

A comparison of postoperative pain scales in neonates

S. Suraseranivongse¹*, R. Kaosaard², P. Intakong², S. Pornsiriprasert²,
Y. Karnchana², J. Kaopinpruck² and K. Sangjeen²

¹Department of Anesthesiology and ²Department of Nursing, Faculty of Medicine,
Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

*Corresponding author. E-mail: sisur@mahidol.ac.th

Background. Practical, valid and reliable pain measuring tools in neonates are required in clinical practice for effective pain management and prevention of the evaluator bias.

Methods. This prospective study was designed to cross-validate three pain scales: CRIES (cry, requires O₂, increased vital signs, expression, sleeplessness), CHIPPS (children's and infants' postoperative pain scale) and NIPS (neonatal infant pain scale) in terms of validity, reliability and practicality. The pain scales were translated. Concurrent validity, predictive validity and interrater reliability in postoperative pain were studied in 22 neonates after major surgery. Construct validity and concurrent validity in procedural pain were determined in 24 neonates before and during frenulectomy under topical anaesthesia.

Results. All scales had excellent interrater reliability (intraclass correlation >0.9). Construct validity was determined for all pain scales by the ability to differentiate the group with low pain scores before surgery and high scores during surgery ($P<0.001$). The positive correlations among all scales, ranging between $r=0.30$ and $r=0.91$, supported concurrent validity. CRIES showed the lowest correlation with other scales with correlation coefficients of $r=0.30$ and $r=0.35$. All scales yielded very good agreement ($K>0.9$) with routine decisions to treat postoperative pain. High sensitivity and specificity (>90%) for postoperative pain from all scales were achieved with the same cut-off point of 4. In terms of practicality, NIPS was the most acceptable (65%).

Conclusions. Based on our findings, we recommended NIPS as a valid, reliable and practical tool.

Br J Anaesth 2006; **97**: 540–4

Keywords: neonates; pain, procedural; pain, postoperative; pain, scale; tools, validity

Accepted for publication: June 4, 2006

Evidence from studies of neonatal neuroanatomy and neurochemistry and of functional ability to respond to painful stimuli has led us to believe that pain in neonatal patients should be assessed and treated.^{1,2} In the absence of objective tools to measure pain in this age group, assessment and treatment will be necessarily influenced by the knowledge base and personal bias of the evaluator.³ Several scoring systems validated for measuring postoperative pain in neonates have been developed. These include CRIES (cry, requires O₂, increased vital signs, expression, sleeplessness), CHIPPS (children's and infants' postoperative pain scale) and NIPS (neonatal infant pain scale).

CRIES⁴ is an acronym of five physiological and behavioural variables previously shown to be associated with neonatal pain (see Appendix in *British Journal of Anaesthesia* online). CHIPPS⁵ has high values of sensitivity (0.92–0.96) and specificity (0.74–0.95) to determine

postoperative analgesic demand. Physiological parameters were excluded because of unreliability and the lack of discriminating power to detect analgesic requirement (see Appendix in *British Journal of Anaesthesia* online). NIPS⁶ is a behavioural observational scale with a very simple scoring system. Six items, all are scored by 0 (no) or 1 (yes) except cry which has a 3 graded scale (0, 1, 2) (see Appendix in *British Journal of Anaesthesia* online).

This study aimed to identify the most valid, reliable, sensitive, specific and practical pain scale for routine use in clinical practice.

Methods

The three pain scores were translated. Cross-validations were performed to test the validity (construct validity, concurrent validity and predictive validity), reliability and

practicality of measures. Comparisons of validity, reliability and practicality among three pain scales—CRIES, CHIPPS and NIPS were also performed.

After obtaining approval from the Institution Review Board and informed written consent from parents, we conducted the study in four phases.

Phase 1: translation

The three pain scales were translated from English into Thai by an anaesthetist who was fluent in both languages. Then, another bilingual anaesthetist, who was not associated with the translation phase, translated the Thai version back into English. Finally, the back-translated scales were rechecked with the original scales by another translator whose mother tongue was English. Alterations were made on the basis of the third expert's opinion in order to produce the same meaning as the original scales.

Phase 2: testing of concurrent validity, predictive validity and interrater reliability in postoperative pain

Newborn infants admitted to the neonatal surgical intensive care unit postoperatively were enrolled in the study. Patients who received neuromuscular blocking agents were excluded. Patient characteristic data, operation, type of analgesic received and ventilator management were recorded. All nurses in this unit were trained to use CRIES, CHIPPS and NIPS by scoring the 12 neonatal behaviours of pain used in the scores from videotape until achieving a reliability of more than 0.9 as measured by intraclass correlations. Each infant was assessed hourly by three nurses. The period of assessment ranged between 24 and 72 h depending on the severity of the surgical procedure. At each assessment time, two nurses independently assessed the same infant by using three pain scales. The third nurse evaluated the infants making routine decisions to give analgesics based on inconsolable crying even after the other causes of distress were relieved.

Concurrent validity. Correlations between CRIES, CHIPPS and NIPS were tested at the same time points in all patients and separately for intubated and non-intubated patients.

Predictive validity in postoperative pain. The agreement kappa (K) between the second and third nurse for each of three pain scales was assessed. By considering the third nurse's decision to treat pain as a gold standard, sensitivity and specificity of each pain scale from the second nurse's score were determined. For each of the three scores the cut-off point for the treatment of pain that yielded the highest value of kappa, sensitivity and specificity was selected.

Interrater reliability. Reliability is a measure of consistency. Scores from each of the pain scales assessed by two observers (first and second nurse) were analysed for interrater reliability.

Phase 3: testing of construct validity and concurrent validity in procedural pain

Construct validity. This is an assessment of the meaning of the instrument in terms of its theoretical basis by comparison with external variables related to this construction. Neonates who underwent frenulectomy for tongue-tie under topical anaesthesia, as usual practice,⁷ were enrolled in this study. Lidocaine jelly was applied at frenulum ~3 min before starting the procedure. All behaviour during the placing of electrodes for electrocardiographic (ECG) monitoring and finger probe for SpO_2 monitoring was videotaped and recorded as a pain-free situation. Then, all behaviour during mouth opening and stretching, frenulum clamping and excision were also videotaped and recorded as a situation in which some discomfort is to be expected even with the use of local anaesthetic.

The chronological sequence of videotapes was rearranged into a new random sequence by using a random number table in order to blind the raters. Two nurses were trained to rate all pain scales. Pain scores from three pain scales in a pain-free period during monitoring were expected to be lower than those during an operation.

Concurrent validity in procedural pain. In order to minimize the concern about the contamination of scoring during repeated observation in the postoperative period, correlation of CRIES, CHIPPS and NIPS were tested before and during frenulectomy from the random sequence of videotapes.

Phase 4: practicality of measures

Nurses from a neonatal surgical intensive care unit were asked to rank the scores from the least likely '(0)' to 'the most likely' (10) according to the ease of use, time consumed, feasibility of their use in clinical situations, ability of the scales to differentiate the severity of pain and help in the decision to treat pain including general satisfaction with the scales. All comments regarding the content of pain scales were also recorded.

Statistical data analyses. Sample size estimation was based on estimates of the 95% confidence intervals (CI) of the true sensitivity and specificity of CHIPPS. A previous study of CHIPPS reported sensitivities from six studies ranging from 0.92 to 0.96 and specificities ranging from 0.74 to 0.95. It was expected that 95% CI of sensitivity of CHIPPS in this study would be 0.94 (0.05) and specificity of 0.85 (0.05). Using the formula $n = \frac{z_{\alpha}^2 pq}{d^2}$, where p =estimated sensitivity or specificity, $q=1-p$, d =allowable error (precision)=0.05 and α =probability of type I error=0.05 (2-sided); the required sample sizes in the painful and pain-free groups were 87 and 196, respectively. The value of d was set at 0.05 to obtain a width of 0.10 for the 95% CIs of sensitivity and specificity, i.e. 0.89–0.99, 0.80–0.90, respectively. However, as the incidence of pain-free observations was 25%,⁸ to get 196 pain-free observations, a total of 784 observations were needed. That is, 17 newborn infants were recruited and

each infant was observed every hour during a 48 h period. For the purposes of the power calculation observation at each time using three instruments, i.e. CRIES, CHIPPS, NIPS were treated as being independent. This was felt to be justified on the basis of the interval of 1 h between observations and the fact that significant pain was treated.

Patient characteristic data were presented as mean (SD) and median (range). Interrater reliability was analysed by intraclass correlation using a two-way random effect model. An intraclass correlation of >0.8 was considered acceptable. As all pain scores were ordinal data, construct validity was determined by using the Wilcoxon matched-pair signed-rank test to assess the difference in pain scores before and during surgery. The correlations among CRIES, CHIPPS and NIPS were analysed with a Spearman's rank correlation. The agreement of all pain scales from the second and third nurse at various cut-off points, corresponding to the decision to treat pain after surgery were analysed by using the Kappa (K) statistic. Values of K were interpreted as follows: <0.2, poor agreement; 0.21–0.4, fair agreement; 0.41–0.6, moderate agreement; 0.61–0.8, good agreement; and 0.81–1.0, very good agreement.⁹ The practicality of the scales was analysed by using descriptive statistics. All analyses were performed with SPSS for Windows V.11.5 (SPSS, Chicago, IL, USA).

Results

Among the 22 neonates enrolled in the study, there were 13 boys (59.1%), median age of 1 day (range 1–23 days), mean gestational age of 39.9 weeks (SD 2.3 weeks) and mean body weight of 2409 g (SD 488 g). One thousand and twenty-seven observations were performed. Fifty per cent of patients were intubated and ventilatory support was provided after surgery.

Concurrent validity was assessed in terms of correlations between scores in all patients, intubated patients and non-intubated patients (Table 1). In the postoperative

period, the correlation of CHIPPS with NIPS was good, ranging between $r=0.84$ and $r=0.88$ for the various comparisons. However, the correlations of CRIES with CHIPPS (ranging between $r=0.30$ and $r=0.38$) or NIPS (ranging between $r=0.32$ and $r=0.39$) were fair. Strong correlations between all scales ($r>0.8$) were recorded before and during frenulectomy (Table 2). The interrater reliability of all pain scales was excellent. Intraclass correlation coefficient (95% CI): CRIES=0.98 (0.97–0.98), CHIPPS=0.93 (0.93–0.94), NIPS=0.98 (0.98–0.98).

Construct validity was demonstrated in 24 neonates, 14 boys (58.3%), median age of 5 days (range 3–27 days) who underwent frenulectomy for correction of tongue-tie. There was a significant difference in pain scores while monitoring before surgery and during surgery. The median pain scores (Interquartile range, IQR) during monitoring were lower than during surgery (Table 3).

Table 2 Spearman's rank correlation among CRIES, CHIPPS and NIPS in all patients for procedural pain ($n=48$ episodes). *All P -values were <0.001

	Nurse A*			Nurse B*		
	CRIES	CHIPPS	NIPS	CRIES	CHIPPS	NIPS
CRIES	1.00	0.89	0.80	1.00	0.88	0.85
CHIPPS		1.00	0.85		1.00	0.91
NIPS			1.00			1.00

Table 3 Construct validity of pain scores before and during surgery. Wilcoxon matched-pair signed-rank test. *All P -values were <0.001

Pain scales	Observer	Pain scores: median (IQR)	
		Before surgery	During surgery
CRIES	A	3 (1–5)	8 (7–8)*
	B	3.5 (2–5)	8 (6–8)*
CHIPPS	A	2.5 (1–5)	9 (8–9)*
	B	2.5 (1–4)	8.5 (7–9)*
NIPS	A	0 (0–4)	5 (5–5)*
	B	1 (0–3)	6 (5–6)*

Table 4 Cut-off points which yielded the highest agreement with the clinical decision to treat postoperative pain including sensitivity and specificity in all patients, intubated and non-intubated patients. Agreement between second and third nurse. †The third nurse's decision was considered as a gold standard

Patients	Pain scale	Clinical decision to treat pain			
		Cut-off point	Kappa	Sensitivity (%) [†]	Specificity (%) [†]
All patients	CRIES	4	0.92	96.5	99.6
	CHIPPS	4	0.93	96.6	99.7
	NIPS	4	0.93	96.6	99.7
Intubated patients	CRIES	4	0.96	100	99.8
	CHIPPS	4	1.00	100	100
	NIPS	4	0.92	92.3	99.8
Non-intubated patients	CRIES	4	0.88	93.8	99.4
	CHIPPS	4	0.88	93.8	99.4
	NIPS	4	0.94	100	99.6

Table 1 Spearman's rank correlation among CRIES, CHIPPS and NIPS in patients during postoperative period. *All P -values were <0.001

	Nurse A*			Nurse B*		
	CRIES	CHIPPS	NIPS	CRIES	CHIPPS	NIPS
All patients (1027 observations)						
CRIES	1.00	0.35	0.35	1.00	0.30	0.34
CHIPPS		1.00	0.88		1.00	0.85
NIPS			1.00			1.00
Intubated patients (528 observations)						
CRIES	1.00	0.31	0.32	1.00	0.33	0.32
CHIPPS		1.00	0.88		1.00	0.88
NIPS			1.00			1.00
Non-intubated patients (499 observations)						
CRIES	1.00	0.30	0.37	1.00	0.38	0.39
CHIPPS		1.00	0.84		1.00	0.88
NIPS			1.00			1.00

Table 5 Practicality scores of pain scales (0 = least, 10 = most)

Items	Mean (sd)		
	CRIES	CHIPPS	NIPS
Easy of use	7.0 (1.9)	8.5 (1.2)	9.1 (0.7)
Duration of rating	6.2 (2.2)	4.5 (3.0)	4.5 (3.0)
Difficulty in assessment	6.5 (2.5)	4.5 (3.3)	4.5 (3.2)
Help in decision to treat	7.9 (1.0)	8.2 (1.4)	8.7 (0.9)
Feasible for clinical practice	7.5 (2.0)	8.0 (1.2)	8.7 (1.1)
Able to differentiate the severity of pain	8.0 (1.6)	7.9 (1.7)	8.7 (1.1)

The predictive validity of the pain scales during persistent pain after surgery is reported in Table 4. The cut-off point which yielded the best agreement with the clinical decision to treat postoperative pain was four for all three pain scales in all neonates both intubated and non-intubated.

In terms of practicality, NIPS was reported to be superior to CHIPPS and CRIES against all criteria (Table 5). Furthermore, from the global rating for routine use, NIPS was the most selected ($N=20$: CRIES=20%, CHIPPS=15% and NIPS=65%).

The content of NIPS was accepted totally by all nurses. The usefulness of items of CRIES was questioned. They were 'the requirement of oxygen to maintain $Sp_{O_2} >95\%$ ' and 'increased vital signs'. Several nurses doubted the value of 'posture of the trunk: rear up' in the CHIPPS score.

Discussion

The three pain scales had excellent interrater reliability, demonstrable concurrent validity and construct validity and good predictive validity.

The reproducibility and consistency of all three pain scales were demonstrated by excellent interrater reliability; all intraclass correlation coefficients were >0.8 . The positive correlations of the scales with each other support concurrent validity. In the postoperative period with persistent pain, the behavioural observational scales such as CHIPPS and NIPS showed good correlation, whereas CRIES showed fair correlation with the other two scales. The difference might be as a result of physiological measurement of CRIES.

There could have been cross-over or contamination of scoring in repeated observations after surgery which might increase the strength of correlation. This effect should be minimal because the content of three pain scales was not difficult to score. Furthermore, we tried to avoid these effects by testing correlation from a new random sequence of behaviour from videotapes before and during frenulectomy. The results also supported concurrent validity with higher correlation of all measures in procedural discomfort.

The construct validity of all pain scales was determined by comparing the group experiencing no pain in a baseline situation before surgery with the group experiencing a discomfort during minor surgery. Even though some infants showed emotional distress behaviour during placing of the ECG electrode and Sp_{O_2} finger probe which might

have falsely increased the pain score. The scores of all scales during surgery were still clinically and statistically higher.

In our study, predictive validity was tested in terms of sensitivity, specificity and for predicting the routine clinical decision to treat pain after surgery. The nurses' assessments and their decision to treat pain are not a perfect gold standard, but they were the best standard of treatment available in our routine practice. The predictive validity of all measures in postoperative pain yielded very good agreement, high sensitivity and specificity in both intubated and non-intubated patients.

Pain assessment and treatment decisions may be influenced by practice settings¹⁰ and the characteristics of the providers such as age, education and personal pain experience.¹¹ Research has shown that the use of a standardized pain assessment tool results in providing ratings of pain that more closely match the child ratings.¹² In order to implement pain scales in clinical practice, cut-off points are necessary for the decision to treat. From the original study⁵ CHIPPS has a cut-off point of 4, which meant that in the case of definite pain the score for CHIPPS was never below 4 points. Our study also showed similar results, the CHIPPS score was 2.5 in a pain-free situation before surgery and in a painful situation it was 8.5–9. Cut-off points were 4 for postoperative pain in both intubated and non-intubated neonates. The sensitivity and specificity of CHIPPS from our result were just as good as in the previous study.⁵

The cut-off points of CRIES and NIPS were not reported in the original studies. Our findings showed a cut-off point of 4 in both scales which also yielded very good sensitivity and specificity.

There were some limitations in this study. First, 50% of our postoperative patients ($n=22$) were intubated and ventilated during the period of study. In our practice, all neonates who underwent major operations received a fentanyl infusion for 12–24 h. After stopping the infusion, 4 of the 11 intubated patients were diagnosed to have pain and received treatment. Even though all these pain scales have not been previously validated in intubated neonates, in non-paralysed patients, grimaces and cries could be assessed. The score from 'cry' might be reduced. Nevertheless, the total scores were still higher than the cut-off point and indicated a requirement for analgesia. Our results demonstrated similar concurrent and predictive validity in intubated and non-intubated neonates.

Second, we could not eliminate the bias of observers in rating pain-free and painful situations despite using rearranged pictures of videotape because we were not able to blind all the procedures during videotape recording.

Concerning practicality, NIPS was the most satisfactory pain scale which nurses selected to use in routine clinical practice because of the ease and feasibility of use. This included the ability to differentiate the severity of pain.

The nurses' comments suggested that NIPS was the only tool which was appropriate on the basis of content, relevance and coverage. CRIES was disliked on the basis of two items. First, 'the requirement of oxygen to maintain

SpO₂ >95%' was invalid because in our hospital newborn infants were routinely taken care of in an incubator with 30% oxygen after surgery; therefore, saturations might not be related to pain. Moreover, vigorous movement of the limbs might falsely reduce the saturation. Second, 'increased vital signs: arterial pressure or heart rate' was felt to be invalid as these changes were unreliable, they could be caused by other factors. Furthermore, the baseline level of preoperative vital signs varied as a result of several factors, such as hunger, discomfort and normal physiological variation. CHIPPS was felt to be unsatisfactory in respect of 'posture of the trunk: rear up' as this behaviour could be seen in neonates without pain who lay in a prone position.

On the basis of our findings, CRIES, CHIPPS and NIPS were all valid and reliable. However, NIPS was the most practical scale because the items were easy to score and there was no need to calculate the change of vital signs, which was an obstacle in a busy clinical practice with limitations of manpower. Furthermore, pulse oximeter is not commonly available in our hospital with limited resource. Therefore, we recommended using NIPS to assess pain in newborn infants in the postoperative period.

Supplementary data

The Appendix can be found at *British Journal of Anaesthesia* online.

Acknowledgements

The authors wish to thank the Siriraj Research Fund for financial support and Dr Chulaluk Komoltri for her invaluable suggestions in statistical analyses.

References

- 1 Fitzgerald M, Anand KJS. Development neuroanatomy and neurophysiology of pain. In: Schechter NL, Berde CB, Yaster M, eds. *Pain in Infants, Children and Adolescents*. Baltimore: Williams and Wilkin, 1993; 11–3
- 2 Giannakouloupoulos X, Sepulveda W, Kourtis P, Glover V, Fisk NM. Fetal plasma cortisol and beta-endorphin response to intrauterine needling. *Lancet* 1994; **344**: 77–81
- 3 Page GG, Halvorsen M. Pediatric nurses: the assessment and control of pain in pre-verbal infants. *J Pediatr Nurs* 1991; **6**: 99–106
- 4 Krechel SW, Bildner J. CRIES: a new neonatal postoperative pain measurement score. Initial setting of validity and reliability. *Paediatr Anaesth* 1995; **5**: 53–61
- 5 Buttner W, Finke W. Analysis of behavioural and physiological parameters for the assessment of postoperative analgesic demand in newborns, infants and young children: a comprehensive report on several consecutive studies. *Paediatr Anaesth* 2000; **10**: 303–18
- 6 Lawrence J, Alcock D, McGrath P, Kay J, MacMurray SB, Dulberg C. The development of a tool to assess neonatal pain. *Neonatal Netw* 1993; **12**: 59–65
- 7 Amir LH, James JP, Beatty J. Review of tongue-tie release at a tertiary maternity hospital. *J Paediatr Child Health* 2005; **41**: 243–5
- 8 Mather L, Mackie J. The incidence of postoperative pain in children. *Pain* 1983; **15**: 271–82
- 9 Altman DG. Some common problems in medical research. In: Altman DG, ed. *Practical Statistics for Medical Research*. London: Chapman and Hall, 1991; 404–8
- 10 Burokas L. Factors affecting nurses decision to medicate pediatric and adult patients after surgery. *Heart Lung* 1985; **14**: 373–9
- 11 Bradshaw C, Zeanah PD. Pediatric nurses assessment of pain in children. *J Pediatr Nurs* 1986; **1**: 314–22
- 12 Colwell C, Clarke L, Perkins R. Postoperative use of pediatric pain scale: children's self report versus nurse assessment of pain intensity and affect. *J Pediatr Nurs* 1996; **11**: 375–82