BMJ Open

Genital tract infections, the vaginal microbiome and gestational age at birth among pregnant women in South Africa: a cohort study protocol

Ranjana M S Gigi [1,2] Mandisa M Mdingi [2] Hyunsul Jung [3]
Shantelle Claassen-Weitz,[4] Lukas Bütikofer,[5] Jeffrey D Klausner,[6]
Christina A Muzny [7] Christopher M Taylor [8]
Janneke H H M van de Wijgert [9] Remco P H Peters,[2,3] Nicola Low [1]

For numbered affiliations see end of article.

Correspondence to
Professor Nicola Low;
nicola.low@unibe.ch

ABSTRACT

Introduction Preterm birth complications are the most common cause of death in children under 5 years. The presence of multiple microorganisms and genital tract inflammation could be the common mechanism driving early onset of labour. South Africa has high levels of preterm birth, genital tract infections and HIV infection among pregnant women. We plan to investigate associations between the presence of multiple lower genital tract microorganisms in pregnancy and gestational age at birth.

Methods and analysis This cohort study enrols around 600 pregnant women at one public healthcare facility in East London, South Africa. Eligible women are ≥18 years and at <27 weeks of gestation, confirmed by ultrasound. At enrolment and 30–34 weeks of pregnancy, participants receive on-site tests for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*, with treatment if test results are positive. At these visits, additional vaginal specimens are taken for: PCR detection and quantification of *Trichomonas vaginalis*, *Candida* spp., *Mycoplasma genitalium, M. hominis*, *Ureaplasma urealyticum* and *U. parvum*; microscopy and Nugent scoring; and for 16S ribosomal RNA gene sequencing and quantification. Pregnancy outcomes are collected from a postnatal visit and birth registers. The primary outcome is gestational age at birth. Statistical analyses will explore associations between specific microorganisms and gestational age at birth. To explore the association with the quantity of microorganisms, we will construct an index of microorganism load and use mixed-effects regression models and classification and regression tree analysis to examine which combinations of microorganisms contribute to earlier gestational age at birth.

Ethics and dissemination This protocol has approvals from the University of Cape Town Research Ethics Committee and the Canton of Bern Ethics Committee. Results from this study will be uploaded to preprint servers, submitted to open access peer-reviewed journals and presented at regional and international conferences.

Trial registration number NCT06131749; Pre-results.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This cohort study takes a holistic approach, investigating both the presence and quantity of multiple lower genital tract microorganisms, including vaginal microbiota, in pregnancy and their associations with gestational age at birth.

⇒ The study is set in a location where the prevalence of genital tract infections and adverse pregnancy outcomes is high, uses ultrasound scans to assess gestational age at enrolment accurately, and uses state-of-the-art molecular diagnostic methods.

⇒ The study setting is limited to one research site, which may affect the generalisability of the findings.

⇒ The use of gestational age at birth as a continuous outcome, instead of preterm birth as a dichotomous outcome, might limit comparability with other studies, but we will also examine the binary outcome preterm birth in secondary analyses.

INTRODUCTION

Preterm birth complications are the most common cause of death in children under 5 years.[1] Close to 1 million infants die every year because they are born preterm (before 37 completed weeks of gestation), mainly from infectious, respiratory and neurological complications, and those that survive can experience long-term morbidity.[1 2] South Africa has a high incidence of preterm birth at around 10%,[3] around 30% of women have one or more curable sexually transmitted infections during pregnancy[4 5] and about 30% of pregnant women are living with HIV.[6]

Microbial colonisation or infection during pregnancy, in the lower or upper genital tract, has been reported to predispose to preterm birth, as do anatomical, biochemical, endocrinological, immunological, nutritional, environmental and psychosocial factors.[7 8] The presence of microorganisms

may contribute to early onset of labour directly, through presumed ascension from the lower to the upper genital tract, or indirectly, through a pathway of inflammatory response, or a combination of both.[7 9] Inflammation may be the common pathway, even if infection has not reached the amniotic cavity.[10]

Much of the research reporting on the role of sexually transmitted infections in pregnancy and preterm birth has focused on single infections, such as *Chlamydia trachomatis*,[11] *Neisseria gonorrhoeae*[12] and *Trichomonas vaginalis*.[13] *Mycoplasma genitalium* is the most recently recognised bacterial sexually transmitted infection and, while an association with preterm birth has been reported, there are few studies with prospective data collection.[14] Bacterial vaginosis is the most common vaginal microbiota dysbiosis and is associated with adverse pregnancy outcomes, either alone or in combination with other sexually transmitted infections.[15–17] Associations with adverse birth outcomes have also been observed for other genital mycoplasmas, *M. hominis*, *Ureaplasma urealyticum* and *U. parvum*.[18] For individual sexually transmitted infections, bacterial vaginosis and colonisation by other genital mycoplasmas, summary ORs for the association with adverse birth outcomes in meta-analyses of univariable data are generally around 1.3–2.0.[11–14 16 18] *Candida* spp have not been found to be associated with preterm birth, but an association with more inflammatory, symptomatic yeast infection cannot be ruled out.[19] Most studies about these microorganisms do not present analyses that examine the role of co-occurrence or control for confounding factors, so the presence or strength of the causal association cannot be assessed.[20] It is also important to include women living with HIV, among whom there are fewer studies about associations between genital tract infections and adverse birth outcomes than among women without HIV infection.[21 22]

The importance of the quantity of different microorganisms as a driver of preterm birth has not been extensively studied,[23–25] but might be as, or more, relevant than their presence.[23] Together with inflammation or immune activation in the genital tract during pregnancy, organism load could be an important driver of the early onset of labour and preterm birth.[8 23 24 26 27] This calls for a holistic approach to research studies, which combines information about the presence of different microorganisms, the quantified load and the microbiome, sociodemographic factors and HIV among women living with infection, most of whom are receiving antiretroviral therapy. The overall aim of this study is to investigate associations between the presence of lower genital tract microorganisms in pregnancy and preterm birth and other adverse pregnancy outcomes. This will be achieved through three objectives to explore: (1) the association between the presence of specific lower genital tract microorganisms and gestational age at birth (primary outcome), as well as secondary adverse pregnancy outcomes; (2) the association between quantified load of vaginal and sexually transmitted microorganisms and gestational age at birth (primary outcome) as well as secondary adverse pregnancy outcomes; and (3) the combinations of microorganisms that are most strongly associated with earlier gestational age at birth.

METHODS AND ANALYSIS
Study design and setting
This prospective closed cohort study follows women enrolled during pregnancy until after they give birth (figure 1). The study is conducted at the antenatal clinic

Enrolment visit
<27 weeks of gestation

Eligibility assessment
↓
Written consent
↓
Questionnaire
↓
Physical exam with ultrasound
↓
Vaginal specimen collection:
2x smear for microscopy
1x swab for CT/NG testing
2x swabs for microbiome
1x swab for microbiology
1x swab for storage

Third trimester visit
30-34 weeks of gestation

Questionnaire
↓
Physical exam
↓
Vaginal specimen collection:
2x smear for microscopy
1x swab for CT/NG testing
2x swabs for microbiome
1x swab for microbiology
1x swab for storage

Post-natal visit
3-6 days post-natal

Questionnaire
for collection of birth outcome

Birth register at birth clinic
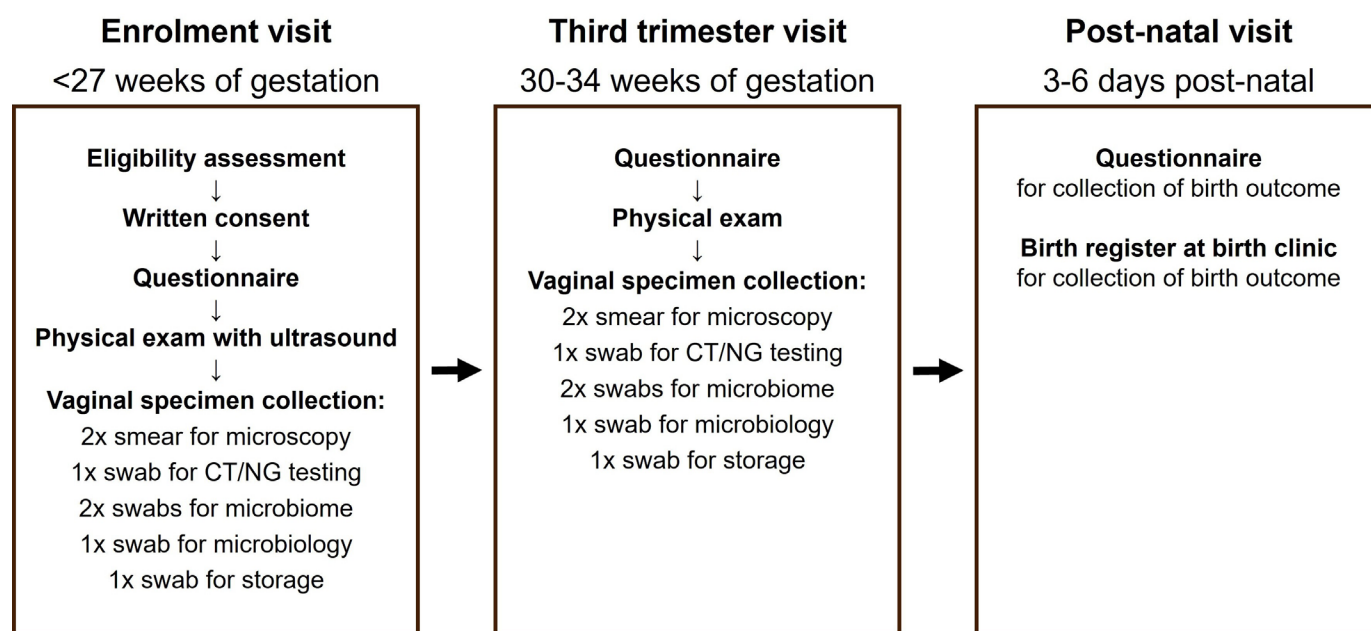for collection of birth outcome

Figure 1 Study visits and specimen collection. CT, *Chlamydia trachomatis*; NG, *Neisseria gonorrhoeae*.

of one primary healthcare facility in Buffalo City Metropolitan Municipality, Eastern Cape Province, South Africa. This cohort study is part of a larger project, called Philani Ndiphile (meaning 'be healthy and I will be healthy' in isiXhosa), which includes a randomised implementation-effectiveness trial of screening strategies for sexually transmitted infections in pregnancy[28] and a case–control study about persistent *C. trachomatis* infection.

Participants

Inclusion criteria: Pregnant women aged 18 years or older, who live in Buffalo City Metropolitan Municipality, intend to deliver in the same municipality and provide written informed consent to take part in the study. The eligible gestational age at enrolment, confirmed by ultrasound, was below 20 weeks at the start of the study in March 2021 and was increased to 27 weeks in September 2021 to increase enrolment and to align with another trial.[29]

Exclusion criteria: Participation in any other research study or inability to understand and speak a local language (English, Afrikaans or isiXhosa).

Enrolment

A trained study field worker approaches all pregnant women attending an antenatal care visit at the clinic and individually informs them about the study. If a potential participant shows interest in the study, the study field worker checks for eligibility. The date of the last menstrual period is used initially to estimate gestational age. If all eligibility criteria are met, a study field worker obtains written informed consent from the participant.

Study procedures and visits

At the enrolment visit, study field workers administer a questionnaire to record sociodemographic, behavioural and clinical information in an online Research Electronic Data Capture (REDCap)[30] software database. The study nurse examines the woman, according to the South African government standard of care.[31] As an additional procedure, a study nurse with training in obstetric ultrasound performs an abdominal ultrasound to estimate the gestational age. If this is later than the eligibility criterion, the participant is excluded from any further study activity. A study nurse collects vaginal samples (figure 1) for on-site testing for *C. trachomatis* and *N. gonorrhoeae* using the Xpert CT/NG assay on the Gene Xpert platform (Cepheid, Sunnyvale, California, USA) and for further off-site laboratory testing (see 'Specimen collection and analysis').

If the test result for *C. trachomatis* or *N. gonorrhoeae* is positive, the woman receives immediate antibiotic treatment if still on-site or is contacted by telephone and asked to return to the clinic for treatment. Antibiotic treatments are first-line regimens according to South African guidelines: for *C. trachomatis*, 1 g oral azithromycin and for *N. gonorrhoeae*, 500 mg intramuscular ceftriaxone (250 mg until South African treatment guidelines for sexually transmitted infections changed in December 2022).[32] Women with vaginal discharge syndrome but with negative Xpert test results for *C. trachomatis* and *N. gonorrhoeae* receive empirical treatment for trichomoniasis with metronidazole 400 mg twice a day for 7 days. The study nurse gives advice to women with *C. trachomatis* or *N. gonorrhoeae* on safe disclosure of her diagnosis to her partner(s) and gives her a notification slip(s) to request her partner(s) to attend a clinic for treatment.

A follow-up visit at 30–34 weeks (third trimester visit) is scheduled at which clinical and obstetric information, as well as the same vaginal specimens, is collected and treatment given, if indicated.

A postnatal visit is scheduled for 3–6 days after giving birth, according to the South African government standard.[31] A study nurse collects information about the birth outcome and perinatal period through a questionnaire with the mother, a patient-held medical record of the baby (the Road to Health card) and/or the birth register from the public birth clinics within the study area. If the participant does not attend the postnatal visit, study staff telephone her to ask her to return to the clinic. If the participant is not able to return to the clinic, the study physician collects the information by telephone or from the birth register.

Outcomes

The primary outcome is gestational age at birth, measured in days, based on the ultrasound assessment at the enrolment visit. Secondary outcomes are preterm birth (<37 completed weeks of gestation), low birth weight (birth weight <2500 g), miscarriage (dead fetus delivered before 28 completed weeks of pregnancy or with birth weight below 1000 g) and stillbirth (dead fetus delivered at or after 28 completed weeks of pregnancy or with birth weight above 1000 g).[33 34] We chose gestational age at birth as the primary outcome because, while the cut-off of 37 weeks is the standard definition of preterm birth, dichotomisation of a continuous variable results in a loss of statistical power.[35]

Specimen collection and analysis
Data sources and variables

The source data are case report forms recording questionnaire data for the enrolment, third trimester and postnatal visits, and forms for laboratory and specimen results, which are stored in REDCap, a secure web-based database[30] (online supplemental file), hosted by the Foundation for Professional Development, Pretoria, South Africa.

Specimen collection

At the enrolment and the third trimester visits, a study nurse collects two vaginal smears using inoculation loops and air dries them on glass slides. She then collects vaginal specimens by inserting swabs into the vagina up to a mark at 4 cm and rotating around the vaginal wall. Five swabs are collected in the following order: one Cepheid GeneXpert Xpert Vaginal/Endocervical Swab

in a tube with Xpert Swab Transport Reagent (Cepheid); two Qiagen digene Female Swabs in a single tube with digene Specimen Transport Medium (Qiagen, Hilden, Germany); and two dry FLOQSwabs (COPAN, Brescia, Italy) each in a separate sterile tube (figure 1).

Transport and storage of specimens

Vaginal smear glass slides are stored and transported in plastic slide carriers at room temperature. All vaginal swabs are initially stored at the clinic in a refrigerator (2–8°C with daily temperature checks). All vaginal swabs, except the Xpert swab, which is tested on-site, are transported on ice packs once a week by overnight road courier to the laboratory at the Department of Medical Microbiology, University of Pretoria, where they are also stored in a refrigerator until DNA extraction.

Microbiological analyses

The Xpert vaginal swabs are tested on-site using the Xpert CT/NG assay (Cepheid) to detect *C. trachomatis* and *N. gonorrhoeae*, as per manufacturer's instructions. At the University of Pretoria, air-dried vaginal smears are heat fixed and Gram stained.[36] Two qualified people record the Nugent scores (0–3: normal; 4–6: intermediate; 7–10: bacterial vaginosis) and the presence of yeasts.[37] In case of discrepancies, a third person assesses the slide and consensus is reached by discussion. At the University of Pretoria, one vaginal FLOQSwabs is used for PCRs. The genomic DNA is extracted using the High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany) as per manufacturer's instructions. Real-time PCR assays are then performed using the Light-Cycler 480 Probes Master Kit (Roche Diagnostics) on the LightCycler 480 II instrument (Roche Diagnostics). Previously published primer and hydrolysis probe sequences and cycling conditions are used for detection and quantification of *M. genitalium*,[38] *M. hominis*,[39] *U. parvum*,[40] *U. urealyticum*,[40] *T. vaginalis*[41] and *Candida* spp.[42 43] The load for each assayed microorganism detected in vaginal swab specimens by real-time PCR or GeneXpert is obtained from the cycle threshold value.

Vaginal microbiome laboratory analyses

The vaginal swabs stored in Qiagen digene Specimen Transport Medium will be used for DNA extraction and subsequent 16S ribosomal RNA (rRNA) amplicon sequencing targeting the V3–V4 hypervariable regions for vaginal microbiota analyses at the Division of Medical Microbiology, University of Cape Town.

A commercial DNA extraction kit will be used and a bead-beating step included.[44] A DNA isolation control will be prepared from an unused vaginal swab specimen during this process. Two PCR rounds will be conducted to prepare amplicon libraries.[45] The aim of the first PCR round is to amplify 16S rRNA gene V3–V4 regions using the 319F 5′-ACTCCTACGGGAGGCAGCAG-3′ forward primer and 806R 5′-GGACTACHVGGGTWTCTAAT-3′ reverse primer. The aim of the second PCR round is to barcode the V3–V4 amplicons by a dual-index approach, permitting multiplexing of up to 384 samples (including controls). Amplicon concentrations for all sample libraries are measured and normalised to form a mixed loading library. The libraries will be sequenced on an Illumina MiSeq instrument (Illumina, San Diego, California, USA), 2×300 bp. To quantify the number of 16S rDNA copies per swab, a quantitative PCR using the same forward and reverse primers as described above will be used. Samples from enrolment and third trimester visits from the same woman will be processed in the same run.

Vaginal microbiota bioinformatics

Raw sequencing reads will be processed using an established bioinformatics pipeline.[46] Taxonomic assignment of amplicon sequence variants will be done in DADA2[47] with SILVA[48] as the reference database. Vaginal microbiome composition data will be visualised in heatmaps and diagrams. For each vaginal sample, we will calculate diversity measures (alpha diversity), relative abundances and estimated concentrations of key vaginal bacteria and bacterial groups, as described.[46] We will use the entire sequencing dataset to design vaginal microbiota types by hierarchical clustering, and each sample will be assigned to one vaginal microbiota type.

Sample size calculation

The sample size has been calculated for objective 1, with a univariable comparison between the presence of a genital tract microorganism in the mother and gestational age at birth. Figure 2 shows that, for any vaginal or sexually transmitted microorganism, or vaginal microbiota type that has a prevalence of 10% or more among all enrolled women, about 500–600 patients provide adequate power (80%) to detect a 1-week difference (with SD 2) in mean gestational age between the two groups using Student's t-test. Specifying an alpha of 0.83% allows for multiple hypothesis testing (six hypotheses, using a Bonferroni correction). We enrol around 600 women and aim to have complete follow-up and outcome data on at least 550 women.

Study timeline

Enrolment began on 28 March 2021, with an estimated date for reaching the target sample size in August 2023. Follow-up of all participants until the postnatal visit is expected to be completed by March 2024.

Statistical analysis

This description gives an overview of the statistical methods for each objective. A detailed statistical analysis plan will be published separately and made publicly available.

We will describe the numbers of women enrolled and available at each follow-up visit in a flow chart. We will present descriptive tables of sociodemographic, behavioural and clinical characteristics and compare women with complete follow-up with those lost to follow-up.
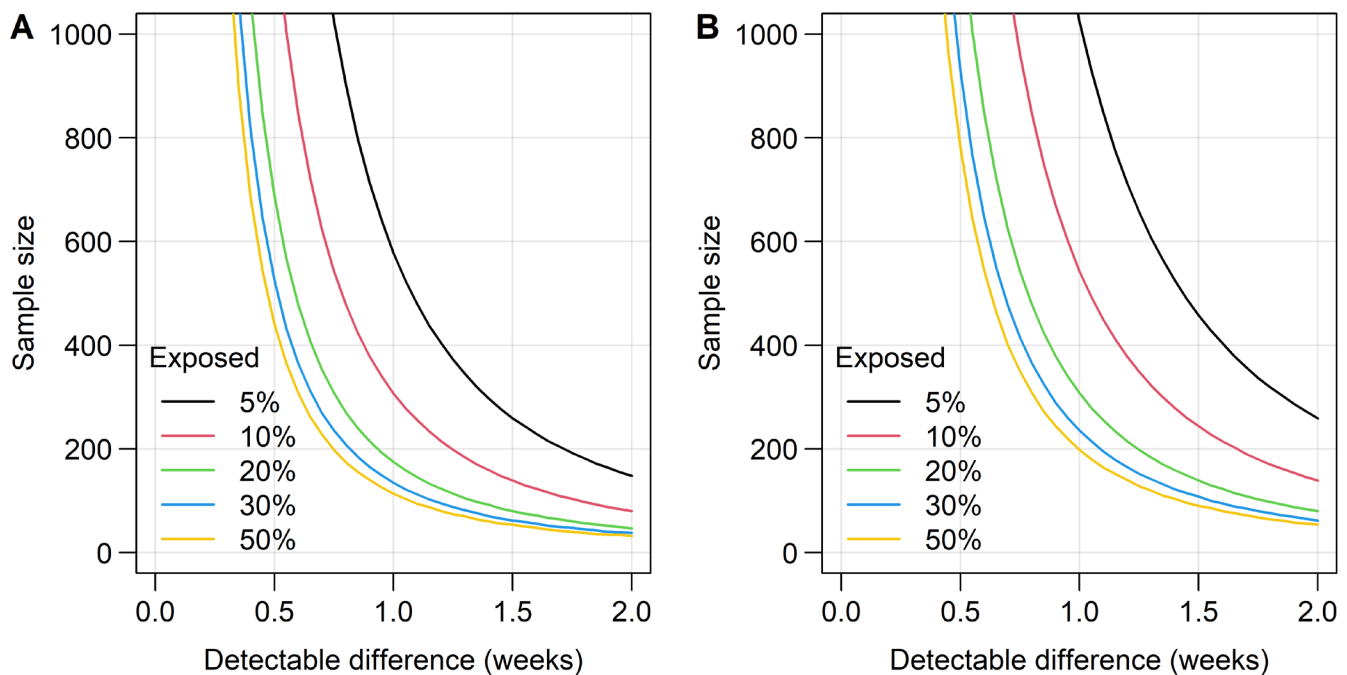
Figure 2 Sample size requirements at different levels of exposure prevalence with power of 80% and alpha 0.83% based on Student's t-test. (A) SD 1.5. (B) SD 2.0. The curves for % exposed are symmetrical around a prevalence of 50%, that is, curve for 10% exposed is same as that for 90% exposed.

Objective 1: association between specified exposures and pregnancy outcomes

1. We will examine a primary set of microorganisms as exposures, detected at either enrolment or at the third trimester visit: *M. genitalium, M. hominis, U. urealyticum, U. parvum, T. vaginalis* and *Candida* spp are the microorganisms for which women did not receive diagnostic tests and treatment during study visits. We will use the mean and SD for the continuous outcome (gestational age) and absolute and relative frequencies for the binary outcomes (all secondary outcomes). Gestational age at birth for each exposure will be compared using Student's t-test and a mean difference with 95% CIs (continuous outcome) and Fisher's exact test and a risk difference with 95% CIs (binary outcomes).

To control for confounding, multivariable regression models will be fitted to all outcomes with a set of prespecified potential confounders (age, educational level, alcohol consumption, HIV infection, prior preterm birth) for all organisms. For analyses of *M. genitalium, T. vaginalis* and *Candida* spp, we will also control for bacterial vaginosis (Nugent score 7–10). The other genital mycoplasmas can be identified from 16S rRNA amplicon sequencing in women with vaginal dysbiosis, so are sometimes considered part of bacterial vaginosis. For these organisms, we will conduct descriptive analyses, stratified by the presence of bacterial vaginosis. For continuous confounders, a linear relationship will be assumed by default but transformations (eg, log) or more flexible approaches (eg, splines or fractional polynomials) will be considered if there is evidence for non-linearity. For the continuous outcome, we will use linear mixed-effects regression models (including data from either visit and the participant as random effect) and report the result as mean difference with 95% CIs. For the binary outcome, we will use logistic mixed-effects regression and report the result as ORs with 95% CIs.

2. Comparisons for associations with timing of detection, other microorganism exposures and birth outcomes will be considered secondary analyses. Associations between vaginal microbiota composition and pregnancy outcomes will be assessed. We will use compositional multivariable analysis methods to identify bacterial taxa that are differentially abundant between binary pregnancy outcome groups at the level of individual taxon relative abundances. We will use mixed-effects models (with the individual participant as the random effect and including data from both visits) to assess associations between continuous and binary pregnancy outcome and the following fixed effects derived from the vaginal microbiota data: alpha diversity, vaginal microbiota types and absolute abundances of predefined bacterial groups.[46] These models will be adjusted for confounding as described in the previous paragraph.

Objective 2: association between quantified microorganism load and pregnancy outcomes

We will investigate the hypothesis that the quantity of microorganisms with inflammatory potential is associated with gestational age at birth. For this, we will analyse the vaginal microbiota data jointly with sexually transmitted infections and *Candida* spp diagnostic test results during pregnancy (these will be considered as

additional covariates in the above-mentioned regression models). We will develop a 'vaginal inflammation index' based on quantification of the vaginal microbiota and their inflammatory potential[49] and of yeasts. This vaginal inflammation index will also be analysed as a fixed effect in mixed-effects models with pregnancy outcomes as the outcomes; these models will not include any of the infection parameters that were used to design the index.

Objective 3: classification and regression tree analysis for the primary outcome

We will conduct exploratory analyses to examine the combination of microorganisms that best predicts earlier gestational age at birth using classification and regression tree analysis.[50] This method belongs to the family of decision tree machine learning algorithms and allows for non-parametric analyses of a large number of binary, categorical or continuous predictors. They are typically easy to interpret and can detect predictors with small marginal effects when there are strong interaction effects. We will make use of the predictive potential for gestational age at birth of all sexually transmitted and genital tract microorganisms, including individual bacterial taxa or bacterial groups identified by 16S rRNA gene amplicon sequencing (as binary or continuous variables) and confounding variables identified in objective 1. We will present variable importance scores and curves of marginal effects to show how prediction of the outcome changes at different levels of the exposure of each variable in the model. To avoid overfitting, we will consider bootstrap aggregating via random forests.[51]

Data management and confidentiality
Data management

Each potential participant screened for eligibility is assigned a unique participant identification number, which does not include any personal identifying information. Personal identification numbers are used to link records, specimens and laboratory test results of the participants. Data are stored in a REDCap database,[30] which is only accessible to authorised project staff. Paper records are kept in lockable fire-resistant filing cabinets. Laboratory records and journals are kept at the University of Pretoria and University of Cape Town. Forms with personal identifying information are kept separately from demographic, clinical and other data. The data manager maintains a separate, access-controlled database that links the personal identification number with identifying information. Data quality checks are conducted by study staff on-site and data administrators at the office of the Foundation for Professional Development. All study data are stored securely at the offices of the Foundation for Professional Development in East London for up to 5 years after the completion of the study or as required by the institutional review board.

Confidentiality

The research team is trained to adhere to guidelines on the Protection of Human Research Participants and Good Clinical Practice and fully protects the confidentiality

of participants. Besides the measures described under data management, interviews are conducted in a private setting. In reports and publications, data will not be presented in a way where it could be linked to individual participants.

Patient and public involvement

There was no involvement of patients or the public in the development of the research questions or the study methods. The research findings will be shared through open access publications and in dissemination meetings with local stakeholders, healthcare providers and communities.

DISCUSSION

This project is important because of its holistic approach, which considers associations between different genital tract infections, their quantity and the vaginal microbiota on earlier gestational age at birth. Many studies in this field have focused on only one or two microorganisms and few studies involve women in sub-Saharan Africa. Strengths of this study include the study setting, where the prevalence of both genital tract infections and adverse pregnancy outcomes is high, the use of ultrasound scans at enrolment for accurate assessment of gestational age and the use of state-of-the-art molecular diagnostic tests and 16S rRNA sequencing. The residual DNA from samples collected in this study will be available for future studies, including joint analyses with other studies of the influence of vaginal microbiota on adverse pregnancy outcomes.

There are limitations to the study design. First, this study involves participants from one clinic, which might limit the generalisability of the findings. Second, using gestational age at delivery as a continuous outcome instead of preterm birth as a dichotomous outcome might limit comparability with other studies. We will, however, examine the binary outcome preterm birth in secondary analyses. Third, the vaginal samples are taken in a fixed sequence at each visit, which might reduce the microorganism load of later samples. Fourth, the development of the vaginal inflammation index will use information about the inflammatory potential of microorganisms,[49] rather than direct concentrations of inflammatory markers.

This study has the potential to generate new evidence about the role of different microorganisms in earlier gestational age at birth through analyses of the presence and quantity of individual and combinations of microorganisms, relative abundance of bacterial genera and microbiota on gestational age at birth. This study will generate new hypotheses, which can be investigated in future studies.

ETHICS AND DISSEMINATION

This protocol and the informed consent forms are approved by the University of Cape Town Research Ethics

Committee (reference: 676/2019), which includes activities at the University of Southern California, University of Alabama at Birmingham and Louisiana State University. Authorisation to analyse deidentified data at the University of Bern has been granted by the Canton of Bern Ethics Committee (reference: 2021-01209). Results from this study will be submitted to regional and international conferences and to open access peer-reviewed journals and preprint servers.

DATA STATEMENT

The research team will prepare datasets used in analyses, in accordance with data sharing requirements of open access journals in which manuscripts are published and in compliance with local Protection of Personal Information Act requirements. These data files will be archived with codebooks as .csv documents or R datasets and stored in REDCap. The final data files will not contain any personal identifying information of participants.

Author affiliations
¹Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
²Research Unit, Foundation for Professional Development, East London, South Africa
³Department of Medical Microbiology, University of Pretoria, Pretoria, South Africa
⁴Department of Pathology, University of Cape Town, Cape Town, South Africa
⁵CTU Bern, Department of Clinical Research, University of Bern, Bern, Switzerland
⁶Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, USA
⁷Division of Infectious Diseases, University of Alabama at Birmingham, Birmingham, Alabama, USA
⁸Department of Microbiology, Immunology, and Parasitology, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA
⁹Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

Twitter Jeffrey D Klausner @drklausner, Christina A Muzny @ChristinaMuzny and Nicola Low @nicolamlow

Contributors RMSG, NL, JHHMvdW and RPHP conceived and designed the study. JDK supported the study through the design of the parent study of the Philani Ndiphile project. RMSG, NL, JHHMvdW, RPHP, HJ, CMT, CAM, SC-W and LB contributed to the data analysis plan. RMSG, NL, RPHP, MMM and HJ were involved in the implementation and management of the study. RMSG and MMM managed the data acquisition. RMSG, NL, JHHMvdW and RPHP drafted the manuscript and all authors revised it. NL, JHHMvdW, RPHP and JDK supervised the study. All authors read and approved the final manuscript.

Competing interests CAM has received research grant funding to her institution by Gilead, Abbott Molecular, Visby and Lupin Pharmaceuticals. She is a consultant to BioNTech, Cepheid and BioFire Diagnostics. She has received honoraria for educational presentations and review activities from Scynexis, Visby, Abbott, Elsevier, UpToDate and DynaMed Plus.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Ranjana M S Gigi http://orcid.org/0000-0001-7926-4153
Mandisa M Mdingi http://orcid.org/0000-0002-2795-0343
Hyunsul Jung http://orcid.org/0000-0001-9331-6300
Christina A Muzny http://orcid.org/0000-0002-4005-3858
Christopher M Taylor http://orcid.org/0000-0002-7983-5275
Janneke H H M van de Wijgert http://orcid.org/0000-0003-2728-4560
Nicola Low http://orcid.org/0000-0003-4817-8986

REFERENCES

1. UN Inter-agency Group for Child Mortality Estimation. *Levels & Trends in Child Mortality: Report 2019*. New York: United Nations Children's Fund, 2019. Available: https://www.unicef.org/reports/levels-and-trends-child-mortality-report-2019
2. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 2008;371:261–9.
3. World Health Organization. *Born too soon: decade of action on preterm birth*. Geneva: World Health Organization, 2023. Available: https://www.who.int/publications/i/item/9789240073890
4. Joseph Davey DL, Nyemba DC, Gomba Y, *et al*. Prevalence and correlates of sexually transmitted infections in pregnancy in HIV-infected and- uninfected women in Cape Town, South Africa. *PLoS One* 2019;14:e0218349.
5. Moodley D, Moodley P, Sebitloane M, *et al*. High prevalence and incidence of asymptomatic sexually transmitted infections during pregnancy and postdelivery in KwaZulu Natal, South Africa. *Sex Transm Dis* 2015;42:43–7.
6. Woldesenbet SA, Lombard C, Manda S, *et al*. *The 2019 National Antenatal Sentinel HIV Survey (ANCHSS) Key Findings*. South Africa: National Department of Health, 2021. Available: https://www.nicd.ac.za/wp-content/uploads/2019/07/Antenatal_survey-report_24July19.pdf
7. Romero R, Espinoza J, Kusanovic JP, *et al*. The preterm parturition syndrome. *BJOG* 2006;113:17–42.
8. Hunter PJ, Awoyemi T, Ayede AI, *et al*. Biological and pathological mechanisms leading to the birth of a small vulnerable newborn. *Lancet* 2023;401:1720–32.
9. Goldenberg RL, Culhane JF, Iams JD, *et al*. Epidemiology and causes of preterm birth. *Lancet* 2008;371:75–84.
10. MacIntyre DA, Sykes L, Teoh TG, *et al*. Prevention of preterm labour via the modulation of inflammatory pathways. *J Matern Fetal Neonatal Med* 2012;25 Suppl 1:17–20.
11. Tang W, Mao J, Li KT, *et al*. Pregnancy and fertility-related adverse outcomes associated with Chlamydia trachomatis infection: a global systematic review and meta-analysis. *Sex Transm Infect* 2020;96:322–9.
12. Vallely LM, Egli-Gany D, Wand H, *et al*. Adverse pregnancy and neonatal outcomes associated with Neisseria gonorrhoeae: systematic review and meta-analysis. *Sex Transm Infect* 2021;97:104–11.
13. Van Gerwen OT, Craig-Kuhn MC, Jones AT, *et al*. Trichomoniasis and adverse birth outcomes: a systematic review and meta-analysis. *BJOG* 2021;128:1907–15.
14. Frenzer C, Egli-Gany D, Vallely LM, *et al*. Adverse pregnancy and perinatal outcomes associated with Mycoplasma genitalium: systematic review and meta-analysis. *Sex Transm Infect* 2022;98:222–7.
15. Hillier SL, NugentRP, Eschenbach DA, *et al*. Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. *N Engl J Med* 1995;333:1737–42.
16. Leitich H, Bodner-Adler B, Brunbauer M, *et al*. Bacterial vaginosis as a risk factor for preterm delivery: a meta-analysis. *Am J Obstet Gynecol* 2003;189:139–47.

17 van de Wijgert JHHM. The vaginal microbiome and sexually transmitted infections are interlinked: consequences for treatment and prevention. *PLoS Med* 2017;14:e1002478.

18 Jonduo ME, Vallely LM, Wand H, *et al*. Adverse pregnancy and birth outcomes associated with Mycoplasma hominis, Ureaplasma urealyticum and Ureaplasma parvum: a systematic review and meta-analysis. *BMJ Open* 2022;12:e062990.

19 Gigi RMS, Buitrago-Garcia D, Taghavi K, *et al*. Vulvovaginal yeast infections during pregnancy and perinatal outcomes: systematic review and meta-analysis. *BMC Womens Health* 2023;23:116.

20 Low N. Chlamydia trachomatis and reproductive health: what can we learn from systematic reviews of observational studies *Sex Transm Infect* 2020;96:315–7.

21 Moodley D, Sartorius B, Madurai S, *et al*. Pregnancy outcomes in association with STDs including genital HSV-2 shedding in a South African cohort study. *Sex Transm Infect* 2017;93:460–6.

22 Adachi K, Klausner JD, Bristow CC, *et al*. Chlamydia and gonorrhea in HIV-infected pregnant women and infant HIV transmission. *Sex Transm Dis* 2015;42:554–65.

23 Goodfellow L, Verwijs MC, Care A, *et al*. Vaginal bacterial load in the second trimester is associated with early preterm birth recurrence: a nested case-control study. *BJOG* 2021;128:2061–72.

24 Bretelle F, Rozenberg P, Pascal A, *et al*. High Atopobium vaginae and Gardnerella vaginalis vaginal loads are associated with preterm birth. *Clin Infect Dis* 2015;60:860–7.

25 Freitas AC, Bocking A, Hill JE, *et al*. Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. *Microbiome* 2018;6:117.

26 Goldenberg RL, Hauth JC, Andrews WW. Intrauterine infection and preterm delivery. *N Engl J Med* 2000;342:1500–7.

27 Andrews WW, Hauth JC, Goldenberg RL, *et al*. Amniotic fluid Interleukin-6: correlation with upper genital tract microbial colonization and gestational age in women delivered after spontaneous labor versus indicated delivery. *Am J Obstet Gynecol* 1995;173:606–12.

28 Medina-Marino A, Cleary S, Muzny CA, *et al*. Sexually transmitted infection screening to prevent adverse birth and newborn outcomes: study protocol for a randomized-controlled hybrid-effectiveness trial. *Trials* 2022;23:441.

29 Vallely AJ, Pomat WS, Homer C, *et al*. Point-of-care testing and treatment of sexually transmitted infections to improve birth outcomes in high-burden, low-income settings: study protocol for a cluster randomized crossover trial (the WANTAIM trial, Papua New Guinea). *Wellcome Open Res* 2019;4:53.

30 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDcap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.

31 South African National Department of Health. *Guidelines for Maternal Care in South Africa: A manual for clinics, community health centres and district hospitals*. Pretoria: Department of Health, 2016. Available: https://knowledgehub.health.gov.za/elibrary/guidelines-maternity-care-south-africa-2016

32 Peters RPH, Garrett N, Chandiwana N, *et al*. Southern African HIV Clinicians Society 2022 guideline for the management of sexually transmitted infections: Moving towards best practice. *South Afr J HIV Med* 2022;23:1450.

33 World Health Organization. *ICD-10 International statistical classification of diseases and related health problems*. Geneva: World Health Organization, 2016. Available: https://icd.who.int/training/icd10training/ICD-10%20training/ICD-10_Resources/ICD-10_Volume_1.pdf

34 March of Dimes, PMNCH, Save the Children, WHO. *Born Too Soon: The Global Action Report on Preterm Birth*. Geneva: World Health Organization, 2012. Available: https://www.who.int/publications/i/item/9789241503433

35 Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006;332:1080.

36 Jordan J. *Laboratory Procedure Manual: Bacterial Vaginosis in Swabs*. Atlanta: Centers for Disease Control and Prevention, 2004. Available: https://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/l34_c_met_bv.pdf

37 Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J Clin Microbiol* 1991;29:297–301.

38 Edberg A, Jurstrand M, Johansson E, *et al*. A comparative study of three different PCR assays for detection of Mycoplasma genitalium in urogenital specimens from men and women. *J Med Microbiol* 2008;57:304–9.

39 Férandon C, Peuchant O, Janis C, *et al*. Development of a real-time PCR targeting the yidC gene for the detection of Mycoplasma hominis and comparison with quantitative culture. *Clin Microbiol Infect* 2011;17:155–9.

40 Frølund M, Björnelius E, Lidbrink P, *et al*. Comparison between culture and a multiplex quantitative real-time polymerase chain reaction assay detecting Ureaplasma urealyticum and U. parvum. *PLoS One* 2014;9:e102743.

41 Pillay A, Radebe F, Fehler G, *et al*. Comparison of a Taqman-based real-time polymerase chain reaction with conventional tests for the detection of Trichomonas vaginalis. *Sex Transm Infect* 2007;83:126–9.

42 Fujita S, Lasker BA, Lott TJ, *et al*. Microtitration plate enzyme immunoassay to detect PCR-amplified DNA from Candida species in blood. *J Clin Microbiol* 1995;33:962–7.

43 Shin JH, Nolte FS, Holloway BP, *et al*. Rapid identification of up to three Candida species in a single reaction tube by a 5' exonuclease assay using fluorescent DNA probes. *J Clin Microbiol* 1999;37:165–70.

44 Villette R, Autaa G, Hind S, *et al*. Refinement of 16S rRNA gene analysis for low biomass biospecimens. *Sci Rep* 2021;11:10741.

45 Holm JB, Humphrys MS, Robinson CK, *et al*. Ultrahigh-throughput multiplexing and sequencing of >500-base-pair Amplicon regions on the Illumina Hiseq 2500 platform. *mSystems* 2019;4.

46 van de Wijgert JHHM, Verwijs MC, Agaba SK, *et al*. Intermittent lactobacilli-containing vaginal probiotic or metronidazole use to prevent bacterial vaginosis recurrence: A pilot study incorporating microscopy and sequencing. *Sci Rep* 2020;10:3884.

47 Callahan BJ, McMurdie PJ, Rosen MJ, *et al*. Dada2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581–3.

48 Pruesse E, Quast C, Knittel K, *et al*. SILVA: a comprehensive online resource for quality checked and aligned Ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 2007;35:7188–96.

49 Kyongo JK, Crucitti T, Menten J, *et al*. Cross-sectional analysis of selected genital tract immunological markers and molecular vaginal microbiota in sub-Saharan African women, with relevance to HIV risk and prevention. *Clin Vaccine Immunol* 2015;22:526–38.

50 Breiman L, Friedman JH, Olshen RA. *Classification and regression trees*. Monterey: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

51 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.