Risk factors associated with preterm birth: identification, prediction and evaluation in the BRISA cohort

Gabriel Bazo, Ricardo Zorzetto N. Vêncio, Paulo Ricardo H. Rocha, Ricardo Cavalli, Alexandre Ferraro, Heloisa Bettiol, Marco Antonio Barbieri

https://doi.org/10.1590/SciELOPreprints.7882

Submitted on: 2024-01-11
Posted on: 2024-01-15 (version 1)
(YYYY-MM-DD)

Title

Risk factors associated with preterm birth: identification, prediction and evaluation in the BRISA cohort

Authors

Gabriel Bazo MSc[1], Ricardo Z. N. Vêncio PhD[2], Paulo R. H. Rocha PhD[1], Ricardo C. Cavalli PhD[3], Alexandre A. Ferraro PhD[4], Heloisa Bettiol PhD[1] and Marco A. Barbieri PhD[1]

[1]Department of Pediatrics, Ribeirão Preto Medical School, University of São Paulo, Brazil.
[2]Department of Computing and Mathematics, Faculty of Philosophy, Science and Letters of Ribeirão Preto, University of São Paulo, Brazil.
[3]Department of Gynecology and Obstetrics, Ribeirão Preto Medical School, University of São Paulo, Brazil.
[4]Department of Pediatrics, Medical School, University of São Paulo, Brazil.

Corresponding Author

Gabriel Bazo, Department of Pediatrics, Ribeirão Preto Medical School, University of São Paulo – Brazil, Av. Bandeirantes 3900, email: gabrielbazo@usp.br; phone: +55 163315-3316

Funding: This study was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil (Grant 88882.378513/2019-01). The BRISA study was funded by the São Paulo Research Foundation (FAPESP) (Grant 2008/53593-0).

Conflicts of interest: None declared.

Ethics approval: This study was approved by the Research Ethics Committee of Hospital das Clínicas, Ribeirão Preto Medical School (process number 3.842.040). The BRISA study was approved by the same Ethics Committee (process number 11157/2008).

Authors' contributions:

Gabriel Bazo: Conceptualization, formal analysis, investigation, methodology, visualization, writing – original draft, and writing – review & editing.
https://orcid.org/0000-0002-2205-9420

Ricardo Zorzetto N. Vêncio: Conceptualization, formal analysis, investigation, methodology, supervision, visualization, writing – original draft, and writing – review & editing.
https://orcid.org/0000-0003-0425-7877

Paulo Ricardo H. Rocha: Methodology, visualization, writing – original draft, and writing – review & editing.
https://orcid.org/0000-0002-4238-9603

Ricardo C. Cavalli: Methodology, visualization, and writing – review & editing.
http://orcid.org/0000-0001-5010-4914

Alexandre A. Ferraro: Methodology, visualization, and writing – review & editing.
https://orcid.org/0000-0003-1439-8091

Heloisa Bettiol: Conceptualization, methodology, visualization, writing – original draft, and writing – review & editing.
https://orcid.org/0000-0001-8744-4373

Marco Antonio Barbieri: Conceptualization, formal analysis, investigation, methodology, project administration, resources, supervision, visualization, writing – original draft, and writing – review & editing.
https://orcid.org/0000-0001-8060-1428

ABSTRACT

Problem: Preterm birth is the leading cause of death and can result in significant long-term loss of physical and psychological capacity among survivors.

Background: An estimated 15 million babies are born preterm every year. Prediction models based on machine learning methods have reported promising results.

Aims: To identify risk factors associated with preterm birth and to develop and validate a prediction model for this outcome in a Brazilian birth cohort.

Methods: Cross-sectional study of all births that occurred in Ribeirão Preto-SP and of one in three births that occurred in São Luís-MA, Brazil, in 2010. Questionnaires were applied to obtain pregnancy-related data. Multivariate adaptive regression splines were used to determine the independent variables. Preterm birth, defined as birth before 37 weeks gestational age, was the dependent variable. A random forest model was developed and its performance was evaluated by ROC analysis.

Findings: The preterm birth rates were 12.7% (RP) and 14.1% (SL). The prediction and validation accuracies of the RF-based model were 91.3% and 85.5% respectively. The model can be applied starting in the third month of gestation and is more effective in identifying preterm infants with GA<31 weeks and 6 days (AUC=0.98).

Discussion: It was possible to build a prediction model based on easily accessible low-cost data, without the need for complementary tests, providing results similar to those of other studies.

Conclusions: Previous preterm birth and prenatal care were determinants. The use of an application for individualized patient monitoring an early stage can have positive effects on the quality of life of mother and child.

KEY WORDS

Preterm birth, perinatal health, health policies, prediction models, machine learning

INTRODUCTION

The World Health Organization (WHO) defines preterm birth (PTB) as the result of a pregnancy in which childbirth occurs before 37 completed (1). The incidence of PTB ranges from 5 to 10% in high income countries and can reach 18% in low/middle countries. In Brazil, according to data from the Live Birth Information System (SINASC), this indicator showed a slight increase in recent years, with a PTB rate of 11.11% in 2017, 11.15% in 2018, and 11.23% in 2019 (2).

Preterm birth accounts for 35% of neonatal deaths and for approximately 16% of all deaths. Surviving newborns are at higher risk of developing short- and long-term morbidities (3). However, according to the WHO, more than three-quarters of preterm newborns can be saved, given that care is provided during the prenatal period and childbirth (4).

Since the etiology of PTB involves different sociodemographic, nutritional, biological and environmental factors, its causes are not fully understood (5). For this reason, the development of computer models to predict and aid in decision-making, with a focus on pregnancy outcome, is a matter of great interest for health professionals and public administrators. Identifying risk factors and determining which women are a priority for obstetric management, intensified monitoring, and/or some type of intervention are essential to reduce the rates and costs of PTB (6), as is the identification of women at low risk of PTB for standard follow-up.

In order to achieve more precise, personalized outcomes, the application of predictive and prognostic instruments, which rely on advanced statistical and computational methodologies, have proven to enhance clinical decision-making by effectively estimating probable patient outcomes. (7). The prediction of PTB has been a matter of great interest since the 1970s (8). Several studies have explored different combinations of analysis methods and data types in order to improve the prediction result (9–11). In recent years, studies using machine learning-based methods have reported promising results, with accuracies ranging from 85 to 99% (8, 12–14). In Brazil, studies applying prediction methods to local data are still incipient. Predictions are made based on the individual clinical characteristics of each pregnant woman, such as exams, laboratory indicators, and ultrasonography (15–17).

The aim of the present study was to identify determinants of PTB in a population-based Brazilian birth cohort, and to develop a machine learning model for predicting prematurity.

METHODS

Participants

This is a cross-sectional that used data from the BRISA birth cohort (Brazilian Ribeirão Preto and São Luís Birth Cohort Studies), which was conducted in 2010 simultaneously in two cities, Ribeirão Preto (State of São Paulo) and São Luís (State of Maranhão), the details of recruitment and cohort attrition have been described elsewhere (18). In the present study, data from the Ribeirão Preto (RP) birth cohort were used for analysis and development of the predictive model and data from the São Luís (SL) cohort for validation of the model.

The dataset used for development of the model included 257 variables related to the pregnant woman from RP. Among childbirth and newborn characteristics, only gestational age (GA) was included in the analysis. A total of 7,699 mothers were interviewed, including mothers of stillborn infants and twins. The exclusion criteria were the lack of information about GA (19 participants) and mothers who did not answer the questionnaire (29 participants). The study thus included the information from 7,651 women, with a loss of 0.5%.

The BRISA thematic project was approved by the local Ethics Committees. The present study was approved by the same local Ethics Committee.

Dependent variable

The PTB variable was created by categorization of GA, with birth occurring before 37 completed weeks of gestation or 259 days being defined as preterm. The GA was calculated using two sources of information: GA calculated from the date of the last menstruation reported by the mother during the interview, and/or GA calculated by gestational ultrasound (USG) performed early during pregnancy and provided by the mother at the time of the interview (19,20).

Selection of predictors

Feature selection was applied to obtain the predictors of a possible risk of PTB and to subsequently create the prediction model. In this analysis, all variables present in the database that precede the outcome are used and the result obtained is a list of predictor variables. Multivariate adaptive regression splines (MARS) were used (21). To measure how important a particular predictor is for the outcome, the generalized cross-validation (GCV) parameter was calculated (22).

Other variables that have been well established in the literature as risk determinants of PTB were added to these first variables, namely maternal smoking before and during pregnancy, mother's age (23), number of weeks without work at the end of pregnancy (24), month of onset of prenatal care, and the number of prenatal ultrasounds (25). This combination of variables composes the model developed in the present study.

Statistical analysis

Exploratory analyses were performed using the Stata 14 software (College Station, TX, USA). The R software (version 3.6.1) was used for computational modeling, the Earth package (version 5.1) for feature selection, the Random Forest package (version 4.6) for model development, and the Caret package (version 6.0) to generate the evaluation metrics of the model. Receiver operating characteristic (ROC) curves were constructed and the areas under the curve (AUC) were calculated to determine the best time during pregnancy for application of the model, as well as the detection performance for each PTB category. Descriptive statistics were used for characterization of the sample. Differences between the PTB and term groups were evaluated by the chi-squared test and the unpaired Student *t*-test.

Prediction model development

The prediction problem was established as a binary classification, i.e., whether or not the pregnant woman will have a PTB. The chosen algorithm, random forest (RF), belongs to the class of supervised learning algorithms. This algorithm "learns" the mapping function from the input to the output (26). To assess the generalization ability of the model, *k*-fold cross-validation was applied, in this study, five sets were used.

After development of the model, a confusion matrix was created and the positive attribute was defined as 1 to indicate that class PTB in the variable. The following metrics were evaluated: sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV). Validation was performed with a subset of women from the BRISA birth cohort of SL (n = 5,212) using the same set of variables and the same evaluation metrics as employed in the RP model.

RESULTS

Considering all births that occurred in the BRISA cohort in 2010, there were 12.7% and 14.1% PTB in RP and SL, respectively. Feature selection analysis indicated the predictors of PTB risk described in Table 1. The main predictor was a history of PTB. The RF-based prediction model achieved 90.8% accuracy, 68.5% sensitivity and 93.1% specificity, with PPV and NPV of 50.5% and 96.6%, respectively, for identifying possible PTB. Validation showed 85.5% accuracy, 48% sensitivity, 89.5% specificity, PPV of 32.6%, and NPV of 94.2% (Table 2).

Among the most important characteristics of mothers from RP, 61% of those with PTB reported previous PTB versus only 7.3% of term mothers (p<0.001) and 52% had threatened PTB during the current pregnancy versus 8.7% of term mothers (p<0.001) (Table 3).

The absence of prenatal care was approximately six times greater in the group of PTB mothers compared to the term group (4.7% vs. 0.8%, p<0.001). Prenatal care in the public sector was more common among mothers of children born at term and private care was more common among mothers of PTB infants (p<0.001). The first prenatal consultation of mothers of PTB infants occurred before that of mothers in the term group (p=0.02). However, the mean number of visits was higher for term mothers in the second (p<0.001) and third trimester of gestation (p=0.002).

With respect to maternal smoking, pregnant women with PTB smoked more before (p<0.001) and during pregnancy (p<0.001) than mothers of term infants. The need for hospitalization during pregnancy was three times higher among other with PTB (21.4% vs. 7.45%, p<0.001). Arterial hypertension during pregnancy was more frequent among pregnant women with PTB compared to women in the term group (22.9% vs. 11%, p<0.001), as were other diseases during pregnancy (22.5% vs. 18.6%, p<0.001). The characteristics of mother from SL are shown in Supplemental Table 1.

ROC analysis showed that the performance of the prediction model increased with advancing pregnancy, 3 months (AUC = 0.78), 6 months (AUC = 0.82), and 9 months (AUC = 0.83). Regarding its performance in detecting PTB based on GA category, the prediction model was more accurate in identifying extremes and very preterm infants (GA < 31 weeks and 6 days, AUC = 0.98), followed by moderately preterm infants (GA between 32 and 33 weeks and 6 days, AUC = 0.85), finally, late preterm infants at GA between 34 and 35 weeks and 6 days (AUC = 0.84) and GA within the interval of 36 weeks only (AUC = 0.74).

DISCUSSION

This study used the RF model to analyze PTB based on the data of a population birth cohort. The accuracy of this model was 90.8%. The determinants of the model were obtained by predictor selection. The study explored a large set of sociodemographic, educational, health-related, obstetric and lifestyle determinants of the pregnant woman and her partner. The result of predictor selection showed a predominance of factors directly related to PTB (e.g., a history of PTB, maternal morbidities and factors related to prenatal care) over indirect factors (e.g., number of weeks without work at the end of pregnancy and number of living children).

The results indicate that the necessary preventive measures come down to adequate prenatal monitoring and that this prediction model would be ideally used to support the decision-making of health professionals in primary and secondary care during prenatal consultations. Also, since the prediction model had a high NPV (98%), it may routinely be used to identify false positives that can result in unnecessary follow-up and/or treatment of pregnant women.

In a study of 596 patients, Lee and Ahn (2019) obtained an accuracy of the RF model of 89.2%. The model included information about sociodemographic factors, lifestyle, morbidities during pregnancy, previous PTB, and obstetric data (27). A second study by the same group (2020) involving a sample of 731 patients obtained an accuracy of 86.8% also using RF. In addition to the data used in the previous study, the authors included education, income, periodontal data, gastroesophageal reflux disease, pelvic inflammation, and medicalization history (28). Using a model similar to those mentioned above, the accuracy obtained in the present study was higher (90.8%). Although the present study and the other two cited reports used gestational data, differences in the combination of factors in the prediction models, in the size of the samples studied, and in the adjustments of the models may explain the differences between metrics.

Since information on prenatal characteristics, morbidities, socioeconomic data and lifestyle used in the prediction model is already part of routine monitoring and is collected during the consultations of pregnant women, there are no costs of data collection, a fact permitting the use of this prediction model in primary care at national level. A total of 18 items are required for use of the model; all items are reported in the literature to have a strong relationship with PTB. Previous PTB and threatened PTB during the current pregnancy (29), social class, month of onset and number of prenatal visits in the second and third trimester of gestation, type of service where delivery occurred and number of ultrasounds performed throughout prenatal care (30,31), number of fetuses (32), hospitalization during pregnancy, occurrence of diseases during the current pregnancy and gestational hypertension (33,34), smoking before and during pregnancy and mother’s age (23,35), recommendation of anemia medication (36), number of weeks without work at the end of pregnancy (37), and number of living children. It should be noted that some of these characteristics change during pregnancy; thus, the information available at the time of consultation should be used for modeling (Figure 1).

Validation of the predictive model using the information of pregnant women from SL showed 85.5% accuracy, with high specificity and a high NPV, as also observed in RP. Differences in values might be attributed to the higher prevalence of PTB in SL compared to RP (14.1% vs. 12.7%), in addition to possible cultural and economic differences between the two Brazilian regions.

The present study has some limitations. Its cross-sectional design may have interfered with the prediction accuracy. This study did not consider temporal information about some variables. Also, did not consider possible correlation and mediation effects between variables in the analyses since both RF and MARS are techniques that are not affected by these effects. However, even with limitations, it was possible to explore the situations during the care of the pregnant woman and to build a prediction model based on easily accessible low-cost data, without the need for complementary tests, providing results similar to those of other studies (7,12,14). Model simulations at different times during the prenatal period (Figure 1) showed that the result is already acceptable by the third month of pregnancy and improves over the course of pregnancy as new prenatal data are obtained.

A web-based computer application was developed for use of the prediction model. The information of the pregnant woman is entered into the app, which then performs the prediction based on the model developed and answers whether or not the woman will have a PTB. The use of the prediction model with its high specificity and NPV during pregnancy monitoring may avoid the overuse of medications and treatments and reduce PTB rates and high healthcare costs during pregnancy and during neonatal ICU stays (38).

Studies indicate that obtaining information in primary care for the correct distinction between clinical care and the prevention of future diseases is essential for quaternary prevention (39,40). Changes in classification criteria, cutoff points for the definition of risk states or even programmed medical interventions can generate confusion and harm for healthy pregnant women (40,41), diverting resources for the care of individuals with real needs, in addition to increasing the risk of iatrogenic damage. The performance of the predictive model based on GA categories corroborates this information (Figure 2), with the model showing a better performance in identifying PTB at a younger GA and the real need for hospitalization (AUC = 0.98). With advancing GA, the performance of the predictive model decreases, close to the transition between PT and term delivery (thirty-sixth week) the predictive model shows a reduction in performance (AUC = 0.74), the time when the risk of iatrogenic damage is highest (41).

In conclusion, factors such as previous and threatened PTB, as well as prenatal care, were determinants of PTB in this study. Despite a practically universal prenatal follow-up rate in this population, the health-related indicators and place where the service is performed seem to influence PTB. The use of a web-based application for individualized patient monitoring can have positive effects on the quality of life of the mother-infant binomial and possibly reduce healthcare costs.

REFERENCES

1. WHO. WHO: recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths. Modifications recommended by FIGO as amended October 14, 1976. Acta Obstet Gynecol Scand [Internet]. 1977;56(3):247–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/560099
2. Brasil. Datasus. Informações de Saúde [Internet]. Ministério da Saúde. 2019 [cited 2019 Dec 7]. Available from: http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6936
3. Oskovi Kaplan ZA, Ozgu-Erdinc AS. Prediction of Preterm Birth: Maternal Characteristics, Ultrasound Markers, and Biomarkers: An Updated Overview. J Pregnancy. 2018;2018.
4. WHO. Preterm birth [Internet]. 2018 [cited 2019 Mar 18]. Available from: https://www.who.int/news-room/fact-sheets/detail/preterm-birth
5. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet [Internet]. 2008 Jan;371(9606):75–84. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673608600744
6. Koullali B, Oudijk MA, Nijman TAJ, Mol BWJ, Pajkrt E. Risk assessment and management to prevent preterm birth. Semin Fetal Neonatal Med [Internet]. 2016 Apr;21(2):80–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1744165X16000159
7. Lee K-S, Ahn KH. Application of Artificial Intelligence in Early Diagnosis of Spontaneous Preterm Labor and Birth. Diagnostics [Internet]. 2020 Sep 22;10(9):733. Available from: https://www.mdpi.com/2075-4418/10/9/733
8. Papiernik-Berkhauer E. Coefficient de risque d'accouchement prématuré (C.R.A.P.). La Press médicale [Internet]. 1969;77(21):793–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4895392
9. Catley C, Frize M, Walker RC, Petriu DC. Predicting High-Risk Preterm Birth Using Artificial Neural Networks. IEEE Trans Inf Technol Biomed [Internet]. 2006 Jul;10(3):540–9. Available from: http://ieeexplore.ieee.org/document/1650509/
10. Lee KJ, Yoo J, Kim YH, Kim SH, Kim SC, Kim YH, et al. The Clinical Usefulness of Predictive Models for Preterm Birth with Potential Benefits: A KOrean Preterm collaboratE Network (KOPEN) Registry-Linked Data-Based Cohort Study. Int J Med Sci. 2020;17(1):1–12.
11. Vovsha I, Salleb-Aouissi A, Raja A, Koch T, Rybchuk A, Radeva A, et al. Using Kernel Methods and Model Selection for Prediction of Preterm Birth. 2016 Jul 27; Available from: http://arxiv.org/abs/1607.07959
12. Shaniba Asmi P, Subramaniam K, Iqbal N V. Classification of Nonlinear Features of Uterine Electromyogram Signal Towards the Prediction of Preterm Birth. IETE J Res [Internet]. 2019;0(0):1–10. Available from: https://doi.org/03772063.2019.1634491
13. Bedeeuzzaman M, Anju S. Preterm Birth Prediction Using EHG Signals. Int J Sci Res Eng Trends. 2019;5(3):3–6.
14. Peng J, Hao D, Yang L, Du M, Song X, Jiang H, et al. Evaluation of electrohysterogram measured from different gestational weeks for recognizing preterm delivery: a preliminary

study using random Forest. Biocybern Biomed Eng [Internet]. 2020 Jan;40(1):352–62. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0208521619304899
15. Dória MT, Spautz CC. Trabalho de parto prematuro predição e prevenção. Femina. 2011;
16. Souza RT, Cecatti JG. A Comprehensive Integrative Review of the Factors Associated with Spontaneous Preterm Birth, Its Prevention and Prediction, Including Metabolomic Markers. Rev Bras Ginecol e Obs / RBGO Gynecol Obstet [Internet]. 2020 Jan 27;42(01):051–60. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0040-1701462
17. Bittar RE, Fonseca EB da, Zugaib M. Predição e prevenção do parto pré-termo. Femina. 2010;
18. Confortin SC, Ribeiro MRC, Barros AJD, Menezes AMB, Horta BL, Victora CG, et al. RPS Brazilian Birth Cohorts Consortium (Ribeirão Preto, Pelotas and São Luís): history, objectives and methods. Cad Saude Publica [Internet]. 2021;37(4). Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2021000400601&tlng=en
19. Hutcheon JA, Jacobsen GW, Kramer MS, Martinussen M, Platt RW. Small Size at Birth or Abnormal Intrauterine Growth Trajectory: Which Matters More for Child Growth? Am J Epidemiol [Internet]. 2016 Jun 15;183(12):1107–13. Available from: https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwv310
20. Medeiros MNL, Cavalcante NCN, Mesquita FJA, Batista RLF, Simões VMF, Cavalli R de C, et al. Validity of pre and post-term birth rates based on the date of last menstrual period compared to early obstetric ultrasonography. Cad Saude Publica [Internet]. 2015 Apr;31(4):885–90. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2015000400885&lng=en&tlng=en
21. Friedman JH. Multivariate Adaptive Regression Splines. Ann Stat [Internet]. 1991 Mar;19(1):1–67. Available from: http://projecteuclid.org/euclid.aos/1176347963
22. Pan W, Shen X, Liu B. Cluster Analysis: Unsupervised Learning via Supervised Learning with a Non-convex Penalty. J Mach Learn Res. 2013;14(7).
23. Silva AAM da, Bettiol H, Barbieri MA, Brito LGO, Pereira MM, Aragão VMF de, et al. Which factors could explain the low birth weight paradox? Rev Saude Publica [Internet]. 2006 Aug;40(4):648–55. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102006000500014&lng=en&tlng=en
24. McDonald AD, McDonald JC, Armstrong B, Cherry NM, Nolin AD, Robert D. Prematurity and work in pregnancy. Occup Environ Med [Internet]. 1988 Jan 1;45(1):56–62. Available from: http://oem.bmj.com/cgi/doi/10.1136/oem.45.1.56
25. Brasil. PORTARIA Nº 570, DE 1º DE JUNHO DE 2000 [Internet]. Brasil; 2000. Available from: http://bvsms.saude.gov.br/bvs/saudelegis/gm/2000/prt0570_01_06_2000_rep.html
26. Duda R, Stork D, Hart P. Pattern Classification. 2nd ed. Wiley-Interscience; 2000. 680 p.
27. Lee K-S, Ahn KH. Artificial Neural Network Analysis of Spontaneous Preterm Labor and Birth and Its Major Determinants. J Korean Med Sci [Internet]. 2019;34(16). Available from: https://jkms.org/DOIx.php?id=10.3346/jkms.2019.34.e128
28. Lee K-S, Song I-S, Kim E-S, Ahn KH. Determinants of Spontaneous Preterm Labor and Birth Including Gastroesophageal Reflux Disease and Periodontitis. J Korean Med Sci [Internet]. 2020;35(14). Available from: https://jkms.org/DOIx.php?id=10.3346/jkms.2020.35.e105

29. Goodfellow L, Care A, Sharp A, Ivandic J, Poljak B, Roberts D, et al. Effect of QUiPP prediction algorithm on treatment decisions in women with a previous preterm birth: a prospective cohort study. BJOG An Int J Obstet Gynaecol. 2019;0–2.
30. Do Carmo Leal M, Esteves-Pereira AP, Nakamura-Pereira M, Torres JA, Theme-Filha M, Domingues RMSM, et al. Prevalence and risk factors related to preterm birth in Brazil. Reprod Health [Internet]. 2016;13(Suppl 3). Available from: http://dx.doi.org/10.1186/s12978-016-0230-0
31. Pan I, Nolan LB, Brown RR, Khan R, Van Der Boor P, Harris DG, et al. Machine learning for social services: A study of prenatal case management in Illinois. Am J Public Health. 2017;107(6):938–44.
32. Razaz N, Avitan T, Ting J, Pressey T, Joseph KS. Perinatal outcomes in multifetal pregnancy following fetal reduction. Cmaj. 2017;189(18):E652–8.
33. Bertagnolli M, Luu TM, Lewandowski AJ, Leeson P, Nuyt AM. Preterm birth and hypertension: is there a link? Curr Hypertens Rep. 2016;18(4):28.
34. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet [Internet]. 2008 Jan;371(9606):75–84. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673608600744
35. Lamy S, Houivet E, Marret S, Hennart B, Delavenne H, Benichou J, et al. Risk factors associated to tobacco and alcohol use in a large French cohort of pregnant women. Arch Womens Ment Health [Internet]. 2019 Apr 3;22(2):267–77. Available from: http://link.springer.com/10.1007/s00737-018-0892-4
36. Xiong X, Buekens P, Alexander S, Demianczuk N, Wollast E. Anemia during pregnancy and birth outcome: a meta-analysis. Am J Perinatol. 2000;17(03):137–46.
37. Buen M, Amaral E, Souza RT, Passini R, Lajos GJ, Tedesco RP, et al. Maternal Work and Spontaneous Preterm Birth: A Multicenter Observational Study in Brazil. Sci Rep [Internet]. 2020 Dec 16;10(1):9684. Available from: http://www.nature.com/articles/s41598-020-66231-2
38. Mwamakamba LW, Zucchi P. Cost estimate of hospital stays for premature newborns of adolescent mothers in a Brazilian public hospital. Einstein (São Paulo) [Internet]. 2014 Jun;12(2):223–9. Available from: http://www.scielo.br/scielo.php ?script=sci_arttext&pid=S1679-45082014000200223&lng=en&tlng=en
39. Martins C, Godycki-Cwirko M, Heleno B, Brodersen J. Quaternary prevention: an evidence-based concept aiming to protect patients from medical harm. Br J Gen Pract [Internet]. 2019 Dec 28;69(689):614–5. Available from: http://bjgp.org/lookup/doi/10.3399/bjgp19X706913
40. Tesser CD, Norman AH. Differentiating clinical care from disease prevention: a prerequisite for practicing quaternary prevention. Cad Saude Publica [Internet]. 2016 Oct;32(10). Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2016001003001&lng=en&tlng=en
41. Silva AA, Barbieri MA, Gomes UA, Bettiol H. Trends in low birth weight: a comparison of two birth cohorts separated by a 15-year interval in Ribeirão Preto, Brazil. Bull World Health Organ [Internet]. 1998;76(1):73–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9615499

Table 1: Variables identified by predictor selection. BRISA cohort, Ribeirão Preto, 2010.

Variable (attribute)	Variable importance based on GCV*
Previous preterm birth	100
Threatened preterm birth during current pregnancy	58.1
Number of fetuses	42.4
Number of prenatal visits in the third trimester	34.8
Type of prenatal care service	28.4
Hospitalization during pregnancy	25.4
Morbidities during pregnancy	20.4
Number of living children	18.7
Iron supplementation for anemia treatment	17.3
Number of weeks without work at the end of pregnancy	13.9
Arterial hypertension during pregnancy	12.2
Number of prenatal visits in the second trimester	8.1

*GCV – generalized cross-validation.

Table 2: Performance of the model for predicting preterm birth. BRISA cohort, Ribeirão Preto, 2010.

	Accuracy	Sensitivity	Specificity	PPV	NPV
RP	90.8	68.5	93.1	50.5	96.6
SL	85.5	48.0	89.5	32.6	94.2

RP – Ribeirão Preto; SL – São Luís; PPV – positive predictive value; NPV – negative predictive value.

Table 3: Characteristics of the pregnant women. BRISA cohort, Ribeirão Preto, 2010.

Variable	Term, N (%)		Preterm, N (%)		p-value
	6,679 (87.3)		972 (12.7)		
Previous preterm birth					< 0.001
Yes	487	7.3	593	61.0	
No	6,104	91.4	363	37.4	
Threatened preterm birth during pregnancy					< 0.001
Yes	581	8.7	505	52	
No	6,084	91.1	451	46.4	
Prenatal care					< 0.001
Yes	6,625	99.2	926	95.3	
No	54	0.8	42	4.7	
Type of prenatal service					< 0.001
Public	3,653	54.7	440	45.3	
Private	2,958	44.3	482	49.6	
Did not receive prenatal care	54	0.8	50	4.7	
Hospitalization during pregnancy					< 0.001
Yes	497	7.45	208	21.4	
No	6,178	92.5	762	78.4	
Morbidities during pregnancy					< 0.001
Yes	1,242	18.6	219	22.5	
No	5,423	81.2	748	77	
Hypertension during pregnancy					< 0.001
Yes	735	11	223	22.9	
No	5,930	88.8	742	76.4	
Recommendation of anemia medication					< 0.001
Yes	4,374	65.5	569	58.6	
No	2,237	33.5	349	35.9	
Did not receive prenatal care	54	0.8	46	4.7	
Maternal smoking before pregnancy					< 0.001
Yes	1,296	19.4	247	25.4	
No	5,383	80.6	724	74.5	
Maternal smoking during pregnancy					< 0.001
Yes	735	11	164	16.9	
No	5,944	89	807	83	
Number of weeks without work at the end of pregnancy					< 0.001
First trimester	33	0.5	5	0.5	
Second trimester	123	1.85	30	3.1	
Third trimester	2,274	34	254	26.2	
Worked throughout pregnancy	1,104	16.5	225	23.2	
Did not work during pregnancy	3,112	46.6	446	46	
			Mean (SD)		
Number of prenatal visits in the second trimester	2.6	1.12	2.5	1.5	< 0.001
Number of prenatal visits in the third trimester	4.35	2.05	2.9	2.1	0.002
Number of fetuses	1	0.06	1.1	0.3	< 0.001
Number of living children	1.8	2.6	3.01	1.2	< 0.001
Maternal age (years)	27	6.23	27.2	6.44	0.32
Month of onset of prenatal care	2.5	1.41	2.4	1.5	0.02
Number of ultrasounds during pregnancy	4.1	1.65	4.4	2	< 0.001

* The total number varies because of missing information for some categories of the variables.

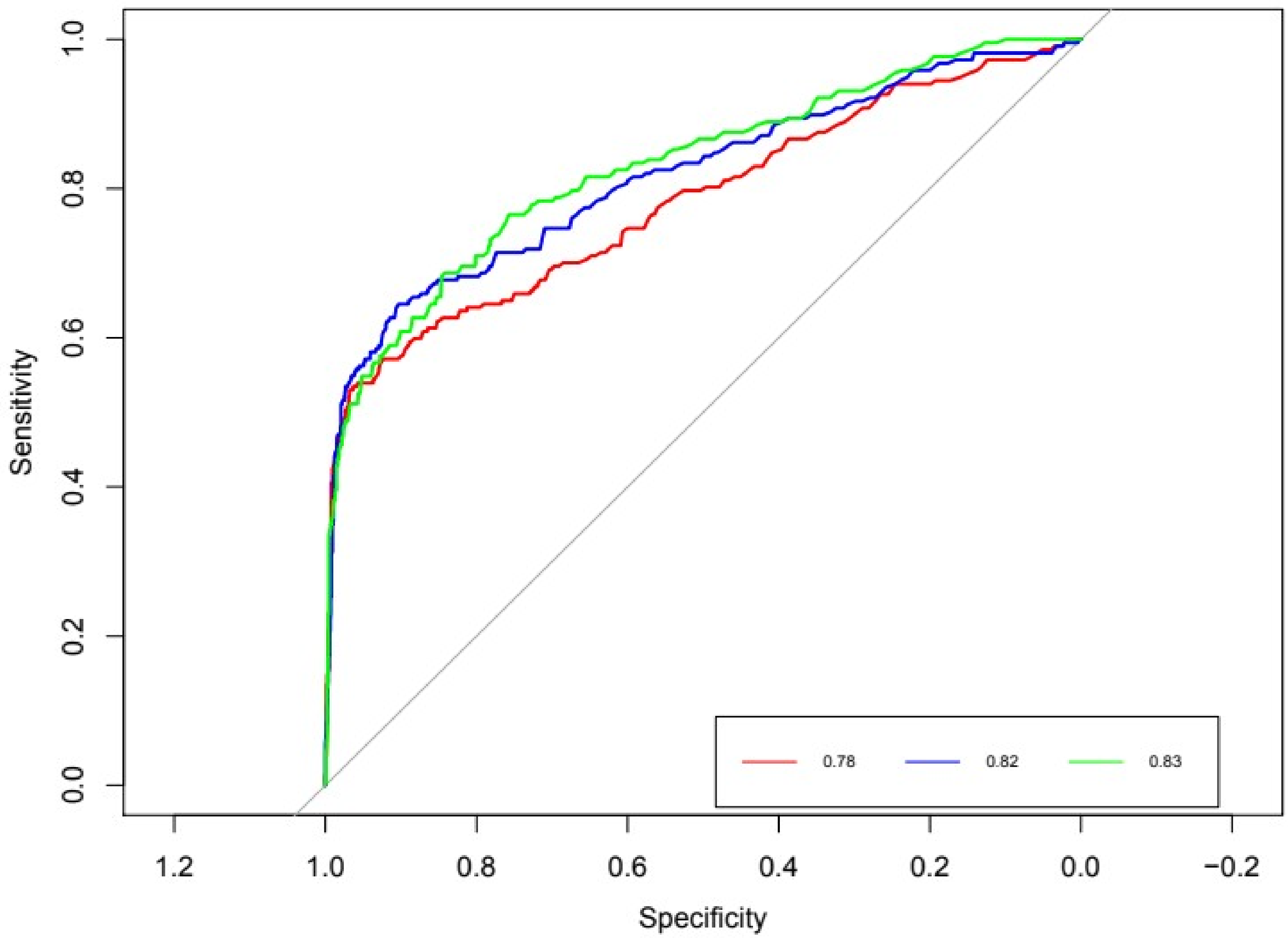

Figure 1: Receiver operating characteristic (ROC) curves for use of the prediction model at three different times during pregnancy: 3 months (red line, AUC = 0.78), 6 months (blue line, AUC = 0.82), and 9 months (green line, AUC = 0.83).

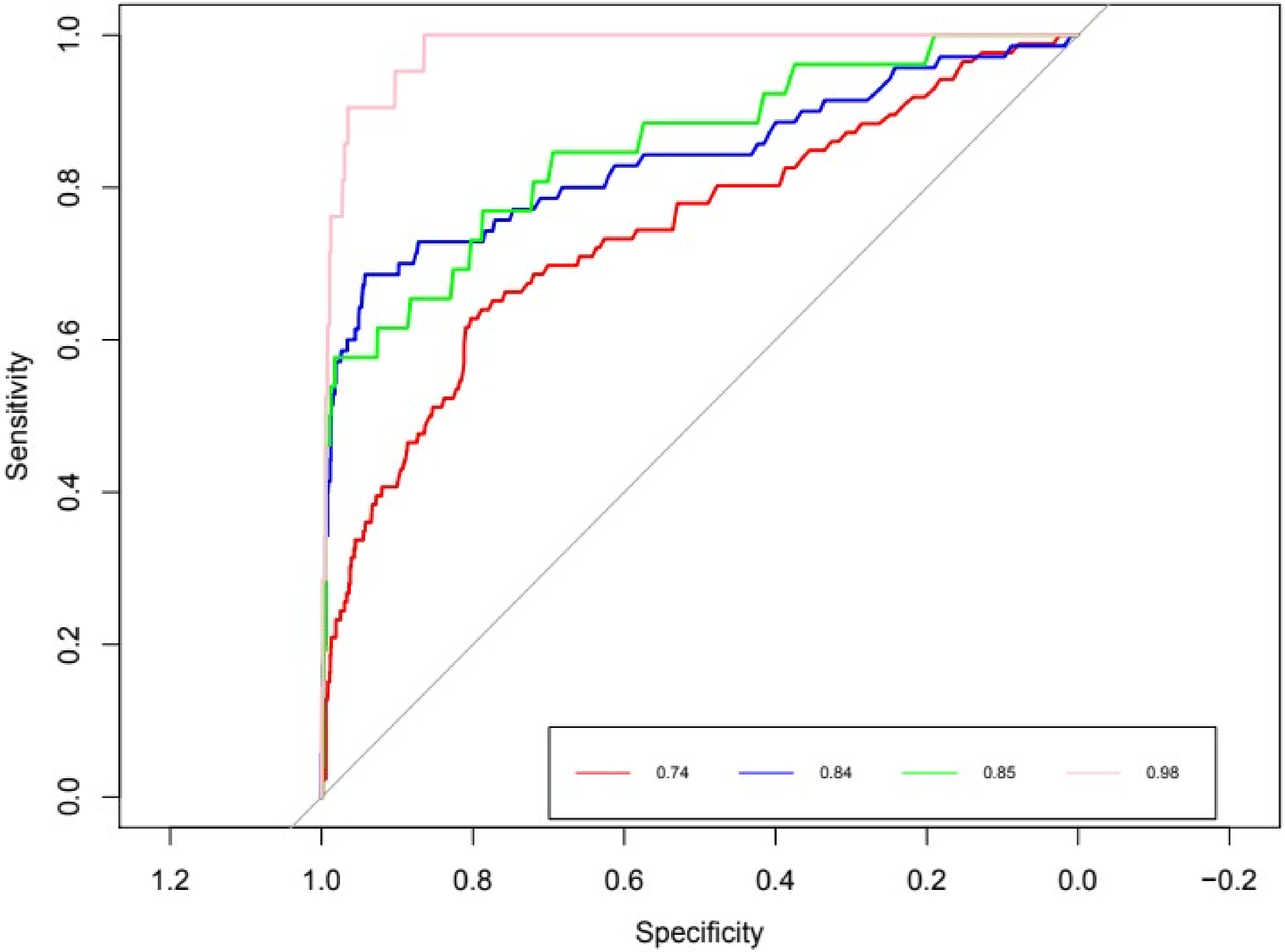

Figure 2: Receiver operating characteristic (ROC) curves for assessing the performance of the model in detecting preterm birth according to gestational age (GA). Pink line (GA < 31 weeks and 6 days, AUC = 0.98), green line (GA > 32 and < 33 weeks and 6 days, AUC = 0.85), blue line (GA > 34 and < 35 weeks and 6 days, AUC = 0.84), and red line (GA within the interval of 36 weeks, AUC = 0.74).

Supplemental Table 1: Characteristics of the pregnant women. BRISA cohort, São Luís, 2010.

Variables	**Term, N(%)**		**Preterm, N(%)**		**p-value**
	4,477 (85.9)		735 (14.1)		
Previous preterm birth					< 0.001
Yes	437	9.8	369	50.2	
No	4,024	90.2	366	49.8	
Threatened preterm birth during pregnancy					< 0.001
Yes	417	9.3	313	42.6	
No	4,055	90.7	421	57.4	
Prenatal care					< 0.001
Yes	4,403	98.3	702	95.5	
No	74	1.7	33	4.5	
Type of prenatal service					0.609
Public	3,596	80.3	567	77.2	
Private	806	18.0	134	18.3	
Did not receive prenatal care	74	1.7	33	4.5	
Hospitalization during pregnancy					< 0.001
Yes	371	8.3	107	14.6	
No	4,105	91.7	628	85.4	
Morbidities during pregnancy					0.173
Yes	270	6.0	35	4.8	
No	4,204	94.0	700	95.2	
Hypertension during pregnancy					< 0.001
Yes	687	15.3	176	24.0	
No	3,789	84.7	558	76.0	
Recommendation of anemia medication					0.901
Yes	4,009	89.6	640	87.2	
No	389	8.7	61	8.3	
Did not receive prenatal care	74	1.7	33	4.5	
Maternal smoking before pregnancy					0.163
Yes	415	9.3	80	10.9	
No	4,061	90.7	654	89.1	
Maternal smoking during pregnancy					0.001
Yes	167	3.7	46	6.3	
No	4,310	96.3	689	93.7	
Number of weeks without work at the end of pregnancy					0.001
First trimester	29	0.65	0	0	
Second trimester	73	1.6	12	1.6	
Third trimester	970	21.7	132	18	
Worked throughout pregnancy	365	8.2	87	11.8	
Did not work during pregnancy	3,028	67.8	503	68.5	
			Mean (SD)		
Number of prenatal visits in the second trimester	2.54	0.97	2.47	1.05	0.123
Number of prenatal visits in the third trimester	3.1	1.36	2.53	1.3	< 0.001
Number of fetuses	1	0.1	1.1	0.27	< 0.001
Number of living children	1.86	1.16	1.88	1.18	0.626
Maternal age (years)	25.2	5.95	24.9	6.32	0.35
Month of onset of prenatal care	3.01	1.44	3.03	1.43	0.77
Number of ultrasounds during pregnancy	3.1	1.74	3.4	2.2	0.004

* The total number varies because of missing information for some categories of the variables.