BMJ Open

nSeP: immune and metabolic biomarkers for early detection of neonatal sepsis—protocol for a prospective multicohort study

Mallinath Chakraborty [1,2] Patrícia R S Rodrigues,[2] W John Watkins,[3] Angela Hayward,[1] Alok Sharma,[1] Rachel Hayward,[1] Elisa Smit,[1] Rebekka Jones,[1] Nitin Goel,[1] Amar Asokkumar,[1] Jennifer Calvert,[1] David Odd,[1] Ian Morris,[1] Cora Doherty,[1] Sian Elliott,[1] Angela Strang,[2] Robert Andrews,[2] Summia Zaher,[4] Simran Sharma [5,6] Sarah Bell,[7] Siva Oruganti,[1] Claire Smith,[8] Judith Orme,[8] Sarah Edkins,[2] Marie Craigon,[9] Daniel White,[2] Widad Dantoft [2] Luke C Davies,[2] Linda Moet,[2] James E McLaren,[2] Samantha Clarkstone,[2] Gareth L Watson,[2] Kerenza Hood,[2] Sailesh Kotecha [10] B. Paul Morgan,[2] Valerie B O'Donnell,[2] Peter Ghazal [11]

For numbered affiliations see end of article.

Correspondence to
Dr Mallinath Chakraborty; chakrabortym@cardiff.ac.uk

ABSTRACT

Introduction Diagnosing neonatal sepsis is heavily dependent on clinical phenotyping as culture-positive body fluid has poor sensitivity, and existing blood biomarkers have poor specificity.

A combination of machine learning, statistical and deep pathway biology analyses led to the identification of a tripartite panel of biologically connected immune and metabolic markers that showed greater than 99% accuracy for detecting bacterial infection with 100% sensitivity. The cohort study described here is designed as a large-scale clinical validation of this previous work.

Methods and analysis This multicentre observational study will prospectively recruit a total of 1445 newborn infants (all gestations)—1084 with suspected early—or late-onset sepsis, and 361 controls—over 4 years. A small volume of whole blood will be collected from infants with suspected sepsis at the time of presentation. This sample will be used for integrated transcriptomic, lipidomic and targeted proteomics profiling. In addition, a subset of samples will be subjected to cellular phenotype and proteomic analyses. A second sample from the same patient will be collected at 24 hours, with an opportunistic sampling for stool culture. For control infants, only one set of blood and stool sample will be collected to coincide with clinical blood sampling. Along with detailed clinical information, blood and stool samples will be analysed and the information will be used to identify and validate the efficacy of immune-metabolic networks in the diagnosis of bacterial neonatal sepsis and to identify new host biomarkers for viral sepsis.

Ethics and dissemination The study has received research ethics committee approval from the Wales Research Ethics Committee 2 (reference 19/WA/0008) and operational approval from Health and Care Research Wales. Submission of study results for publication will involve making available all anonymised primary and processed data on public repository sites.

Strengths and limitations of this study

► Early diagnosis of sepsis working towards a rapid point-of-care test using a single drop of blood.
► Identification of host–response patterns for discriminating various infections (bacterial, viral and fungal).
► Simultaneous study on maternal sepsis (ongoing) with linking of data between the mother–infant dyad.
► Involvement of multiple centres for extending the applicability of findings to other populations.

Trial registration number NCT03777670

INTRODUCTION

Neonatal sepsis: definition and epidemiology

Neonatal sepsis remains a major cause of morbidity and mortality, especially in preterm and low birth weight infants.[1] Globally, 2.6 million newborn infants are estimated to die every year (about 7000 everyday), accounting for 46% of deaths in children younger than 5 years.[2] Estimated rates of death vary from 17.6 per 1000 live births to 20 per 1000 live births.[2] While the precise cause of death is difficult to ascertain in many cases, complications after preterm birth, birth asphyxia and neonatal sepsis remain the major causes of neonatal deaths worldwide.[3] The majority of deaths occur in low-to-middle-income countries, where the impact of neonatal infection, in particular, is significant.[4 5] In the UK, around 10% of live-born infants are treated empirically with

antibiotics, as estimated by the National Institute for Clinical Excellence (NICE) in the UK, although reports from Switzerland[6] (2.2%–2.6%) and Norway[7] (2.3%) suggest lower estimates. The actual incidence of culture-proven sepsis is much lower, between 1 and 4 per 1000 live births in the USA[8] to 6.1 per 1000 live births in the UK,[1] making neonatal sepsis is a major cause of morbidity in resource-rich countries and mortality globally.

Neonatal sepsis is defined as a culture-positive infection from a normally sterile site (blood, cerebrospinal fluid (CSF), urine) in the first 28 days of life. Early-onset sepsis (EOS) occurs within the first 72 hours after birth and late-onset sepsis occurs after the first 3 days of life.

Diagnosis of neonatal sepsis: current approach and challenges

Clinicians use a combination of physiological, biochemical (C reactive protein (CRP), procalcitonin), haematological (full blood count (FBC), white cell count, neutrophils) and microbiological investigations to diagnose sepsis. However, the biochemical[9] or haematological tests fare poorly as sensitive or specific markers of sepsis.[10] Currently, the gold standard for diagnosing sepsis, including neonatal sepsis, remains a culture-positive body fluid that is normally sterile (eg, blood, CSF, urine). However, sepsis is a dynamic clinical condition and has been difficult to define even in adults.[11] Traditional laboratory culture methods are not able to grow all bacteria[12] and are not always positive even in the presence of clinical signs.[13]

Newborn infants may have reduced clinical signs and symptoms suggestive of sepsis; a problem exacerbated in EOS after exposure to potentially pathogenic organisms during delivery or established maternal infection. Consequently, clinical practice is often to have a low threshold for starting antibiotics in infants, often based on a combination of risk factors. Overall, the incidence of proven EOS remains low in clinically well infants started on treatment based on risk factor analysis,[14] while clinical signs and conventional laboratory tests also have a low positive predictive value when compared with positive blood cultures, the current gold standard.[15 16] A combination of features (clinical and laboratory) taken together improves the sensitivity and specificity but still falls short of the ideal.[17] Consequently, the lack of a definitive diagnostic test for neonatal sepsis has led to a reliance on risk factor-based models in deciding clinical treatment of newborn infants suspected of sepsis. Due to an imperfect definition and poor diagnostic criteria for neonatal sepsis, a significant proportion of newborn infants receives empirical treatment with antibiotics, leading to the related problem of antibiotic resistance, which is already a worldwide problem.[18] In preterm infants, recent studies have demonstrated the presence of a wide spectrum of antibiotic resistance genes in the gut microbiome after antibiotic use.[19 20]

Significantly, there have been attempts at antibiotic stewardship in the neonatal community. Using a Bayesian approach, the Kaiser Permanente group from California created a multivariable prediction model for EOS incorporating existing risk factors along with the infant's clinical status at birth.[21 22] Implementation of this model has significant potential for reducing the empirical use of antibiotics for EOS[23] and has been incorporated in the recent NICE guidance (NG195). However, these initiatives do not have an impact on the fundamental problem: the lack of a rapid and definitive test to conclusively diagnose neonatal sepsis. Until that goal is achieved, significant progress in managing newborn infants with suspected sepsis is unlikely.[14]

Pathway network analysis of whole blood——a system biology approach

Several studies have shown that changes in host gene expression occur in response to infection, allowing blood cells to act as biosensors for the changes.[24–26] Here, the analyses of unbiased high throughput measurements, such as the transcriptome of whole blood, describe a systems biology approach that can be further extended to integrate other multiple types of 'omic' data such as metabolomics and proteomics data as well as integrative cellular phenotyping.

Earlier studies from our group investigating optimal methods for the sampling and extraction of neonatal whole-blood transcription products demonstrated the first feasibility studies for using genome wide RNA analysis as a methodological approach for identifying host biomarkers of infection and vaccination in early life.[27 28] The sampling methods were further refined in 2015 with the development of a single drop (50 µl) methodology. Furthermore, virtual clinical trials showed the requirement for multiple markers; optimally 24 biomarkers, in discrete biological pathways underpinning causality[29 30] using blood samples taken at the first presentation of clinical infection.[26 31] A combination of machine learning, statistical and deep pathway biology analyses led to the identification of a 52-gene probe panel of biologically connected network modules. Nineteen biomarkers were found to provide a zero per cent misclassification, a number very close to the simulation prediction of 24. The modules comprise three central pathways, innate-immune or inflammatory, adaptive-immune and, unexpectedly at that time, metabolic and in combination, give an unusually high diagnostic quality. Despite patient heterogeneity and possible confounding issues with gestational age, the pathway biomarkers had greater than 99% accuracy for detecting bacterial infection with 100% sensitivity (figure 1). Furthermore, these specific combinations of biomarkers allowed the detection of neonatal sepsis in samples that had displayed blood culture negative results, illustrating the specific diagnostic benefits of the combinations of biomarkers. The high accuracy and sensitivity values could not have resulted from the investigation of any of the individual biomarkers alone. A critical part of these findings is the integration of metabolic pathways for increasing both sensitivity and specificity. To date, these
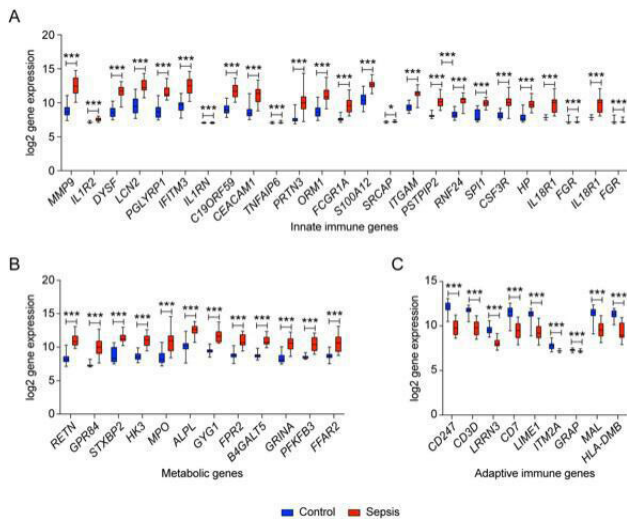
Figure 1 Box and Whisker plots of normalised log2 expression values of control non-infected neonates (N=35) and blood culture positive neonatal sepsis cases (N=26) using the tri-pathway classifier for neonatal bacterial infection, comprising probes for (A) 25 genes for the innate (inflammatory and danger signals), (B) 12 gene probes for the metabolic and (C) nine gene probes comprising adaptive immune pathways. Control, blue; bacterial sepsis, red.

studies provide a proof of concept but need independent confirmatory studies as well as investigating specificity against non-bacterial (viral and fungal) infections and sterile inflammation. The urgent unmet medical question is whether predictive host pathways can be used to first identify whether a patient is infected at or before the clinical presentation, and to further discriminate between the type of infection (in particular, bacterial, viral or fungal) and predictability of progression to sepsis. Multiplexed mass spectroscopic and multiplexed PCR platforms provide emerging platforms for translating the validated biomarkers from this study for clinically evaluating rapid and potentially point-of-care testing.

METHODS
Hypothesis and specific aims
We hypothesise that integrative analytical measurements of networked immune and metabolic pathways that selectively change on infection can accurately identify neonatal sepsis and discriminate from bacterial, viral or fungal infection. Our specific aims include:
1. To conduct prospective observational studies on validation cohorts of infants with suspected sepsis (cases) and controls and to collect blood and stool samples for multiparameter (cell, gene, protein and metabolite products) pathway analyses.
2. To apply clinical phenotyping to identify bacterial, viral, fungal and non-infective inflammatory conditions (eg, necrotising enterocolitis, hypoxic ischaemic encephalopathy, intraventricular haemorrhage) and to apply machine learning and deep systems network and pathway analyses.

Inclusion criteria
The study is open to the inclusion of neonates admitted to neonatal units in the UK and international centres. To qualify for inclusion, they must be:
1. Screened with traditional tests (FBC, inflammatory markers like CRP and blood culture) for suspected sepsis (including non-infective inflammatory conditions) and started on antibiotics (suspected sepsis).
2. Being sampled for non-septic conditions (blood sampling for routine monitoring, jaundice, hypoglycaemia, etc) (Controls).
3. Have informed consent from parents/guardians to use blood and stool samples (initial sample and 24-hour sample) and clinical data for the study.

Exclusion criteria
Although no specific clinical exclusion criteria will be used, the research team may consider excluding some infants if faced with the following situations:
1. Language and communication issues, which impairs informed consent even after using interpreters.
2. When, in the opinion of the attending clinician, an infant is not expected to survive the next 24 hours (to avoid undue distress to the family).

Group stratification criteria

The control group will comprise neonates being sampled for non-septic conditions (blood sampling for routine monitoring, jaundice, hypoglycaemia, etc).

The 'suspected sepsis' group will include neonates suspected of having an infection due to risk factors for EOS, clinical signs or other features suggestive of sepsis (NICE CG149 and NG195). These infants will undergo standard clinical investigations (including blood/CSF culture, FBC, biochemistry including CRP as appropriate) and be started on antibiotic treatment. Infants will subsequently be further classified into three groups (figure 2):
1. Confirmed sepsis: infants with a positive blood/CSF culture with a pathogenic organism agreed to be the cause of clinical symptoms.
2. Confirmed no sepsis: infants with no growth on their blood culture, no change in their differential cell counts or rise on CRP and become clinically well within 24 hours. These infants will usually have short-term antibiotics (36 hours), which will be discontinued once culture results (including blood) are available.
3. Clinical sepsis (judgement made by two senior clinicians blinded to blood culture outcome): infants who have no growth of a pathogenic organism in blood/CSF culture but fulfil at least one criterion from each of the following three domains or ≥3 criteria in total[17 32 33]:
   – Ongoing clinical signs in keeping with sepsis and unexplained by other causes, including irregular temperature (<36.5°C, >37.5°C); lethargy; poor feeding/feed intolerance/vomiting; impaired central or peripheral perfusion; haemodynamic signs
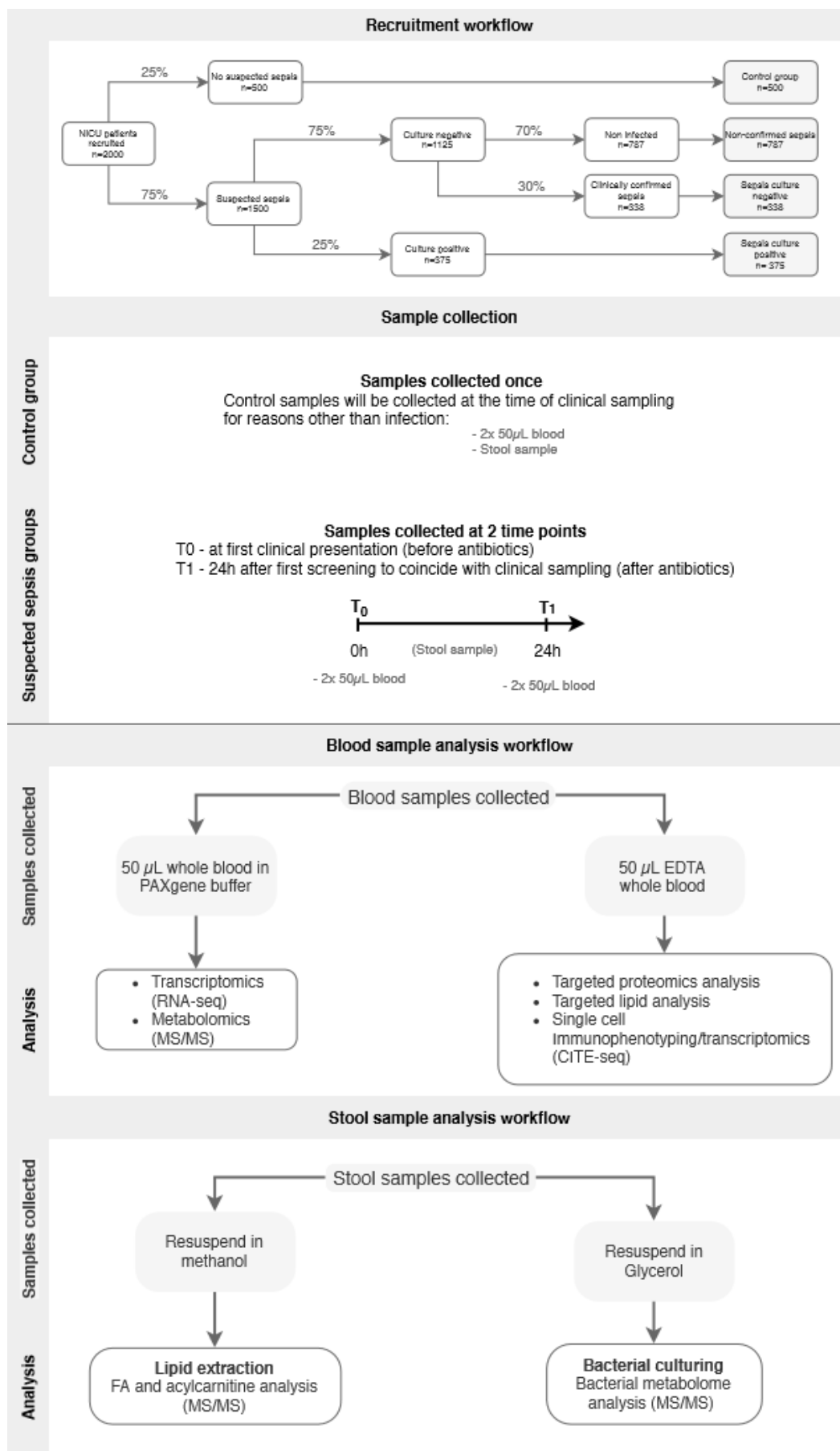
Figure 2 Details of study workflow. MS = mass spectrometry, FA = fatty acids

(tachycardia, hypotension); hypoglycaemia/hyperglycaemia; apnoea; increased ventilatory support; seizures; poor urine output.

- Change in blood results from normal range, including white cell count $<4 \times 10^9/L$ or $>20 \times 10^9/L$, platelets count $<100 \times 10^9/L$, CRP $>15\,mg/L$, hyperlactataemia $>2.0\,mmol/L$.
- Continued use of antibiotics by the clinical team for at least 5 days due to ongoing suspicion of sepsis or clinical concerns.

Medical notes and charts from these infants will be examined by two senior clinicians independently to reach a diagnosis of clinical sepsis. If there is a disagreement between senior clinicians in diagnosing clinical sepsis, then a third senior clinician will act as the adjudicator to reach a final diagnosis of sepsis.

Recruitment and consent

Infants in the neonatal unit can be suspected of having sepsis at any point during their admission. This is considered an emergency, and the intervention (screening and starting antibiotics) is delivered to infants often without prior discussion with parents/guardians.

For our study, sample collection along with other screening tests at the beginning of the infective episode and before starting antibiotics would be most appropriate. However, if parents/guardians are not present at that time to provide informed consent, this could bias the study towards less significant sepsis episodes. Thus, we will collect and store the first blood sample during the initial screening and seek deferred consent[34–37] from parents/guardians within 72 hours of the sample collection. At this time, parents/guardians will be asked to provide written consent for use of their infant's previously collected blood and stool samples and have routine health-related information of mother and infant collected from their medical records. Consent will be taken by a trained doctor, nurse, or research staff member.

For control infants, where sepsis is *not* suspected, there is usually sufficient time to discuss the study with the infant's parents/guardians. Written consent will be sought before any samples are taken; consent will also cover the collection of routine health-related information while their infant is in the NICU. Consent will be taken by an appropriately trained doctor, nurse or research staff member. Deferred consent will only be sought in

emergencies where sepsis *is* suspected as described above, and ethical approval was granted to do this (REC reference 19/WA/0008).[34–37] For infants with possible sepsis, consent will be sought from parents/guardians to collect samples from three distinct episodes in total, each separated by at least 1 week.

Wherever consent has not been granted by parents/guardians, all samples will be discarded, and the infant will not be entered into the research study.

We will follow the CONsent methods in childreN's emerGEncy medicine and urgent Care Trials (CONNECT) guidance related to deferred consenting in paediatric trials (https://www.liverpool.ac.uk/psychology-health-and-society/research/connect/).

The clinical pathway will involve the following steps (figure 3):

1. Suspected sepsis in an admitted newborn infant on the neonatal unit, leading to screening tests (FBC, CRP, blood culture), or control infants who are undergoing phlebotomy for routine monitoring, jaundice, hypoglycaemia, etc. Along with the clinical screening tests, two additional small samples of blood (50 µl or one drop of blood each for immune-metabolite and transcriptomic analysis, respectively, a total of 100 µl) and a sample of stool will be collected for the research study in a minivette or capillary tube and stored in a vial at −20ºC on the unit.
2. At the earliest opportunity, the clinical team will approach parents/guardians for permission for the research team members to discuss the study with them.
3. A research team member will give detailed written information to parents/guardians and inform them of the sample collection. At this stage, parents/guardians will be reassured that the child is not yet in the research study and will only do so if parents/guardians give consent. They will also be reassured that participation is voluntary; if they decline, this will not affect the clinical care of their child in any way, and the samples of blood collected for the research study will be discarded on the ward using the usual pathway (as per Human Tissue Act).
4. After giving parents/guardians reasonable time to consider the information, and an opportunity to ask questions, written consent will be requested from them for their child to participate in the study
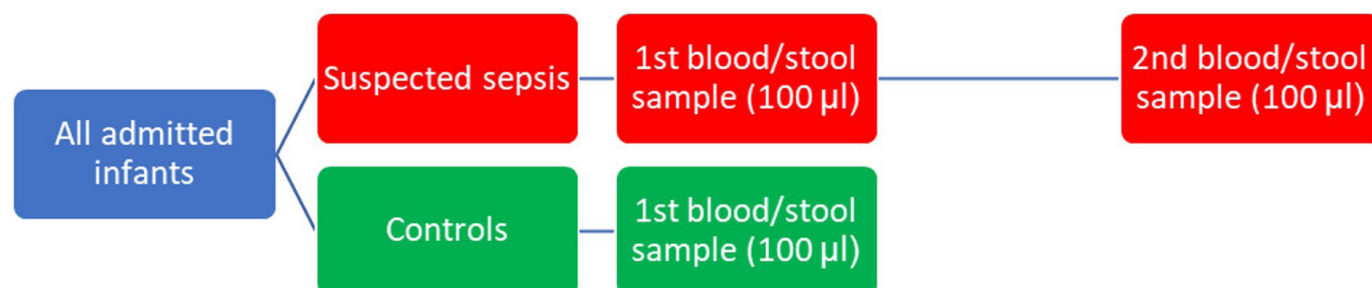
Figure 3 Study flow diagram for the collection of samples.

(research samples and clinical data). Consent will be sought specifically for the following:

– The current episode of sepsis.
– Two further episodes of sepsis.
– For long-term storage of samples in the Biobank

Parents can opt to participate in any or all the above options.

5. Once written consent is gained, demographic and clinical data from the infant will be collected for the research study and entered into an electronic database (see below). Each infant will receive a study number to allow for anonymisation, and this number will be used to label the sample before sending it for analysis. Thus, no identifiable data will be shared with the wider research team outside the neonatal unit.

6. A second identical set of research blood samples (two 50 µl samples, total of 100 µl) and stool samples will be collected after 24 hours from all cases of suspected sepsis, along with routine blood samples used clinically to monitor the infant. The sample will be labelled with an anonymised study number before transfer for analysis.

7. For control infants who are undergoing blood samples for monitoring purposes unrelated to sepsis, prior informed consent will be sought from parents/guardians for collecting two drops of blood (100 µl) and clinical information.

8. Opportunistic stool samples will be collected from cases and controls around the time of blood sampling.

9. There will be a limit to the number of samples that can be collected from each infant (figure 3):
– Suspected sepsis: 100 µl+100 µl per episode of suspected sepsis per week, up to a maximum of three episodes of suspected sepsis.
– Controls: 100 µl once only per episode of stay.

10. Detailed clinical data, along with the final diagnosis and outcome, will be collected for the study. These data are normally collected as part of the clinical care of the infant and are not specific to the research study.

Storage and sample analysis

The blood samples, at clinical sites, will be stored at −20°C, except for a subset of samples that will be stored at 4°C for certain proteomic and metabolite analyses. Stool samples will be stored at 4°C or −20ºC, within 15 mins of collection, in an upright position using the storage box provided.

Blood samples transferred to the laboratory will be processed in batches and initially retained by the nSeP team onsite at Cardiff University in dedicated lock secured −80°C. For long-term storage and access for future research, samples will be housed by the Cardiff Biobank. For the transcriptomic and lipidomic analysis, the blood is mixed with the stabilising reagent potassium amyl xanthate(PAX) in the collection tubes, all cells are immediately lysed and are not considered to be human tissue. The second tube will be stored as an EDTA-whole blood sample (with or without a stabiliser for cellular phenotyping) and after cell, and targeted proteomics and metabolomics analyses will be banked for future use, for validating future diagnostic platforms. Stool samples initially stored at 4°C are processed within 24 hours for later bacterial cultivation by suspension and homogenization into PBS supplemented with 25% glycerol and subsequent freezing in cryogenic vials at −80°C degrees. Stool samples initially stored at −20°C are intended for lipid quantification and are transferred within 24 hours to −80°C storage in the original stool sample tube. Stool samples will be processed for metabolite and microbial analysis using standard methods and are no longer considered human tissue.

We aim to apply a systems biology multiomic analysis of blood samples. This will involve using microarray and RNAseq methodology for probing the transcriptome of whole blood. While for specific immune cell characterisation, we will use techniques such as the Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-Seq) methods that perform RNAseq along with quantitative and qualitative information on surface proteins at a single-cell level. For metabolite and proteomic analyses, we will use the methods of liquid chromatography with tandem mass spectrometry (LC-MS/MS) in a targeted high-throughput manner, and we will use quantitative lipidomic profiling (LC-MS/MS) for comprehensive screening of specific pathways such as the complement system.

Clinical data

Detailed clinical data will be collected from the participant, including the following:

1. Demographic data (gestation, birth weight, age in days, mode of delivery, sex).
2. Antenatal clinical data (chorioamnionitis, antenatal corticosteroids, intrapartum antibiotics and type of delivery).
3. Reason for suspecting sepsis:
– Risk factors for EOS.
– Clinical signs.
4. Standard markers of infection at presentation (CRP, white cell count, neutrophil count).
5. Vital physiological monitoring data (continuous and discrete).
6. Final diagnosis
– Infection
– (1) confirmed bacterial sepsis (blood culture), (2) confirmed non-bacterial sepsis (viral/fungal by immunology, PCR, blood culture), (3) clinical sepsis (consensus of two senior attending neonatal clinicians).
– Non-infective inflammation.
– Not infected and not inflamed.
7. Outcome of sepsis episode defined as the duration of antibiotic treatment:
– Complete clinical resolution.
– Clinical resolution with sequalae.

– Death.

Statistical analysis plan
Sample size calculation and data analyses plan

Prior transcriptomic power calculations for sampling neonates and infants from a variety of different populations have shown a power sample calculation that ranges from N=10 to 30 for 90% power to detect twofold change with a false discovery rate-adjusted type-1-error (alpha) of 0.01 in genome wide gene expression.[26 27] Both data processing and analysis will be performed with R Language and Environment for Statistical Computing (R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing (2016). Available from http://www.R-project.org) followed by quality control and data processing using Bioconductor packages.[38] Preliminary lipidomic investigations using a new platform suggest that for 1084 cases and 361 controls (total 1445), 70% of the biomarkers would distinguish cases from controls with a power of 0.95 or better with the standard alpha of 0.05. As the platform technology for these measurements is new, further power calculations for lipid metabolite data (using independent data sets) will be reassessed during the study applying an adaptive approach, which may change the sample size calculation further. For RNAseq and microarray transcriptomic data, a per-gene hypothesis of differential average expression will be tested using a negative binomial generalised linear model—DESeq2 package and the empirical Bayes moderated t test—limma package, respectively,[39] and resulting pvalues will be adjusted for multiple testing using the Benjamini-Hochberg and Bonferroni-Holm algorithm, respectively.[40] For classification, a variety of machine learning and statistical pathway biology approaches, as described in Smith *et al*,[26] will be used. Pathway analyses will be carried out stepwise using a pathway biology approach, becoming more focused. For metabolomic and proteomic data, absolute concentrations (determined by LC-MS/MS) of analytes in extracts from blood samples will apply validated tools such as MetaboAnalyst V.4.0.[41] Further multivariate statistical testing (principal component analysis, partial least squares-discriminant analysis, random forest, analysis of variance) using group assignment derived from clinical phenotyping (types of sepsis and between sepsis and non-septic controls), will also be used to determine metabolites or proteins that are significantly different in abundance between clinical phenotypes. For single-cell analyses (CITE-seq[42] using surface protein and RNA libraries (10× Genomics) and next-generation sequencing, it will be multiplexed by cell 'hashing'.[43] Demultiplexed sequencing data will be aligned to the reference transcriptome using *CellRanger* (10× Genomics) and the number of unique molecular identifiers per cell will be quantified. Computational analyses and quality control will be performed using R packages including *Seurat*[44] to integrate hashtag, protein and RNA libraries while also enabling demultiplexing of donors, doublet detection and cell clustering.

For the final pathway biomarker assessment of the predictive success of the model, receiver operator characteristics curves will be applied.

Patient and public involvement

Patients and advisory groups were directly involved in the planning and preparatory stages of this research study, and study results will be disseminated through sepsis advisory groups.

FUNDING, ETHICS AND DISSEMINATION

DISCUSSION

The nSeP study is the first large-scale independent validation for a new blood test to diagnose neonatal sepsis based on a combination of transcriptomic and lipidomic data. We hypothesise that this test will discriminate bacterial from viral sepsis and non-infected patients. The potential strengths of the study are:

1. Early diagnosis, offering a rapid point-of-care test in time.
2. A clear distinction between infected and non-infected patients.
3. Identification of pattern of various infections (bacterial, viral and fungal).
4. Simultaneous study on maternal sepsis with linking of data between the mother–infant dyad.
5. Potentially change clinical practice by identifying or ruling out infection early.

Recruitment to the study started in March 2020. However, due to the global pandemic, recruitment to non-COVID-19-related studies was temporarily halted. Recruitment was restarted again in October 2020 but slowed down in the first 3 months of 2021 due to the redeployment of key staff. Since April 2021, we have been consistently recruiting to the study. From collaborating centres, we have now recruited 603 infants (blinded to cases and controls) so far.

The investigation of transcriptomic data will provide the necessary validation of the original 52-biomarker classifier. More importantly, the study will allow a further

reduction to a minimal set with high reproducibility, sensitivity and specificity in multiple cohorts; we anticipate from 3 to 19 biomarkers dependant on performance. The high dimensionality of the primary data allows for further computationally intensive work using machine learning and artificial intelligence to identify completely new biomarkers for increasing overall reliability and accuracy. New biomarkers would be derived from the different data modalities, comprising gene transcript and lipid species levels. The validated biomarkers would be implemented in the near term in clinical settings using a computationally efficient scoring algorithm applied to existing well-established diagnostic platforms (point-of-care or laboratory based), such as PCR and mass spectroscopy. Optimisation for the negative predictive value for the test will be a key feature of the scoring algorithm.

Author affiliations
[1]Regional Neonatal Intensive Care Unit, University Hospital of Wales, Cardiff, UK
[2]Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, UK
[3]Department of Statistics, Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, UK
[4]Department of Obstetrics and Gynaecology, University Hospital of Wales, Cardiff, UK
[5]Infection and Immunity, Cardiff University, Cardiff, UK
[6]Women's unit, Cardiff and Vale NHS Trust, Cardiff, UK
[7]Department of Anaesthetics, University Hospital of Wales, Cardiff, UK
[8]Simpsons Special Cary Baby Unit, Royal Infirmary of Edinburgh, Edinburgh, UK
[9]Infection Medicine, Deanery of Biomedical Sciences, Edinburgh Medical School, The University of Edinburgh, Edinburgh, UK
[10]Department of Child Health, Institute of Molecular & Experimental Medicine, Cardiff University School of Medicine, Cardiff, UK
[11]Department of Systems Medicine, Medical School, Cardiff University, Cardiff, UK

ORCID iDs
Mallinath Chakraborty http://orcid.org/0000-0002-1721-6532
Simran Sharma http://orcid.org/0000-0002-6647-9355
Widad Dantoft http://orcid.org/0000-0002-8546-5514
Sailesh Kotecha http://orcid.org/0000-0003-3535-7627
Peter Ghazal http://orcid.org/0000-0003-0035-2228

REFERENCES

1 Cailes B, Kortsalioudaki C, Buttery J. Epidemiology of UK neonatal infections: the neonIN infection surveillance network. *Arch Dis Child Fetal Neonatal Ed*2017.
2 Hug L, Sharrow D, You D. Levels and Trends in Child Mortality. In: Leston N, ed. New York, USA: United Nations Children's Fund, 2017: 1–40.
3 GBD 2016 Mortality Collaborators. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970-2016: a systematic analysis for the global burden of disease study 2016. *Lancet* 2017;390:1084–150.
4 Liu L, Oza S, Hogan D, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the sustainable development goals. *Lancet* 2016;388:3027–35.
5 Zaidi AKM, Darmstadt GL, Stoll BJ. Neonatal Infections: a Global Perspective. In: Wilson CB, Nizet V, Maldonado YA, eds. *Remington and Klein's infectious diseases of the fetus and newborn infant*. 8 ed. Philadelphia, PA, USA: Elsevier Saunders, 2016: 24–53.
6 Zihlmann-Ji J, Braun C, Buettcher M, et al. Reduction of duration of antibiotic therapy for suspected early-onset sepsis in Late-Preterm and term newborns after implementation of a Procalcitonin-Guided algorithm: a population-based study in central Switzerland. *Front Pediatr* 2021;9:702133.
7 Fjalstad JW, Stensvold HJ, Bergseng H, et al. Early-onset sepsis and antibiotic exposure in term infants: a nationwide population-based study in Norway. *Pediatr Infect Dis J* 2016;35:1–6.
8 Shane AL, Sánchez PJ, Stoll BJ. Neonatal sepsis. *Lancet* 2017;390:1770–80.
9 Brown JVE, Meader N, Wright K, et al. Assessment of C-reactive protein diagnostic test accuracy for late-onset infection in newborn infants: a systematic review and meta-analysis. *JAMA Pediatr* 2020;174:260–8.
10 Iroh Tam P-Y, Bendel CM. Diagnostics for neonatal sepsis: current approaches and future directions. *Pediatr Res* 2017;82:574–83.
11 Angus DC. Opening the debate on the new sepsis definition defining sepsis: a case of bounded rationality and fuzzy thinking? *Am J Respir Crit Care Med* 2016;194:14–15.
12 De SK, Shetty N, Kelsey M. How to use. blood cultures. *Arch Dis Child Educ Pract Ed* 2014;99:144–51.
13 Martin GS, Mannino DM, Eaton S, et al. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med* 2003;348:1546–54.
14 Benitz WE, Wynn JL, Polin RA. Reappraisal of guidelines for management of neonates with suspected early-onset sepsis. *J Pediatr* 2015;166:1070–4.
15 Fanaroff AA, Korones SB, Wright LL, et al. Incidence, presenting features, risk factors and significance of late onset septicemia in very low birth weight infants. The National Institute of child health and human development neonatal research network. *Pediatr Infect Dis J* 1998;17:593–8.
16 Hornik CP, Fort P, Clark RH, et al. Early and late onset sepsis in very-low-birth-weight infants from a large group of neonatal intensive care units. *Early Hum Dev* 2012;88 Suppl 2:S69–74.
17 Modi N, Doré CJ, Saraswatula A, et al. A case definition for national and international neonatal bloodstream infection surveillance. *Arch Dis Child Fetal Neonatal Ed* 2009;94:F8–12.
18 Li G, Bielicki JA, Ahmed ASMNU, Ahmed A, et al. Towards understanding global patterns of antimicrobial use and resistance in neonatal sepsis: insights from the NeoAMR network. *Arch Dis Child* 2020;105:26–31.
19 Gibson MK, Wang B, Ahmadi S, et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat Microbiol* 2016;1:16024.
20 Gasparrini AJ, Crofts TS, Gibson MK, et al. Antibiotic perturbation of the preterm infant gut microbiome and resistome. *Gut Microbes* 2016;7:443–9.

21 Escobar GJ, Puopolo KM, Wi S, *et al*. Stratification of risk of early-onset sepsis in newborns ≥ 34 weeks' gestation. *Pediatrics* 2014;133:30–6.

22 Kuzniewicz MW, Puopolo KM, Fischer A, *et al*. A quantitative, Risk-Based approach to the management of neonatal early-onset sepsis. *JAMA Pediatr* 2017;171:365–71.

23 Goel N, Shrestha S, Smith R, *et al*. Screening for early onset neonatal sepsis: NICE guidance-based practice versus projected application of the Kaiser Permanente sepsis risk calculator in the UK population. *Arch Dis Child Fetal Neonatal Ed* 2020;105:118–22.

24 Manger ID, Relman DA. How the host 'sees' pathogens: global gene expression responses to infection. *Curr Opin Immunol* 2000;12:215–8.

25 Liew C-C, Ma J, Tang H-C, *et al*. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med* 2006;147:126–32.

26 Smith CL, Dickinson P, Forster T, *et al*. Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat Commun* 2014;5:4649.

27 Smith CL, Dickinson P, Forster T, *et al*. Quantitative assessment of human whole blood RNA as a potential biomarker for infectious disease. *Analyst* 2007;132:1200–9.

28 Flanagan KL, Noho-Konteh F, Ghazal P, *et al*. Transcriptional profiling technology for studying vaccine responses: an untapped goldmine. *Methods* 2013;60:269–74.

29 Khondoker MR, Bachmann TT, Mewissen M, *et al*. Multi-factorial analysis of class prediction error: estimating optimal number of biomarkers for various classification rules. *J Bioinform Comput Biol* 2010;8:945–65.

30 Watterson S, Ghazal P. Use of logic theory in understanding regulatory pathway signaling in response to infection. *Future Microbiol* 2010;5:163–76.

31 Dickinson P, Smith CL, Forster T, *et al*. Whole blood gene expression profiling of neonates with confirmed bacterial sepsis. *Genom Data* 2015;3:41–8.

32 Dorling J, Abbott J, Berrington J, *et al*. Controlled trial of two incremental Milk-Feeding rates in preterm infants. *N Engl J Med* 2019;381:1434–43.

33 ELFIN trial investigators group. Enteral lactoferrin supplementation for very preterm infants: a randomised placebo-controlled trial. *Lancet* 2019;393:423–33.

34 Woolfall K, Young B, Frith L, *et al*. Doing challenging research studies in a patient-centred way: a qualitative study to inform a randomised controlled trial in the paediatric emergency care setting. *BMJ Open* 2014;4:e005045.

35 Woolfall K, Frith L, Gamble C, *et al*. How experience makes a difference: practitioners' views on the use of deferred consent in paediatric and neonatal emergency care trials. *BMC Med Ethics* 2013;14:45.

36 Gilbert RE, Mok Q, Dwan K. Impregnated central venous catheters for prevention of bloodstream infection in children (the catch trial): a randomised controlled trial. *The Lancet* 2016;387:1732–42.

37 Songstad NT, Roberts CT, Manley BJ, *et al*. Retrospective consent in a neonatal randomized controlled trial. *Pediatrics* 2018;141.

38 Huber W, Carey VJ, Gentleman R, *et al*. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 2015;12:115–21.

39 Ritchie ME, Phipson B, Wu D, *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.

40 Holm S. A simple sequentially Rejective multiple test procedure. *Scand J Stat* 1979;6:65–70.

41 Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr Protoc Bioinformatics* 2019;68:e86.

42 Stoeckius M, Hafemeister C, Stephenson W, *et al*. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017;14:865–8.

43 Stoeckius M, Zheng S, Houck-Loomis B, *et al*. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol* 2018;19:224.

44 Butler A, Hoffman P, Smibert P, *et al*. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36:411–20.