DIGS STUDY PROTOCOL

Detecting Intra-uterine Group B Streptococcus (DIGS)

Grant title: Pre-labour invasion of the human uterus by *Streptococcus agalactiae*

Chief Investigator:
Professor Gordon CS Smith, Dept. of Obstetrics & Gynaecology, University of Cambridge

Co-investigators:
Professor D.S. Charnock-Jones, Dept. of Obstetrics & Gynaecology, University of Cambridge
Professor Julian Parkhill, Department of Veterinary Medicine, University of Cambridge
Professor Sharon Peacock, Dept. of Medicine, University of Cambridge
Professor David Henry Rowitch, Dept. of Paediatrics, University of Cambridge
Dr Paul Kirk, MRC Biostatistics Unit, Cambridge.

Statistician:
Dr Ulla Sovio, Dept. of Obstetrics & Gynaecology, University of Cambridge

Sponsor (joint):
Cambridge University Hospitals NHS Trust (CUHFT) & Cambridge University

Funder:
Medical Research Council

Centre (single):
The Rosie Hospital, Cambridge, UK

Table of contents

1. Introduction

1.1 The problem of GBS

The most common microbe causing severe neonatal sepsis is *Streptococcus agalactiae*, referred to as Group B Streptococcus (GBS)[1]. Invasive GBS in the infant is classified into early onset (within first week of life) and late onset disease (EOD and LOD, respectively). About 60% of invasive GBS is EOD. It is established that GBS EOD occurs through vertical transmission from the mother during the process of labour and delivery[1]. The mean colonisation rate in pregnant women is ~20%[2]. In the absence of intervention, the risk of vertical transmission if the mother is colonised is ~50% and ~1-2% of these infants will end up with a clinical infection. Hence, even in the absence of medical intervention, EOD is rare, affecting about 1 in 1000 neonates.

1.2 Current screening and intervention

Universal screening of all pregnant women for GBS is not currently recommended in the UK. The primary method for identifying women colonised by GBS in current UK practice is from a vaginal swab obtained in response to symptoms, such as vaginal discharge or suspected premature rupture of the fetal membranes. Detection of GBS in the mother is an indication for maternal intrapartum antibiotic prophylaxis (IAP) and this reduces the risk of neonatal infection by about 80%[3]. Antibiotic prophylaxis to prevent GBS EOD is not recommended for women being delivered by planned caesarean section as the risk of EOD is very low with this mode of delivery.

1.3 Organism factors determining pathogenicity

GBS isolates can be categorised into 10 serotypes, based on capsular polysaccharides, or into 11 major clonal complexes (CC), based on multilocus sequence typing (analysis of SNPs in bacterial housekeeping genes) and there is overlap between the two approaches. The combination of serotype III and CC17 may be over-represented in EOD as about 40% of GBS isolates from infants had this combination[4], which was about 5 times greater than observed in colonising strains. Moreover, all of the serotype III + CC17 isolates in that study expressed the *hvgA* gene, which is one of multiple GBS virulence proteins which promote invasive disease[5].

1.4 GBS and fetal inflammation

GBS can cause fetal organ damage indirectly through stimulating release of pro-inflammatory cytokines[5]. Cytokines associated with GBS infection include interleukin (IL)-1β, IL-6, IL-8, and tumour necrosis factor (TNF)-α[6]. Elevation of these cytokines is central to fetal inflammatory response syndrome (FIRS)[7] which is a state of systemic inflammation in the fetus. Current data support the concept that small numbers of GBS organisms may stimulate a disproportionate inflammatory response and that the inflammatory response can itself cause harm, separate from any direct effect of the bacteria[8].

1.5 Pre-labor invasion of the uterus by GBS

As described above, vertical transmission of GBS at the time of labour and birth is well recognised. However, we have demonstrated that GBS was also present in the placenta prior to

the onset of labour and rupture of the fetal membranes[9]. We also have unpublished preliminary data which demonstrates a number of additional key observations:
(i) GBS DNA is more commonly detected in the placenta among women who had a positive culture result for GBS during the pregnancy
(ii) there is a strong association between the presence of GBS in the placenta and the risk of neonatal morbidity
(iii) the association between placental GBS and morbidity is mediated through FIRS.
A key element in our preliminary data is that we used a novel, ultrasensitive GBS assay, which is ~350 times more sensitive than current molecular methods of detection. By using this assay we were able to demonstrate that there are many more cases of neonatal morbidity caused by clinically undetected GBS than there are cases of culture-proven neonatal sepsis. However, the preliminary data were obtained from women with both vaginal and caesarean deliveries and many of the latter procedures were performed after the onset of labour and rupture of the fetal membranes.

1.6 Over-arching aim of the DIGS study
This study will, for the first time, systematically analyse pre-labour invasion of the uterus by GBS by studying a large number of women being delivered by planned, pre-labour caesarean section, and will address the following questions:
(i) what is the relationship between maternal colonisation by GBS and pre-labour invasion of the intra-uterine tissues by GBS?
(ii) how does invasion affect the infant in terms of markers of intra-uterine inflammation and complications in the neonatal period?
(iii) can we identify genomic characteristics of GBS which promote invasion?
The findings of the study could have multiple positive impacts. We could discover cases of neonatal morbidity caused by occult GBS infection, including infants delivered by planned caesarean section. Our work could lead to the development of tests to better predict infants at risk of neonatal morbidity and could result in better targeting of neonatal antibiotic therapy. Finally, better ascertainment of the true burden of GBS could positively impact on trials of vaccination during pregnancy and trials of screening for GBS and treating women with IAP, which are currently hampered by the relative rarity of GBS EOD.

1.7 Other studies of samples obtained at the time of planned caesarean section
Planned caesarean section, like any surgical procedure, allows access to human tissue samples by a wide range of researchers. In the Rosie Hospital, Cambridge, there are two main groups of researchers who have an interest in collecting samples at this point: (i) the Cambridge Blood and Stem Cell Biobank based in the Haematology Department, and (ii) The Centre for Trophoblast Research (CTR) Placental Biobank. Neither of these groups need access to the mother or child clinical record and simply require anonymous biological samples obtained from the placenta and umbilical cord after delivery, combined with basic clinical and demographic information in relation to the pregnancy.

2. Overview of study design

We plan to study 1,800 women being delivered by planned, pre-labour caesarean section at the Rosie Hospital, Cambridge, UK between 2022 and 2025. We will obtain swabs from the mother's anogenital tract prior to the caesarean section to determine colonisation with GBS. We will obtain tissue samples at the time of the caesarean section, specifically, placenta, fetal membranes, amniotic fluid, umbilical cord and fetal (i.e. umbilical cord) blood. We will then review the case record of the baby and determine any neonatal complications and obtain maternal data to allow us to adjust analyses of neonatal outcome for obstetric and maternal characteristics (such as gestational age at delivery and maternal age). Using these data and biological samples, we will address the main aims of the study:

(1) To determine the relationships between maternal colonisation with GBS at the time of caesarean section and the presence of GBS in the placenta, fetal membranes, amniotic fluid and fetal blood.

(2) To determine the interrelationships between the presence of GBS in the intra-uterine tissues, pro-inflammatory cytokine levels in the same tissues, and the risk of neonatal morbidity.

(3) To compare the presence of genes associated with GBS pathogenicity in those with and without evidence of GBS DNA in the intra-uterine tissues among women colonised by GBS.

(4) To obtain anonymised samples of umbilical cord blood, and placenta, plus basic maternal and infant data, for the Cambridge Blood and Stem Cell Biobank and the CTR Placenta Biobank.

3. Selection, recruitment and withdrawal of participants

3.1 Eligibility criteria

Eligible women are those attending the Rosie Hospital, Cambridge, for a planned pre-labour caesarean section who provide written informed consent for participation. We will exclude women who are unable to consent (including those who are non-fluent English speakers), those aged <16 years, any woman who has had antibiotic treatment within the preceding week, and any women with an infectious condition which could represent a hazard (e.g. high risk carrier of viral hepatitis B).

3.2 Recruitment and consent

Women will be sent a letter of invitation and a patient information leaflet prior to attending the Rosie Hospital for their planned caesarean section. The information provided will describe the study and its rationale and it will explain what their participation in the study would involve if they provided consent. The invitation letter will explain that we would like to approach them to discuss the project in more detail when they attend hospital for their delivery. The communication will provide the recipient with information about how to decline to be approached on the day of their procedure. Specifically, they will be given study team communication details including an email address, a mobile phone number, a landline, and a postal address. If they do not wish to be approached on the day of their caesarean section, they can opt out of being approached by sending the instruction through any one of these methods. Women who do not opt out may be approached by a research nurse or midwife on the morning

of the procedure. Again, they will be able to decline this discussion when approached. But for those who do not opt out, a research nurse or midwife will explain the study and, for those who agree, will obtain their written informed consent for participation in the study. The consent form will also have additional optional check boxes to indicate the participant's approval for some specific elements of sample collection and sample utilisation, specifically (i) are obtaining maternal blood before the caesarean section, (ii) obtaining a rectal swab, (iii) using material for the creation of cell lines, and (iv) use of samples in animal studies.

The consent form will specify that the participant is passing on their samples as a "gift". The patient information leaflet and the consent form will specify that samples may be used in collaboration with external collaborators and this could include commercial entities and/or overseas entities, but there would be no capacity for financial gain for participants in any way. For example, a number of pharmaceutical companies have developed novel vaccines for GBS and they may be interested in understanding the relationship between levels of maternal antibodies and the likelihood that GBS is detected in the placenta. These companies have developed clinical grade assays that can differentiate antibodies to different serotypes of GBS. Understanding this relationship is a secondary aim of the present study. Collaboration with a pharmaceutical company may allow the generation of very high quality data without incurring the cost of performing the assays. Hence, women consenting to the study will be made aware of the potential for such collaborations.

3.3 Withdrawal

Women will be informed that they can withdraw from the study at any point and that this will not affect their care. At the time of withdrawal the participant can request that all of their data and biological samples will be destroyed. However, the patient information leaflet explains that if the participant withdraws after samples have been passed on to the biobanks that it will not be possible to destroy their data and samples as they were completely anonymised before being passed on.

3.4 Patient and public involvement

The protocol for approaching participants was informed by questionnaires given to 32 women attending for planned caesarean section (97% response rate). Additionally, we conducted face to face interviews with five women attending for planned caesarean section to determine their attitudes to participation and to the aims of the study. These surveys demonstrated that the population attending for planned caesarean section were aware of the issue of GBS in pregnancy and they were very positive about the goals of the research. Overall, >90% of the women surveyed stated that they would have likely consented for the study had it been active at the time of their own procedure. Women were given a free text box in the questionnaire and there were two responses, both positive: "sounds like a brilliant study" and "excellent study, important topic". The consent process was informed by two specific themes that arose from these data. First, women wanted the opportunity to discuss it with their partner prior to consenting, hence we aim to send the patient information leaflet to their home, in advance of

their caesarean section date. Second, one woman was unsure about having a rectal swab hence we have included the capacity to opt out of this element of the determination of colonisation status.

4. Outcomes

4.1 Maternal data

We will collect relevant maternal data from the hospital's electronic medical record (EPIC). This could include positive microbial culture results obtained during the pregnancy, results of other relevant laboratory tests (such as serological tests or markers of inflammation) and relevant data on maternal demographics (such as age, body mass index and smoking status), obstetric history (such as the indication for the caesarean section and any previous pregnancies affected by GBS) and prescriptions (e.g. recent antibiotic use).

4.2 Neonatal data

The primary indication for admission to the neonatal unit (NNU) will be classified using the hospital's Electronic Medical Record (EPIC) based on the strength of supporting evidence for sepsis, defined on the basis of clinical, imaging and laboratory assessments. The most common reason for admission following pre-labour CS at term is likely to be for respiratory complications (e.g. transient tachypnoea of the newborn) and previous studies have reported associations between FIRS and perinatal respiratory function[7]. An analysis of morbidity following term planned repeat caesarean section demonstrated a 7% rate of neonatal unit admission, where 52% had no sepsis, 46% had suspected sepsis and 2% had confirmed sepsis[10].

5. Collection of biological samples

5.1 Maternal blood

We will ask the anaesthetist to obtain a sample of maternal blood at the time that they establish intravenous access. Where it is not possible for the anaesthetist to do so, a nurse, midwife or doctor may obtain a sample by phlebotomy. Participants will be informed that they can decline to have phlebotomy without having to withdraw from the study.

5.2 Maternal swabs

Women have an indwelling urinary catheter sited in theatre prior to sterilisation of the surgical field and application of sterile surgical drapes. The theatre nurse or midwife performing the catheterisation will take a swab from the lower vagina and a swab from the rectum to establish the mother's GBS colonisation status. Placing the swab in the rectum will be optional, as indicated by yes/no check boxes on the consent form and participants will be informed that they can decline to have the rectal swab without having to withdraw from the study.

5.3 Amniotic fluid

Amniotic fluid will be collected using a syringe after the surgeon has made a hole in the fetal membranes.. If it is impossible for the surgeon to collect a sample of amniotic fluid by syringe for this (or any other) reason, the fetus will be delivered as routine and the surgeon may be asked to gather amniotic fluid from any pool that forms on the sterile surgical drapes. It is probable that in a proportion of cases it may not be possible to collect a sample of amniotic fluid.

5.4 Cord blood and placenta

After delivery of the placenta, usual practice is that the surgeon will allow a time for transfer of blood from the placenta to the baby (this is called "delayed cord clamping"). After this is complete, the cord is clamped and cut, the neonate is passed to the neonatal team and the placenta and membranes are delivered and passed to the scrub nurse. The placenta with attached and clamped umbilical cord and membranes will be passed directly to a member of the research team for subsequent processing. Cord blood may be obtained from the placenta. Approximately 5-10ml of blood will be retained for the DIGS study and the remainder of the collection will be obtained for the Cambridge Blood and Stem Cell Biobank. We will obtain multiple biopsies of the maternal surface of the placenta (the basal plate) plus samples of the fetal membranes and the umbilical cord tissue, as previously described[11]. Samples of the maternal surface of the placenta, fetal membranes, amniotic fluid and fetal blood may be sent for culture for GBS. Culture of tissue samples may be preceded by homogenisation using glass beads. Samples from the same sites may also be flash frozen for molecular studies (e.g. detection of GBS DNA or measurement of cytokine levels) and samples of the placenta, fetal membranes and umbilical cord may also be obtained for microscopy. In addition, placental samples may be obtained for the CTR Tissue Bank. The size of samples will depend on the request from the Tissue Bank and could vary, depending on the project, from very small (pea sized samples) which are flash frozen for molecular analysis to providing them with the whole of the remaining placenta, which could be used for the isolation and culture of primary cells (trophoblast).

6. Analysis of samples

6.1 Culture for GBS

Samples will be cultured to detect the presence of GBS. Current practice is that samples are incubated in Todd-Hewitt broth supplemented with antibiotics to select for GBS. After culture, identification of GBS is confirmed using matrix-assisted laser desorption/ionisation time-of flight (MALDI-TOF) mass spectrometry. Antibiotic disc susceptibility testing is performed according to EUCAST guidelines. GBS isolates may be stored at -80°C until required for further analysis. Microbial cultures will be performed in Prof Peacock's lab in the Department of Medicine at Cambridge University. This research lab provides clinical grade facilities with appropriate microbiological safety systems in place (CL2, as defined by the COSHH regulations 2002). Within this facility, microbes can be cultured from human samples, pathogens identified, their antibiotic sensitivities determined, and isolates stored for future study.

6.2 Identification of GBS DNA

We will determine the presence of GBS DNA in the placenta, fetal membranes, fetal blood and amniotic fluid using molecular methods. For example, we have developed a nested PCR quantitative PCR (PCR-qPCR) assay for the GBS 16S rRNA gene. Although we have confirmed the specificity of the assay by spiking samples with genomic DNA for *S pyogenes* and *S pneumoniae* (up to 100-times more than the GBS DNA present), we may confirm the identity of any amplified product by DNA sequencing and we may perform a PCR-qPCR assay for other GBS specific genes (e.g. *sip*) in parallel to mitigate the risk of false positive results with the 16S rRNA PCR-qPCR assay. The assays can run in multiplex format (16S rRNA, *sip* and a human DNA control). We may also identify GBS DNA using sequencing based approaches.

6.3 Analysis of vaginal GBS isolates

Culture will be the primary method for detecting GBS in the maternal swabs. We may also perform sequencing of GBS from isolates grown from maternal swabs to identify genes or variants associated with presence of GBS in the placenta. First, we may test associations between placental invasion and the presence of previously described virulence factors[12]. For example, isolates from invasive disease in infants have demonstrated higher proportions with the hyper-virulent adhesin, *hgvA.* Secondly, we may perform discovery-based, hypothesis-generating analyses of the WGS data to identify any novel associations using bacterial-adapted GWAS approaches such as PySEER[13].

6.4 Quantifying markers of inflammation

We may compare cytokine levels in relation to the presence or absence of GBS DNA in the intra-uterine tissues. In order to select the optimal cytokines, we may first analyse fetal blood using a proteomics platform to screen a large number of possible candidates to measure, for example using the Olink proximity extension assay[14]. This provides relative quantification of a large number of proteins. We may use this to identify further markers of inflammation in addition to the pro-inflammatory cytokines identified in our preliminary analysis. We may then assay relevant cytokines using ELISA, e.g. the R&D Systems Ella Automated Immunoassay platform in cases which are positive for GBS DNA in a given tissue and compare levels with those in GBS negative controls. We may also compare inflammation using microscopy, for example, by histological examination of sections or using immunohistochemistry to determine infiltration by immune cells.

6.5 Analysis of maternal serology

We may quantify the levels of antibodies to GBS in the mother's and baby's blood, for example using ELISA based methods. We may use methods which assess total antibody levels and methods which quantify antibodies to the GBS capsular polysaccharides for specific serotypes of GBS.

6.6 Further analysis of samples

The samples collected are a valuable resource and will represent three years of intense research activity. While the primary aim of the sample collection funded by the MRC is to address the hypotheses outlined above, the successful grant application to the MRC also stated that the samples would be stored and could be used to address other research questions. This could be other projects led by the current investigators or it could be through the supply of samples to collaborators. It is not possible, therefore, to list every possible method that might be used to study the samples and this will be explained in the patient information leaflet and reiterated in the consent process.

In the event that we supplied external research groups with data and/or biological samples, all transfers will be under a formal contract, specifically, a Data Transfer Agreement or a Materials Transfer Agreement. These are written and approved by the University of Cambridge Clinical School Research Operations Office. They require the signature of the recipient researcher and a commitment from the researcher's institution to be bound by the terms of the agreement. Samples would be identified only by a unique study number and no identifying information would be sent to any external collaborator. Moreover, the agreements will require the receiving institution and researcher not to undertake any attempt to identify individuals based on the data or samples sent. These collaborations could include commercial entities. Study participants will consent to giving their samples as a "gift" to be used in the ways described above.

We anticipate that the data and samples from the study may continue to be used for some years after the MRC funded element has completed. For this reason the study end date is 31st August 2032. At this point if further analysis of the data and samples is intended, a new IRAS application may be submitted or the samples may be transferred to a biobank, with appropriate HTA approvals. However, if no alternative arrangements are made, the data and samples would be securely destroyed.

Additionally, the staff collecting samples as part of the DIGS study will also collect anonymous samples for two Cambridge based biobanks, the Haematology Department Cord Blood Biobank and the CTR Placental Biobank, and these are described below. The samples provided to these biobanks will be completely anonymous, i.e. there will be no identifier which allows a connection between the individual donating the sample and the tissues stored. Tissues will be supplied with relevant demographic and clinical data, such as age, smoking status, birth weight, gestational age and experience of pregnancy complications. Any exclusion criteria for either biobank will be assessed and samples will only be passed on to the biobanks which lack a given biobank's exclusion criteria.

6.7 Use of samples by the Cambridge Blood and Stem Cell Biobank

The Cambridge Blood and Stem Cell Biobank has current active ethics approval (18/EE/0199) and is registered in the UK Tissue Directory (GBR-1-9). The resource aims to facilitate research in the fields of normal development and ageing, inherited blood disorders, haematological malignancies, and autoimmune disorders. This may include but is not limited to investigating the

cellular and molecular mechanisms of normal development and ageing, modelling of haematological disorders and discovery and testing of diagnostic/prognostic markers and novel therapies. Samples obtained in this study will primarily be used in research into ageing, normal blood and stem cell development/function and modelling haematological malignancies. This may require genetic analysis (DNA, RNA) and generation of cell lines. Samples may be genetically altered to model haematological disease and transferred into an animal recipient to study the disease in vivo; or may be transplanted directly to study the cellular and molecular dynamics of the resulting repopulated bone marrow/blood system. These experiments are important to understand the role of specific genetic changes in malignancy or the clonal composition of bone marrow and contribution to the clonal diversity of the blood system with age. Potential participants with current diagnosis of systemic infection, cancer, a haematological or immune system disorder, or undergoing immune suppression will be excluded since these may result in stressed, atypical haematological and/or immunological systems.

6.8 Use of samples by the CTR Placental Biobank

The CTR Placental Biobank has current active ethics approval (22/EE/0133) and is registered in the UK Tissue Directory (GBR-1-162). The resource exists to support translational research in normal and abnormal function of the placenta. Samples collected for the CTR biobank will principally be placental tissue, but could also include the fetal membranes, umbilical cord and cord blood. The nature of the collection of samples to be made available through a biobank precludes the possibility that all possible uses of the samples can be described. However, for illustration, some of the experimental techniques would include the analysis of flash frozen samples of placenta or digestion of fresh tissues and isolation and subsequent culture of cells. Subsequent analysis could include fluorescent sorting of isolated cells into unique populations for gene analysis or functional studies. Analysis of proteins and nucleic acids could also be performed, including sequencing of total RNA, DNA methylation and micro-RNAs. Analysis of proteins could be measurement of specific proteins using targeted techniques, or it could be analysis of multiple different proteins from the same sample using a method such as proteomics. Other “omic” methods could also be employed. Such analyses could be performed on trophoblast and related cell types as well as their differentiated derivatives to understand the gene networks that regulate proliferation and differentiation into the various cellular sub-types. Analyses may also be performed to identify molecules secreted by cell types cultured from these tissues. Studies could include co-culture of isolated trophoblast with other cell types (fibroblast, immune cells) to investigate co-stimulation, and how functions such as trophoblast invasion and maternal arterial remodelling are regulated. Cytotoxicity assays of immune cells and target cells may be performed, as well as infection assays to analyse pathogen–cell interactions. Microscopic analysis of tissues sections and isolated cells may also be performed to further characterize cell types and their differentiated derivatives.

7. Sample size, power calculations and statistical methods

We aim to collect samples from women attending the Rosie Hospital for planned caesarean delivery over a three year period (this allows a further one year to complete analysis of the last

collected samples). If we assume that two thirds of eligible women consent, we would obtain samples from a total of 1,800 women. Although our survey indicated that >90% of women would be prepared to take part in the study, we have used the more conservative estimate for this calculation. Moreover, we have re-estimated our power calculations with an even more conservative estimate of 50% recruitment. We have used the existing literature and our preliminary data to estimate some key proportions and relative risks, and these are employed in the table below. We have then calculated the statistical power of the analysis to address the hypothesis tests described above.

Power calculations[§] for main hypothesis tests (alpha=0.05 [two-sided])

Hypothesis	Exposure	Outcome	n1	n2	% in 1	RR	Power A	Power B
1	GBS +ve HVS	GBS +ve placenta	1476	324	5%	3	>99%	>99%
2	GBS +ve placenta	NNU admission	1710	90	7%	2.5	89%	81%
2	GBS +ve placenta	Cytokine storm	270	90	1%	10	93%	90%
3	hvgA positive HVS	GBS +ve placenta	292*	32	5%	5	90%	83%

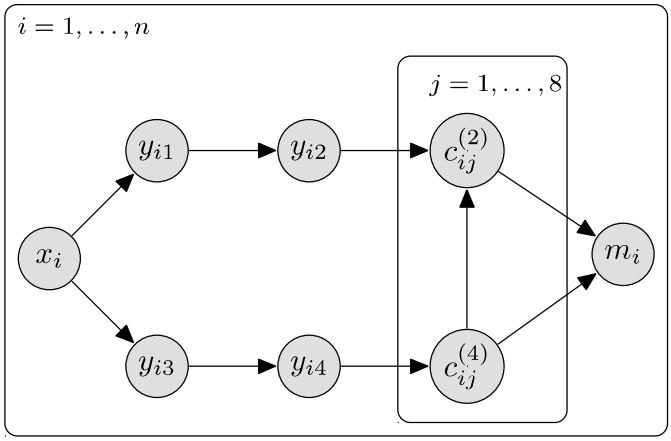
*[§] "power twoproportions" in Stata v15 *HVS cultured a GBS strain lacking hvgA*
Power A assumes 67% recruitment, n=1,800; Power B assumes 50% recruitment, n=1350.
n1=number negative for exposure and n2=number exposed for Power A;
HVS denotes high vaginal swab; NNU denotes neonatal unit admission; RR=predicted relative risk associated with exposure (all relative risks used to estimate power are conservative and at or below those described in our preliminary data, or previous publications). Cytokine storm is defined as two or more pro-inflammatory cytokines elevated to >10 multiples of the median value among controls.

On the basis of our target sample size, we will have approximately 90% power or higher to address the main hypothesis tests. However, even with the ultraconservative 50% recruitment the study would still have approximately 80% power or higher to address the main hypothesis tests.

We will use appropriate statistical methods. For example, we may compare continuous data using Student's t test, paired or unpaired as appropriate, testing for a normal distribution and using appropriate transformations where the assumption of a normal distribution is violated. Where transformation to a normal distribution is not possible, we may employ non-parametric methods (e.g. Mann-Whitney U test). We may compare proportions using the chi square test, with use of McNemar's test for paired data. Multivariate analysis may employ the general linear model for continuous data and logistic regression for dichotomous data. Again, appropriate statistical methods will employed where the assumption of independence of observations is violated.

We may also perform higher level modelling to take into account more complex associations existing within the whole dataset. An example of this is Bayesian network analysis. An interpretable *Bayesian network* (BN) model (simplified illustration) is presented in **Fig. 1**. In this

example we would obtain maximum likelihood estimates for the parameters i.e. the probabilities $P(X|\mathrm{Pa}(X))$, where $\mathrm{Pa}(X)$ denoted the parents of the variable X in the network. We would assess model robustness by considering plausible alternative topologies for the Bayesian network[15], and would adjust for potential confounders by modelling $P(M|\mathrm{Pa}(M), Z)$ using a logistic regression model, where M is a random variable indicating whether or not the baby was admitted to the NNU, and Z denotes the set of potential confounding factors, e.g. maternal age, BMI etc. To simplify modelling, we would initially assume in our model that the binarized abundances of cytokines *j* and *j'* are conditionally independent of one another (for all $j \neq j'$), and then augment our understanding of how the cytokines covary by performing separate analyses in which we model the continuous cytokine measurements using Gaussian graphical models (GGMs)[16]. For the GGM analyses, we would model cases and controls separately, to allow us to perform differential network analyses between the two groups and thereby obtain a more complete description of the fetal "cytokine storm" resulting from GBS colonisation. We note that due to the anticipated dependencies illustrated in **Fig. 1**, hypothesis tests (e.g. for association between GBS +ve HVS and each of the intrauterine tissues) are not independent, hence we can gain power relative to the values calculated above by considering hypotheses such as association between GBS +ve HVS and having any GBS +ve intrauterine tissues. All analyses would be performed in appropriate statistical software packages, such as the R statistical programming language. Scripts required to reproduce analyses may be made available via GitHub.



Notation:
For the *i*th pregnancy:

- x_i: GBS +ve HVS
- y_{i1}: GBS +ve placenta
- y_{i2}: GBS +ve bacteraemia
- y_{i3}: GBS +ve fetal membranes
- y_{i4}: GBS +ve amniotic fluid
- $c_{ij}^{(k)}$: *j*th cytokine extremely elevated in *k*th tissue
- m_i: NN morbidity

where $x_i, y_{i1}, y_{i2}, y_{i3}, y_{i4}, c_{ij}, m_i \in \{0,1\}\ \forall\ i$ are the observed variables.

Fig 1: Directed acyclic graph (DAG) representation of the assumed conditional dependencies between the observed variables (unobserved latent states and potential confounders are omitted for simplicity). The cytokine measurements will be binarized by setting $c_{ij} = 1$ if the measured abundance of the *j*th cytokine in the UC blood of the *i*th individual is >10 multiples of the median of control values.

8. Confidentiality, data handling and data storage

Personal data will be stored on secure servers compliant with NHS information governance requirements. Hard copies of data including maternal identifiers will be kept in locked cabinets

or drawers in a security controlled area. Samples will be stored in locked freezers, held in locked labs which are themselves embedded in a security controlled area. For the purpose of the laboratory work in the DIGS study, samples will be identified with a unique identifier and all laboratory data will be identified only by this means. Hence, researchers working with the bio-samples will be unaware of the mother's identifiers. These pseudo-anonymised data will be stored on password protected computers. If the samples are stored beyond the duration of active ethics approval, the storage will be licensed by the Human Tissue Authority (HTA). All samples passed on to Cambridge Blood and Stem Cell Biobank and CTR Placenta Biobank will be wholly anonymised. i.e. it will be impossible for the recipients to identify the mother or child. The samples in both biobanks will be stored under an HTA license and in accordance with the HTA standards.

9. Study insurance, management, regulation and governance

Cambridge University Hospitals NHS Foundation Trust, as a member of the NHS Clinical Negligence Scheme for Trusts, will accept full financial liability for harm caused to participants in the study caused through the negligence of its employees and honorary contract holders. There are no specific arrangements for compensation should a participant be harmed through participation in the study, but no-one has acted negligently. The University of Cambridge will arrange insurance for negligent harm caused as a result of protocol design and for non-negligent harm arising through participation in the study.

The study will be approved by IRAS and the CUHFT R&D department and will be registered with the ISRCTN. The study will be overseen by a Study Management Group, consisting of the PI and Co-PIs which will meet regularly over the course of the study. Sensitive personal information will be housed in an ISO 27001 compliant system. We currently use the Secure Data Hosting Service (SDHS) provided by the University of Cambridge Clinical School Computing Service. SDHS operational policy can be seen here:
https://www.medschl.cam.ac.uk/research/information-governance/sdhs-security-policy/
Cambridge University is making further provision for the large-scale storage of secure data and it is possible that data from the project may be stored on a different server, but fulfilling equivalent security criteria.

Anonymised extracts of the clinical data may be made available to bona fide researchers and sharing requires execution of a Data Transfer Agreement. The recipient's institution makes a number of commitments, including an agreement to prohibit any attempt to identify study participants using the released data or otherwise breach confidentiality, or make unapproved contact with study participants. This agreement will be consistent with the responsibilities laid out in the "MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies". The DTA requires signatures from both the receiving researcher and their host institution.

The Council of the School of Clinical Medicine defines policies in respect of information governance, taking into account legal, University and NHS requirements. The Council is also responsible for ensuring that sufficient resources are provided to support the requirements of the Policy. The Secretary of the School of Clinical Medicine has responsibility for all Information Governance protocols, for communication of such policies within the School and for ensuring that they are managed responsibly. The Research Governance Office is responsible for overseeing day-to-day information governance issues, including developing and maintaining policies, standards, procedures and guidance, co-ordinating information governance in the School and raising awareness of information governance. Investigators and departmental data managers are responsible for ensuring that the Policy and its supporting standards and guidelines are built into local processes, and provide evidence of compliance when requested by either the Research Governance Officer or their authorised representative, as part of any audit. All staff, whether permanent, visiting, temporary or contracted, and students, are responsible for ensuring that they are aware of the requirements incumbent upon them and for ensuring that they comply with these on a day to day basis.

See:

Data Management Policy & Procedures	https://researchgovernance.medschl.cam.ac.uk/information-governance-storage-of-research-participant-data/
Data Security Policy	https://researchgovernance.medschl.cam.ac.uk/information-governance-storage-of-research-participant-data/secure-data-hosting-policy/
Data Sharing Policy	https://researchgovernance.medschl.cam.ac.uk/information-governance-storage-of-research-participant-data/data-transfer-agreements/
Institutional Information Policy	https://researchgovernance.medschl.cam.ac.uk/information-governance-storage-of-research-participant-data/information-governance-policy/

10. References

1. ACOG. Prevention of Group B Streptococcal Early-Onset Disease in Newborns: ACOG Committee Opinion, Number 797. *Obstet Gynecol* 2020;135(2):e51-e72. doi: 10.1097/AOG.0000000000003668 [published Online First: 2020/01/25]
2. Russell NJ, Seale AC, O'Driscoll M, et al. Maternal Colonization With Group B Streptococcus and Serotype Distribution Worldwide: Systematic Review and Meta-analyses. *Clin Infect Dis* 2017;65(suppl_2):S100-S11. doi: 10.1093/cid/cix658 [published Online First: 2017/11/09]
3. Ohlsson A, Shah VS. Intrapartum antibiotics for known maternal Group B streptococcal colonization. *Cochrane Database Syst Rev* 2014(6):CD007467. doi: 10.1002/14651858.CD007467.pub4 [published Online First: 2014/06/11]
4. McGee L, Chochua S, Li Z, et al. Multistate, population-based distributions of candidate vaccine targets, clonal complexes, and resistance features of invasive Group B Streptococci within the US: 2015-2017. *Clin Infect Dis* 2020 doi: 10.1093/cid/ciaa151 [published Online First: 2020/02/16]
5. Vornhagen J, Adams Waldorf KM, Rajagopal L. Perinatal Group B Streptococcal Infections: Virulence Factors, Immunity, and Prevention Strategies. *Trends Microbiol* 2017;25(11):919-31. doi: 10.1016/j.tim.2017.05.013 [published Online First: 2017/06/22]
6. Landwehr-Kenzel S, Henneke P. Interaction of Streptococcus agalactiae and Cellular Innate Immunity in Colonization and Disease. *Front Immunol* 2014;5:519. doi: 10.3389/fimmu.2014.00519 [published Online First: 2014/11/18]
7. Jung E, Romero R, Yeo L, et al. The fetal inflammatory response syndrome: the origins of a concept, pathophysiology, diagnosis, and obstetrical implications. *Semin Fetal Neonatal Med* 2020;25(4):101146. doi: 10.1016/j.siny.2020.101146 [published Online First: 2020/11/10]
8. Mancuso G, Midiri A, Beninati C, et al. Dual role of TLR2 and myeloid differentiation factor 88 in a mouse model of invasive group B streptococcal disease. *J Immunol* 2004;172(10):6324-9. doi: 10.4049/jimmunol.172.10.6324 [published Online First: 2004/05/07]
9. de Goffau MC, Lager S, Sovio U, et al. Human placenta has no microbiome but can contain potential pathogens. *Nature* 2019;572(7769):329-34. doi: 10.1038/s41586-019-1451-5 [published Online First: 2019/08/02]
10. Tita AT, Landon MB, Spong CY, et al. Timing of elective repeat cesarean delivery at term and neonatal outcomes. *N Engl J Med* 2009;360(2):111-20.
11. Pasupathy D, Dacey A, Cook E, et al. Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth* 2008;8(1):51.
12. Shabayek S, Spellerberg B. Group B Streptococcal Colonization, Molecular Characteristics, and Epidemiology. *Front Microbiol* 2018;9:437. doi: 10.3389/fmicb.2018.00437 [published Online First: 2018/03/30]
13. Lees JA, Vehkala M, Valimaki N, et al. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat Commun* 2016;7:12797. doi: 10.1038/ncomms12797 [published Online First: 2016/09/17]

14. Colicino E, Cowell W, Bozack A, et al. Association between prenatal immune phenotyping and cord blood leukocyte telomere length in the PRISM pregnancy cohort. *Environ Res* 2020;191:110113. doi: 10.1016/j.envres.2020.110113 [published Online First: 2020/08/26]
15. Babtie AC, Kirk P, Stumpf MP. Topological sensitivity analysis for systems biology. *Proc Natl Acad Sci U S A* 2014;111(52):18507-12. doi: 10.1073/pnas.1414026112 [published Online First: 2014/12/17]
16. Opgen-Rhein R, Strimmer K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst Biol* 2007;1:37. doi: 10.1186/1752-0509-1-37 [published Online First: 2007/08/09]

.