

# Agreement of Revised Premature Infant Pain Profile Scoring between Healthcare Providers and Laypersons: A Cross-sectional Study

DIPEN V PATEL<sup>1</sup>, MEGHA J ANANT<sup>2</sup>, BHADRA Y TRIVEDI<sup>3</sup>, AJAY GAJANAN PHATAK<sup>4</sup>, MAYUR K SHINDE<sup>5</sup>, RESHMA K PUJARA<sup>6</sup>, SOMASHEKHAR M NIMBALKAR<sup>7</sup>



## ABSTRACT

**Introduction:** The experience of pain during the neonatal period has short and long-term consequences. The Revised Premature Infant Pain Profile (PIPP-R) is a globally accepted and validated tool for assessing pain in neonates. Adequate pain management measures can be implemented using the PIPP-R, even in the absence of consultants.

**Aim:** To assess the agreement among healthcare providers and laypersons in scoring the PIPP-R.

**Materials and Methods:** A cross-sectional study was conducted at Shree Krishna Hospital, a rural Tertiary Care Teaching Hospital in central Gujarat, India. The duration of the study was one year and six months, from January 2021 to June 2022. The study included 12 volunteers from various fields, such as consultant neonatologists, first year postgraduate students in Department of Paediatrics, neonatal nurses, social workers, Bachelor of Medicine, Bachelor of Surgery (MBBS) interns, and mothers of newborns. A neonatology consultant provided training on the PIPP-R scoring system using handouts and a presentation. The volunteers then evaluated 100 prerecorded videos of newborns undergoing painful procedures. Agreement between volunteers for the total PIPP-R score and its subcomponents was assessed using Bland-Altman analysis and Cohen's Kappa statistics.

**Results:** A total of 100 videos of newborns (51 girls, 49 boys) undergoing painful procedures were evaluated for the PIPP-R score. The mean age, gestational age, and birth weight of the newborns were  $2.21 \pm 1.55$  days,  $37 \pm 2.44$  weeks, and  $2.56 \pm 0.72$  kg, respectively. The procedures included heel prick for Random Blood Sugar (RBS) (44%), intravenous sampling/insertion (34%), and intramuscular vitamin K injection (22%). The mean difference with 95% Confidence Limits (CL) of total PIPP-R scores between the two consultants (neonatologists) was  $-0.640$  ( $-5.196, 3.916$ ). The length of the CL was  $-9.112$ , which fell outside the defined CL of  $4.2$  (20% of the total score), indicating unacceptable agreement between the two consultants. Similarly, agreement between each consultant and any of the other participants, including residents, nurses, interns, mothers, and social workers, regarding the total PIPP-R score, as well as its subcomponents, was also deemed unacceptable.

**Conclusion:** The present study concluded that the inter-rater reliability of the PIPP-R score and its subcomponents was unacceptable between consultants and with any of the other participants.

**Keywords:** Agreement, Neonates, Painful procedures, Pre-recorded video, Reliability

## INTRODUCTION

Preterm babies are delivered prematurely before their anatomy and physiology are capable of sustaining them in the extrauterine environment. In the Neonatal Intensive Care Unit (NICU), infants are subjected to a hostile environment and various tissue damaging procedures as part of their clinical care [1]. Pain is a continuous or periodic unpleasant feeling that can be dull, acute, or piercing in nature. Newborns experience pain when they are sick or when they undergo diagnostic and therapeutic treatments [2,3]. On average, babies undergo 14 painful procedures in the first two weeks of life [4]. These procedures elicit varying degrees of pain, which can have short-term, as well as long-term consequences [5-8]. Evaluating pain in newborns and young children is more complex and challenging than in adults. There are several techniques and approaches for measuring pain in newborns like Neonatal Infant Pain Scale (NIPS), Neonatal Facial Coding System (NFCS), Neonatal Pain Agitation and Sedation Scale (N-PASS), and PIPP-R, among others [2,9].

The PIPP-R is a feasible and validated tool for pain assessment in preterm and term neonates. Although the name is "PIPP-R," it is validated for neonates with a gestational age of 26-40 weeks. PIPP-R consists of three behavioural, two physiological, and two contextual indicators. Each indicator is assessed on a 4-point scale. The instrument requires physicians to evaluate the neonate's behavioural

state and monitor physiological changes in Heart Rate (HR), oxygen saturation, and facial expression as potential markers of pain [10-12].

The primary objective of all caregivers should be to evaluate the newborn baby's pain and take steps to minimise it in order to avoid these adverse effects [5]. The mother (caretaker) and healthcare providers should be able to interpret the information expressed by the neonate to assess pain. Pain evaluation should be part of a holistic approach to the child's care, and clinicians and other healthcare professionals must regularly measure pain in real-time. It was hypothesised that PIPP-R scoring has a low learning curve, allowing anyone (volunteer/participant) to master this skill with short training. However, the reliability of PIPP-R between investigators has not yet been examined. Therefore, the current study was carried out with the aim of studying the inter-rater reliability of the total PIPP-R score and its components among healthcare providers/laypersons.

## MATERIALS AND METHODS

A cross-sectional study was conducted at Shree Krishna Hospital, a rural Tertiary Care Teaching Hospital in central Gujarat, India. The duration of the study was one year and six months, from January 2021 to June 2022. The study was approved by the Institutional Ethics Committee (IEC), registered with the Central Drugs Standard Control Organisation (CDSCO), on 19 November 2020 (IEC/HMPCMCCE/2020/

Ex.34/279/20). Informed written consent was taken from each volunteer for the study.

**Inclusion criteria:** Videos of 100 physiologically stable newborns were included for assessment. The study included two volunteers (participants) from different fields, namely consultant neonatologists, social workers, MBBS interns, 1<sup>st</sup> year postgraduate students from paediatrics, nurses, and mothers of newborns, to score the PIPP-R from the videos. Thus, a total of 12 volunteers were selected for the assessment of PIPP-R scoring.

**Exclusion criteria:** Neonates requiring any respiratory support, sedatives, or analgesics, having hypoxic ischaemic encephalopathy, or any congenital anomaly were excluded from the study.

**Sample size calculation:** Bland JM and Altman DG, suggested a minimum of 100 records to provide reasonably stable estimates of the 95% Confidence Interval (CI) for agreement studies [13]. So, from the video collections of previous studies, a total of 100 videos of neonates undergoing a pain procedure were selected for the current study.

### Study Procedure

A training session was conducted for the participants by two consultants from the Neonatology Department, who were co-investigators of the study. The participants were educated about the importance of identifying neonatal pain, the components of the PIPP-R scoring system, and the calculation of the total PIPP-R score using handouts for one hour. The participants then independently assessed 10 videos, and these videos were individually discussed with each participant by the consultant neonatologist to ensure accurate scoring. Any questions or difficulties raised by the participants were addressed. By the end of assessing the 10 videos, all participants were found competent in scoring the videos for PIPP-R. This process took approximately two hours for each participant.

Within 15 days after the completion of the training sessions, five tablets were arranged, each containing 100 pre-recorded videos of newborns who had undergone painful procedures at the Institute. Each participant group (social workers, interns, residents, nurses, and mothers) received one tablet each, except for the consultants. The scoring process of the 100 videos took each participant about three weeks before the tablet was transferred to the next participant in the group. This scoring process was completed in approximately two months. After the tablets were returned by the other participant groups, the consultants were provided with one tablet each. Due to their busy schedules, the consultants took about three months to perform the scoring.

The PIPP-R scoring system includes indicators viz., changes in heart rate, decreases in oxygen saturation, brow bulge, eye squeeze, nasolabial furrow, gestational age, and baseline behavioural state. Each indicator was scored on a scale of 0-3. Therefore, the total PIPP-R score ranges from 0-21 [11]. The entire procedure of measuring the PIPP-R score was divided into four steps.

- Step 1: Observing an infant at first for 15 seconds just before the procedure to record the highest HR, lowest oxygen saturation, and behavioural state.
- Step 2: Observing an infant for 30 seconds immediately after the procedure to record changes in the form of the highest HR, lowest oxygen saturation, and duration of each facial action.
- Step 3: Scoring for contextual items based on the changes.
- Step 4: Calculating the total score by adding up the scores of all the items.

The total score represents the intensity of pain, with a higher score indicating a higher degree of pain. The agreement among assessors was evaluated for the total score. However, for clinical decision-making, the total PIPP-R score is categorised as follows: scores of 6 or less generally indicate minimal or no pain, scores between 6-12 are considered mild pain, and scores greater than 12 reflect moderate to severe pain [10]. It is important to note that these categories were not used to classify the intensity of the pain.

### STATISTICAL ANALYSIS

Bland-Altman analysis was used to assess agreement among different volunteers [13]. It was decided that the mean difference in the total score should be between -1 and +1, and the length of the CL should be within 20% to 25% of the total PIPP-R score. The maximum total PIPP-R score for preterm newborns is 21, and for term newborns, it is 18. Thus, a range of CI below 4.2 (i.e., 20% for the preterm newborn's PIPP-R score and 23.33% for the term newborn's PIPP-R score) was considered as acceptable agreement. For facial expressions, the total score is 9, so 20% of that (i.e., 1.8) was considered an acceptable range for the CL. Cohen's Kappa statistic measures the inter-rater reliability of categorical data [14]. Cohen's Kappa was used to assess agreement between different components of PIPP-R among consultants, as well as between consultants and other volunteers. It was interpreted as none, minimal, weak, moderate, strong, and almost perfect agreement if the Kappa was in the range of 0 to 0.20, 0.21 to 0.39, 0.40 to 0.59, 0.60 to 0.79, 0.80 to 0.90, and 0.91 to 1.00, respectively. STATA (14.2), Stata Corp LLC, Texas, United States of America (USA) was used to analyse the data.

### RESULTS

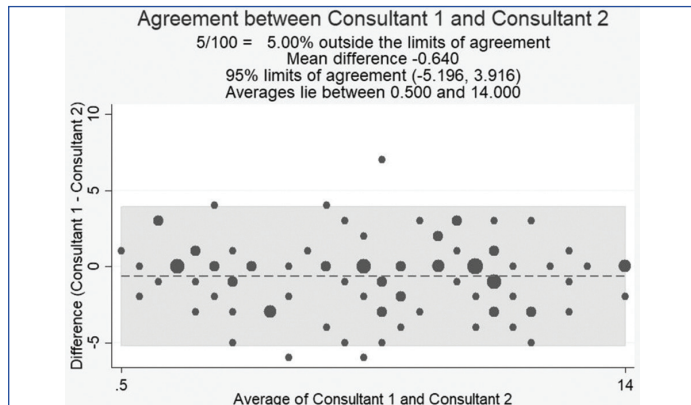
A total of 100 recorded videos of newborns undergoing painful procedures were evaluated for PIPP-R scores. The videos included 51 girls and 49 boys. The mean age, gestational age, and birth weight of the newborns were 2.21±1.55 days, 37±2.44 weeks, and 2.56±0.72 kg, respectively. The procedures involved heel prick for RBS estimation (44%), intravenous sampling/line insertion (34%), and intramuscular vitamin K administration (22%). All the assessors were middle-aged (25-40 years). Both consultants had more than a decade of experience in paediatrics, while both nurses had 5+ years of experience. In all participant groups except for nurses, mothers, and interns, one male and one female assessor were included. The

Individuals	Behaviour state	Fall in SpO <sub>2</sub> score	Increase in HR score	Brow bulge score	Eye squeeze score	Nasolabial furrow score	Gestational age score	Total PIPP score
Mean±SD								
Consultant 1	1.08±1.22	0.53±0.88	0.56±0.67	1.48±1.15	1.48±1.15	1.48±1.15	0.43±0.56	7.04±3.70
Consultant 2	1.23±1.21	0.48±0.85	0.62±0.74	1.64±1.17	1.64±1.17	1.64±1.17	0.43±0.56	7.68±3.81
Resident 1	1.38±1.23	0.40±0.72	0.62±0.80	2.17±0.88	2.17±0.88	2.17±0.88	0.43±0.56	9.34±2.94
Resident 2	1.25±1.19	0.36±0.66	0.59±0.81	2.03±0.98	2.03±0.98	2.03±0.98	0.43±0.56	8.72±3.18
Intern 1	1.22±1.27	0.58±0.88	0.71±0.77	2.08±1.08	2.08±1.08	2.08±1.08	0.43±0.56	9.18±3.73
Intern 2	1.34±1.25	0.56±0.89	0.80±0.77	2.13±0.98	2.13±0.98	2.13±0.98	0.43±0.56	9.52±3.55
Mother 1	1.37±1.22	0.41±0.77	0.65±0.83	2.31±0.66	2.31±0.66	2.31±0.66	0.43±0.56	9.79±2.58
Mother 2	1.18±1.22	0.61±0.90	0.77±0.80	2.22±0.86	2.22±0.86	2.22±0.86	0.43±0.56	9.65±3.26
Social worker 1	1.77±1.13	0.59±0.95	0.73±0.86	1.46±1.31	1.53±1.34	1.44±1.31	0.43±0.56	7.95±4.43

Social worker 2	1.7±1.18	0.57±0.95	0.72±0.87	1.43±1.32	1.45±1.31	1.52±1.34	0.43±0.56	7.82±4.44
Nurse 1	1.23±1.21	0.50±0.81	0.67±0.78	2.06±0.85	2.06±0.85	2.06±0.85	0.43±0.56	9.02±3.05
Nurse 2	1.23±1.21	0.57±0.92	0.63±0.76	1.91±0.95	1.91±0.95	1.91±0.95	0.43±0.56	8.58±3.44

**[Table/Fig-1]:** Comparison of Revised Premature Infant Pain Profile (PIPP-R) scoring amongst assessors.  
 Note: The gestational age score was calculated using a predetermined criteria and not assessed and also, same for all assessors. All the other components of PIPP-R were assessed independently by the assessors and hence there is difference in these scores across assessors; HR: Heart rate

mean±SD oxygen saturation (SpO<sub>2</sub>) in percentage (%) and HR at baseline for the babies were 94.61±3.54 and 147.60±18.81 Beats Per Minute (BPM), respectively. The mean±SD values of all the components of PIPP-R, along with the total score, are provided in [Table/Fig-1]. The mean difference (95% CI) in PIPP-R scores between the consultants was -0.640 (-5.196, 3.916). This means that the mean difference between consultants in PIPP-R scores of the 100 videos is -0.64 units, and about 95% of the differences are within -5.196 and +3.916 units [Table/Fig-2].



**[Table/Fig-2]:** Bland-Altman agreement between two consultants.

The length of the confidence interval is 9.112, which is outside the defined confidence limit of 4.2 (20% of the total score), suggesting unacceptable agreement. The mean difference for agreement between consultants and other assessors ranged from -2.75 to -0.14, indicating that other assessors probably underestimated the pain. Although unacceptable, nurses assessed PIPP-R better than others [Table/Fig-3].

Individuals	Consultant 1		Consultant 2	
	Mean difference	95% limits of agreement (range of CL)	Mean difference	95% limits of agreement (range of CL)
Consultant 1	-----	-----	-0.640	(-5.196, 3.916) (9.112)
Resident 1	-2.300	(-8.919, 4.319) (13.238)	-1.660	(-8.747, 5.427) (14.174)
Resident 2	-1.680	(-8.254, 4.894) (13.148)	-1.040	(-7.926, 5.846) (11.930)
Intern 1	-2.750	(-9.604, 4.104) (13.708)	-2.110	(-9.206, 4.986) (7.096)
Intern 2	-2.610	(-8.181, 2.961) (11.142)	-1.970	(-7.750, 3.810) (11.56)
Mother 1	-2.140	(-7.739, 3.459) (11.198)	-1.50	(-7.465, 4.465) (11.93)
Mother 2	-2.480	(-8.220, 3.260) (11.48)	-1.840	(-8.136, 4.456) (12.592)
Social worker 1	-0.910	(-7.251, 5.431) (12.682)	-0.270	(-7.070, 6.530) (13.600)
Social worker 2	-0.780	(-7.075, 5.515) (-12.59)	-0.140	(-6.998, 6.718) (13.716)
Nurse 1	-1.980	(-7.396, 3.436) (-10.832)	-1.340	(-5.006, 2.326) (-7.332)
Nurse 2	-1.540	(-7.121, 4.041) (-11.162)	-0.900	(-5.129, 3.329) (-8.458)

**[Table/Fig-3]:** Bland-Altman agreement of total PIPP-R between consultants and rest of the assessors.

The agreement of facial expression parameters (combined score of brow bulge, eye squeeze, and nasolabial furrow) of PIPP-R scores between consultants and the rest of the assessors was also found to be unacceptable with a similar trend. Most assessors underestimated the pain compared to consultants, and nurses had better agreement with consultants, though still unacceptable [Table/Fig-4].

Individuals	Consultant 1		Consultant 2	
	Mean difference	95% limits of agreement	Mean difference	95% limits of agreement
Consultant 1	-----	-----	-0.480	(-4.718, 3.758)
Resident 1	-1.650	(-7.280, 3.980)	-1.170	(-7.014, 4.674)
Resident 2	-2.070	(-7.842, 3.702)	-1.590	(-7.700, 4.520)
Intern 1	-1.800	(-6.601, 3.001)	-1.320	(-6.287, 3.647)
Intern 2	-1.950	(-6.990, 3.090)	-1.470	(-6.788, 3.848)
Mother 1	-2.490	(-8.462, 3.482)	-2.010	(-8.154, 4.134)
Mother 2	-2.220	(-6.993, 2.553)	-1.740	(-6.697, 3.217)
Social worker 1	0.010	(-5.757, 5.777)	0.490	(-5.330, 6.513)
Social worker 2	0.040	(-5.696, 5.776)	0.520	(-5.522, 6.562)
Nurse 1	-1.740	(-6.626, 3.146)	-1.260	(-4.724, 2.204)
Nurse 2	-1.290	(-5.962, 3.382)	-0.810	(-4.031, 2.411)

**[Table/Fig-4]:** Bland-Altman agreement of facial expression parameters of PIPP-R score between consultants and the other assessors.

Even after categorising behaviour state and changes in SpO<sub>2</sub> and HR according to the PIPP-R scoring instructions, a weighted Kappa (with quadratic weights) showed poor inter-rater reliability, although nurses exhibited acceptable Kappa values with consultants [Table/Fig-5].

Individuals	Consultant 1			Consultant 2		
	BS	SpO <sub>2</sub>	HR	BS	SpO <sub>2</sub>	HR
	Weighted Kappa with quadratic weights					
Consultant 1	----	----	----	0.79	0.83	0.57
Resident 1	0.64	0.41	0.48	0.53	0.37	0.39
Resident 2	0.61	0.23	0.50	0.51	0.20	0.40
Mother 1	0.60	0.34	0.36	0.56	0.31	0.32
Mother 2	0.65	0.62	0.26	0.61	0.47	0.46
Intern 1	0.72	0.65	0.33	0.68	0.50	0.53
Intern 2	0.60	0.67	0.19	0.60	0.56	0.41
Social worker 1	0.44	0.81	0.40	0.45	0.76	0.50
Social worker 2	0.41	0.83	0.41	0.44	0.76	0.50
Nurse 1	0.78	0.81	0.44	0.64	0.93	0.85
Nurse 2	0.79	0.63	0.50	0.84	0.78	0.86

**[Table/Fig-5]:** Inter-rater reliability of behaviour state (BS), change in oxygen saturation (SpO<sub>2</sub>) and change in Heart Rate (HR) categories between consultants and other assessors.

Subtle observations: During the process of assessing heart rate and SpO<sub>2</sub>, it was observed that in some newborns, after the procedure, the heart rate dropped (21%), and SpO<sub>2</sub> increased (17%) according to the consultant's assessment.

## DISCUSSION

The present study was conducted to assess the agreement between consultants and other healthcare workers, as well as laypersons, for



PIPP-R. In the present study, 100 prerecorded videos of newborns undergoing painful procedures were examined by study participants (volunteers). Overall, there was unacceptable agreement between the consultants and the rest of the participants. There are many one-dimensional and multidimensional pain evaluation measures for newborns [15]. The PIPP-R, an upgraded version of the original PIPP, is a multidimensional pain assessment instrument. Since the item statements on the scale were altered to make them more comprehensible, pain evaluation in disadvantaged groups is considered more objective due to the improved scoring system and the broad range of gestational ages for which it may be used for pain assessment [16]. Results and subtle observations from the current study indicate that recording PIPP-R is not easy, and it is not a straightforward process for everyone, as even the scores did not agree between the experienced consultants.

There have been a few attempts to assess the concordance between two assessors in the PIPP-R scoring. The reliability of PIPP-R scores, in terms of Intraclass Correlation (ICC), was found to be good when the scoring was performed by two competent nurses [16]. Similar findings were reported when three specialists assessed the PIPP-R scores [17]. Another validation study also reported very high ICC among three nurses for PIPP-R scores [18]. These studies indicated that within a subspecialty, PIPP-R is a reliable tool in terms of ICC. In contrast, the agreement between two consultants was found to be unacceptable in the current study. This discrepancy might be due to the fact that ICC is mathematically equal to Kappa, which is a measure of agreement for categorical data. There have been a few attempts to check the concordance between different groups (nurses/parents/physicians, etc.) in assessing pain. A study conducted in the pediatric emergency department reported discordance between nurses and parents on the Face Legs Activity Cry Consolability (FLACC) scale in children below four years of age [19]. Another study from the pediatric emergency department reported poor agreement between patients and caregivers in pain assessed through the Wong-Baker FACES (WBF) and Faces Pain Scale-Revised (FPS-R) scales [20]. These findings corroborate with the results of the current study.

Perception of pain and pain scores may vary from person to person based on their previous experiences and their relationship with the patients [19,21]. Zhou H et al., conducted a meta-analysis of 12 studies investigating the association between self-reported pain ratings for dyads consisting of a child and parent, a child and nurse, and a parent and nurse. They concluded that assessments of children's pain by nurses and parents provide rough estimates rather than an accurate reflection of what children are actually experiencing [21]. The authors found that the assessment of changes in heart rate and SpO<sub>2</sub> did not agree between the two individuals, even among the consultants. Although changes in physiological markers are detected in newborns undergoing painful procedures, it is doubtful if they accurately assess pain, as they are a result of sympathetic nervous system activation and may represent general discomfort rather than specific pain. These markers are also reported in response to non painful stimuli, making it challenging to interpret them solely as indicators of pain. Nevertheless, they are recognised as objective markers in composite pain measurements [22].

Participants observed that while assessing the PIPP-R score, the assessor has to simultaneously focus on multiple parameters, including behaviour, identifying maximum heart rate and minimum SpO<sub>2</sub> before and immediately after the procedure, all within a strict time-bound manner. Placing the pulse oximetry probe is necessary to record pulse rate and SpO<sub>2</sub>. Due to the painful procedure, newborns often move their hands and legs, causing changes in the waveform and heart rate on all types of SpO<sub>2</sub> monitors, including Masimo pulse oximeters. This reduces the accuracy of the PIPP-R score. The current study was based on video assessments, allowing the ability to replay the video to carefully evaluate the individual components, which might be very difficult in real-time assessments.

An alternative to PIPP-R could be the use of simpler scales that contain fewer components, have better inter-rater reliability, and are less time-bound [21]. The Neonatal Infant Pain Scale (NIPS) is a multidimensional pain scale designed for use in newborns. It contains indications for facial expressions, crying, breathing patterns, arm and leg movements, level of arousal, as well as one physiological signal [23]. The NIPS can be considered a reliable, valid, and clinically relevant instrument with high practical importance [24]. Oliveira NRG et al., assessed the correlation, internal consistency, and reliability between two experts in physical therapy who have extensive technical experience in neonatology, in assessing pain using NIPS and PIPP-R. They found high internal consistency for NIPS ( $r=0.824$ ) and moderate for PIPP-R (0.655) [25]. Similarly, Bellieni CV et al., assessed the agreement of NIPS and PIPP between three nurses and found that NIPS had better interobserver reliability than PIPP [26].

Most neonatal pain assessment scales assess babies' facial expressions, although some also include elements like crying, limb movement, and vital indicators. Real-time pain assessment requires dynamic nursing monitoring rather than an instantaneous operation. As a result, frequent pain assessment is time consuming and labour-intensive. The results can be influenced by various factors, including subjective differences in observers, interruptions from other clinical procedures, a lack of time, gender differences, neonatal activity interference, etc., [27,28]. Therefore, another alternative could be to utilise Artificial Intelligence (AI) as a neonatal pain-expression-recognition technology. The automated detection of newborn pain expressions has progressed from static photos to dynamic films and from theoretical research to system implementation, making AI-based Neonatal Pain Assessment (AI-NPA) possible. On one hand, AI-NPA may compensate for the inadequacies of onsite NPA performed by medical staff, and it may offer the benefits of simplicity and efficiency. To create a model with strong anti-interference capabilities and great resilience for real-world data, AI-NPA requires a huge amount of precisely classified neonatal pain data. Cheng XC et al., developed an automated NPA system for NIPS and found highly consistent readings with onsite measurement [27].

The authors considered the total PIPP-R score, as well as its subcomponents to assess agreement. The sample size is reasonably good for the present study. Videos were assessed so that the evaluators had ample time to do the scoring. Different care professionals were involved in the present study.

### Limitation(s)

The present study was a single-centre study. Purposive sampling of assessors was done to select the volunteers. Having only two participants from each profession may not be representative. The videos were evaluated for agreement rather than real-time bedside assessment.

### CONCLUSION(S)

The agreement between consultants and other healthcare workers, as well as lay persons, was deemed unacceptable in both the total PIPP-R score and facial expression score. Even after considering the subcomponents of the PIPP-R score, such as behaviour state, change in oxygen saturation, and change in HR, the Kappa value was not impressive, confirming poor agreement. Contrary to expectations, the learning curve for PIPP-R scoring appears to be steep, suggesting that persistent efforts and experience are required to master this skill, hence rejecting the hypothesis. This fact is supported by the better agreement observed between experienced nurses and consultants. Conducting multicentric agreement studies utilising different frontline healthcare workers will help strengthen the evidence.

## REFERENCES

- [1] Ionio C, Colombo C, Brazzoduro V, Mascheroni E, Confalonieri E, Castoldi F, et al. Mothers and fathers in NICU: The impact of preterm birth on parental distress. *Eur J Psychol.* 2016;12(4):604-21.
- [2] Williams MD, Lascelles BDX. Early neonatal pain-a review of clinical and experimental implications on painful conditions later in life. *Front Pediatr.* 2020;8:30.
- [3] Page GG. Are there long-term consequences of pain in newborn or very young infants? *J Perinat Educ.* 2004;13(3):10-17.
- [4] Hall RW, Anand KJ. Pain management in newborns. *Clin Perinatol.* 2014;41(4):895-924.
- [5] Orovec A, Disher T, Caddell K, Campbell-Yeo M. Assessment and management of procedural pain during the entire neonatal intensive care unit hospitalization. *Pain Manag Nurs.* 2019;20(5):503-11.
- [6] Walker SM. Long-term effects of neonatal pain. *Semin Fetal Neonatal Med.* 2019;24(4):101005.
- [7] Maxwell LG, Fraga MV, Malavolta CP. Assessment of pain in the newborn: An update. *Clin Perinatol.* 2019;46(4):693-707.
- [8] Anand KJ, Scalzo FM. Can adverse neonatal experiences alter brain development and subsequent behaviour? *Biol Neonate.* 2000;77(2):69-82.
- [9] Campbell-Yeo M, Eriksson M, Benoit B. Assessment and management of pain in preterm infants: A practice update. *Children (Basel).* 2022;9(2):244.
- [10] Stevens B, Johnston C, Petryshen P, Taddio A. Premature infant pain profile: Development and initial validation. *Clin J Pain.* 1996;12(1):13-22.
- [11] Stevens BJ, Gibbins S, Yamada J, Dionne K, Lee G, Johnston C, et al. The premature infant pain profile-revised (PIPP-R): Initial validation and feasibility. *Clin J Pain.* 2014;30(3):238-43.
- [12] Ballantyne M, Stevens B, McAllister M, Dionne K, Jack A. Validation of the premature infant pain profile in the clinical setting. *Clin J Pain.* 1999;15(4):297-303.
- [13] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-10.
- [14] McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med (Zagreb).* 2012;22(3):276-82.
- [15] De Melo GM, Lélis AL, de Moura AF, Cardoso MV, da Silva VM. Escalas de avaliação de dor em recém-nascidos: Revisão integrativa [Pain assessment scales in newborns: Integrative review]. *Rev Paul Pediatr.* 2014;32(4):395-402.
- [16] Sadeghi A, Rassouli M, Gharehdaghi FA, Nourian M, Nasiri M, Farhadi R, et al. Validation of the Persian version of premature infant pain profile-revised in hospitalized infants at the neonatal intensive care units. *Iran J Pediatr.* 2017;27(5):e10056.
- [17] Taplak AŞ, Bayat M. Psychometric testing of the Turkish version of the premature infant pain profile revised-PIPP-R. *J Pediatr Nurs.* 2019;48:e49-e55.
- [18] Fitri SYR, Lusmilasari L, Juffrie M. The Indonesian version of the premature infant pain profile-revised: Translation and adaptation of a neonatal pain assessment. *Int J Nurs Sci.* 2019;6(4):439-44.
- [19] Rybojad B, Sieniawski D, Rybojad P, Samardakiewicz M, Aftyka A. Pain evaluation in the paediatric emergency department: Differences in ratings by patients, parents and nurses. *Int J Environ Res Public Health.* 2022;19(4):2489.
- [20] Lawson SL, Hogg MM, Moore CG, Anderson WE, Osipoff PS, Runyon MS, et al. Pediatric pain assessment in the emergency department: Patient and caregiver agreement using the Wong-Baker FACES and the Faces Pain Scale-Revised. *Pediatric Emergency Care.* 2021;37(12):e950-54.
- [21] Zhou H, Roberts P, Horgan L. Association between self-report pain ratings of child and parent, child and nurse and parent and nurse dyads: Meta-analysis. *J Adv Nurs.* 2008;63(4):334-42.
- [22] Bellieni CV. Pain assessment in human fetus and infants. *AAPS J.* 2012;14(3):456-61.
- [23] 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(6):59-66.
- [24] Malavizhi G, Vatsa M, Roseline M, Nithin S, Paul S. Inter-rater reliability of neonatal infant pain scale (NIPS) as a multidimensional behavioural pain tool. *Journal of Health and Allied Sciences NU.* 2012;2(01):26-30.
- [25] Oliveira NRG, Formiga CMR, Ramos BA, Santos RND, Moreira NNDS, Marçal PGE, et al. Correlation between the ratio Pipp-R and nips and inter-rater reliability to evaluate pain during the procedure suction in newborn premature. *Research Square.* 2021.
- [26] Bellieni CV, Cordelli DM, Callani C, Palazzi C, Franci N, Perrone S, et al. Inter-observer reliability of two pain scales for newborns. *Early Hum Dev.* 2007;83(8):549-52.
- [27] Cheng X, Zhu H, Mei L, Luo F, Chen X, Zhao Y, et al. Artificial intelligence based pain assessment technology in clinical application of real-world neonatal blood sampling. *Diagnostics (Basel).* 2022;12(8):1831.
- [28] Salekin MS, Mouton PR, Zamzmi G, Patel R, Goldgof D, Kneusel M, et al. Future roles of artificial intelligence in early pain management of newborns. *Paediatr Neonatal Pain.* 2021;3(3):134-45.

### PARTICULARS OF CONTRIBUTORS:

1. Professor, Department of Neonatology, Bhaikaka University, Karamsad, Anand, Gujarat, India.
2. Junior Resident, Department of Paediatrics, Bhaikaka University, Karamsad, Anand, Gujarat, India.
3. Assistant Professor, Department of Paediatrics, Bhaikaka University, Karamsad, Anand, Gujarat, India.
4. Professor, Department of Central Research Services, Bhaikaka University, Karamsad, Anand, Gujarat, India.
5. Assistant Professor, Department of Central Research Services, Bhaikaka University, Karamsad, Anand, Gujarat, India.
6. Assistant Professor, Department of Neonatology, Bhaikaka University, Karamsad, Anand, Gujarat, India.
7. Professor, Department of Neonatology, Bhaikaka University, Karamsad, Anand, Gujarat, India.

### NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:

Mr. Ajay Gajanan Phatak,  
Professor, Department of Central Research Services (Academic Centre),  
Bhaikaka University, Karamsad, Anand-388325, Gujarat, India.  
E-mail: ajaygp@charutarhealth.org

### PLAGIARISM CHECKING METHODS: [Jain H et al.]

- Plagiarism X-checker: Mar 02, 2023
- Manual Googling: May 12, 2023
- iThenticate Software: Sep 11, 2023 (5%)

### ETYMOLOGY: Author Origin

**EMENDATIONS: 9**

### AUTHOR DECLARATION:

- Financial or Other Competing Interests: None
- Was Ethics Committee Approval obtained for this study? Yes
- Was informed consent obtained from the subjects involved in the study? Yes
- For any images presented appropriate consent has been obtained from the subjects. NA

Date of Submission: **Feb 27, 2023**

Date of Peer Review: **Apr 22, 2023**

Date of Acceptance: **Sep 13, 2023**

Date of Publishing: **Nov 01, 2023**