manifestations of pain to avoid negative response from the clinical staff. The prevailing culture, not only of the hospital ward or clinic but also of the family and community, can influence pain expression and hence observational measurement of pain. As yet there are no systematic measures of such influences, but they should be taken into account into conducting and interpreting observational measurements of pain. Such cultural influences may be particularly important to take into account in multinational trials where different modes of behavioral and social response to pain prevail in different sites.

4.3. Developmental variations in pain behavior and age-specificity of scales

Over the course of child development, systematic transformations can be observed in the way pain is expressed or communicated (von Baeyer and Spagrud, 2003). During the toddler and preschool years the expression of pain becomes shaped increasingly by the child's growing understanding of emotions and the ability to anticipate outcomes and feelings. About 50% of 3-year-olds cry before an injection, suggesting that because of their previous experience they anticipate and fear the pain of their imminent injection (Negayama, 1999). By preschool age, children are also developing an ability to feign, exaggerate or suppress outward signs of pain, if doing so carries some gain for the child (e.g., avoiding an injection or painful therapy, or getting out of bed).

The most rapid developmental changes in the way pain is expressed probably occur before age 3 years, a fact which is acknowledged in the organization of the Ped-IMMPACT consensus process in that it provides separate consideration of the 0 to 2 age span. At age 3 and beyond, there are continued changes in pain expression, e.g., less crying in older children given the same physical stimulus. These developmental changes in pain expression should in principle influence the process of pain measurement. However, pain intensity measures are not as yet age-normed, and there are no separate versions of observational scales for different ages. Most observational instruments were initially designed to accommodate a particular age range, but have been applied to a broader age range. For example, the CHEOPS (McGrath et al., 1985) was initially developed for age 1 to 7 years, but it has since been applied successfully in studies of children from 1 month through 17 years of age. It remains for research to determine whether there would be any gain in adjusting observational pain intensity scores by age, or in creating separate forms for different ages as is normal practice for broad-band child behavior scales (Behavior Assessment Scale for Children; Child Behavior Check List).

4.4. Pain behavior in relation to duration of pain

A strong painful stimulus elicits an immediate, robust, instinctual behavioral response including withdrawal, vocalization, and grimacing. That overt response attenuates rapidly (over the course of minutes or hours) in the face of continuous or chronic pain, and may be replaced by more covert responses accompanied by rigidity, silence, and guarding the affected body part. In long-term acute and chronic pain, an absence of visible signs of pain is common, except during exacerbations that occur when the patient moves or perceives a social cue, e.g., someone asking about the pain (von Baeyer and Spagrud, 2003). The methods for observational measurement of brief acute pain (e.g., procedural pain) are much better established than methods for measurement of long-lasting and chronic pain.

What signs of long-term or chronic pain are visible to an observer? Chronic pain, except during acute exacerbations, is likely to be manifested in complex changes such as increased irritability, low mood, difficulty with sleep, hostility, changes in appetite, and school performance, all of which require knowledge of the child's baseline condition and temperament. Of the

observational measures reviewed here, only the DEGR, DEGRr, and PPPM explicitly address these observable manifestations of longer-term pain.

4.5. Directions for further research

Several directions for further research were identified through the course of this review. There is need for the development and validation of observational measures of chronic or recurrent pain. These could be based on use of structured physical activity, instructions, or social cues to elicit signs of pain that might otherwise remain covert. It may also be possible to combine or integrate observational and self-report measures of pain for this purpose, although further investigation would be needed to develop and validate integrated pain measures. More consideration of the context of pain (i.e., physical, cultural, and social) is needed. In particular, researchers should seek to develop and validate systematic ways to account for the impact of context in pain scales. Researchers might also consider measurement of other aspects of pain (affective, evaluative, sensory quality) as additional outcome variables in clinical trials.

In relation to psychometric considerations, further investigation is warranted to validate the weighting of items within observational checklists. Should items all be weighted equally, as is done in most scales, or should a more complex weighting system be employed to reflect the varying salience of different behaviors in judging pain? A related issue is the level of measurement employed in scales (Stevens, 1946). Frequently scale developers assign an arbitrary numerical value to levels of behavior items (e.g., 0, 1, 2, 3) and then treat the resulting scale sum as an interval or ratio scale, which is not justified without further investigation. Burden could be minimized and utility maximized by use of readily understood metrics and short forms, and by retaining only those items that demonstrably contribute strongly to the full scale score. Researchers should also seek to determine whether age-adjusted scales, or separate forms for different ages, would improve the validity of pain intensity measurement.

This review has identified six observational pain scales that can currently be recommended on the basis of established criteria (Tables 1 and 2) for use in clinical trials. This field of measurement is, however, in its infancy. Before any scale can be regarded as a gold standard, further work on scale development and validation is needed as outlined above.

Acknowledgements

This project was commissioned and funded by the Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (Ped-IMMPACT). The assistance of the following people is gratefully acknowledged (in alphabetic order): Marieke Blom, Christine Chambers, Jacqueline Cummine, Darby Eakins, Sasha Forsyth, Tara Gokavi, Kelly Hayton, Patrick McGrath, Jane Scott, and Jennifer Stinson.

References

Ambuel B, Hamlett KW, Marx CM, Blumer JL. Assessing distress in pediatric intensive care environments: the COMFORT scale. J Pediatr Psychol 1992;17:95–109.

Arts SE, Abu-Saad HH, Champion GD, Crawford MR, Fisher RJ, Juniper KH, et al. Age-related response to lidocaine-prilocaine (EMLA) emulsion and effect of music distraction on the pain of intravenous cannulation. Pediatrics 1994;93:797–801.

Beyer JE, McGrath PJ, Berde CB. Discordance between self-report and behavioral pain measures in children aged 3–7 years after surgery. J Pain Symptom Manage 1990;5:350–6.

Boelen-van der Loo WJ, Scheffer E, de Haan RJ, de Groot CJ. Clinimetric evaluation of the pain observation scale for young children in children aged between 1 and 4 years after ear, nose, and throat surgery. J Dev Behav Pediatr 1999;20:222–7.

Bosenberg A, Thomas J, Lopez T, Kokinsky E, Larsson LE. Validation of a six-graded faces scale for evaluation of postoperative pain in children. Paediatr Anaesth 2003;13:708–13.

Bournaki MC. Correlates of pain-related responses to venipunctures in school-age children. Nurs Res 1997;46:147–54.

Broome ME, Endsley RC. Maternal presence, childrearing practices, and children's responses to an injection. Res Nurs Health 1989;12:229–35.

Büttner W, Finke W. Analysis of behavioural and physiological parameters for the assessment of postoperative analgesic demand in newborns, infants and young children: a comprehensive report on seven consecutive studies. Paediatr Anaesth 2000;10:303–18.

Byrne A, Morton J, Salmon P. Defending against patient's pain: A qualitative analysis of nurses's responses to children's pain. J Psychosom Res 2001;50:69–76.

Calamandrei M, Messeri A, Busoni P, Bernini G, Lippi A, Tucci F. Comparison of two application techniques of EMLA and pain assessment in pediatric oncology patients. Reg Anesth 1996;21:557–60.

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

Chambers CT, Finley GA, McGrath PJ, Walsh TM. The parents' postoperative pain measure: replication and extension to 2-6-year-old children. Pain 2003;105:437–43.

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

Chambers CT, Hardial J, Craig KD, Court C, Montgomery C. Faces Scales for the Measurement of Postoperative Pain Intensity in Children Following Minor Surgery. Clin J Pain 2005;21:277–85.

Chambers CT, Reid GJ, McGrath PJ, Finley GA. Development and preliminary validation of a postoperative pain measure for parents. Pain 1996;68:307–13.

Cohen LL, La Greca AM, Blount RL, Kazak AE, Holmbeck GN, Lemanek KL. Introduction: Evidence-based assessment in pediatric psychology. J Pediatr Psychol 2006; e-pub ahead of print; doi:10.1093/jpepsy/jsj115.

Craig KD. The facial expression of pain: Better than a thousand words. Am Pain Soc J 1992;1:153–62.

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

Elliott CH, Jay SM, Woody P. An observation scale for measuring children's distress during medical procedures. J Pediatr Psychol 1987;12:543–51.

Gaffney A, McGrath PJ, Dick B. Measuring pain in children: developmental and instrument issues. In: Schechter NL, Berde CB, Yaster M, editors. Pain in infants, children and adolescents. Baltimore, MD: Williams & Wilkins Lippincott; 2002. p. 128–41.

Gauvain-Piquard A, Rodary C, Rezvani A, Lemerle J. Pain in children aged 2-6 years: a new observational rating scale elaborated in a pediatric oncology unit–preliminary report. Pain 1987;31:177–88.

Gauvain-Piquard A, Rodary C, Rezvani A, Serbouti S. The development of the DEGR(R): A scale to assess pain in young children with cancer. Eur J Pain 1999;3:165–76.

Gilbert CA, Lilley CM, Craig KD, McGrath PJ, Court CA, Bennett SM, et al. Postoperative pain expression in preschool children: Validation of the child facial coding system. Clin J Pain 1999;15:192–200.

Goodenough B, Addicoat L, Champion GD, 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 other self-report and behavioral measures. Clin J Pain 1997;13:60–73.

Goodenough B, Champion GD, Laubreaux L, Tabah L, Kampel L. Needle pain severity in children: Does the relationship between self-report and observed behaviour vary as a function of age? Aust J Psychol 1998;50:1–9.

Goodenough B, Thomas W, Champion GD, Perrott D, Taplin JE, von Baeyer CL, 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.

Hannallah RS, Broadman LM, Belman AB, Abramowitz MD, Epstein BS. Comparison of caudal and ilioinguinal/iliohypogastric nerve blocks for control of post-orchiopexy pain in pediatric ambulatory surgery. Anesthesiology 1987;66:832–4.

Hester NO. The preoperational child's reaction to immunization. Nurs Res 1979:250–5.

Hodgins MJ. Interpreting the meaning of pain severity scores. Pain Res Manag 2002;7:192–8.

Humphrey GB, Boon CM, van Linden van den Heuvell GF, van de Wiel HB. The occurrence of high levels of acute behavioral distress in children and adolescents undergoing routine venipunctures. Pediatrics 1992;90:87–91.

Jaworski TM, Bradley LA, Heck LW, Roca A, Alarcon GS. Development of an observation method for assessing pain behaviors in children with juvenile rheumatoid arthritis. Arthritis Rheum 1995;38:1142–51.

Jay SM, Ozolins M, Elliott CH. Assessment of children's distress during painful medical procedures. Health Psychol 1983;2: 133–47.

Joyce BA, Schade JG, Keck JF, Gerkensmeyer J, Raftery T, Moser S, et al. Reliability and validity of preverbal pain assessment tools. Issues Compr Pediatr Nurs 1994;17:121–35.

Katz ER, Kellerman J, Siegel SE. Behavioral distress in children with cancer undergoing medical procedures: Developmental considerations. J Consult Clin Psychol 1980;48:356–65.

Krane EJ, Jacobson LE, Lynn AM, Parrot C, Tyler DC. Caudal morphine for postoperative analgesia in children: a comparison with caudal bupivacaine and intravenous morphine. Anesth Analg 1987;66:647–53.

LeBaron S, 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;52:729–38.

McGrath PA, Gillespie J. Pain assessment in children and adolescents. In: Turk DC, Melzack R, editors. Handbook of pain assessment. New York: Guilford Press; 2001. p. 97–118.

McGrath PJ. There is more to pain measurement in children than ''ouch. Can Psychol 1996;27:63–75.

McGrath PJ. Behavioral measures of pain. In: Finley GA, McGrath PJ, editors. Measurement of Pain in Infants and Children. Progress in Pain Reseach and Management 1998;Vol. 10. Seattle, WA: IASP Press; 1998. p. 83–102.

McGrath PJ, Johnson G, Goodman JT, Schillinger J, Dunn J, Chapman J. CHEOPS: A behavioral scale for rating postoperative pain in children. In: Fields HL, Dubner R, Cervero F, editors. Advances in pain research and therapy. vol. 9. New York: Raven Press. 1985. pp. 395–402.

Merkel S, Malviya S. Pediatric pain, tools, and assessment. J Perianesth Nurs 2000;15:408–14.

Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S. The FLACC: A behavioral scale for scoring postoperative pain in young children. Pediatr Nurs 1997;23:293–7.

Myatt HM, Myatt RA. The development of a paediatric quality of life questionnaire to measure post-operative pain following tonsillectomy. Int J Pediatr Otorhinolaryngol 1998;44:115–23.

Negayama K. Development of reactions to pain of inoculation in children and their mothers. Int J Behav Dev 1999;23:731–46.

Peden V, Vater M, Choonara I. Validating the Derbyshire Children's Hospital Pain Tool: A pilot study. Paediatr Anaesth 2003;13:109–13.

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

Royal College of Nursing. Clinical Practice Guidelines. The recognition and assessment of acute pain in children. Technical Report. 2001. www.rcn.org.uk/publications/pdf/guidelines/cpg_contents.pdf (Available July 15, 2006).

Schneider EM, LoBiondo-Wood G. Perceptions of procedural pain: parents, nurses, and children. Child Health Care 1992;21:157–62.

Stevens B. Development and testing of a pediatric pain management sheet. Pediatr Nurs 1990;16:543–8.

Stevens SS. On the theory of scales of measurement. Science 1946;103:677–80.

Stewart B, Lancaster G, Lawson J, Williams K, Daly J. Validation of the Alder Hey Triage Pain Score. Arch Dis Child 2004;89:625–30.

Stinson JN, Kavanagh T, Yamada J, Gill N, Stevens B. Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. Pain 2006; e-pub ahead of print; doi:10.1016/j.pain.2006.05.006.

Tarbell SE, Cohen IT, Marsh JL. The Toddler-Preschooler Postoperative Pain Scale: an observational scale for measuring postoperative pain in children aged 1–5. Preliminary report.. Pain 1992;50:273–80.

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

van Dijk M, Koot HM, Saad HH, Tibboel D, Passchier J. Observational visual analog scale in pediatric pain assessment: useful tool or good riddance? Clin J Pain 2002;18:310–6.

Voepel-Lewis T, Merkel S, Tait AR, Trzcinka A, Malviya S. The reliability and validity of the Face, Legs, Activity, Cry, Consolability observational tool as a measure of pain in children with cognitive impairment. Anesth Analg 2002;95:1224–9.

von Baeyer CL, Spagrud LJ. Social development and pain in children. In: McGrath PJ, Finley GA, editors. The context of pediatric pain: biology, family, culture. Seattle, WA: IASP Press; 2003. p. 8–97.

Walco GA, Conte PM, Labay LE, Engel R, Zeltzer LK. Procedural distress in children with cancer: Self-report, behavioral observations, and physiological parameters. Clin J Pain 2005;21:484–90.

Williams AC, Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. Pain 2000;85:457–63.