

Newborn skin maturity model for gestational age prediction: a clinical trial for a novel medical device validation

Zilma Silveira Nogueira Reis (✉ zilma@ufmg.br)

Department of Gynecology and Obstetrics, Faculty of Medicine, Universidade Federal de Minas Gerais, Brazil. Center for Artificial Intelligence, Innovation and Health, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0001-6374-9295>

Rodney Nascimento Guimarães

Department of Gynecology and Obstetrics, Faculty of Medicine, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0001-5617-4470>

Roberta Maia de Castro Romanelli

Department of Pediatrics, Faculty of Medicine, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0002-1660-0751>

Juliano de Souza Gaspar

Department of Gynecology and Obstetrics, Faculty of Medicine, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0003-0670-9021>

Gabriela Silveira Neves

Hospital Sofia Feldman, Brazil. Department of Pediatrics, Faculty of Medicine, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0001-6765-2968>

Marynea Silva do Vale

Maternal and Child Unit, University Hospital, Universidade Federal do Maranhão, EBSEH, Brazil. <https://orcid.org/0000-0002-1544-5470>

Paulo de Jesus H Nader

University Hospital of Canoas, Universidade Luterana do Brasil, Brazil.

Martha David Rocha de Moura

Hospital Materno Infantil de Brasília, Brazil. <https://orcid.org/0000-0002-7851-375X>

Gabriela Luíza Nogueira Vitral

Department of Pediatrics, Faculty of Medicine, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0001-7306-8776>

Marconi Augusto Aguiar dos Reis

Department of Pediatrics, Faculty of Medicine, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0001-5294-1361>

Marcia Margarida Mendonça Pereira

Maternal and Child Unit, University Hospital, Universidade Federal do Maranhão, EBSEH, Brazil.

<https://orcid.org/0000-0002-2699-2963>

Patrícia Franco Marques

Maternal and Child Unit, University Hospital, Universidade Federal do Maranhão, EBSEH, Brazil.

<https://orcid.org/0000-0001-6812-6687>

Silvana Salgado Nader

University Hospital of Canoas, Universidade Luterana do Brasil, Brazil. <https://orcid.org/0000-0001-9360-6201>

Augusta Luize Harff

University Hospital of Canoas, Universidade Luterana do Brasil, Brazil. <https://orcid.org/0000-0003-3220-1643>

Ludmylla de Oliveira Beleza

Hospital Materno Infantil de Brasília, Brazil. <https://orcid.org/0000-0001-9975-562X>

Maria Eduarda Canellas de Castro

Hospital Materno Infantil de Brasília, Brazil. <https://orcid.org/0000-0001-8727-8328>

Rayner Guilherme Souza

Department of Gynecology and Obstetrics, Faculty of Medicine, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0003-2440-2242>

Gisele Lobo Pappa

Center for Artificial Intelligence, Innovation and Health, Universidade Federal de Minas Gerais, Brazil. Computer Science Department, Universidade Federal de Minas Gerais, Brazil. <https://orcid.org/0000-0002-0349-4494>

Regina Amélia Pessoa Lopes de Aguiar

State Health Department of Minas Gerais, Brazil <https://orcid.org/0000-0003-2470-3539>

Research Article

Keywords: gestational age, prematurity, childbirth, skin physiological phenomena, machine learning, equipment and supplies

Posted Date: April 13th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1216628/v3>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The early access to prenatal care and high-cost technologies for pregnancy dating challenge the early neonatal risks assessment at birth in resource-constrained settings. To overcome the absence or low accuracy of postnatal gestational age, we developed a frugal innovation based on the photobiological properties of the newborn's skin and predictive models.

Methods: A multicenter single-blinding, and single-arm clinical trial intention-to-diagnosis evaluated the accuracy of a novel device to detect gestational age and preterm newborns. The first-trimester ultrasound (US), a second comparator US, and the last menstrual period (LMP) data from antenatal reports were the references for gestational age at birth. A portable multiband reflectance photometer assessed 781 newborns' skin maturity and used machine learning models to predict gestational age, adjusted to birth weight and antenatal corticosteroid therapy exposure.

Results: As the primary outcome, the predicted gestational by the new test had high agreement with the reference gestational age calculated with the intraclass correlation coefficient (0.970 [95%CI: 0.965, 0.974]) similar values to the comparator-US and better than the comparator-LMP gestational ages. As secondary outcomes, the new test achieved 97.7% (95%CI: 96.5%, 98.6%) agreement with the reference gestational age within one-week error. This value surpassed those of comparator-US (91.3% [95%CI: 89.2%, 93.1%]), and of comparator-LMP gestational ages (64.1% [60.7% to 67.5%]). Bland-Altman limits of the new test were -7.1 to 4.7 days. Prematurity discrimination with the novel device had the area under the receiver operating characteristic curve (AUROC) (0.998 [95%CI: 0.997, 1.000]), similar to comparator-US (0.996 [95% CI: 0.993, 0.999]); and superior to comparator-LMP gestational ages (0.957 [95%CI:0.941, 0.974]). In newborns with absent or unreliable LMP (n=451), the intent-to-discriminate analysis showed correct classifications with the new test of 96.5% (95%CI: 94.3%, 98.0%), while with the comparator-LMP gestational age was 69.6% (95% CI: 65.3%, 73.7%).

Interpretation: The assessment of the newborn's skin maturity adjusted by learning models promises accurate pregnancy dating at birth even without the antenatal ultrasound reference.

Funding: Grand Challenges Exploration from the Bill & Melinda Gates Foundation, *Fundação de Amparo a Pesquisa de Minas Gerais*, Brazilian Ministry of Health, CNPq - *Conselho Nacional de Desenvolvimento Científico e Tecnológico*.

Introduction

Public policies and best practices advise pregnant women to plan pregnancy with early access to prenatal care for pregnancies to be safely monitored until birth.¹ However, many barriers to covering all pregnancies and births with due care remain unsurpassed, particularly in scenarios without well-equipped facilities.² Adverse maternal and neonatal outcomes affect newborns unevenly according to the birth scenario and gestational age.³ First-day-of-life mortality is 30 times higher in low- and medium-income countries than in high-income countries.⁴ Preterm neonates are more vulnerable to death or survival with neurological sequelae. The need to recognize early risks at birth is faced with reduced early access to prenatal care and high-cost technologies for pregnancy dating, such as obstetric echography in resource-constrained settings.⁵ Early obstetric ultrasound currently offers the best gestational age method.⁶ However, access to high-cost equipment, poor training, lack of skills of health professionals, or late prenatal care limit pregnancy dating and, consequently, detection of prematurity.^{5,7} Improving preterm birth outcomes demands accurate gestational age assessment to direct opportune decisions regarding neonatal care.⁸ Approaches to enhance the reliability of pregnancy dating through more accurate and accessible technologies can improve pregnancy outcomes and neonatal survival.^{6,9}

Health technology development is critical for supporting healthcare systems. Medical devices and digital health technologies have brought innovative solutions with the potential to save lives,¹⁰ mitigating gaps of quality among disparate healthcare scenarios.¹¹ Furthermore, digital health technologies emerge with the potential to impact the equality of healthcare, creating new landscapes of opportunities such as applied data science to improve prediction models.¹² Currently, computer science has advanced with improvements to medical practice, detecting patterns by processing datasets through layered mathematical models,¹³ fostering skills and competencies of professionals to support the best healthcare decisions.¹⁰

The new test explored in this study is an innovative approach used to estimate gestational age based on the photobiological properties of the newborn's skin and by means of learning predictive models enhanced with clinical variables.¹⁴ Usable as a medical device, we developed this technology to easily assist health professionals in making better decisions regarding newborn care whenever the pregnancy dating is unknown or doubtful. This study aims to validate the photobiological model of skin maturity adjusted to the clinical data to promptly detect gestational age and determine its accuracy in detecting prematurity. We tested the hypothesis of equivalence between the gestational age measured by this new test, the pregnancy dating comparators calculated using ultrasound exams, and the last menstrual period (LMP).

Methods

Study design and Participants

A multicenter prospective clinical trial intention-to-diagnosis study by singlegroup, singleblinding, and singlearm with a reference standard. This article adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prediction or Diagnosis (TRIPOD) for completeness and clearness.¹⁵ To assess the risk of bias and applicability, the development and validation methods followed guidance from the Prediction Model Risk of Bias Assessment Tool (PROBAST).¹⁶ The clinical trial protocol is disclosed in the WHO's International Clinical Trial Platform - Brazilian Clinical Trials Registry RBR-3f5bm5.

This report examined primary and secondary outcomes of data concerning gestational age prediction and clinical safety of the novel device. Secondary outcomes related to lung maturity prediction are under analysis for further publication. Five Brazilian referral centers participated in the study: Clinical Hospital – Universidade Federal de Minas Gerais (as coordinator); Hospital Sofia Feldman – Minas Gerais State; Hospital da Universidade Luterana do Brasil – Rio Grande do Sul State; Hospital Materno-infantil de Brasília – Distrito Federal; and Hospital Universitário da Universidade Federal do Maranhão – Maranhão State. The local independent ethics review board approved the research protocol, registered under the number CAAE 81347817.6.1001.5149 at the Brazilian National Research Council. In addition, parents signed an informed consent form on behalf of the newborn before participating.

A prospective concurrent and sequential process enrolled newborns during the first 24 hours of life. The first enrollment occurred on 2019-01-02, and the last enrollment occurred on 2021-05-30. Eligibility criteria, participants' timeline, and the procedures followed the clinical protocol.¹⁷ In short, we assessed the skin maturity of live newborns with at least 24 weeks or more of gestational age. All of them had reports of antenatal ultrasound, one from 7 and 13 weeks and 6 days, and the other from 14 and 23 weeks and 6 days of gestation. Anhydramnios, hydrops, congenital skin diseases, or chorioamnionitis were the exclusion criteria due to their potential to modify the skin structure. Detailed procedures are available in appendix pp 2 and the published protocol.¹⁷

The Intervention

The intervention is a novel device that processes the backscattered signal acquired from the skin of the newborn's sole with clinical variables to predict gestational age. Its development included steps from the workbench to the clinical experimentation already described.¹⁴ Likewise, we previously analyzed the best body position to assess the skin reflectance for pregnancy dating, as well as environmental influences such as humidity, temperature, ambient light, and the newborn's skin color.^{14,18} Regarding the characteristics of the components, wavelengths from 400 nm to 1200 nm of the light emitter categorized the safety level of this medical device as Class II (noninvasive and medium risk) according to the regulatory agency in Brazil. When the light-emitting sensor touches the skin over the sole for a few seconds, it triggers 10 automated measurements. The device emits alerts regarding measurement errors caused by the involuntary movement of the newborn or examiner under a set of known constraints of the skin reflectance of a newborn, events that emit alerts and requires a new

attempt.¹⁴ The device output was blinded to the examiners. The reliability of the skin reflection acquisition was assessed during the certification visit of a senior researcher in the collaborator centers (appendix p 8).

The skin assessment occurred with the newborn inside incubators, incubators-radiant warmer, warming pad-bassinet, standard crib, or in the mother's lap to ensure minimum manipulation and avoid unbalancing the clinical conditions. The sensor touched the sole three times, following complete disinfection with alcohol. Fourteen minimum viable products were produced for the study (figure 1). At the beginning and the end of the clinical trial, the irradiance emitted by each device and the reflection against a standard-white Spectralon® offered values for calibration. The adjusted value was the raw value of the acquisition divided by the irradiance of the Light Emitting Diode of each device.

Improvements on the gestational age estimation model

The new test had an algorithm to predict gestational age, previously described elsewhere and patented.¹⁴ We assessed the Pearson coefficient to homologate the correlation between skin reflectance with the reference gestational age. Skin reflectance had a strong positive correlation with reference gestational age; $r = 0.79$, $p < 0.001$, figure 2.

The standalone newborn skin reflectance value was the raw variable, adjusted by clinical variables. The current data set the groundwork for improvements in the algorithm predicting gestational age based on machine learning models as part of the research protocol. The analytical pipeline is detailed in (appendix pp 8-10). The nonlinear machine learning method eXtreme Gradient Boosting (XGBoost) created at most 50 trees with a maximum depth of three. The models were validated using a 10-fold cross-validation approach ten-fold, repeated 30 times. All clinical variables used as predictors of the models were available at the test time. Therefore, they could be used in real scenarios from user input into the medical device interface.

The set of models' performance is presented as supplementary material (appendix pp 9-12), including intermediary analysis with incubator stay, gender, and jaundice variables to verify the elimination of intervenient variables after technological improvements since the early version of the device.¹⁴ The skin acquisition, adjusted for birth weight, achieved an R^2 of 0.983 and a low mean absolute error (MAE), 0.232 weeks which means 1.62 days. Adding antenatal corticosteroid therapy for fetal maturation exposure (ACTFM) information, the model achieved even better performance in terms of R^2 of 0.986, with an MAE of 0.295 weeks which means 2.07 days. Such three predictive variables model was the new test under validation in this study. However, three gestational age predictions had ACTFM data imputation due to missing information related to failures in the prenatal record available at maternity admission.

Outcomes

The primary outcome was the agreement between the gestational age predicted by the new test and the reference gestational age. The secondary endpoint was the accuracy of the new test to discriminate

preterm newborns considering thresholds at 37, 32, and 28 weeks of pregnancy. Moreover, the proportion of preterm newborns correctly detected at birth within a one-week error margin. Another secondary endpoint was comparing the difference between predicted gestational age and the gestational age calculated by a second ultrasound exam after 13 weeks and 6 days of gestation and before 22 weeks via comparator-ultrasound gestational age (comparator-US GA) and with the comparator-last menstrual period gestational age (comparator-LMP GA). This outcome was intended to simulate the performance of the test in scenarios without the reference and compare the agreement between established methods for gestational age calculation and the new test. The safety of the new device is still a derived endpoint, which refers to the report of unexpected medical events, unintended illness or injury, or unfortunate clinical signs in subjects, users, or others, whether or not they are related to the investigated product. The users answered nine questions about issues with the medical device, after each skin acquisition (appendix pp 4).

Statistical analysis

Intention-to-diagnosis analysis guaranteed all newborns who were included in the statistical analysis, regardless of any result of the new test, and even inconclusive ones. Descriptive analysis of the newborn's clinical characteristics and the intervention measurements were performed.

Regarding the primary endpoint, the agreement among different methods for gestational age at birth determination was calculated using the intraclass coefficient correlation and Bland & Altman intervals and the day-difference to reference gestational age. Regarding the accuracy of predicted gestational age by the models in identifying premature newborns, the area under the receiver operating characteristic curve (AUROC) at 95% CI described the model's discrimination and diagnostic parameters. Qui-square Test, Mann Whitney Test, and Mean-paired differences compared interest groups. The significance level for hypothesis tests is 5%. The SPSS software 19.0 was used in the statistical analysis of the data.

Role of the funding source

No sponsor had any role in the study design, data collection, data analysis, data interpretation, writing of, or decision to submit the manuscript.

Results

Of the 791 potential eligible newborns, two were under Rh alloimmunization during pregnancy, which was considered an exclusion criterion (figure 3). Among the 789 newborns with their skin assessed with the medical device, eight had no reference standard to allow the dependent variable; four had no antenatal first-trimester ultrasound; three had no comparator ultrasound, and one had an unsolved digit date error. All 781 newborns who met the eligibility criteria of the clinical trial were included in the analysis.

Seventy hundred two pregnant women gave birth to 781 newborns. Despite early access to prenatal care (median 12, IQR 4 weeks, table 1), only 296 (42.2%) women fulfilled the criteria for reliable LMP, among 613 who were able to provide such a date. According to reference gestational age at birth, 415 of 781 (53.1%) newborns were term. Among 366 (46.9%) preterm newborns, 235 (30.1%) had a gestational age at birth less than 37 to 32 weeks, 131 (16.8%) had a gestational age less than 32 weeks, and 42 (5.4%) less than 28 weeks. A number of the newborns (n= 273, 35.1%) received ACTFM, following local protocols, and three were missing. The frequency of fetal abnormal growth classification at birth was 115 (14.7%) who were small for gestational age and 59 (7.6%) who were large for gestational age. About a third (35.9%) of newborns were in intensive care at the time of the skin assessment.

Gestational age estimation at birth

The distribution of gestational age calculated according to the different bases corroborated some differences among established methods of antenatal dating figured in the overlapped histogram, in weeks of gestation (figure 4). Reference gestational age had a median of 37.3 (IQR 6.3) weeks, lower than the reference the comparator-US-GA who had a median of 37.1 (IQR 6.1) weeks, $p < 0.001$ (paired-Wilcoxon Test). However, the comparator-LMP-GA, when available had a median of 37.4 (IQR 6.8) weeks, similar to the reference gestational age $p = 0.282$ (paired-Wilcoxon Test). Preterm birth had a frequency of 46.9% (366/781), 47.1% (368/781), and 45.6% (310/680) according to reference gestational age, comparator-US-GA, and comparator-LMP-GA, respectively. Post-term birth had a frequency of 0.1% (1/781), 0.3% (2/781), and 4.0% (27/680) regarding reference gestational age, comparator-US-GA, and comparator-LMP-GA, respectively. Otherwise, data quality of LMP recall revealed the most frequent digit preferences were for day-5 51 (8.3%), day-15 (6.7%), day-20 (7.2%), and day-25 (4.7%). Those frequencies had significant differences when compared to the day adjusted to the reference gestational age $p < 0.008$ (Cochran Q test for k-related samples) (appendix pp 6).

The new test provided by the medical device predicted gestational age based on newborn skin reflectance values, the raw variable, adjusted by clinical variables, birth weight, and ACTFM exposure using the XGBoost algorithm. Details are in supplementary data (appendix p12). Agreement between the predicted gestational age, reference, and comparators was high considering the intraclass coefficient (ICC), table 2. Considering 95% CI, such a predictive model provided by the new test had a similar ICC to those calculated between established references with the reference gestational age. Besides, the coefficient of correlation of predicted gestational age with established methods had high exceptional values (figure 5).

The new test underestimated the reference gestational age in -1.30 (95% CI: -1.53 to -1.11) days, as well as in -0.79 (95% CI: -1.11 to -0.47) days, and -2.21 (95% CI: -3.11 to -1.30) days the comparators the US and LMP GA, respectively. In the meantime, the comparator-US GA underestimated the reference gestational age in -0.53 (95% CI: -0.88 to -0.19) days. The endpoints of the Bland-Altman 95% limits of agreement

were the 2.5th percentile and 97.5th percentile for the distribution of the difference between paired measurements (figure 5). Therefore 95% of the differences, between the new test and reference gestational age, were between -7.1 and 4.7 days. This range was shorter than those for the Comparator-US GA, -10.0 to 8.0 days, and comparator-LMP GA, -25.0 to 29.0, in relation to the reference gestational age.

The boxplots in figure 6 present the proportion of preterm newborns correctly detected at birth within a one-week error. The new test achieved 97.7 % (95% CI: 96.5% to 98.6%) of one-week error agreement with the reference pregnancy dating. This value exceeded the value of 91.3% (95% CI: 89.2% to 93.1%) of the comparator-US GA one-week error, and the value of 64.1% (60.7% to 67.5%) of the comparator-LMP GA one-week error considering the intention-to-diagnose analysis.

Accuracy of the new test to discriminate preterm newborns

Considering an overlapping of 95% CIs in AUROCs, the new test using a medical device had similar performance to comparator-US gestational age in discriminating preterm against term newborns at all cutoffs, respectively AUROC 0.998 (95% CI: 0.997, 0.999) versus 0.993 (95% CI: 0.988, 0.997) at 37 weeks; and superior to comparator-US 0.957 (95% CI: 0.941, 0.974), figure 7. At 32 weeks and 28 cutoffs, the new test had similar performance compared to the comparator-US gestational age again.

A comprehensive analysis of prediction accuracy for preterm newborns by the methods of gestational age estimation for different prematurity cutoffs is shown in appendix pp 13. We emphasize relevant likelihood ratio positive at 37 weeks 13.8 (9.6 to 19.9) when the medical device predicts gestational age, overlapping the comparators in terms of 95% CI: 25.0 (15.4 to 40.4) for -US GA, and 17.1 (11.0 to 26.6) for -LPM GA.

Intent to preterm newborn discrimination by the new test

Birth care settings where the new technology is to be applied deserve an intent to preterm newborn discriminant analysis, simulating the existence of baseline references for the gestational age calculation. Therefore, we considered newborns whose mothers had no recall of LMP or unreliable information as scenario one, corresponding to 451 (57.7%) of the newborns. In scenario two, we gathered newborns whose mothers had reliable LMP (table 3). Concerning missing data, three test values for gestational age obtained with ACTFM machine learning imputation were valid results for this analysis. At the same time, 101 missing data for LMP were newborns who had no comparator-LMP gestational age due to unknown menstrual dates.

The lack of an unreliable LMP of scenario-one resulted in low discrimination accuracy (69.6% [95%CI: 65.3% to 73.7%]) with the comparator-LMP GA. Nevertheless, 96.5% (95% CI: 94.3% to 98.0%) of the

newborns were correctly classified as preterm or term using the new test. Great accuracy using any available method for gestational age estimation was observed in scenario-two, where the LMP was reliable. In such a scenario, we stand similar accuracy of the new test (97.9%, [95% CI: 95.7% to 99.1%]), when compared to the comparator-US GA (97.0%, [95% CI: 94.5% to 98.5%]) and to the comparator-LMP GA (96.1%, [95% CI: 93.4% to 97.9%]). The overall analysis includes the crosstabs in appendix pp 15-16.

Safety of the device

There were no reports of unexpected medical events, unintended illness or injury, or unfortunate clinical signs in subjects, users, or others related to the investigational product or otherwise. Two devices were replaced due to an unintentional drop.

Discussion

A reliable antenatal age is a prerequisite for preterm newborn classification in birth care settings and constitutes the first step to delivering the necessary care, considering the risks of prematurity. Term newborn, allied with good tone, breathing, or crying, are essential elements to determine steps of newborn resuscitation.¹⁹ Although that statement in itself seems very simple, the reality is far from it. Without the certainty of the day in the female cycle on which conception occurred, ultrasound measurement of the crown-rump-length is a standard consensual reference for redating pregnancy in comparison with the LMP.⁶ This dependence on early echographic scans has deprived many pregnant women and their babies of reliable gestational age.⁸ Such a technological gap causes even more disparities than the difference between childbirth scenarios in fully equipped facilities and those ill-equipped with scarce technology. Moreover, it can impair the correct classification of infants as premature or growth-restricted.²⁰ The main contribution of this clinical trial is to validate a new approach for gestational age estimation independent of fetal ultrasound measures by demonstrating highly accurate outcomes. Based on birth weight, ACTFM exposure data, and the use of a frugal medical device to assess skin maturity and process algorithms, 364 of 366 preterm neonates with less than 37 weeks of gestation were detected, with 97.1% (95% CI: 95.6 to 98.1) being correctly classified.

In this combined study covering enhancing and validating prediction model for postnatal gestational age information, we believe the application of k-fold cross-validation with the use of machine learning algorithms provided accurate predictions.²¹ While large data samples are unavailable, the process of training and testing are able to estimate the performance of algorithms until we have finished other ongoing clinical trials for external validation.²² Furthermore, the quantification of uncertainty intervals regarding the predicted gestational age (calculated in days) and comparisons with established references allowed the simulation of realistic scenarios for application. Furthermore, the confidence intervals accompanying AUROC's accuracy contributed to revealing the forecast's limits as to discriminating terms from preterms newborns at different cutoff points, with clinical relevance. Such strengths are critical to

assure the potential value of the new test in facing the challenges of postnatal prematurity identification.²³ In fact, data science algorithms have thus emerged with the aim of qualifying pregnancy dating. High-performance reports using learning models based on antenatal ultrasound predictors²⁴ contradistinguished meager outcomes from those using other morphometric postnatal predictors.²⁵ Moreover, valuable algorithms with postnatal combinations on maturity scores of newborns are promising, demanding special skills to apply.²⁶ Underqualified birth attendants represent a challenge in developing countries, further limiting the use of existing birth care solutions.²⁷ One advantage of our new device is the skin assessment automation that alerts measurement errors caused by the movement of the newborn or examiner. Previous reports have detailed the human skin's light-skin interaction and optical properties that benefit this technology.^{14,18}

The predictive XGBoost algorithm used information that health professionals could quickly obtain in childbirth settings—the birth weight and the ACTFM exposure—and which could add value to the physical data of skin maturity. The gestational age estimated with the novel device had in this clinical trial a better agreement with the reference gestational age at birth than the antenatal comparators. The Bland-Altman, 95% limit, was lower than the comparators-US and -LMP gestational age. Working with more flexible forecasts within seven days of error, 97.7 % (95% CI: 96.5% to 98.6%) of newborns had a valuable prediction with the new test. Moreover, the device can provide a gestational age to overcome situations without ACTFM information as a potential tool in low resources birth settings. The prediction performance of the skin reflection and birth weight model remains with high R2 and low MAE values (appendix pp 12).

Considering the simulated scenario with absent or unreliable LMP (n=451), the new test had a better performance than the comparator-LMP gestational age and was similar to the comparator-US. This result highlighted the context of use for this medical device since the gestational age based on memory recall of the LMP missed 68 out of 199 preterm newborns, expressing a lower sensitivity (69.6%, [65.3% to 73.7%]) when we applied the intent-to-discriminate analysis.

Antenatal corticosteroid therapy for fetal maturation exposure regressor played an uncertain role in the predictive model. The model's ability to explain gestational age (R2) and MAE were similar with or without ACTFM as a predictive variable (appendix p12). Antenatal corticosteroids to improve newborn outcomes are an evidence-based intervention recommended for women at risk of preterm birth.²⁸ However, even with the acceleration of lung maturity, the effect of the drug occurs in other organs. The early fetal presence of receptors of corticosteroid hormone receptors in skin epithelial cells indicates that glucocorticoids may play an important role in the differentiation and development of human skin.²⁹ Regardless, clinical evidence of the effect of ACTFM exposure on skin maturity is weak, and the topic remains unsubstantiated.³⁰ Thus, the adoption of the new test deserves caution. Thus, until proven otherwise, we interpreted that the importance of ACTFM exposure data to better adjust the gestational age modeling is related to an effect on skin maturity. Even so, we cannot deny that antenatal exposure to corticotherapy is more common in premature infants—72.3% of preterm newborns in this sample. In this

respect, this regressor variable could imply a bias favoring preterm newborn detection. The aforementioned ongoing study for external validation of the algorithms could further elucidate this issue because the enrollment process of newborns introduced the Mozambican birth scenario, where unfortunately ACTFM is not guaranteed for every woman at risk of preterm birth.²²

Birth weight is a known estimator of risks to newborns. As part of primary routines in childbirth settings, this information has practical applicability, even in facilities with scarce high-cost technologies.¹ Meanwhile, predicting preterm birth based on birth weight when lacking a gold standard is an imperfect solution.⁷ Additionally, the LMP reference and the postnatal scores of newborn maturity have demonstrated low accuracy in determining gestational age and identifying prematurity.³¹ Later prenatal care and unqualified date recollection justify efforts to enhance the reliability of pregnancy dating through more accurate and accessible technologies, seeking to improve pregnancy outcomes and neonatal survival.⁸ In our study, qualifying the LMP at birth with questions about memory of date, menstrual cycles, and checking antenatal clinical documents at birth provided a gestational age able to identify 160/167 (95.8%, [91.6% to 98.3%]) preterm newborns, when available.

Current approaches to calculating gestational age are sensitive to data quality, resulting in a misplaced classification of prematurity.⁷ The present study was committed to representing a realistic scenario in terms of data quality, as stated in the research protocol, with data collection and curation to assure the best reference and comparators for the analysis. Before opening the blinding of the trial, a consistent process confronted data entries with digital images of clinical documents taken during the enrollment. Furthermore, dedicated software was developed exclusively for the clinical trial, considering the variables' quality and constraints. Part of the enrollment occurred during the COVID-19 pandemic, causing a minimal amount of missing data, such as the yes or no for ACMF information (3/781 newborns).

Regarding the generalizability of outcomes, this multicenter trial gathered perinatal centers from the northern, central, southwestern, and southern regions of Brazil. This collaborative evaluation contributed to sampling a mixed population of newborns with high miscegenation and involved 15 examiners who attended good clinical practice training. The intraobserver error and interobserver error of measures were slow, corroborating previous results.¹⁴ The number of preterm newborns was enough to analyze subcategories of prematurity as extreme preterm (n=42); however, the overall rate of preterm newborns was 46.9%, values observed in referral facilities and not in the general population of Brazilian newborns.³² The number of neonatal deaths during 72 h of follow-up was 14 (1.8%), with 12 deaths occurring in newborns with gestational age below 28 weeks due to extreme prematurity complications. We expect to target worse childbirth scenarios for this technology implementation.²⁷ In addition, the safety of this device is similar to other optical technologies already used in neonatal care.¹⁹

Identifying preterm newborns is the first step to attending to their needs. The global rate of neonatal mortality corresponds to 6,700 neonatal daily deaths, mostly from preventable or treatable conditions in scenarios of healthcare scarcity.³³ Without proper comparisons, the new test had a lower error range

than after birth scores of maturity. For the future, comparisons are expected based on postnatal approaches for gestational age estimation, such as scores of maturity and foot length or image combinations.²⁵ We hope that strengthening the data sources of healthcare facilities with a reliable gestational age can help in identifying vulnerable newborns in situations with the absence or lack of such information.

Declarations

Acknowledgments

This research was supported by the Grand Challenges Exploration from the Bill & Melinda Gates Foundation (Grant number OPP1128907, Contract), <http://www.gatesfoundation.org/>, and *Fundação de Amparo a Pesquisa de Minas Gerais*, Brazil, <http://www.fapemig.br/en/>, nonprofit sectors. The clinical trial is funded by the Brazilian Ministry of Health, Program of Development of the Industrial Health Complex (PROCIS), project 23072.052747/2017-51. ZSNR is a researcher with a grant from *Conselho Nacional de Pesquisa* (CNPq 305837/2021-4), Brazil. The authors would like to thank the Center for Clinical Research of Hospital das Clínicas, Universidade Federal de Minas Gerais for their general support and good clinical practice training. We also thank all the clinical staff at the involved hospitals who cared for the patients and undergraduate students who helped with the data collection.

Competing interests

The authors declare a patent deposit on behalf of the Universidade Federal de Minas Gerais and Fundação de Amparo a Pesquisa de Minas Gerais, Brazil, <http://www.fapemig.br/en/>. The inventors were ZSNR, RNG, BR1020170235688 (CTIT-PN862). BirthTech, a spin-off company, received a license to produce and commercialize this technology, and RNG is its founder.

Data Availability Section, Responsibility, and Analysis

The lead authors (ZSNR, RMCR, RNG, JSG, and RAPLA) had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Correspondence and requests for materials should be addressed to ZSNR. Data is available upon reasonable request and after anonymization to allow sharing of data ethically and legally, preserving the confidentiality of the persons who participated in this study. The study protocol, statistical analysis plan, and informed consent form are available in previous publications. Standard operational procedures with detailed methods are deposited in protocols.io

References

1. Tunçalp, Ö. *et al.* WHO recommendations on antenatal care for a positive pregnancy experience—going beyond survival. *BJOG Int J Obstet Gynaecol* 2017. **124**, 860–862.
2. Lawn, J. E. *et al.* Every Newborn: progress, priorities, and potential beyond survival. *Lancet Lond Engl* 2014. **384**, 189–205.
3. Walani, S. R. Global burden of preterm birth. *Int J Gynecol Obstet* 2020. **150**, 31–33.
4. Oza, S., Cousens, S. N. & Lawn, J. E. Estimation of daily risk of neonatal death, including the day of birth, in 186 countries in 2013: a vital-registration and modelling-based study. *Lancet Glob Health* 2014. **2**, e635-644.
5. Kim, E. T., Singh, K., Moran, A., Armbruster, D. & Kozuki, N. Obstetric ultrasound use in low and middle-income countries: a narrative review. *Reprod Health* 2018. **15**, 129.
6. ACOG. Committee Opinion No 700: Methods for Estimating the Due Date. *Obstet Gynecol* 2017. **129**, e150–e154.
7. Miller, L. *et al.* Working with what you have: How the East Africa Preterm Birth Initiative used gestational age data from facility maternity registers. *PloS One* 2020. **15**, e0237656.
8. Karl, S. *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. **10**, e0124286.
9. Kullinger, M., Granfors, M., Kieler, H. & Skalkidou, A. Discrepancy between pregnancy dating methods affects obstetric and neonatal outcomes: a population-based register cohort study. *Sci Rep* 2018. **8**, 6936.
10. WHO. Global strategy on digital health 2020-2025. *Geneva: World Health Organization* 2021. <https://apps.who.int/iris/handle/10665/344249>, (accessed March 25, 2022).
11. Nelson, G. A. & Holschuh, C. Evaluation of Telehealth Use in Prenatal Care for Patient and Provider Satisfaction: A Step Toward Reducing Barriers to Care. *J Nurse Pract* 2021. **17**, 481–484.
12. Meskó, B., Drobni, Z., Bényei, É., Gergely, B. & Gyórfy, Z. Digital health is a cultural transformation of traditional healthcare. *mHealth* 2017. **3**, 38.
13. Miller, D. D. & Brown, E. W. Artificial Intelligence in Medical Practice: The Question to the Answer? *Am J Med* 2018. **131**, 129–133.
14. Reis, Z. S. N., Vitral, G. L. N., de Souza, I. M. F., Rego, M. A. S. & Guimaraes, R. N. Newborn skin reflection: Proof of concept for a new approach for predicting gestational age at birth. A cross-sectional study. *PloS One* 2017. **12**, e0184734.
15. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015.

13, 1.

16. Venema, E. *et al.* Large-scale validation of the prediction model risk of bias assessment Tool (PROBAST) using a short form: high risk of bias models show poorer discrimination. *J Clin Epidemiol* 2021. **138**, 32–39.
17. Reis, Z. S. N. *et al.* Prematurity detection evaluating interaction between the skin of the newborn and light: protocol for the preemie-test multicentre clinical trial in Brazilian hospitals to validate a new medical device. *BMJ Open* 2019. **9**, e027442.
18. Silva, P. C., Guimarães, R. N., Souza, R. G. & Reis, Z. S. N. A quantitative cross-sectional analysis of the melanin index in the skin of preterm newborns and its association with gestational age at birth. *Skin Res Technol Off J Int Soc Bioeng Skin ISBS Int Soc Digit Imaging Skin ISDIS Int Soc Skin Imaging ISSI* 2020. **26**, 356–36.
19. Wyckoff, M. H. *et al.* Neonatal life support: 2020 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. *Circulation* 2020. **142**, n. 16_suppl_1, p. S185-S221.
20. Scott, K. *et al.* “I can guess the month ... but beyond that, I can’t tell” an exploratory qualitative study of health care provider perspectives on gestational age estimation in Rajasthan, India. *BMC Pregnancy Childbirth* 2020. **20**, 529.
21. Weiss, S. M. & Indurkha, N. Rule-based machine learning methods for functional prediction. *J Artif Intell Res* 1995. **3**, 383–403.
22. Reis, Z. *et al.* Premature or Small for Gestational Age Discrimination: International Multicenter Trial Protocol for Classification of the Low-Birth-Weight Newborn Through the Optical Properties of the Skin. *JMIR Res Protoc* 2020. **9**, e16477.
23. Ananth, C. V. & Brandt, J. S. Fetal growth and gestational age prediction by machine learning. *Lancet Digit Health* 2020. **2**, e336–e337.
24. Fung, R. *et al.* Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study. *Lancet Digit Health* 2020. **2**, e368–e375.
25. Torres, M. T., Valstar, M. F., Henry, C., Ward, C. & Sharkey, D. Small sample deep learning for newborn gestational age estimation. *12th IEEE International Conference on Automatic Face & Gesture Recognition* 2017. pp. 79–86. doi:10.1109/FG.2017.19.
26. Rittenhouse, K. J. *et al.* Improving preterm newborn identification in low-resource settings with machine learning. *PloS One* 2019. **14**, e0198919.

27. Soubeiga, D., Gauvin, L., Hatem, M. A. & Johri, M. Birth Preparedness and Complication Readiness (BPCR) interventions to reduce maternal and neonatal mortality in developing countries: systematic review and meta-analysis. *BMC Pregnancy Childbirth* 2014. **14**, 129.
28. Emeruwa, U. N., Krenitsky, N. M., & Sheen, J. J. Advances in Management for Preterm Fetuses at Risk of Delivery. *Clinics in Perinatology* 2020. **47**, 685-703.
29. Condon, J. *et al.* Expression of type 2 11 β -hydroxysteroid dehydrogenase and corticosteroid hormone receptors in early human fetal life. *J Clin Endocrinol Metab* 1998. **83**, 4490–4497.
30. August, D. & Kandasamy, Y. The effects of antenatal glucocorticoid exposure on fetal and neonatal skin maturation. *J Perinat Med* 2017. **45**.
31. Lee, A. C. *et al.* Diagnostic Accuracy of Neonatal Assessment for Gestational Age Determination: A Systematic Review. *Pediatrics* 2017. **140**, e20171423.
32. Leal, M. do C. *et al.* Saúde reprodutiva, materna, neonatal e infantil nos 30 anos do Sistema Único de Saúde (SUS). *Ciênc Saúde Coletiva* 2018. **23**, 1915–1928.
33. Ezbakhe, F., Pérez-Foguet, A. Child mortality levels and trends. *Demographic Research* 2020. **43**, p. 1263–1296.

Tables

Table 1 Baseline characteristics of the pregnancy and newborns

	N*	Statistics
Maternal data	702	..
Maternal age (years), median (IQR)	702	27 (9)
First prenatal care assessment (weeks), median (IQR)	616	12 (4)
Absent recall of last menstrual period	702	89 (12.7)
Reliable last menstrual period, n/N (%)	613	296 (42.20)
Diabetes, n/N (%)	701	103 (14.7)
Hypertensive disturbance during pregnancy, n/N (%)	702	148 (21.1)
ACMF, n/N (%)	698	273 (35.1)
Multiple gestation, n/N (%)	702	74 (10.5)
Neonatal data	781	..
Reference gestational age at birth (weeks)	781	37.3 (6.3)
Gestational age at the first ultrasound assessment (weeks), median (IQR)	781	10.1 (3.6)
Gestational age at the second ultrasound assessment (weeks), median (IQR)	781	19.4 (4.3)
ACMF exposure, n/N (%)	777	273 (35.1)
Major malformation, n/N (%)	781	8 (1.1)
1-minute Apgar score, median (IQR)	775	8 (1)
5-minute Apgar score, median (IQR)	777	9 (1)
Birthweight (g), median (IQR)	781	2740 (1498)
Sex, male, n/N (%)	781	390 (49.9)
Incubator accomodation at the skin assessment, n/N (%)	781	239 (30.6)
NICU at the skin assessment, n/N (%)	781	280 (35.9)
Jaundice at the skin assessment, n/N (%)	779	255 (32.7)
Phototherapy at the skin assessment, n/N (%)	774	32 (4.1)
Newborn mortality until 72 hours, n/N (%)	781	14 (1.8)
Respiratory distress syndrome until 72 hours, n/N (%)	781	215 (27.5)
Newborn classifications based on reference gestational age
Preterm ^a , n/N (%)	781	366 (46.9)
Moderate to late preterm ^b , n/N (%)	781	235 (30.2)
Very preterm ^c , n/N (%)	781	89 (11.4)
Extremely preterm ^d , n/N (%)	781	42 (5.4)
Small for gestational age, n/N (%)	781	115 (14.7)
Appropriate for gestational age, n/N (%)	781	607 (77.7)

Large for gestational age, n/N (%)	781	59 (7.6)
------------------------------------	-----	----------

ACMF: Antenatal Corticosteroid Therapy for Fetal Maturation. NICU: neonatal intensive care unit. IQR: interquartile range. *781 newborns from 702 pregnant women.

^aLess than 37 weeks. ^bMore than 32 to less than 37 weeks. ^cMore than 28 to less than 32 weeks. ^dLess than 28 weeks.

Table 2 Agreement between predicted gestational age with the established references

	Test (medical device)	Reference GA
ICC with reference GA (95% CI)	0.997 (0.997 to 0.998)	1
ICC with comparator-US GA (95% CI)	0.994 (0.993 to 0.995)	0.993 (0.992 to 0.994)
ICC with comparator-LMP GA (95% CI)	0.957 (0.950 to 0.963)	0.958 (0.951 to 0.964)
Day paired-difference with reference GA (95% CI)	-1.30 (-1.53 to -1.11) p < 0.001	0
Day paired-difference with comparator-US GA (95% CI)	-0.79 (-1.11 to -0.47) p < 0.001	-0.53 (-0.88 to -0.19) p = 0.002
Day paired-difference with LMP GA (95% CI)	-2.21 (-3.11 to -1.30) p < 0.001	0.83 (-0.07 to 1.74) p = 0.071
Bland-Altman 95% limits for the medical device (days)		-7.1 to 4.7
Bland-Altman 95% limits for comparator-US (days)	-8.7 to 8.4	-10.0 to 8.0
Bland-Altman 95% limits for comparator-LPM (days)	-30.0 to 23.4	-25.0 to 29.0

CI: confidence interval. ICC: intraclass correlation coefficient. GA: gestational age. p-values for the paired-t test. LMP: last menstrual period. Reference GA is the best due date. Comparator-US GA is the gestational age calculated with a second antenatal ultrasound exam after 13 weeks and 6 days of gestation and before 22 weeks. Comparator-LMP GA is the gestational age calculated with the last menstrual period. Test using the medical device is the gestational age-predicted with skin reflectance, birth weight, and ACTFM exposure predictors with the XGBoost model.

Table 3. Intent to preterm newborn discrimination according to simulated scenarios of care

	Scenario-one: Absent or unreliable LMP (n = 451, 57.7%)				Scenario-two: Reliable LMP (n = 330, 42.3%)			
	Preterm newborns	Sens [95% CI]	Spec [95% CI]	ACU [95% CI]	Preterm newborns	Sens [95% CI]	Spec [95% CI]	ACU [95% CI]
Reference GA (n=781)	199/451	167/330
Test, medical device (n=781)	199/451	199/199 (100%, [98.2% to 100%])	236/252 (93.7%, [89.9% to 96.3%])	435/451 (96.5%, [94.3% to 98.0%])	167/330	165/167 (98.8%, [95.7% to 99.9%])	158/163 (96.9%, [93.0% to 99.0%])	323/330 (97.9%, [95.7% to 99.1%])
Comparator-US GA (n=781)	199/451	190/199 (95.5%, [91.6% to 97.9%])	241/252 (95.6%, [92.3% to 97.8%])	431/451 (95.6%, [93.2% to 97.3%])	167/330	162/167 (97.0%, [93.2% to 99.0%])	158/163 (96.9%, [93.0% to 99.0%])	320/330 (97.0%, [94.5% to 98.5%])
Comparator-LMP GA (n=680)	154/451	131/199* (65.8%, [59.1% to 72.2%])	183/252* (72.6%, [66.9% to 77.9%])	314/451* (69.6%, [65.3% to 73.7%])	167/330	160/167 (95.8%, [91.6% to 98.3%])	157/163 (96.3%, [93.0% to 99.0%])	317/330 (93.4%, [94.5% to 97.9%])

*We consider the absence of comparator-LMP gestational age, when 45 newborns were preterm, 56 newborns were term.

ACU: newborn correctly classified (accuracy). LMP: last menstrual period. CI: confidence interval. GA: gestational age. Reference GA is the best due date. Comparator-US GA is the gestational age calculated with a second antenatal ultrasound exam after 13 weeks and 6 days of gestation and before 22 weeks. Comparator-LMP GA is the gestational age calculated with the last menstrual period. Test using the medical device is the gestational age-predicted with skin reflectance, birth weight, and ACTFM exposure predictors with the XGBoost model.

Figures



Figure 1

The new device and its simulated application on a newborn-doll

Source: the authors

Figure 2

Correlation plot between the skin reflectance of the newborn and the reference gestational age at birth

Figure 3

Flow diagram of participants throughout the study with results for the predictive model

Notes: The Test is the predicted gestational age with the XGBoost algorithm, including skin reflectance, birth weight, and antenatal corticosteroid therapy for fetal maturation exposure predictors (ACTFM) for gestational age. (+) positive. (-) negative. Three missing ACTFM data received imputation.

Figure 4

The distribution of estimated gestational age at birth by established methods evaluated in this study

Notes: distribution of reference GA (n=781). distribution of comparator-LMP GA (n=680). distribution of comparator-US GA (n=781). The red hatched line corresponds to the limit between preterm and term newborns. The green hatched line corresponds to the limit between term and post-term newborns.

Figure 5

Correlation between medical device gestational age and established methods of pregnancy dating and Bland-Altman plots

GA: gestational age. LMP: last menstrual period. US: ultrasound. Reference GA is the best due date. Comparator-US GA is the gestational age calculated with a second antenatal ultrasound exam after 13 weeks and 6 days of gestation and before 22 weeks. Comparator-LMP GA is the gestational age calculated with the last menstrual period. The new test is gestational age-predicted with the XGBoost algorithm, based on newborn skin reflectance values, birth weight, and ACTFM exposure information.

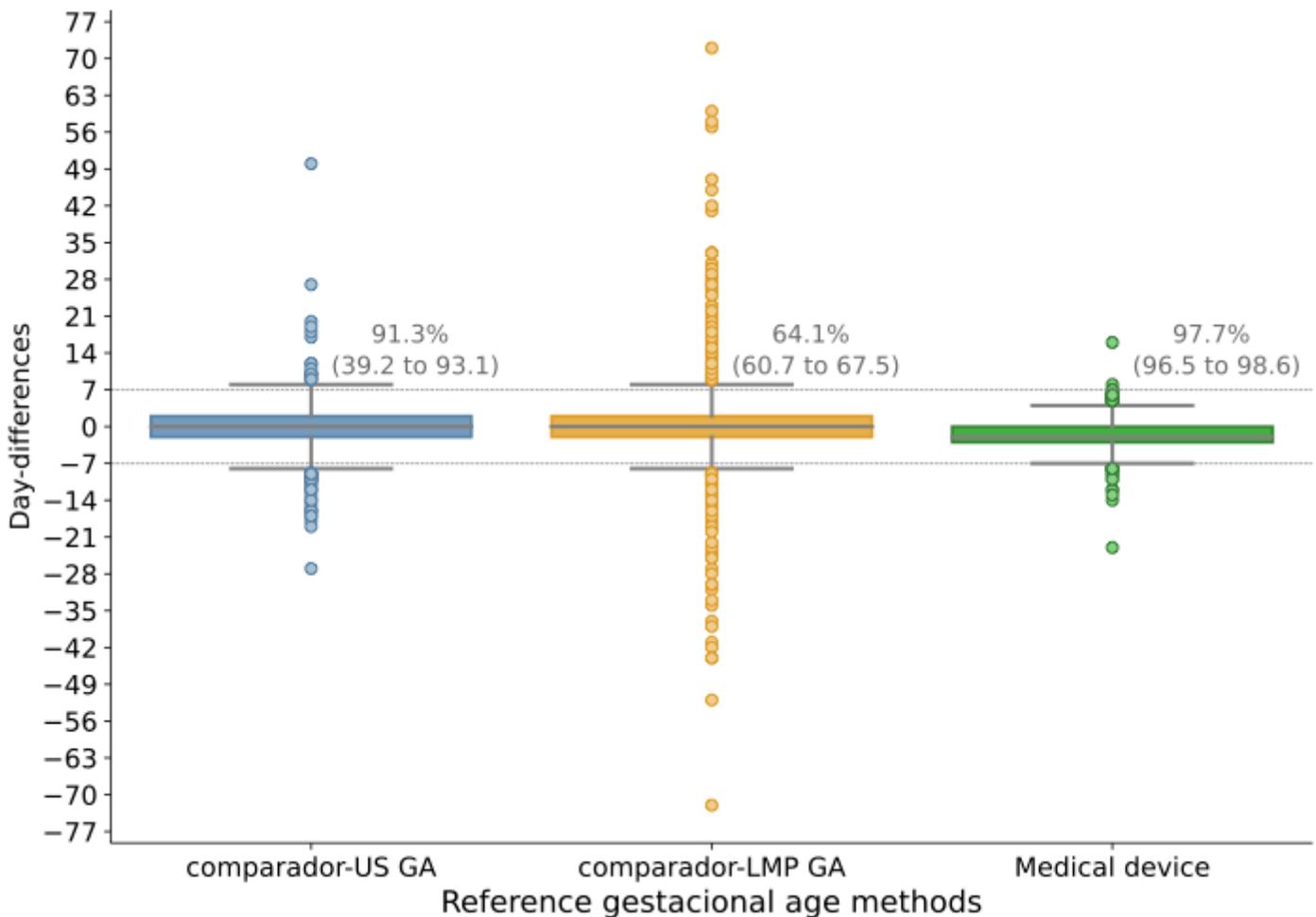


Figure 6

Box-plot of day-differences between methods and reference gestational age, with the proportion of agreement within seven days agreement

GA: gestational age. LMP: last menstrual period. US: ultrasound. The values are the proportion of agreement within seven days of agreement (95% CI). Reference GA is the best due date. Comparator-US GA is the gestational age calculated with a second antenatal ultrasound exam after 13 weeks and 6 days of gestation and before 22 weeks. Comparator-LMP GA is the gestational age calculated with the last menstrual period. The new test is gestational age-predicted with the XGBoost algorithm, based on newborn skin reflectance values, birth weight, and ACTFM exposure information. We included 101 missing data in the calculation of rate agreement for the comparator-LMP GA.

Figure 7

Receiver operating characteristic for the models to discriminate terms from preterm newborns

AUROC: the area under the receiver operating characteristic curve. CI: confidence interval. GA: gestational age. LMP: last menstrual period. US: ultrasound. Comparator-US GA is the gestational age calculated with a second antenatal ultrasound exam after 13 weeks and 6 days of gestation and before 22 weeks. Comparator-LMP GA is the gestational age calculated with the last menstrual period. The new test is gestational age-predicted with the XGBoost algorithm, based on newborn skin reflectance values, birth weight, and ACTFM exposure information.

A - Preterm discrimination at 37 weeks of gestation

AUROC: New medical device (test): 0.998 (95% CI: 0.997, 0.999). Comparators: US GA, 0.993 (95% CI: 0.988, 0.997); LMP GA, 0.957 (95% CI: 0.941, 0.974).

B - Preterm discrimination at 32 weeks of gestation

AUROC: New medical device (test): 0.999 (95% CI: 0.997, 1.000). Comparators: US GA, 0.996 (95% CI: 0.993, 0.999); LMP GA, 0.977 (95% CI: 0.962, 0.992).

C - Preterm discrimination at 28 weeks of gestation

AUROC: New medical device (test): 1.000 (95% CI: 0.999, 1.000). Comparators: US GA, 0.997 (95% CI: 0.994, 1.000); LMP GA, 0.981 (95% CI: 0.964, 0.998).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialUFMG.docx](#)