BMJ Open

Cohort profile: Colombian Cohort for the Early Prediction of Preterm Birth (COLPRET): early prediction of preterm birth based on personal medical history, clinical characteristics, vaginal microbiome, biophysical characteristics of the cervix and maternal serum biochemical markers

Carlos Hernan Becerra-Mojica [1,2,3] Miguel Antonio Parra-Saavedra [4]
Luis Alfonso Diaz-Martinez [5] Raigam Jafet Martinez-Portilla [6]
Bladimiro Rincon Orozco [7]

[Check for updates]

ABSTRACT

Purpose Preterm birth (PTB) is a public health issue. Interventions to prolong the length of gestation have not achieved the expected results, as the selection of population at risk of PTB is still a challenge. Cervical length (CL) is the most accepted biomarker, however in the best scenario the CL identifies half of the patients. It is unlikely that a single measure identifies all pregnant women who will deliver before 37 weeks of gestation, considering the multiple pathways theory. We planned this cohort to study the link between the vaginal microbiome, the proteome, metabolome candidates, characteristics of the cervix and the PTB.

Participants Pregnant women in the first trimester of a singleton pregnancy are invited to participate in the study. We are collecting biological samples, including vaginal fluid and blood from every patient, also performing ultrasound measurement that includes Consistency Cervical Index (CCI) and CL. The main outcome is the delivery of a neonate before 37 weeks of gestation.

Findings to date We have recruited 244 pregnant women. They all have measurements of the CL and CCI. A vaginal sample for microbiome analysis has been collected in the 244 patients. Most of them agreed to blood collection, 216 (89%). By August 2021, 100 participants had already delivered. Eleven participants (11 %) had a spontaneous PTB.

Future plans A reference value chart for the first trimester CCI will be created. We will gather information regarding the feasibility, reproducibility and limitations of CCI. Proteomic and metabolomic analyses will be done to identify the best candidates, and we will validate their use as predictors. Finally, we plan to integrate clinical data, ultrasound measurements and biological profiles into

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ To the best of our knowledge, this is the first Colombian cohort designed to obtain differential demographic and biological characteristics from the early stages of pregnancy in women who deliver preterm.

⇒ The biomarkers identified will allow us to create algorithms to establish an individualised risk of preterm birth (PTB). The study includes stratified risk based on the antecedents of PTB.

⇒ We will have a significant repository of clinical data, biological samples and ultrasound images to extend the research.

⇒ The study includes biological markers evaluated only at one point in the first trimester, and we will not obtain trends over time.

an algorithm to obtain a multidimensional biomarker to identify the individual risk for PTB.

INTRODUCTION

Preterm birth (PTB) is a public health issue around the world. It has been estimated that at least 15 million babies are born preterm each year.[1] Prematurity is the leading cause of neonatal and under 5 years old mortality. Approximately one million babies die each year due to PTB complications.[2] Hyaline membrane disease, necrotizing enterocolitis and intracerebral hemorrhage are the most significant consequences of prematurity, and these conditions are the basis of increased

BMJ

mortality and short-term disabilities.[3] According to the Global Burden of Disease report, PTB accounts for 3.1% of the years of healthy life lost due to disability.[4] The frequency of PTB varies worldwide, with rates ranging from 9% to 12%.[1] The global report on PTB published by the WHO shows a rise in frequency worldwide. In Colombia, the rate of prematurity is 14%, higher than the world's average.[5] According to the Colombian National Institute of Health, in 2019, 45.2% of perinatal and late neonatal deaths were caused by prematurity, asphyxia and related causes.[6]

Some strategies have contributed to the reduction of the morbidity related to prematurity. These include the use of uterine inhibitors, steroid-induced lung development, neuroprotection by magnesium sulfate, technological and care improvements in paediatric intensive care units and 'kangaroo mother' programmes.[7] However, none of these measures, alone or combined, have reduced de incidence of preterm delivery, and therefore, the prematurity.[8–10]

One of the objectives of the WHO is the reduction of perinatal mortality and morbidity of children under 5 years old; this ambitious objective involves diligent work in reducing PTB, which requires basic sciences research to predict and prevent this condition, as well as clinical research in a translational programme to help translate all biological findings into a valuable clinical scenario for the patients.[7]

The most reliable risk factor for PTB prediction is a history of previous PTB.[11] In a recent meta-analysis, the authors concluded that the risk of having PTB before 37 weeks in a subsequent pregnancy in a woman with a medical history of PTB is 30%.[12] The benefit of progesterone therapy was demonstrated for the first time in pregnant women with a history of PTB, reducing PTB risk in the subsequent pregnancy by 34%.[13] This result was confirmed later with vaginal progesterone reducing the PTB frequency before 34 or 37 weeks compared with the placebo group.[14] However, only 15% of pregnant women have a history of PTB, which does not allow us correctly to classify the other 85% at risk of PTB.

Regarding the biomarkers for PTB prediction, cervical changes throughout gestation have been studied, and cervical shortening is the most evaluated characteristic. These changes can be identified through the measurement of CL by transvaginal ultrasound and, when they occur at an early gestational age, can predict the occurrence of PTB.[15] A short cervix, with a length less than 25 mm (10th percentile), was identified in 18% of patients who had PTB before 37 weeks.[16] When combined with a history of PTB, a CL less than 25 mm between 20 and 24 weeks may identify up to 28% of patients who will deliver between 34 and 37 weeks.[17] A meta-analysis of individual patient data showed that the vaginal use of progesterone in patients with a short cervix could reduce the risk of PTB by 38%.[18] More recently, two proposals to increase the sensitivity (Se) of the CL, one in nulliparous changing the cut-off of the cervical length (CL) and the other with

customized charts based on the maternal characteristics, have been published[19 20]

Other cervical measurements have been proposed.[21] Parra-Saavedra *et al* postulated the Consistency Cervical Index (CCI) as a predictor of spontaneous PTB. The authors demonstrated greater Se and positive predictive value (PPV) than CL measurement.[21] However, more evidence is needed to validate the reproducibility of this measure as an effective tool to predict PTB.

The new sequencing techniques have allowed us to better understand bacterial communities present in the vagina and their characterization throughout gestation to predict PTB.[22 23] Showing differences among women who delivered preterm and those who delivered at term, and racial differences that show the variability of these bacterial communities among populations and their potential role as a predictor of PTB.[24]

The proteome characterization in the serum of pregnant women has improved the understanding of the pathways involved in PTB.[25] Inflammation has been recognized as one of the critical processes associated with the changes in maternal and fetal tissues during delivery. Inflammatory mediators released from senescent fetal tissues during pregnancy spread to maternal gestational tissues to signal preparation for delivery. The increased load of immune cells and inflammatory proteins is reflected in the tissues with accumulation in the cervical stroma favouring rapid remodelling and releasing proteolytic enzymes that digest the proteins of the extracellular matrix to culminate in the process of cervical dilation, which is a step before expulsion.[26]

The search for biomarkers using proteomics has been carried out on different types of samples, in which some differences have been found between the groups of patients with PTB and full-term delivery. From cervical samples, it has been determined that the proteins expressed at the highest levels in women with PTB are fibronectin, extracellular matrix protein 1, laminin and calsintenin.[27] Other studies found proteins such as desmoplakin 1, which participates in intercellular junctions and cell–cell communication, localized in human fetal membranes.[28]

From the beginning of the pregnancy, many physiological changes and metabolic adaptations in the mother are present, and they are necessary for gestational development and maintaining maternal–fetal well-being.[29] In recent years, interest has increased regarding understanding how these processes occur in the maternal system to identify possible candidate biomarkers for the early detection of pathologies in pregnancy, such as preeclampsia and PTB, through metabolomic analysis. This branch of science is responsible for studying the molecules present in cells, tissues and body fluids and how these are reflected in the state of health and well-being. The metabolic profile based on gene expression analysis provides a quantifiable reading of the biochemical state from normal physiology to various pathologies. This information allows a realistic approach to the phenotype

of the disease and the possible identification of a set of predictive markers with higher reliability.[30]

Advances in biomarker detection technology related to metabolomics proteomics have resulted in the generation of a large amount of multidimensional data, which requires developing more efficient analytical techniques. Machine learning methods represent the ability of algorithms to learn from a set of data and apply this knowledge to a new set of data to make accurate classifications and predictions. This approach allows better identification of the essential characteristics present in the data and, therefore, a better understanding of the factors relevant to the problem studied than traditional approaches. The use of machine learning methods in omics data analysis has recently increased, and the strategies developed for each specific study.[31 32]

Because several routes contribute to the onset of labour,[33] it is unlikely that a single marker will identify pregnant women who will deliver before term, which explains the limited results of current interventions for the containment of PTB. In this regard, we believe that it is necessary to integrate the clinical, epidemiological and environmental aspects, as well as the characteristics of the cervix and biological markers of the different pathways involved in the onset of labour; this task will result in a large volume of data which require, not only the clinical expertise but also the integration of all the information in artificial intelligence algorithms, This will allow us to establish an individual risk, derived from biological and environmental determinants, on the way to personalised medicine.[34]

Cohort description

From January 2020, all consecutive pregnant women who attended their first-trimester screening ultrasound at the maternal–fetal medicine units of two centres, Hospital Universitario de Santander and Centro de Atencion Materno Fetal INUTERO in Bucaramanga Colombia, have been invited to participate in the study.

Eligibility criteria

Inclusion criteria were women with a singleton pregnancy between 11+0 and 13+6 week's gestation, planning to stay in the nearby area during gestation and delivery. Exclusion criteria were participants with a medical history of cancer, HIV infection, or history of cervical surgery. Also, pregnant women whose gestation ended before 37 weeks on medical indication of fetal or maternal causes were excluded. Patients were invited consecutively by a research team member in the recruitment centres. After acceptance and signing the consent, the pre-established format for collecting data was filled out for the following categories: age, parity, gestational age, marital status, labour status, educational level, social security type, PTB history, a medical condition associated (HTA, diabetes), smoking status, health insurances, height and weight to calculate the body mass index.
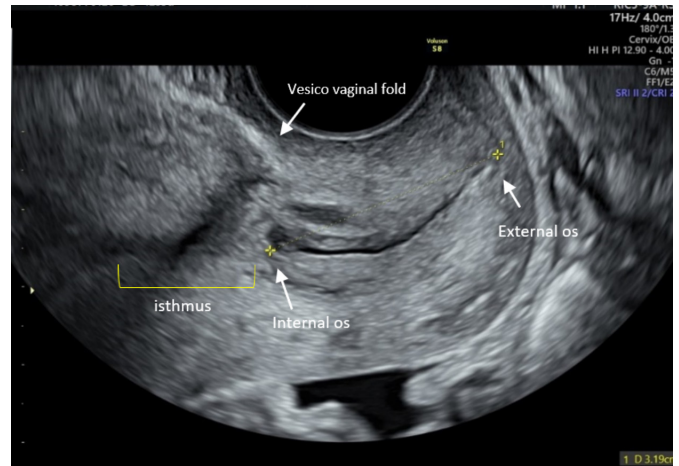
Figure 1 Image showing the basal characteristics for the Consistency Cervical Index measurement.

METHODS

Main outcome

The main outcome is spontaneous PTB which is defined as delivering the baby before 37 weeks of gestation, excluding those indicated by maternal or fetal causes.

All women selected to participate in the study underwent the following procedures:

Ultrasound measurements

A medical team of specialists in maternal–fetal medicine, who carried out the screening studies (11+0–13+6 weeks) at these institutions, received training from Dr Miguel Parra-Saavedra in the measurement technique of the CCI. Subsequently, the physicians made images with the requested characteristics and a predesigned format, which was qualified as appropriate after achieving compliance greater than 90%.

Proposed protocol for the first trimester cervical consistency index measurement

1. The maternal bladder must be empty.
2. With the woman in lithotomy position, the vaginal transducer is directed toward the anterior vaginal fornix.
3. To obtain the adequate image (figure 1)
   a. Use the vision angle of 180°.
   b. Obtain a sagittal view of the cervix, visualising the endocervical canal.
   c. Identify the internal and the external os.
   d. Magnify the image until the cervix occupies 75% of the screen.
4. To obtain cervical measurements (figure 2)
   a. Measure the CL without including the isthmus, place one calliper in the external os and the second calliper where the cervical glands end, at the level of the vesicovaginal fold. (the line will serve as a reference).
   b. Freeze the image and press the double image screen button, be sure that the anterior lip of the cervix
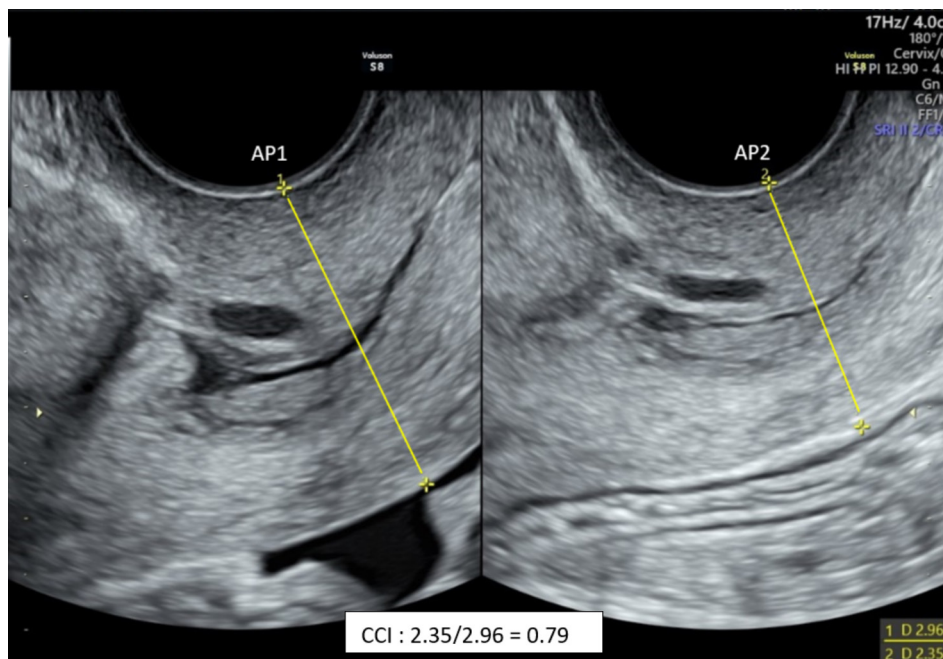
Figure 2 Anteroposterior diameter of the cervix without probe pressure (left), and the anteroposterior diameter after pressure (right). CCI, Consistency Cervical Index.

is equal to the posterior lip (without over press the cervix with the vaginal probe).

c. Press the transducer toward the posterior vaginal wall until the cervix displacement starts, use the cineloop to center the image with similar characteristics to the left image and freeze.

d. Measure the anterior–posterior diameter of the cervix in each image, tracing a perpendicular line to the cervical canal.

e. The callipers must be placed on the bright line on the anterior and posterior edge of the cervix.

5. Obtain the cervical consistency index dividing the anteroposterior diameter under pressure (AP2) by the basal anterior-posterior diameter without pressure (AP1). Obtain a number between 0 and 1.

6. Repeat step 4 to obtain a second measurement, use the minor number obtained (the minor number represents the maximum compressibility of the cervix).

According to the national recommendations, all patients underwent a screening study that includes a CRL measurement to establish gestational age.[35 36] In addition, a CL measurement was performed, using the described technique,[37] also two CCI measurements were performed.[21] The images of these measurements are saved in the ultrasound equipment (General Electric Voluson E8, General Electric Voluson E6, General Electric Voluson S8). Additionally, 50 patients were evaluated by two operators. The images and the study information are exported to the external hard disc to protect the data.

Vaginal sample for the microbiome

At the time of screening, a sample of the middle third of the vagina is taken with a device designed for this purpose(Patent Ref. No:NC2016/0002338. Res. 1214 of the Colombian Industry and Commerce Superintendency). The women's sample acceptability was previously proved.[38] The single-use device has a brush to collect the vaginal sample; the brush is subsequently separated from the device and dispensed in a sterile container; two ml of cell-conserving liquid is added to maintain the sample's integrity. The vials with the brushes inside are stored at a temperature of 4°C until DNA extraction.

DNA from vaginal samples is extracted with the salting-out technique. The DNA is re suspended in 50 µL of Tris-Edta buffer for its preservation. DNA quantification and purity are carried out in Scientific Nano Drop One (Thermo Scientific, USA), and the DNA sample is stored at 20°C until the sequencing.

Blood sample for omics analysis

Two blood samples collected with EDTA anticoagulant are taken from each patient. The extraction of the sample is carried out through the vacuum system technique in the veins located in the antecubital area of the forearm. The samples are centrifuged at 3000 rpm for 10 min in the laboratory with the Sorvall ST 16 Centrifuge (Thermo Scientific, USA)to separate its components and obtain blood plasma. From each patient, four aliquots of 800 µL are obtained and collected in Eppendorf tubes and are stored in Ultra-Low Temperature Freezer (Thermo Scientific, USA) at −70°C.

Follow-up

After the first visit that includes the registration of the initial information, the ultrasound study and the collection of the vaginal and blood samples, a research team member makes a monthly telephone call to monitor the evolution of the pregnancy until delivery. A record is kept

with the date of delivery, the delivery route, the delivery characteristics, and if it was spontaneous or occurred on medical indication. In the case of medical indication, the reason for it is also recorded. These data are confirmed in the file of each patient.

Data collection

Clinical data are stored in a password-protected, web-based electronic database, RED Cap, with the deidentification capability to protect patient information. After extraction from the ultrasound machine, the deidentified ultrasound images are stored in a repository and linked through the unique assigned code. The information will be analysed with the Stata statistical software.

STATISTICAL PLAN

Sample size

Calculations will be based on the main outcome, which is PTB before 37 weeks. However, this is a diagnostic test analysis on which we will calculate the needed number of patients according to the expected increase in Se for a fixed 15% false positive rate (FPR) using the current best predictive model for PTB as the baseline Se. According to Celik et al from the Fetal Medicine Foundation Model,[39] the Se at a 15% FPR for a comprehensive model using CL, obstetric history and maternal characteristics for the prediction of PTB below 37 weeks is 34.7%. We expect to increase the prediction to 50% for the same 15% FPR. For a delta of 15.30% of the expected increase using an 80% probability of finding a statistically significant difference for a threshold of 0.05 and an expected 15% attrition, we would need a total of 384 patients in the cohort. In addition, we plan to increase the cohort as required by the results of the exploratory omics studies.

Analysis plan

Sociodemographic data will be analysed in proportions or central tendency measures: mean and SD to normally distributed data, and the not-normal data will be reported with median and IQR. The population included for analysis will be divided into two groups: with spontaneous premature birth (<37 weeks) and with term birth (≥37 weeks). The differences between groups will be evaluated via the $\chi2$, $\chi2$ of lineal tendency ($\chi2LT$), or Wilcoxon tests. According to the situation, differences between these dichotomic categories will be evaluated with $\chi2$ or Fisher's exact tests, considered statistically significant when $\alpha<0.05$. For continuous variables, ROC analysis will be performed to find the better cut-off points to predict premature labour, and those distributions will be grouped into two to four categories according to the association relationships.

Regarding the predictive model, we decided to take a classical approach for a multivariate logistic model and add a machine learning approach with automatic variable selection and penalisation of non-significant variables to avoid overfitting and lack of parsimoniously and

robustness for the model. A classical and machine learning approach will be used for the construction of predictive models for PTB. First, a training and testing dataset will be constructed in a 50%/50% ratio. The training dataset will be used for the construction of the models and the testing dataset for testing and calibration of the models. In the training dataset, an AI elastic net model will be used as the primary approach for model creation. The elastic net uses a ridge regression that penalises non-statistically significant variables and reduces their coefficients to zero to avoid a non-robust model. Elastic net also uses the lasso regression, which performs an automatic selection of the variables that best predicts the outcome (PTB). The final model will be selected based on the lowest lambda which includes a 10-fold cross-validation automatically chosen by the lowest Akaike information criterion and Bayesian information criteria. Because of automatic selection of predictors achieved by L1-penalty, no previous subset selection, which typically has been used in previous methods, need to be performed, thereby reducing the variance and instability of the prediction model. Automatic selection of predictors performed in elastic net results in a simpler, sparse model that includes only a subset of variables, thereby allowing for better interpretation of the model.

For the classic approach, we will use the training dataset to create a multivariate nested model using forward and backwards stepwise regression to assess the association between several predictors and the main outcome. A verification inflation factor (VIF) analysis will be performed to stablish possible multicollinearity among variables in the model, where a VIF greater than 10 will be considered as highly multicollinear and will be excluded from the model, and a VIF of 4 or more will be explored. A multiple model will be created using a nested logistic approach in which the base model will comprise all possible maternal characteristics found statistically significant at the univariate analysis, then a second model will be added to the first model using biomarkers such as CL, CCI, to evaluate the added value of the addition of these markers for the prediction of PTB. A Nagelkerke R^2 by X^2 analysis will be used to calculate is there a statistically significant difference among models. And as a third step, a third model will be created using the first two models and adding new markers for PTB such as microbiome.

The testing dataset will be used for validation of both models using an ROC (receiver operating characteristic) curve analysis (compared by DeLong method), Se, specificity (Sp), PPV, negative predictive value and accuracy as discrimination methods. Calibration of the models will be performed by plotting the observed versus predicted using the testing dataset and compared using the Hosmer-Lomeshow test for goodness-of-fit. Finally, a Kaplan-Meier survival analysis will be used to stablish the difference in time among women selected as high and low risk using the previous model approach. Data will be analysed using STATA V.17 for Mac and R statistics.

16S/18S NGS SEQUENCING

Vaginal samples are processed using Illumina MiSeq Sequencing System. It has kits for adding index (Nextera XT index) and a flow cell (500-Cycle MiSeq Nano), to obtain a 2×250 paired-end configuration considering the Illumina MiSeq V2 platform. Primers are directed to region V3 and V4 regions of the 16S rRNA gene for the identification of bacteria and archaea, suggested by Illumina (F-16S 5'-TCGTCGGCAGCGTCAGATGTGTAT AAGAGACAGCCTACGGGNGGCWGCAG and R-16S 3'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGAC AGGACTACHVGGGTATCTAATCC). Likewise, primers targeting the 18S SSU rRNA gene targeting eukaryotes are used based on Amaral-Zettler et al[40] and Stoeck et al[41] (F-1391 5'- GTACACACCGCCCGTC and R-EukBr 3'- TGATCCTTCTGCAGGTTCACCTAC). The protocol Part No. 15044223 Rev. B is used to generate amplicons and index addition, as the company suggests.

Metagenomic analysis

The initial analysis of the sequences is carried out using the MG-RAST server, which performs rRNA detection by initial BLAST search using SILVA database reduced data set. Subsequently, the rRNA sequences are grouped using a 97% identity percentage and finally identified the samples with the SILVA SSU database.[42] Processing options to be used include deleting the readings of artificial replication, Bowtie to eliminate the contamination caused by the host using A. thaliana, TAIR, TAIR9 as a reference, and filtering low-quality sequences using 15 as the lowest Phred score.[42] The MEGAN software will be used for the taxonomic and phylogenetic analysis of the data.[43] Available from: https://dx.plos.org/10.1371/journal.pcbi.1004957. For statistical hypothesis testing, taxonomic profile analysis, calculation of effect sizes and CIs, the STAMP software (statistical analysis of taxonomic and functional profiles) V.2.1.3. will be used, the software allows comparing between pairs of samples or between groups of samples arranged in two or more treatment groups.[44]

Findings to date

From January 2020 to August 2021, 261 pregnant women have been invited to participate in the cohort, 251 (96.2%) agreed to participate and signed the informed consent; in the case of minors, the assent was signed too. Seven patients, who reported being under antibiotic treatment, were excluded (figure 3). Three of them were under systemic treatment for urinary tract infection, and the rest were under topical treatment for genital infections. Four patients who did not answer follow-up calls during pregnancy and up to 30 days after the probable delivery date were also excluded.

From the initial phase of the study, 244 patients are included for this analysis, the main results are: The mean age of the patients is 28 years (min 16 and max 45 years). The marital status of the patients is distributed as follows: 73 (28.6%) were married, 35 (13.7%) were single and 147
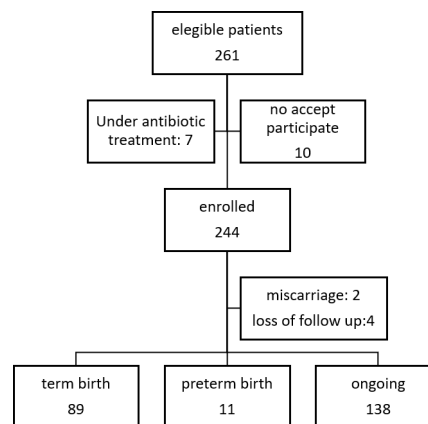
Figure 3 Patients flow chart.

(57.6%) were in a civil union. The occupation reported by the patients is 110 (43.3%) housewives, 86 (33.8%) employees, 44 (17.3%) freelancers and 14 (5.5%) students (for a total of 254). Concerning the social security plan, 141 (39.2%) are in the contributive system, 90 (35.35) in the subsidised, 15 (5.9%) in the special regimen and 9 (3.5%) are not in any social security plan (for a total of 255). The median body mass index is 25.3 kg/m^2 (min 16.5 and max 43.3). The mean gestational age at the time of the study is 13.2 weeks. The median CL is 38 mm (min 22 mm and max 56 mm). The median CCI: median 0.82 (min 0.64 and max 0.95).

In all cases, the amount of DNA extracted from vaginal fluid samples has been considered sufficient, with a mean DNA quantification of 275 ng/μL.

At the time of writing, 100 (39 %) patients had delivered, 11 patients (11%) delivered before 37 weeks.

Author affiliations

[1]Obstetrics and Gynecology Department, Health Faculty, Universidad Industrial de Santander, Bucaramanga, Colombia
[2]Maternal-fetal Medicine Unit, Hospital Universitario de Santander, Bucaramanga, Colombia
[3]Centro de Atencion Materno-fetal INUTERO, Bucaramanga, Colombia
[4]Obstetrics and Gynecology Department, Universidad Libre, Barranquilla, Colombia
[5]School of Medicine . Health Faculty, Universidad Industrial de Santander, Bucaramanga, Colombia
[6]Clinical Research Division, National Institute of Perinatology, Mexico City, Mexico
[7]Basics Sciences Department, Health Faculty, Universidad Industrial de Santander, Bucaramanga, Colombia

ORCID iDs
Carlos Hernan Becerra-Mojica http://orcid.org/0000-0003-3188-530X
Miguel Antonio Parra-Saavedra http://orcid.org/0000-0003-2287-000X
Luis Alfonso Diaz-Martinez http://orcid.org/0000-0002-4498-6639
Raigam Jafet Martinez-Portilla http://orcid.org/0000-0003-4711-3857
Bladimiro Rincon Orozco http://orcid.org/0000-0001-5323-0422

REFERENCES

1 March of Dimes, PMNCH, Save the children, WhoHowson CP, Kinney MV, Lawn JE, eds. *Born too soon: the global action report on preterm birth*. Geneva: World health organization, 2012. https://www.who.int/en/news-room/fact sheets/detail/pretermbirth

2 Liu L, Oza S, Hogan D, *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. *The Lancet* 2016;388:3027–35.

3 Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 2008;371:261–9.

4 Lozano R, Fullman N, Abate D, *et al*. Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related sustainable development goals for 195 countries and territories: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;392:2091–138.

5 Chawanpaiboon S, Vogel JP, Moller A-B, *et al*. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob Health* 2019;7:e37–46.

6 Sivigila. *Instituto Nacional de Salud*. Colombia, 2019.

7 WHO. *Who recommendations on interventions to improve preterm birth outcomes*, 2015: 98. ISBN: 9789241508988. https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/preterm-birth-guideline/en/

8 Han S, Crowther CA, Moore V. Magnesium maintenance therapy for preventing preterm birth after threatened preterm labour. *Cochrane Database Syst Rev* 2013;5:CD000940.

9 Naik Gaunekar N, Raman P, Bain E. Maintenance therapy with calcium channel blockers for preventing preterm birth after threatened preterm labour. *Cochrane Database Syst Rev* 2013;10:CD004071.

10 Papatsonis DNM, Flenady V, Liley HG. Maintenance therapy with oxytocin antagonists for inhibiting preterm birth after threatened preterm labour. *Cochrane Database Syst Rev* 2013;10:CD005938.

11 Passini R, Cecatti JG, Lajos GJ, *et al*. Brazilian multicentre study on preterm birth (EMIP): prevalence and factors associated with spontaneous preterm birth. *PLoS One* 2014;9:e109069.

12 Phillips C, Velji Z, Hanly C, *et al*. Risk of recurrent spontaneous preterm birth: a systematic review and meta-analysis. *BMJ Open* 2017;7:e015402.

13 Meis PJ, Klebanoff M, Thom E. National institute of child health and human development maternal-fetal medicine units network. Prevention of recurrent preterm delivery by 17 alpha hydroxyprogesterone caproate. *N Engl J Med* 2003;348:2379–85.

14 da Fonseca EB, Bittar RE, Carvalho MHB, *et al*. Prophylactic administration of progesterone by vaginal suppository to reduce the incidence of spontaneous preterm birth in women at increased risk:

15 Iams JD, Goldenberg RL, Meis PJ, *et al*. The length of the cervix and the risk of spontaneous premature delivery. National Institute of child health and human development maternal fetal medicine unit network. *N Engl J Med* 1996;334:567–72.

16 Heath VC, Southall TR, Souka AP, *et al*. Cervical length at 23 weeks of gestation: prediction of spontaneous preterm delivery. *Ultrasound Obstet Gynecol* 1998;12:312–7.

17 Celik E, To M, Gajewska K. Fetal medicine foundation second trimester screening group. cervical length and obstetric history predict spontaneous preterm birth: development and validation of a model to provide individualized risk assessment. *Ultrasound Obstet Gynecol* 2008;31:549–54.

18 Romero R, Conde-Agudelo A, Da Fonseca E, *et al*. Vaginal progesterone for preventing preterm birth and adverse perinatal outcomes in singleton gestations with a short cervix: a meta-analysis of individual patient data. *Am J Obstet Gynecol* 2018;218:161–80.

19 Guerby P, Girard M, Marcoux G, *et al*. Midtrimester cervical length in low-risk nulliparous women for the prediction of spontaneous preterm birth: should we consider a new definition of short cervix? *Am J Perinatol* 2021. doi:10.1055/s-0041-1728818. [Epub ahead of print: 03 May 2021].

20 Gudicha DW, Romero R, Kabiri D, *et al*. Personalized assessment of cervical length improves prediction of spontaneous preterm birth: a standard and a percentile calculator. *Am J Obstet Gynecol* 2021;224:288.e1–288.e17. doi:10.1016/j.ajog.2020.09.002

21 Parra-Saavedra M, Gómez L, Barrero A, *et al*. Prediction of preterm birth using the cervical consistency index. *Ultrasound Obstet Gynecol* 2011;38:44–51.

22 Chu DM, Seferovic M, Pace RM, *et al*. The microbiome in preterm birth. *Best Pract Res Clin Obstet Gynaecol* 2018;52:103–13.

23 Staude B, Oehmke F, Lauer T, *et al*. The microbiome and preterm birth: a change in paradigm with profound implications for pathophysiologic concepts and novel therapeutic strategies. *Biomed Res Int* 2018;2018:1–12.

24 Callahan BJ, DiGiulio DB, Goltsman DSA, *et al*. Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women. *Proc Natl Acad Sci U S A* 2017;114:9966–71.

25 D'Silva AM, Hyett JA, Coorssen JR. Proteomic analysis of first trimester maternal serum to identify candidate biomarkers potentially predictive of spontaneous preterm birth. *J Proteomics* 2018;178:31–42.

26 Dixon CL, Sheller-Miller S, Saade GR, *et al*. Amniotic fluid exosome proteomic profile exhibits unique pathways of term and preterm labor. *Endocrinology* 2018;159:2229–40.

27 Parry S, Leite R, Esplin MS, *et al*. Cervicovaginal fluid proteomic analysis to identify potential biomarkers for preterm birth. *Am J Obstet Gynecol* 2020;222:493.e1–493.e13.

28 Pereira L, Reddy AP, Jacob T, *et al*. Identification of novel protein biomarkers of preterm birth in human cervical-vaginal fluid. *J Proteome Res* 2007;6:1269–76.

29 Liang L, Rasmussen M-LH, Piening B, *et al*. Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women. *Cell* 2020;181:1680–92.

30 Beger RD, Dunn W, Schmidt MA, *et al*. Metabolomics enables precision medicine: "a white paper, community perspective". *Metabolomics* 2016;12:149.

31 Li Y, Kuhn M, Gavin A-C, *et al*. Identification of metabolites from tandem mass spectra with a machine learning approach utilizing structural features. *Bioinformatics* 2020;36:1213–8.

32 Melo CFOR, Navarro LC, de Oliveira DN, *et al*. A machine learning application based in random forest for integrating mass spectrometry-based metabolomic data: a simple screening method for patients with Zika virus. *Front Bioeng Biotechnol* 2018;6:31.

33 Romero R, Dey SK, Fisher SJ. Preterm labor: one syndrome, many causes. *Science* 2014;345:760–5.

34 Stevenson DK, Wong RJ, Aghaeepour N, *et al*. Towards personalized medicine in maternal and child health: integrating biologic and social determinants. *Pediatr Res* 2021;89:252–8.

35 Salomon LJ, Alfirevic Z, Bilardo CM, *et al*. ISUOG practice guidelines: performance of first-trimester fetal ultrasound scan. *Ultrasound Obstet Gynecol* 2013;41:102–13.

36 The fetal medicine foundation. The 11-13 weeks scan | Education | Welcome to the Fetal Medicine Foundation [Internet]; 2019. https://fetalmedicine.org/education/the-11-13-weeks-scan [Accessed cited 2019 Apr 23].

37 Heath VC, Southall TR, Souka AP, *et al*. Cervical length at 23 weeks of gestation: relation to demographic characteristics and previous obstetric history. *Ultrasound Obstet Gynecol* 1998;12:304–11.

38 Torrado-García LM, Martínez-Vega RA, Rincon-Orozco B. A novel strategy for cervical cancer prevention using Cervical-Vaginal self-collected samples shows high acceptability in women living in low-income conditions from Bucaramanga, Colombia. *Int J Womens Health* 2020;12:1197–204.

39 Celik E, To M, Gajewska K, *et al*. Cervical length and obstetric history predict spontaneous preterm birth: development and validation of a model to provide individualized risk assessment. *Ultrasound Obstet Gynecol* 2008;31:549–54.

40 Amaral-Zettler LA, McCliment EA, Ducklow HW, *et al*. A method for studying protistan diversity using massively parallel sequencing of v9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One* 2009;4. doi:10.1371/journal.pone.0006372. [Epub ahead of print: 23 Apr 2019].

41 Stoeck T, Bass D, Nebel M, *et al*. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* 2010;19 Suppl 1:21–31.

42 Keegan KP, Glass EM, Meyer F. MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. In: *Methods in molecular biology*. Clifton, NJ, 2016: 207–33. http://www.ncbi.nlm.nih.gov/pubmed/26791506

43 Huson DH, Beier S, Flade I, *et al*. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 2016;12. doi:10.1371/journal.pcbi.1004957. [Epub ahead of print: 23 Apr 2019].

44 Parks DH, Tyson GW, Hugenholtz P, *et al*. Stamp: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 2014;30:3123–4.