BMJ Open

Developing and validating a risk score for prediction of preterm birth at Felege Hiwot Comprehensive Specialized Hospital, Northwest, Ethiopia: Retrospective follow up study, 2021

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-061061
Article Type:	Original research
Date Submitted by the Author:	01-Feb-2022

Complete List of Authors:	Feleke, Sefineh; Woldia University, Public health Anteneh, Zelalem; Bahir Dar University College of Medical and Health Sciences, Epidemiology Wassie, Gizachew; Bahir Dar University College of Medical and Health Sciences, Epidemiology and Biostatistics Yalew, Anteneh ; Wolkite University, Department of Public Health Dessie , Anteneh; Debre Tabor University, Reproductive Health

Keywords:	OBSTETRICS, PUBLIC HEALTH, PERINATOLOGY

SCHOLARONE™
Manuscripts

Developing and validating a risk score for prediction of preterm birth at Felege Hiwot Comprehensive Specialized Hospital, Northwest Ethiopia: Retrospective follow- up study

Sefineh F.Feleke[1], Zelalem A.Anteneh[2], Gizachew T.Wassie[2], Anteneh K.Yalew[3], Anteneh M.Dessie[4]

1. *Woldia University, Collage of Health Science, Department of Public Health*
2. *Bahir Dar University, College of Medicine and Health Science, School of Public Health, Department of Epidemiology and Biostatistics.*
3. *Wolkite University, Collage of Health Science, Department of Public Health*
4. *Debre Tabor University, Collage of Medicine and Health science, Department of Public Health*

Correspondence: Sefineh Fenta

Collage of Health Science, Department of Public Health, Woldia University,Woldia, Ethiopia

Email address: fentasefineh21@gmail.com

Phone number: +251928573882

Objective: To develop and validate a risk score for the prediction of preterm birth using maternal characteristics.

Method: A retrospective follow-up study was conducted on March (1- 30) 2021 at Felege Hiwot comprehensive specialized hospital. The sample size was determined by assuming 10 events per predictor, based on this assumption total sample size was 1308. Data were collected using a structured checklist through chart review. Data were coded and entered into Epidata, version 3.02, and was analyzed by using R statistical programming language version 4.0.4 for further processing and analysis. Bivariable logistic regression was done to identify the relationship between each predictor and preterm birth. Variables with ($p \leq 0.25$) from the bivariable analysis were entered into a backward stepwise multivariable logistic regression model, and significant variables ($p < 0.05$) were retained in the multivariable model. Model accuracy and goodness of fit were assessed by computing the area under the ROC curve (discrimination) and calibration plot (calibration) respectively

Results: Residence, gravidity, hemoglobin < 11 mg/dl, early rupture of membranes, antepartum hemorrhage, and pregnancy-induced hypertension remained in the final multivariable prediction model. The AUC of the model was 0.816 (95% confidence interval: 0.779 – 0.856).

Conclusion: These results suggest the possibility of predicting preterm birth using a simple prediction model constructed from maternal characteristics.

Strength and Limitations of the study

- ✓ Adequate number of participants with the outcome, which helped us to construct the model using a sufficient number of predictor variables.
- ✓ Prediction model is constructed from easily obtainable maternal characteristics that make it applicable in primary care settings.
- ✓ A single-site study, it is confined to a single area, which needs external validation before using it in another context.
- ✓ Furthermore, data were collected from each mother's card; due to this, some important variables were missed, such as previously highlighted factors with preterm birth in different studies

Background

Preterm birth is described as babies that are born alive before the end of 37 weeks of pregnancy[1]. Preterm birth can be accidental (due to spontaneous preterm labor and/or preterm membrane rupture) or induced by the provider (by cesarean or labor induction)[2]. Most preterm births happen spontaneously[3].

An estimated 15 million babies worldwide are born too early per year. That's more than 1 in 10 infants. About 1 million newborns die per year because of preterm birth complications[4].

Across 184 countries, the rate of preterm birth ranges from 5% to 18% of babies born [5]. However, there are stark disparities in survival rates around the world. Half of the babies born at or below 32 weeks die in low-income settings due to a lack of practical, cost-effective, and critical care, such as comfort, breastfeeding assistance, basic infection care, and trouble Breathing[6].

In Ethiopia, every year, 320,000 babies are born too early and because of direct preterm complications, 24,400 children under five die [7]. According to the 2019 Mini Ethiopia Demographic and Health Survey, the neonatal mortality rate was 30 deaths per 1,000 live births and prematurity was the major cause of death[8]

Furthermore, the effect of preterm birth is also prolonged beyond the neonatal phase and throughout life[9]. Hence, the largest risk of severe health issues, including cerebral palsy, intellectual disability, chronic lung disease, and vision and hearing loss, is faced by babies born before maturity. This introduces a lifelong disability dimension. At some point in their lives, most people will face the struggles and potential disasters of preterm birth either directly in their families or indirectly through events for the nations[9, 10].

To alleviate this burden in the past few decades, numerous methods have been attempted internationally, including in Ethiopia, to prevent and enhance the treatment of preterm births [11-13]. Although several efforts were undertaken to prevent and reduce preterm birth, its rate appears to have increased over time [10, 14]. As part of the strategy, it is essential to diagnose or predict preterm birth earlier in pregnancy to take appropriate measures for high-risk groups.

There are clinical prediction models that attempt to predict the probability of preterm birth, however, all include laboratory tests that are generally not accessible in low-resource settings, like fetal fibronectin, insulin-like growth factor binding protein-1 (IGFBP-1), interleukin-6, and placental alpha-macroglobulin-1 to predict preterm birth[15-20].

Hence, because of limited resources, the use of easily accessible data to forecast preterm birth seems to be appealing in low- and middle-income areas. Although there are prediction models for preterm birth, variation in the occurrence of preterm birth globally is relevant, indicating variations in exposure to psychosocial, sociodemographic, and medical risk factors and genetic differences [21-23].

There is no prediction model for preterm birth in Ethiopia. Therefore, developing and validating a risk score for prediction of preterm birth using maternal(clinical and non-clinical) characteristics based on the available measurement is paramount to allow early preterm birth intervention such as utero transfer to tertiary care centers, appropriate corticosteroid administration while preventing excessive use, neuroprotective magnesium sulfate therapy, and antibiotic treatment in the event of infection[15, 24]

Methods and Materials

Study setting

This retrospective study was conducted among 1260 pregnant women who did prenatal care and finally delivered at Felege Hiwot Comprehensive Specialized Hospital, Bahir Dar city, Northwest Ethiopia from January 30, 2019, to January 30, 2021. Bahir Dar is the capital city of Amhara national regional state and is found 575 km northwest of Addis Ababa.

Felege Hiwot comprehensive specialized hospital was established with the German State government during the regime of Emperor H/ Selassie I in April 1963 G.C and is one of the oldest public hospitals in the Northwestern part of the country and located at the northern end of the city near Lake Tana and aspires to see a healthy, productive and prosperous society and become a center of medical service Excellency by 2029. During its establishment, it was planned to serve 25,000 people. The hospital has currently a total of 1431 manpower (5 obstetricians and gynecologists and 63 midwives among others) in different disciplines. It has a total of 500 formal beds, 11 wards (emergency ward and Inpatient wards such as Gynecological &Obstetric, Surgical, orthopedics, Medical, Pediatric, L&D, Eye unit, NICU, psychiatric, oncology, and 22 OPDS), 39 clinical and non-clinical departments /service units / providing laboratory, Diagnostic, curative & Rehabilitation service at outpatient & inpatient bases as well as disease prevention & health promotion services.

Sample size determination

The sample size required for model development was determined based on the minimum standard of 10 events per candidate predictor considered, according to the formula N = (n × 10)/I where N is the sample size, n is the number of candidate predictor variables and I is the estimated event rate in the population[25]. Since there were 17 candidate predictors considered and 10 events per candidate predictor, the estimated number of events for the study was 170. Based on a study done on the prevalence of preterm birth in Debre Tabor hospital was 13%[26], so taking into account this the required sample size was calculated as follows, n= 170*100/13= 1308.

Patient and public involvement

There was no direct interaction with patients in this study and no direct patient involvement in the design or conduct of this study.

Study Design and Participants

The theoretical design of the present study was; the incidence of preterm birth as a function of multiple predictors during pregnancy. The source population of the study was all pregnant mothers who gave birth at FHCSH. To be included in this study, mothers must meet all of the following eligibility criteria; All medical records of mothers who gave birth and had at least one ANC follow-up in FHCSH from January 30/2019 to January 30/2021.

Sampling method and procedures

A simple random sampling technique was employed to select participants using the medical registration number of a delivered mother from the delivery registration book. First, all mother delivered at FHCSH from January 30/2019 to January 30/20201 was identified from the delivery registration book. After that records of mothers who meet the inclusion criteria were included in the study. Subsequently, a sampling frame was prepared. Finally, the study unit was selected by using a computer-generated random number.

Data Collection

Outcome assessment: The outcome variable was attributed to women whose medical records indicated a physician or midwife diagnosis of preterm birth and delivery between 28 and 36 completed weeks of gestation. The gestational age (GA) was measured using either LNMP, which is found to be a more reliable measure of GA in a low-resource setting[27, 28], or an early ultrasound result.

Predictor assessment: Data was collected using a structured checklist through chart review. Checklists were developed after reviewing various relevant literature [29-33]. It consists of socio-demographic (Maternal age, Residence), Maternal obstetric characteristics : (History of preterm birth, History of abortion, history of stillbirth gravidity, Parity, Multiple pregnancy, APH, PROM, Gestational DM, and PIH), Maternal medical condition : (HGB level, Diabetic Mellitus, Chronic Hypertension, UTI and HIV).

Quality Assurance Mechanisms

To maintain the quality of data, the data collectors and supervisors were trained for a day on the objective of the study, the content of the checklists, how to fill the checklists. Afterward, reviewing 15 charts on medical records of mothers who gave birth at Felege Hiwot Comprehensive Specialized Hospital which is found in Northwest Ethiopia were done. After that, some adjustments were done accordingly. The checklist was developed in English.

Data Processing and Analysis

Data were entered into a software application (EPI DATA, version 3.02) and was analyzed by using R statistical programming language version 4.0.4 for further processing and analysis. There were 13(1%), 2(0.2 %), 11 (0.9 %),15 (2.5%), 21 (1.7%) ,29(2.3%),20(1.6%) and 20 (1.6%) missing values for premature rupture of membranes , residence, chronic hypertension, multiple pregnancy gestational diabetes Mellitus, pregnancy-induced hypertension ,antepartum hemorrhage and hemoglobin respectively. We assumed data were missing at random, and we, therefore, performed a multivariate imputation by chained equations for all variables evaluated in the prediction model [34]. Sensitivity analysis was performed to assess whether the assumption of missing at random (MAR) is valid or not, and the results were reasonably comparable (**Table1**). Descriptive statistics including median, inter-quartile range (IQR), and percentages, were carried out.

Table 1. Sensitivity analysis of the model to predict preterm birth: Comparison of the regression coefficients, standard errors (SE), and p-values for complete case analysis (CCA) and multiple imputed data (MI).

Predicator variables	Complete case analysis			Multiple imputations		
	B	SE	P value	B	SE	P value
Chronic hypertension	0.7313	0.6297	0.24	0.581	0.6285	0.92

(yes)						
Residence (rural)	0.815	0.1946	<0.001	1.154	0.1958	<0.001
GDM(yes)	0.709	0.4028	0.07	0.472	0.4236	0.26
HGB(<11g/dl)	0.497	0.2185	0.02	0.642	0.2153	0.001
PROM (yes)	1.898	0.2080	<0.001	2.097	0.2129	<0.001
APH (yes)	1.194	0.2858	<0.001	1.298	0.2874	<0.001
PIH (yes)	1.353	0.2600	<0.001	1.368	0.2523	<0.001
Multiple pregnancy (yes)	0.539	0.3173	0.08	0.446	0.3257	0.17
Gravidity(primigravida)	0.426	0.1944	0.02	0.711	0.1976	<0.001

Model Development and Validation

For model development, bivariable logistic regression was done to obtain insight into the association of each potential predictor and preterm birth. Variables with ($p < 0.25$) from the bivariable analysis were entered into a backward stepwise multivariable logistic regression model, and significant variables ($p < 0.05$) were retained in the multivariable model. The results of significant predictors were reported as coefficients with 95% confidence intervals (CI). To check for the model accuracy and goodness of fit, we computed the area under the ROC curve (discrimination) and calibration plot (calibration) using "classifierplots" and "givitiR" packages of R respectively. The AUC ranged from 0.5 (no predictive ability) to 1 (perfect discrimination)[35]. The regression coefficients and their 95% confidence levels, and the AUC were adjusted for overfitting or over-optimism using bootstrapping technique. To make internal validation, we computed 1000 random bootstrap [36]samples with the replacement on all predictors in the data. The model's predictive performance after bootstrapping is considered as the performance that can be expected when the model is applied to future similar populations. To evaluate the clinical and public health impact of the model, we performed a decision curve analysis (DCA) [37] of standardized net benefit across a range of threshold probabilities (0 to 1). In the DCA, the model was compared against two extreme scenarios; "intervention for all" and "no intervention". In our case, the intervention considered is the referral of high-risk pregnant women to facilities where appropriate corticosteroid administration, antibiotic treatment.

Risk Score Development

To construct an easily applicable preterm birth prediction score, we transformed each coefficient from the model to a rounded number by dividing it by the lowest coefficient. The number of points was subsequently rounded to the nearest integer. We determined the total score for each individual by assigning the points for each variable present and adding them up. The score was transformed to a dichotomous, allowing each pregnant woman to be classified as having a high or low risk of preterm birth. The receiver operating characteristic curve (ROC) was plotted and the area under the curve (AUC) was calculated to measure the discriminatory power of the scoring system.

Ethical Approval

Ethical clearance was obtained from the Institutional Review Boards (IRB) of Bahir Dar University, College of Medicine and Health Sciences with Protocol number 083/ 2021) on February 26, 2021. Confidentiality was maintained by omitting the personal identifier of the participant during the data collection procedure and information was used only for research purposes. Data were collected from the register, which was kept in a secure place and all data were fully anonymized before we access them. After the collection of data, all the patient records and patients' cards were placed back into a secure place. Data were entered into a password-protected computer.

Result

Demographic, Obstetric, and Clinical Characteristics of mothers who gave birth at Felege Hiwot Comprehensive Specialized Hospital.

A total of 1260 study cards were reviewed from a sample of 1308, about 48 cards were not reviewed due to the outcome of intrauterine fetal death, abortion. ***Table (2)*** shows the demographic, obstetric, and clinical characteristics of mothers who gave birth included in the analysis. The median age of the study participants was 26 years with IQR (24-30years); the majority of the participants 1086 (86.2%) were in the age group of 20-34 years.

More than three fourth of the participants 926 (73.49%) were urban residents. From the total of mothers who delivered at FHCSH, more than two-third 841 (66.7%) were multigravida. About parity, above half of them713 (56.6%) were multipara. Concerning past obstetric history, 55 (6.5%) of them had a history of previous preterm birth, 76 (9%) of them had a previous history of stillbirth and 162 (19.3%) of them had a previous history of abortion.

Table 2. Demographic, obstetric, and clinical characteristics of mothers who gave birth at FHCSH from January 30/2019 to January 30/2021, in Northwest Ethiopia.

Characteristics	Category	Frequency	Percent
Gravidity	Primigravida	419	33.3
	Multigravida	841	66.7
Residence	Urban	926	73.5
	Rural	334	26.5
GDM	Yes	44	3.5
	No	1216	96.5
APH	Yes	84	6.7
	No	1176	93.3
PIH	Yes	110	8.73
	No	1150	91.27
HGB level	<11d/dl	236	18.7
	>=11g/dl	1024	81.3
Chronic hypertension	Yes	21	1.7
	No	1239	98.3

PROM	Yes	195	15.5
	No	1065	84.5
Multiple pregnancies	Yes	90	7.2
	No	1170	92.8

PROM: Premature rupture of membrane, HGB: hemoglobin, PIH: pregnancy-induced hypertension, APH: antepartum hemorrhage, GDM: gestational diabetes mellitus

Development of prediction model for preterm birth

Out of 1260 delivered neonates, 169 (13.4%) (95%, CI (11.6%, 15.4%) was preterm infants.

The bivariable logistic regression analysis found several factors were eligible to be included in the prediction model. Variables with $P \leq 0.25$ in the bivariable logistic regression analysis were hemoglobin level, Gravidity, residence, gestational diabetes mellitus, APH, PIH, chronic hypertension, PROM, and multiple pregnancies. Using the results, a prediction model was developed an equation for the prediction model was obtained. ***(Table 3)***

Table 3: Coefficients and risk-scores of each predictor included in the model to predict preterm birth (*n* = 1260)

Predictors Variables*	Multivariable analysis			
	Original β (95 % CI)	Bootstrap β	P-value	Risk score
Residence (rural)	1.161 (0.780, 1.545)	1.148	<0.001	2
Gravidity (primigravida)	0.675 (0.291, 1.061)	0.666	0.01	1
PROM (yes)	2.081 (1.669 , 2.50)	2.051	<0.001	3
APH (yes)	1.364 (0.806 , 1.915)	1.348	<0.001	2
PIH (yes)	1.387 (0.887 , 1.879)	1.368	<0.001	2
HGB <11g/dl	0.676 (0.255 , 1.09)	0.677	<0.001	1

**Variables retained in the reduced model are; residence, APH, hemoglobin, PIH, gravidity, and PROM. Both backward and forward selection showed the same results. β after internal validation with bootstrapping bootstrapping is shown. Simplified risk score: we divided the coefficient of predictors* *included in the reduced model by the smallest (0.666). The probability or risk of preterm birth = 1/ (1 + exp – (-3.517+ 1.148 * Residence (rural) + 0.666 *gravidity (primigravida) + 2.051*PROM (yes) + 1.348 * APH (yes) + 1.387*PIH +0.677*HGB (<11g/dl)..*

The AUC of the final reduced model was 0.816 (95% confidence interval: 0.779 – 0.856) **(Figure 1a)**. The calibration test had a p-value of 0.6228, indicating that the model does not misrepresent the data or calibration of the model was visually accurate since observed and predicted probabilities were similar **(Figure 1b).**

Validation of the model with the bootstrap technique showed hardly any indication of undue influence by particular observations, with an optimism coefficient of 0.085, resulting AUC of 0.789 (corrected 95% CI: 0.748–0.83).

Using the coefficients (β) the predicted risk cutoff point was a probability of (SpEqualSe > 0.1320), the model has a sensitivity of 75.74%, specificity of 72.87%, a positive predictive value of 30.2%, and a negative predictive value of 95.1%.

When applying DCA, we first evaluate whether our model understudy has a higher net benefit than the default strategies (referring all and none). This model outperforms the default strategies across the relevant threshold range. The model has the highest net benefit across the entire range of threshold probabilities, which indicates that the model has the highest clinical and public health value. Hence, referral decision made using the model has a higher net benefit than not referring at all or referring all regardless of their risk thresholds as shown in ***figure (2)***

Risk Classification Using a Simplified Risk Score

We created a simplified risk score from the model for practical use. The reduced model's prediction score was simplified by rounding all regression coefficients. The simplified score had a considerably comparable prediction accuracy with the original β coefficients, with an AUC of 0.786 (95%CI: 0.729–0.827) **(figure 3).** The possible minimum and maximum scores a mother can have are 0 and 11, respectively.

When dichotomized to low risk (<3) and high risk (≥3) based on the risk score, 278 (14.36%) were categorized as high risk and 982 (77.9%) as low risk for preterm birth. Using "SpEqualSe", the suggested threshold score to predict preterm birth using risk scores is ≥3with a sensitivity of 75.14 % and specificity of 67.46% (**table 4**).

Table 4: Risk classification of preterm birth using simplified prediction score (n = 1260)

Score*(risk category)	Prediction Model Based on Maternal Characteristics	
	Number of mothers	Incidence of preterm birth
<3 (Low)	982 (77.9%)	72 (7.9%)
>=3 (High)	278 (14.36%)	97 (53.59%)
Total	1260 (100%)	169 (13.4%)

** Score = (2*PIH) + (3*PROM) + (hemoglobin < 11 mg/dl) + 2*residence + (2*APH) + gravidity*.

Discussion

In this study, the incidence of preterm birth was found to be 13.4%. Maternal characteristics were identified in this retrospective study to build a preterm birth prediction risk score. The optimal combination of maternal factors to predict preterm birth include residency, gravidity, hemoglobin < 11 mg/dl, early rupture of membranes, antepartum hemorrhage, and pregnancy-induced hypertension, according to the prediction model. The model has an AUC of 0.816 (95%CI: 0.776 – 0.856). Predicting the probability of preterm birth in pregnant women is essential to take appropriate measures accordingly. Identifying women at risk of preterm birth is an important task for clinical care providers. However, in low and middle-income countries, there are only a few methods available for reliably predicting actual preterm labor in women. Previously, the focus of the research was to explain the maternal and fetal determinants of preterm birth. In recent years, the focus shifted to predicting preterm birth optimally using a combined set of characteristics.

Without any advanced laboratory or imaging testing, this study measured the predicted performance of a model based on maternal features during pregnancy. Furthermore, we discovered that utilizing SpEqualSe as an optimal cut point, the sensitivity and specificity of this prediction model achieved 75.14 percent and 67.46 percent, respectively, at the score threshold of 3.

In our study, a combination (residency, gravidity, hemoglobin < 11 mg/dl, early rupture of membranes, antepartum hemorrhage, and pregnancy-induced hypertension) of maternal characteristics results in an AUC of 0.816 (95%CI: 0.776 – 0.856), has an excellent accuracy according to diagnostic accuracy classification[38].

A study conducted in China showed that a model developed using advanced maternal age, lower maternal height, history of preterm delivery, amount of vaginal bleeding during pregnancy, and lack of folic acid intake before pregnancy for the prediction of overall preterm birth with AUC of (0.6)[39].

This difference may be due to some of the predictors they used such as lower maternal height, lack of folic acid intake before pregnancy, and advanced maternal age. However predictors they used such as lack of folic acid intake before pregnancy not easily obtainable information in routine clinical practice, which makes their model less practical in our setting. This prediction model constitutes variables that are easily obtainable and have reasonable accuracy to be used by both mid-and lower-level health professionals in the primary care settings. Among the maternal characteristics included in our model, five can be easily found from history taking and one by test for hemoglobin.

The model's accuracy is consistent with a retrospective study done in China that established a preterm birth prediction model based on maternal characteristics, including demographics and clinical characteristics, and a model with predictors (gravidity, educational status, residency, previous history of preterm birth, twin pregnancy, pre-gestational diabetes mellitus (type I or II), chronic hypertension, and place of birth) with AUC of 0.749 (95%CI: 0.732–0.767) [40].

On the other hand, a model incorporating four predictors (cervical length at admission, gestational age, amniotic fluid glucose, and IL-6) has an area under the curve (AUROC) of 0.86[41] and similarly, the combination of biophysical, biochemical, immunological, microbiological, fetal cell, exosomal, or cell-free RNA at different gestational ages, integrated as part of a multivariable predictor model may be necessary to advance our attempts to predict sPTL and preterm birth. In the prediction of spontaneous preterm birth within 48 hours, a prognostic model including qfFN and clinical risk factors showed excellent results[42, 43]. Both models have higher discriminatory performance. The reason for the lower discriminatory performance in our study as compared to the studies described above could be because we used secondary data available from the register and as this dataset is limited and some variables that require advanced laboratory tests were not included in the model.

Hence, predictors necessitate laboratory testing, which is often unavailable in low-resource settings. As a result, such predictors are difficult to come by in ordinary clinical and public health practice, making the model less useful.

In our prediction score, using 3 as a cutoff point has an acceptable level of specificity, sensitivity, PPV, and NPV to predict preterm birth. It is also possible to shift the cutoff point to increase either of the accuracy measures depending on the program aim and availability of resources.

Conclusion and recommendation

This study shows the possibility of predicting preterm birth using a simple prediction model constructed from maternal characteristics. Thus, the optimal combination of maternal characteristics such as residence, gravidity, hemoglobin < 11 mg/dl, premature rupture of membrane, antepartum hemorrhage, and pregnancy-induced hypertension shows the possibility of predicting preterm birth using a simple prediction model constructed from maternal characteristics. In addition, risk score calculations based on a combination of predictors were effective and had comparable accuracy with the model-based approach of original β coefficients. This score may assist in clinical decision-making. In addition, incorporating this convenient and easily applicable score in the health care system to be used by clinicians to inform pregnant mothers about the future course of their outcome after external validation. Doing further research is needed to validate the prediction tool using prospective follow-up studies in another context before introducing it to the clinical and public health practices.

Data Sharing Statement

The data will be available upon request from the corresponding author.

Author Contributions: S.F.F. conceived the study and wrote the manuscript. Z.A.A, S.F.F, G.T.W, A.K.Y, and A.M.D, all contribute to data analysis, study design, and supervision of data collection. All authors participated in manuscript revision for intellectual content and approval of the final version. All authors have read and agreed to the published version of the manuscript.

Funding statement: This work was supported by Bahir Dar University grant number (2500ETB). This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

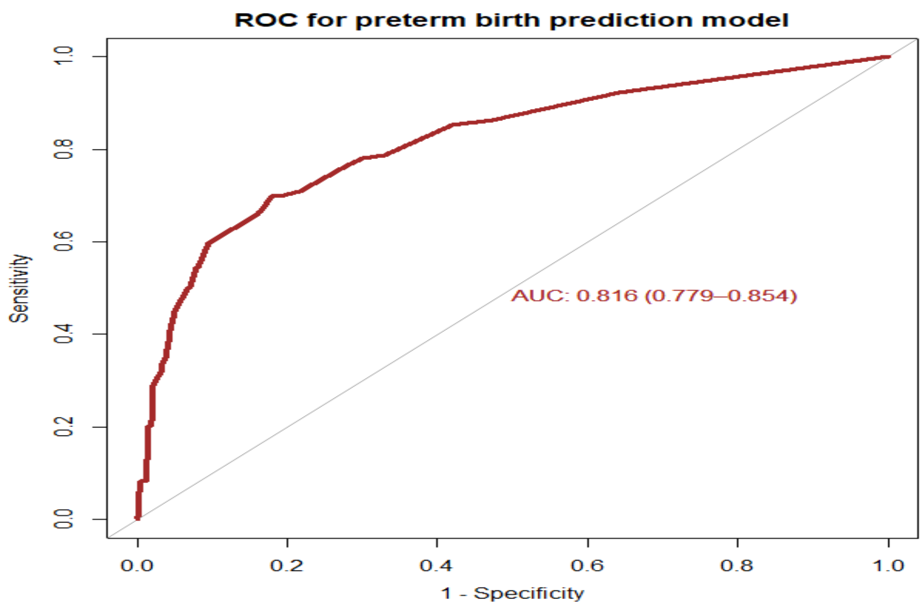
Competing interest's statement: The author reports no conflicts of interest in this work

References

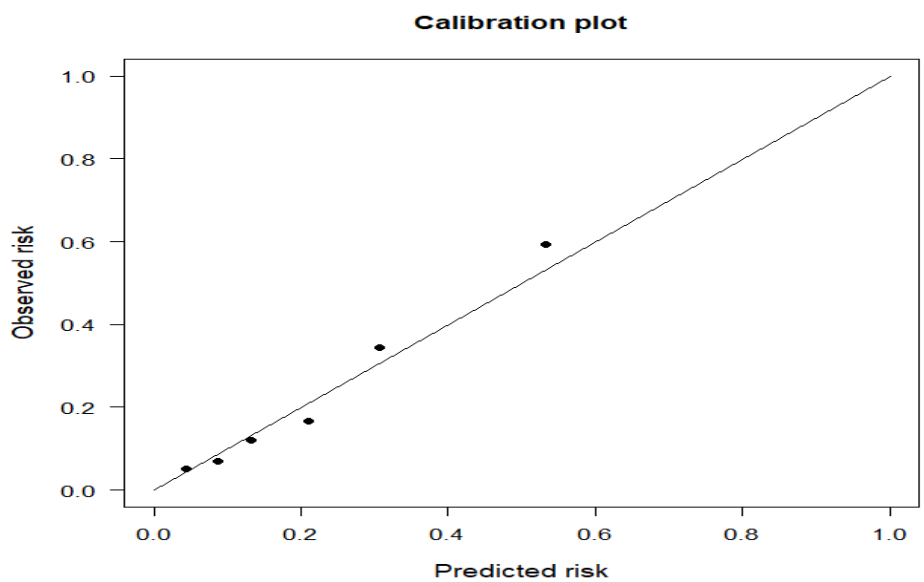
1. World Health Organization: **Preterm birth and low birth weight**. 2020.
2. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet 2008;371(9606):75e84
3. Althabe F: **Born too soon: the global action report on preterm birth**: World Health Organization; 2012.
4. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet. 2016;388(10063):3027-35.
5. World Health Organization. WHO fact sheet on preterm birth. Available from: http://www.who.int/mediacentre/factsheets/fs363/en/
6. World Health Organization: **WHO fact sheet: Preterm birth**. *World Health Organization, Geneva, Switzerland http://www who int/media centre/factsheets/fs363/en/ Accessed* 2018, **26**.
7. https://reliefweb.int/report/ethiopia/ethiopia-profile-preterm-and-low-birth-weight-prevention-and-care
8. Rockville (MD): **DHS Program. Mini Ethiopia Demographic and Health Survey**. ICF; 2019– 2019.
9. Li S, Xi B: **Preterm birth is associated with risk of essential hypertension in later life**. *International journal of cardiology* 2014, **172**(2):e361-e363.
10. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller A-B, Kinney M, Lawn J: **Born too soon: the global epidemiology of 15 million preterm births**. *Reproductive health* 2013, **10**(S1): S2.
11. Soon BT: **The global action report on preterm birth**. *Geneva: World Health Organization* 2012.
12. Griffin JB, Jobe AH, Rouse D, McClure EM, Goldenberg RL, Kamath-Rayne BD: **Evaluating WHO-recommended interventions for preterm birth: a mathematical model of the potential reduction of preterm mortality in Sub-Saharan Africa**. *Global Health: Science and Practice* 2019, **7**(2):215-227.
13. Victora CG, Rubens CE, Group GR: **Global report on preterm birth and stillbirth (4 of 7): delivery of interventions**. *BMC Pregnancy and Childbirth* 2010, **10**(S1): S4.
14. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller A-B, Narwal R, Adler A, Garcia CV, Rohde S, Say L: **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**. *The lancet* 2012, **379**(9832):2162-2172.
15. Oskovi Kaplan ZA, Ozgur-Erdinc AS: **Prediction of Preterm Birth: Maternal Characteristics, Ultrasound Markers, and Biomarkers: An Updated Overview**. 2018, **2018**:8367571.
16. Lucaroni F, Morciano L, Rizzo G: **Biomarkers for predicting spontaneous preterm birth: an umbrella systematic review**. 2018, **31**(6):726-734.
17. Lee KA, Chang MH, Park M-H, Park H, Ha EH, Park EA, Kim YJ: **A model for prediction of spontaneous preterm birth in asymptomatic women**. *Journal of Women's Health* 2011, **20**(12):1825-1831.

18. Georgiou HM, Di Quinzio MK, Permezel M, Brennecke SP: **Predicting preterm labour: current status and future prospects**. *Disease markers* 2015, **2015**.
19. Shennan AH: **Prediction and prevention of preterm birth: a quagmire of evidence**. *Ultrasound Obstet Gynecol* 2018, **51**(5):569-570.
20. Son M, Miller ES: **Predicting preterm birth: Cervical length and fetal fibronectin**. *Semin Perinatol* 2017, **41**(8):445-451.
21. York TP, Strauss III JF, Neale MC, Eaves LJ: **Racial differences in genetic and environmental risk to preterm birth**. *PloS one* 2010, **5**(8):e12391.
22. Culhane JF, Goldenberg RL: **Racial disparities in preterm birth**. In: *Seminars in perinatology: 2011*: Elsevier; 2011: 234-239.
23. Raglan GB, Lannon SM, Jones KM, Schulkin J: **Racial and ethnic disparities in preterm birth among American Indian and Alaska Native women**. *Maternal and child health journal* 2016, **20**(1):16-24.
24. Koullali B, Oudijk M, Nijman T, Mol B, Pajkrt E: **Risk assessment and management to prevent preterm birth**. In: *Seminars in Fetal and Neonatal Medicine: 2016*: Elsevier; 2016: 80-88.
25. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: **A simulation study of the number of events per variable in logistic regression analysis**. *Journal of clinical epidemiology* 1996, **49**(12):1373-1379.
26. Mekonen DG, Yismaw AE, Nigussie TS, Ambaw WM: **Proportion of Preterm birth and associated factors among mothers who gave birth in Debretabor town health institutions, northwest, Ethiopia**. *BMC Research Notes* 2019, **12**(1):2.
27. Tigist B, Abdela A, Zenebe G: **Preterm birth and associated factors among mothers who gave birth in Debre Markos Town Health Institutions**. *Institutional Based Cross sectional study* 2013.
28. Rosenberg RE, Ahmed ANU, Ahmed S, Saha SK, Chowdhury MA, Black RE, Santosham M, Darmstadt GL: **Determining gestational age in a low-resource setting: validity of last menstrual period**. *Journal of health, population, and nutrition* 2009, **27**(3):332.
29. Wassie M, Manaye Y, Abeje G, Tifrie M, Worku G: **Determinants of Preterm Birth among Newborns Delivered in Bahir Dar City Public Hospitals, North West Ethiopia**. 2020.
30. Woday A, Muluneh MD, Sherif S: **Determinants of preterm birth among mothers who gave birth at public hospitals in the Amhara region, Ethiopia: A case-control study**. *PloS one* 2019, **14**(11):e0225060.
31. Wudie F, Tesfamicheal F, Fisseha H, Weldehawaria N, Misgena K, Alema H, Gebregziabher Y, Fisseha G, Woldu M: **Determinants of preterm delivery in the central zone of Tigray, northern Ethiopia: A case-control study**. *South African Journal of Child Health* 2019, **13**(3):108-114.
32. Woldeyohannes D, Kene C, Gomora D, Seyoum K, Assefa T: **Factors Associated with Preterm Birth among Mothers Who gave Birth in Dodola Town Hospitals, Southeast Ethiopia: Institutional Based Cross-Sectional Study**. *Clinics Mother Child Health* 2019, **16**(317):2.
33. Muchie KF, Lakew AM, Teshome DF, Yenit MK, Sisay MM, Mekonnen FA, Habitu YA: **Epidemiology of preterm birth in Ethiopia: systematic review and meta-analysis**. *BMC pregnancy and childbirth* 2020, **20**(1):1-12.

34. Kwak SK, Kim JH: **Statistical data preparation: management of missing values and outliers**. *Korean Journal of anesthesiology* 2017, **70**(4):407.
35. Grobbee DE, Hoes AW. Clinical epidemiology: principles, methods, and applications for clinical research: Jones & Bartlett Publishers; 2014.
36. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 2012;98:683-90. [Crossref] [PubMed]
37. Vickers AJ, Elkin EB: **Decision curve analysis: a novel method for evaluating prediction models**. *Medical decision making: an international journal of the Society for Medical Decision Making* 2006, **26**(6):565-574.
38. Mandrekar JN: **Receiver Operating Characteristic Curve in Diagnostic Test Assessment**. *Journal of Thoracic Oncology* 2010, **5**(9):1315-1316.
39. He J-R, Ramakrishnan R, Lai Y-M, Li W-D, Zhao X, Hu Y, Chen N-N, Hu F, Lu J-H, Wei X-L: **Predictions of preterm birth from early pregnancy characteristics: born in Guangzhou cohort study**. *Journal of clinical medicine* 2018, **7**(8):185.
40. Chen M, Xie N, Liang Z, Qian T, Chen D: **Early Prediction Model for Preterm Birth Combining Demographic Characteristics and Clinical Characteristics**. 2020.
41. Cobo T, Aldecoa V, Figueras F, Herranz A, Ferrero S, Izquierdo M, Murillo C, Amoedo R, Rueda C, Bosch J: **Development and validation of a multivariable prediction model of spontaneous preterm delivery and microbial invasion of the amniotic cavity in women with preterm labor**. *American Journal of Obstetrics and Gynecology* 2020.
42. Lamont R, Richardson L, Boniface J, Cobo T, Exner M, Christensen I, Forslund S, Gaba A, Helmer H, Jørgensen J: **Commentary on a combined approach to the problem of developing biomarkers for the prediction of spontaneous preterm labor that leads to preterm birth**. *Placenta* 2020.
43. Stock SJ, Horne M, Bruijn M, Morris R, Dorling J, Jackson L, Chandiramani M, David AL, Khalil A, Shennan A: **793: A new prediction model for birth within 48 hours in women with preterm labour symptoms**. *American Journal of Obstetrics & Gynecology* 2020, **222**(1): S502.



a)



b)

Figure 1: (a) Area under the ROC curve for the prediction model, and (b) Predicted versus observed preterm birth probability in the sample. This analysis includes mothers who gave birth at FHCSH from January 30/2019 to January 30/2021(n = 1260). Calibration plot created using "plot Calibration" in R programming.

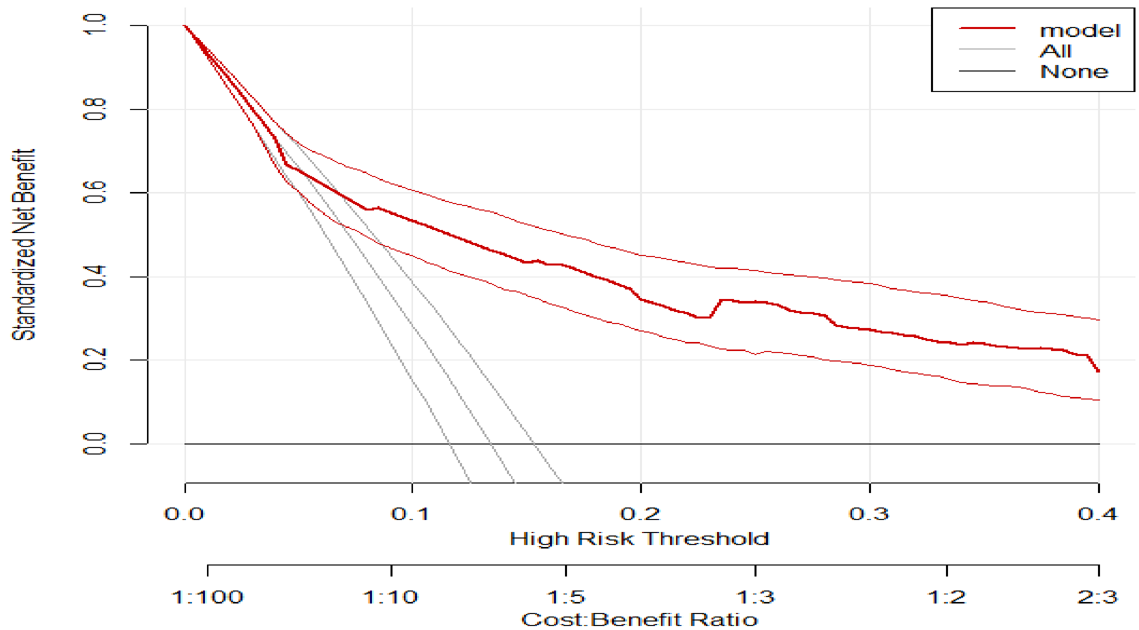


Figure 2: A decision curve plotting showing the net benefit of the model against threshold probability.

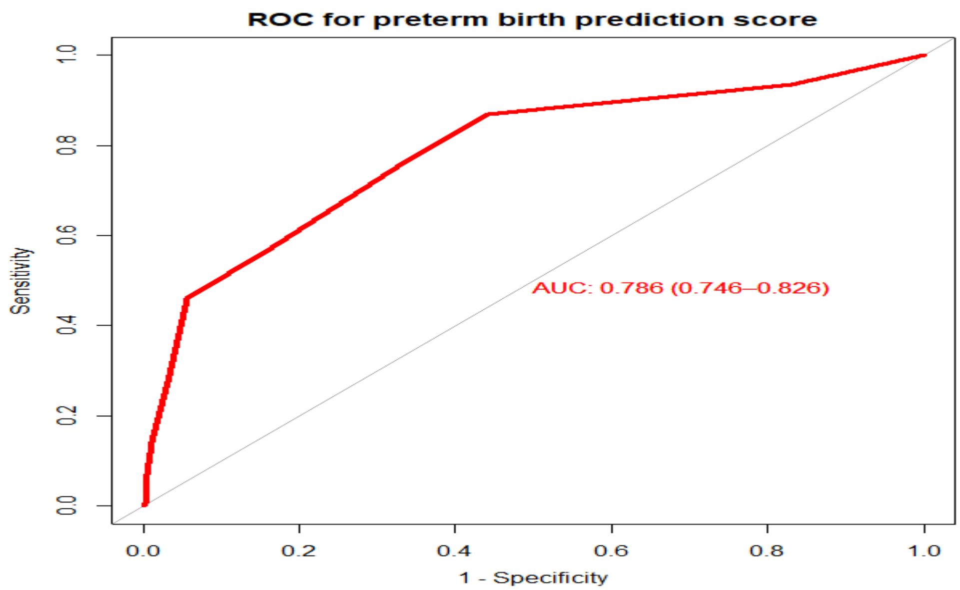


Figure 3: Area under the ROC curve for the simplified risk score to predict the risk of preterm birth among mothers who gave birth at FHCSH from January 30/2019 to January 30/2021.

BMJ Open

Developing and validating risk prediction model for preterm birth at Felege Hiwot comprehensive specialized hospital, Northwest Ethiopia: A retrospective follow-up study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-061061.R1
Article Type:	Original research
Date Submitted by the Author:	16-Jul-2022
Complete List of Authors:	Feleke, Sefineh; Woldia University, Public health Anteneh, Zelalem; Bahir Dar University College of Medical and Health Sciences, Epidemiology Wassie, Gizachew; Bahir Dar University College of Medical and Health Sciences, Epidemiology and Biostatistics Yalew, Anteneh ; Wolkite University, Department of Public Health Mengist, Anteneh; Debre Tabor University, Reproductive Health
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Obstetrics and gynaecology, Public health, Evidence based practice, Paediatrics, Diagnostics
Keywords:	OBSTETRICS, PUBLIC HEALTH, PERINATOLOGY

SCHOLARONE™
Manuscripts

Developing and validating risk prediction model for preterm birth at Felege Hiwot comprehensive specialized hospital, Northwest Ethiopia: A retrospective follow-up study

Sefineh F.Feleke*[1], Zelalem A.Anteneh[2], Gizachew T.Wassie[2], Anteneh K. Yalew[3], Anteneh M.Dessie[4]

1. Woldia University, Collage of Health Science, Department of Public Health
2. Bahir Dar University, College of Medicine and Health Science, School of Public Health, Department of Epidemiology and Biostatistics
3. Wolkite University, Collage of Health Science, Department of Public Health
4. Debre Tabor University, Collage of Health Science, Department of Public Health

*Correspondence: Sefineh Fenta
Email address: fentasefineh21@gmail.com
Phone number: +251928573882

Abstract

Objective: To develop and validate a risk prediction model for the prediction of preterm birth using maternal characteristics.

Design: A retrospective follow-up study was conducted. Data were coded and entered into Epidata, version 3.02, and were analyzed by using R statistical programming language version 4.0.4 for further processing and analysis. Bivariable logistic regression was done to identify the relationship between each predictor and preterm birth. Variables with ($p \leq 0.25$) from the bivariable analysis were entered into a backward stepwise multivariable logistic regression model, and significant variables ($p < 0.05$) were retained in the multivariable model. Model accuracy and goodness of fit were assessed by computing the area under the ROC curve (discrimination) and calibration plot (calibration) respectively.

Setting and participants: This retrospective study was conducted among 1260 pregnant women who did prenatal care and finally delivered at Felege Hiwot Comprehensive Specialized Hospital, Bahir Dar city, Northwest Ethiopia from January 30, 2019, to January 30, 2021.

Results: Residence, gravidity, haemoglobin < 11 mg/dl, early rupture of membranes, antepartum haemorrhage, and pregnancy-induced hypertension remained in the final multivariable prediction model. The AUC of the model was 0.816 (95% confidence interval: 0.779 – 0.856).

Conclusion: This study showed the possibility of predicting preterm birth using maternal characteristics during pregnancy. Thus, using this model could help to identify pregnant women at a higher risk of having a preterm birth to be linked to a center

Keywords: Prediction Model, Preterm birth, Risk score, Ethiopia

Strength and Limitations of the study

- ✓ An adequate number of participants with the outcome helped us to construct the model using a sufficient number of predictor variables and inclusion of sensitivity analyses.
- ✓ Multiple imputations was used to address missing data, which has been shown to be a valid technique for dealing with missing data within logistic regression models, resulting in less bias than excluding all women with missing data[1].
- ✓ The prediction model is constructed from easily obtainable maternal characteristics that make it applicable in primary care settings.

- ✓ A single-site study, it is confined to a single area, which needs external validation before using it in another context.
- ✓ Furthermore, data were collected from each mother's card; due to this, some important variables were missed, such as previously highlighted factors with preterm birth in different studies.

Introduction

Preterm birth is described as babies that are born alive before the end of 37 weeks of pregnancy[2]. Preterm birth can be accidental (due to spontaneous preterm labor and/or preterm membrane rupture) or induced by the provider (by cesarean or labor induction)[3]. Most preterm births happen spontaneously[4].

An estimated 15 million babies worldwide are born too early per year. That's more than 1 in 10 infants. About 1 million newborns die per year because of preterm birth complications[5].

Across 184 countries, the rate of preterm birth ranges from 5% to 18% of babies born [6]. However, there are stark disparities in survival rates around the world. Half of the babies born at or below 32 weeks die in low-income settings due to a lack of practical, cost-effective, and critical care, such as comfort, breastfeeding assistance, basic infection care, and trouble Breathing[7].

Furthermore, the effect of preterm birth is also prolonged beyond the neonatal phase and throughout life[8]. Hence, the largest risk of severe health issues, including cerebral palsy, intellectual disability, chronic lung disease, and vision and hearing loss, is faced by babies born before maturity. This introduces a lifelong disability dimension. At some point in their lives, most people will face the struggles and potential disasters of preterm birth either directly in their families or indirectly through events for the nations[8, 9].

To alleviate this burden in the past few decades, numerous methods have been attempted internationally, including in Ethiopia, to prevent and enhance the treatment of preterm births [10-12]. As part of the strategy, it is essential to diagnose or predict preterm birth earlier in pregnancy to take appropriate measures for high-risk groups.

However, in most nations, predicting preterm birth is still largely based on subjective clinical experience. This approach may increase unnecessary hospital admissions and unnecessary but potentially harmful treatments, such as the use of steroids for the maturation of the fetal lung and tocolysis[13, 14]

There were clinical prediction models that aim to estimate the likelihood of preterm birth that include laboratory tests that are typically inaccessible in low-resource settings, such as fetal fibronectin, insulin-like growth factor binding protein-1 (IGFBP-1), interleukin-6, and placental alpha-macroglobulin-1[15-20].

Although there were prediction models for preterm birth, variation in the occurrence of preterm birth globally is relevant, indicating variations in exposure to psychosocial, sociodemographic, and medical risk factors and genetic differences [21-23].

Hence, because of limited resources, the use of easily accessible data to forecast preterm birth seems to be appealing in low- and middle-income areas.

Therefore, developing and validating a risk prediction model for prediction of preterm birth using maternal(clinical and non-clinical) characteristics based on the available measurement is paramount to allow early preterm birth intervention such as utero transfer to tertiary care centers, appropriate corticosteroid administration while preventing excessive use, neuroprotective magnesium sulfate therapy, and antibiotic treatment in the event of infection[15, 24]

Methods and Materials

Study setting

This retrospective study was conducted among 1260 pregnant women who did prenatal care and finally delivered at Felege Hiwot Comprehensive Specialized Hospital, Bahir Dar city, Northwest Ethiopia from January 30, 2019, to January 30, 2021. Bahir Dar is the capital city of Amhara national regional state and is found 575 km northwest of Addis Ababa. The hospital has currently a total of 1431 manpower (5 Obstetricians and Gynaecologists and 63 midwives among others) in different disciplines. It has a total of 500 formal beds, 11 wards (emergency ward and Inpatient wards such as Gynecological &Obstetric, Surgical, Orthopaedics, Medical, Pediatric, L&D, Eye unit, NICU, psychiatric, oncology, and 22 OPDS), 39 clinical and non-clinical departments /service units / providing laboratory, Diagnostic, curative & Rehabilitation service at outpatient & inpatient bases as well as disease prevention & health promotion services.

Sample size determination

The sample size required for model development was determined based on the minimum standard of 10 events per candidate predictor considered, according to the formula N = (n × 10)/I where N is the sample size, n is the number of candidate predictor variables and I is the estimated event rate in the population[25]. Since there were 17 candidate predictors considered and 10 events per candidate predictor, the estimated number of events for the study was 170. Based on a study done on the prevalence of preterm birth in Debre Tabor hospital was 13%[26], so taking into account this the required sample size was calculated as follows, n= 170*100/13= 1308.

Study Design and Participants

The theoretical design of the present study was; the incidence of preterm birth as a function of multiple predictors during pregnancy. The source population of the study was all pregnant mothers who gave birth at FHCSH. To be included in this study, mothers must meet all of the following eligibility criteria; All medical records of mothers who gave live birth and had at least one ANC follow-up in FHCSH from January 30/2019 to January 30/2021.

Sampling method and procedures

A simple random sampling technique was employed to select participants using the medical registration number of a delivered mother from the delivery registration book. First, all mother delivered at FHCSH in the last two years was identified from the delivery registration book. After that records of mothers who meet the inclusion criteria were included in the study. Subsequently, a sampling frame was prepared. Finally, the study unit was selected by using a computer-generated random number.

Data Collection

Outcome assessment: The outcome variable was attributed to women whose medical records indicated a physician or midwife diagnosis of preterm birth and delivery between 28 and 36 completed weeks of gestation. The gestational age (GA) was measured using either LNMP, which is found to be a more reliable measure of GA in a low-resource setting[27, 28], or an early ultrasound result(12 weeks).

Predictor assessment: Data were collected using a structured checklist through chart review. Checklists were developed after reviewing various relevant literatures [29-33]. It consists of socio-demographic (Maternal age, Residence), Maternal obstetric characteristics : (History of preterm birth, History of abortion, history of stillbirth gravidity, Parity, Multiple pregnancy, APH, PROM, Gestational DM, and PIH), Maternal medical condition : (HGB level, Diabetic Mellitus, Chronic Hypertension, UTI and HIV).

Quality Assurance Mechanisms

To maintain the quality of data, the data collectors and supervisors were trained for a day on the objective of the study, the content of the checklists, how to fill the checklists. Afterward, reviewing 15 charts on medical records of mothers who gave birth at Felege Hiwot Comprehensive Specialized Hospital which is found in Northwest Ethiopia were done. After that, some adjustments (removing variables that were not available in medical record of mothers) were done accordingly. The checklist was developed in English.

Data Processing and Analysis

Data were entered into a software application (EPI DATA, version 3.02) and was analyzed by using R statistical programming language version 4.0.4 for further processing and analysis. There were 13(1%), 2(0.2 %), 11 (0.9 %),15 (2.5%), 21 (1.7%) ,29(2.3%),20(1.6%) and 20 (1.6%) missing values for premature rupture of membranes , residence, chronic hypertension, multiple pregnancy gestational diabetes Mellitus, pregnancy-induced hypertension ,antepartum hemorrhage and hemoglobin respectively.

We assumed data were missing at random, and we, therefore, performed a multivariate imputation by chained equations for all variables evaluated in the prediction model [34]. Sensitivity analysis was performed to assess whether the assumption of missing at random (MAR) is valid or not, and the results were reasonably comparable table **(1)**. Descriptive statistics including median, inter-quartile range (IQR), and percentages, were carried out.

Table 1. Sensitivity analysis of the model to predict preterm birth: Comparison of the regression coefficients, standard errors (SE), and p-values for complete case analysis (CCA) and multiple imputed data (MI).

Predicator variables	Complete case analysis			Multiple imputations		
	B	SE	P value	B	SE	P value
Chronic hypertension (yes)	0.7313	0.6297	0.24	0.581	0.6285	0.92

Residence (rural)	0.815	0.1946	<0.001	1.154	0.1958	<0.001
GDM(yes)	0.709	0.4028	0.07	0.472	0.4236	0.26
HGB(<11g/dl)	0.497	0.2185	0.02	0.642	0.2153	0.001
PROM (yes)	1.898	0.2080	<0.001	2.097	0.2129	<0.001
APH (yes)	1.194	0.2858	<0.001	1.298	0.2874	<0.001
PIH (yes)	1.353	0.2600	<0.001	1.368	0.2523	<0.001
Multiple pregnancy (yes)	0.539	0.3173	0.08	0.446	0.3257	0.17
Gravidity(primigravida)	0.426	0.1944	0.02	0.711	0.1976	<0.001

Model Development and Validation

For model development, bivariable logistic regression was done to obtain insight into the association between each potential predictor and preterm birth. Variables with ($p \leq 0.25$) from the bivariable analysis were entered into a backward stepwise multivariable logistic regression model, and significant variables ($p < 0.05$) were retained in the multivariable model. The results of significant predictors were reported as coefficients with 95% confidence intervals (CI). To check for the model accuracy and goodness of fit, we computed the area under the ROC curve (discrimination) and calibration plot (calibration) using "classifierplots" and "givitiR" packages of R respectively. The AUC ranged from 0.5 (no predictive ability) to 1 (perfect discrimination)[35]. The regression coefficients and their 95% confidence levels, and the AUC were adjusted for overfitting or over-optimism using bootstrapping technique. To make internal validation, we computed 1000 random bootstrap [36]samples with the replacement on all predictors in the data. The model's predictive performance after bootstrapping is considered as the performance that can be expected when the model is applied to future similar populations. To evaluate the clinical and public health impact of the model, we performed a decision curve analysis (DCA) [37] of standardized net benefit across a range of threshold probabilities (0 to 1). In the DCA, the model was compared against two extreme scenarios; "intervention for all" and "no intervention". In our case, the intervention considered is the referral of high-risk pregnant women to facilities where appropriate corticosteroid administration, antibiotic treatment.

Risk Score Development

To construct an easily applicable preterm birth prediction score, we transformed each coefficient from the model into a rounded number by dividing it by the lowest coefficient. The number of points was subsequently rounded to the nearest integer. We determined the total score for each individual by assigning the points for each variable present and adding them up. The score was transformed to a dichotomous, allowing each pregnant woman to be classified as having a high or low risk of preterm birth. The receiver operating characteristic curve (ROC) was plotted and the area under the curve (AUC) was calculated to measure the discriminatory power of the scoring system.

Patient and public involvement

There was no direct interaction with patients in this study and no direct patient involvement in the design or conduct of this study.

Result

Demographic, Obstetric, and Clinical Characteristics of mothers

A total of 1260 study cards were reviewed from a sample of 1308, about 48 cards were not reviewed due to the outcome of intrauterine fetal death, and abortion. Table (**2**) shows the demographic, obstetric, and clinical characteristics of mothers who gave birth included in the analysis. The median age of the study participants was 26 years with IQR (24-30years); the majority of the participants 1086 (86.2%) were in the age group of 20-34 years. More than three fourth of the participants 926 (73.49%) were urban residents. Of the total of mothers who delivered at FHCSH, more than two-thirds of 841 (66.7%) were multigravida. About parity, above half of them713 (56.6%) were multipara. Concerning past obstetric history, 55 (6.5%) of them had a history of previous preterm birth, 76 (9%) of them had a previous history of stillbirth and 162 (19.3%) of them had a previous history of abortion.

Table 2. Demographic, obstetric, and clinical characteristics of mothers who gave birth at FHCSH , Northwest Ethiopia, 2021.

Characteristics	Category	Frequency	Percent
Gravidity	Primigravida	419	33.3
	Multigravida	841	66.7
Residence	Urban	926	73.5
	Rural	334	26.5
GDM	Yes	44	3.5
	No	1216	96.5
APH	Yes	84	6.7
	No	1176	93.3
PIH	Yes	110	8.73
	No	1150	91.27
HGB level	<11d/dl	236	18.7
	>=11g/dl	1024	81.3
Chronic hypertension	Yes	21	1.7
	No	1239	98.3
PROM	Yes	195	15.5

	No	1065	84.5
Multiple pregnancies	Yes	90	7.2
	No	1170	92.8

PROM: Premature rupture of membrane, HGB: hemoglobin, PIH: pregnancy-induced hypertension, APH: antepartum hemorrhage, GDM: gestational diabetes mellitus

Development of prediction model for preterm birth

Out of 1260 delivered neonates, 169 (13.4%) (95%, CI (11.6%, 15.4%) was preterm infants. The bivariable logistic regression analysis found several factors were eligible to be included in the prediction model. These variables were haemoglobin level, Gravidity, residence, gestational diabetes mellitus, APH, PIH, chronic hypertension, PROM, and multiple pregnancies. Using the results, a prediction model was developed an equation for the prediction model was obtained table **(3).**

Table 3: Coefficients and risk-scores of each predictor included in the model to predict preterm birth (*n* = 1260)

Predictors Variables*	**Multivariable analysis**			
	Original β (95 % CI)	Bootstrap β	P-value	Risk score
Residence (rural)	1.161 (0.780, 1.545)	1.148	<0.001	2
Gravidity (primigravida)	0.675 (0.291, 1.061)	0.666	0.01	1
PROM (yes)	2.081 (1.669 , 2.50)	2.051	<0.001	3
APH (yes)	1.364 (0.806 , 1.915)	1.348	<0.001	2
PIH (yes)	1.387 (0.887 , 1.879)	1.368	<0.001	2
HGB <11g/dl	0.676 (0.255 , 1.09)	0.677	<0.001	1

**Variables retained in the reduced model are; residence, APH, hemoglobin, PIH, gravidity, and PROM. Both backward and forward selection showed the same results. β after internal validation with bootstrapping is shown. Simplified risk score: we divided the coefficient of predictors included in the* *reduced model by the smallest (0.666). The probability or risk of preterm birth = 1/ (1 + exp – (-3.517+ 1.148 * Residence (rural) + 0.666 *gravidity (primigravida) + 2.051*PROM (yes) + 1.348 * APH (yes) + 1.387*PIH +0.677*HGB (<11g/dl)..*

The AUC of the final reduced model was 0.816 (95% confidence interval: 0.779 – 0.856) **(Figure 1a)**. The calibration test had a p-value of 0.492, indicating that the model does not misrepresent the data or calibration of the model was visually accurate since observed and predicted probabilities were similar **(Figure 1b).**

Validation of the model with the bootstrap technique showed hardly any indication of undue influence by particular observations, with an optimism coefficient of 0.085, resulting AUC of 0.789 (corrected 95% CI: 0.748–0.83).

Using the coefficients (β) the predicted risk cutoff point was a probability of (SpEqualSe > 0.1320), the model has a sensitivity of 75.74%, specificity of 72.87%, a positive predictive value of 30.2%, and a negative predictive value of 95.1%.

When applying DCA, we first evaluate whether our model understudy has a higher net benefit than the default strategies (referring all and none). This model outperforms the default strategies across the relevant threshold range. The model has the highest net benefit across the entire range of threshold probabilities, which indicates that the model has the highest clinical and public health value. Hence, referral decision made using the model has a higher net benefit than not referring at all or referring all regardless of their risk thresholds as shown in ***figure (2).***

Risk Classification Using a Simplified Risk Score

We created a simplified risk score from the model for practical use. The reduced model's prediction score was simplified by rounding all regression coefficients. The simplified score had a considerably comparable prediction accuracy with the original β coefficients, with an AUC of 0.786 (95%CI: 0.729–0.827) **(figure 3).** The possible minimum and maximum scores a mother can have are 0 and 11, respectively.

Using "SpEqualSe", the suggested threshold score to predict preterm birth using risk scores is ≥3with a sensitivity of 75.14 % and specificity of 67.46% table (**4**).

When dichotomized to low risk (<3) and high risk (≥3) based on the risk score, 278 (14.36%) were categorized as high risk and 982 (77.9%) as low risk for preterm birth.

Table 4: Risk classification of preterm birth using simplified prediction score (n = 1260)

Score*(risk	Prediction Model Based on Maternal Characteristics

category)	Number of mothers	Incidence of preterm birth
<3 (Low)	982 (77.9%)	72 (7.9%)
>=3 (High)	278 (14.36%)	97 (53.59%)
Total	1260 (100%)	169 (13.4%)

**** Score = (2*PIH) + (3*PROM) + (hemoglobin < 11 mg/dl) + 2*residence + (2*APH) + gravidity***.

Discussion

In this study, the incidence of preterm birth was found to be 13.4%. Maternal characteristics were identified in this retrospective study to build a preterm birth prediction risk score. The optimal combination of maternal factors to predict preterm birth include residency, gravidity, hemoglobin < 11 mg/dl, early rupture of membranes, antepartum hemorrhage, and pregnancy-induced hypertension, according to the prediction model. The model has an AUC of 0.816 (95%CI: 0.776 – 0.856). Predicting the probability of preterm birth in pregnant women is essential to take appropriate measures accordingly. Identifying women at risk of preterm birth is an important task for clinical care providers. However, in low and middle-income countries, there are only a few methods available for reliably predicting actual preterm labor in women. Previously, the focus of the research was to explain the maternal and fetal determinants of preterm birth. In recent years, the focus shifted to predicting preterm birth optimally using a combined set of characteristics.

Without any advanced laboratory or imaging testing, this study measured the predicted performance of a model based on maternal features during pregnancy. Furthermore, we discovered that utilizing SpEqualSe as an optimal cut point, the sensitivity and specificity of this prediction model achieved 75.14 percent and 67.46 percent, respectively, at the score threshold of 3.

In our study, a combination (residency, gravidity, hemoglobin < 11 mg/dl, early rupture of membranes, antepartum hemorrhage, and pregnancy-induced hypertension) of maternal characteristics results in an AUC of 0.816 (95%CI: 0.776 – 0.856), has an excellent accuracy according to diagnostic accuracy classification[38].

A study conducted in China showed that a model developed using advanced maternal age, lower maternal height, history of preterm delivery, amount of vaginal bleeding during pregnancy, and lack of folic acid intake before pregnancy for the prediction of overall preterm birth with AUC of (0.6)[39].

This difference may be due to some of the predictors they used such as lower maternal height, lack of folic acid intake before pregnancy, and advanced maternal age. However predictors they used such as lack of folic acid intake before pregnancy not easily obtainable information in routine clinical practice, which makes their model less practical in our setting. This prediction model constitutes variables that are easily obtainable and have reasonable accuracy to be used by both mid-and lower-level health professionals in the primary care settings. Among the maternal characteristics included in our model, five can be easily found from history taking and one by test for hemoglobin.

The model's accuracy is consistent with a retrospective study done in China that established a preterm birth prediction model based on maternal characteristics, including demographics and clinical characteristics, and a model with predictors (gravidity, educational status, residency, previous history of preterm birth, twin pregnancy, pre-gestational diabetes mellitus (type I or II), chronic hypertension, and place of birth) with AUC of 0.749 (95%CI: 0.732–0.767) [40].

On the other hand, a model incorporating four predictors (cervical length at admission, gestational age, amniotic fluid glucose, and IL-6) has an area under the curve (AUROC) of 0.86[41] and similarly, the combination of biophysical, biochemical, immunological, microbiological, fetal cell, exosomal, or cell-free RNA at different gestational ages, integrated as part of a multivariable predictor model may be necessary to advance our attempts to predict sPTL and preterm birth. In the prediction of spontaneous preterm birth within 48 hours, a prognostic model including qfFN and clinical risk factors showed excellent results[42, 43]. Both models have higher discriminatory performance. The reason for the lower discriminatory performance in our study as compared to the studies described above could be because we used secondary data available from the register and as this dataset is limited and some variables that require advanced laboratory tests were not included in the model.

Hence, predictors necessitate laboratory testing, which is often unavailable in low-resource settings. As a result, such predictors are difficult to come by in ordinary clinical and public health practice, making the model less useful.

A study conducted in the UK found that data on maternal characteristics and obstetric history at 11–13 weeks of gestation were predictive of spontaneous early preterm deliveries; this model had an AUC of 0.67[44] which had lower discriminatory performance than the present study. This difference may be difference in study population.

A model that predict a risk of preterm delivery in women with a multiple pregnancy incorporates previous preterm delivery, monochorionicity, smoking, educational level, and triplet pregnancy for preterm and very preterm delivery had a c-index of 0.68 (95% CI 0.63 to 0.72) and 0.68 (95% CI 0.62 to 0.75) respectively[45]. It had lower discriminatory performance than the present study. This might be due to study population difference. In the present study the study populations were both women who had multiple pregnancies and singleton pregnancy.

In our prediction score, using 3 as a cutoff point has an acceptable level of specificity, sensitivity, PPV, and NPV to predict preterm birth. It is also possible to shift the cutoff point to increase either of the accuracy measures depending on the program aim and availability of resources.

Conclusion and recommendation

This study shows the possibility of predicting preterm birth using a simple prediction model constructed from maternal characteristics. Thus, the optimal combination of maternal characteristics such as residence, gravidity, haemoglobin < 11 mg/dl, premature rupture of membrane, antepartum haemorrhage, and pregnancy-induced hypertension shows the possibility of predicting preterm birth using a simple prediction model constructed from maternal characteristics. In addition, risk score calculations based on a combination of predictors were effective and had comparable accuracy with the model-based approach of original β coefficients. This score may assist in clinical decision-making. In addition, incorporating this convenient and easily applicable score in the health care system to be used by clinicians to inform pregnant mothers about the future course of their outcome after external validation. Doing further research is needed to validate the prediction tool using prospective follow-up studies in another context before introducing it to the clinical and public health practices.

Data Sharing Statement

The data will be available upon request from the corresponding author.

Author Contributions: S.F.F. conceived the study and wrote the manuscript. Z.A.A, S.F.F, G.T.W, A.K.Y, and A.M.D, all contribute to data analysis, study design, and supervision of data collection. All authors participated in manuscript revision for intellectual content and approval of the final version. All authors have read and agreed to the published version of the manuscript.

Funding statement: This work was supported by Bahir Dar University grant number (2500ETB).

Competing interest's statement: The author reports no conflicts of interest in this work.

Ethical approval

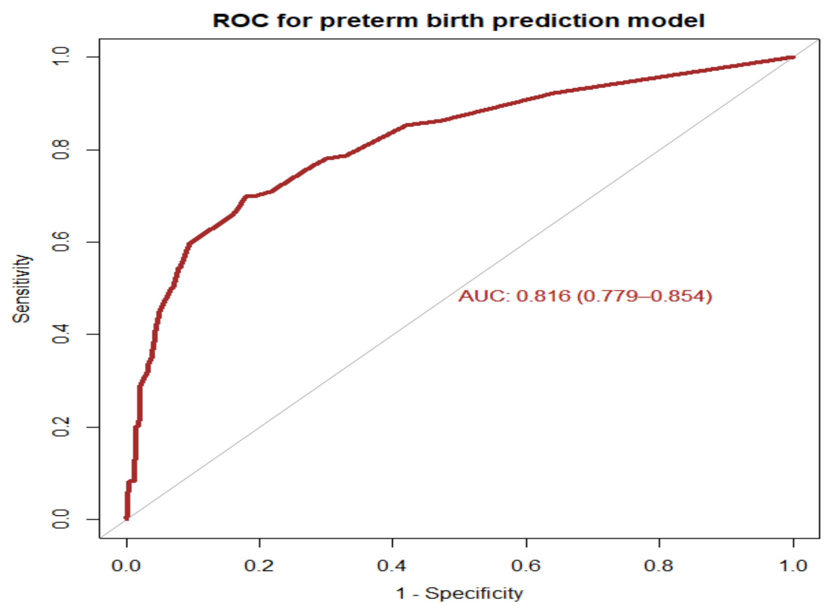
Ethical clearance was obtained from the Institutional Review Boards (IRB) of Bahir Dar University, College of Medicine and Health Sciences with Protocol number 083/ 2021) on February 26, 2021. Confidentiality was maintained by omitting the personal identifier of the participant during the data collection procedure and information was used only for research purposes. Data were collected from the register, which was kept in a secure place and all data were fully anonymized before we access them. After the collection of data, all the patient records and patient cards were placed back in a secure place. Data were entered into a password-protected computer.

References

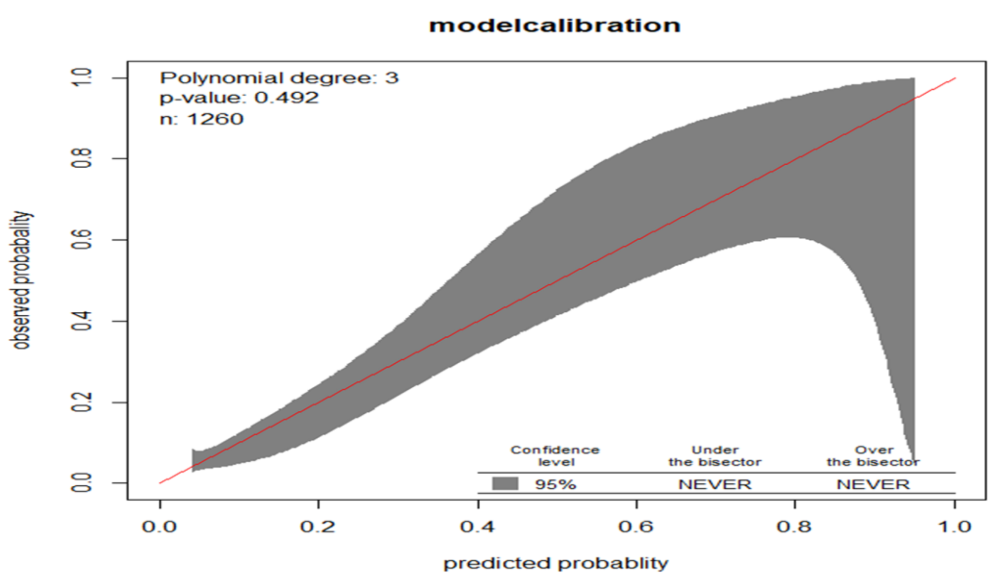
1. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009; b2393:338. https://doi. org/10.1136/bmj.b2393 PMID: 19564179
2. Organization WH: **Preterm birth and low birth weight.** 2020.
3. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet 2008;371(9606):75e84
4. Althabe F: **Born too soon: the global action report on preterm birth:** World Health Organization; 2012.
5. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet. 2016;388(10063):3027-35.
6. World Health Organization. WHO fact sheet on preterm birth. Available from: http://www.who.int/mediacentre/factsheets/fs363/en/
7. Organization WH: **WHO fact sheet: Preterm birth.** *World Health Organization, Geneva, Switzerland http://www who int/mediacentre/factsheets/fs363/en/ Accessed 2018*, 26.
8. Li S, Xi B: **Preterm birth is associated with risk of essential hypertension in later life.** *International journal of cardiology* 2014, 172**(2)**:e361-e363.
9. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller A-B, Kinney M, Lawn J: **Born too soon: the global epidemiology of 15 million preterm births.** *Reproductive health* 2013, 10**(S**1):S2.
10. Soon BT: **The global action report on preterm birth.** *Geneva: World Health Organization 2012*.
11. Griffin JB, Jobe AH, Rouse D, McClure EM, Goldenberg RL, Kamath-Rayne BD: **Evaluating WHO-recommended interventions for preterm birth: a mathematical model of the potential reduction of preterm mortality in Sub-Saharan Africa.** *Global Health: Science and Practice 2019*, 7**(2)**:215-227.
12. Victora CG, Rubens CE, Group GR: **Global report on preterm birth and stillbirth (4 of 7): delivery of interventions.** *BMC Pregnancy and Childbirth* 2010, 10**(S**1):S4.
13. Kemp M, Newnham J, Challis J, Jobe A, Stock S: **The clinical use of corticosteroids in pregnancy.** *Human reproduction update* 2016, 22**(2)**:240-259.
14. Lorthe E, Goffinet F, Marret S, Vayssiere C, Flamant C, Quere M, Benhammou V, Ancel P-Y, Kayem G: **Tocolysis after preterm premature rupture of membranes and neonatal outcome:**

a propensity-score analysis. *American Journal of Obstetrics and Gynecology* 2017, 217(**2**):212. e211-212. e212.
15. Oskovi Kaplan ZA, Ozgu-Erdinc AS: **Prediction of Preterm Birth: Maternal Characteristics, Ultrasound Markers, and Biomarkers: An Updated Overview.** 2018, 20**18:8**367571.
16. Lucaroni F, Morciano L, Rizzo G: **Biomarkers for predicting spontaneous preterm birth: an umbrella systematic review.** 2018, 31(**6**):726-734.
17. Lee KA, Chang MH, Park M-H, Park H, Ha EH, Park EA, Kim YJ: **A model for prediction of spontaneous preterm birth in asymptomatic women.** *Journal of Women's Health* 2011, 20(**12**):1825-1831.
18. Georgiou HM, Di Quinzio MK, Permezel M, Brennecke SP: **Predicting preterm labour: current status and future prospects.** *Disease markers* 2015, 20**15.**
19. Shennan AH: **Prediction and prevention of preterm birth: a quagmire of evidence.** *Ultrasound Obstet Gynecol* 2018, 51(**5**):569-570.
20. Son M, Miller ES: **Predicting preterm birth: Cervical length and fetal fibronectin.** *Semin Perinatol* 2017, 41(**8**):445-451.
21. York TP, Strauss III JF, Neale MC, Eaves LJ: **Racial differences in genetic and environmental risk to preterm birth.** *PloS one* 2010, 5(**8**):e12391.
22. Culhane JF, Goldenberg RL: **Racial disparities in preterm birth.** In: *Seminars in perinatology: 2011:* Elsevier; 2011: 234-239.
23. Raglan GB, Lannon SM, Jones KM, Schulkin J: **Racial and ethnic disparities in preterm birth among American Indian and Alaska Native women.** *Maternal and child health journal* 2016, 20(**1**):16-24.
24. Koullali B, Oudijk M, Nijman T, Mol B, Pajkrt E: **Risk assessment and management to prevent preterm birth.** In: *Seminars in Fetal and Neonatal Medicine: 2016:* Elsevier; 2016: 80-88.
25. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: **A simulation study of the number of events per variable in logistic regression analysis.** *J Clin Epidemiol* 1996, 49(**12**):1373-1379.
26. Mekonen DG, Yismaw AE, Nigussie TS, Ambaw WM: **Proportion of Preterm birth and associated factors among mothers who gave birth in Debretabor town health institutions, northwest, Ethiopia.** *BMC Research Notes* 2019, 12(**1**):2.
27. Tigist B, Abdela A, Zenebe G: **Preterm birth and associated factors among mothers who gave birth in Debre Markos Town Health Institutions.** *Institutional Based Cross sectional study* 2013.
28. Rosenberg RE, Ahmed ANU, Ahmed S, Saha SK, Chowdhury MA, Black RE, Santosham M, Darmstadt GL: **Determining gestational age in a low-resource setting: validity of last menstrual period.** *Journal of health, population, and nutrition* 2009, 27(**3**):332.
29. Wassie M, Manaye Y, Abeje G, Tifrie M, Worku G: **Determinants of Preterm Birth among Newborns Delivered in Bahir Dar City Public Hospitals, North West Ethiopia.** 2020.
30. Woday A, Muluneh MD, Sherif S: **Determinants of preterm birth among mothers who gave birth at public hospitals in the Amhara region, Ethiopia: A case-control study.** *PloS one* 2019, 14(**11**):e0225060.
31. Wudie F, Tesfamicheal F, Fisseha H, Weldehawaria N, Misgena K, Alema H, Gebregziabher Y, Fisseha G, Woldu M: **Determinants of preterm delivery in the central zone of Tigray, northern Ethiopia: A case-control study.** *South African Journal of Child Health* 2019, 13(**3**):108-114.
32. Woldeyohannes D, Kene C, Gomora D, Seyoum K, Assefa T: **Factors Associated with Preterm Birth among Mothers Who gave Birth in Dodola Town Hospitals, Southeast Ethiopia: Institutional Based Cross Sectional Study.** *Clinics Mother Child Health* 2019, 16(**3**17):2.

33. Muchie KF, Lakew AM, Teshome DF, Yenit MK, Sisay MM, Mekonnen FA, Habitu YA: **Epidemiology of preterm birth in Ethiopia: systematic review and meta-analysis.** BMC *pregnancy and childbirth* 2020, 20(**1**):1-12.
34. Kwak SK, Kim JH: **Statistical data preparation: management of missing values and outliers.** *Korean journal of anesthesiology* 2017, 70(**4**):407.
35. Grobbee DE, Hoes AW. Clinical epidemiology: principles, methods, and applications for clinical research: Jones & Bartlett Publishers; 2014.
36. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 2012;98:683-90. [Crossref] [PubMed]
37. Vickers AJ, Elkin EB: **Decision curve analysis: a novel method for evaluating prediction models.** *Med Decis Making* 2006, 26(**6**):565-574.
38. Mandrekar JN: **Receiver Operating Characteristic Curve in Diagnostic Test Assessment.** *Journal of Thoracic Oncology* 2010, 5(**9**):1315-1316.
39. He J-R, Ramakrishnan R, Lai Y-M, Li W-D, Zhao X, Hu Y, Chen N-N, Hu F, Lu J-H, Wei X-L: **Predictions of preterm birth from early pregnancy characteristics: born in guangzhou cohort study.** *Journal of clinical medicine* 2018, 7(**8**):185.
40. Chen M, Xie N, Liang Z, Qian T, Chen D: **Early Prediction Model for Preterm Birth Combining Demographic Characteristics and Clinical Characteristics.** 2020.
41. Cobo T, Aldecoa V, Figueras F, Herranz A, Ferrero S, Izquierdo M, Murillo C, Amoedo R, Rueda C, Bosch J: **Development and validation of a multivariable prediction model of spontaneous preterm delivery and microbial invasion of the amniotic cavity in women with preterm labor.** *American Journal of Obstetrics and Gynecology* 2020.
42. Lamont R, Richardson L, Boniface J, Cobo T, Exner M, Christensen I, Forslund S, Gaba A, Helmer H, Jørgensen J: **Commentary on a combined approach to the problem of developing biomarkers for the prediction of spontaneous preterm labor that leads to preterm birth.** *Placenta* 2020.
43. Stock SJ, Horne M, Bruijn M, Morris R, Dorling J, Jackson L, Chandiramani M, David AL, Khalil A, Shennan A: 79**3: A new prediction model for birth within 48 hours in women with preterm labour symptoms.** *American Journal of Obstetrics & Gynecology* 2020, 22**2**(**1**):S502.
44. Beta, J.; Akolekar, R.; Ventura, W.; Syngelaki, A.; Nicolaides, K.H. Prediction of spontaneous preterm delivery from maternal factors, obstetric history and placental perfusion and function at 11–13 weeks. Prenat. Diagn. 2011, 31, 75–83
45. van de Mheen L, Schuit E, Lim AC, Porath MM, Papatsonis D, Erwich JJ, van Eyck J, Van Oirschot CM, Hummel P, Duvekot JJ: **Prediction of preterm birth in multiple pregnancies: development of a multivariable model including cervical length measurement at 16 to 21 weeks' gestation. Jo**urnal *of obstetrics and gynaecology Canada 2014*, 36(4**)**:309-319.



a)



b)

Figure 1: (a) Area under the ROC curve for the prediction model, and (b) Predicted versus observed preterm birth probability in the sample. This analysis includes mothers who gave birth at FHCSH, 2021(n = 1260). Calibration plot created using "givitiCalibrationBelt" in R programming.

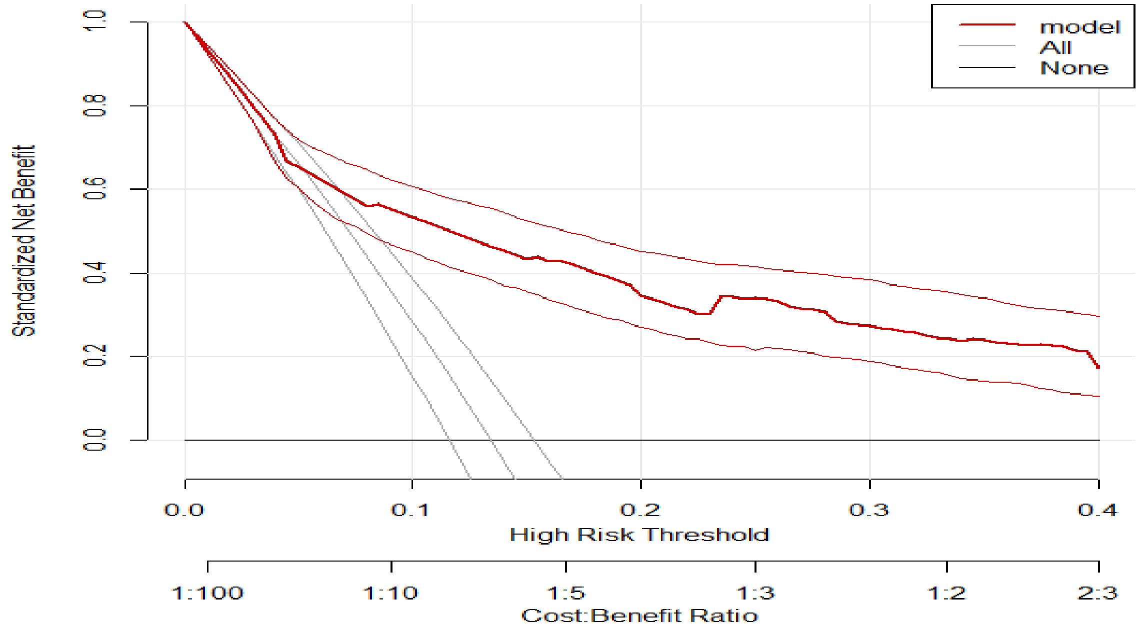


Figure 2: A decision curve plotting the net benefit of the model against threshold probability.

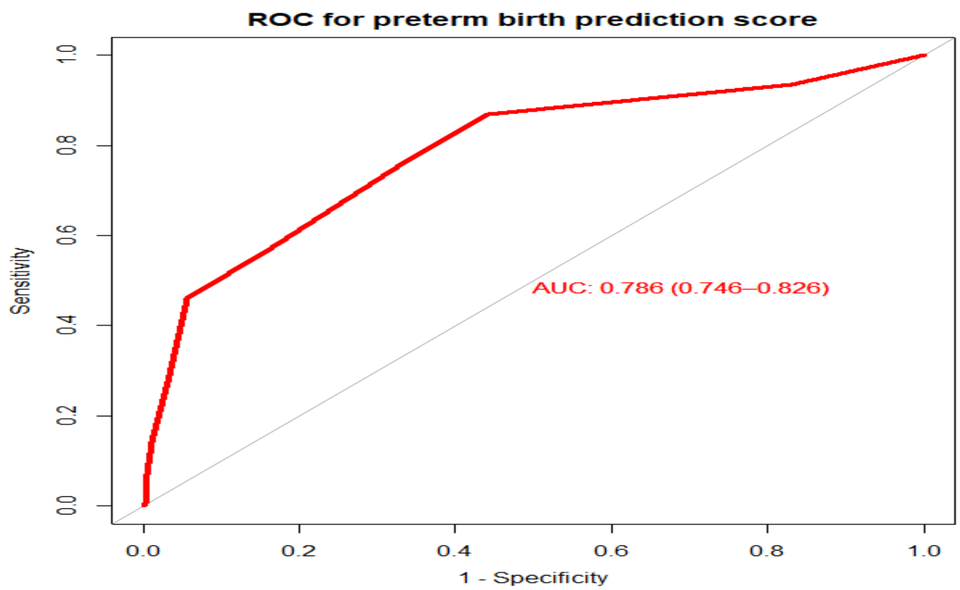


Figure 3: Area under the ROC curve for the simplified risk score to predict the risk of preterm birth among mothers who gave birth at FHCSH, 2021.

The RECORD statement – checklist of items, extended from the STROBE statement that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Line 1-59	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Line 1-59
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Line 60-100		
Objectives	3	State specific objectives, including any prespecified hypotheses	Line 96-98		
Methods					
Study Design	4	Present key elements of study design early in the paper	Line 103		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Line 103-113 Line 123-128		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the	Line 123-128	RECORD 6.1: The methods of study population selection (such as codes or	Line 129-135

		sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Line 137-147	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Line 137-147
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Line 142-147		
Bias	9	Describe any efforts to address potential sources of bias	Line 123-141		
Study size	10	Explain how the study size was	Line 114-122		

		arrived at			
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Line 155-198		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Line 155-198		
Data access and cleaning methods		Line 148-154		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Line 130-132
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-	Line 130-132

				level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	Line 130-135	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Line 130-135
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	Line 232-241		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	Line 247		
Main results	16	(a) Give unadjusted estimates	Line 246-286		

		and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Line 162-166		
Discussion					
Key results	18	Summarise key results with reference to study objectives	Line 351-362		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Line 50-59	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	………………
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Line 351-362		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Line 351-362		

Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Line 369		
Accessibility of protocol, raw data, and programming code		Line 364		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Line 364

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Developing and validating risk prediction model for preterm birth at Felege Hiwot comprehensive specialized hospital, Northwest Ethiopia: A retrospective follow-up study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-061061.R2
Article Type:	Original research
Date Submitted by the Author:	29-Aug-2022

Complete List of Authors:	Feleke, Sefineh; Woldia University, Public health Anteneh, Zelalem; Bahir Dar University College of Medical and Health Sciences, Epidemiology Wassie, Gizachew; Bahir Dar University College of Medical and Health Sciences, Epidemiology and Biostatistics Yalew, Anteneh ; Wolkite University, Department of Public Health Mengist, Anteneh; Debre Tabor University, Reproductive Health

Primary Subject Heading:	Research methods
Secondary Subject Heading:	Obstetrics and gynaecology, Public health, Evidence based practice, Paediatrics, Diagnostics
Keywords:	OBSTETRICS, PUBLIC HEALTH, PERINATOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Developing and validating risk prediction model for preterm birth at Felege Hiwot Comprehensive Specialized Hospital, Northwest Ethiopia: A retrospective follow-up study

Sefineh Fenta Feleke*[1], Zelalem Alamrew Anteneh[2], Gizachew Tadesse Wassie[2], Anteneh Kassa Yalew[3], Anteneh Mengist Dessie[4]

1. Woldia University, Collage of Health Science, Department of Public Health
2. Bahir Dar University, College of Medicine and Health Science, School of Public Health, Department of Epidemiology and Biostatistics
3. Wolkite University, Collage of Health Science, Department of Public Health
4. Debre Tabor University, Collage of Health Science, Department of Public Health

*Correspondence: Sefineh Fenta

Email address: fentasefineh21@gmail.com

Phone number: +251928573882

P.O.Box: 400

Abstract

Objective: To develop and validate a risk prediction model for the prediction of preterm birth using maternal characteristics.

Design: A retrospective follow-up study was conducted. Data were coded and entered into Epidata, version 3.02, and were analyzed by using R statistical programming language version 4.0.4 for further processing and analysis. Bivariable logistic regression was done to identify the relationship between each predictor and preterm birth. Variables with ($p \leq 0.25$) from the bivariable analysis were entered into a backward stepwise multivariable logistic regression model, and significant variables ($p < 0.05$) were retained in the multivariable model. Model accuracy and goodness of fit were assessed by computing the area under the ROC curve (discrimination) and calibration plot (calibration), respectively.

Setting and participants: This retrospective study was conducted among 1260 pregnant women who did prenatal care and finally delivered at Felege Hiwot Comprehensive Specialized Hospital, Bahir Dar city, Northwest Ethiopia, from January 30, 2019, to January 30, 2021.

Results: Residence, gravidity, haemoglobin < 11 mg/dl, early rupture of membranes, antepartum haemorrhage, and pregnancy-induced hypertension remained in the final multivariable prediction model. The AUC of the model was 0.816 (95% confidence interval: 0.779 – 0.856).

Conclusion: This study showed the possibility of predicting preterm birth using maternal characteristics during pregnancy. Thus, using this model could help to identify pregnant women at a higher risk of having a preterm birth to be linked to a center

Keywords: Prediction Model, Preterm birth, Risk score, Ethiopia

Strength and Limitations of the study

- ✓ An adequate number of participants with the outcome helped us to construct the model using a sufficient number of predictor variables and the inclusion of sensitivity analyses.
- ✓ Multiple imputation were used to address missing data, which has been shown to be a valid technique for dealing with missing data within logistic regression models, resulting in less bias than excluding all women with missing data.
- ✓ The prediction model is constructed from easily obtainable maternal characteristics that make it applicable in primary care settings.

- ✓ A single-site study, it is confined to a single area, which needs external validation before using it in another context.
- ✓ Furthermore, data were collected from each mother's card; due to this, some important variables were missed, such as previously highlighted factors of preterm birth in different studies.

Introduction

Preterm birth is described as babies that are born alive before the end of 37 weeks of pregnancy[1]. Preterm birth can be accidental (due to spontaneous preterm labor and/or preterm membrane rupture) or induced by the provider (by cesarean or labor induction)[2]. Most preterm births happen spontaneously[3].

An estimated 15 million babies worldwide are born too early per year. That is more than 1 in 10 infants. About 1 million newborns die per year because of preterm birth complications[4].

Across 184 countries, the rate of preterm birth ranges from 5% to 18% of babies born [5]. However, there are stark disparities in survival rates around the world. Half of the babies born at or below 32 weeks die in low-income settings due to a lack of practical, cost-effective, and critical care, such as comfort, breastfeeding assistance, basic infection care, and trouble Breathing[6].

Furthermore, the effect of preterm birth is also prolonged beyond the neonatal phase and throughout life[7]. Hence, the largest risk of severe health issues, including cerebral palsy, intellectual disability, chronic lung disease, and vision and hearing loss, is faced by babies born before maturity. This introduces a lifelong disability dimension. At some point in their lives, most people will face the struggles and potential disasters of preterm birth either directly in their families or indirectly through events for the nations[7, 8].

To alleviate this burden, in the past few decades, numerous methods have been attempted internationally, including in Ethiopia, to prevent and enhance the treatment of preterm births [9-11]. As part of the strategy, it is essential to diagnose or predict preterm birth earlier in pregnancy to take appropriate measures for high-risk groups. However, in most nations, predicting preterm birth is still largely based on subjective clinical experience. This approach may increase unnecessary hospital admissions and unnecessary but potentially harmful treatments, such as the use of steroids for the maturation of the fetal lung and tocolysis[12, 13].

There were clinical prediction models that aim to estimate the likelihood of preterm birth that include laboratory tests that are typically inaccessible in low-resource settings, such as fetal fibronectin, insulin-like growth factor binding protein-1 (IGFBP-1), interleukin-6, and placental alpha-macroglobulin-1[14-19]. Most current research on PTB prediction focuses on finding PTB risk factors using a hypothesis-testing methodology in highly controlled environments. PTB has been linked to a number of risk factors, including previous preterm labor, multiple gestation (carrying several children), and diabetes, problems with the cervix, uterus, or placenta, smoking, and infections [20-22]. However, women who have preterm delivery often have no known risk factors[23]. In addition, some of the predictors (such as prior PTB) do not apply for first-time mothers.

Predicting the risk of PTB in pregnant women has been the subject of numerous studies[24], but no model exists that is accurate enough to be used in clinical settings. Most research (e.g., cervical length or fetal fibronectin) have concentrated on predictors during the second or third trimester[25].These predictors, however, can only forecast PTB at intermediate risk and have only been shown to be reliable in high-risk populations. Unfortunately, the majority of women who give birth early have no evident risk factors, and more than half of PTBs happen in low-risk pregnancies, indicating the limited usefulness of using fetal fibronectin or cervical length in the general population[26].

Due to scarce resources, using readily available data to predict PTB seems appealing in low- and middle-income communities. But relatively few models have been made public. The considerable range in PTB occurrence across the globe, which suggests differences in exposure to psychosocial, sociodemographic, and medical risk factors as well as genetic variations, is also significant [27-29]. As a result, it is necessary to develop and evaluate PTB prediction models in various populations.

Therefore, developing and validating a risk prediction model for the prediction of preterm birth using maternal(clinical and nonclinical) characteristics based on the available measurements is paramount to allow early preterm birth interventions such as in utero transfer to tertiary care centers, appropriate corticosteroid administration while preventing excessive use, neuroprotective magnesium sulfate therapy, and antibiotic treatment in the event of infection[14, 30]

Methods and Materials

Study setting

This retrospective study was conducted among 1260 pregnant women who did prenatal care and finally delivered at Felege Hiwot Comprehensive Specialized Hospital, Bahir Dar city, Northwest Ethiopia, from January 30, 2019, to January 30, 2021. Bahir Dar is the capital city of Amhara national regional state and is found 575 km northwest of Addis Ababa. The hospital has currently a total of 1431 manpower (5 Obstetricians and Gynaecologists and 63 midwives among others) in different disciplines. It has a total of 500 formal beds, 11 wards (emergency ward and Inpatient wards such as Gynecological &Obstetric, Surgical, Orthopaedics, Medical, Pediatric, L&D, Eye unit, NICU, psychiatric, oncology, and 22 OPDS), 39 clinical and non-clinical departments /service units / providing laboratory, Diagnostic, curative & Rehabilitation service at outpatient & inpatient bases as well as disease prevention & health promotion services.

Sample size determination

The sample size required for model development was determined based on the minimum standard of 10 events per candidate predictor considered, according to the formula N = (n × 10)/I, where N is the sample size, n is the number of candidate predictor variables and I is the estimated event rate in the population[31]. Since there were 17 candidate predictors considered and 10 events per candidate predictor, the estimated number of events for the study was 170. Based on a study done on the prevalence of preterm birth in Debre Tabor hospital was 13%[32], so taking into account this the required sample size was calculated as follows, n= 170*100/13= 1308.

Study Design and Participants

The theoretical design of the present study was; the incidence of preterm birth as a function of multiple predictors during pregnancy. The source population of the study was all pregnant mothers who gave birth at FHCSH. To be included in this study, mothers must meet all of the following eligibility criteria; all medical records of mothers who gave live birth and had at least one ANC follow-up in FHCSH from January 30/2019 to January 30/2021.

Sampling method and procedures

A simple random sampling technique was employed to select participants using the medical registration number of a delivered mother from the delivery registration book. First, all mothers delivered at FHCSH in the last two years was identified from the delivery registration book. After that, records of mothers who met the inclusion criteria were included in the study. Subsequently, a sampling frame was prepared. Finally, the study unit was selected by using a computer-generated random number.

Data Collection

Outcome assessment: The outcome variable was attributed to women whose medical records indicated a physician or midwife diagnosis of preterm birth and delivery between 28 and 36 completed weeks of gestation. The gestational age (GA) was measured using either LNMP, which is found to be a more reliable measure of GA in a low-resource setting[33, 34], or an early ultrasound result(12 weeks).

Predictor assessment: Data were collected using a structured checklist through chart review. Checklists were developed after reviewing various relevant literatures [35-39]. It consists of socio-demographic (Maternal age, Residence), Maternal obstetric characteristics : (History of preterm birth, History of abortion, history of stillbirth, gravidity, Parity, Multiple pregnancy, APH, PROM, Gestational DM, and PIH), Maternal medical condition : (HGB level, Diabetic Mellitus, Chronic Hypertension, UTI and HIV).

Quality Assurance Mechanisms

To maintain the quality of data, the data collectors and supervisors were trained for a day on the objective of the study, the content of the checklists, and how to fill the checklists. Afterward, reviewing 15 charts medical records of mothers who gave birth at Felege Hiwot Comprehensive Specialized Hospital which is found in Northwest Ethiopia were done. After that, some adjustments (removing variables that were not available in the medical records of mothers) were done accordingly. The checklist was developed in English.

Data Processing and Analysis

Data were entered into a software application (EPI DATA, version 3.02) and was analyzed by using R statistical programming language version 4.0.4 for further processing and analysis. There were 13(1%), 2(0.2 %), 11 (0.9 %),15 (2.5%), 21 (1.7%) ,29(2.3%),20(1.6%) and 20 (1.6%) missing values for premature rupture of membranes , residence, chronic hypertension, multiple pregnancy gestational diabetes Mellitus, pregnancy-induced hypertension ,antepartum hemorrhage and hemoglobin respectively.

We assumed the data were missing at random, and we, therefore, performed a multivariate imputation by chained equations for all variables evaluated in the prediction model [40]. Sensitivity analysis was performed to assess whether the assumption of missing at random (MAR) is valid or not, and the results were reasonably comparable table **(1)**. Descriptive statistics including median, interquartile range (IQR), and percentages, were carried out.

Table 1. Sensitivity analysis of the model to predict preterm birth: Comparison of the regression coefficients, standard errors (SE), and p-values for complete case analysis (CCA) and multiple imputed data (MI).

Predicator variables	Complete case analysis			Multiple imputations		
	B	SE	P value	B	SE	P value
Chronic hypertension (yes)	0.7313	0.6297	0.24	0.581	0.6285	0.92
Residence (rural)	0.815	0.1946	<0.001	1.154	0.1958	<0.001
GDM(yes)	0.709	0.4028	0.07	0.472	0.4236	0.26
HGB(<11g/dl)	0.497	0.2185	0.02	0.642	0.2153	0.001
PROM (yes)	1.898	0.2080	<0.001	2.097	0.2129	<0.001
APH (yes)	1.194	0.2858	<0.001	1.298	0.2874	<0.001
PIH (yes)	1.353	0.2600	<0.001	1.368	0.2523	<0.001
Multiple pregnancy (yes)	0.539	0.3173	0.08	0.446	0.3257	0.17
Gravidity(primigravida)	0.426	0.1944	0.02	0.711	0.1976	<0.001

Model Development and Validation

For model development, bivariable logistic regression was done to obtain insight into the association between each potential predictor and preterm birth. Variables with ($p \leq 0.25$) from the bivariable analysis were entered into a backward stepwise multivariable logistic regression model, and significant variables ($p < 0.05$) were retained in the multivariable model. The results of significant predictors were reported as coefficients with 95% confidence intervals (CI). To check for the model accuracy and goodness of fit, we computed the area under the ROC curve (discrimination) and calibration plot (calibration) using "classifierplots" and "givitiR" packages of R respectively. The AUC ranged from 0.5 (no predictive ability) to 1 (perfect discrimination)[41]. The regression coefficients and their 95% confidence levels, and the AUC were adjusted for overfitting or over optimism using the bootstrapping technique. To make internal validation, we computed 1000 random bootstrap [42]samples with the replacement of all predictors in the data. The model's predictive performance after bootstrapping is considered as the performance that can be expected when the model is applied to future similar populations. To evaluate the clinical and public health impact of the model, we performed a decision curve analysis (DCA) [43] of standardized net benefits across a range of threshold probabilities (0 to 1). In the DCA, the model was compared with two extreme scenarios; "intervention for all" and "no intervention". In our case, the intervention considered is the referral of high-risk pregnant women to facilities where appropriate, corticosteroid administration, antibiotic treatment.

Risk Score Development

To construct an easily applicable preterm birth prediction score, we transformed each coefficient of the model into a rounded number by dividing it by the lowest coefficient. The number of points was subsequently rounded to the nearest integer. We determined the total score for each individual by assigning points for each variable present and adding them up. The score was transformed to dichotomous, allowing each pregnant woman to be classified as having a high or low risk of preterm birth. The receiver operating characteristic curve (ROC) was plotted and the area under the curve (AUC) was calculated to measure the discriminatory power of the scoring system.

Patient and public involvement

There was no direct interaction with patients in this study and no direct patient involvement in the design or conduct of this study.

Result

Demographic, Obstetric, and Clinical Characteristics of mothers

A total of 1260 study cards were reviewed from a sample of 1308, about 48 cards were not reviewed due to the outcome of intrauterine fetal death, and abortion. Table (**2**) shows the demographic, obstetric, and clinical characteristics of mothers who gave birth included in the analysis. The median age of the study participants was 26 years with IQR (24-30years); the majority of the participants 1086 (86.2%) were in the age group of 20-34 years. More than three-fourth of the participants 926 (73.49%) were urban residents. Of the total of mothers who delivered at FHCSH, more than two-thirds of 841 (66.7%) were multigravida. About parity, above, half of them713 (56.6%) were multipara. Concerning past obstetric history, 55 (6.5%) of them had a history of previous preterm birth, 76 (9%) of them had a history of stillbirth and 162 (19.3%) of them had a history of abortion.

Table 2. Demographic, obstetric, and clinical characteristics of mothers who gave birth at FHCSH , Northwest Ethiopia, 2021.

Characteristics	Category	Frequency	Percent
Gravidity	Primigravida	419	33.3
	Multigravida	841	66.7
Residence	Urban	926	73.5
	Rural	334	26.5
GDM	Yes	44	3.5
	No	1216	96.5
APH	Yes	84	6.7
	No	1176	93.3
PIH	Yes	110	8.73
	No	1150	91.27
HGB level	<11d/dl	236	18.7
	>=11g/dl	1024	81.3
Chronic hypertension	Yes	21	1.7
	No	1239	98.3
PROM	Yes	195	15.5

	No	1065	84.5
Multiple pregnancies	Yes	90	7.2
	No	1170	92.8

PROM: Premature rupture of membrane, HGB: hemoglobin, PIH: pregnancy-induced hypertension, APH: antepartum hemorrhage, GDM: gestational diabetes mellitus

Development of prediction model for preterm birth

Out of 1260 delivered neonates, 169 (13.4%) (95%, CI (11.6%, 15.4%) was preterm infants. The bivariable logistic regression analysis found several factors were eligible to be included in the prediction model. These variables were haemoglobin level, Gravidity, residence, gestational diabetes mellitus, APH, PIH, chronic hypertension, PROM, and multiple pregnancies. Using the results, a prediction model was developed, and equation for the prediction model was obtained table **(3).**

Table 3: Coefficients and risk scores of each predictor included in the model to predict preterm birth (*n* = 1260)

Predictors Variables*	**Multivariable analysis**			
	Original β (95 % CI)	Bootstrap β	P-value	Risk score
Residence (rural)	1.161 (0.780, 1.545)	1.148	<0.001	2
Gravidity (primigravida)	0.675 (0.291, 1.061)	0.666	0.01	1
PROM (yes)	2.081 (1.669 , 2.50)	2.051	<0.001	3
APH (yes)	1.364 (0.806 , 1.915)	1.348	<0.001	2
PIH (yes)	1.387 (0.887 , 1.879)	1.368	<0.001	2
HGB <11g/dl	0.676 (0.255 , 1.09)	0.677	<0.001	1

**Variables retained in the reduced model are; residence, APH, hemoglobin, PIH, gravidity, and PROM. Both backward and forward selection showed the same results. β after internal validation with bootstrapping is shown. Simplified risk score: we divided the coefficient of predictors included in the reduced model by the smallest (0.666). The probability or risk of preterm birth = 1/ (1 + exp – (-*

*3.517+ 1.148 * Residence (rural) + 0.666 *gravidity (primigravida) + 2.051*PROM (yes) + 1.348 * APH (yes) + 1.387*PIH +0.677*HGB (<11g/dl).*

The AUC of the final reduced model was 0.816 (95% confidence interval: 0.779 – 0.856) **(Figure 1a)**. The calibration test had a p-value of 0.492, indicating that the model does not misrepresent the data or the calibration of the model was visually accurate since the observed and predicted probabilities were similar **(Figure 1b).**

In addition, to verify whether any maternal characteristics were used as a specific predictor of preterm birth we performed an ROC analysis. The analysis indicated that, residence (AUC=0.604, 95% CI 0.564 to 0.643), gravidity (AUC=0.59, 95% CI 0.571 to 0.628), PROM (AUC=0.580, 95% CI 0.544 to 0.616), APH (AUC= 0.695, 95% CI 0.661 to 0.729), PIH (AUC= 0.721, 95% CI 0.685 to 0.757), and HGB (AUC=0.630, 95% CI 0.591 to 0.668) emerged as better predictors of preterm birth (**Figure 2**).

Validation of the model with the bootstrap technique showed hardly any indication of undue influence by particular observations, with an optimism coefficient of 0.085, resulting AUC of 0.789 (corrected 95% CI: 0.748–0.83).

Using the coefficient (β), the predicted risk cutoff point was a probability of (SpEqualSe > 0.1320), the model has a sensitivity of 75.74%, specificity of 72.87%, a positive predictive value of 30.2%, and a negative predictive value of 95.1%.

When applying DCA, we first evaluate whether our model understudy has a higher net benefit than the default strategies (referring all and none). This model outperforms the default strategies across the relevant threshold range. The model has the highest net benefit across the entire range of threshold probabilities, which indicates that the model has the highest clinical and public health value. Hence, the referral decision made using the model has a higher net benefit than not referring at all or referring all regardless of their risk threshold as shown in ***figure (3).***

Risk Classification Using a Simplified Risk Score

We created a simplified risk score from the model for practical use. The reduced model's prediction score was simplified by rounding all regression coefficients. The simplified score had a considerably comparable prediction accuracy to the original β coefficients, with an AUC of 0.786 (95%CI: 0.729–0.827) **(figure 4).** The possible minimum and maximum scores a mother can have are 0 and 11, respectively.

Using "SpEqualSe", the suggested threshold score to predict preterm birth using risk scores is ≥3with a sensitivity of 75.14 % and specificity of 67.46% table (**4**).

When dichotomized into low risk (<3) and high risk (≥3) based on the risk score, 278 (14.36%) were categorized as high risk and 982 (77.9%) as low risk for preterm birth.

Table 4: Risk classification of preterm birth using simplified prediction score (n = 1260)

Score*(risk category)	Prediction Model Based on Maternal Characteristics	
	Number of mothers	Incidence of preterm birth
<3 (Low)	982 (77.9%)	72 (7.9%)
>=3 (High)	278 (14.36%)	97 (53.59%)
Total	1260 (100%)	169 (13.4%)

**** Score = (2*PIH) + (3*PROM) + (hemoglobin < 11 mg/dl) + 2*residence + (2*APH) + gravidity***.

Discussion

In this study, the incidence of preterm birth was found to be 13.4%. Maternal characteristics were identified in this retrospective study to build a preterm birth prediction risk score. We intended to employ maternal features that are easily accessible and pertinent to clinical practice in countries with constrained resources, including Ethiopia. These nations may not have the financial resources to pay for ultrasound exams and laboratory tests. The optimal combination of maternal factors to predict preterm birth includes residency, gravidity, and hemoglobin < 11 mg/dl, early rupture of membranes, antepartum hemorrhage, and pregnancy-induced hypertension, according to the prediction model. The model has an AUC of 0.816 (95%CI: 0.776 – 0.856). Predicting the probability of preterm birth in pregnant women is essential to take appropriate measures accordingly. Identifying women at risk of preterm birth is an important task for clinical care providers. However, in low and middle-income countries, there are only a few methods available for reliably predicting actual preterm labor in women. Previously, the focus of the research was to explain the maternal and fetal determinants of preterm birth. In recent years, the focus shifted to predicting preterm birth optimally using a combined set of characteristics.

Without any advanced laboratory or imaging testing, this study measured the predicted performance of a model based on maternal features during pregnancy. Furthermore, we discovered that utilizing SpEqualSe as an optimal cut point, the sensitivity and specificity of this prediction model achieved 75.14 percent and 67.46 percent, respectively, at the score threshold of 3.

In our study, a combination (residency, gravidity, hemoglobin < 11 mg/dl, early rupture of membranes, antepartum hemorrhage, and pregnancy-induced hypertension) of maternal characteristics resulted in an AUC of 0.816 (95%CI: 0.776 – 0.856), has an excellent accuracy according to diagnostic accuracy classification[44].

We found that early rupture of membrane is strong predictors of preterm birth. Similar evidence was found in different studies [36, 37, 45, 46]. The effect of a burst membrane on uterine contraction could explain this. Existing scientific evidence confirms that when a membrane ruptures, natural uterotonic chemicals are released, and these uterotonic chemicals drive uterine contraction, resulting in PTB. This finding suggested that due attention should be given to women with premature rupture of membrane.

In our study, pregnancy-induced hypertension is strong predictors of preterm birth. Similar studies have demonstrated that pregnancy-induced hypertension was predictive of subsequent preterm birth[47, 48]. This could be related to vascular injury to the placenta caused by pregnancy-induced hypertension issues or iatrogenesis caused by the severity of hypertension or its complications. As a result, the oxytocin receptors are activated, resulting in preterm labor and delivery. Or else this conclusion could be explained by current scientific evidence suggesting that PIH is linked to vascular and placental injury, which causes oxytocin receptors to be activated, resulting in PTB. Therefore, it is imperative to identify populations at risk pregnancy-induced hypertension and introduce risk lowering interventions.

Another strong predictor of preterm birth is the place of residence. Existed evidence shows that there is an association between preterm birth and rural residence [49-53]. This gap may be explained by the greater accessibility and availability of maternal health service in metropolitan regions. It has long been understood that social deprivation and the nuanced interactions between them affect prenatal outcomes, including premature birth[54]. Hence, accessing maternal health services targeted to rural women could improve prenatal outcomes including the risk of preterm birth.

Antepartum hemorrhage is the predictor of preterm birth which is supported by different studies[55]. Identification of groups at risk for antepartum hemorrhage and the introduction of risk-reducing measures are therefore essential. Other predictors of preterm birth are gravidity and hemoglobin <11 g/dl (anemia) which is in line with different studies[32, 56]. The molecular factors that could explain how anemia, iron deficiency, or both, could result in preterm delivery. In reality, a number of plausible molecular processes have linked anemia to a higher risk of premature birth. Accordingly, maternal and fetal stress can be caused by anemia (by resulting in hypoxia) and iron deficiency (by increasing serum nor-epinephrine concentrations), which in turn induces the production of corticotrophin-releasing hormone (CRH). Additionally, iron deficiency may raise the risk of maternal infections, which can again boost the synthesis of CRH. High levels of CRH are known to be a risk factor for PTB since they increase the likelihood of PTB [57]. Thus, we can conclude that, in order to prevent PTB, routine ANC services need to place a greater emphasis on anemia prevention.

A study conducted in China showed that a model developed using advanced maternal age, lower maternal height, history of preterm delivery, amount of vaginal bleeding during pregnancy, and lack of folic acid intake before pregnancy for the prediction of overall preterm birth with AUC of (0.6)[58]. Which had lower discriminatory performance than the present study, this difference may be due to some of the predictors they used such as lower maternal height, lack of folic acid intake before pregnancy, and advanced maternal age. However, the predictors they used such as lack of folic acid intake before pregnancy are not easily obtainable information in routine clinical practice, which makes their model less practical in our setting. This prediction model constitutes variables that are easily obtainable and have reasonable accuracy to be used by both mid-and lower-level health professionals in primary care settings. Among the maternal characteristics included in our model, five can be easily found by history taking and one by test for hemoglobin. The model's accuracy is consistent with a retrospective study done in China that established a preterm birth prediction model based on maternal characteristics, including demographics and clinical characteristics, and a model with predictors (gravidity, educational status, residency, history of preterm birth, twin pregnancy, pre-gestational diabetes mellitus (type I or II), chronic hypertension, and place of birth) with AUC of 0.749 (95%CI: 0.732–0.767) [48].

On the other hand, a model incorporating four predictors (cervical length at admission, gestational age, amniotic fluid, glucose, and IL-6) has an area under the curve (AUROC) of 0.86[59] and similarly, the combination of biophysical, biochemical, immunological, microbiological, fetal cell, exosomal, or cell-free RNA at different gestational ages, integrated as part of a multivariable predictor model may be necessary to advance our attempts to predict sPTL and preterm birth. In the prediction of spontaneous preterm birth within 48 hours, a prognostic model including qfFN and clinical risk factors showed excellent results[60, 61]. Both models have higher discriminatory performance. The reason for the lower discriminatory performance in our study compared to the studies described above could be because we used secondary data available from the register and as this dataset is limited and some variables that require advanced laboratory tests were not included in the model.

Hence, predictors necessitate laboratory testing, which is often unavailable in low-resource settings. As a result, such predictors are difficult to come by in ordinary clinical and public health practice, making the model less useful.

A study conducted in the UK found that data on maternal characteristics and obstetric history at 11–13 weeks of gestation were predictive of spontaneous early preterm delivery; this model had an AUC of 0.67[62] which had lower discriminatory performance than the present study. This difference may be the difference in the study population.

A model that predicts a risk of preterm delivery in women with multiple pregnancy incorporating previous preterm delivery, monochorionicity, smoking, educational level, and triplet pregnancy for preterm and very preterm delivery had a c-index of 0.68 (95% CI 0.63 to 0.72) and 0.68 (95% CI 0.62 to 0.75) respectively[63]. It had lower discriminatory performance than the present study. This might be due to the study population difference. In the present study, the study populations were both women who had multiple pregnancies and singleton pregnancy.

In our prediction score, using 3 as a cutoff point has an acceptable level of specificity, sensitivity, PPV, and NPV to predict preterm birth. It is also possible to shift the cutoff point to increase either of the accuracy measures depending on the program aim and availability of resources.

The strength of the study was using an adequate number of participants with the outcome, which helped us to construct the model using a sufficient number of predictor variables. In addition, our prediction model was constructed from easily obtainable maternal characteristics that make it applicable in primary care setting and multiple imputation were used to address missing data, which has been shown to be a valid technique for dealing with missing data within logistic regression models, resulting in less bias than excluding all women with missing data.

However, the findings from this study should be interpreted with the perspective of the following limitations. As a single-site study, it is confined to a single area, which needs external validation before using it in another context. Furthermore, data were collected from each mother's card; due to this, some important variables were missed, such as previously highlighted factors with preterm birth in different studies.

Conclusions and recommendations

This study shows the possibility of predicting preterm birth using a simple prediction model constructed from maternal characteristics. Thus, the optimal combination of maternal characteristics such as residence, gravidity, haemoglobin < 11 mg/dl, premature rupture of membrane, antepartum haemorrhage, and pregnancy-induced hypertension shows the possibility of predicting preterm birth using a simple prediction model constructed from maternal characteristics. In addition, risk score calculations based on a combination of predictors was effective and had comparable accuracy with the model-based approach of the original β coefficients. This score may assist in clinical decision-making. In addition, incorporating this convenient and easily applicable score in the health care system to be used by clinicians to inform pregnant mothers about the future course of their outcome after external validation. Doing further research is needed to validate the prediction tool using prospective follow-up studies in another context before introducing it to clinical and public health practices.

Data Sharing Statement

Data will be available upon request from the corresponding author.

Author Contributions: S.F.F. conceived the study and wrote the manuscript. Z.A.A, S.F.F, G.T.W, A.K.Y, and A.M.D, all contribute to data analysis, study design, and supervision of data collection. All authors participated in manuscript revision for intellectual content and approval of the final version. All authors have read and agreed to the published version of the manuscript.

Funding statement: This work was supported by Bahir Dar University grant number (25000ETB).

Competing interest's statement: The author reports no conflicts of interest in this work.

Ethical approval

Ethical clearance was obtained from the Institutional Review Board (IRB) of Bahir Dar University, College of Medicine and Health Sciences with Protocol number 083/ 2021) on

February 26, 2021. It is a retrospective study of medical records and personal identifiers were not used on the data collection checklist. So, the IRB waived the requirement for informed consent from each participant. Confidentiality was maintained by omitting the personal identifier of the participant during the data collection procedure and the information was used only for research purposes. Data were collected from the register, which was kept in a secure place and all data were fully anonymized before we accessed them. After the collection of data, all patient records and patient cards were placed back in a secure place. Data were entered into a password-protected computer.

References

1. Organization WH: **Preterm birth and low birth weight**. 2020.
2. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet 2008;371(9606):75e84
3. Althabe F: **Born too soon: the global action report on preterm birth**: World Health Organization; 2012.
4. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet. 2016;388(10063):3027-35.
5. World Health Organization. WHO fact sheet on preterm birth. Available from: http://www.who.int/mediacentre/factsheets/fs363/en/
6. Organization WH: **WHO fact sheet: Preterm birth**. *World Health Organization, Geneva, Switzerland http://www who int/mediacentre/factsheets/fs363/en/ Accessed* 2018, **26**.
7. Li S, Xi B: **Preterm birth is associated with risk of essential hypertension in later life**. *International journal of cardiology* 2014, **172**(2):e361-e363.
8. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller A-B, Kinney M, Lawn J: **Born too soon: the global epidemiology of 15 million preterm births**. *Reproductive health* 2013, **10**(S1):S2.

9. Soon BT: **The global action report on preterm birth**. *Geneva: World Health Organization* 2012.

10. Griffin JB, Jobe AH, Rouse D, McClure EM, Goldenberg RL, Kamath-Rayne BD: **Evaluating WHO-recommended interventions for preterm birth: a mathematical model of the potential reduction of preterm mortality in Sub-Saharan Africa**. *Global Health: Science and Practice* 2019, **7**(2):215-227.

11. Victora CG, Rubens CE, Group GR: **Global report on preterm birth and stillbirth (4 of 7): delivery of interventions**. *BMC Pregnancy and Childbirth* 2010, **10**(S1):S4.

12. Kemp M, Newnham J, Challis J, Jobe A, Stock S: **The clinical use of corticosteroids in pregnancy**. *Human reproduction update* 2016, **22**(2):240-259.

13. Lorthe E, Goffinet F, Marret S, Vayssiere C, Flamant C, Quere M, Benhammou V, Ancel P-Y, Kayem G: **Tocolysis after preterm premature rupture of membranes and neonatal outcome: a propensity-score analysis**. *American Journal of Obstetrics and Gynecology* 2017, **217**(2):212. e211-212. e212.

14. Oskovi Kaplan ZA, Ozgu-Erdinc AS: **Prediction of Preterm Birth: Maternal Characteristics, Ultrasound Markers, and Biomarkers: An Updated Overview**. 2018, **2018**:8367571.

15. Lucaroni F, Morciano L, Rizzo G: **Biomarkers for predicting spontaneous preterm birth: an umbrella systematic review**. 2018, **31**(6):726-734.

16. Lee KA, Chang MH, Park M-H, Park H, Ha EH, Park EA, Kim YJ: **A model for prediction of spontaneous preterm birth in asymptomatic women**. *Journal of Women's Health* 2011, **20**(12):1825-1831.

17. Georgiou HM, Di Quinzio MK, Permezel M, Brennecke SP: **Predicting preterm labour: current status and future prospects**. *Disease markers* 2015, **2015**.

18. Shennan AH: **Prediction and prevention of preterm birth: a quagmire of evidence**. *Ultrasound Obstet Gynecol* 2018, **51**(5):569-570.

19. Son M, Miller ES: **Predicting preterm birth: Cervical length and fetal fibronectin**. *Semin Perinatol* 2017, **41**(8):445-451.

20. Cobo T, Kacerovsky M, Jacobsson B: **Risk factors for spontaneous preterm delivery**. *International Journal of Gynecology & Obstetrics* 2020, **150**(1):17-23.

21. Ren H: **Du M. Role of maternal periodontitis in preterm birth. Front Immunol. 2017; 8: 139**. In.; 2017.

22. Oskovi Kaplan ZA, Ozgu-Erdinc AS: **Prediction of preterm birth: maternal characteristics, ultrasound markers, and biomarkers: an updated overview**. *Journal of pregnancy* 2018, **2018**.

23. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller A-B, Narwal R, Adler A, Garcia CV, Rohde S, Say L, Lawn JE. 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. The Lancet. 2012;379(9832):2162–72. https://doi.org/ 10.1016/s0140-6736(12)60820-4.

24. Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratinam S, Khan KS, Mol BWJ, Pajkrt E, Moons KG, Schuit E: Pro**gnostic models in obstetrics: available, but far from applicable. A**merican *journal of obstetrics and gynecology 201*6, 214**(1)**:79-90. e36.

25. Sananes N, Langer B, Gaudineau A, Kutnahorsky R, Aissi G, Fritz G, Boudier E, Viville B, Nisand I, Favre R: Pre**diction of spontaneous preterm delivery in singleton pregnancies: where are we and where are we going? A review of literature. J**o*urnal of Obstetrics and Gynaecology 201*4, 34(**6)**:457-461.

26. Catley C, Frize M, Walker RC, Petriu DC: Pre**dicting high-risk preterm birth using artificial neural networks. I**EE*E Transactions on information technology in biomedicine 200*6, 10(**3)**:540-549.

27. Raglan GB, Lannon SM, Jones KM, Schulkin J: Rac**ial and ethnic disparities in preterm birth among American Indian and Alaska Native women. M**at*ernal and child health journal 201*6, 20(**1)**:16-24.

28. Culhane JF, Goldenberg RL: Rac**ial disparities in preterm birth. I**n: Sem*inars in perinatology: 2011: E*lsevier; 2011: 234-239.

29. York TP, Strauss III JF, Neale MC, Eaves LJ: Rac**ial differences in genetic and environmental risk to preterm birth. P**lo*S one 201*0, 5(8**)**:e12391.

30. Koullali B, Oudijk M, Nijman T, Mol B, Pajkrt E: Ris**k assessment and management to prevent preterm birth. I**n: Sem*inars in Fetal and Neonatal Medicine: 2016: E*lsevier; 2016: 80-88.

31. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A s**imulation study of the number of events per variable in logistic regression analysis. J** C*lin Epidemiol 199*6, 49(**12)**:1373-1379.

32. Mekonen DG, Yismaw AE, Nigussie TS, Ambaw WM: Pro**portion of Preterm birth and associated factors among mothers who gave birth in Debretabor town health institutions, northwest, Ethiopia. B**MC *Research Notes 201*9, 12(**1)**:2.

33. Tigist B, Abdela A, Zenebe G: Pre**term birth and associated factors among mothers who gave birth in Debre Markos Town Health Institutions. I**ns*titutional Based Cross sectional study 201*3.

34. Rosenberg RE, Ahmed ANU, Ahmed S, Saha SK, Chowdhury MA, Black RE, Santosham M, Darmstadt GL: Det**ermining gestational age in a low-resource setting: validity of last menstrual period. J**ou*rnal of health, population, and nutrition 200*9, 27(**3)**:332.

35. Wassie M, Manaye Y, Abeje G, Tifrie M, Worku G: Det**erminants of Preterm Birth among Newborns Delivered in Bahir Dar City Public Hospitals, North West Ethiopia. 2**020.

36. Woday A, Muluneh MD, Sherif S: Det**erminants of preterm birth among mothers who gave birth at public hospitals in the Amhara region, Ethiopia: A case-control study. P**lo*S one 201*9, 14(**11)**:e0225060.

37. Wudie F, Tesfamicheal F, Fisseha H, Weldehawaria N, Misgena K, Alema H, Gebregziabher Y, Fisseha G, Woldu M: Det**erminants of preterm delivery in the central zone of Tigray, northern Ethiopia: A case-control study. S**ou*th African Journal of Child Health 201*9, 13(**3)**:108-114.

38. Woldeyohannes D, Kene C, Gomora D, Seyoum K, Assefa T: Fac**tors Associated with Preterm Birth among Mothers Who gave Birth in Dodola Town Hospitals, Southeast Ethiopia: Institutional Based Cross Sectional Study. C**li*nics Mother Child Health 201*9, 16(**31**7):2.

39. Muchie KF, Lakew AM, Teshome DF, Yenit MK, Sisay MM, Mekonnen FA, Habitu YA: Epi**demiology of preterm birth in Ethiopia: systematic review and meta-analysis. B**MC *pregnancy and childbirth 202*0, 20(**1)**:1-12.

40. Kwak SK, Kim JH: Sta**tistical data preparation: management of missing values and outliers. K**or*ean journal of anesthesiology 201*7, 70(**4)**:407.

41. Grobbee DE, Hoes AW. Clinical epidemiology: principles, methods, and applications for clinical research: Jones & Bartlett Publishers; 2014.

42. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 2012;98:683-90. [Crossref] [PubMed]

43. Vickers AJ, Elkin EB: Dec**ision curve analysis: a novel method for evaluating prediction models. M**ed *Decis Making 200*6, 26(**6)**:565-574.

44. Mandrekar JN: Rec**eiver Operating Characteristic Curve in Diagnostic Test Assessment. J**ou*rnal of Thoracic Oncology 201*0, 5(9**)**:1315-1316.

45. Brhane M, Hagos B, Abrha MW, Weldearegay HG: Does **short inter-pregnancy interval predicts the risk of preterm birth in Northern Ethiopia? 20**19, 12(**1)**:405.

46. Sifer S, Kedir B, Demisse G, Sisay Y: Det**erminants of preterm birth in neonatal intensive care units at public hospitals in Sidama zone, South East Ethiopia; case control study. J** P*ediatr Neonatal Care 201*9, 9(6**)**:180-186.

47. Xu X, Tan H, Zhou S, He Y, Shen L, Liu Y, Hu L, Wang X, Li X: Study on the application of Back-Propagation Artificial Neural Network used the model in predicting preterm birth. Zhonghua liu Xing Bing xue za zhi= Zhonghua Liuxingbingxue Zazhi 2014, 35(9):1028-1031.

48. Chen M, Xie N, Liang Z, Qian T, Chen D: Early Prediction Model for Preterm Birth Combining Demographic Characteristics and Clinical Characteristics. 2020.

49. Kebede EB, Terfa YB, Geleta BA, Akuma AO: Predictors of preterm birth in Jimma town public hospitals, Jimma, Ethiopia. Journal of Pediatric and Neonatal Individualized Medicine (JPNIM) 2021, 10(1):e100125-e100125.

50. Belaynew W, Teumay A, Getachew G, Mohamed K: Effects of inter pregnancy interval on preterm birth and associated factors among postpartum mothers who gave birth at Felege Hiwot referral hospital. World J Pharm Pharm Sci 2015, 4(4):12-25.

51. Muhumed II, Kebira JY, Mabalhin MO: Preterm Birth and Associated Factors Among Mothers Who Gave Birth in Fafen Zone Public Hospitals, Somali Regional State, Eastern Ethiopia. Research and Reports in Neonatology 2021, 11:23-33.

52. Zhang Y-P, Liu X-H, Gao S-H, Wang J-M, Gu Y-S, Zhang J-Y, Zhou X, Li Q-X: Risk factors for preterm birth in five Maternal and Child Health hospitals in Beijing. PloS one 2012, 7(12):e52780.

53. Wagura P: factors associated with preterm birth at kenyatta national Hospital. BMC Pregnancy Childbirth (18).

54. Panaretto K, Lee H, Mitchell M, et al. Risk factors for preterm, low birth weight and small for gestational age birth in urban Aboriginal and Torres Strait Islander women in Townsville. Aust N Z J Public Health. 2006;30(2):163–170. doi:10.1111/j.1467-842X.2006.tb00111.

55. Olusanya BO, Ofovwe GE: Predictors of preterm births and low birthweight in an inner-city hospital in sub-Saharan Africa. Maternal and child health journal 2010, 14(6):978-986.

56. Mekuriyaw AM, Mihret MS, Yismaw AE: Determinants of Preterm Birth among Women Who Gave Birth in Amhara Region Referral Hospitals, Northern Ethiopia, 2018: Institutional Based Case Control Study. International Journal of Pediatrics 2020, 2020.

57. !!! INVALID CITATION !!! .

58. He J-R, Ramakrishnan R, Lai Y-M, Li W-D, Zhao X, Hu Y, Chen N-N, Hu F, Lu J-H, Wei X-L: Predictions of preterm birth from early pregnancy characteristics: born in guangzhou cohort study. Journal of clinical medicine 2018, 7(8):185.

59. Cobo T, Aldecoa V, Figueras F, Herranz A, Ferrero S, Izquierdo M, Murillo C, Amoedo R, Rueda C, Bosch J: Development and validation of a multivariable prediction model of

spontaneous preterm delivery and microbial invasion of the amniotic cavity in women with preterm labor. Americ*an Journal of Obstetrics and Gynecology 2020.*

60. Lamont R, Richardson L, Boniface J, Cobo T, Exner M, Christensen I, Forslund S, Gaba A, Helmer H, Jørgensen J: Commen**tary on a combined approach to the problem of developing biomarkers for the prediction of spontaneous preterm labor that leads to preterm birth. Plac**en*ta 2020.*

61. Stock SJ, Horne M, Bruijn M, Morris R, Dorling J, Jackson L, Chandiramani M, David AL, Khalil A, Shennan A: 793: A **new prediction model for birth within 48 hours in women with preterm labour symptoms. Amer**ic*an Journal of Obstetrics & Gynecology 2020,* 222(1)**:S5**02.

62. Beta, J.; Akolekar, R.; Ventura, W.; Syngelaki, A.; Nicolaides, K.H. Prediction of spontaneous preterm delivery from maternal factors, obstetric history and placental perfusion and function at 11–13 weeks. Prenat. Diagn. 2011, 31, 75–83

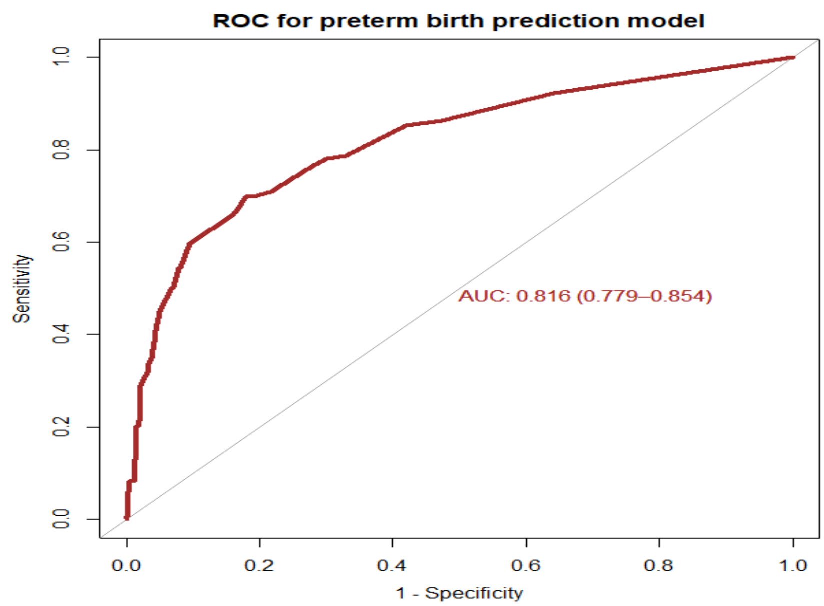
63. van de Mheen L, Schuit E, Lim AC, Porath MM, Papatsonis D, Erwich JJ, van Eyck J, Van Oirschot CM, Hummel P, Duvekot JJ: Predicti**on of preterm birth in multiple pregnancies: development of a multivariable model including cervical length measurement at 16 to 21 weeks' gestation. Journa**l *of obstetrics and gynaecology Canada 2014, 36*(4):30**9-**319.

Figure 1: (a) Area under the ROC curve for the prediction model, and (b) Predicted versus observed preterm birth probability in the sample. This analysis includes mothers who gave birth at FHCSH, 2021(n = 1260). Calibration plot created using "givitiCalibrationBelt" in R programming.

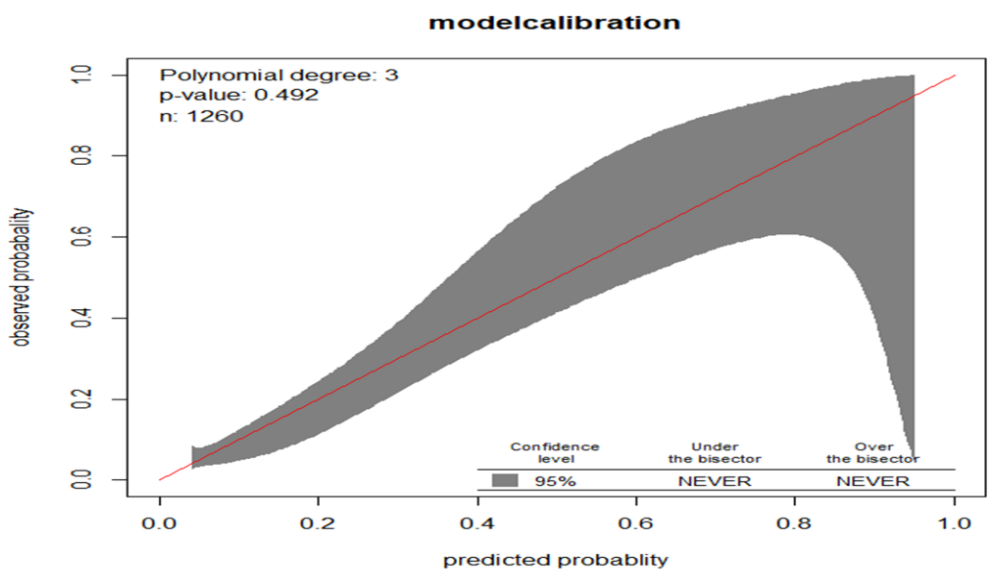
Figure 2: Receiver operating characteristic curve of maternal parameters for prediction of postpartum glucose intolerance. Residence, PROM, APH, PIH, HGB and Gravidity.

Figure 3: A decision curve plotting the net benefit of the model against threshold probability.

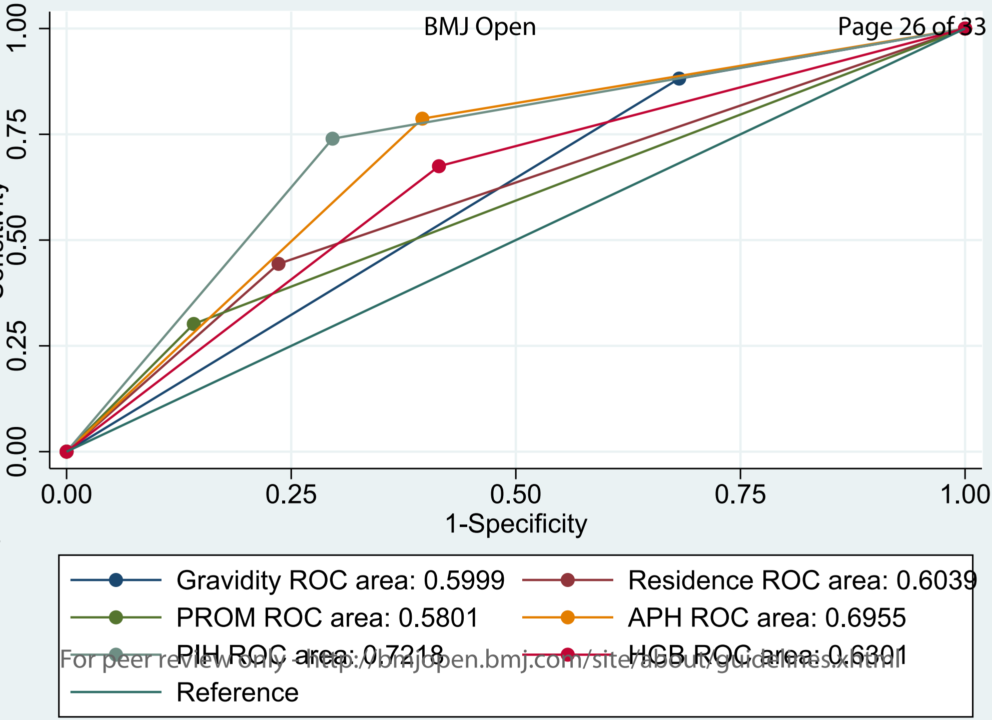
Figure 4: Area under the ROC curve for the simplified risk score to predict the risk of preterm birth among mothers who gave birth at FHCSH, 2021.

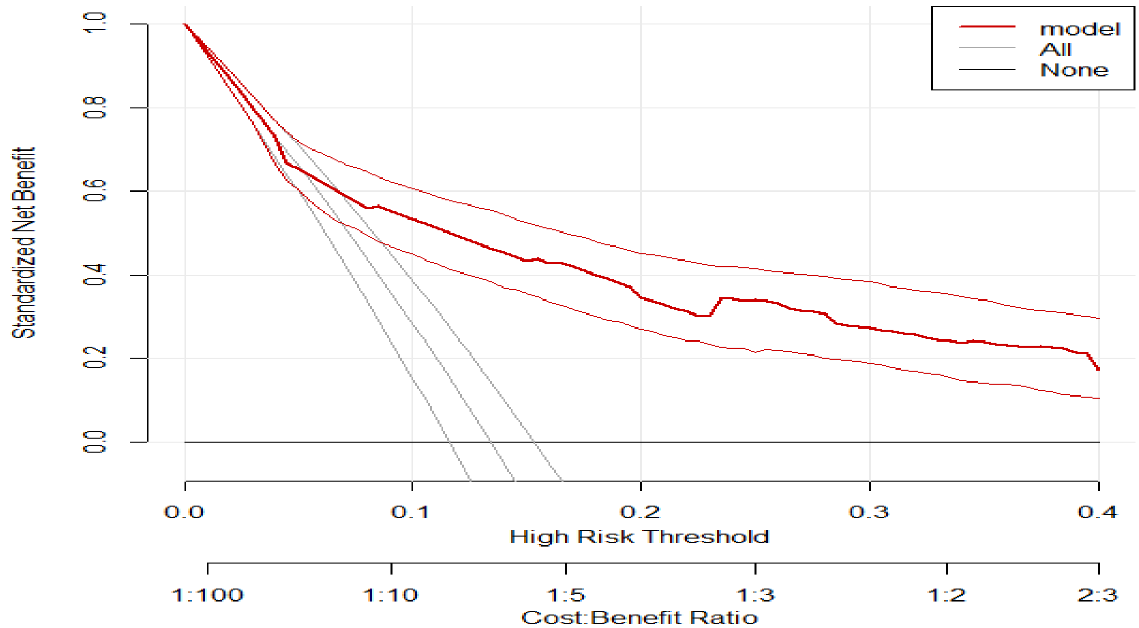

ROC for preterm birth prediction model
Sensitivity
1 - Specificity
AUC: 0.816 (0.779–0.854)
0.0
0.2
0.4
0.6
0.8
1.0

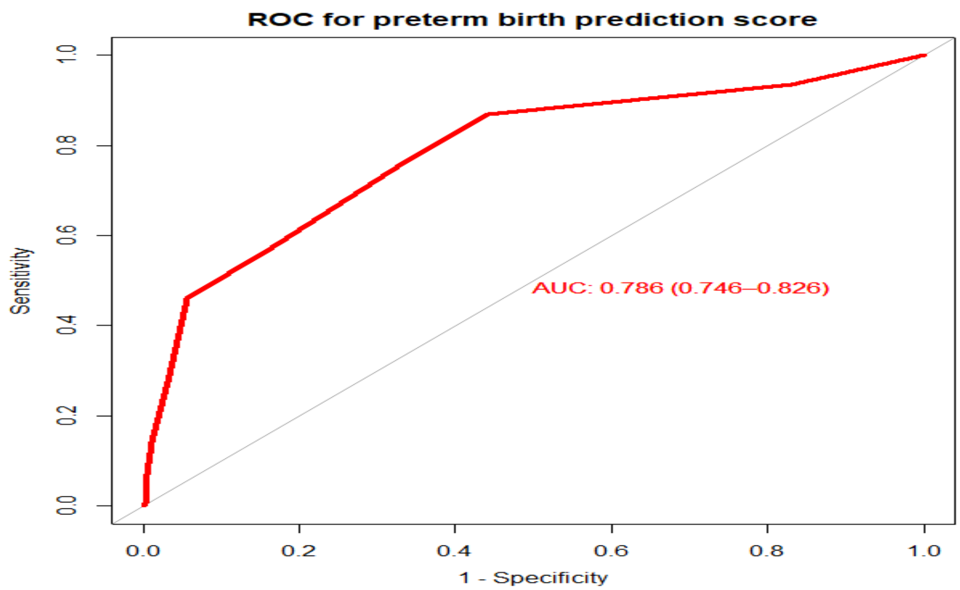
a)


modelcalibration
Polynomial degree: 3
p-value: 0.492
n: 1260
observed probabality
predicted probablity
Confidence level
Under the bisector
Over the bisector
95%
NEVER
NEVER
0.0
0.2
0.4
0.6
0.8
1.0

b)


1.00
0.75
0.50
0.25
0.00
Sensitivity
0.00
0.25
0.50
0.75
1.00
1-Specificity
Gravidity ROC area: 0.5999
Residence ROC area: 0.6039
PROM ROC area: 0.5801
APH ROC area: 0.6955
PIH ROC area: 0.7218
HGB ROC area: 0.6301
Reference


model
All
None
Standardized Net Benefit
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
High Risk Threshold
1:100
1:10
1:5
1:3
1:2
2:3
Cost:Benefit Ratio


ROC for preterm birth prediction score
Sensitivity
1 - Specificity
AUC: 0.786 (0.746–0.826)
0.0
0.2
0.4
0.6
0.8
1.0

The RECORD statement – checklist of items, extended from the STROBE statement that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Line 1-59	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Line 1-59
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Line 60-100		
Objectives	3	State specific objectives, including any prespecified hypotheses	Line 96-98		
Methods					
Study Design	4	Present key elements of study design early in the paper	Line 103		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Line 103-113 Line 123-128		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the	Line 123-128	RECORD 6.1: The methods of study population selection (such as codes or	Line 129-135

		sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Line 137-147	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Line 137-147
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Line 142-147		
Bias	9	Describe any efforts to address potential sources of bias	Line 123-141		
Study size	10	Explain how the study size was	Line 114-122		

		arrived at			
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Line 155-198		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Line 155-198		
Data access and cleaning methods		Line 148-154		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Line 130-132
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-	Line 130-132

				level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	Line 130-135	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Line 130-135
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	Line 232-241		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	Line 247		
Main results	16	(a) Give unadjusted estimates	Line 246-286		

		and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Line 162-166		
Discussion					
Key results	18	Summarise key results with reference to study objectives	Line 351-362		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Line 50-59	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	………………
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Line 351-362		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Line 351-362		

Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Line 369		
Accessibility of protocol, raw data, and programming code		Line 364		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Line 364

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.