Accepted Manuscript online: 2022-07-20 Article published online: 2022-09-12

Development and Validation of a Neonatal Physical Maturity Score for Low- and Middle-Income Countries

Vaisakh Krishnan, MD, DNB[1] Vijay Kumar, MD, DNB[1] Hemadri Vegda, DNB[1] Vidya Ujjanappa, MD[1] Anju Manari, MSc[2] Ajithkumar V. Thamunni, MD, DNB[3] Ashraf T. P., MD, DNB[3] Sahana Devadas, MD[4] Sudindrashayana Fattepur, MD[5] Paul Basett, MSc[6] Sudhin Thayyil, MD, PhD[1]

[1] Department of Brain Sciences, Centre of Perinatal Neuroscience, Imperial College London, London, United Kingdom
[2] International Perinatal Epidemiology Unit, Bengaluru, Karnataka, India
[3] Department of Paediatrics Institute of Maternal and Child Health, Government Medical College, Kozhikode, Kerala, India
[4] Department of Paediatrics Bangalore Medical College, Bengaluru, Karnataka, India
[5] Department of Paediatrics Karnataka Institute of Medical Sciences, Hubballi, Karnataka, India
[6] Stasconsultancy, London, United Kingdom

Address for correspondence Vaisakh Krishnan, MD, DNB (Paediatrics), Department of Brain Sciences, Centre of Perinatal Neuroscience, Level 5, Hammersmith house, Du Cane road, Imperial College, London W12 0HS, United Kingdom (e-mail: vp21@ic.ac.uk).

Am J Perinatol 2024; 41(suppl S1):e514–e519.

Abstract

Objective Currently available gestational age scoring systems are complex and inaccurate for wider use in low- and middle-income countries (LMIC), particularly in infants with neonatal encephalopathy. Here, we aimed to develop a scoring system based on physical characteristics for identifying late preterm infants from term infants.

Study Design This was a prospective observational study conducted in 2 phases- the discovery phase and validation phase. In the first phase, we examined the accuracy of 10 objective physical characteristics in a prospective cohort of 1,006 infants recruited from three hospitals in South India. A weighted scoring system and a photo card were then developed based on the six best performing characteristics which were validated in another prospective cohort of 1,004 infants.

Results The final score had a sensitivity of 66.0% (95% confidence intervals [CIs], 58.4–73.8%), specificity of 80.0% (95% CI, 77.2–82.7%), and a negative predictive value of 93.0% (95% CI, 90.5–94.5%).

Conclusion This scoring system may have wider applications in LMIC, particularly in community settings and in infants with neonatal encephalopathy.

Keywords
- newborn
- gestational assessment
- preterm infants
- low- and middle-income countries

Key Points

- This is an easily administered scoring system using physical characters to identify late preterm infants.
- The scoring is not affected by neurological injury and can be used in encephalopathic infants.
- Overall accuracy is better than previous scores encompassing the physical criteria alone.

received
March 2, 2022
accepted
July 6, 2022
accepted manuscript online
July 20, 2022
article published online
September 12, 2022

DOI https://doi.org/10.1055/a-1905-5334.
ISSN 0735-1631.

© 2022. The Author(s).
This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Thieme Medical Publishers, Inc., 333 Seventh Avenue, 18th Floor, New York, NY 10001, USA

Identifying late preterm infants from term infants is important to guide the clinical care. This is often challenging in low- and middle-income countries (LMIC) settings, where last menstrual period (LMP) date is unknown or uncertain, and first-trimester ultrasound dating scan is unavailable. Well-established scoring systems for newborns such as the new Ballard et al's[1] or Dubowitz et al's score[2] predict gestational age with good accuracy but require the use of complex neurologic criteria and trained specialists to administer the scoring. This limits the applicability of these scores in the assessment of newborns with neurological abnormalities affecting the tone and posture such as hypoxic ischemic encephalopathy, metabolic disorders, epileptic encephalopathies, or structural brain malformations and in growth-restricted infants. These complex scores are also less generalizable to the community settings in LMIC where specialist neonatology services might be limited.

Although 18 different gestational age scoring systems have been described previously, the ones that perform the best involve these complex neurologic criteria. A systematic review done by Lee et al[3] on the performance of various gestational assessment scores found that the studies involving physical criteria alone were of poor quality and performed poorly in terms of accuracy. This is a gap in the existing scoring systems and emphasizes the need for a simpler and more generalizable gestational assessment score for the LMIC setting. In this study, we developed and validated a gestational assessment scoring system to identify term infants from late preterm infants based on easily identifiable physical characteristics that has potential for wider use in LMIC.

Materials and Methods

We conducted a nested study within a multicenter prospective observational study involving infants born at or after 35 weeks of gestation between October 2020 and August 2021 at three tertiary neonatal units in South India—Government Medical College, Kozhikode; Karnataka Institute of Medical Sciences, Hubballi; and Bangalore Medical College, Karnataka, India (trial registration number: NCT04054453). Prior to the start of the study, we critically reviewed all the 18 published gestational assessment scoring systems and selected 10 objective physical maturity components (►**Supplementary Table S1**; available in the online version) that can be easily used by nonphysicians.[3]

In the first phase, we examined the accuracy of all the 10 components in 1,006 infants (discovery cohort) and identified the best performing components which were subsequently used to develop a weighted score. In the second phase, we developed a photo chart (►**Fig. 1**) of the components of this weighted and optimized score, trained neonatal and obstetric nursing staff at each study site in using this chart, and then validated the weighted score in another cohort of 1,004 infants (validation cohort).

All the assessments were performed within 72 hours after birth by an examiner (neonatal or obstetric nurse) masked to the gestational age. We included only infants whose accurate gestational age was available, either from the LMP or the first-trimester ultrasound scan. Mothers with irregular cycles and whose LMP is unknown were excluded from the study.

Statistical Analysis

In the discovery phase, associations between each physical characteristic and a late preterm birth were assessed using the chi-square test and the joint associations with multiple logistic regression analysis. A backward selection procedure was used to retain only the statistically significant variables which was used to create a weighted scoring system to predict a preterm birth by considering the size of the regression coefficients from the model (►**Table 1**). This score was then validated in the second phase using receiver operating characteristics (ROC) curves, and the diagnostic predictive ability of the score was calculated with sensitivity, specificity, and positive and negative predictive values.

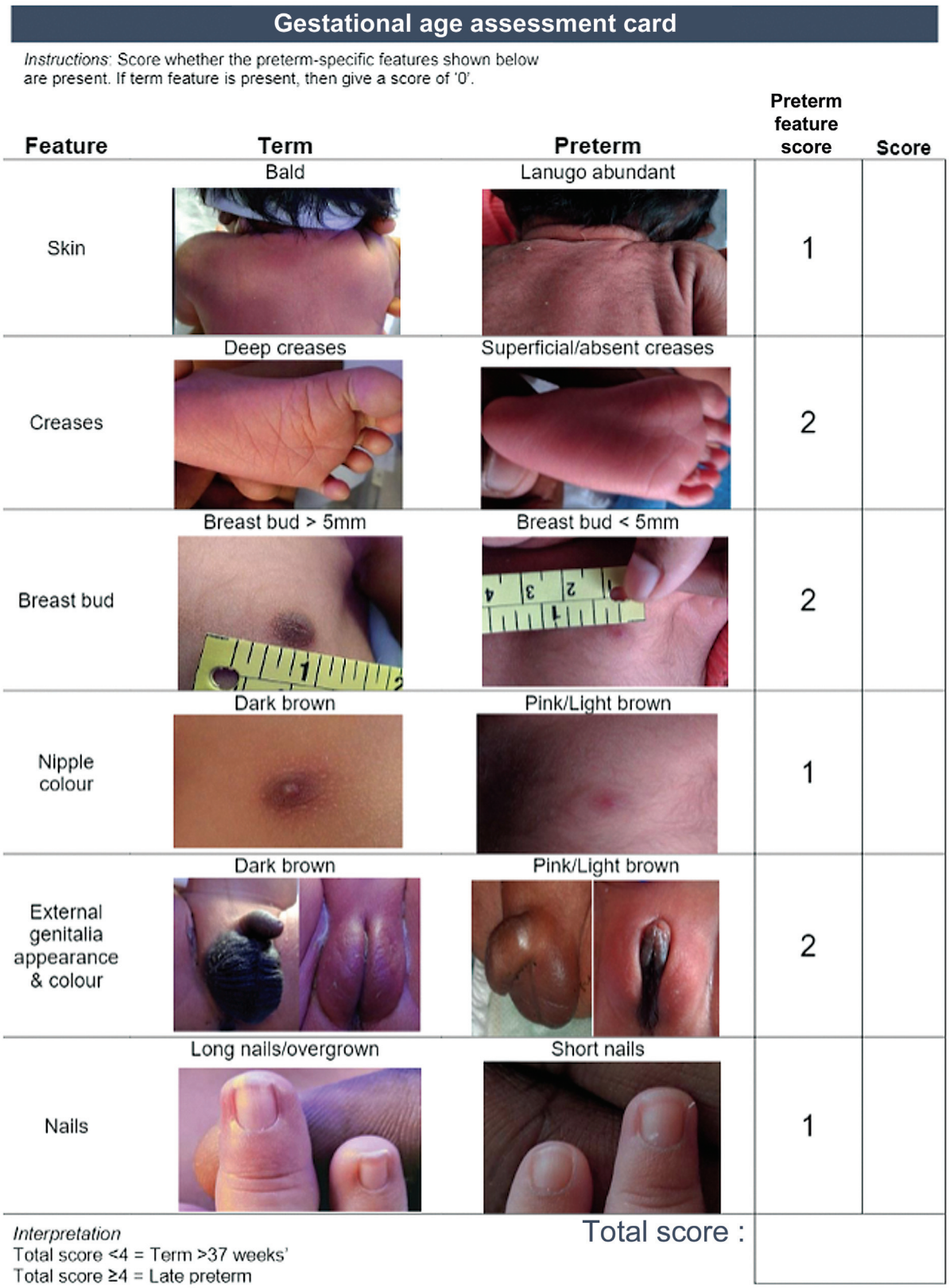
Assuming a specificity of 80% for the new score, precision of 8%, $1-\alpha$ of 95%, with a prevalence of preterm gestation of 15%, 886 infants were required to validate the score. We increased the sample size to 1,000 infants to account for a potentially lower specificity.

Results

Of the 1,006 discovery cohort infants, data on 928 infants were analyzed (►**Fig. 2**). The mean (standard deviation [SD]) gestational age as per LMP or first dating scan was 37.9 (1.5) weeks. Gestational age was estimated by first-trimester ultrasound in 414 (44.6%) and by regular LMP in 514 (55.4%) infants. Although all the 10 components were associated with prematurity on univariate analysis ($p<0.001$), only 6 showed a significant association on multiple logistic regression analysis. These were lanugo hairs ($p=0.001$), plantar crease ($p \leq 0.001$), breast bud ($p \leq 0.001$), nipple color ($p=0.04$), genitalia color ($p \leq 0.001$), and nail size ($p=0.03$). The area under ROC curve was 0.83 (95% confidence interval [CI], 0.80–0.87), and a value of ≥ 4 for the score had a sensitivity of 82.1% and specificity of 74.6%.

Of the 1,004 validation cohort infants, data on 977 were analyzed (►**Fig. 2**). The clinical characteristics of validation cohort were similar to the discovery cohort (►**Table 2**). The mean (SD) gestational age was 38.0 (1.5) weeks. Gestational age was estimated by first-trimester ultrasound in 572 (58.5%) and by regular LMP in 405 (41.5%) infants. The optimized score including six components had an area under ROC of 0.73 (95% CI, 0.69–0.78). A score of 4 or more had a sensitivity and specificity of 66.7 and 80.0%, respectively, for identifying late preterm infants from term infants. Positive predictive value and negative predictive value were 38.6 and 92.7%, respectively (►**Table 3**). The overall accuracy of the score was 77.8%. In the subgroup of infants with neonatal encephalopathy ($n=28$), accuracy was 75.0% and in small for gestational age infants ($n=254$), the accuracy was 64.9%.

 © 2022. The Author(s).



Gestational age assessment card

Instructions: Score whether the preterm-specific features shown below are present. If term feature is present, then give a score of '0'.

Feature	Term	Preterm	Preterm feature score	Score
Skin	Bald	Lanugo abundant	1	
Creases	Deep creases	Superficial/absent creases	2	
Breast bud	Breast bud > 5mm	Breast bud < 5mm	2	
Nipple colour	Dark brown	Pink/Light brown	1	
External genitalia appearance & colour	Dark brown	Pink/Light brown	2	
Nails	Long nails/overgrown	Short nails	1	
		Total score :		

Interpretation
Total score <4 = Term >37 weeks'
Total score ≥4 = Late preterm

Fig. 1 Photo card for assessment of the physical maturity scoring.

Table 1 Final multivariable regression model results

Preterm characteristic	Multivariable analysis		Score	
	OR (95% CI)	*p*-Value	Coefficient	Points
Abundant lanugo	2.01 (1.35–3.00)	0.001	0.70	1
Superficial or absent plantar crease	2.41 (1.60–6.63)	<0.001	0.88	2
Breast bud less than 5 mm	2.74 (1.74–4.30)	<0.001	1.01	2
Nipple color pinkish/light brown	1.61 (1.01–2.56)	0.04	0.48	1
Genitalia color pinkish/light brown	2.53 (1.65–3.89)	<0.001	0.93	2
Short nails	1.65 (1.06–2.55)	0.03	0.50	1

Abbreviations: CI, confidence interval; OR, odds ratio.

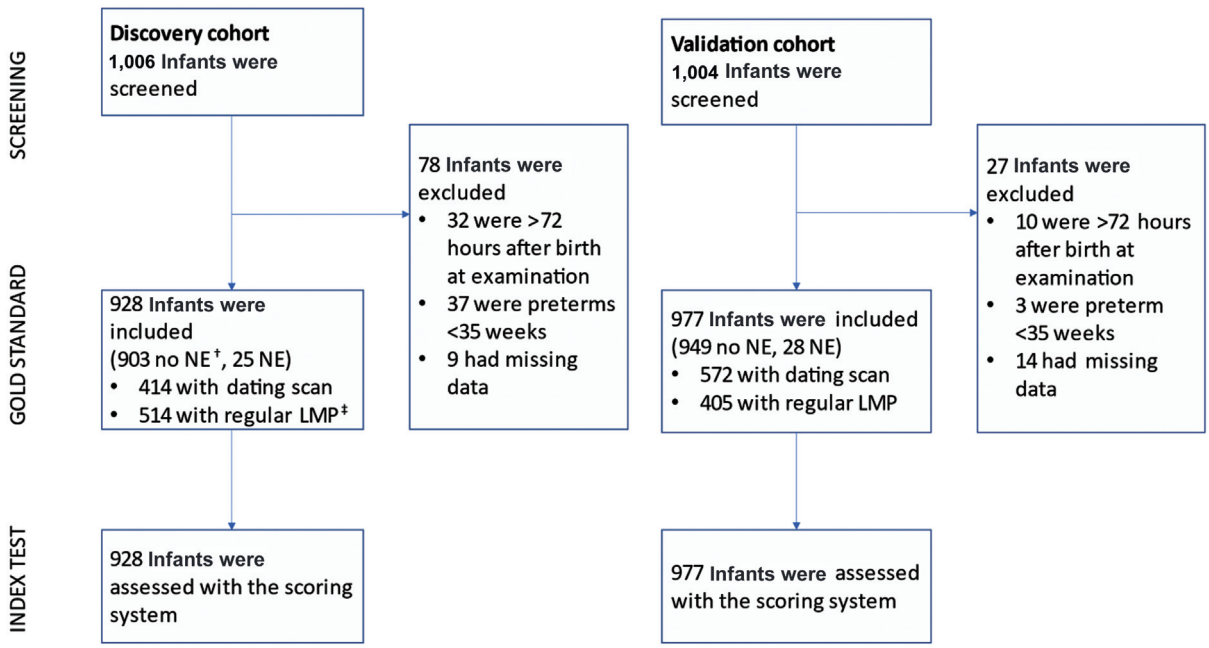


Fig. 2 Flow diagram of recruitment of infants to the study. †NE, neonatal encephalopathy; ‡LMP, last menstrual period.

Discussion

In this large prospective study involving a total of 2,010 infants from South India, we found that an objective and weighted physical maturity gestation score accurately identified late preterm infants from full term infants. Our scoring system can be used with minimal training using a small photo card and does not require any complex neurological assessments. Hence, this scoring system has wider applicability in LMIC especially in community settings, and in infants with neonatal encephalopathy where systems based on neurological criteria cannot be used. Each of the 10 physical characteristics in the discovery phase was significantly associated with the prediction of preterm gestation. However, we found a combination of six physical features—lanugo, plantar creases, breast bud, nipple color, genitalia color, and nail size, as the best to differentiate a late preterm infant from a term infant.

Although gestational assessment scores based on complex and detailed neurological assessments such as the Dubowitz et al's[2] and new Ballard et al's[1] are more accurate than our scoring system, these scores are of little use in community settings in LMICs and for assessments of infants with neonatal encephalopathy. A large community-based cohort validation study from Bangladesh reported that the published scoring systems had poor accuracy in community settings with sensitivities and specificities as low as 5%.[4] In contrast, our scoring system with only six components is simple and feasible to perform with minimal training and have much higher accuracy (78%).

Strengths and Limitations

The key strengths of the study are the large number of infants included and the rigorous methodology followed. Our assessors included nurses from both obstetric and neonatal disciplines, highlighting the generalizability of the scoring administration. The ease of assessment was facilitated with the help of a pocket-sized photo card describing the physical characteristics to be identified. As

Table 2 Clinical characteristics of the discovery and validation cohorts

	Discovery phase ($n = 928$)	Validation phase ($n = 977$)
Gestational age, *n* (%)		
35 wk	72 (7.7)	76 (7.8)
36 wk	10 (11.5)	79 (8.1)
37 wk	167 (17.9)	163 (16.7)
38 wk	227 (24.5)	268 (27.4)
39 wk	224 (24)	242 (24.8)
40 wk	116 (12.5)	132 (13.5)
41 wk	11 (1.2)	13 (1.3)
42 wk	4 (0.4)	4 (0.4)
Mean birth weight, g (SD)	2,857 (565)[a]	2,805 (571)[b]
Small for gestational age, *n* (%)	166 (23.8)	254 (27.1)
Neonatal encephalopathy, *n* (%)	25 (2.7)	28 (2.9)

Abbreviation: SD, standard deviation.
[a]Data available only in 696 infants.
[b]Data available only in 937 infants.

Table 3 Diagnostic performance of score for the prediction of a preterm birth in discovery and validation phases

Statistic	Discovery estimate (95% CI)	Validation estimate (95% CI)
Sensitivity	82.1% (75.7–87.4%)	66.7% (58.4–73.8%)
Specificity	74.6% (71.4–77.7%)	80.0% (77.2–82.7%)
Positive predictive value	43.6% (38.3–49.1%)	38.6% (32.7–44.7%)
Negative predictive value	94.6% (92.4–96.3%)	92.7% (90.5–94.5%)
Area under the curve	0.83 (0.80–0.87)	0.73 (0.69–0.78)

Abbreviations: CI, confidence interval,

there are no neurological components, this score may be suitable in infants with neonatal encephalopathy. Even though the numbers of infants with neonatal encephalopathy were small, the accuracy of scoring system in this subgroup (75.0%) was similar to the overall accuracy (77.8%).

Our study has some limitations. First, we tested the score only in late preterm and term infants and not across the entire range of prematurity. This was intentional, as we wanted to develop a scoring system for wider use outside neonatal intensive care settings, where extremely premature infants would be cared for by specialist physicians. Second, although we recognize that fetal ultrasound is the gold standard for validation, we had to rely on regular, well-recalled LMP in 55.4% infants in discovery cohort and 41.5% infants in the validation cohort. This limitation was due to unavailability of first-trimester dating scan in those mothers. We did take care to include cases where LMP was accurately known, and mother reported regular menstrual cycles. Finally, we did not examine the interrater reliability test for our scoring system in this study.

Conclusion

Our new gestational assessment scoring system consisting of six physical characteristics can identify late preterm infants from term infants with reasonable accuracy and has wider applicability in community settings and for assessment of infants with neonatal encephalopathy.

Authors' Contributions

V.K. conceived the idea, designed the study, developed the study protocol, trained the research nurses in collecting the data, analyzed and interpreted the data, and wrote the first draft of the manuscript. V.K., H.V., V.U., and A.M. recruited and examined the infants, and A.V.T., A.T.P., S.D., and S.F. supervised the study at each of the trial sites. P.B. analyzed the data. S.T. supervised the entire project, including protocol development, interpretation of the data and preparation of the manuscript, and is the guarantor for the study. All authors have participated in development of the manuscript and have approved the final version for submission. V.K. and S.T. have full access to the data.

Funding

The study was funded by a program grant from the National Institute of Health Research (NIHR), United Kingdom, and was supported by the NIHR Biomedical Research Centre based at Imperial College London. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health.

Conflict of Interest

None declared.

Acknowledgments

We acknowledge the contribution of research nurses—Anju Manari, Anitha S. Hiremath, Bharathi K., Bhavya B., Carol D'Souza, Christine Rajiv Geete, Meenu Benny, Deepthi T.V., Elisha Ashalatha, Indiramma, Lakshmi Balachandran, Megha A., Neethu C.M., Ramya C., Reenal D'Souza, Rekhavathi, Shadiya K., Sharon Priyanka, Suja Mullassery, Suma S.V., Sreerenjini N., Vimal A. Madhale, and Manu Mohan (International Perinatal Epidemiology Unit, Bengaluru, Karnataka, India) for their active participation in data collection, and Phoebe Ivain (Imperial College London) in preparing the photo cards. We thank Munirah Mazlan (Imperial College, London) for assisting with the manuscript editing.

References

1 Ballard JL, Khoury JC, Wedig K, Wang L, Eilers-Walsman BL, Lipp R. New Ballard score, expanded to include extremely premature infants. J Pediatr 1991;119(03):417–423
2 Dubowitz LM, Dubowitz V, Goldberg C. Clinical assessment of gestational age in the newborn infant. J Pediatr 1970;77(01):1–10
3 Lee AC, Panchal P, Folger L, et al. Diagnostic accuracy of neonatal assessment for gestational age determination: a systematic review. Pediatrics 2017;140(06):e20171423
4 Lee AC, Mullany LC, Ladhani K, et al; Projahnmo Study Group. Validity of newborn clinical assessment to determine gestational age in Bangladesh. Pediatrics 2016;138(01):e20153303

 © 2022. The Author(s).