medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

TITLE

Development of risk prediction models for preterm delivery in a rural setting in Ethiopia

AUTHORS

Clara Pons-Duran, PhD[1]*, Bryan Wilder, PhD[1,2]*, Bezawit Mesfin Hunegnaw, MD MPH[3], Sebastien Haneuse, PhD[4], Frederick G. B. Goddard, MS PhD[1], Delayehu Bekele, MD MPH[1,5], Grace J. Chan, MD MPH PhD[1,3,6]

AFFILIATIONS

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[2] Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

[3] Department of Pediatrics and Child Health, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia

[4] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[5] Department of Obstetrics and Gynecology, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia

[6] Division of Medical Critical Care, Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, MA, USA.

* Joint first authors

CORRESPONDING AUTHOR

Clara Pons-Duran, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Kresge 913, Boston, MA 02115

cponsduran@hsph.harvard.edu

WORD COUNT

3072

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

ABSTRACT

Background. Preterm birth complications are the leading causes of death among children under five years. A key practical challenge, however, is the inability to accurately identify pregnancies that are at high risk of preterm delivery, especially in resource-limited settings where there is limited availability of biomarkers assessment.

Methods. We evaluated whether risk of preterm delivery can be predicted using available data from a pregnancy and birth cohort in Amhara region, Ethiopia. All participants were enrolled in the cohort between December 2018 and March 2020. The study outcome was preterm delivery, defined as any delivery occurring before week 37 of gestation regardless of vital status of the fetus or neonate. A range of sociodemographic, clinical, environmental, and pregnancy-related factors were considered as potential inputs. Cox and accelerated failure time models, and decision tree ensembles were used to predict risk of preterm delivery. Model discrimination was estimated using the area-under-the-curve (AUC). Additionally, the conditional distributions of cervical length (CL) and fetal fibronectin (FFN) were simulated to ascertain whether those factors could improve model performance.

Results. A total of 2493 pregnancies were included. Of those, 138 women were censored due to loss-to-follow-up before delivery. Overall, predictive performance of models was poor. The AUC was highest for the tree ensemble classifier (0.60, 95%CI [0.57, 0.63]). When models were calibrated so that 90% of women who experienced a preterm delivery were classified as high risk, at least 75% of those classified as high risk did not experience the outcome. The simulation of CL and FFN distributions did not significantly improve models' performance.

Conclusions. Prediction of preterm delivery remains a major challenge. In resource-limited settings, predicting high-risk deliveries would not only save lives, but also inform resource allocation. It may not be possible to accurately predict risk of preterm delivery without investing in novel technologies to identify genetic factors, immunological biomarkers or the expression of specific proteins.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

INTRODUCTION

Globally, almost 15 million babies are born preterm before 37 weeks of gestation each year [1]. A variety of factors are known to be associated with risk of preterm birth, including obstetrics history, anthropometric measurements, infections, ultrasound measurements and biological and genetic markers [2,3]. Accurate prediction tools to identify women at an increased risk of preterm delivery would allow policy makers, practitioners and researchers to target interventions designed to reduce preterm deliveries. Some studies conducted in high income settings developed predictive models to classify women based on their risk for preterm birth considering multiple maternal characteristics [4,5]. Their discriminative performance was modest (area under the receiving operator characteristic curve (AUC) from 0.62 to 0.70), with generally lower performance when performing external validation [5].

Targeting interventions to high-risk pregnancies is a critical challenge because of the lack of accurate prediction tools. Some published models were developed for women with *a priori* known risk factors like preterm labor or multiple pregnancy [6-8]. Other models used predictors that are not readily available in resource-limited settings such as cervical length (CL), bacterial vaginosis, fetal fibronectin (FFN), cytokine concentration and other biomarkers [9-11]. Additionally, to our knowledge, no published model handled competing risks of stillbirth, or considered a combined outcome of preterm delivery regardless of vital status of the fetus or neonate. However, preterm and stillbirth share common causes and risk factors, and it is likely that the biological mechanisms that trigger preterm labor or rupture of membranes may lead to the delivery of a preterm stillborn in extreme cases [12,13]. Overall, there is a gap in the development of prediction tools that are accurate and applicable to the general population, with and without *a priori* risk, especially in low-resource countries where data on biomarkers that could contribute to improve model performance are not commonly available.

Most preterm birth cases occur among women without known risk factors [14,15]. Limited availability of promising biomarkers to predict preterm birth in low-resource settings make it critical to develop context-specific predictive tools. We use large datasets from the Birhan pregnancy cohort [16], to test whether it is possible to predict risk of preterm delivery in rural Ethiopia. Additionally, we aim to ascertain whether it would be effective to invest in the

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

collection of known key predictors such as CL or FFN to improve accuracy of predictions by simulating the conditional distribution of those two factors.

METHODS

Study design and setting

We conducted a cohort study in the Birhan field site, including 16 villages in Amhara region, Ethiopia, covering a population of 77,766, to estimate morbidity and mortality outcomes among 17,108 women of reproductive age and 8,554 children under-five with house-to-house surveillance every three months. The site is a platform for community and facility-based research and training that was established in 2018, with a focus on maternal and child health [17]. Nested in the site is an open pregnancy and birth cohort that enrolls approximately 2,000 pregnant women and their newborns per year with rigorous longitudinal follow-up over the first two years of life and household data linked with health facility information [16]. The catchment area is rural and semi-urban, covers both highland and lowland areas, and includes two different districts, Angolela Tera, and Kewet/Shewa Robit.

We used data from the Birhan Health and Demographic Surveillance System (HDSS) and the nested pregnancy and birth cohort, Birhan Maternal and Child Health (MCH) cohort to develop a series of risk prediction models for preterm delivery [16,17]. The HDSS provides estimates and trends of health and demographic outcomes including morbidity among women of reproductive age and children under two years and births, deaths, marriages, and migration in the entire population [17]. The pregnancy and birth cohort, generates evidence on pregnancy, birth, and child outcomes using clinical and epidemiological data at both the community and health facility level [16].

Study participants

The sample for this study included women enrolled during pregnancy in the MCH cohort between December 2018 and March 2020. They were followed-up in home and facility visits through delivery. Women who were enrolled during pregnancy and were followed up beyond 28 weeks of gestation were included in the study. This gestational age cut-off was used because in

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Ethiopia stillbirths are considered ≥28 weeks. We excluded newborns with implausible gestational ages at birth: <28 weeks due to the definition of stillbirth, and ≥46 weeks.

Study variables and definitions

The outcome of this study was *preterm delivery*, a composite indicator defined as any delivery occurring before 37 completed weeks of gestation, regardless of vital status of the fetus or neonate. This includes both *preterm births* (live birth prior to completion of week 37 of gestation) [18], as well as *stillbirths* (any fetal death after 28 completed weeks of gestation) [19] which occurred before 37 weeks of gestation.

Gestational age was estimated using the best available method from ultrasound measurements, reported date of last menstrual period, fundal height or maternal recall of gestational age in months. Detailed information on these estimations can be found elsewhere [20].

The selection of potential predictors was guided by literature review and expert knowledge from study obstetricians and pediatricians. Predictors with low prevalence rates in the sample (rare events with ≤ 5 cases) were dropped. Over 70 socio-demographic, biological, environmental and pregnancy-related predictors were included in the initial models. The complete list of assessed predictors can be found in Table S1 (Online Supplementary Document 1). Dummy variables indicating missingness for each predictor were included as additional variables, an approach that is justified for predictive models because it reflects the complete state of knowledge available at the time of prediction.

Analysis

Descriptive statistics

A descriptive analysis of the background characteristics of women who experienced term compared to preterm delivery was performed using t-test for continuous variables, chi-square test for most of the binary variables and Fisher's exact test for multiple gestations, to test for statistically significant differences between groups.

Prediction models

Five models were fit to predict risk of preterm delivery, including linear models and nonlinear decision tree approaches. All five strategies were designed to predict the outcome of preterm

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

delivery using information available at 28 weeks of gestation. The first four models were time-to-event methods which modelled the time until delivery from the 28th week gestation mark, accounting for left truncation and right censoring of person-time. Left truncation arises when women are enrolled beyond 28 weeks of gestation while right censoring arises when follow-up ceases prior to observation of the event of interest (e.g. due to outmigration or loss-to-follow-up). The first model was a Cox proportional hazards model, fit using the R package *survival* [21]. Second, an accelerated failure time model was fit with a log-logistic distribution using the R package *flexsurv* [22]. Third, a decision tree was fit using the R package *LTRtrees* that extends previous uses of a decision tree in survival analysis to account for left truncation and right censoring (LTRCART, left truncation right censoring classification and regression trees) [23]. Fourth, a decision tree ensemble was implemented using the *eXtreme Gradient Boosting* (XGB) R package which uses a Poisson likelihood function proposed by Fu and Simonoff (2017) to account for right censoring and left truncation [23,24]. Finally, a fifth analysis based on a XGB classification model was fit using a binary outcome (i.e. whether delivery was preterm or not) instead of the time-to-event. During fitting this last model, data which was either right censored or left truncated was excluded.

Models were fit and evaluated using 5-fold cross validation due to the need to evaluate models on out-of-sample data while reserving as much data as possible for fitting [25]. All models were evaluated on the same held-out dataset within each fold, regardless of which data or methods were used while fitting the model. Model performance was assessed using the AUC to assess accuracy at binary classification, and the c-index to assess the fraction of pairs for which predicted risk was concordant with delivery time. For both metrics, a value of 0.5 represents a random prediction which is uncorrelated with the true outcome. Larger values indicate more accurate predictions, and a value of 1 represents predictions which are perfectly concordant with the true outcome.

Simulation of cervical length and fetal fibronectin

A final analysis was performed to simulate the potential impact of including CL and FFN as predictors. These two variables were found in past work to be significantly associated with preterm birth [26-28]. Since they are not regularly collected in the study region, we used simulation to assess the potential gain from collecting them. The simulation used data from the

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

MFMU PREDS study [29], a study which screened 2929 women for risk factors for preterm birth in the United States. PREDS study identified CL and FFN as key predictors for preterm birth [30,31]. Details on the simulation model and comparison between the simulated and real measurements can be found in the Supplemental Methods and Results (Online Supplementary Document 1).

RESULTS

The sample composed 2834 pregnancies. Among those we excluded 75 (2.6%) records with gestational age at delivery <28 and ≥46 weeks, and a further 266 (9.4%) pregnancies whose follow-up did not go beyond 28th gestational week. A total of 2493 pregnancies were included in the study. Of those, 138 (5.5%) women were lost to follow-up before delivery or did not have a recorded gestational age at delivery, so they were treated as censored observations in the time-to-event models and excluded from the binary classification model. A total of 968 (38.8%) women were enrolled in the cohort after 28 weeks of gestation (left-truncation), thus, time-varying predictors were considered missing for them since no information on those factors was available at the time of prediction. These women were also excluded from the binary classification model.

Among the 2355 women who were included in the study and followed until delivery, 14% had a preterm delivery. There was no difference in some background characteristics like age, body mass index, parity or history of previous preterm births among women with term deliveries compared to women with preterm deliveries (Table 1). However, the two groups showed significant differences in literacy (43.3% of women with term delivery were illiterate, compared to 50.5% of those who delivered prematurely), geographic location (42.8% of term deliveries occurred in the highland district within Birhan field site, compared to 52.4% of preterm deliveries), and multiple gestation (1.1% of term deliveries were multiple, compared to 3.4% of preterm deliveries).

Some predictors had high levels of missing data, particularly where our study relied on facility visits. Approximately 25% of participants did not attend any antenatal care visit in the study health facilities after being enrolled in the cohort, thus, variables collected at antenatal care visits, such as current infections or concomitant diseases were missing for these women. Further,

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

over 70% of women who attended at least one antenatal care visit had missing data on lab and point-of-care results such as white blood cell counts, proteinuria or bacteriuria.

The predictive performance of all models was generally poor (Table 2). The c-statistic and AUC were highest for the XGB classification model, with an AUC of 0.60. The receiver operating characteristic curves (ROC) depict the trade-off between the false and true positive rates achieved by varying the threshold for classifying delivery as preterm or term (Figure 1). As an example, at the point on this curve corresponding to a 90% true positive rate, all models had a false positive rate of at least 75%, indicating a lack of specificity in picking out women who are truly at higher risk.

There was substantial heterogeneity in the factors that were ultimately retained in the five models (Table 3). Both biological and socio-demographic factors were among the top contributors of standard time-to-event models. Regarding decision tree models, the top five predictors are mainly biological, with neonatal sex being the predictor with the greatest importance.

The performance of several individual models improved when simulated measurements of CL and FFN were included as features for each individual, particularly the accelerated failure time and LTRCART decision tree models (Table 4). However, no model exceeded an estimated AUC of 0.60 indicating that the overall predictability of preterm delivery did not change substantially from the inclusion of these additional predictors.

DISCUSSION

Our study shows that risk prediction of preterm delivery remains a challenge in the absence of data on biomarkers. Despite using a wide range of methodological approaches to adjust for missing data, competing risk of stillbirths, and late cohort enrollments, both traditional epidemiological and machine learning models performed poorly and had low specificity in identifying women who delivered before 37 weeks of gestation. This low predictive performance differs with existing models with higher predictive ability that were designed to predict preterm birth among women at an already high risk due to obstetric conditions such as twin pregnancy [7,8], short cervix, cervical insufficiency [10,32], or hospital admission due to preterm labor [6,33,34]. Other higher performing algorithms used predictors that were not available or

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

applicable in low-resource settings such as amniotic and cervical fluids [33], inflammatory markers [35], or method of conception [5,8]. Lack of a fixed prediction time point and competing risk of stillbirth are methodological gaps of most published studies.

To our knowledge, our study is among the few in developing risk prediction models in a low-resource setting. Only one published study presented the development of a model for preterm birth prediction in Ethiopia [36]. Despite reporting good model performance, this study presents an important limitation; it predicts preterm birth retrospectively using all information available in hindsight (e.g. events such as premature rupture of membranes), while our aim is to assess whether early prediction is possible to inform preventive interventions, leading us to fix a time point for prediction (28 weeks of gestation). Moreover, their study was conducted using data from a hospital-based cohort, likely to be composed of women at higher *a priori* risk of adverse outcomes.

Studies carried out in Ethiopia identified risk factors for preterm delivery including obstetric conditions, socio-demographic characteristics, urinary and vaginal infections, and hypertensive disorders [37,38]. We included all these factors to build the most accurate models with the available data. Among all predictors, neonatal sex was assigned high importance in the decision tree models despite the small difference in prevalence between boys and girls. This is consistent with higher rates of preterm birth for male fetuses in other studies [39,40]. Nevertheless, prediction studies like this aim to characterize prognosis and to anticipate or forecast an outcome.

We simulated the conditional distribution of CL and FFN to test whether access to these measurements may improve models' predictive ability. However, the improvement of the models was negligible, indicating that women who are identified as high-risk via CL or FFN could also be identified as high risk via other predictors. Although both CL and FFN are among the most used indicators to identify high risk pregnancies for preterm delivery in clinical practice, their measurement is not always recommended in *a priori* low-risk populations [15,41]. Our results do not support the allocation of resources to CL and FFN measurement in low-resource settings to predict risk of preterm birth. Similarly, other studies observed a poor predictive power of CL and FFN in the absence of additional maternal predictors [42].

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

The results of this study have important research-related and public health implications. Given the poor performance of all available predictive models, it is fundamental to continue research on the underlying causes of preterm delivery, better understanding the pathways between different risk factors and preterm birth, in order to predict and prevent preterm births in the future. Most cases still occur among women without any known risk factor [14,15]. It is crucial to look for new indicators and biomarkers of preterm delivery. Genetic factors, immunological biomarkers and protein expression are showing promising results [43,44]. There may be value in exploring the use of 'omics' since no biomarkers predictive of preterm birth have yet been identified [45]. While all settings can benefit from such technologies, from an equity perspective it may be especially important to ensure availability in low-resource settings where the survival of preterm infants is lower and identifying high-risk women can enable targeted preventative interventions.

Predictive algorithms with modest performance could be used to identify pregnancies at a very low risk of preterm delivery, thus excluding them from interventions. However, a large proportion of women at a low risk will still be targeted in those interventions due to the low specificity of the models. In Ethiopia, recommending pregnant women to stay in maternity waiting homes is part of the birth preparedness strategy, though it has not been demonstrated to improve pregnancy outcomes [46,47]. Targeting the recommendation of staying in maternity waiting homes to a reduced number of individuals would increase the cost-effectiveness of the intervention and improve the pregnancy experience of some women.

Our findings should be interpreted considering some limitations. Similar to most longitudinal studies, there was study attrition. To address loss to follow-up, we adjusted time-to-event models for censoring and created a "missing" category for all predictors with missing data. The use of a composite outcome that included all preterm deliveries regardless of vital status of the fetus or neonate did not enable the models to separately predict the risk of having a live preterm baby from the risk of having a preterm stillbirth. However, the use of a combined outcome allows us to address competing risks of preterm stillbirths, a common limitation of other available prediction models. Despite these limitations, our study fills an evidence gap by exploring prediction of preterm delivery during the first 28 weeks of gestation in a resource-limited setting with important restrictions in data availability. We considered a comprehensive selection of >70 predictors and tested five different algorithms: Accelerated Failure Time, Cox, LTRCART and two decision tree ensembles. We acknowledge that the classification tree ensemble is not

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

recommended for data with censoring or truncation. However, we fit this model together with the other four algorithms for the purpose of being fully exhaustive in our effort to explore all potential methodological options to achieve our study aim of developing an accurate algorithm. The results of all of them show that the difficulty of predicting preterm delivery is robust to potential variation in the process of constructing risk models.

In settings with low coverage of antenatal care and limited resources to perform ultrasound and biomarker measurements, predicting risk of preterm delivery remains a major challenge. New indicators of preterm delivery may be necessary to enable targeted interventions.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

REFERENCES

1 Chawanpaiboon S, Vogel JP, Moller A-B, Lumbiganon P, Petzold M, Hogan D, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. The Lancet Global Health. 2019;7:e37-e46.

2 Cobo T, Kacerovsky M, Jacobsson B. Risk factors for spontaneous preterm delivery. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics. 2020;150:17-23.

3 Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet (London, England). 2008;371:75-84.

4 Kim JI, Lee JY. Systematic Review of Prediction Models for Preterm Birth Using CHARMS. Biol Res Nurs. 2021;23:708-22.

5 Meertens LJE, van Montfort P, Scheepers HCJ, van Kuijk SMJ, Aardenburg R, Langenveld J, et al. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. Acta obstetricia et gynecologica Scandinavica. 2018;97:907-20.

6 Stock SJ, Horne M, Bruijn M, White H, Boyd KA, Heggie R, et al. Development and validation of a risk prediction model of preterm birth for women with preterm labour symptoms (the QUIDS study): A prospective cohort study and individual participant data meta-analysis. PLoS Med. 2021;18:e1003686.

7 van de Mheen L, Schuit E, Lim AC, Porath MM, Papatsonis D, Erwich JJ, et al. Prediction of Preterm Birth in Multiple Pregnancies: Development of a Multivariable Model Including Cervical Length Measurement at 16 to 21 Weeks' Gestation. Journal of Obstetrics and Gynaecology Canada. 2014;36:309-19.

8 Zhang J, Pan M, Zhan W, Zheng L, Jiang X, Xue X. Two-stage nomogram models in mid-gestation for predicting the risk of spontaneous preterm birth in twin pregnancy. Arch Gynecol Obstet. 2021;303:1439-49.

9 Lee KJ, Yoo J, Kim YH, Kim SH, Kim SC, Kim YH, et al. The Clinical Usefulness of Predictive Models for Preterm Birth with Potential Benefits: A KOrean Preterm collaboratE Network (KOPEN) Registry-Linked Data-Based Cohort Study. Int J Med Sci. 2020;17:1-12.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

10 Lee SM, Park KH, Jung EY, Cho SH, Ryu A. Prediction of spontaneous preterm birth in women with cervical insufficiency: Comprehensive analysis of multiple proteins in amniotic fluid. J Obstet Gynaecol Res. 2016;42:776-83.

11 Zhu YZ, Peng GQ, Tian GX, Qu XL, Xiao SY. New model for predicting preterm delivery during the second trimester of pregnancy. Scientific reports. 2017;7:11294.

12 Gravett MG, Rubens CE, Nunes TM, Group TGR. Global report on preterm birth and stillbirth (2 of 7): discovery science. BMC pregnancy and childbirth. 2010;10.

13 Lawn JE, Gravett MG, Nunes TM, Rubens CE, Stanton C, Group GR. Global report on preterm birth and stillbirth (1 of 7): definitions, description of the burden and opportunities to improve data. BMC pregnancy and childbirth. 2010;10:S1.

14 Mercer BM, Goldenberg RL, Moawad AH, Meis PJ, Iams JD, Das AF, et al. The preterm prediction study: effect of gestational age and cause of preterm birth on subsequent obstetric outcome. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Am J Obstet Gynecol. 1999;181:1216-21.

15 Rozenberg P. Universal cervical length screening for singleton pregnancies with no history of preterm delivery, or the inverse of the Pareto principle. BJOG : an international journal of obstetrics and gynaecology. 2017;124:1038-45.

16 Chan GJ, Hunegnaw BM, Van Wickle K, Mohammed Y, Hunegnaw M, Bekele C, et al. Birhan maternal and child health cohort: a study protocol. BMJ Open. 2021;0.

17 Bekele D, Hunegnaw BM, Bekele C, Van Wickle K, Tadesse F, Goddard FGB, et al. Cohort Profile: The Birhan Health and Demographic Surveillance System. Int J Epidemiol. 2022;51:e39-e45.

18 WHO. Preterm birth. 2018. Available: https://www.who.int/news-room/fact-sheets/detail/preterm-birth. Accessed: 16 December 2022.

19 Blencowe H, Cousens S, Jassir FB, Say L, Chou D, Mathers C, et al. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. The Lancet Global Health. 2016;4:e98-e108.

20 Chan GJ, Goddard FGB, Hunegnaw BM, Mohammed Y, Hunegnaw M, Haneuse S, et al. Estimates of Stillbirths, Neonatal Mortality, and Medically Vulnerable Live Births in Amhara, Ethiopia. JAMA Netw Open. 2022;5:e2218534.

21 Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

22 Jackson C. flexsurv: A Platform for Parametric Survival Modeling in R. Journal of Statistical Software. 2016;70:1-33.

23 Fu W, Simonoff JS. Survival trees for left-truncated and right-censored data, with application to time-varying covariate data. Biostatistics. 2017;18:352-69.

24 Chen T, Guestrin C. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining2016. p. 785-94.

25 Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. 1st edition ed. New York: Springer Science; 2001.

26 Iams JD, Casal D, McGregor JA, Goodwin TM, Kreaden US, Lowensohn R, et al. Fetal fibronectin improves the accuracy of diagnosis of preterm labor. American Journal of Obstetrics & Gynecology. 1995;173:141-5.

27 Kagan KO, To M, Tsoi E, Nicolaides KH. Preterm birth: the value of sonographic measurement of cervical length. BJOG : an international journal of obstetrics and gynaecology. 2006;113 Suppl 3:52-6.

28 Society for Maternal-Fetal Medicine . Electronic address pso, McIntosh J, Feltovich H, Berghella V, Manuck T. The role of routine cervical length screening in selected high- and low-risk women for preterm birth prevention. Am J Obstet Gynecol. 2016;215:B2-7.

29 Goldenberg RL, Iams JD, Mercer BM, Meis PJ, Moawad AH, Copper RL, et al. The Preterm Prediction Study: The Value ofNew vs Standard Risk Factors in Predicting Early and All Spontaneous Preterm Births. American Journal of Public Health. 1998;88:233-8.

30 Goldenberg RL, Mercer BM, Meis PJ, Copper RL, Das A, McNellis D. The Preterm Prediction Study: Fetal Fibronectin Testing and Spontaneous Preterm Birth. Obstetrics & Gynecology. 1996;87:643-8.

31 Iams JD, Goldenberg RL, Meis PJ, Mercer BM, Moawad A, Das A, et al. The length of the cervix and the risk of spontaneous premature delivery. New England Journal of Medicine. 1996;334:567-72.

32 Yoo HN, Park KH, Jung EY, Kim YM, Kook SY, Jeon SJ. Non-invasive prediction of preterm birth in women with cervical insufficiency or an asymptomatic short cervix (</=25 mm) by measurement of biomarkers in the cervicovaginal fluid. PloS one. 2017;12:e0180878.

33 Holst RM, Hagberg H, Wennerholm UB, Skogstrand K, Thorsen P, Jacobsson B. Prediction of Spontaneous Preterm Delivery in Women With Preterm Labor: Analysis of

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Multiple Proteins in Amniotic and Cervical Fluids. Obstetrics & Gynecology. 2009;114:268-77.

34 Vivanti AJ, Maraux B, Bornes M, Darai E, Richard F, Rouzier R. Threatened preterm birth: Validation of a nomogram to predict the individual risk of very preterm delivery in a secondary care center. J Gynecol Obstet Hum Reprod. 2019;48:501-7.

35 Vogel I, Goepfert AR, Thorsen P, Skogstrand K, Hougaard DM, Curry AH, et al. Early second-trimester inflammatory markers and short cervical length and the risk of recurrent preterm birth. J Reprod Immunol. 2007;75:133-40.

36 Feleke SF, Anteneh ZA, Wassie GT, Yalew AK, Dessie AM. Developing and validating a risk prediction model for preterm birth at Felege Hiwot Comprehensive Specialized Hospital, North-West Ethiopia: a retrospective follow-up study. BMJ Open. 2022;12:e061061.

37 Hassen JA, Handiso MN, Admassu BW. Predictors of Preterm Birth among Mothers Who Gave Birth in Silte Zone Public Hospitals, Southern Ethiopia. Journal of pregnancy. 2021;2021:1706713.

38 Wakeyo D, Addisu Y, Mareg M. Determinants of Preterm Birth among Mothers Who Gave Birth in Dilla University Referral Hospital, Southern Ethiopia: A Case-Control Study. BioMed research international. 2020;2020:7031093.

39 Vatten LJ, Skjaerven R. Offspring sex and pregnancy outcome by length of gestation. Early human development. 2004;76:47-54.

40 Zeitlin J, Saurel-Cubizolles MJ, De Mouzon J, Rivera L, Ancel PY, Blondel B, et al. Fetal sex and preterm birth: are males at greater risk? Human Reproduction. 2002;17:2762-8.

41 Dos Santos F, Daru J, Rogozinska E, Cooper NAM. Accuracy of fetal fibronectin for assessing preterm birth risk in asymptomatic pregnant women: a systematic review and meta-analysis. Acta obstetricia et gynecologica Scandinavica. 2018;97:657-67.

42 Esplin MS, Elovitz MA, Iams JD, Parker CB, Wapner RJ, Grobman WA, et al. Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women. JAMA. 2017;317:1047-56.

43 Leow SM, Di Quinzio MKW, Ng ZL, Grant C, Amitay T, Wei Y, et al. Preterm birth prediction in asymptomatic women at mid-gestation using a panel of novel protein

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

biomarkers: the Prediction of PreTerm Labor (PPeTaL) study. Am J Obstet Gynecol MFM. 2020;2:100084.

44 Zhang G, Feenstra B, Bacelis J, Liu X, Muglia LM, Juodakis J, et al. Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. The New England journal of medicine. 2017;377:1156-67.

45 Hornaday KK, Wood EM, Slater DM. Is there a maternal blood biomarker that can predict spontaneous preterm birth prior to labour onset? A systematic review. PloS one. 2022;17:e0265853.

46 Federal Democratic Republic of Ethiopia MoH. National Antenatal Care Guideline. 2021. Available: https://e-library.moh.gov.et/library/wp-content/uploads/2022/03/ANC-GUIDELINE_Feb-24-2022.pdf. Accessed: 16 December 2022.

47 van Lonkhuijzen L, Stekelenburg J, van Roosmalen J. Maternity waiting facilities for improving maternal and neonatal outcome in low-resource countries. The Cochrane Library. 2012;10(10):CD006759.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

TABLES

Table 1. Characteristics of the study sample

Variables	Total*			Term delivery			Preterm delivery			
	N	**mean**	**SD**	**N**	**mean**	**SD**	**N**	**mean**	**SD**	**p-value**
Maternal age (years)	2487	27.3	6.1	2030	27.3	6	319	27.2	6.4	0.69
Body mass index (preconception)	1108	22.0	3.1	930	22.0	3.2	138	21.9	2.6	0.60
	N	**n**	**%**	**N**	**n**	**%**	**N**	**n**	**%**	**p-value**
Illiterate	2488	1105	44.4	2031	880	43.3	319	161	50.5	0.02
Location: highland district (Angolela Tera)	2489	1075	43.2	2032	870	42.8	319	167	52.4	<0.01
Primiparous	2493	793	31.8	2034	627	30.8	321	108	33.6	0.34
History of a previous preterm birth	2493	47	1.9	2034	38	1.9	321	7	2.2	0.87
Multiple gestation	2408	35	1.5	2034	23	1.1	321	11	3.4	<0.01
Neonatal sex: female	2355	1144	48.6	2034	993	48.8	321	151	47.0	0.59

**Total counts include censored study participants with unknown pregnancy outcome or date of delivery*

SD – Standard Deviation

Table 2. Performance metrics of the different predictive models

Model	AUC (95%CI)		c statistic (95% CI)	
Accelerated failure time	0.57	(0.54, 0.61)	0.53	(0.52, 0.55)
Cox	0.51	(0.47, 0.54)	0.52	(0.50, 0.53)
LTRCART	0.54	(0.51, 0.58)	0.53	(0.51, 0.55)
XGBoost (classification)	0.60	(0.57, 0.63)	0.54	(0.52, 0.55)
XGBoost (survival)	0.52	(0.49, 0.56)	0.52	(0.50, 0.53)

AUC – Area Under the receiver operating characteristic Curve; CI – Confidence Interval; LTRCART – Left Truncation Right Censoring Classification And Regression Trees; XGBoost – eXtreme Gradient Boosting

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Table 3. Top 5 predictors for each model: model coefficients and feature importance scores

Model	**Top predictors**				
	1	**2**	**3**	**4**	**5**
Accelerated failure time: *Accelerator factor*	Multiple gestation 0.74	Lymphocyte % = NA 1.26	Vaginal bleeding 1.22	History of pre-eclampsia 0.83	Maternal age = NA 0.83
Cox: *Hazard ratio (95%CI)*	Multiple gestation 3.15 (0.61, 16.21)	History of diabetes 0.33 (0.03, 3.61)	Maternal age = NA 2.79 (1.72, 4.55)	Lymphocyte % = NA 2.68 (0.10, 1.63)	History of pre-eclampsia 2.29 (1.44, 3.62)
LTRCART: *Split level*	Neonatal sex 1	Multiple gestation 2	NA -	NA -	NA -
XGBoost classification: *Gain*	Location (district) 0.11	Parity 0.10	Family size = top quintile 0.10	Neonatal sex 0.09	Systolic blood pressure 0.08
XGBoost survival: *Gain*	Neonatal sex 1.00	NA -	NA -	NA -	NA -

LTRCART – Left Truncation Right Censoring Classification And Regression Trees; NA – Not Applicable/missing data; XGBoost – eXtreme Gradient Boosting

Table 4. Performance metrics of the models with simulated measurements

Model	**AUC (95%CI)**		**c statistic (95% CI)**	
Accelerated failure time	0.60	(0.56, 0.63)	0.55	(0.53, 0.57)
Cox	0.55	(0.52, 0.59)	0.53	(0.50, 0.55)
LTRCART	0.60	(0.56, 0.64)	0.55	(0.52, 0.57)
XGBoost (classification)	0.60	(0.56, 0.64)	0.55	(0.52, 0.56)
XGBoost (survival)	0.48	(0.45, 0.51)	0.50	(0.48, 0.51)

AUC – Area Under the receiver operating characteristic Curve; CI – Confidence Interval; LTRCART – Left Truncation Right Censoring Classification And Regression Trees; XGBoost – eXtreme Gradient Boosting

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

FIGURES

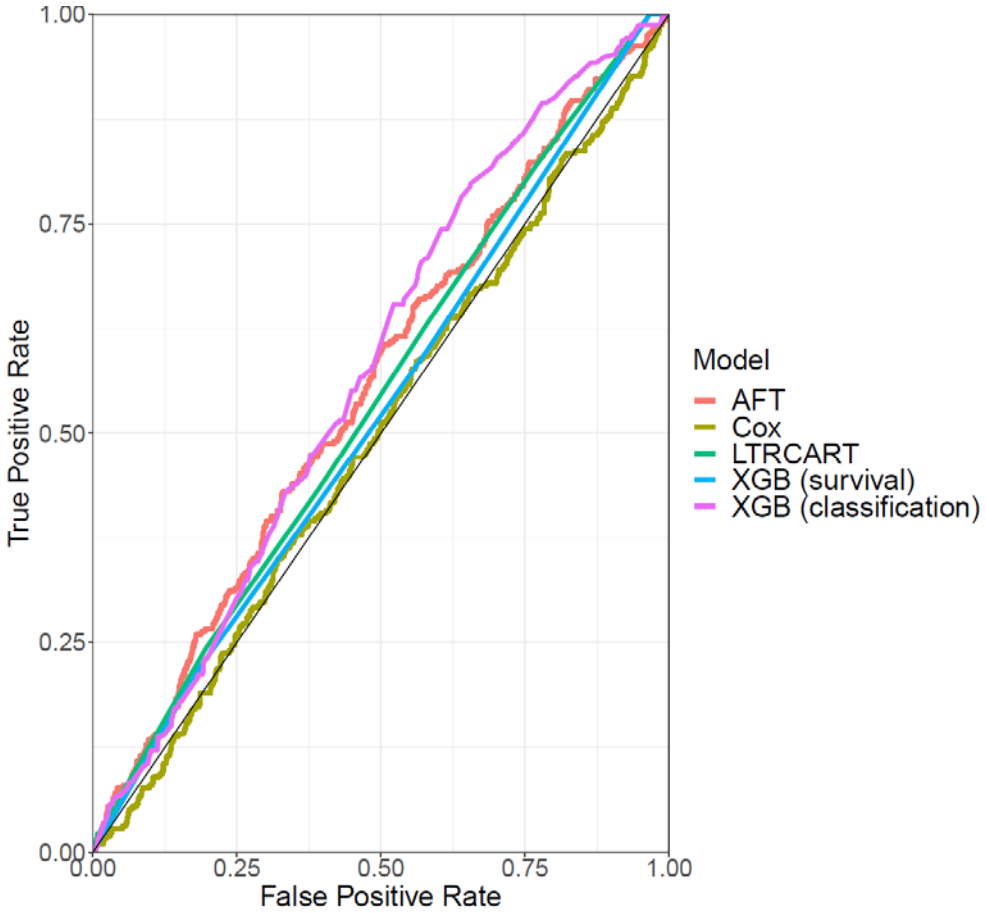


Figure 1. ROC curves for each model.

AFT – Accelerated Failure Time; LTRCART – Left Truncation Right Censoring Classification And Regression Trees; ROC - Receiver Operating Characteristic Curve; XGB – eXtreme Gradient Boosting

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

AUTHORSHIP CONTRIBUTIONS

CPD, BW, SH and GJC conceptualized and designed the study. GJC is the study PI, she obtained funding for the study, and supervised all study activities. DB is the co-PI of the study, he helped to obtain funding for the study and led data collection team. BMH, DB and GJC participated in data collection. CPD and BW conducted the data analysis. SH oversaw the data analysis. FGBG curated the data for this study. CPD and BW drafted the first version of the manuscript. All authors critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank all the mothers and children who participated in the study (HDSS and MCH cohort) and the community of the Birhan field site. We also thank data collectors, supervisors, coordinators, and the HaSET team for their contributions.

ETHICAL CONSIDERATIONS

Ethical clearance was obtained from the Ethics Review Board (IRB) of Saint Paul's Hospital Millennium Medical college, (Addis Ababa, Ethiopia), and Harvard T.H. Chan School of Public Health (Boston, United States). Signed informed consent was obtained from all participants.

COMPETING INTERESTS

No competing interests.

FUNDING

This work has been supported by the Bill & Melinda Gates Foundation (grants INV-010382 and INV-003612 to Dr Chan).

medRxiv preprint doi: https://doi.org/10.1101/2022.11.04.22281948; this version posted December 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

ACCESS TO DATA

Data are available upon reasonable request. Data use is governed by the Birhan Data Access Committee (DAC) and follows Birhan's data sharing policy. All researchers who wish to access Birhan data can complete a Birhan data request form and submit it for decision by the Birhan DAC. Datasets will only be provided with deidentified data to maintain confidentiality of study participants.

SUPPLEMENTARY DOCUMENTS

- Online Supplementary Document 1

Table S1. Predictors assessed for inclusion in the models

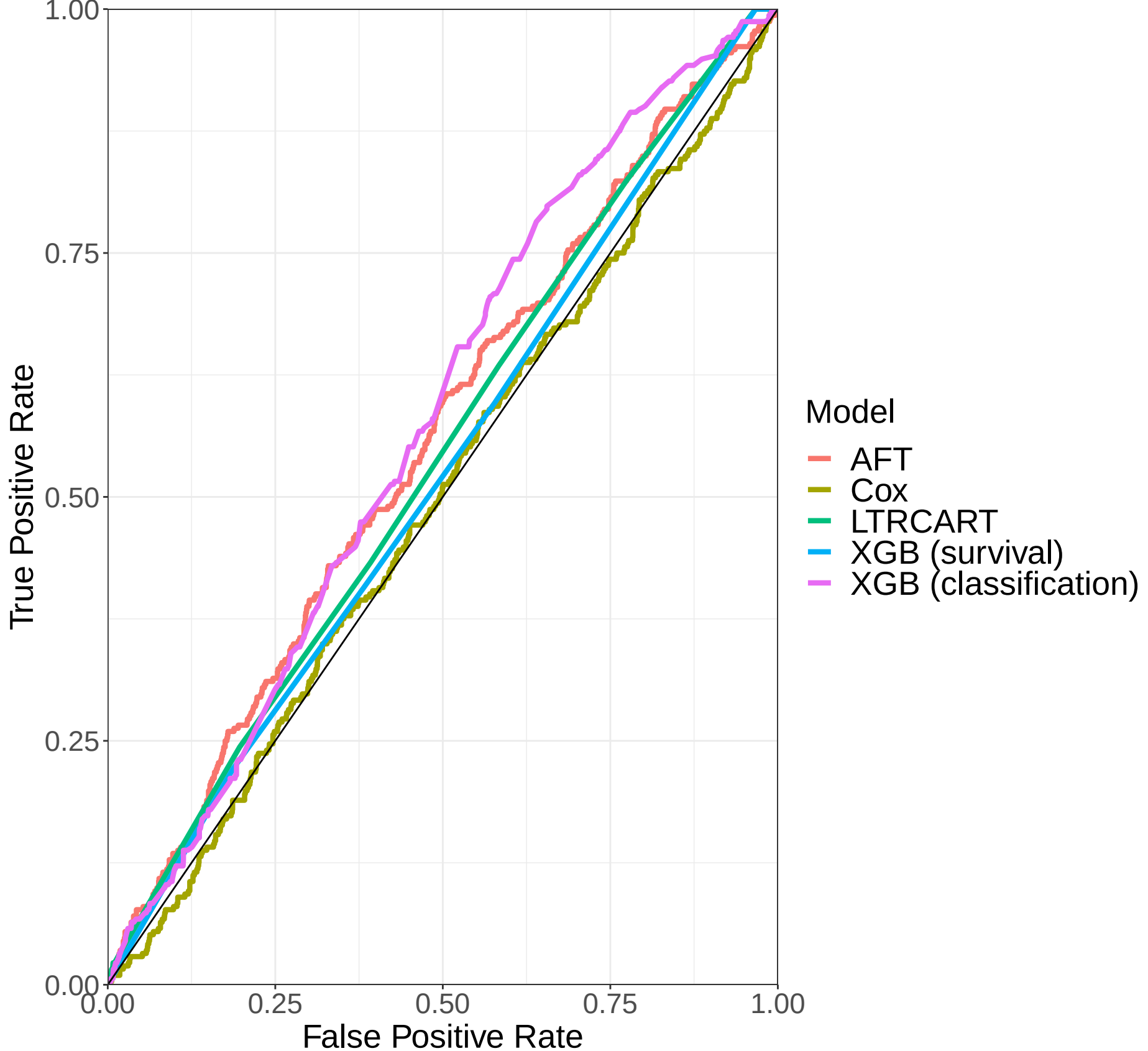
Supplemental Methods and Results. Simulation of cervical length and foetal fibronectin: methods and fit between simulated and real values:

Table S2. Coefficients for simulated model of cervical length and foetal fibronectin

Figure S1. Joint distribution of gestational age and cervical length in the PREDS dataset (left) and simulated values (right). The blue line gives a LOWESS smoothing

Figure S2. Joint distribution of foetal fibronectin (1: positive, 0: negative) and cervical length in the PREDS dataset (left) and simulated values (right). The blue line gives a LOWESS smoothing

- TRIPOD Checklist


1.00
0.75
0.50
0.25
0.00
True Positive Rate
0.00
0.25
0.50
0.75
1.00
False Positive Rate
Model
AFT
Cox
LTRCART
XGB (survival)
XGB (classification)