ORIGINAL ARTICLE

Postoperative pain assessment in the neonatal intensive care unit

C McNair, M Ballantyne, K Dionne, D Stephens, B Stevens

Arch Dis Child Fetal Neonatal Ed 2004;89:F537–F541. doi: 10.1136/adc.2003.032961

See end of article for authors' affiliations

Correspondence to:
C McNair, Hospital for Sick Children,
Neonatology, 555
University Avenue,
Toronto, Ontario, M5G
1X8, Canada; carol.mcnair@sickkids.ca

Accepted 11 May 2004

Objectives: To compare the convergent validity of two measures of pain (premature infant pain profile (PIPP) and crying, requires oxygen, increased vital signs, expression, and sleepless (CRIES)) in real life postoperative pain assessment in infants.

Methods: This study was a prospective, repeated measures, correlational design. Two staff nurses were randomly assigned either the PIPP or CRIES measure. An expert rater assessed each infant after surgery, and once a day using the visual analogue scale (VAS).

Setting: A level III neonatal intensive care unit in a metropolitan university affiliated paediatric hospital.

Results: Pain was assessed in 51 neonates (28–42 weeks of gestational age) after surgery. There was no significant difference in the rates of change between the pain assessment measures across time using repeated measures analysis of variance ($F_{50,2} = 0.62$, $p = 0.540$), indicating correlation between the measures. Convergent validity analysis using intraclass correlation showed correlation, most evident in the first 24 hours (immediately, 4, 8, 20, and 24 hours after the operation). Correlations were more divergent at 40 and 72 hours after surgery. No significant interactions were found between gestational age and measure ($F_{304,4} = 0.75$, $p = 0.563$) and surgical group and measure ($F_{304,2} = 0.39$, $p = 0.680$).

Conclusions: PIPP and CRIES are valid measures that correlate with pain for the first 72 hours after surgery in term and preterm infants. Both measures would provide healthcare professionals with an objective measure of a neonatal patient's pain.

Evidence supporting the existence of pain in infants has increased significantly in the last few decades. However, the subjective nature of pain makes measurement in infants challenging. New infant pain measures that have been used in research require further validation in the clinical setting. The premature infant pain profile (PIPP), developed by Stevens *et al.*,¹ is an example of a composite measure for assessing acute procedural pain in preterm and term infants which has been developed and validated for use in research^{2–5} and, to a limited extent, clinical practice.⁶ The crying, requires oxygen, increased vital signs, expression, and sleepless (CRIES) measure, developed by Bildner and Krechel^{7,8} to assess postoperative pain, has also been primarily developed for research but has undergone limited clinical testing.^{7,8} Further validation of CRIES and PIPP in the clinical setting is required.⁹

Infants experience pain during the postoperative period as a result of the surgical procedure as well as from continuing postoperative interventions. It is of primary importance to determine the most useful way to assess postoperative pain. To date, no composite measures of infant pain have been assessed for construct—for example, convergent—validity in the clinical context of real time postoperative pain management.

Review of literature

A number of infant pain assessment measures have been developed over the past decade.^{10–12} However, most have had minimal testing outside of research, leaving clinicians with the challenge of determining if such measures could be valid, reliable, and feasible in clinical practice.⁹

Bours *et al.*⁹ reviewed and described all current measures (published and unpublished) designed to assess pain in preterm and term infants. After reviewing these measures and their reliability, validity, and clinical utility, they rated CRIES^{7,8} and PIPP¹ as two of the better multidimensional

instruments. To determine concurrent validity of newly developed pain measures, further comparisons have been made with measures such as the visual analogue scale (VAS), the psychometric properties of which have been established.¹³

PIPP consists of seven indicators including assessment of gestational age and behavioural state (contextual indicators), heart rate and oxygen saturation (physiological indicators), and facial actions—brow bulge, eye squeeze, and nasolabial furrow (behavioural indicators).¹ PIPP creates a score from 18 to 21 depending on gestational age, with 0–6 reflecting no pain, 6–12 reflecting mild-moderate pain, and above 12 indicating severe pain.

CRIES includes similar indicators to PIPP: crying, oxygen requirements, increases in heart rate or blood pressure, facial expression, and sleep behaviour. CRIES creates a score from 0 to 10, similar to most self report or observational measures of pain.

VAS is a continuous 10 cm line with the potential for scoring pain at any point on the scale—from no pain at 0 cm to severe pain at 10 cm. VAS is used to measure pain by observation in children below 4 years of age and self report in adults and older children.¹³

MATERIALS AND METHODS

Purpose

The primary purpose of this study was to prospectively compare (a) the convergent validity of PIPP and CRIES and (b) the convergent validity of PIPP and CRIES to the observational VAS in the context of real life postoperative pain assessment in infants.

The secondary purposes were to describe the patterns of pain intensity and resultant management strategies during

Abbreviations: CRIES, crying, requires oxygen, increased vital signs, expression, and sleepless; NICU, neonatal intensive care unit; PIPP, premature infant pain profile; VAS, visual analogue scale

Table 1 Basic details of the three gestational age groups studied

	28–31 weeks (n = 6)	32–35 weeks (n = 10)	≥36 weeks (n = 35)
Birth weight (g)	1471 (1203)	1665 (492)	3020 (720)
1 minute Apgar	4.5 (3.02)	6.0 (2.67)	7.42 (1.58)
5 minute Apgar	6.33 (2.94)	8.1 (1.10)	8.65 (0.80)
Male (69%)	5 (83%)	5 (50%)	25 (71%)
Female (31%)	1 (17%)	5 (50%)	10 (29%)

Values are mean (SD) except for the sex distribution. Significant differences were found in weight and Apgar scores.

the first 72 hours after an operation for infants of various gestational ages and to contribute to the overall construct validity of the pain assessment measures.

Design

A prospective, repeated measures, cohort design (with random assignment of raters) was used to assess the pain responses of infants in the first 72 hours immediately after surgery.

Setting

The study was conducted in a 42 bed, outborn, level III neonatal/surgical neonatal intensive care unit (NICU) at a metropolitan university affiliated paediatric hospital.

Sample

The convenience sample consisted of 51 infants who had received some sort of surgery. Criteria for inclusion were: between 28 and 42 weeks of gestation at birth; within the first 30 days of life; having surgery; not known to have neurological abnormalities or anomalies. The infants were stratified into the following groups by gestational age at birth: 28–31 weeks; 32–35 weeks; ≥ 36 weeks. These strata were developed in consideration of infant development and to be consistent with the gestational age indicators specified in the infant pain measures.

The sample size was calculated on the basis of a minimal clinically significant difference between the measures of 20% and standard deviation of 1. With an alpha of 0.017, a sample size of 45 would provide 80% power.

Data collection

The hospital's research ethics review board approved the study. If parents agreed to be approached by the research assistant, the study was explained and consent obtained.

Figure 1 Categories of neonatal surgery. NEC, Necrotising enterocolitis; TEF, tracheo-oesophageal fistula.

Table 2 Measures of correlation between CRIES and PIPP

Time after surgery (h)	n	ICC	95% CI	
			Lower	Upper
Immediately	51	0.6008	0.39	0.75
4	50	0.5449	0.32	0.71
8	51	0.5602	0.34	0.72
12	50	0.3483	0.08	0.57
16	49	0.3404	0.07	0.57
20	47	0.6329	0.42	0.77
24	48	0.6139	0.40	0.76
32	50	0.3080	0.04	0.54
40	49	0.2049	-0.08	0.46
48	45	0.3864	0.11	0.61
56	47	0.4883	0.24	0.68
64	44	0.4495	0.18	0.66
72	42	0.2523	-0.05	0.51

CRIES, Crying, requires oxygen, increased vital signs, expression, and sleepless; PIPP, premature infant pain profile; ICC, intraclass correlation.

Raters

Three raters each observed each infant's pain and rated the pain independently. The primary pain measures (PIPP and CRIES) were randomly assigned to nurses 1 (the infant's nurse) and 2 (another staff nurse or charge nurse). Randomisation was determined before the start of the study using a table of random numbers and remained confidential until the time of data collection.

The expert rater consistently used the observational VAS to assess pain. At any given time interval, the expert rater could be one of two clinical nurse specialists (authors of this paper), both of whom had extensive experience in real life infant pain assessment and in researching the psychometric properties of pain measures. Inter-rater reliability ($r = 0.90-0.95$) was established before the study between the two expert raters and an infant pain expert.

Timing of measurements

The times at which the two staff nurses assessed pain in the postoperative period using PIPP and CRIES were as follows:

- (1) During the first 24 postoperative hours: immediately after surgery (on return to the NICU) and every four hours thereafter.
- (2) For the subsequent 24–72 postoperative hours as per the following protocol:
 - (a) every eight hours when receiving continuous infusion analgesia
 - (b) every eight hours when receiving no analgesia

Table 3 Measures of correlation between CRIES, PIPP, and VAS

Time after surgery (h)	n	ICC	95% CI	
			Lower	Upper
Immediately	41	0.55	0.37	0.70
4	4	0.87	0.43	0.99
8	3	0.88	0.28	0.99
12	4	0.80	0.24	0.98
24	15	0.69	0.42	0.87
48	16	0.07	-0.18	0.43
72	23	0.26	0.01	0.54

CRIES, Crying, requires oxygen, increased vital signs, expression, and sleepless; PIPP, premature infant pain profile; VAS, visual analogue scale; ICC, intraclass correlation.

- (c) every eight hours and before all analgesia administration and at the analgesic's peak effect time (one to two hours after administration) when receiving intermittent bolus analgesia.

The observational VAS score was completed during the preoperative assessment, immediately after surgery (on return to the NICU), and once every 24 hours for the subsequent 72 hours. VAS measurements were consistently recorded at a time when PIPP and CRIES were pre-scheduled. The timing of daily VAS measurements depended on the availability of the expert rater and the assessment schedule and thus varied across the study.

Data analysis

CRIES, PIPP, and VAS scores were calculated and double checked for each postoperative event for each infant. Basic data were analysed using SPSS and correlational data and repeated measures analysis of variance using SAS. A descriptive data analysis was performed to determine measures of central tendency (means, medians, ranges, and standard deviations) and distribution of the data.

Convergent validity was evaluated by comparing the within subject variation for both the PIPP and CRIES scores and the PIPP, CRIES, and VAS scores using intraclass correlation (covariance matrix method). PIPP was scaled differently from CRIES and VAS and hence needed adjustment to be comparable with the other two.

The pattern of pain intensity across time was established by comparing the slopes for each measure. A slope was calculated for each subject using a linear regression model. This was done for each of the three pain measures. Hence the data analysed using a repeated measures analysis of variance were rates of change rather than raw pain scores. Distributions of the slopes from all three measures were assessed, and histograms showed minor deviations from normality.

The data were examined further to determine the interaction between pain scores and other factors. The measures (CRIES, PIPP, and VAS) were considered to be the between subject factors, and gestational age and type of surgery were the within subject factors.

RESULTS

A total of 51 infants were separated into three gestational age groups, with six in the 28–31 weeks group, 10 in the 32–35 weeks group, and 35 in the greater than 36 weeks group. Most were in the last group, which reflects the outborn, largely term, newborn surgical NICU population. As expected, there were significant differences in weight and Apgar scores between the three groups, which can be attributed to gestational age (table 1).

For analysis, the infants were grouped according to whether they had received minor or major surgery. Minor surgery included procedures for the following diagnoses: cataracts, one (2%); small sacral teratoma, one (2%); urology (small bladder extrophy), one (2%); pyloric stenosis, two

Figure 2 Pain scores after surgery. Values are means.

(4%). Major surgery consisted of intra-abdominal procedures, including gastroschisis and omphalocele repair (six (12%)), necrotising enterocolitis (three (6%)), and bowel resections due to congenital atresia, meconium plug, etc (20 (39%)), and intrathoracic procedures such as patent ductus arteriosis ligation (four (8%)), lobectomy (one (2%)), and cystic hygroma and tracheo-oesophageal fistula repair (12 (23%)) (fig 1). The minor and major categories were determined through consultation with local experts (surgeon and anaesthesiologist) and based on the following: site of surgery; duration of surgery; amount of tissue damage; and associated stress factors (criteria adapted from Anand and Aynsley-Green¹⁴).

Convergent validity

Convergent validity is defined as the extent to which two or more instruments that purport to be measuring the same construct agree with each other.^{15 16} The convergent validity of the pain assessment measures used in this study was established by comparing the within subject scores determined for the CRIES and PIPP measures and the within subjects scores for CRIES, PIPP, and VAS at each measurement time.

On examination of the overall intraclass correlation profiles, there was no difference between the measures, indicating correlation. Some measurement time points (immediately, 4, 8, 20, and 24 hours after surgery) show moderate correlation, whereas others are more divergent (40 and 72 hours after surgery) indicating fair correlation (tables 2 and 3). The classification of correlation was based on categories developed by Landis and Koch.¹⁷ The categories are described as: 0.81–1.0, almost perfect; 0.61–0.80, substantial; 0.41–0.60, moderate; 0.21–0.40, fair; 0.00–0.20, poor.

Repeated measures

Repeated measures analysis of variance of the main effects and interactions among gestational age, type of surgery, and the pain measures was performed. There was no significant difference in the slopes of change between the measures ($F_{50,2} = 0.62$, $p = 0.540$). There were no significant differences between gestational age groups ($F_{151,2} = 1.37$, $p = 0.265$) or between surgical groups ($F_{151,1} = 2.87$, $p = 0.973$) (table 4). Also, no significant interactions were detected when gestational age and measure ($F_{304,4} = 0.75$, $p = 0.563$) and surgical group and measure ($F_{304,2} = 0.39$, $p = 0.680$) were compared.

Patterns of pain response over time

When the change in pain scores over time were examined, a consistent pattern emerged across all three measures (fig 2). The highest pain scores occurred immediately after surgery, followed by a gradual decrease over the first 12 hours. Pain

Table 4 Between group differences

	F value	p Value
Main effects		
Age groups	1.37	0.2651
Surgery type	2.86	0.973
Measurement tool	0.62	0.5407
Interactions		
Age × measurement	0.75	0.5629
Surgery × measurement	0.39	0.6798

scores remained relatively low until about 48 hours, and then rose slightly between 48 and 72 hours. The small increase during the third day coincides with the clinical practice of converting analgesic management from opioid to non-opioid drugs.

The slopes of change for all subjects showed correlation as no difference was detected between pain measurements using PIPP, CRIES, and VAS. For ease of display, fig 2 is presented using raw data versus slopes of change.

Patterns of analgesic pain management over time

The primary analgesic used in this study for postoperative pain management was morphine (46, 92%). Of the remaining infants, four (8%) received acetaminophen for pain relief, and one who had a minor procedure did not receive analgesia. No patients received epidural infusions or caudal blocks during or after surgery. Most infants returned to the NICU unreversed and ventilated. No other sedatives or muscle relaxants (except those given during surgery) were used with any of our patients at the time of the study.

Forty five infants received continuous morphine infusion. The mean (SD) doses across time were 15.5 (9.8) $\mu\text{g/kg/h}$ immediately after surgery, 13.4 (8.7) $\mu\text{g/kg/h}$ 24 hours after, 12.6 (11.1) $\mu\text{g/kg/h}$ 48 hours after, and 12.8 (11.0) $\mu\text{g/kg/h}$ 72 hours after. Three infants (6.5%) were weaned from morphine by 24 hours, seven (15%) by 48 hours, and 22 (48%) by 72 hours.

DISCUSSION

The CRIES pain measure has been previously validated in term infants after surgery^{7,8} whereas the PIPP measure has been validated in term and preterm infants with procedural pain for treatment and diagnostic purposes.^{1,5,6} This study shows that CRIES and PIPP are valid measures for assessing postoperative pain in neonates, both term and preterm.

Correlation was established between CRIES, PIPP, and observational VAS for the first 24 hours after surgery. Thereafter, at 48 and 72 hours, correlation was limited (0.07 and 0.26 respectively). Similar conflicting findings are evident in the neonatal literature where the observational VAS correlated^{18,19} and failed to show correlation with other pain assessment.^{20,21}

There are several possible explanations for the limited correlation with the VAS after 24 hours. Firstly, the timing of the VAS measurements depended on the availability of the expert raters and thus varied across the study. The only consistent measurement time between PIPP, CRIES, and VAS was immediately after surgery. Thereafter the VAS measurements occurred at any of the three potential times in each 24 hour period. This resulted in inconsistent time comparisons, small sample sizes, and insufficient power for adequate comparison of the measures.

Secondly, pain assessment using the observational VAS is subjective, and thus pain intensity ratings may have varied between the two expert raters in spite of the restriction to two raters and the establishment of high inter-rater reliability before the study ($r = 0.90-0.95$). Furthermore, the VAS scores determined by the expert raters in this study were noted to be lower than the PIPP and CRIES scores. This is similar to the findings of Buchholz *et al*²¹ where experienced raters assigned lower pain scores using the observational VAS than less experienced raters using the modified infant pain scale when assessing pain in neonates after surgery. The discrepancy in scores was thought to reflect bias of experienced observers.²¹

Finally, there are limitations when the VAS is used in an observational capacity in children under 4 years of age, such as in this study. The observational VAS relies on a single item, whereas CRIES and PIPP are multivariate composite pain

assessment measures which may result in a more objective comprehensive assessment. Lawrence *et al*²⁰ had a similar finding when comparing the neonatal infant pain scale, a structured multivariate measure, with the observational VAS. When comparing the inter-rater reliability, they found that the unstructured observations of the VAS resulted in more variability in scores than the structured ratings of the neonatal infant pain scale ($r = 0.53-0.84$), and correlation was not established.

When the change in pain scores was examined over time, a consistent pattern emerged across all three measures (not influenced by surgical procedure or gestational age group). The highest pain scores occurred immediately after surgery, followed by a gradual decrease over the first 12 hours. Thereafter, pain scores remained relatively low throughout the study (< 2.2). It is unlikely that the raters had any bearing on the low scores. The raters were randomly assigned the pain measure just before assessment and performed assessments independently. The low pain scores are most likely due to the method of postoperative pain management. All infants received analgesics during the operative procedure (either fentanyl or morphine). Also, all infants received additional analgesics on return to the NICU, primarily continuous intravenous infusion morphine at a dose of 10–15 $\mu\text{g/kg/h}$. The influence of pain management strategies became more evident in the 48–72 hour postoperative period when scores started to rise slightly at the same time as 48% of infants had their analgesia either discontinued or converted from opioid to non-opioid treatment (from intravenous morphine to oral Tylenol).

Our division of infants into the two surgical categories, although carried out in consultation with our surgeons and using components of the criteria of Anand and Aynsley-Green,¹⁴ would not necessarily be easily replicated, as it was biased by our institution's practices and surgical techniques. This is a limitation of the study.

There were no significant main effects between gestational age groups or surgical groups. Furthermore no significant interactions were detected when gestational age groups and measure and surgical group and measure were compared. These results may have been influenced by the diverse nature of the study population, including diverse diagnoses and surgical procedures, and a convenience sample that was largely distributed in the greater than 36 weeks gestational age group.

CONCLUSIONS

In conclusion, no significant difference was detected in the slopes of change between the three measures. CRIES and PIPP correlate with VAS in the first 24 hours after surgery when term and preterm neonates are assessed. In addition, CRIES and PIPP are valid measures and are correlated when pain is assessed for the first 72 hours after surgery in term and preterm neonates. Both measures would provide an objective measure of a neonatal patient's pain after an operation. This is the first step to ensuring adequate postoperative pain management.

The pain scores in our study population were low, probably because of effective use of pharmacological methods of pain management. Data were not collected on non-pharmacological methods, and it would be interesting to determine how much the combined implementation of pharmacological and non-pharmacological interventions contributed to our low pain scores. Because of the nature of our NICU, our study population was heavily weighted towards the larger, closer to term infant. Therefore both measures require further evaluation with neonates of lower gestational age (< 32 weeks).

Authors' affiliations

C McNair, M Ballantyne, K Dionne, D Stephens, B Stevens, Hospital for Sick Children, Faculty of Nursing, University of Toronto, Toronto, Ontario, Canada

We would like to acknowledge funding received from the Grace Evelyn Simpson Reeves Award, Hospital for Sick Children Foundation, Toronto, Ontario, Canada

REFERENCES

- 1 Stevens B, Johnston C, Petryshen P, *et al*. The premature infant pain profile: development and initial validation. *Clin J Pain* 1996;**12**:13–22.
- 2 Gibbins S. *Safety and efficacy of sucrose for procedural pain relief in preterm and term neonates*. Toronto: University of Toronto, Faculty of Nursing, 2001.
- 3 Gibbins S, Stevens B, Hodnett E, *et al*. Efficacy and safety of sucrose for procedural pain relief in preterm and term neonates. *Nurs Res* 2002;**51**:375–82.
- 4 Johnston CC, Stremmler R, Horton L, *et al*. Effect of repeated doses of sucrose during heel stick procedure in preterm neonates. *Biol Neonate* 1999;**75**:160–6.
- 5 Stevens B, Johnston C, Franck L, *et al*. The efficacy of developmentally sensitive interventions and sucrose for relieving procedural pain in very low birth weight neonates. *Nurs Res* 1999;**48**:35–43.
- 6 Ballantyne M, Stevens B, McAllister M, *et al*. Validation of the premature infant pain profile in the clinical setting. *Clin J Pain* 1999;**15**:297–303.
- 7 Bildner J, Krechel S. Increasing staff nurse awareness of postoperative pain management in the NICU. *Neonatal Netw* 1996;**15**:11–16.
- 8 Krechel SW, Bildner J. CRIES: a new neonatal postoperative pain measurement score. Initial testing of validity and reliability. *Paediatr Anaesth* 1995;**53**:53–61.
- 9 Bours GJ, Abu-Saad HH, Hamers JP, *et al*. *Pain assessment in neonates: a state of the art study*. Limberg: University of Limberg, Faculty of Nursing, 1996.
- 10 Abu-Saad HH, Bours GJ, Stevens B, *et al*. Assessment of pain in the neonate. *Semin Perinatol* 1998;**22**:402–16.
- 11 Franck LS, Greenberg CS, Stevens B. Pain assessment in infants and children. *Pediatr Clin North Am* 2000;**47**:487–512.
- 12 Stevens B, Gibbins S, Franck LS. Treatment of pain in the neonatal intensive care unit. *Pediatr Clin North Am* 2000;**47**:633–50.
- 13 van Dijk M. Pain unheard? *Post-operative pain assessment in neonates and infants*. Rotterdam: De Erasmus Universiteit Rotterdam, Faculty of Nursing, 2001.
- 14 Anand KJS, Aynsley-Green A. Measuring the severity of surgical stress in newborn infants. *J Pediatr Surg* 1988;**23**:297–305.
- 15 McDowell I, Newell C. *Measuring health*. In: *A Guide to rating scales and questionnaires*. 2nd ed. New York: Oxford University Press, 1996.
- 16 Streiner DL, Norman G. *Health measurement scales*. Oxford: Oxford University Press, 1991.
- 17 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
- 18 van Dijk M, Bouwmeester NJ, Duivenvoorden HJ, *et al*. Efficacy of continuous versus intermittent morphine administration after major surgery in 0–3 year-old infants; a double-blind randomized controlled trial. *Pain* 2002;**98**:305–13.
- 19 Taddio A, Shah V, Gilber-MacLeod C, *et al*. Conditioning and hyperalgesia in newborns exposed to repeated heel lance. *JAMA* 2002;**288**:857–61.
- 20 Lawrence J, Alcock D, McGrath P, *et al*. The development of a tool to assess neonatal pain. *Neonatal Netw* 1993;**12**:59–66.
- 21 Buchholz M, Karl H, Pomietto M, *et al*. A. Pain scores in infants: a modified infant pain scale versus visual analogue. *J Pain Symptom Manage* 1998;**15**:117–24.