STUDY PROTOCOL

The Childhood Acute Illness and Nutrition (CHAIN)
network nested case-cohort study protocol: a multi-omics
approach to understanding mortality among children in sub-
Saharan Africa and South Asia [version 2; peer review: 1
approved, 1 approved with reservations]

James M. Njunge1,2, Kirkby Tickell3, Abdoulaye Hama Diallo4,
Abu Sadat Mohammad Sayeem Bin Shahid5, Md. Amran Gazi5, Ali Saleem6,
Zaubina Kazi6, Syed Ali6, Caroline Tigoi1,2, Ezekiel Mupere7, Christina L. Lancioni8,
Emily Yoshioka3, Mohammod Jobayer Chisti5, Moses Mburu1,2, Moses Ngari1,2,
Narshion Ngao1,2, Bonface Gichuki1,2, Elisha Omer1,2, Wilson Gumbi1,2,
Benson Singa9, Robert Bandsma10,11, Tahmeed Ahmed5, Wieger Voskuijl12,13,
Thomas N. Williams2,14, Alex Macharia2, Johnstone Makale2, Anna Mitchell15,
Jessica Williams15, Joe Gogain15, Nebojsa Janjic15, Rupasri Mandal16,
David S. Wishart16, Hang Wu17, Lei Xia17, Michael Routledge17,18,
Yun Yun Gong17, Camilo Espinosa19-21, Nima Aghaeepour19, Jie Liu22,
Eric Houpt22, Trevor D. Lawley23, Hilary Browne23, Yan Shao23, Doreen Rwigi24,
Kevin Kariuki24, Timothy Kaburu24, Holm H. Uhlig25,26, Lisa Gartner26,
Kelsey Jones27,28, Albert Koullman29,30, Judd Watson31*, James Berkley1,2,31*

1The Childhood Acute Illness and Nutrition Network, Nairobi, Kenya
2KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya
3Global Health and Epidemiology, University of Washington, Seattle, Seattle, USA
4Department of Public Health, Faculty of Health Sciences, University of Ouagadougou, Ouagadougou, Burkina Faso
5Nutrition and Clinical Services Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh
6Department of Pediatrics and Child Health, Aga Khan University Hospital, Karachi, Karachi, Pakistan
7Department of Paediatrics and Child Health, College of Health Sciences, Makerere University, Kampala, Uganda
8Department of Pediatrics, Oregon Health and Science University, Portland, OR, USA
9Kenya Medical Research Institute, Nairobi, Kenya
10Centre for Global Child Health, The Hospital for Sick Children, Toronto, Ontario, Canada
11Department of Biomedical Sciences, University of Malawi College of Medicine, Blantyre, Malawi
12Amsterdam UMC location, University of Amsterdam, Amsterdam, The Netherlands
13Amsterdam Centre for Global Child Health & Emma Children's Hospital, Amsterdam, The Netherlands
14Institute of Global Health Innovation, Department of Surgery and Cancer, Imperial College London, London, UK
15SomaLogic, Boulder, Colorado, USA
16Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada
17School of Food Science and Nutrition, University of Leeds, Leeds, UK
18School of Food and Biological Engineering, Jiangsu University, Zhenjiang, China
Abstract

Introduction: Many acutely ill children in low- and middle-income settings have a high risk of mortality both during and after hospitalisation despite guideline-based care. Understanding the biological mechanisms underpinning mortality may suggest optimal pathways to target for interventions to further reduce mortality. The Childhood Acute Illness and Nutrition (CHAIN) Network (www.chainnnetwork.org) Nested Case-Cohort Study (CNCC) aims to investigate biological mechanisms leading to inpatient and post-discharge mortality through an integrated multi-omic approach.

Methods and analysis: The CNCC comprises a subset of participants from the CHAIN cohort (1278/3101 hospitalised participants, including 350 children who died and 658 survivors, and 270/1140 well community children of similar age and household location) from nine sites in six countries across sub-Saharan Africa and South Asia. Systemic proteome, metabolome, lipidome, lipopolysaccharides, haemoglobin variants, toxins, pathogens, intestinal microbiome and biomarkers of enteropathy will be determined. Computational systems biology analysis will include machine learning and multivariate predictive modelling with stacked generalization approaches accounting for the different characteristics of each biological modality. This systems approach is anticipated to yield mechanistic insights, show interactions and behaviours of the components of biological entities, and help develop interventions to reduce mortality among acutely ill children.

Ethics and dissemination. The CHAIN Network cohort and CNCC was approved by institutional review boards of all partner sites. Results will be published in open access, peer reviewed scientific journals and presented to academic and policy stakeholders. Data will be made publicly available, including uploading to recognised omics databases.
Keywords
Case-Cohort, Mortality, Children, LMIC, Omics, Systems Biology

Trial registration NCT03208725.
Introduction

In 2019, an estimated 5.2 million children under five years died worldwide, predominantly in sub-Saharan and South Asia and mainly due to infectious causes. When children in low- and middle-income countries (LMIC) experience an acute illness, their average mortality risk during and after a hospitalisation event remains unacceptably high and can exceed 20%. This high mortality risk may be due to illness severity, underlying comorbidities such as malnutrition, limited healthcare access and resources, and poverty. The results of the CHAIN cohort, as well as other studies, suggest that half of all mortality in children admitted for acute illness in low- and middle-income countries (LMIC) experience an infectious syndrome, positive bacterial blood cultures are found in only <10% of cases. Studies including The Global Enteric Multicenter Study (GEMS) and The Pneumonia Etiology Research for Child Health (PERCH) have identified epidemiological associations between specific bacterial and viral pathogens and mortality. However, the pathways to mortality during such infections, other than via overwhelming sepsis, are unclear.

Systemic inflammation and altered metabolism

Mortality has been associated with a sepsis-like host-response profile characterized by systemic inflammation and altered metabolism linked to bioenergetic deficits. While deaths are typically associated with these infectious syndromes, positive bacterial blood cultures are found in only <10% of cases. Studies including The Global Enteric Multicenter Study (GEMS) and The Pneumonia Etiology Research for Child Health (PERCH) have identified epidemiological associations between specific bacterial and viral pathogens and mortality. However, the pathways to mortality during such infections, other than via overwhelming sepsis, are unclear.

Amendments from Version 1

Version 2 has been modified based on the reviewers’ comments and suggestions. In the introduction, we have added text to clarify that the biomedical and the social factors influencing mortality risk have already been described in the recent CHAIN cohort publication on childhood mortality during and after acute illness in Africa and south Asia. In the methods section, we have clarified that community children are considered to be well and typical for the community rather than healthy and the criteria for inclusion in the CHAIN cohort. We have also highlighted that the CHAIN study focused on describing the upstream biomedical and social pathways underlying vulnerability since these may be amenable to interventions at an earlier time-point and not on the immediate causes of death which is an objective of the Child Health and Mortality Prevention Surveillance (CHAMPS) Network. Lastly, under the samples size section, we have indicated that increased mortality risk after admission (including >1 year) is now considered a well-recognised phenomenon based on multiple studies in both low- and high-income settings.

Any further responses from the reviewers can be found at the end of the article.
The study protocol was reviewed and expected to deliver more comprehensive mechanistic insights into the biological pathways leading to mortality. The specific focus of CNCC is on generating new hypotheses. Systemic proteome, metabolome, intestinal microbiome, lipidome, lipopolysaccharides, haemoglobin variants, toxins, pathogens and biomarkers of enteric pathology and identifying intervenable mechanisms leading to death among acutely ill children in LMIC may include pre-existing conditions such as sickle cell disease and exposure to toxins including aflatoxins.

Host genetics and toxins
Other contributors to morbidity and mortality among acutely ill children in LMIC may include pre-existing conditions such as sickle cell disease and exposure to toxins including aflatoxins.

The Childhood Acute Illness and Nutrition (CHAIN) Network aims to improve child survival by improving understanding of the biological pathways leading to death among acutely ill children in sub-Saharan Africa and South Asia. The CHAIN cohort was a prospective observational study at nine hospitals which enrolled children aged two to 23 months admitted with acute illness, stratified by anthropometric status, to investigate mortality despite guideline-based care associated with their admission event and for 6 months after discharge. Some biomedical and the social factors influencing mortality risk among acutely ill children within 30 days of hospital admission and during the first 180 days following hospital discharge have recently been described. Briefly, besides lower nutritional status and higher signs of illness severity and HIV, underlying pre-existing medical conditions, adverse caregiver and household-level characteristics and access to health care had both direct and indirect associations with mortality.

This protocol describes the CHAIN Nested Case-Cohort (CNCC) study that brings together a consortium of laboratories to efficiently utilise systematically collected samples from the CHAIN cohort to examine biological pathways leading to mortality. The specific focus of CNCC is on generating mechanistic insights into the biological pathways leading to mortality, including those driven by infectious, nutritional and social factors. This analysis will be both targeted based on prior literature and untargeted discovery work that will generate new hypotheses. Systemic proteome, metabolome, intestinal microbiome, lipidome, lipopolysaccharides, haemoglobin variants, toxins, pathogens and biomarkers of enteropathy will be determined. Computational systems biology analyses across these meta-dimensional data modalities are expected to deliver more comprehensive mechanistic insights into the biological pathways and converging pathological processes leading to mortality. These data will inform the development of novel interventions aimed at mortality risk reduction among acutely ill children.

Methods
Ethical approval: The study protocol was reviewed and approved by the Oxford Tropical Research Ethics Committee, UK; the Kenya Medical Research Institute, Kenya; the University of Washington and Oregon Health and Science University, USA; Makerere University School of Biomedical Sciences Research Ethics Committee and The Uganda National Council for Science and Technology, Uganda; Aga Khan University, Pakistan; the International Centre for Diarrhoeal Disease Research, Bangladesh; The University of Malawi; The University of Ouagadougou and Centre Muraz, Burkina Faso; the Hospital for Sick Children, Canada; and University of Amsterdam, The Netherlands.

Participant enrolment and data collection
The objectives of the CHAIN cohort study, participating sites, patient and public involvement, eligibility, screening, enrolment and follow-up procedures, and specimen collection have been detailed previously. Briefly, 3101 acutely ill children were recruited from nine sites in six countries in sub-Saharan Africa and South Asia. Children were enrolled at admission to hospital and followed up for 180 days after discharge from hospital. Participating sites included rural and urban hospitals in Bangladesh (icddr,b Dhaka Hospital and Matlab Hospital), Burkina Faso (Banfora Regional Referral Hospital), Kenya (Kilifi County Hospital, Mbagathi Sub-County Hospital and Migori County Referral Hospital), Malawi (Queen Elizabeth Central Hospital), Pakistan (Civil Hospital, Karachi), and Uganda (Mulago Hospital). The mid-upper arm circumference (MUAC) was used to enrol children into three strata that included not wasted (MUAC ≥12.5 cm for age ≥six months or MUAC ≥12 cm for age <six months), moderately wasted (MUAC 11.5 to <12.5 cm for age ≥six months or MUAC 11 to <12 cm for age <six months), and severely wasted or kwashiorkor (oedematous malnutrition) (MUAC <11.5 cm for age ≥six months or MUAC <11 cm for age <six months or bilateral pedal oedema unexplained by other medical causes). In addition, children in the same communities as hospitalised cases who were not acutely ill: community participants (CP, N=1140) were recruited at a single time-point to establish community norms for demographic and biological factors. Community children were included into the study if they had no hospital admission in the 14 days prior to contact with the study team and were not ill and therefore are considered to be well and typical for the community rather than ‘healthy’. These community children were from the same neighbourhoods as acutely ill children and therefore, may have had anthropometric deficits, micronutrient deficiencies, helminth infections among others and therefore would not be regarded as ‘healthy’.

CHAIN focused on describing the upstream biomedical and social pathways underlying vulnerability since these may be amenable to interventions at an earlier timepoint and not on the immediate causes of death which is an objective of the Child...
Health and Mortality Prevention Surveillance (CHAMPS)² Network. Causes of death were allocated in CHAIN using the standard WHO methods to identify immediate and underlying causes of death, as published³.

Samples collected at admission and discharge from hospitalised children and from community participants will be utilised in the CNCC together with characteristics of study participants: demographic (age, sex, and site), clinical (shock, severe inflammatory response syndrome, consciousness level, pneumonia, diarrhoea, malaria, HIV, anaemia, blood glucose), anthropometric (nutritional oedema/kwashiorkor, MUAC, and z scores for weight-for-length, weight-for-age, and length-for-age), social data (caregiver characteristics, household-level exposures, and access to health care) and other laboratory tests (clinical chemistry, complete blood count, rectal swab and blood culture) that were systematically undertaken in real-time at each site. Case report forms used to collect data and clinical and sample handling SOPs can be downloaded at www.chainnetwork.org/resources.

Samples collection, processing, and storage
Samples were collected at admission to hospital, discharge and at follow up timepoints. Standard operating procedures (SOPs) for collecting and handling biological samples were prepared and harmonized training was provided at each site to ensure comparability. Time from collection to freezing at -80°C was recorded for each sample with a target window of 30 minutes (Figure 1). Sample types were stool, faecal swabs, whole blood, serum, plasma, and dry blood spots. For each type of sample, one or more aliquots were archived and no further aliquoting will be carried out before shipment.

Figure 1. A. Duration of time from sample collection to storage at -80°C. ~80% of the samples (whole blood, plasma, serum, and rectal swabs) achieved a median of 44 minutes from collection to storage at a -80°C freezer. Except for stool, the rest of the samples were frozen at a -80°C freezer within 1 hour of collection. B. Proportion of samples available for analysis within the nested case cohort study. Greater than 90% of the selected study children have available samples for analysis. Approximately 80% of the study participants have stool samples available for analysis. Adm = admission, disch = discharge.
Samples collected have been transferred from sites to the KEMRI/Wellcome Trust Research Programme biorepository in Kilifi, Kenya.

**Stool:** Faecal samples were collected in faecal pots at home or in hospital by the caregivers or study health workers and transported to the laboratory within 30 minutes in a temperature controlled cool box maintained at 2 – 8°C using ice packs. At the laboratory, samples were aliquoted into sterile, prelabelled and barcoded two mL cryovials and transferred into a -80°C freezer. No additives, preservatives or media was added to the faecal samples.

**Faecal swabs:** Copan (Copan Diagnostics, USA) flocked rectal swabs were collected when study participants were lying on the left lateral position. The swab was inserted gently into the anal canal until it reached the stopper or until resistance was felt. The swab was rotated three times and then slowly removed and inserted into the barcoded retaining tube.

**Blood:** Venepuncture and blood draw was done using aseptic techniques. For plasma, blood collected into EDTA tubes was centrifuged at ~100x g for ten minutes at 6°C and the sample divided into two aliquots and frozen at -80°C freezer. For serum, blood was collected into red top tubes and allowed to clot for ~30 minutes. Serum samples were obtained by spinning the clotted sample at 100x g for ten minutes at 4°C.

**Dried blood spot:** A drop of 75 – 80 µL of whole blood was spotted on a barcoded Whatman Protein Saver 903 card (Merck, Germany), dried for 15 minutes and transferred into a zip lock bag with a desiccant and stored in a -80°C freezer.

The target turnaround time from sample collection to storage was ~30 minutes and a median of 44 minutes (interquartile range; IQR 35 – 57) was achieved (Figure 1). All samples were collected and processed using aseptic techniques while maintaining cold chain from collection to storage. The median samples storage time was 2.2 years (IQR 1.8 – 2.7) before shipment for analysis.

Samples were shipped from Kenya to consortium laboratories for analysis (Figure 2). SOPs used at sites for sample collection, processing and storage can be downloaded at [www.chainnetwork.org/resources](http://www.chainnetwork.org/resources).

**Study design**

A nested case-cohort design drawing from a larger cohort provides estimates that are representative of the entire cohort while also yielding slightly more statistical power than a case control design. Case-cohort selection involves a random subsample of the original cohort (sub-cohort), independently of how cases are defined, followed by additional inclusion of all cases outside the sub-cohort creating the case-cohort set. This results in overrepresentation of the cases in the case-cohort set compared to the original cohort which is dealt with in analysis by weighting. A further advantage of this design over a case-control design is that the sub-cohort can also be used in analysis of other outcomes. In CHAIN, this could include outcomes of readmission or growth among survivors in future analyses, as well comparisons between acutely ill children and community children and between nutritional strata among acutely ill children in the current analysis.

The original CHAIN cohort stratification was maintained in the CNCC design; thus, a random 24% sub-cohort of children enrolled in the CHAIN cohort stratified by site was selected. This included 658 survivors (sub-cohort non-cases) and 109 deaths (sub-cohort cases). Thereafter, all remaining deaths (n=241 cases) that were not selected in the random 24% sub-cohort (Figure 3) were added, giving a total of 350 cases. An additional 30 randomly selected community participants from each site (total N=270) were chosen.

**Sample size**

Multiple studies in both low- and high-income settings indicate a prolonged increased mortality risk after admission (including >1 year) compared to the community peers which is now considered a well-recognised phenomenon. The primary analyses will be of acute mortality occurring within 30-days from admission and post-discharge mortality within 180-days from discharge using samples collected at admission and discharge timepoints. Power and effect size are a function of the number of cases, correlation between cases and survivors, and the ratio of cases to survivors. It is likely that the strength of association between exposures of interests and death will vary in magnitude between the 30-day and 180-day mortality periods. Sample size calculations suggest that a ratio of one case to two non-cases is adequate to reliably detect an HR of 1.5 at >80% power. The design selected all available deaths and simulation-based power calculations suggested that adding more non-cases per cases would yield minimal gains in power while incurring substantial additional expense.

**Laboratory analyses**

The full laboratory protocol is described in the detailed methods section. Briefly, the aptamer based SomaScan assay will be used to quantify ~7000 proteins in plasma. We will also apply both targeted and untargeted mass spectrometry-based approaches to quantify metabolites including amino acids, acylcarnitines, biogenic amines, lipids, sugars, vitamins and minerals in serum. Lipopolysaccharide levels in plasma will be measured via a limulus amoebocyte lysate-based, quantitative chromogenic endpoint assay. The gut microbiome will be determined through metagenomic sequencing of DNA extracted from stool samples. Further, faecal biomarkers of enteropathy including myeloperoxidase, alpha-1-antitrypsin, and calprotectin will be quantified in stool using ELISA. We will use quantitative polymerase chain TaqMan Array Card based assay to test for a custom panel of 35 common enteric pathogens using nucleic acids extracted from rectal swabs. Finally, we will quantify haemoglobin variants using dried blood spots and aflatoxin-albumin in serum samples.
Detailed methods

**Plasma proteomics:** The SomaScan assay that uses SOMAmer (Slow Off-rate Modified Aptamers) reagents to selectively bind and quantify proteins will be used for plasma proteomics profiling. The SomaScan Assay quantitatively transforms the protein epitope availability in a biological sample into a specific SOMAmer-based DNA signal.

A SOMAmer-protein binding step is followed by a series of partitioning and wash steps to convert relative epitope concentrations into measurable nucleic acid signals that are quantified using DNA-hybridisation microarrays. Briefly, diluted plasma samples are incubated with the respective SOMAmer reagent mixes that have been attached to streptavidin (SA)-coated beads (catch 1). Each of the SOMAmers is

---

**Figure 2. Samples collection, processing, storage and shipment workflow.** Consent and sampling was carried out at the site’s ward or clinic. Samples were collected, pre-processed and stored at the study site -80 °C freezers. The duration from sampling to storage was ~44 minutes. Data curation and visualisation was done in R and the shiny application where queries and reports were transmitted online to sites for prompt and timed resolution. Data for samples was captured in the Kilifi Integrated Data Management System (KIDMS) that was linked to patient data in the clinical database captured on Research Electronic Data Capture (REDCap). Samples were subsequently shipped to the central biobank facility at KEMRI Wellcome Research Programme. CHAIN NCC study samples were analysed in the KEMRI-Wellcome Trust Research laboratories or shipped to consortium partners. CHAIN = Childhood Acute Illness and Nutrition, NCC = Nested Case Cohort, EED = Environmental Enteric Dysfunction, Hb = Haemoglobin, KEMRI-CMR = Kenya Medical Research Institute – Centre for Microbiology Research. TMIC = The Metabolomics Innovation Centre, ICDDR, B = International Centre for Diarrhoeal Disease Research, Bangladesh.
We will apply a targeted quantitative metabolomics approach to analyse serum samples using a combination of direct injection mass spectrometry with a reverse-phase liquid chromatography tandem mass spectrometry (LC−MS/MS) custom assay. This custom assay, in combination with an ABSciex 4000 QTrap (Applied Biosystems/MDS Sciex, USA) mass spectrometer, can be used for the targeted identification and quantification of up to 150 different endogenous metabolites including amino acids, acylcarnitines, biogenic amines & derivatives, uremic toxins, glycerophospholipids, sphingolipids and sugars. The method combines the derivatization and extraction of analytes, and the selective mass-spectrometric detection using multiple reaction monitoring (MRM) pairs. Isotope-labelled internal standards and other internal standards are used for metabolite quantification. The custom assay contains a 96 deep-well plate with a filter plate attached with sealing tape, and reagents and solvents used to prepare the plate assay. First 14 wells are used for one blank, three zero samples, seven standards and three quality control samples. For all metabolites except organic acid, samples are thawed on ice and vortexed and centrifuged at 13,000x g. Ten (10) µL of each sample is loaded onto the centre of the filter on the upper 96-well plate and dried in a stream of nitrogen. Subsequently, phenyl-isothiocyanate is added for derivatization. After incubation, the filter spots are dried again using an evaporator. Extraction of the metabolites is achieved by adding 300 µL of extraction solvent. The extracts are obtained by centrifugation into the lower 96-deep well plate, followed by a dilution step with MS running solvent.

For organic acid analysis, 150 µL of ice-cold methanol and 10 µL of isotope-labelled internal standard mixture are added to 50 µL of serum sample for overnight protein precipitation and subsequently centrifuged at 13000x g for 20 minutes. Fifty (50) µL of supernatant is loaded into the centre of a 96-deep well plate, followed by the addition of 3-nitrophenylhydrazine reagent. After incubation for 2h, Butylated hydroxytoluene stabilizer and water are added before LC-MS injection.

Mass spectrometric analysis is performed on an ABSciex 4000 Qtrap® tandem mass spectrometry instrument (Applied Biosystems/MDS Analytical Technologies, Foster City, USA) equipped with an Agilent 1260 series UHPLC system (Agilent Technologies, Palo Alto, USA). The samples are delivered to the mass spectrometer by a LC method followed by a direct injection method. Data pre-processing is done using Analyst 1.6.2.

Serum Micronutrients: Both water-soluble and fat-soluble vitamins found in serum samples are analysed via targeted LC−MS/MS methods using previously described methods. To quantify the water-soluble vitamins in serum samples, 7-point calibration curves are generated by adding 10 µL of the isotopically labelled internal standard mixture to 50 µL of the calibration solutions in eppendorf tubes. Serum samples are also prepared by adding the isotopically labelled internal standard mixture to 50 µL of serum. A total of 60 µL of an aqueous Trichloroacetic acid solution (50 mg/mL) is pipetted to each eppendorf tube containing the calibrants or serum samples. Each tube is vortexed for 30 s for thorough mixing and then left on ice for 1 h. After cooling, each tube
is centrifuged at 13 000 rpm for 20 min, and 100 µL of the supernatant transferred to a new HPLC vial. A volume of 10 µL is injected for LC–MS/MS analysis. An Agilent 1260 series UHPLC system (Agilent Technologies, Palo Alto, USA) coupled with an AB Sciex QTRAP 4000 mass spectrometer (Sciex Canada, Concord, Canada) is used to analyse water-soluble vitamins. An Agilent reversed-phase Zorbax Eclipse XDB C18 column (3.0 mm × 100 mm, 3.5 µm particle size, 80 Å pore size) coupled to a Phenomenex (Torrance, CA) SecurityGuard C18 precolumn (4.0 mm × 3.0 mm) is used for the separation of all water-soluble vitamins in the serum samples.

For the LC–MS/MS analysis of fat-soluble vitamins, 50 µL of calibration solutions or serum samples is pipetted into glass vials, followed by addition of 50 µL of internal standard mixture solution. Subsequently, 300 µL of methanol and 0.2 M ZnSO₄ mixture solution (1:1 v/v) is added to precipitate the serum proteins and to facilitate the release of 25-hydroxyvitamin D₃ from vitamin D binding protein. After this precipitation step, 1 mL of hexane is added to every sample to extract the fat-soluble vitamins. All the samples are vortexed for 10 minutes and centrifuged at 13 000 rpm for 20 minutes. Subsequently 650 µL of the hexane layer is then transferred to a new HPLC vial for evaporation under nitrogen gas at 40°C until dried. Finally, 200 µL of methanol is added to each dried sample to reconstitute the analytes, and 10 µL is injected for LC–MS/MS analysis. An Agilent 1260 series UHPLC system (Agilent Technologies, Palo Alto, CA) coupled with an AB Sciex QTRAP 4000 mass spectrometer (Sciex Canada, Concord, Canada) is used to analyse the fat-soluble vitamins in samples. A Phenomenex Kinetex C18 column (3.0 mm × 100 mm, 2.6 µm particle size, 100 Å pore size) connected to a Phenomenex SecurityGuard C18 precolumn (4.0 mm × 3.0 mm), is used to separate the fat-soluble vitamins. Data pre-processing is done using Sciex Analyst 1.6.2.

Metals and other trace elements in biological samples are measured using inductively coupled plasma mass spectrometry (ICP-MS) which is a powerful and sensitive approach. Trace elemental analysis is done using previously described methods. All metal analysis will be performed on a Perkin-Elmer NexION 350x ICP–MS (Perkin-Elmer, Woodbridge, ON, Canada), operating in a kinetic energy discrimination (KED) mode. Argon (ICP/MS grade, 99.999%) is used as a nebulizer (0.96 mL minute⁻¹), an auxiliary (1 mL minute⁻¹) and a plasma gas (16 mL minute⁻¹). Helium (He) is used as a non-reactive collision gas (Cell gas A: 4.3) to eliminate/minimize chemical interference. The dwell time for each metal ion is set to 50 ms with a total integration time of 500 ms (10 sweeps per reading and three replicates). The uptake of samples/standards/QCs is done by a peristaltic pump using the following protocol: (1) sample flush for 60 s at 48 rpm, (2) read delay for 15 s at 20 rpm, (3) spectral analyses at 20 rpm, and (4) washing for 60 s at 45 rpm. All samples are diluted using 1% HNO₃, 5% H₂O₂, and MilliQ water (grade 1) by a factor of 10. Indium (In) is added to the dilution solvent as an internal standard. The final concentration of indium after mixing with the samples/standards is 20 ppb. An external calibration curve is used for the quantitation of all metal ions using a six- to seven-point calibration curve (for each metal) and linear regression. The performance of the ICP–MS is checked daily using a commercially prepared Perkin Elmer Nexion calibration solution to evaluate the sensitivity of the instrument. The Nexion solution is also used to calibrate the mass spectrometer at low (Be), mid (In), and high (U) masses. The accuracy of the ICP–MS analytical protocol will be evaluated in every sequence by the analysis of standard reference materials (SRMs)—i.e., serum and water QCs. Continuing calibration verification (CCV) is run every 15 samples to monitor the validity of each calibration curve throughout the sequence.

**Serum lipidomics:** For lipid profiling, the serum samples are extracted using the protein precipitation liquid extraction (PPLE) protocol and analysed by liquid chromatography with mass spectrometry detection (LC–MS) as described previously. All solvents and additives are of HPLC grade or higher and purchased from Sigma Aldrich (Haverhill, Suffolk, UK) unless otherwise stated. The PPLE protocol has been described previously. Briefly, 50 µL of plasma is transferred into a 2 mL screw cap Eppendorf plastic tube (Eppendorf, Stevenage, UK). Immediately, 650 µL of chloroform is added to each sample, followed by thorough mixing. Then, 100 µL of the LIPID-IS (5 µM in methanol), 100 µL of the CARNITINE-IS (5 µM in methanol) and 150 µL of methanol are added to each sample, followed by thorough mixing. Then, 400 µL of acetone is added to each sample. The samples are then vortexed and centrifuged for 10 minutes at 20,000 g to pellet any insoluble material. The supernatant is pipetted into separate 2 mL screw cap amber-glass auto-sampler vials (Agilent Technologies, Cheadle, United Kingdom). The organic extracts are dried down to dryness using a Concentrator Plus system (Eppendorf, Stevenage, UK) run for 60 minutes at 60°C. The samples are then reconstituted in 100 µL of 2: 1: 1 (propan-2-ol, acetonitrile and water, respectively) then thoroughly vortexed. The reconstituted sample is transferred to glass-coated 384-well plate (Esslab Plate+™), ready for LC-MS analysis.

Full chromatographic separation of intact lipids is achieved using Waters Acquity H-class UPLC System (Waters, Hertfordshire, United Kingdom) with the injection of 10 µL onto a Waters Acquity UPLC® CSH C18 column (Waters, Hertfordshire, United Kingdom): 1.7 µm, 2.1 mm × 50 mm, maintained at 55°C. Mobile phase A is 6:4, acetonitrile and water with 10 mM ammonium formate. Mobile phase B is 9:1, propan-2-ol and acetonitrile with 10 mM ammonium formate. The flow is maintained at 500 µL per minute through the following gradient: 0.00 minutes, 40% mobile phase B; 0.40 minutes, 43% mobile phase B; 0.45 minutes, 50% mobile phase B; 2.40 minutes, 54% mobile phase B; 2.45 minutes, 70% mobile phase B; 7.00 minutes, 99% mobile phase B; 8.00 minutes, 99% mobile phase B; 8.3 minutes, 40% mobile phase B; 10 minutes, 40% mobile phase B. The sample injection needle is washed using 9:1, 2-propan-2-ol...
and acetonitrile. The Thermo Scientific Exactive Orbitrap mass spectrometer with a heated electrospray ionization source (Thermo Fisher Scientific, Hemel Hempstead, UK) is used. The mass spectrometer is calibrated immediately before sample analysis using positive and negative ionization calibration solution (recommended by Thermo Scientific). Additionally, the mass spectrometer scan rate is set at 4 Hz, giving a resolution of 25,000 (at 200 m/z) with a full-scan range of m/z 100 to 1,800 with continuous switching between positive and negative mode.

The instrument responses of the analytes, selection based on retention time and accurate m/z, are normalized to the relevant internal standard response (producing area ratios). These area ratios correct the intensity for any extraction and instrument variations. The area ratios are then blank corrected where intensities less than three times the blank samples are set to a ‘Not Found’ result (i.e., zero concentration). The accepted area ratios are then multiplied by the concentration of the internal standard to give the analyte semi-quantitative concentrations.

**Plasma lipopolysaccharides:** Endotoxin levels in EDTA plasma samples are measured via a Limulus Amoebocyte lysate-based, quantitative chromogenic endpoint assay (ThermoFisher, UK). The assay is performed based on a protocol described previously. Briefly, 50 μL of plasma is diluted 1:3 in 100 μL 10 mM MgCl₂ solution (Lonza, Switzerland) to prevent inhibition of the Limulus Amoebocyte lysate assay by the EDTA present in the sample. The diluted samples are heated to 70°C for 30 minutes to inactivate interfering proteins, then further diluted 1:10 in 10 mM MgCl₂ solution (Lonza, Switzerland), and plated (50 μL) in pyrogen-free 96-well plates for the Endotoxin detection assay. All samples are run in duplicates. The chromogenic endotoxin detection assay is conducted according to the manufacturer’s instructions. Absorbance is measured at 405 nm with the SPECTROstar Nano (BMG Labtech, Germany) against a standard curve of pure lipopolysaccharide (from *E. coli* O111:B4 (InVivogen, USA)) ranging from 0.05 EU/ml to 1.00 EU/ml. The detection range of the assay is 0.01 EU/ml to 1.00 EU/ml. Due to the variable baseline turbidity of plasma samples interfering with the absorbance measurement, a sample-specific blank is used alongside the standards. The absorbance of the sample-specific blank is deducted from the samples’ absorbance measurements before reading the values against the standard curve. Means of duplicate measurements are taken. Endotoxin concentrations are calculated from the standard curve using a linear model.

**Stool metagenomics:** Faecal samples in cryotubes are removed from the -80°C freezer, allowed to defrost and 30mg is used for DNA extraction. The FastDNA Spin Kit for Soil also incorporates a bead beating step to lyse bacterial cells which is performed using the FastPrep-24™ Classic bead beating grinder and lysis system (also supplied by MP Biomedicals). Negative controls using sterile Phosphate-buffered saline (PBS) are included for each extraction batch (24 samples per extraction batch). Total DNA is eluted in DNase/Pyrogen-free water and stored at -80°C until metagenomic sequencing. DNA samples, including negative controls, are quantified and subjected to paired-end (2 x 151bp) metagenomic sequencing on the NovaSeq (Illumina, USA) S4 XP with 192 samples per flow cell (targeting 13 million reads or 4 Gb / sample). Low-quality bases are trimmed using Trim Galore v0.5.0 with default parameters. Potential human contaminant sequences are removed by screening against the human genome (GRCh38) with BMTagger v1.1.0. Quality-controlled, paired-end reads will be independently assembled with metaSPAdes23 v3.13.0 using default parameters. Taxonomic classification from metagenomics reads will be performed using a reference genome mapping approach using the Human Gastrointestinal Bacteria Genome Collection enriched with Metagenome Assembled Genomes (MAGs) generated from the study metagenomes.

**Faecal pathogen detection:** Total nucleic acid will be extracted from rectal swab and tested using the enteric TaqMan Array Card (TAC, Life Technologies, USA) developed by the Houpt Lab at The University of Virginia. The custom enteric TAC targets multiple pathogens simultaneously including bacteria, viruses, helminths, and protozoa. The card is a 384-well microfluidic real-time PCR format for TAC-compatible instrument platforms. Briefly, both DNA and RNA are extracted from faecal swab specimens using the QIAamp Fast DNA Stool MiniKit (Qiagen, Germany) as previously described. The faecal sample undergo a lysate preparation process that includes mechanical disruption (bead beating), removal of inhibitors, purification and elution of DNA and RNA using spin columns. The extracted total nucleic acid is then used for testing. 40 μL of total nucleic acid extract is mixed with 50 μL of AgPath One Step RT-PCR buffer and 4 μL of enzyme mix (Life Technologies) in a 100 μL reaction, then loaded onto the microfluidic card by centrifugation at 1,200 rpm, 1 min for twice, followed by sealing and cutting off the loading ports. The cards are analysed in a ViiA7 or QuantStudio 7 Flex instrument (Life Technologies, USA). The cycling conditions are set as follows: 45°C for 20 minutes and 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 s and 60°C for 1 minute. Extrinsic controls (MS2 bacteriophage and Phocine herpesvirus) are added to each sample during the lysate preparation to evaluate extraction and amplification efficiency while blank controls help rule out contamination during the extraction process. The assay uses predefined thresholds although the baseline for each target may require adjustment. A sample is deemed positive for a particular target when an appropriately shaped amplification curve crosses the analytical threshold for positivity before or at cycle 34.9 and negative if the curve does not cross the threshold line or crosses the threshold line at cycle 35.0 or later. Pathogen quantities in copy number are derived based on the standard curve for each target.

**Faecal biomarkers of enteropathy:** Stool myeloperoxidase, alpha-1-antitrypsin, and calprotectin will be quantified using the ELISA assay (Immundiagnostik AG, Germany) according...
to manufacturer’s instructions. The absolute concentration of the three biomarkers in samples will be calculated for 15 mg of stool using a dose response curve generated from the manufacturer’s standards.

**Haemoglobin variants:** Ion-exchange high-performance liquid chromatography (HPLC) will be used for separation and quantification of HB variants in dry blood spots (DBS) using the D-10 Haemoglobin Testing System (Bio-Rad, USA). DBS are resuspended on a lysing buffer to release haemoglobin. The samples are then diluted on the D-10 Haemoglobin Testing System (Bio-Rad, USA) and injected into the analytical cartridge. The D-10 delivers a programmed buffer gradient of increasing ionic strength to the cartridge, where the haemoglobin variants are separated based on their ionic interactions with the cartridge material. The separated haemoglobin then pass through the flow cell of the filter photometer, where changes in the absorbance at 415 nm are measured. The D-10 System will be used for quantification of haemoglobins variants namely: A, S, F and A2 and other minor variants and diagnosis of B-thalassaemia and sickle cell anaemia among study children.

**Serum aflatoxin:** The analysis of aflatoxin-albumin (AF-alb) will follow the method first described by Chapot and Wild and detailed by Xu and colleagues. In brief, albumin is extracted from serum samples, hydrolysed overnight using pronase and purified with Sep-pak C-18 cartridges (Waters, MO, USA). AF-alb is quantified using a competitive ELISA with a limit of detection (LOD) of 3pg AF-lysine equivalents per mg albumin. Each ELISA includes positive and negative control samples for quality control. All samples are measured in duplicate for each ELISA batch and repeated at least two times on separate days to confirm the results. Results are accepted when values within each ELISA have a %CV below 10% and samples tested on separate occasions had a %CV below 15%. Samples with concentrations above the linear range of the standard curve are repeated at a suitable dilution. This ELISA has been validated for measuring aflatoxin albumin biomarker levels against dietary aflatoxin intake, and also against an LC-MS assay for measuring aflatoxin lysine. Whilst the ELISA consistently gives AF-alb values that are approximately three-fold higher than aflatoxin lysine values, the results between these methods correlate for the same samples. The AF-alb values are, therefore, suitable for use in analyses to compare variations in aflatoxin exposure in relation to health outcomes or other variables such as child growth.

**Statistical analysis**

**Descriptive analysis.** The distribution and prevalence of biomarkers in laboratory assays which have clear clinical interpretations will be described at admission and discharge timepoints across the nutritional strata and compared with community participants. These more targeted assays will include the serum micronutrients, serum aflatoxin biomarkers, plasma lipopolysaccharides, haemoglobin variants, faecal biomarkers of enteropathy, and faecal pathogens. For these assays, the crude and confounder adjusted case fatality rates, which will be appropriately weighted for selection into the NCC analysis and the original CHAIN cohort, will be described.

The appropriate non-parametric statistical test will be employed to calculate the significance of the association between each feature and each outcome or measure of interest in every relevant subset of patients. For binary outcomes (i.e., the primary outcome of mortality), the Wilcoxon rank sum test will be used to determine the significance of the association while Hedge’s g will be used to calculate the effect size of the difference between groups. For continuous measures, the Spearman’s rank correlation coefficient will be used to assess the strength and significance of the association between the feature and the measure. Linear regression models will be fit for each feature against the confounding variables of interest and the residuals taken as the confounder-adjusted features.

**Integrated analysis**

Advanced data integration and pathway analysis using systems biology data modelling to investigate integrated mechanistic pathways and reveal novel interactions between different biological modalities will be employed. Cross-validated multivariate models predicting the outcomes or measures of interest will be derived from the full dataset using a gradient-boosted tree algorithm (XGBoost). A two-layer repeated cross-validation scheme will be used to prevent overfitting, optimize the parameters of the model, and get an estimate of the performance of the model on unseen patients. In brief, patients will be randomly and evenly split into a training set and a test set. An XGBoost model will be trained on the training set and then used to make predictions for the patients in the test set. This procedure will be performed 50 times using different and random train and test splits in each iteration. Final test predictions for each patient will be generated by integrating the model’s predictions for the patient using only the predictions from iterations in which the relevant patient was in the test set. Classifier model performance will be assessed with the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC). Regression model performance will be assessed using Pearson’s correlation coefficient (Pearson’s R).

**Stacked generalization**

To account for the differing information content between biological modalities and allow each modality to participate equally in the predictive models for the outcomes of interest, a stacked generalization approach will be employed. An independent predictive model will be trained using only the data from each modality while maintaining the same cross-validation folds across all modalities. A meta-learner will be subsequently trained to combine the predictions from each independent biological modality. Ultimately, each modality’s predictive model and the global meta-learner will be used to generate the final cross-validated test predictions for each patient in the dataset.
Pathway enrichment analysis
Gene set overrepresentation analysis of Gene Ontology (GO) Terms for the top genes and proteins significantly associated with the outcomes of interest will be performed in R using the topGO package (Version 2.42.0) [7]. The Biological Process ontology will be used to define GO terms for query. Fisher’s exact test will be used to calculate the p-value of enrichment for each GO term queried. Where relevant, top genes and proteins will be split into upregulated and downregulated sets before analysis.

Adjustment for multiple comparisons
All p-values will be corrected for multiple hypothesis testing using Bonferroni’s correction where appropriate. In brief, the raw p-value will be multiplied by the total number of statistical tests performed to generate the adjusted p-value.

Conclusions
The CHAIN Nested Case-Cohort Study is a unique opportunity to understand the mechanisms leading to early and later deaths among acutely ill children. The study brings together a consortium of multiple world-leading laboratories and integrated systems biology analyses of meta-dimensional data to provide a novel perspective on molecular mechanisms and converging pathological processes leading to mortality among acutely ill children. These data will inform the development of novel interventions aimed at mortality risk reduction among vulnerable children.

Declarations
Data availability
Where appropriate, the raw data will be deposited to public open access repositories. The processed data and analysis code will be deposited on the Harvard Dataverse under the Biosciences subtheme (https://dataverse.harvard.edu/dataverse/kwirp) and may be requested via email to dgc@kemri-wellocme.org.

Acknowledgments
This paper is published with permission from the Director of KEMRI.

References


Open Peer Review

**Current Peer Review Status:** ✔️ ✨

---

**Version 1**

**Reviewer Report 29 July 2022**

https://doi.org/10.21956/gatesopenres.14914.r32265

© 2022 Bassat Q. 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.

**Quique Bassat**

Barcelona Institute for Global Health (ISGlobal), Hospital Clínica de Barcelona, University of Barcelona, Barcelona, Spain

This is a good manuscript reflecting rather accurately the protocol that the researchers have drafted for this nested case-control study. I have only a few comments to make which I believe would make the overall manuscript slightly clearer.

1. **Abstract:** "...270/1140 well community children". Would use "healthy" rather than "well".

2. **Rationale/intro:** Biological factors will surely have an enormous weight, but it is critical not to disregard "social" (i.e. poverty-related; access-related, etc.) factors. The authors mention this, but it is not clear to me how deaths will be analysed taking them into account. This should be better explained. In other words, it may very well be that investigators find novel mechanistic hypotheses of mortality, but how much will those be put in the context of the overwhelming social determinants of death?

3. **One of the missing pieces in this project is the link with the more thorough cause of death estimates. In other words, no postmortem work considered the gold standard approach for CoD investigation was part of or is planned as part of CHAIN, if I am not wrong. Doing all the mechanistic work without a clear description of the chain of events leading to death seems to me a bit of a futile exercise as investigators may be convinced that a given child died of disease X but may not have the clear postmortem confirmation of that supposition. How can this project link with already existing efforts (CHAMPS, etc.) in relation to CoD ascertainment and child mortality surveillance?**

4. **In sepsis (irrespective of underlying etiology), one of the common pathways and transversal to different underlying infections is endothelial and immune activation, leading to loss of endothelial stability and eventual end-organ dysfunction. How many of these pathways are going to be investigated? How will their functionality be measured?**

5. **What will the methods be for CoD ascertainment in CHAIN? What methods have been used**
to investigate the underlying etiologies?

6. In Box 1, and understanding that you only discuss biological pathways, it may be useful to add a brief paragraph on their interrelation with social determinants.

7. Among the social data considered, is there any way of assessing socio-economic status? If so, which way?

8. Among the clinical data described, add "outcome data" and "Cause of Death" data, and specify by which methods they were determined, whether verbal autopsy, clinical records, etc.

9. I notice post-discharge period goes up to 6 months (180 days). How will you take into account mortality due to any other non-related causes (e.g. car accident) after discharge? Will you correct this? Wouldn't it make more sense to choose a shorter timeframe (i.e. 1 month or 3 months) rather than such a long period (6 months)?

10. I still find it unclear by reading the manuscript what the primary endpoints/measurements/objectives of this study are. Perhaps adding a final box at the end of the paper with "Expected results" may help it be more clear?

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?
Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology, cause of death, mortality surveillance, malaria

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 03 Aug 2022

james njunge, KEMRI/Wellcome Trust Research Programme, Nairobi, Kenya

Reviewer: Quique Bassat, Barcelona Institute for Global Health (ISGlobal), Hospital Clinic
This is a good manuscript reflecting rather accurately the protocol that the researchers have drafted for this nested case-control study. I have only a few comments to make which I believe would make the overall manuscript slightly clearer.

Abstract: "...270/1140 well community children". Would use "healthy" rather than "well".

Community children\(^1\) were included in the study if they had no hospital admission in the 14 days prior to contact with the study team and were not ill and therefore are considered to be well and typical for the community rather than healthy. These community children were from the same neighbourhoods as acutely ill children and therefore, may have had anthropometric deficits, micronutrient deficiencies, helminths, etc. and so we would not regard them as optimally healthy.

**Rationale/intro:** Biological factors will surely have an enormous weight, but it is critical not to disregard "social" (i.e. poverty-related; access-related, etc.) factors. The authors mention this, but it is not clear to me how deaths will be analysed taking them into account. This should be better explained. In other words, it may very well be that investigators find novel mechanistic hypotheses of mortality, but how much will those be put in the context of the overwhelming social determinants of death?

The Childhood Acute Illness and Nutrition (CHAIN) cohort study aimed to characterise the biomedical and social risk factors for mortality in acutely ill children in hospitals and after discharge to identify targeted interventions to reduce mortality\(^1\) (ref 58). Some of the biomedical and social factors which we examined that influence mortality among acutely ill children within 30 days of hospital admission and during the first 180 days following hospital discharge have already been described within the published CHAIN epidemiological paper\(^2\) (ref 12). Briefly, lower nutritional strata at hospital admission, higher signs of illness severity at admission, HIV, underlying pre-existing medical conditions, adverse caregiver and household-level exposures, access to health care, and discharge against medical advice were associated with mortality. The specific focus of the CHAIN Network nested case-cohort (NCC) study is on generating mechanistic insights into the biological pathways leading to mortality, including those driven by nutritional and social factors.

One of the missing pieces in this project is the link with the more thorough cause of death estimates. In other words, no postmortem work considered the gold standard approach for CoD investigation was part of or is planned as part of CHAIN, if I am not wrong. Doing all the mechanistic work without a clear description of the chain of events leading to death seems to me a bit of a futile exercise as investigators may be convinced that a given child died of disease X but may not have the clear postmortem confirmation of that supposition. How can this project link with already existing efforts (CHAMPS, etc.) in relation to CoD ascertainment and child mortality surveillance?

We would like to thank the reviewer for this comment. CHAIN\(^2\) focussed on describing the upstream biomedical and social pathways underlying vulnerability since these are more amenable to interventions at a much earlier timepoint and not on the immediate causes of death which is the objective of the Child Health and Mortality Prevention Surveillance.
(CHAMPS)$^3$ Network. Unlike CHAMPS, CHAIN has considerable data in life in most cases long before the specific illness that caused death.

Causes of death were allocated in CHAIN using the standard WHO methods to identify immediate and underlying causes of death, as presented in the published epidemiological paper$^2$. We recognize that such an approach has limitations.

**In sepsis (irrespective of underlying etiology), one of the common pathways and transversal to different underlying infections is endothelial and immune activation, leading to loss of endothelial stability and eventual end-organ dysfunction. How many of these pathways are going to be investigated? How will their functionality be measured?**

Our previous work on mortality has shown that immune activation, endothelial dysfunction, and metabolic disturbances underlie mortality among acutely ill children with severe malnutrition$^4,5$. CHAIN NCC is combining data collected within the CHAIN cohort and conducting a large panel of omics laboratory investigations (systemic proteome, metabolome, lipidome, intestinal microbiome, etc.) to characterise multiple interconnected pathways including, but not limited to, endothelial function, immune function and inflammation, microbial translocation, micronutrient deficiency, altered metabolism and markers of organ dysfunction. We aim to use a systems biology data modelling approach to investigate these integrated mechanistic pathways and reveal novel interactions between different biological modalities.

**What will the methods be for CoD ascertainment in CHAIN? What methods have been used to investigate the underlying etiologies?**

Details are provided in response to comment 4 above. Some underlying etiologies were determined during the cohort study, as presented in the published epidemiological paper$^2$. Others are being determined within the NCC, as described in this manuscript.

**In Box 1, and understanding that you only discuss biological pathways, it may be useful to add a brief paragraph on their interrelation with social determinants.**

Details are provided in response 2 above. The focus of the CHAIN NCC is on generating mechanistic insights into the biological pathways leading to mortality with an overall aim of identifying novel biological interventions to reduce mortality among children in LMIC. The social determinants of death were described in the published epidemiological paper$^2$.

**Among the social data considered, is there any way of assessing socio-economic status? If so, which way?**

This has already been described within the published epidemiological paper$^2$ with extensive details in the supplementary appendix.

**Among the clinical data described, add “outcome data” and “Cause of Death” data, and specify by which methods they were determined, whether verbal autopsy, clinical records, etc.**
This has already been described within the published epidemiological paper\textsuperscript{2}.

\textit{I notice post-discharge period goes up to 6 months (180 days). How will you take into account mortality due to any other non-related causes (e.g. car accident) after discharge? Will you correct this? Wouldn't it make more sense to choose a shorter timeframe (i.e. 1 month or 3 months) rather than such a long period (6 months)?}

Multiple studies in both low- and high-income settings indicated a prolonged increased mortality risk after admission (including >1 year) compared to the community peers\textsuperscript{6-8} which is now a well-recognised phenomenon. This is one of the key puzzles that CHAIN is trying to unravel in LMIC settings. Of note, this may include a higher rate of traumatic deaths, possible because a child with impaired metabolic and immunological resilience or with underlying comorbidities may be more likely to die as a result of an injury than a child without these vulnerabilities.

\textit{I still find it unclear by reading the manuscript what the primary endpoints/measurements/objectives of this study are. Perhaps adding a final box at the end of the paper with "Expected results" may help it be more clear?}

We have indicated in the text under the sample size section that “The primary analyses will be of acute mortality occurring within 30-days from admission and post-discharge mortality within 180-days from discharge using samples collected at admission and discharge timepoints.” Additionally, in the conclusion section, we have indicated that the analysis will “provide a novel perspective on molecular mechanisms and converging pathological processes leading to mortality among acutely ill children. These data will inform the development of novel interventions aimed at mortality risk reduction among vulnerable children.”

References


**Competing Interests:** None

---

**Reviewer Report 28 July 2022**

https://doi.org/10.21956/gatesopenres.14914.r32267

© 2022 Freemark M. 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.

Michael Scott Freemark  
Division of Pediatric Endocrinology and Diabetes, Duke University School of Medicine, Durham, NC, USA

This should be a comprehensive analysis of the factors associated with mortality in malnourished children.

Two questions:

1. I presume HIV status will be incorporated into the statistical analysis; is this correct?

2. The authors do not specify the measurement of inflammatory cytokines or hormones/growth factors shown previously to associate with mortality; is there a reason for that?

**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:** No competing interests were disclosed.

**Reviewer Expertise:** Pathogenesis of malnutrition, obesity and insulin resistance

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.

---

Author Response 03 Aug 2022

**james njunge,** KEMRI/Wellcome Trust Research Programme, Nairobi, Kenya

**Reviewer:** **Michael Scott Freemark,** Division of Pediatric Endocrinology and Diabetes, Duke University School of Medicine, Durham, NC, USA

*This should be a comprehensive analysis of the factors associated with mortality in malnourished children.*

**Two questions:**

**I presume HIV status will be incorporated into the statistical analysis; is this correct?**

Yes, the HIV status will be included in the analysis.

**The authors do not specify the measurement of inflammatory cytokines or hormones/growth factors shown previously to associate with mortality; is there a reason for that?**

We will use the SomaScan platform that quantifies approximately 7000 proteins. The analytes detected include a large array of inflammatory proteins including cytokines and chemokines, hormones, and growth factors among others. See Plasma proteomics under the detailed methods section. The specific list of analytes will be provided in subsequent manuscripts.

**Competing Interests:** None