



Published in final edited form as:

Pain. 2007 December 5; 132(3): 264–272.

Initial Validation of the Behavioral Indicators of Infant Pain (BIIP)

Liisa Holsti, PhD^{*,†,‡,§} and Ruth E. Grunau, PhD^{*,†,§}

^{*}*Centre for Community Child Health Research, Child and Family Research Institute, Vancouver, Canada*

[†]*School of Rehabilitation Sciences, University of British Columbia, Vancouver, Canada*

[‡]*Department of Pediatrics, University of British Columbia, Vancouver, Canada*

[§]*Children's and Women's Health Centre of British Columbia, Vancouver, Canada*

Abstract

Accurate pain assessment in preterm infants in the neonatal intensive care unit (NICU) is complex. Infants who are born at early gestational ages (GA), and who have had greater early pain exposure, have dampened facial responses which may lead to under-treatment. Since behavioral and physiological responses to pain in infants are often dissociated, using multidimensional scales which combine these indicators into a single score may limit our ability to determine the effects of interventions on each system. Our aim was to design a unidimensional scale which would combine the relatively most specific, individual, behavioral indicators for assessing acute pain in this population.

The Behavioral Indicators of Infant Pain (BIIP) combines sleep/wake states, 5 facial actions and 2 hand actions. Ninety two infants born between 23-32 weeks GA were assessed during 3, one minute Phases of blood collection. Outcome measures included changes in BIIP and in Neonatal Infant Pain Scale (NIPS) scores coded in real time from continuous bedside video recordings; changes in heart rate (HR) were obtained using custom physiological processing software.

Scores on the BIIP changed significantly across Phases of blood collection ($p < 0.01$). Internal consistency (0.82) and inter-rater reliability (0.80-0.92) were high. Correlations between the BIIP and NIPS were modest ($r = 0.64$, $p < 0.01$) as were correlations between the BIIP and mean heart rate ($r = 0.45$, $p < 0.01$). In this initial study, the BIIP has been shown to be a reliable, valid scale for assessing acute pain in preterm infants in the NICU.

Keywords

premature; infant; pain; assessment

1.0 Introduction

Preterm infants undergo repeatedly painful diagnostic and therapeutic procedures to ensure their survival. In order to provide appropriate pain management in this population, accurate assessment of pain is necessary (Anand et al., 2005), yet complex because we are limited to indirect measures of pain responses. And although infant pain responses have both behavioral

Corresponding author: Dr. Liisa Holsti, Centre for Community Child Health Research, F611, 4480 Oak Street, Vancouver, British Columbia, V6H 3V4. Canada, Phone: 604-875-2000, local 5200, Fax: 604-875-2384, E-mail address: lholsti@cw.bc.ca.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

and physiological components, dissociations between behavioral and physiological pain responses are common (Barr, 1998; Morison et al., 2001). Furthermore, facial actions are the relatively most specific behavioral pain indicators in preterm infants, yet these responses may be dampened. (Johnston et al., 1996; Grunau et al., 2001). Thus, using scales such as the Neonatal Facial Coding System (NFCS; Grunau and Craig, 1987) and the Premature Infant Pain Profile (PIPP; Stevens et al., 1996) which rely solely on facial actions as the behavioral indicator may lead clinicians to the erroneous conclusion that pain is not present and that treatment is not needed. In addition, preterm infants born at earlier gestational ages may display different pain behaviors than infants at later gestational ages due to neurological immaturity. Such behaviors may not be captured in the current pain scales because the behaviors chosen are based on behaviors seen in term infants.

A number of scales have been developed for assessing acute pain in preterm infants; few have had adequate validation (Duhn and Medves, 2004). Currently, the two scales used most widely are the Neonatal Infant Pain Scale (NIPS; Lawrence et al., 1993) and the PIPP (Stevens et al., 1996). The NIPS includes crying as one of the indicators, but many preterm infants require mechanical ventilation, a procedure which precludes using cry as an indicator. Furthermore, the extremity action descriptors on the NIPS are very general. On the other hand, the PIPP (Stevens et al., 1996) uses weightings to adjust for varying gestational ages and sleep/wake states to accommodate these contextual factors. However, the weightings applied to the scoring of the sleep/wake states, in particular, may obscure important information related to infant arousal. For example higher weightings are applied to infants in deep sleep; however, it is not possible to determine whether or not infants in deep sleep feel greater pain.

More recently, using an assessment tool based on the Synactive Theory of Development (Als, 1984), we have shown that two developmentally relevant hand movements are reliable indicators of pain/stress in preterm infants (Grunau et al., 2000; Holsti et al., 2004; Holsti et al., 2005b). Unlike facial responses, hand and other body movements are heightened in infants who have had greater exposure to painful procedures or when handling is clustered (e.g. Holsti et al., 2005a). Thus, our objective was to improve on current pain scales by combining into a single unidimensional scale the relatively most specific and well studied individual facial actions and sleep/waking states with 2 hand movements. Specifically, the aims of this study were to determine the inter-rater reliability, the internal consistency, the construct and the concurrent validity of this new scale, the Behavioral Indicators of Infant Pain (BIIP).

2.0 Methods

This study is a repeated measures cohort study. The study sample comprised 92 preterm infants (49 male and 43 female) born between 24 - 31 completed weeks gestational age (GA) and admitted to the level-III NICU in the Children's & Women's Health Centre of British Columbia, Vancouver, Canada. Infants who had received analgesics or sedatives within 72 hours of the assessment, who had a major congenital anomaly, or who were exposed to maternal illicit drug use during pregnancy were omitted. Eighty eight assessments were completed at 32 weeks (+/- 7 days) postconceptional age (PCA). An additional 12 assessments were completed between 25-28 weeks PCA. Of these 12 assessments, 4 assessments were on infants seen at the early age only, 8 assessments were on infants assessed again at 32 weeks PCA. Thus, the total number of assessments was 100. Infant characteristics are presented in Table 1. The status of the infants on the test day are presented in Table 2. Sample size estimates were calculated using GPOWER (Faul and Erdfelder, 1998). Effect sizes entered into the program were based on changes in NFCS scores during blood collection at 31-33 weeks (Craig et al., 1993). Using this method, 15 infants were needed to detect differences between each Phase for a power of 0.90 with the statistical significance set at 0.05.

2.1 Background Data

A NICU-trained research nurse who completed the prospective clinical chart review obtained information from birth to day of testing including, but not limited to the following: birth weight, gestational age at birth, illness severity using the Score for Neonatal Acute Physiology (SNAP-II; Richardson et al., 2001), head ultrasound scan results, daily opioid and other analgesic and sedative exposure, numbers and types of invasive skin breaking procedures, respiratory support, type and time of last handling just prior to blood collection. As in our previous studies, procedural pain exposure was defined as the sum of every skin breaking procedure from birth to the testing day (e.g. heel lance, intramuscular injection, chest tube insertion, central line insertion). As is protocol in our nursery, each attempt at a procedure is documented in the medical chart; therefore, the total reflected all skin breaks (Grunau et al., 2001; Holsti et al., 2004). Total intravenous (iv) morphine exposure was calculated from birth to the test day by multiplying the average daily dose of iv morphine, adjusted for daily weight, by the number of days of iv morphine (Grunau et al., 2001; Holsti et al., 2004).

2.2 Procedures

The infants were recruited by a NICU-trained research nurse; written informed consent was obtained from the mother or other legal guardian according to a protocol approved by the Clinical Research Ethics Board of the University of British Columbia and the Children's and Women's Health Centre of British Columbia Research Review Committee. A video camera was positioned for a close-up view of the face and body and was attached to a custom made recording set-up on a moveable cart. The video signals were fed directly to a VCR and a time code was imprinted automatically. Each study phase was marked with an inaudible event cue signal recorded on the videotape. During the recording, the incubator was partially covered with a blanket, and the infant's position was supported (nested) using a continuous roll around both sides and feet.

2.3 Measures

2.3.1 Behavioral Indicators of Infant Pain (BIIP)—The BIIP combines sleep/wake state indicators, 5 facial actions and 2 hand actions (See Table 3 for definitions of each of the indicators). All of these indicators have been validated individually for assessing acute pain in preterm infants (Grunau et al., 2000; Stevens et al., 2000; Morison et al., 2003; Holsti et al., 2004). Moreover, the 5 facial actions we selected have been validated for assessing post-operative pain in infants born preterm and up to 18 months of age (Peters et al., 2003). Sleep/wake states were coded according to the definitions provided in the Newborn Individualized Developmental Care and Assessment Program (NIDCAP) (Als, 1984). The 5 facial actions were coded according to the definitions for the Neonatal Facial Coding System (Grunau and Craig, 1987) and the two hand actions coded according to the NIDCAP definitions (Als, 1984) (See Table 3). Deep sleep, active sleep, drowsy and active awake, states understood generally to be indicative of low distress, were assigned a score of 0. Crying was given the highest score (2). Facial and hand actions were given a score of 1. Figure 1 shows the BIIP scoring sheet.

2.3.2 Neonatal Infant Pain Scale (NIPS; Lawrence et al., 1993)—The NIPS is a multidimensional pain scale for use in preterm infants which includes indicators for facial expression, cry, breathing patterns, arm and leg movements and state of arousal. Each indicator is given a score of 1 if present. This scale has high inter-rater reliability and internal consistency, and has established content, concurrent and construct validity (Lawrence et al., 1993).

2.4 Heart Rate

Continuous electrocardiographic (ECG) activity was recorded from a single lead of surface ECG and was digitally sampled at 360 Hz off-line using a specially adapted computer acquisition system. Custom physiologic signal processing software (HRView Software, 1996) was used to acquire, process and analyze heart rate (HR). R waves were detected from the sampled ECG, and were used to form a smoothed instantaneous 4-Hz time series as described in Berger et al., 1989. Epochs of heart rate (HR; 2.2 minutes each) were selected for Baseline, Lance/squeeze and Recovery. The epoch selection criteria were based on quantitative signal stationarity, the presence of a stable behavioral state, and the absence of gross movement artifact (Oberlander and Saul, 2002).

2.5 Video Coding

Infant behaviors were videotaped continuously across three, one minute Phases of blood collection (Baseline, Lance/squeeze, Recovery) required for the clinical management of the infants. Videotapes were edited for coding in random order of Phases; coders were blind to all clinical information about the infants. In addition, the order of scoring the BIIP and NIPS was assigned randomly. One of the coders had over 5 years experience coding preterm infant behavior, while the other coder had no previous experience in working with preterm infants in the NICU or in behavioral coding. The BIIP and the NIPS were scored in real time across the three, one minute Phases of blood collection

2.6 Data Analysis

2.6.1. Preliminary Analyses—Although significant brain injury does not appear to alter facial responses during acute procedural pain (Oberlander et al., 2002), both pain scales used in this study include extremity actions which may be altered by CNS injury. Therefore, we explored the data to determine whether significant brain injury influenced the pain scale results. Nineteen infants did not have a head ultrasound scan; all these infants were born ≥ 29 weeks gestational age. Using the Kruskal-Wallis test, no differences in BIIP scores were found between those infants with and those infants without a scan. Next, we compared graphically the scores of the six infants who had significant CNS injury (defined as intraventricular hemorrhage [IVH], Grade III or IV and/or periventricular leukomalacia) versus infants with normal – Grade II IVH. No differences in BIIP scores were found across the three procedure Phases; therefore, these six infants were retained in subsequent analyses.

2.6.2 Primary Analyses

2.6.2.1 Internal Consistency and Inter-rater Reliability: The internal consistency of the BIIP was evaluated using Cronbach's α . The two-way random absolute agreement method of obtaining intra-class correlations (ICC) was used to determine inter-rater reliability. In addition, the Bland and Altman (1986; 1999) approach was used to examine whether a differential level of bias was present between raters.

2.6.2.2. Construct Validity: Changes in Pain Scale Scores across Blood Collection Phases:

To assess changes in BIIP, NIPS scores and mean heart rate repeated measures ANOVA was carried out across the three Phases of blood collection with sex examined as a between subjects variable. Bonferroni corrections were used to correct for overall Type I error. Repeated measures data were examined for sphericity; when present, Greenhouse-Geisser Epsilon values were used to determine significance. The significance level for each test was set at $p < 0.05$. Statistically significant ANOVA was followed by planned Student's t tests for paired comparisons to identify differences between Phases. Since the number of assessments at < 28 weeks GA was small (12), the data were graphed to determine whether the pattern of responses was similar to that of the infants tested at the later age.

2.6.2.3 Concurrent Validity of the BIIP: Spearman rank correlations were used to assess concurrent validity between the BIIP and the NIPS and between the BIIP and mean heart rate.

2.6.3 Secondary Analyses by Gestational Age at Birth—The infants included in the primary analyses had heterogeneous neonatal courses. To ensure the BIIP was reliable and valid for all the infants, for the 88 infants tested at 32 weeks gestational age, the data were re-analyzed with our sample divided into two gestational age groups at birth (earlier born <29 weeks [EB: n=24] versus later born 29-32 weeks [LB: n=64]) (Holsti et al, 2006). Gestational age at birth was chosen as the grouping variable because birth weight, illness severity, gestational age and pain exposure are variables which are highly inter-correlated (e.g. Grunau et al, 2005). All analyses remained the same with the exception of the construct validity analysis; to determine whether there were differences between BIIP scores between the two groups across the three phases (Baseline, Lance/squeeze, Recovery), repeated measures analysis of variance with gestational age group as a between subjects factor was completed.

3. Results

3.1 Internal Consistency and Inter-rater Reliability of the BIIP

The standardized item α for the BIIP during the Lance/squeeze Phase was 0.82. For the inter-rater reliability, levels of agreement between raters ranged from high to very high, depending on the Phase. For the Baseline Phase, the ICC was 0.80 (CI: 0.60-0.90). During the Lance/squeeze Phase, the ICC was 0.92 (CI 0.85-.0.97); and during the Recovery Phase, the ICC was 0.88 (CI: 0.76-0.94). In addition, no tendency of differential bias between raters was found ($\text{Adj } R^2 = -0.01$, $F [1,93] = 0.02$, $p = 0.87$).

3.2 Construct Validity: Changes in Behavioral and Physiological Measures across Phases of Blood Collection

3.2.1 BIIP—For the group of infants as a whole, a significant main effect for Phase was found ($F [1, 98] = 109.6$, $p < 0.0001$) (See Figure 2). Post hoc tests revealed that BIIP scores increased significantly from Baseline to the Lance/Squeeze Phase (Mean scores \pm [SD]: Baseline 1.0 ± 1.8 ; Lance/squeeze 5.3 ± 2.6 ; CI -4.8 to - 3.7, $p < 0001$) and then decreased significantly during the Recovery Phase (1.8 ± 0.3 ; CI 2.8 to 4.1, $p < 0.0001$). No sex effects were found, nor were differences observed in the graphical pattern of responses in the infants tested < 28 weeks compared to infants tested at 32 weeks GA. During the Lance/squeeze phase, 9 infants (10%) had no behavioral response, 46% of the infants who were not intubated cried, 84% of the infants had facial responses and 38% had hand responses. Four infants had hand responses only. Only 34% of the infants showed responses in all three areas (behavioral state, facial and hand actions).

3.2.2 NIPS—An overall Phase effect was found ($F [1,98] = 61.7$, $p < 0.0001$) for the NIPS scores which increased significantly from Baseline to Lance/squeeze (Mean scores \pm [SD]: Baseline: 2.0 ± 1.2 ; Lance/squeeze 4.0 ± 1.7 ; CI -2.3 to - 1.6, $p < 0001$) and then decreased significantly during the Recovery (2.3 ± 1.8 ; CI 1.3 to 2.1, $p < 0.0001$).

3.3 Mean Heart Rate

Of the 100 assessments, twelve (one assessment from the early tested group, 11 from the later tested group) had one or more Phases of heart rate (HR) which could not be analyzed due to technical difficulties. Changes in mean HR were significant across Phases ($F [1,87] = 87.0$, $p < 0.0001$). Mean HR increased significantly from Baseline to Lance/squeeze (Mean, SD Baseline: 157 ± 11 , Lance/squeeze: 175 ± 14 ; CI - 20.7 to - 15.3, $p < 0.0001$) and decreased significantly during Recovery (Recovery: 159 ± 14 ; CI 12.9 to 19.4, $p < 0.0001$).

3.4 Concurrent Validity of the BIIP

The correlation between the BIIP and NIPS during the Lance/squeeze Phase was moderate ($r = 0.64, p < 0.0001$). The correlations between the BIIP and mean HR during the Lance/squeeze Phase was also moderate ($r = 0.45, p < 0.0001$). A lower correlation between the NIPS and mean heart rate was found ($r = 0.28, p = 0.01$). However, the differences between the magnitude of these correlations was not significant ($z = 1.2, p = 0.11$)

3.5 Secondary Analyses of Psychometric Properties by Gestational Age at Birth

3.5.1 Internal consistency and Inter-rater Reliability—When the data were reanalyzed split by gestational age at birth, the standardized item α for the BIIP during the Lance/squeeze Phase was 0.89 for the EB group and 0.76 for the LB Group. For the inter-rater reliability, with the exception of the baseline reliability for the EB infants which was moderate, the levels of agreement between raters ranged from high to very high (Baseline Phase: EB ICC = 0.64 [CI: -0.25 - 0.90], LB ICC = 0.83 [CI: 0.59-0.93]; Lance/squeeze Phase: EB ICC = 0.92 [CI 0.73-.0.97], LB ICC = 0.94 [CI: 0.86-0.98]; Recovery Phase: EB ICC = 0.93 [CI: 0.80-0.98], LB ICC = 0.86 [CI: 0.58-0.95]).

3.5.2. Construct Validity—For the group of infants as a whole, a significant main effect for Phase was found ($F [1,86] = 85.9, p < 0.0001$); however, no gestational age effect was found.

3.5.3 Concurrent Validity—Correlations between the BIIP and NIPS remained moderate and significant irrespective of gestational age group (EB $r = 0.64, p < 0.0001$; LB $r = 0.60, p < 0.0001$). Correlations between the BIIP and mean heart rate also remained moderate (EB $r = 0.33, p < 0.05$; LB $r = .50, p < 0.001$).

4. Discussion

Although a number of scales are available for measuring acute pain in preterm infants, most have not had thorough psychometric testing, do not include theoretically derived, developmentally relevant indicators for preterm infants, or add behavioral and physiological indicators together into a single score making decisions regarding pain management more complex. Combining the relatively most specific and valid behavioral pain indicators into a single scale, the Behavioral Indicators of Infant Pain (BIIP) is a reliable and valid tool for assessing acute pain associated with procedures in preterm infants. In addition to assessing behavioral state changes, including both facial and hand responses is important to be able to capture the range of responses in these infants. Like others (e.g. Johnston et al, 1996), we found that almost half the infants did not cry during the skin-breaking phase of the procedure. Moreover, a small number of infants did not show facial responses, but did have extremity responses to the heel lance.

In addition, because this scale is scored easily in real time by both those experienced and those inexperienced in observing preterm infant behavior, it is practical for clinical use. Indeed, inter-rater reliability was high to extremely high across procedure Phases. Furthermore, despite the heterogeneity of the sample, the internal consistency of the scale was high (0.82). What is unique about this scale is the addition of two clearly defined, developmentally relevant, hand actions (finger splay and fisting) which are derived from theory and which have been shown previously to be clinically reliable indicators of stress in this population (Grunau et al., 2000; Holsti et al., 2004; Holsti et al., 2005b). Importantly, we found that our scale discriminates between procedure Phases with low scores occurring before and after the procedure and the highest scores occurring during the Lance/squeeze Phase. Finally, the BIIP had concurrent validity with total BIIP scores having a moderate, significant relationship with mean heart rate.

In the past, much care has been taken to include in pain scales facial actions which are anatomically derived (e.g. Grunau and Craig, 1987; Stevens et al., 1996); however, much less attention has been paid to determine which body movements would be specific and developmentally relevant stress indicators for preterm infants. For example, the descriptors of the extremity actions on the NIPS are very general (“flexed/extended”) and can reflect a static extremity position rather than an extremity action. Others include so many actions that their specificity is questionable (Craig et al, 1993). Scoring static extremity positions may not be a useful method of quantifying pain intensity, a construct important for measuring chronic pain states. Alternatively, providing extremity indicators which describe specific movement patterns, such as those on the BIIP, whose frequency can be counted may be important not only for measuring pain intensity, but for quantifying levels of activity, both of which may be helpful in assessing persistent pain states, an area of pain assessment which is needed urgently (Boyle et al., 2006; Stevens and Riddell, 2006).

Some may argue that including extremity actions on pain scales for use in preterm infants is not practical since pain management interventions, such as swaddling (e.g. Huang et al, 2004), preclude visualization of the arms and legs. Although we advocate strongly for the judicious use of non-pharmacological interventions to manage pain in the NICU, the long-term developmental effects of swaddling have yet to be determined. Indeed, tight and constant swaddling does not allow the infant to bring hands to mouth, an action which allows the infant to self-regulate (Als, 1984). Importantly, even when constrained, muscles continue to contract isometrically with microscopic changes occurring in the muscle length in response to the stimulus (Oatis, 2004); therefore, despite the outward appearance of less activity in swaddled infants undergoing painful procedures, the neuromotor signal continues to be sent to the brain. Along with multiple factors, such as chronic stress system activation (Grunau et al., 2006) and early illness, these signals may contribute over time to the functional changes in motor development observed in extremely preterm children at school age (e.g. Holsti et al., 2002).

Not surprisingly, we found only a moderate correlation between the BIIP and a validated multidimensional pain scale, the NIPS. The NIPS includes physiological indicators, indicators which, on other valid scales, provide unique, but complementary information (Stevens et al., 1996). Despite the inclusion of physiological parameters on the NIPS, a weaker (but not significantly lower) correlation was found between NIPS scores and mean heart rate during the Lance/squeeze Phase. This finding underscores the importance of not only measuring separately both behavioral and physiological changes observed during painful procedures, but also of the difficulties in combining divergently related indicators into a single score. And although physiological indicators, such as heart rate, vagal tone and oxygen saturations, may provide additional information indicating infant stress (for review see Franck and Miaskowski, 1997), some suggest that behavioral indicators are more ecologically salient (Barr, 1998). That is, from an evolutionary perspective, behavioral indicators are designed specifically to elicit caregiving.

Finally, measuring independently behavioral and physiological responses to pain is critical particularly when we are trying to determine the effects of pharmacological and non-pharmacological interventions. For example, skin-to-skin holding reduces effectively behavioral manifestations of pain, but has little effect on heart rate or on oxygen saturations (Johnston et al., 2003). Similarly, although sucrose reduces pain behavior, varying effects on different physiologic indices have been found (e.g. Boyer et al., 2004; reviewed by Stevens et al, 2004).

In a recent review of pain scales available for use in infants, none of the existing tools appear to be ideal for all situations and many have had limited psychometric testing (Duhn and Medves, 2004). Of the unidimensional scales, the Neonatal Facial Coding System is the most

well developed (Grunau and Craig, 1987). The NFCS can be used for assessing acute procedural and post-operative pain in infants between 23 weeks gestational age up to 18 months of age. Indeed, we retained 5 facial actions for use on the BIIP, including the 2 lower facial actions found to be valid for assessing post-operative pain in infants beyond the newborn period (Peters et al., 2003) for these reasons. However, as we described, the NFCS includes facial actions only, actions which may be dampened in infants who are sicker or who have had ongoing exposure to skin-breaking procedures. Finally, although the BIIP does contain proportionately more facial indicators than body movement indicators, rather than attempt to achieve scoring “balance” by including equal numbers of indicators in each category, we chose to include the two hand actions which were relatively more specific and which were most likely to occur (Holsti et al., 2004). Using this method, we were able to capture the varied responses of those infants in our sample who had been exposed to greater procedural pain.

Of the multidimensional scales, the PIPP (Stevens et al., 1996) is the most well-validated tool for use with premature infants; however, conceptual issues surrounding this scale, such as the weightings applied to sleep/wake states, may hide important information with regard to arousal in these infants. The Neonatal Infant Pain Scale (NIPS; Lawrence et al., 1993), the COMFORT scale (Ambuel et al., 1992), the Premature Infant Pain Assessment (PIPA; Jorgensen et al., 1999), the Pain Assessment Tool (PAT; Spence et al., 2003) and the Distress Scale for Ventilated Newborn Infants (Sparshott, 1996) either have been validated using only older preterm infants (i.e. > 32 weeks), or have not been evaluated for their clinical utility. The PAIN Scale (Hudson-Barr et al., 2002) is an adapted version of the NIPS scale. Although the validity of this scale has been assessed, inter-rater reliability has not been established and little variation exists between basal and stimulus-response scores.

A number of limitations of our study should be mentioned. First, although our study includes infants assessed at early gestational ages (<29 weeks), the number of infants tested at that age is small. Assessing pain in infants born at extremely low gestational ages is a challenge since many are placed on sedation or analgesic medications. Second, the infants in our study were scored from videotape rather than at bedside. This method of scoring is practical for research purposes, but does not establish the feasibility of our scale. Research is underway currently to establish the feasibility, clinical utility and validity of this scale for use in measuring post-operative pain and to determine the ability of the BIIP to discriminate between skin-breaking and non-skin breaking procedures.

In conclusion, measuring pain in preterm infants is necessary for appropriate pain management. Many pain scales are available for use in this population, but few have had thorough psychometric evaluation. The BIIP improves the accuracy of pain assessment for preterm infants by combining theoretically derived, developmentally relevant hand movements and sleep/wake states with anatomically derived facial actions. Moreover, the BIIP provides not only non-adjusted sleep/wake state ratings which allow for more detailed evaluation of levels of arousal, but it includes two relatively specific hand actions and, most importantly, it allows measurement of behavioral and physiological indices separately.

Acknowledgements

Grant support: National Institutes of Health grant HD39783 (REG), Canadian Institutes of Health Research grant MOP42469 (REG), British Columbia Ministry of Children and Family Development grant through the Human Early Learning Partnership (LH); Child and Family Research Institute Establishment Grant (LH: 06-2426 REG: 02-2403;03-3112), Canadian Child Health Clinician Scientist Career Award (LH), Michael Smith Foundation for Health Research (REG).

We would like to thank the reviewers for their constructive comments on an earlier version of this manuscript; the staff and families of the Neonatal Intensive Care Unit at B.C. Children’s Hospital for their participation in this study; Patrician Mortenson and Colleen Jantzen for carrying out behavioral coding in the Early Human Experience Unit,

Centre for Community Child Health Research, Child and Family Research Institute; and the CANDO Research Unit in the Division of Occupational Therapy, School of Rehabilitation Sciences, University of British Columbia.

References

- Als, H. Manual for the naturalistic observation of newborn behavior (preterm and fullterm). Boston: The Children's Hospital; 1984.
- 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. [PubMed: 1545324]
- Anand KJS, Aranda JV, Berde CB, Buckman A, Caparelli EV, Carlo WA, Hummel P, Lantos J, Johnston CC, Lehr VT, Lynn AM, Maxwell LG, Oberlander TF, Raju TNK, Soriano SG, Tadio, Walco GA. Analgesia and anesthesia for neonates: study design and ethical issues. *Clin Ther* 2005;27:814–843. [PubMed: 16117988]
- Barr R. Reflections on measuring pain in infants: dissociation in responsive systems and “honest signalling”. *Arch Dis Child* 1998;79:F152–F156.
- Berger RD, Paul P, Cohen RJ. Transfer function analysis of autonomic regulation: canine atrial rate response. *Am J Physiol* 1989;256:H142–H152. [PubMed: 2912176]
- Bland JM, Altman DG. Statistical methods for assessing the agreement between two methods of clinical measurement. *Lancet* 1986;8:307–310. [PubMed: 2868172]
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 1999;8:135–160. [PubMed: 10501650]
- Boyer K, Johnston C, Walker CD, Filion F, Sherrard A. Does sucrose analgesia promote physiologic stability in preterm neonates? *Biol Neonate* 2004;85:26–31. [PubMed: 14631163]
- Boyle EM, Freer Y, Mae Wong C, McIntosh N, Anand KJS. Assessment of persistent pain or distress and adequacy of analgesia in preterm ventilated infants. *Pain* 2006;124:87–91. [PubMed: 16725260]
- Craig KD, Whitfield MF, Grunau RVE, Linton J, Hadjistavropoulos HD. Pain in the preterm neonate: Behavioral and physiological indices. *Pain* 1993;52:287–300. [PubMed: 8460047]
- Duhn LJ, Medves JM. A systematic integrative review of infant pain assessment tools. *Adv Neonatal Care* 2004;4:126–140. [PubMed: 15273943]
- Faul, F.; Erdfelder, E. GPOWER: A priori-,post hoc-, and compromise power analyses for MS-DOS (Computer program). Bonn Germany: Bonn University; 1998.
- Franck LS, Miaskowski C. Measurement of neonatal responses to painful stimuli: a research review. *J Pain Sympt Manage* 1997;14(6):343–378.
- Grunau RV, Craig KD. Pain expression in neonates: facial action and cry. *Pain* 1987;28:395–410. [PubMed: 3574966]
- Grunau RE, Holsti L, Haley D, Oberlander TF, Weinberg J, Solimano A, Whitfield MF, Fitzgerald C, Yu W. Neonatal procedural pain exposure predicts lower cortisol and behavioral reactivity in preterm infants in the NICU. *Pain* 2005;113:293–300. [PubMed: 15661436]
- Grunau, RE.; Holsti, L.; Peters, J. Long-term effects of infant pain. In: Tibboel, D.; Bhat, R., editors. *Pain Control and Sedation. Sem Fetal Neonatal Medicine*. 11. 2006. p. 268-275.
- Grunau RE, Holsti L, Whitfield MF, Ling E. Are twitches, startles and body movements pain indicators in extremely low birth weight infants? *Clin J Pain* 2000;16:37–45. [PubMed: 10741817]
- Grunau RE, Oberlander TF, Whitfield MF, Fitzgerald C, Lee SK. Demographic and therapeutic determinants of pain reactivity in very low birth weight neonates at 32 weeks' postconceptional age. *Pediatrics* 2001;107:105–112. [PubMed: 11134442]
- Holsti L, Grunau RE, Oberlander TF, Whitfield MF. Prior pain induces heightened motor responses during clustered care in preterm infants in the NICU. *Early Hum Devel* 2005a;81:293–302. [PubMed: 15814212]
- Holsti L, Grunau RE, Oberlander TF, Whitfield MF. Specific NIDCAP® movements help identify acute pain in preterm infants in the NICU. *Pediatrics* 2004;114:65–72. [PubMed: 15231909]
- Holsti L, Grunau RVE, Whitfield MF. Developmental coordination disorder in extremely low birth weight children at nine years. *J Dev Behav Pediatr* 2002;23:9–15. [PubMed: 11889346]
- Holsti L, Grunau RE, Whitfield MF, Oberlander TF, Lindh V. Behavioral responses to pain are heightened after clustered care in preterm infants. *Clin J Pain* 2006;22:757–764. [PubMed: 17057556]

- Holsti L, Grunau RE, Whitfield MF, Oberlander TF, Weinberg J. Body movements, an important additional factor in discriminating pain from stress in preterm infants. *Clin J Pain* 2005b;21:491–498. [PubMed: 16215334]
- HR View Software. Brighton, MA: Boston Medical Technologies; 1996.
- Huang CM, Tung WS, Kuo LL, Ying-Ju C. Comparison of pain responses of premature infants to the heelstick between containment and swaddling. *J Nurs Res: JNR* 2004;12:31–34.
- Hudson-Barr D, Capper-Michel B, Lambert S, Palermo TM, Morbeto K, Lombardo S. Validation of the Pain in Neonates (PAIN) Scale with the Neonatal Infant Pain Scale (NIPS). *Neonat Network* 2002;21:15–21.
- Johnston CC, Stevens B, Pinelli J, Gibbins S, Filion F, Jack A, Steele S, Boyer K, Veilleux A. Kangaroo care is effective in diminishing pain response in preterm neonates. *Arch Pediatr Adolesc Med* 2003;157:1084–1088. [PubMed: 14609899]
- Johnston CC, Stevens BJ. Experience in a neonatal intensive care unit affects pain response. *Pediatrics* 1996;98:925–903. [PubMed: 8909487]
- Jorgensen K, Watt LB, Pearson C, Bauchner H, Mirochnick M. Preterm Infant Pain Assessment (PIPA): a new tool to accurately assess pain in the preterm infant. *Pediatr Res* 1999;45:203A.
- Lawrence J, Alcock D, McGrath P, Kay J, MacMurray SB, Dulberg C. The development of a tool to assess neonatal pain. *Neonatal Network* 1993;12:59–66. [PubMed: 8413140]
- Morison SJ, Grunau RE, Oberlander TF, Whitfield MF. Relations between behavioral and cardiac autonomic reactivity to acute pain in preterm neonates. *Clin J Pain* 2001;17:350–358. [PubMed: 11783816]
- Morison SJ, Holsti L, Grunau RE, Whitfield MF, Oberlander TF, Chan HWP, Williams L. Are there developmentally distinct motor indicators of pain in preterm infants? *Early Hum Devel* 2003;72:131–146. [PubMed: 12782425]
- Oatis, CA. *Kinesiology. The Mechanics and Pathomechanics of Human Movement*. New York: Lippincott Williams & Wilkins. A Wolters Kluwer Company; 2004. Biomechanics of skeletal muscle; p. 44-65.
- Oberlander TF, Grunau RE, Fitzgerald C, Whitfield MF. Does parenchymal brain injury affect biobehavioral pain responses in very low birth weight infants at 32 weeks' postconceptional age? *Pediatrics* 2002;110:570–576. [PubMed: 12205262]
- Oberlander T, Saul JP. Methodological considerations for the use of heart rate variability as a measure of pain reactivity in vulnerable infants. *Clin Perinatol* 2002;29:427–443. [PubMed: 12380467]
- Peters JWB, Koot HM, Grunau RE, de Boer J, van Druensen MJ, Tibboel D, Duivenvoorden HJ. Neonatal Facial Coding System for assessing postoperative pain in infants: item reduction is valid and feasible. *Clin J Pain* 2003;19:353–363. [PubMed: 14600535]
- Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. *J Pediatr* 2001;138:92–100. [PubMed: 11148519]
- Sparshott M. The development of a clinical distress scale for ventilated newborn infants: Identification of pain and distress based on validated behavioural scores. *J Neonatal Nurs* 1996;2:5–11.
- Spence K, Gillies D, Harrison D, Johnston L, Nagy S. A reliable pain assessment tool for clinical assessment in the neonatal intensive care unit. *JOGNN* 2005;34:80–86. [PubMed: 15673649]
- Stevens, B.; Johnston, C.; Gibbins, S. Pain assessment in neonates. In: Anand, KJ.; Stevens, BJ.; McGrath, PJ., editors. *Pain in Neonates 2nd Revised and Enlarged Edition Pain Research and Clinical Management*. 10. Amsterdam: Elsevier Science; 2000. p. 101-134.
- Stevens BJ, Ridell RP. Looking beyond acute pain in infancy. *Pain* 2006;124:11–12. [PubMed: 16839690]
- Stevens B, Johnston CC, Petryshen P, Taddio A. Premature infant pain profile: development and initial validation. *Clin J Pain* 1996;12:13–22. [PubMed: 8722730]
- Stevens, B.; Yamada, J.; Ohlsson, A. *The Cochrane Library*. Chichester: Wiley; 2004. Sucrose for analgesia in newborn infants undergoing painful procedures (Cochrane review).

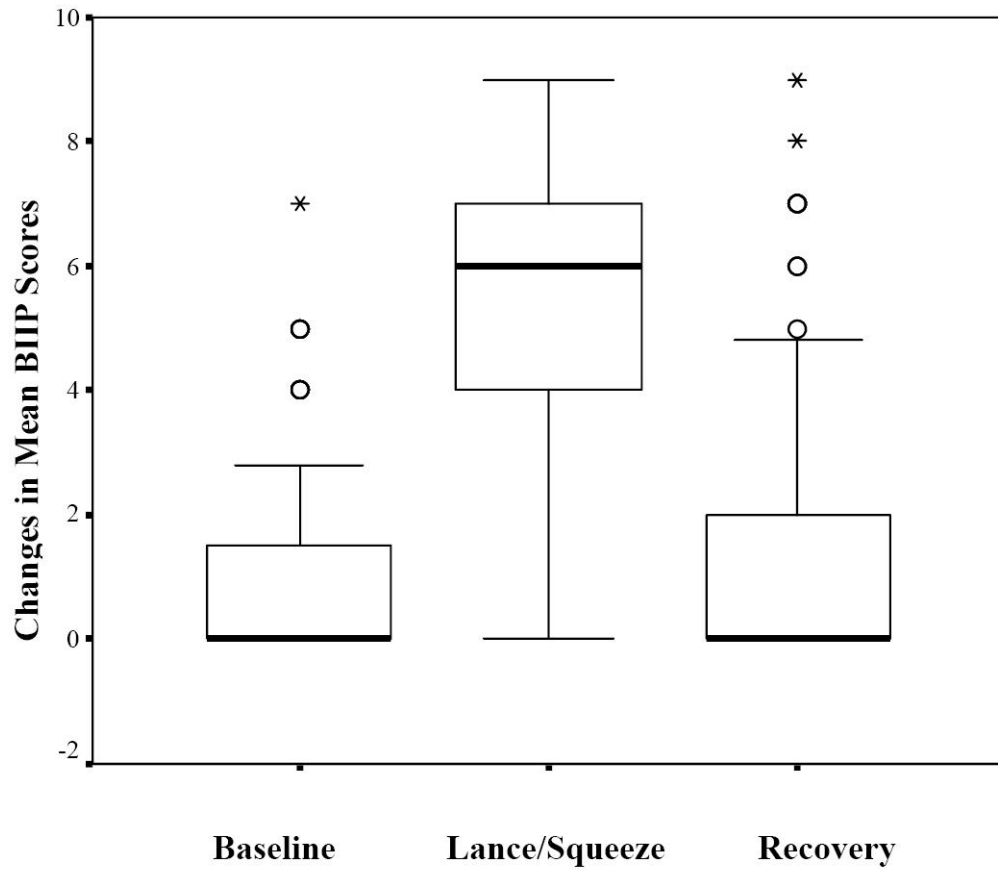


Figure 2. Changes in mean Behavioral Indicators of Infant Pain (BIIP) scores across phases of blood collection.

Table 1

Infant Characteristics n = 92

	Mean (sd)	Range	N (%)
Birth weight (grams)	1280 (445)	500 - 2525	
Gestational age at birth (weeks)	29 (2)	24 - 32	
SNAP- II Day 1*	12 (10)	0 - 46	
Ventilation (days)	8 (13)	0 - 50	
Other respiratory support (days)	8 (9)	0 - 36	
Pain exposure at 32 weeks PCA [†]	76 (58)	8 - 276	
Intravenous morphine exposure at 32 weeks PCA [‡]	0.90 (2.1)	0-9.6	
Grade III, IV Intraventricular hemorrhage or periventricular leukomalacia (n=82) [§]			6 (0.07)
Days of Dexamethazone	0.3(2)	0-19	
Ethnicity (Caucasian)			63 (69)
Maternal age (years)	30 (6)	19-47	

* Score for Neonatal Acute Physiology (Richarson et al, 2001)

[†] Number of invasive (skin breaking) procedures from birth to the test day

[‡] Morphine exposure = daily average intravenous mg/kg) X days

[§] Infants born weighing > 1250 grams did not receive head ultrasound scans (n=19).

Table 2

Infant Characteristics on Test Day n = 92

	Mean (sd)	Range	N (%)
Postnatal age on test day (days) - (a. Early Assessment, b. Later Assessment)	13 (8) 19 (13)	5 – 34 3 – 59	
Mechanical ventilation on test day			20 (20)
Number of painful procedures in 24 hours before the test day - (a. Early Assessment, b. Later Assessment)	a. 0.5 (0.5) b. 1.0 (1.0)	a. 0 – 1 b. 0 – 8	
Number of minutes since last handling before blood collection (a. Early Assessment, b. Later Assessment)	a. 90 b. 106	a. 45-182 [*] b. 5 – 324 [*]	

* Two infants required gentle stimulation to help them recover from a dip in heart rate in the 30 minute Baseline phase before the blood collection.

Table 3
Behavioural Indicators of Infant Pain (Preterm and Fullterm) Definitions

Sleep/Wake States	Description
Deep Sleep	Eyes closed, regular breathing, no movements of extremities
Active Sleep	Eyes closed, twitches or startles of extremities, rapid eye movements, irregular breathing
Drowsy	Eyes open (but roving or not focused) or closed, irregular breathing, some body movements
Quiet Awake	Eyes open, focused, very few or no body movements
Active Awake	Eyes open, active extremity movements
Agitated/Crying	Upset, fussing, highly aroused, crying
Face & Hand Actions	Description
Brow Bulge	Bulging, creasing and/or vertical furrows above and between brows occurring as a result of lowering and drawing together of the eyebrows.
Eye Squeeze	Squeezing and/or bulging of the eyelids
Naso-labial Furrow	Pulling upwards and deepening of the naso-labial furrow (a line or wrinkle which begins adjacent to the nostril wings and runs down and outwards beyond the lip corners).
Horizontal Mouth	A distinct horizontal stretch pull at the corners of the mouth sometimes accompanied by a taut upper lip.
Taut Tongue	Raised, cupped tongue with sharp tensed edges. The first occurrence of taut tongue is usually easy to see, often occurring with a wide open mouth. After this first occurrence, the mouth may close slightly. Taut tongue can be scored on the basis of the still visible tongue edges.
Finger Splay	Sudden opening of the hands with fingers extended and separated from each other
Fisting	Tight closing and flexing of the fingers to form a fist