Journal of Perinatology (2010) 30, 474–478

www.nature.com/jp

ORIGINAL ARTICLE

Validity and reliability of the N-PASS assessment tool with acute pain

P Hummel[1], P Lawlor-Klean[2] and MG Weiss[3]

[1]Department of Neonatology and Pediatrics, Gloria Smith/American Nurses Foundation Scholar, Loyola University Medical Center, Maywood, IL, USA; [2]Department of Neonatology, Neonatal Clinical Nurse Specialist, Advocate Christ Medical Center, Oak Lawn, IL, USA and [3]Department of Neonatology, Loyola University Medical Center, Maywood, IL, USA

Objective: To establish evidence of clinical validity and reliability of the Neonatal Pain, Agitation, and Sedation Scale (N-PASS) in neonates with acute heelstick pain.

Study Design: Prospective psychometric evaluation, randomized crossover design. Two nurses administered the N-PASS simultaneously and independently during an actual and sham heelstick done in randomized order. One nurse also administered the Premature Infant Pain Profile (PIPP) concurrently with the N-PASS. Heelsticks were videotaped for repeat analysis.

Result: Construct (discriminate) validity was established through the Wilcoxon Signed-ranks test, comparing the distribution of the heelstick and sham N-PASS scores. The mean pain scores were 3.93 (2.30) and 0.81 (1.21) for the heelstick and sham procedures, respectively ($Z = -6.429$, $P<0.0001$). Convergent validity was demonstrated by correlation with the PIPP scores (Spearman rank correlation coefficient of 0.75 and 0.72 for raters 1 and 2, respectively). Inter-rater reliability was high, measured by intra-class coefficients; the ICC estimates (95% CI) of the pain scale were 0.86 (0.78, 0.92) and 0.93 (0.88, 0.96) for a single rating and average of two independent ratings, respectively ($P<0.0001$). Internal consistency, measured by Cronbach's alpha, was evident (0.84 to 0.89). Test–retest reliability was demonstrated by repeat scoring of videotaped heelsticks, measured by Spearman's rho correlation (0.874, $P<0.0001$).

Conclusion: This research provides beginning evidence that the N-PASS is a valid and reliable tool for assessing acute heelstick pain in infants 0 to 30 days of age, 23 to 40 weeks gestation.
Journal of Perinatology (2010) **30,** 474–478; doi:10.1038/jp.2009.185; published online 19 November 2009

Keywords: neonatal pain assessment; newborn pain; premature infant pain; infant pain

Correspondence: P Hummel, Department of Neonatology and Pediatrics, Loyola University Medical Center, 2160 South First Avenue, Maywood, IL 60153, USA.
E-mail: phummel@lumc.edu
Received 29 April 2009; revised 4 October 2009; accepted 11 October 2009; published online 19 November 2009

Introduction

Infants in the neonatal intensive care unit (NICU) experience prolonged pain and distress postoperatively and due to mechanical ventilation, as well as from painful procedures, occurring a mean of 14 times per day, and up to over 400 times during the hospitalization of an extremely preterm infant.[1,2] Heelstick blood collection is the most commonly performed painful procedure in the NICU.

Behavioral pain assessment is an essential nursing function, performed multiple times daily, contributing to clinical management and evaluation of therapies. Several behavioral neonatal pain measurement tools, developed over the past 20 years, are currently used in research and in the clinical setting. Neonatal pain assessment reviews concur that pain assessment tools should be multidimensional as well as valid and reliable.[3–5] Pain assessment tools score the infant's pain by assigning numeric values to the presence and/or intensity of selected pain behaviors.

Preterm infants are capable of discriminating between sham and actual heelsticks. Johnston *et al*. found that preterm infants' maximum heart rate and upper facial action contributed significantly to the behavioral differences found when comparing the response to real and sham heelsticks. Infants of lower gestational age had lower response than the more mature infants.[6] More recent research by Gibbins *et al*.[7] found that extremely preterm infants responded to acutely painful stimuli with increased facial action and heart rate, and decreased oxygen saturation; the response was dampened as compared with more mature infants.

Pain in the NICU can be categorized into three types (1) acute-procedural pain, (2) acute-prolonged pain, and (3) chronic pain.[8] The Neonatal Pain, Agitation, and Sedation Scale (N-PASS) was developed as a clinically relevant tool to assess prolonged pain and sedation in infants, as well as acute-procedural pain. Validity and reliability of the N-PASS with prolonged pain has been described.[9]

Procedural pain assessment has been well studied in the neonate; most pain assessment tools were validated with procedural pain.[10] The Premature Infant Pain Profile (PIPP)[11] has been researched and used extensively with procedural pain. Although

there is no need for another neonatal acute-procedural pain scale, there is a recognized need to validate existing behavioral pain and distress scales for procedural use.[12] This research was designed to investigate the validity and reliability of the N-PASS tool with acute-procedural pain, specifically, heelstick blood collection.

Methods

Subjects

This is a convenience sample of infants 23 to 42 weeks gestation at birth and <30 days old at the time of the procedure. An infant was evaluated up to two times for this study. Exclusion criteria included (1) critical condition or medical instability, where the observation and sham process is deemed to possibly cause undue stress, as determined by the bedside nurse, (2) pharmacologic paralysis, (3) sedation deemed significant enough to decrease the infant's response to the heelstick procedure, as determined by an N-PASS sedation score of $\leqslant -2$ at the previous sedation assessment point, and (4) neurologic status that causes abnormal pain perception and/or response, such as hypoxic-ischemic-encephalopathy or spina bifida.

Setting

Data were collected at two Level III NICUs, 50 and 40 beds, respectively, both caring for infants 22/23 weeks gestation at birth and above. Eighty-three percent of the data collections occurred in one center.

Measures/instruments

The N-PASS. The N-PASS was developed as a clinically relevant tool to assess ongoing or acute pain in neonates and infants, as well as sedation levels. The five indicators are (1) crying/irritability (silent cry observed in the intubated infant is scored as a cry), (2) behavior/state, (3) facial expression, (4) extremities/tone, and (5) vital signs. Points are added to the preterm infant's pain score to approximate the normal response of a full-term infant. These points were derived from the PIPP pain assessment tool.[11] Gestational age groups are <28 weeks (3 points), 28 to 31 weeks (2 points), 32 to 35 weeks (1 point), and >35 weeks (0 points). Scores range from 0 to 13. Behavioral state as a modifier was not included in the development of N-PASS, as it was designed initially for ongoing pain assessment. The pain assessment portion of the tool was tested in this research study of acute pain.

The PIPP. The PIPP tool was used to establish criterion validity of the N-PASS. The PIPP was developed and validated for the assessment of acute pain.[11] Indicators include heart rate elevation and oxygen desaturation, plus three components of facial activity. Two modifiers, gestational age and pre-procedural behavioral state are included. Scores range from 0 to 21.

Human subjects

Institutional review board approval was obtained. At least one parent was contacted in person or by phone by a researcher to obtain consent. Signed consent was obtained after the study risks and benefits were explained. Confidentiality with videotaping was maintained by marking the tape with the study number, with no other identifiers.

Procedure

The N-PASS was studied with routine heelstick procedures in the NICU, compared with a sham heelstick procedure. The bedside nurse randomly determined the order of events through coin toss. A researcher and a trained staff nurse scored the infant with both the sham and real heelstick procedures using the N-PASS tool. One observer also scored the infant using the PIPP tool. The infant was assessed before and during each procedure. The observers were trained in the use of both tools. When scoring both tools, the raters were instructed to watch the entire procedure, then score using the N-PASS and PIPP.

An infant blanket or screen was placed over the incubator or crib, obstructing the view of the lower half of the baby's body, to blind the observers to the procedure, while still allowing the bedside nurse to perform the procedure without additional stress to the infant. Videotaping included the upper half of the infant's body only.

The sham heelstick consisted of the nurse holding the foot as if doing a heelstick, wiping the heel gently with a dry 2×2 gauze, and holding the heel in place for at least 30 s. No lancing, squeezing of the heel, or alcohol/other prep was allowed with the sham procedure. The blood collection supplies were used as if in an actual heelstick, but no blood was obtained. The actual heelstick consisted of the standard NICU procedure of holding the foot in the usual position, prepping and collecting blood.

The second procedure, whether real or sham, occurred after the infant was quieted following the first procedure. The interval between procedures was between 5 and 10 min. The observers watched the infant's behavior and monitored vital signs, to enable scoring with the N-PASS and the PIPP tools. The assessment occurred before any intervention (baseline) and was repeated with each real and sham heelstick. Thirteen heelstick and sham procedures were videotaped with parental consent to enable intra-rater reliability (test–retest) analysis.

Statistical procedures

Sample size. The sample size and power calculations were done in PASS2002. Data from earlier research of the N-PASS tool indicated high inter-rater reliability as measured by the intra-class correlation coefficient (ICC) (0.92 to 0.97).[4] A sample size of 11 neonates (and two raters) will achieve 80% power to detect a difference of 0.15 between the null hypothesis ICC of 0.8 and the alternative hypothesis ICC of 0.95 using an F-test with a significance level of 0.05.

A sample size of 11 achieves 80% power to detect a difference of 3 between the null hypothesis that both the real and sham heelstick treatments have an average N-PASS pain score of 5.0 and the alternative hypothesis that the sham group has a mean of 2.0 with estimated standard deviations of 3.0 and 3.0 and with a significance level (alpha) of 0.05 using a two-sided Wilcoxon Signed-ranks test. Testing the reliability and validity of the N-PASS instrument in each of four age groups necessitated 11 observations in each age group. Thus, a total sample size of 44 observations will provide an adequate power for this study.

Statistical analyses

The primary analysis focused on the reliability and validity of the sum score derived from the N-PASS instrument. Two measures of instrument validity were applied (1) Spearman rank correlation between the N-PASS and the PIPP as a measure of convergent validity and (2) Wilcoxon Signed-ranks test to compare the distribution of N-PASS scores between the heelstick and sham heelstick procedures, as a measure of discriminate construct validity.

Three measures of instrument reliability were used (1) ICC, as a measure of inter-rater reliability; (2) Spearman's rho as a measure of test–retest reliability; and (3) Cronbach's alpha, as a measure of internal consistency. The ICC corresponds to a one-way random effects model, with each neonate representing a level of the random person factor. The ICC(1,2) model produces two different ICC estimates: one for the reliability of a single rating, and one for the reliability of the mean of two ratings. Recorded videos were used to evaluate test–retest reliability of the N-PASS instrument. A nurse viewed the video of each neonate and an N-PASS pain score was recorded. One week and also at one year later, the same nurse re-evaluated the video of each neonate and recorded a second N-PASS pain score. The first and second N-PASS scores were analyzed using Spearman's rho statistic.

Counts and percentages are reported for categorical variables and range, median, mean, and standard deviation for continuous/ordinal data. A Spearman's rank correlation coefficient was used to evaluate associations between continuous/ordinal variables and N-PASS score. The Kruskal–Wallis or Mann–Whitney test was applied as appropriate to compare N-PASS scores between categorically defined groups.

Results

Fifty-nine observations were collected on 42 infants. Some infants were observed twice, on different days. As the observers were not the same for the two observations, the data were considered independent and analyzed as such. Infant age ranged from 1 to 30 days. Males comprised 50.8%, females 49.2%. Gestational age ranged from 23 to 40 weeks at birth, with each gestational age group represented, consisting of 13 to 16 per group (Table 1). Twenty-seven percent of the subjects were mechanically ventilated at the time of observation, mostly in the youngest gestational age group. Thirty-six percent were on continuous positive airway pressure or high-flow humidified nasal cannula.

Discriminate validity

Wilcoxon Signed-ranks test was used to compare heelstick to sham scores. The mean heelstick pain score was 3.93 (2.30); the mean sham score was 0.81 (1.21) ($Z = -6.429$, $P<0.001$). See Table 2 for the Wilcoxon Signed-ranks test calculated for each gestational age group.

Convergent validity

The Spearman rho correlation between the heelstick scores of the N-PASS pain scale and the PIPP scale was 0.743 ($P<0.001$). See Table 3 for Spearman rho correlations calculated for each gestational age group.

Inter-rater reliability

For the heelstick observation of a neonate, the ICC estimates (95% CI) of the pain scale were 0.86 (0.78, 0.92) and 0.93 (0.88, 0.96) for a single rating and average of two independent ratings,

Table 1 Sample size by gestational age categories

Weeks gestation	*Frequency*	*Percent*	*Cumulative percent*
36–40	14	23.7	23.7
32–35	16	27.1	50.8
28–31	13	22.0	72.9
<28	16	27.1	100.0
Total	59	100.0	

Table 2 Wilcoxon Signed-ranks test comparing heelstick to sham per gestational age group

Gestational age group	*Wilcoxon Signed-ranks test Z (p)*
0 (>35 weeks)	−3.391 (0.001)
1 (32–35 weeks)	−3.450 (0.001)
2 (28–31 weeks)	−3.089 (0.002)
3 (<28 weeks)	−3.193 (0.001)

Table 3 Spearman rho correlations between the N-PASS and PIPP per gestational age group (p)

Gestational age group	*Spearman rho correlation*
0 (>35 weeks)	0.774 (0.001)
1 (32–35 weeks)	0.708 (0.002)
2 (28–32 weeks)	0.800 (0.001)
3 (<28 weeks)	0.767 (0.001)

Abbreviations: N-PASS, Neonatal Pain, Agitation, and Sedation Scale; PIPP, Premature Infant Pain Profile.

respectively. Similarly, for the sham observation, the ICC estimates (95% CI) of the pain scale were 0.79 (0.67, 0.87) and 0.88 (0.80, 0.93) for a single rating and average of two independent ratings, respectively. F-tests, carried out to test the null hypothesis that the ICC (single rating) was 0.6 or less (versus >0.6) were highly significant (heelstick $F = 13.70$, $P<0.0001$ and sham $F = 8.56$, $P<0.0001$).

Internal consistency

The mean sham heelstick pain scores were 0.83 (1.25) and 1.09 (1.71) for raters 1 and 2, respectively. The mean heelstick pain scores were 3.74 (2.35) and 3.5 (2.48) for raters 1 and 2, respectively. Cronbach's alpha for the 5-item pain scale was 0.835 and 0.887 for raters 1 and 2, respectively.

Test–retest reliability

Scores of 13 videotaped heelstick and sham procedures 1 week apart yielded a Spearman's rho correlation coefficient of 0.874, ($P<0.0001$). The videotapes were scored again 1 year after the initial evaluation, yielding a Spearman's rho correlation coefficient of 0.846 ($P<0.0001$).

Prematurity

Table 4 shows the mean heelstick N-PASS pain scores for the infants grouped by gestational age, without adding the prematurity points. Differences between the four gestational age groups were not significant, as analyzed with the Kruskal–Wallis test ($H = 4.818$, $P = 0.186$). Mann–Whitney analysis of the most and least mature gestational age groups showed a nonsignificant difference in the pain score, with mean ranks of 18.64 and 12.75, respectively ($Z = -1.852$, $P = 0.07$). Mann–Whitney analysis comparing the two highest age groups with the two lowest age groups showed a significant difference, with mean ranks of 34.55 and 25.29, respectively ($Z = -2.09$, $P = 0.037$). Pearson correlation between gestational age in weeks and the pain score was significant, 0.292, ($P = 0.025$).

Order of procedures

The N-PASS scores were compared between the group of infants receiving the heelstick first ($n = 28$), as compared with those receiving the sham procedure first ($n = 31$) using the Mann–Whitney *U*-test. The mean ranks were not significantly different, 28.74 and 29.43, respectively ($Z = -0.114$, $P = 0.910$).

Table 4 N-PASS scores by gestational age groups

Weeks gestation	*Mean heelstick N-PASS score*	*N-PASS score range*	*Standard deviation*
36–40	4.714	2–7	1.978
32–35	4.375	1–10	2.527
28–31	3.385	1–9	2.468
<28	3.250	0–7	2.049

Abbreviation: N-PASS, Neonatal Pain, Agitation, and Sedation Scale.

Discussion

Many acute-procedural pain assessment tools have been validated; there is no need for the development of additional tools to assess acute pain in the neonate. The N-PASS is widely used for the assessment of ongoing pain in neonates; therefore it is appropriate to validate the scale for use in acute-procedural pain, contributing to the clinical ease of using one tool for both ongoing and acute pain.

The order of the procedure did not have a significant effect on the pain score, despite the relatively short interval between procedures. The impact of prior painful stimuli or stressful events on the reactivity of infants to acute pain has been researched. In one study, infants who were younger, asleep, and had a more recent painful event were less likely to mount a behavioral response to pain.[13] Conversely, other research with healthy preterm and full-term newborns showed an increased reactivity to subsequent painful procedures.[14]

The addition of points to the premature infant's pain score is based on the research-supported premise that premature infants are less able to exhibit signs of pain than the term infant.[6,13] Correlation between gestational age and N-PASS pain score supports these findings, with a low but significant correlation. Mean scores for each gestational age group are similar without prematurity points added, with no significant differences in mean pain scores between gestational age groups. Earlier research of the N-PASS with ongoing pain also did not support the current method of adding points for prematurity.[9] These statistics do not support the current method of adding points for prematurity. There was a significant difference per Mann–Whitney analysis between the scores of the groups when divided at a gestational age of 30 weeks, with the younger group having small but significantly lower scores ($Z = -2.52$, $P = 0.012$). The examination of the application of gestational age as a modifier of pain response is appropriate in this research, since the pain source, the heelstick, is the same for all infants. The findings indicate that the points for prematurity, if used, should be restructured so that infants under 30 weeks receive a 1-point correction for prematurity, as compared with those 30 weeks and greater. An alternative to using gestational age as a modifier in the clinical setting is to score the N-PASS without adding the point for prematurity, but using a lower threshold for treatment for the extremely premature infant. The updated N-PASS is available at http://www.n-pass.com.

Limitations

The N-PASS was studied in the clinical setting rather than laboratory methods due to the goal of developing a clinically useable tool. Bias was controlled as much as possible by blinding

the observers to the procedures. The findings should be interpreted cautiously due to the moderate sample size. The generalizability is limited to two NICUs in the Midwest. Further testing is required in other populations and settings and with larger samples to validate the findings.

Further research

The N-PASS tool requires validity and reliability testing with various nonverbal populations and management styles. Pain assessment of pharmacologically paralyzed and neurologically compromised infants is clinically challenging, requiring further research.

Conclusion

This research provides beginning evidence that the N-PASS is a clinically useable, reliable, and valid tool to assess acute-procedural heelstick pain in neonates. The previously used gestational age modulation scores are modified to reflect that infants under 30 weeks gestation have a slightly lower behavioral response.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We thank the staff nurses, nurse practitioners, and physicians in the two NICU's for their cooperation in this research, especially those nurses who collected data. The primary author was funded by the American Nurses Foundation, Gloria Smith Scholar.

References

1 Barker DP, Rutter N. Exposure to invasive procedures in neonatal intensive care unit admission. *Arch Dis Child Fetal Neonatal Ed* 1995; **72**: 47–48.
2 Simons SH, van Dijk M, Anand KS, Roofthooft D, van Lingen RA, Tibboel D. Do we still hurt newborn babies? A prospective study of procedural pain and analgesia in neonates. *Arch Pediatr Adolesc Med* 2003; **157**: 1058–1064.
3 Duhn LJ, Medves JM. A systematic integrative review of infant pain assessment tools. *Adv Neonatal Care* 2004; **4**: 126–140.
4 Byers JF, Thornley K. Cueing into infant pain. *MCN* 2004; **29**: 84–89.
5 Hummel P, van Dijk M. Pain assessment: current status and challenges. *Semin Fetal Neonatal Med* 2006; **11**: 237–245.
6 Johnston CC, Stevens BJ, Yang F, Horton L. Differential response to pain by very premature neonates. *Pain* 1995; **61**: 471–479.
7 Gibbins S, Stevens B, McGrath PJ, Yamada J, Beyene J, Breau L *et al.* Comparison of pain responses in infants of different gestational ages. *Neonatology* 2008; **93**: 10–18.
8 Stevens BJ, Pillai RR, Oberlander TE, Gibbins S. Assessment of pain in neonates and infants. In: Anand KJS, Stevens BJ, McGrath PJ (eds). *Pain in Neonates and Infants*. 3rd edn. Elsevier: Edinburgh, 2007. pp 67–90.
9 Hummel P, Puchalski M, Creech SD, Weiss MG. Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. *J Perinatol* 2008; **28**: 55–60.
10 Duhn LJ, Medves JM. A systematic integrative review of infant pain assessment tools. *Adv Neonatal Care* 2004; **4**: 126–140.
11 Stevens B, Johnston C, Petryshen P, Taddio A. Premature Infant Pain Profile: development and initial validation. *Clin J Pain* 1996; **12**: 13–22.
12 Crellin D, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. *Paediatr Anaesth* 2007; **17**: 720–733.
13 Johnston CC, Stevens BJ, Franck LS, Jack A, Stremler R, Platt R. Factors explaining lack of response to heel stick in preterm newborns. *J Obstet Gynecol Neonatal Nurs* 1999; **28**: 587–594.
14 Porter FL, Wolf CM, Miller JP. The effect of handling and immobilization on the response to acute pain in newborn infants. *Pediatrics* 1998; **102**: 1383–1389.