Vaginal bacterial load in the second trimester is associated with early preterm birth recurrence: a nested case-control study

Authors:
Laura Goodfellow, Marijn C. Verwijs, Angharad Care, Andrew Sharp, Jelena Ivandic, Borna Poljak, Devender Roberts, Christina Bronowski, A. Christina Gill, Alistair C. Darby, Ana Alfirevic, Bertram Muller-Myhsok, Zarko Alfirevic, Janneke H.H.M. van de Wijgert

Affiliations:
1. Harris Wellbeing Research Centre, Department of Women’s and Children’s Health, University of Liverpool, United Kingdom.
2. Department of Clinical Infection, Microbiology and Immunology, Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom.
3. Liverpool Women’s NHS Foundation Trust, Liverpool, United Kingdom.
4. Centre for Genomic Research, University of Liverpool, Liverpool.
5. Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany.
6. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.

*Corresponding author. Address: University Department, Liverpool Women’s Hospital, Crown Street, Liverpool, L8 7S5. Tel: 0151 7959557. Fax: 0151 795 9599 email: l.goodfellow@liverpool.ac.uk
†Contributed equally

Running title (60 characters): Vaginal microbiota & recurrent early preterm birth

Word Counts:
Abstract: 249 words (max. 250)
Introduction: 371 words (max. 400)
Text: 3,491 words (max. 3,500)
Tables/figures: 3 tables/1 figure (max. 4)
Appendices: 2

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
**Abstract**

**Objective**: To assess the association between vaginal microbiome (VMB) composition and recurrent early spontaneous preterm birth (sPTB)/preterm prelabour rupture of membranes (PPROM).

**Design**: Nested case-control study.

**Setting**: UK tertiary referral hospital.

**Sample**: High-risk women with previous sPTB/PPROM <34\(^0\) weeks gestation who had a recurrence (n=22) or delivered at ≥37\(^0\) weeks without PPROM (n=87).

**Methods**: Vaginal swabs collected between 15-22 weeks gestation were analysed by 16S rRNA gene sequencing and 16S quantitative PCR.

**Main outcome measure**: Recurrent early sPTB/PPROM.

**Results**: 28/109 high-risk women had anaerobic vaginal dysbiosis, with the remainder dominated by lactobacilli (*L. iners* 36/109, *L. crispatus* 23/109, or other 22/109). VMB type, diversity, and stability were not associated with recurrence. Women with a recurrence, compared to those without, had a higher median vaginal bacterial load (8.64 vs. 7.89 log\(_{10}\) cells/μl, adjusted odds ratio (aOR)=1.90, 95% confidence interval (CI)=1.01-3.56, p=0.047) and estimated *Lactobacillus* concentration (8.59 vs. 7.48 log\(_{10}\) cells/μl, aOR=2.35, CI=1.20-4.61, p=0.013). A higher recurrence risk was associated with higher median bacterial loads for each VMB type after stratification, although statistical significance was reached only for *L. iners*-domination (aOR=3.44, CI=1.06-11.15, p=0.040). Women with anaerobic dysbiosis or *L. iners*-domination had a higher median vaginal bacterial load than women with a VMB dominated by *L. crispatus* or other lactobacilli (8.54, 7.96, 7.63, and 7.53 log\(_{10}\) cells/μl, respectively).

**Conclusions**: Vaginal bacterial load is associated with early sPTB/PPROM recurrence. Domination by lactobacilli other than *L. iners* may protect women from developing high bacterial loads. Future PTB studies should quantify vaginal bacteria and yeasts.

**Funding**: Wellbeing of Women, London, UK

**Keywords**: vaginal microbiome; *Lactobacillus*; spontaneous preterm birth; preterm premature rupture of membranes.

**Tweetable abstract**: Increased vaginal bacterial load in the second trimester may be associated with recurrent early spontaneous preterm birth.
Introduction

Previous spontaneous preterm birth (sPTB) or preterm prelabour rupture of membranes (PPROM) is the strongest risk factor for recurrent sPTB/PPROM.\(^1\) This may implicate a genetic contribution\(^2\) but does not rule out a contribution of persistent microbes and/or inflammation in the female reproductive tract.\(^3\)

16S ribosomal RNA (16S rRNA) sequencing was first used to assess the relationship between the bacterial vaginal microbiota (VMB) and PTB in 2014.\(^4\) Since then over 2000 pregnant women have been assessed in 12 studies (Appendix A, Table A1). The majority of these studies suggested a contribution of the VMB composition to PTB, but results have been inconsistent. The healthy vagina in pregnancy is dominated by *Lactobacillus* species but contains up to 80 different *Lactobacillus* and other bacterial taxa, and the optimal ways of summarising and categorising sequencing data to determine clinically relevant associations with PTB are yet to be determined.\(^5,6\) Studies to date have shown that increased relative abundances of *L. crispatus*, *L. gasseri*, and *L. jensenii* were either not associated with PTB or associated with a reduced PTB risk, and increased relative abundances of bacterial vaginosis (BV)-associated anaerobes were either not associated with PTB or associated with an increased PTB risk (Table A1). Results for *L. iners* relative abundance were inconsistent, with 6 studies showing no association, two studies an increased risk, and one study a decreased risk (Table A1). Importantly, all studies except two\(^7,8\) relied on compositional data, meaning that an increase in the proportion of one taxon always leads to a decrease in the proportions of other taxa.\(^9\) Furthermore, an increase in a proportion does not necessarily indicate an increase in absolute quantity. VMB researchers have only recently begun to semi-quantify 16S rRNA sequencing data by combining sequencing data with a quantitative PCR (qPCR) of
the 16S rRNA gene. The two studies to date that included some quantification reported opposite results, with one reporting a positive association between overall vaginal bacterial load and PTB risk, and the other a negative association.

Our aim was to assess associations between VMB compositions and recurrent early sPTB/PPROM (delivery at <34\textsuperscript{0} weeks compared to delivery at ≥37\textsuperscript{0} weeks) among high-risk pregnant women who had had a previous early sPTB/PPROM, using compositional and semi-quantitative 16S sequencing data.

**Methods**

**Study design and population**

Two cohorts of women at high and low-risk of PTB were enrolled at Liverpool Women’s Hospital, UK, from 1 June 2015 until 31 December 2017. The hospital’s patient and public engagement group helped guide the research and recruitment (Appendix A). High-risk was defined as a history of sPTB or PPROM at 16\textsuperscript{0}-33\textsuperscript{6} weeks gestation, and low-risk as having had at least one previous term birth (≥37\textsuperscript{0} weeks) without PPROM and never having had a PTB or PPROM. In the low-risk cohort, but not in the high-risk cohort, women were excluded if they had had significant cervical surgery or previous obstetric or medical problems (Appendix A). For this nested case-control study, we selected 109 women from the high-risk cohort (N=137) and 145 women from the low-risk cohort (N=223) based on the pregnancy outcome (sPTB or PPROM 16\textsuperscript{0}-33\textsuperscript{6} weeks or birth ≥37\textsuperscript{0} weeks in the high-risk cohort and birth ≥39\textsuperscript{0} weeks in the low-risk cohort) and availability of valid VMB data (Figure 1). Birth outcomes were classified independently by two obstetricians experienced in PTB.
management. When there was a discrepancy between them, the case was reviewed by a
third experienced obstetrician until the team reached consensus. Participants with medically
indicated PTBs, or for whom it was unclear whether a birth was truly spontaneous, were
excluded. We used relevant parts of the CROWN initiative core outcome set for PTB
research\textsuperscript{11} (Appendix A).

Clinical procedures

Participants were invited to two study visits at approximately 16 weeks (15\textsuperscript{+1}-18\textsuperscript{+6} weeks)
and 20 weeks (19\textsuperscript{+0}-23\textsuperscript{+0}) gestation. At each visit, a study obstetrician interviewed the
participant; collected high vaginal swabs during a speculum examination for VMB
assessments at the University of Liverpool Center for Genomic Research (HydraFlock
standard tapered swabs, Medical Wire and Equipment, Corsham, UK) and for quantitative
assessment of foetal fibronectin and culture by the local NHS laboratory; determined
cervical length; and did a 3D foetal ultrasound. All other procedures during pregnancy, and
all treatments, were in accordance with UK national guidelines.\textsuperscript{12} Preterm birth prevention
therapy (cerclage, vaginal pessary, or vaginal progesterone) was offered if cervical length
was \leq 25\text{mm} or based on clinician and patient preference in case of a large change in cervical
length between measurements. None of the low-risk women required an intervention but
71.6\% (78/109) of the high-risk women did (Table 1). Initiation of this treatment was always
after the VMB sampling visits.

VMB laboratory assessments

DNA was extracted from one swab per woman per visit using lysozyme lysis and bead-
beating procedures combined with the Qiagen DNeasy Blood and Tissue kit (Qiagen,
Manchester, UK) (Appendix A). The V3-V4 region of the 16S rRNA gene was amplified and sequenced on an Illumina HiSeq 2500 instrument (Illumina, San Diego, CA) run in rapid mode, 2x300bp using a 250PE and 50E kit. The panbacterial 16S rRNA gene copy concentration per sample was determined using the BactQuant quantitative PCR (qPCR) assay. Molecular data processing steps are described in Appendix A.

We only used the first set of valid sequencing results for each participant in the analyses because of the short time period between the two study visits, with one exception. In the subset of participants with valid sequencing results for both visits (88 high-risk and 129 low-risk women) we used data from both visits for the VMB stability analyses.

Creation and selection of VMB variables

Existing approaches

We systematically reviewed the published literature to identify VMB variables that were associated with PTB in at least one study (Appendix A, Table A1). These included: richness and Simpson diversity (1-D) as continuous variables; two categorical VMB composition variables with each participant assigned to one category (the community state types (CSTs) described by Ravel et al\(^{15}\) and \textit{Lactobacillus} groups (dominant, intermediate, and deplete) based on \textit{Lactobacillus} relative abundance);\(^{16}\) stability groups, slightly modified from Romero \textit{et al};\(^{17}\) and the presence/absence of specific taxa of interest (listed in Table A1).

New approaches

The final three sets of VMB variables are based on our previous work and were applied to PTB research for the first time.\(^{18-21}\) The first is a categorical VMB composition variable that
we named ‘VMB type’. These VMB types are similar to Ravel’s CSTs\textsuperscript{15} but with an increased emphasis on the relative abundance of pathobionts (Appendix A). We defined pathobionts as bacteria that are considered more pathogenic than BV-anaerobes and that often co-occur with lactobacilli instead of BV-anaerobes, such as streptococci. Secondly, each non-minority amplicon sequence variant (ASV; equivalent to a taxon) in each sample was allocated to one of four ‘bacterial groups’ based on the published literature (Appendix B): lactobacilli; BV-anaerobes; pathobionts; and a rest group of ‘other bacteria’.\textsuperscript{20} Within each sample, read counts of ASVs belonging to the same bacterial group were summed. This resulted in four continuous relative abundance variables (one for each bacterial group) per sample, which sum to 1.0 for each sample. Finally, we converted these bacterial group relative abundances into estimated concentrations (again four continuous variables, one for each bacterial group) making use of the BactQuant results (Appendix A). The BactQuant data were also used to determine estimated concentrations of taxa of interest (listed in Table A1) and the total bacterial load in each sample.

**Statistical analysis**

Our primary comparisons were between high-risk women who gave birth ≥37\textsuperscript{\textdegree}0 weeks gestation and high-risk women who had a recurrent early sPTB or PPROM <34\textsuperscript{\textdegree}0 weeks. Low-risk women who delivered at ≥39\textsuperscript{\textdegree}0 weeks gestation without PPROM were used as the normal reference group because we had never before characterised VMB compositions in pregnant women residing in Liverpool. Characteristics between these three groups were compared by student’s t-test for age, Mann-Whitney U test for other continuous variables, and Fisher’s exact test for binary and categorical variables. The various VMB variables described above were compared between the two high-risk groups using unadjusted and adjusted logistic regression, with adjustments for body mass index (BMI; as a quadratic term...
due to the bimodal association with PTB), history of cervical surgery (none, single LLETZ, or multiple LLETZ or knife cone biopsy), and smoking at the time of enrolment (yes/no).

Quartiles of the expected distributions of total vaginal bacterial load and total estimated Lactobacillus concentration were generated using data from the low-risk reference group, and high-risk participants were allocated to one of these quartiles. These newly created categorical variables were also compared between the two high-risk groups of interest by logistic regression. Finally, in an effort to differentiate between the effects of total vaginal bacterial load and the types of bacteria that make up this load, women were stratified by VMB type, and the logistic regression analyses were repeated for each stratum.

Results

We recruited 137 high-risk women and 223 low-risk women. Of 131 high-risk women with VMB data that passed quality control, 22 had a recurrent early sPTB/PPROM and 87 had term births ≥37⁰ weeks without PPROM (Figure 1); the remaining women were excluded because they delivered between 34⁰ and 37⁰ weeks or had a birth initiated by a caregiver. Of 214 low-risk participants with VMB data that passed quality control, 145 gave birth at ≥39⁰ weeks without PPROM and were used as the normal reference group.

The participant characteristics of the three groups were similar (Tables 1 and A2), except for those that are known risk factors for sPTB/PPROM: a higher proportion of high-risk women with a recurrence, compared to those who delivered at term, had two or more previous early sPTB/PPROM events (31.8% vs. 9.2%) and multiple previous LLETZ or knife cone biopsies (13.6% vs. 1.1%). The median gestational age when the first valid VMB sample was taken was slightly later in the low-risk group (16⁰⁵ weeks) than in the high-risk group (16³³
The VMBs of over 70% of pregnant women, high- and low-risk, were dominated by lactobacilli (defined as ≥75% relative abundance; Tables 2 and A3, heatmap in Figure A1). *L. iners* domination was most common overall (VMB type Li; 33.0% of high-risk and 22.1% of low-risk pregnant women), followed by *L. crispatus* domination (VMB type Lcr; 21.1% and 26.9%), and either domination by other lactobacilli (most commonly *L. gasseri* or *L. jensenii*) or ≥50% bifidobacteria (VMB type Lo+BL; 20.1% and 25.5%). In addition, 12.4% of high-risk and 12.8% of low-risk women had mild anaerobic dysbiosis (VMB type LA; defined as 25%-75% lactobacilli with the remainder BV-anaerobes), and 12.8% of high-risk and 13.1% of low-risk women had severe anaerobic dysbiosis (VMB type BV; ≥75% BV-anaerobes). Pathobionts and ‘other bacteria’ were rarely present, and if present, only in low relative abundance or estimated concentration (Tables 2 and A3).

The VMB of the majority of participants with samples available from both visits between 15th and 23rd weeks gestation was stable (Table 2). About two thirds of the high-risk (65.9%) and low-risk women (64.3%) were dominated by the same *Lactobacillus* species at both visits, and 14.8% of high-risk and 22.5% of low-risk women continued to have some degree of anaerobic dysbiosis. The remaining women switched *Lactobacillus* species (n=5) or shifted from lactobacilli-domination to anaerobic dysbiosis (n=12) or vice-versa (n=17).

The median vaginal bacterial load in the low-risk reference cohort was 7.68 log_{10} cells/μl (interquartile range (IQR) 6.80-8.35) and in the high-risk cohort 8.07 (IQR 7.28-8.92) log_{10}
High-risk women with mild or severe anaerobic dysbiosis (pooled together) had higher median vaginal bacterial loads than women with domination by *L. iners*, *L. crispatus*, or other lactobacilli/bifidobacteria (8.54, 7.79, 7.63, 7.53 log\(_{10}\) cells/μl, respectively; p-values are <0.05 for all comparisons between dysbiosis and the other groups, but are not significant for the comparisons between *L. iners*-domination and the other two lactobacilli groups; Table A4).

**Associations with early sPTB/PPROM recurrence – existing approaches**

In the high-risk women, we did not identify any associations between recurrence and VMB richness, Simpson (1-D) diversity, CSTs\(^\text{15}\) categorical *Lactobacillus* groups\(^\text{16}\) categorical stability groups\(^\text{17}\) (Table 2), and presence or relative abundance of common *Lactobacillus* species (Table A3). *Ureaplasma* species and BVAB TM7-H1 were present in higher proportions of high-risk women with a recurrence than in high-risk women with a term birth, and *Dialister* species vice versa, but their relative abundances when present were very low (Table A3). In contrast, *Bifidobacterium breve* was present in a lower proportion of women with a recurrence, but its relative abundance when present was also low. The proportions of low-risk women with these taxa present were similar to those in high-risk women without a recurrence (Table A3, Figure A2A).

**Associations with early sPTB/PPROM recurrence – new approaches**

We did not identify any associations between recurrence and VMB type or estimated BV-anaerobes or pathobionts concentrations (Table 2). However, high-risk women with a recurrence had a higher overall vaginal bacterial load (8.64 vs. 7.89 log\(_{10}\) cells/μl, adjusted odds ratio (aOR) 1.90, 95% confidence interval (CI) 1.01-3.56, p=0.047) and a higher total...
estimated Lactobacillus concentration (8.59 vs. 7.48 log_{10} cells/μl, aOR 2.35, 95% CI 1.20-4.61, p=0.013) than high-risk women without a recurrence (Table 2). The aOR was 1.63 (95% CI 0.93-2.86, p=0.093) for each increase in total vaginal bacterial load quartile, and 2.61 (95% CI 1.27-5.33, p=<0.001) for each estimated Lactobacillus concentration quartile (Table 2).

Table 3 shows estimated concentrations of total vaginal bacteria, the four bacterial groups, L. crispatus, L. iners, and other lactobacilli in high-risk women with and without a recurrence after stratification by VMB type. A trend towards higher recurrence risk with higher median bacterial loads was seen for each VMB type after stratification but was only significant in women with L. iners-domination: the median vaginal bacterial loads were 9.03 (IQR 8.02-9.12) log_{10} cells/μl in women with a recurrence compared to 7.79 (IQR 6.93-8.65) log_{10} cells/μl in women with a term birth (aOR 3.44, 95% CI 1.06-11.15, p=0.011). The association was no longer significant after adjustment for confounders (aOR 2.38, 95% CI 0.71-7.95, p=0.160). Within each of the other VMB type strata, the median estimated concentrations in high-risk women who had a term birth were similar to those in low-risk women, and non-significantly lower than those in high-risk women who had a recurrence.

The majority of high-risk participants with an early sPTB/PPROM recurrence (17/22; 77.3%), and about half of the high-risk participants without a recurrence (44/87; 50.6%), had at least one VMB characteristic that was positively associated with recurrence. The following positively associated VMB characteristics overlapped considerably in both high-risk groups (Venn diagrams in Figure A2): highest quartile of total bacterial load, highest quartile of estimated total Lactobacillus concentration, highest quartile of estimated L. iners concentration, and presence of BVAB TM7-H1. The only positively associated VMB characteristic
characteristic that did not completely overlap with the highest quartile of bacterial and estimated total \textit{Lactobacillus} concentration was the presence of \textit{Ureaplasma} species (this was the case in high-risk pregnancies with and without a recurrence).

**Discussion**

**Main Findings**

This exploratory study suggests that high vaginal bacterial load in the second trimester is associated with early sPTB/PPROM recurrence in high-risk pregnant women, irrespective of the bacterial composition. Our statistical power was limited: only 22 women had an early sPTB/PPROM recurrence. Therefore, not all associations in line with this conclusion reached statistical significance, especially after reducing statistical power further by stratification.

However, we believe that this new hypothesis warrants further research for various reasons. First, the recurrence risk increased for each increase in bacterial and \textit{Lactobacillus} concentration quartile (on a log$_{10}$ scale). Second, the findings were consistent for each VMB type after stratification: higher recurrence in high-risk women with higher median bacterial loads for all VMB types (reaching significance for \textit{L. iners}-domination), and similar median bacterial loads in high-risk and low-risk women who had a term birth for all VMB types.

Third, estimated concentrations reached higher levels for anaerobic dysbiosis and \textit{L. iners}-domination than for domination by other lactobacilli or bifidobacteria. Our findings therefore do not contradict earlier findings by other groups that indicated increased PTB risk for anaerobic dysbiosis and high \textit{L. iners} relative abundance.$^{8,22-24}$ In fact, the inconsistent findings related to \textit{L. iners} in the literature might be due to the lack of quantification in most studies.
Interpretation in light of other evidence

To our knowledge, the potential association between the vaginal bacterial load and PTB has only been assessed in two previous studies. Freitas et al\(^7\) also showed a higher 16S gene concentration in the second trimester (approximately 16 weeks gestation) in women who had an sPTB <37 weeks compared to women with a term birth ≥37 weeks (8.1 vs. 7.8 log\(_{10}\) copies/swab, respectively). Conversely, Elovitz et al\(^8\) reported a similar 16S gene concentration in women who had an sPTB <37 weeks compared to women with a term birth ≥37 weeks in samples collected before 24 weeks gestation (approximately 8 log\(_{10}\) copies/swab in each group), and a lower concentration (approximately 7.5 vs 7.9 log\(_{10}\) copies/swab) in samples collected at 28 weeks gestation. High vaginal bacterial load may only be associated with PTB when present prior to 24 weeks gestation, but additional quantitative studies are needed.

Female genital tract inflammation is considered one potential cause of PTB.\(^{25}\) BV (a clinical condition caused by anaerobic dysbiosis) is associated with vaginal inflammation, and is a very common condition in non-pregnant women.\(^{26}\) Pregnant women are protected from BV due to the high oestrogen levels during pregnancy.\(^{27,28}\) In our study, the prevalence of anaerobic dysbiosis, and the relative abundances and estimated concentrations of BV-associated bacteria and pathobionts, were indeed low. This likely explains why we did not identify any statistically significant associations between the presence or extent of anaerobic dysbiosis and early sPTB/PPROM recurrence. However, the non-significant trends consistently suggested increased risk. Another common inflammatory vaginal condition is vulvovaginal candidiasis, for which pregnant women are at increased risk.\(^{29}\) In our study, only one woman (low-risk) reported having been treated for vulvovaginal candidiasis prior to
study enrolment. Future studies should systematically assess vaginal yeasts concentrations in addition to bacterial concentrations in all participants, and determine their potential contributions to early sPTB/PPROM recurrence.

In our study, *L. iners* was able to achieve higher bacterial loads than the other lactobacilli, and was associated with early sPTB/PPROM recurrence. *L. iners* has a genome of just 1.3 Mbp,\(^{30}\) strikingly smaller than the other common vaginal *Lactobacillus* species, and seems fully adapted to the vaginal niche.\(^{31}\) This may explain why it is able to grow to such high densities in the vagina. *Lactobacillus* species are considered to be non-inflammatory when present in the vagina,\(^{32}\) but they might cause inflammation when they travel through the cervical mucus plug into the upper genital tract. Ascension to the upper genital tract may be increased when vaginal density is high. Studies have shown that *L. iners* with high expression of clustered regularly inter-spaced short palindromic repeat (CRISPR) genes are present in women with BV but not in women without BV.\(^{33}\) CRISPR genes are the primary bacterial defence against bacteriophages. The *L. iners* strains associated with high bacterial loads and recurrent sPTB/PPROM in our study may differ from strains associated with lower bacterial loads. This requires further research.

Two uncommon taxa (present in fewer than 20% of the women in the low risk group) were non-significantly associated with early sPTB/PPROM recurrence in our study and were also associated with PTB in other studies: *Ureaplasma*\(^{7,24}\) and BVAB TM7-H1\(^{6}\). In women with early sPTB/PPROM recurrence, the presence of *Ureaplasma* species in the vagina did not completely overlap with having high bacterial loads and lactobacilli concentrations. *Ureaplasma* species may therefore play a role in PTB that is independent from overall
vaginal bacterial composition and load. Our data confirms that *Bifidobacterium breve* may have a protective effect against early sPTB/PPROM recurrence.\(^{16}\)

**Strengths and Limitations**

Our study had some limitations, most notably limited statistical power and insufficient data on vaginal yeasts, but also had strengths. Pregnancy outcomes were very carefully selected and assessed, many VMB composition variables were created covering different aspects of VMB composition, and most of these VMB variables were (semi)quantified. The latter turned out to be crucial, as our most important findings concern bacterial concentrations. The quantification method that we used has recently been validated for non-minority taxa in two different studies.\(^{7,8}\)

We had initially planned to analyse early sPTB and early PPROM as two separate outcomes because the aetiological pathways may differ,\(^ {34}\) and to exclude women who used PTB prevention treatments,\(^ {35}\) but the limited statistical power did not permit this. A final limitation is the lack of VMB data in the third trimester and/or closer to the birth.

**Conclusion**

Vaginal bacterial load in the second trimester, irrespective of the bacterial composition, was associated with early sPTB/PPROM recurrence. Domination by lactobacilli other than *L. iners* may protect women from developing high bacterial loads. These findings should be confirmed in larger, longitudinal studies that incorporate quantification of vaginal bacteria and yeasts. If they are confirmed, interventions that maintain a non-*iners* lactobacilli/bifidobacteria-dominated VMB may protect women from inflammation-associated PTB.
Acknowledgements

We would like to thank all participants for their enthusiastic involvement in the study, in particular members of the Harris-Wellbeing Patient and Public Engagement group. We would also like to thank the Liverpool Women’s Hospital for hosting the research, Tracy Ricketts at the Liverpool Women’s Hospital for administrative support, University of Liverpool Centre for Genomic Research staff for 16S sequencing, and Jacques Ravel and Mike Humphrys at the University of Maryland School of Medicine, Institute for Genome Sciences, for the 16S qPCR assays.

Disclosure of interests

All authors have completed the ICMJE uniform disclosure forms and declare: LG, AC, AS, AA, BM-M, AA, ZA and JW received a grant from Wellbeing of Women charity to establish the Harris Wellbeing Research Centre that funded the submitted work. MW, aCG, CB, DR, JI and BP received no support from any organisation for the submitted work. All authors declare no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Contribution to authorship

JW, AS, DR, BM-M, AC, AA and ZA conceived the study, wrote the protocol and obtained funding. AC, JI, BP and LG contributed to the protocol and recruited participants. AD, JW, aCG, and CB developed the DNA extraction, sequencing, and bioinformatics methods. LG, AC and AS extracted the clinical data. LG performed the laboratory analysis. JW developed the VMB analysis methods, and LG and MV performed the data processing. LG performed the data analysis and wrote the initial draft, and JW, ZA and MV contributed to data interpretation and revised the paper. All authors approved the final manuscript.

Details of Ethics Approval

The study was approved by North West Research Ethics Committee - Liverpool Central, reference 11/NW/0720 on 4 November 2011. The collection of vaginal fluid for this component of the study was approved in amendment 4 on 19 October 2015.

Funding

The prospective cohort study was funded Wellbeing of Women as part of a charitable donation from Lord and Lady Harris to establish the Harris-Wellbeing PTB Research Centre, University of Liverpool. This covered administrative costs, laboratory analysis, salary for AC, study support costs for AC and LG. No additional funding was used. Wellbeing of Women were not involved in the conduct of the research or writing the paper.
References


11 January 2021


**Tables, Figures, and Supplementary Materials Caption List**

- Table 1: Participant characteristics by pregnancy risk group and outcome
- Table 2: Associations between VMB variables and recurrent early sPTB/PPROM in high-risk women
- Table 3: Associations between estimated median concentrations of vaginal bacteria and early recurrent sPTB/PPROM stratified by VMB type
- Figure 1: Participant flow
- Appendix A: Supplementary methods, figures, and tables
- Appendix B: Assignment of amplicon sequence variants to bacterial groups
## Table 1: Participant characteristics by pregnancy risk group and outcome

<table>
<thead>
<tr>
<th></th>
<th>Low-risk pregnancy (LR)</th>
<th>High-risk pregnancy (HR)</th>
<th>P* value LR vs HR</th>
<th>P* value HR term vs HR preterm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Term birth n=145</td>
<td>Term birth n=87</td>
<td>Early sPTB or PPROM n=22</td>
<td></td>
</tr>
<tr>
<td><strong>Age (years) mean (SD)</strong></td>
<td>31.1 (4.5)</td>
<td>30.0 (4.5)</td>
<td>31.4 (5.4)</td>
<td>0.145</td>
</tr>
<tr>
<td><strong>BMI (kg/m²) median (IQR)</strong></td>
<td>24.5 (22-29)</td>
<td>25 (22-28)</td>
<td>27.5 (22-35)</td>
<td>0.619</td>
</tr>
<tr>
<td><strong>Current smoker number (%)</strong></td>
<td>15 (10.6)</td>
<td>15 (14.6)</td>
<td>5 (23.8)</td>
<td>0.096</td>
</tr>
<tr>
<td><strong>Ethnicity (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>139 (95.9)</td>
<td>81 (93.1)</td>
<td>22 (100.0)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>3 (2.1)</td>
<td>4 (4.6)</td>
<td>12 (50.0)</td>
<td>0.61</td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>22 (100.0)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>2 (1.4)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Not recorded</td>
<td>1 (0.1)</td>
<td>2 (2.3)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of previous PPROM or sPTB 16-0-33+6 weeks (%)</strong></td>
<td>79 (90.8)</td>
<td>79 (90.8)</td>
<td>17 (73.3)</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Previous cervical surgery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nil significant</td>
<td>131 (90.3)</td>
<td>79 (90.8)</td>
<td>17 (73.3)</td>
<td></td>
</tr>
<tr>
<td>Single LLETZ</td>
<td>14 (9.7)</td>
<td>7 (8.0)</td>
<td>2 (9.1)</td>
<td></td>
</tr>
<tr>
<td>Multiple LLETZ or knife cone</td>
<td>1 (1.1)</td>
<td>3 (13.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gestational age at first valid VMB sample (weeks)</strong></td>
<td>16+5</td>
<td>16+2</td>
<td>16+3</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Cervical length (mm) median and IQR</strong></td>
<td>36 (26-43)</td>
<td>36 (26-43)</td>
<td>0.000</td>
<td>0.58</td>
</tr>
<tr>
<td>16 weeks</td>
<td>41 (36-47.1)</td>
<td>35.5 (31-42)</td>
<td>36 (26.5-43.5)</td>
<td></td>
</tr>
<tr>
<td>20 weeks</td>
<td>41 (37-45)</td>
<td>30 (30-40)</td>
<td>30 (26-37)</td>
<td></td>
</tr>
<tr>
<td><strong>pNN (ng/ml) median and IQR</strong></td>
<td>7 (5-16)</td>
<td>7 (5-22)</td>
<td>9 (5-23.5)</td>
<td>0.766</td>
</tr>
<tr>
<td>16 weeks</td>
<td>7 (5-16)</td>
<td>9 (6-12)</td>
<td>7 (6-12)</td>
<td></td>
</tr>
<tr>
<td>20 weeks</td>
<td>7 (5-12)</td>
<td>7 (5-14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Preterm birth prevention treatment used after study visit</strong></td>
<td>139 (100.0)</td>
<td>61 (70.7)</td>
<td>17 (77.3)</td>
<td>na</td>
</tr>
<tr>
<td>None</td>
<td>139 (100.0)</td>
<td>61 (70.7)</td>
<td>17 (77.3)</td>
<td></td>
</tr>
<tr>
<td>Cervical Cerclage</td>
<td>4 (4.6)</td>
<td>2 (3.3)</td>
<td>7 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Progesterone</td>
<td>2 (2.3)</td>
<td>1 (4.5)</td>
<td>3 (13.6)</td>
<td></td>
</tr>
<tr>
<td>Arabin Pessary</td>
<td>20 (33.0)</td>
<td>2 (3.3)</td>
<td>6 (3.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Antimicrobial in pregnancy prior to sampling (%)</strong></td>
<td>115 (79.3)</td>
<td>66 (74.7)</td>
<td>17 (77.3)</td>
<td>0.038</td>
</tr>
<tr>
<td>None</td>
<td>115 (79.3)</td>
<td>66 (74.7)</td>
<td>17 (77.3)</td>
<td></td>
</tr>
<tr>
<td>Metronidazole or clindamycin</td>
<td>0 (0.0)</td>
<td>4 (4.6)</td>
<td>1 (4.5)</td>
<td></td>
</tr>
<tr>
<td>Other antibiotic</td>
<td>21 (14.5)</td>
<td>9 (10.3)</td>
<td>3 (13.6)</td>
<td></td>
</tr>
<tr>
<td>Clotrimazole</td>
<td>1 (0.7)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Not recorded/ unsure</td>
<td>8 (5.5)</td>
<td>9 (10.3)</td>
<td>1 (4.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Gestational age at PPROM (weeks)</strong></td>
<td>31+1</td>
<td>18+0-33+6</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td><strong>Gestational age at birth (weeks)</strong></td>
<td>40+1</td>
<td>38+6</td>
<td>31+5</td>
<td>18+0-35+5</td>
</tr>
<tr>
<td>Birthweight (g) median and range</td>
<td>3594 (439)</td>
<td>3234 (489)</td>
<td>1778 (673)</td>
<td>na</td>
</tr>
</tbody>
</table>
Abbreviations: BMI, body mass index; HR, high-risk pregnancy; LR, low-risk pregnancy; na, not applicable; PPROM, preterm prelabour rupture of membranes; qfFN, quantitative foetal fibronectin level in cervicovaginal fluid; sPTB, spontaneous preterm birth.

* P values by Student’s t-test for age, Mann-Whitney U test for BMI, cervical length, qfFN and gestational age at sampling; and Fisher’s exact test for remainder of variables.

¶ Missing values: BMI 1 each for LR and HR term; smoking 3 for LR and 1 each for HR term and early preterm; birthweight 2 each for HR term and HR preterm. Cervical length at 16 weeks: HR term n=80, HR preterm n=20, LR n=136. Cervical length at 20 weeks: HR term n=79, HR preterm n=14, LR n=135. qfFN at 16 weeks: HR term n=77, HR preterm n=20, LR n=142. qfFN at 20 weeks: HR term n=75, HR preterm n=19, LR n=136.

parity was defined as number of previous pregnancies with live births or stillbirths ≥24+0 weeks gestation. Eligibility for the HR cohort included ≥1 sPTB or PPROM at 16+0 to 33+6 weeks gestation. Therefore, all HR women with parity=0 had at least one previous late miscarriage. Antimicrobial usage was only assessed in the pregnancy prior to sampling. Clindamycin is available without medical review as an ‘over the counter’ medication in the UK, and so usage after the study visit was not possible to assess within our study design.
Table 2: Associations between VMB variables and recurrent early sPTB/PPROM in high-risk women

<table>
<thead>
<tr>
<th>VMB type</th>
<th>Low-risk pregnancy (LR)</th>
<th>High-risk pregnancy (HR)</th>
<th>p-value* difference between HR vs LR</th>
<th>Adjusted odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=145</td>
<td>n=109</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Birth ≥ 39 weeks</td>
<td>Birth ≤ 37 weeks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median (IQR)</td>
<td>Median (IQR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 (6-25)</td>
<td>7 (6-25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.31 (0.04-0.52)</td>
<td>0.31 (0.04-0.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adjusted¶ logistic regression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Birth ≥ 37 vs. &lt;34 weeks</td>
<td>Median (IQR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 (6-25)</td>
<td>8 (6-25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.31 (0.04-0.52)</td>
<td>0.29 (0.04-0.31)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations**: HR, high-risk pregnancy; LR, low-risk pregnancy; PPROM, preterm prelabour rupture of membranes; sPTB, spontaneous preterm birth; VMB, vaginal microbiota.

*P values by Mann-Whitney U test for continuous variables and Fisher’s exact test for remainder. Logistic regression adjusted for body mass index as quadratic term, history of cervical surgery (none/single LLETZ/multiple LLETZ and/or knife cone biopsy) and smoking (yes/no at time of visit). Adjusted analyses excluded two HR women with missing smoking data, one with missing BMI data, and variable numbers because of collinearity within the regression. VMB type definitions: Li = L. iners-dominated (≥ 75% lactobacilli with L. iners the most common); Ltr = L. crispatus-dominated (≥ 75% lactobacilli with L. crispatus the most common); Lo +B = other lactobacilli or Bifidobacterium dominated (≥ 25% lactobacilli); Lo +B +L = either ≥ 75% lactobacilli with L. iners/L. crispatus the most common, or ≥ 75% Bifidobacterium; lactobacilli and anaerobes (≥ 25% lactobacilli) and BV mixture of BV-anaerobes (≥ 25% lactobacilli). VMB changes between 16 and 20 weeks were only assessed in the subset of participants with valid VMB assessments at both time points LR n=129; HR n=71 and H R preterm n=17.

11 January 2021
Table 3: Associations between estimated median concentrations of vaginal bacteria and early recurrent sPTB/PPROM stratified by VMB type

<table>
<thead>
<tr>
<th>VMB type</th>
<th>Low risk pregnancy (LR)</th>
<th>High risk pregnancy (HR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Birth ≥39 weeks n=145</td>
<td>Birth ≥37 weeks n=7</td>
</tr>
<tr>
<td></td>
<td>Median (ICR)</td>
<td>Median (ICR)</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>HR</td>
</tr>
<tr>
<td></td>
<td>Adjusted* logistic regression model of HR</td>
<td>Unadjusted logistic regression model of HR</td>
</tr>
<tr>
<td></td>
<td>Birth ≥37 vs &lt;34 weeks (early preterm) n=109</td>
<td>Birth ≥37 vs ≤34 weeks (≥37 weeks) n=22</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial load</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total lactobacillus</td>
<td>7.73 (6.42-8.46)</td>
<td>6.73 (7.14-8.23)</td>
</tr>
<tr>
<td>Total BV-associated</td>
<td>4.74 (0.62-6.9)</td>
<td>4.62 (1.64-7.07)</td>
</tr>
<tr>
<td>Total pathogens</td>
<td>0.00 (0.0)</td>
<td>0.00 (0.0)</td>
</tr>
<tr>
<td>Total other bacteria</td>
<td>0.00 (0.0)</td>
<td>0.00 (0.0)</td>
</tr>
<tr>
<td>Type of Lactobacillus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. crispatus</td>
<td>7.22 (6.14-8.45)</td>
<td>6.73 (7.14-8.23)</td>
</tr>
<tr>
<td>L. iners</td>
<td>0.00 (0.0)</td>
<td>0.00 (0.0)</td>
</tr>
<tr>
<td>Other lactobacilli</td>
<td>5.09 (3.9-6.94)</td>
<td>5.34 (3.02-7.16)</td>
</tr>
</tbody>
</table>

11 January 2021
Abbreviations: 95% CI, 95% confidence interval; HR, high-risk pregnancy; IQR, interquartile range; LR, low-risk pregnancy; OR, odds ratio; PPROM, preterm prelabour rupture of membranes; sPTB, spontaneous preterm birth; VMB, vaginal microbiota. *Logistic regression adjusted for body mass index as quadratic term, history of cervical surgery (none/single LLETZ/multiple LLETZ and/or knife cone biopsy) and smoking (yes/no at time of visit). Adjusted analyses excluded two HR women with missing smoking data, one with missing BMI data, and variable numbers because of collinearity within the regression. Participants were included in these analyses even if a taxa was not present in that sample (with concentration=0). ¶VMB type definitions: LinL. iners-dominated (≥ 75% lactobacilli with L. iners the most common); LcrL. crispatus-dominated (≥75% lactobacilli with L. crispatus the most common); Lo+ BLother lactobacilli- or Bifidobacterium dominated (Lo+ BL: either ≥75% lactobacilli with L. jensenii or L. gasseri the most common, or ≥50% Bifidobacterium); lactobacilli and anaerobes (LA: 25%-75% lactobacilli); and BV=mixture of BV-anaerobes (<25% lactobacilli).
High-risk pregnancy
(Previous sPTB or PPROM <34 weeks)

N=137

Numbers | Exclusions | Numbers
--- | --- | ---
NA | Early term delivery 37\(^{+0}\)-38\(^{+6}\) (low risk only) | 65
14 | sPTB/PPROM 34\(^{+0}\)-37\(^{+0}\) | 2
NA | sPTB/PPROM <34\(^{+0}\) (low risk only) | 2
8 | Caregiver initiated preterm birth | 0
6 | VMB data failed quality control | 9

N=109

High-Risk Early Preterm
n=22
sPTB or PPROM <34\(^{+0}\) weeks

High-Risk Term
n=87
Birth ≥37\(^{+0}\) weeks

Low-risk pregnancy
(Parous women with all previous births at term)

N=223

N=145

Low-Risk Early Preterm
n=22
sPTB or PPROM <34\(^{+0}\) weeks

Low-Risk Term
n=145
Birth ≥39\(^{+0}\) weeks