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

AFFILIATIONS

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

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

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

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

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

* Equal contributions

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

3024

KEY POINTS

Question: Is it possible to accurately predict risk of preterm delivery in a rural setting in Ethiopia?

Findings: In this nested cohort study that used data of 2493 pregnancies, the predictive performance of the five predictive models fit was poor, with the highest area under the receiving operator characteristic curve being 0.60 for the tree ensemble classifier.

Meaning: More research is needed to look for new markers and predictors of preterm delivery, and to evaluate the cost-effectiveness and feasibility of accessing such information in resource-limited settings.

ABSTRACT

Importance: There is a lack of accurate predictive models to identify pregnancies that are at high risk of preterm delivery, especially in resource-limited settings.

Objective: To assess whether it is possible to develop highly accurate prognostic predictive models for preterm delivery using data from a resource-limited setting in Ethiopia.

Design: Nested cohort study using data from the Birhan Health and Demographic Surveillance System and its associated pregnancy and child cohort.

Setting: The study was conducted at the Birhan field site, North Shewa zone, Ethiopia, a platform for community and facility-based research and training, with a focus on maternal and child health.

Participants: Women enrolled during pregnancy in the pregnancy and child cohort between December 2018 and March 2020, who are followed-up in home and facility visits beyond 28 weeks of gestation and through delivery.

Exposures: A wide range of socio-demographic, economic, clinical, environmental, and pregnancy-related factors were considered as potential predictors for preterm delivery. The selection of predictors was guided by literature review and expert knowledge. Predictors that were not available in the surveillance system and cohort datasets were not considered.

Main Outcome Measure: 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. For time-to-event models, time to delivery in weeks was the main outcome measure.

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 was poor. The area under the receiving operator characteristic curve 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.

Conclusions and Relevance: 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.

Introduction

Globally, almost 15 million babies are born preterm before 37 weeks of gestation each year.[1] Premature birth complications are the leading causes of death among children under five years of age,[2] and lead to a range of chronic consequences.[3] While numerous studies have investigated interventions designed to reduce the risk of preterm delivery, they have typically been conducted in populations of pregnancies without regard to underlying risk.[4,5] This is in part because of a lack of accurate prediction tools with which high-risk pregnancies, that have the greatest need and are likely to benefit the most, can be identified.

Numerous factors are associated with risk of preterm birth, including: socio-demographic characteristics; environmental exposures; obstetric history; anthropometric measurements; personal attitudes and behaviors; infections and concomitant diseases; ultrasound measurements; and biological and genetic markers.[6,7] Some studies developed predictive models to classify women based on their risk for preterm birth.[8,9] 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.[9] Better performing predictive models are available; however, for a number of reasons, their findings are not generalizable to pregnant women living in resource-limited settings. First, some published models focused on specific population groups, such as women at high risk of preterm birth due to previous premature deliveries, multiple pregnancy, or primigravidity.[10-13] Second, studies included predictors that are not readily available in resource-limited settings such as cervical length (CL), bacterial vaginosis, fetal fibronectin (FFN), or cytokine concentration.[14-16] Third, many models suffer from methodologic issues including poor data quality, timing of data collection, varying definitions of predictors and lack of validation.[9,17] Finally, 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.[18,19]

In settings where obtaining information on highly-predictive factors for preterm delivery entails challenges, it is fundamental to investigate the use of large datasets from pregnancy cohorts and demographic surveillance systems designed to identify socio-demographic and clinical factors related to birth outcomes. The goal of this study was to assess whether it is possible to develop highly accurate prognostic predictive models using data from a resource-limited setting in Ethiopia. Additionally, the conditional distribution of known key predictors such as CL or FFN was simulated to ascertain whether it would be effective to invest in the collection of those factors to accurately predict risk of preterm delivery.

Methods

Study design and setting

This is a nested cohort study using data from the Birhan Health and Demographic Surveillance System (HDSS) and its associated Maternal and Child Health (MCH) cohort.[20,21] The study was conducted at the Birhan field site, Ethiopia, a platform for community and facility-based research and training, with a focus on maternal and child health. The site catchment area covers two districts (Angolela Tera, and Kewet/Shewa Robit woredas) and 16 villages in the North Shewa Zone, Amhara region. The area is mainly rural with two semi-urban villages, and it covers both highland and lowland areas.[20]

Data collection

In 2018, a baseline survey was conducted to collect information on household composition, socio-demographic status, living conditions, and asset ownership. A total of 19,957 households and 79,653 inhabitants were included in the baseline survey. Subsequently, as part of the HDSS, data is collected every three months on morbidity among women of reproductive age and children under two years, and demographic events (births, deaths, and migration).[20]

Clinical and epidemiological data are collected through community and facility-based visits as part of the MCH cohort. Standardized data is collected on nutritional status, infectious disease, obstetric and medical history, anthropometrics, physical examination, laboratory results, and pregnancy outcomes. Pregnant women are enrolled in the MCH cohort if they live in the Birhan catchment area, provide informed consent and are enrolled in the HDSS.[21]

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. Only women who were followed up beyond 28 weeks of gestation were included in the study. This gestational age cut-off was used because in Ethiopia stillbirths are considered ≥28 weeks. The range of gestational age at delivery considered logical considering was between 28 and less than 46 complete weeks of gestation. Women with a gestational age at delivery outside of this range were excluded.

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) [22], as well as *stillbirths* (any fetal death after 28 completed weeks of gestation)[23] 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 the hierarchy of these estimations can be found elsewhere.[24]

The selection of potential predictors was guided by literature review and expert knowledge from study obstetrician and pediatricians. Predictors that were not available in the HDSS and MCH cohort datasets were not considered. Reasons for dropping predictors from the analysis included ambiguous definitions and low prevalence rates in the sample (rare events with ≤ 5 cases). Socio-demographic, biological, environmental and pregnancy-related predictors were considered for inclusion in the models, while ultrasound and laboratory measurements were unavailable. The complete list of assessed predictors can be found in eTable 1 (Supplement).

A varying number of pregnancies of the study sample had missing data on the different predictors. 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 delivery using information available at 28 weeks of gestation. The first four models were time-to-event methods which modeled 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*.[25] Second, an accelerated failure time model (AFT) was fit with a log-logistic distribution using the R package *flexsurv*.[26] 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).[27] 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.[27,28] 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.[29] 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.[30-32] 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 MFMU PREDS study,[33] 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.[34,35] Details on the simulation model and comparison between the simulated and real measurements can be found in the eMethods and Results (Supplement).

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.

Results

The sample composed 2834 pregnancies. Among those we excluded 75 (2.6%) records with gestational age <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.

Women who experienced term and preterm deliveries showed no significant difference in relation to background characteristics such as age, anthropometrics, and obstetric history (Table 1). However, the two groups showed significant differences in terms of other features including literacy, geographic location, and multiple gestation.

Some predictors had high levels of missing data, particularly where our study relied on information collected by the health system. Variables that were collected at antenatal care visits, such as current infections or concomitant diseases were missing for the approximately 25% of cohort participants who did not attend any antenatal care visits in the study health facilities. Further, 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 tradeoff 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). Among the five predictors with the highest contribution to the standard time-to-event models, there were both biological and socio-demographic factors. Regarding decision tree models, the top five predictors are mainly biological, with neonatal sex being the predictor with the greatest importance. The full prediction models to allow predictions for individuals are available upon request.

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

This nested cohort study used data from 2493 pregnant women to fit a comprehensive series of models attempting to predict preterm delivery in North Shewa Zone, Ethiopia. Both traditional epidemiological and machine learning models performed poorly (AUC and c-statistic <0.60). Moreover, none of the five models had high specificity in identifying women who delivered before 37 weeks of gestation. This low predictive performance contrasts with some related work. However, most well-performing models were designed to predict preterm birth among women at an already high risk due to obstetric conditions such as twin pregnancy,[36,37] short cervix, cervical insufficiency,[14,38] or hospital admission due to preterm labor.[7,10,39,40] In addition, some published algorithms used predictors that were not available or applicable in our setting such as amniotic and cervical fluids,[39] inflammatory markers,[41] or method of conception.[9,36]

To our knowledge, no prediction models for preterm delivery have been developed in Ethiopia. However, association studies carried out in the country identified risk factors including obstetric

conditions, socio-demographic characteristics, urinary and vaginal infections, and hypertensive disorders [42,43]. All these risk factors were assessed for inclusion in our models. Interestingly, 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.[44,45] Further research on associations between predictors and preterm delivery is needed to understand etiologies; prediction studies aim to characterize prognosis, and a different approach is required to inform risk factor choices.

To inform policy decisions on the need to collect data on key known predictors of preterm birth, we simulated the impact of CL and FFN on predictive performance. Contrary to what we expected, 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. Currently, both measurements are among the most used indicators to identify high risk pregnancies for preterm delivery in clinical practice, although their measurement is not always recommended in *a priori* low-risk populations.[46,47] Their association with preterm birth was found to be strongly significant.[34,35] Nonetheless, other studies observed a poor predictive power of CL and FFN in the absence of additional maternal predictors.[48]

Our findings should be interpreted considering some limitations. First, the data had a considerable level of missing information, and a high proportion of lost to follow-up individuals. Both caveats were partially overcome accounting for censoring and including a missing category for all predictors. Second, our models were not externally validated. However, our objective was to create a model that could inform local and regional health authorities. Third, the use of a composite outcome that included all preterm deliveries regardless of vital status of the fetus or neonate made it impossible to separately predict the risk of having a live premature baby. Nonetheless, stillbirths compete in risk with preterm live births, thus our approach addressed this limitation. Finally, binary classification models are not recommended for data with censoring or truncation. We included this method only for the purpose of being fully exhaustive.

Despite the limitations, our study fills an evidence gap by exploring prediction of preterm delivery in a resource-limited setting with important restrictions in data availability. The comprehensive selection of predictors and the variety of tested algorithms allows conclusions to be drawn with clinical and public health implications. First, it is fundamental to continue looking for new indicators and markers of preterm delivery. Genetic factors, immunological biomarkers and protein expression show promising results.[49,50] However, there may be value in exploring the use of 'omics' no biomarker(s) predictive of preterm birth have yet been identified.[51] In addition, preterm birth is thought to be multifactorial and related to infection and inflammation processes.[6] The causal pathway between risk factors and preterm birth is still unclear, and it is critical to understand those mechanisms to predict and prevent this outcome. Most cases still occur among low-risk women without any known risk factor.[46]

Preterm delivery does not appear to be predictable with high accuracy using clinical and socio-demographic data. The use of larger high-quality datasets with data on rare complications that

are predictors of preterm delivery could improve model performance, although a significant increase is not expected. To drive substantial improvements in predictive accuracy, our results suggest it is important to increase the availability of 'omic' technologies to identify gene variants or protein expression. 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. This implies that still a large proportion of women at a low risk will receive such interventions. In Ethiopia, keeping pregnant women in maternity waiting homes is part of the birth preparedness strategy, though it has not been demonstrated to improve pregnancy outcomes.[5,52] 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.

To conclude, in settings with low coverage of antenatal care attendance and limited resources to perform ultrasound and biomarker measurements, it is fundamental to develop tools that can be applied to all pregnant women, and that include the possibility of having missing data for key potential predictors. New indicators of preterm delivery risk are likely necessary to enable targeted interventions.

REFERENCES

1. Chawanpaiboon S, Vogel JP, Moller A-B, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob Health*. 2019;7(1):e37-e46. doi:10.1016/S2214-109X(18)30451-0
2. WHO. Child mortality (under 5 years). Accessed Jun 28, 2022. https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-under-5-mortality-in-2020
3. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet*. 2008;371(9608):261-269. doi:10.1016/s0140-6736(08)60136-1
4. Sosa CG, Althabe F, Belizan JM, Bergel E. Bed rest in singleton pregnancies for preventing preterm birth. *Cochrane Database Syst Rev*. Mar 30 2015;(3):CD003581. doi:10.1002/14651858.CD003581.pub3
5. van Lonkhuijzen L, Stekelenburg J, van Roosmalen J. Maternity waiting facilities for improving maternal and neonatal outcome in low-resource countries. *Cochrane Database Syst Rev*. Oct 17 2012;10(10)doi:10.1002/14651858.CD006759.pub3
6. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet*. Jan 5 2008;371(9606):75-84. doi:10.1016/s0140-6736(08)60074-4
7. Cobo T, Kacerovsky M, Jacobsson B. Risk factors for spontaneous preterm delivery. *Int J Gynaecol Obstet*. Jul 2020;150(1):17-23. doi:10.1002/ijgo.13184
8. Kim JI, Lee JY. Systematic Review of Prediction Models for Preterm Birth Using CHARMS. *Biol Res Nurs*. Oct 2021;23(4):708-722. doi:10.1177/10998004211025641
9. Meertens LJE, van Montfort P, Scheepers HCJ, et al. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. *Acta Obstet Gynecol Scand*. Apr 17 2018;97(8):907-20. doi:10.1111/aogs.13358
10. Stock SJ, Horne M, Bruijn M, 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*. Jul 2021;18(7):e1003686. doi:10.1371/journal.pmed.1003686
11. Weber A, Darmstadt GL, Gruber S, et al. Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women. *Ann Epidemiol*. Nov 2018;28(11):783-789 e1. doi:10.1016/j.annepidem.2018.08.008
12. Arabi Belaghi R, Beyene J, McDonald SD. Prediction of preterm birth in nulliparous women using logistic regression and machine learning. *PLoS One*. 2021;16(6):e0252025. doi:10.1371/journal.pone.0252025
13. Schaaf JM, Ravelli AC, Mol BW, Abu-Hanna A. Development of a prognostic model for predicting spontaneous singleton preterm birth. *Eur J Obstet Gynecol Reprod Biol*. Oct 2012;164(2):150-5. doi:10.1016/j.ejogrb.2012.07.007
14. 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*. Jul 2016;42(7):776-83. doi:10.1111/jog.12976
15. Zhu YZ, Peng GQ, Tian GX, Qu XL, Xiao SY. New model for predicting preterm delivery during the second trimester of pregnancy. *Scientific reports*. Sep 12 2017;7(1):11294. doi:10.1038/s41598-017-11286-x
16. Lee KJ, Yoo J, 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):1-12. doi:10.7150/ijms.37626
17. Honest H, Bachmann LM, Sundaram R, Gupta JK, Kleijnen J, Khan KS. The accuracy of risk scores in predicting preterm birth--a systematic review. *J Obstet Gynaecol*. Jun 2004;24(4):343-59. doi:10.1080/01443610410001685439

18. 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 Childbirth*. 2010;10(1):S1. doi:10.1186/1471-2393-10-S1-S1
19. Gravett MG, Rubens CE, Nunes TM, Group TGR. Global report on preterm birth and stillbirth (2 of 7): discovery science. *BMC Pregnancy Childbirth*. 2010;10(1)doi:10.1186/1471-2393-10-S1-S2
20. Bekele D, Hunegnaw BM, Bekele C, et al. Cohort Profile: The Birhan Health and Demographic Surveillance System. *Int J Epidemiol*. May 9 2022;51(2):e39-e45. doi:10.1093/ije/dyab225
21. Chan GJ, Hunegnaw BM, Wickle KV, et al. Birhan maternal and child health cohort: a study protocol. *BMJ Open*. 2021;0(e049692.)doi:doi:10.1136/bmjopen-2021-049692
22. WHO. Preterm birth. Accessed Jun 28, 2022. https://www.who.int/news-room/fact-sheets/detail/preterm-birth
23. Blencowe H, Cousens S, Jassir FB, et al. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. *Lancet Glob Health*. 2016;4(2):e98-e108. doi:10.1016/s2214-109x(15)00275-2
24. Chan GJ, Goddard FGB, Hunegnaw BM, et al. Estimates of Stillbirths, Neonatal Mortality, and Medically Vulnerable Live Births in Amhara, Ethiopia. *JAMA Netw Open*. Jun 1 2022;5(6):e2218534. doi:10.1001/jamanetworkopen.2022.18534
25. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Springer; 2000.
26. Jackson C. flexsurv: A Platform for Parametric Survival Modeling in R. *J Stat Softw*. 2016;70(8):1-33. doi:10.18637/jss.v070.i08
27. Fu W, Simonoff JS. Survival trees for left-truncated and right-censored data, with application to time-varying covariate data. *Biostatistics*. Apr 1 2017;18(2):352-369. doi:10.1093/biostatistics/kxw047
28. Chen T, Guestrin C. XGBoost. presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016.
29. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. 1st edition ed. Springer Series in Statistics. Springer Science; 2001.
30. Iams JD, Casal D, McGregor JA, et al. Fetal fibronectin improves the accuracy of diagnosis of preterm labor. *Am J Obstet Gynecol*. 1995;173(1):141-5. doi:10.1016/0002-9378(95)90182-5
31. 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*. Sep 2016;215(3):B2-7. doi:10.1016/j.ajog.2016.04.027
32. Kagan KO, To M, Tsoi E, Nicolaides KH. Preterm birth: the value of sonographic measurement of cervical length. *BJOG*. Dec 2006;113 Suppl 3:52-6. doi:10.1111/j.1471-0528.2006.01124.x
33. Goldenberg RL, Iams JD, Mercer BM, et al. The Preterm Prediction Study: The Value ofNew vs Standard Risk Factors in Predicting Early and All Spontaneous Preterm Births. *Am J Public Health*. 1998;88(2):233-8. doi:10.2105/ajph.88.2.233
34. Goldenberg RL, Mercer BM, Meis PJ, Copper RL, Das A, McNellis D. The Preterm Prediction Study: Fetal Fibronectin Testing and Spontaneous Preterm Birth. *Obstet Gynecol*. 1996;87(5):643-8. doi:10.1016/0029-7844(96)00035-x
35. Iams JD, Goldenberg RL, Meis PJ, et al. The length of the cervix and the risk of spontaneous premature delivery. *NEJM*. 1996;334(9):567-72. doi:10.1056/NEJM199602293340904
36. 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*. Jun 2021;303(6):1439-1449. doi:10.1007/s00404-020-05872-0
37. van de Mheen L, Schuit E, Lim AC, et al. Prediction of Preterm Birth in Multiple Pregnancies: Development of a Multivariable Model Including Cervical Length Measurement at 16 to 21 Weeks' Gestation. *J Obstet Gynaecol Can*. 2014;36(4):309-319. doi:10.1016/s1701-2163(15)30606-x

38. 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(7):e0180878. doi:10.1371/journal.pone.0180878
39. Holst RM, Hagberg H, Wennerholm UB, Skogstrand K, Thorsen P, Jacobsson B. Prediction of Spontaneous Preterm Delivery in Women With Preterm Labor: Analysis of Multiple Proteins in Amniotic and Cervical Fluids. *Obstet Gynecol*. 2009;114(2):268-277. doi:10.1097/AOG.0b013e3181ae6a08
40. 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*. Sep 2019;48(7):501-507. doi:10.1016/j.jogoh.2019.04.004
41. Vogel I, Goepfert AR, Thorsen P, et al. Early second-trimester inflammatory markers and short cervical length and the risk of recurrent preterm birth. *J Reprod Immunol*. Oct 2007;75(2):133-40. doi:10.1016/j.jri.2007.02.008
42. Hassen JA, Handiso MN, Admassu BW. Predictors of Preterm Birth among Mothers Who Gave Birth in Silte Zone Public Hospitals, Southern Ethiopia. *J Pregnancy*. 2021;2021:1706713. doi:10.1155/2021/1706713
43. 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 Res Int*. 2020;2020:7031093. doi:10.1155/2020/7031093
44. Zeitlin J, Saurel-Cubizolles MJ, De Mouzon J, et al. Fetal sex and preterm birth: are males at greater risk? *Human Reproduction*. 2002;17(10):2762-8. doi:10.1093/humrep/17.10.2762
45. Vatten LJ, Skjaerven R. Offspring sex and pregnancy outcome by length of gestation. *Early Hum Dev*. Jan 2004;76(1):47-54. doi:10.1016/j.earlhumdev.2003.10.006
46. Rozenberg P. Universal cervical length screening for singleton pregnancies with no history of preterm delivery, or the inverse of the Pareto principle. *BJOG*. Jun 2017;124(7):1038-1045. doi:10.1111/1471-0528.14392
47. 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 Obstet Gynecol Scand*. Jun 2018;97(6):657-667. doi:10.1111/aogs.13299
48. Esplin MS, Elovitz MA, Iams JD, et al. Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women. *JAMA*. Mar 14 2017;317(10):1047-1056. doi:10.1001/jama.2017.1373
49. Zhang G, Feenstra B, Bacelis J, et al. Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *NEJM*. Sep 21 2017;377(12):1156-1167. doi:10.1056/NEJMoa1612665
50. Leow SM, Di Quinzio MKW, Ng ZL, et al. Preterm birth prediction in asymptomatic women at mid-gestation using a panel of novel protein biomarkers: the Prediction of PreTerm Labor (PPeTaL) study. *Am J Obstet Gynecol*. May 2020;2(2):100084. doi:10.1016/j.ajogmf.2019.100084
51. 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(4):e0265853. doi:10.1371/journal.pone.0265853
52. Federal Democratic Republic of Ethiopia MoH. National Antenatal Care Guideline. 2021. March.

AUTHOR 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.

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).

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.

The models developed for this study are available upon request to the authors.

FIGURES

Figure 1. ROC curves for each model.

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

SUPPLEMENT

eTable 1. Predictors assessed for inclusion in the models

eMethods and Results. Simulation of cervical length and fetal fibronectin: methods and fit between simulated and real values:

eTable 2. Coefficients for simulated model of cervical length and fetal fibronectin

eFigure 1. Joint distribution of gestational age and cervical length in the PREDS dataset (left) and simulated values (right). The blue line gives a LOWESS smoothing

eFigure 2. Joint distribution of fetal fibronectin (1: positive, 0: negative) and cervical length in the PREDS dataset (left) and simulated values (right). The blue line gives a LOWESS smoothing

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
Woreda: 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*

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)

LTRCART – Left Truncation Right Censoring Classification And Regression Trees; XGBoost – eXtreme Gradient Boosting

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

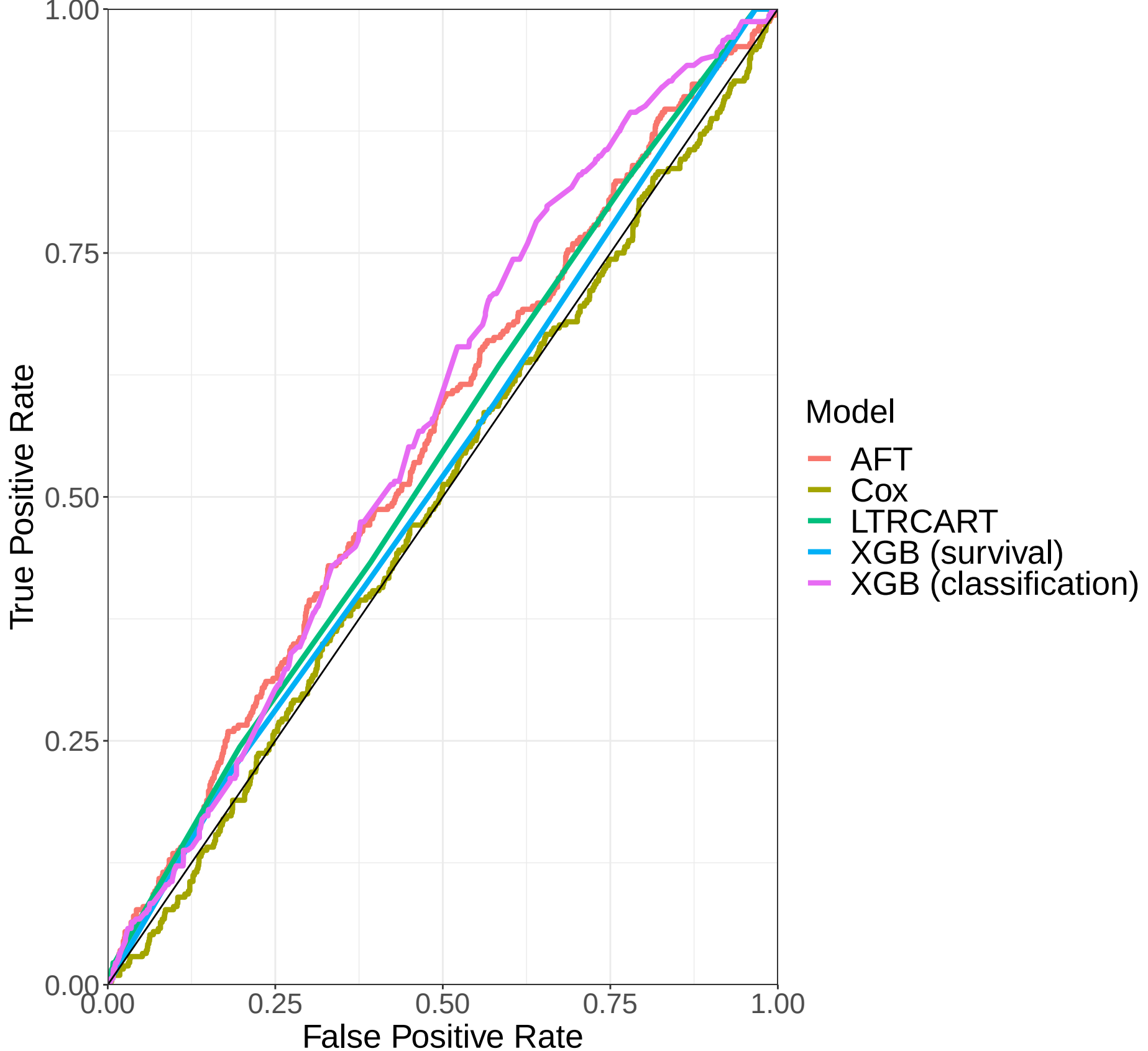
Model	Top predictors				
	1	**2**	**3**	**4**	**5**
Accelerated failure time: *Accelerator factor*	Multiple gestation 0.74	Lymphocyte % = NA 1.26	Vaginal bleeding 1.22	History of pre-eclampsia 0.83	Maternal age = NA 0.83
Cox: *Hazard ratio (95%CI)*	Multiple gestation 3.15 (0.61, 16.21)	History of diabetes 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*	Woreda 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)

LTRCART – Left Truncation Right Censoring Classification And Regression Trees; XGBoost – eXtreme Gradient Boosting


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)