STUDY PROTOCOL

Whole blood genome-wide transcriptome profiling and metagenomics next-generation sequencing in young infants with suspected sepsis in low-and middle-income countries: A study protocol [version 1; peer review: 2 approved with reservations]

Constantin R. Popescu¹,², Bentry Tembo³, Rhoda Chifisi³, Miranda M.M. Cavanagh⁴, Amy Huei-Yi Lee⁵, Blessings Chiluzi³, Emily J. Ciccone⁶, Gerald Tegha⁷, Esther Alonso-Prieto⁴,⁸, Jennifer Claydon¹, Dustin Dunsmuir¹,⁹, Mike Irvine¹, Guy Dumont¹,¹⁰, J. Mark Ansermino¹,⁴,⁹, Matthew O Wiens¹,⁹,¹¹,¹², Jonathan J. Juliano¹³-¹⁵, Niranjan Kissoon¹,⁴,⁸, Tisungane Mvalo⁷,¹⁶, Norman Lufesi¹⁷, Msandeni Chiume-Kayuni³, Pascal M Lavoie¹,⁴,⁸

¹BC Children’s Hospital Research Institute, Vancouver, BC, Canada  
²Department of Pediatrics, Université Laval, Québec, QC, Canada  
³Kamuzu Central Hospital, Lilongwe, Malawi  
⁴BC Children’s & Women’s Health Centre, Vancouver, BC, Canada  
⁵Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada  
⁶Department of Medicine, University of North Carolina, Chapel Hill, NC, USA  
⁷University of North Carolina Project Malawi, Lilongwe, Malawi  
⁸Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada  
⁹Department of Anesthesiology, Pharmacology & Therapeutics, University of British Columbia, Vancouver, BC, Canada  
¹⁰Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada  
¹¹Walimu, Kampala, Uganda  
¹²Mbarara University of Science and Technology, Mbarara, Uganda  
¹³Division of Infectious Diseases, School of Medicine, University of North Carolina, Chapel Hill, NC, USA  
¹⁴Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA  
¹⁵Curriculum in Genetics and Molecular Biology, School of Medicine, University of North Carolina, Chapel Hill, NC, USA  
¹⁶Department of Pediatrics, University of North Carolina, Chapel Hill, NC, USA  
¹⁷Clinical Services Directorate, Ministry of Health, Lilongwe, Malawi

First published: 24 Sep 2020, 4:139  
https://doi.org/10.12688/gatesopenres.13172.1  
Latest published: 24 Sep 2020, 4:139  
https://doi.org/10.12688/gatesopenres.13172.1

Abstract  
Conducting collaborative and comprehensive epidemiological research on neonatal sepsis in low- and middle-income countries (LMICs) is challenging due to a lack of diagnostic tests. This
A prospective study protocol aims to obtain epidemiological data on bacterial sepsis in newborns and young infants at Kamuzu Central Hospital in Lilongwe, Malawi. The main goal is to determine if the use of whole blood transcriptome host immune response signatures can help in the identification of infants who have sepsis of bacterial causes. The protocol includes a detailed clinical assessment with vital sign measurements, strict aseptic blood culture protocol with state-of-the-art microbial analyses and RNA-sequencing and metagenomics evaluations of host responses and pathogens, respectively. We also discuss the directions of a brief analysis plan for RNA sequencing data. This study will provide robust epidemiological data for sepsis in neonates and young infants in a setting where sepsis confers an inordinate burden of disease.

**Keywords**
Newborns, young infants, sepsis, prospective study, low- and middle-income country, epidemiology, diagnosis, RNA sequencing, metagenomics, blood culture.

---

**Corresponding author:** Pascal M Lavoie (plavoie@cw.bc.ca)

**Author roles:** Popescu CR: Conceptualization, Methodology, Resources, Software, Writing – Original Draft Preparation, Writing – Review & Editing; Tembo B: Methodology, Resources, Writing – Review & Editing; Chifisi R: Resources, Writing – Review & Editing; Cavanagh MMM: Resources, Writing – Review & Editing; Lee A HY: Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Chiluzi B: Supervision, Writing – Review & Editing; Ciccone EJ: Conceptualization, Funding Acquisition, Methodology, Project Administration, Writing – Review & Editing; Tegha G: Project Administration, Resources, Writing – Review & Editing; Claydon J: Project Administration, Resources, Writing – Review & Editing; Alonso-Prieto E: Project Administration, Resources, Writing – Review & Editing; Dunsmuir D: Software, Writing – Review & Editing; Irvine M: Methodology, Writing – Review & Editing; Dumont G: Funding Acquisition, Supervision, Writing – Review & Editing; Ansermino JM: Conceptualization, Funding Acquisition, Software, Supervision, Writing – Review & Editing; Wiens MO: Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; Juliano JJ: Conceptualization, Funding Acquisition; Kissoon N: Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; Mvalo T: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; Lavoie PM: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; Lavoie PM: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; Chime-Kayuni M: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing;

**Competing interests:** The investigators in this study are academic researchers in the public sector. Investigators in Malawi receive salary compensation for their time invested carrying this study. Many investigators on this study are clinicians involved in caring for sick babies and have a deep, vested interest in generating new knowledge as well as ensuring safety and well-being of research participants.

**Grant information:** This study is funded by a Stars in Global Health grant: Improving the Early Diagnosis of Neonatal Sepsis in Malawi from Grand Challenges Canada (grant number R-ST-POC-1701-04630). The mNGS component of this study is funded by a Grand Challenges Exploration grant: Improving Diagnosis of Sepsis in Young Infants in Africa: Are We Using Antibiotics Appropriately? from the Bill & Melinda Gates Foundation to Tisungane Mvalo [INV-008947/OPP1211936].

**Copyright:** © 2020 Popescu CR et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Popescu CR, Tembo B, Chifisi R et al. Whole blood genome-wide transcriptome profiling and metagenomics next-generation sequencing in young infants with suspected sepsis in low- and middle-income countries: A study protocol [version 1; peer review: 2 approved with reservations] Gates Open Research 2020, 4:139

https://doi.org/10.12688/gatesopenres.13172.1

**First published:** 24 Sep 2020, 4:139 https://doi.org/10.12688/gatesopenres.13172.1

---

**Invited Reviewers**

<table>
<thead>
<tr>
<th>Report Number</th>
<th>Report Version</th>
<th>Report Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>version 1</td>
<td>report</td>
</tr>
<tr>
<td>2</td>
<td>version 1</td>
<td>report</td>
</tr>
</tbody>
</table>

1. Julia Johnson*, Johns Hopkins University, Baltimore, USA

2. Andrew Argent*, University of Cape Town, Cape Town, South Africa

Any reports and responses or comments on the article can be found at the end of the article.

---
Introduction

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection. Each year, over 11 million people die of sepsis worldwide, with young infants accounting for nearly a quarter of these deaths\(^1\). By far, the greatest burden of infants’ sepsis mortality occurs in low- and middle-income countries (LMIC). Concerted global efforts have resulted in a 50% reduction in sepsis-related mortality in children below 5 years of age, over three decades. However, health gains in young infants lag behind older children\(^2\). Furthermore, precise epidemiological data on the incidence and risk factors for neonatal sepsis are still lacking especially in LMIC where the vastmajority of cases occur\(^3\).

Effective public health measures and treatment guidelines rely on robust evidence and case identification using clear diagnostic criteria, both of which are lacking in young infants\(^4\). This makes the management of sepsis in this age group challenging. Early clinical signs are insconsiously non-specific and often overlap with other health conditions, making sepsis difficult to recognize based on clinical criteria alone. This is compounded by the need for prompt antiobiotic treatment to ensure survival from bacterial sepsis in infants. Currently, blood cultures are the gold standard for diagnosis of bacterial sepsis. However, the specificity and sensitivity of blood cultures are low in young infants. Also, blood culture contamination can be frequent in low resource settings due to a lack of resources and disinfection policies\(^5\). As such, positive blood culture results do not necessarily indicate a true infection because of the high possibility of sample contamination during blood collection when aseptic procedures are not strictly followed. Furthermore, differentiating a blood culture contamination from a true infection can be problematic particularly due to the common occurrence of bacteremia involving commensal skin pathogens in newborns\(^6\). In addition to limiting the epidemiological understanding of the problem, these challenges greatly complicate the management of infants with sepsis in LMIC\(^7\) and seriously limit the development of effective prevention and treatment guidelines for antibiotic use.

Ribonucleic acid (RNA) profiling using deep sequencing in whole blood detects the host immune response produced during an infection. This approach provides useful information about the etiology and, potentially, the severity of sepsis in humans. Studies in the United Kingdom\(^8\), Spain\(^9\) and the United States of America\(^10\) have shown that host transcriptome signatures can discriminate between bacterial and viral causes of sepsis, as well as other syndromes in infants under 3 months of age. To date, very few studies have used RNA-sequencing (RNA-Seq) for case identification of bacterial sepsis in neonates, and to understand the etiology of sepsis in LMIC which globally represent over 90% of the disease burden. Therefore, we hypothesize that RNA-Seq of host immune responses in whole blood will inform the “true” prevalence and epidemiology of bacterial sepsis in infants in a LMIC setting. In addition, we hypothesize that metagenomic next-generation sequencing (mNGS) technologies may be useful in this study population for detection of etiologic pathogens and genes conferring antimicrobial resistance (AMR) further augmenting our epidemiological knowledge\(^11\).

To help fill these knowledge gaps, we developed a prospective study protocol that aims to:

1) Determine the prevalence of bacterial sepsis in infants under three months evaluated for suspected sepsis at a regional hospital in Lilongwe, Malawi, Africa.

2) Establish whether blood molecular RNA signatures in this setting can more accurately identify young infants with bacterial causes among those in whom sepsis is suspected.

3) Provide proof-of-concept that mNGS can be used to detect pathogens and AMR gene detection in infants with bacterial sepsis.

Study protocol

Study design

This will be a prospective, longitudinal cohort study.

Setting

Kamuzu Central Hospital (KCH), based in Lilongwe, is the largest referral hospital for the central region of Malawi (population ~18 millions), delivering adult and pediatric clinical care for ~5 million inhabitants. At KCH, infants under 2 weeks of age are admitted to a dedicated Neonatal Unit from the maternity ward, home, or another hospital facility. Infants between 2 weeks and 3 months are admitted from home or another hospital to the Special Care Nursery in the main pediatric ward or a High-Dependency Unit if they require oxygen therapy.

Participants

Infants less than 3 months with suspected sepsis from all gender and ethnic groups who present to the Neonatal Unit, pediatric Special Care Nursery or High-Dependency Unit at KCH are eligible for inclusion if they have received antibiotics for less than 4 hours prior to enrollment. Infants less than 3 months of age in whom sepsis is sufficiently unlikely such that antibiotics will not be administered, who have not received antibiotics within 72 hours prior to enrollment, and who require blood sampling for clinical indications are included as controls. Written informed consent (see Consent Form in Extended data\(^20\)) will be obtained from parents/legal guardians in English or Chichewa (local dialect).

Study procedures

Active recruitment will begin in June 2018, after a 2-week period of study training with the Malawi team. Infants will be screened daily for eligibility at the time of initial presentation for suspected sepsis. Eligibility will be determined by a trained study nurse or clinical officer on-site, who will also obtain consent in English or Chichewa, collect the history of the presenting illness and perform a physical examination...
(with vital signs), following a pre-specified Data Collection Form (see Extended data). Consent will take place in a location as private as possible, near the bedside or in a separate room. Acknowledging the typical congestion in the hospital environment, complete privacy is not always possible. Research staff will be trained in International Conference on Harmonization Good Clinical Practices and will follow the highest possible standards of privacy and confidentiality.

A log of all infants approached for consent will be collected. Additionally, unit census data from all admissions will be reviewed at the end of the study to determine the overall number of eligible infants. For participants, a separate paper log of participant ID and name will be maintained.

Once consent has been granted, additional verbal permission will be obtained from parents to record a short video (~30 seconds) of the infant, using a High Definition iPad camera. Vital signs (heart rate, respiratory rate, oxygen saturation, temperature and blood pressure) will be recorded by a study nurse or clinical officer on admission, and daily thereafter by a dedicated vital sign assistant. Heart rate, respiratory rate and oxygen saturation will be captured using a custom application developed by the Digital Health Innovation Centre for International Child Health (DD, GD, JMA) that employs a saturation probe connected to an Android device. Blood pressure will be measured via automated monitors (General Electric Dinamap Pro 300V2 monitor) using a standard blood pressure protocol (see Extended data). Axillary body temperatures will be obtained using electronic thermometers (Welch Allyn SureTemp, model 692) by trained staff. Blood pressure monitors and electronic thermometers were provided by the study investigators at the beginning of the study.

A complete blood count with differential, blood glucose, blood culture and lumbar puncture (as determined by the medical team) will be collected at the time of initial assessment, as per the standard of care for infants with suspected sepsis in Malawi. A blood sample (0.5 mL) for RNA studies will also be collected in RNALater™ (Invitrogen), at the time of blood sampling to minimize infants’ discomfort.

An additional 2 mL of blood, a rectal swab from the infant, and a vaginal swab from the mother will be collected in 100 infants weighing more than 2.5 kg (for safety reasons), for mNGS analyses.

Prior to initiating the study, clinical staff will be trained on a specific blood culture sampling protocol (see Extended data) designed and implemented in conjunction with the clinical team at KCH to minimize blood culture contamination and provide at least 2 mL of blood for pathogen detection.

Blood will be inoculated into BACTEC PEDS Plus bottles, incubated in a BD BACTEC 9050. Cerebrospinal fluid samples will be plated to sheep blood and chocolate agar media and a thioglycollate broth tube and incubated for five days. Gram stain (Fisher Healthcare protocol Gram stain set with stabilized iodine) will be performed according to manufacturer.

Blood culture bottles flagged by the instrument as possible growth will be further analyzed by Gram stain performed on the sample. Samples will be plated to appropriate media based on organism morphology seen on the Gram stain. Identification of organisms for cultures of cerebrospinal fluid with growth will be completed using biochemical tests, bioMerieux API, and BD Crystal kits. Antimicrobial susceptibility testing will be performed by disk diffusion (BD BBL Susceptibility Disks) and/or MIC (bioMerieux E-Test) in accordance with Clinical and Laboratory Standards Institute (CLSI) M100 Performance Standards for Antimicrobial Susceptibility Testing guidelines and according to manufacturer. If no growth is detected after five days, the culture result will be finalized.

Complete blood counts with cell differential will be performed in an EDTA whole blood sample using a Beckman Coulter AcT5 Diff analyzer. Samples will be tested within 24 hours of collection. Samples that were clotted or demonstrated 3–4+ hemolysis will be rejected.

The processing of blood counts, blood culture and cerebrospinal fluids (as indicated), as well as the storage (~80°C) of the RNA-protected research whole blood samples, will be done on-site by the University of North Carolina (UNC)-Project Malawi laboratory. At the end of the study, research blood samples and all bacterial isolates will be shipped in a single batch, on dry ice to Vancouver, Canada.

The standard of treatment for sepsis at KCH is to use intravenous (IV) benzylpenicillin 50,000 International Units (IU) per kg of body weight twice a day for neonates younger than 7 days and 4 times a day for infants between 7 days and 3 months of age. In addition, IV gentamicin is used at 3 mg per kg of body weight once a day for low birth weight infants, or 5 mg per kg of body weight once a day in appropriately grown term infants, and 7.5 mg per kg of body weight once per day for infants between 7 days and 3 months of age. The total duration of antibiotic treatment is dependent on the clinical course. If the infant is able to tolerate feeds orally, is afebrile and otherwise clinically well, oral antibiotics are administered after 3 days of IV treatment, using either amoxicillin 125 mg every 8 hours or erythromycin 125 mg every 6 hours for an additional 5 days. In cases of atypical pneumonia, azithromycin 10 mg once per day for 3 days is used. For suspected meningitis, the dose of penicillin is increased (100,000 IU – same dosing interval). Intravenous ceftriaxone 100 mg per kg once a day is used as a second-line antibiotic treatment if there is no response to the first-line, or in cases of suspected meningitis. During hospitalization, clinical interventions, including antibiotic treatment will be provided as per the medical team and will follow the standard of care at KCH.
Ethical considerations
The study received approval on October 3, 2017 from the National Health Sciences Research Committee at the Ministry of Health in Lilongwe, Malawi (under study #17/8/1819; title: “Improving the Early Diagnosis of Neonatal Sepsis”), amended Oct 4th, 2019 to include mNGS in a subgroup of infants), and on October 18, 2017 from the UBC Children’s & Women’s Research Ethics Board (certificate #H16-02639; Vancouver Canada).

The study will provide a standardized blood culture and complete blood count with differential to all participants, as these tests are often not available clinically due to laboratory resource limitations. Cerebrospinal fluid and urine cultures will also be provided whenever the clinical team determines these tests are indicated as per the standards of clinical care in Malawi.

Data collection
Data will be prospectively collected at the time of presentation and during hospitalization until final disposition, as detailed in the Data Collection Form (see Extended data), using standardized paper and electronic forms. Variables are designed to be largely self-explanatory with no attempt made at prespecifying definitions. However, all history and physical exam data will be captured by clinical staff members trained to the study protocol during the 2-week run-in period, under the supervision of a single clinical officer (BT). Gestational maturity will be estimated visually using a Ballard assessment when the information cannot be provided from the caregiver or the chart.

De-identified data will be entered into a password-protected REDCap database using a password-protected iPad device\(^1\). Electronic REDCap data (including videos) will be uploaded weekly via a dedicated wi-fi network onto a secured database hosted at the BC Children’s Hospital Research Institute (Vancouver, Canada). No information that discloses the identity of participants will be recorded on the mobile study devices during data collection. No personal information will be published. A list of study personnel and their delegated tasks will also be maintained by the study coordinator in Vancouver.

During the study, paper-based data collection forms, including consent forms, will be stored at KCH in a secured place, under the responsibility of the site PI (MC). Access to the data will be limited to co-investigators and study members directly involved in the study via secured access to the main server in Vancouver, Canada. At the end of the study, records will be reviewed and the data verified by at least two study investigators for accuracy and completeness. All study-related documents will be kept for at least 5 years according to policies from the University of British Columbia. De-identified data will be made publicly available, following approval by the Vancouver and Malawi research ethics boards.

Partnerships
Study staff will be hired from a pool of clinical staff dedicated to the neonatal unit at KCH, through a partnership with the Pediatric and Child Health Initiative (PACHI). The metagenomic sequencing for pathogen and AMR gene detection sub-study component is conducted in partnership with UNC Project-Malawi and the UNC. MNGS will be analyzed through the Chan-Zuckerberg Initiative.

Study size
Precise a priori power calculations are difficult due to the absence of transcriptome data in similar LMIC cohorts. However, based on communications with local study investigators, we expect that ~20% of the infants in the study will have a positive blood culture. Therefore, we estimate that enrolling 300 infants will yield about 60 bacterial sepsis cases. This will provide 80% power to detect differences in expression for ~40 gene markers, considering previous studies\(^{24,25}\), using a 5% false-discovery rate method of adjustment for multiple comparisons. An additional 100 infants will be enrolled for the mNGS objective. As this is exploratory, no formal sample size calculation was performed for the mNGS sub-study.

Data analysis
Data will be coded to facilitate analysis. The cohort will initially be analyzed descriptively, listing baseline demographic and clinical variables with mean ± standard deviation, median with interquartile range, and proportions (with 95% confidence intervals) depending on the data distribution. Bacterial species and antimicrobial resistance patterns for positive blood cultures will also be reported.

Definitions. The following definitions will be used to classify sepsis cases in the study:

- **Culture-proven bacterial sepsis**: Infants with a positive blood or cerebrospinal fluid culture for a known bacterial pathogen, who present with at least one of the following clinical signs: ill-looking (based on physician assessment), not feeding well (according to parent/caregiver), severe recesions with breathing, convulsions, abdominal distension or lethargy. Infants who present the above-listed criteria and are severely ill (based on physician assessment) are classified as having severe sepsis.

- **Clinical sepsis**: Infants who meet the above-listed criteria in absence of a positive blood culture.

- **Contaminants**: Infants who have a positive blood culture, but who are feeding well (according to parent/caregiver), appear clinically well (by physician assessment), and do not show severe recesions with breathing, lethargy, convulsions or abdominal distention.

- **Non-sepsis controls**: Infants who evolve clinically well without having received antibiotics.
Differences in baseline demographics (gestational age, birth weight, age at presentation, etc.) between the aforementioned groups (sepsis, severe sepsis, clinical sepsis, contaminants, and non-sepsis) will be compared. Significant independent association between culture-proven bacterial and/or clinical sepsis (versus controls), or mortality will be determined using multivariable models adjusting for gestational age or birth weight, sex and age at presentation, plus other significant co-variables.

Matching. To identify a gene signature of bacterial sepsis, RNA-Seq will first be run on a subset of infants with culture-proven bacterial sepsis and matched controls. Matching will be performed using a semi-parametric propensity score algorithm26, by identifying the main confounders to the outcome of sepsis. Propensity scores will be estimated from a generalized linear model (GLM) and a nearest neighbor propensity matching with replacement algorithm will be performed to generate a 1:1 match between controls and sepsis cases. Sensitivity of matches will be assessed by using a variable number of potential confounders from the following: sex, gestational age, age, and birth weight.

For RNA-seq, total RNA will be extracted from whole blood using the RiboPure RNA Purification kit. Quantification and quality assessment of total RNA will be performed on an Agilent 2100 Bioanalyzer. Samples with sufficiently high RNA Integrity Number will be considered for sequencing. Poly-adenylated RNA will be captured using the NEBNext Poly (A) mRNA Magnetic Isolation Module. Strand-specific cDNA libraries will be generated from polyadenylated RNA using the KAPA Stranded RNA-Seq Library Preparation kit and sequenced on a HiSeq 2500 (Illumina; San Diego, CA). Sequence quality will be assessed using FastQC and MultiQC1.8.1. The FASTQ sequence reads will be aligned to the human genome (Ensembl GRCh38.98) using STAR v2.7 and mapped to Ensembl GRCh38 transcripts. Read-counts will be generated using htseq-count (HTSeq 0.11.2-r1). Data processing and subsequent differential gene expression will be performed using the latest versions of R and DESeq227. Genes with very low counts (with less than 10 counts in the smallest number of biological replicates within each group) and globin transcripts will be filtered out prior to analysis.

Classifiers. We will derive a set of gene classifiers from the RNA-Seq data obtained from matched culture-proven bacterial sepsis and control cases. These classifiers will be derived, first, from differentially expressed genes identified using the Wald statistics test to identify the top 100 differentially expressed genes between groups. Differentially expressed genes will be compared to published literature (Table 1) to define a final list of curated markers. Additionally, we will apply machine learning approaches to identify potential biomarkers specific to neonatal sepsis from the blood transcriptome. Performance of models from different machine learning approaches5,6,29 will be assessed to compare model accuracy, precision and recall. These classifiers will then be applied to culture-negative clinical sepsis cases. Given that sepsis outcomes are strongly linked to infants’ sex30, post-natal age31,32 and other factors such as breastfeeding we will also explore how gene signatures are influenced by these variables, and how sex-related transcript profiles may alter disease severity.

For mNGS, both DNA and RNA will be extracted using Zymo Quick DNA/RNA kits and sequenced on an Illumina iSeq platform. We will target an 8 million-reads depth for DNA and 4 million-reads depth for RNA. The data will be analyzed using IDSeq. To determine if bacterial AMR genes can be linked to maternal vaginal flora, we will sequence the DNA from the vaginal swab to determine if the same bacterial AMR genes are present in maternal flora. In an exploratory analysis, we will assess if we can identify the same strain of bacteria, using approaches similar to StrainSifter31.

Prediction models. We will test the ability of clinical variables, but also a limited set of top-discriminating gene markers to predict in-hospital mortality from bacterial sepsis. Clinical variables will include features extracted from the point-of-care vital signs photoplethysmogram and infants’ videos. Univariate analyses will first be carried out to determine their level of association with the mortality outcome. Continuous variables will be assessed for model fit using the Hosmer-Lemeshow test32. Missing data will be imputed by the method of multivariate imputation by chained equations33. Following univariate analysis candidate models will be generated using a step-wise selection procedure minimizing Akaike’s Information Criterion (AIC). This method is considered asymptotically equivalent to cross-validation and bootstrapping34,35. All models generated in this sequence having AIC values within 10% of the lowest value will be considered as reasonable candidates. The final selection of a model will be judged on model parsimony (the simpler the better), availability of the predictors (with respect to minimal resources and cost), and the attained sensitivity. We will aim for a predictive model with a ROC of >0.75–0.8, favoring sensitivity over specificity whenever required. Analyses will be conducted using SAS 9.4 (Carey, NC, USA) and R 3.1.3 (Vienna, Austria; http://www.R-project.org).

Future data availability
At the end of the study period, de-identified data will be made publicly accessible, following approvals from the University of British Columbia Children’s and Women’s Research Ethics Board, and the National Health Science Research Council of Malawi. A demonstration version of the data collection (no upload to REDCap) of the data collection Android app will be available from the Pediatric Sepsis Data CoLab website: https://dataverse.scholarsportal.info/data/verse/Pedi_SepsisCoLab. RNA-Seq will be deposited with National Center for Biotechnology Information Gene Expression Omnibus.
<table>
<thead>
<tr>
<th>GEO / EBI accession #</th>
<th>Citation</th>
<th>Type of data</th>
<th>Platform</th>
<th># infants</th>
<th>Study population</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE69686</td>
<td>15</td>
<td>Gene array</td>
<td>Affymetrix GeneChip GGH3 Transcriptome Array</td>
<td>68</td>
<td>Newborns admitted to NICU within 24h of sepsis evaluation (US)</td>
</tr>
<tr>
<td>GSE25504</td>
<td>11,12</td>
<td>Gene array</td>
<td>Affymetrix Human Genome U133 Plus 2.0 Array; Illumina Human HT-12 V3.0 expression beadchip; Affymetrix Human Genome U219 Array</td>
<td>62/30/69</td>
<td>Newborns admitted to NICU and i) evaluated for sepsis either i) confirmed or ii) suspected (cases) or iii) &quot;well&quot; (controls) (Scotland)</td>
</tr>
<tr>
<td>EMTAB4785</td>
<td>13</td>
<td>Gene array</td>
<td>Affymetrix GeneChip Human 1.0 ST GeneChip</td>
<td>17/19</td>
<td>VLBW neonates with risk factors* and clinical signs of sepsis, or matched controls* (Spain)</td>
</tr>
<tr>
<td>GSE53471</td>
<td>36</td>
<td>Gene array</td>
<td>Affymetrix Human Genome U219 Array</td>
<td>113</td>
<td>Cord blood from infants born vaginally &gt;36 weeks of gestation (Finland, Estonia, Russia)</td>
</tr>
<tr>
<td>GSE64456</td>
<td>25</td>
<td>Gene array</td>
<td>Illumina Human HT-12 V4.0 expression beadchip</td>
<td>279/19</td>
<td>Term infants &lt;60 days evaluated for i) fever (emergency department; cases) or ii) healthy control infants aged 0-6 months (US)</td>
</tr>
<tr>
<td>GSE70461</td>
<td>37</td>
<td>Gene array</td>
<td>Affymetrix Human Genome U133 Plus 2.0 Array</td>
<td>20</td>
<td>Newborns &lt;33 weeks in NICU (Norway)</td>
</tr>
<tr>
<td>GSE111404, GSE123070</td>
<td>27</td>
<td>RNA-Seq</td>
<td>Illumina HiSeq 2500</td>
<td>60</td>
<td>Healthy term (&gt;36 weeks; &gt;2.5 kg birth weight) (Gambia and Papua New Guinea)</td>
</tr>
<tr>
<td>GSE138712</td>
<td>38</td>
<td>RNA-Seq</td>
<td>Illumina HiSeq 2500</td>
<td>18</td>
<td>Preterm neonates born &lt;30 weeks of gestation admitted to the NICU investigated for sepsis in the first 42 days of life (Australia)</td>
</tr>
<tr>
<td>GSE40396</td>
<td>39</td>
<td>Gene array</td>
<td>Illumina Human HT-12 V4.0 expression beadchip</td>
<td>4/0/1</td>
<td>Children 2-36 months with i) fever without a source, ii) bacterial infection or iii) well/afebrile, having outpatient surgery (US)</td>
</tr>
<tr>
<td>GSE6269</td>
<td>40</td>
<td>Gene array</td>
<td>Affymetrix Human Genome U133A Array; Affymetrix Human Genome U133 Plus 2.0 Array; Sentrix Human-6 Expression BeadChip</td>
<td>16/0/1/3</td>
<td>Children/adults with i) E. coli, ii) S. aureus, iii) S. pneumonia or iv) influenza A infection (US)</td>
</tr>
</tbody>
</table>

*risk factors: maternal chorioamnionitis, incomplete group B streptococcus (GBS) perinatal prophylaxis, unknown GBS status, rupture of membranes >18 hours; maternal fever during labor; premature labor incompletely treated with antibiotics, neonates with indwelling devices, neonates with recent surgery (<72 hours); "matched for gestational age, birth weight, gender, ethnicity, mode of delivery, antenatal steroids, postnatal age and "clinical status".

GEO: Gene Expression Omnibus; EBI: European Bioinformatics Institute; NICU, neonatal intensive care unit; VLBW: Very low birth weight (1,500 g birth weight).
Supplemental material.pdf contains the analysis plan. The discussion considers the impact of the study purpose of this study are outlined followed by a potential data using RNA-Seq and mNGS. Specific protocols derived for the purpose of this study are outlined followed by a potential data analysis plan. The discussion considers the impact of the study as well as the strengths and limitations. Ultimately, the data generated from the study provides an opportunity to advance the knowledge of sepsis in infants, particularly in LMICs where it has the most substantial impact.

**Data availability**

**Underlying data**

No underlying data are associated with this article.

**Extended data**

Scholars Portal Dataverse: Improving the early diagnosis of neonatal sepsis in Malawi [supplemental material]. https://doi.org/10.5683/SP2/QXBYDX20.

Popescu et al. - Supplemental material.pdf contains the following extended data:

- Blood Culture Procedure
- IV Device Access Policy
- Blood Pressure Procedure
- Data Collection Form

Extended data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

**Authors’ contributions**

NK, NL, MCK and PML conceived the main study, obtained funding and provided study oversight; EJC, JJJ and TM conceptualized and obtained funding for the mNGS study component; BT, CRP, RC and MMMC helped develop the study protocol including clinical procedures and collected the data; EAP, JC administered the resources and finances of the main study, in partnership with PACHI; DD developed the customized REDCap app for vital signs data collection; GT developed the mNGS analysis and coordinated the sample collection for the main study; GD, MJA supervised the development of the data transfer platform used in this study; CRP, AL, MI, MOW, PML developed the statistical analysis plan. All authors contributed through continuous support and provided input at various levels during the entire study.

**Acknowledgments**

We are grateful to the BC Children’s Hospital Foundation for ongoing resource and financial support, the Data Management team at the BC Children’s Hospital Research Institute for support with creation of the REDCap database, and the BC Women’s Hospital for temperature and blood pressure monitor equipment donations. We also thank the University of British Columbia Centre of International Child Health and the University of North Carolina for diverse logistics support during study. CRP is currently supported by the BC Children’s Hospital Research Institute Sue Carruthers Graduate Studentship. PML received a salary award from the Michael Smith Foundation for Health Research and is currently funded by the BC Children’s Hospital Research Institute through an Investigator Grant Program Award.

---

**Discussion**

This study will provide robust epidemiological data in a high risk area for sepsis. This will help address an important global health problem that affects the lives of millions of infants around the world. In the long term, these data could help improve the triaging, diagnosis, and immediate clinical management of young infants with suspected sepsis in both a local and global context. It could potentially also inform more judicious antibiotic use. As antimicrobial resistance is a major rising global health concern, identifying truly septic patients may reduce unnecessary empirical antibiotic use in infants with non-bacterial infections. This study will also directly benefit infants at KCH, by enabling access to standard of care investigations for suspected sepsis and providing human resource support for all infants admitted to the neonatal unit, given the current staffing limitations.

**Strengths**

The prospective nature of this study is a major strength. The data collection is informed by rigorous literature reviews3,4. Studies of neonatal sepsis in LMIC have been mostly retrospective, often starting with case selection by a positive blood culture. In addition, a variety of different definitions for sepsis have been used, without congruence or consistent biological confirmation9. In our study, the use of robust micro-biological methods informed by the complementary use of whole blood RNA-Seq may help address these gaps and allow a more precise estimation of the incidence of sepsis. RNA-seq has not been reported in full-term infants with sepsis and this study will provide these data in LMIC. As the mNGS component of this study will be carried out locally as part of a UNC Project, in Lilongwe, the study will build capacity for using this technology in Malawi, for research and eventually for diagnosis.

**Limitations**

There are some limitations in the study protocol. First, the choice of a single regional hospital for recruitment may not be representative of infants assessed for suspected sepsis, for example, in a rural setting. Second, following an infant’s disposition only until discharge will limit an assessment of long-term mortality and morbidity post-discharge. Third, we anticipate a number of challenges for this study: some lead investigators are located in geographically remote time zones, which could make challenging; access to blood sampling outside normal business hours; ensuring a sufficient supply of study supplies and staff in a resource-limited hospital environment; and inconsistent wi-fi/cell network which could complicate data transfer. These challenges will be specifically considered, discussed and addressed throughout the study duration.

**Conclusion**

In conclusion, this study protocol aims to address the gap of epidemiological data on the prevalence of sepsis in infants in a LMIC and to contribute to advancing diagnostic precision using RNA-Seq and mNGS. Specific protocols derived for the purpose of this study are outlined followed by a potential data analysis plan. The discussion considers the impact of the study...
References


Page 9 of 15
Andrew Argent
Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

General Comments:
○ This is an important study, that has probably been underway for some time at the point of review. It will provide important epidemiological information on sepsis in infants (<3 months of age) in Lilongwe in Malawi.

○ The data collected will also open the door to future studies evaluating the potential tools for diagnosis of bacterial infections in small infants. Importantly it may also be possible to evaluate the possibility of excluding bacterial sepsis on the basis of tests done in sick infants. In the longer term that may have important implications for antibiotic stewardship in poorer countries with limited resources for investigation and monitoring of therapy.

○ I have not been able to find references in this document to infections such as syphilis. Will this be considered in the study, and is it likely that there would be different responses to intra-uterine infections such as syphilis and more acute bacterial infections?

Specific Comments:
○ Title: Given that this study will be entirely at a single centre in Malawi, I wonder how appropriate it is to title this “in low- and middle- income countries”. It may not be reasonable to assume that these features will be generalizable to such a wide group of countries across the world.

○ Introduction: One of the challenges of severe infection and related illness is the overlap of terms such as sepsis / severe infection / bacterial infection. These terms are used interchangeably in multiple settings, when the implications of the different nuances may be significant.

○ I would really appreciate it if the authors could: Make it clear that severe infections in neonates could be the consequences of infection with a variety of pathogens including: viruses (e.g. parechovirus infection, herpes infection); bacterial infections (including usual
pathogens, but potentially bacteria that are generally associated with normal commensals on the skin and contamination of cultures).

- Para 1: The opening statement of “Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection” is made without a reference. Given that the issue of sepsis definitions in infants and children is under review by a variety of working groups, it may be important to provide the reference (from adult sepsis groups).

- Methods - Participants: The authors will be including infants with no indication for antibiotic therapy as control subjects. The protocol states that these infants will be admitted to the study if they need blood sampling for clinical indications. It would be really useful to understand what possible clinical indications there will be for taking blood, and specifically how consent for the study will be taken from the parents of these infants.

- Study procedures: I would appreciate explanation of how active recruitment for the study can be started in June 2018, when the protocol is up for review now.

- Para 4: It is noted that differential counts will be done using coulter counts, however, is there any capacity to check very high counts using manual techniques? In patients with high red cell precursor counts, will white cell counts be corrected for this (I have noted that patients with evidence of extensive hemolysis will have those specimens rejected)?

- Data collection: Given that patients up to 3 months of age will be admitted to the study, how will gestational age be estimated in older infants (I am not concerned that Ballard scores may not be valid after a few days of age)?

- Definitions:
  - It would be interesting to consider the items included and not included in the development of the definitions. Factors such as apnea (may overlap with the lethargy, but not necessarily), abnormalities of temperature (either hyper- or hypothermia) have not been included.
  - It is clear that the group being defined as clinical sepsis could be infected with non-bacterial pathogens. However, the authors have not addressed the group who may have been given antibiotics prior to the collection of blood culture specimens (they would be admitted to the study if antibiotics have been given within 4 hours prior to being consented). How likely are patients to fall into this category, and if there are patients in this category, how will they be defined? Is there a reason for a 4-hour cut-off, and how likely are antibiotics prior to culture to adversely impact on culture positivity rates?
  - Clearly infants with factors such as hypoglycemia, dehydration with associated electrolyte and acid-base abnormalities, congenital cardiac problems (probably a very small group of infants) could fall into the clinical sepsis group, but potentially do not have sepsis (as defined by bacterial infection).
Classifiers: How will “breast-feeding” be classified? What I am addressing is whether the authors will separate total breast-feeding (not other food intake), partial breast-feeding (additional nutrients provided – including oils, porridge etc.) and non-breastfeeding? Will breast-feeding simply be attributed on the basis of the mother’s history?

Data Analysis:
- It would be useful to have more information as to how patients will be classified. As an example, if the RNA patterns are compatible with bacterial infection but the cultures are negative (and vice versa) – how will those patients be categorized?
- It is not clear how other non-bacterial pathogens (including malaria and viral pathogens) will be factored into the analysis and categorization of these patients.
- To what extent would it be possible during the analysis of all the material collected to consider whether there may be genetic factors in this population that are different to populations in the USA and Europe? To what extent is it possible to evaluate the effect of maternal exposure to pathogens prior to delivery on the infections suffered by the infants?

Conclusions: This study has substantial strengths, although a potential challenge will be the practicalities of completing the study given the constraints of the particular clinical environment.

Supplementary Material: It is interesting that bacterial infections of the urinary tract are referred to in the data collection forms, but this is not addressed in the text. Diagnosis of bacterial urinary infections requires close attention to adequacy of specimen collection, and interpretation. It may be important to bring some commentary on this into the text of the main article.

Is the rationale for, and objectives of, the study clearly described?
Yes

Is the study design appropriate for the research question?
Yes

Are sufficient details of the methods provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: I have been part of the Pediatric Surviving Sepsis working group, and I am currently part of the Pediatric Sepsis definitions working group (as supported by the SCCM).
**Reviewer Expertise:** Pediatric critical care; sepsis (particularly in children); low and middle income areas.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

A few comments:
- Control population could be described in more detail. It seems that only infants with suspected sepsis will be considered for this study, and that the subset of infants who initially present with suspected sepsis but are interpreted as unlikely to have it and not given antibiotics will serve as controls. This does potentially lead to misclassification of infants, including those who initially are fairly well appearing but become ill during admission and are ultimately treated with antibiotics and/or are found to have culture-confirmed bacterial sepsis. How would these infants be handled? An alternative approach would be to include infants who are presenting with other chief complaints, without concern for SBI or sepsis, to serve as true controls.

- Would suggest inclusion of urine samples in young infants, as urinary tract infections are a common cause of serious bacterial infection and associated sepsis in this age group. Urine samples should be obtained by catheter specimen to be of utility.

- The protocol states that rectal swabs will only be obtained in infants weighing greater than 2.5 kg for safety reasons, but do not state what specifically the safety concerns are. Rectal or peri-rectal swabs have been obtained in smaller infants as part of research studies and clinical care, and with appropriate procedures, this should not be an issue. Avoidance in very small extremely preterm infants may be indicated, but this does not seem to be the likely population for this study. If there is concern, perhaps stool samples could be obtained instead in infants below threshold of a weight or age cutoff, if available.
It may be beneficial to describe what is known about neonate/infant sepsis at the study site based on microbiology data, which would help the reader interpret whether the described microbiologic procedures (and standard antibiotic therapy) are likely to capture the most common causes of sepsis in this population. This may not be necessary for the protocol but should be included in a future manuscript.

This may exceed the scope of the study, but to adequately capture etiologic pathogens of infants with clinical sepsis without identified bacterial pathogen, the authors could consider adding at least a limited investigation for viral pathogens.

Would recommend review of this protocol by an additional reviewer with specific expertise in NGS.

**Comments on supplemental materials:**
- Would recommend providing specific time recommended for betadine/iodine to air dry (as provided for alcohol), as inadequate time to dry is a common lapse in IPC practices for blood culture collection that leads to contamination.
- Would specify acceptable sites for blood draws.
- For BP measurements, it currently states to recheck if low in 30 minutes. Depending on how low the BP is, this may be a critical finding and waiting 30 minutes could be life-threatening. Recommend rechecking within 5 minutes or less to assess validity of finding. Also, consider providing BP norms by age/weight to assist providers, unless these are readily known.
- For the birth history, it states to enter full term as 40 weeks. Full term includes 37-40 weeks. This granularity of data may not be needed, but if descriptive statistics, such as median GA, are performed, inclusion of 40 weeks for all term infants may be misleading. Consider creating a separate question of preterm (yes/no) and using a GA question as follow-up for only preterm instead.
- For type of delivery, what does "bre" stand for? Breech? Would this mean breech extraction/vaginal delivery or C-section for breech? Would expand most abbreviations to avoid any confusion.
- Describe what resuscitation at birth means - any intervention at all by medical team? Only need for respiratory intervention such as oxygen, PPV, or intubation, or need for compressions? Unclear, for example, whether need for suctioning would be considered needing resuscitation for this form's purpose.
- Number of days in hospital: consider collection of precise admission date and discharge date to avoid any errors in calculation or inadvertent incorrect interpretation of partial days.

**Is the rationale for, and objectives of, the study clearly described?**
Yes

**Is the study design appropriate for the research question?**
Yes

Are sufficient details of the methods provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Neonatology, neonatal sepsis, infection prevention and control, healthcare-associated infections, clinical research; US and LMICs

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.