PLOS ONE

RESEARCH ARTICLE

Improving preterm newborn identification in low-resource settings with machine learning

Katelyn J. Rittenhouse[1,2]*, Bellington Vwalika[3], Alexander Keil[1], Jennifer Winston[1], Marie Stoner[1], Joan T. Price[1,2], Monica Kapasa[3], Mulaya Mubambe[3], Vanilla Banda[3], Whyson Muunga[3], Jeffrey S. A. Stringer[1]

1 University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States, **2** University of North Carolina Global Projects Zambia, Lusaka, Zambia, **3** University of Zambia School of Medicine, Lusaka, Zambia

* katelyn_rittenhouse@med.unc.edu

Check for updates

OPEN ACCESS

Citation: Rittenhouse KJ, Vwalika B, Keil A, Winston J, Stoner M, Price JT, et al. (2019) Improving preterm newborn identification in low-resource settings with machine learning. PLoS ONE 14(2): e0198919. https://doi.org/10.1371/journal.pone.0198919

Editor: Chelsea Dobbins, University of Queensland, AUSTRALIA

Received: May 24, 2018

Accepted: February 12, 2019

Published: February 27, 2019

Copyright: © 2019 Rittenhouse et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data have been de-identified and uploaded to the Open Science Framework, at DOI: 10.17605/OSF.IO/RZG2E.

Funding: This work was supported by a Bill and Melinda Gates Foundation grant to the Global Alliance to Prevent Prematurity and Stillbirth (Seattle Children's Hospital/GAPPS 13008/OPP1033514). Additional support was provided by the US National Institutes of Health through the UNC Center for AIDS Research (P30 AI50410) and

Abstract

Background

Globally, preterm birth is the leading cause of neonatal death with estimated prevalence and associated mortality highest in low- and middle-income countries (LMICs). Accurate identification of preterm infants is important at the individual level for appropriate clinical intervention as well as at the population level for informed policy decisions and resource allocation. As early prenatal ultrasound is commonly not available in these settings, gestational age (GA) is often estimated using newborn assessment at birth. This approach assumes last menstrual period to be unreliable and birthweight to be unable to distinguish preterm infants from those that are small for gestational age (SGA). We sought to leverage machine learning algorithms incorporating maternal factors associated with SGA to improve accuracy of preterm newborn identification in LMIC settings.

Methods and findings

This study uses data from an ongoing obstetrical cohort in Lusaka, Zambia that uses early pregnancy ultrasound to estimate GA. Our intent was to identify the best set of parameters commonly available at delivery to correctly categorize births as either preterm (<37 weeks) or term, compared to GA assigned by early ultrasound as the gold standard. Trained midwives conducted a newborn assessment (<72 hours) and collected maternal and neonatal data at the time of delivery or shortly thereafter. New Ballard Score (NBS), last menstrual period (LMP), and birth weight were used individually to assign GA at delivery and categorize each birth as either preterm or term. Additionally, machine learning techniques incorporated combinations of these measures with several maternal and newborn characteristics associated with prematurity and SGA to develop GA at delivery and preterm birth prediction models. The distribution and accuracy of all models were compared to early ultrasound dating. Within our live-born cohort to date (n = 862), the median GA at delivery by early ultrasound was 39.4 weeks (IQR: 38.3–40.3). Among assessed newborns with complete data included in this analysis (n = 468), the median GA by ultrasound was 39.6 weeks (IQR: 38.4–40.3). Using machine learning, we identified a combination of six accessible

trainee / mentor support: T32 HD075731 (JTP), K01 TW010857 (JTP), and D43 TW009340 (KR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

parameters (LMP, birth weight, twin delivery, maternal height, hypertension in labor, and HIV serostatus) that can be used by machine learning to outperform current GA prediction methods. For preterm birth prediction, this combination of covariates correctly classified >94% of newborns and achieved an area under the curve (AUC) of 0.9796.

Conclusions

We identified a parsimonious list of variables that can be used by machine learning approaches to improve accuracy of preterm newborn identification. Our best-performing model included LMP, birth weight, twin delivery, HIV serostatus, and maternal factors associated with SGA. These variables are all easily collected at delivery, reducing the skill and time required by the frontline health worker to assess GA.

Trial registration

ClinicalTrials.gov Identifier: NCT02738892

Introduction

Preterm birth affects more than one in ten live births worldwide.[1] It is the single largest cause of neonatal death and the second leading cause of death in children under the age of 5 years.[2] Many babies who survive a preterm birth face life-long morbidity, including cognitive disability, poor motors skills, behavioral problems, hearing loss, chronic lung disease, and decreased economic productivity.[3–5] The greatest burden of preterm birth falls on low-and middle-income countries (LMICs), where more than 90% of the global 15 million preterm deliveries occur each year[6] and where preterm infants carry a 7-fold higher risk of neonatal mortality and a 2.5-fold higher risk of post-neonatal mortality compared to their full-term counterparts.[7] In these settings, preterm infants often go unrecognized due to inaccurate estimation of gestational age (GA). On the individual level, this can result in missed opportunities for clinical intervention; on the population level, this can limit the ability to monitor preterm birth rates and make informed decisions around policy and resource allocation.

Early prenatal ultrasound, widely regarded as the gold standard for GA dating, is unavailable in many LMIC settings. In its absence, providers must rely on other methods, such as last menstrual period (LMP), newborn assessment, or birthweight to classify infant GA at delivery. Each of these approaches has limitations. Reported LMP is subject to patient recall and can be very unreliable in settings where women present late for care.[8–11] Newborn assessment, including the commonly used New Ballard Score (NBS),[12] suffers from poor inter-rater reliability[13, 14] and tends to overestimate GA, particularly in LMICs[15] and settings with high rates of small-for-gestational age (SGA).[16–21] Finally, birthweight, while an easily obtained and reliable indicator, does not distinguish between an infant that is preterm and one that is SGA.

We sought to develop a machine learning algorithm that can estimate GA at birth from readily obtained indicators in a setting where early ultrasound is not available. We were particularly interested in the simple, binary classification of preterm (i.e., <37 weeks) versus term. We hypothesized that a model combining LMP, individual elements of the NBS, birthweight, and key pregnancy risk factors associated with SGA, would outperform any individual approach.

Methods

This study was conducted using data from the Zambian Preterm Birth Prevention Study (ZAPPS; ClinicalTrials.gov identifier: NCT02738892), an ongoing prospective obstetrical cohort at the Women and Newborn Hospital of the University Teaching Hospital (UTH) in Lusaka, Zambia. The rationale for our study, its procedures, and cohort characteristics have been described elsewhere.[22] Briefly, women are enrolled in early pregnancy and followed through delivery and the postpartum period. Written informed consent is obtained from all participants prior to study enrollment for collection of maternal and newborn data. GA is established by ultrasound (Sonosite M-Turbo; Fuji Sonosite, Inc, Bothell, WA) at study screening using the fetal crown rump length (if <14 weeks gestation)[23] or head circumference and femur length (if ≥14 weeks).[24] All fetal biometry measurements are measured twice and then averaged for gestational age calculations.

The study employs midwives who attend to participants admitted to the labor ward or postpartum unit at UTH. Their duties include ensuring that relevant clinical information is captured in the study record, that babies are weighed at birth or shortly thereafter, and that the NBS is performed within 72 hours of delivery.

Newborns were included in this analysis if they were live-born and they had a complete set of characteristics and metrics assessed in this study. We defined preterm birth as birth prior to 37 weeks of gestation and SGA as a birthweight less than the 10^{th} percentile for its corresponding GA.[25] The NBS sums assessments of 5 domains of neuromuscular maturity and 7 domains of physical maturity into a composite score that is used to assign GA at delivery.[12] We evaluated both the composite score and its 12 individual components in this study.

In our analyses, we assessed eight models: three single parameter GA dating methods and five multiple parameter novel machine learning GA dating models (Table 1). We were primarily interested in classifying preterm birth as a binary outcome (i.e., <37 weeks or not) to identify newborns at highest risk of complications from preterm delivery, but we also wished to

Table 1. Components of gestational age dating models.

GA Dating Model	GA at delivery (composite NBS*)	GA at delivery (LMP)	GA at delivery (birth weight 50% ile)^	NBS (individual components)†	Birth weight	Maternal height	HTN in labor	Maternal HIV infection	Twin gestation
NBS	•								
LMP		•							
Birth weight			•						
Optimized NBS§				•					
NBS(-)LMP (-)§					•	•	•	•	•
NBS(-)LMP (+)§		•			•	•	•	•	•
NBS(+)LMP (-)§				•	•	•	•	•	•
NBS(+)LMP (+)§		•		•	•	•	•	•	•

GA: gestational age; NBS: New Ballard Score; LMP: last menstrual period; HTN: hypertension

*Composite NBS: Sum of neuromuscular and physical maturity domains

^Birth weight 50%ile: Intergrowth 50th birthweight-for-age centiles used to convert birthweights to GA

†Individual NBS components: 5 Neuromuscular maturity domains and 7 physical maturity domains

§Machine learning models

https://doi.org/10.1371/journal.pone.0198919.t001

assess how the models might estimate GA as a continuous outcome. We restricted our models to maternal and newborn characteristics that are accessible to health workers in resource-limited settings at the time of delivery, either through direct assessment or review of the medical record.

The single parameter GA dating methods assessed include 1) LMP, 2) NBS, and 3) birth weight. GA dating by NBS was calculated from the composite NBS using the formula for GA conversion, as described by Ballard et al.[12] GA dating by birth weight was calculated under the naïve assumption that all infants are born at the 50th birthweight-for-age centile and used INTERGROWTH standards[26] to convert these birthweights to GA.

The multiple parameter machine learning models assessed include 1) Optimized NBS, 2) NBS(-)LMP(-), 3) NBS(-)LMP(+), 4) NBS(+)LMP(-), and 5) NBS(+)LMP(+). In all machine learning models incorporating NBS, including Optimized NBS, all 12 individual NBS components were included. With the exclusion of Optimized NBS, all machine learning models included an additional five maternal and newborn parameters with various combinations of NBS and LMP. Maternal and newborn parameters were identified by stepwise regression and included birth weight in addition to parameters with an *a priori* association with preterm birth (twin delivery, maternal HIV serostatus) and SGA (maternal height, maternal hypertension). We used hypertension in labor as a surrogate marker for maternal hypertension because, although imperfect, it is a readily accessible metric at delivery in the maternal delivery case file. Hypertension in labor was defined as systolic blood pressure $\geq$140 and/or diastolic blood pressure $\geq$90 recorded in the maternal delivery case file. For twin deliveries, we included only baby A (the first baby to be delivered) in our dataset. This approach to twin deliveries reduced bias toward artificially increased model accuracy by including two newborns with closely matched characteristics. HIV serostatus was determined by rapid ELISA performed according to local protocol at first antenatal care visit.[27]

We used super learner[28] to generate five GA and prematurity prediction models (components described above). In brief, super learner is a machine learning approach for combining the strengths of multiple predictive models or learners. Super learner finds the weighted, convex combination of these algorithms that minimizes the cross-validated mean squared error of predictions of GA and preterm birth. To reduce concerns about over-fitting the data, we utilized k-fold cross validation (with 10 folds) to select the combination of learners. K-fold cross validation ensures that the learner is not fit (trained) to the same data that are used to make predictions and judge performance. We used super learner computational macro (arXiv:1805.08058 [stat.ML]) developed in SAS version 9.4 (Cary, North Carolina) along with the SAS procedures HPFOREST,[29] GENMOD,[30] and GAM?[31] to perform random forest algorithms, generalized linear and logistic modeling, and generalized additive modeling, respectively. For each algorithm included in our Super Learner library, the hyper-parameters were the defaults given by the SAS macro,[32] which were based off defaults from the R super learner package.[28] For our continuous GA prediction modeling of super learner models, we combined linear regression, random forest regression, and generalized additive models. For binary preterm birth classification modeling, we combined logistic regression, random forest classification, and generalized additive model.

Kernel density plots and Pearson's correlation coefficients were generated to compare the predicted GAs from each continuous outcome model to GAs by early ultrasound. For our primary analysis, we based accuracy of each predictive model on the model fit to the data in which we had complete data (n = 468). In a subset of births without NBS that were not used to train predictive models, we subsequently estimated predictive accuracy. We note that, while super learner utilizes cross-validation to reduce over-fit, the super learner predictions do not, themselves, estimate cross-validated accuracy; thus, this latter step is necessary to produce fair

estimates of the out-of-sample accuracy of our approach. Receiver operating curves (ROCs) were generated and area under the curve (AUC) calculated for the diagnostic accuracy of preterm birth for each binary classification model. We also calculated the positive predictive value, negative predictive value, and percent correct classification for the identification of preterm infants using the best cutoff point for each model, as determined using the Youden method.[33] The Youden method determines a cutoff point by optimizing the differentiating ability of a test or model when equal weight is given to sensitivity and specificity. A subsequent sensitivity analysis including women enrolled in the first trimester (<14 weeks gestation) was conducted on our best-performing model to further assess stability and validity in women with the most accurate gestational age dating. All super learner modeling was performed in SAS as described above; all other analyses were performed using STATA release 14 (College Station, TX). This study was approved by the University of Zambia Biomedical Research Ethics Committee and the University of North Carolina Institutional Review Board.

Results

Between August 2015 and September 2017, 1450 pregnant women were consented and enrolled into the ZAPPS cohort. To date, 862 (59.4%) participants have had live births with deliveries captured by a study midwife. A total of 468 (53.1%) of these live births had newborns assessed at <72 hours of life by a trained nurse midwife and had complete data available to be included in subsequent preterm birth predictive modeling (Table 2). Among assessed live births, median ultrasound-based GA was 39.6 weeks (IQR: 38.4–40.3), with preterm birth prevalence 6.8%. The median birth weight was 3100g (IQR: 2855–3400). The prevalence of SGA in this population was 14.1%. NBS assessment was the most common missing parameter, causing study exclusion (n = 300; 76.1% of live births not assessed).

Kernel density plots, with accompanying Pearson correlation coefficients, comparing the continuous GA distributions of the 8 models evaluated in this study to those calculated by ultrasound are shown in Fig 1. The models generated by the super learner program for GA as a continuous outcome clustered estimated GAs around the mean, resulting in a loss of outliers and less accurate estimation of GA as a continuous outcome (Fig 1D–1H). Despite clustering

Table 2. Maternal and newborn parameters included in gestational age modeling.

Parameter		All Live Births *Median (IQR) or n (%)*	Live Births Assessed *Median (IQR) or n (%)*	Live Births Not Assessed *Median (IQR) or n (%)*	p-value*
	n (%)	862	468 (54.3)	394 (45.7)	-
	GA at delivery (by ultrasound)	39.4 (38.3–40.3)	39.6 (38.4–40.3)	39.4 (38.1–40.1)	0.055
	Preterm Birth (<37 weeks)	95 (11.0)	32 (6.8)	63 (16.0)	<0.001
	Small for Gestational Age (<10%ile)	129 (15.0)	66 (14.1)	63 (16.0)	0.439
Maternal	Height (m)	1.60 (1.56–1.64)	1.60 (1.56–1.64)	1.60 (1.56–1.64)	0.916
	Hypertension labor^	171 (22.9)	69 (24.8)	97 (21.7)	0.476
	GA at delivery (LMP)	39.1 (37.3–40.1)	39.3 (37.4–40.4)	38.9 (37.1–40.0)	0.019
	HIV-infected	217 (25.2)	123 (26.3)	94 (23.9)	0.426
Newborn	Twin delivery	25 (3.0)	10 (2.1)	15 (4.1)	0.102
	Birth weight (g)	3063 (2800–3390)	3100 (2855–3400)	3000 (2700–3300)	0.015
	GA at delivery (composite NBS†)	39.2 (37.6–40.8)	39.2 (38.0–40.8)	39.2 (37.6–41.2)	0.762

*p-values calculated by Mann-Whitney test for continuous variables or chi-square test for dichotomous categorical variables

^Hypertension in labor was defined as systolic blood pressure ≥140 and/or diastolic blood pressure ≥90 recorded in the maternal delivery case file

https://doi.org/10.1371/journal.pone.0198919.t002

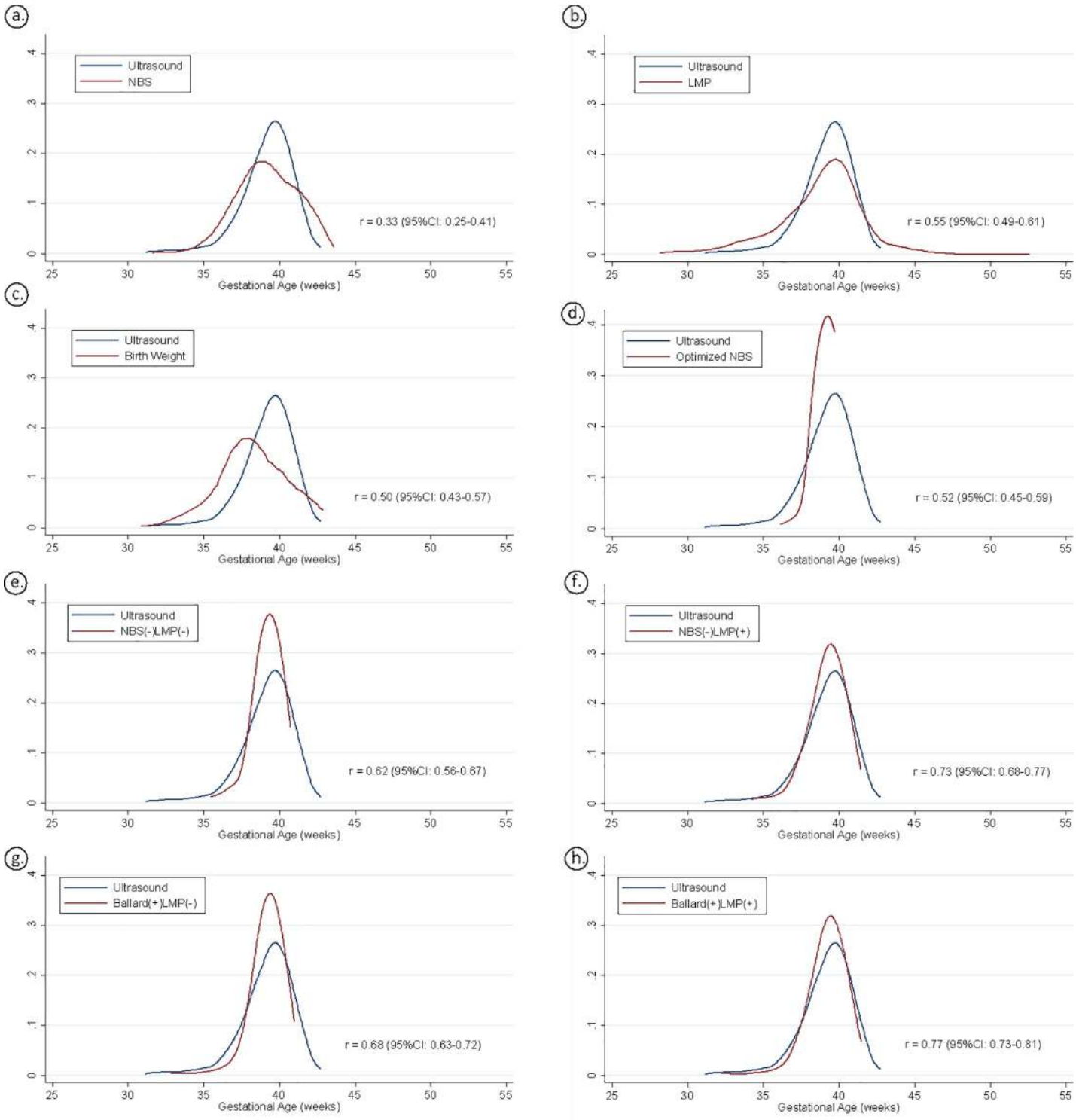

Fig 1. Distribution of gestational age at birth by all continuous models. r = Pearson's correlation coefficient.

https://doi.org/10.1371/journal.pone.0198919.g001

around the mean, the NBS(-)LMP(+) and NBS(+)LMP(+) machine learning models (Fig 1F and 1H) were found to best approximate the distribution of GA at delivery as compared to ultrasound dating (Pearson correlation coefficients 0.73 and 0.77, respectively).

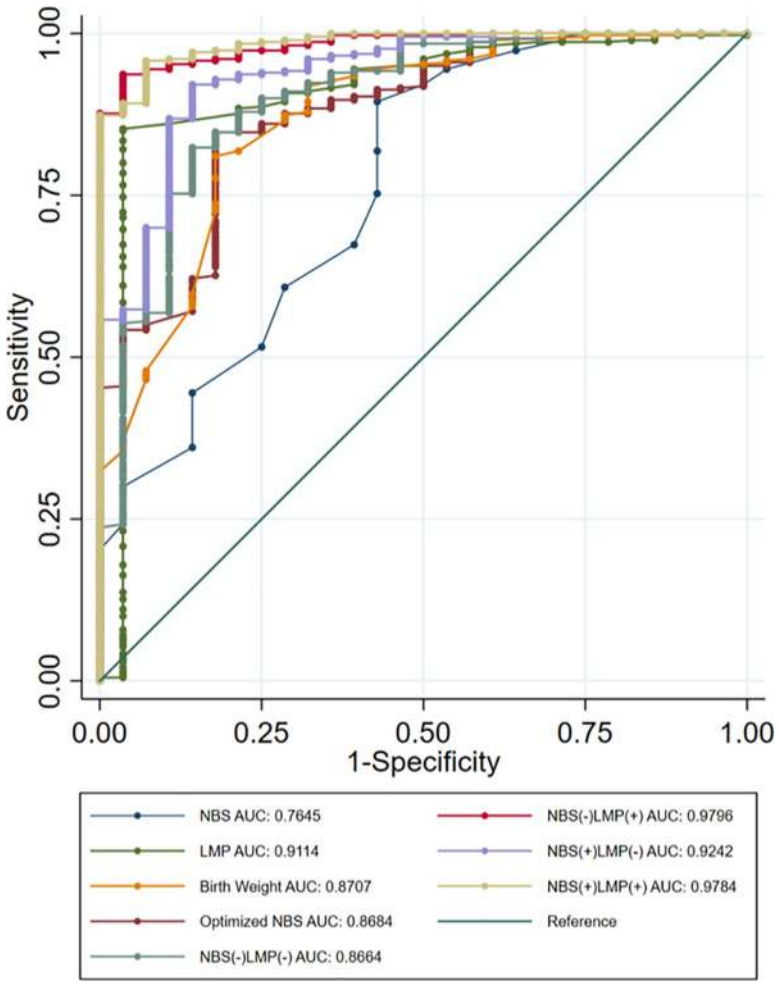

Fig 2. Diagnostic accuracy of binary models to identify preterm newborns. AUC: Area Under Curve.

https://doi.org/10.1371/journal.pone.0198919.g002

The accuracy of preterm birth classification by each GA dating method was assessed using ROCs and associated AUCs (Fig 2). The AUC for Optimized NBS using super learner (0.8684) was improved compared to the single parameter NBS model (0.7645). The NBS(-)LMP(-) super learner model incorporating maternal and newborn parameters without LMP and NBS had an AUC of 0.8664, outperforming the NBS model and performing similarly to Optimized NBS. Adding LMP to this model, NBS(-)LMP(+), improved the AUC (0.9796) more than adding NBS or both NBS and LMP (0.9242 and 0.9784, respectively).

Positive predictive value (PPV), negative predictive value (NPV), and correct classification of all models predicting prematurity are shown in Table 3. In concordance with our ROC analysis, the NBS(-)LMP(+) super learner model incorporating LMP without NBS had the highest percent correct classification (94.0%). Additionally, this model had a NPV of 98.9%, similar to other models, and a PPV of 53.6%, substantially out-performing all other GA dating methods.

In a sensitivity analysis, we tested NBS(-)LMP(+), our best-performing model, on the subset of women who were enrolled in the first trimester (<14 weeks; n = 204), as their ultrasound dating is expected to be most accurate. We found a PPV of 73.3%, NPV of 98.8%, correct classification of 94.6%, and AUC of 0.9679. Additionally, because our best-performing model excluded NBS–the variable most likely to cause study exclusion due to missingness–we were

Table 3. Prediction of preterm birth by binary models.

GA Dating Method	Positive Predictive Value	Negative Predictive Value	Correct Classification
NBS	28.6%	96.3%	88.0%
LMP	32.3%	99.7%	85.9%
Birth Weight	24.0%	98.4%	82.1%
Optimized NBS	28.9%	98.4%	85.0%
NBS(-)LMP(-)	26.8%	98.4%	83.6%
NBS(-)LMP(+)	53.6%	99.5%	94.0%
NBS(+)LMP(-)	32.6%	98.9%	86.8%
NBS(+)LMP(+)	37.3%	99.7%	88.7%

Best cutoff point for calculations by each GA Dating Method determined using the Youden method

https://doi.org/10.1371/journal.pone.0198919.t003

able to test our model in this population. In 245 newborns not included in our initial analysis, we found a PPV of 71.9%, NPV of 95.7%, correct classification of 90.2%, and AUC of 0.8776.

Discussion

In our urban Zambian cohort with early pregnancy ultrasound dating, we used machine learning to identify a parsimonious set of maternal and newborn variables associated with prematurity and small for gestational age that can improve discrimination between preterm and term newborns as compared to common gestational age dating methods (New Ballard Score, last menstrual period, and birth weight). This exploratory study demonstrates the promising utility of machine learning techniques to optimize algorithms for the identification of preterm birth and other adverse birth outcomes in low-resource settings.

Although a positive correlation between the number of parameters and accuracy of GA assessment has been established,[15] increasing parameter collection has negative feasibility of use, particularly in LMIC settings. In sub-Saharan Africa, up to one-half of all deliveries occur outside of the hospital and have no skilled birth attendant,[34, 35] limiting the utility of GA dating methods requiring numerous maternal and newborn metrics and characteristics. A significant strength of our best-performing model is that it incorporates only six maternal and newborn characteristics and metrics available at delivery: LMP, birth weight, twin gestation, maternal HIV serostatus, hypertension at delivery, and maternal height.

An interesting finding of our analysis was the strength of LMP as a predictor of GA and preterm birth. Limitations of LMP as a dating method are well-documented.[9] Women with lower educational attainment[9, 36] and later presentation to care[8, 37] tend to have less accurate recall of LMP, and the measure is subject to number preference (e.g. rounding to zero or five, or preference for 1st or 10th of month) and recall bias.[9, 10] Consequently, GA estimates by LMP alone suffer imprecision, with some estimates differing by weeks when compared to ultrasound.[38–41] Indeed, data from the Zambia Perinatal Record System, an electronic system that captured more than 250,000 births over a 6 year period in Lusaka, suggests an impossibly high preterm birth rate of 35% when LMP is used to determine GA.[11, 42, 43] Despite these limitations, we demonstrate that LMP is a useful predictor of prematurity and GA at delivery when incorporated into a model that allows obviously implausible estimates to be overridden by other parameters.

Further, our best-performing prematurity prediction model excluded NBS. In fact, the addition of NBS components to our best-performing list of covariates decreased the PPV, percent correct classification, and AUC of the model. LMP outperformed NBS in prematurity prediction, both when assessed individually and in combination with other parameters. This

finding supports a recent systematic review indicating that NBS has lower agreement with ultrasound dating than LMP.[15] The exclusion of NBS from our best-performing model has the benefit of omitting lengthy and technical neonatal assessment procedures. Despite only including one newborn measurement, our model achieves an excellent AUC and correctly classifies more than 94% of newborns as preterm and term in our well-dated Zambian cohort. Additionally, when excluding NBS, we were able to apply our model to a cohort of an additional 245 births with a 20% preterm birth rate, demonstrating the broader applicability of methods excluding NBS. Implementation of this model using six accessible maternal and newborn characteristics may increase the accuracy and rapidity of preterm newborn identification in LMIC settings as well as decrease the time and level of training required by frontline health workers to assess preterm birth.

The calculated PPVs for all models demonstrate the limitations of our currently utilized, single parameter methods to correctly identify preterm newborns. Only 32.3% and 28.6% of newborns predicted to be preterm by LMP and NBS, respectively, were also classified as preterm by early pregnancy ultrasound dating. Many of our multiple parameter, machine learning models performed similarly, with comparable PPVs. Only our best-performing model, NBS(-) LMP(+), had a PPV greater than 50%. All GA dating models demonstrated a propensity for overestimating preterm birth rates and misclassifying term newborns as preterm; however, our best-performing model had a decreased tendency to misidentify preterm newborns in this way. Additionally, this model correctly classified 94% of newborns, including 30 out of 32 preterm newborns. Further, in both our sensitivity analysis of newborns with ultrasound dating $<$14 weeks gestation (the most accurate GA dating) and in our external validation cohort of newborns without NBS, our best-performing model had a PPV $>$70%. Although far from perfect as a preterm newborn algorithm, our parsimonious model significantly improves upon currently available identification methods, allowing clinicians to better direct care and resources to newborns in need, especially in low-resource, LMIC settings.

A significant limitation of this current study is survival bias of assessed newborns, as demonstrated by the significant differences between our assessed and not assessed populations (Table 2). Preterm, especially early preterm, newborns were sometimes not assessed by study midwives because they were deemed too ill for the assessment or because of parental or neonatal provider objection to the exam. These early preterm newborns would likely have been identified as preterm by models included in this analysis. Thus, our estimates likely underestimate the performance of preterm birth identification in all models. Even with continuous staffing of the labor ward by midwives trained on NBS performance, many newborns were not evaluated within 72 hours. As many early preterm and critically ill newborns are never assessed, newborn assessments may not be the most effective measure of GA for these babies. In our cohort, we would be able to assess significantly more newborns in our model (81% vs. 54%) if we included newborns on whom NBS was not collected, indicating that GA dating methods excluding newborn assessment may be more efficacious in LMIC settings.

A further limitation of our current model is that it was developed to optimize the accuracy the average GA for a given set of covariates, which was then used to infer term ($\geq$37 weeks) versus preterm ($<$37 weeks) births. This model results in limited accuracy in the prediction of the complete distribution of GA. Utilizing super learner capabilities to better model the distribution of GA, rather than just the mean, as a continuous outcome may be a helpful next step for neonatal providers desiring to better estimate accurate GA. Additionally, our current model requires all characteristics and measurements to assess preterm birth status be present for study inclusion. Consequently, if a woman does not know her LMP, her newborn is omitted from this model. Future work to assess novel GA and preterm newborn prediction models using machine learning techniques should include methods to impute missing data. Further,

as validation of our model was limited to internal k-fold cross validation, preventing over-fitting the data, in addition to a small model validation cohort of newborns missing NBS, external validation in a larger dataset should be pursued in future work.

Conclusion

In summary, by leveraging the capacity of cutting-edge machine learning algorithms and maternal parameters associated with prematurity and SGA newborns, we identified a parsimonious list of covariates that improves accuracy of preterm newborn identification. Our model incorporates six accessible maternal and newborn characteristics and metrics, reducing the skill and time required to assess gestational age. This exploratory study supports the need for further research into the use of machine learning techniques to improve the accuracy of gestational age assessment in low resource settings and to assist frontline health workers in identifying newborns who may require special care.

Author Contributions

Conceptualization: Katelyn J. Rittenhouse, Bellington Vwalika, Alexander Keil, Jennifer Winston, Joan T. Price, Monica Kapasa, Mulaya Mubambe, Vanilla Banda, Whyson Muunga, Jeffrey S. A. Stringer.

Data curation: Katelyn J. Rittenhouse, Jennifer Winston, Jeffrey S. A. Stringer.

Formal analysis: Katelyn J. Rittenhouse, Alexander Keil, Jennifer Winston, Marie Stoner, Joan T. Price, Jeffrey S. A. Stringer.

Funding acquisition: Jeffrey S. A. Stringer.

Investigation: Katelyn J. Rittenhouse, Monica Kapasa, Mulaya Mubambe, Vanilla Banda, Whyson Muunga.

Methodology: Katelyn J. Rittenhouse, Bellington Vwalika, Alexander Keil, Jennifer Winston, Marie Stoner, Jeffrey S. A. Stringer.

Project administration: Katelyn J. Rittenhouse.

Resources: Alexander Keil, Jennifer Winston.

Software: Alexander Keil.

Supervision: Bellington Vwalika, Alexander Keil, Jennifer Winston, Marie Stoner, Joan T. Price, Jeffrey S. A. Stringer.

Validation: Katelyn J. Rittenhouse.

Writing – original draft: Katelyn J. Rittenhouse.

Writing – review & editing: Katelyn J. Rittenhouse, Bellington Vwalika, Alexander Keil, Jennifer Winston, Marie Stoner, Joan T. Price, Monica Kapasa, Mulaya Mubambe, Vanilla Banda, Whyson Muunga, Jeffrey S. A. Stringer.

References

1. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, Narwal R, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet. 2012; 379(9832):2162–72. https://doi.org/10.1016/S0140-6736(12)60820-4 PMID: 22682464

2. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet. 2015; 385(9966):430–40. https://doi.org/10.1016/S0140-6736(14)61698-6 PMID: 25280870
3. Wang ML, Dorer DJ, Fleming MP, Catlin EA. Clinical outcomes of near-term infants. Pediatrics. 2004; 114(2):372–6. PMID: 15286219
4. Woythaler MA, McCormick MC, Smith VC. Late preterm infants have worse 24-month neurodevelopmental outcomes than term infants. Pediatrics. 2011; 127(3):e622–9. https://doi.org/10.1542/peds.2009-3598 PMID: 21321024
5. Mwaniki MK, Atieno M, Lawn JE, Newton CR. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. Lancet. 2012; 379(9814):445–52. https://doi.org/10.1016/S0140-6736(11)61577-8 PMID: 22244654
6. Lawn JE, Davidge R, Paul VK, von Xylander S, de Graft Johnson J, Costello A, et al. Born too soon: care for the preterm baby. Reprod Health. 2013; 10 Suppl 1:S5.
7. Katz J, Lee AC, Kozuki N, Lawn JE, Cousens S, Blencowe H, et al. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. Lancet. 2013; 382(9890):417–25. https://doi.org/10.1016/S0140-6736(13)60993-9 PMID: 23746775
8. Kramer MS, McLean FH, Boyd ME, Usher RH. The validity of gestational age estimation by menstrual dating in term, preterm, and postterm gestations. JAMA. 1988; 260(22):3306–8. PMID: 3054193
9. Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. Paediatr Perinat Epidemiol. 2007; 21 Suppl 2:86–96.
10. Savitz DA, Terry JW Jr., Dole N, Thorp JM Jr., Siega-Riz AM, Herring AH. Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. American journal of obstetrics and gynecology. 2002; 187(6):1660–6. PMID: 12501080
11. Chi BH, Vwalika B, Killam WP, Wamalume C, Giganti MJ, Mbewe R, et al. Implementation of the Zambia electronic perinatal record system for comprehensive prenatal and delivery care. Int J Gynaecol Obstet. 2011; 113(2):131–6. https://doi.org/10.1016/j.ijgo.2010.11.013 PMID: 21315347
12. Ballard JL, Khoury JC, Wedig K, Wang L, Eilers-Walsman BL, Lipp R. New Ballard Score, expanded to include extremely premature infants. J Pediatr. 1991; 119(3):417–23. PMID: 1880657
13. 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. Pediatrics. 2016; 138(1).
14. Taylor RA, Denison FC, Beyai S, Owens S. The external Ballard examination does not accurately assess the gestational age of infants born at home in a rural community of The Gambia. Ann Trop Paediatr. 2010; 30(3):197–204. https://doi.org/10.1179/146532810X12786388978526 PMID: 20828452
15. 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; 140(6).
16. Roberfroid D, Huybregts L, Lanou H, Henry MC, Meda N, Menten J, et al. Effects of maternal multiple micronutrient supplementation on fetal growth: a double-blind randomized controlled trial in rural Burkina Faso. Am J Clin Nutr. 2008; 88(5):1330–40. https://doi.org/10.3945/ajcn.2008.26296 PMID: 18996870
17. Huybregts L, Roberfroid D, Lanou H, Menten J, Meda N, Van Camp J, et al. Prenatal food supplementation fortified with multiple micronutrients increases birth length: a randomized controlled trial in rural Burkina Faso. Am J Clin Nutr. 2009; 90(6):1593–600. https://doi.org/10.3945/ajcn.2009.28253 PMID: 19812173
18. Schmiegelow C, Minja D, Oesterholt M, Pehrson C, Suhrs HE, Bostrom S, et al. Factors associated with and causes of perinatal mortality in northeastern Tanzania. Acta Obstet Gynecol Scand. 2012; 91(9):1061–8. https://doi.org/10.1111/j.1600-0412.2012.01478.x PMID: 22676243
19. Malaba LC, Iliff PJ, Nathoo KJ, Marinda E, Moulton LH, Zijenah LS, et al. Effect of postpartum maternal or neonatal vitamin A supplementation on infant mortality among infants born to HIV-negative mothers in Zimbabwe. Am J Clin Nutr. 2005; 81(2):454–60. https://doi.org/10.1093/ajcn.81.2.454 PMID: 15699235
20. Baumann C, Huppi P, Amato M. [Prenatal and postnatal determination of gestational age of small newborn infants]. Z Geburtshilfe Perinatol. 1993; 197(3):135–40. PMID: 8396289
21. Constantine NA, Kraemer HC, Kendall-Tackett KA, Bennett FC, Tyson JE, Gross RT. Use of physical and neurologic observations in assessment of gestational age in low birth weight infants. J Pediatr. 1987; 110(6):921–8. PMID: 3585608
22. Castillo MC FN, Rittenhouse K, Price JT, Freeman BL, Mwape H, Winston J, Sindano N, Baruch-Gravett C, Chi BH, Kasaro MP, Litch JA, Stringer JS, Vwalika B. The Zambian Preterm Birth Prevention Study (ZAPPS): Cohort characteristics at enrollment. Gates Open Res. 2018; 2(25).

23. Papageorghiou AT, Kennedy SH, Salomon LJ, Ohuma EO, Cheikh Ismail L, Barros FC, et al. International standards for early fetal size and pregnancy dating based on ultrasound measurement of crown-rump length in the first trimester of pregnancy. Ultrasound Obstet Gynecol. 2014; 44(6):641–8. https://doi.org/10.1002/uog.13448 PMID: 25044000

24. Papageorghiou AT, Kemp B, Stones W, Ohuma EO, Kennedy SH, Purwar M, et al. Ultrasound-based gestational-age estimation in late pregnancy. Ultrasound Obstet Gynecol. 2016; 48(6):719–26. https://doi.org/10.1002/uog.15894 PMID: 26924421

25. Villar J, Cheikh Ismail L, Staines Urias E, Giuliani F, Ohuma EO, Victora CG, et al. The satisfactory growth and development at 2 years of age of the INTERGROWTH-21(st) Fetal Growth Standards cohort support its appropriateness for constructing international standards. American journal of obstetrics and gynecology. 2018; 218(2S):S841–S54 e2. https://doi.org/10.1016/j.ajog.2017.11.564 PMID: 29273309

26. Stirnemann J, Villar J, Salomon LJ, Ohuma E, Ruyan P, Altman DG, et al. International estimated fetal weight standards of the INTERGROWTH-21(st) Project. Ultrasound Obstet Gynecol. 2017; 49(4):478–86. https://doi.org/10.1002/uog.17347 PMID: 27804212

27. Zambia consolidationed guidelines for treatment and prevention of HIV infection. In: Health Mo, editor. Lusaka Zambia2016.

28. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol. 2007; 6:Article25.

29. The HPFOREST Procedure. SAS® Enterprise Miner™ 142: High-Performance Procedures. Cary, NC: SAS Institute Inc.; 2016.

30. The GENMOD Procedure. SAS/STAT® 143 User's Guide. Cary, NC: SAS Institute Inc.; 2017.

31. The GAM Procedure. SAS/STAT® 143 User's Guide. Cary, NC: SAS Institute Inc.; 2017.

32. Keil A. SuperLearnerMacro 2018 [Available from: https://cirl-unc.github.io/SuperLearnerMacro/.

33. Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3(1):32–5. PMID: 15405679

34. Montagu D, Yamey G, Visconti A, Harding A, Yoong J. Where do poor women in developing countries give birth? A multi-country analysis of demographic and health survey data. PLoS One. 2011; 6(2): e17155. https://doi.org/10.1371/journal.pone.0017155 PMID: 21386886

35. UNICEF. The State of the World's Children. New York: NY; 2017.

36. Hoffman CS, Messer LC, Mendola P, Savitz DA, Herring AH, Hartmann KE. Comparison of gestational age at birth based on last menstrual period and ultrasound during the first trimester. Paediatr Perinat Epidemiol. 2008; 22(6):587–96. https://doi.org/10.1111/j.1365-3016.2008.00965.x PMID: 19000297

37. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller AB, et al. Born too soon: the global epidemiology of 15 million preterm births. Reprod Health. 2013; 10 Suppl 1:S2.

38. Gernand AD, Paul RR, Ullah B, Taher MA, Witter FR, Wu L, et al. A home calendar and recall method of last menstrual period for estimating gestational age in rural Bangladesh: a validation study. J Health Popul Nutr. 2016; 35(1):34. https://doi.org/10.1186/s41043-016-0072-y PMID: 27769295

39. Jehan I, Zaidi S, Rizvi S, Mobeen N, McClure EM, Munoz B, et al. Dating gestational age by last menstrual period, symphysis-fundal height, and ultrasound in urban Pakistan. Int J Gynaecol Obstet. 2010; 110(3):231–4. https://doi.org/10.1016/j.ijgo.2010.03.030 PMID: 20537328

40. Neufeld LM, Haas JD, Grajeda R, Martorell R. Last menstrual period provides the best estimate of gestation length for women in rural Guatemala. Paediatr Perinat Epidemiol. 2006; 20(4):290–8. https://doi.org/10.1111/j.1365-3016.2006.00741.x PMID: 16879501

41. Rosenberg RE, Ahmed AS, Ahmed S, Saha SK, Chowdhury MA, Black RE, et al. Determining gestational age in a low-resource setting: validity of last menstrual period. J Health Popul Nutr. 2009; 27 (3):332–8. PMID: 19507748

42. Stringer EM, Vwalika B, Killam WP, Giganti MJ, Mbewe R, Chi BH, et al. Determinants of stillbirth in Zambia. Obstetrics and gynecology. 2011; 117(5):1151–9. https://doi.org/10.1097/AOG.0b013e3182167627 PMID: 21508755

43. Vwalika B, Stoner MC, Mwanahamuntu M, Liu KC, Kaunda E, Tshuma GG, et al. Maternal and newborn outcomes at a tertiary care hospital in Lusaka, Zambia, 2008–2012. Int J Gynaecol Obstet. 2017; 136 (2):180–7. https://doi.org/10.1002/ijgo.12036 PMID: 28099725