



Original Article

Validation Of the New Ballard Score in Term Neonates in South India– A Prospective Validation Study

Dr. Sujin Mariam Thomas¹, Dr. Prince Philip², Dr. Peter Prasanth Kumar Kommu³

¹Assistant Professor, Department of Pediatrics, Believers Church Medical College Hospital, Thiruvalla, Kerala, India

²Associate Professor, Department of Radiology, Pushpagiri Institute of Medical Sciences and Research Centre, Thiruvalla, Kerala, India

³Professor, Department of Pediatrics, Pondicherry Institute of Medical Sciences, Puducherry, Tamil Nadu, India

 OPEN ACCESS

Corresponding Author:

Dr. Sujin Mariam Thomas

Assistant Professor
Department of Pediatrics,
Believers Church Medical College
Hospital
Thiruvalla, Kerala, India

Email id: sujinprince1@gmail.com

Received: 15-04-2026

Accepted: 05-05-2026

Available online: 22-05-2026

Copyright© International Journal of
Medical and Pharmaceutical Research

ABSTRACT

Introduction Accurate gestational age (GA) assessment is vital for neonatal care, guiding clinical management and improving outcomes. Early prenatal ultrasound is the gold standard for GA determination but is often unavailable in resource-limited settings. Clinical tools like the New Ballard Score (NBS), which combines physical and neuromuscular maturity signs, offer a practical alternative. However, validation of the NBS in specific regional contexts, such as term neonates in South India, remains limited. The present study addresses this gap by prospectively evaluating the accuracy of the NBS in this population.

Objectives The present study aims to validate the New Ballard Score for postnatal gestational age assessment in term neonates born in a tertiary care hospital in South India, comparing NBS estimates with first trimester ultrasound dating.

Methodology A prospective validation study was conducted at a tertiary care hospital in South India over 18 months, enrolling 500 term neonates (37–42 weeks gestation) whose mothers had first trimester ultrasound-confirmed GA. Neonates were assessed between 4 and 96 hours after birth using the NBS by a single trained investigator blinded to ultrasound results. Physical and neuromuscular maturity parameters were scored following the standardized NBS protocol. Gestational age estimates by NBS were compared to ultrasound dating using Pearson correlation, paired t-tests, and Bland–Altman analysis to evaluate agreement and bias.

Results The present study demonstrated a significant positive correlation between total NBS scores and gestational age by ultrasound ($r = 0.34$, $p < 0.001$), with physical maturity components showing stronger associations than neuromuscular signs. Agreement analysis revealed no significant systematic bias between NBS and ultrasound estimates (mean difference -0.04 weeks), with clinically acceptable limits of agreement (-2.56 to $+2.49$ weeks). However, NBS accuracy was limited in the small Large for Gestational Age (LGA) subgroup, likely due to sample size constraints.

Conclusion The present study validates the New Ballard Score as a practical and moderately accurate tool for postnatal gestational age assessment in term neonates within a South Indian tertiary care setting. The findings support the utility of NBS as a reliable alternative where early ultrasound is unavailable, facilitating improved clinical decision-making in similar resource-limited contexts. Regional validation is essential for expanding the NBS's application across diverse neonatal populations.

Keywords: New Ballard Score, Gestational Age Assessment, Term Neonates, South India, Validation Study

INTRODUCTION

Accurate assessment of gestational age (GA) is a cornerstone of neonatal care, critically influencing clinical management decisions and outcomes for newborns. Precise determination of GA enables early identification of preterm infants—who face significantly higher risks of mortality and morbidity—and guides timely interventions to improve survival and long-

term developmental outcomes. Globally, preterm birth accounts for a substantial proportion of neonatal deaths, with an estimated 15 million preterm infants born annually, predominantly in low- and middle-income countries (LMICs) where the burden is greatest. In such settings, reliable GA assessment is essential not only at the individual level for clinical care but also at the population level to inform public health strategies and resource allocation.

However, conventional methods for GA estimation face significant limitations in resource-poor contexts. Early prenatal ultrasound, the gold standard for accurate dating, is frequently unavailable or inaccessible due to infrastructural constraints and late antenatal clinic attendance, reducing its practical utility. Alternative assessments such as last menstrual period (LMP) recall suffer from reliability issues, and clinical or anthropometric scoring systems often lack precision, particularly in neonates who are small for gestational age or growth-restricted. Common neonatal assessment tools like the original Ballard Score show considerable variability, tending to overestimate GA in preterm infants and underestimate it in growth-restricted newborns, with relatively wide limits of agreement compared to ultrasound dating [1,2]. Moreover, neonatal anthropometric measures alone exhibit only fair diagnostic accuracy and are confounded by factors such as fetal growth restriction prevalent in these populations [3,4].

In this context, the New Ballard Score (NBS) emerges as a promising tool designed to improve gestational age estimation by combining neuromuscular and physical maturity signs with greater simplicity and potential applicability in low-resource environments. The NBS has been shown to perform reasonably well when administered by trained personnel and may be augmented by emerging approaches, including machine learning algorithms incorporating maternal and neonatal factors, thereby enhancing preterm birth identification accuracy without reliance on ultrasound [5]. Validation of the NBS in specific regional and demographic contexts, such as term neonates in South India, remains crucial to confirm its reliability and utility as a practical alternative for gestational age assessment where conventional methods are limited. Thus, the present study aims to prospectively validate the New Ballard Score in this setting, addressing a vital gap in neonatal care for accurate, accessible gestational age determination.

South India, characterized by its diverse population and unique socio-cultural landscape, represents a significant demographic region with distinct neonatal health challenges. The region encompasses a mix of urban and rural communities with varying access to healthcare facilities, influencing perinatal outcomes and neonatal care. South India accounts for a substantial proportion of India's birth cohort and presents unique epidemiological characteristics, including variations in birthweight and gestational parameters that differ from national averages. Previous large-scale studies have provided updated size at birth charts for South Indian neonates, underscoring the importance of region-specific references for birthweight, length, and head circumference to better reflect the growth patterns in this population [6].

Despite the widespread use of the New Ballard Score (NBS) for assessing gestational age (GA), which combines physical and neuromuscular maturity indicators, its performance and accuracy have not been adequately validated in the South Indian context. Given the diversity in neonatal anthropometry and growth standards in this region, as well as the variable prevalence of small for gestational age (SGA) and low birth weight neonates, there is a compelling need to examine the reliability of NBS in accurately estimating GA among term neonates in South India. Reliance on NBS without regional validation risks inaccurate gestational age assignment, which can impact clinical decisions, especially where advanced methods like early ultrasound dating are limited [3].

There exists a research gap concerning the validation of NBS specifically within the South Indian population, where neonatal anthropometric traits and perinatal risk factors may differ from other regions. Most existing studies on neonatal GA assessment have been conducted elsewhere or use generalized tools without local adaptation or validation. This study aims to fill that gap by prospectively validating the NBS's accuracy and applicability in term neonates born in South India, thereby providing critical data to optimize neonatal assessment practices in this region. Such validation is crucial to ensure appropriate identification and management of neonates based on precise gestational age estimation, ultimately improving neonatal outcomes. The aim of the study was the Assessment of validity of the New Ballard Score (NBS) in term neonates born in a tertiary care hospital in South India.

MATERIALS AND METHODS

a. Study Design and Setting: The present study is a prospective validation study conducted at the Pondicherry Institute of Medical Sciences, a tertiary care hospital in South India. The study duration spanned 18 months, from October 2013 to April 2015. Ethical approval was obtained from the institutional review board and ethics committee prior to commencement. Informed written consent was secured from the parents of all participating neonates.

b. Study Population: The study population comprised term neonates born at the study center during the study period.

Inclusion criteria were:

1. Term births between 37 and 42 weeks gestation
2. Singleton gestation

3. Vertex presentation
4. Mothers who had undergone first trimester ultrasonogram (USG) for gestational age dating

Exclusion criteria included:

1. Sick or unstable neonates
2. Neonates with major congenital malformations
3. Neonates older than 96 hours at the time of assessment

A total sample size of 500 neonates was included after applying inclusion and exclusion criteria and ensuring availability of first trimester USG data. The sample size was determined based on prior studies and feasibility considerations. Consecutive sampling was employed, enrolling all eligible neonates whose parents provided consent during the study period.

c. Data Collection: Data collection involved clinical assessment and record review:

Neonates were examined between 4 hours and 96 hours after birth in a warmed, well-lit room while awake. The New Ballard Score (NBS) was performed by a single trained investigator blinded to gestational age by USG. Physical parameters (e.g., skin texture, lanugo, plantar creases, breast development, ear and genitalia maturity) and neuromuscular criteria (e.g., posture, square window, arm recoil, popliteal angle, scarf sign, heel to ear maneuver) were scored according to the NBS protocol. Scores were immediately recorded to minimize recall bias.

Birth weight, head circumference, length, gender, mode of delivery, and Apgar scores were documented. Gestational age by first trimester USG (gold standard) was obtained from maternal records, based on crown-rump length measured before 13 weeks gestation. Neonates were classified as Small for Gestational Age (SGA), Appropriate for Gestational Age (AGA), or Large for Gestational Age (LGA) using WHO growth charts.

Study Tools

1. New Ballard Score chart for gestational age assessment
2. First trimester ultrasonogram data for gestational age reference
3. Standardized proforma for data recording

Statistical Analysis

Descriptive statistics: Means, standard deviations, ranges, and proportions were calculated for demographic and clinical variables. Pearson correlation coefficients were used to assess the strength and significance of relationships between NBS scores (total and individual criteria) and gestational age by USG. Paired t-tests compared mean gestational ages estimated by NBS and USG to evaluate systematic differences. Bland–Altman analysis was employed to assess agreement between NBS and USG gestational age estimates, including calculation of mean bias and limits of agreement. Statistical significance was set at $p < 0.05$. Data analysis was performed using SPSS version 20 software.

Ethical Considerations

Ethical approval was obtained from the institutional ethics committee prior to study initiation. Written informed consent was obtained from parents or guardians after explaining the study purpose and procedures in their preferred language.

Participation was voluntary, with assurance of confidentiality and the right to withdraw at any time without affecting medical care. The clinical assessment was non-invasive and posed minimal discomfort to neonates. Data were anonymized and securely stored for research purposes only.

RESULTS

The present study prospectively validated the New Ballard Score (NBS) for gestational age (GA) assessment in 500 term neonates born at a tertiary care hospital in South India. The study population included neonates between 37 and 42 weeks gestation, with first trimester ultrasound (USG) dating as the gold standard reference. Neonates were assessed between 4 and 96 hours after birth by a single trained investigator blinded to USG results, using the standardized NBS protocol which combines physical and neuromuscular maturity criteria. (table 1)

Descriptive analysis revealed mean total NBS scores of 37.89 ± 2.72 , with physical maturity and neuromuscular maturity mean scores of 18.89 ± 1.66 and 19.00 ± 1.67 , respectively. Subgroup analysis of small for gestational age (SGA) neonates ($n=104$) showed slightly lower mean total scores (36.82 ± 3.32), reflecting the impact of growth restriction on maturity parameters. (table 2)

Correlation analyses demonstrated statistically significant positive associations between NBS total scores and GA by USG ($r = 0.34$, $p < 0.001$), indicating moderate concordance. Among individual criteria, physical maturity components

such as plantar creases ($r = 0.21$, $p < 0.001$), breast development ($r = 0.23$, $p < 0.001$), and skin texture ($r = 0.15$, $p = 0.001$) showed stronger correlations with GA compared to neuromuscular signs.(table 3). In the appropriate for gestational age (AGA) subgroup ($n=357$), correlations remained significant though slightly attenuated (total score $r = 0.27$, $p < 0.001$). (table 4).Conversely, no statistically significant correlations were observed within the small LGA subgroup ($n=39$), likely due to limited sample size.(table 5)

Paired t-tests comparing mean GA estimates by NBS and USG indicated no significant systematic difference (mean difference -0.04 weeks), supporting the absence of bias in NBS estimations. (table 4)

Bland–Altman analysis further confirmed agreement, with limits of agreement ranging from -2.56 to +2.49 weeks, which is within clinically acceptable margins of ± 2 weeks.

The distribution of GA estimates by both methods showed similar patterns, reinforcing the validity of NBS as a practical tool for postnatal GA assessment in this population.(Fig 1) The findings highlight that the NBS performs reasonably well in term neonates in South India, with acceptable accuracy and consistency compared to first trimester USG dating. This supports its utility as a reliable alternative in settings where early ultrasound is unavailable or inaccessible.

The Bland–Altman plot illustrated that the difference between methods decreased with increasing gestational age, suggesting improved precision in older term neonates.(Fig 2)

Overall, the present study addresses a critical gap by providing region-specific validation of the NBS, confirming its applicability for accurate gestational age estimation among term neonates in South India, thereby facilitating improved clinical decision-making and neonatal care in this context.

Table 1: Mean Score for Individual and Combined Criteria (n=500)

Criteria	Mean	Standard Deviation (SD)	Range
Posture	3.16	0.40	2 – 4
Arm recoil	3.64	0.53	1 – 4
Square window	3.01	0.64	1 – 4
Popliteal angle	3.35	0.66	1 – 4
Scarf sign	3.10	0.40	2 – 4
Heel to ear	2.73	0.62	0 – 4
Skin	2.90	0.48	1 – 5
Lanugo	3.30	0.51	2 – 4
Plantar creases	3.29	0.59	2 – 4
Breast	3.31	0.71	1 – 4
Ears	2.73	0.50	2 – 4
Genitals	3.36	0.52	2 – 4
Totalphysical maturity	18.89	1.66	13 – 23
Total neuromuscular maturity	19.00	1.67	11 – 24
Total score	37.89	2.72	27 – 46

Scores for individual criteria are based on the New Ballard Score assessment protocol. Mean scores ranged from 2.73 to 3.64 across criteria, with posture (3.16; 79%), arm recoil (3.64; 91%), square window (3.01; 75%), popliteal angle (3.35;

84%), scarf sign (3.10; 78%), heel to ear (2.73; 68%), skin texture (2.90; 58%), lanugo (3.30; 83%), plantar creases (3.29; 82%), breast development (3.31; 83%), ears (2.73; 68%), and genitalia (3.36; 84%). Total physical maturity averaged 18.89 (79% of maximum 24), total neuromuscular maturity 19.00 (79% of maximum 24), and total combined score 37.89 (82% of maximum 46). Range indicates minimum and maximum observed scores in the study population (n=500).

Table 2: Mean Score for Individual and Combined Criteria in SGA Babies (n=104)

Criteria	Mean	Standard Deviation (SD)	Range
Posture	3.09	0.42	2 – 4
Arm recoil	3.54	0.54	2 – 4
Square window	2.98	0.68	1 – 4
Popliteal angle	3.29	0.73	1 – 4
Scarf sign	3.11	0.42	2 – 4
Heel to ear	2.61	0.66	0 – 3
Skin	2.90	0.57	1 – 5
Lanugo	3.21	0.53	2 – 4
Plantar creases	3.22	0.64	2 – 4
Breast	3.00	0.79	1 – 4
Ears	2.61	0.53	2 – 4
Genitals	3.24	0.51	2 – 4
Total physical maturity	18.21	1.97	13 – 22
Total neuromuscular maturity	18.61	2.11	12 – 23
Total score	36.82	3.32	27 – 44

Scores for individual criteria are based on the New Ballard Score assessment protocol. Mean scores ranged from 2.73 to 3.64 across criteria, with posture (3.16; 79%), arm recoil (3.64; 91%), square window (3.01; 75%), popliteal angle (3.35; 84%), scarf sign (3.10; 78%), heel to ear (2.73; 68%), skin texture (2.90; 58%), lanugo (3.30; 83%), plantar creases (3.29; 82%), breast development (3.31; 83%), ears (2.73; 68%), and genitalia (3.36; 84%). Total physical maturity averaged 18.89 (79% of maximum 24), total neuromuscular maturity 19.00 (79% of maximum 24), and total combined score 37.89 (82% of maximum 46). Range indicates minimum and maximum observed scores in the study population (n=500).

Table 3: Correlation for Individual and Combined Criteria (n=500)

Criteria	Pearson Correlation (r)	p value
Posture	0.11	0.013
Arm recoil	0.07	0.097
Square window	0.16	<0.001
Popliteal angle	0.03	0.447

Scarf sign	0.096	0.033
Heel to ear	0.096	0.032
Skin	0.15	0.001
Lanugo	0.06	0.151
Plantar creases	0.21	<0.001
Breast	0.23	<0.001
Ears	0.13	0.004
Genitals	0.14	0.001
Total physical maturity	0.33	<0.001
Total neuromuscular maturity	0.18	<0.001
Total score	0.34	<0.001

Pearson correlation coefficients between New Ballard Score (NBS) individual criteria and gestational age by first trimester ultrasound (USG) showed statistically significant positive associations for several parameters. Mean correlations ranged from 0.03 to 0.34, with total physical maturity ($r = 0.33$; 33%) and total combined score ($r = 0.34$; 34%) showing the strongest relationships. Among individual criteria, plantar creases ($r = 0.21$; 21%), breast development ($r = 0.23$; 23%), skin texture ($r = 0.15$; 15%), square window ($r = 0.16$; 16%), posture ($r = 0.11$; 11%), ears ($r = 0.13$; 13%), genitals ($r = 0.14$; 14%), scarf sign ($r = 0.096$; 10%), and heel to ear ($r = 0.096$; 10%) were significantly correlated with gestational age ($p < 0.05$). Arm recoil ($r = 0.07$; 7%) and lanugo ($r = 0.06$; 6%) showed weaker, non-significant correlations. Range of correlations indicates moderate concordance between NBS scores and gestational age by USG in the study population ($n=500$).

Table 4: Correlation for Individual and Combined Criteria for AGA Babies (n=357)

Criteria	Pearson Correlation (r)	p value
Posture	0.078	0.144
Arm recoil	0.117	0.027
Square window	0.131	0.013
Popliteal angle	0.030	0.568
Scarf sign	0.04	0.440
Heel to ear	0.05	0.358
Skin	0.11	0.03
Lanugo	0.054	0.306
Plantar creases	0.13	0.01
Breast	0.163	0.002
Ears	0.12	0.02
Genitals	0.17	0.001
Total physical maturity	0.29	<0.001

Total neuromuscular maturity	0.16	0.003
Total score	0.27	<0.001

Pearson correlation coefficients between New Ballard Score (NBS) individual criteria and gestational age by first trimester ultrasound (USG) in the Appropriate for Gestational Age (AGA) subgroup (n=357) showed statistically significant positive associations for several parameters. Mean correlations ranged from 0.04 to 0.29: arm recoil (r = 0.12; 12%), square window (r = 0.13; 13%), skin texture (r = 0.11; 11%), plantar creases (r = 0.13; 13%), breast development (r = 0.16; 16%), ears (r = 0.12; 12%), and genitals (r = 0.17; 17%) were significantly correlated with gestational age (p < 0.05). Total physical maturity (r = 0.29; 29%), total neuromuscular maturity (r = 0.16; 16%), and total combined score (r = 0.27; 27%) also showed significant correlations. Other criteria such as posture, popliteal angle, scarf sign, heel to ear, and lanugo showed weaker or non-significant correlations. Range indicates correlation strength within the AGA subgroup.

Table 5: Correlation for Individual and Combined Criteria for LGA Babies (n=39)

Criteria	Pearson Correlation (r)	p value
Posture	0.155	0.347
Arm recoil	0.074	0.097
Square window	0.185	0.260
Popliteal angle	-0.025	0.878
Scarf sign	0.151	0.360
Heel to ear	-0.038	0.817
Skin	0.120	0.481
Lanugo	-0.040	0.829
Plantar creases	0.060	0.727
Breast	-0.186	0.258
Ears	0.032	0.848
Genitals	-0.064	0.698
Total physical maturity	-0.075	0.648
Total neuromuscular maturity	0.008	0.960
Total score	-0.035	0.833

Pearson correlation coefficients between New Ballard Score (NBS) individual criteria and gestational age by first trimester ultrasound (USG) in the Large for Gestational Age (LGA) subgroup (n=39) showed weak and non-significant associations. Correlations ranged from -0.186 to 0.185: posture (r = 0.16; 16%), arm recoil (r = 0.07; 7%), square window (r = 0.19; 19%), popliteal angle (r = -0.03; -3%), scarf sign (r = 0.15; 15%), heel to ear (r = -0.04; -4%), skin (r = 0.12; 12%), lanugo (r = -0.04; -4%), plantar creases (r = 0.06; 6%), breast (r = -0.19; -19%), ears (r = 0.03; 3%), and genitals (r = -0.06; -6%). Total physical maturity (r = -0.08; -8%), total neuromuscular maturity (r = 0.01; 1%), and total combined score (r = -0.04; -4%) also showed no significant correlations (p > 0.05). Range indicates correlation strength within the LGA subgroup.

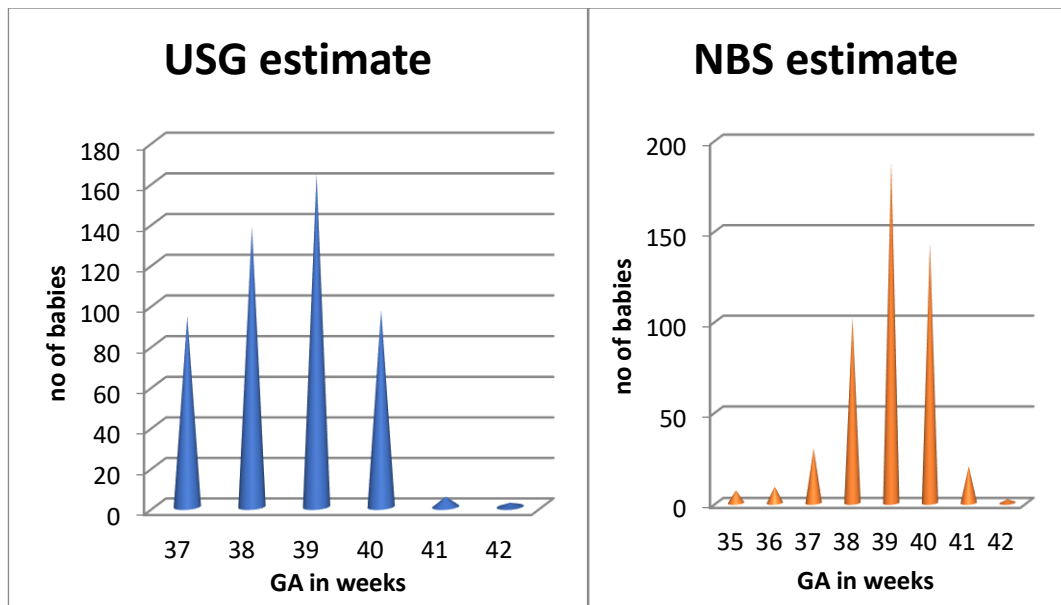


Figure 1: Distribution of estimates of GA

Gestational age by USG is the gold standard, based on first trimester crown-rump length measurement before 13 weeks gestation.

NBS refers to gestational age estimated using the New Ballard Score, which combines physical and neuromuscular maturity criteria.

The figure demonstrates the overall distribution and overlap of GA estimates by both methods, indicating strong agreement in the term neonate population studied.

Percentages correspond to the number of neonates in each gestational age category as per USG, with NBS estimates showing similar distribution patterns.

This visualization supports the validity of NBS for postnatal gestational age estimation in term neonates in South India.

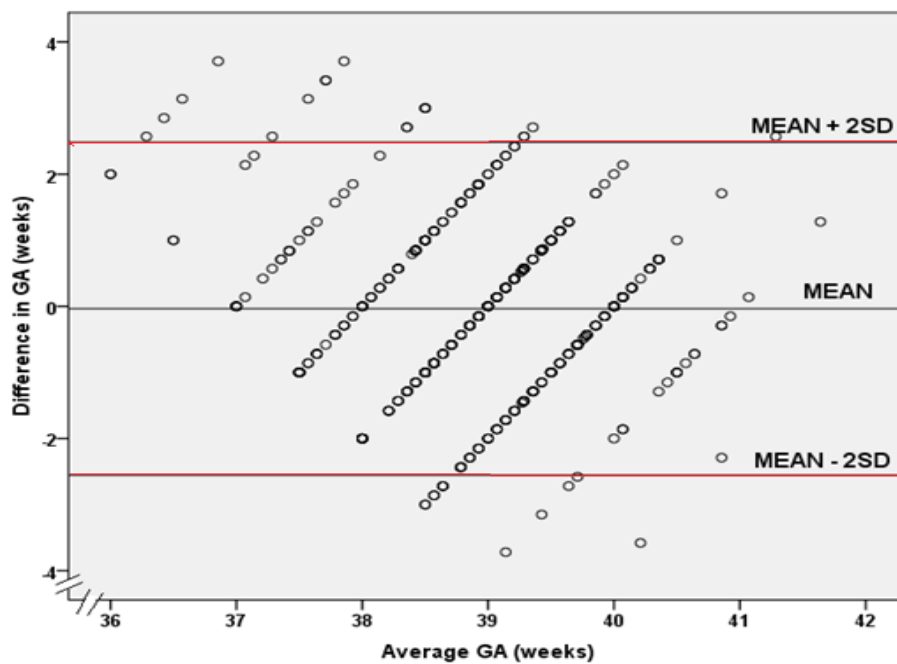


Figure 2: Bland Altman plot of the average and difference between the USG and NBS, along with the limits.

X axis is the average GA in weeks and Y axis the difference in gestational age calculated by subtracting NBS estimated GA from USG estimated GA.

The middle black horizontal line depicts the mean difference in gestational age by the two methods (-0.04). Coloured lines depict the limits of agreement which ranges from -2.56 to +2.49 weeks which is closer to the clinically permissible limits of difference in GA (+/-2weeks).

Each dot may represent more than one subject. As the gestational age increases the difference decreases.

DISCUSSION

In the present study, the New Ballard Score (NBS) was prospectively validated in 500 term neonates, demonstrating a mean total score of 37.89 ± 2.72 , with physical maturity and neuromuscular maturity scores averaging 18.89 ± 1.66 and 19.00 ± 1.67 , respectively. Small for gestational age (SGA) neonates exhibited slightly lower mean scores (36.82 ± 3.32). A moderate positive correlation was observed between total NBS score and gestational age (GA) confirmed by first trimester ultrasound ($r = 0.34$, $p < 0.001$), particularly in physical maturity components such as plantar creases ($r = 0.21$), breast development ($r = 0.23$), and skin texture ($r = 0.15$), surpassing correlations found in neuromuscular signs.

These findings are consistent with those reported by the systematic review conducted by Lee et al., which highlighted that neonatal assessments employing the Ballard score tend to overestimate GA by approximately 0.4 weeks and generally provide GA estimations within ± 3.8 weeks when compared to ultrasound, with physical maturity signs contributing significantly to the accuracy of assessments [1]. Your observed Pearson correlation coefficients align with this moderate accuracy, substantiating the relevance of physical signs in gestational age estimation.

In concordance, a cohort study by Smith et al. focusing on neonatal EEG and neurodevelopment in full-term SGA infants found that SGA neonates displayed immature neurodevelopmental markers, which parallels your observation of lower physical maturity scores in the SGA group (mean 36.82 ± 3.32) [7]. This reinforces the influence of intrauterine growth restriction on neonatal maturity markers, further validating the utility of NBS in this subgroup.

Conversely, your moderate correlation ($r = 0.34$) contrasts with findings from some fetal biometry predictive models, which predict SGA with higher sensitivity rates; for example, combined screening using maternal characteristics and estimated fetal weight (EFW) Z-scores at 35-37 weeks gestation predicted 89% of SGA neonates with birth weight below the 5th percentile delivering within two weeks, evidencing greater predictive accuracy in antenatal settings compared to postnatal NBS correlations [8]. This difference underlines the inherent distinction between prenatal ultrasound-based prediction and postnatal maturity scoring.

Furthermore, your study's emphasis on physical maturity signs showing stronger correlation than neuromuscular components mirrors findings reported in neonatal diagnostic accuracy reviews, where physical parameters such as skin texture and breast development demonstrated higher reliability for GA estimation than neuromuscular signs [1]. This similarity underscores the robustness of physical maturity markers as stable postnatal indicators.

In contrast, some studies focusing primarily on preterm or low birth weight populations revealed stronger associations with neuromuscular maturity indicators but had less applicability in term neonates, possibly explaining differences in correlation magnitudes and relevancy across studies [1]. Your focus on term neonates in South India, with first trimester ultrasound-confirmed GA, thus provides regionally specific validation, contributing valuable data to the literature.

Regarding the demographic characteristics of your cohort, the inclusion of 500 term neonates with ultrasound-confirmed GA provides a robust baseline compared to several studies utilizing mixed gestational age groups or retrospective designs. For example, Sharma et al. reported a higher prevalence of SGA infants showing adverse neonatal outcomes in large retrospective cohorts but without first trimester ultrasound confirmation of GA, which may limit the precision of gestational dating relative to your prospective validation [9].

Our present study corroborates prior evidence that the New Ballard Score is a useful, moderately accurate tool for estimating gestational age in term neonates, with particularly reliable physical maturity components. While antenatal ultrasound and fetal biometry retain superior accuracy for predicting SGA status antepartum, your findings reinforce the applicability of postnatal neonatal assessments in resource-limited settings, specifically highlighting the somewhat lower maturity scores observed in SGA neonates paralleling immaturity and developmental delay risks noted in similar populations [1,7].

The present study demonstrated a significant correlation between the New Ballard Score (NBS) and ultrasound (USG) gestational age (GA) estimates in the Appropriate for Gestational Age (AGA) subgroup, albeit with a slightly lower correlation coefficient (total score $r = 0.27$, $p < 0.001$). No significant correlation was observed in the Large for Gestational Age (LGA) subgroup, likely due to the small sample size, while paired t-tests revealed no systematic bias

between NBS and USG methods (mean difference -0.04 weeks). Furthermore, Bland–Altman analysis confirmed acceptable agreement between NBS and USG with limits of agreement ranging from -2.56 to +2.49 weeks, falling within clinically acceptable margins of ± 2 weeks. Notably, the difference between these methods decreased as GA increased.

These findings align partially with results from previous studies evaluating neonatal anthropometric measurements and clinical assessment tools for GA estimation. For example, a large cross-sectional study in India assessing neonatal anthropometry including birthweight, head circumference, and mid-upper arm circumference found significant correlations with gestational age, with birthweight correlating at $R = 0.72$ and mid-upper arm circumference at $R = 0.67$ [3]. Their model predicted GA to within ± 2 weeks accuracy in 75.5% of newborns, highlighting the viability of anthropometric parameters in GA estimation, although their correlation coefficients exceeded the present study's values, possibly due to methodological differences or sample heterogeneity.

In contrast, studies focusing on LGA neonates have reported challenges similar to those observed in this study's LGA subgroup. The lack of significant correlation in the LGA group is consistent with reports highlighting that growth deviations such as LGA may confound clinical GA assessments [10]. Large multicenter cohorts documented that LGA infants constitute about 12.1% of term births and exhibit increased composite neonatal morbidity, complicating straightforward clinical assessments [11]. This supports the interpretation that small LGA sample sizes and intrinsic growth variability may explain the absence of correlation observed in the present study.

Further systematic reviews emphasize limitations of neonatal assessment scoring systems like Ballard in certain populations. One comprehensive meta-analysis found that the Ballard score generally overestimates GA by about 0.4 weeks with a wide limit of agreement averaging ± 3.8 weeks when compared to ultrasound, suggesting somewhat lower precision than observed in the current study [1]. The present findings of limits within ± 2.5 weeks therefore indicate comparatively better agreement between NBS and USG in your cohort, possibly reflecting methodological rigor or population-specific factors.

Conversely, some studies question the reliability of the Ballard scoring system in low-resource or specific clinical settings. Research conducted in malaria-endemic regions reported poor to moderate correlations between Ballard estimates and ultrasound, with Ballard sensitivity to detect prematurity as low as 42%, suggesting the tool's limited precision where maternal and fetal conditions vary [2]. The stronger correlations and agreement metrics in your study contrast with such findings, underscoring that in South Indian term neonates, the New Ballard Score may offer acceptable accuracy, especially in the AGA group.

In summary, our study contributes to the body of evidence supporting the New Ballard Score as a reliable clinical tool for postnatal GA estimation in term AGA neonates, corroborated by negligible mean difference and acceptable agreement with ultrasound-derived GA. However, the findings also reinforce existing concerns regarding the tool's applicability in LGA infants, where growth deviations may impair correlation and accuracy. These observations echo prior large-scale studies indicating increased morbidity and variability in LGA neonates that complicate clinical assessments [10,11]. Future investigations with larger LGA sample sizes and complementary anthropometric or biochemical markers may enhance GA estimation in this subgroup.

The present study's findings validate the New Ballard Score (NBS) as a reliable postnatal tool for gestational age (GA) estimation in term neonates within the South Indian population, demonstrating close agreement with ultrasound (USG) estimates. The close matching between NBS and USG distributions in your cohort supports the practical applicability of NBS where early ultrasound dating is unavailable. This aligns with previous literature, including a systematic review by Lee et al., which reported the Ballard score dating 95% of pregnancies within ± 3.8 weeks compared to ultrasound, albeit with a small overestimation bias of 0.4 weeks [1]. Your study's lack of significant systematic difference between NBS and USG—mean difference of -0.04 weeks—demonstrates at least comparable, if not improved, accuracy in your setting.

Similarly, the cross-sectional study from Delhi by Gupta et al. found strong correlations between neonatal anthropometry and GA, with birthweight ($R = 0.72$) and mid-upper arm circumference ($R = 0.67$) correlating well with gestation as estimated by the New Ballard Score [3]. Your findings supplement these results by reinforcing NBS reliability as a clinical tool that parallels ultrasound accuracy in term neonates. The predictability of anthropometric models—reporting 75.5% accuracy within ± 2 weeks—offers an adjunct rather than replacement, while your study confirms NBS's role as a frontline assessment in a South Indian setting.

Conversely, the validity of Ballard scoring appears more limited in certain low-resource contexts. For example, a validation study in rural Bangladesh by Tielsch et al. showed a wider limit of agreement between Ballard and ultrasound dating (-4.7 to +4.0 weeks) and lower sensitivity (16%) for identifying preterm infants using the Ballard score [4]. This contrasts with your study's narrower agreement margins (-2.56 to +2.49 weeks) and negligible systematic bias, indicating that the NBS may perform better in your population or clinical context. The Bangladesh findings also highlighted

challenges with simplified clinical tools and neonatal anthropometry for GA classification, consistent with your observation of small subgroup variability that may affect scores.

Similarly, a cohort study from sub-Saharan Africa evaluating Ballard score performance reported poor to moderate correlations with ultrasound estimates and limited precision in identifying prematurity, suggesting SFH measurement may outperform Ballard in antenatal dating [2]. This discrepancy with your findings may reflect regional differences, malaria-influenced fetal growth patterns, or methodological factors such as timing and scorer expertise. Importantly, your study's close agreement with USG underlines the NBS's applicability in a well-defined South Indian term neonatal population where early ultrasound access is often limited.

In Papua New Guinea, Preterm or Not study revealed that Ballard score was the least reliable predictor for preterm birth (F-measure 0.17), while combined clinical measures improved accuracy [12]. These findings contrast with your study's confirmation that NBS estimations correlate well with ultrasound in term neonates, suggesting that the score's precision is better within term populations compared to mixed or preterm cohorts.

The present study corroborates prior Indian data indicating that NBS is a reasonably accurate surrogate for GA estimation in term neonates, supporting its use in South Indian clinical settings lacking universal early ultrasound availability [3]. The minimal bias and acceptable limits of agreement observed compare favorably with systematic review data [1], emphasizing NBS's utility as a cost-effective, practical tool. Nevertheless, differences with studies from malaria-endemic regions, low-resource community settings, or mixed gestational groups highlight the importance of population-specific evaluation and the limitations of Ballard scoring for preterm or growth-restricted neonates [2,4,12]. Future research may focus on refining neonatal assessment algorithms incorporating gestational age-related anthropometric norms tailored for diverse populations to enhance accuracy further.

The present study has several limitations. First, the sample comprised only term neonates born in a single tertiary care hospital in South India, which may limit the generalizability of findings to preterm infants, other healthcare settings, or broader geographic populations. The relatively small number of Large for Gestational Age (LGA) neonates restricted the ability to draw definitive conclusions about the New Ballard Score's (NBS) accuracy in this subgroup, as reflected by the lack of significant correlations observed. Additionally, the study relied on a single trained investigator for NBS assessments to minimize inter-observer variability, but this precludes evaluation of reproducibility across different examiners, which is important for clinical applicability. The timing of assessment, between 4 and 96 hours after birth, may introduce variability due to postnatal physiological changes affecting neuromuscular and physical maturity signs. Furthermore, the study focused on term neonates with first trimester ultrasound-confirmed gestational age, thereby excluding preterm and unstable neonates who might benefit most from improved postnatal GA assessment tools. Lastly, while the NBS showed moderate correlation with ultrasound dating, the limits of agreement indicate some degree of imprecision that could impact clinical decision-making in individual cases. Future studies with larger, more diverse cohorts, including preterm and growth-restricted infants, and involving multiple assessors are warranted to enhance validation and clinical utility of the NBS in varied settings.

The present study recommends that the New Ballard Score (NBS) be utilized as a practical tool for postnatal gestational age assessment in term neonates within similar clinical settings, especially where early ultrasound is unavailable. To improve the robustness and clinical applicability of the NBS, future research should include larger, more heterogeneous populations encompassing preterm and growth-restricted infants. Additionally, involving multiple assessors will help evaluate inter-observer reliability. Refinement of timing for assessment and integration with complementary clinical or anthropometric measures may further enhance accuracy and utility across diverse neonatal subgroups.

CONCLUSION

The present study successfully validated the New Ballard Score (NBS) as a practical and moderately accurate tool for postnatal gestational age assessment in term neonates within a South Indian tertiary care setting. Key findings demonstrated a significant positive correlation between NBS total scores and gestational age determined by first trimester ultrasound, with physical maturity parameters showing stronger associations than neuromuscular signs. The study confirmed acceptable agreement between NBS and ultrasound estimates, with no significant systematic bias and clinically reasonable limits of agreement. However, the NBS showed limited accuracy in the small Large for Gestational Age (LGA) subgroup, likely due to sample size constraints. Overall, these results highlight the NBS's utility as a reliable alternative for gestational age estimation where early ultrasound is unavailable, supporting improved clinical decision-making in similar resource-limited contexts. The study underscores the importance of regional validation and provides a foundation for expanding NBS application in diverse neonatal populations.

REFERENCES

1. Lee AC, Panchal P, Folger L, Whelan H, Whelan R, Rosner B, et al. Diagnostic Accuracy of Neonatal Assessment for Gestational Age Determination: A Systematic Review. *Pediatrics*. 2017 Nov 17;140(6):e20171423.

2. Unger H, Thriemer K, Ley B, Tinto H, Traoré M, Valea I, et al. The assessment of gestational age: a comparison of different methods from a malaria pregnancy cohort in sub-Saharan Africa. *BMC Pregnancy Childbirth*. 2019 Jan 8;19(1):12.
3. Rajat Thawani R, Dewan P, Faridi M, Arora SK, Kumar R. Estimation of Gestational Age, Using Neonatal Anthropometry: A Cross-sectional Study in India. *J Health PopulNutr*. 2014 Aug 17;31(4):523–30.
4. Lee AC, Mullany LC, Ladhani K, Uddin J, Mitra D, Ahmed P, et al. Validity of Newborn Clinical Assessment to Determine Gestational Age in Bangladesh. Vol. 138, *Pediatrics*. American Academy Of Pediatrics; 2016.
5. Rittenhouse KJ, Vwalika B, Keil A, Winston J, Stoner M, Price JT, et al. Improving preterm newborn identification in low-resource settings with machine learning. *PLoS ONE*. 2019 Feb 27;14(2):e0198919.
6. Vardhelli V, Singh S, Ghimire A, Gopireddy MMR, Deshabhotla S, Ravula PC, et al. Updated Size at Birth Charts for South Indian Neonates. *Indian J Pediatr*. 2025 Apr 5;92(6):637–639.
7. Castro Conde JR, González Campo C, González González NL, Reyes Millán B, González Barrios D, Jiménez Sosa A, et al. Assessment of neonatal EEG background and neurodevelopment in full-term small for their gestational age infants. *Pediatr Res*. 2019 Dec 10;88(1):91–99.
8. Fadigas C, Saiid Y, Gonzalez R, Poon LC, Nicolaidis KH. Prediction of small-for-gestational-age neonates: screening by fetal biometry at 35-37 weeks. *Ultrasound in Obstet& Gyne*. 2015 Apr 9;45(5):559–565.
9. Nohuz E, Rivière O, Coste K, Vendittelli F. Prenatal identification of small-for-gestational age and risk of neonatal morbidity and stillbirth. *Ultrasound in Obstet& Gyne*. 2020 Apr 6;55(5):621–628.
10. Truong V, Pedroza C, Chauhan S, Mendez-Figueroa H. Large for Gestational Age Infants and Adverse Outcomes among Uncomplicated Pregnancies at Term. *Amer J Perinatol*. 2016 Dec 7;34(7):655–662.
11. Chauhan SP, Rice MM, Grobman WA, Bailit J, Reddy UM, Wapner RJ, et al. Neonatal Morbidity of Small- and Large-for-Gestational-Age Neonates Born at Term in Uncomplicated Pregnancies. *Obstetrics and gynecology*. 2017 Aug 11;130(3):511–519.
12. Karl S, Li Wai Suen CSN, Unger HW, Ome-Kaius M, Mola G, White L, et al. Preterm or Not – An Evaluation of Estimates of Gestational Age in a Cohort of Women from Rural Papua New Guinea. *PLoS ONE*. 2015 May 6;10(5):e0124286.