










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. Popescu ^{1,2}, Bentry Tembo³, Rhoda Chifisi³, Miranda M.M. Cavanagh⁴, Amy Huei-Yi Lee⁵, Blessings Chiluzi³, Emily J. Ciccone⁶, Gerald Tegha⁷, Esther Alonso-Prieto^{4,8}, Jennifer Claydon¹, Dustin Dunsmuir^{1,9}, Mike Irvine ¹, Guy Dumont ^{1,10}, J. Mark Ansermino ^{1,4,9}, Matthew O. Wiens^{1,9,11,12}, Jonathan J. Juliano ¹³⁻¹⁵, Niranjana Kissoon^{1,4,8}, Tisungane Mvalo ^{7,16}, Norman Lufesi¹⁷, Msandeni Chiume-Kayuni³, Pascal M. Lavoie ^{1,4,8}

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

V1 First published: 24 Sep 2020, 4:139
<https://doi.org/10.12688/gatesopenres.13172.1>

Latest published: 26 Nov 2020, 4:139
<https://doi.org/10.12688/gatesopenres.13172.2>

Open Peer Review

Reviewer Status  

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 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.

Invited Reviewers		
	1	2
<hr/>		
version 2		
(revision)		
26 Nov 2020	report	report
version 1		
24 Sep 2020		
	report	report
<hr/>		
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.		

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 AHY:** Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Chiluzi B:** Supervision, Writing – Review & Editing; **Cicccone EJ:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Writing – Review & Editing; **Tegha G:** Project Administration, Resources, Writing – Review & Editing; **Alonso-Prieto E:** Project Administration, Resources, Writing – Review & Editing; **Claydon J:** 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; **Lufesi N:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Chiume-Kayuni M:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Lavoie PM:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, 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].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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>

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 deaths¹⁻³. 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 children⁴. Furthermore, precise epidemiological data on the incidence and risk factors for neonatal sepsis are still lacking especially in LMIC where the vast majority of cases occur^{1,2,5,6}.

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 infants⁷. This makes the management of sepsis in this age group challenging. Early clinical signs are inconspicuously 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 antibiotic 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 policies⁸. 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 newborns⁹. In addition to limiting the epidemiological understanding of the problem, these challenges greatly complicate the management of infants with sepsis in LMICs¹⁰ 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^{11,12}, Spain¹³ and the United States of America^{14,15} 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 knowledge¹⁶⁻¹⁹.

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*²⁰). 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 Lab at the BC Children's Hospital Research Institute and Centre for International Child Health (DD, GD, JMA) that employs a saturation probe connected to an Android device²¹. Blood pressure will be measured via automated monitors (General Electric Dinamap Pro 300V2 monitor) using a standard blood pressure protocol (see *Extended data*²⁰). 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 Malawi²². A blood sample (0.5 mL) for RNA studies will also be collected in *RNAlater*TM (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*²⁰) 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 KCH²².

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*²⁰), using standardized paper and electronic forms. Variables are designed to be largely self-explanatory with no attempt made at pre-specifying 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 device²³. 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 \pm 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 recessions 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 recessions 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 poly-adenylated 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 0.11.2-1). 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 approaches^{28,29} 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 age^{15,27} 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 bootstrapping^{34,35}. 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/dataverse/Pedi_SepsisCoLab. RNA-Seq will be deposited with National Center for Biotechnology Information Gene Expression Omnibus.

Table 1. Other published datasets that included human infants aged <3 months (whole blood).

GEO / EBI accession #	Citation	Type of data	Platform	# infants	Study population
GSE69686	15	Gene array	Affymetrix GeneChip GCh3 Transcriptome Array	68	Newborns admitted to NICU within 24h of sepsis evaluation (US)
GSE25504	11,12	Gene array	Affymetrix Human Genome U133 Plus 2.0 Array; Illumina Human HT-12 V3.0 expression beadchip; Affymetrix Human Genome U219 Array	62/30/69	Newborns admitted to NICU and i) evaluated for sepsis either i) confirmed or ii) suspected (cases) or iii) "well" (controls) (Scotland)
EMTAB4785	13	Gene array	Affymetrix GeneChip Human 1.0 ST GeneChip	17/19	VLBW neonates with risk factors* and clinical signs of sepsis, or matched controls* (Spain)
GSE53471	36	Gene array	Affymetrix Human Genome U219 Array	113	Cord blood from infants born vaginally >36 weeks of gestation (Finland, Estonia, Russia)
GSE64456	25	Gene array	Illumina Human HT-12 V4.0 expression beadchip	279/19	Term infants <60 days evaluated for i) fever (emergency department; cases) or ii) healthy control infants aged 0-6 months (US)
GSE70461	37	Gene array	Affymetrix Human Genome U133 Plus 2.0 Array	20	Newborns <33 weeks in NICU (Norway)
GSE111404 GSE123070	27	RNA-Seq	Illumina HiSeq 2500	60	Healthy term (>36 weeks; >2.5 kg birth weight) (Gambia and Papua New Guinea)
GSE138712	38	RNA-Seq	Illumina HiSeq 2500	18	Preterm neonates born <30 weeks of gestation admitted to the NICU investigated for sepsis in the first 42 days of life (Australia)
GSE40396	39	Gene array	Illumina Human HT-12 V4.0 expression beadchip	4/0/1	Children 2-36 months with i) fever without a source, ii) bacterial infection or iii) well/afebrile, having outpatient surgery (US)
GSE6269	40	Gene array	Affymetrix Human Genome U133A Array Affymetrix Human Genome U133 Plus 2.0 Array Sentrix Human-6 Expression BeadChip	16/0/1/3	Children/adults with i) E. coli, ii) S. aureus, iii) S. pneumoniae or iv) influenza A infection (US)

* 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).

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 reviews^{41,42}. 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 confirmation¹⁰. In our study, the use of robust microbiological 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 troubleshooting and real-time study monitoring more 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

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/QXBYDX>²⁰.

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.

References

- Rudd KE, Johnson SC, Agesa KM, *et al.*: **Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study.** *Lancet.* 2020; **395**(10219): 200–11.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fleischmann-Struzek C, Goldfarb DM, Schlattmann P, *et al.*: **The global burden of paediatric and neonatal sepsis: a systematic review.** *Lancet Respir Med.* 2018; **6**(3): 223–30.
[PubMed Abstract](#) | [Publisher Full Text](#)
- 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**(10063): 3027–3035.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liu L, Johnson HL, Cousens S, *et al.*: **Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000.** *Lancet.* 2012; **379**(9832): 2151–61.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kissoon N, Carapetis J: **Pediatric sepsis in the developing world.** *J Infect.* 2015; **71 Suppl 1**: S21–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tan B, Wong JJ, Sultana R, *et al.*: **Global Case-Fatality Rates in Pediatric Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis.** *JAMA Pediatr.* 2019; **173**(4): 352–62.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lavoie PM, Popescu CR, Molyneux EM, *et al.*: **Rethinking management of neonates at risk of sepsis.** *Lancet.* 2019; **394**(10195): 279–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ombelet S, Barbe B, Affolabi D, *et al.*: **Best Practices of Blood Cultures in Low- and Middle-Income Countries.** *Front Med (Lausanne).* 2019; **6**: 131.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kan B, Razzaghian HR, Lavoie PM: **An Immunological Perspective on Neonatal Sepsis.** *Trends Mol Med.* 2016; **22**(4): 290–302.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Popescu CR, Cavanagh MMM, Tembo B, *et al.*: **Neonatal sepsis in low-income countries: epidemiology, diagnosis and prevention.** *Expert Rev Anti Infect Ther.* 2020; **18**(5): 443–52.
[PubMed Abstract](#) | [Publisher Full Text](#)
- 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.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cernada M, Serna E, Bauerl C, *et al.*: **Genome-wide expression profiles in very low birth weight infants with neonatal sepsis.** *Pediatrics.* 2014; **133**(5): e1203–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wynn JL, Cvijanovich NZ, Allen GL, *et al.*: **The influence of developmental age on the early transcriptomic response of children with septic shock.** *Mol Med.* 2011; **17**(11–12): 1146–56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wynn JL, Guthrie SO, Wong HR, *et al.*: **Postnatal Age Is a Critical Determinant of the Neonatal Host Response to Sepsis.** *Mol Med.* 2015; **21**(1): 496–504.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ramesh A, Nakielny S, Hsu J, *et al.*: **Etiology of fever in Ugandan children: identification of microbial pathogens using metagenomic next-generation sequencing and IDseq, a platform for unbiased metagenomic analysis.** *bioRxiv.* 2018.
[Publisher Full Text](#)
- Forbes JD, Knox NC, Peterson CL, *et al.*: **Highlighting Clinical Metagenomics for Enhanced Diagnostic Decision-making: A Step Towards Wider Implementation.** *Comput Struct Biotechnol J.* 2018; **16**: 108–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Greninger AL, Naccache SN: **Metagenomics to Assist in the Diagnosis of Bloodstream Infection.** *J Appl Lab Med.* 2019; **3**(4): 643–53.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Grumaz S, Stevens P, Grumaz C, *et al.*: **Next-generation sequencing diagnostics of bacteremia in septic patients.** *Genome Med.* 2016; **8**(1): 73.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Popescu CR, Tembo B, Chifisi R, *et al.*: **Improving the early diagnosis of neonatal sepsis in Malawi [supplemental material].** 2020; 1–19.
<http://www.doi.org/10.5683/SP2/QXBYDX>
- Dunsmuir D: **MalawiVitals Android App for Study Data Collection.** 2020.
[Publisher Full Text](#)
- Kawaza K, Mzikamanda RR, Molyneux L: **Care of the infant and newborn in Malawi: The COIN course.** 2017.
[Reference Source](#)
- Harris PA, Taylor R, Thielke R, *et al.*: **Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support.** *J Biomed Inform.* 2009; **42**(2): 377–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Herberg JA, Kaforou M, Wright VJ, *et al.*: **Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children.** *JAMA.* 2016; **316**(8): 835–45.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mahajan P, Kuppermann N, Mejias A, *et al.*: **Association of RNA Biosignatures With Bacterial Infections in Febrile Infants Aged 60 Days or Younger.** *JAMA.* 2016; **316**(8): 846–57.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ho DE, Imai K, King G, *et al.*: **Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference.** *Political Analysis.* 2007; **15**(3): 199–236.
[Publisher Full Text](#)
- Lee AH, Shannon CP, Amenyogbe N, *et al.*: **Dynamic molecular changes during the first week of human life follow a robust developmental trajectory.** *Nat Commun.* 2019; **10**(1): 1092.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Leclercq M, Vittrant B, Martin-Magniette ML, *et al.*: **Large-Scale Automatic Feature Selection for Biomarker Discovery in High-Dimensional OMICS Data.** *Front Genet.* 2019; **10**: 452.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rohart F, Gautier B, Singh A, *et al.*: **mixOmics: An R package for 'omics feature selection and multiple data integration.** *PLoS Comput Biol.* 2017; **13**(11): e1005752.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- O'Driscoll DN, Greene CM, Molloy EJ: **Immune function? A missing link in the gender disparity in preterm neonatal outcomes.** *Expert Rev Clin Immunol.* 2017; **13**(11): 1061–71.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tamburini FB, Andermann TM, Tkachenko E, *et al.*: **Precision identification of diverse bloodstream pathogens in the gut microbiome.** *Nat Med.* 2018; **24**(12): 1809–14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hosmer DW, Lemeshow S: **Goodness of fit tests for the multiple logistic regression model.** *Commun Stat - Theory Methods.* 1980; **9**(10): 1043–69.
[Publisher Full Text](#)
- Van Buuren S, Groothuis-Oudshoorn K: **Multivariate Imputation by Chained Equations.** *J Stat Softw.* 2011; **45**(3): 1–67.
[Reference Source](#)
- Stone M: **An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion.** *J R Stat Soc Ser B-Methodological.* 1977; **39**(1): 44–7.
[Publisher Full Text](#)
- Shibata R: **Bootstrap estimate of Kullback-Leibler information for model selection.** *Stat Sin.* 1997; **7**(2): 375–94.
[Publisher Full Text](#)
- Kallionpaa H, Laajala E, Oling V, *et al.*: **Standard of hygiene and immune adaptation in newborn infants.** *Clin Immunol.* 2014; **155**(1): 136–47.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kalixstad B, Kultima HG, Andersstuen TK, *et al.*: **Gene expression profiles in preterm infants on continuous long-term oxygen therapy suggest reduced oxidative stress-dependent signaling during hypoxia.** *Mol Med Rep.* 2017; **15**(4): 1513–26.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ng S, Strunk T, Lee AH, *et al.*: **Whole blood transcriptional responses of very preterm infants during late-onset sepsis.** *PLoS One.* 2020; **15**(6): e0233841.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hu X, Yu J, Crosby SD, *et al.*: **Gene expression profiles in febrile children with defined viral and bacterial infection.** *Proc Natl Acad Sci U S A.* 2013; **110**(31): 12792–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ramilo O, Allman W, Chung W, *et al.*: **Gene expression patterns in blood leukocytes discriminate patients with acute infections.** *Blood.* 2007; **109**(5): 2066–77.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nemetchev BR, Liang LD, Kissoon N, *et al.*: **Predictor variables for post-discharge mortality modelling in infants: a protocol development project.** *Afr Health Sci.* 2018; **18**(4): 1214–25.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liang LD, Kotadia N, English L, *et al.*: **Predictors of Mortality in Neonates and Infants Hospitalized With Sepsis or Serious Infections in Developing Countries: A Systematic Review.** *Front Pediatr.* 2018; **6**: 277.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 19 October 2020

<https://doi.org/10.21956/gatesopenres.14371.r29656>

© 2020 **Argent A.** This is an open access peer review report 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.



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.

Author Response 09 Nov 2020

Pascal Lavoie, BC Children's Hospital Research Institute, Vancouver, Canada

We thank the reviewer for their insightful comments which greatly help us improve and clarify important aspects of the protocol. Below are point-by-point responses to each comment:

Reviewer #2

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?

Reply: Thank you for this comment. We will consider the exposure and possibility of congenital syphilis in analyses of the transcriptomic data. Routine testing for syphilis and HIV is done in Malawi during antenatal visits. According to the WHO Global AIDS Monitoring in 2018, 1% of women tested on antenatal visits were positive for syphilis. Therefore, we expect few subjects in our cohort to have been exposed during pregnancy.

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.

Reply: Good suggestion. We have changed the title to: "**Whole blood genome-wide transcriptome profiling and metagenomics next-generation sequencing in young infants with suspected sepsis in a low- and middle-income country: A study protocol**". The study is meant to provide the first transcriptome dataset in a LMIC, realizing of course that generalizability will need to be confirmed in future studies in other LMICs.

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.

Reply: In neonates and young infants, these are often used interchangeably due to a lack of precise operational definitions. In this article, we have chosen to use the term “sepsis” which is commonly used in the neonatal literature (for lack of better option), but at the same time we understand that this term is somewhat imprecise.

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).

Reply: Thank you for the suggestion. We have added this clarification to the introduction.

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).

Reply: We have added a reference from the adult literature as suggested.

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.

Reply: See answer to similar comment from 1st reviewer. The need for clinical indication simply refers to the justification for adding an extra blood sample. Informed consent is also obtained from the parent/caregivers 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.

Reply: Thank you for the comment. Indeed, recruitment for our study effectively started in June 2018. We considered publishing the protocol earlier but due to unforeseen delays, we have been unable to do this. In the end, we strongly believe in open access research and in having pre-specified analysis principles and bases before we undertake any analysis of the data, so this is why we insist on publishing this protocol now. This is also in line with our study sponsor's requirement to make all information about the study available publicly as early as possible, including details in the protocol which would likely not be published elsewhere in the future. Realizing also the importance of structured and comparable datasets for cohort studies, especially in LMICs, we feel very proud to finally see this protocol submitted for review.

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)?

Reply: Indeed, the UNC lab has the capacity for manual differential on complete blood counts as needed and will be able to provide this data.

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)?

Reply: We agree that the Ballard score will not be feasible outside the newborn period. This has been clarified (Page 6). In those circumstances, we rely on available medical records or parent/caregiver recall of information.

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.

Reply: We have based our definitions on the WHO list of danger signs, that are also part of the COIN manual in Malawi. Temperature and respiratory rates are recorded automatically on admission. It was very hard for us to conceive that we could accurately capture apneas due to a lack of continuous monitoring or even visual observations due to a profound lack of staff in this setting.

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?

Reply: The 4-hour cut-off was extensively discussed within our study group. First, based on experience we expected that only a small proportion of infants will have received antibiotic treatment for less than 4 hours prior to initial presentation in this setting. We also considered data suggesting that blood cultures positivity rapidly declines after antibiotic administration (Rand KH et al, Open Forum Infect Dis 2019). Finally, we discussed with our Malawian colleagues that it could be ethically challenging not to offer the study to infants who have received antibiotics for less than 4 hours considering that these infants may benefit from the information provided by blood cultures that were only available as part of this study protocol. Beyond 4 hours, we estimated that the benefit of blood cultures would be sufficiently low to ethically and scientifically justify excluding those infants. This

clarification was added in the *Participants* section (Page 4). We have added that: "If necessary, we will separately analyze infants who have received antibiotics prior to enrollment from those who have not" (Definition section; Page 7).

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).

Reply: Correct.

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?

Reply: Breastfeeding is classified as exclusive, mixed (with formula) or formula alone and will be recorded from the caregiver's history. We did not consider other forms of nutrition, as our study population will include mostly young infants in a hospital setting. Virtually 100% infants at KCH are dependent on breastfeeding as there is little formula milk available (as widely advocated by the WHO).

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?

Reply: This is a very good question that we hope will be informed by this study. One way to address this situation is by providing "upper" and "lower" estimates of "bacterial sepsis" and precision assuming the RNA is wrong versus the blood culture is wrong (i.e. false negatives). In absence of satisfying diagnostic gold standard test for bacterial sepsis we can only assess the differences/agreement between these two approaches.

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.

Reply: We will not perform viral testing in our cohort due to testing availabilities. Malaria testing will be performed if deemed indicated by the treating medical team (e.g. during high seasons), but as per our understanding that although the neonatal mortality from malaria remains high the population incidence is relatively low in this age group (Deribew A et al. *Malar J.* 2017).

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?

Reply: Thank you for the comment. Available transcriptomic data sets from infants suspected of sepsis originate mostly from high-income countries. Therefore, our cohort will be unique for exploring the molecular host responses to infections in a low-income country, and consequently provide the opportunity of comparing the two contexts. Although the mNGS portion of the study may provide focused bacteriological data from the mothers, extensively evaluating the maternal exposure to pathogens will be difficult within the scope of this study.

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.

Reply: See similar comment made by the 1st reviewer, we added urine cultures to the protocol as suggested.

Competing Interests: None

Reviewer Report 12 October 2020

<https://doi.org/10.21956/gatesopenres.14371.r29653>

© 2020 Johnson J. This is an open access peer review report 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.



Julia Johnson

Division of Neonatology, Department of Pediatrics, Johns Hopkins University, Baltimore, MD, USA

This is an overall well-articulated and clear study protocol for a research study seeking to assess whole blood genome-wide transcriptome profiling and metagenomics NGS in infants with suspected sepsis. The authors lay out the rationale for the study and have selected the appropriate study population and design to answer this question. Thank you for the opportunity to review the protocol of this interesting study.

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.

Author Response 09 Nov 2020

Pascal Lavoie, BC Children's Hospital Research Institute, Vancouver, Canada

We thank the reviewer for her insightful comments which greatly help us improve and clarify important aspects of the protocol. Below are point-by-point responses to each comment:

Reviewer #1

This is an overall well-articulated and clear study protocol for a research study seeking to assess whole blood genome-wide transcriptome profiling and metagenomics NGS in infants with suspected sepsis. The authors lay out the rationale for the study and have selected the appropriate study population and design to answer this question. Thank you for the opportunity to review the protocol of this interesting study. 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.

Reply: Our initial plan was to recruit infants with other chief complaints without concern for SBI. However, the Malawi IRB felt strongly that this control group was not scientifically adequate and requested that we include in our control group “only infants in whom sepsis was sufficiently unlikely so that antibiotics were not to be started”. In the end, we are collecting antibiotic administration during the hospital stay so we will have the possibility of excluding infants who initially are fairly well appearing but become ill during admission from the control group. Therefore, the risk of misclassification is likely to be extremely low.

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.

Reply: As suggested by the reviewer we have included it, but it is likely that this will not be available in all infants. Urine cultures were part of our initial protocol, but the feasibility in this setting was uncertain.

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 cut-off, if available.

Reply: We have reworded this section to make this clearer (Page 5). The safety concern is about taking an extra research blood sample in smaller, potentially ill babies, i.e. not the swab itself. This was expressed by our clinician colleagues at the Kamuzu Central Hospital. Culturally, they felt that this may be difficult to accept for parents. In making the decision to exclude babies <2.5 kg, we considered that the main objective of the mNGS component of the study is to provide feasibility data so in this context it was not clearly warranted to include smaller babies. The rectal and vaginal swabs are to be done and analyzed in parallel to the blood sample and therefore were not planned to be done in babies <2.5kg for the same reasons.

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.

Reply: Thank you for the comment. We have added this information to the Study size method section (Page 6-7). Blood cultures are not routinely done for neonates suspected of sepsis at KCH, so it is challenging to have baseline microbiological data. Our study will provide blood cultures for the infants suspected of sepsis enrolled in the study and therefore detail the microbiological landscape in this population. However, based on studies done at the Queen Elizabeth Central Hospital, Blantyre, Malawi (also a regional referral centre), we expect a relatively high incidence of Gram-negative pathogens displaying resistance patterns to first-line antimicrobials (Iroh Tam et al, Clin Infect Dis 2019). We will also explore resistance to antibiotics for pathogens detected in our cohort.

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.

Reply: Unfortunately, the logistics of running viral samples in this setting is complex, precluding this as a possibility. There are simply no adequate resources to do this. We have added this to the limitation section of the discussion.

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.

Reply: This has been reworded in the Supplemental Material, reuploaded on the dataverse. In fact, during our on-site study training, we intended to applying the same scrubbing and air drying times for betadine/iodine as for alcohol.

Would specify acceptable sites for blood draws.

Reply: For practical and safety reasons, we entrust the decision to the clinical team at the KCH, as they are best placed to evaluate the infant's clinical status, medical management and level of comfort with the procedure of the healthcare worker performing the blood draw.

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.

Reply: The reviewer is absolutely right commenting that blood pressure q30 min may be too long to provide safe clinical care in a sick baby. However, this is an observational study that aims to document, not intervene. At KCH, measuring blood pressure in neonates has not been integrated in routine clinical care, in absence of proper equipment. Our study provides the first blood pressure monitors for newborns, as well as training for the medical team on their use. We also provide support for the clinical team with BP targets if needed. However, the goal here is to begin integrating vital signs monitoring for infants, considering the challenges in managing low BP in LMICs (lack of continuous monitoring or inotropic

support etc.). In deciding to suggest repeating blood pressure 30 minutes after, we also considered that nurses are profoundly lacking and overstretched in this setting, and IV infusions are largely unavailable.

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.

Reply: Age is recorded as number of completed gestational weeks (the electronic form actually encourages the data collector to enter the exact GA, which we realize is not obvious looking at the material provided).

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.

Reply: That is correct, it stands for breech vaginal delivery. We have had to abbreviate the term in the case report form due to space constraints in the data collection forms.

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.

Reply: Correct, resuscitation is defined as any intervention by medical team. Due to a lack of formal chart documentation this information will often be collected from parents/caregivers who have no idea what these different interventions are. Resuscitation include suctioning, bag and mask ventilation and chest compressions. Presently, neonates are not intubated at KCH. Suctioning alone is not considered resuscitation.

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.

Reply: As per our IRB, it was not possible to collect "dates" (for privacy reasons). We collect the admission date in the infant's chart, and the duration of hospitalization is calculated at the moment of discharge. We trust that this method will be able to provide accurate hospital stays.

Competing Interests: None

Comments on this article

Version 1

Author Response 09 Nov 2020

Tisungane Mvalo, University of North Carolina Project Malawi, Lilongwe, Malawi

Comment from Dr. Tisungane Mvalo, on behalf of Corresponding Author Dr. Pascal Lavoie and rest of investigator team:

We thank both reviewers for their time and for their comments which are very insightful and will tremendously help improve the impact of this study. Below are specific responses to each of their points/comments:

Reviewer #1

This is an overall well-articulated and clear study protocol for a research study seeking to assess whole blood genome-wide transcriptome profiling and metagenomics NGS in infants with suspected sepsis. The authors lay out the rationale for the study and have selected the appropriate study population and design to answer this question. Thank you for the opportunity to review the protocol of this interesting study. 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.

Reply: Our initial plan was to recruit infants with other chief complaints without concern for SBI. However, the Malawi IRB felt strongly that this control group was not scientifically adequate and requested that we include in our control group "only infants in whom sepsis was sufficiently unlikely so that antibiotics were not to be started". In the end, we are collecting antibiotic administration during the hospital stay so we will have the possibility of excluding infants who initially are fairly well appearing but become ill during admission from the control group. Therefore, the risk of misclassification is likely to be extremely low.

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.

Reply (from the authors): As suggested by the reviewer we have included it, but it is likely that this will not be available in all infants. Urine cultures were part of our initial protocol, but the feasibility in this setting was uncertain.

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 cut-off, if available.

Reply: We have reworded this section to make this clearer (Page 5). The safety concern is about taking an extra research blood sample in smaller, potentially ill babies, i.e. not the swab itself. This was expressed by our clinician colleagues at the Kamuzu Central Hospital. Culturally, they felt that this may be difficult to accept for parents. In making the decision to exclude babies <2.5 kg, we considered that the main objective of the mNGS component of the study is to provide feasibility data so in this context it was not clearly warranted to include smaller babies. The rectal and vaginal swabs are to be done and analyzed in parallel to the blood sample and therefore were not planned to be done in babies <2.5kg for the same reasons.

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.

Reply: Thank you for the comment. We have added this information to the Study size method section (Page 6-7). Blood cultures are not routinely done for neonates suspected of sepsis at KCH, so it is challenging to have baseline microbiological data. Our study will provide blood cultures for the infants suspected of sepsis enrolled in the study and therefore detail the microbiological landscape in this population. However, based on studies done at the Queen Elizabeth Central Hospital, Blantyre, Malawi (also a regional referral centre), we expect a relatively high incidence of Gram-negative pathogens displaying resistance patterns to first-line antimicrobials (Iroh Tam et al, Clin Infect Dis 2019). We will also explore resistance to antibiotics for pathogens detected in our cohort.

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.

Reply: Unfortunately, the logistics of running viral samples in this setting is complex, precluding this as a possibility. There are simply no adequate resources to do this. We have added this to the limitation section of the discussion.

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.

Reply: This has been reworded in the Supplemental Material, reuploaded on the dataverse. In fact, during our on-site study training, we intended to applying the same scrubbing and air drying times for betadine/iodine as for alcohol.

Would specify acceptable sites for blood draws.

Reply: For practical and safety reasons, we entrust the decision to the clinical team at the KCH, as they are best placed to evaluate the infant's clinical status, medical management and level of comfort with the procedure of the healthcare worker performing the blood draw.

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.

Reply: The reviewer is absolutely right commenting that blood pressure q30 min may be too long to provide safe clinical care in a sick baby. However, this is an observational study that aims to document, not intervene. At KCH, measuring blood pressure in neonates has not been integrated in routine clinical care, in absence of proper equipment. Our study provides the first blood pressure monitors for newborns, as well as training for the medical team on their use. We also provide support for the clinical team with BP targets if needed. However, the goal here is to begin integrating vital signs monitoring for infants, considering the challenges in managing low BP in LMICs (lack of continuous monitoring or inotropic support etc.). In deciding to suggest repeating blood pressure 30 minutes after, we also considered that nurses are profoundly lacking and overstretched in this setting, and IV infusions are largely unavailable.

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.

Reply: Age is recorded as number of completed gestational weeks (the electronic form actually encourages the data collector to enter the exact GA, which we realize is not obvious looking at the material provided).

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.

Reply: That is correct, it stands for breech vaginal delivery. We have had to abbreviate the term in the case report form due to space constraints in the data collection forms.

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.

Reply: Correct, resuscitation is defined as any intervention by medical team. Due to a lack of formal chart documentation this information will often be collected from parents/caregivers who have no idea what these different interventions are. Resuscitation include suctioning, bag and mask ventilation and chest compressions. Presently, neonates are not intubated at KCH. Suctioning alone

is not considered resuscitation.

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.

Reply: As per our IRB, it was not possible to collect “dates” (for privacy reasons). We collect the admission date in the infant’s chart, and the duration of hospitalization is calculated at the moment of discharge. We trust that this method will be able to provide accurate hospital stays.

Reviewer #2

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?

Reply: Thank you for this comment. We will consider the exposure and possibility of congenital syphilis in analyses of the transcriptomic data. Routine testing for syphilis and HIV is done in Malawi during antenatal visits. According to the WHO Global AIDS Monitoring in 2018, 1% of women tested on antenatal visits were positive for syphilis. Therefore, we expect few subjects in our cohort to have been exposed during pregnancy.

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.

Reply: Good suggestion. We have changed the title to: “**Whole blood genome-wide transcriptome profiling and metagenomics next-generation sequencing in young infants with suspected sepsis in a low- and middle-income country: A study protocol**”. The study is meant to provide the first transcriptome dataset in a LMIC, realizing of course that generalizability will need to be confirmed in future studies in other LMICs.

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.

Reply: In neonates and young infants, these are often used interchangeably due to a lack of precise operational definitions. In this article, we have chosen to use the term “sepsis” which is commonly used in the neonatal literature (for lack of better option), but at the same time we understand that this term is somewhat imprecise.

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).

Reply: Thank you for the suggestion. We have added this clarification to the introduction.

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).

Reply: We have added a reference from the adult literature as suggested.

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.

Reply: See answer to similar comment from 1st reviewer. The need for clinical indication simply refers to the justification for adding an extra blood sample. Informed consent is also obtained from the parent/caregivers 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.

Reply: Thank you for the comment. Indeed, recruitment for our study effectively started in June 2018. We considered publishing the protocol earlier but due to unforeseen delays, we have been unable to do this. In the end, we strongly believe in open access research and in having pre-specified analysis principles and bases before we undertake any analysis of the data, so this is why we insist on publishing this protocol now. This is also in line with our study sponsor's requirement to make all information about the study available publicly as early as possible, including details in the protocol which would likely not be published elsewhere in the future. Realizing also the importance of structured and comparable datasets for cohort studies, especially in LMICs, we feel very proud to finally see this protocol submitted for review.

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)?

Reply: Indeed, the UNC lab has the capacity for manual differential on complete blood counts as needed and will be able to provide this data.

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)?

Reply: We agree that the Ballard score will not be feasible outside the newborn period. This has been clarified (Page 6). In those circumstances, we rely on available medical records or parent/caregiver recall of information.

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.

Reply: We have based our definitions on the WHO list of danger signs, that are also part of the COIN manual in Malawi. Temperature and respiratory rates are recorded automatically on admission. It was very hard for us to conceive that we could accurately capture apneas due to a lack of continuous monitoring or even visual observations due to a profound lack of staff in this setting.

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?

Reply: The 4-hour cut-off was extensively discussed within our study group. First, based on experience we expected that only a small proportion of infants will have received antibiotic treatment for less than 4 hours prior to initial presentation in this setting. We also considered data suggesting that blood cultures positivity rapidly declines after antibiotic administration (Rand KH et al, Open Forum Infect Dis 2019). Finally, we discussed with our Malawian colleagues that it could be ethically challenging not to offer the study to infants who have received antibiotics for less than 4 hours considering that these infants may benefit from the information provided by blood cultures that were only available as part of this study protocol. Beyond 4 hours, we estimated that the benefit of blood cultures would be sufficiently low to ethically and scientifically justify excluding those infants. This clarification was added in the *Participants* section (Page 4).

We have added that: "If necessary, we will separately analyze infants who have received antibiotics prior to enrollment from those who have not" (Definition section; Page 7).

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).

Reply: Correct.

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?

Reply: Breastfeeding is classified as exclusive, mixed (with formula) or formula alone and will be recorded from the caregiver’s history. We did not consider other forms of nutrition, as our study population will include mostly young infants in a hospital setting. Virtually 100% infants at KCH are dependent on breastfeeding as there is little formula milk available (as widely advocated by the WHO).

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?

Reply: This is a very good question that we hope will be informed by this study. One way to address this situation is by providing “upper” and “lower” estimates of “bacterial sepsis” and precision assuming the RNA is wrong versus the blood culture is wrong (i.e. false negatives). In absence of satisfying diagnostic gold standard test for bacterial sepsis we can only assess the differences/agreement between these two approaches.

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.

Reply: We will not perform viral testing in our cohort due to testing availabilities. Malaria testing will be performed if deemed indicated by the treating medical team (e.g. during high seasons), but as per our understanding that although the neonatal mortality from malaria remains high the population incidence is relatively low in this age group (Deribew A et al. Malar J. 2017).

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?

Reply: Thank you for the comment. Available transcriptomic data sets from infants suspected of sepsis originate mostly from high-income countries. Therefore, our cohort will be unique for exploring the molecular host responses to infections in a low-income country, and consequently provide the opportunity of comparing the two contexts. Although the mNGS portion of the study may provide focused bacteriological data from the mothers, extensively evaluating the maternal exposure to pathogens will be difficult within the scope of this study.

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.

Reply: See similar comment made by the 1st reviewer, we added urine cultures to the protocol as suggested.

Competing Interests: None
