Do pain measurement instruments detect the effect of pain-reducing interventions in neonates? A systematic review on responsiveness

Do pain measurement instruments detect the effect of pain-reducing interventions in neonates? A systematic review on responsiveness

Naomi Meesters[a], Tinne Dilles[b], Sinno Simons[a], Monique van Dijk[a,c]

Running title: Review on responsiveness to pain in neonates

Affiliations: [a]Department of Pediatrics, Division of Neonatology, Erasmus MC –Sophia Children's Hospital, Rotterdam, the Netherlands; [b]Nursing and Midwifery Sciences, Centre for Research and Innovation in Care, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; and [c]Intensive Care and Department of Pediatric Surgery, Erasmus MC –Sophia Children's Hospital, Rotterdam, the Netherlands

Address correspondence to:

Naomi Meesters Erasmus MC - Sophia Children's Hospital.
Department of Pediatrics, Division of Neonatology
Wytemaweg 80, 3015 CN Rotterdam, the Netherlands.
Phone: 0031-10-7040704
n.meesters@erasmusmc.nl

Research funding: None

Conflicts of interest: None

Abstract

The effectiveness of pain-reducing interventions in newborns can only be determined if pain measurement instruments are responsive; i.e. able to detect a decrease in pain intensity after the pain-reducing intervention. This review assesses the methodological quality of studies on this measurement property – the responsiveness. We searched the literature published until January 2018 for validation studies of pain measurement instruments focusing on responsiveness to pain-reducing treatment in neonates. Methodological quality of the included studies was rated using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. Nine studies were included involving 10 pain measurement instruments. These studies differed with respect to the population, setting and type of pain-reducing intervention. In all studies, pain scores were significantly lower after a pain-reducing intervention and the instrument used was therefore considered responsive. We rated four studies as having poor methodological quality, five as fair quality and none as good quality. In conclusion, the responsiveness was studied for only ten of the 43 existing pain measurement instruments for the use in neonates. As this is an important property of a pain instrument, more research on this topic is needed, with attention for blinding and formulating a specific hypothesis before start of data collection.

Perspective: This review focuses on the property of measurement instruments to detect changes in pain intensity after a pain-reducing intervention in neonates. We concluded that this property – the responsiveness – is under studied and that the methodological quality of the included studies was low. Future high-quality validation studies should focus on responsiveness.

Keywords: Pain assessment; responsiveness; sensitivity to change; neonates; psychometrics

Introduction

Worldwide, more than 15 million children a year are born prematurely and this number is rising [43]. In the NICU setting they are at risk of acute pain; one study reported a mean number of 11.4 painful procedures per day during the first 14 day [35]. They are also at risk of prolonged or chronic pain from, for example, skin conditions, gastro-intestinal conditions, inflammation and the aftermath of surgery [2; 3].

From the moment it became clear that neonates are capable of experiencing pain - in the late 1980s – some 40 pain measurement instruments for the use in neonates have been developed [9]. While the Premature Infant Pain Profile (PIPP) is one of the best known instruments to measure acute pain, the Échelle Douleur Inconfort Nouveau-Né (EDIN) and COMFORTneo are two of the instruments developed to measure prolonged pain [12; 36; 41]. The American Association of Pediatrics recommends to consistently use validated instruments to assess the need for pain management throughout a neonate's hospitalization [25]. The quality of such an instrument is largely determined by its clinimetric properties [9; 13]. One of these properties is responsiveness. Its definition varies in the literature; in some sources it is defined as the ability to detect a change in general, also called "sensitivity to change" [38]. In this definition, the change could be of any magnitude. Other definitions add that this change should be clinically important [38]. A third group of definitions consider an instrument responsive only if score change reflects a real change in the construct to be measured [38].

The Consensus-based Standard for the selection of health Measurement Instruments (COSMIN) guidelines adhere to the latter definition: "the ability of an instrument to detect change over time in the construct to be measured" [11]. Because there is no gold standard for the measurement of pain in neonates [33], a construct approach is recommended according to these guidelines and hypothesis testing is the correct way to determine responsiveness [11]. The hypotheses should be based on the

expected mean difference between the changes in scores before and after an intervention in groups or the expected correlation between the changes in scores of the studied instrument compared to an instrument of which the responsiveness has been established.

Pain-reducing interventions are necessary to prevent detrimental effects on brain development and developmental outcomes [16; 34; 40; 42]. Analgesia or sedation could be necessary to adequately treat prolonged pain. A European cohort study showed that this was the case for 34% of the NICU patients [7]. The consistent use of a pain measurement instrument is important to guide pain-reducing treatment – either pharmacological or non-pharmacological. and the instrument therefore should be able to detect real changes in pain intensity. With respect to the measurement of pain in infants, the responsiveness of an instrument is based on either the infant's reaction to a painful intervention or the infant's reaction to a pain-reducing intervention [10]. It is clinically important to evaluate patients' pain intensity and the appropriate pain-reducing intervention should be decided based on this evaluation.

Because NICU patients are frequently exposed to pain and because neonates respond to pain differently than do older infants [18], we wanted to know if the current pain measurement instruments are validly measuring the effect of pain-reducing interventions in neonates. We therefore performed a systematic review of the literature, focusing on this measurement property.

Methods

Inclusion- and exclusion criteria:

Eligible for inclusion were original validation studies evaluating the responsiveness of pain measurement instruments in neonates, defined as newborns up to a postnatal age of 28 days. If the postnatal age was not specifically described, articles using the concept "neonates" were also considered eligible for

inclusion. A study was included only if the study population at least partially consisted of neonates. Studies in which interventions were pharmacologic, non-pharmacologic or both were included.

The definition of responsiveness for this review was adapted from the COSMIN definition as follows: "The ability of an instrument to detect changes in pain intensity over time after a pain-reducing intervention". The selected studies did not always apply this definition but did evaluate responsiveness. All studies had a longitudinal aspect and evaluated the change in pain score before and after a pain-reducing intervention within the same patient. Because efficacy trials designed to evaluate the effect of interventions with the use of validated pain measurement instruments serve a different purpose from that of validation studies, we excluded these trials.

Only validation studies of which a full text version was published were included. The pain had to be assessed in a hospital. The year of publication was considered to be irrelevant for this research question; therefore there were no restrictions with respect to the publication date. Articles published in other languages than Dutch or English were excluded. The English-language abstracts of articles published in other languages than Dutch or English were screened to determine their possible eligibility.

Search strategy

The electronic bibliographic databases EMBASE, MEDLINE, Web of Science, Scopus, Cochrane, Cinahl, PsychINFO and google scholar were searched until January 2018 by a biomedical information specialist (WB) from the Erasmus University Medical Center Medical Library.

Search terms used were combinations of the following words:

- Neonates: "newborn" or "neonat*" or " premature" or "preverbal" or "0-year-old" or "young children" or "baby" or "low birthweight"
- Validation: "valid*" or "sensitivity to change" or "responsiveness" or "psychometric*"

- Pain: “pain” or “distress” or “comfort”
- Measurement: “scale” or “score” or “measurement” or “profile”

Complete search strategies are included in appendix 1.

We also searched the literature to determine the number of instruments published until January 2018. We included the instruments mentioned in the systematic reviews on pain assessment in neonates of Cong et al., Stevens et al. and Duhn et al. [9; 13; 37] that met our inclusion criteria, as well as the ones additionally found in the database for this review. Furthermore, we hand-searched the reference lists of the articles for additional relevant articles.

Selection of articles

References of the different databases were combined and duplicates were removed. Two reviewers (MvD and NM) selected potentially eligible articles based on title and abstract. Articles of which the title and abstract made clear that the study did not meet the inclusion criteria were excluded. The full texts of the remaining studies, if available, were read and assessed on eligibility: first, whether the study population met the inclusion criteria and second, whether a change in the pain score before and after a pain-reducing intervention was analyzed. Also, the reference lists of these studies were manually searched to identify other potentially eligible studies. If the two independent reviewers did not agree on inclusion, a third reviewer (CT) was consulted and whose verdict was accepted.

Data extraction

Information on the measurement instrument, study department, study population and intervention was extracted for each study. The specific method used to calculate responsiveness and the number of measurements, the time interval between the pain reducing intervention and the assessment of pain, the (blinding of the) observers and the statistical analysis were described. Methodological quality was

the primary outcome and described together with the results of the studies, as the degree to which pain scores were lower after the intervention served as secondary outcome.

Quality assessment

The COSMIN group, an international multidisciplinary team of experts, reached consensus on taxonomy, terminology and definitions of measurement properties following an international Delphi study in 2010 [31]. Three different quality domains are described, namely reliability, validity and responsiveness. The accompanying critical appraisal tool, the COSMIN checklist, guides the evaluation of the methodological quality of studies on measurement properties [30]. For systematic reviews, a scoring system has been developed to obtain an overall quality rating of the different COSMIN items [39]. Four response options (excellent, good, fair, poor) are defined for each item. The lowest rating of any item determines the quality score for each measurement property. For the evaluation of the included studies in this review we used only the items referring to the quality of determining the responsiveness. These items and the accompanying scoring system have been described by Mokkink et al [30]. The total scores before and after a pain reducing intervention were compared using Cohen's d. If the mean and standard deviation were absent, we took the median to represent the mean and calculated the standard deviation by multiplying the IQR by 0.741. The effect size was considered small at Cohens' d of 0.20, medium at 0.50, and large at 0.80 [8].

Results

The search strategy yielded 1915 citations. After having removed duplicates, we screened the titles and abstracts of 811 citations on eligibility. Thirty-seven articles were written in another language then English or Dutch, but all those studies would also have been excluded based on the English abstract.

Thirty-three full texts were read, of which nine validation studies met all the inclusion and exclusion criteria and were included in this review (figure 1). After having consulted a third reviewer (CT) for two studies, the two reviewers agreed on all studies. One study was excluded because no data was provided about the change in scores [19]. Checking the reference lists did not result in additional studies to include.

Study description

Next to the 35 instruments that were included in the three reviews and met our inclusion criteria reviews [9; 13; 37], the search strategy yielded another 8 instruments (see appendix 2). The responsiveness was evaluated for 10 of these 43 instruments (Figure 2). Table 1 gives an overview of the nine included studies that studied the responsiveness of 10 different pain measurement instruments. Five of these studies included only neonates [12; 22-24; 41], of which the EDIN study only included premature infants [12]. The four other studies included not only neonates but also older children [6; 15; 26; 32]. One study did not describe the postnatal age range but was not excluded because one quarter of the children were younger than 0.1 years [6]. Four studies only focused on the postoperative period [22-24; 32]. The EDIN study concerned preterm born neonates diagnosed with respiratory distress syndrome and assigned an EDIN score of 7 or higher before the pain-reducing intervention [12]. In the four other studies, the source of the pain was not mentioned [6; 15; 26; 41].

In all but one study the intervention consisted of pharmacological analgesic treatment. Only the study on the responsiveness of the COMFORTneo scale also included non-pharmacological interventions in the analysis [41]. Dosing of the analgesics was described in only two studies, on the EDIN and MAPS [12; 32]. In the remaining six studies the evaluated intervention was the administration of analgesics and/or sedatives, but dosing was not specified [6; 15; 22-24; 26].

Study quality

Table 2 describes the methods used for determining the responsiveness of the pain measurement instrument, the results, and the quality score according to the COSMIN checklist. The quality of five studies was rated as poor for the following reasons.

In three of these studies the sample size was smaller than 30 [22; 24; 32], and thus are of poor methodological quality according to the COSMIN checklist [39]. In two of these studies the required power had not been calculated [22; 24]; in the MAPS study, the sample size was calculated for the interrater reliability instead of the responsiveness [32]. Regarding the LIDS, only the p-value of the difference was reported and no additional information was given [22]. The article on the COMFORTneo did not describe the time between the two assessments and for another pain instrument, the EVENDOL, the statistical analysis used was not clearly described[15; 41].

The quality of the four other studies was rated as fair. Figure 3 shows the complete assessment of all nine studies. Because of the lack of a gold standard, items 15 to 17 need not be assessed [24].

Missing items

Six studies only used paired pain assessments, both before and after the intervention, and did not describe missing assessments [12; 22; 23; 26; 32; 41]. Items 1 and 2 of the COSMIN checklist were considered not applicable for these studies.

In the study on the COMFORT behavior scale, the pain scores were retrieved retrospectively and the authors specified the number of lacking re-assessments [6]. In the EVENDOL study the number of pain scores after the intervention was lower than that before the intervention in two different situations (in rest and during mobilization), but the reason was not described [15]. The authors of the study on the OPS and CRIES stated they included all infants admitted to the NICU and that each was assessed hourly

with both pain scales; the reason why there were more OPS-scores (N=77) than CRIES-scores (N=74) is unclear [24].

Hypothesis formulation

All studies hypothesized that the mean group score after a pain-reducing intervention would be lower than that before the intervention. In two studies the size of the expected absolute or relative difference was predefined [6; 23]. For the NPASS, the power analysis was based on an expected difference of 1.2 points, but the authors did not support this with evidence [23]. An expected difference of five points for the COMFORT behavior scale was based on a previous study and the authors also aimed for a COMFORT behavior score below 17, which is taken to be the cutoff value for treatment [6].

Blinding

Only in the LIDS study, pain scores were assigned bedside by nurses who were not aware of the situation – before or after the intervention [22].

Timing

Responsiveness was based on two [6; 12; 23; 24; 41] or more [15; 22; 26; 32] pain scores per patient. The time interval between the before and after assessments varied between studies, from one hour [23; 24] to eight hours [12]. In three studies with multiple pain scores, the time interval between assessments was pre-defined but increased between the assessments [22; 26; 32]. The EVENDOL was scored two times in two different situations, whereby the time interval was specified [15].

Analysis

All nine studies tested the significance of the difference between the score before and after a pain-reducing intervention. In the EVENDOL study it was not described which test was used [15], the eight

other studies used a paired t-test, one-way ANOVA or the equivalent non-parametric tests. In six studies either the mean difference per patient was calculated [12; 24] or mean scores before and after the intervention were presented [15; 23; 26; 41]. The studies on the MAPS and COMFORT behavior scale presented also the number of scores above (MAPS) or below (COMFORT behavior) a certain value [6; 32]. Regarding the LIDS, only the p-value and actual scores are presented [22].

Responsiveness

All studies found significant decreases of the scores after the intervention. With respect to the a priori formulated hypotheses, after the intervention 63% of the COMFORT behavior scores had decreased at least 5 points and 74% of the pain scores were below 17 [6]. While the mean difference for the N-PASS was not calculated in the study, the difference between the scores before and after the intervention was 3.0 points [23]. It was concluded for all included pain measurement instruments that they were responsive.

Table 3 shows the calculated Cohen's d for eight measurement instruments. For seven instruments the difference between before and after intervention scores was large and for one instrument (LIDS) medium. It was not possible to calculate the effect size for the CRIES and OPS, because only the mean decline was presented.

Discussion

Despite the enormous increase in neonatal pain and pain assessment research, very little is still known about the responsiveness of pain measurement instruments. This review evaluated the property of neonatal pain measurements to detect changes after a pain-reducing intervention as well as the quality of the studies evaluating this measurement property. While the responsiveness of the neonatal pain

measurement instruments included in this review proved satisfactory, the quality of the studies in question as judged by the COSMIN checklist was only poor to fair. Responsiveness to treatment is an important clinimetric property of pain measurement instruments in general [1; 9]. This review points out some important issues to improve future research on this measurement property.

Poor quality of the validation studies was primarily due to the lack of clearly formulated a priori hypotheses. Because no high quality evidence on the responsiveness of pain measurement instruments in neonates was found, formulating hypotheses in future studies on the correlation between change in scores on another instrument is currently not appropriate. Hypotheses should be based on the expected mean difference between the changes in scores before and after an intervention. Only in the study on the COMFORT behavior scale the expected mean difference was explicitly stated in the hypothesis[6]. In the N-PASS study, this was reflected in the power analysis [23]. Another option would be to aim for a score below the cut-off score for pain, as in the study on the COMFORT behavior scale [6]. In that case it is also necessary to determine the expected size of the change, for example expressed as the percentage of scores below this value after the intervention. None of the included studies formulated hypotheses on the expected effect size or the mean difference between two groups whose change scores are expected to differ in change.

With respect to the eight instruments in the current study for which Cohen's d could be calculated, only for the LIDS the Cohen's d was below 0.80. However, according to the COSMIN checklist, this calculation should always be compared with that stated in a priori hypothesis.

Seven of the included studies determined the responsiveness by calculating p-values for the differences between the scores without the formulation of any hypothesis [12; 15; 22; 24; 26; 32; 41]. This means that only the statistical significance of the difference is calculated, but not the relevance of the change score for clinical practice. According to the COSMIN checklist, this is not a correct method if not

combined with a beforehand formulated hypothesis. Furthermore, the p-value is also dependent on the standard deviation of the change and the sample size.

Another problem is the risk of bias when the assessor is not blinded to the situation – before or after the intervention. Blinding is notably difficult in this type of study design. It seems impossible to blind assessors and adhere to a fixed time interval after the intervention if pain is assessed at the bedside. Still, in one study assessors could be blinded for the situation [22]. In this study the time interval between the intervention and pain assessment varied greatly, however. A possible solution for blinding is having assessors assign pain scores afterwards when watching video recordings of the child [6]. These recordings will be randomized for the before and after the intervention situation and will typically not reveal any patient identification . This practice may also result in more consistent and objective pain scores [5].

For the N-PASS, on the other hand, the authors deliberately chose not to do this because their goal was to develop a measurement instrument usable in daily practice [23]. Moreover, it is not easy to obtain high-quality footage of premature neonates lying in an incubator, especially those with the lowest gestational age: it is relatively dark, the incubator's surface is reflective, the face is partly covered by fixation material for respiratory support and sometimes a pacifier. Three previous studies used videotapes to validate pain measurement instruments in extremely premature neonates [20; 21; 28]. In two of these, however, these neonates were filmed at a postconceptional age of 32 weeks [20; 21] and the third only included neonates whose face was well visible because they were not intubated or on non-invasive ventilation [28]. Pain needs to be assessed in NICU patients of all ages and in all circumstances. Therefore, the question remains if it is possible to reliably assess pain using videotapes in all premature neonates.

The evaluation of the responsiveness of all the included instruments focused on prolonged pain treated primarily by analgesics and sedatives. The definition of responsiveness suggests that a change over time must occur. With respect to a change in pain after pain reducing interventions, this definition seems to be inseparably linked to a prolonged aspect of pain. Validation studies that focus on acute pain measurement instruments, such as the PIPP [36], often evaluate the responsiveness by comparing the pain score during a painful intervention to the baseline pain score or a score during a non-painful intervention. These studies were excluded because the concept of responsiveness to pain differs from the concept of responsiveness to a pain reducing intervention.

In general, the importance of evaluating responsiveness in clinimetric studies only became more clear after the publication of the COSMIN guidelines in 2010 [31]. Before this publication, there was no consensus regarding the definition of responsiveness and the appropriate methods to evaluate this property. With respect to neonatal pain measurement instruments, most of the studies included in this review were published before 2010. This could explain why this property was not evaluated for most instruments . However, only one of the five instruments that were developed after 2010 evaluated the responsiveness in their validation studies. We hypothesize that this measurement property might be considered difficult to study in the neonatal population. Still, t~~T~~his is not the first review that found that studies on the responsiveness of pain measurement instruments lack quality. A systematic review on the ~~responsiveness~~ clinimetric properties of the FLACC in children also concluded that 18 out of the 20 included studies had a poor or fair methodological quality with respect to responsiveness [10].

In 2011, Angst criticized the COSMIN initiative and suggested it should be possible to use more traditional ways to determine the responsiveness [4]. He raised the question of how to deal with existing literature if the methods used in those studies are considered inappropriate. The authors of the COSMIN checklist responded that the guidelines are not meant to question previous studies, but rather

to help researchers to improve studies on measurement properties [29]. We believe this is also how the results of the current review should be interpreted. Next to this, NICU personnel should keep in mind that high-quality research on the responsiveness of pain measurement instruments is lacking. They use these instruments in clinical practice and should always consult other professionals if they feel an assessment does not correspond to their clinical judgement.

In the past few years research efforts have focused on more objective measures to assess pain in neonates, such as heart rate variability,[14] skin conductance,[9; 27] and "brain-oriented" approaches such as near-infrared spectroscopy (NIRS) and electroencephalography (EEG) [9; 17; 27]. A multimodal approach, possibly automated, has been suggested to improve pain assessment [44; 45].
By comparing noxious-evoked activity during a venipuncture with topical local anesthetic to a venipuncture without this anesthetic, Hartley et al. concluded that their EEG template was sensitive to analgesic modulation [17]. It is important that future validation studies on these physiological pain indicators also include the responsiveness. More evidence is needed before these indicators can be reliably implemented in clinical practice [9].

It is possible that unpublished data is available on other pain measurement instruments, resulting in publication bias. As none of the measurement instruments was addressed more than once, data pooling was not possible. Also, the included studies are heterogeneous with a large variation in study population, setting and type of intervention, which prevents comparison of the measure of responsiveness of the included pain measurement instruments. Finally, this review concentrates on only one measurement property and therefore is not suitable for determining the optimal measurement instrument to assess pain in neonates.

We may have missed relevant publications in other languages than English and Dutch, although none of the English-language abstracts of studies in other languages met our inclusion and exclusion criteria.

Furthermore, we excluded all studies not designed to evaluate the clinimetric properties. These included efficacy trials in which the effect of a pain reducing intervention in neonates was evaluated with the use of a pain measurement instrument. For these trails, the instrument's responsiveness should have been ascertained first. The studies on the COMFORT behavior, FLACC, MAPS and EVENDOL also included children beyond the neonatal age. In these studies, subgroup analysis of the data on neonates was not performed; it is not known, therefore, whether these instruments are more or less responsive in this specific population.

There is an urgent need for pain measurement instruments that could tell whether a pain-reducing intervention has the desired effect in newborns. This is the first review that describes the quality of validation studies on neonatal pain measurement instruments that included responsiveness as a measurement property. We found that for only 10 out of the 43 existing instruments the change in pain intensity in neonates after a pain-reducing intervention has been studied. Methodological quality of the studies in question was found to be poor to fair. Therefore, and because alternative methods to assess pain in clinical practice are lacking, more research on this topic is needed, with more attention to blinding and formulating a specific hypothesis before start of data collection.

Acknowledgements: We would like to thank Wichor Bramer, biomedical information specialist from the Medical Library of Erasmus University, Rotterdam for assisting in the searches. Ko Hagoort is thanked for editorial assistance. Also, we would like to thank Caroline Terwee from the VU University Medical Center, Amsterdam as a COSMIN expert for being the third reviewer for the articles we were not sure to include.

References

[1] Abu-Saad HH, Bours GJ, Stevens B, Hamers JP: Assessment of pain in the neonate. Semin Perinatol 22:402-416 1998.

[2] Anand KJS: Defining pain in newborns: need for a uniform taxonomy? Acta Paediatr 106:1438-1444 2017.

[3] Anand KJS, Eriksson M, Boyle EM, Avila-Alvarez A, Andersen RD, Sarafidis K, Polkki T, Matos C, Lago P, Papadouri T, Attard-Montalto S, Ilmoja ML, Simons S, Tameliene R, van Overmeire B, Berger A, Dobrzanska A, Schroth M, Bergqvist L, Courtois E, Rousseau J, Carbajal R, Europain survey working group of the NeoOpioid Consortium: Assessment of continuous pain in newborns admitted to NICUs in 18 European countries. Acta Paediatr 106:1248-1259 2017.

[4] Angst F: The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Med Res Methodol 11:152 2011.

[5] Black RE, Lord S, Wright IMR: The premature infant pain profile (PIPP) scores-who does them counts. J Paediatr Child Health 50:71 2014.

[6] Boerlage AA, Ista E, Duivenvoorden HJ, de Wildt SN, Tibboel D, van Dijk M: The COMFORT behaviour scale detects clinically meaningful effects of analgesic and sedative treatment. Eur J Pain 19:473-479 2015.

[7] Carbajal R, Eriksson M, Courtois E, Boyle E, Avila-Alvarez A, Andersen RD, Sarafidis K, Polkki T, Matos C, Lago P, Papadouri T, Montalto SA, Ilmoja ML, Simons S, Tameliene R, van Overmeire B, Berger A, Dobrzanska A, Schroth M, Bergqvist L, Lagercrantz H, Anand KJ, Europain Survey Working Group: Sedation and analgesia practices in neonatal intensive care units (EUROPAIN): results from a prospective cohort study. Lancet Respir Med 3:796-812 2015.

[8] Cohen J. CHAPTER 2 - The t Test for Means. In: Statistical Power Analysis for the Behavioral Sciences (Revised Edition). Academic Press, 1977, pp. 19-74.

[9] Cong X, McGrath JM, Cusson RM, Zhang D: Pain assessment and measurement in neonates: an updated review. Adv Neonatal Care 13:379-395 2013.

[10] Crellin DJ, Harrison D, Santamaria N, Babl FE: Systematic review of the Face, Legs, Activity, Cry and Consolability scale for assessing pain in infants and children: is it reliable, valid, and feasible for use? Pain 156:2132-2151 2015.

[11] de Vet HC, Terwee CB, Mokkink LB, Knol DL: Measurement in Medicine. New York, Cambridge University Press, 2011.

[12] Debillon T, Zupan V, Ravault N, Magny J, Dehan M: Development and initial validation of the EDIN scale, a new tool for assessing prolonged pain in preterm infants. Arch Dis Child Fetal Neonat Ed 85:F36-41 2001.

[13] Duhn LJ, Medves JM: A systematic integrative review of infant pain assessment tools. Adv Neonatal Care 4:126-140 2004.

[14] Faye PM, De Jonckheere J, Logier R, Kuissi E, Jeanne M, Rakza T, Storme L: Newborn infant pain assessment using heart rate variability analysis. Clin J Pain 26:777-782 2010.

[15] Fournier-Charriere E, Tourniaire B, Carbajal R, Cimerman P, Lassauge F, Ricard C, Reiter F, Turquin P, Lombart B, Letierce A, Falissard B: EVENDOL, a new behavioral pain scale for children ages 0 to 7 years in the emergency department: design and validation. Pain 153:1573-1582 2012.

[16] Grunau RE: Neonatal pain in very preterm infants: long-term effects on brain, neurodevelopment and pain reactivity. Rambam Maimonides Med J 4:e0025 2013.

[17] Hartley C, Duff EP, Green G, Mellado GS, Worley A, Rogers R, Slater R: Nociceptive brain activity as a measure of analgesic efficacy in infants. Sci Transl Med 9 2017.

[18] Hatfield LA: Neonatal pain: What's age got to do with it? Surg Neurol Int 5:S479-489 2014.
[19] Hodgkinson K, Bear M, Thorn J, Van Blaricum S: Measuring pain in neonates: evaluating an instrument and developing a common language. Aust J Adv Nurs 12:17-22 1994.
[20] Holsti L, Grunau RE: Initial validation of the Behavioral Indicators of Infant Pain (BIIP). Pain 132:264-272 2007.
[21] Holsti L, Grunau RE, Oberlander TF, Osiovich H: Is it painful or not? Discriminant validity of the Behavioral Indicators of Infant Pain (BIIP) scale. Clin J Pain 24:83-88 2008.
[22] Horgan MF, Glenn S, Choonara I: Further development of the Liverpool Infant Distress Scale. J Child Health Care 6:96-106 2002.
[23] Hummel P, Puchalski M, Creech SD, Weiss MG: Clinical reliability and validity of the N-PASS: Neonatal pain, agitation and sedation scale with prolonged pain. J Perinatol 28:55-60 2008.
[24] Krechel SW, Bildner J: CRIES: a new neonatal postoperative pain measurement score. Initial testing of validity and reliability. Paediatr Anaesth 5:53-61 1995.
[25] Lim Y, Godambe S: Prevention and management of procedural pain in the neonate: an update, American Academy of Pediatrics, 2016. Arch Dis Child Educ Pract Ed 102:254-256 2017.
[26] Manworren RC, Hynan LS: Clinical validation of FLACC: preverbal patient pain scale. Pediatr Nurs 29:140-146 2003.
[27] Maxwell LG, Malavolta CP, Fraga MV: Assessment of pain in the neonate. Clin Perinatol 40:457-469 2013.
[28] Milesi C, Cambonie G, Jacquot A, Barbotte E, Mesnage R, Masson F, Pidoux O, Ferragu F, Thevenot P, Mariette JB, Picaud JC: Validation of a neonatal pain scale adapted to the new practices in caring for preterm newborns. Arch Dis Child Fetal Neonatal Ed 95:F263-266 2010.
[29] Mokkink LB, Terwee CB, Knol DL, Vet dHCW: The new COSMIN guidelines regarding responsiveness. Author's response. BMC Med Res Methodol 11:152 2011.
[30] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC: The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 19:539-549 2010.
[31] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 63:737-745 2010.
[32] Ramelet A-S, Rees NW, McDonald S, Bulsara MK, Abu-Saad HH: Clinical validation of the multidimensional assessment of pain scale. Pediatric Anesthesia 17:1156-1165 2007.
[33] Ranger M, Johnston CC, Anand KJ: Current controversies regarding pain assessment in neonates. Semin Perinatol 31:283-288 2007.
[34] Ranger M, Zwicker JG, Chau CM, Park MT, Chakravarthy MM, Poskitt K, Miller SP, Bjornson BH, Tam EW, Chau V, Synnes AR, Grunau RE: Neonatal Pain and Infection Relate to Smaller Cerebellum in Very Preterm Children at School Age. J Pediatr 167:292-298 e291 2015.
[35] Roofthooft DW, Simons SH, Anand KJ, Tibboel D, van Dijk M: Eight years later, are we still hurting newborn infants? Neonatology 105:218-226 2014.
[36] Stevens B, Johnston C, Taddio A, Gibbins S, Yamada J: The premature infant pain profile: evaluation 13 years after development. Clin J Pain 26:813-830 2010.
[37] Stevens B, Pillai Riddel R, Oberlanderr T, Gibbins S. Assessment of pain in neonates and infants. In: Pain in Neonates and Infants. United Kingdom. Elsevier Health Sciences, 2007.
[38] Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM: On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 12:349-362 2003.

[39] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC: Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 21:651-657 2012.
[40] Valeri BO, Holsti L, Linhares MB: Neonatal pain and developmental outcomes in children born preterm: a systematic review. Clin J Pain 31:355-362 2015.
[41] van Dijk M, Roofthooft DW, Anand KJ, Guldemond F, de Graaf J, Simons S, de Jager Y, van Goudoever JB, Tibboel D: Taking up the challenge of measuring prolonged pain in (premature) neonates: the COMFORTneo scale seems promising. Clin J Pain 25:607-616 2009.
[42] Vinall J, Grunau RE: Impact of repeated procedural pain-related stress in infants born very preterm. Pediatr Res 75:584-587 2014.
[43] World Health Organization: Premature Birth (Fact sheet). Available at: http://www.who.int/news-room/fact-sheets/detail/preterm-birth. Accessed May 31, 2018
[44] Worley A, Fabrizi L, Boyd S, Slater R: Multi-modal pain measurements in infants. J Neurosci Methods 205:252-257 2012.
[45] Zamzmi G, Pai CY, Goldgof D, Kasturi R, Ashmeade T, Sun Y. An approach for automated multimodal analysis of infants' pain, Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 4148-4153, 2016.

Figure 1 **Flowchart**

Figure 2 **Evaluation of responsiveness in neonatal pain measurement instruments**

Figure 3 **COSMIN checklist quality scores for responsiveness**

1 Excellent; 2 Good; 3 Fair; 4 Poor; ◯ Not applicable

Items 1 to 14: see study Mokkink et al. for complete description [30]

Short description:

1:	Percentage missing items		8:	Hypotheses formulated
2:	Handling missing items		9:	Direction of change in hypotheses
3:	Sample size		10:	Absolute or relative magnitude of change
4:	Longitudinal design		11:	Description comparator instrument
5:	Time interval		12:	Properties comparator instrument
6:	Description intervention		13:	Important flaws design or method
7:	Change in patients		14:	Adequate design and statistical methods

Color should not be used for any images in the print version.

Appendix 1

Embase.com

('pain measurement'/exp OR 'pain assessment'/exp OR ((pain/exp OR comfort/de) AND ('rating scale'/exp OR 'scoring system'/exp OR 'clinical assessment tool'/exp OR 'facial expression'/exp OR 'monitoring device'/de)) OR ((pain* OR comfort*) NEAR/6 (measure* OR assess* OR scale* OR Profile OR validit* OR score* OR quantif* OR rating OR intensit* OR monitor*)):ab,ti) AND (newborn/de OR prematurity/exp OR 'newborn intensive care'/exp OR 'low birth weight'/exp OR 'newborn period'/exp OR 'gestational age'/exp OR 'newborn care'/exp OR 'newborn disease'/de OR 'newborn surgery'/de OR baby/de OR (newborn* OR (new* NEXT/1 born*) OR neonat* OR (prematur* NEAR/3 (infant* OR child*)) OR preterm* OR nicu OR nicus OR 'low birth weight' OR 'low birthweight' OR lbw OR elbw OR vlbw OR 'gestational age' OR baby OR babies OR (0 NEAR/3 year* NEXT/1 old*) OR (first* NEXT/2 month* NEXT/2 life)):ab,ti) AND ('validation study'/exp OR 'validation process'/exp OR validity/exp OR psychometry/de OR 'internal consistency'/de OR (validat* OR validit* OR psychometr* OR responsiv* OR (sensitivit* NEAR/3 change*)):ab,ti)

Medline ovid

("pain measurement"/ OR ((exp pain/) AND ("facial expression"/)) OR ((pain* OR comfort*) ADJ6 (measure* OR assess* OR scale* OR Profile OR validit* OR score* OR quantif* OR rating OR intensit* OR monitor*)).ab,ti.) AND (exp "Infant, Newborn"/ OR "Intensive Care Units, Neonatal"/ OR "Intensive Care, Neonatal"/ OR "gestational age"/ OR (newborn* OR (new* ADJ born*) OR neonat* OR (prematur* ADJ3 (infant* OR child*)) OR preterm* OR nicu OR nicus OR "low birth weight" OR "low birthweight" OR lbw OR elbw OR vlbw OR "gestational age" OR baby OR babies OR (0 ADJ3 year* ADJ old*) OR (first* ADJ2 month* ADJ2 life)).ab,ti.) AND ("validation study"/ OR "Validation Studies as Topic"/ OR Psychometrics/ OR (validat* OR validit* OR psychometr* OR responsiv* OR (sensitivit* ADJ3 change*)).ab,ti.)

psycINFO ovid

("pain measurement"/ OR ((exp pain/) AND ("facial expressions"/ OR "Rating Scales"/)) OR ((pain* OR comfort*) ADJ6 (measure* OR assess* OR scale* OR Profile OR validit* OR score* OR quantif* OR rating OR intensit* OR monitor*)).ab,ti.) AND (120.ag. OR "Neonatal Intensive Care"/ OR (newborn* OR (new* ADJ born*) OR neonat* OR (prematur* ADJ3 (infant* OR child*)) OR preterm* OR nicu OR nicus OR "low birth weight" OR "low birthweight" OR lbw OR elbw OR vlbw OR "gestational age" OR baby OR babies OR (0 ADJ3 year* ADJ old*) OR (first* ADJ2 month* ADJ2 life)).ab,ti.) AND ("Test Validity"/ OR Psychometrics/ OR (validat* OR validit* OR psychometr* OR responsiv* OR (sensitivit* ADJ3 change*)).ab,ti.)

Cinahl ebsco

(MH "pain measurement+" OR ((MH pain+) AND (MH "facial expression+")) OR ((pain* OR comfort*) N5 (measure* OR assess* OR scale* OR Profile OR validit* OR score* OR quantif* OR rating OR intensit* OR monitor*))) AND (MH "Infant, Newborn+" OR MH "Intensive Care Units, Neonatal+" OR MH "Intensive Care, Neonatal+" OR MH "gestational age+" OR (newborn* OR (new* N1 born*) OR neonat* OR (prematur* N2 (infant* OR child*)) OR preterm* OR nicu OR nicus OR "low birth weight" OR "low birthweight" OR lbw OR elbw OR vlbw OR "gestational age" OR baby OR babies OR (0 N2 year* N1 old*) OR (first* N2 month* N2 life))) AND (MH "validation studies+" OR MH Psychometrics+ OR (validat* OR validit* OR psychometr* OR responsiv* OR (sensitivit* N2 change*)))

Cochrane

(((pain* OR comfort*) NEAR/6 (measure* OR assess* OR scale* OR Profile OR validit* OR score* OR quantif* OR rating OR intensit* OR monitor*)):ab,ti) AND ((newborn* OR (new* NEXT/1 born*) OR neonat* OR (prematur* NEAR/3 (infant* OR child*)) OR preterm* OR nicu OR nicus OR 'low birth

weight' OR 'low birthweight' OR lbw OR elbw OR vlbw OR 'gestational age' ORoreiorien baby OR babies OR (0 NEAR/3 year* NEXT/1 old*) OR (first* NEXT/2 month* NEXT/2 life)):ab,ti) AND ((validat* OR validit* OR psychometr* OR responsiv* OR (sensitivit* NEAR/3 change*)):ab,ti)

Web-of-science

TS=((((pain* OR comfort*) NEAR/5 (measure* OR assess* OR scale* OR Profile OR validit* OR score* OR quantif* OR rating OR intensit* OR monitor*))) AND ((newborn* OR (new* NEAR/1 born*) OR neonat* OR (prematur* NEAR/2 (infant* OR child*)) OR preterm* OR nicu OR nicus OR "low birth weight" OR "low birthweight" OR lbw OR elbw OR vlbw OR "gestational age" OR baby OR babies OR (0 NEAR/2 year* NEAR/1 old*) OR (first* NEAR/2 month* NEAR/2 life))) AND ((validat* OR validit* OR psychometr* OR responsiv* OR (sensitivit* NEAR/2 change*))))

Scopus

TITLE-ABS-KEY((((pain* OR comfort*) W/5 (measure* OR assess* OR scale* OR Profile OR validit* OR score* OR quantif* OR rating OR intensit* OR monitor*))) AND ((newborn* OR (new* W/1 born*) OR neonat* OR (prematur* W/2 (infant* OR child*)) OR preterm* OR nicu OR nicus OR "low birth weight" OR "low birthweight" OR lbw OR elbw OR vlbw OR "gestational age" OR baby OR babies OR (0 W/2 year* W/1 old*) OR (first* W/2 month* W/2 life))) AND ((validat* OR validit* OR psychometr* OR responsiv* OR (sensitivit* W/2 change*))))

Pubmed publisher

("pain measurement"[mh] OR ((pain[mh]) AND ("facial expression"[mh])) OR ((pain*[tiab] OR comfort*[tiab]) AND (measure*[tiab] OR assess*[tiab] OR scale*[tiab] OR Profile OR validit*[tiab] OR score*[tiab] OR quantif*[tiab] OR rating OR intensit*[tiab] OR monitor*[tiab]))) AND ("Infant, Newborn"[mh] OR "Intensive Care Units, Neonatal"[mh] OR "Intensive Care, Neonatal"[mh] OR

"gestational age"[mh] OR (newborn*[tiab] OR (new born*[tiab] OR newly born*[tiab]) OR neonat*[tiab] OR (prematur*[tiab] AND (infant*[tiab] OR child*[tiab])) OR preterm*[tiab] OR nicu OR nicus OR "low birth weight" OR "low birthweight" OR lbw OR elbw OR vlbw OR "gestational age" OR baby OR babies)) AND ("validation study"[mh] OR "Validation Studies as Topic"[mh] OR Psychometrics[mh] OR (validat*[tiab] OR validit*[tiab] OR psychometr*[tiab] OR responsiv*[tiab] OR (sensitivit*[tiab] AND change*[tiab]))) AND publisher[sb]

Google scholar

"pain|comfort measurement|assessment|scale|validity|score"

newborn|neonate|neonatal|prematurity infant|infants|nicu|nicus

validation|validity|psychometry|responsiveness|responsivity|"sensitivity to change"

Appendix 2

Year	Author	Scale	Type*	Resp**	Comments
1987	Grunau	NFCS	1	○	
1990	Stevens	NAPI	1	○	Validated in neonates by Schade et al., 1996
1993	Craig	IBCS	1	○	
1993	Lawrence	NIPS	1	○	
1994	Hodgkinson	PAT	1	○	
1995	Krechel	CRIES	1	●	
1996	Sparshot	DSVNI	1	○	
1996	Horgan	LIDS	1	●	
1996	Stevens	PIPP	1	○	
1997	Carbajal	DAN or APN	1	○	
1998	Blauer	SUN	1	○	
2001	Debillon	EDIN	1	●	
2002	Hudson-Barr	PAIN	1	○	
2003	Marceau	NNICUPAT	1	○	
2004	Cignacco	BPSN	1	○	
2007	Bellieni	ABC	1	○	
2007	Holsti	BIIP	1	○	
2008	Hummel	NPASS	1	●	
2009	van Dijk	COMFORTneo	1	●	
2010	Hand	COVERS	1	○	
2010	Milesi	FANS	1	○	
2012	Liaw	PASPI	1	○	
2014	Lundquist	ALPS	1	○	
2014	Pollki	NIAPAS	1	○	
2014	Gibbins	PIPP-R	1	○	
1987	Hannalah	OPS	2	●	Validated in neonates by Krechel et al., 1995
1989	Barrier	POPS or CSS	2	○	Validated in neonates by Schade et al., 1996
1992	Ambuel	COMFORT	2	○	
1996	Schade	RIPS	2	○	Validated in neonates by Schade et al., 1996
1997	Merkel	FLACC	2	●	Validated in neonates by Manworren et al., 2003
2000	Buttner	CHIPPS	2	○	
2004	Suominen	CAAS	2	○	
2005	Ista	COMFORT-B	2	●	
2007	Ramelet	MAPS	2	●	
2012	Fournier-Charriere	EVENDOL	2	●	
1993	Bell	NPAS	3	○	
1994	Pokela	BPS	3	○	
1995	Friedrichs	NPAT	3	○	
1998	Guinsberg	MPCS	3	○	
1989	Johnson	PAIN	4	○	
1994	Lin	IPEC	4	○	
1994	Wielenga	WOPP	4	○	
1999	Jorgensen	PIPA	4	○	

* Type of pain measurement instrument

1: Validated neonatal pain measurement instrument
2: Infant/pediatric pain measurement instrument, validated in neonates
3: Pain measurement instrument used in neonates, not validated
4: Pain measurement instrument used in neonates, no full text published

** Responsiveness determined
○ No
● Yes