The association between the vaginal microbiota and recurrent early spontaneous preterm birth

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Medicine

by

Laura Rachel Goodfellow

June 2021

Thesis Abstract

The association between the vaginal microbiota and recurrent early spontaneous preterm birth

Laura Goodfellow

Preterm birth is the leading cause of death of children under 5 worldwide. Despite over 30 years investment in basic science and clinical research the rate has remained relatively static. This has been attributed to the multifactorial nature of preterm birth.

Infection and/or inflammation are estimated to contributing to approximately a third of preterm births, however the exact organism(s) frequently remain elusive and are often attributed to bacteria that are not cultured using standard laboratory techniques. Vaginal bacteria are a presumed source of inflammatory organisms due to the close anatomical relationship with the uterus and cervix. The advent of culture-independent methods of assessment of the microbiota has enabled new interrogation of this theory. In the past seven years over 2000 women have undergone assessment of the vaginal microbiota (VMB) in pregnancy. Most studies have identified components of the VMB that relate to preterm birth. However, poor understanding of the distribution of VMB characteristics in healthy pregnancies, differences in almost every aspect of study design, and reliance on proportional data rather than quantification have hampered efforts to translate these findings into clinically meaningful information.

This project advances the field of VMB in preterm birth work in two ways. Firstly we summarised VMB characteristics that have previously been associated with preterm birth and assessed whether these are associated with recurrence of preterm birth under 34 weeks gestation, contextualised by reference to the expected distribution of the VMB characteristics in our locality.

Secondly, a criticism of the initial next-generation sequencing studies was over-reliance on data solely describing the proportions of bacteria present in a given sample, without consideration of the overall bacterial load. We were able to address this by complementing our analysis with semi-quantitative analysis using quantitative PCR of the 16S rRNA gene.

We found that four VMB characteristics (out of 29 assessed) showed trends towards reproduction of previous work. These associations did not meet traditional statistical significance (<0.05%), most likely because we could only evaluate 22 preterm births.

The most striking finding of our analysis was that vaginal bacterial load in the second trimester, irrespective of the bacterial composition, was associated with early spontaneous preterm birth or preterm prelabour rupture of membranes recurrence. Women with a recurrence, compared to those without, had a fivefold higher median vaginal bacterial load. There was a gradient effect of bacterial load on risk of recurrence of preterm birth. Interestingly, domination by lactobacilli other than *L. iners* protected women from developing high bacterial loads. This could explain the inconsistent association between *L. iners* and preterm birth in previous studies; if *L. iners* dominance is only associated with PTB when there is a concurrent high bacterial load, then women with *L. iners* domination and a normal bacterial load would not show this association. We also found that the women who have high bacterial loads and a preterm birth have a higher rate of early onset chorioamnionitis, strengthening the evidence that the bacterial load is clinically important.

This study provides a strong foundation on which to base further studies incorporating assessment of bacterial load alongside traditional 16S rRNA analysis to fully elucidate the contribution of the VMB to preterm birth. Once fully understood the VMB is potentially modifiable and could provide a novel mode of preterm birth prevention.

Acknowledgements

This project was made possible by a generous donation from Lord and Lady Harris through the charity Wellbeing of Women, to found the Harris Wellbeing Research Centre at the University of Liverpool. Thank you for this fantastic opportunity.

I would like to thank my supervisors, Professor Zarko Alfirevic and Professor Janneke van de Wijgert. Zarko is a truly inspirational human being; his ability to cut to the essence of a problem in a fraction of a second, ask the pertinent questions, and make the impossible happen amaze me. I feel truly honoured to have worked with him. Janneke has been the voice of reason and steadfast direction that this project has required. Throughout this year marred by the COVID pandemic, with immense pressure as part of the scientific community working to overcome this disease, she has still made time for me. I admire her direct approach and have really enjoyed her teaching about epidemiological methods.

I have met lovely people along this journey whom I would like to thank. Firstly my preterm buddy, Dr Angharad Care, who put her heart and soul into this project and I have learnt so much from, mostly by attempting to emulate. Next, the ever thorough Dr Marijn Verwijs, who fastidiously assisted me with the in silico analysis, I really enjoyed our lab chats too. I next need to thank the people who have walked this path before me, in particular Dr Christina Gill and Dr Christina Brownowski who were fantastic sources of support during the early phase of this study whilst I was grappling with the laboratory analysis. I also wish to extend my thanks to my colleagues and friends Dr Andrew Sharp, Dr Borna Poljak and Dr Jelena Ivandic, for their great help with participant recruitment, and pep talks.

I was overcome by the readiness to participate in this project by the families of Liverpool. This was both from families who had experienced the heartache of preterm birth and were bravely embarking on another pregnancy, and those who had had previous term births and were juggling the demands of parenting, pregnancy and expectations outside the home. Families from both groups went out of their way to altruistically participate and I hope that the outcomes from this work do justice to your participation. In particular Mrs Joanne Johnson and Dr Adrian O'Hara for their enthusiastic input into the patient and public engagement work and peer to peer support.

I would like to thank my family. My mum has no idea what this thesis is about (and likewise I have not scratched the surface of her work) but has afforded me the respect to get on with it unimpeded. To my dad who made the mistake of asking about my work over Christmas dinner, and has never quite overcome the discomfort of realising his daughter has dedicated years of her life to assessing 'bugs in the vagina', sorry. But to both of you, thank you for all your support to get me to this stage.

My lovely husband has been a constant support throughout this process, despite coming from a baseline similar to my parents. He has carefully considered all my dilemmas, encouraged, re-focused and hugged as necessary. As well as providing really fantastic IT support and being a great lockdown buddy. Finally, the two most perfect people in the world, my boys, Hugh and Arlo, you are both fantastic. When Hugh started discussing baby loss during his home schooling question about the saddest thing in the world, I simultaneously realised how much of this you are taking in whilst I am oblivious, and how quickly you are maturing. Arlo's loving nature and fun outlook mean that we never have a dull day. Looking forward to our next chapter together.

Abbreviations

ASV	Amplicon Sequence Variant
aOR	adjusted Odds Ratio
BAME	Black, Asian and Minority Ethnic
BAPM	British Association of Perinatal Medicine
BMI	Body Mass Index
BV	Bacterial Vaginosis
BVAB	BV-associated Bacterium type (followed by number)
CI	Confidence Interval
CL	Cervical Length
CRP	C-Reactive Protein
CST	Community State Type
Ct	Cycle threshold
cpn60	60 kDa chaperonin protein
DNA	Deoxyribonucleic acid
FDR	False Discovery Rate
GA	Gestational Age
GROW	Gestation Related Optimal Weight
HIV	Human Immunodeficiency Virus
HR	High-Risk pregnancy (for preterm birth)
HVS	High Vaginal Swab
IQR	Inter-Quartile Range
IMD	Index of Multiple Deprivation
LD	Lactobacillus Dominance
LEfSe	Linear discriminant analysis effect size
LLETZ	Large Loop Excision of the Transformation Zone

LR	Low-Risk pregnancy (for preterm birth)
Medline	Medical Literature Analysis and Retrieval System Online
NA	Not Applicable
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NIH	National Institutes of Health
NS	Not Significant
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PPROM	Preterm Prelabour Rupture of Membranes
PREBIC	Preterm Birth International Collaborative
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTB	Preterm Birth
qfFN	quantitative Fetal Fibronectin
qPCR	quantitative Polymerase Chain Reaction
RA	Relative Abundance
RCT	Randomised Control Trial
RDP	Ribosomal Database Project
RNA	Ribonucleic acid
ROM	Rupture of Membranes
RR	Risk Ratio
rrn	ribosomal RNA operons
SEN	Special Educational Need
PCR	Polymerase Chain Reaction
sPTB	Spontaneous Preterm Birth
UK	United Kingdom

USA	United States of America
VMB	Vaginal Microbiota
WHO	World Health Organisation
3D	3 Dimensional
16S rRNA	Small subunit ribosomal RNA

1 Introduction

1.1 Definitions

1.1.1 Microbiota and microbiome

A microbiota describes the microorganisms that colonize a biological niche, in this case the vagina. The term microorganism refers to taxa that can be viewed down a microscope as discrete organisms and includes bacteria, fungi, archaea and protists. Bacteria form the most abundant part of the human microbiome and are the focus of most work assessing the vaginal microbiota in pregnancy.

The term 'microbiome' refers to an 'entire habitat' including the microorganisms, their genomes, and the surrounding environmental conditions, whereas the 'microbiota' consists of only 'microorganisms present in a defined environment'.[1] Therefore the vaginal microbiota shall be the focus of this thesis.

1.1.2 Spontaneous early preterm birth

The World Health Organisation (WHO) defines preterm birth as birth less than 37 completed weeks of gestation.[2] This is subclassified into extremely preterm (<28^{+0} weeks), early preterm (28^{+0} to 33^{+6} weeks), and late preterm (34^{+0} to 36^{+6} weeks).

This thesis will focus on births between 16^{+0} and 33^{+6} weeks (inclusive). This will include births of spontaneous onset and those preceded by preterm prelabour rupture of membranes (PPROM), but not medically indicated preterm births.

1.1.3 Preterm prelabour rupture of membranes

PPROM refers to the rupture of the amniotic membranes prior to the onset of labour and prior to 37^{+0} weeks gestation. In some cases the rupture of membranes is the first symptom a woman will notice of the labour process; in other cases it occurs remotely from the labour

process. PPROM can also precipitate spontaneous preterm labour. In order to achieve a differentiation of membranes rupturing early in the labour process (which was actually preterm labour), and isolated PPROM we used a definition of rupture of membranes at least 12 hours prior to the onset of uterine activity. This was in keeping with previous research from our group.[3] A small number of cases of PPROM are preceded by instrumentation of the uterus, mostly for the purposes of amniocentesis or chorionic villous sampling. Only cases of spontaneous PPROM were eligible for this study.

1.2 Impact of preterm birth

Worldwide, approximately 10.6% of births are preterm,[4] with a rate ranging from 5% to 18% by country.[5] Preterm birth complications are the leading cause of death in children under 5 years old globally, and were responsible for approximately 1 million deaths of children in 2015.[6] Despite improvements in neonatal care the mortality attributed to preterm birth between 2000 and 2015 remained static, due to the increases in the incidence of preterm birth.[5,7]

Preterm birth is recognised as an important public health problem and national and international bodies have committed to reducing this burden. The WHO set a goal of a 50% reduction in mortality related to preterm birth in resource poor countries from 2010-2025 [8] and this is seen as crucial to the Sustainable Development Goal of ending preventable newborn and child deaths by 2030.[9,10] In the UK the Department of Health has committed to reducing the preterm birth rate from 8% to 6% by 2025.[11]

The morbidity and mortality associated with preterm birth are inversely proportional to gestational age at birth,[12] as shown in Figure 1.1. The EPICure 2 study found that babies born at 24 weeks gestation in 2006 in England had only a 40% chance of survival to hospital discharge, and these infants had a 48% chance of major morbidity.[13] At the other end of the

scale later preterm births have very good neonatal survival rates (Figure 1.1[14]). However, infants born in the late preterm period do have a higher risk of long term health sequelae compared to babies born at term: the risk of cerebral palsy for infants born at 32-36 weeks is 0.7%, compared to 0.1% in term infants.[15] Later preterm birth is more common than earlier preterm birth, therefore more children with cerebral palsy are survivors of later preterm births than early preterm births.[12] Ultimately a reduction in all preterm births is needed at both ends of this scale to make a significant impact on the related morbidity and mortality.

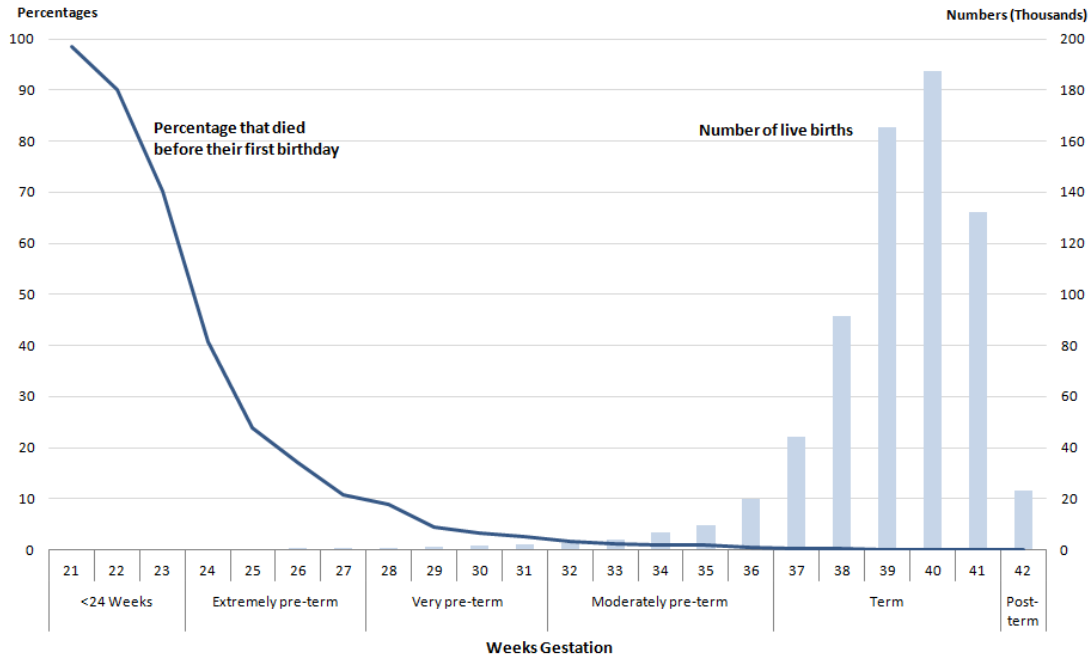


Figure 1.1:Percentage of infant deaths and number of live births by week gestation, 2013, in England and Wales. From Office of National Statistics, 2017[14]

The reasons for the rising incidence of preterm births are more complex than it initially appears for several reasons. Firstly, the rate of preterm birth attributed to multiple birth is increasing globally, driven by the increase in assisted reproductive technologies. Twin pregnancies have approximately a 50% chance of preterm birth,[16] and almost all higher order multiple pregnancies deliver preterm.[17] Secondly, with the improvements in neonatal medicine the lower limit of viability is reducing, and so pregnancies that may have

previously been classified as miscarriages are now being classified as extremely preterm births.[18] Finally, interventions to reduce stillbirth can lead to an apparent increase in preterm live birth rate.[19] Preterm birth as a pregnancy outcome can be seen as a success when it replaces pregnancy loss, or an adverse outcome preventing further developmental time in utero. [20] Either way, it is desirable to understand the pathophysiology related to preterm birth in order to understand how to optimise gestational length.

The ill-health consequences of preterm birth mean that a strategy to reduce the burden even by a small amount could potentially equate to great improvements in healthcare costs, wellbeing of families, and most importantly improved health for the next generation.

1.3 The multifactorial nature of preterm birth

Preterm labour is a pathological condition with multiple aetiologies.[21] Figure 1.2 visualises some possible underlying aetiologies, and the concept that, individually, these could be contributing in varying amounts.

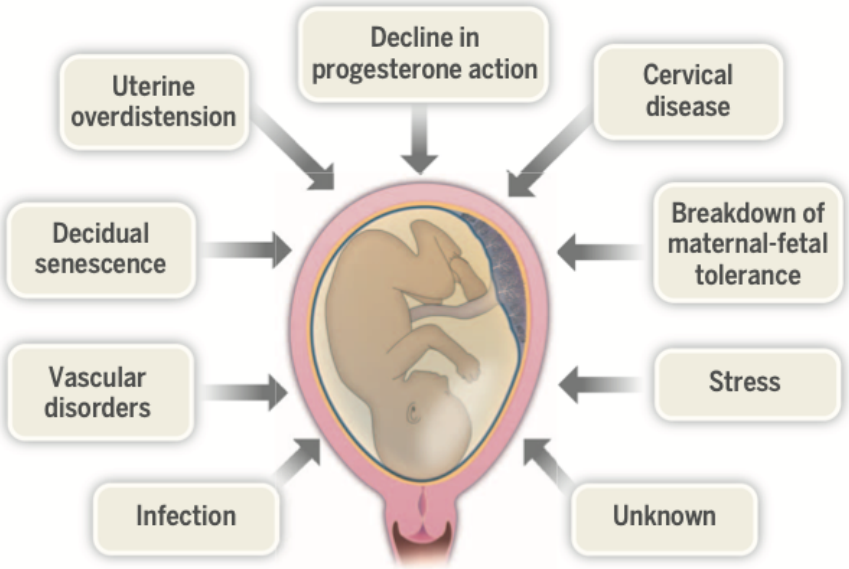


Figure 1.2: Possible mechanisms of preterm birth[22]

When a woman presents with symptoms of preterm labour the process of birth may have already started. Tocolytics have only shown benefit in delaying birth to allow time for administration of corticosteroids and in-utero transfer; in the best cases they only delay birth by an average of 4 days.[23,24]

In order to prevent preterm birth the current clinical strategy focuses on an attempt to identify women destined to have preterm labour antenatally and initiate preterm birth prevention treatment, psychological support and optimisation of the fetal condition. In 2019 NHS England published the Saving Babies Lives Version II Care Bundle recommending that every maternity provider has a designated clinician with an interest in preterm birth prevention and access to transvaginal ultrasound assessment of cervix.[25] Preterm birth prevention treatment is then initiated based on obstetric history and/ or a short cervix.[26] The primary preterm birth prevention treatment options available in the UK are cervical cerclage, vaginal progesterone and cervical pessary. Cervical cerclage and pessary are viewed as 'mechanical' therapies to strengthen a cervix and keep the cervix closed enabling the pregnancy to continue. These may also improve the 'barrier' between the vagina and the uterus thereby protecting the pregnancy from ascending microorganisms. The mechanism by which progesterone works has been suggested as a combination of averting a functional progesterone withdrawal[27] (which would have otherwise led to labour) and reducing inflammation.[28] Progesterone is also the treatment with the strongest evidence for preterm birth prevention in women with a short cervix.[29]

Clinically, these treatments are primarily targeted towards treating or preventing cervical shortening. However, in our hospital 9% of women with a previous spontaneous preterm birth under 34 weeks and a cervical length that would be considered adequate, ≥25mm in the midtrimester, still have a recurrent early preterm birth.[30] This suggests that a fair proportion of recurrent early preterm births are mediated by other pathologies that we

currently do not target. In support of this sentiment an individual patient data analysis of 4.1 million singleton births in 5 high income countries concluded that "a strong emphasis on new basic research is ultimately more likely to mitigate the problem of preterm birth than either policy or public health actions or more widespread use of currently available clinical interventions and medications".[31]

This thesis describes the conduct and findings of a cohort study which aimed to identify and understand putative causes of recurrent preterm birth that are not currently targeted by preterm birth prevention treatments. The study was designed to target two cohorts of women: a 'high-risk' cohort with a history of preterm prelabour rupture of membranes (PPROM) or spontaneous preterm birth (sPTB) under 34 weeks gestation and a 'low-risk' cohort who were parous women with all previous births at term (≥ 37 weeks gestation). The high-risk cohort was used to understand the contribution of the vaginal microbiota and clinically available data to recurrent early spontaneous preterm birth. The low-risk cohort provided reference ranges for the expected distribution of the variables studied.

1.4 Infection and preterm birth

It has been recognised for over 30 years that infection contributes to the causation of some preterm births.[32] Evidence for infection causing preterm birth includes:

- In animal experiments the introduction of both systemic and intrauterine microbes results in preterm birth.[33]
- Extra uterine infections, such as pyelonephritis are associated with preterm birth.[34]
- Antibiotics for asymptomatic bacteriuria may reduce the chance of preterm birth (95% CI 0.13-0.88). [35]
- Bacterial vaginosis is associated with preterm birth, independent of other risk factors of preterm birth.[36]

- Women delivering by Caesarean section after spontaneous preterm labour with intact membranes are much more likely to have positive chorioamniotic fluid cultures than either women who had a Caesarean section ≤30 weeks without preterm labour, or a Caesarean section after 37 weeks with labour.[37]

However, the relationship between infection and preterm birth is at times unclear. As detailed in Table 1.1 treatment of infection does not necessarily confer a reduction in preterm birth risk. Also, it is not understood how infection outside of the genital tract, for example in the dental cavity, travels to the intrauterine cavity and initiates preterm birth. It had been postulated that this was mediated via the placental microbiome,[38,39] however more recent research refutes the description of the placental microbiome and appears to comprehensively demonstrate that most placentas are sterile.[40] It is still possible that a placental microbiota may exist in some women; with a potential contribution to preterm birth in these women. Recent commentaries call for a consensus on a careful technique for further analysis of this field.[41]

The true picture of the vaginal microbiota in a given woman is likely to be much more complex than the result that can be obtained from a single swab culture. For example there will be a range of species that are not amenable to laboratory culture, and the 'load' of a particular species maybe as important as the presence or absence, but this detail is lost with culture alone. The advent of large scale, affordable culture independent techniques to assess the human microbiota was hoped to address this.

Table 1.1: Infection related risk factors for preterm birth and effects of intervention.[42]

Risk factor for preterm birth	OR or RR of preterm birth (under 37 weeks)	Risk reducing interventions	Effect of intervention
Bacterial vaginosis	Bacterial vaginosis OR of preterm birth 2.19 (95%CI 1.54-3.12)[43]	a) Screening and treatment for bacterial vaginosis within a programme of screening for infections in pregnancy b) Screening and treatment only for bacterial vaginosis	a) RR of 0.55 for preterm birth (95% CI 0.41-0.75; 2058 participants, *moderate quality evidence*) [44] No evidence of benefit in low income settings[45] b) RCT in low risk pregnancies showed no difference in PTB risk with treatment of bacterial vaginosis RR 1.10 (95%CI 0.53-2.32)[46]
Chlamydia trachomatis	OR 2.28 (95% CI:1.64–3.16)[47]	Screening and treatment for *Chlamydia trachomatis*	Treatment under 20 weeks gestation associated with lower risk of preterm birth compared to after 20 weeks, RR 0.54 (95 % CI 0.37–0.80).[48] Risk of PTB with infection in Australia 2001-12 similar to background rate, attributed to treatment. [49] No evidence of superiority of particular treatment [50]
Asymptomatic bacteriuria	OR 2.10 (95%CI 1.56-2.85)[51]	Antibiotics for asymptomatic bacteriuria	May be associated with a reduction in preterm birth RR of 0.34 for preterm birth (95% CI 0.13-0.88; two studies, 327 women, *low certainty evidence*) [52] Single study in low risk women showed no benefit in treatment for birth <34 weeks (risk 2.5% in treated vs 1.0% in untreated, risk difference -1.5% 95% CI -15.3-18.5) [53]
Periodontal disease in pregnancy	OR from 2.04 to 4.19 [54]	Periodontal treatment during pregnancy	Reduction in risk of preterm birth OR 0.65 (95% CI 0.45-0.93)[55] but uncertain [56]

Table 1.2 (continued): Infection related risk factors for preterm birth and effects of intervention.[42]

Risk factor for preterm birth	OR or RR of preterm birth (under 37 weeks)	Risk reducing interventions	Effect of intervention
Trichomonas vaginalis	RR 1.42 (95% CI, 1.15-1.75) [57]	Screening and treatment for *Trichomonas vaginalis*	RCT found increased risk of PTB with metronidazole treatment in USA (RR 1.8; 95%CI 1.2 - 2.7)[58], subsequent Cochrane review found no benefit of treatment [59]
Vaginal candidiasis	No increased risk in cohort studies[60,61]	a) Screening and treatment for vaginal candidiasis within a programme of screening for infections in pregnancy b) Screening and treatment only for vaginal candidiasis	a) RR of 0.55 for preterm birth (95% CI 0.41-0.75; 2058 participants, *moderate quality evidence*) [44] b) Pilot study showed non-significant trend towards prevention of preterm birth (RR = 0.33, 95%CI 0.04-3.03) [62]

1.4.1 The human microbiota

The human body is a complex ecosystem and hosts taxa from the entire phlogenic tree of life including eukarya, bacteria, archaea and viruses.[63–65] The human body has roughly the same number of bacterial cells as human cells,[66] and these are the component of the human microbiota that have had the most scientific interest.

The study of the human microbiota has been revolutionised by the application of culture independent sequencing technologies over the past 20 years. This project used 16S rRNA gene sequencing, which is possible because bacteria have areas of the 16S rRNA gene that have been conserved for long periods of evolution.[67] The 16S rRNA gene can be used as a target site for application of primers to amplify the number of reads using polymerase chain reaction (PCR), and then sequenced using high throughput technologies. In between the primer target sites are 'hypervariable' regions that are unique to each bacterial taxon. The 16S rRNA gene sequencing result can then cross referenced to libraries to identify the taxa present.[68]

16S rRNA sequencing was developed and initially applied to environmental samples,[68] but by the early 2000s the technique had been developed enough for application to human health and disease.[69] With a reduction in price and increase in capacity, the use of gene sequencing for microbial identification has vastly increased over the past decade. In parallel, complementary culture-independent ways of assessing the human microbiota have also been developed, such as metagenomic shotgun sequencing and metatranscriptomic sequencing. These are collectively referred to as 'next generation sequencing'. Each technique has positives and negatives, as described in Table 1.3.

Table 1.3: Techniques for assessing the human microbiota

Technique	Positives	Negatives
16S rRNA sequencing	Sequences only DNA with 16S rRNA gene (bacteria and archaea) Works with very small number of copies of gene in initial sample (as low as 10 copies) More widely used and so better reference libraries available Lower risk of false positives Lower cost Can be combined with quantitative PCR (qPCR) to estimate concentration/absolute abundance	In isolation can only be used for taxa identification and relative abundance By sequencing only a limited component of the DNA potential to miss differences in strain Not possible to sequence other microbes (for example fungi) Reliance on reference databases for taxonomy, which may be inaccurate or not optimised for use in the vaginal microbiome Poor negative predictive value of rarer members of the microbiome (due to preferential PCR expansion of more abundant species) Dependent on DNA extraction protocol appropriately lysing all cells
Shotgun metagenomic sequencing	Sequences all genetic information in the sample Possible to use for additional genetic analysis such as metabolic function profiling or antibiotic resistance profiling Able to achieve better resolution because it sequences all genetic variations	Higher risk of false positives Sequences host DNA too, so need for host DNA depletion if this is not desired, with potential to deplete DNA of interest too Higher bacterial load needed in initial sample Less widely used and so reference libraries are less complete Higher cost and computational power needed
Metatranscriptomic sequencing	Sequences all actively transcribed genes from microbes and host Lower host fraction means that is can be more cost effective than metagenomic sequencing Provides stronger evidence for functional activity than 16S rRNA sequencing or metagenomic sequencing	Unable to capture those microbes that are insufficiently transcriptionally active at the time of sampling Higher cost than 16S rRNA sequencing Reliance on reference databases

Over the past two decades next generation sequencing has been utilised in both large projects and specialised niches to characterise the human microbiota in health and disease. The largest of these was the US National Institute for Health(NIH) project 'The Human Microbiome Project'. [65,70] These projects have now generated a wealth of information, as well as 'more questions than answers',[71] which are beginning to be applied to human diseases.

An initial finding from the Human Microbiome Project was that each person appears to have a largely unique gut microbiota,[72] which appears to be influenced by geographical location, age and lifestyle.[73,74] There appears to be characteristics of the gut microbiota that are associated with conditions as diverse as inflammatory bowel disease,[75] autism,[76] and obesity.[77] Intriguingly, there is increasing evidence of a 'microbiota–gut–brain communication'[78] which links neurological conditions to microbiota 'signatures'. Features of the microbiota have been described for psychiatric conditions such as depression[79] and even neurosurgical pathology such as vascular malformations (cavernous angiomas).[80] These findings have prompted trials of gut microbiota transplants, with as yet mixed results.[81] Outside of the gut, different microbiological niches have been studied in specific diseases, for example, lung microbiota correlate with disease severity in cystic fibrosis.[82]

However, we are still searching for the best method to characterise a microbiota, and how to understand the interaction between a microbiota, host and environment. The NIH funded Human Microbiome Project finished in 2019 and their concluding summary recommended that more work is needed to understand these interactions before related therapeutics are likely to be developed.[71,83] The legacy of these initial projects is that we now have access to robust libraries to characterise individual taxa, an initial framework for characterisations of the human microbiota and initial descriptions of human microbiota in health.

1.4.2 The vaginal microbiota

The vaginal microbiota of most women of reproductive age has low diversity and is dominated by a single species of *Lactobacillus*.[84] In the non-pregnant state the microbiota of healthy women changes with the menstrual cycle[85] and sexual activity. [85–87] The vaginal microbiota also differs with ethnicity[88] and influences from external sources, such as vitamin D have been suggested recently.[89]

Vaginal dysbiosis is characterised by a deviation from the low diversity, *Lactobacillus* dominated state. Bacterial vaginosis (BV) is the most common form of vaginal dysbiosis, in which there is an increase in a range of anaerobic bacteria. There is a highly diverse bacterial community in BV, which is why the bacterial richness (total number of species observed in a given sample) and the alpha diversity index (a measure to quantify the diversity of the vaginal community in a given sample) of the vaginal microbiota in these women is higher than in Lactobacillus-dominated women.

Vaginal dysbiosis is associated with increasing inflammation, irrespective of whether the woman experiences symptoms such as itching or pain.[90] Women with BV have an increased chance of acquiring human immunodeficiency virus (HIV) and other sexually transmitted infections, with a correspondingly increased risk of pelvic inflammatory disease. [91,92]

The more complete understanding of the vaginal microbiota afforded by next generation sequencing has prompted a renewed interest in the opportunity to modify the vaginal microbiota away from a dysbiotic state using probiotics. However, probiotics have struggled to achieve sustained modifications of the vaginal microbiota, suggesting that colonisation of the vagina has not been achieved.[93] As a consequence, antibiotics remain the mainstay of treatment of bacterial vaginosis, both in the pregnant and non-pregnant state.[94]

1.4.3 The vaginal microbiota in pregnancy

The vaginal microbiota in healthy pregnancy was first characterised by 16S rRNA sequencing in the United States in 2014,[95] and shortly afterwards in the European population.[96] Both of these studies described *Lactobacillus* dominance in the majority of pregnant women, which increases as pregnancy progresses with a corresponding reduction in diversity at later gestations. The dominant species of *Lactobacillus* appears relatively stable on an individual level throughout pregnancy but appears to differ by ethnicity. In 2019 similar findings were replicated on a larger scale (n=613) in the pregnancy component of the US Human Microbiome Project.[97]

1.4.4 The vaginal microbiota and preterm birth

Given the long standing association between infection, inflammation and preterm birth the application of 16S rRNA sequencing to pregnancy was swiftly applied to the field of preterm birth research.[98] Over the past 7 years over 2000 women have been involved in studies assessing the relationship between the VMB and PTB.[99,100] The majority of these studies have suggested a contribution of the VMB to PTB, but the particular relationship remains contentious. As described above the vaginal microbiota of healthy pregnancies differs by ethnicity and gestation, and so this needs to be considered within descriptors and comparisons.

The optimal way of describing and analysing VMB compositions is yet to be determined for the PTB field. The difficulty with this is understandable; the healthy vagina in pregnancy has up to 80 different bacterial taxa,[97] and between 10,000,000 and 100,000,000 bacterial cells are expected on a high vaginal swab.[101]

Initial work assessing the association between the vaginal microbiome and preterm birth classified the vaginal microbiome into five community state types (CST) according to the

Ward hierarchical clustering. This suggested an association between community state type II (*Lactobacillus* iners dominance) and preterm birth.[102] However, the primers used in this work were not optimised to differentiate the different species of *Lactobacillus*. More recent work has moved to grouping participants based on the relative abundance of *Lactobacillus*, and attempting to identify other species or global assessments of the microbiome such as 'richness' and 'diversity' that can be used to quantify the whole microbiome and attempt to predict adverse outcome. [103]

The 'relative abundance' of a particular microbial species within a particular participant is calculated as the percentage of the 16S rRNA gene that corresponds to that species in a particular participant. This masks the differences in the numbers of 16S rRNA genes that a particular species has (range 1-9 copies), and differences in the 'density' of the bacterial load between participants. This means that one participant with 10% relative abundance of a given species and a 'bacterial load' of 10^9 microbial species in her vagina could have 10x more copies of this species in her microbiota than a participant with a total bacterial load of 10^8, but this would be masked by only assessing relative abundance. The studies of lung and gut microbiota suggest that the 'bacterial load' may be a significant part of the assessment of the microbiome.[104]

It is possible to semi-quantify 16S rRNA sequencing data by combining sequencing with a quantitative PCR (qPCR) of the 16S rRNA gene.[105] Assuming that sampling techniques are similar between participants, and using published data about the number of copies of 16S rDNA that a particular species has, it is then possible to estimate the quantity of a particular species for each participant. This technique has not been as widely used in pregnancy, but is mentioned as an important next step in a number of recent publications. [103,106,107] This technique has the additional benefit of avoiding compositional statistics, which strictly should avoid traditional regression analyses, although they are commonly applied.[108]

The BactQuant assay is a commonly used qPCR assay of the 16S rRNA gene that can be used to semi quantify the amount of bacterial DNA in a sample, and in turn the bacterial load within that sample.[105] In order to assess reproducibility of the assay the developers used plasmid standards and reported a coefficient of variance below 15% their intra-run analysis.[105] However, a logical concern with regard to estimation of bacterial load in the vagina is the effect of variance in sampling technique on the result. It would be logical to suppose that a sample obtained with more rotations, or more pressure may obtain a higher bacterial load. To my knowledge only one published paper has assessed the reproducibility of the BactQuant assay of cervico-vaginal swabs (and no studies of specifically high vaginal swabs). This study found no difference in bacterial load between two swabs in 30 women with a histologically-proven diagnosis of high-grade squamous intra-epithelial lesions of the cervix, once the volume of carrier fluid was accounted for.[109] This study was primarily designed to compare Cytobrush and swab sampling of the VMB, and therefore the amount of pressure applied/method of sampling of each participant was likely to be similar for each swab. Therefore, whilst the reproducibility between two swabs from the same participant is reassuring, the uncertainty with regard to the potential for variation in result by sample collection method should be borne in mind as a potential limitation of this assay.

1.4.5 Meta-analysis of the vaginal microbiota in preterm birth studies

Two recent attempts to perform a metanalysis of VMB in PTB studies were abandoned due to 'differences in every technical aspect of study design'.[99,110] The PREBIC preterm birth collaborative recently published recommendations for a minimum dataset for future research as a first step to address these issues.[99] However, until a universal method of performing the laboratory analysis and characterisation of the vaginal microbiota is adopted, individual patient data meta-analyses will not be feasible.

We are now entering the ‘next era’ of VMB and PTB analysis, whereby in order to advance the field it is necessary to be informed by the initial work and use this to focus the next steps to assess which initial findings are reproducible and clinically important.

2 Justification for project

The aetiology of preterm birth is still poorly understood. Women with recurrent early preterm birth/PPROM may have an underlying tendency to infection, low level inflammation, or consistent environmental influences that alter the VMB and in turn contribute to the recurrent early PTB. This project was performed to assess the contribution of the vaginal microbiota to this medley.

2.1 Rationale for components of the study

2.1.1 Rationale for focus on early spontaneous preterm birth

The focus on extremely preterm and early preterm births (<34^{+0} weeks) was chosen for two reasons. Firstly, this is the cohort with the highest morbidity and mortality, and so there is the strongest desire to understand, and ideally prevent, these events. Secondly, although preterm birth is normally defined by an arbitrary timepoint that birth occurs, the relationship between the gestation at birth and neonatal outcome may be best understood as a continuum.[111] Pregnancies with a late preterm birth are more likely to include not only pregnancies with pathology relating to preterm birth, but also 'normal' pregnancies in which the normal occurrence of labour happened to start on the earlier side of the normal distribution of spontaneous labour for that population. Therefore, we postulated that by focusing on births under 34 weeks gestation there would be a better chance of identifying pregnancies affected by pathology(ies) causing preterm birth.

The lower gestational limit of preterm birth is contentious. The WHO definition of preterm birth does not include a lower limit, and individual countries report variably from all live births or gestations between 20 and 28 weeks.[5] In the UK if a birth occurs at less than 24^{+0} completed weeks of pregnancy, and the infant shows no signs of life at birth, then the pregnancy is classified as a 'miscarriage'; and correspondingly will not be included in birth

data. However, in 2019 the British Association of Perinatal Medicine (BAPM) published a framework advocating active management for appropriately selected cases of birth from 22^{+0} weeks gestation onwards.[112] Therefore, UK classifications of preterm birth will fluctuate dependent on resuscitation practices of births between 22^{+0} and 23^{+6} weeks gestation. Risk factors and outcomes of subsequent pregnancies, are similar in women with pregnancy losses at 14^{+0} -23^{+6} weeks and women with spontaneous preterm birth after 24 weeks gestation.[113–115] In 2012 an expert working group suggested that second trimester births could be included within the analysis of preterm birth.[111] Correspondingly this thesis will focus on early preterm birth and use the gestational range 16^{+0} to 33^{+6} weeks.

The prospective nature of the cohort study meant that some participants did have late spontaneous preterm birth or PPROM. These samples were analysed and the results assessed, to help understand the complete picture, but they were not the primary focus of the project.

2.1.2 Rationale for including early PPROM

Once the membranes are ruptured at early gestations, any ongoing pregnancy remains vulnerable to serious complications including chorioamnionitis, cord prolapse and placental abruption. A genetic pathway analysis performed by our group prior to the study suggested that sPTB and PPROM have some common causes, and also some unique pathways[116] and so analysis was planned individually and combined.

2.1.3 Rationale for exclusion of medically indicated preterm births

Approximately two thirds of preterm births are either of spontaneous onset or following preterm prelabour rupture of membranes (PPROM).[117] The remaining third are medically indicated, and often related to pregnancy pathologies such as hypertensive disorders, diabetes or small for gestational age.[118–120] There is considerable overlap between these

pregnancy pathologies and preterm birth, indeed having had a previous pregnancy pathology is a risk factor for preterm birth in a subsequent pregnancy.[121] However, these pathologies are studied separately and this work will focus on preterm birth of spontaneous onset, or following PPROM.

2.1.4 Rationale for study of recurrent preterm births

Previous PTB is the strongest risk factor for spontaneous PTB.[122] This has commonly been quoted as an example of the genetic contribution to preterm birth.[123] However, this does not rule out other possible reasons for recurrence such as a microbial contribution due to low level inflammation, a susceptibility to infection as demonstrated in the first pregnancy.[124] Alternatively environmental contributions that are relatively constant between pregnancies, such as maternal nutritional state may in turn impact upon her microbiota.[125]

We targeted recruitment at women who had had a sPTB or PPROM under 34 weeks (high-risk cohort), expecting a reoccurrence rate between 10% and 20% based on our previous internal audit and external cohorts.[126]

Previous preterm birth research has been criticised for heterogeneous phenotypes contributing to heterogeneity in outcomes.[127] Focusing on this group with a high reoccurrence rate a had several advantages. Firstly, a narrower inclusion criteria reduced the heterogeneity in our population. Secondly, the high reoccurrence rate meant we needed to recruit less participants to achieve enough preterm births for analysis, and so we had the resources to focus on carefully characterising all preterm births in the cohort. This meant we were able to include only those with carefully phenotyped sPTB or PPROM. Finally the risk factor of a previous preterm birth means that that any positive findings of the research could potentially be considered as contributing to the primary preterm birth

too and lead to targeted further work looking at preterm births in primiparous women. Primiparous women make up the group with the largest amount of preterm babies born, but on an individual level without specific risk factors have a low chance of preterm birth.[117] If a targeted test and treatment could be applied in this group then large population benefits could be achieved in preterm birth reduction.

The focus on recurrent early preterm birth does have a drawback; in this study the low risk cohort was recruited only to describe the expected distribution of the variables in the local population and was not expected to provide enough participants with early sPTB or PPROM to analyse the contribution of the variables to preterm birth in this group. As such, our findings can only directly be related to recurrent early sPTB or PPROM, and these might differ from contributions to preterm birth in the population as a whole found in other literature.

2.1.5 Rationale for reference to low-risk cohort

The 'healthy' VMB in pregnancy is well described as varying by location and ethnicity,[97] but had not previously been characterised in the Liverpool population. In order to place our VMB findings within the local context we therefore characterised the VMB of parous women who delivered at term as our normal reference group.

Since our high risk cohort were, by definition of having a previous preterm birth, parous women, we chose to recruit only parous women to the low risk cohort too. There is some evidence that having had a previous pregnancy, whether or not this resulted in a livebirth, is associated with an alteration in the VMB,[128] so it was desirable to overcome this confounder. In order to identify parous women at low risk of preterm birth we recruited only those who had had all previous births at term (≥ 37^{+0} weeks gestation).

The analysis was restricted to only those low-risk women who gave birth ≥ 39^{+0} weeks gestation. This was because of previous work suggesting that this should be considered the gestation of absolute 'normality'.[129] This is because the risk of more subtle neurodevelopmental adverse outcomes actually extends into the 'early term' gestational period. Figure 2.1 illustrates that in a study of Scottish school children aged 4 to 19 years, gestation at birth has a dose-dependent relationship across the whole range of gestation.[130] This scale is exaggerated by the log function on the y axis and as an observational population study could not completely control for the current obstetric practice of recommending early term birth to women with complex pregnancies (for example diabetes, preeclampsia, growth restriction). Nevertheless it is consistent with other studies demonstrating higher mortality and morbidity amongst infants born at 37 to 38 weeks gestation compared to 39 to 40 weeks.[113,131,132]

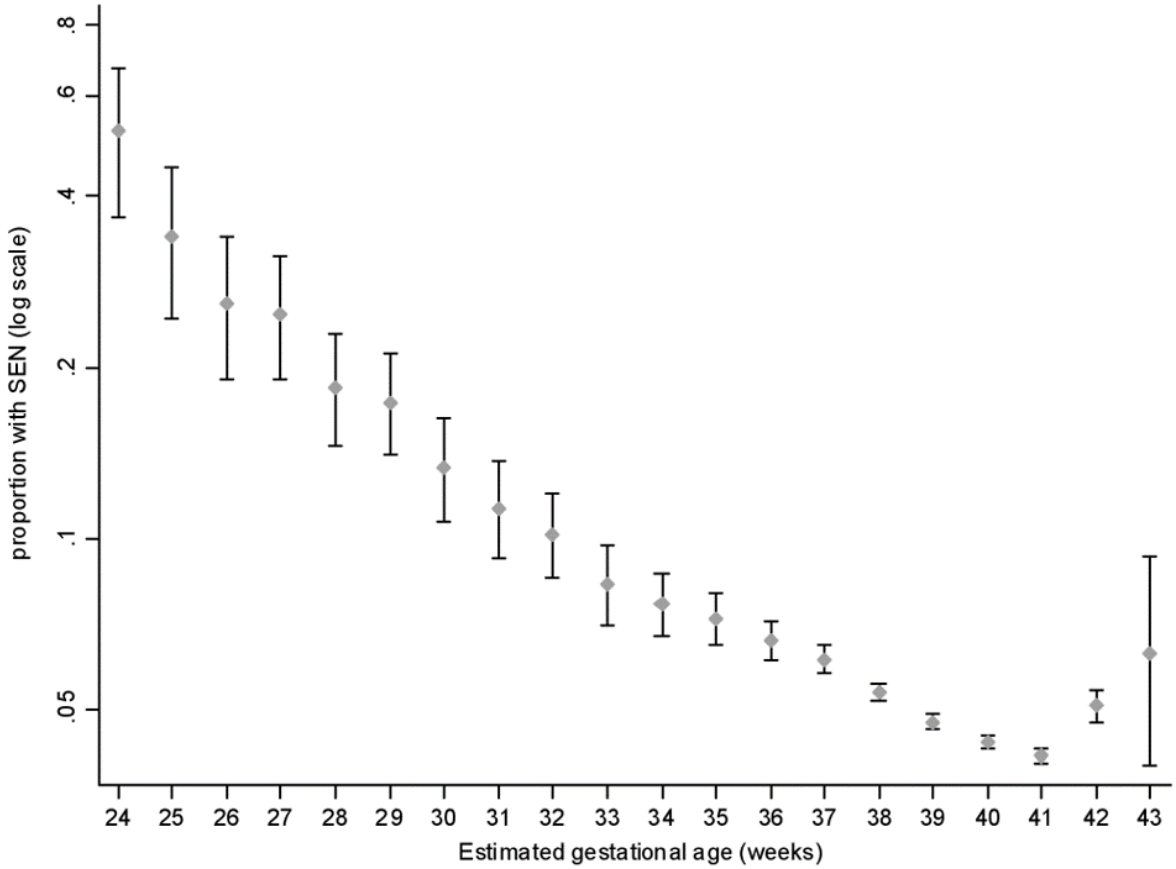


Figure 2.1: Gestational age at delivery and special educational need (SEN): retrospective cohort study of 407,503 school children[130]

2.2 Justification for reference to previous vaginal microbiota studies

This project is submitted seven years after the first assessment of the contribution of the VMB to PTB.[133] At present, most VMB in PTB studies suggest an association between a particular VMB component and PTB. However, differences in almost every technical aspect of study design make comparability effectively impossible. In order to advance this field, VMB characteristics were developed with reference to previous methods and positive findings.

2.3 Aims

1) Identify and develop methods of VMB characterisation in the second trimester that are potentially associated with sPTB/PPROM based on previous literature
2) Describe the distribution of the VMB in 'healthy' pregnancies in the second trimester in our local population, using the VMB characteristics described in (1)
3) Assess the contribution of the vaginal microbiota in the second trimester to recurrent early sPTB/PPROM, with reference to the expected distribution of VMB characteristics in 'healthy' pregnancies
4) Assess whether any VMB characteristics associated with recurrent early sPTB/PPROM identified in (3) are also associated with late sPTB/PPROM, in high-risk women

2.4 Objectives

1) To review published studies that have assessed the association between the VMB and PTB, alongside studies of the VMB in non-pregnant women that have characterised pathologies known to be associated with PTB. This body of literature will be used to select previously used VMB characteristics for inclusion in this project, and develop new VMB characteristics for analysis
2) Vaginal swab samples in the second trimester from parous women with all previous births at term, and birth $\geq 39^{+0}$ weeks in the index pregnancy, will be analysed by 16S rRNA gene sequencing and qPCR of the 16S gene. These samples will be used to determine the distribution of the VMB characteristics identified in (1) in the 'healthy' pregnancies
3) Vaginal swab samples will be collected in the second trimester from women with a history of sPTB or PPROM at 16^{+0} to 33^{+6} weeks gestation. Those women who go on to have a recurrent sPTB or PPROM at 16^{+0} to 33^{+6} weeks gestation will form our

'cases'. Participants who go on to give birth at ≥37^{+0} weeks without PPROM will form our 'controls'. These two groups will be used to perform a nested case-control study, with aetiological modelling, to assess the contribution of the VMB (using the characteristics identified in (1)) to recurrent early sPTB/PPROM. These results will be compared to the expected distribution of the VMB characteristics in the low-risk reference group.

4) Any VMB characteristics that are identified as having an association with recurrent early sPTB or PPROM in (3) will also be assessed for within high-risk participants who have a late sPTB or PPROM. Aetiological modelling will then be used to assess the association between these characteristics and late sPTB or PPROM, in comparison to high-risk women who give birth at ≥37^{+0} weeks without PPROM, in the high-risk group.

3 Participants and Methods

3.1 Introduction

The Harris-Wellbeing Preterm Birth Research Centre team conducted a prospective observational study entitled "The development of novel biomarkers for prediction of preterm labour in a high-risk population" study. Recruitment to the first stage of the study started on 1 April 2012. The vaginal microbiota and clinically available data components were added on 19th October 2015 with an ethical amendment that approved the collection of vaginal swabs and the recruitment of the low-risk cohort. I joined the study team on 1st February 2016. Recruitment was completed on 31 December 2017 and all participants had given birth by the end of June 2018.

This thesis describes a nested case-control study to assess the relationship between the vaginal microbiota and recurrent early preterm birth in women at high-risk of sPTB, with reference to the expected distribution of the vaginal microbiota in women at low-risk of sPTB.

3.2 Ethical approval

The study was approved by North West Research Ethics Committee - Liverpool Central, reference 11/NW/0720, on 4th November 2011. The amendment in October 2015 retained the original reference code. All participants gave written informed consent.

3.3 Funding

The prospective cohort study was funded as part of a charitable donation from Lord and Lady Harris, gained in an open competition that was facilitated by the charity Wellbeing of Women. This donation founded the Harris-Wellbeing PTB Research Centre, University of Liverpool. This covered administrative costs, laboratory analysis, and study support costs

for myself.

3.4 Study design

Pregnant women were recruited from Liverpool Women's Hospital, UK, in two groups:

- Healthy parous women with no risk factors for preterm birth. These women formed our low-risk reference group.
- Women with a history of at least one previous sPTB or PPROM at 16^{+0}-33^{+6} weeks gestation. These women formed our high-risk group.

3.4.1 Whole study inclusion criteria

Inclusion criteria for both groups of the study were:

- Singleton pregnancy
- Having had early pregnancy 'dating' scan[134] before 14^{+0} weeks gestation showing no fetal abnormalities
- Between 15^{+1} and 23^{+0} weeks gestation at first study visit
- Standard antenatal care in the UK includes offering all women HIV and hepatitis B testing at their 'booking' appointment with a midwife.[134] Potential participants were required to have taken part in this programme, and the results to have been negative.
- Participants were only eligible for study participation for one pregnancy, so they could not be recruited a second time during a subsequent pregnancy.
- Not having had vaginal sex within 48 hours prior to the first study visit

The early pregnancy 'dating' scan was used to calculate gestational age at sampling and birth. The HIV and hepatitis testing was stipulated as part of the laboratory risk assessment for the parallel analysis of biomarkers present in the blood that were also assessed in this project (separate manuscripts in preparation). The uptake of HIV testing in UK antenatal

services in 2015 was 98.2%, with only 0.15% of women testing positive[135] so this is unlikely to have had a substantial impact upon recruitment.

3.4.2 Low-risk group recruitment

These participants were women who were planning to give birth at Liverpool Women's Hospital, or with the hospital's linked homebirth team.

In the UK all pregnant women are encouraged to 'book' with an antenatal care provider in the first trimester, and it is anticipated that the majority of women will receive antenatal and intrapartum care from the same provider.[134] Liverpool Women's Hospital[136] is one such care provider, with the largest stand-alone maternity unit in the UK and a dedicated midwifery led birthing suite. As such it provides care for many women who have 'low-risk' pregnancies and 'midwifery led antenatal care'[134] (as well as being a tertiary referral centre for complex pregnancies).

Low-risk participants were selected to reflect the 'healthy pregnant' population as much as possible. Additional inclusion criteria for this group were:

- Having had at least 1 previous pregnancy that progressed to at least 37^{+0} weeks gestation

Exclusion criteria for this group were:

- Previous late miscarriage (≥ 16^{+0} weeks gestation)
- Previous preterm birth (<37^{+0} weeks gestation), including medically induced preterm birth
- Previous PPROM (<37^{+0} weeks gestation)
- History of significant cervical surgery*
- Medical condition considerably affecting health†

- Previous pregnancy affected by an obstetric complication with an increased chance of reoccurrence[#]

* Significant cervical surgery was defined as either: 1 large loop excision of the transformation zone (LLETZ) over 10mm; 2 or more LLETZ; or a knife cone biopsy

[†] A medical condition considerably affecting health was defined as women who had been referred to a specialist antenatal clinic for a medical condition by the 'booking midwife' (first healthcare professional who met woman in pregnancy and took a detailed medical history). For example (not limited to): epilepsy, inflammatory bowel disease, or renal impairment.

[#] A previous pregnancy affected by an obstetric complication with an increased chance of reoccurrence was defined as previous gestational diabetes, preeclampsia requiring magnesium sulphate, or gestational hypertension requiring antihypertensives.

The low-risk participants were recruited as follows:

1. Study clinicians used hospital electronic records to identify women who had 'booked' for antenatal care[134] to be provided by Liverpool Women's Hospital and fitted the eligibility criteria as described above.
2. Study staff then telephoned or sent a text message to potential participants to offer them involvement in the study. The offer of involvement was given between the early pregnancy 'dating' scan (described above) and 15^{+0} weeks gestation. Each potential participant was only contacted once, with an answerphone message left if possible. A smaller number of the low-risk cohort were recruited by self-identification from study leaflets (Appendix A) distributed by the community midwives or the same information from the leaflet that was posted on social media by the research group and Liverpool Women's Hospital. Potential participants were

invited to telephone or send a text message to a dedicated research number and then eligibility was confirmed by the research team.

3. Potential participants were spoken to on the telephone and the research study was briefly explained, including the rationale and need for speculum examination with vaginal swabs and cervical length scan. Participants were offered a 3D ultrasound scan of the pregnancy at each study visit.
4. Potential participants who wanted to proceed with study recruitment were then scheduled for their first study visit at approximately 16 weeks gestation (15^{+1}-18^{+6} weeks). Study visits then proceeded as described below.

The process for identification and approach of the both groups of pregnant women was approved by the ethics committee with input from the Harris-Wellbeing patient and public involvement group. The use of electronic case notes for participant identification predated the changes to UK general data protection legislation.[137]

3.4.3 High-risk group recruitment

Additional eligibility criteria to this group were:

- Having had at least one previous preterm spontaneous preterm birth (sPTB), preterm prelabour rupture of membranes (PPROM) or spontaneous late miscarriage at 16^{+0}-33^{+6} weeks gestation.
- Not using preterm birth preventative treatment (cervical cerclage/vaginal progesterone/ vaginal pessary) at the time of first study visit.

There were no additional exclusion criteria for the high risk cohort, and in particular women could take part in this group if they had had significant cervical surgery, previous obstetric or medical problems.

These participants were prospectively recruited from the Liverpool Women's Hospital Preterm Labour Prevention Clinic. Study visits were timed to coincide with the participant's scheduled visits to the preterm birth prevention clinic (within the study schedule).

3.4.4 Study schedule

Participants for were invited to two study visits at approximately 16 weeks gestation (15^{+1}-18^{+6} weeks) and approximately 20 weeks gestation (19^{+0}-23^{+0} weeks). Participants who were recruited after 19^{+0} weeks gestation only had a single study visit.

Participants who had a second study visit were asked to abstain from sex for 48 hours prior to the study visit.

3.4.5 Procedure for study visits

3.4.5.1 *First visit*

At the first study visit an obstetric doctor confirmed eligibility and provided a participant information leaflet (Appendix B low-risk and Appendix C high-risk). The doctor and participant then completed a written consent form (Appendix D) and a study number was allocated.

The doctor then completed a data collection form (Appendix E low risk and Appendix F high-risk) detailing the following:

- Age in years (at time of first study visit)
- Height and weight (these were used to calculate the body mass index)
- Home postcode

- Smoking history
- Self-reported ethnicity
- Current gestation and estimated due date
- Method used to estimate due date
- Number of previous pregnancies affected by sPTB or PPROM between 16^{+0} and 33^{+6} weeks gestation (only relevant for the high-risk cohort)
- Whether the participant had previous cervical surgery, and if so, this was classified into one of three groups: (1) nil significant (for example no cervical surgery or punch biopsy); (2) single LLETZ <10mm (large loop excision of the transformation zone); or (3) a single LLETZ ≥10mm, multiple LLETZ or at least one knife cone biopsy
- Medication that the participant had taken in the previous 72 hours (including vitamin supplements)
- Whether antimicrobials had been used in the pregnancy, and if so the name, indication and dates of administration
- Whether the participant had had vaginal sex in the previous 48 hours (if so the participant was not recruited to this part of the study)

The participant then had blood taken for additional components of the study. Next, the participant had a speculum examination, with a chaperone present, and five vaginal swabs were taken as below. A cervical length scan was performed after the speculum examination.

The 3D ultrasound scan was then performed, and photographs provided to the participant. Any questions were addressed as necessary, and the second clinic visit was scheduled. High-risk participants were managed in the standard manner for the preterm birth prevention clinic. Preterm birth prevention therapy (cerclage, vaginal pessary or vaginal progesterone) was offered if cervical length was ≤ 25mm or based on clinician and patient

preference if there was considered to be a large change in cervical length between measurements.

3.4.5.2 *Collection and storage of cervico-vaginal swabs*

Firstly, a sample of cervico-vaginal fluid was taken from the posterior fornix (10 seconds of rotation) for quantitative fetal fibronectin (qfFN) assessment using Rapid fFN 10Q System (HOLOGIC, Marlborough, MA, USA). Secondly, three high vaginal swab samples were collected using HydraFlock standard tapered swabs (Medical Wire and Equipment, Corsham, England). Two of these were frozen dry and the other was placed in 1ml of 'RNAlater' (Merck Life Science UK Ltd, Dorset, UK). Thirdly a high vaginal swab was used to collect a sample of fluid from the posterior fornix and sent to the NHS laboratory for standard culture and sensitivity.

The qfFN swab was processed by laboratory technicians and the result recorded. Obstetric doctors were blind to the qfFN result until all participants in the study had given birth. The HydraFlock standard tapered swabs were stored in 1.2ml cryogenic tubes (Fisher scientific, Loughborogh, UK) at -80°C within an hour of sampling.

3.4.5.3 *Second study visit*

At the second study visit the obstetric doctor verbally confirmed ongoing willingness to participate in the study, then completed the second section of the data collection form (Appendix E low-risk and Appendix F high-risk).

Information gathered at the second visit included:

- Whether any new concerns or pathologies had been identified with the pregnancy
- Whether preterm birth prevention treatment was currently being used (if so what, and when started)
- Medication usage within the previous 72 hours (including vitamins)

- Use of antimicrobials since the previous study visit, and if so the name, indication and dates of administration
- Whether the participant had had vaginal sex in the previous 48 hours

The blood sampling, speculum examination with vaginal swabs, cervical length scan and fetal scan were then repeated in the same way as the first visit. At the end of the second visit patient contact details were confirmed and verbal consent was re-iterated to contact the patient for pregnancy outcome data if this wasn't possible on the hospital electronic system.

3.4.5.4 *Further data collection*

Preterm birth clinic records were reviewed to ascertain whether preterm birth prevention treatment (cervical cerclage, pessary or vaginal progesterone) was used after enrolment in the study. No low-risk participants required preterm birth prevention therapy, so this was only relevant to the high-risk group.

Hospital records were used to ascertain delivery details for all women delivering at our unit. If participants gave birth elsewhere then clinicians at the delivering unit were contacted and asked to provide delivery details. Where this wasn't possible the participants were contacted directly by telephone. The GROW birthweight centile[138] was calculated using the mother's height, weight, ethnicity, parity and the infant's birthweight to facilitate comparison of birthweights across the groups despite differences in gestation at birth.

3.4.5.5 *Data handling*

Study data was inputted into a Microsoft Excel spreadsheet, stored on the University of Liverpool's secure server. The Index of Multiple Deprivation (IMD) was obtained using the woman's home postcode on the UK government website.[139] The IMD ranks every neighbourhood in England from 1 (most deprived) to 32844 (least deprived).[140] The IMD is a

collective score summarising income deprivation, employment deprivation, health deprivation and disability, education skills and training deprivation, barriers to housing and services, living environment deprivation, and crime.

3.5 Classification of birth outcome

3.5.1 Low-risk group

When a participant in the low-risk group gave birth ≥39^{+0} weeks gestation, the hospital electronic records were reviewed to confirm there was no evidence of PPROM <37 weeks gestation (or if the woman delivered elsewhere this was confirmed with the unit where she gave birth). If no PPROM <37^{+0} weeks was confirmed, then the participant was allocated to the low-risk reference group.

The remainder of participants within the low-risk group were excluded from the analysis. However, their births were still classified in the same way as the high risk women for completeness and to facilitate later analysis, as described in Figure 3.1. In cases of birth or PPROM <37^{+0} weeks detailed review was undertaken as described in section 3.5.3.

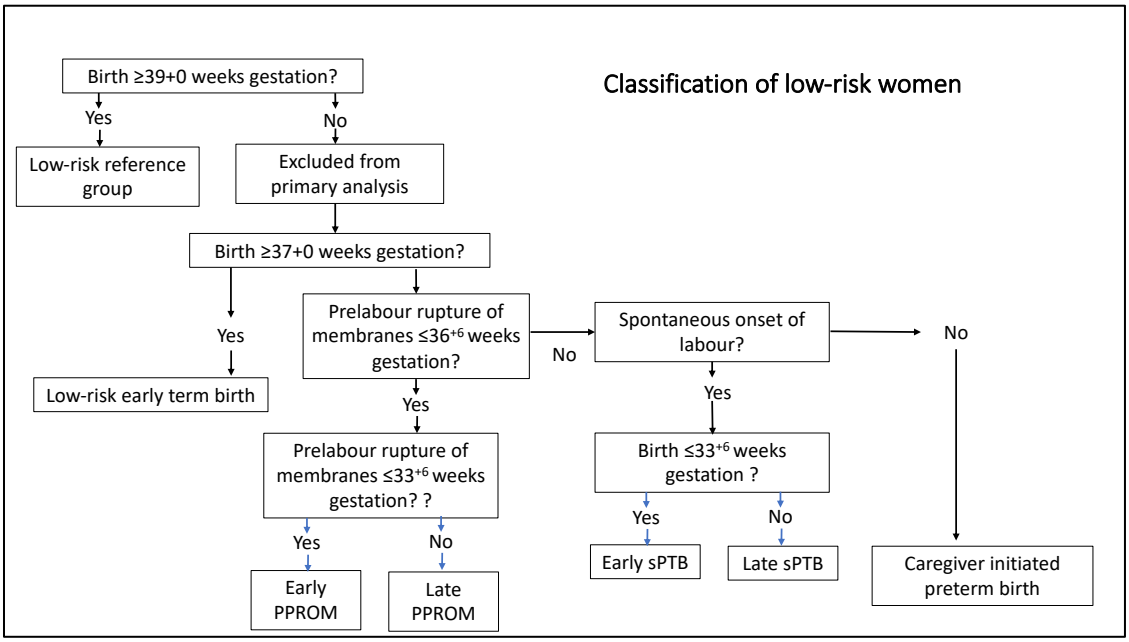


Figure 3.1:Workflow for classification of births in the low-risk group. Only those low-risk participants who gave birth ≥39+0 weeks gestation formed the 'low risk reference group' and were included in the primary analysis. The remainder of the low-risk births were still classified for pregnancy outcome purposes but not included in the primary analysis.

3.5.2 Term births in the high-risk group

When a participant in the high-risk group gave birth ≥37^{+0} weeks gestation the hospital electronic records were reviewed to confirm there was no evidence of PPROM <37^{+0} weeks gestation (or if the woman gave birth elsewhere this was confirmed with the unit where birth occured). If no PPROM <37^{+0} weeks was confirmed, then the participant was allocated to the high-risk reference group.

3.5.3 Preterm births and PPROM

Dr Angharad Care, Harris Wellbeing clinical research fellow, and I, independently reviewed all available data relating to all births or PPROM <37^{+0} weeks gestation, including hospital records and supplementary reports from other hospitals if participants had care elsewhere. In the cases of discrepancy the case was reviewed by Professor Alfirevic until the team reached a consensus on classification of the birth.

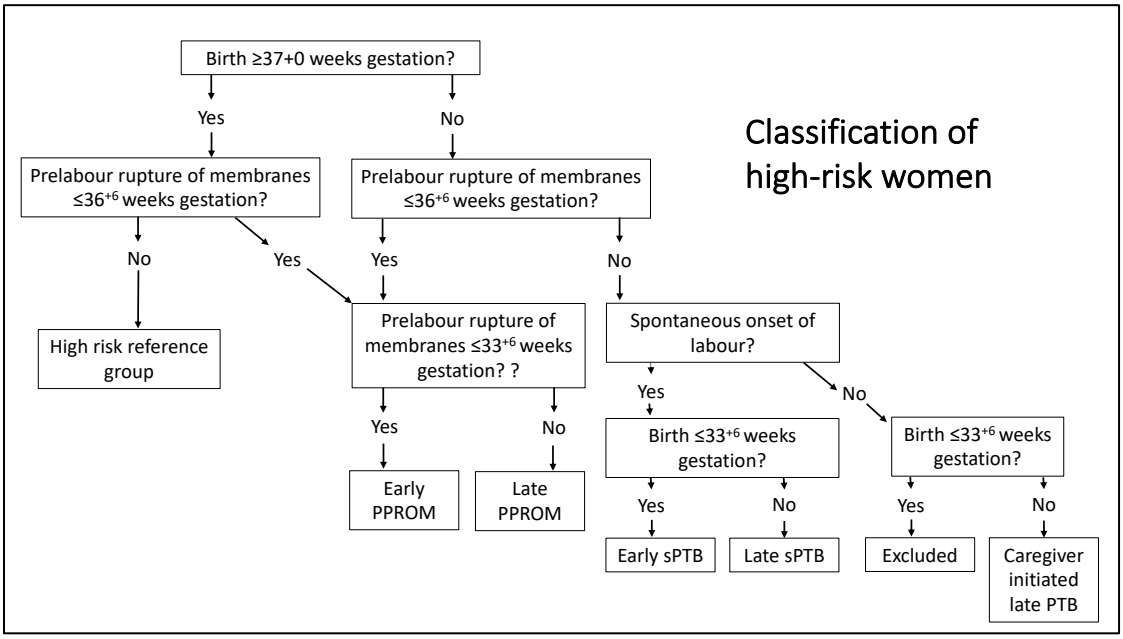


Figure 3.2: Workflow for classification of births in the high-risk group

Figure 3.2 shows the workflow for the classification of births, and Table 3.1 shows how clinical diagnoses were defined. The judgement of the contemporaneously recorded treating clinicians was used to record diagnoses unless further information became available at a later time that refuted this. For example, if a diagnosis of PPROM was made by the treating clinicians based on history and examination, then this diagnosis was kept unless further information later refuted this. Further information that could refute this would be ultrasound scans showing normal amniotic fluid levels and the absence of ongoing vaginal leakage of fluid. In order to achieve as homogeneous population of sPTB/PPROM as possible participants where the diagnosis of sPTB or PPROM was uncertain were excluded. High-risk participants with obstetric complications, for example preeclampsia or gestational diabetes were included if they also clearly had sPTB or PPROM.

Table 3.1: Definitions used for classification of birth outcomes

Classification	Description
Spontaneous labour	Regular uterine activity with cervical shortening or dilation such that the treating clinicians judged labour to be present
Prelabour rupture of membranes	Rupture of membranes confirmed either by speculum examination or the AmniSure ROM test (QIAGEN, UK) without onset of spontaneous labour in the following 12 hours
Caregiver initiated preterm birth	Induction of labour or Caesarean section ≤36^{+6} weeks gestation *without* evidence of PPROM or spontaneous preterm labour

3.5.4 Additional classification for cases of PPROM

In cases of PPROM we were particularly interested to ascertain whether chorioamnionitis was thought to be present at the time of PPROM (and so could have been implicated in the pathway to PPROM), or whether PPROM occurred in the absence of chorioamnionitis. We therefore further classified cases of PPROM according to whether there was evidence of chorioamnionitis within 7 days of PPROM, or not (Figure 3.3, Table 3.2).

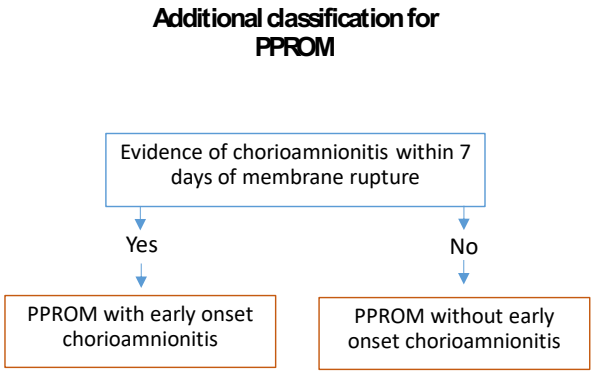


Figure 3.3: Additional classification for women with PPROM

Table 3.2: Definition of chorioamnionitis developed for this study

Term	Description
Chorioamnionitis	Contemporaneous clinical notes documenting concern about chorioamnionitis due to one or more of: raised white cell count; raised C-reactive protein (CRP); maternal pyrexia; or the use of broad spectrum antibiotics to treat chorioamnionitis based on the clinical situation. *Not histologic evidence of chorioamnionitis on placental histology alone (subclinical chorioamnionitis)*

3.5.5 Conclusion of major classification of birth outcomes

The review of all births culminated in all participants being classified into one of 15 mutually exclusive groups as described in Table 3.3.

Table 3.3: Major classification of all births

Participant population	Major classification	Description of birth	Classification code
High-risk *Previous PPROM or spontaneous preterm birth 16^{+0}- 33^{+6} weeks gestation*	Term	Birth ≥37+0 weeks gestation	A
	Caregiver initiated preterm birth	Caregiver initiated preterm birth	B
	Early PPROM	PPROM ≤33^{+6} weeks gestation with early chorioamnionitis	C
		PPROM ≤33^{+6} weeks gestation without early chorioamnionitis	D
	Late PPROM	PPROM 34^{+0}-≤36^{+6} weeks gestation with early chorioamnionitis	E
		PPROM 34^{+0}-≤36^{+6} weeks gestation without early chorioamnionitis	F
	Early sPTB	sPTB ≤33^{+6}	G
	Late sPTB	sPTB 34^{+0}-≤36^{+6}	H
	Unknown	Unable to ascertain birth details	I
Low-risk *Parous women with all previous births ≥37^{+0} weeks gestation*	Term	Birth ≥39^{+0} weeks gestation	LA
	Caregiver initiated preterm birth	Caregiver initiated preterm birth	LB
	Early PPROM	PPROM ≤33^{+6} weeks gestation with early chorioamnionitis	LC
		PPROM ≤33^{+6} weeks gestation without early chorioamnionitis	LD
	Late PPROM	PPROM 34^{+0}-≤36^{+6} weeks gestation with early chorioamnionitis	LE
		PPROM 34^{+0}-≤36^{+6} weeks gestation without early chorioamnionitis	LF
	Early sPTB	sPTB ≤33^{+6}	LG
	Late sPTB	sPTB 34^{+0}-≤36^{+6}	LH
	Unknown	Unable to ascertain birth details	LI
	Early term birth	Birth 37^{+0}-≤38^{+6}	LJ

3.5.6 Contributing factors to preterm birth

In addition to the major classification of birth Dr Care and I reviewed the medical information and attributed 'contributing factors' to the women with sPTB and PPROM (classification codes C-H and LC-LH). This was based on the Clinical Phenotyping Tool developed by Villar *et al* (2012).[141] Of note the contributing factors are not mutually exclusive, and some women had no discernible contributing factors to preterm birth. The tool developed by Villar *et al* also included maternal stress and family history in the classification system. We did not measure maternal stress around the time of delivery, and the project includes a genetic component, and so these were excluded from our classification system. The contributing factors that we did use are detailed in Table 3.4.

Table 3.4: Potential contributing factors to sPTB/PPROM that were assessed for within the medical records and recorded if present

Contributing factors to preterm birth	Description
Chorioamnionitis	Contemporaneous clinical notes documenting concern about chorioamnionitis due to one or more of; raised white cell count, raised C-reactive protein (CRP), maternal pyrexia, or the use of broad spectrum antibiotics to treat chorioamnionitis. *Not histologic evidence of chorioamnionitis on placental histology alone (subclinical chorioamnionitis)*
Placental dysfunction	Evidence of placental dysfunction contributing to preterm labour. Defined as evidence of placental abruption at time of delivery or on placental histology report, birthweight under 5th customised centile, or severe preeclampsia
Extra amniotic infection	Contemporaneous clinical notes documenting concern about major systemic infection due to one or more of; raised white cell count, raised CRP, maternal pyrexia, microbiological culture of pathological organism from a normally sterile site, or the use of broad spectrum antibiotics for presumed extra amniotic infection.
Polyhydramnios	Maximum pool depth ≥10cm on ultrasound assessment
Uterine anomaly	Documented uterine anomaly
Maternal comorbidities	Maternal medical condition that affects a major organ system (for example chronic hypertension, chronic renal failure, epilepsy, pre-existing and gestational diabetes)
Cervical insufficiency	Received treatment for short cervical length <28^{+0} weeks gestation

3.6 Final study inclusion and exclusion criteria

This nested case-control study retained six groups of participants for analysis. These were:

1) *Low-risk reference group*: Low-risk participants (as described in 3.4.2) who gave birth ≥39^{+0} weeks gestation (LA classification code, Table 3.3)
2) *High-risk reference group*: High-risk participants (as described in 3.4.3) who gave birth ≥37^{+0} weeks gestation, without PPROM (A classification code, Table 3.3)
3) *High-risk late sPTB group*: High-risk participants (as described in 3.4.3) who had an sPTB 34^{+0}-36^{+6} weeks gestation, (H classification code, Table 3.3)
4) *High-risk late PPROM group*: High-risk participants (as described in 3.4.3) who had PPROM 34^{+0}-36^{+6} weeks gestation (E and F classification codes, Table 3.3)
5) *High-risk early sPTB group*: High-risk participants (as described in 3.4.3) who had an sPTB 16^{+0}-33^{+6} weeks gestation, (G classification code, Table 3.3)
6) *High-risk early PPROM group*: High-risk participants (as described in 3.4.3) who had PPROM 16^{+0}-33^{+6} weeks gestation (C and D classification codes, Table 3.3)

In addition to the global inclusion criteria (section 3.4.1) and low and high risk group specific criteria (sections 3.4.2 and 3.4.3) the final study had the additional inclusion criteria:

A vaginal microbiota sample result available for analysis (having passed the quality control process) with:

- The participant having abstained from vaginal intercourse in the 48 hours prior to the study visit

and

- The participant not using preterm birth prevention therapy at the time of study visit

We only used the first set of valid sequencing results for each participant for the majority of 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 we used data from both visits for the stability analyses.

3.7 Laboratory process for assessment of the vaginal microbiota

3.7.1 DNA extraction

DNA extraction and sequencing were performed at the University of Liverpool Centre for Genomic Research. DNA was extracted from one sample per participant per visit. Birth outcomes were not known at the time of DNA extraction. A total of n=706 swabs from 364 participants underwent DNA extraction, PCR amplification and 16S rRNA sequencing and rarefaction. After this the analysis was restricted to only the participants eligible for the case-control study as described in section 3.6.

The samples were thawed, and DNA was extracted by adding 180 µl of enzymatic lysis buffer containing lysozyme to the sample (Sigma-Aldrich, Dorset, UK); incubation for 30 minutes at 37 ^{0}C; 25 µl of proteinase K and 200 µl of buffer AL were added from the Qiagen DNeasy Blood and Tissue kit (Qiagen, Manchester, UK); incubation for 30 minutes at 56^{0} C; addition of 200 mg of 0.1 mm zirconia/silica beads (Thistle Scientific, Glasgow, UK) then bead-beating for 5 minutes at 25 Hz on a Qiagen TissueLyser II (Qiagen, Manchester, UK). Next 200 µl of 100% ethanol was added to the sample and it was centrifuged. After centrifugation the swab head was discarded, and the pellet was purified in four subsequent centrifugation steps after adding one-by-one 200 µl 100% ethanol, 500 µl buffer AW1, 500 µl buffer AW2 and 75 µl buffer AE as per manufacturer's instructions (Qiagen, Manchester,

UK). In order to facilitate detection of contaminants downstream we included one negative control (an empty tube) with each DNA extraction round of 23 study samples. The DNA concentration of all samples was measured by Qubit (Invitrogen, Thermo Scientific, Paisley, UK) and the DNA quality of all samples by Nanodrop (Thermo Scientific, Paisley, UK). Samples with a particularly low DNA concentration or quality were discarded (n=25) and the DNA extraction process was performed on the second sample taken during that study visit (these samples subsequently achieved acceptable concentration and quality assessments).

3.7.2 PCR amplification and 16S rRNA gene sequencing

Two PCR rounds were performed for each DNA sample (study samples and negative controls) for 16S rRNA gene amplification and barcoding. Firstly, the V3-V4 region of the 16S rRNA gene was amplified as described previously.[142] DNA was amplified using 1.25 µl of a 10 µM concentration of 319F 5'-ACTCCTACGGGAGGCAGCAG-3' forward primer and 1.25 µl of a 10 µM concentration of 806R 5'-GGACTACHVGGGTWTCTAAT-3' reverse primer, 12.5 µl NEB Next HF 2x PCR Master Mix (New England Biolabs, Hitchin, UK), 9 µl of nuclease-free water and 1 µl of DNA extraction product to make a 25 µl reaction volume. The first denaturation cycle was performed for 30 seconds at 98 °C, followed by 10 cycles with a denaturation cycle of 10 seconds (at 98 °C), an annealing cycle of 30 seconds (at 58 °C), an extension cycle of 30 seconds (at 72 °C), and finally an extension cycle of 5 minutes at 72 °C. PCR products were then purified and size-selected using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) in a 1:1 bead-to-sample ratio. The final PCR round used the standard Illumina Nextera XT index kit v2 (Illumina, San Diego, CA, USA), aimed at V3-V4 sequences by a dual-index approach. This permits multiplexing of up to 384 samples at a time (two rounds were performed to accommodate all samples). The barcoding used 2.5 µl of Index 1 primer, 2.5 µl of Index 2 primer, 12.5 µl NEB Next HF 2x PCR Master Mix

and 7.5 µl sample making a 25 µl reaction volume. The first denaturation cycle was performed for 3 minutes at 98 ^{0}C, followed by 15 cycles with a denaturation cycle of 30 seconds (at 98 °C), an annealing cycle of 30 seconds (at 55 °C), an extension cycle of 30 seconds (at 72 °C), and a final extension cycle of 5 minutes at 72 ^{0}C. AMPure beads were then used to purify PCR products as explained above, also using a 1:1 bead-to-sample ratio. Each PCR run also had a negative control (10 µl of nuclease-free water instead of 9 µl of nuclease-free water and 1 µl of DNA) to identify contaminants, and 10 µl of 0.2 ng/µl ZymoBiomics Microbial Community DNA standard (Zymo Research Corp, Irvine, CA, USA), a commercially available positive control. The PCR runs also included the DNA extraction negative controls. DNA collected from the same participant at different visits were included in the same PCR run. The Qubit Fluorometer with the dsDNA HS Assay kit (Invitrogen, Thermo Scientific, Paisley, UK) was used to measure PCR product DNA concentrations of each sample (including negative and positive controls). Two negative controls were not successfully amplified and so were excluded from the subsequent steps, all participants samples, positive controls and the remainder of the negative controls were used for subsequent steps.

Amplicons from samples were evenly pooled into sequencing libraries at a mass of 0.8 ng DNA per amplicon. To achieve this, Qubit DNA concentrations and Fragment Analyzer (Agilent, Santa Clara, USA) quality control information were combined for pooling before size-selection using Pippin Prep (Sage Scientific, Beverly, Massachusetts, USA). Samples with a DNA concentration of <0.30 ng/µl (such as the negative controls) were added in a fixed volume of 1 µl. The two libraries were sequenced on an Illumina HiSeq instrument (Illumina, San Diego, CA, USA), run in rapid mode, 2x300bp using a 250PE and 50PE kit. DNA collected from the same participant at different visits was included in the same library.

3.7.3 Panbacterial 16S rRNA gene qPCR

Extracted DNA from all participant samples (n=706 samples from n=364 participants) was sent to the Institute for Genome Sciences of the University of Maryland (Baltimore, MD, USA) for estimation of the panbacterial 16S rRNA gene copy concentration using the BactQuant qPCR assay. This assay is based on an analyses of 4,938 16S rRNA gene sequences in the Greengenes database.[105,143] The analysis was performed as described previously.[105,144] Briefly, 1.5 µl of template (1:10 diluted DNA) was added to 3.5 µl of reaction mix, with the final reaction containing 1.8 µM each of the forward (341F) and reverse (806R) primer targeting the 16S V3-V4 region, 225 nM of the TaqManW probe, 1X Platinum Quantitative PCR SuperMix-UDG with ROX (Invitrogen, Thermo Scientific, Waltham, MA, USA) and molecular-grade water. For each experiment there was an in-run standard curve (ranging from 10 to 10^8, with 10^2–10^8 in 10-fold serial linear dilutions) and no-template controls performed in triplicate. Amplification and real-time fluorescence detections were performed on the Bio-Rad CFX 384 instrument (Bio-Rad Inc., Hercules, CA, USA). The PCR conditions were: 3 minutes at 50 °C for UDG treatment, 10 minutes at 95 °C for Taq activation, 15 seconds at 95 °C for denaturation and 1 minute at 60 °C for annealing and extension, times 40 cycles. Cycle threshold (Ct) value for each 16S qPCR reaction were obtained using a manual Ct threshold of 0.05 and automatic baseline. The 16S rRNA gene concentration was reported in copies/µL for each sample.

Quality control of the BactQuant assay was performed by excluding samples that did not amplify in two of three, or all three, of the triplicate qPCR cycles, or had skewed low 16S rRNA gene concentration results of <1,000 copies/µl.

3.7.4 Molecular data processing

A mean raw unpaired read count of 368,205 reads per study sample (95% confidence interval (CI): 353,388 – 383,022 reads) was obtained. Cutadapt v1.16[145] was used to

demultiplex the reads, and remove primer sequences. All subsequent steps were performed DADA2 version 1.8 package for large paired end datasets in R version 3.5.1 (R core team, 2015).[146] DADA2 was chosen because of its ability to resolve reads to a single nucleotide. The *fastqFilter* command was used for error correction with parameter settings aiming to maximize read retention. The minimum read lengths (*truncLen*) were set to 255 for forward reads and 210 for reverse reads based on the quality plots, *maxEE* to a maximum of 5 for forward read and 8 for reverese read expected errors, *maxN* to zero ambiguous bases allowed, and *truncQ* to zero. Approximately 10% of reads were discarded after error correction. The *learnErrors* command was then used to determine the read error rates. The reads were then assigned to unique amplicon sequence variants (ASVs, equivalent to a taxon) using the *derepFastq* command, and ASVs with higher than average error rates were discarded (denoised) using the *dada* command.[146,147] The *mergePairs* command was then used to merge forward and reverse reads. The *removeBimeraDenovo* command was used to remove chimeric compositions of two separate parent ASVs (Bimeras) with the Silva version 132 database as the reference database;[148] 6.3% of ASVs were identified as bimeric and removed. Overall, a median of 28% of the raw reads per study sample were removed during these DADA2 clean-up process.

DADA2 was used to perform taxonomic assignment in two steps. Firstly *assignTaxonomy* was used to map ASVs to taxa at genus level or above using the RDP classifier with a minimum bootstrap value of 50% and the Silva v132 database as the reference database.[148,149] Secondly *addSpecies* was used to map ASVs to species level. Only ASVs with exact (100%) identity matches with species in the Silva database were assigned to each species.

3.7.5 Further data processing

A spreadsheet containing the sequences, taxonomic assignments, and read counts for each ASV per sample was imported into Microsoft Excel for Mac version 16. ASVs with a read count in all samples combined of less than 100 were removed, as well as two non-bacterial ASVs, and one likely contaminant ASV. The likely contaminant was a *Cutibacterium* genus that was present in two negative controls at a relative abundance that was higher than in any study sample. The Silva v132 database[148] did not include the vaginal taxa BV-associated bacterium 1 (BVAB1), BVAB2 and BVAB TM7 sequences, but these have been published elsewhere[100,150] and were identified manually in our dataset. The sequences for *Mageeibacillus indolicus* (BVAB3) and *Fenollaria massiliensis* were also evaluated but not present in our dataset. The taxonomic assignment derived from the Silva database[148] was double-checked for the 112 ASVs with a relative abundance of at least 0.05% of the read count of all samples combined (out of a total of 1646 ASVs) using the *Microbial Nucleotide BLAST (BLASTn)* function on the National Center for Biotechnology Information NCBI website.[151] In cases of discrepancy the Vaginal 16S rDNA Reference Database[152] was used as a tiebreaker. This resulted in 45 *Lactobacillus* genus ASVs being reassigned to various *Lactobacillus* species, two *Streptococcus* genus ASVs being reassigned to *S. agalactiae,* a *Staphylococcus* genus being reassigned to *S.aureus,* a *Gardnerella* genus ASV being reassigned to *G. vaginalis*, an *Atopobium* genus ASV being reassigned to *A. vaginae,* a *Sneathia* genus ASV being reassigned to *S. amnii* and a *Enterococcus* genus ASV being reassigned to *E. faecalis*. Read counts for ASVs assigned to the exact same taxonomy were then summed for each sample. The lowest total read count for any specific sample above 1000 reads was 1101, so *GuniFrac* 1.1 package in R was used to rarefy to 1101 reads. The rarefied ASV table contained 690 samples and 290 unique ASVs. The *prop.table* function in R was then used to transform rarefied read counts into relative abundances.

3.8 Selection of participants and samples to retain for analysis

3.8.1 Selection based on pregnancy outcome

277 of the 364 participants had a pregnancy outcome as described in section 3.6 and were eligible for analysis in this project. These participants contributed 484 samples.

3.8.2 Failed quality control of 16S rRNA analysis

16 out of 706 samples (from n=364 participants) became invalid due to rarefaction, 11 of which were samples eligible for inclusion in this project. Ten of the participants that provided these samples had an alternative study visit sample suitable for analysis (i.e. study visit attended, not using preterm birth prevention therapy at the time of study visit and no sex in previous 48 hours), so only one participant for the current analysis was excluded at this stage.

3.8.3 Failed quality control of the panbacterial 16S rRNA gene qPCR

31 out of 706 samples (from n=364 participants) did not produce valid qPCR results from the BactQuant assay, 19 of which were samples eligible for inclusion in this project. Fifteen of the participants that provided these samples had an alternative study visit sample suitable for analysis, so four participants for the current analysis were excluded at this stage.

3.8.4 Final inclusion

Of the 484 samples from 277 participants eligible for analysis, 454 (93.8%) of samples from 272 (98.2%) of participants had valid 16S rRNA sequencing and valid BactQuant assays. Only these samples and participants with valid BactQuant assays and 16S rRNA data were retained for further analysis.

The rarefied ASV relative abundance table consisted of 276 ASVs in 454 samples (for 272 individual participants), mapping to species (181; 65.6%), genus (82; 29.7%), or higher

taxonomic levels (13; 4.7%) (Appendix G). This was used for all methods of data reduction and analysis.

3.9 Estimation of bacterial load and taxon concentration

The ASV-specific concentrations per sample and overall bacterial load were estimated by combining using the sample-specific 16S rRNA gene concentration data (BactQuant assay, section 3.7.3) with the rarefied ASV table (16S rRNA sequencing data, section 3.7.5).

For each of the 276 unique ASVs included within the rarefied ASV table the 16S rDNA gene copy number was identified within in the NCBI version of the rrn database,[153] and in the case of missing data the Greengenes database.[143] In situations where ASVs were mapped to multiple species at genus level, then the mean 16S gene copy number based on all potential species was calculated and used. If the database didn't detail the mean 16S gene copy number of a species then the mean copy number of the corresponding genus was used. For BVAB1 and BVAB2 only 'order' level taxonomic information (*Clostridiales* order). We therefore used the *Clostridiales* order mean copy number (=4.62). The ASV-specific copy-normalized rarefied relative abundance was multiplied by the sample-specific 16S rRNA gene copies concentration to estimate the concentration of each ASV in cells/µl per sample. This method has previously been shown correlate with species-specific quantitative PCR results for non-minority species.[154,155] The concentrations were then $\log_{10}$-transformed. To prevent skewed negative values concentration results less than one cell/µl were set to one prior to $\log_{10}$-transformation.

3.10 Statistical analysis

3.10.1 Statistical packages

Statistical analysis, bar charts, and scatter plots were made in STATA version 15.1 (StataCorp, College Station, TX, USA). Heatmaps showing the twenty ASVs with highest

median relative abundance were made using the *gplots* package in R. The eucalan chart was made using a website produced by Meta-Chart.[156]

3.10.2 Analysis planned

3.10.2.1 *Early sPTB/PPROM*

Descriptive statistics were used to describe the distribution of VMB characteristics in the low-risk women who delivered at ≥39^{+0} weeks gestation without PPROM. This group formed the low-risk reference group.

Our primary comparisons were between high-risk women who gave birth ≥37^{+0} weeks gestation and high-risk women who had a recurrent early sPTB or PPROM <34^{+0} weeks. Characteristics between these groups of interest 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. Initially the analysis was performed for the separate outcomes of early PPROM and early sPTB, and then these groups were combined for more detailed analysis.

The VMB variables that were selected were compared between the high-risk who delivered at term and the high-risk early sPTB/PPROM preterm groups using unadjusted and adjusted logistic regression, with adjustments for body mass index (BMI) as a quadratic term, history of cervical surgery, and smoking. If a significant association between a continuous VMB variable and early sPTB/PPROM recurrence was identified, and the variable was present in over 30% of participants, then quartiles of the expected distribution of that VMB variable 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 high-risk term and preterm groups 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.

For categorical outcomes, or if the prevalence of the taxa was so low that quartiles were not practical (n<5 early PTB participants with the taxon present) then the low-risk participants are reported alongside the high-risk, in order to give a reference range for the expected distribution.

Euclan diagrams (proportional Venn diagrams) were produced to visualise the co-location or independence of VMB features that were associated with a higher incidence of early PTB for the high-risk term births and high-risk early sPTB/PPROMs. This was performed to help understand whether our multiple analysis methods were identifying the same participants in multiple ways, or different participants with disparate VMB features associated with PTB.

3.10.2.2 *High-risk late preterm births and low-risk preterm birth groups*

If VMB characteristics were identified as associated with an increased or decreased risk of early sPTB/PPROM then these characteristics were also assessed for within the high-risk late preterm birth group, and the low-risk preterm birth groups. This was a targeted exploratory analysis to assess whether those characteristics were replicable. The design of the study was such that samples were donated by participants, and VMB analysis performed, prior to the outcome of the pregnancy being known, therefore the VMB analysis had already been performed for these participants in the course of the study.

3.10.3 Calculation of sample size

No formal calculation of sample size was undertaken as this was primarily an exploratory analysis and pragmatically planned addition to the already running “The development of novel biomarkers for prediction of preterm labour in a high-risk population study”.

3.10.4 Management of multiple comparisons

In order to comprehensively assess the relationships between the VMB and recurrent early PTB that had been identified in previous work this study inevitably had to perform a large number of comparisons. 88 primary comparisons were performed. Using the traditional 5% significance level this would be expected to give a positive result in 4-5 comparisons. However, the picture is a little more complicated because many of the classifications of the VMB have significant overlap (for example VMB type/CST/group according to *Lactobacillus* dominance and the interdependencies of individual taxa). Therefore, a positive or negative finding using one classification system would be expected to correlate to a positive or negative finding using another (similar) classification system.

The statistical analysis is further complicated because this study involved a combination of confirmatory analysis (for methods of VMB data reduction that had previously been linked to risk of PTB, such as those described in Table 4.3), and exploratory analysis using the bacterial types, bacterial groups and the BactQuant assay to estimate concentrations of bacterial groups which was applied to the field of PTB research for the first time in this study. This means that for some components an exploratory analysis is more appropriate, but for others a confirmatory analysis is more appropriate.

Given that preterm labour is best described as a pathological condition with multiple aetiologies,[21] it would not be expected that a single VMB characteristic would have a strikingly strong association with PTB, and small contributions to PTB risk in limited numbers of participants may best explain the contribution of the VMB to PTB. Therefore, in order to fairly present the data, the hypothesis-driven variables that we designed (VMB types, and presence/absence, relative abundance and concentration of bacteria in each 'bacterial group') were not adjusted for multiple comparisons (16 calculations).

In order to account for the large number of analyses performed within the confirmatory analysis component of the study (75 calculations), the Benjamini-Hochberg false discovery rate (FDR) procedure[157] was performed for associations with early sPTB/PPROM within the high-risk cohort, with an FDR rate of 25%. However, mindful of the propensity of this method to produce type II statistical errors the results were also viewed holistically and findings that appeared to confirm trends seen in previous studies were also highlighted, and a full description of the association observed described even if this did not meet the significance level at 0.05%, before or after correction for multiple testing.

3.10.5 Management of missing data

Within the high-risk group the co-variates of BMI and smoking had one and two missing values respectively. The adjusted bimodal logistic regression was performed both without these participants, and with the participants assuming they did, then didn't, smoke and had a low and high BMI. The results did not materially change, and so the adjusted bimodal logistic regressions are presented with exclusion of participants with missing data in order to present the raw data.

3.10.6 Consideration of core outcome sets

The CROWN initiative has defined a core outcome set for PTB.[158] We report on PPROM, birthweight, and gestational age at birth. There were no cases of maternal mortality. The remainder of the core outcome set is not relevant to this study. During the course of the research, the PREBIC consortium[159] made recommendations for a minimum dataset for research on the VMB in PTB.[99] Despite the recommendations not having been formulated during data collection, 24/26 of the 'essential' criteria are presented, and 11/18 of the 'desirable' criteria (Tables 3.5 and 3.6).

Table 3.5: Essential criteria for VMB in PTB studies as defined by PREBIC consortium[99]

Item required	Detail or reason for absence
Age	Table 5.3
Race/ethnicity	Table 5.3
Parity	Table 5.3
BMI	Table 5.3
Smoking status	Table 5.3
History of sexual transmitted infection	Information not collected
History of PTB	Table 5.3
Indication of previous PTB	All previous PTB either sPTB <34 weeks or after PPROM <34 weeks. Details in Table 5.3
Information on included singleton/multiple pregnancy	All singleton pregnancies
Exclusion of other complications in pregnancy leading to PTB	Study team cases of PTB other than sPTB or PPROM. Figure 5.1
Information on use of antibiotics before sampling	Table 5.3
Information on use of antibiotics after sampling	Data not collected
Adequate assessment of gestational age	All pregnancies had 'dating' scan prior to 14 weeks gestation
Gestational age at sampling	Table 5.3
Single or longitudinal sampling	Two sampling timepoints, approximately 16 and 20 weeks gestation
Information on primary swab	HydraFlock standard tapered swabs, (Medical Wire and Equipment, Corsham, England)
Sample location	All swabs taken and stored at Liverpool Women's Hospital, UK
Information on primers used	'PCR amplification and 16S rRNA gene sequencing' section 3.7.2
Range of bacteria covered by primers	Good coverage of vaginal microbiota, as described by Van der Pol et al[160]
Attrition record	Figure 5.1
Statement of outcome measures	Section 4.3

Table 3.5 continued

Item required	Detail or reason for absence
Definition of PTB	Section 3.5
Indication for PTB of index pregnancy	Table 5.3
Stratification of PTB by phenotype	Attempted, but as described in section 7.3 no difference identified and data presented as a whole to improve readability and statistical power
Races/ethnicities analysed separately	Not applicable because over 90% of population white.
Analysis of lactobacilli to species level	Table 6.6 and Table 7.6
Other interventions for PTB excluded (eg cerclage, pessary)	Women with interventions for PTB not eligible for recruitment. Participants with interventions after VMB analysis are included, as shown in Table 5.3

Table 3.6: Desirable criteria for VMB in PTB studies as defined by PREBIC consortium[99]

Item required	Detail or reason for absence
Marital status	Not collected
Socioeconomic status	Table 5.3
Alcohol use	Not collected
Substance use	Not collected
Gestational age of previous PTB	All previous PTB were sPTB or PPROM 16^{+0} -33^{+6} weeks
History of late miscarriage	Table 5.3
History of LLETZ	Table 5.3
Date of last sexual intercourse prior to sampling	Participants asked to abstain for 48 hours prior to sampling
Reported history of douching	Not collected
Self collected or physician collected samples	Physician collected
Measurement of cervical length	Table 5.3
Use of fetal fibronectin	Table 5.3
Measurement of pH	Not collected
Simultaneous cultivation	Not performed by research team
Objective measurement of BV by microscopy	Not performed by research team
Consent to use specimens	Consent documented for 'gifting' of samples to other ethically approved research
Other subtypes of PTB (< 34 weeks, <28 weeks)	Primary outcome sPTB or PPROM <34 weeks. Insufficient sample size to justify further subtypes
Quantitative/qualitative analysis	Performed comprehensively. Chapters 6, 7 and 8

4 Vaginal microbiota characteristics

In order to assess whether VMB characteristics that had previously been associated with preterm birth were reproducible in our population we summarised VMB in PTB studies.

Three further methods of VMB characterisation were also developed based on published literature on the VMB in non-pregnant women and applied to PTB research for the first time.

4.1 Previous work assessing the vaginal microbiota and preterm birth

In order to identify previously published associations between VMB characteristics and PTB a PubMed search was performed on 14.1.20 as shown in Table 4.1

Table 4.1: PubMed search terms (all fields)

Search number	Search terms	Number of studies identified
1	Preterm* OR premature*	220182
2	Vagin* AND microbiota	1491
3	Birth OR labour	367152
4	1 AND 2 AND 3 limit to English language	131

Abstracts and/or full text papers were then reviewed to select studies using the following criteria:

Inclusions:

- Observational cohorts, case-control or cross-sectional studies that used next generation sequencing or 16S rRNA qPCR to primarily assess the relationship between vaginal microbiota and preterm birth, preterm premature rupture of membranes and/or late miscarriage

Exclusions:

- Case reports and case series that analysed the effects of interventions, such as vaginal pessary and cervical cerclage, on the lower genital tract microbiota or on PTB
- Studies that included women with symptoms of preterm labour or preterm premature rupture of membranes at the time of sampling
- Studies that included first trimester miscarriage within the PTB outcome

Studies that included women at high and low risk of preterm birth, with or without preterm birth prevention treatment were included. This methodology identified 14 articles from which data was extracted, as shown in the PRISMA flow diagram[161] (Figure 4.1). Characteristics of the included studies are shown in Table 4.2.

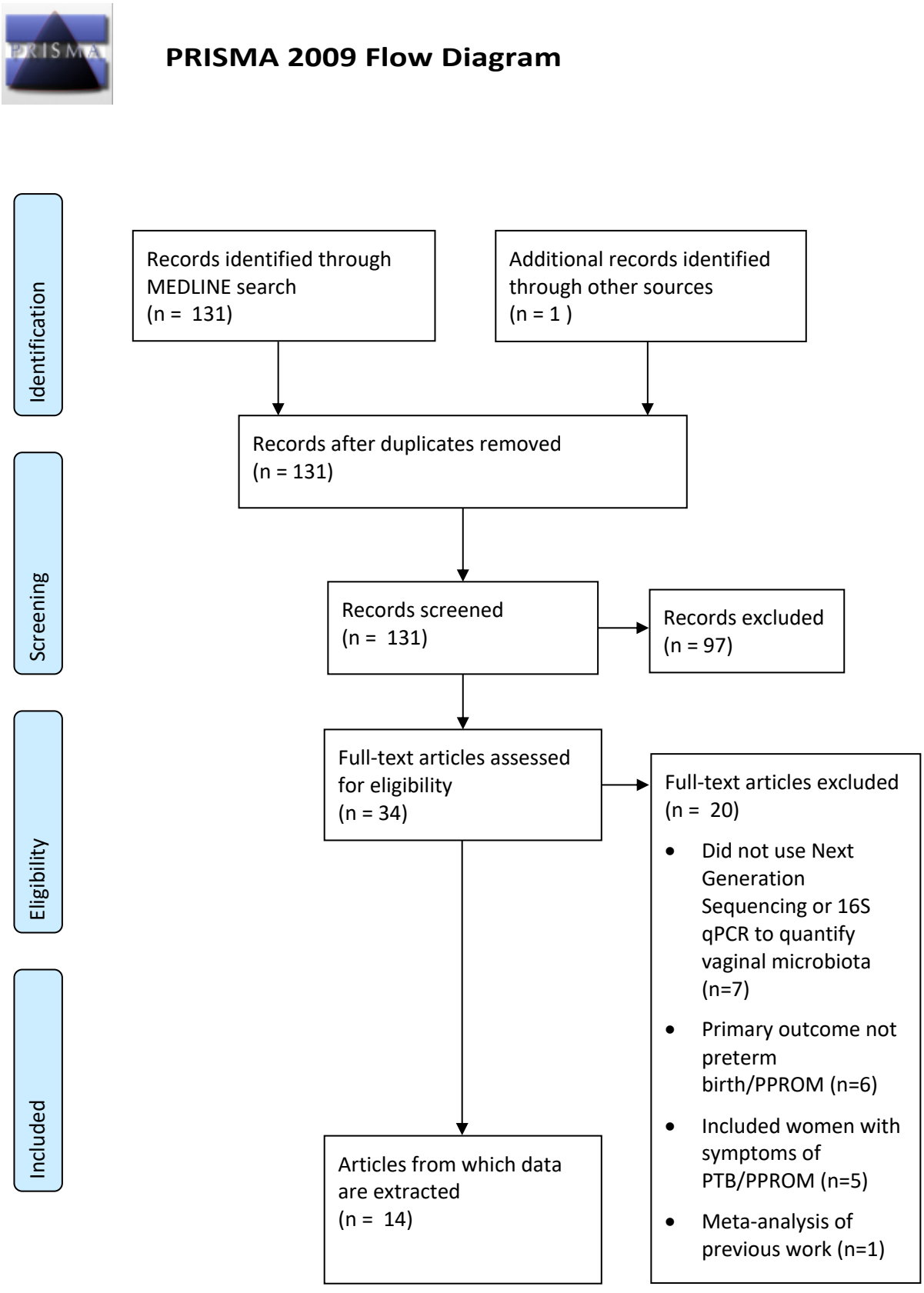


Figure 4.1: PRISMA flow diagram showing selection of studies used to identify VMB characteristics previously associated with PTB

Table 4.2: Study Characteristics		Romero 2014[98]	DiGiulio 2015[162]	Subramania 2016[163]	Nelson 2016[164]	Kindinger 2017[102]	Callahan 2017[165]	Stout 2017[166]	Wheeler 2018[167]	Freitas 2018[168]	Tabatanaei 2019[106]	Brown 2019[103]	Elovitz 2019[101]	Fettweis 2019[100]	Blostein 2020[169]
Population	Number of term birth participants	72	34	20	27	184	85	53	37	170	356	36	432	90	100
	Number of PTB participants	18	15	20	13	44	50	24	14	46	94	60	107	45	25
	Inclusion of sPTB cases	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
	Inclusion of PPROM cases	✓	✓		✓	✗	✓	✓		✓		✓	✓	✓	✓
	Inclusion of medically indicated PTB	✗	✓	✗	✗	✗	✓	✓		✗	✗	✗	✗	✗	✗
	Percent of PTB under 34 weeks	100%	20%	100% (<35)		52%	58%	33%		13% (<32)	18%		75% (<35)		
	Gestation of sampling (weeks)	6-birth	10-birth	21-25^{+6}	9-24	16-34	6-birth	6-36	6-25^{+6}	11-16	8-13^{+6}	6-36^{+6}	16-28	6-birth	<16
Laboratory methods	16S rRNA sequencing	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
	Variable region sequenced	1-3	3-5, 4	4	4	1-3	4	1-3, 3-5	4		4	1-2	3-4	1-3	4
	qPCR of 16S rRNA gene	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗
	Alternate NGS of eukaryotic DNA*	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗
Ethnicity	Caucasian/White/European ancestry	6%	59%	50%	0%	65%	23%		24%	60%	73%	51%	21%	14%	
	Black/African American	87%	4%	50%	100%	19%	59%	69%	76%	2%	7%	23%	75%	78%	
	Asian		12%	0%	0%	17%	4%		0%	14%	4%	26%			
	Other/mixed	6%	25%	0%	0%	0%	13%	31%	0%	24%	16%	0%	4%	7%	
	Control for confounders (either statistically or by matching)	✗	✗	✗	Omitted previous antibiotics	Age, BMI, ethnicity	✗	GA at sampling	Ethnicity	✗	GA at sampling	Age, BMI, ethnicity	Ethnicity	Age, race, income	Parity, ethnicity
Relationship between VMB and PTB suggested		✗	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓

*Table 4.2: Characteristics of studies used to identify VMB characteristics and taxa of interest. Abbreviations: qPCR, quantitative PCR; NGS, next generation sequencing; BMI body mass index; GA gestational age. *Alternate NGS of eukaryotic DNA was performed by cpn60 universal target sequencing[168] and Whole shotgun metagenomic/metatranscriptomic sequencing[110]. ✓='yes', ✗='no'. Grey=not available/not applicable.*

4.2 Identification of vaginal microbiota characteristics of interest

For each of the 14 included studies (Table 4.2) the main manuscript was first assessed, and all VMB characteristics that the authors identified within their main manuscript as showing a relationship with PTB were recorded. The 'relationship with PTB' was included if the authors had identified a difference in the VMB parameter between participants who had term and preterm births, using each individual study's method for participant recruitment, PTB classification, VMB data analysis, and test of significance. The remaining 13 studies were then reviewed to identify whether a comparable method of VMB data reduction was used in the other studies, and if so whether this identified an association with PTB, or not. This secondary analysis utilised all available parts of the manuscripts including supplementary results where available. For individual taxa where it was apparent that the study had identified this taxon within their sequencing data (either by reviewing heatmap results or lists of taxa identified), and then a biomarker discovery method was used, such as LEfSe[170] (Linear discriminant analysis effect size), if the biomarker discovery method did not highlighted this taxon as having a relationship with PTB then this was classified as having been measured and showing no affect. When there was no evidence either within the main document or supplementary material of a taxon having been identified then this was judged to be not measured. For some individual taxa testing of the association between the taxon and PTB was inferred within the manuscript or supplementary material and no association was reported, but the statistical test results are not shown, or the level of assessment is uncertain. This discrepancy is noted within the resultant table, but did not alter the decision about whether a taxon was selected for analysis, or not.

Taxa of interest and VMB data reduction methods were retained for analysis if the method had shown an association with PTB in two or more previous studies, irrespective of whether this relationship was in the same direction, for example the opposing findings about

bacterial load in Freitas *et al*[168] and Elovitz *et al.*[101] In addition it was acknowledged that the sequences for BVAB1, BVAB2 and BVAB TM7 have only recently become available, and so despite single studies showing a relationship between these taxa they were retained because historical studies would not have been able to assess these. Two further taxa were retained as 'taxa of interest' despite a relationship with PTB only having been identified in single studies: *Lactobacillus jensenii*, retained to maintain assessment all of the major vaginal *Lactobacillus* species; and *Bifidobacterium breve*, retained because of the plausible relationship with a reduction in inflammation[171] and its candidacy as a non-*Lactobacillus* 'protective' taxon. This produced six VMB data reduction methods and 23 taxa of interest for analysis in the present study (Table 4.3)

Table 4.3: Findings associated with PTB		Romero 2014[98]	DiGiulio 2015[162]	Subramania 2016[163]	Nelson 2016[164]	Kindinger 2017[102]	Callahan 2017[165]	Stout 2017[166]	Freitas 2018[168]	Wheeler 2018[167]	Tabatanae 2018[106]	Brown 2019[103]	Elovitz 2019[101]	Fettweis 2019[100]	Blostein 2020[169]
Global sample assessments	Species richness				⇔	⇔		↑	↑	⇔		↑			
	Species diversity (alpha diversity)	⇔	↑	⇔	⇔	⇔		↑	↑	⇔	⇔	⇔	⇔	↑	⇔
	Community State Types (CSTs)	⇔	IV ↑		⇔	CST III ↑ CST I ↓			⇔		IV ↑		IV ↑	CST I ↓	⇔
	Group according to *Lactobacillus* relative abundance					⇔	Low *lactos*					Low *lactos*			⇔
	Instability						↑	↑				↑		⇔	
	Total bacterial load								↑			⇔	↓		
Relative abundance of specific taxa	*Lactobacillus* species	⇔	↓		⇔	⇔	↓	⇔	↑			↓			↓
	Lactobacillus crispatus	⇔				↓	↓	⇔	⇔		↓	⇔	⇔	↓	
	Lactobacillus iners	⇔				↑	⇔	⇔	⇔		↓	⇔	↑	⇔	
	Lactobacillus jensenii	⇔				⇔	↓	⇔	⇔		⇔	⇔	⇔	⇔	
	Lactobacillus gasseri	⇔				⇔	↓	⇔	⇔		↓	⇔	⇔	⇔	
	Aerococcus	⇔			↓	⇔	↑	⇔	⇔		⇔		⇔	↑	
	Atopobium vaginae	⇔				⇔	↑	⇔	⇔		⇔	⇔	↑	⇔	
	Bifidobacterium breve				⇔	⇔	⇔	⇔	⇔		↓	⇔	⇔		
	Clostridiales BVAB2	⇔					⇔	⇔					⇔	↑	
	Dialister	⇔			⇔	⇔	↑	⇔	↑		⇔	↑	⇔	↑	
	Gardnerella vaginalis	⇔	↑		⇔	⇔	↑	⇔	⇔		⇔	⇔	⇔	⇔	
	Lachnospiracea BVAB1	⇔					⇔	⇔					⇔	↑	
	Mageeibacillus indolicus												↑		
	Megasphaera	⇔			⇔		⇔	⇔	↑		⇔	⇔	↑	↑	
	Mobiluncus					⇔	↑	⇔	↑				↑		
	Mycoplasma					⇔	↑	⇔	↑				⇔	⇔	
	Parvimonas	⇔			⇔	⇔		⇔	↑			↑		↑	

Table 4.3 continued	Romero 2014[98]	DiGiulio 2015[162]	Subramania 2016[163]	Nelson 2016[164]	Kindinger 2017[102]	Callahan 2017[165]	Stout 2017[166]	Freitas 2018[168]	Wheeler 2018[167]	Tabatanaei 2018[106]	Brown 2019[103]	Elovitz 2019[101]	Fettweis 2019[100]	Blostein 2020[169]
Peptoniphilus					⇔	↑	⇔	↑		⇔	↑	⇔		↑
Prevotella	⇔			↓	⇔	↑	⇔	↑		⇔	↑	⇔	↑	
Porphyromonas species				⇔			⇔	↑				↑		
Sneathia species	⇔			↓	⇔	⇔	⇔			⇔	⇔	↑	↑	
Streptococcus species	⇔			⇔	⇔	↑	⇔	↑		⇔	↑	⇔	⇔	
BVAB TM7-H1	⇔					⇔	⇔						↑	
Ureaplasma species	⇔	↑			⇔	⇔	⇔	↑			⇔	⇔	⇔	

Table 4.3: VMB characteristics and taxa of interest identified as having an association with PTB in previous literature. Abbreviations: BVAB1, BV-associated bacterium type 1; BVAB TM7, BV-associated bacterium (phylum TM7); Lactos, Lactobacillus species; CST, community state type, as first described by Ravel et al[98]. CST is variably defined over the studies but in the positive associations CST I is characterised by dominance of L. crispatus, CST III is characterised by dominance of L. iners and CST IV is characterised by non-dominance of Lactobacillus and a VMB made up of mixed bacterial species. Findings associated with preterm birth in previous studies. ↓=reduced in preterm birth cases compared to term, ↑=increased in preterm birth cases compared to term, dark grey= not assessed or not available ⇔on white background = tested and no difference clearly shown in manuscript and/or supplementary material , ⇔on light grey background= testing is inferred within manuscript and/or supplementary material and no difference shown but statistical tests not shown and/or level of assessment is unclear

4.3 Vaginal microbiota characterisation methods chosen

Based on Table 4.3 six methods of VMB data reduction that had previously shown an association with PTB were selected.

4.3.1.1 *Richness*

This was calculated by counting the number of discrete ASVs, identified in each sample.

4.3.1.1.1 Simpson Diversity Index

This is a value from 0-1 calculated by:

$$D = 1 - (\sum n(n-1)/N(N-1))$$

Key

D Simpson diversity

n Relative abundance of each taxa

N Total number of taxa in the sample

Simpson's diversity index is the probability that two taxa drawn at random from an infinitely large community will be different species. Simpson's diversity index is expressed as the reciprocal (1 minus the diversity), meaning that higher values represent higher diversity.

4.3.1.2 *Community State Types*

Each sample was allocated to one of five mutually exclusive groups defined by the dominance of a particular species of *Lactobacillus*, or a group with non-dominance of a *Lactobacillus* species. Table 4.4 shows the community state type definitions used, as first described by Ravel *et al*.[84]

Table 4.4: Description of the Community State Types (CSTs)

Community State Type (CST)	Dominant species
I	*Lactobacillus crispatus*
II	*Lactobacillus gasseri*
III	*Lactobacillus iners*
IV	Group without dominance of a *Lactobacillus* species
V	*Lactobacillus jensenii*

4.3.1.3 *Groups based on relative abundance of Lactobacillus species*

Each sample was allocated to one of three mutually exclusive groups based on the relative abundance of *Lactobacillus* species as described by Brown and MacIntyre *et al.*[103,172,173] These were: (1) *Lactobacillus* dominant (≥75% *Lactobacillus* species); (2) Intermediate (50-74.9% *Lactobacillus* species); and (3) *Lactobacillus* deplete (≤50% *Lactobacillus* species).

4.3.1.4 *Stability groups*

Only participants who had valid VMB sample results available for both study visits (approximately 16 and 20 weeks gestation) were eligible for this analysis. Each eligible participant was allocated to a mutually exclusive stability group (Table 4.5). These groups were based on the method used by Romero *et al,* [95] modified to account for only two visits.

Table 4.5: VMB stability groups. LD, Lactobacillus Dominant

Group name	Descriptor
Same Lactobacillus	Both visits ≥75% relative abundance of Lactobacillus with the same Lactobacillus dominance at each (either L.iners, L.crispatus or other Lactobacillus)
Different Lactobacillus	Both visits ≥75% relative abundance of Lactobacillus with different Lactobacillus dominance at each (either L.iners, L.crispatus or other Lactobacillus)
Remain non-LD	Both visits <75% relative abundance of Lactobacillus
Non-LD to LD	First visit <75% relative abundance of Lactobacillus to second visit ≥75% relative abundance of Lactobacillus
LD to non-LD	First visit ≥75% relative abundance of Lactobacillus to second visit <75% relative abundance of Lactobacillus

4.3.1.5 *The presence/absence and relative abundance of taxa of interest*

Taxa that had previously shown a relationship to PTB, as described in section 4.1, were identified and two variables created: the presence/absence of each taxon and; if present, the relative abundance each taxon.

4.3.2 New approaches

The final three sets of VMB variables are based on previous work of the study team[174–177] and were applied to PTB research for the first time.

4.3.2.1 *Vaginal microbiota types*

These VMB types are similar to Ravel's CSTs[84] but with an increased emphasis on pathobionts. These were based on VMB types described previously,[175] and modified to apply to the pregnant population which is expected to have high relative abundances of *Lactobacillus* (Table 4.6). 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.

Table 4.6: Description of VMB types

VMB type name	Abbreviation	Relative abundance (RA) of *Lactobacillus*	Further details
L. crispatus	Lcr	≥75%	Relative abundance of *L. crispatus* > relative abundance of other *Lactobacillus* species
L. iners	Li	≥75%	Relative abundance of *L. iners* > relative abundance of other *Lactobacillus* species
Other *Lactobacillus*-dominated	Lo	≥75%	Relative abundance of *Lactobacillus* >75% but not fitting Lcr or Li categories (mostly *L. jensenii or L. gasseri* dominated)
Bifidobacterium	BL	<75%	Relative abundance of *Bifidobacterium* > 50%
***Lactobacillus* and anaerobes**	LA	≥25% & <75%	With relative abundance of *Bifidobacterium* < 50%
Bacterial Vaginosis	BV	<25%	With or without *Gardenerella* genus

4.3.2.2 *Bacterial groups*

Each non-minority (present in at least 1 sample with a relative abundance of >1%) ASV in each sample was allocated to one of four 'bacterial groups'[176] based on the published literature (Appendix G): (1) lactobacilli; (2) BV-anaerobes (Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, and Tenericutes except those included in the other 3 groups); (3) pathobionts (most Proteobacteria, and streptococci, staphylococci, enterococci, Spirochaetaceae, *Listeria*, *C. trachomatis*, and *N. gonorrhoeae*); and (4) "other bacteria" (a rest group, containing Actinobacteria that are known to be [facultative] aerobic skin bacteria, *Bifidobacterium* species, and difficult-to-classify minority species). 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, expressed as a percentage within 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.

4.3.2.3 *Bacterial load*

The total bacterial load was estimated by summing the concentration of all ASVs present in each sample.

4.4 Summary

We systematically reviewed the published literature to identify VMB variables that have been associated with PTB previously, and developed three new VMB variables for assessment based on the previous work of our group outside of pregnancy. We will now assess the distribution of these variables in low-risk pregnancies with term birth, and then whether these variables are able to detect characteristics associated with sPTB/PPROM in high-risk pregnancies.

5 Results- Demographic and clinical outcomes

5.1 Recruitment

A total of 227 low-risk women and 137 high-risk women were recruited between 2/6/2015 and 23/11/2017. Two low-risk and three high-risk participants were excluded from the analysis because they did not have a sample that passed the quality control process as described in section 3.8.

5.2 Pregnancy outcomes

Birth outcomes for the low and high-risk populations are shown in Table 5.1 and Table 5.2, respectively. Based on the pregnancy outcomes 145 low-risk and 109 high-risk participants were retained for the primary analysis and a further 4 low-risk and 14 high-risk for supplementary analysis (Figure 5.1).

Eleven high-risk participants (11/137, 8%) were excluded due to medically indicated preterm birth, two of which were prior to 34 weeks gestation (one case of maternal cancer requiring cytotoxic chemotherapy at 32^{+0} weeks gestation and one case of severe maternal anxiety due to previous poor pregnancy outcomes at 33^{+6} weeks gestation). There were no medically indicated preterm births in the low risk group.

Table 5.1: Birth outcomes for low-risk participants

Major classification	Description of birth	Number (n=227)	Percentage
Term	Birth ≥39+0 weeks gestation	145	63.8
Caregiver initiated preterm birth	Caregiver initiated preterm birth	0	0
Early PPROM	PPROM ≤33+6 weeks gestation with early chorioamnionitis	0	0
	PPROM ≤33+6 weeks gestation without early chorioamnionitis	2	0.9
Late PPROM	PPROM 34+0-≤36+6 weeks gestation with early chorioamnionitis	0	0
	PPROM 34+0-≤36+6 weeks gestation without early chorioamnionitis	0	0
Early sPTB	sPTB ≤33+6	0	0
Late sPTB	sPTB 34+0-≤36+6	2	0.9
Unknown	Unable to ascertain birth details	4	1.8
Early term birth in low risk	Birth 37+0-≤38+6 weeks in low risk population without PPROM	72	31.7
No sample	All samples failed quality control	2	0.9

Table 5.2: Birth outcomes for high-risk participants

Major classification	Description of birth	Number (n=137)	Percentage
Term	Birth ≥37+0 weeks gestation	87	63.5
Caregiver initiated preterm birth	Caregiver initiated preterm birth	11	8.0
Early PPROM	PPROM 16+0-33+6 weeks gestation with early chorioamnionitis	2	1.5
	PPROM 16+0-33+6 weeks gestation without early chorioamnionitis	10	7.3
Late PPROM	PPROM 34+0-≤36+6 weeks gestation with early chorioamnionitis	0	0
	PPROM 34+0-≤36+6 weeks gestation without early chorioamnionitis	2	1.5
Early sPTB	sPTB 16+0-33+6	10	7.3
Late sPTB	sPTB 34+0-≤36+6	12	8.8
Unknown	Unable to ascertain birth details	0	0
No sample	All samples failed quality control	3	2.2

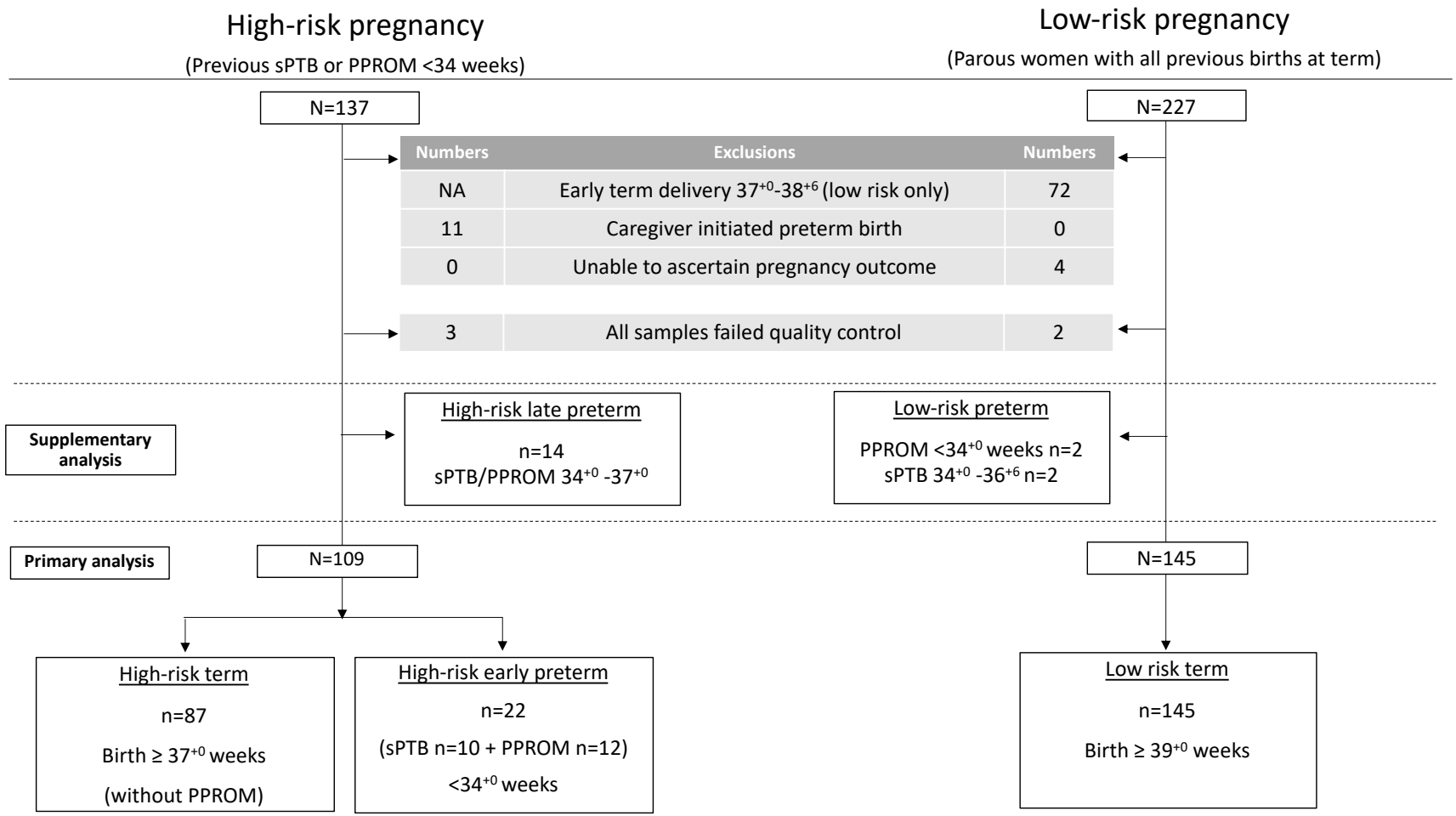


Figure 5.1: Selection of participants to retain for further analysis

5.3 Participant demographics

The participant characteristics of the three groups retained for the primary analysis were broadly similar (Table 5.3), 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 at sampling was slightly later in the low-risk group (16^{+5} weeks) than in the high-risk group (16^{+3} weeks).

Table 5.3: Participant characteristics by pregnancy risk group and outcome for primary analysis

		Low-risk pregnancy (LR)	High-risk pregnancy (HR)		P value LR vs HR	P value HR term vs HR
		Term birth n=145	Term birth n=87	Early sPTB or PPROM n=22		
Age (years)	mean (SD)	31.1 (4.5)	30.0 (4.5)	31.4 (5.4)	0.145	0.211
BMI (kg/m2)*	median (IQR)	24.5 (22-29)	25 (22-28)	27.5 (22-35)	0.619	0.139
Current smoker*	number (%)	15 (10.6)	15 (17.4)	5 (23.8)	0.096	0.537
Ethnicity (%)	White	139 (95.9)	81 (93.1)	22 (100.0)	0.540	0.732
	Black	3 (2.1)	4 (4.6)	0 0.0		
	Asian	0 0.0	0 0.0	0 0.0		
	Other	2 (1.4)	0 0.0	0 0.0		
	Not recorded	1 (0.1)	2 (2.3)	0 0.0		
Parity (%)	0	0 0.00	9 (10.34)	1 (4.55)	0.000	0.017
	1	102 (70.34)	43 (49.43)	10 (45.45)		
	2	36 (24.83)	27 (31.03)	4 (18.18)		
	≥ 3	7 (4.83)	8 (9.20)	7 (31.82)		
Number of previous PPROM or sPTB 16+0-33+6 weeks (%)	1		79 (90.8)	15 (68.2)	na	0.012
	≥ 2		8 (9.2)	7 (31.8)		
Previous cervical surgery	Nil significant	131 (90.3)	79 (90.8)	17 (77.3)	na	0.033
	Single LLETZ	14 (9.7)	7 (8.0)	2 (9.1)		
	Multiple LLETZ or knife cone		1 (1.1)	3 (13.6)		
Gestational age at first valid VMB sample (weeks)	median and IQR	16+5 (16+2-17+1)	16+2 (16+0-16+6)	16+3 (15+6-17+3)	0.000	0.794
Cervical length* (mm) (median and IQR)	16 weeks	41 (36-47.1)	35.5 (31-42)	36 (26.5-43.5)	0.000	0.578
	20 weeks	41 (37-45)	35 (30-40)	30 (26-37)	0.000	0.097
qfFN* (ng/ml) (median and IQR)	16 weeks	7 (5-16)	7 (5-22)	9 (6-23.5)	0.766	0.410
	20 weeks	7 (5-12)	7 (5-14)	7 (6-12)	0.274	0.914
Preterm birth prevention treatment used after study visit	None	139 (100.0)	61 (70.1)	17 (77.3)	na	0.097
	Cervical cerclage		4 (4.6)	2 (9.1)		
	Progesterone		2 (2.3)	1 (4.5)		
	Arabin pessary		20 (23.0)	2 (9.1)		
Antimicrobial in pregnancy prior to sampling (%)	None	115 (79.3)	65 (74.7)	17 (77.3)	0.038	0.895
	Metronidazole or clindamycin	0 0.0	4 (4.6)	1 (4.5)		
	Other antibiotic	21 (14.5)	9 (10.3)	3 (13.6)		
	Clotrimazole	1 (0.7)	0 0.0	0 0.0		
	Not recorded/ unsure	8 (5.5)	9 (10.3)	1 (4.5)		

Table 5.3 Participant characteristics by pregnancy risk group and outcome for primary analysis (continued)

		Low-risk pregnancy		High-risk pregnancy				P value LR vs HR	P value HR term vs HR preterm
		Term birth n=145		Term birth n=87		Early sPTB or PPROM n=22			
Index of multiple deprivation score quintile (%)	1 (most deprived)	74	(51.0)	50	(57.5)	15	(68.2)	0.317	0.830
	2	19	(13.1)	6	(6.9)	2	(9.1)		
	3	24	(16.6)	9	(10.3)	2	(9.1)		
	4	20	(13.8)	11	(12.6)	2	(9.1)		
	5 (least deprived)	7	(4.8)	6	(6.9)	0	0.0		
	Not recorded	1	(0.7)	5	(5.7)	1	(4.5)		
Type of previous PTB (all under 34 weeks) (%)	None	145	(100.0)	0	0.0	0	0.0	na	0.003
	sPTB (≥1)			56	(64.4)	6	(27.3)		
	PPROM (≥1)			28	(32.2)	13	(59.1)		
	≥1 pregnancy with PPROM and ≥1 pregnancy with sPTB			3	(3.4)	3	(13.6)		
Previous late miscarriage (16+0-23+6 weeks) (%)	None	145	(100.0)	69	(79.3)	20	(90.9)	na	0.134
	1			17	(19.5)	1	(4.5)		
	≥2			1	(1.1)	1	(4.5)		
Gestational age at PPROM (weeks+days)	median and range					31+1	(18+0-33+6)	na	na
Gestational age at birth (weeks+days)	median and range	40+1	(39+0-41+6)	38+6	(37+0-41+5)	31+5	(18+0-35+5)		
Birthweight (g)*	mean (SD)	3594	(439)	3234	(489)	1778	(673)		
GROW birthweight centile*	median and IQR	45.0	(23.2-71.9)	35.9	(15.3-50.8)	35.5	(18.5-63.1)	0.102	0.760

Notes about Table 5.3
**BMI has 1 missing value for low-risk and 1 for HR term birth. Smoking has 3 missing values for LR, and 1 each for HR term and early sPTB or PPROM . Cervical length at 16 weeks n=136 LR, n=80 for HR-term, n=20 HR early preterm. CL at 20 weeks , LR n=135, HR term n=79, HR early preterm n=14. qfFN at 16 weeks LR n=142, HR term n=77, HR early preterm n=20. qfFN at 20 weeks LR n=136, HR term n=75, HR early preterm n=19. Birthweight has 2 missing values for HR term and 2 missing values for HR early sPTB or PPROM. GROW birthweight centile[138] LR n=138, HR term birth n=83 and HR early sPTB/PPROM n=19. Parity was recorded as number of previous pregnancies with live births, or stillbirths ≥ 24+0 weeks gestation. Eligibility for the high-risk cohort included ≥1 sPTB or PPROM at 16^{+0} to 33^{+6} weeks gestation. Therefore all high-risk participants with parity=0 had at least one previous late miscarriage. 16 weeks gestation is abbreviated for the first study visit that was carried out at 15^{+1}-18^{+6} weeks and 20 weeks gestation indicates the second study visit that was carried out at 19^{+0}-23^{+0} weeks P values calculated using 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.*

5.4 Birthweight

The GROW birthweight centile[138] was calculated to facilitate comparison of birthweights across the groups despite differences in gestation at birth (Table 5.3 and Figure 5.2). High-risk early sPTB/PPROM and high-risk term birth babies had a similar distribution of GROW birthweight centiles. There was a non-significant trend towards a lower GROW birthweight centile in the high-risk pregnancies than the low-risk pregnancies (p=0.102, Table 5.3).

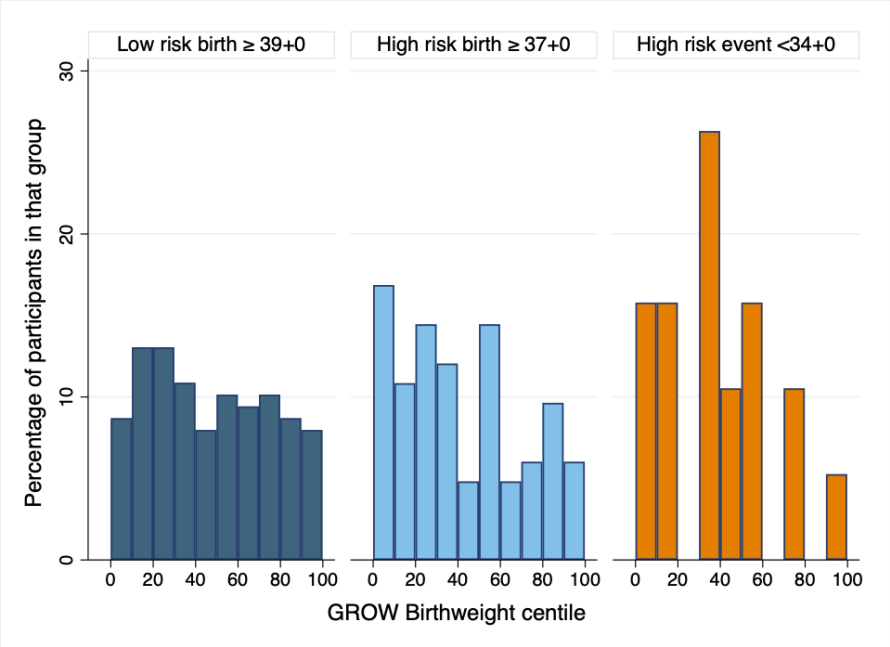


Figure 5.2: Birthweight centile by pregnancy outcome. Women were only included in this analysis if all components of the GROW calculation[138] were available, as such low risk n=138, high-risk term birth n=83 and high risk early sPTB/PPROM n=19.

5.5 Discussion about demographics

5.5.1 Ethnicity

Over 90% of participants in each group were white ethnicity. The 2011 UK census showed 88.9% of Liverpool city residents to be of white ethnicity, so this is broadly in keeping with the local population.[178] The vaginal microbiome has been reported as differing by ethncity,[97] and analysis stratified by ethnicity has been advocated[99] and performed.[100]

However, with so few Black, Asian and minority Ethnic (BAME) participants this analysis was not possible in our population. The study team considered restricting the analysis to participants of white ethnicity, however any positive findings would be expected to be applied to the population as a whole and was judged inappropriate. As such all participants were included in the analysis.

5.5.2 Parity and number of previous PPROM or sPTB

Parity is reported because it is recommended by the PREBIC consortium,[99,159] however it is a little misleading because in the high-risk population all participants who were recorded as nulliparous had had a previous pregnancy loss after 16^{+0} weeks gestation. Therefore the number of previous sPTB/PPROM pregnancies is also given (Table 5.3).

High risk participants with a recurrent sPTB/PPROM were more likely to have: a parity ≥3; ≥2 previous sPTB/PPROM; and a history of both sPTB and PPROM (p=0.017, p=0.012 and p=0.003 respectively). Having had more previous preterm births is a recognised risk factor for preterm birth.[179] Whether having had more previous sPTB/PPROM alters the VMB is unknown, but differences in the VMB during the first trimester by pregnancy history have been previously described,[128] so this is possible.

5.5.3 Gestation of sampling

The low-risk group had a slightly later median gestation of first valid VMB sample than either of the high-risk groups (low-risk group 16^{+5} weeks, IQR 16^{+2}-17^{+1}; high-risk term birth 16^{+2} weeks, IQR16^{+0}-16^{+6}; high-risk early sPTB/PPROM group 16^{+3} weeks, IQR 15^{+6}-17^{+3} weeks, p value for difference between high and low risk groups <0.000). This is likely to be because the high-risk cohort were offered recruitment at the preterm birth prevention clinic, which although targeted at 16 weeks gestation was often scheduled to fit with patient and clinic availability and in practice recruitment of some women was performed

during 15th gestational week. Conversely low-risk participants were offered recruitment by telephone and a specific study visit was arranged. This was arranged during the 16th or 17th gestational week. The vaginal microbiota is recognised to alter as pregnancy progresses.[97,173] However, the median gestational age differences were only 2 and 3 days respectively between the low-risk and high-risk term/high-risk early sPTB/PPROM groups. Such a short difference in gestational age was felt to be unlikely to have a material effect on the VMB, and so no adjustment for gestational age at sampling was performed.

5.5.4 Cervical length and quantitative fetal fibronectin

These were primarily measured for other components of the 'Biomarkers of Preterm Birth' study,[180] and reported within this work as recommended by the PREBIC consortium.[99,159] The cervical length of high-risk participants at both 16 and 20 weeks gestation was shorter than the low-risk group, as expected.[181] The median cervical length at approximately 20 weeks gestation (19^{+0}-23^{+0} weeks) was 30mm in the high-risk early sPTB/PPROM group and 35mm in the high-risk term birth group. Importantly this study excluded women who were using PTB prevention treatment at the time of their first visit/valid sample, in order to exclude alterations in the VMB that could be due to the PTB prevention treatment. PTB prevention treatment is normally offered to women with a short cervical length, as per UK standard practice.[26] Therefore women with short cervical length are likely to be under-represented in this study.

5.5.5 Preterm birth prevention treatment used after the study visit

The use of preterm birth prevention treatment (cervical cerclage, vaginal progesterone or Arabin pessary) after the study visit was broadly similar between the high-risk term (29.9%, 21/87) and high-risk early sPTB/PPROM groups (22.7%, 5/22) (Table 5.3). We chose to retain women who used PTB prevention treatment after their vaginal sample collection in our analysis. This approach has recently been advised against[182] as it is possible that women

destined to deliver preterm may have had their delivery delayed by the therapy, weakening the ability of our analysis to detect a relationship between the VMB and sPTB/PPROM. However, any future therapy developed based upon VMB analysis would likely be applied in combination with current treatments, and so by retaining these participants we were able to assess the contribution of the VMB to recurrent sPTB/PPROM within current clinical practice.

5.5.6 Antimicrobial therapy in pregnancy prior to sampling

High-risk participants were slightly more likely to have used an antimicrobial in the pregnancy prior to sampling (24.8%, 82/109) compared to the low-risk group (21.7%, 40/145), p=0.038. However, there were only two participants (both high-risk) in whom the antibiotic was metronidazole or clindamycin; the antibiotics with the most evidence to support their use for the treatment of bacterial vaginosis.[44] Supplementary analysis was performed to account for use of antimicrobials when a positive association was found between a VMB parameter and PTB.

5.5.7 Index of multiple deprivation

Liverpool Women's Hospital is situated in the 4th most deprived local authority area in England (out of 343).[140] A woman's lived environment, including her diet, is likely to influence her vaginal microbiota.[89] Given the differing recruitment processes for the low and high risk groups, including the need for extra hospital visits in the low risk group, we were concerned that participants with lower IMD scores may have been under-represented in the low-risk group. Reassuringly the IMD scores did not show a statistically significant difference between participants in each pregnancy outcome group (Table 5.3).

5.5.8 Birthweight

The median GROW[138] birthweight centile for the high-risk cohort who delivered at term was not statistically different to those who had an early sPTB/PPROM at 35.5 (IQR 18.5-63.1) and 35.9 (IQR 15.3-50.8) respectively (p=0.760) (Table 5.3). When comparing the low and high risk women the median birthweight centiles were 45.0 (IQR 23.2-71.9) and 35.5 (IQR18.2-58.6) respectively. This difference was not statistically significant (p= 0.102) (Figure 5.2). The difference between the high and low risk groups is in keeping with previous literature showing common risk factors for preterm birth and growth restriction.[121] Conversely in our high-risk group recurrence of early sPTB/PPROM did not seem to be associated with growth restriction, potentially suggesting that risk factors for recurrence are different for risk factors for growth restriction.

5.5.9 Selection of covariates for multiple regression analysis

The proposed multiple regression analysis was designed to understand the contribution of the VMB characteristics to recurrent early sPTB/PPROM, this is referred to as aetiological modelling.[183] This contrasts to predictive modelling in which the aim is to design the best model to detect cases, irrespective of whether the variables are causative of the event.

In order to optimise aetiological modelling the aim is to control for covariates which may have an effect upon both the variable (VMB) and the outcome (early sPTB/PPROM). Importantly the variable should not have an effect on the covariate. In aetiological modelling the covariates should be determined based on existing knowledge, rather than the data. It is irrelevant whether there is a direct relationship between the covariate and the outcome, although of course the covariate does need to have been reliably recorded.

The study team considered including history of ≥2 previous sPTB/PPROM as a covariate within the logistic regression models, however it is uncertain how and why the number of

previous sPTB/PPROM would affect the VMB. Using obstetric history within aetiological modelling has previously been discussed as statistically problematic.[184] Therefore obstetric history was not selected as a covariate.

Cervical length may have an effect on the VMB, or conversely the VMB may affect the cervical length, so this was also considered an inappropriate co-variate. However, previous cervical surgery was deemed to be an appropriate covariate, because it may have an effect on the VMB, but the present VMB cannot have an effect on whether or not previous cervical surgery has happened,[185] and previous cervical surgery has been comprehensively been shown to have an effect on risk of preterm birth.[186]

Body mass index (BMI) was selected as a covariate because it may have an association with the VMB,[187,188] and does have an effect on the risk of preterm birth.[189,190] BMI was converted to a quadratic term due to the bimodal association with PTB.

The final covariate selected was smoking at the time of enrolment. Smoking has been suggested to have an effect on the VMB outside of pregnancy,[191] and is associated with preterm birth.[192]

Given the small number of participants with our primary outcome (sPTB/PPROM <34^{+0} weeks gestation), and therefore the limited number of covariates we could feasibly control for, the study team decided that three covariates was the maximum appropriate. Each of the selected co-variates had two or less missing values for the high-risk participants (Table 5.3).

5.6 Contributory factors to preterm birth

5.6.1 Allocation of contributory factors to PTB

All participants who had PPROM or sPTB <37^{+0} weeks gestation had a review of the clinical notes to ascertain whether there were clinically evident factors that may have contributed to the event. 54.2% of participants who had early sPTB/PPROM at 16^{+0}-33^{+6} weeks (Table 5.4) and 50% of participants who had late sPTB/PPROM (Table 5.5) had at least one contributory factor identified. The most common contributory factors to sPTB/PPROM <34^{+0} weeks were chorioamnionitis and cervical insufficiency.

In the low-risk population there were two cases of PPROM without chorioamnionitis <34^{+0} weeks gestation. In both cases there was a placental contribution to birth. Of the remaining two low-risk women who delivered preterm there was one late spontaneous PTB with a uterine abnormality, and a further late sPTB with no contributing factors.

Table 5.4: Details of factors that were identified as contributing to the event for women with PPROM or sPTB 16^{+0}-33^{+6} weeks gestation. Note contributing factors are not mutually exclusive, so total number of contributing factors are more than total number of pregnancies in each pregnancy outcome group.

Contributing factors at birth	**High-risk participants** PPROM 16^{+0}-33^{+6} weeks With early (within 7 days) chorio-amnionitis n=2	Without early chorio-amnionitis n=10	sPTB 16^{+0}-33^{+6} weeks n=10	**Low-risk participants** PPROM 16^{+0}-33^{+6} weeks without early chorio-amnionitis n=2	**Total n=24**
None	0	6 (60%)	5 (50%)	0	11(45.8%)
Chorioamnionitis	2 (100%)	2 (20%)	1 (10%)	0	5 (20.8%)
Placental dysfunction	0	0	2 (20%)	2 (100%)	4 (16.7%)
Extra amniotic infection	0	0	0	0	0
Polyhydramnios	0	0	1 (10%)	0	1 (4.2%)
Uterine anomaly	0	1 (10%)	0	0	1 (4.2%)
Maternal comorbidities	0	1 (10%)	1 (10%)	0	2 (8.3%)
Cervical insufficiency	0	3 (30%)	2 (20%)	0	5 (20.8%)
Multiple contributing factors	0	2 (20%)	1 (10%)	0	3 (12.5%)

Table 5.5:Details of factors that were identified as contributing to the event for women with PPROM or sPTB 34+0-36+6 weeks gestation.

Contributing factors at birth	High-risk participants		Low-risk participants	Total n=16
	PPROM 34^{+0}-36^{+6} weeks without early chorioamnionitis n=2	sPTB 34^{+0}-36^{+6} weeks n=12	sPTB 34^{+0}-36^{+6} weeks n=2	
None	1 (50%)	6 (50%)	1 (50%)	8 (50%)
Chorioamnionitis	0	0	0	0
Placental dysfunction	0	1 (8.3%)	0	1 (6.3%)
Extra amniotic infection	0	0	0	0
Polyhydramnios	0	0	0	0
Uterine anomaly	0	0	1 (50%)	1 (6.3%)
Maternal comorbidities	0	0	0	0
Cervical insufficiency	1 (50%)	5 (41.7%)	0	6 (37.5%)
Multiple contributing factors	0	0	0	0

Note: There were no cases of late PPROM with chorioamnionitis within 7 days.

5.7 Discussion of contributory factors to preterm birth

We used a modified version of the preterm birth classification system proposed by Villar *et al* in 2012[141] to allocate contributory factors to the preterm birth cases. Villar *et al* developed the system for use in the INTERGROWTH-21st project to identify phenotypes of preterm birth across 8 countries between 2009 and 2014. The system was applied to 5828 births under 37^{+0} weeks and was able to allocate a phenotype in 70% of births,[193] slightly more than the 52.5% allocation in our study. The INTERGROWTH project included multiple pregnancies (which accounted for 10.4% of their preterm births) and medically indicated preterm births (which, by definition identify a contributory factor/phenotype). This is likely

to be why they were able to identify a contributory factor in more cases than we were. Infection, approximately equally split between chorioamnionitis and extrauterine infection, accounted for approximately 15% of INTERGROWTH preterm births, similar to the 12.5% (5/40) of all births under 37^{+0} weeks with chorioamnionitis as a contributing factor in the current study. Of note in the current study all of these were births under 34^{+0} weeks gestation, accounting for 20.8% of births <34^{+0} weeks. The INTERGROWTH project did not have access to cervical assessment data, so we are not able to compare this.

A modified version of the Villar *et al*[141] classification system has also been applied to preterm births in Ontario, Canada between 2012 and 2014.[194] When isolated to spontaneous recurrent preterm births (n=1274) the authors identified phenotypes (comparable to our contributory factors) in 32% of preterm births (under 37^{+0} weeks gestation). This study did not identify any cases of chorioamnionitis as contributing to preterm birth but did attribute extra-amniotic infection to 18.4% (235/1274) of spontaneous recurrent preterm births. The current study had a wider definition of chorioamnionitis, and so these are likely to be different labels for a similar pathology. The Villar *et al*[141] study also did not have access to cervical length assessments.

An alternative phenotype allocation system was developed by Manuck *et* al to account for the likelihood that a particular factor contributed to the PTB (graded as strong/moderate/possible evidence).[195] This was applied to 1025 births (of both high and low-risk women) under 34^{+0} weeks gestation in the United States of America and described possible cervical insufficiency in 11.6% of cases.[195] Given that cervical insufficiency is a risk factor for recurrent preterm birth the higher rate in our population (5/22, 20.8%) is unsurprising. Manuck *et* al described possible and strong evidence of infection in 18% of births, and possible evidence of infection in a further 27% of births. The differential grading makes further comparison difficult, but in summary there appears to be some evidence of

infection contributing to preterm birth in between 10 and 40% of cases, and the rate appears higher in the earlier preterm births, as summarised by Lamont in 2015.[196]

5.8 Conclusion of recruitment, pregnancy outcomes and demographics

This study was able to meet the recruitment targets set within the initial funding bid.[197] There were 22 cases of recurrent early sPTB/PPROM available for analysis after recruitment of 137 high-risk women, along with 87 cases of high-risk term birth and 145 cases of low-risk term birth. 14 cases of late sPTB/PPROM were available for supplementary analysis.

The research team selected the covariates smoking, previous cervical surgery and BMI to be used in the multiple regression aetiological modelling based on previous literature. These were well recorded, with two or less missing values in the high-risk group for each covariate.

Eleven high-risk participants were excluded from the analysis due to medically indicated late preterm birth. Without medical intervention these pregnancies may have progressed to term births, which potentially could have changed the composition of high-risk term birth group slightly. However previous early sPTB/PPROM is a risk factor for common obstetric complications such as growth restriction,[198] indeed in this study babies of the high-risk participants had approximately a 10% lower birthweight centile than the low-risk participants. The poor obstetric history inevitably increases maternal anxiety, and often clinician anxiety too, increasing the chance of intervention, therefore this attrition is likely to be inevitable. The high rate of medically indicated late preterm births gives further justification for the decision to focus the primary analysis on early sPTB/PPROM; to both focus on the pregnancies with the worst morbidity and mortality and obtain a cleaner phenotype of sPTB/PPROM. Infection appears to play a stronger role in early sPTB/PPROM

than late sPTB/PPROM, and so it is plausible that the VMB may have a stronger link to the earlier events too.

The research team attempted to attribute contributory factors to the cases of preterm birth. This was not possible in approximately half of both early and late sPTB/PPROM cases. When it was possible the most common contributory factors identified were cervical insufficiency and chorioamnionitis. The higher rate of cervical insufficiency in this group with recurrent sPTB/PPROM is unsurprising compared to previous literature which has used mixed populations with primary and secondary sPTB/PPROM cases, often without knowledge of the cervical length in pregnancy. The other common contributory factor identified was chorioamnionitis. The vaginal microbiota is often thought to be the source of chorioamnionitis,[196] further justifying its examination in this thesis.

6 Results- vaginal microbiota in low-risk women

6.1 Overview

This group was designed to describe the expected distribution of VMB characteristics in our local population in healthy, multiparous women who gave births at ≥ 39 weeks gestation. 145 low-risk women were eligible for inclusion in the low-risk term birth group. 136 participants had a VMB sample with a valid result at approximately 16 weeks gestation (15^{+1}-18^{+6} weeks) and 129 participants had a VMB sample with a valid result at approximately 20 weeks gestation (19^{+0}-23^{+0} weeks). For the majority of the analysis the first available sample was used, with the exception of two parts. Heatmaps were produced for samples at each visit to visualise the distribution of the VMB at each visit, and the stability analysis was performed on the subset of participants with valid results at each visit.

6.2 Heatmaps

The four most common taxa at both study visits were the *Lactobacillus* species of *L. iners, L. crispatus, L. gasseri* and *L. jenseni.* As described by Ravel *et al*[95] the distribution of VMB characteristics broadly fell into 5 groups according to dominant species of lactobacilli, and so these were used to organise heatmaps (Figure 6.1 and Figure 6.2).

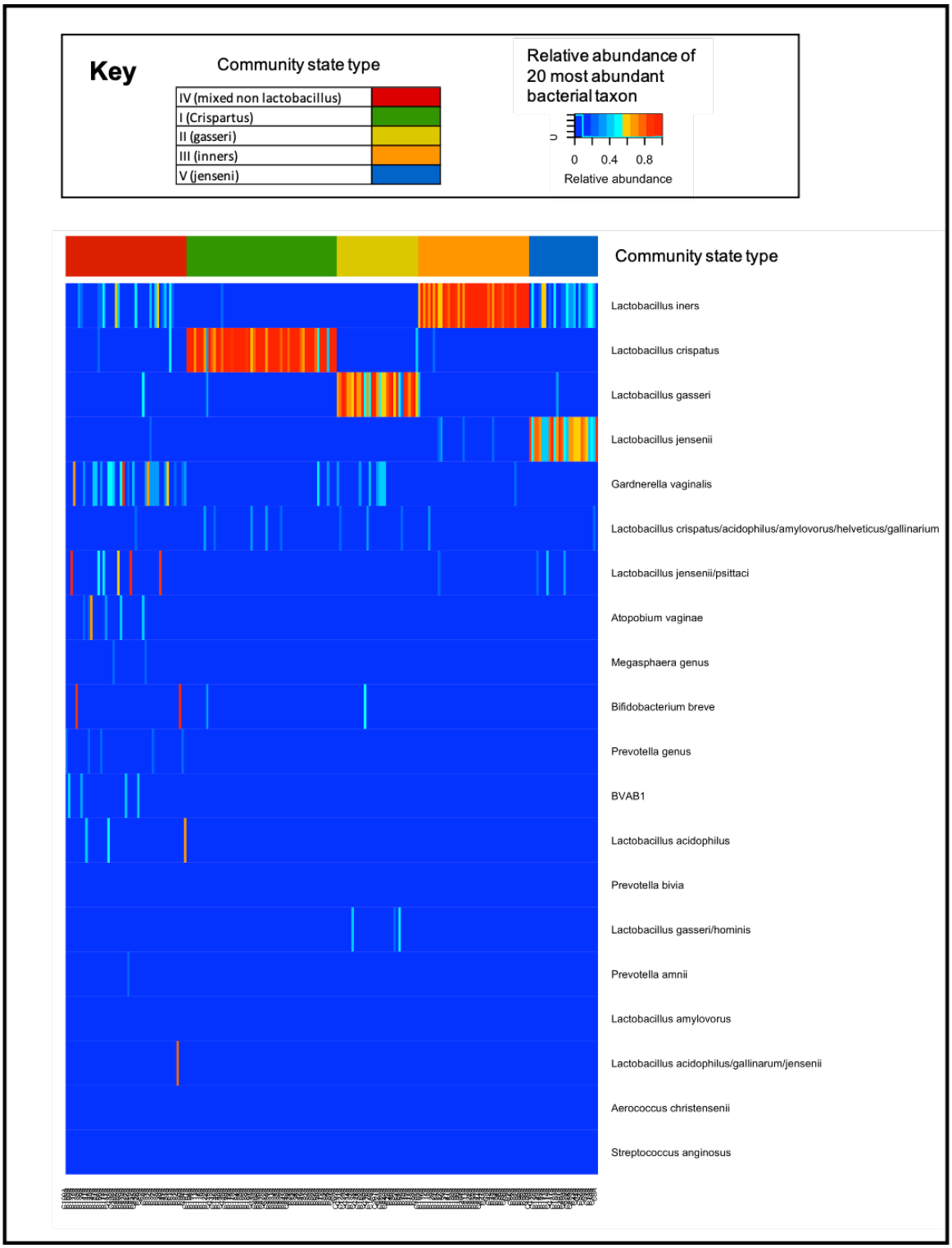


Figure 6.1: Heatmap of 20 most abundant taxa in low-risk term birth group at 16 weeks gestation, n=136.

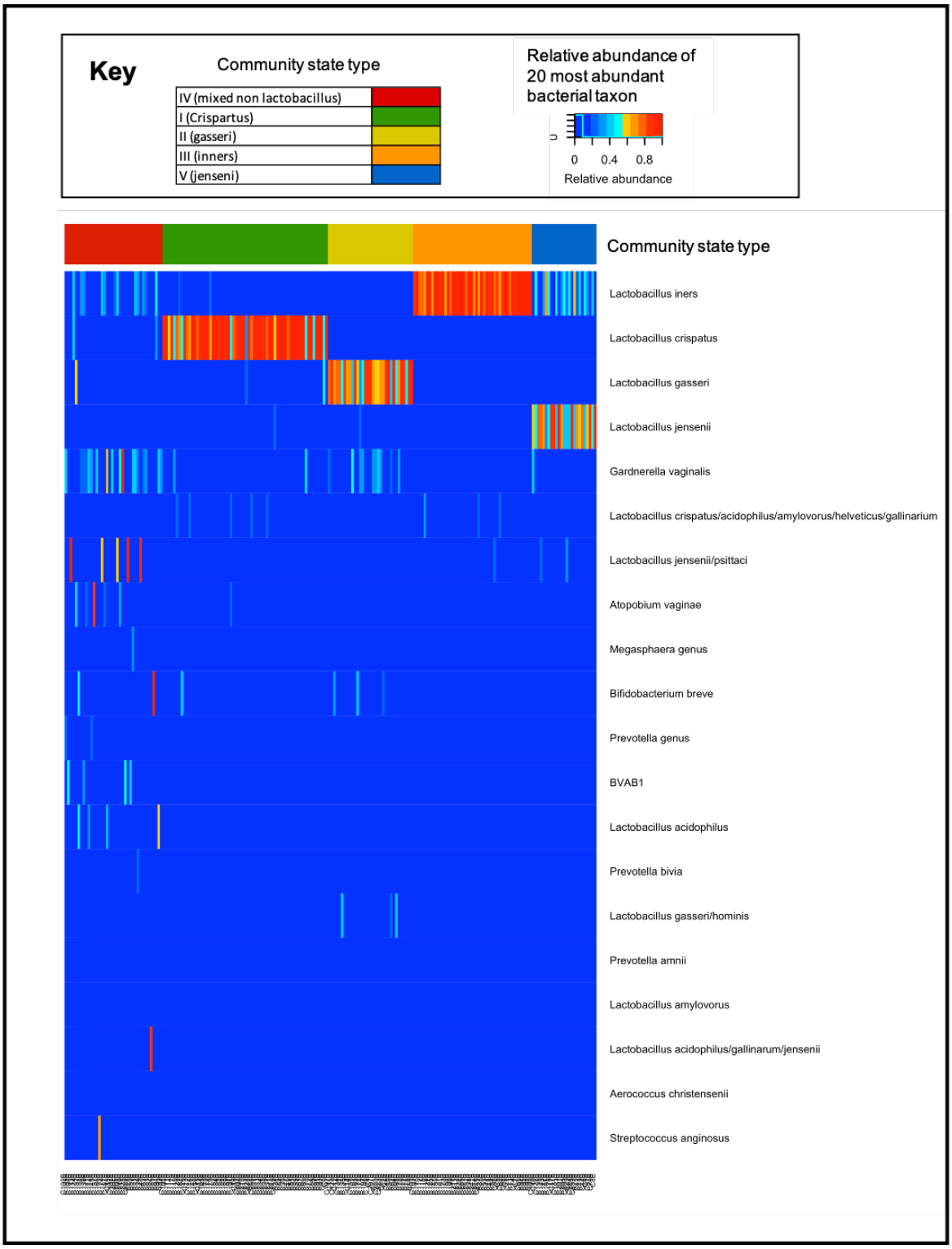


Figure 6.2: Heatmap of 20 most abundant taxa in low-risk term birth group at 20 weeks gestation, n=129

6.3 Richness

The median richness was 7, IQR 4-14, with a left skew to the data as shown in Figure 6.3.

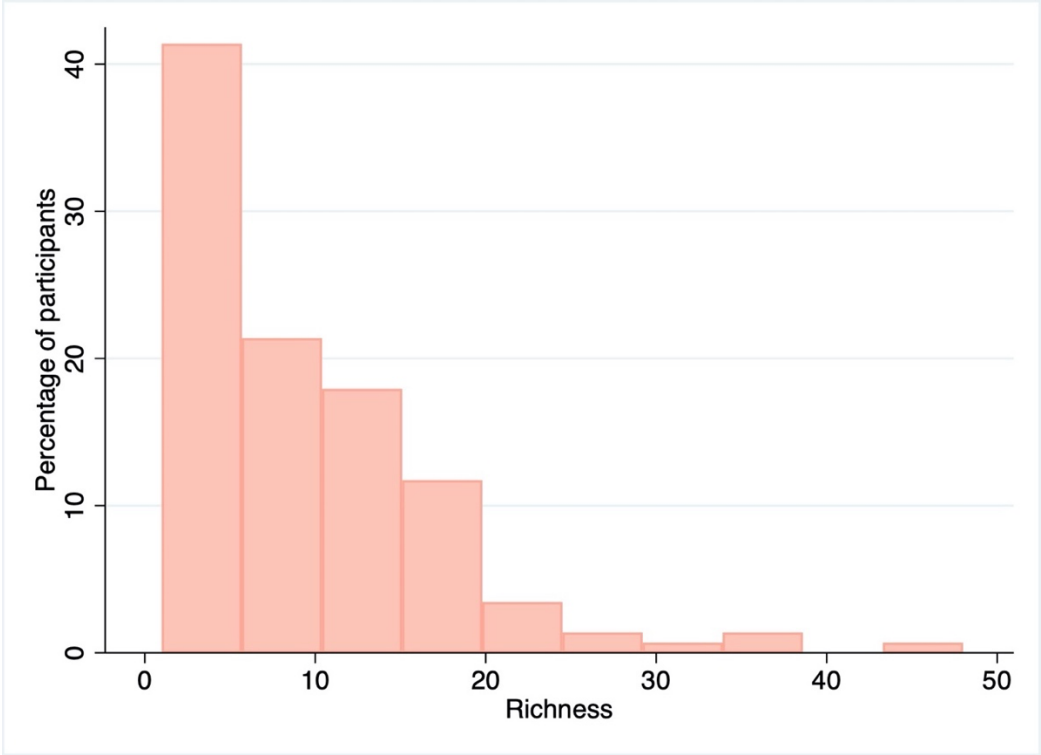


Figure 6.3: Richness in low-risk term birth group at 16 weeks gestation, n=145

6.4 Simpson diversity index

The median Simpson diversity index was 0.38, IQR 0.07-0.53. The distribution showed a left skew, as displayed in Figure 6.4.

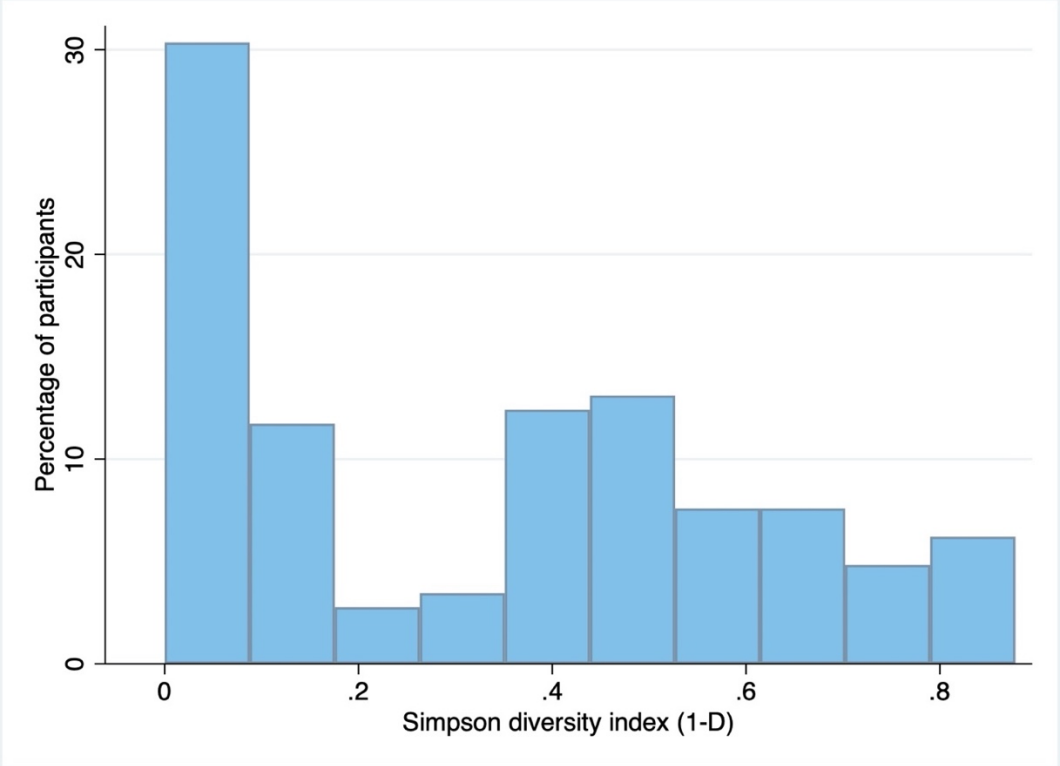


Figure 6.4: Simpson Diversity index (1-D) in the low-risk term birth group, n=145

6.5 Community State Types

The most common CST within the LR term birth group was I (*L. crispatus* dominance), corresponding to 29.7% (43/145) of participants. The distribution of all CSTs is shown in Table 6.1.

Table 6.1: Distribution of CSTs in low-risk term birth group n=145

Community State Type	n	(%)
I *(L. crispatus)*	43	(29.7)
II *(L. gasseri)*	16	(11.0)
III *(L. iners)*	31	(21.4)
IV (mixed non Lactos)	35	(24.1)
V *(L. jensenii)*	20	(13.8)
Total	145	(100)

6.6 Group according to *Lactobacillus* relative abundance

The majority of participants (72.5%, 105/145) had a relative abundance of *Lactobacillus* over 75% (*Lactobacillus* dominance group). The distribution of all participants is shown in Table 6.2.

Table 6.2: Distribution of group according to Lactobacillus relative abundance in low-risk term birth group, n=145

Group according to *Lactobacillus* relative abundance	n	(%)
Lactobacillus dominance	105	(72.4)
Intermediate	14	(9.7)
Lactobacillus deplete	26	(17.9)
Total	145	(100.0)

6.7 Stability groups

The majority of participants remained dominated by the same *Lactobacillus* species, or remained with a non- *Lactobacillus* dominant microbiota (Table 6.3). Only 5.4% (7/129) swapped from a state of *Lactobacillus* dominance to not, or vice versa.

Table 6.3: VMB stability groups in the subset of low-risk term birth participants who had a VMB sample available for both study visits

Change in VMB between 16 and 20 weeks gestation (stability)	n	(%)
Same *Lactobacillus*	83	(64.3)
Different *Lactobacillus*	3	(2.3)
Remain non-LD	29	(22.5)
Non-LD to LD	7	(5.4)
LD to non-LD	7	(5.4)
Total	129	(100)

n=129. LD, Lactobacillus *Dominant*

6.8 Vaginal microbiota types

The most common VMB type in the low-risk term birth group was Lcr (*L. crispatus* dominance), 26.9% of participants, closely followed by Lo (other *Lactobacillus* species) and Li (*L. iners* dominance). The VMB type 'BL' was notably rare, with only 2.8% (4/145) participants meeting this criterion. Bifidobacterium are proposed to be either protective of preterm birth,[106] or at least not associated with an increased risk of PTB (Table 4.3). Therefore, a decision was made to combine the VMB types of BL (*Bifidobacterium* dominated) and Lo (other lactobacillus) for the purposes of analysis of the preterm births.

Table 6.4: VMB type in low-risk term birth participants, n=145.

VMB type	n	(%)
Lcr	39	(26.9)
Li	32	(22.1)
Lo	33	(22.8)
BL	4	(2.8)
LA	18	(12.4)
BV	19	(13.1)
Total	145	(100.0)

VMB type was defined based on relative abundance of applicable taxa as follows: (1) L. crispatus (Lcr; ≥75% lactobacilli of which L. crispatus was the most common); (2) L. iners-dominated (Li; >75% lactobacilli of which L. iners was the most common; (3) other lactobacilli-dominated (Lo; ≥75% lactobacilli of which L. jensenii or L. gasseri were the most common; (4) Bifidobacterium dominated (Bifidobacterium ≥50%); (5) lactobacilli, and anaerobes (LA; 25%- 75% lactobacilli); and (6) polybacterial G. vaginalis-containing (BV_GV; <25% lactobacilli and <50% Bifidobacterium).

6.9 Taxa of interest

We assessed the presence/absence, and if present the relative abundance of the 4 species of *Lactobacillus* and 19 taxa of interest identified in Table 4.3. All species of *Lactobacillus* and 18 of the taxa of interest were identified in at least one participant. *Mageeibacillus indolicus* was not identified in any participants and so excluded from further analysis. The percentage of participants with each taxon present, and median relative abundance (with IQR) is displayed in Table 6.5.

Table 6.5: Percentage of low-risk term birth participants with taxa of interest present, and when present median and IQR of relative abundance (%)

		Number of samples with taxon	% of samples with taxon	Median (%)	IQR
Relative abundance of species of *Lactobacillus* (%)	*Lactobacillus crispatus*	125	(86.21)	3.45	(0.82-94.19)
	Lactobacillus gasseri	47	(32.41)	20.44	(0.18-58.13)
	Lactobacillus iners	82	(56.55)	38.87	(5.09-86.28)
	Lactobacillus jensenii	53	(36.55)	21.16	(1.81-69.57)
Relative abundance of taxon of interest (%)	*Aerococcus*	24	(16.55)	0.82	(0.27-3.27)
	Atopobium vaginae	44	(30.34)	3.04	(0.27-14.17)
	Bifidobacterium breve	14	(9.66)	1.45	(0.27-31.70)
	Clostridiales BVAB2	11	(7.59)	2.36	(0.54-6.12)
	Dialister	34	(23.45)	0.27	(<0.1-0.72)
	Gardnerella vaginalis	57	(39.31)	15.44	(0.91-33.51)
	Lachnospiracea BVAB1	3	(2.07)	0.18	(<0.1-32.97)
	Megasphaera	21	(14.48)	9.90	(6.09-12.81)
	Mobiluncus	4	(2.76)	0.18	(<0.1-0.50)
	Mycoplasma	2	(1.38)	0.91	(0.82-1.00)
	Parvimonas	15	(10.34)	0.63	(0.18-2.36)
	Peptoniphilus	26	(17.93)	0.18	(0.18-0.45)
	Prevotella	57	(39.31)	0.91	(0.27-6.99)
	Porphyromonas species	7	(4.83)	0.18	(<0.1-0.99)
	Sneathia species	14	(9.66)	4.77	(1.18-10.90)
	Streptococcus species	19	(13.10)	0.27	(<0.1-10.8)
	BVAB TM7-H1	2	(1.38)	7.20	(2.0-12.3)
	Ureaplasma species	26	(17.93)	0.22	(0.18-0.45)

6.10 Bacterial groups

All taxa present in the ASV table (section 3.7.5) were allocated to one of four bacterial groups (Appendix G) as follows: (1) lactobacilli; (2) BV-anaerobes (Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, and Tenericutes except those included in the other 3 groups); (3) pathobionts (most Proteobacteria, and streptococci, staphylococci, enterococci, Spirochaetaceae, *Listeria*, *C. trachomatis*, and *N. gonorrhoeae*); and (4) "other

bacteria" (a rest group, containing Actinobacteria that are known to be [facultative] aerobic skin bacteria, *Bifidobacterium* species, and difficult-to-classify minority species). The presence/absence and if present relative abundance and concentration of taxa in each bacterial group is shown in Table 6.6

Table 6.6: Presence/ absence and if present relative abundance and estimated concentration of taxa in each bacterial group in low-risk women with term births, n=145

Bacterial group	Presence/absence of bacterial groups		Relative abundance of bacterial groups (%)		Bacterial group concentration in log^{10} cells/µl	
	n	(%)	Median	IQR	Median	IQR
Total Lactobacillus	143	(98.62)	98.64	(71.75-99.90)	7.31	(6.46-8.03)
Total BV associated	109	(75.17)	3.00	(0.54-45.32)	6.40	(4.85-7.58)
Total pathobionts	34	(23.45)	0.18	(<0.1-0.82)	4.33	(3.63-4.99)
Total other bacteria	50	(34.48)	0.27	(0.18-0.64)	4.72	(4.14-5.63)

6.11 Total bacterial load

The total bacterial load was estimated by summing the estimated concentration of all taxa in each sample. The median bacterial load was 7.68 log_{10} cells/µL, IQR 6.80-8.35 log_{10} cells/µL. The distribution is displayed in Figure 6.5.

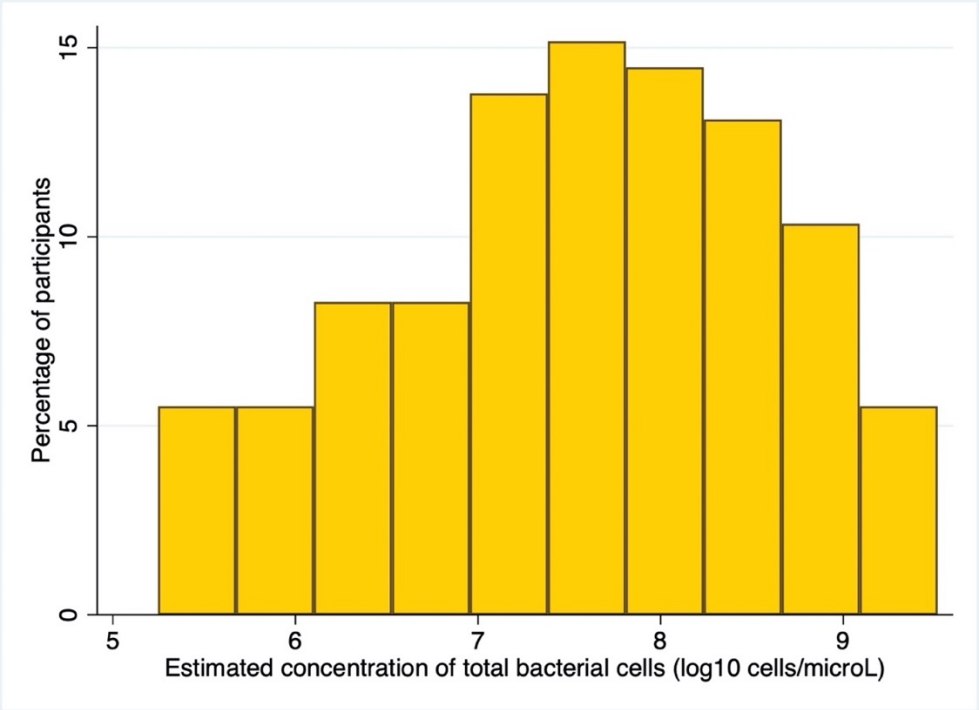


Figure 6.5: : Distribution of total bacterial load in low-risk women with term births, n=145

6.12 Discussion of vaginal microbiota in low-risk term group

The VMB of the low-risk term birth group in the second trimester was predominantly comprised of Lactobacillus. Over 98% of participants had at least one species of lactobacilli present, and 72% of participants met our definition of *Lactobacillus* dominance by having over 75% relative abundance of *Lactobacillus*. This is consistent with findings within the NIH microbiome project that found 78% of women of European ancestry had *Lactobacillus* dominance in pregnancies that progress to term births in the USA (n=40), at a similar gestation to our cohort.[97] Analysis of the metadata available online for the largest group of pregnancies that have undergone VMB assessment in women that had term births shows a very similar picture, with 76% of white women having had a VMB dominated by *Lactobacillus* at 16-20 weeks gestation (68/90).[101] This contrasts to 51% (161/313) of VMB assessments at 16-20 weeks having *Lactobacillus* dominance in women of black ethnicity in

the same study. This is consistent with the previously recorded differences in VMB composition by ethnicity.[97]

The VMB of the majority of our low-risk participants had relatively low species richness, with a median of only 7 species observed (IQR 4-14), although there was a left skew to this data with a maximum richness of 48 species. This is consistent with the original study assessing the VMB in uncomplicated pregnancies with term births, in which analysis of metadata available online shows a median richness of 9 species, IQR 7-14 (n=22).[95] This is markedly less than the median richness of non-pregnant women of reproductive age, which was 19, IQR 10-30 (n=32) in the same study, and consistent with later studies that have confirmed a lower richness in pregnancy compared to non-pregnant women.[199] Comparisons of species richness between VMB studies are particularly vulnerable to be differences inherent in the DNA extraction, sequencing techniques and downstream processing of each study.[107] This is because an extraction technique may not lyse certain types of cells, a primer set may preferentially amplify certain types of bacteria more than others, and each study will choose to rarefy at a different depth. However, those that have measured richness in a similar manner to the present study have found that participants who had a term birth have similar median richness to our study of approximately 4.5,[102] 8,[103] and 12[166] species per sample, suggesting our findings are broadly in keeping with the published literature.

We chose to express the Simpson diversity index as the reciprocal (1-D) whereas other recent studies have expressed this variable as an inverse (1/D). By converting the reciprocal to the inverse (inverse Simpson=1/(1-D)) it is possible to make comparisons with Brown *et al*[103] and Fettweis *et al.*[100] Our median Simpson diversity index (1-D) was 0.38, IQR 0.07-0.53. This is equivalent to inverse Simpson index (1/D) of median 1.61, IQR 1.07-2.13. This is very similar to the median inverse Simpson diversity at 18^{+0}-23^{+6} weeks of approximately

1.5 in pregnancies to progressed to term birth as sampled by Brown *et al*[103] and approximately 1.8 in in pregnancies to progressed to term birth as sampled by Fettweis *et al.*[100]

The most common VMB types and CSTs in the low-risk term birth participants were those dominated by *L. crispartus,* closely followed by the CSTs and VMB types with domination of *L. iners* and other species of lactobacillus. These findings are consistent with other VMB assessments in the second trimester of pregnancies that have progressed to term in Caucasian populations.[97,101,106] In keeping with the differences in *Lactobacillus* dominance, previous work has also noted differences in the normal distribution of CSTs by ethnicity,[96,97,101] and indeed the distribution of *Lactobacillus* dominance and CSTs in the current study was closer to that of previous studies in Caucasian populations[106] than in women of African American ethnicity.[95]

The VMB was notably stable for the majority of low-risk term birth participants who had both study visits available for analysis. Between approximately 16 and 20 weeks gestation 64.3% (83/129) of participants remained dominated by the same lactobacilli, and a further 22.5% (29/129) remained in a state of non-lactobacilli dominance. Only 2.3% (3/129) of participants became dominated by a different *Lactobacillus* and 7 (5.4%) of participants switched from lactobacilli dominance to non-lactobacilli dominance, or vice versa (Table 6.3). This stability is also highlighted visually by the very similar heatmaps of the two visits (Figure 6.1 and Figure 6.2). The stability of the VMB in the second trimester is consistent with previous studies.[95,97]

The stability of the VMB between the 16 and 20 week sampling has three benefits. Firstly, it suggests good reproducibility of our technique used to assess the VMB. The current study only used a single sample and VMB analysis method to assess the VMB at each time point, and early work into the VMB had raised concerns about poor reproducibility within VMB

assessment.[200] Indeed, optimal assessment of the VMB would utilise multiple samples and sequencing techniques.[110] However, this is rarely achieved in practice due to pressures of laboratory time and costs. The high VMB stability in this study indirectly suggests that there is good reproducibility within the current work, increasing the confidence in the accuracy of the results.

Secondly the relative stability between the 16 and 20 week assessments suggests that it is reasonable to combine the VMB assessments at each time point. This increases the number of early sPTB/PPROM participants available for analysis from 17 to 22, increasing our statistical power. Thirdly the relative stability of the VMB between approximately 16 and 20 weeks gestation suggests that, if there is a VMB characteristic associated with early sPTB/PPROM, there is potentially a window of time during which therapy could be given to modify the VMB, aimed at improving the pregnancy outcome.

There was a notable minority of participants who had low-risk term births and a VMB that could be considered to have unfavourable characteristics. 21.4% of low-risk term birth participants had a VMB in CST III (*L. iners* domination), that has been associated with sPTB[102] and 24.1% had a VMB in CST IV (*Lactobacillus* deplete), that has been associated with both sPTB and PPROM.[101,106,162] Likewise, 32% (46/145) of low-risk term birth participants had a Simpson diversity index over 0.5, higher diversity has been associated with sPTB and PPROM in three studies.[110,162,168] 7.5% (11/145) of low-risk term birth participants also had a richness of 20 or greater, another VMB characteristic that has been (inconsistently) associated with sPTB/PPROM. [103,168] However, different analytical measures inhibit direct comparison. Likewise differential computation of the bacterial load[101,168] inhibited direct comparison of this metric with previous work.

It is possible to directly compare the distribution of the group according to *Lactobacillus* relative abundance in the low-risk term birth group with that described in the term birth

group, as described by Brown *et al*.[103] In the present study 17.9% (26/145) of low-risk term birth participants were *Lactobacillus* deplete (<50% relative abundance of lactobacillus), this is slightly higher than the 10% (3/30)[173] at 20-22 weeks gestation and 2.7% (1/36)[103] and of term birth participants at 23-26 weeks gestation as described by Brown *et al*. This may partly be accounted for by the earlier sampling gestation, as the VMB tends to *Lactobacillus* dominance as pregnancy progresses.[95,96] The difference could also be accounted for by different sequencing methods, the MacIntyre group[103,173] at this time were using primers targeted at the V1-2 hypervariable regions, which could have preferentially sequenced the lactobacillui and led to under- sequencing of non-lactobacilli species such as *Gardnerella vaginalis* and *Bifidobacterium bifidum.*[201] In contrast the current study used primers targeted at the V3-4 hypervariable regions, which recent research indicates gives a more accurate representation of non-lactobacilli species.[201] Therefore, whether the current population truly has more low-risk participants who go on to have term births with lactobacilli depletion than the London population, as sampled by the MacIntyre group, is unclear.

It is also noteworthy that a small proportion of participants in the low-risk term birth group had rarer taxa that have previously been associated with PTB. For example, the recently sequenced BVAB TM7-H1 was reported as associated with preterm birth in the NIH funded vaginal microbiome study, and indeed appears not to have been present prior to 24 weeks gestation in pregnancies that delivered at term in that study (n=59).[110] BVAB TM7-H1 was present in two (2/145, 1.3%) low-risk term birth participants in the present study (at relative abundances of 2% and 12.3% respectively). When assessing rare taxa in 16S rRNA sequencing studies it is important to consider the chance of spurious taxa due to undetected contamination or taxonomic mis-assignment.[107] Nevertheless the presence of

BVAB TM7-H1 in 1.3% of our low-risk term birth participants is plausible given our larger sample size than the NIH funded vaginal microbiome study.[110]

Six taxa that have been shown to be associated with sPTB/PPROM in three or more previous studies were present in a notable proportion of the low-risk term birth group: *Dialister*, *Mobiluncus, Parvimonas, Peptoniphilus, Prevotella* and *Streptococcus*. These taxa were present in 23.5%, 2.76%, 10.34%, 17.9%, 39.3% and 13.1% respectively of low-risk term birth pregnancies. Of note, when present the median relative abundances of these taxa were under 1% (Table 6.6), so in the majority of participants with these taxa present they comprised a small part of the whole VMB.

It is important to be mindful that association does not imply causation. Within clinical practice it is fairly common for women at low-risk of preterm birth to be prescribed antibiotic treatment for asymptomatic bacterial vaginosis, despite lack of evidence of benefit in this situation.[46,202] 16S rRNA sequencing is not currently clinically applicable because of the long sequencing timeline. However, if this data had been available during pregnancy there could have potentially been un-necessary clinician and patient anxiety caused by the presence of unfavourable VMB characteristics. This group was particularly low-risk of preterm birth because they had all given birth before, and all births had been at term and without major obstetric complications.[121] In the future, if more detailed VMB characteristics are clinically available, then this group of participants, who ultimately all had live births at ≥ 39^{+0} weeks gestation, may provide reassurance to clinicians and women faced with such information in pregnancy.

Based on current evidence[196] there is unlikely to be a single taxon, or even VMB characterisation method, that directly causes sPTB/PPROM. There are likely to be multiple protective factors that lead to term birth in the majority of pregnancies, as demonstrated in this low-risk term birth group.

6.13 Conclusion

The low-risk term birth group have demonstrated the distribution of VMB characteristics in our population. This data will be used to compare to the high-risk group in the following chapter.

The low-risk term birth VMB characteristics are in keeping with other pregnancy cohorts with term birth sampled in the second trimester, particularly in those of predominantly Caucasian ethnic origin. The VMB showed a high degree of stability between the two study visits (approximately 16 and 20 weeks gestation), this therefore justified the combination of study samples for both visits to increase statistical power. There was a notable minority of low-risk term birth participants who demonstrated VMB characteristics that previous studies have high-lighted as associated with preterm birth. These women had a very low baseline risk of preterm birth and are likely to possess multiple protective factors that facilitated term birth.

7 Results- vaginal microbiota in high-risk women

7.1 Overview

This chapter describes the outcome of our primary analysis: a nested case-control study of the contribution of the VMB to recurrent early sPTB/PPROM. The VMB is characterised of the 109 high-risk participants who had either term birth (≥37^{+0} weeks), sPTB <34^{+0} weeks or PPROM <34^{+0} weeks. Assessment is made of whether the VMB characteristics identified in chapter 4 are able to identify a contribution of the VMB to early sPTB/PPROM, using aetiological modelling. The VMB characteristics of the high-risk participants are presented alongside those of the low-risk term births in order to give a frame of reference for the expected distribution of our parameters within the local population.

For the majority of analysis the first valid sample from each participant is used, with the exception of the stability analysis whereby only the subset of participants with samples available at both visits (approximately 16 and 20 weeks gestation) are included, and both samples used.

7.2 Heatmap

The relative abundance of taxa in each bacterial group, and the 13 taxa with the highest relative abundance assessed visually, did not show a clear difference by pregnancy outcome group (Figure 7.1). The most common taxa were the species of Lactobacillus: *L. iners* and *L. crispatus* followed by all other types of *Lactobacillus* pooled.

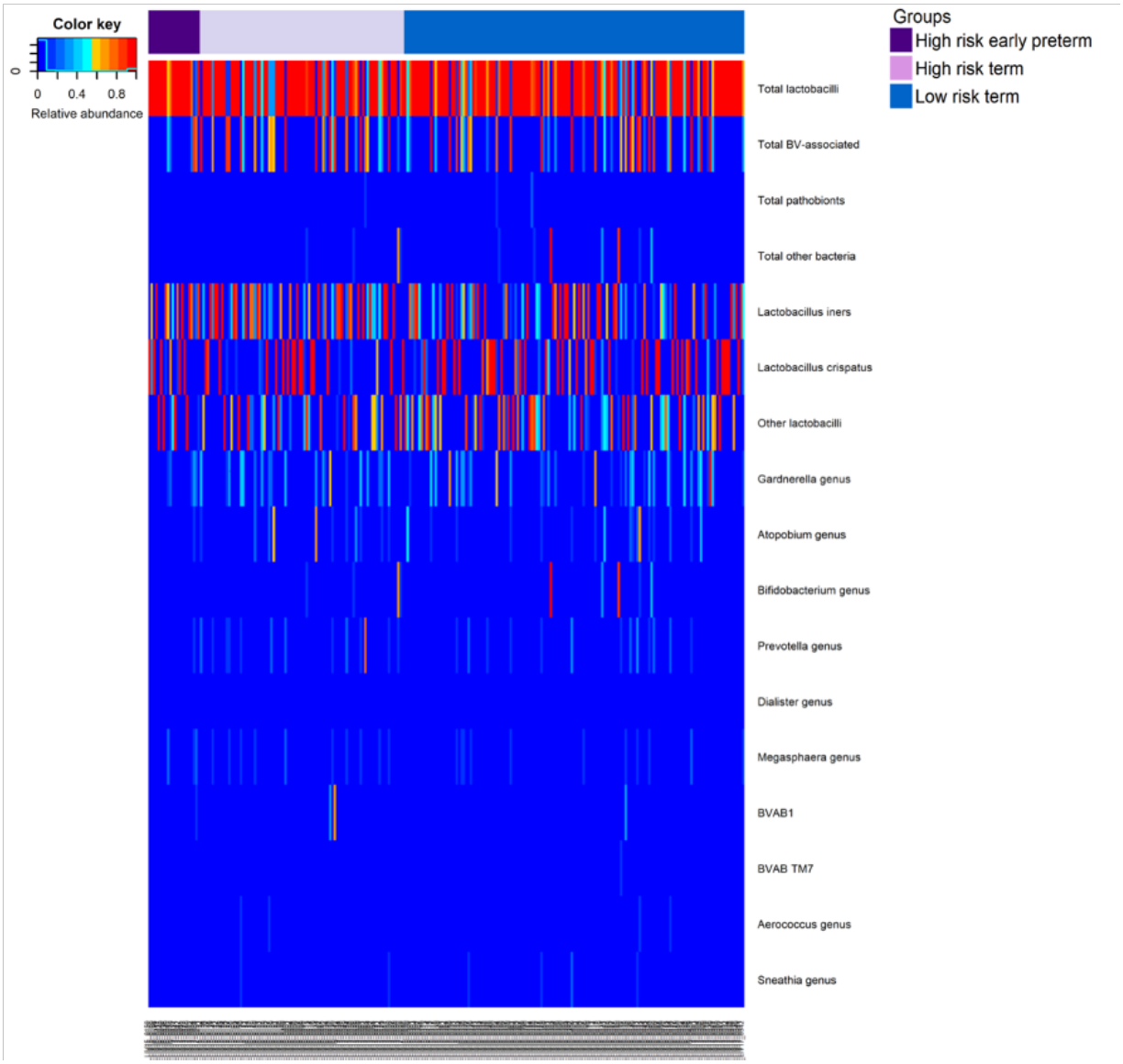


Figure 7.1: Heatmap. This diagram depicts all samples (n = 254) on the x-axis. The top four rows of the y-axis show each amplicon sequence variants (ASVs) attributed to one of four mutually exclusive groups (appendix G). The lower bars on the y-axis show the 13 most common ASVs. The top bar depicts pregnancy outcome group. Abbreviations: BV, bacterial vaginosis; BVAB1, BV-associated bacterium type 1; BVAB TM7, BV-associated bacterium (phylum TM7).

7.3 Assessment as to whether to analyse phenotypes together or separately

In order to assess whether the early sPTB and early PPROM groups should be analysed together or separately, the VMB characteristics of richness, diversity, VMB type and bacterial load were initially visualised, looking for differences between early sPTB and early PPROM. These four VMB characteristics were selected because they assess different components of the VMB.

7.3.1 VMB characteristics separated by type of preterm birth

Graphical representation showed no difference in species richness (Figure 7.2), Simpson diversity (Figure 7.3), VMB type (Figure 7.4), or bacterial load (Figure 7.5) between high-risk participants with early sPTB or early PPROM.

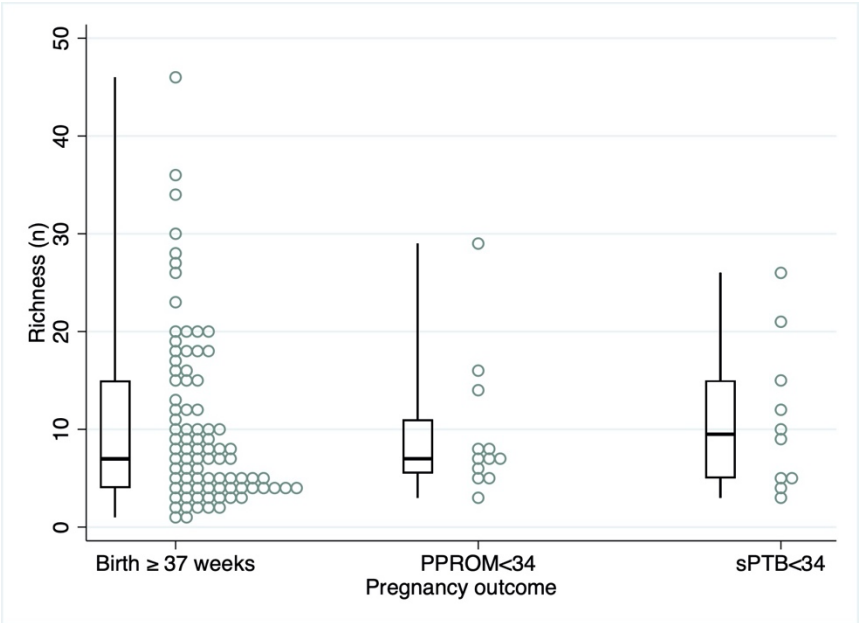


Figure 7.2: Richness of high-risk participants by pregnancy outcome in high-risk group. Dots show individual values, box plot shows median, interquartile range and whiskers show 2.5% and 97.5% percentiles. Difference between early sPTB and early PPROM groups p-value= 0.766, Mann-Whitney U test

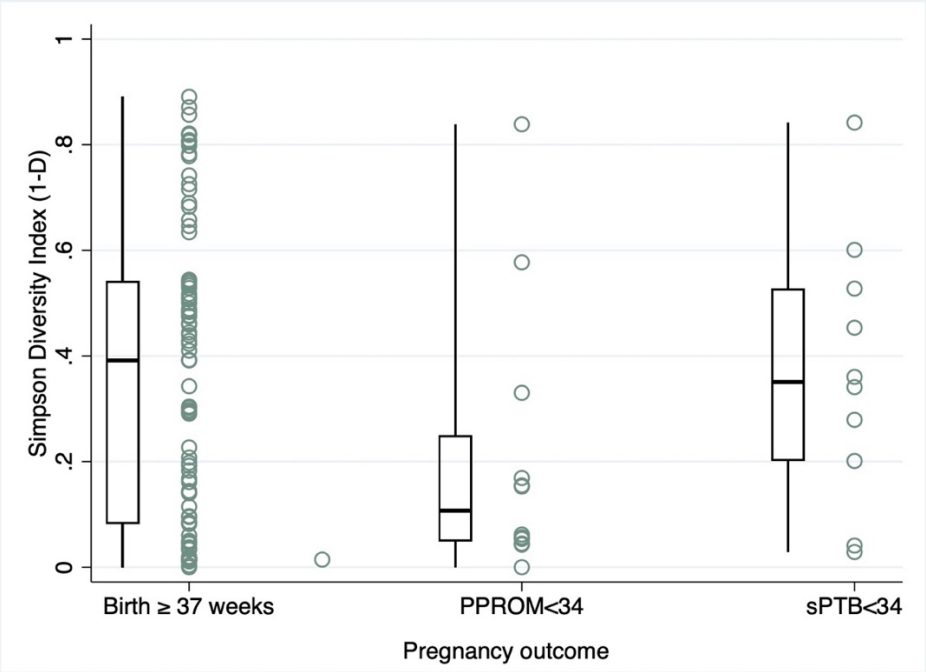


Figure 7.3: Simpson Diversity according to pregnancy outcome in high-risk group. Dots show individual values, box plot shows median, interquartile range and whiskers show 2.5% and 97.5% percentiles. Difference between early sPTB and early PPROM groups p-value= 0.129, Mann-Whitney U test

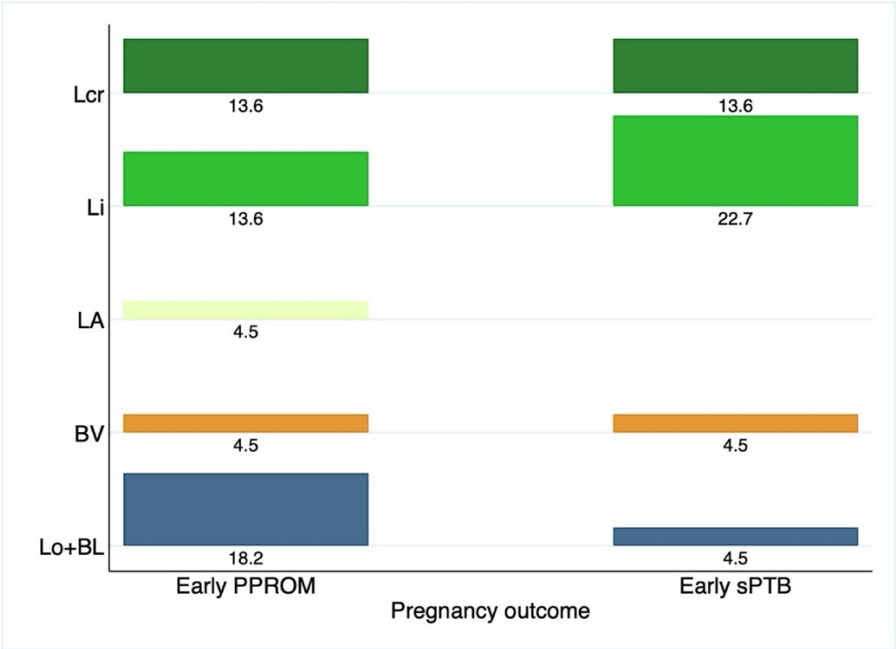


Figure 7.4: Distribution of VMB types by in participants with early PPROM (n=12) and early sPTB (n=10) VMB type was defined based on relative abundance of applicable taxa as follows: (1) L. iners-dominated (Li; >75% lactobacilli of which L. iners was the most common); (2) L. crispatus (Lcr; ≥75% lactobacilli of which L. crispatus was the most common); (3) other lactobacilli-dominated or Bifidobacterium dominated (Lo+BL; either≥75% lactobacilli of which L. jensenii or L. gasseri were the most common, or Bifidobacterium ≥50%); (4) lactobacilli, and anaerobes (LA; 25%- 75% lactobacilli); and (5) polybacterial G. vaginalis-containing (BV; <25% lactobacilli and <50% Bifidobacterium). Difference between early sPTB and early PPROM groups p-value= 0.680, Fisher's exact test

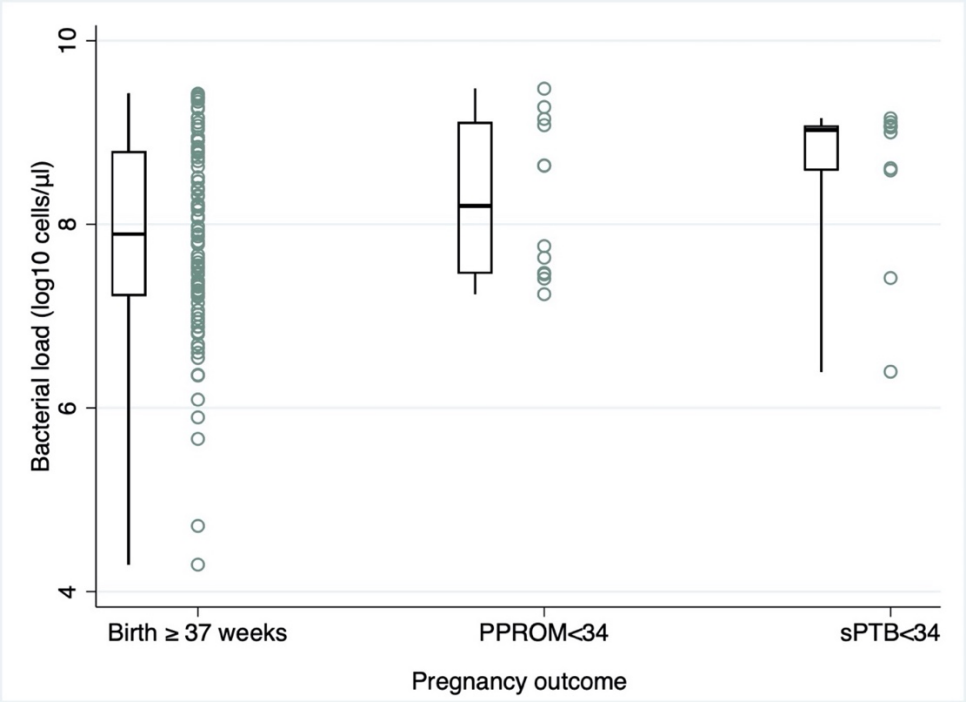


Figure 7.5: Bacterial load according to pregnancy outcome in high-risk group. Dots show individual values, box plot shows median, interquartile range and whiskers show 2.5% and 97.5% percentiles. Difference between early sPTB and early PPROM groups p-value= 0.742, Mann-Whitney U test

Based on the similar distribution of VMB characteristics for high-risk participants with early sPTB and early PPROM, and in order to reduce the risk of false positive results due to multiple testing and false negative results due to lack of statistical power, the research team decided to combine these groups.

7.4 Richness and Diversity

In the high-risk women we did not identify any association between recurrent sPTB/PPROM and VMB richness or Simpson (1-D) diversity (Table 7.1)

7.5 Community State Type

In the high-risk women we did not identify any association between recurrent sPTB/PPROM and allocation to CST[95] group (Table 7.2)

7.6 Group according to relative abundance of *Lactobacillus*

In the high-risk women we did not identify an association between recurrent sPTB/PPROM and allocation to group according to relative abundance of *Lactobacillus*[103] (Table 7.2).

7.7 Vaginal microbiota stability

The VMB of the majority of participants with samples available from both visits between 15 and 23 weeks gestation was stable (Table 7.3). 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 fluctuated between lactobacilli-domination and anaerobic dysbiosis (n=12) or vice-versa (n=17).

In the high-risk women we did not identify an association between recurrent sPTB/PPROM and allocation to categorical stability groups.[95]

7.8 Vaginal microbiota types

We did not identify an association between VMB type and reoccurrence of sPTB/PPROM (Table 7.4).

Table 7.1: VMB richness and diversity and the association with sPTB/PPROM in high risk women

	Low-risk pregnancy (LR)	**High-risk pregnancy (HR)**				Unadjusted logistic regression model of HR ≥ 37 vs <34 weeks (n=109)			Adjusted** logistic regression model of HR ≥ 37 vs <34 weeks (n=106)		
	Birth ≥39 weeks n=145	All high-risk pregnancies n=109	Birth ≥37 weeks n=87	sPTB or PPROM <34 weeks n=22	p value* difference between HR ≥ 37 vs. <34 weeks						
	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)		OR	95% CI	P	OR	95% CI	P
Richness (n)	7 (4-14)	7 (4-15)	7 (4-15)	8 (5-14)	0.351	1.00	(0.94-1.06)	0.976	0.99	(0.93-1.06)	0.785
Simpson diversity (1-D)	0.38 (0.07-0.53)	0.30 (0.06-0.53)	0.39 (0.08-0.54)	0.19 (0.05-0.45)	0.599	0.34	(0.06-1.99)	0.231	0.29	(0.04-2.18)	0.232

*Table 7.1 notes *P values based on Mann-Whitney U test for continuous variables **Logistic regression adjusted for body mass index (BMI) as a quadratic term, history of cervical surgery (none/single LLETZ/multiple LLETZ and/or knife cone biopsy) and smoking (yes/no at time of visit). The adjusted analyses excluded two high risk women with missing smoking data, and one high risk woman with missing BMI data, the remainder of the exclusions were because of collinearity within the regression.*

Table 7.2: CST and group according to Lactobacillus relative abundance and the association with sPTB/PPROM in high-risk women

	Low-risk pregnancy (LR)	**High-risk pregnancy (HR)**				Unadjusted logistic regression model of HR ≥ 37 vs <34 weeks (n=109)			Adjusted** logistic regression model of HR ≥ 37 vs <34 weeks (n=106)		
	Birth ≥39 weeks n=145	All high-risk pregnancies n=109	Birth ≥37 weeks n=87	sPTB/PPROM <34 weeks n=22	p value* difference between HR ≥ 37 vs. <34 weeks						
	Number (%)	Number (%)	Number (%)	Number (%)		OR	95% CI	P	OR	95% CI	P
Community State Type											
I *(L. crispatus)*	43 (29.7)	23 (21.1)	17 (19.5)	6 (27.3)		Comparator			Comparator		
II *(L. gasseri)*	16 (11.0)	11 (10.1)	9 (10.3)	2 (9.1)		0.63	(0.10-3.78)	0.613	0.38	(0.04-3.95)	0.421
III *(L. iners)*	31 (21.4)	33 (30.3)	26 (29.9)	7 (31.8)	0.848	0.76	(0.22-2.66)	0.671	0.55	(0.13-2.24)	0.402
IV (mixed non Lactos)	35 (24.1)	29 (26.6)	25 (28.7)	4 (18.2)		0.45	(0.11-1.85)	0.271	0.43	(0.09-2.01)	0.287
V *(L. jensenii)*	20 (13.8)	13 (11.9)	10 (11.5)	3 (13.6)		0.85	(0.17-4.17)	0.841	0.94	(0.17-5.12)	0.944
Group according to *Lactobacillus* relative abundance											
Lactobacillus dominance (>75%)	105 (72.4)	80 (73.4)	61 (70.1)	19 (86.4)		Comparator			Comparator		
Intermediate (50-74.9%)	14 (9.7)	6 (5.5)	5 (5.7)	1 (4.5)	0.277	0.64	(0.07-5.84)	0.69	0.91	(0.09-9.36)	0.940
Lactobacillus deplete (<50%)	26 (17.9)	23 (21.1)	21 (24.1)	2 (9.1)		0.31	(0.07-1.42)	0.13	0.29	(0.05-1.53)	0.140

*Table 7.2 notes:.*P value calculated by Fisher's exact test **Logistic regression adjusted for body mass index (BMI) as a quadratic term, history of cervical surgery (none/single LLETZ/multiple LLETZ and/or knife cone biopsy) and smoking (yes/no at time of visit). The adjusted analyses excluded two high risk women with missing smoking data, and one high risk woman with missing BMI data, the remainder of the exclusions were because of collinearity within the regression.*

Table 7.3: Stability assessment and the association between VMB stability and recurrent sPTB/PPROM in high-risk women

Stability: change in VMB between 16 and 20 weeks gestation	**Low-risk pregnancy (LR)**		**High-risk pregnancy (HR)**					Unadjusted logistic regression model of HR ≥ 37 vs <34 weeks (n=88)			Adjusted** logistic regression model of HR ≥ 37 vs <34 weeks (n=84)		
	Birth ≥39 weeks n=129		All high-risk pregnancies n=88		Birth ≥37 weeks n=71		sPTB/PPROM <34 weeks n=17	p value* difference between HR ≥ 37 vs. <34 weeks					
	Number	(%)	Number	(%)	Number	(%)	Number	(%)		OR	95% CI	P	OR	95% CI	P
Same lactobacilli	83	(64.3)	58	(65.9)	46	(64.8)	12	(70.6)		Comparator			Comparator		
Different lactobacilli	3	(2.3)	2	(2.3)	1	(1.4)	1	(5.9)		3.83	(0.22-65.85)	0.354	4.40	(0.19-104)	0.359
Remain non-LD	29	(22.5)	13	(14.8)	12	(16.9)	1	(5.9)	0.579	0.32	(0.04-2.71)	0.295	0.42	(0.04-4.11)	0.457
Non-LD to LD	7	(5.4)	10	(11.4)	8	(11.3)	2	(11.8)		0.96	(0.18-5.11)	0.960	0.96	(0.16-5.90)	0.965
LD to non-LD	7	(5.4)	5	(5.7)	4	(5.6)	1	(5.9)		0.96	(0.10-9.38)	0.971	0.41	(0.02-7.21)	0.543

*Table 7.3 notes: LD, Lactobacillus dominant.*P value calculated by Fisher's exact test **Logistic regression adjusted for body mass index (BMI) as a quadratic term, history of cervical surgery (none/single LLETZ/multiple LLETZ and/or knife cone biopsy) and smoking (yes/no at time of visit). The adjusted analyses excluded two high risk women with missing smoking data, and one high risk woman with missing BMI data, the remainder of the exclusions were because of collinearity within the regression.*

Table 7.4 Distribution of VMB types by pregnancy outcome and the association between VMB type and recurrent sPTB/PPROM in high risk women:

	Low-risk pregnancy (LR)	**High-risk pregnancy (HR)**				Unadjusted logistic regression model of HR ≥ 37 vs <34 weeks (n=109)			Adjusted** logistic regression model of HR ≥ 37 vs <34 weeks (n=106)		
	Birth ≥39 weeks n=145	All high-risk pregnancies n=109	Birth ≥37 weeks n=87	sPTB/PPROM <34 weeks n=22	p value* difference between HR ≥ 37 vs. <34 weeks						
	Number (%)	Number (%)	Number (%)	Number (%)		OR	95% CI	P	OR	95% CI	P
Lcr	39 (26.9)	23 (21.1)	17 (19.5)	6 (27.3)		Comparator			Comparator		
Li	32 (22.1)	36 (33.0)	28 (32.2)	8 (36.4)		0.810	(0.24-2.74)	0.734	0.592	(0.15-2.34)	0.455
Lo+BL	37 (25.5)	22 (20.2)	17 (19.5)	5 (22.7)	0.696	0.885	(0.23-3.48)	0.862	0.814	(0.17-3.79)	0.793
LA	18 (12.4)	14 (12.8)	13 (14.9)	1 (4.5)		0.202	(0.02-1.89)	0.161	0.156	(0.01-1.98)	0.154
BV	19 (13.1)	14 (12.8)	12 (13.8)	2 (9.1)		0.472	(0.08-2.75)	0.404	0.480	(0.07-3.24)	0.451

*Table 7.4 notes: P value calculated by Fisher's exact test **Logistic regression adjusted for body mass index (BMI) as a quadratic term, history of cervical surgery (none/single LLETZ/multiple LLETZ and/or knife cone biopsy) and smoking (yes/no at time of visit). The adjusted analyses excluded two high risk women with missing smoking data, and one high risk woman with missing BMI data, the remainder of the exclusions were because of collinearity within the regression. VMB type was defined based on relative abundance of applicable taxa as follows: (1) L. iners-dominated (Li; >75% lactobacilli of which L. iners was the most common); (2) L. crispatus (Lcr; ≥75% lactobacilli of which L. crispatus was the most common); (3) other lactobacilli-dominated or Bifidobacterium dominated (Lo+BL; either≥75% lactobacilli of which L. jensenii or L. gasseri were the most common, or Bifidobacterium ≥50%); (4) lactobacilli, and anaerobes (LA; 25%- 75% lactobacilli); and (5) polybacterial G. vaginalis-containing (BV_GV; <25% lactobacilli and <50% Bifidobacterium).*

Table 7.5: Relative abundance of bacterial groups and taxa of interest by pregnancy outcome and association with sPTB/PPROM in the high-risk group

		Low-risk pregnancy				High-risk pregnancy									
		Birth ≥ 39 weeks n=145				Birth ≥ 37 weeks n=87				sPTB or PPROM <34 weeks n=22				P value HR ≥ 37 vs <34 weeks	
		Number of samples with taxa	% of samples with taxa	Median	IQR	Number of samples with taxa	% of samples with taxa	Median	IQR	Number of samples with taxa	% of samples with taxa	Median	IQR	Presence/ absence of taxa	Relative abdundance of taxa
Relative abundance of bacterial groups (%)	Total Lactobacillus	143	(98.62)	98.55	(69.75-99.90)	85	(97.70)	99.18	(53.04-99.90)	22	(100.00)	98.46	(91.64-99.82)	1.00	0.91
	Total BV associated	109	(75.17)	0.82	(0.09-20.44)	62	(71.26)	0.64	(0-46-32)	19	(86.36)	1.14	(0.18-8.23)	0.18	0.74
	Total pathobionts	34	(23.45)	0	(0-0)	20	(22.99)	0	(0-0)	4	(18.18)	0	(0-0)	0.78	0.73
	Total other bacteria	50	(34.48)	0	(0-0.18)	35	(40.23)	0	(0-0.09)	10	(45.45)	0	(0-0.09)	0.81	0.89
Relative abundance of species of Lactobacillus (%)	*Lactobacillus crispatus*	125	(86.21)	2.00	(0.18-79.29)	71	(81.61)	1.09	(0.17-16.71)	18	(81.82)	1.73	(0.09-68.84)	1.00	0.84
	Lactobacillus gasseri	47	(32.41)	0	(0-0.18)	31	(35.63)	0	(0-0.64)	5	(22.73)	0	(0-0)	0.32	0.28
	Lactobacillus iners	82	(56.55)	0	(0-48.59)	63	(72.41)	21.98	(0-78.11)	13	(59.09)	16.21	(0-79.93)	0.30	0.76
	Lactobacillus jensenii	53	(36.55)	0	(0-3.09)	24	(27.59)	0	(0-0.02)	9	(40.91)	0	(0-4.81)	0.30	0.30
Relative abundance of taxon of interest (%)	*Aerococcus*	24	(16.55)	0.82	(0.27-3.27)	18	(20.69)	1.27	(0.54-1.81)	3	(13.64)	0.36	(0.18-0.36)	0.34	0.08
	Atopobium vaginae	44	(30.34)	3.04	(0.27-14.17)	28	(32.18)	7.58	(0.77-13.81)	4	(18.18)	2.04	(1.14-7.81)	0.30	0.53
	Bifidobacterium breve	14	(9.66)	1.45	(0.27-31.70)	7	(8.05)	3.63	(0.36-10.90)	0	0.00			0.34	
	Clostridiales BVAB2	11	(7.59)	2.36	(0.54-6.12)	10	(11.49)	1.68	(0.27-3.63)	2	(9.09)	2.45	(0.27-0.46)	1.00	0.83
	Dialister	34	(23.45)	0.27	(<0.1-0.72)	29	(33.33)	0.27	(0.18-0.82)	2	(9.09)	0.86	(0.18-1.54)	**0.03**	0.78
	Gardnerella vaginalis	57	(39.31)	15.44	(0.91-33.51)	33	(37.93)	18.62	(0.64-35.06)	10	(45.45)	11.26	(1.72-19.16)	0.63	0.22
	Lachnospiracea BVAB1	3	(2.07)	0.18	(<0.1-32.97)	4	(4.60)	16.03	(0.14-48.82)	1	(4.55)	16.26		1.00	1.00
	Megasphaera	21	(14.48)	9.90	(6.09-12.81)	18	(20.69)	10.40	(3.54-12.26)	3	(13.64)	19.26	(11.90-21.44)	0.56	**0.04**
	Mobiluncus	4	(2.76)	0.18	(<0.1-0.50)	2	(2.30)	0.59	(0.27-0.91)	2	(9.09)	6.90	(0.27-13.5)	0.18	0.68
	Mycoplasma	2	(1.38)	0.91	(0.82-1.00)	4	(4.60)	0.45	(<0.1-0.90)	1	(4.55)	0.18		1.00	1.00
	Parvimonas	15	(10.34)	0.63	(0.18-2.36)	12	(13.79)	0.45	(<0.1-1.18)	2	(9.09)	0.59	(0.27-0.91)	0.73	0.85
	Peptoniphilus	26	(17.93)	0.18	(0.18-0.45)	21	(24.14)	0.18	(<0.1-0.54)	6	(27.27)	<0.1	(<0.1-0.18)	0.79	0.22
	Prevotella	57	(39.31)	0.91	(0.27-6.99)	42	(48.28)	2.86	(0.18-11.35)	7	(31.82)	0.54	(0.18-0.91)	0.23	0.40
	Porphyromonas species	7	(4.83)	0.18	(<0.1-0.99)	12	(13.79)	0.18	(<0.1-1.45)	1	(4.55)	0.45		0.46	0.41
	Sneathia species	14	(9.66)	4.77	(1.18-10.90)	15	(17.24)	1.81	(0.36-6.17)	2	(9.09)	3.45	(0.18-6.72)	0.52	0.82
	Streptococcus species	19	(13.10)	0.27	(<0.1-10.8)	10	(11.49)	0.13	(<0.1-0.27)	2	(9.09)	<0.1	(<0.1-<0.1)	1.00	0.23
	BVAB TM7-H1	2	(1.38)	7.20	(2.0-12.3)	0	0.00			1	(4.55)	0.36		0.20	
	Ureaplasma species	26	(17.93)	0.22	(0.18-0.45)	11	(12.64)	<0.1	(<0.1-0.36)	7	(31.82)	<0.1	(<0.1-0.45)	0.05	0.84

Table 7.5 notes: Grey box, value not applicable due to no sample fitting this criterion. P values based on Fisher's exact test for presence/absence and Mann-Whitney U test for relative abundance. Test of significance based on relative abundance includes all participants for relative abundance of bacterial groups and species of lactobacilli but is limited to only those participants with each taxon present for the taxa of interest (to avoid skewing of data in cases of rare taxa).Bold indicates p<0.05

7.9 Taxa of interest

We assessed the presence/absence and, if present, the relative abundance of the 4 species of *Lactobacillus* and 19 taxa of interest identified in Table 4.3. All species of *Lactobacillus* and 18 of the taxa of interest were identified in at least one participant. *Mageeibacillus indolicus* was not identified in any participants and was, therefore, excluded from further analysis. The percentage of participants with each taxa present, median relative abundance (with IQR) and association between presence/absence and early sPTB/PPROM in the high-risk group are displayed in Table 7.5.

Contrary to previous studies, the *Dialister* species was present in more high-risk participants who had term birth (29/87, 33.3%) than early sPTB/PPROM (2/22, 9.1%), p=0.03. *Megasphaera* species was only present in 3 high-risk participants with recurrent sPTB/PPROM, but when present, the relative abundance was higher (median 19.3%, IQR 11.9-21.4) than in the high-risk participants with term births (median 10.5%, IQR 3.5-12.3), p=0.04 (Figure 7.6).

The median relative abundance of *Ureaplasma* and BVAB TM7-H1 were very low, but *Ureaplasma* species and BVAB TM7-H1 were present in a higher proportion of women with recurrent early sPTB/PPROM (31.8% and 4.6%) than those without (12.6% and 0%;Table 7.5, Figure 7.7). 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 7.5, Figure 7.7). These differences did not reach traditional statistical significance of $p<0.05$ (Figure 7.7).

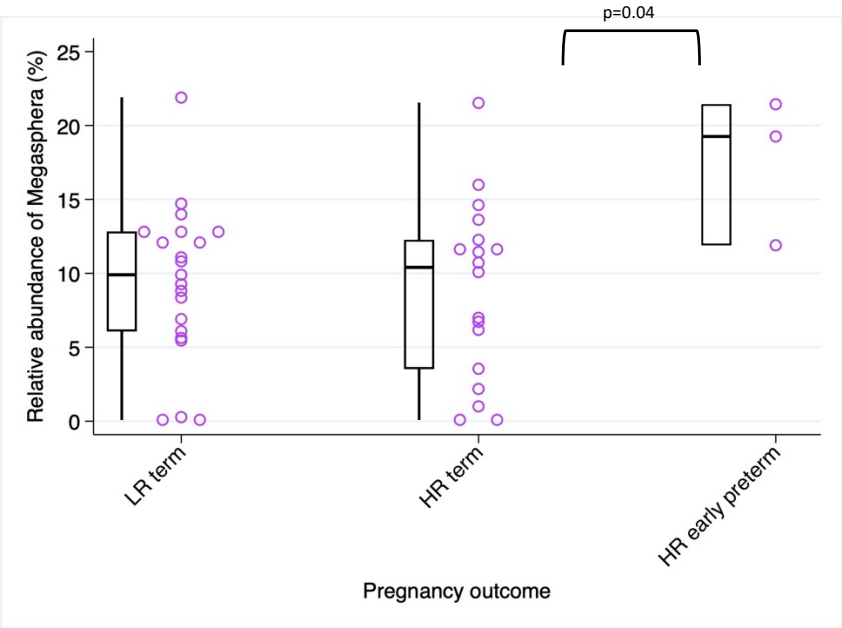


Figure 7.6: Relationship between relative abundance of Megasphaera and pregnancy outcome. Analysis restricted to the subset of participants with Megasphaera present. Dots show individual values, box plot shows median, interquartile range and whiskers show 2.5% and 97.5% percentiles. P value calculated by Mann-Whitney U test.

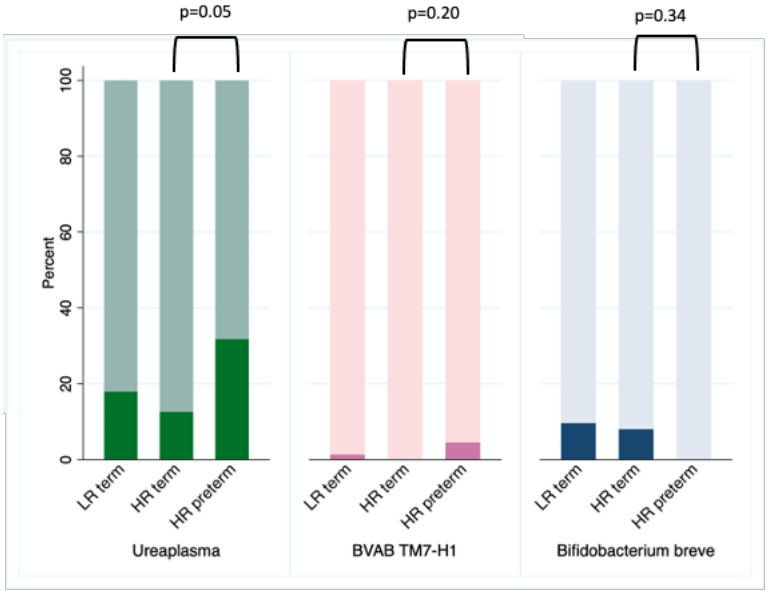


Figure 7.7: Positive trends identified between presence/absence of taxa of interest and recurrent early (<34 weeks) sPTB or PPROM (abbreviated to HR preterm for diagram), and low risk cohort for comparison. Dark colour= taxon present. Light colour= taxon not present.

7.9.1 Benjamini-Hochberg false discovery rate (FDR) procedure

The Benjamini- Hochberg false discovery rate (FDR) procedure[157] was performed for the positive associations identified in Table 7.5 with an FDR rate of 25%. The associations between *Ureaplasma* and *Megasphaera* and high-risk pregnancy outcome were no longer statistically significant after accounting for multiple comparisons.

7.10 Concentration of taxa in bacterial groups and total bacterial load

High-risk women with an early PTB recurrence had a higher overall vaginal bacterial load (8.64 vs. 7.89 $\log_{10}$ cells/µl, adjusted odds ratio (aOR) 1.90, 95% 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 an early PTB recurrence (Table 7.6, Figure 7.8). These findings remained significant after adjustment for smoking, previous cervical surgery and BMI. The remainder of the bacterial groups and major species of *Lactobacillus* were not statistically different according to pregnancy outcome in the high-risk group (Table 7.6).

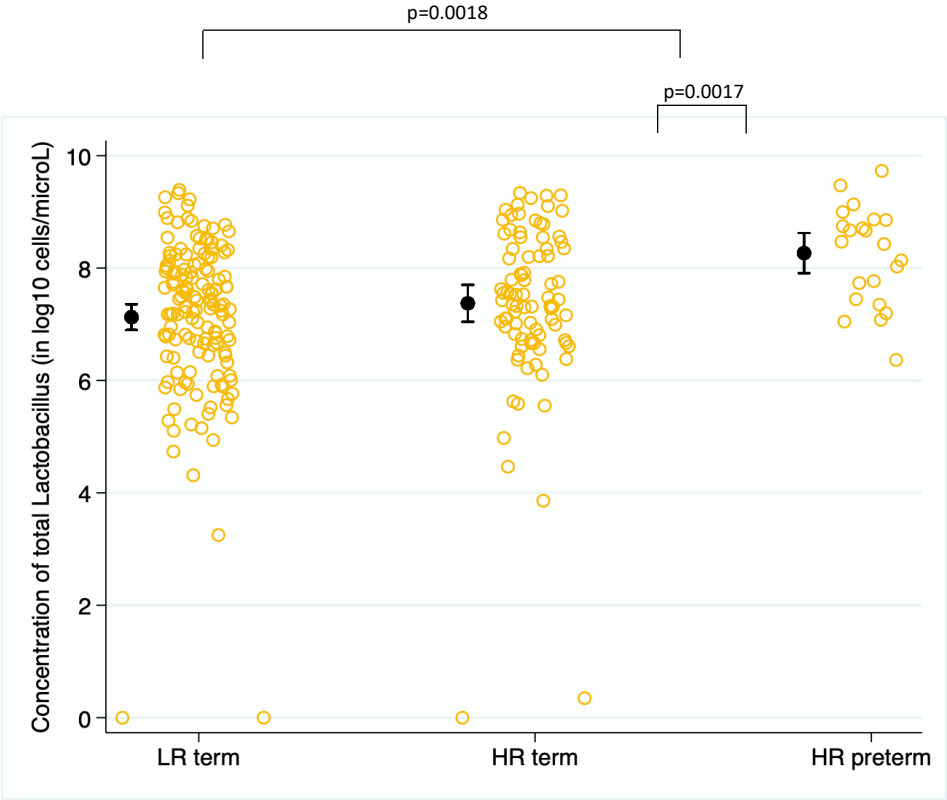


Figure 7.8: Concentration of total Lactobacillus by pregnancy outcome. Dots show individual values, scale shows mean and 95% CI of mean. P values quoted as difference between HR and LR cohort and between HR term and HR preterm within HR cohort (Mann-Whitney test).

Table 7.6: Total bacterial concentration, concentration of taxa within each bacterial group and concentration of major species of Lactobacillus by pregnancy outcome and relationship to early sPTB/PPROM in the high risk group.

		Low risk pregnancy (LR)		**High risk pregnancy (HR)**					Unadjusted logistic regression model of HR ≥ 37 vs <34 weeks (n=109)			Adjusted * logistic regression model of HR ≥ 37 vs <34 weeks (n=106)		
		Birth ≥ 39 weeks n=145		Birth ≥ 37 weeks n=87		sPTB or PPROM <34 weeks (early preterm) n=22		P value* diference HR≥ 37 vs. <34 weeks n=109						
		Median (IQR)		Median (IQR)		Median (IQR)		P	OR	95% CI	P	OR	95% CI	P
Overall assessment														
	Bacterial load	7.68	(6.80-8.35)	7.89	(7.22-8.80)	8.64	(7.47-9.08)	**0.030**	1.78	(1.04-3.04)	**0.035**	1.90	(1.01-3.56)	**0.047**
Bacterial group														
	Total Lactobacillus	7.31	(6.42-8.01)	7.48	(6.87-8.45)	8.59	(7.47-8.87)	**0.002**	2.30	(1.28-4.13)	**0.005**	2.35	(1.20-4.61)	**0.013**
	Total BV associated	5.41	(1.91-7.03)	5.20	(0-7.65)	6.65	(5.09-8.36)	0.111	1.14	(0.97-1.33)	0.102	1.15	(0.96-1.38)	0.126
	Total pathobionts	0.00	(0-0)	0.00	(0-0)	0.00	(0-0)	0.639	0.94	(0.74-1.19)	0.596	0.93	(0.712-1.22)	0.596
	Total other bacteria	0.00	(0-4.24)	0.00	(0-4.84)	0.00	(0-5.12)	0.612	1.04	(0.88-1.23)	0.671	1.04	(0.86-1.25)	0.698
Type of Lactobacillus														
	L. crispatus	5.88	(4.41-6.83)	6.17	(4.08-6.91)	6.32	(5.79-7.76)	0.188	1.10	(0.91-1.32)	0.319	1.02	(0.83-1.25)	0.843
	L. iners	4.65	(0-7.14)	6.55	(0-7.98)	7.42	(0-8.81)	0.496	0.97	(0.85-1.11)	0.684	0.97	(0.84-1.12)	0.644
	Other Lactobacilli	5.49	(3.77-6.91)	5.81	(4.21-7.09)	5.95	(5.34-7.61)	0.154	1.17	(0.93-1.48)	0.172	1.14	(0.88-1.47)	0.311

*Table 7.6 notes: * P values based on Mann-Whitney U test **Logistic regression adjusted for body mass index (BMI) as a quadratic term, history of cervical surgery (none/single LLETZ/multiple LLETZ and/or knife cone biopsy) and smoking (yes/no at time of visit). The adjusted analyses excluded two high risk women with missing smoking data, and one high risk woman with missing BMI data, the remainder of the exclusions were because of collinearity within the regression. Bold indicates P<0.05.*

Table 7.7: : High risk participants were allocated to quartiles based on the concentration of total bacterial load, total Lactobacillus, L. iners and L. crispartus.

		Concentration (log10 cells/µl)	Low-risk pregnancy (LR)		High-risk pregnancy (HR)						p value* difference between HR ≥ 37 vs. <34 weeks	Unadjusted logistic regression model of HR ≥ 37 vs <34 weeks (n=109)			Adjusted ** logistic regression model of HR ≥ 37 vs <34 weeks (n=106)		
			Birth ≥39 weeks n=145		All high-risk pregnancies n=109		Birth ≥37 weeks n=87		sPTB or PPROM <34 weeks n=22								
			n	(%)	n	(%)	n	(%)	n	(%)		OR	95% CI	P	OR	95% CI	P
Quartile of total bacterial load	1	<6.80	37	(25.5)	12	(11.0)	11	(12.6)	1	(4.5)	0.093	1.59	(0.98-2.58)	0.060	1.63	(0.93-2.86)	0.087
	2	≥6.80-<7.68	36	(24.8)	34	(31.2)	28	(32.2)	6	(27.3)							
	3	≥7.68- <8.35	35	(24.1)	18	(16.5)	17	(19.5)	1	(4.5)							
	4	≥8.35	37	(25.5)	45	(41.3)	31	(35.6)	14	(63.6)							
Quartile of total Lactobacillus	1	<6.42	36	(24.8)	13	(11.9)	12	(13.8)	1	(4.5)	**0.018**	2.33	(1.28-4.24)	**0.006**	2.61	(1.27-5.33)	**0.009**
	2	≥6.42- >7.31	36	(24.8)	25	(22.9)	24	(27.6)	1	(4.5)							
	3	≥7.31- >8.01	36	(24.8)	29	(26.6)	23	(26.4)	6	(27.3)							
	4	≥8.01	37	(25.5)	42	(38.5)	28	(32.2)	14	(63.6)							
Quartile of L. crispatus	1	<5.21	32	(22.1)	10	(9.2)	10	(11.5)	0	0.0	0.240	1.91	(1.00-3.63)	**0.049**	2.32	(1.09-4.97)	**0.030**
	2	≥5.21 - <6.03	30	(20.7)	19	(17.4)	16	(18.4)	3	(13.6)							
	3	≥6.03- <7.12	31	(21.4)	37	(33.9)	29	(33.3)	8	(36.4)							
	4	≥7.12	32	(22.1)	23	(21.1)	16	(18.4)	7	(31.8)							
	nil present	not applicable	20	(13.8)	20	(18.3)	16	(18.4)	4	(18.2)							
Quartile of. L. iners	1	<5.78	21	(14.5)	12	(11.0)	11	(12.6)	1	(4.5)	**0.041**	2.35	(1.10-5.04)	**0.028**	2.13	(0.96-4.71)	0.061
	2	≥5.78- <7.04	20	(13.8)	17	(15.6)	17	(19.5)	0	0.0							
	3	≥7.04- <7.63	21	(14.5)	12	(11.0)	10	(11.5)	2	(9.1)							
	4	≥7.63	20	(13.8)	35	(32.1)	25	(28.7)	10	(45.5)							
	nil present	not applicable	63	(43.4)	33	(30.3)	24	(27.6)	9	(40.9)							

*Table 7.7 notes: * P values based Fisher's exact test**Logistic regression adjusted for body mass index (BMI) as a quadratic term, history of cervical surgery (none/single LLETZ/multiple LLETZ and/or knife cone biopsy) and smoking (yes/no at time of visit). The adjusted analyses excluded two high risk women with missing smoking data, and one high risk woman with missing BMI data, the remainder of the exclusions were because of collinearity within the regression. The lower two logistic regressions by quartile of species of Lactobacillus were restricted only to samples with the species present (top two included all participants). Bold indicates P<0.05.*

7.11 Allocation of high-risk participants to quartiles of low-risk participants

In order to contextualise the VMB characteristics in the high-risk group, the low-risk group was used to create quartiles of VMB characteristics when the variable was present in over 30% of participants. The VMB characteristics selected for analysis were: 1) total bacterial load; 2) concentration of total Lactobacillus; 3) concentration of *L. iners* and; 4) concentration of *L. crispartus*.

Assessment was then made of the contribution that each increase in quartile of the VMB characteristics had to risk of early sPTB/PPROM in the high-risk group (Table 7.7, Figure 7.9.). This showed that high-risk participants had an aOR of 2.61 (95%CI 1.27-5.33, p=0.006) of early sPTB/PPROM for each increase in quartile of total *Lactobacillus* (visualised in Figure 7.10), and an aOR of 2.32 (95%CI 1.09-4.97, p=0.030) for each increase in quartile of *L. crispatus.*

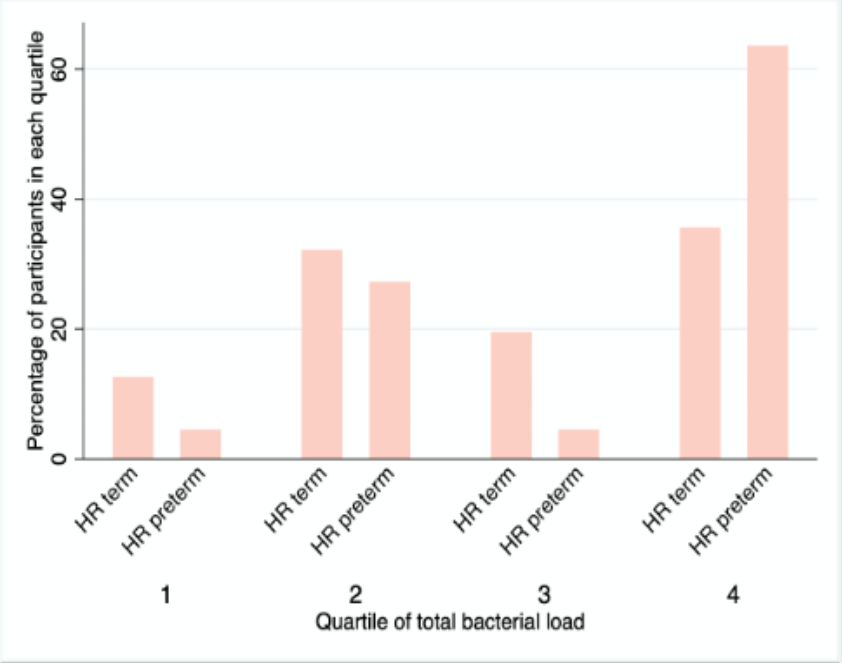


Figure 7.9: Percentage of high-risk participants for each pregnancy outcome in each quartile of total bacterial load

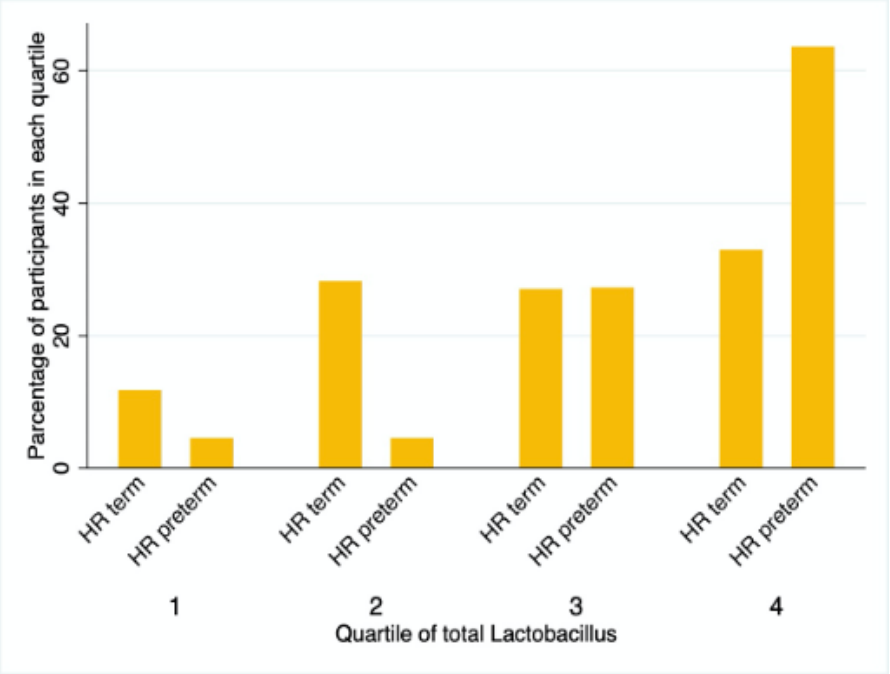


Figure 7.10: Percentage of high-risk participants for each pregnancy outcome in each quartile of total lactobacillus

7.12 Exploratory analysis

Based on the finding of the association between a higher bacterial load, and higher concentration of *Lactobacillus* and sPTB/PPROM in the high-risk group an exploratory analysis was developed by the research team to assess the VMB characteristics that contributed to the relationship between a high bacterial load and recurrent sPTB/PPROM. All participants were stratified by VMB type (Table 4.6), and the logistic regression analyses were repeated for each stratum. For this analysis the research team decided to combine the VMB types 'LA' (25%- 75% *Lactobacillus* and <50% Bifidobacterium) and 'BV' (<25% *Lactobacillus* and <50% Bifidobacterium) in order to conserve statistical power for these rarer groups (each group had only 14/109, 12.8%, of high risk participants) (Table 7.4). Participants in both of these groups have vaginal dysbiosis (non-dominance of

Lactobacillus), and so this combination was felt to be biologically plausible. Table 7.8 shows estimated concentrations of total vaginal bacteria, the four bacterial groups, *L. cripatus*, *L. iners*, and other *Lactobacillus* in high-risk women with and without a recurrence after stratification by VMB type. In women with the VMB type 'Li' (domination by *L. iners*) participants with a recurrent sPTB/PPROM had a significantly higher median bacterial loads, and estimated concentrations of total *Lactobacillus* and *L. iners* compared to high-risk participants with a term birth. However, the associations were no longer significant after adjustment for confounders. Within each of the other VMB type strata there were similar, but non-significant, trends for participants with recurrent early sPTB/PPROM to have higher bacterial loads, and concentrations of both total *Lactobacillus* and total BV-associated bacteria, than high-risk participants with term birth. The median estimated concentrations in high-risk women who had a term birth were similar to those in low-risk women.

High-risk women with anaerobic dysbiosis (LA and BV group) had higher median vaginal bacterial loads than women with domination by *L. iners, L. crispatus*, or other *Lactobacillus*/bifidobacteria (8.54, 7.79, 7.63, 7.53 $\log_{10}$ cells/µl, respectively (Table 7.8 and Table 7.9). High-risk participants with domination of *L. iners* had the highest median bacterial load of women with the *Lactobacillus* dominated VMB types, but these differences did not reach $P<0.05$ in comparison to the other *Lactobacillus* dominated VMB types (Table 7.9).

.

Table 7.8: Stratified Assessment of the relationship between concentration of total bacterial load, taxa in bacterial groups and types of Lactobacillus by VMB types

VMB type¶			Low risk pregnancy (LR)		High risk pregnancy (HR)						Unadjusted logistic regression model of HR ≥ 37 vs <34 weeks (n=109)			Adjusted* logistic regression model of HR ≥ 37 vs <34 weeks (n=106)		
			Birth ≥ 39 weeks n=145		All high-risk pregnancies n=109		Birth ≥ 37 weeks n=87		sPTB or PPROM <34 weeks (early preterm) n=22							
			Median (IQR)		Median (IQR)		Median (IQR)		Median (IQR)		OR	95% CI	P	OR	95% CI	P
LCr	n		39		23		17		6			23			18	
	Bacterial load		7.75	(6.44-8.46)	7.63	(7.14-8.63)	7.55	(7.14-8.32)	8.20	(7.42-9.06)	1.67	(0.57-4.89)	0.349	1.35	(0.41-4.40)	0.623
	Bacterial group	Total Lactobacillus	7.73	(6.42-8.46)	7.63	(7.14-8.59)	7.55	(7.14-8.25)	8.20	(7.41-8.86)	1.64	(0.55-4.85)	0.375	1.32	(0.40-4.30)	0.649
		Total BV associated	4.74	(0-6.29)	4.62	(0-6.07)	4.29	(0-5.36)	5.24	(0-6.71)	1.09	(0.79-1.50)	0.601	0.87	(0.55-1.36)	0.535
		Total pathobionts	0.00	(0-0)	0.00	(0-0)	0.00	(0-0)	0.00	(0-0)	0.93	(0.49-1.75)	0.825	1.07	(0.48-2.39)	0.866
		Total other bacteria	0.00	(0-3.19)	0.00	(0-0)	0.00	(0-0)	0.00	(0-0)	1.01	(0.59-1.71)	0.980	0.97	(0.56-1.68)	0.904
	Type of Lactobacillus	*L. crispatus*	7.72	(6.41-8.45)	7.63	(7.12-8.37)	7.55	(7.12-8.22)	8.19	(7.41-8.80)	1.66	(0.55-5.02)	0.365	1.31	(0.40-4.27)	0.658
		L. iners	0.00	(0-4.23)	0.00	(0-4.81)	0.00	(0-5.32)	0.00	(0-0)	0.71	(0.41-1.21)	0.204	0.84	(0.49-1.42)	0.514
		Other Lactobacilli	5.09	(3.39-5.94)	5.34	(3.02-7.16)	4.64	(3.02-6.14)	6.20	(5.34-7.39)	1.17	(0.77-1.76)	0.457	1.03	(0.93-1.80)	0.930
Li	n		32		36		28		8			36			34	
	Bacterial load		7.64	(6.94-8.25)	7.79	(6.93-8.65)	7.79	(6.93-8.65)	9.03	(8.02-9.12)	3.44	(1.06-11.15)	**0.040**	2.38	(0.71-7.95)	0.160
	Bacterial group	Total Lactobacillus	7.61	(6.90-8.25)	7.96	(7.23-8.83)	7.79	(6.93-8.60)	8.91	(8.02-9.10)	3.40	(1.04-11.11)	**0.042**	2.37	(0.70-8.04)	0.165
		Total BV associated	4.55	(0-6.37)	5.16	(0-6.60)	4.75	(0-5.79)	6.77	(5.33-7.73)	1.45	(0.98-2.15)	0.061	1.57	(0.95-2.59)	0.079
		Total pathobionts	0.00	(0-0)	0.00	(0-0)	0.00	(0-0)	0.00	(0-2.31)	1.21	(0.84-1.74)	0.307	1.15	(0.69-1.91)	0.590
		Total other bacteria	0.00	(0-3.61)	0.00	(0-4.26)	0.00	(0-3.38)	4.07	(0-5.69)	1.39	(1.00-1.94)	0.050	1.43	(0.95-2.16)	0.085
	Type of Lactobacillus	*L. crispatus*	5.89	(3.75-6.60)	6.19	(4.45-6.84)	6.02	(3.83-6.78)	6.38	(6.12-7.79)	1.26	(0.82-1.94)	0.295	1.09	(0.69-1.72)	0.695
		L. iners	7.52	(6.90-8.12)	7.93	(7.13-8.79)	7.73	(6.93-8.59)	8.82	(8.02-9.07)	3.35	(1.03-10.86)	**0.044**	2.33	(0.70-7.76)	0.169
		Other Lactobacilli	4.97	0-6.75)	5.59	(4.08-6.59)	5.40	(4.41-7.30)	5.81	(4.08-6.59)	1.20	(0.78-1.85)	0.398	1.06	(0.69-1.63)	0.782
Lo or BL	n		37		22		17		5			22			19	
	Bacterial load		7.12	(6.32-7.69)	7.53	(7.24-8.64)	7.48	(7.21-8.73)	7.64	(7.46-8.61)	1.36	(0.48-3.86)	0.563	3.47	(0.61-19.7)	0.161
	Bacterial group	Total Lactobacillus	7.12	(6.08-7.69)	7.45	(7.10-8.61)	7.36	(6.94-7.89)	7.61	(7.42-8.61)	1.68	(0.54-5.29)	0.371	5.96	(0.65-54.5)	0.114
		Total BV associated	4.64	(0-5.59)	4.29	(0-6.19)	0.00	(0-4.94)	6.36	(6.31-6.58)	1.84	(0.97-3.50)	0.063	model failed		
		Total pathobionts	0.00	(0-2.98)	0.00	(0-4.11)	0.00	(0-4.11)	0.00	(0-0)	0.87	(0.53-1.41)	0.560	0.66	(0.23-1.87)	0.430
		Total other bacteria	0.00	(0-5.01)	1.13	(0-5.93)	0.00	(0-5.95)	5.09	(0-5.12)	1.05	(0.76-1.44)	0.769	1.21	(0.77-1.91)	0.410
	Type of Lactobacillus	*L. crispatus*	5.17	(4.00-5.88)	5.55	(4.44-6.54)	5.52	(5.26-6.54)	5.59	(0-6.16)	0.85	(0.59-1.21)	0.362	0.83	(0.53-1.29)	0.410
		L. iners	0.00	(0-6.41)	0.00	(0-7.09)	4.42	(0-7.09)	0.00	(0-0)	0.84	(0.61-1.16)	0.297	0.87	(0.58-1.30)	0.497
		Other Lactobacilli	6.99	(5.97-7.51)	7.46	(6.84-8.51)	7.24	(6.93-7.84)	7.61	(7.42-8.30)	1.87	(0.58-6.05)	0.293	8.97	(0.66-121)	0.099
LA or BV	n		37		28		25		3			28			27	
	Bacterial load		7.93	(7.44-8.52)	8.54	(8.11-9.22)	8.31	(8.07-9.16)	9.15	(9.11-9.28)	9.46	(0.28-318)	0.210	12.42	(0.31-501)	0.182
	Bacterial group	Total Lactobacillus	6.98	(6.30-7.45)	7.38	(6.63-8.37)	7.25	(6.56-8.19)	8.26	(8.15-8.86)	4.57	(0.54-38.5)	0.162	5.49	(0.17-63.1)	0.172
		Total BV associated	7.93	(7.19-8.51)	8.41	(8.07-9.19)	8.18	(8.07-9.15)	9.05	(8.83-9.24)	6.58	(0.33-133)	0.219	9.93	(0.42-237)	0.156
		Total pathobionts	0.00	(0-0)	0.00	(0-3.92)	0.00	(0-4.49)	0.00	(0-0)	model failed			model failed		
		Total other bacteria	0.00	(0-4.40)	4.57	(0-6.07)	4.59	(0-5.96)	0.00	(0-6.30)	0.83	(0.54-1.28)	0.402	0.70	(0.30-1.62)	0.407
	Type of Lactobacillus	*L. crispatus*	4.98	(0-5.93)	5.21	(0-6.41)	4.88	(0-6.60)	5.79	(0-5.90)	1.04	(0.69-1.55)	0.859	0.84	(0.48-1.46)	0.537
		L. iners	5.65	(0-7.11)	6.68	(4.80-8.21)	6.56	(4.74-7.98)	8.26	(8.15-8.86)	4.74	(0.49-45.6)	0.179	9.82	(0.35-274)	0.179
		Other Lactobacilli	4.79	(2.37-5.84)	5.32	(4.12-6.04)	5.34	(4.01-6.06)	5.20	(5.10-5.44)	1.09	(0.64-1.85)	0.756	1.03	(0.55-1.93)	0.916

*Table 7.8 notes: 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. Concentration data quoted in log10 cells/µl.*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 analysis even if a taxa wasn't present in that sample (with concentration =0). ¶VMB type definitions: Li=L. iners-dominated (≥75% lactobacilli with L. iners the most common); Lcr=L. crispatus-dominated (≥75% lactobacilli with L. crispatus the most common); Lo+BL=other 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).*

Table 7.9: P-values for mean bacterial load comparisons between VMB types in high-risk women.

	N	Lcr	Li	Lo+BL	LA+BV
***L. crispatus*-dominated (Lcr)**	23		0.475	0.946	0.008
***L. iners*-dominated (Li)**	36			0.361	0.033
Other *Lactobacillus* - or *Bifidobacterium*-dominated (Lo+BL)	22				0.007
***Lactobacillus* with anaerobes or BV (LA+BV)**	28				

Table 7.9 notes: The p-values (by Mann-Whitney U test) are for mean bacterial load comparisons between VMB types in high-risk women (n=109). The mean bacterial loads for each VMB type are reported in Table 7.8

7.13 Correlation with clinical phenotype

The distribution of VMB types and participants in the highest quartile of bacterial load by clinical phenotype is shown in Table 7.10. Of note, both participants with early PPROM and early chorioamnionitis had bacterial loads in the highest quartile, as did the participant with early sPTB and chorioamnionitis. Conversely, the two participants who had early PPROM and developed chorioamnionitis more than 7 days after PPROM (as described in Table 3.2) did not have bacterial loads in the highest quartile. Visually, there did not appear to be a clustering of VMB types with a particular phenotype.

Table 7.10: Distribution of VMB types and top quartile of bacterial load by clinical phenotype of early sPTB/PPROM

Contributing factors at birth	High-risk participants PPROM 16^{+0}-33^{+6} weeks With early (within 7 days) chorioamnionitis n=2	Without early chorioamnionitis n=10	sPTB 16^{+0}-33^{+6} weeks n=10
None	0	LCr $LCr^{Bact\ cells}$ Li $LA+BV^{Bact\ cells}$ Lo+BL Lo+BL	LCr $Li^{Bact\ cells}$ $Li^{Bact\ cells}$ $Li^{Bact\ cells}$ $LA+BV^{Bact\ cells}$
Chorioamnionitis	$LA+BV^{Bact\ cells}$ $Lo+BL^{Bact\ cells}$	Li Lo+BL	$Li^{Bact\ cells}$
Placental dysfunction	0	0	$Li^{Bact\ cells}$ $Lo+BL^{Bact\ cells}$
Extra amniotic infection	0	0	0
Polyhydramnios	0	0	$Li^{Bact\ cells}$
Uterine anomaly	0	$Li^{Bact\ cells}$	0
Maternal comorbidities	0	$LCr^{Bact\ cells}$	$Li^{Bact\ cells}$
Cervical insufficiency	0	$LCr^{Bact\ cells}$ Li Lo+BL	LCr $Li^{Bact\ cells}$
Multiple contributing factors	0	Li $LCr^{Bact\ cells}$	$Li^{Bact\ cells}$

Table 7.10 notes:Li=L. iners-dominated (≥75% lactobacilli with L. iners the most common); LCr =L. crispatus-dominated (≥75% lactobacilli with L. crispatus the most common); Lo+BL=other lactobacilli- or Bifidobacterium dominated (either≥75% lactobacilli with L. jensenii or L. gasseri the most common, or ≥50% Bifidobacterium); lactobacilli and anaerobes (LA+BV; either LA, 25%- 75% lactobacilli, without ≥50% Bifidobacterium or BV, mixture of BV-anaerobes (<25% lactobacilli).) Each participant is represented by the symbol for their VMB type, and if that participant also had a concentration of bacterial cells in the top quartile then the VMB type has the superscript 'Bact cells'.

7.15 Characteristics associated with recurrent early preterm birth

The VMB characteristics identified throughout Chapter 7 as potentially associated with recurrent sPTB/PPROM in the high-risk group are:

1) Highest quartile of total bacterial load
2) Highest quartile of total *Lactobacillus* concentration
3) Highest quartile of *L. iners* concentration
4) Highest quartile of *L. crispatus* concentration
5) Presence of *Ureaplasma* species
6) Highest quartile of *Megasphaera* species
7) Presence of BVAB TM7-H1

18/22 (81.2%) of HR early sPTB/PPROM and 63/87 (72.4%) of HR term participants had at least one VMB characteristic associated with early sPTB/PPROM recurrence (Figure 7.11).

All high-risk early sPTB/PPROM patients in the highest quartile of *Lactobacillus* concentration were also in the highest quartile of bacterial load (n=14) (Table 7.7). The majority (25/28, 89.2%) of high-risk term participants in the highest quartile of *Lactobacillus* concentration were also in the highest quartile of total bacterial load. Over 97% of participants in all pregnancy outcome groups had *Lactobacillus* present, and it was by far the most common species identified in the population, with a median relative abundance over 98% in all pregnancy outcome groups (Table 7.5). Therefore, as expected, the most common species, Lactobacillus, contributes to the most to the total bacterial load.

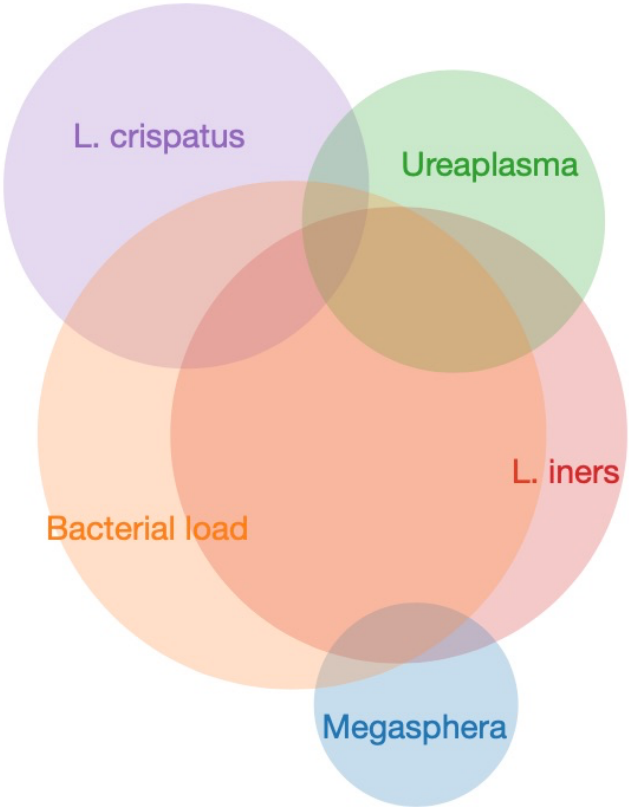
In both the high-risk term and the high-risk early sPTB/PPROM there was a high degree of overlap between highest quartile of *L. iners* and total bacterial load (Figure 7.11).

This is consistent with the findings in Table 7.8 which showed that women with domination of *L. iners* (VMB type Li) had higher bacterial loads. Eight high-risk participants with sPTB/PPROM had the VMB type Li (Table 7.8). These participants had particularly high total bacterial loads (median 9.03, IQR 8.02-9.12 log_{10} cells/µl); this is over 10-fold higher than the low-risk term birth participants with the same domination of *L. iners* (median bacterial load 7.52, IQR 6.90-8.12 log_{10} cells/µl).

Compared to the over-lap with *L. iners* there was less overlap between highest quartile of *L. crispatus* and highest quartile of total bacterial load (Figure 7.11). Only 57% (4/7) of high-risk early sPTB/PPROM participants with the highest quartile of *L. crispatus* also had highest quartile of total bacterial load, and 37.5% (6/16) of high-risk term birth participants shared these traits.

When *L. crispatus* was dominant there were very low concentrations of *L. iners* (median concentration of *L. iners* 0 for all pregnancy outcome groups in Table 7.8). Conversely when *L. iners* was dominant *L. crispatus* was also present (median concentration of *L. crispatus* in *L. iners* dominated pregnancies 5.89 and 6.19 log_{10} cells/µl in the low and high-risk groups respectively). This appears to account for the relationship between quintile of *L. crispatus* and recurrent sPTB/PPROM when the whole high-risk population is assessed (Table 7.7), which is no longer seen within the *L. crispatus* dominant VMB type (aOR of recurrent sPTB/PPROM in the high risk group 1.31, 95% CI 0.40-4.27, p=0.658, Table 7.8).

High risk term n=63



High risk early sPTB/PPROM n=18

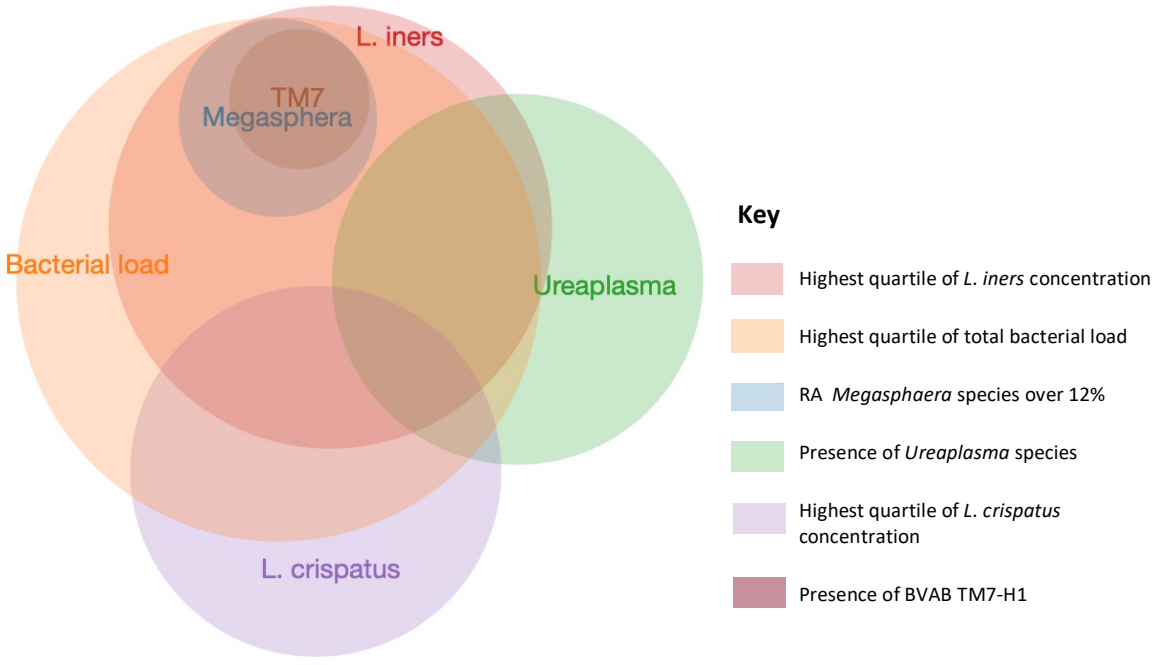


Figure 7.11: Euclan diagrams showing the co-location of VMB characteristics that were positively associated with early sPTB/PPROM within the high-risk cohort. 18/22 (81.2%) of HR early sPTB/PPROM and 63/87 (72.4%) of HR term participants had at least one VMB characteristic associated with early sPTB/PPROM recurrence. The highest quartile of Lactobacillus was omitted from this diagram because of almost complete overlap with highest quartile of bacterial load (see section 7.9.1) RA, relative abundance.

The proportion of high-risk participants with Ureaplasma present and other VMB features associated with sPTB/PPROM was similar for both the high-risk term (6/11, 54.5%) and high-risk preterm groups (3/7, 42.9%) (Figure 7.11). This suggests that presence of Ureaplasma does not closely correlate with other VMB characteristics associated with sPTB/PPROM, and it may be acting independently to increase the risk of sPTB/PPROM.

Two (9.1%) of high-risk early sPTB/PPROM participants had a high relative abundance of Megasphaera species (defined as over 12% based on the distribution in the low-risk term group, Figure 7.6). Both these participants were also in the highest quartile of bacterial load and *L. iners* concentration and one also had BVAB-TM7 present (Figure 7.11). Both of these participants had anaerobic dysbiosis (one VMB type BV and one LA).

Five (5.75%) high-risk term birth participants had a high relative abundance of Megasphaera species, all of whom also had anaerobic dysbiosis (four VMB type BV and one LA). However, only one of these participants also had the combination of highest quartile of bacterial load and *L. iners* (Figure 7.11). Consistent with Table 7.8 and Table 7.9 it appears that anaerobic dysbiois is associated with an overall higher bacterial load, but it is when the bacterial load is particularly high that the risk of recurrence of sPTB/PPROM appears to be the raised.

7.16 Characteristics protective of term birth

Two VMB characteristics were potentially associated with term birth:

1) Presence of Bifidobacterium
2) Presence of *Dialister* species

The presence of *Dialister* species was associated with term birth in this analysis (Table 7.5). This is contrary to previous work showing this taxa associated with PTB/PPROM (Table

4.3).[100,103,165,168] The assessment of taxa associated with sPTB/PROM was designed as a confirmatory analysis based on previous literature (section 3.10.4). On the basis of this association being in the opposite direction to previous work, and no longer significant after correction for multiple comparisons the research team deemed this to be likely to be a type-1 statistical error and no further analysis of this characteristic was carried out.

The only other VMB characteristic associated with a reduction in risk of early sPTB/PPROM was the presence of *Bifidobacterium breve.* This taxon was totally absent from the group with early sPTB/PPROM. Within the low-risk term birth group and the high-risk term birth group women with *Bifidobacterium breve* present tended to have a lower concentration of total *Lactobacillus*; only 1.38% (2/145) of the low-risk group had *Bifidobacterium breve* present and were in the highest quartile of total *Lactobacillus*, compared to 5.52% (8/145) of in the lowest quartile of total *Lactobacillus* (Table 7.11). However less than 10% of these groups had *Bifidobacterium breve* present overall.

Table 7.11: Number of pregnancies with Bifidobacterium present by quartile of total Lactobacillus concentration

			Low-risk pregnancy (LR) Birth ≥ 39 weeks n=145	High-risk pregnancy (HR) Birth ≥ 37 weeks n=87
		Concentration of total Lactobacillus (log10 cells/µl)	n (%)	n (%)
Quartile of total Lactobacillus	1	<6.42	8 (5.52)	1 (1.15)
	2	≥6.42- >7.31	3 (2.07)	3 (3.45)
	3	≥7.31- >8.01	1 (0.69)	3 (3.45)
	4	≥8.01	2 (1.38)	0 0.00
Absence of *Bifidobacterium breve*		not applicable	131 (90.34)	80 (91.95)
Fisher's exact test P value			0.043	0.185
Total			145 (100.00)	87 (100.00)

Table 7.11 notes: P value shows that within the low-risk group there was a significant association between presence of Bifidobacterium and quartile of total Lactobacillus (p=0.043).

7.17 Effect of antimicrobials

In order to assess whether the use of an antimicrobial in pregnancy, prior to VMB sampling, affected the bacterial load an assessment was made of bacterial load according to antimicrobial usage in pregnancy prior to VMB sampling. 20.1% of low-risk and 24.8% of high-risk participants had used antimicrobials in the current pregnancy (Table 7.12). The most common indication for antibiotic prescription was urinary tract infection, accounting for 11/21 (52.4%) of the low-risk group prescriptions and 5/12 (41.7%) of the high-risk group prescriptions. The majority of participants received cefalexin for treatment of urinary tract infections, with two participants in each group receiving nitrofurantoin, and two low-risk participants receiving trimethoprim (before pregnancy was known about). The other common antibiotic used was amoxicillin; prescribed for chest, ear and dental infections (8/21, 38% of low-risk prescriptions and 3/12, 33% of high-risk prescriotions). Only high-risk participants had received treatment targeted at bacterial vaginosis (5/12, 41.7%

participants, one of whom also received cephalexin for a UTI) in the form of either oral metronidazole (n=1) or vaginal clindamycin cream (n=4). A single low-risk participant received vaginal clotrimazole for symptomatic vaginal candidiasis. No association was noted, in either the high or low-risk pregnancy groups, between bacterial load and use of antimicrobials (whether targeted at bacterial vaginosis, or not) (Table 7.12).

Table 7.12: Bacterial load (in log_{10} cells/µl) based on antimicrobial use in the pregnancy prior to VMB assessment.

	Low-risk				High-risk				P value of difference in bacterial load compared to no antimicrobial	
	n	Percentage of participants	Bacterial load (log10 cells/µl)		n	Percentage of participants	Bacterial load (log10 cells/µl)			
			Median	IQR			Median	IQR	Low-risk	High-risk
No antimicrobial	115	79.9	7.69	6.72-8.41	82	75.2	7.96	7.28-8.95	Comparator	Comparator
Antimicrobial not clindamycin or metronidazole	21	14.6	7.60	6.75-8.45	12	11.0	8.19	7.29-8.77	0.940	0.991
Clindamycin or metronidazole	0	0.0			5	4.6	8.47	8.06-8.64		0.412
Not known/unsure	8	5.6	7.42	7.18-7.65	10	9.2	7.42	6.97-8.92	0.460	0.452
Total	144				109					

Table 7.12 notes: A single low-risk participant who used clotrimazole prior to VMB sampling is excluded from this analysis. One high-risk participant received multiple antimicrobials including metronidazole and was included in the clindamycin or metronidazole group only. P value calculated by Mann-Whitney U test.

7.18 Relationship with cervical length

Visual assessment was performed to assess whether there was a relationship between cervical length and bacterial load.

Visually no association was noted between bacterial load and cervical length at either the time of VMB assessment (15^{+1}-23^{+0} weeks gestation, Figure 7.12), or at approximately 20 weeks gestation days gestation (19^{+0}-21^{+6} weeks, Figure 7.13).

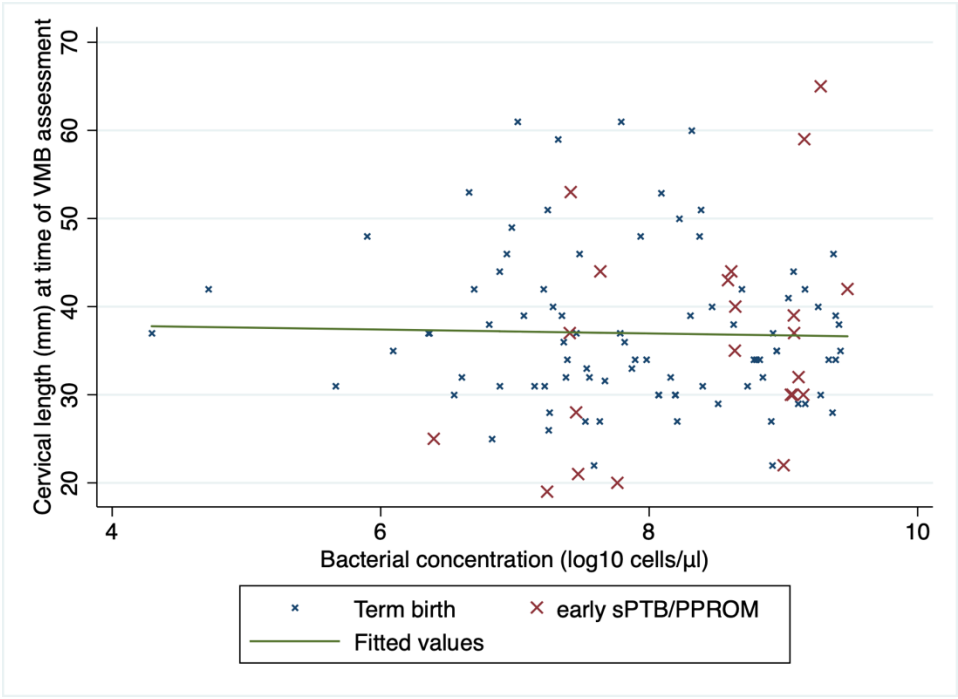


Figure 7.12: Relationship between bacterial load and cervical length at time of VMB assessment (15^{+1}-23^{+0} weeks gestation) gestation in the high-risk group. High-risk term n=87, high-risk preterm n=22. PTB, preterm birth.

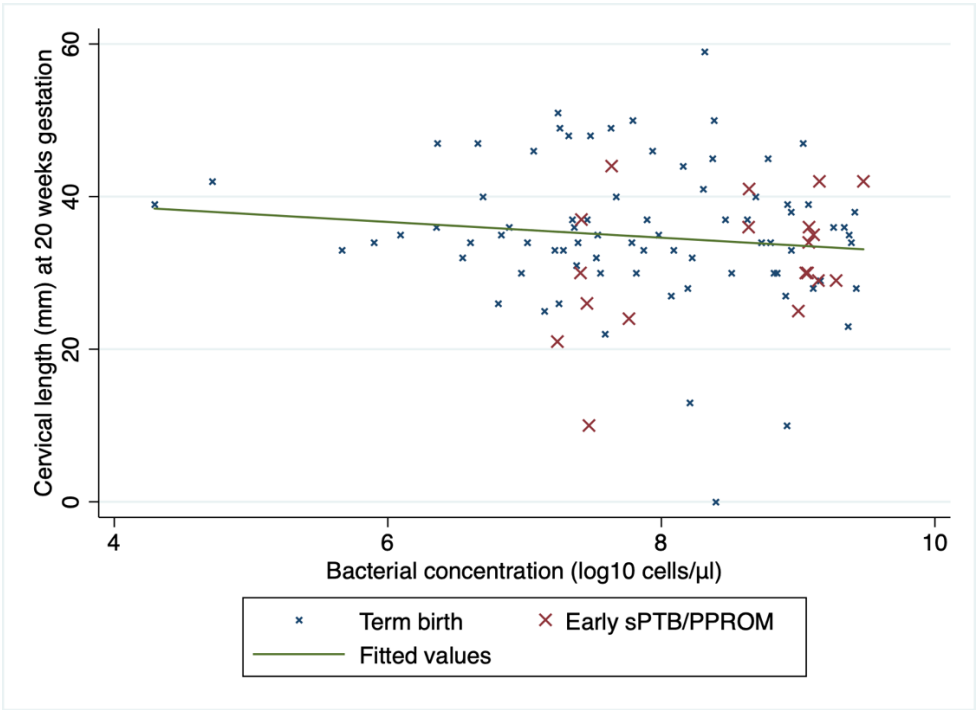


Figure 7.13: Relationship between bacterial load and cervical length at approximately 20 weeks (19^{+0}-23^{+0} weeks gestation) gestation in the high-risk group. High-risk term n=79, high-risk preterm n=19 (participants without cervical length assessment at 20 weeks omitted). PTB, preterm birth.

7.19 Discussion

We identified a positive association between the total vaginal bacterial load, predominantly comprised of *Lactobacillus*, and early sPTB/PPROM recurrence. Women with a recurrence, compared to those without, had a fivefold higher median vaginal bacterial load (after inverse $\log_{10}$ transformation). There was a gradient effect with a 2.61-fold increased adjusted odds of recurrence for each increase of total *Lactobacillus* concentration quartile.

When we went on to divide our population by 'VMB type'; a classification system based on the dominant species of lactobacilli, or non-dominance of lactobacilli. Particularly high overall bacterial loads were present in participants with recurrent sPTB/PPROM and dominance of *L. iners,* or non-dominance of lactobacilli. The high bacterial loads were comprised of high concentrations of lactobacilli, along with high concentrations of BV associated bacteria and 'other' bacteria (as defined in Appendix G). The groups were small and after controlling for confounders the associations did not reach pre-specified statistical significance. The VMB types dominated by *L. crispatus*, other *Lactobacillus* and *Bifidobacterium*, had slightly higher total bacterial loads and concentrations of *Lactobacillus* in the early sPTB/PPROM group compared to the high-risk term birth group, but the differences were not as striking and were much further from statistical significance. High-risk and low-risk participants who had term births had similar bacterial loads within VMB types.

We postulate that some women with *L. iners* domiance*,* or non-dominance of lactobacilli, reach higher overall bacterial loads than women dominated by *L. crispatus* or other species of lactobacilli, or conversely the later women are protected from higher bacterial loads. The combination of dominance of *L. iners* or non-dominance of lactobacilli, along with a high bacterial load, shows stronger association with recurrent sPTB/PPROM. This would be consistent with previous work suggesting an association between the relative abundance of

L. iners and sPTB.[101,102] This would also explain the lack of consistency in the reproduction of this finding (Table 4.3); if *L. iners* dominance is only associated with PTB when there is a concurrent high bacterial load, then women with *L. iners* domination and a normal bacterial load would not show this association. We also postulate that the women who have high bacterial loads and do go on to have sPTB/PPROM could be at higher risk of early onset chorioamnionits (Table 7.10).

Brown *et al* assessed the relationship between bacterial load and PPROM prior to 37 weeks gestation.[172] This work showed no difference in bacterial load at 8-12 weeks gestation between pregnancies destined for PPROM $<37^{+0}$ weeks and pregnancies destined for term birth. This study is not directly comparable to the current work because the gestation of sampling was earlier and the gestation of PPROM later; giving a longer latency period between sampling and event. Brown *et al* also found no difference in bacterial load after PPROM when comparing women with PPROM who did, and did not, go on to develop chorioamnionitis with funisitis. The current study's equivalent definition of the chorioamnionitis with funisitis cases could be the cases with early onset chorioamnionitis (Table 3.2). We only had three early sPTB/PPROM cases meeting this definition, but interestingly all of these were in the highest quartile of bacterial load in the second trimester (Table 7.10). This difference in findings is not surprising; in the Brown *et al* study the bacteria contributing to PPROM may have already ascended into the uterine cavity in order to contribute to the PPROM,[203] and after the rupture of membranes there will be a substantial shift in vaginal fluids, so the crucial time when the high bacterial load contributes to the pathology may be prior to the PPROM.

To our knowledge, assessment of the association between the vaginal bacterial load and PTB has been applied in two other studies. Freitas *et al*[168] found a higher concentration of 16S rRNA genes per swab (taken at 16 weeks gestation) in women who went on to have an

sPTB <37 weeks compared to a term birth ≥37 weeks (8.1 vs. 7.8 $\log_{10}$ copies/swab, respectively). Elovitz *et al*[101] found similar concentrations of 16S rRNA genes in women who went on to have sPTB <37 weeks and term births ≥37 weeks when they assessed samples taken at less than 24 weeks gestation (approximately 8 $\log_{10}$ copies/swab in each group), and found a lower concentration (approximately 7.5 vs 7.9 $\log_{10}$ copies/swab) of 16S rRNA genes in women who went on to have a PTB in samples taken at 28 gestation. This may indicate that high vaginal bacterial load is only associated with PTB when present prior to 24 weeks gestation, but additional studies are needed to replicate these findings.

Bacteria have between one and fourteen 16S rRNA genes per bacterial cell, and the average gene copy number per cell varies by taxon (Appendix G). The concentration of 16S rRNA genes in a vaginal swab is therefore not directly analogous to the concentration of bacterial cells (bacterial load) in that swab, but only provides a crude estimate. We corrected our bacterial load estimations by average 16S gene copy number per cell for each taxon present in the sample in an effort to improve the estimations.

In non-pregnant women, a higher vaginal bacterial load is usually associated with bacterial vaginosis.[176] Pregnant women are protected from bacterial vaginosis due to the high oestrogen levels during pregnancy.[204,205] Oestrogen increases vaginal glycogen stores in the vagina, which provide *Lactobacillus* with a competitive advantage compared to bacterial vaginosis-associated bacteria. *Lactobacillus* very efficiently convert glycogen into lactic acid, creating an acidic vaginal environment in which they thrive and most competing bacteria do not.

In our study, the prevalence of anaerobic dysbiosis, and the relative abundances and estimated concentrations of bacterial vaginosis-associated bacteria and pathobionts, were indeed low. In the high-risk group, the median estimated concentration of BV-associated bacteria (6.76 $\log_{10}$ cells/µl) was 68 times lower, and the median concentration of

pathobionts (4.81 $\log_{10}$ cells/µl) more than 6,000 times lower, than the median estimated concentration of *Lactobacillus* (8.59 $\log_{10}$ cells/µl). These very low concentrations of BV-associated bacteria and pathobionts in the pregnant women in our cohorts may explain why we did not identify any associations between these bacterial groups and pregnancy outcome.

We found that *L. iners* domination was associated with higher bacterial loads than the other lactobacilli, and the combination of *L. iners* domination and high bacterial load was associated with early sPTB/PPROM recurrence. *L. iners* was first described in 1999.[206] It has a genome of just 1.3 Mpb,[207] strikingly smaller than the other common vaginal species of lactobacillus, for example *L. crispatus*, with a genome of about 2.3 Mbp.[208] Species with such small genomes are normally dependent on exogenous sources for vital nutrients, and therefore either have a symbiotic or parasitic relationship with their host.[209] However, which of these *L. iners* holds is still debated.[210] Additionally, there are multiple strains of most bacterial species. The technology to differentiate bacteria to strain level is relatively new,[211] but a recent study of 195 pregnant women in the USA found 21 different strains of *L. iners* present throughout their population, with 19% of participants having more than one strain of *L. iners* present at any one time.[212]

Macklaim *et al* observed that that *L.iners* strains with high expression of CRISPR (clustered regularly inter- spaced short palindromic repeat) genes were present in bacterial vaginosis, but not in healthy conditions.[213] CRISPR genes are the primary bacterial defence against phages (viruses that infect bacteria). This could suggest that the *L.iners* associated with high bacterial loads and recurrent sPTB/PPROM in our study may be different strains, or behaving in a different way, to the *L.iners* associated with lower bacterial loads in the high-risk term birth and low-risk pregnancy groups.

We had postulated that high-risk women, given the nature of their previous poor obstetric

history may seek more healthcare and be more likely to have used antimicrobials in pregnancy. This in turn could have had an effect on the bacterial load. We did find that slightly more high-risk participants had used antimicrobials in pregnancy prior to their study visit compared to low-risk participants (24.8% vs. 20.1%, Table 7.12). However, no difference was shown in the bacterial load between women who had used clindamycin or metronidazole (antibiotics recommended for treatment of bacterial vaginosis[202]), other antibiotics, or no antibiotics in pregnancy. There are some limitations with this analysis. We relied on participant recall only, the time when the antibiotics were administered was inconsistently recorded and could not be accounted for in the analysis, and 8/144 (5.5%) of low-risk and 10/109 (9.2%) of high-risk participant has missing data. Nevertheless, the proportions of participants using antimicrobials in pregnancy were similar for the high and low-risk groups, and there was no difference in bacterial load by antimicrobial usage. Therefore, the impact of antimicrobials on the bacterial loads in this study (and in turn the association identified between bacterial load and early sPTB/PPROM) is unlikely to be accounted for by differential antimicrobial usage in the highest risk pregnancies.

In agreement with previous work[95,97] we found that the VMB was relatively stable between study visits (approximately 16 and 20 weeks gestation). Like Fettweis *et al*[100] we did not identify any association between VMB instability and PTB. This VMB stability in pregnancy means that, if a VMB characteristic were to be associated with PTB, there is a window in which an intervention to target this association could be applied.

Three uncommon taxa (present in fewer than 20% of the women in the low risk group) showed non-significant trends towards an association with early sPTB/PPROM recurrence in our study and were also associated with PTB in studies by others: the presence of *Ureaplasma*[162,168] and BVAB TM7-H1[100], and the relative abundance of *Megasphaera.*[101,168] In women with early sPTB/PPROM recurrence, the presence of these taxa in the vagina did

not completely overlap with having high bacterial loads and lactobacilli concentrations. While uncommon, the presence or relative abundance of these types of BV-associated bacteria may therefore also play a role in PTB, for example by contributing to inflammatory pathways.[21] We also found an association in the opposite direction to previous research,[103,110,168,214] an increased prevalence of *Dialister* in high-risk participants with a term birth compared to recurrent sPTB/PPROM. This is likely to be a type-1 statistical error (false positive finding).

Our data confirms that *Bifidobacterium breve* may have a protective effect against early sPTB/PPROM recurrence.[103] This helped guide our decision to place women with *Bifidobacterium* ≥50% relative abundance within a 'VMB type' with lactobacillus.

With regards to the remainder of the taxa of interest, as identified in Table 4.3, this study did not show an association between these taxa and early sPTB/PPROM. Most of these are rare taxa, also described as minority species, and comprise less that 1% of the bacterial species in a sample. Minority species are liable to under replication in the PCR steps necessary for 16S rRNA sequencing.[107] This because the genetic material from species that make up proportionally more of the microbiome is a able to out-compete the genetic material of the rarer taxa for the limited sequencing reagents added to each sample. As such 16S rRNA sequencing has been shown to have good positive predictive value, but poorer negative predictive value for minority species when compared to culture based assays.[215] Additionally minority taxa can be less well characterised within sequencing libraries, and so at higher risk of taxonomic misassignment.[107] Finally different study protocols, including DNA extraction methods and primer sets may be better, or worse at amplifying particular microbial species. Therefore, whilst notable that the remainder of the taxa of interest did not show an association with early sPTB/PPROM in this study, this methodology (even if we had an appropriate sample size and power calculation) is not the

optimal mode of ruling out an association. Panels of PCR based assays targeted at the taxa of interest could be developed to better address this whether there is an association between these taxa and early sPTB/PPROM.

Short cervical length in pregnancy, particularly in women with a previous preterm birth, was the predominant clinical parameter used to initiate preterm birth prevention treatment during the course of the study. [26,216,217] The relationship between cervical length and preterm birth has the strongest evidence after 18 weeks gestation,[218] we therefore visualised the relationship between bacterial load and cervical length at both the study visit used for the majority of the analysis (median 16 weeks gestation, Table 5.3), and approximately 20 weeks gestation. No association was identified between cervical length and bacterial load in the high-risk participants at either time point.

The cervical length of high-risk participants in this study is a little unusual because women with treatment for a short cervix at study initiation (approximately 16 weeks gestation) were not eligible for recruitment. Therefore we have may have preferentially recruited women with a longer cervix, in whom the contribution to preterm birth of cervical insufficiency may be less marked. This may be demonstrated by the finding of less women in the early PTB group having received preterm birth prevention treatment after recruitment than the high-risk term birth group (conversely the high-risk early sPTB/PPROM group may have had a short cervical length that our current practice did not detect).

Possible explanations for the lack of association between short cervical length and bacterial load are either: a) this group of participants were not the predominant group with a cervical contribution to sPTB/PPROM, and so we did not have the power to detect an association which is present or b) bacterial load may act independently of cervical length and contribute to the approximately 9% recurrence rate of early sPTB/PPROM in women who maintain a normal cervical length.[30]

7.19.1.1 *Strengths and limitations*

We have identified a recurrent early sPTB/PPROM phenotype that is as pure as possible, and we applied methods of VMB data reduction that had been shown to have an association with PTB previously. In addition, we have developed and applied new methods of data reduction, that incorporate concentration estimation, for the first time in this area of research. These methods overcome some of the previous criticisms of using standard biostatistical methods for compositional data.[219] Furthermore, recent work with human faecal samples found a stronger microbe- metabolite relationship when bacterial taxa concentrations were estimated by combining relative abundances with 16S rRNA qPCR compared to relative abundances alone.[220]

We had limited statistical power due the small number of women with early sPTB/PPROM recurrence. This means that not all the associations consistent with our conclusion reached statistical significance. We had initially planned to analyse early sPTB and early PPROM, which may have different aetiological pathways,[116] as two separate endpoints but these were combined to preserve statistical power. In addition, we chose to retain women who used PTB prevention treatment (cervical cerclage, Arabin pessary, or vaginal progesterone) after their 16 and/or 20-week vaginal sample collection in our analysis. This approach has recently been advised against[182] as it is possible that women destined to deliver preterm may have had their delivery delayed by the therapy, weakening the ability of our analysis to detect a relationship between the VMB and sPTB/PPROM. However, any future therapy developed based upon VMB analysis is likely to be applied in combination with current treatments, and so by retaining these participants we were able to assess the contribution of the VMB to recurrent sPTB/PPROM within current clinical practice.

The primary way in which the VMB is thought to affect the risk of PTB is female genital tract inflammation.[221] Another common inflammatory vaginal condition is vulvovaginal

candidiasis, for which pregnant women are at increased risk.[222] Unfortunately the current study did not assess vaginal yeasts concentrations.

We were not able to collect data about antimicrobial use after involvement in the study because of the wide range of sources that the participants could have obtained these from. This includes hospitals, primary care practitioners, walk-in centres and, in the case of antifungals, over the counter purchases. A final limitation is the lack of VMB data in the third trimester and/or closer to the birth.

7.20 Conclusion

Among women who had a previous early sPTB/PPROM, recurrence was associated with increased total vaginal bacterial load, which was mostly made up of lactobacilli. Women with a previous early sPTB/PPROM were especially likely to have a recurrent event if they had a VMB dominated by *L.iners* or anaerobic dysbiosis and a high bacterial load. 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.

8 Association of vaginal microbiota with late preterm birth

The primary focus of this thesis was to assess the association between the VMB in the second trimester and recurrent early sPTB/PPROM. Our findings suggest that eight VMB characteristics may be associated with recurrence (chapter 7), the most striking being a high vaginal bacterial load, predominantly comprised of lactobacillus.

Women with a previous early sPTB/PPROM who avoid a recurrent event have previously been shown to be at an increased risk of late sPTB/PPROM compared to parous women without previous sPTB/PPROM.[223] This study concurred with the previous work; 10.2% (14/137) of high-risk women had a late sPTB/PPROM, compared to only 0.9% (2/227) of low-risk women (Figure 5.1). This chapter takes the opportunity provided by the these participants to assesses whether the VMB characteristics described in chapter 7 are also associated with late sPTB/PPROM in our high-risk population.

The authors are mindful that a large number of comparisons have already been performed throughout this process, and to repeat the analysis within the late sPTB/PPROM group would further increase our chance of false-positive results. Therefore, a targeted analysis was performed focusing only on the contribution the VMB characteristics identified in chapter 7 to late sPTB/PPROM. We chose to maximise statistical power by combining the outcomes of late sPTB and late PPROM, based on the work in the early sPTB/PPROM groups (section 7.3).

8.1 Demographics of participants with late preterm birth

The high-risk late sPTB/PPROM group had similar demographics to the high-risk early sPTB/PPROM group (Table 8.1). When compared to the high-risk term birth group those

with a late sPTB/PPROM were more likely to have had ≥2 previous sPTB/PPROM pregnancies (42.9% vs 9.2%).

8.2 Association between vaginal microbiota characteristics and late preterm birth

8.2.1 Distribution of vaginal microbiota characteristics associated with early preterm birth in the late preterm birth group

The distribution of these VMB characteristics by pregnancy outcome, comparing the late sPTB/PPROM group to both the other high-risk groups are shown in Table 8.2. There was no statistically significant difference between the distribution of the VMB characteristics assessed in the high-risk term and high-risk late sPTB/PPROM groups. The only non-significant trend noted was the presence of BVAB TM7-H1 in a single high-risk late sPTB/PPROM participant, and no high-risk term birth participants.

When comparing the early and late sPTB/PPROM groups the only statistically significant differences noted were more participants in the top quartiles of bacterial load and concentration of *Lactobacillus* in the early sPTB/PPROM groups (63.6% vs. 21.4%, p=0.019 and 63.6% vs. 14.3%, p=0.006) (Table 8.2).

Table 8.1: Participant characteristics by pregnancy outcome group.

		Low-risk pregnancy (LR)		High-risk pregnancy (HR)						P value late sPTB/PPROM	
		Term birth n=145		Term birth n=87		Late sPTB/PPROM n=14		Early sPTB or PPROM n=22		vs. HR term	vs. HR early sPTB/PPROM
Age (years)	mean (SD)	31.1	(4.5)	30.0	(4.5)	30.4	(5.0)	31.4	(5.4)	0.750	0.589
BMI (kg/m2) *	median (IQR)	24.5	(22-29)	25	(22-28)	25.5	(21-30)	27.5	(22-35)	0.679	0.505
Current smoker *	number (%)	15	(10.6)	15	(17.4)	6	(42.9)	5	(23.8)	0.069	0.283
Ethnicity (%)	White	139	(95.9)	81	(93.1)	12	(85.7)	22	(100.0)	0.306	0.144
	Black	3	(2.1)	4	(4.6)	1	(7.1)	0	0.0		
	Asian	0	0.0	0	0.0	0	0.0	0	0.0		
	Other	2	(1.4)	0	0.0	0	0.0	0	0.0		
	Not recorded	1	(0.1)	2	(2.3)	1	(7.1)	0	0.0		
Index of multiple deprivation score quintile (%)	1 (most deprived)	74	(51.0)	50	(57.5)	8	(57.1)	15	(68.2)	0.477	0.865
	2	19	(13.1)	6	(6.9)	3	(21.4)	2	(9.1)		
	3	24	(16.6)	9	(10.3)	1	(7.1)	2	(9.1)		
	4	20	(13.8)	11	(12.6)	1	(7.1)	2	(9.1)		
	5 (least deprived)	7	(4.8)	6	(6.9)	0	0.0	0	0.0		
	Not recorded	1	(0.7)	5	(5.7)	1	(7.1)	1	(4.5)		
Number of previous PPROM or sPTB <34 weeks (%)	1			79	(90.8)	8	(57.1)	15	(68.2)	0.004	0.723
	≥ 2			8	(9.2)	6	(42.9)	7	(31.8)		
Previous cervical surgery *	Nil significant	131	(90.3)	79	(90.8)	11	(78.6)	17	(77.3)	0.080	1.000
	Single LLETZ	14	(9.7)	7	(8.0)	1	(7.1)	2	(9.1)		
	Multiple LLETZ or knife cone			1	(1.1)	2	(14.3)	3	(13.6)		
Gestational age at sample (weeks)	median and range**	16+5	(16+0-21+2)	16+2	(15+1-21+6)	16+3	(15+4-20+6)	16+3	(15+2-19+4)	0.871	0.935
Preterm birth prevention treatment used after study visit	None	139	(100.0)	61	(70.1)	8	(57.1)	17	(77.3)	0.215	0.393
	Cervical cerclage			4	(4.6)	1	(7.1)	1	(4.5)		
	Progesterone			2	(2.3)	2	(14.3)	1	(4.5)		
	Arabin pessary			20	(23.0)	3	(21.4)	2	(9.1)		
	Progesterone and cerclage			0	0.0	0	0.0	1	(4.5)		
Gestational age at PPROM (weeks+ days)	median and range					35+6	34+0-36+6	31+1	18+0-33+6	not applicable	not applicable
Gestational age at birth (weeks+days)	median and range	40+1	(39+0-41+6)	38+6	37+0-41+5	35+6	34+4-37+1	31+5	18+0-35+5		
Birthweight (g)*	mean (SD)	3594	439.0	3234	489.0	2602	379.0	1778	673.0		
GROW birthweight centile	median and IQR	44.7	(23.2-71.0)	35.5	(18.6-62.1)	24.6	(9.6-52.3)	36.9	(16.2-54.6)	0.337	0.484
Cervical length median and IQR	16 weeks	41	(36-47.1)	35.5	(31-42)	35	(29-40)	36	(26.5-43.5)	0.591	0.956
	20 weeks	41	(37-45)	35	(30-40)	31	(23-35)	30	(26-37)	0.018	0.465
qfFN median and IQR	16 weeks	7	(5-16)	7	(5-22)	12	(6-45)	9	(6-23.5)	0.298	0.696
	20 weeks	7	(5-12)	7	(5-14)	20	(7-42)	7	(6-12)	0.221	0.183

Table 8.1 notes: BMI has 1 missing value for low-risk and 1 for HR term birth. Smoking has 3 missing values for LR, and 1 each for HR term and early sPTB or PPROM . Cervical length at 16 weeks n=136 LR, n=80 for HR-term, n=13 HR late preterm, n=20 HR early preterm. CL at 20 weeks , LR n=135, HR term n=79, HR late preterm n=14, HR early preterm n=14. qfFN at 16 weeks LR n=142, HR term n=77, HR late preterm n=12, HR early preterm n=20. qfFN at 20 weeks LR n=136, HR term n=75, HR late preterm n=14, HR early preterm n=19. Birthweight has 2 missing values for HR term and 2 missing values for HR early sPTB or PPROM. GROW birthweight centile LR n=138, HR term birth n=83, HR late preterm n=14, and HR early sPTB/PPROM n=19. 16 weeks gestation is abbreviated for the first study visit that was carried out at 15^{+1}-18^{+6} weeks and 20 weeks gestation indicates the second study visit that was carried out at 19^{+0}-23^{+0} weeks P values calculated using 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.

Table 8.2: Distribution of VMB characteristics associated with early sPTB/PPROM by pregnancy outcome.

	Low-risk pregnancy (LR)		High-risk pregnancy (HR)						P value late sPTB/PPROM	
	Term birth n=145		Term birth n=87		Late sPTB/PPROM n=14		Early sPTB or PPROM n=22		vs. HR term	vs. HR early sPTB/PPROM
	n	(%)	n	(%)	n	(%)	n	(%)		
Top quartile of bacterial load	37	(25.5)	31	(35.6)	3	(21.4)	14	(63.6)	0.373	**0.019**
Top quartile of total Lactobacillus concentration	37	(25.5)	28	(32.2)	2	(14.3)	14	(63.6)	0.221	**0.006**
Highest quartile of *L. iners* concentration (if present)	20	(13.8)	25	(28.7)	4	(28.6)	10	(45.5)	1.000	0.485
Highestet quartile of *L. crispatus* concentration (if present)	32	(22.1)	16	(18.4)	4	(28.6)	7	(31.8)	0.469	1.000
Presence of Ureaplasma	26	(17.9)	11	(12.6)	1	(7.1)	7	(31.8)	1.000	0.115
Highest quartile of *Megasphera* relative abundance	8	(5.5)	5	(5.7)	0	0.0	2	(9.1)	1.000	0.511
Prescence of BVAB TM7-H1	2	(1.4)	0	0.0	1	(7.1)	1	(4.5)	0.139	1.000
Prescence of Bifidobacertium	14	(9.7)	7	(8.0)	2	(14.3)	0	0.0	0.608	0.144

Table 8.2 notes: P value by Fisher's exact test comparing the distribution of each VMB characteristic in the late sPTB/PPROM group and both the high-risk term birth group and the high-risk early sPTB/PPROM groups. Bold denotes p<0.05

8.2.2 Association between vaginal microbiota characteristics and clinical phenotypes of late preterm birth

There does not appear to be a pattern of distribution of the VMB types and highest quartile of bacterial load by clinical phenotype of late sPTB/PPROM (Table 8.3). Of potential interest all the participants with the highest quartile of bacterial load had cervical insufficiency. One of the participants with cervical insufficiency, VMB type LA+BL and the highest quartile of bacterial load was also the participant with BVAB-TM7 H1 present.

We modified our definition of clinical phenotypes from the work of Villar *et al* (2012).[141] Cervical insufficiency was defined as treatment for short cervical length offered prior to 28^{+0} weeks gestation. Our participants were not eligible for the study if they were using PTB prevention treatment at the time of potential enrolment, so in this study all participants who received treatment did so after VMB assessment. Three out of five (60%) of the participants who went on to have late sPTB/PPROM after PTB prevention treatment were in the top quartile of bacterial load, compared to none of the participants who had late sPTB/PPROM and did not have PTB prevention treatment (n=9) (Table 8.4). This difference is significant, p=0.028, Fisher's exact test. The three participants who had high bacterial loads and subsequent cervical shortening prompting initiation of preterm birth prevention treatment all had different preterm birth prevention treatments; one each of vaginal progesterone, cervical pessary and cervical cerclage.

The difference in initiation of preterm birth prevention treatment, which implies detected cervical shortening, in the late sPTB/PPROM group by quartile of bacterial load prompted similar assessments in the early sPTB/PPROM and high-risk term birth groups (Table 8.5 and Table 8.6). In these pregnancy outcome groups there was no significant association noted between bacterial load and subsequent use of PTB prevention treatment, although a trend

in the same direction as the late sPTB/PPROM group was seen in the high-risk term birth group.

Table 8.3: Distribution of VMB types and top quartile of bacterial load by clinical phenotype of late sPTB/PPROM

Contributing factors at birth	High-risk participants	
	PPROM 34^{+0}-36^{+6} weeks without early chorioamnionitis n=2	**sPTB 34^{+0}-36^{+6} weeks n=12**
None	Li	LCr LCr Li Li LA+BV Lo+BL
Chorioamnionitis	0	0
Placental dysfunction	0	LA+BV
Extra amniotic infection	0	0
Polyhydramnios	0	0
Uterine anomaly	0	0
Maternal comorbidities	0	0
Cervical insufficiency	LCr	LCr $LCr^{Bact\ cells}$ Li $LA+BV^{Bact\ cells}$ $LA+BV^{Bact\ cells}$
Multiple contributing factors	0	0

Table 8.3 notes :Li=L. iners-dominated (≥75% lactobacilli with L. iners the most common); LCr =L. crispatus-dominated (≥75% lactobacilli with L. crispatus the most common); Lo+BL=other lactobacilli- or Bifidobacterium dominated (either≥75% lactobacilli with L. jensenii or L. gasseri the most common, or ≥50% Bifidobacterium); lactobacilli and anaerobes (LA+BV; either LA, 25%- 75% lactobacilli, without ≥50% Bifidobacterium or BV, mixture of BV-anaerobes (<25% lactobacilli).) Each participant is represented by the symbol for their VMB type, and if that participant also had a concentration of bacterial cells in the top quartile then the VMB type has the superscript 'Bact cells'.

Table 8.4: : Distribution of high-risk participants who had late sPTB/PPROM by whether they had PTB prevention treatment after their study visit and quartile of bacterial load

	Top quartile bacterial load	Not top quartile bacterial load	Total n
PTB prevention treatment	3 (60%)	2 (40%)	5
No PTB prevention treatment	0	9 (100%)	9
Total	3	11	14

Table 8.4 statistics: p=0.028, Fisher's exact test.

Table 8.5: : Distribution of high-risk participants who had term birth by whether they had PTB prevention treatment after their study visit and quartile of bacterial load

	Top quartile bacterial load	Not top quartile bacterial load	Total n
PTB prevention treatment	12 (48%)	13 (52%)	25
No PTB prevention treatment	19 (30.6%)	43 (69.4%)	62
Total	31	56	87

Table 8.5 statistics: p=0.144, Fisher's exact.

Table 8.6: Distribution of high-risk participants who had early sPTB/PPROM by whether they had PTB prevention treatment after their study visit and quartile of bacterial load

	Top quartile bacterial load	Not top quartile bacterial load	Total n
PTB prevention treatment	2 (40%)	3 (60%)	5
No PTB prevention treatment	12 (70%)	5 (30%)	17
Total	14	8	22

Table 8.6 statistics: p=0.309, Fisher's exact

8.3 Discussion

No association was identified between the VMB characteristics associated with recurrent early sPTB/PPROM and late sPTB/PPROM in the high-risk group. On an individual level, there were three notable participants who had VMB characteristics that could be identified as similar to those that were associated with early sPTB/PPROM. These three participants had bacterial loads in the highest quartile, one also had BVAB TM7 present, and two were in the VMB group LA+BV (which was associated with particularly high bacterial loads in the early sPTB/PPROM group). All of these participants received PTB prevention treatment for short cervical length. It is possible that the PTB treatment averted the early sPTB/PPROM under 34^{+0} weeks gestation in these participants.

In order to assess whether high bacterial load in the second trimester was associated with subsequent PTB treatment initiation (implying subsequent cervical shortening) the association between highest quartile of total bacterial load and subsequent PTB treatment was assessed for all high-risk pregnancy outcomes. 60% of the high-risk late sPTB/PPROM participants who had subsequent PTB prevention treatment were in the top quartile of bacterial load, compared to none of the late sPTB/PPROM participants who didn't have PTB treatment. A similar trend was seen within the high-risk term birth group; 48% (12/25) of the participants who had subsequent PTB prevention treatment were in the highest quartile of bacterial load, compared to 30.6% (19/62) of participants who didn't have PTB prevention treatment. This difference did not reach statistical significance.

Within the high-risk early sPTB/PPROM group an opposite trend was noted, with more participants who didn't have PTB prevention treatment having a higher bacterial load (12/17, 70% vs. 2/5, 40%, Table 8.6). The 17 participants who had an early sPTB/PPROM without PTB prevention treatment are those that our current practice of cervical length surveillance didn't detect as having a short cervix, but they did nevertheless have an early

PTB/PPROM. This is consistent with previous work from our group showing that 9% of women with a previous sPTB/PPROM <34^{+0} weeks gestation have a recurrent event despite normal cervical length screening.[30] If the current work is replicable then assessment of bacterial load may be able to complement cervical length assessment in preterm birth risk stratification. Further work would then be required to assess whether this is also able to help with targeting preterm birth prevention treatment.

To our knowledge only a single study has previously assessed the association between the VMB by separating early and late PTB in the same population. Tabatabaei *et al*[106] also found VMB characteristics that were associated with early, but not late PTB, also in a predominantly Caucasian population. Despite different laboratory and bioinformatic techniques there are some similarities in the taxa of interest between this study and our own. Bifidobacterium species showed a similar distribution to our work; it was only present in a single participant, at low relative abundance with early sPTB/PPROM (n=1/17, 5.8%), but was present in 4/77 (5.2%) of participants with late sPTB/PPROM and about 27/356 (7.6%) of participants with term births. This is comparable to the 0%/14.3%/8.0% in our high-risk early sPTB/PPROM/late sPTB/PPROM /term groups respectively.

8.3.1 Conclusion

The VMB characteristics that this work identified as associated with early sPTB/PPROM did not replicate in the late sPTB/PPROM group. It was noted that the late sPTB/PPROM participants who were in the highest quartile of bacterial load all subsequently had cervical shortening and received PTB prevention treatment. It is possible that this PTB prevention treatment averted an early sPTB/PPROM in these participants.

9 Discussion and conclusion

This thesis assesses the relationship between the VMB in the second trimester and recurrent sPTB/PPROM. This chapter provides a summary and discussion of the key findings.

9.1 Vaginal microbiota characteristics of interest

Molecular assessment of the vaginal microbiota by next generation sequencing has only been possible in the past decade. Previous attempts to meta-analyse prior studies have been hampered by technical differences in every aspect of study design.[110,182] Therefore, instead of re-attempting meta-analysis we focused on the findings interpreted as associated with PTB, or term birth, by each previous research team. Through this method we were able to identify six global VMB characteristics and 18 taxa that had shown an association with preterm birth in at least two previous studies and we selected an additional five taxa of interest based on scientific plausibility (Table 4.3).

In addition we developed three VMB variables specifically for this study. The first variable was a modified version of the mutually exclusive VMB types that had been previously been correlated with Nugent score and risk of STI acquisition outside of pregnancy by our research group.[176] The modifications accounted for the expected increased relative abundance of *Lactobacillus* in pregnancy and corresponding reduction in BV-anaerobes and pathobionts. Secondly the distribution in each sample of broad groups of taxa was assessed by allocating each taxon to one of four 'bacterial groups', based on the published literature; lactobacilli; BV-anaerobes; pathobionts; and a rest group of 'other bacteria'. This allowed assessment of the percentage of each bacterial group in each sample. The final new variable was the concentration of taxa in each bacterial group, estimated by combining the BactQuant assay with the relative abundance of taxa in each bacterial group. This was

expressed as a continuous variable on a log scale$_{10}$, and also by quartile of the distribution in the low risk group.

Individual research groups have developed and used their own VMB characterisation methods across research studies,[103,172,173] and the community state type grouping[133] has been commonly used, with some adaptions. However, the optimal method of VMB assessment in pregnancy is yet to be determined, and no previous studies have focused on replicability of previous work from outside of their own research group. This study gave the opportunity to assess replicability of VMB characteristics associated with PTB by an independent group.

9.2 Vaginal microbiota in healthy pregnancies

We found that the 2nd trimester VMB was dominated by *Lactobacillus* in 72.5% of low-risk participants with term birth. The most common VMB type and CST was that dominated by *L. crispartus,* closely followed by the CSTs and VMB types with domination of *L. iners* and other species of *Lactobacillus.* These findings are consistent with other VMB assessments in the second trimester of pregnancies that have progressed to term in Caucasian populations.[97,101,106,110]

The VMB was stable for the majority of low-risk term birth participants who had two study visits available for analysis. The global sample characteristics such of sample richness, diversity and bacterial load of our low-risk term birth group were broadly in keeping with results from participants who have had term births in previous studies.

There were a notable minority of participants who had low-risk term births and a VMB that could be considered to have unfavourable characteristics. There are likely to be multiple protective factors that lead to term birth in the majority of pregnancies, even in the presence of unfavourable VMB characteristics.

9.3 Contribution of the vaginal microbiota to recurrent early preterm birth

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. High-risk participants who had an early sPTB/PPROM had approximately a five times higher bacterial load in the second trimester than high-risk participants who had a term birth.

Our statistical power was limited because only 22 women had an early sPTB/PPROM recurrence. Not all associations in line with our conclusion reached statistical significance, especially after stratification by bacterial composition.

However, there are a number of reasons why we believe the hypothesis warrants further research. Firstly, the recurrence risk of early sPTB/PPROM increased for each increase in bacterial and *Lactobacillus* concentration quartile, when compared to the distribution of these parameters within the low-risk term birth participants. Secondly, within each VMB type, the high-risk term birth and low-risk term birth participants had similar median bacterial loads, suggesting that it is the bacterial load, independent of the VMB type, that is contributing to the increased risk of early sPTB/PPROM. Finally, bacterial load reached higher levels in participants dominated by *L. iners* or with non-dominance of *Lactobacillus* than for participants with other lactobacilli or bifidobacteria. Therefore our findings do not contradict earlier findings that showed increased PTB risk with non-dominance of lactobacilli and high *L. iners* relative abundance.[101,102,162,165] In fact, the inconsistent findings related to *L. iners* in the literature might be due to the lack of quantification in most studies.

We attempted correlation between VMB findings and clinical phenotype of sPTB/PPROM. This was hampered by small numbers in each subgroup, however we did note that both

participants who had early PPROM with early onset chorioamnionitis had high bacterial loads. It is possible that the high bacterial loads increase the chance of ascension of bacteria into the upper genital tract and adverse outcome.

Three taxa were non-significantly associated with early sPTB/PPROM recurrence in our study and were also associated with PTB in other studies: *Ureaplasma,*[162,168] *Megasphera*[101,168] and BVAB TM7-H1[100]. 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 concurs with earlier work suggesting that *Bifidobacterium breve* may have a protective effect against early sPTB/PPROM recurrence.[103]

The remainder of the VMB characteristics assessed did not show an association with early sPTB/PPROM. Lack of an association, particularly in this small study, does not imply no association, although the magnitude of association appears too small to be detected with recruitment of 137 high-risk women.

9.4 Contribution of the vaginal microbiota to late preterm birth

The VMB characteristics that were associated with early sPTB/PPROM were not replicable in the high-risk late sPTB/PPROM group.

There were only three (3/14, 21%) participants in the highest quartile of bacterial load in the late sPTB/PPROM group. No participants were using preterm birth prevention therapy at the time of VMB sampling, but preterm birth prevention therapy was offered if there was cervical shortening, in line with standard UK care.[26] The three participants in the highest quartile of bacterial load in the late sPTB/PPROM group received preterm birth

prevention treatment after study involvement. It is possible that this treatment averted early sPTB/PPROM in these participants.

This prompted assessment of preterm birth prevention therapies in the other high-risk pregnancy outcome groups. The high-risk term group had a similar trend, with more participants who received PTB prevention treatment having a higher bacterial load (suggesting the treatment might have averted PTB in some of these cases too). When assessing the early sPTB/PPROM group the opposite was true, there was a non-significant trend for more participants without PTB prevention treatment to have a high bacterial load. The participants who were high-risk, but our current cervical surveillance methods do not identify as having cervical shortening, in order to target PTB prevention treatment, are arguably the women who we currently least well served by PTB prevention clinics. If the current work is replicable then it would be desirable to utilise bacterial load as another method of risk stratification for preterm birth, and ideally develop an intervention to target this.

9.5 Strengths and limitations

Strengths of this study are that we carefully selected and assessed pregnancy outcomes to achieve as clean a phenotype as possible of recurrent early sPTB/PPROM. VMB characteristics were carefully selected to cover different aspects of VMB composition and when possible were semi-quantified.

The most notable limitations are our limited statistical power and lack of data about vaginal yeasts. We are also unable to comment on the functional capacity of the bacteria identified, nor identify the bacteria to species level. The wider availability of metatranscriptomics and metagenomics may make this possible in future work. We had initially planned to analyse early sPTB and PPROM as separate outcomes but limited

statistical power did not permit this. There were 11 participants excluded from the high-risk group due to medically indicated PTB. This high rate of medically indicated PTB may be inevitable within such a high-risk cohort, but nevertheless changes the composition of the term group by removing a section of the participants who are likely to have progressed to term without intervention.

9.6 Implications for future research

Larger, longitudinal studies should assess whether the findings of this work are replicable. In particular assessment should be made of the gestational period during which increased bacterial load is associated with early sPTB/PPROM, so that any intervention can be optimally timed. Concomitantly quantification should be made of the yeasts present, and functional assessment undertaken to detect the transcriptional activity of bacteria present. As our ability to detect different strains of lactobacilli improves this should also be incorporated into future work to better understand how and why some VMB types appear to reach higher bacterial loads. If they are confirmed, interventions that maintain a non-iners lactobacilli/ bifidobacteria-dominated VMB may protect women from inflammation-associated PTB.

References

1. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome [Internet]. 2015;3(1):1–3. Available from: http://dx.doi.org/10.1186/s40168-015-0094-5
2. WHO. Who: Recommended Definitions, Terminology and Format for Statistical Tables Related to The Perinatal Period And Use of A New Certificate For Cause of Perinatal Deaths. Acta Obstet Gynecol Scand [Internet]. 1977 Jan 1 [cited 2019 Mar 11];56(3):247–53. Available from: https://obgyn.onlinelibrary.wiley.com/doi/abs/10.3109/00016347709162009?sid=nlm%3Apubmed
3. Care A, Muller-Myhsok B, Olearo E, Todros T, Caradeux J, Goya M, et al. Should phenotype of previous preterm birth influence management of women with short cervix in subsequent pregnancy? Comparison of vaginal progesterone and Arabin pessary. Ultrasound Obstet Gynecol. 2019;53(4):529–34.
4. Chawanpaiboon S, Vogel JP, Moller AB, Lumbiganon P, Petzold M, Hogan D, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. Lancet Glob Heal. 2019;7(1):e37–46.
5. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, Narwal R, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. Lancet [Internet]. 2012;379(9832):2162–72. Available from: http://dx.doi.org/10.1016/S0140-6736(12)60820-4
6. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet [Internet]. 2016;388(10063):3027–35. Available from: http://dx.doi.org/10.1016/S0140-6736(16)31593-8
7. Howson CP, Kinney M V, McDougall L, Lawn JE. Born too soon: preterm birth matters. Reprod Health. 2013;10 Suppl 1(Suppl 1):S1.
8. WHO. Born Too Soon: A Global Action Report on Preterm Birth [Internet]. WHO. World Health Organization; 2012 [cited 2019 Mar 11]. Available from: https://www.who.int/pmnch/media/news/2012/preterm_birth_report/en/
9. United Nations. Sustainable Development Goals: Sustainable Development Knowledge Platform [Internet]. 2016 [cited 2019 Mar 11]. Available from: https://sustainabledevelopment.un.org/?menu=1300
10. Lee AC, Blencowe H, Lawn JE. Small babies, big numbers: global estimates of preterm birth. Lancet Glob Heal [Internet]. 2019 Jan 1 [cited 2019 Mar 11];7(1):e2–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30389450
11. Department of Health. Safer Maternity Care The National Maternity Safety Strategy-Progress and Next Steps 2 DH ID box Title: Safer Maternity Care-The National Maternity Safety Strategy-Progress and Next Steps Author: Acute Care and Workforce / Acute Care and Quality / Resoluti. 2017;(November). Available from: www.nationalarchives.gov.uk/doc/open-government-licence/
12. Platt MJ. Outcomes in preterm infants. Public Health [Internet]. 2014;128(5):399–403. Available from: http://dx.doi.org/10.1016/j.puhe.2014.03.010
13. Costeloe KL, Hennessy EM, Haider S, Stacey F, Marlow N, Draper ES. Short term outcomes after extreme preterm birth in England: Comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). BMJ [Internet]. 2012 Dec 4 [cited 2019 Mar 11];345(7886):e7976. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/23212881
14. Office for National Statistics. Births by Parents' Characteristics in England and Wales, 2016. Off Natonal Statsitics [Internet]. 2017;1–12. Available from: http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsbyparentscharacteristicsinenglandandwales/2014/pdf
15. Himpens E, Van Den Broeck C, Oostra A, Calders P, Vanhaesebrouck P. Prevalence, type, distribution, and severity of cerebral palsy in relation to gestational age: A meta-analytic review. Dev Med Child Neurol. 2008;50(5):334–40.
16. Murray SR, Stock SJ, Cowan S, Cooper ES, Norman JE. Spontaneous preterm birth prevention in multiple pregnancy. Obstet Gynaecol. 2018;20(1):57–63.
17. National Insitute for Health and Care Excellence. Twin and triplet pregnancy. NICE Guidel (NG 137). 2019;(September):1–61.
18. Bonet M, Cuttini M, Piedvache A, Boyle E, Jarreau P, Kollée L, et al. Changes in management policies for extremely preterm births and neonatal outcomes from 2003 to 2012: two population-based studies in ten European regions. BJOG An Int J Obstet Gynaecol [Internet]. 2017 Sep 1 [cited 2019 Mar 11];124(10):1595–604. Available from: http://doi.wiley.com/10.1111/1471-0528.14639
19. Morisaki N, Ganchimeg T, Vogel J, Zeitlin J, Cecatti J, Souza J, et al. Impact of stillbirths on international comparisons of preterm birth rates: a secondary analysis of the WHO multi-country survey of Maternal and Newborn Health. BJOG An Int J Obstet Gynaecol [Internet]. 2017 Aug 1 [cited 2019 Mar 11];124(9):1346–54. Available from: http://doi.wiley.com/10.1111/1471-0528.14548
20. Vogel JP, Chawanpaiboon S, Moller AB, Watananirun K, Bonet M, Lumbiganon P. The global epidemiology of preterm birth. Best Pract Res Clin Obstet Gynaecol [Internet]. 2018;52:3–12. Available from: https://doi.org/10.1016/j.bpobgyn.2018.04.003
21. Romero R, Espinoza J, Kusanovic JP, Gotsch F, Hassan S, Erez O, et al. The preterm parturition syndrome. BJOG. 2006 Dec;113 Suppl:17–42.
22. Romero R, Dey SK, Fisher SJ. Preterm labor: One syndrome, many causes. Science (80-). 2014;345(6198).
23. Flenady V, Wojcieszek AM, Papatsonis DNM, Stock OM, Murray L, Jardine LA, et al. Calcium channel blockers for inhibiting preterm labour and birth. Cochrane Database Syst Rev [Internet]. 2014;(6). Available from: https://doi.org//10.1002/14651858.CD002255.pub2
24. Flenady V, Reinebrant HE, Liley HG, Tambimuttu EG, Papatsonis DNM. Oxytocin receptor antagonists for inhibiting preterm labour. Cochrane Database Syst Rev [Internet]. 2014;(6). Available from: https://doi.org//10.1002/14651858.CD004452.pub3
25. NHS England. Saving babies ' lives version two: A care bundle for reducing perinatal mortality. NHS Engl [Internet]. 2019;1–72. Available from: https://www.england.nhs.uk/wp-content/uploads/2019/03/Saving-Babies-Lives-Care-Bundle-Version-Two-Final-Version2.pdf
26. National Institute for Clinical Excellence. Preterm Labour and Birth, Full Guideline [Internet]. 2015. Available from: https://www.nice.org.uk/guidance/ng25
27. Nadeem L, Shynlova O, Matysiak-Zablocki E, Mesiano S, Dong X, Lye S. Molecular evidence of functional progesterone withdrawal in human myometrium. Nat Commun [Internet]. 2016;7(May):11565. Available from: http://dx.doi.org/10.1038/ncomms11565
28. Shah NM, Lai PF, Imami N, Johnson MR. Progesterone-related immune modulation of pregnancy and labor. Front Endocrinol (Lausanne). 2019;10(MAR):1–19.
29. Jarde A, Lutsiv O, Beyene J, McDonald SD. Vaginal progesterone, oral progesterone, 17-OHPC, cerclage, and pessary for preventing preterm birth in at-risk singleton

pregnancies: an updated systematic review and network meta-analysis. BJOG An Int J Obstet Gynaecol. 2019;126(5):556–67.
30. Care AG, Sharp AN, Lane S, Roberts D, Watkins L, Alfirevic Z. Predicting preterm birth in women with previous preterm birth and cervical length >/= 25 mm. Ultrasound Obstet Gynecol. 2014 Jun;43(6):681–6.
31. Martin JN, D'Alton M, Jacobsson B, Norman JE. In pursuit of progress toward effective preterm birth reduction. Obstet Gynecol. 2017;129(4):715–9.
32. Romero R. Infection in the pathogenesis of preterm labor. Semin Perinatol. 1988;12:262–79.
33. Elovitz MA, Mrinalini C. Animal models of preterm birth. Trends Endocrinol Metab. 2004;15(10):479–87.
34. Dotters-Katz SK, Heine RP, Grotegut CA. Medical and infectious complications associated with pyelonephritis among pregnant women at delivery. Infect Dis Obstet Gynecol [Internet]. 2013 [cited 2019 Mar 11];2013:124102. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24194632
35. Smaill F, Vazquez J. Antibiotics for asymptomatic bacteriuria in pregnancy (Review) Antibiotics for asymptomatic bacteriuria in pregnancy. Cochrane Database Syst Rev. 2015;2(8).
36. Hillier SL, Nugent RP, Eschenbach DA, Krohn MA, Gibbs RS, Martin DH, et al. Association between Bacterial Vaginosis and Preterm Delivery of a Low-Birth-Weight Infant. N Engl J Med [Internet]. 1995 Dec 28 [cited 2019 Mar 11];333(26):1737–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7491137
37. Goldenberg RL, Hauth JC, Andrews WW. Intrauterine Infection and Preterm Delivery. Epstein FH, editor. N Engl J Med [Internet]. 2000 May 18 [cited 2019 Mar 11];342(20):1500–7. Available from: http://www.nejm.org/doi/10.1056/NEJM200005183422007
38. Aagaard, Kjersti, Jun Ma, Kathleen M. Antony, Radhika Ganu, Joseph Petrosino and JV. The Placenta Harbors a Unique Microbiome Kjersti. Sci T ransl Med. 2016;6(237):1–22.
39. Vinturache AE, Gyamfi-Bannerman C, Hwang J, Mysorekar IU, Jacobsson B. Maternal microbiome - A pathway to preterm birth. Vol. 21, Seminars in Fetal and Neonatal Medicine. 2016.
40. De Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, et al. Human placenta has no microbiome but can harbour potential pathogens Europe PMC Funders Group. Nature. 2019;572(7769):329–34.
41. Dudley DJ. The placental microbiome: Yea, nay, or maybe? BJOG An Int J Obstet Gynaecol. 2019;2019.
42. Goodfellow L, Care A, Alfirevic Z. Controversies in prevention of spontaneous preterm birth in asymptomatic women: an evidence summary and expert opinion. BJOG An Int J Obstet Gynaecol. 2020;1–18.
43. Leitich H, Bodner-Adler B, Brunbauer M, Kaider A, Egarter C, Husslein P. Bacterial vaginosis as a risk factor for preterm delivery: A meta-analysis. Am J Obstet Gynecol [Internet]. 2003 Jul 1;189(1):139–47. Available from: https://doi.org/10.1067/mob.2003.339
44. Sangkomkamhang US, Lumbiganon P, Prasertcharoensuk W, Laopaiboon M. Antenatal lower genital tract infection screening and treatment programs for preventing preterm delivery. Cochrane Database Syst Rev [Internet]. 2015 Feb 1 [cited 2020 Apr 2];(2). Available from: http://doi.wiley.com/10.1002/14651858.CD006178.pub3
45. Lee AC, Mullany LC, Quaiyum M, Mitra DK, Labrique A, Christian P, et al. Effect of population-based antenatal screening and treatment of genitourinary tract

infections on birth outcomes in Sylhet, Bangladesh (MIST): a cluster-randomised clinical trial. Lancet Glob Heal [Internet]. 2019;7(1):e148–59. Available from: http://dx.doi.org/10.1016/S2214-109X(18)30441-8
46. Subtil D, Brabant G, Tilloy E, Devos P, Canis F, Fruchart A, et al. Early clindamycin for bacterial vaginosis in pregnancy (PREMEVA): a multicentre, double-blind, randomised controlled trial. Lancet. 2018;392(10160):2171–9.
47. Ahmadi A, Ramazanzadeh R, Sayehmiri K, Sayehmiri F, Amirmozafari N. Association of Chlamydia trachomatis infections with preterm delivery; a systematic review and meta-analysis. BMC Pregnancy Childbirth. 2018;18(1):1–7.
48. Folger AT. Maternal Chlamydia trachomatis Infections and Preterm Birth: The Impact of Early Detection and Eradication During Pregnancy. Matern Child Health J [Internet]. 2014;18(8):1795–802. Available from: https://doi.org/10.1007/s10995-013-1423-6
49. Reekie J, Roberts C, Preen D, Hocking JS, Donovan B, Ward J, et al. Chlamydia trachomatis and the risk of spontaneous preterm birth, babies who are born small for gestational age, and stillbirth: a population-based cohort study. Lancet Infect Dis. 2018;18(4):452–60.
50. Cluver C, Novikova N, Eriksson DOA, Bengtsson K, Lingman GK. Interventions for treating genital Chlamydia trachomatis infection in pregnancy. Cochrane Database Syst Rev [Internet]. 2017;(9). Available from: https://doi.org//10.1002/14651858.CD010485.pub2
51. Meis PJ, Michielutte R, Peters TJ, Bradley H, Sands RE, Coles EC, et al. Factors associated with preterm birth in Cardiff , Wales birth II. 1995;
52. Smaill FM, Vazquez JC. Antibiotics for asymptomatic bacteriuria in pregnancy. Vol. 2019, Cochrane Database of Systematic Reviews. 2019.
53. Kazemier BM, Koningstein FN, Schneeberger C, Ott A, Bossuyt PM, Miranda E De, et al. Maternal and neonatal consequences of treated and untreated asymptomatic bacteriuria in pregnancy : a prospective cohort study with an embedded randomised controlled trial. 2015;15(November).
54. Teshome A, Yitayeh A. Relationship between periodontal disease and preterm low birth weight: Systematic review. Pan Afr Med J. 2016;24:1–10.
55. George A, Shamim S, Johnson M, Ajwani S, Bhole S, Blinkhorn A, et al. Periodontal treatment during pregnancy and birth outcomes: a meta-analysis of randomised trials. Int J Evid Based Healthc [Internet]. 2011 Jun 1;9(2):122–47. Available from: https://doi.org/10.1111/j.1744-1609.2011.00210.x
56. López NJ, Uribe S, Martinez B. Effect of periodontal treatment on preterm birth rate: a systematic review of meta-analyses. Periodontol 2000 [Internet]. 2015 Feb 1;67(1):87–130. Available from: https://doi.org/10.1111/prd.12073
57. Silver BJ, Guy RJ, Kaldor JM, Jamil MS, Rumbold AR. Trichomonas vaginalis as a cause of perinatal morbidity: A systematic review and Meta-analysis. Sex Transm Dis. 2014;41(6):369–76.
58. Klebanoff MA, Carey JC, Hauth JC, Hillier SL, Nugent RP, Thom EA, et al. Failure of metronidazole to prevent preterm delivery among pregnant women with asymptomatic Trichomonas vaginalis infection. N Engl J Med. 2001 Aug;345(7):487–93.
59. Gülmezoglu AM, Azhar M. Interventions for trichomoniasis in pregnancy. Cochrane Database Syst Rev [Internet]. 2011;(5). Available from: https://doi.org//10.1002/14651858.CD000220.pub2
60. McGregor JA, French JI, Parker R, Draper D, Patterson E, Jones W, et al. Prevention of premature birth by screening and treatment for common genital tract infections: Results of a prospective controlled evaluation. Am J Obstet Gynecol [Internet]. 1995

Jul 1;173(1):157–67. Available from: https://doi.org/10.1016/0002-9378(95)90184-1
61. Cotch MF, Hillier SL, Gibbs RS, Eschenbach DA. Epidemiology and outcomes associated with moderate to heavy Candida colonization during pregnancy. Am J Obstet Gynecol [Internet]. 1998 Feb 1;178(2):374–80. Available from: https://doi.org/10.1016/S0002-9378(98)80028-8
62. Roberts CL, Rickard K, Kotsiou G, Morris JM. Treatment of asymptomatic vaginal candidiasis in pregnancy to prevent preterm birth: An open-label pilot randomized controlled trial. BMC Pregnancy Childbirth [Internet]. 2011;11(1):18. Available from: http://www.biomedcentral.com/1471-2393/11/18
63. Bang C, Schmitz RA. Archaea associated with human surfaces: Not to be underestimated. FEMS Microbiol Rev. 2015;39(5):631–48.
64. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey H V., Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A. 2013;110(9):3229–36.
65. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature. 2007;449(7164):804–10.
66. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. 2016;14(8):1–14.
67. Woese CR, Stackebrandt E, Macke TJ, Fox GE. A phylogenetic definition of the major eubacterial taxa. Syst Appl Microbiol. 1985;6:143–51.
68. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci U S A. 1985;82(20):6955–9.
69. Clarridge Jill E. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. Clin Microbiol Rev. 2004;17(4):840–62.
70. The Integrative HMP (iHMP) Research Network Consortium. The Human Microbiome Project. 2014.
71. Proctor LM, Creasy HH, Fettweis JM, Lloyd-Price J, Mahurkar A, Zhou W, et al. The Integrative Human Microbiome Project. Nature. 2019;569(7758):641–8.
72. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. Nature [Internet]. 2013;493(7430):45–50. Available from: https://doi.org/10.1038/nature11711
73. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell [Internet]. 2019;176(3):649-662.e20. Available from: https://doi.org/10.1016/j.cell.2019.01.001
74. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature [Internet]. 2012;486(7402):222–7. Available from: https://doi.org/10.1038/nature11053
75. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655–62.
76. Kang D-W, Ilhan ZE, Isern NG, Hoyt DW, Howsmon DP, Shaffer M, et al. Differences in fecal microbial metabolites and microbiota of children with autism spectrum disorders. Anaerobe [Internet]. 2018;49:121–31. Available from: http://www.sciencedirect.com/science/article/pii/S1075996417302305
77. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Ley RE, Sogin ML, et al. A core gut microbiome between lean and obesity twins. Nature. 2009;457(7228):480–4.

78. Zhu S, Jiang Y, Xu K, Cui M, Ye W, Zhao G, et al. The progress of gut microbiome research related to brain disorders. J Neuroinflammation. 2020;17(1):1–20.
79. Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J, Tito RY, et al. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol [Internet]. 2019;4(4):623–32. Available from: https://doi.org/10.1038/s41564-018-0337-x
80. Polster SP, Sharma A, Tanes C, Tang AT, Mericko P, Cao Y, et al. Permissive microbiome characterizes human subjects with a neurovascular disease cavernous angioma. Nat Commun [Internet]. 2020;11(1):2659. Available from: http://dx.doi.org/10.1038/s41467-020-16436-w
81. Zhang Z, Mocanu V, Cai C, Dang J, Slater L, Deehan EC, et al. Impact of fecal microbiota transplantation on obesity and metabolic syndrome- a systematic review. Nutrients. 2019;11(10).
82. Cuthbertson L, Walker AW, Oliver AE, Rogers GB, Rivett DW, Hampton TH, et al. Lung function and microbiota diversity in cystic fibrosis. Microbiome. 2020;8(1):1–13.
83. NIH Human Microbiome Portfolio Analysis Team. A review of 10 years of human microbiome research activities at the US National Institutes of Health, Fiscal Years 2007-2016. 2019;1–19. Available from: https://doi.org/10.1186/s40168-019-0620-y
84. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011;108(SUPPL. 1):4680–7.
85. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. Sci Transl Med. 2012 May;4(132):132ra52.
86. Vodstrcil LA, Twin J, Garland SM, Fairley CK, Hocking JS, Law MG, et al. The influence of sexual activity on the vaginal microbiota and Gardnerella vaginalis clade diversity in young women. PLoS One. 2017;12(2):1–15.
87. Plummer EL, Vodstrcil LA, Fairley CK, Tabrizi SN, Garland SM, Law MG, et al. Sexual practices have a significant impact on the vaginal microbiota of women who have sex with women. Sci Rep. 2019;9(1):1–14.
88. Borgdorff H, Van Der Veer C, Van Houdt R, Alberts CJ, De Vries HJ, Bruisten SM, et al. The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. PLoS One. 2017;12(7).
89. Jefferson KK, Parikh HI, Garcia EM, Edwards DJ, Serrano MG, Hewison M, et al. Relationship between vitamin D status and the vaginal microbiome during pregnancy. J Perinatol [Internet]. 2019;39(6):824–36. Available from: https://doi.org/10.1038/s41372-019-0343-8
90. Borgdorff H, Gautam R, Armstrong SD, Xia D, Ndayisaba GF, Van Teijlingen NH, et al. Cervicovaginal microbiome dysbiosis is associated with proteome changes related to alterations of the cervicovaginal mucosal barrier. Mucosal Immunol. 2016;9(3).
91. van de Wijgert JHHM. The vaginal microbiome and sexually transmitted infections are interlinked: Consequences for treatment and prevention. PLoS Med. 2017;14(12):2–5.
92. Lewis FMT, Bernstein KT, Aral SO. Vaginal Microbiome and Its Relationship to Behavior, Sexual Health, and Sexually Transmitted Diseases. Obstet Gynecol [Internet]. 2017 Apr;129(4):643–54. Available from: https://pubmed.ncbi.nlm.nih.gov/28277350
93. van de Wijgert JHHM, Verwijs MC. Lactobacilli-containing vaginal probiotics to cure or prevent bacterial or fungal vaginal dysbiosis: a systematic review and recommendations for future trial designs. BJOG An Int J Obstet Gynaecol. 2019;

94. National Institute for Health and Care Excellence. Bacterial vaginosis. 2018.
95. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Nikita L, et al. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. Microbiome. 2014;2(1):1–19.
96. MacIntyre DA, Chandiramani M, Lee YS, Kindinger L, Smith A, Angelopoulos N, et al. The vaginal microbiome during pregnancy and the postpartum period in a European population. Sci Rep. 2015;5(Cst Iv):1–9.
97. Serrano MG, Parikh HI, Brooks JP, Edwards DJ, Arodz TJ, Edupuganti L, et al. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. Nat Med [Internet]. 2019;25(June). Available from: http://dx.doi.org/10.1038/s41591-019-0465-8
98. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Bieda J, et al. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. Microbiome [Internet]. 2014 Dec 27 [cited 2019 Mar 19];2(1):18. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/2049-2618-2-18
99. Peelen MJ, Luef BM, Lamont RF, de Milliano I, Jensen JS, Limpens J, et al. The influence of the vaginal microbiota on preterm birth: A systematic review and recommendations for a minimum dataset for future research. Placenta [Internet]. 2019;79(March):30–9. Available from: https://doi.org/10.1016/j.placenta.2019.03.011
100. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The vaginal microbiome and preterm birth. Nat Med. 2019;25(6):1012–21.
101. Elovitz MA, Gajer P, Riis V, Brown AG, Humphrys MS, Holm JB, et al. Cervicovaginal microbiota and local immune response modulate the risk of spontaneous preterm delivery. Nat Commun [Internet]. 2019;10(1):1–8. Available from: http://dx.doi.org/10.1038/s41467-019-09285-9
102. Kindinger LM, Bennett PR, Lee YS, Marchesi JR, Smith A, Cacciatore S, et al. The interaction between vaginal microbiota, cervical length, and vaginal progesterone treatment for preterm birth risk. Microbiome [Internet]. 2017;5(1):1–14. Available from: http://dx.doi.org/10.1186/s40168-016-0223-9
103. Brown RG, Al-Memar M, Marchesi JR, Lee YS, Smith A, Chan D, et al. Establishment of vaginal microbiota composition in early pregnancy and its association with subsequent preterm prelabor rupture of the fetal membranes. Transl Res [Internet]. 2019;1–14. Available from: https://doi.org/10.1016/j.trsl.2018.12.005
104. Cox MJ, Turek EM, Hennessy C, Mirza GK, James PL, Coleman M, et al. Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-cystic fibrosis bronchiectasis patients. PLoS One. 2017;12(2):1–17.
105. Liu CM, Aziz M, Kachur S, Hsueh PR, Huang YT, Keim P, et al. BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay. BMC Microbiol. 2012;12.
106. Tabatabaei N, Eren A, Barreiro L, Yotova V, Dumaine A, Allard C, et al. Vaginal microbiome in early pregnancy and subsequent risk of spontaneous preterm birth: a case-control study. BJOG An Int J Obstet Gynaecol [Internet]. 2018;8–11. Available from: http://doi.wiley.com/10.1111/1471-0528.15299
107. Berman HL, McLaren MR, Callahan BJ. Understanding and interpreting community sequencing measurements of the vaginal microbiome. BJOG An Int J Obstet Gynaecol. 2020;127(2):139–46.
108. Nayfach S, Pollard KS. Toward Accurate and Quantitative Comparative Metagenomics. Cell [Internet]. 2016 Aug 25;166(5):1103–16. Available from: https://pubmed.ncbi.nlm.nih.gov/27565341

109. Mitra A, MacIntyre DA, Mahajan V, Lee YS, Smith A, Marchesi JR, et al. Comparison of vaginal microbiota sampling techniques: Cytobrush versus swab. Sci Rep. 2017;7(1):1–10.
110. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The vaginal microbiome and preterm birth. Nat Med. 2019;25(6).
111. Kramer MS, Papageorghiou A, Conde-Agudelo A, Waller S, Villar J, Bhutta Z, et al. Challenges in defining and classifying the preterm birth syndrome. Am J Obstet Gynecol [Internet]. 2011;206(2):108–12. Available from: http://dx.doi.org/10.1016/j.ajog.2011.10.864
112. MacTier H, Bates SE, Johnston T, Lee-Davey C, Marlow N, Mulley K, et al. Perinatal management of extreme preterm birth before 27 weeks of gestation: A framework for practice. Arch Dis Child Fetal Neonatal Ed. 2020;105(3):F232–9.
113. Bastek JA, Sammel MD, Paré E, Srinivas SK, Posencheg MA, Elovitz MA. Adverse neonatal outcomes: examining the risks between preterm, late preterm, and term infants. Am J Obstet Gynecol. 2008;199(4):367.e1-367.e8.
114. Ancel PY, Saurel-Cubizolles MJ, Di Renzo GC, Papiernik E, Bréart G. Risk factors for 14-21 week abortions: a case-control study in Europe. The Europop Group. Hum Reprod [Internet]. 2000 Nov [cited 2019 Mar 11];15(11):2426–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11056147
115. Monson MA, Gibbons KJ, Esplin MS, Varner MW, Manuck TA. Pregnancy Outcomes in Women With a History of Previable, Preterm Prelabor Rupture of Membranes. Obstet Gynecol [Internet]. 2016 Nov [cited 2019 Mar 11];128(5):976–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27741176
116. Capece A, Vasieva O, Meher S, Alfirevic Z, Alfirevic A. Pathway analysis of genetic factors associated with spontaneous preterm birth and pre-labor preterm rupture of membranes. PLoS One. 2014;9(9).
117. Martin JN, D'Alton M, Jacobsson B, Norman JE. In Pursuit of Progress Toward Effective Preterm Birth Reduction. Obstet Gynecol [Internet]. 2017 Apr 1 [cited 2019 Mar 11];129(4):715–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28277357
118. Goldenberg RL, Culhane JF, Iams JD, Romero R. Preterm Birth 1 Epidemiology and causes of preterm birth. 2008;75–84.
119. Souza RT, Cecatti JG, Passini R, Tedesco RP, Lajos GJ, Nomura ML, et al. The Burden of Provider-Initiated Preterm Birth and Associated Factors: Evidence from the Brazilian Multicenter Study on Preterm Birth (EMIP). Zakar T, editor. PLoS One [Internet]. 2016 Feb 5 [cited 2019 Mar 11];11(2):e0148244. Available from: https://dx.plos.org/10.1371/journal.pone.0148244
120. Vogel JP, Lee ACC, Souza JP. Maternal morbidity and preterm birth in 22 low- and middle-income countries: a secondary analysis of the WHO Global Survey dataset. BMC Pregnancy Childbirth [Internet]. 2014 Jan 31 [cited 2019 Mar 11];14:56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24484741
121. Kvalvik LG, Wilcox AJ, Skjærven R, Østbye T, Harmon QE. Term complications and subsequent risk of preterm birth: registry based study. BMJ. 2020;369:m1007.
122. Ferrero DM, Larson J, Jacobsson B, Di Renzo GC, Norman JE, Martin JN, et al. Cross-Country Individual Participant Analysis of 4.1 Million Singleton Births in 5 Countries with Very High Human Development Index Confirms Known Associations but Provides No Biologic Explanation for 2/3 of All Preterm Births. Luo Z-C, editor. PLoS One [Internet]. 2016 Sep 13 [cited 2019 Mar 11];11(9):e0162506. Available from: https://dx.plos.org/10.1371/journal.pone.0162506
123. Zhang G, Srivastava A, Bacelis J, Juodakis J, Jacobsson B, Muglia LJ. Genetic studies of gestational duration and preterm birth. Best Pract Res Clin Obstet Gynaecol

[Internet]. 2018 Oct [cited 2019 Mar 11];52:33–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30007778
124. Hwang J, Vinturache AE, Jacobsson B, Gyamfi-Bannerman C, Mysorekar IU. Maternal microbiome – A pathway to preterm birth. Semin Fetal Neonatal Med. 2016;21(2):94–9.
125. Barrientos-Durán A, Fuentes-López A, de Salazar A, Plaza-Díaz J, García F. Reviewing the composition of vaginal microbiota: Inclusion of nutrition and probiotic factors in the maintenance of eubiosis. Nutrients. 2020;12(2):1–30.
126. Koullali B, Ravelli ACJ, Kazemier BM, Pajkrt E, Mol BWJ, Oudijk MA. 851: Recurrence rate of spontaneous preterm birth. Am J Obstet Gynecol [Internet]. 2016;214(1):S443–4. Available from: http://dx.doi.org/10.1016/j.ajog.2015.10.901
127. Villar J, Papageorghiou AT, Knight HE, Gravett MG, Iams J, Waller SA, et al. The preterm birth syndrome : a prototype phenotypic classification. Am J Obstet Gynecol [Internet]. 2012;206(2):119–23. Available from: http://dx.doi.org/10.1016/j.ajog.2011.10.866
128. Nasioudis D, Forney LJ, Schneider GM, Gliniewicz K, France M, Boester A, et al. Influence of Pregnancy History on the Vaginal Microbiome of Pregnant Women in their First Trimester. Sci Rep [Internet]. 2017;7(1):1–6. Available from: http://dx.doi.org/10.1038/s41598-017-09857-z
129. Goldenberg RL, Gravett MG, Conde-Agudelo A, Bhutta ZA, Barros F, Villar J, et al. The preterm birth syndrome: issues to consider in creating a classification system. Am J Obstet Gynecol [Internet]. 2011;206(2):113–8. Available from: http://dx.doi.org/10.1016/j.ajog.2011.10.865
130. MacKay DF, Smith GCS, Dobbie R, Pell JP. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. PLoS Med. 2010 Jun;7(6):e1000289.
131. Zhang X, Kramer MS. Variations in Mortality and Morbidity by Gestational Age among Infants Born at Term. J Pediatr. 2009;154(3).
132. Tita ATN, Landon MB, Spong CY, Lai Y, Leveno KJ, Varner MW, et al. Timing of Elective Repeat Cesarean Delivery at Term and Neonatal Outcomes. N Engl J Med [Internet]. 2009 Jan 8 [cited 2019 Mar 11];360(2):111–20. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa0803267
133. Ravel J, Chaemsaithong P, Tarca AL, Chaiworapongsa T, Bieda J, Romero R, et al. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. Microbiome. 2014;2(1):1–15.
134. National Institute for Health and Care Excellence. Antenatal care for uncomplicated pregnancies. Clin Guidel [Internet]. 2008;(March):1–55. Available from: https://www.nice.org.uk/guidance/cg62/resources/antenatal-care-for-uncomplicated-pregnancies-975564597445
135. Public Health England (PHE). HIV Testing in England: 2017 Report. 2018;57. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/759270/HIV_testing_in_England_2017_report.pdf
136. Liverpool Women's NHS Foundation Trust. Liverpool Women's Hosital [Internet]. [cited 2020 Aug 29]. Available from: https://www.liverpoolwomens.nhs.uk/our-services/maternity/
137. UK Government. UK General Data Protection Regulation [Internet]. 2018 [cited 2020 Jun 17]. Available from: https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation
138. Gardosi J. GROW documentation. Chart [Internet]. 2015;(July):1–9. Available from:

http://www.gestation.net/GROW_documentation.pdf
139. Ministry of Housing Communities & Local Government. English indices of deprivation 2015 [Internet]. 2015 [cited 2019 Apr 4]. Available from: http://imd-by-postcode.opendatacommunities.org/imd/2015
140. Department for Communities and Local Government. The English Index of Multiple Deprivation (IMD) 2015- Guidance [Internet]. 2015 [cited 2019 Apr 4]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/464430/English_Index_of_Multiple_Deprivation_2015_-_Guidance.pdf
141. Villar J, Papageorghiou AT, Knight HE, Gravett MG, Iams J, Waller SA, et al. The preterm birth syndrome: A prototype phenotypic classification. Am J Obstet Gynecol [Internet]. 2012 Feb [cited 2019 Mar 11];206(2):119–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22177191
142. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome [Internet]. 2014;2(1):6. Available from: https://doi.org/10.1186/2049-2618-2-6
143. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl Environ Microbiol [Internet]. 2006 Jul 1;72(7):5069 LP – 5072. Available from: http://aem.asm.org/content/72/7/5069.abstract
144. Nowak RG, Randis TM, Desai P, He X, Robinson CK, Rath JM, et al. Higher Levels of a Cytotoxic Protein, Vaginolysin, in Lactobacillus-Deficient Community State Types at the Vaginal Mucosa. Sex Transm Dis [Internet]. 2018 Apr;45(4):e14–7. Available from: https://pubmed.ncbi.nlm.nih.gov/29465671
145. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal; Vol 17, No 1 Next Gener Seq Data Anal - 1014806/ej171200 [Internet]. 2011 May 2; Available from: https://journal.embnet.org/index.php/embnetjournal/article/view/200
146. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods [Internet]. 2016 May 23;13:581. Available from: https://doi.org/10.1038/nmeth.3869
147. Rosen MJ, Callahan BJ, Fisher DS, Holmes SP. Denoising PCR-amplified metagenome data. BMC Bioinformatics [Internet]. 2012;13(1):283. Available from: https://doi.org/10.1186/1471-2105-13-283
148. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35(21):7188–96.
149. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol [Internet]. 2007 Aug 15;73(16):5261 LP – 5267. Available from: http://aem.asm.org/content/73/16/5261.abstract
150. Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. N Engl J Med. 2005;353(18):1899–911.
151. US National Library of Medicine. Basic Local Alignment Search Tool [Internet]. [cited 2019 Apr 10]. Available from: blast: Basic Local Alignment Search Tool. https://blast.ncbi.nlm.nih.gov/Blast.cgi (accessed%0DAug 15, 2018).
152. Fettweis JM, Serrano MG, Sheth NU, Mayer CM, Glascock AL, Brooks JP, et al. Species-level classification of the vaginal microbiome. BMC Genomics. 2012;13

Suppl 8(Suppl 8):1–9.
153. Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res. 2015 Jan;43(Database issue):D593-8.
154. Boshier F, Srinivasan S, Fredricks DN, Schiffer J. Complementing 16S rRNA gene amplicon sequencing with estimates of total bacterial load can infer absolute bacterial species concentrations as measured by targeted assays in the vaginal microbiome. Cape Town, South Africa. In: Keystone Symposia Role of the Genital Tract Microbiome in Sexual and Reproductive Health [Internet]. Cape Town, South Africa; 2018. p. 11–5. Available from: https://www.keystonesymposia.org/views/Web/Meetings/dsp_PrintAllAbstracts.cfm?MeetingID=1660
155. Jian C, Luukkonen P, Yki-Jarvinen H, Salonen A, Korpela K. Quantitative PCR provides a simple and accessible method for quantitative microbiome profiling. bioRxiv [Internet]. 2018;478685. Available from: https://www.biorxiv.org/content/early/2018/11/27/478685
156. Meta-Chart. Venn Diagram Maker Online.
157. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B [Internet]. 1995 Aug 17;57(1):289–300. Available from: http://www.jstor.org/stable/2346101
158. Van't Hooft J, Duffy JMN, Daly M, Williamson PR, Meher S, Thom E, et al. A core outcome set for evaluation of interventions to prevent preterm birth. In: Obstetrics and Gynecology. 2016. p. 49–58.
159. Prebic consortium [Internet]. [cited 2020 Aug 6]. Available from: https://www.prebicglobal.org
160. Van Der Pol WJ, Kumar R, Morrow CD, Blanchard EE, Taylor CM, Martin DH, et al. In Silico and Experimental Evaluation of Primer Sets for Species-Level Resolution of the Vaginal Microbiota Using 16S Ribosomal RNA Gene Sequencing. J Infect Dis. 2019;219(2):305–14.
161. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 2009;6(7).
162. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al. Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci U S A. 2015;112(35):11060–5.
163. Subramaniam A, Kumar R, Cliver SP, Zhi D, Szychowski JM, Abramovici A, et al. Vaginal Microbiota in Pregnancy: Evaluation Based on Vaginal Flora, Birth Outcome, and Race. Am J Perinatol. 2016;33(4).
164. Nelson DB, Shin H, Wu J, Dominguez-Bello MG. .The Gestational Vaginal Microbiome and Spontaneous Preterm Birth among Nulliparous African American Women. Am J Perinatol. 2016 Jul;33(9):887–93.
165. Callahan BJ, DiGiulio DB, Goltsman DSA, Sun CL, Costello EK, Jeganathan P, 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 [Internet]. 2017;114(37):9966–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28847941%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5604014
166. Stout MJ, Zhou Y, Wylie KM, Tarr PI, Macones GA, Tuuli MG. Early pregnancy vaginal microbiome trends and preterm birth. Am J Obstet Gynecol. 2017;217(3).
167. Wheeler S, Pryor K, Antczak B, Truong T, Murtha A, Seed P. The relationship of cervical microbiota diversity with race and disparities in preterm birth. J Neonatal

Perinatal Med. 2018;11(3).
168. Freitas AC, Bocking A, Hill JE, Money DM. Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. Microbiome. 2018;6(1):1–15.
169. Blostein F, Gelaye B, Sanchez SE, Williams MA, Foxman B. Vaginal microbiome diversity and preterm birth: results of a nested case–control study in Peru. Ann Epidemiol [Internet]. 2020;41:28–34. Available from: https://doi.org/10.1016/j.annepidem.2019.11.004
170. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol [Internet]. 2011;12(6):R60. Available from: https://doi.org/10.1186/gb-2011-12-6-r60
171. Campisciano G, Zanotta N, Licastro D, De Seta F, Comar M. In vivo microbiome and associated immune markers: New insights into the pathogenesis of vaginal dysbiosis. Sci Rep [Internet]. 2018;8(1):1–9. Available from: http://dx.doi.org/10.1038/s41598-018-20649-x
172. Brown RG, Marchesi JR, Lee YS, Smith A, Lehne B, Kindinger LM, et al. Vaginal dysbiosis increases risk of preterm fetal membrane rupture, neonatal sepsis and is exacerbated by erythromycin. BMC Med. 2018;16(1):1–15.
173. Brown RG, Chan D, Terzidou V, Lee YS, Smith A, Marchesi JR, et al. Prospective observational study of vaginal microbiota pre- and post-rescue cervical cerclage. BJOG An Int J Obstet Gynaecol. 2019;126(7):916–25.
174. van de Wijgert JHHM, Verwijs MC. Lactobacilli-containing vaginal probiotics to cure or prevent bacterial or fungal vaginal dysbiosis: a systematic review and recommendations for future trial designs. BJOG An Int J Obstet Gynaecol. 2019;1–13.
175. Verwijs MC, Agaba SK, Darby AC, van de Wijgert JHHM. Impact of Oral Metronidazole Treatment on the Vaginal Microbiota and Correlates of Treatment Failure. Am J Obstet Gynecol [Internet]. 2019;(August). Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002937819310075
176. van de Wijgert JHHM, Verwijs MC, Gill AC, Borgdorff H, van der Veer C, Mayaud P. Pathobionts in the Vaginal Microbiota: Individual Participant Data Meta-Analysis of Three Sequencing Studies. Front Cell Infect Microbiol. 2020;10(April).
177. van de Wijgert JHHM, Verwijs MC, Agaba SK, Bronowski C, Mwambarangwe L, Uwineza M, et al. Intermittent Lactobacilli-containing Vaginal Probiotic or Metronidazole Use to Prevent Bacterial Vaginosis Recurrence: A Pilot Study Incorporating Microscopy and Sequencing. Sci Rep [Internet]. 2020;10(1):1–15. Available from: http://dx.doi.org/10.1038/s41598-020-60671-6
178. Liverpool City Council. 2011 Census Ethnicity in Liverpool [Internet]. Liverpool; 2011 [cited 2020 Dec 7]. Available from: http://liverpool.gov.uk/Images/C4Ethnicity.pdf
179. Iams JD, Berghella V. Care for women with prior preterm birth. Am J Obstet Gynecol. 2010 Aug;203(2):89–100.
180. Goodfellow L, Care A, Sharp A, Ivandic J, Poljak B, Roberts D, et al. Effect of QU i PP prediction algorithm on treatment decisions in women with a previous preterm birth: a prospective cohort study . BJOG An Int J Obstet Gynaecol. 2019;(October 2018):1–7.
181. Meertens LJE, van Montfort P, Scheepers HCJ, van Kuijk SMJ, Aardenburg R, Langenveld J, et al. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. Acta Obstet Gynecol Scand. 2018;97(8):907–20.
182. Peelen MJ, Luef BM, Lamont RF, de Milliano I, Jensen JS, Limpens J, et al. The influence of the vaginal microbiota on preterm birth: A systematic review and recommendations for a minimum dataset for future research. Placenta [Internet].

2019;79(December 2018):30–9. Available from: https://doi.org/10.1016/j.placenta.2019.03.011
183. Van Diepen M, Ramspek CL, Jager KJ, Zoccali C, Dekker FW. Prediction versus aetiology: Common pitfalls and how to avoid them. Nephrol Dial Transplant. 2017;32(February):ii1–5.
184. Howards PP, Schisterman EF, Heagerty PJ. Potential confounding by exposure history and prior outcomes: An example from perinatal epidemiology. Epidemiology. 2007;18(5):544–51.
185. Mitra A, Macintyre D, Lee Y, Smith A, Marchesi J, Lyons D, et al. P34 Cervical intraepithelial neoplasia is associated with an altered vaginal microbiome and innate immune disruption. Int J Gynecol Cancer [Internet]. 2019 Nov 1;29(Suppl 4):A71 LP-A71. Available from: http://ijgc.bmj.com/content/29/Suppl_4/A71.1.abstract
186. Kyrgiou M, Athanasiou A, Paraskevaidi M, Mitra A, Kalliala I, Martin-hirsch P, et al. Adverse obstetric outcomes after local treatment for cervical preinvasive and early invasive disease according to cone depth : systematic review and meta-analysis. 2016;
187. Neggers YH, Nansel TR, Andrews WW, Schwebke JR, Yu KF, Goldenberg RL, et al. Dietary intake of selected nutrients affects bacterial vaginosis in women. J Nutr. 2007;137(9):2128–33.
188. Fettweis JM, Paul Brooks J, Serrano MG, Sheth NU, Girerd PH, Edwards DJ, et al. Differences in vaginal microbiome in African American women versus women of European ancestry. Microbiol (United Kingdom). 2014;160:2272–82.
189. Han Z, Mulla S, Beyene J, Liao G, Mcdonald SD. Maternal underweight and the risk of preterm birth and low birth weight : a systematic review and meta-analyses. 2011;(November 2010):65–101.
190. Torloni MR, Betrán AP, Daher S, Widmer M, Dolan SM, Menon R, et al. Maternal BMI and preterm birth: A systematic review of the literature with meta-analysis. J Matern Neonatal Med. 2009;22(11):957–70.
191. Brotman RM, He X, Gajer P, Fadrosh D, Sharma E, Mongodin EF, et al. Association between cigarette smoking and the vaginal microbiota: A pilot study. BMC Infect Dis. 2014;14(1):1–11.
192. Burguet A, Kaminski M, Abraham-lerat L, Schaal J, Cambonie G, Fresson J, et al. The complex relationship between smoking in pregnancy and very preterm delivery . Results of the Epipage study. 2004;111(March):258–65.
193. Barros FC, Papageorghiou AT, Victora CG, Noble JA, Pang R, Iams J, et al. The distribution of clinical phenotypes of preterm birth syndrome implications for prevention. JAMA Pediatr. 2015;169(3):220–9.
194. Maghsoudlou S, Beyene J, Yu ZM, McDonald SD. Phenotypic Classification of preterm Birth Among Multiparous Women: A Population-Based Cohort Study. J Obstet Gynaecol Canada. 2019;41(10):1433-1443.e12.
195. Manuck TA, Esplin MS, Andrews W, Zhang H, Bukowski R, Reddy UM, et al. The phenotype of spontaneous preterm birth: application of a clinical phenotyping tool. Am J Obstet Gynecol [Internet]. 2015;212(4):487.e1-487.e11. Available from: http://dx.doi.org/10.1016/j.ajog.2015.02.010
196. Lamont RF. Advances in the prevention of infection-related preterm birth. Front Immunol. 2015;6(NOV):1–12.
197. Ricketts T, Goodfellow L, Care A, Alfirevic A. Harris-Wellbeing Preterm Birth Centre - Home [Internet]. [cited 2019 Mar 19]. Available from: https://www.harris-wellbeingptbcentre.co.uk/
198. Svorai M, Aricha B, Erez O. Prior Preterm Birth and Birthweight Below the 5th Percentile are Independent Risk Factors for Recurrence of a Small for Gestational

Age Neonate. Matern Med. 2020;2(1):28–33.
199. Freitas AC, Chaban B, Bocking A, Rocco M, Yang S, Hill JE, et al. The vaginal microbiome of pregnant women is less rich and diverse, with lower prevalence of Mollicutes, compared to non-pregnant women. Sci Rep. 2017;7(1).
200. Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, et al. Experimental and analytical tools for studying the human microbiome. Nat Rev Genet [Internet]. 2011 Dec 16;13(1):47–58. Available from: https://pubmed.ncbi.nlm.nih.gov/22179717
201. Graspeuntner S, Loeper N, Künzel S, Baines JF, Rupp J. Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. Sci Rep. 2018;8(1):4–10.
202. Brocklehurst P, Gordon A, Heatley E, Milan SJ. Antibiotics for treating bacterial vaginosis in pregnancy. Cochrane Database Syst Rev [Internet]. 2013;(1). Available from: https://doi.org//10.1002/14651858.CD000262.pub4
203. Bayar E, Bennett PR, Chan D, Sykes L, MacIntyre DA. The pregnancy microbiome and preterm birth. Semin Immunopathol. 2020;42(4):487–99.
204. Cruickshank R, Sharman A. The Biology of the Vagina in the Human Subject. BJOG An Int J Obstet Gynaecol [Internet]. 1934 Apr 1;41(2):208–26. Available from: https://doi.org/10.1111/j.1471-0528.1934.tb08759.x
205. Hay P. Bacterial Vaginosis as a Mixed Infection. In: Brogden K, Guthmiller J, editors. Polymicrobial Diseases [Internet]. Washington (DC); 2002. Available from: https://www.ncbi.nlm.nih.gov/books/NBK2495
206. Falsen E, Pascual C, Sjbden B, Ohlnl M, Collins MD. Phenotypic and phylogenetic characterization of a novel Lactobacillus species from human sources: description of Lactobacillus iners sp. nov. Int J Syst Bacteriol. 1999;49:217–21.
207. Macklaim JM, Gloor GB, Anukam KC, Cribby S, Reid G. At the crossroads of vaginal health and disease, the genome sequence of Lactobacillus iners AB-1. Proc Natl Acad Sci U S A. 2011;108(SUPPL. 1):4688–95.
208. Van Der Veer C, Hertzberger RY, Bruisten SM, Tytgat HLP, Swanenburg J, De Kat Angelino-Bart A, et al. Comparative genomics of human Lactobacillus crispatus isolates reveals genes for glycosylation and glycogen degradation: Implications for in vivo dominance of the vaginal microbiota. Microbiome. 2019;7(1):1–14.
209. Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, et al. The evolution of host specialization in the vertebrate gut symbiont Lactobacillus reuteri. PLoS Genet. 2011;7(2).
210. Petrova MI, Reid G, Vaneechoutte M, Lebeer S. Lactobacillus iners: Friend or Foe? Trends Microbiol [Internet]. 2017;25(3):182–91. Available from: http://dx.doi.org/10.1016/j.tim.2016.11.007
211. Albanese D, Donati C. Strain profiling and epidemiology of bacterial species from metagenomic sequencing. Nat Commun [Internet]. 2017;8(1):1–14. Available from: http://dx.doi.org/10.1038/s41467-017-02209-5
212. Tortelli BA, Lewis AL, Fay JC. The structure and diversity of strain level variation in vaginal bacteria. bioRxiv [Internet]. 2020 Jan 1;2020.06.26.173922. Available from: http://biorxiv.org/content/early/2020/06/26/2020.06.26.173922.abstract
213. Macklaim JM, Fernandes AD, Di Bella JM, Hammond JA, Reid G, Gloor GB. Comparative meta-RNA-seq of the vaginal microbiota and differential expression by Lactobacillus iners in health and dysbiosis. Microbiome. 2013;1(1):1–11.
214. Callahan BJ, DiGiulio DB, Goltsman DSA, Sun CL, Costello EK, Jeganathan P, 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. 2017;114(37):9966–71.
215. Martinson JNV, Pinkham N V., Peters GW, Cho H, Heng J, Rauch M, et al. Rethinking

gut microbiome residency and the Enterobacteriaceae in healthy human adults. ISME J [Internet]. 2019;13(9):2306–18. Available from: http://dx.doi.org/10.1038/s41396-019-0435-7
216. Iams JD, Goldenberg RL, Meis PJ, Mercer BM, Moawad A, Das A, et al. The Length of the Cervix and the Risk of Spontaneous Premature Delivery. N Engl J Med [Internet]. 1996 Feb 29;334(9):567–73. Available from: https://doi.org/10.1056/NEJM199602293340904
217. Berghella V, Saccone G. Cervical assessment by ultrasound for preventing preterm delivery. Cochrane Database Syst Rev [Internet]. 2019;(9). Available from: https://doi.org//10.1002/14651858.CD007235.pub4
218. Hughes K, Kane SC, Araujo Júnior E, Da Silva Costa F, Sheehan PM. Cervical length as a predictor for spontaneous preterm birth in high-risk singleton pregnancy: current knowledge. Ultrasound Obstet Gynecol. 2016;48(1):7–15.
219. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nat Rev Microbiol [Internet]. 2018;16(7):410–22. Available from: https://doi.org/10.1038/s41579-018-0029-9
220. Jian C, Luukkonen P, Yki-Järvinen H, Salonen A, Korpela K. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLoS One. 2020;15(1).
221. Romero R, Espinoza J, Gonçalves LF, Kusanovic JP, Friel L, Hassan S. The Role of Inflammation and Infection in Preterm Birth. Semin Reprod Med. 2007;25(01):21–39.
222. Fidel PL, Sobel JD. Immunopathogenesis of recurrent vulvovaginal candidiasis. Clin Microbiol Rev. 1996;9(3):335–48.
223. Brown HK, Speechley KN, Macnab J, Natale R, Campbell MK. Biological determinants of spontaneous late preterm and early term birth: A retrospective cohort study. BJOG An Int J Obstet Gynaecol. 2015;122(4):491–9.
224. Whitman W., Rainey F, Kämpfer P. Bergey's Manual of Systematics of Archaea and Bacteria [Internet]. Trujillo ME, Dedysh S, DeVos P, Hedlund B, Kämpfer P, Rainey FA, et al., editors. Chichester: Wiley; 2015. Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118960608

The Development of Novel Biomarkers for Preterm Labour
Project ID number: 11/NW/0720

A group of research doctors, based at Liverpool Women's Hospital, are trying to develop new methods to predict preterm birth. We run a specialist preterm birth prevention clinic where we see ladies who have had babies before between 16 and 34 weeks of pregnancy, and many of these ladies are kindly helping us develop ways to predict preterm birth. We also need a group of healthy mums to compare to. We are looking for ladies who fit the following criteria:

- **Currently under 17 weeks of pregnancy with single baby**
- **Have had at least one baby before**
- **All previous babies have been born at 37 or more weeks of pregnancy (less than 3 weeks early)**
- **No high blood pressure/pre-eclampsia/diabetes in pregnancies before**

The study would involve:

- **Two visits at about 16 and about 20 weeks of pregnancy**
- **Each visit lasts for about an hour**
- **During the visits we would ask for a urine sample, take blood, and do a speculum examination to take vaginal swabs**
- **During the visit we would also do an internal scan to measure the length of your cervix. If the cervix is short this means that you could be at risk of preterm birth, and so we would not continue with the research but we would offer you treatment for this in our specialist clinic.**

In return:

- **You would be contributing to important research that aims to reduce the number of babies born too soon**
- **During each visit we would do an ultrasound scan of your baby, and give you free pictures to take home. This can involve a scan to find out the baby's sex at 16 weeks if you would like, and a 3D scan at 20 weeks.**

If you would like to be involved, or for more information, please text your name and hospital number to 07470003801.

Thank you!

Appendix B

Control group information leaflet

Version: 2.0
Date: 22.01.2016
Project ID number:11/NW/0720

Liverpool Women's NHS
NHS Foundation Trust
Crown Street
Liverpool
L8 7SS
Tel: 0151 708 9988
www.lwh.nhs.uk

The Development of Novel Biomarkers for Preterm Labour

Dr Angharad Care – Harris Preterm Birth Clinical Research Fellow, Liverpool Women's Hospital
Professor Zarko Alfirevic – Professor in Fetal & Maternal Medicine, Liverpool Women's Hospital

We are inviting pregnant women who are very likely to have a normal and healthy pregnancy to take part in a research study to help mums and babies who suffer with prematurity (babies born too early). Before you decide whether or not to take part it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. One of our team will go through the information sheet with you and answer any questions you may have.

Why are we doing the study?

A baby being born too early is a really important area for us to research as we still understand very little about why some mothers go into labour too early and who it is going to affect. Premature infants are at greater risk for death, cerebral palsy, delays in development, hearing problems, and problems seeing. These risks are greater the earlier a baby is born.

The aim of this study is to develop earlier and safer ways of detecting preterm labour. Research has shown that it may be possible to use blood tests or vaginal swabs taken from the mother early in pregnancy to gather information which may help us to predict the likelihood of preterm labour. In this study we are taking blood and vaginal swabs from healthy women to compare them to women who have a baby born too early. If there are signals that can be used to predict early labour we hope to only find them in the mum's that have a preterm labour and will not be in the group that carry their baby to term. For this we need to take blood and swabs from women like you who have had a normal pregnancy before.

There is also evidence that women with less acidity (a higher pH) or certain types of bacteria in the vagina may be more at risk of preterm delivery. We will perform vaginal swabs for pH and bacteria check when you come to clinic at 16 weeks and 20 weeks of your pregnancy. This is not a swab for sexually transmitted diseases, if this is something

you want this should be done at a local family planning or sexual health clinic. We understand that at 16 weeks it would be an additional trip to hospital for you, therefore in addition to the tests we can provide you with a reassurance scan to show you your baby's heartbeat and a picture of your baby. If you would like to know the sex of your baby, we could also tell you this information (but we warn all women a scan can never be 100% accurate). You will still need to attend your booked 20 week scan as this will look at all your baby's organs to check they are developing normally and if you want we can arrange your visit to happen at the same time, (if you have your scan at Liverpool Women's Hospital), so you don't have additional trips to the hospital. However, if you attend a separate appointment we will still scan your baby and provide you with pictures to keep for free.

There is a link between bacteria, infection and preterm birth but it is not clear how the bacteria cause preterm birth or how bacteria get into the womb. Certain "pro-biotics" are being developed for good gut health and help with digestive problems. In this project we are also analysing bacteria in urine (wee) and stool (poo) as well as vaginal bacteria. If there are "bad" bacteria that are linked with preterm birth we may be able to design a "probiotic" to replace the bacteria in the tummy with "good" bacteria to try and prevent preterm birth. We are also collecting stool samples which can be done at 16 and 20 weeks of pregnancy at home with our specially designed kits and posted back to our lab. These samples are optional, if you did not wish to provide them, you may still be included in the research study.

It might be that a combination of these "biomarkers" or signals in the body that might allow us to predict preterm birth, so our statisticians will be putting all this data together to see if combinations of this information that we find can predict if women are protected from preterm birth or are at risk.

Why have I been chosen?

We are inviting all women over 18 years who have previously had a normal healthy pregnancy that delivered at term.

Do I have to take part?

It is up to you to decide whether or not to take part. If you decide to take part you are free to withdraw at any time without giving a reason. If you do not feel able to take part it will not in any way affect the care you or your family receives.

What will happen to me if I take part?

If you do decide to take part you will be given this information sheet to keep and asked to sign a consent form. We will make a special appointment for you to be seen at 16 weeks for a blood test, vaginal swab and a 3 minute internal scan of your cervix to measure the length (short cervix is a risk factor for preterm birth). If your cervix is very short we will provide you with preventative treatment and you will not be able to take part in the research study. We will then perform a scan of your baby for reassurance and provide you with pictures (and sex the baby if you would like to find out). We will also ask that you donate a urine samples and stool samples at 16 weeks and 20 weeks of pregnancy, these can be done at home. The appointment may take up to 55 minutes for us to chat about your pregnancy, collect information about you and perform the swabs and blood tests with you feeling comfortable. You will be given a second appointment at 20 weeks and the same will happen at this appointment.. We will be looking at genes that make it more likely that you might have a preterm birth but we will not be testing for genetic disease or paternity on any of these samples.

Will my taking part in this study be kept confidential?

Yes. We will follow ethical and legal practice and all information will be handled in confidence. Any information you give us will only be used by the research team in the course of the research to develop these new tests. Any samples and data stored will be stored securely. They will be coded, and no personal data (name and address) will be available to the researchers. However, if any analysis provides clinically relevant information we will inform your medical doctor.

What are the possible benefits of taking part?

The results of this research will not benefit the course of your pregnancy. We hope that the results of the study overall will enable us to improve antenatal care provided to women by developing safer prenatal tests that will help us detect pregnancies at risk of preterm birth. We understand that you are doing research to benefit others and will perform an additional scan to give you the opportunity to see your baby and your baby's heartbeat by one of the Obstetric doctors at the 16 week visit.

What are the possible disadvantages and risks of taking part?

We will take additional samples of blood, urine, stool (optional), vaginal swabs. This will be carried out by someone who is skilled in taking blood. Some women find it embarrassing to have a vaginal swab but this will be performed by a female doctor or researcher. Ultrasound has an excellent safety record and will not harm your baby but is optional.

What will happen if I don't want to continue in the study?

You are free to withdraw at anytime. If you withdraw from the study we will not access any further samples and will destroy any of your samples that were collected for the study.

What will happen to any samples I give?

The samples will be coded and no personal data (name and address) will be stored with the sample. The development of these tests will be done in laboratories in the UK, Finland and Germany. We also ask whether you would be willing to gift samples to be used for other ethically approved research studies into pregnancy problems.

What will happen to the results of the research study?

The results from our project will be published as research papers in medical journals. No data will be published that will allow individuals to be identified.

Where can I get further information or discuss any problems?

Please contact a member of the fetal centre on 0151 702 4608 to discuss any questions or worries about the study, or if you have any complaints. If your concerns are not resolved, please contact Patient Advisory Liaison Services (PALS) on 0151 702 4353, if you have any concerns regarding the care you have received, or as an initial point of contact if you have a complaint. You can also visit PALS by asking at the hospital reception.

Who is organising and funding the research?

This research is organised by the Wellbeing-Harris Preterm Birth Research Centre part of the University of Liverpool Centre for Women and Children's Health Research at the Liverpool Women's Hospital.

Who has reviewed the study?

All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by **NRES Committee**.

Thank you for taking the time to read this information leaflet.

Appendix C

PTL Clinic - Patient information
Version:2.2

Date: 22.01.2016

Project ID number:11/NW/0720

Liverpool Women's NHS
NHS Foundation Trust
Crown Street
Liverpool
L8 7SS
Tel: 0151 708 9988
www.lwh.nhs.uk

The Development of Novel Biomarkers for Preterm Labour

Dr Angharad Care – Clinical Research Fellow, Liverpool Women's Hospital

Professor Zarko Alfirevic – Professor in Fetal & Maternal Medicine, Liverpool Women's Hospital

We are inviting many of the women who attend our unit to take part in a research study. Before you decide whether or not to take part it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. One of our team will go through the information sheet with you and answer any questions you may have.

Why are we doing the study?

The aim of this study is to develop earlier and safer ways of detecting problems in pregnancy. Research has shown that it may be possible to use a blood tests, vaginal swabs, urine or stool taken from the mother early in pregnancy to gather information which may help us work out why preterm birth happens and to predict the chances of preterm labour. We will do this by assessing a number of 'biomarkers' within these samples including different proteins, different types of bacteria and DNA. We will be taking samples at approximately 16 and 20 weeks. We want to identify why some women fail to respond to certain treatments for preterm labour, so we can work out which treatments are best for which women.

There is also evidence that women with less acidity (a higher pH) or certain types of bacteria in the vagina may be more at risk of preterm delivery. We wish to evaluate the ability of vaginal pH and types of bacteria to predict preterm birth or a short cervix. We will perform a vaginal swabs for pH and bacterial species check when you come to clinic at 16 weeks and 20 weeks of your pregnancy. This is not a swab for sexually transmitted diseases, if this is something you want this should be done at a local family planning or sexual health clinic.

There is a link between bacteria, infection and preterm birth but it is not clear how the bacteria cause preterm birth or how bacteria gets into the womb. Certain "pro-biotics" are being developed for good gut health and help with digestive problems. In this project we are also analysing bacteria in urine (wee) and stool (poo) on top of vaginal bacteria. If there are "bad" bacteria that are linked with preterm birth we may be able to design a "probiotic" to replace the bacteria in the tummy that may be causing preterm birth. We

are also collecting stool samples which can be done at 16 and 20 weeks of pregnancy at home with our specially designed kits and posted to our lab.

It might be that a combination of these “biomarkers” might allow us to predict preterm birth, so our statisticians will be putting all this data together to see if combinations of this information that we find out about you can predict if women are protected from preterm birth or if women are at risk.

In the future this data may also help us target specific treatments to women to prevent preterm birth if we think we can work out why preterm birth may be happening.

Before we can offer new tests routinely it is important to ensure they work well and are accurate. To do this we need the help of women who are at high risk of preterm labour, either because of a previous preterm birth or cervical surgery.

Why have I been chosen?

We are inviting all women over 18 years of age attending the preterm labour clinic to take part in the study.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and asked to sign a consent form. If you decide to take part you are free to withdraw at any time without giving a reason. If you do not feel able to take part it will not in any way affect the care you or your family receives.

What will happen to me if I take part?

If you agree to take part we will ask you to donate additional blood samples; 24ml (5 teaspoons) of blood will be taken at 16 weeks and 20 weeks at your PTL clinic appointments. We will arrange for you to have vaginal swab at the time of your internal scan to screen for bacterial vaginosis and provide you with treatment if required. As an optional part of the study, we will ask you provide a urine sample and stool sample which can be done either at clinic or at home with our specially designed kits and posted to the lab. We will not be testing for STI’s, genetic disease or paternity on any of these samples.

Will my taking part in this study be kept confidential?

Yes. We will follow ethical and legal practice and all information will be handled in confidence. Any information you give us will only be used by the research team in the course of the research to develop these new tests. Any samples and data stored will be stored securely. They will be coded, and no personal data (name and address) will be available to the researchers. However, if DNA analysis provides clinically relevant information we will inform your medical doctor

What are the possible benefits of taking part?

The results of this research will not be available in the course of your pregnancy and will not directly benefit you. We hope that the results of the study overall will enable us to improve antenatal care provided to women by developing safer prenatal tests that will help us detect pregnancies at risk of preterm birth. We will ensure that your doctor is informed of any progress that means these new tests could be available for you in future pregnancies.

What are the possible disadvantages and risks of taking part?

Providing samples for research will add on additional time to your clinic appointments. An appointment may last for up to 45 mins. Blood taking can be uncomfortable and this will be carried out by someone who is skilled in venepuncture (taking blood). Some people may experience bruising at the site which will resolve over a few days. Ultrasound has an excellent safety record and will not harm your baby.

What will happen if I don't want to continue in the study?

You are free to withdraw at anytime. If you withdraw from the study we will not access any further samples and will destroy any of your samples that were collected for the study.

What will happen to any samples I give?

Samples will be collected specifically for this study. The samples will be coded and no personal data (name and address) will be stored with the sample. The development of these tests will be done in laboratories in the UK, Finland and the Germany. We also ask whether you would be willing to gift samples to be used for other ethically approved research studies into pregnancy problems. Your samples will only be used in research studies designed to develop these new methods for the early detection of pregnancy complications.

What will happen to the results of the research study?

The results from our project will be published as research papers in medical journals. No data will be published that will allow individuals to be identified.

Where can I get further information or discuss any problems?

Please contact a member of the fetal centre on 0151 702 4608 to discuss any questions or worries about the study, or if you have any complaints. If your concerns are not resolved, please contact Patient Advisory Liaison Services (PALS) on 0151 702 4353, if you have any concerns regarding the care you have received, or as an initial point of contact if you have a complaint. You can also visit PALS by asking at the hospital reception.

Who is organising and funding the research?

This research is organised by the Department of Women's and Children's Health, University of Liverpool.

Who has reviewed the study?

All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by NRES Committee North West - Liverpool Central **NRES Committee**.

Thank you for taking the time to read this information leaflet.

Appendix D

THE UNIVERSITY of LIVERPOOL

Liverpool Women's NHS
NHS Foundation Trust

Study Number:
Patient Identification Number for this trial:

CONSENT FORM

Title of Project: **Developing novel biomarkers for the prediction of preterm labour**

Name of Researcher:
Dr Angharad Care, Clinical Research Fellow, University of Liverpool/Liverpool Women's Hospital
Dr Andrew Sharp, Clinical Lecturer, University of Liverpool/Liverpool Women's Hospital
Prof Zarko Alfirevic, Professor of Obstetrics, University of Liverpool/Liverpool Women's Hospital

Please initial box

1. I confirm that I have read and understand the information sheet dated …

(version ……) for the above study and have had the opportunity to ask questions.

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.

3. I give permission for my medical notes and data collected during this research to be looked at by responsible individuals from the University of Liverpool & Liverpool Women's Hospital or from regulatory authorities where it is relevant to my taking part in research.

4. I agree for DNA/RNA testing on my blood samples.

5. I give permission for the researcher to contact my medical practitioner in the event of clinically significant findings from this research.

6. I give permission for my samples to be sent outside of the UK

7. Once we carry out the study on the samples you kindly donate, if there is any surplus sample it will be stored in the Liverpool Women's Research Tissue Bank, of which the University of Liverpool is the custodian.

_____	_____	_____
Name of Patient	Date	Signature
_____	_____	_____
Name of Person taking consent (if different from researcher)	Date	Signature
_____	_____	_____
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with hospital notes

Appendix E

Preterm Labour Biomarker Study

CONTROL Recruitment Data Collection Sheet

Assigned Study Number:.....

Date:...../...../.....

Demographics

Age:.....yrs

Height (cm):

Weight (kg):

Smoking: Yes / No Per day: 1-5 / 6-10 / 10-15 / >15

Ethnicity:

Obstetric History

Gestation:.....wks…….days Confirmed EDD (DD/MM/YYYY): __ __/__ __/__ __ __ __

G.....P..... EDD by scan / menstrual dating / IVF (delete as appropriate)

No. of previous pregnancies complicated by:

Previous spontaneous preterm birth <34 weeks (n)	Gestation:
Previous PPROM <34weeks followed by sPTB(n)	Gestation:
Previous PPROM <34 weeks with IOL (n)	Gestation:
Clinical/placental histopathological chorioamnionitis (n)	
Birthweight centile <10th (n)	
Gestational Hypertension (n)	
Gestational diabetes requiring metformin/insulin therapy (n)	
Placental Abruption (n)	
Caregiver initiated preterm birth < 37 weeks (n)	
Multiple Pregnancy (n)	
Intrauterine Fetal Death (n)	
Polyhydramnios (n)	
Short Cervix (requiring treatment) (n)	
Surgical termination of pregnancy (n)	

History recurrent miscarriage (3 or more consecutive): Yes No

Medical History

History of cervical surgery? (circle) Yes / No

If yes, please specify: Single LLETZ Multiple LLETZ Knife Cone Biopsy

Other / Year:.....

NB. If cervical surgery only indication for referral – NOT eligible for recruitment

Please list any medical conditions:

Medication currently taking / taken in last 72 hours (give time, date and route):

Antibiotic course in pregnancy? Yes / No
If yes, dates taken (if multiple, record date of most recent): ……../……../…….. - ……../……../……..

Type of antibiotic:…….. PO / PV / topical/ IV (delete as appropriate)

Cervical length:……mm

Current Pregnancy Screening

Hepatitis B screening result: Positive Negative Not Done

Hepatitis C screening result: Positive Negative Not Done

HIV screening results: Positive Negative Not Done

Invasive Testing

Invasive testing or procedures performed (CVS, amniodrainage): Performed Not Done

Diet History:

Vegetarian Yes No

Vegan Yes No

Any use of probiotics in pregnancy? Yes No

If so what taken and when (freetext):

Other specific dietary requirement / intolerance (freetext):

Sexual history:

Have you had sex in the last 48 hours?

No

With condom

No condom

Any concerns about vaginal discharge or soreness in past week? Yes/ No

What is the concern? Smell/ excessive volume /itch/sore/Other (free text)

Samples taken (visit A):

Urine sample	Yes	Not done	
Fetal fibronectin:	Yes	Not done	Result (ng/ml):
HVS	Yes	Not done	Result:
			BV: yes no Treatment: yes no

Vaginal pH:…….

Vaginal dry swab:	x1	x2	Not done
Vaginal RNA swab:	yes	Not done	
Stool sample:	yes	no	
Plasma	yes	no	
Serum	yes	no	
RNA	yes	no	
Whole blood	yes	no	

Preterm Labour Biomarker Study

VISIT 2: Data Collection Sheet

Study Number:.....

Date:...../...../.....

Gestation:.....wks ……days

Cervical length:……mm

Samples taken (visit B):

Urine sample Yes Not done

Fetal fibronectin: Yes Not done Result (ng/ml):

HVS Yes Not done Result:

BV: yes no Treatment: yes no

Vaginal pH:…….

Vaginal dry swab: x1 x2 Not done

Vaginal RNA swab: yes Not done

Stool sample: yes no

Plasma yes no

Serum yes no

RNA yes no

Whole blood yes no

Is this patient currently on treatment for short cervical length? Yes / No

If yes, please specify:

Are there any other pathologies or concerns with this pregnancy? Yes / No

What medications have been taken within last 72hrs? (please detail dose, time, route)

Any antibiotic courses since last study visit? Yes / No

Details: Dates taken: ……../……../…….. - ……../……../……..
Name:_____

Any use of probiotics since last study visit? Yes No

If so what taken and when (freetext):

Have you had sex in the last 48 hours?

No / With condom /No condom

Any concerns about vaginal discharge or soreness in past week? Yes/ No

What is the concern? Smell/ excessive volume /itch/sore/Other (free text)

Follow up Data

Gestational hypertension	Yes No	
Pre-eclampsia	Yes No	
PPROM	Yes No	
Low dose aspirin commenced <16+0	Yes No	
Prophylactic dose LMWH	Yes No	
Gestational diabetes requiring metformin/insulin	Yes No	
Antepartum haemorhage > 20+0	Yes No	
Corticosteroids for fetal lung maturity	Yes No	Gestational Age when steroids given:
Treatment with a cervical cerclage	Yes / No	If yes, gestational age at Rx:
Treatment with vaginal progesterone	Yes / No	If yes, gestational age at Rx:
Treatment with arabin pessary	Yes / No	If yes, gestational age at Rx:

Other (please specify):

Delivery Data

Gestational Age at delivery

Birthweight Sex

Appendix F

Preterm Labour Biomarker Study

HIGH RISK CASE Recruitment Data Collection Sheet

Assigned Study Number:.....

Date:...../...../.....

Demographics

Age:.....yrs

Height (cm):

Weight (kg):

Smoking: Yes / No Per day: 1-5 / 6-10 / 10-15 / >15

Ethnicity:

Obstetric History

Gestation:.....wks…….days Confirmed EDD (DD/MM/YYYY): __ __/__ __/__ __ __ __

G.....P..... EDD by scan / menstrual dating / IVF (delete as appropriate)

No. of previous pregnancies complicated by:

Previous spontaneous preterm birth <34 weeks (n)	Gestation:
Previous PPROM <34weeks followed by sPTB(n)	Gestation:
Previous PPROM <34 weeks with IOL (n)	Gestation:
Clinical/placental histopathological chorioamnionitis (n)	
Birthweight centile <10th (n)	
Gestational Hypertension (n)	
Gestational diabetes requiring metformin/insulin therapy (n)	
Placental Abruption (n)	
Caregiver initiated preterm birth < 37 weeks (n)	
Multiple Pregnancy (n)	
Intrauterine Fetal Death (n)	
Polyhydramnios (n)	
Short Cervix (requiring treatment) (n)	
Surgical termination of pregnancy (n)	

History recurrent miscarriage (3 or more consecutive): Yes No

Medical History

History of cervical surgery? (circle) Yes / No

If yes, please specify: Single LLETZ Multiple LLETZ Knife Cone Biopsy

Other / Year:.....

NB. If cervical surgery only indication for referral – NOT eligible for recruitment

Please list any medical conditions:

Medication currently taking / taken in last 72 hours (give time, date and route):

Antibiotic course in pregnancy? Yes / No
If yes, dates taken (if multiple, record date of most recent): ……../……../…….. - ……../……../……..

Type of antibiotic:…….. PO / PV / topical/ IV (delete as appropriate)

Cervical length:…….mm

Current Pregnancy Screening

Hepatitis B screening result: Positive Negative Not Done

Hepatitis C screening result: Positive Negative Not Done

HIV screening results: Positive Negative Not Done

Invasive Testing

Invasive testing or procedures performed (CVS, amniodrainage): Performed Not Done

Diet History:

Vegetarian Yes No

Vegan Yes No

Other specific dietary requirement / intolerance (freetext):

Any use of probiotics in pregnancy? Yes/ No

If so what taken and when (free text)

Sexual history:

Have you had sex in the last 48 hours?

No/ With condom/ No condom

Any concerns about vaginal discharge or soreness in past week? Yes/ No

What is the concern? Smell/ excessive volume /itch/sore/Other (free text)

Samples taken (visit A):

Urine sample Yes Not done

Fetal fibronectin: Yes Not done Result (ng/ml):

HVS Yes Not done Result:

BV: yes no Treatment: yes no

Vaginal pH:…….

Vaginal dry swab: x1 x2 Not done

Vaginal RNA swab: yes Not done

Stool sample: yes no

Plasma yes no

Serum yes no

RNA yes no

Whole blood yes no

Preterm Labour Biomarker Study

VISIT 2: Data Collection Sheet

Study Number:.....

Date:...../...../.....

Gestation:.....wks ……days

Cervical length:…….mm

Samples taken (visit B):

Urine sample Yes Not done

Fetal fibronectin: Yes Not done Result (ng/ml):

HVS Yes Not done Result:

BV: yes no Treatment:

yes no

Vaginal pH:…….

Vaginal dry swab: x1 x2 Not done

Vaginal RNA swab: yes Not done

Stool sample: yes no

Plasma yes no

Serum yes no

RNA yes no

Whole blood yes no

Is this patient currently on treatment for short cervical length? Yes / No

If yes, please specify:

Are there any other pathologies or concerns with this pregnancy? Yes / No

What medications have been taken within last 72hrs? (please detail dose, time, route)

Any antibiotic courses since last study visit? Yes / No

Details: Dates taken: ……../……../…….. - ……../……../……..

Name:_____

Any use of probiotics since last study visit? Yes No

If so what taken and when (freetext):

Have you had sex in the last 48 hours?

No / With condom /No condom

Any concerns about vaginal discharge or soreness in past week? Yes/ No

What is the concern? Smell/ excessive volume /itch/sore/Other (free text)

Follow up Data

Gestational hypertension	Yes No	
Pre-eclampsia	Yes No	
PPROM	Yes No	
Low dose aspirin commenced <16+0	Yes No	
Prophylactic dose LMWH	Yes No	
Gestational diabetes requiring metformin/insulin	Yes No	
Antepartum haemorhage > 20+0	Yes No	
Corticosteroids for fetal lung maturity	Yes No	Gestational Age when steroids given:
Treatment with a cervical cerclage	Yes / No	If yes, gestational age at Rx:
Treatment with vaginal progesterone	Yes / No	If yes, gestational age at Rx:
Treatment with arabin pessary	Yes / No	If yes, gestational age at Rx:

Other (please specify):

Delivery Data

Gestational Age at delivery

Birthweight

Sex

Appendix G

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Acinetobacter genus	Pathobionts	Proteobacteria/Gammaproteobacteria (Pseudomonadales)	yes	Can reside, possibly indigenously, on the human skin and in the human respiratory tract. Can cause nosocomial infections such as bacteremia, secondary meningitis, pneumonia, and urinary tract infections in humans.
Actinobaculum massiliense	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	Gram-stain-positive cocci associated with recurrent cystitis and a case of pelvic inflammatory disease.
Actinomyces europaeus	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	
Actinomyces genus	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	
Actinomyces graevenitzii	BV	Actinobacteria/Actinobacteria (Actinomycetales)	yes	
Actinomyces naeslundii	BV	Actinobacteria/Actinobacteria (Actinomycetales)	yes	
Actinomyces neuii	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	
Actinomyces odontolyticus	BV	Actinobacteria/Actinobacteria (Actinomycetales)	yes	
Actinomyces turicensis	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	
Actinomyces urogenitalis	BV	Actinobacteria/Actinobacteria (Actinomycetales)	yes	
Actinotignum genus	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	Facultative anaerobic bacterium, part of the urinary bacterium of healthy patients but associated with UTIs
Actinotignum schaalii	BV	Actinobacteria/Actinobacteria (Actinomycetales)	yes	
Actinotignum urinale	BV	Actinobacteria/Actinobacteria (Actinomycetales)	yes	
Aerococcus christensenii	BV	Firmicutes/Bacilli (Lactobacillales)	no	Facultative anaerobe that has often been described in molecular studies of BV.
Aerococcus genus	BV	Firmicutes/Bacilli (Lactobacillales)	no	Facultative anaerobe that has often been described in molecular studies of BV.
Aerococcus urinae	BV	Firmicutes/Bacilli (Lactobacillales)	no	
Agathobacter genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Anaerobe found in stool culture
Akkermansia muciniphila	Other	Verrucomicrobia/Verrucomicrobiae (Verrucomicrobiales)	yes	Anaerobe found in gut
Alloprevotella genus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	Often described in molecular studies of BV.

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Alloscardovia omnicolens	Other	Actinobacteria/Actinobacteria (Bifidobacteriales)	no	Closely related to Bifidobacteria, which are considered to have probiotic properties.
Anaerococcus genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Often described in molecular studies of BV.
Anaerococcus hydrogenalis	BV	Firmicutes/Clostridia (Clostridiales)	no	
Anaerococcus lactolyticus	BV	Firmicutes/Clostridia (Clostridiales)	no	
Anaerococcus nagyae	BV	Firmicutes/Clostridia (Clostridiales)	no	
Anaerococcus octavius	BV	Firmicutes/Clostridia (Clostridiales)	no	
Anaerococcus prevotii	BV	Firmicutes/Clostridia (Clostridiales)	no	
Anaerococcus provenciensis	BV	Firmicutes/Clostridia (Clostridiales)	no	
Anaerococcus senegalensis	BV	Firmicutes/Clostridia (Clostridiales)	no	Anaerobe described in stool samples
Anaerococcus vaginalis	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Anaeroglobus geminatus	BV	Firmicutes/Negativicutes (Selenomonadales)	no	Anaerobe described in periodontitis.
Anaeroglobus genus	BV	Firmicutes/Negativicutes (Selenomonadales)	no	Anaerobe described in periodontitis. Occurs in gastrointestinal tract and presumably in oral cavity
Anaerostipes hadrus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Strict anaerobe described in the gut/faeces.
Arcanobacterium genus	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	Often described in molecular studies of BV.
Arcanobacterium urinimassiliense	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	
Atopobium deltae	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	no	
Atopobium genus	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	no	Includes A. vaginae, which is known to be important in BV.
Atopobium minutum	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	no	
Atopobium parvulum	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	yes	
Atopobium rimae	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	yes	
Atopobium vaginae	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	no	
Bacillus subtilis	Other	Firmicutes/Bacilli (Bacillales)	yes	
Bacteroidales order	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	Bacteroidales have commonly been described in BV.
Bacteroides fragilis	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Bacteroides genus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Bacteroides stercoris	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Bacteroides uniformis	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Bacteroides vulgatus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Bergeyella genus	Other	Bacteroidetes/Bacteroidia (Flavobacteriales)	yes	Aeorbic. Rarely animal bites can become infected with these organisms
Bifidobacterium bifidum	Other	Actinobacteria/Actinobacteria (Bifidobacteriales)	no	
Bifidobacterium breve	Other	Actinobacteria/Actinobacteria (Bifidobacteriales)	no	
Bifidobacterium dentium	Other	Actinobacteria/Actinobacteria (Bifidobacteriales)	no	
Bifidobacterium genus	Other	Actinobacteria/Actinobacteria (Bifidobacteriales)	no	
Bifidobacterium longum	Other	Actinobacteria/Actinobacteria (Bifidobacteriales)	no	
Bilophila wadsworthia	BV	Proteobacteria/Deltaproteobacteria (Desulfovibrionales)	yes	Anaerobic, isolated from gastrointestinal, genital, and oral cavities of humans
Blautia faecis	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Blautia genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Blautia massiliensis	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Brevibacterium genus	Other	Actinobacteria/Actinobacteria (Micrococcales)	no	
Brevibacterium ravenspurgense	Other	Actinobacteria/Actinobacteria (Micrococcales)	no	
Bulleidia extructa	BV	Firmicutes/Erysipelotrichia (Erysipelotrichales)	yes	
Bulleidia genus	BV	Firmicutes/Erysipelotrichia (Erysipelotrichales)	no	Strict anaerobe described in periodontal disease.
Burkholderia-Caballeronia-Paraburkholderia genus	Other	Proteobacteria/Gammaproteobacteria (Betaproteobacteriales)	no	Not all betaproteobacteria are pathobionts
BVAB TM7-H1	BV	Unclassified (TM7 division)	yes	First identified in BV patients.
BVAB1	BV	Firmicutes/Clostridia (Clostridiales)	no	
BVAB2	BV	Firmicutes/Clostridia (Clostridiales)	no	Known to be important in BV.
Campylobacter genus	Pathobionts	Epsilonbacteraeota/Campylobacteria (Campylobacterales)	no	
Campylobacter hominis	Pathobionts	Epsilonbacteraeota/Campylobacteria (Campylobacterales)	no	

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Campylobacter ureolyticum	Pathobionts	Epsilonbacteraeota/Campylobacteria (Campylobacterales)	no	
Catonella genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Chlamydia trachomatis	Pathobionts	Chlamydiae/Chlamydiae (Chlamydiales)	yes	
Christensenellaceae_R-7_group genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Often described in molecular studies of BV.
Clostridiaceae family	BV	Firmicutes/Clostridia (Clostridiales)	yes	Many genera in this family have been described in BV.
Clostridiales order	BV	Firmicutes/Clostridia (Clostridiales)	no	Most genera of this order have been described in molecular studies of BV.
Clostridiales XIW5054 genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Strict anaerobe, closely related to Atopobium which is known to be important in BV.
Clostridium sensu_stricto_2 genus	BV	Firmicutes/Clostridia (Clostridiales)	no	
Clostridium_sensu_stricto_1 perfringens	BV	Firmicutes/Clostridia (Clostridiales)	no	
Collinsella aerofaciens	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	no	
Coprococcus_2 eutactus	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Corynebacterium_1 appendicis	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	yes	
Corynebacterium_1 massiliense	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	yes	
Corynebacterium genus	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	yes	Aerobe or facultative anaerobe. Common skin bacterium in humans.
Corynebacterium glucuronolyticum	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium pyruviciproducens	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_1 imitans	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_1 pseudogenitalium	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_1 tuscaniense	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_1 urealyticum	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	yes	
Corynebacterium_2 aurimucosum	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_2 coyleae	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_2 genitalium	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_2 genus	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Corynebacterium_2 jeikeium	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_2 riegelii	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Corynebacterium_2 sundsvallense	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Cutibacterium avidum	Other	Actinobacteria/Actinobacteria (Propionibacteriales)	yes	Skin commensal that can be associatred with opportunistic infections
Cutibacterium genus	Other	Actinobacteria/Actinobacteria (Propionibacteriales)	no	
Dermabacter jinjuensis	Other	Actinobacteria/Actinobacteria (Micrococcales)	yes	Aerobe, occasionally faculative anaerobe. Often present on the skin.
Dermacoccus nishinomiyaensis	Other	Actinobacteria/Actinobacteria (Micrococcales)	yes	Aerobe, skin commensal. Normally considered nonpathogenic for humans
Dialister genus	BV	Firmicutes/Negativicutes (Selenomonadales)	no	Known to be important in BV.
Dialister invisus	BV	Firmicutes/Negativicutes (Selenomonadales)	yes	
Dialister Micraerophilus	BV	Firmicutes/Negativicutes (Selenomonadales)	no	
Dialister pneumosintes	BV	Firmicutes/Negativicutes (Selenomonadales)	yes	
Dialister propionicifaciens	BV	Firmicutes/Negativicutes (Selenomonadales)	no	
Dorea longicatena	BV	Firmicutes/Clostridia (Clostridiales)	no	Strict anaerobe, member of the Lachnospiraceae family.
Eggerthellaceae DNF00810 genus	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	no	Described in BV
Enterobacteriaceae Escherichia/Shigella genus	Pathobionts	Proteobacteria/Gammaproteobacteria (Enterobacteriales)	no	
Enterococcus faecalis	Pathobionts	Firmicutes/Bacilli (Lactobacillales)	no	
Enterococcus genus	Pathobionts	Firmicutes/Bacilli (Lactobacillales)	yes	
Ezakiella genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Strict anaerobe, only recently described in vagina; likely to be BV-related.
Facklamia genus	BV	Firmicutes/Bacilli (Lactobacillales)	yes	
Facklamia hominis	BV	Firmicutes/Bacilli (Lactobacillales)	no	
Facklamia ignava	BV	Firmicutes/Bacilli (Lactobacillales)	no	Facultative anaerobe closely related to Aerococcus, which has often been described in BV.
Faecalibacterium genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Often described in molecular studies of BV.
Faecalibacterium prausnitzii	BV	Firmicutes/Clostridia (Clostridiales)	no	
Fastidiosipila genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Known to be important in BV.

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Finegoldia genus	BV	Firmicutes/Clostridia (Clostridiales)	no	
Finegoldia magna	BV	Firmicutes/Clostridia (Clostridiales)	no	Often described in molecular studies of BV.
Fusicatenibacter genus	BV	Fusobacteria/Fusobacteriia (Fusobacteriales)	no	Strict anaerobes. In same order as Leptotrichia and Sneathia, which have commonly been described in BV.
Fusicatenibacter saccharivorans	BV	Firmicutes/Clostridia (Clostridiales)	no	
Fusobacterium nucleatum	BV	Fusobacteria/Fusobacteriia (Fusobacteriales)	no	
Gallicola genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Strict anaerobe related to Anaerococcus and Peptoniphilus, which have often been described in BV.
Gardnerella genus	BV	Actinobacteria/Actinobacteria (Bifidobacteriales)	no	Known to be important in BV.
Gardnerella vaginalis	BV	Actinobacteria/Actinobacteria (Bifidobacteriales)	no	
Gemella asaccharolytica	BV	Firmicutes/Bacilli (Bacillales)	no	
Gemella genus	BV	Firmicutes/Bacilli (Bacillales)	no	Often described in molecular studies of BV.
Gemella sanguinis	BV	Firmicutes/Bacilli (Bacillales)	yes	
Globicatella genus	BV	Firmicutes/Bacilli (Lactobacillales)	yes	Facultative anaerobe closely related to Aerococcus, which has often been described in BV.
Granulicatella elegans	BV	Firmicutes/Bacilli (Lactobacillales)	yes	Facultative anaerobe closely related to Abiotrophia, which has been described in the mouth, gut, and urogenital tract.
Haemophilus genus	Pathobionts	Proteobacteria/Gammaproteobacteria (Pasteurellales)	no	Known pathobiont.
Haemophilus parainfluenzae	Pathobionts	Proteobacteria/Gammaproteobacteria (Pasteurellales)	yes	
Helcococcus genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Strict anaerobe of Peptostreptococcaceae family, which have often been described in molecular studies of BV.
Howardella genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Strict anaerobe, described in animal and human gut/faeces.
Jonquetella anthropi	BV	Synergistetes/Synergistia (Synergistales)	no	Strict anaerobe often described in periodontitis.
Kocuria genus	Other	Actinobacteria/Actinobacteria (Micrococcales)	no	Aerobe, often present on the skin.
Lachnospiraceae family	BV	Firmicutes/Clostridia (Clostridiales)	yes	Many genera in this family have been described in BV.
Lactobacillus acidophilus	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus acidophilus/gallinarum/jensenii	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus amylovorus	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus coleohominis	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Lactobacillus crispatus	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus crispatus/acidophilus/amylovorus/helveticus/gallinarium	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus delbrueckii	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus fementum	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus gasseri	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus gasseri/hominis	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus helveticus	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus iners	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus jensenii	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus jensenii/psittaci	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus mucosae	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus oris	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus psittaci/amylovorus/crispartus	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus reuteri	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lactobacillus rhamnosus	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus salivarius	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	yes	
Lactobacillus vaginalis	Lactobacillus	Firmicutes/Bacilli (Lactobacillales)	no	
Lawsonella clevelandensis	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	
Lawsonella genus	Other	Actinobacteria/Actinobacteria (Corynebacteriales)	no	Facultative anaerobe closely related to Corynebacterium, which has been isolated from skin abscesses.
Megasphaera genus	BV	Firmicutes/Negativicutes (Selenomonadales)	no	Known to be important in BV.
Megasphaera massiliensis	BV	Firmicutes/Negativicutes (Selenomonadales)	yes	
Megasphaera micronuciformis	BV	Firmicutes/Negativicutes (Selenomonadales)	yes	
Micrococcales order	Other	Actinobacteria/Actinobacteria (Micrococcales)	no	Aerobes, often present on the skin.
Micrococcus genus	Other	Actinobacteria/Actinobacteria (Micrococcales)	yes	Aerobes, often present on the skin.

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Mobiluncus curtisii	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	
Mobiluncus genus	BV	Actinobacteria/Actinobacteria (Actinomycetales)	yes	Known to be important in BV.
Mobiluncus mulieris	BV	Actinobacteria/Actinobacteria (Actinomycetales)	yes	
Mogibacterium timidum	Other	Firmicutes/Clostridia (Clostridiales)	yes	Oxygen-producting environmental bacteria that have occasionally been described in the gut.
Moryella genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Often described in molecular studies of BV.
Moryella indoligenes	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Murdochiella asaccharolytica	BV	Firmicutes/Clostridia (Clostridiales)	no	
Murdochiella genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Strict anaerobes, isolated from wound and BV specimens.
Murdochiella massiliensis	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Muribaculaceae family	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	Anaerobe commonly found in gut
Mycoplasma hominis	BV	Tenericutes/Mollicutes (Mycoplasmatales)	no	
Mycoplasma spermatophilum	BV	Tenericutes/Mollicutes (Mycoplasmatales)	yes	
Negativicoccus genus	BV	Firmicutes/Negativicutes (Selenomonadales)	no	Often described in molecular studies of BV.
Negativicoccus succinicivorans	BV	Firmicutes/Negativicutes (Selenomonadales)	no	
Neisseria genus	Pathobionts	Proteobacteria/Gammaproteobacteria (Betaproteobacteriales)	yes	Known pathogen, but included in the pathobiont group due to small numbers of pathogens identified.
Neisseriaceae family	Pathobionts	Proteobacteria/Gammaproteobacteria (Betaproteobacteriales)	no	Known pathogen, but included in the pathobiont group due to small numbers of pathogens identified.
Nosocomiicoccus genus	Other	Firmicutes/Bacilli (Bacillales)	yes	Aerobe of the Staphylococcaceae family, not known to be pathogenic to humans.
Oligella urethralis	Pathobionts	Proteobacteria/Gammaproteobacteria (Betaproteobacteriales)	no	Aerobe in human genital tract, described in bacteraemia
Olsenella genus	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	yes	Strict anaerobe, described in periodontitis.
Olsenella uli	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	yes	
Parvimonas genus	BV	Firmicutes/Clostridia (Clostridiales)	no	
Parvimonas micra	BV	Firmicutes/Clostridia (Clostridiales)	no	Often described in molecular studies of BV.
Peptococcus genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Often described in molecular studies of BV.
Peptococcus niger	BV	Firmicutes/Clostridia (Clostridiales)	yes	

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Peptoniphilus coxii	BV	Firmicutes/Clostridia (Clostridiales)	no	
Peptoniphilus duerdenii	BV	Firmicutes/Clostridia (Clostridiales)	no	
Peptoniphilus genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Often described in molecular studies of BV.
Peptoniphilus harei	BV	Firmicutes/Clostridia (Clostridiales)	no	
Peptoniphilus lacrimalis	BV	Firmicutes/Clostridia (Clostridiales)	no	
Peptoniphilus massiliensis	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Peptoniphilus obesi	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Peptostreptococcaceae family	BV	Firmicutes/Clostridia (Clostridiales)	no	Includes genera Anaerococcus, Finegoldia, Fusobacter, Parvimonas, Peptoniphilus, and Peptostreptococcus, all of which have been described in BV.
Peptostreptococcus anaerobius	BV	Firmicutes/Clostridia (Clostridiales)	no	
Peptostreptococcus genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Strict anaerobes, described in the gut/faeces.
Porphyromonas asaccharolytica	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Porphyromonas genus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	Known to be important in BV.
Porphyromonas somerae	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Porphyromonas uenonis	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella amnii	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella bivia	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella buccalis	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella disiens	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella genus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	Often described in molecular studies of BV and in periodontitis.
Prevotella ihumii	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Prevotella nigrescens	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Prevotella timonensis	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella_2 conceptionensis	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Prevotella_3 genus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Prevotella_6 corporis	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella_6 salivae	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Prevotella_7 bergensis	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella_7 genus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella_7 histicola	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Prevotella_7 melaninogenica	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotella_8 denticola	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Prevotella_8 genus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	
Prevotellaceae family	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	yes	
Propionimicrobium lymphophilum	Other	Actinobacteria/Actinobacteria (Propionibacteriales)	no	Anaerobe, described in urinary tract infection.
Proteobacteria phylum	Other	Proteobacteria/NA (NA)	no	Not all proteobacteria in the vagina are pathobionts.
Pseudoclavibacter genus	Other	Actinobacteria/Actinobacteria (Micrococcales)	yes	Aerobe mostly isolated from environment
Pseudoglutamicibacter albus	Other	Actinobacteria/Actinobacteria (Micrococcales)	no	
Pseudoglutamicibacter cumminsii	Other	Actinobacteria/Actinobacteria (Micrococcales)	yes	
Pseudoglutamicibacter genus	Other	Actinobacteria/Actinobacteria (Micrococcales)	yes	Aerobe isolated from human blood and urine
Pseudomonas aeruginosa	Pathobionts	Proteobacteria/Gammaproteobacteria (Pseudomonadales)	yes	
Pseudoramibacter alactolyticus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Has been described in BV
Pyramidobacter genus	BV	Synergistetes/Synergistia (Synergistales)	yes	Strict anaerobe, described in gut/faeces and periodontitis.
Raoultella genus	Pathobionts	Proteobacteria/Gammaproteobacteria (Enterobacteriales)	yes	Known pathobiont.
Raoultella ornithinolytica	Pathobionts	Proteobacteria/Gammaproteobacteria (Enterobacteriales)	yes	
Rikenellaceae_RC9_gut_group genus	BV	Bacteroidetes/Bacteroidia (Bacteroidales)	no	Strict anaerobe, described as a common resident of animal and human guts.
Romboutsia genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Strict anaerobe, described as a resident of the human gut.
Rothia dentocariosa	BV	Actinobacteria/Actinobacteria (Micrococcales)	yes	Aerobe, described in periodontitis (and systemic infections in immunocompromised).
Rothia mucilaginosa	BV	Actinobacteria/Actinobacteria (Micrococcales)	yes	

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Ruminococcaceae_UCG-015 genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Often described in molecular studies of BV.
Ruminococcus genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Often described in molecular studies of BV.
Ruminococcus_2 bromii	BV	Firmicutes/Clostridia (Clostridiales)	yes	
S5-A14a genus	Other	Firmicutes/Clostridia (Clostridiales)	no	Oxygen-producting environmental bacteria that have occasionally been described in the gut.
Saccharimonadales order	Other	Patescibacteria/Saccharimonadia (Saccharimonadales)	yes	Rarely described environmental order
Salmonella enterica	Pathobionts	Proteobacteria/Gammaproteobacteria (Enterobacteriales)	yes	
Sarcina genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Strict anaerobe, described in gastrointestinal tract.
Scardovia wiggsiae	Other	Actinobacteria/Actinobacteria (Bifidobacteriales)	yes	Part of the Bifidobacteriales
Shuttleworthia genus	BV	Firmicutes/Clostridia (Clostridiales)	yes	Often described in molecular studies of BV.
Shuttleworthia satelles	BV	Firmicutes/Clostridia (Clostridiales)	yes	
Slackia exigua	BV	Actinobacteria/Coriobacteriia (Coriobacteriales)	no	Often described in molecular studies of BV.
Sneathia amnii	BV	Fusobacteria/Fusobacteriia (Fusobacteriales)	no	
Sneathia genus	BV	Fusobacteria/Fusobacteriia (Fusobacteriales)	no	Known to be important in BV.
Sneathia sanguinegens	BV	Fusobacteria/Fusobacteriia (Fusobacteriales)	no	
Solobacterium moorei	BV	Firmicutes/Erysipelotrichia (Erysipelotrichales)	yes	Strict anaerobes, associated with periodontitis.
Staphylococcus aureus	Pathobionts	Firmicutes/Bacilli (Bacillales)	yes	
Staphylococcus genus	Pathobionts	Firmicutes/Bacilli (Bacillales)	no	Known pathobionts; cause a wide variety of diseases in animals and humans. Not all are equally pathogenic but none are completely harmless.
Staphylococcus haemolyticus	Pathobionts	Firmicutes/Bacilli (Bacillales)	no	
Staphylococcus simulans	Pathobionts	Firmicutes/Bacilli (Bacillales)	yes	
Streptococcus agalactiae	Pathobionts	Firmicutes/Bacilli (Lactobacillales)	no	
Streptococcus anginosus	Pathobionts	Firmicutes/Bacilli (Lactobacillales)	no	
Streptococcus cristatus	Pathobionts	Firmicutes/Bacilli (Lactobacillales)	no	
Streptococcus genus	Pathobionts	Firmicutes/Bacilli (Lactobacillales)	no	Known pathobionts; cause a wide variety of diseases in animals and humans. Some streptococci are harmless but none in this list are.
Streptococcus pneumoniae	Pathobionts	Firmicutes/Bacilli (Lactobacillales)	yes	

List of taxa (alphabetical order)	Classification	Phylum/Class (Order) based on NCBI taxonomy browser	Minority taxon (<1% relative abundance in all study samples)?	Comments based on Bergey's manual and published literature
Streptococcus urinalis	Pathobionts	Firmicutes/Bacilli (Lactobacillales)	yes	
Subdoligranulum genus	BV	Firmicutes/Clostridia (Clostridiales)	no	Strict anaerobes, closely related to Blautia and Ruminococcus.
Succiniclasticum genus	BV	Firmicutes/Negativicutes (Selenomonadales)	yes	Anaerobe commonly found in gut
Sutterella genus	Other	Proteobacteria/Gammaproteobacteria (Betaproteobacteriales)	no	Closely related to Campylobacter but not often described as a pathobiont.
Sutterella wadsworthensis	Other	Proteobacteria/Gammaproteobacteria (Betaproteobacteriales)	yes	
Ureaplasma genus	BV	Tenericutes/Mollicutes (Mycoplasmatales)	no	Often described in molecular studies of BV.
Ureaplasma parvum	BV	Tenericutes/Mollicutes (Mycoplasmatales)	no	
Ureaplasma urealyticum	BV	Tenericutes/Mollicutes (Mycoplasmatales)	no	
Varibaculum anthropi	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	
Varibaculum cambriense	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	
Varibaculum genus	BV	Actinobacteria/Actinobacteria (Actinomycetales)	no	Often described in molecular studies of BV.
Veillonella atypica	BV	Firmicutes/Negativicutes (Selenomonadales)	no	Often described in molecular studies of BV.
Veillonella dispar	BV	Firmicutes/Negativicutes (Selenomonadales)	no	
Veillonella genus	BV	Firmicutes/Negativicutes (Selenomonadales)	no	
Veillonella montpellierensis	BV	Firmicutes/Negativicutes (Selenomonadales)	no	
Veillonella parvula	BV	Firmicutes/Negativicutes (Selenomonadales)	yes	
Veillonellaceae family	BV	Firmicutes/Negativicutes (Selenomonadales)	no	Often described in molecular studies of BV.

Table detailing all 276 ASVs that were identified through sequencing. Each ASV was allocated a classification which describes the assigned bacterial group, as descried in section 4.3.2.2. Classification was based on Bergey's manual[224] and available published literature.