Check for updates

STUDY PROTOCOL

REVISED
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: 2 approved]

James M. Njunge [1,2], Kirkby Tickell[3], Abdoulaye Hama Diallo[4],
Abu Sadat Mohammad Sayeem Bin Shahid[5], Md. Amran Gazi[5], Ali Saleem[6],
Zaubina Kazi[6], Syed Ali[6], Caroline Tigoi[1,2], Ezekiel Mupere[7], Christina L. Lancioni[8],
Emily Yoshioka[3], Mohammod Jobayer Chisti[5], Moses Mburu[1,2], Moses Ngari[1,2],
Narshion Ngao[1,2], Bonface Gichuki [1,2], Elisha Omer[1,2], Wilson Gumbi[1,2],
Benson Singa [9], Robert Bandsma[10,11], Tahmeed Ahmed[5], Wieger Voskuijl [12,13],
Thomas N. Williams[2,14], Alex Macharia[2], Johnstone Makale[2], Anna Mitchel[15],
Jessica Williams[15], Joe Gogain[15], Nebojsa Janjic[15], Rupasri Mandal[16],
David S. Wishart[16], Hang Wu [17], Lei Xia[17], Michael Routledge[17,18],
Yun Yun Gong [17], Camilo Espinosa[19-21], Nima Aghaeepour[19], Jie Liu[22],
Eric Houpt [22], Trevor D. Lawley[23], Hilary Browne[23], Yan Shao[23], Doreen Rwigi[24],
Kevin Kariuki [24], Timothy Kaburu [24], Holm H. Uhlig[25,26], Lisa Gartner[25],
Kelsey Jones[27,28], Albert Koulman[29,30], Judd Walson [3*], James Berkley[1,2,31*]

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

[19]Departments of Anesthesiology, Pain, and Perioperative Medicine, Stanford University School of Medicine, Stanford, CA, 94305, USA

[20]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, 94305, USA

[21]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford University School of Medicine, Stanford, CA, 94305, USA

[22]Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, Virginia, USA

[23]Wellcome Sanger Institute, Hinxton, UK

[24]The Centre for Microbiology Research, Kenya Medical Research Institute, Nairobi, Kenya

[25]Translational Gastroenterology Unit, John Radcliffe Hospital, University of Oxford, Oxford, UK

[26]Department of Paediatrics and Biomedical Research Centre, University of Oxford, Oxford, UK

[27]Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK

[28]Gastroenterology Department, Great Ormond Street Hospital for Children, London, UK

[29]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

[30]NIHR BRC Nutritional Biomarker Laboratory, University of Cambridge, Cambridge, UK

[31]Center for Tropical Medicine and Global Health, University of Oxford, Oxford, UK

* Equal contributors

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.chainnetwork.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.

Corresponding authors: Judd Walson (walson@uw.edu), James Berkley (jberkley@kemri-wellcome.org)

Author roles: Njunge JM: Conceptualization, Funding Acquisition, Methodology, Project Administration, Writing – Original Draft Preparation; **Tickell K**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Diallo AH**: Writing – Review & Editing; **Sayeem Bin Shahid ASM**: Methodology, Writing – Review & Editing; **Gazi MA**: Investigation, Methodology, Writing – Review & Editing; **Saleem A**: Methodology, Writing – Review & Editing; **Kazi Z**: Methodology, Writing – Review & Editing; **Ali S**: Methodology, Writing – Review & Editing; **Tigoi C**: Project Administration, Writing – Review & Editing; **Mupere E**: Writing – Review & Editing; **Lancioni CL**: Writing – Review & Editing; **Yoshioka E**: Project Administration, Writing – Review & Editing; **Chisti MJ**: Methodology, Writing – Review & Editing; **Mburu M**: Project Administration, Writing – Review & Editing; **Ngari M**: Methodology, Software, Writing – Review & Editing; **Ngao N**: Project Administration, Writing – Review & Editing; **Gichuki B**: Investigation, Methodology, Project Administration, Writing – Review & Editing; **Omer E**: Investigation, Methodology, Project Administration, Writing – Review & Editing; **Gumbi W**: Investigation, Methodology, Writing – Review & Editing; **Singa B**: Writing – Review & Editing; **Bandsma R**: Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing; **Ahmed T**: Conceptualization, Funding Acquisition, Writing – Review & Editing; **Voskuijl W**: Writing – Review & Editing; **Williams TN**: Methodology, Writing – Review & Editing; **Macharia A**: Methodology, Writing – Review & Editing; **Makale J**: Investigation, Methodology, Writing – Review & Editing; **Mitchel A**: Investigation, Methodology, Writing – Review & Editing; **Williams J**: Investigation, Methodology, Writing – Review & Editing; **Gogain J**: Methodology, Writing – Review & Editing; **Janjic N**: Methodology, Writing – Review & Editing; **Mandal R**: Investigation, Methodology, Writing – Review & Editing; **Wishart DS**: Investigation, Methodology, Writing – Review & Editing; **Wu H**: Investigation, Methodology, Writing – Review & Editing; **Xia L**: Investigation, Methodology, Writing – Review & Editing; **Routledge M**: Methodology, Writing – Review & Editing; **Gong YY**: Methodology, Writing – Review & Editing; **Espinosa C**: Methodology, Software, Visualization, Writing – Review & Editing; **Aghaeepour N**: Methodology, Software, Writing – Review & Editing; **Liu J**: Investigation, Methodology, Writing – Review & Editing; **Houpt E**: Methodology, Resources, Writing – Review & Editing; **Lawley TD**: Methodology, Software, Writing – Review & Editing; **Browne H**: Investigation, Methodology, Software, Writing – Review & Editing; **Shao Y**: Investigation, Methodology, Software, Writing – Review & Editing; **Rwigi D**: Investigation, Methodology, Writing – Review & Editing; **Kariuki K**: Investigation, Methodology, Writing – Review & Editing; **Kaburu T**: Investigation, Methodology, Writing – Review & Editing; **Uhlig HH**: Investigation, Methodology, Writing – Review & Editing; **Gartner L**: Investigation, Methodology, Writing – Review & Editing; **Jones K**: Investigation, Methodology, Writing – Review & Editing; **Koulman A**: Investigation, Methodology, Writing – Review & Editing; **Walson J**: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; **Berkley J**: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

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

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[1,2]. 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%[3–11]. 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 these settings occurs after discharge[5,11,12] but a large group of children can be identified as low risk. In the past decade, very few novel biological interventions have demonstrated the ability to reduce mortality among children in LMIC with acute illnesses[13,14]. Interventions that have been tested in clinical trials have often been translated from high-income settings rather than being based on a thorough mechanistic understanding of child mortality in LMICs. Finally, typical interventions commonly do not address the multiple biological factors that contribute to acute and post-discharge mortality such as the inciting illness, nutrient deficiencies and other comorbidities including enteropathy and intestinal dysbiosis (Box 1). These observations highlight the need for an improved understanding of the mechanisms leading to death and for the development of new interventions for prevention of morbidity and mortality.

Box 1. Multiple and Interconnected biological pathways lead to mortality

Infection
Acutely ill children in LMICs typically present with respiratory infection, diarrhoea or encephalopathy and across these syndromes, those with undernutrition are at higher risk of mortality. While deaths are typically associated with these

---

infectious syndromes, positive bacterial blood cultures are found in only <10% of cases[15,16]. Studies including The Global Enteric Multicenter Study (GEMs) and The Pneumonia Etiology Research for Child Health (PERCH) have identified epidemiological associations between molecular evidence of specific bacterial and viral pathogens and mortality[17,18]. 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[2,19–22]. In sepsis, mortality has been linked to immunoparalysis, organ damage including the cardiac, respiratory, hepatic and renal systems, uncontrolled inflammation, proteolysis, and defects of organ healing capability[23–34]. Mitochondrial and peroxisomal dysfunction involves kynurenine derivatives including intermediates of redox cofactors nicotinamide adenine dinucleotide (NAD), acylglycerophosphocholines, acyl-carnitines, and bile acids. There is also a well-recognised association between the degree of mitochondrial dysfunction and sepsis outcomes suggesting a role of immunometabolism and suggesting new routes for therapeutic intervention[35–37]. There is much overlap in mechanisms described in sepsis and in severely malnourished children suggesting that deaths in malnutrition are largely due to sepsis and that malnutrition may predispose to these phenomena when sepsis occurs[22]. Finally, the role of hormones in metabolic regulation and reprogramming in sepsis and undernutrition has not been well investigated but recent studies show that leptin is associated with mortality[19,20,38].

Nutrient deficiencies
There is limited understanding of micronutrient pathophysiology and requirements during illness especially among children in LMIC. Micronutrient deficiencies are known underlying contributors to increased risk of morbidity and mortality with iron, iodine, folate, vitamin A, and zinc deficiencies being the most widespread[39,40]. During critical illness, micronutrient deficiency may impair enzyme activity and vital processes such as energy metabolism and contribute to oxidative stress[40,41]. Disrupted energy supply to cellular processes and organ dysfunction may result in final common pathways leading to mortality, or syndrome-specific pathways such as energetic exhaustion due to a prolonged, greatly increased work of breathing in pneumonia. Previous studies have mostly not investigated the mortality occurring in children soon after hospital discharge and prolonged elevated mortality risk, although long-term immune paralysis, metabolic and functional abnormalities are recognised phenomenon post-sepsis or ICU care[42–44].

Enteric dysfunction
It is recognised that many children in LMIC settings often exhibit intestinal mucosal changes, including inflammation, epithelial barrier dysfunction (leaky gut) and blunted villi which are similar to those observed in children with environmental enteric dysfunction[45–47]. Enteric barrier dysfunction may lead to translocation of organisms into the circulation. A circulating microbiome consistent with intestinal microbiota (*Lactobacillus sp., Bacteroides sp., and Streptococcus sp.*) has been correlated with products of gut bacterial translocation such as lipopolysaccharide (LPS) and D-lactate and systemic inflammation and predicted adverse cardiovascular events[48]. Biomarkers of enteric dysfunction[49] have not been assessed in relation to mortality risk or linked to specific infection episodes. However, limited data from LMIC settings suggests that in children with severe malnutrition, intestinal inflammation is linked to mortality via systemic inflammation rather than directly[21].

The gut microbiota
The gut microbiota protects against pathogens and maintains normal immune and metabolic homeostasis[50–52]. Loss of diversity and imbalance in composition of the microbiota, termed dysbiosis, is clearly described but associations with mortality risks are not. Dysbiosis is likely correlated with illness severity and has been linked to increased mortality in critically ill patients including sepsis[53,54], in patients undergoing allogeneic hematopoietic-cell transplantation[55], in hospitalized patients with cirrhosis[56] among others. A marked insult on the microbiome is likely when children are hospitalized and exposed to nosocomial microorganisms, and to selection pressure from antimicrobials, resulting in acquisition of antimicrobial resistance genes. The health consequences of these likely hospital-acquired organisms, which may also differ in virulence from community-derived species, is unclear.

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 and identifying intervenable mechanisms leading to death among acutely ill children in sub-Africa and South Asia[57]. 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[12]. 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[57]. 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 ≥12.5cm for age ≥six months or MUAC ≥12cm 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 <11.5cm for age ≥six months or MUAC <11cm 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[57] 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[12] 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 published[12].

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.
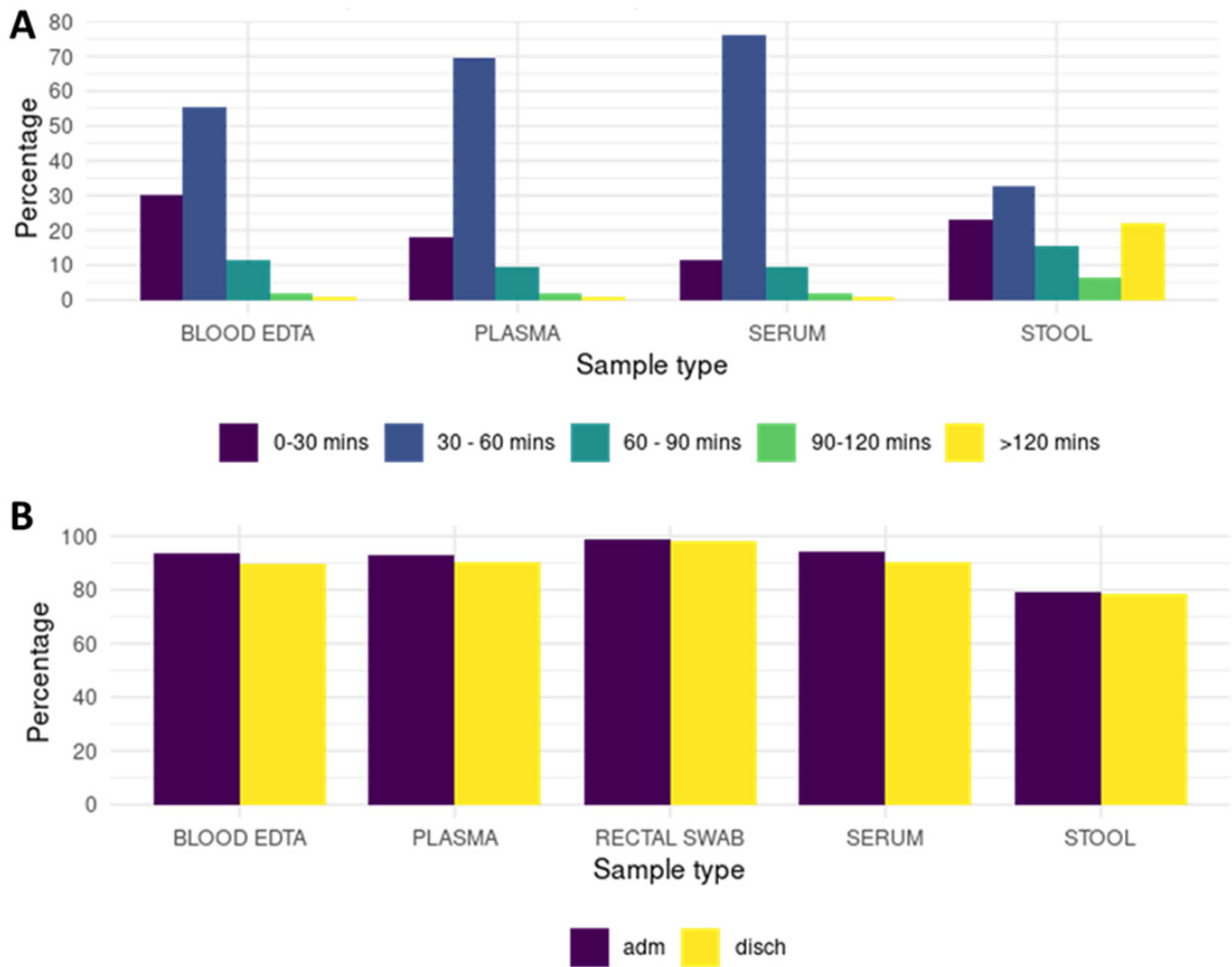
Figure 1. **A.** Duration of time from sample collection to storage at -80°C. ~80% of the samples (whole blood, plasma, serum, and rectal swabs) achieved a median of 44 minutes from collection to storage at a -80°C freezer. Except for stool, the rest of the samples were frozen at a -80°C freezer within 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.

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[58]. 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[59]. 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[10,43,60] 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[61]. 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[62–64]. Lipopolysaccharide levels in plasma will be measured via a limulus amoebocyte lysate-based, quantitative chromogenic endpoint assay[65]. 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[66] 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[67–69] in serum samples.
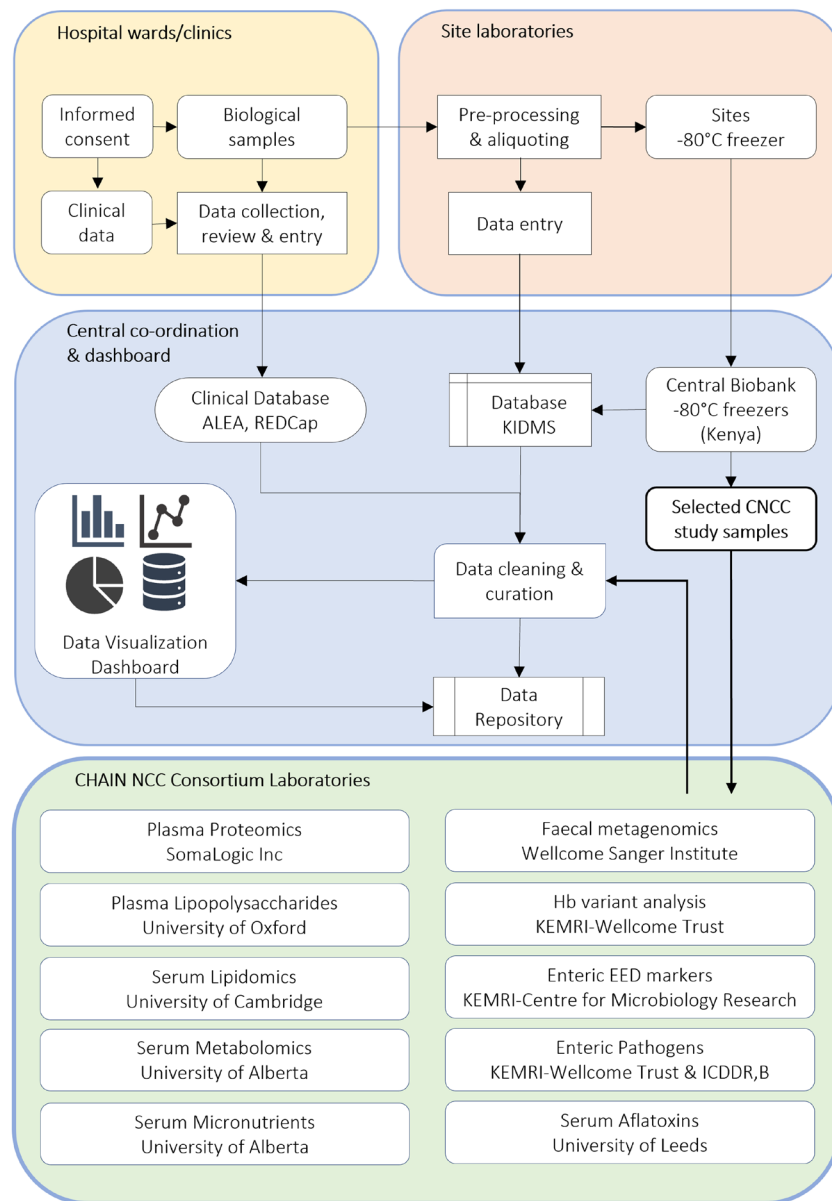
Figure 2. Samples collection, processing, storage and shipment workflow. Consent and sampling was carried out at the site's ward or clinic. Samples were collected, pre-processed and stored at the study site -80 °C freezers. The duration from sampling to storage was ~44 minutes. Data curation and visualisation 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.

Detailed methods

Plasma proteomics: The SomaScan assay that uses SOMAmer (Slow Off-rate Modified Aptamers) reagents to selectively bind and quantify proteins[61] 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 3. Sample selection for the CHAIN Nested Case-Cohort Study. Areas are proportional to the number of observations and excluding community participants are not included.

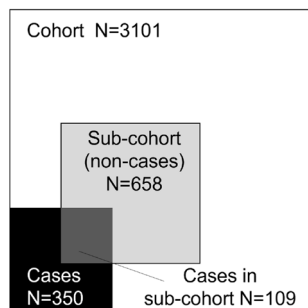synthesized with photo-cleavable biotin, a spacer and cyanine-3 fluorophore at the 5' end. The beads are washed to remove non-specifically associated proteins and other matrix constituents. Proteins that remain bound to SOMAmer reagents are labelled using an NHS-biotin reagent. SOMAmer complexes and unbound SOMAmer reagents are released from the SA beads using ultraviolet light that cleaves the photo-cleavable linker within the SOMAmer reagent construct into a solution containing an anionic competitor (biotin originally incorporated in the SOMAmer is left on the SA beads). Non-specific interactions, which have much faster dissociation rates compared to specific interaction, dissociate, and the polyanionic competitor solution prevents them from rebinding while specific complexes are maintained. The photo-cleavage eluate, which contains all SOMAmer reagents (some bound to a biotin-labelled protein and some free), is separated from the beads and then incubated with a second streptavidin coated bead that binds protein-SOMAmer complexes through biotin on the protein (catch 2). The inversion of polarity of attachment of protein-SOMAmer complexes on the second set of SA beads contributes a substantial degree of additional specificity enhancement by dramatically minimizing non-specific interactions due to multivalent interactions between proteins and SOMAmer-coated SA beads in the first (catch 1) incubation. The free SOMAmer reagents (which no longer contain biotin) are removed during subsequent washing steps. In the final elution step, protein-bound SOMAmer reagents are released from their proteins using denaturing conditions and recovered. These SOMAmer reagents are then quantified by hybridisation to custom DNA microarrays. The cyanine-3 signal from the SOMAmer reagent is detected on microarrays. The readout in relative fluorescent units (RFU) is directly proportional to the amount of target epitope in the initial sample. The resulting measurement and quantitation of ~7000 plasma proteins presented in a proprietary text-based format called ADAT. The *readat* R package which is free, open source, and available on Bioconductor and Bitbucket will be used for working with the SomaLogic ADAT files including importing, transforming and annotating data from these files[70].

Serum Metabolomics: 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, glycerophos-pholipids, sphingolipids and sugars[62,63]. 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 wells 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[62]. 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 $ZnSO_4$ 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[62,63].

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, 99.99%) 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% HNO_3, 5% H_2O_2, and MiliQ 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[64]. 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[64]. 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 Acquite 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, I.D. 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[65]. Briefly, 50 μL of plasma is diluted 1:3 in 100 μL 10 mM $MgCl_2$ 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 $MgCl_2$ 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[71].

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[66]. 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[72]. 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 lysis 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 quantitation of haemoglobins variants namely: A, S, F and A2 and other minor variants and diagnosis of ß-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[73] and detailed by Xu and colleagues[67]. 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[67] 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[68,69], and also against an LC-MS assay for measuring aflatoxin lysine[74]. 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 biomakers, 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[75]. 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)[76]. 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)[77]. 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[75]. 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[78] 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)[79]. 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/kwtrp) and may be requested via email to dgc@kemri-wellcome.org.

Acknowledgments

This paper is published with permission from the Director of KEMRI

References

1. 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**

2. Taylor AW, Blau DM, Bassat Q, *et al.*: **Initial findings from a novel population-based child mortality surveillance approach: a descriptive study.** *Lancet Glob Health.* 2020; **8**(7): e909–e919.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Campbell JD, Sow SO, Levine MM, *et al.*: **The causes of hospital admission and death among children in Bamako, Mali.** *J Trop Pediatr.* 2004; **50**(3): 158–163.
   PubMed Abstract | **Publisher Full Text**

4. Gathara D, Malla L, Ayieko P, *et al.*: **Variation in and risk factors for paediatric inpatient all-cause mortality in a low income setting: data from an emerging clinical information network.** *BMC Pediatr.* 2017; **17**(1): 99.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Nemetchek B, English L, Kissoon N, *et al.*: **Paediatric postdischarge mortality in developing countries: a systematic review.** *BMJ Open.* 2018; **8**(12): e023445.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. English M, Mwaniki P, Julius T, *et al.*: **Hospital Mortality – a neglected but rich source of information supporting the transition to higher quality health systems in low and middle income countries.** *BMC Med.* 2018; **16**(1): 32.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Chisti MJ, Graham SM, Duke T, *et al.*: **Post-Discharge Mortality in Children with Severe Malnutrition and Pneumonia in Bangladesh.** *PLoS One.* 2014; **9**(9): e107663.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Ngari MM, Fegan G, Mwangome MK, *et al.*: **Mortality after Inpatient Treatment for Severe Pneumonia in Children: a Cohort Study.** *Paediatr Perinat Epidemiol.* 2017; **31**(3): 233–242.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Chhibber AV, Hill PC, Jafali J, *et al.*: **Child Mortality after Discharge from a Health Facility following Suspected Pneumonia, Meningitis or Septicaemia in Rural Gambia: A Cohort Study.** *PLoS One.* 2015; **10**(9): e0137095.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Moïsi JC, Gatakaa H, Berkley JA, *et al.*: **Excess child mortality after discharge from hospital in Kilifi, Kenya: a retrospective cohort analysis.** *Bull World Health Organ.* 2011; **89**(10): 725–732, 732A.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Wiens MO, Pawluk S, Kissoon N, *et al.*: **Pediatric Post-Discharge Mortality in Resource Poor Countries: A Systematic Review.** *PLoS One.* 2013; **8**(6): e66698.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. The Childhood Acute Illness and Nutrition (CHAIN) Network: **Childhood mortality during and after acute illness in Africa and south Asia: a prospective cohort study.** *Lancet Glob Health.* 2022; **10**(5): e673–e684.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Kwambai TK, Dhabangi A, Idro R, *et al.*: **Malaria Chemoprevention in the Postdischarge Management of Severe Anemia.** *N Engl J Med.* 2020; **383**(23): 2242–2254.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Dondorp AM, Fanello CI, Hendriksen ICE, *et al.*: **Artesunate versus quinine in the treatment of severe falciparum malaria in African children (AQUAMAT): an open-label, randomised trial.** *Lancet.* 2010; **376**(9753): 1647–1657.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Reddy EA, Shaw AV, Crump JA: **Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis.** *Lancet Infect Dis.* 2010; **10**(6): 417–432.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Deen J, von Seidlein L, Andersen F, *et al.*: **Community-acquired bacterial bloodstream infections in developing countries in south and southeast Asia: a systematic review.** *Lancet Infect Dis.* 2012; **12**(6): 480–487.
    PubMed Abstract | **Publisher Full Text**

17. Levine MM, Nasrin D, Acácio S, *et al.*: **Diarrhoeal disease and subsequent risk of death in infants and children residing in low-income and middle-income countries: analysis of the GEMS case-control study and 12-month GEMS-1A follow-on study.** *Lancet Glob Health.* 2020; **8**(2): e204–e214.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. Pneumonia Etiology Research for Child Health (PERCH) Study Group: **Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study.** *Lancet.* 2019; **394**(10200): 757–779.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Njunge JM, Gwela A, Kibinge NK, *et al.*: **Biomarkers of post-discharge mortality among children with complicated severe acute malnutrition.** *Sci Rep.* 2019; **9**(1): 5981.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Bartz S, Mody A, Hornik C, *et al.*: **Severe acute malnutrition in childhood: hormonal and metabolic status at presentation, response to treatment, and predictors of mortality.** *J Clin Endocrinol Metab.* 2014; **99**(6): 2128–2137.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Attia S, Versloot CJ, Voskuijl W, *et al.*: **Mortality in children with complicated severe acute malnutrition is related to intestinal and systemic inflammation: an observational cohort study.** *Am J Clin Nutr.* 2016; **104**(5):

1441–1449.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Wen B, Njunge JM, Bourdon C, *et al.*: **Systemic inflammation and metabolic disturbances underlie inpatient mortality among ill children with severe malnutrition.** *Sci Adv.* 2022; **8**(7): eabj6779.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Wang J, Sun Y, Teng S, *et al.*: **Prediction of sepsis mortality using metabolite biomarkers in the blood: a meta-analysis of death-related pathways and prospective validation.** *BMC Med.* 2020; **18**(1): 83.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Van Wyngene L, Vandewalle J, Libert C: **Reprogramming of basic metabolic pathways in microbial sepsis: therapeutic targets at last?** *EMBO Mol Med.* 2018; **10**(8): e8712.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

25. Hall MW, Knatz NL, Vetterly C, *et al.*: **Immunoparalysis and nosocomial infection in children with multiple organ dysfunction syndrome.** *Intensive Care Med.* 2011; **37**(3): 525–532.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26. Gardner MM, Kirschen MP, Wong HR, *et al.*: **Biomarkers associated with mortality in pediatric patients with cardiac arrest and acute respiratory distress syndrome.** *Resuscitation.* 2022; **170**: 184–193.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Zinter MS, Orwoll BE, Spicer AC, *et al.*: **Incorporating Inflammation into Mortality Risk in Pediatric Acute Respiratory Distress Syndrome.** *Crit Care Med.* 2017; **45**(5): 858–866.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28. Zarjou A, Agarwal A: **Sepsis and Acute Kidney Injury.** *J Am Soc Nephrol.* 2011; **22**(6): 999–1006.
Publisher Full Text

29. Brealey D, Brand M, Hargreaves I, *et al.*: **Association between mitochondrial dysfunction and severity and outcome of septic shock.** *Lancet.* 2002; **360**(9328): 219–223.
PubMed Abstract | **Publisher Full Text**

30. Langley RJ, Tipper JL, Bruse S, *et al.*: **Integrative "omic" analysis of experimental bacteremia identifies a metabolic signature that distinguishes human sepsis from systemic inflammatory response syndromes.** *Am J Respir Crit Care Med.* 2014; **190**(4): 445–455.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

31. Langley RJ, Tsalik EL, van Velkinburgh JC, *et al.*: **An integrated clinico-metabolomic model improves prediction of death in sepsis.** *Sci Transl Med.* 2013; **5**(195): 195ra195.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

32. Oxenkrug G: **Insulin resistance and dysregulation of tryptophan-kynurenine and kynurenine-nicotinamide adenine dinucleotide metabolic pathways.** *Mol Neurobiol.* 2013; **48**(2): 294–301.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

33. Zhang L, Voskuijl W, Mouzaki M, *et al.*: **Impaired Bile Acid Homeostasis in Children with Severe Acute Malnutrition.** *PLoS One.* 2016; **11**(5): e0155143.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

34. van Zutphen T, Ciapaite J, Bloks VW, *et al.*: **Malnutrition-associated liver steatosis and ATP depletion is caused by peroxisomal and mitochondrial dysfunction.** *J Hepatol.* 2016; **65**(6): 1198–1208.
PubMed Abstract | **Publisher Full Text**

35. Singer M: **The role of mitochondrial dysfunction in sepsis-induced multi-organ failure.** *Virulence.* 2014; **5**(1): 66–72.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

36. Koutroulis I, Batabyal R, McNamara B, *et al.*: **Sepsis Immunometabolism: From Defining Sepsis to Understanding How Energy Production Affects Immune Response.** *Crit Care Explor.* 2019; **1**(11): e0061.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

37. Drosatos K, Khan RS, Trent CM, *et al.*: **Peroxisome Proliferator–Activated Receptor-γ Activation Prevents Sepsis-Related Cardiac Dysfunction and Mortality In Mice.** *Circ Heart Fail.* 2013; **6**(3): 550–562.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

38. Shapiro NI, Khankin EV, Van Meurs M, *et al.*: **Leptin exacerbates sepsis-mediated morbidity and mortality.** *J Immunol.* 2010; **185**(1): 517–524.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

39. Bailey RL, West KP Jr, Black RE: **The Epidemiology of Global Micronutrient Deficiencies.** *Ann Nutr Metab.* 2015; **66**(suppl 2): 22–33.
PubMed Abstract | **Publisher Full Text**

40. Manzanares W, Dhaliwal R, Jiang X, *et al.*: **Antioxidant micronutrients in the critically ill: a systematic review and meta-analysis.** *Crit Care.* 2012; **16**(2): R66.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

41. Depeint F, Bruce WR, Shangari N, *et al.*: **Mitochondrial function and toxicity: role of the B vitamin family on mitochondrial energy metabolism.** *Chem Biol Interact.* 2006; **163**(1–2): 94–112.
PubMed Abstract | **Publisher Full Text**

42. Rahmel T, Schmitz S, Nowak H, *et al.*: **Long-term mortality and outcome in hospital survivors of septic shock, sepsis, and severe infections: The importance of aftercare.** *PLoS One.* 2020; **15**(2): e0228952.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

43. Winters BD, Eberlein M, Leung J, *et al.*: **Long-term mortality and quality of life**

44. Fang WF, Douglas IS, Chen YM, *et al.*: **Development and validation of immune dysfunction score to predict 28-day mortality of sepsis patients.** *PLoS One.* 2017; **12**(10): e0187088.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

45. Attia S, Feenstra M, Swain N, *et al.*: **Starved Guts: Morphologic and Functional Intestinal Changes in Malnutrition.** *J Pediatr Gastroenterol Nutr.* 2017; **65**(5): 491–495.
PubMed Abstract | **Publisher Full Text**

46. Watanabe K, Petri WA Jr: **Environmental Enteropathy: Elusive but Significant Subclinical Abnormalities in Developing Countries.** *EBioMedicine.* 2016; **10**: 25–32.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

47. Crane RJ, Jones KDJ, Berkley JA: **Environmental Enteric Dysfunction: An Overview.** *Food Nutr Bull.* 2015; **36**(1 Suppl): S76–S87.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

48. Zhou X, Guo J, Geng B, *et al.*: **Gut-dependent microbial translocation induces inflammation and cardiovascular events after ST-elevation myocardial infarction.** *Microbiome.* 2018; **6**(1): 66.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

49. Tickell KD, Atlas HE, Walson JL: **Environmental enteric dysfunction: a review of potential mechanisms, consequences and management strategies.** *BMC Med.* 2019; **17**(1): 181.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

50. Ley RE, Peterson DA, Gordon JI: **Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine.** *Cell.* 2006; **124**(4): 837–848.
PubMed Abstract | **Publisher Full Text**

51. Hand BY, Timothy W: **Role of the Microbiota in Immunity and Inflammation.** *Cell.* 2014; **157**(1): 121–141.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

52. Robertson RC, Manges AR, Finlay BB, *et al.*: **The Human Microbiome and Child Growth - First 1000 Days and Beyond.** *Trends Microbiol.* 2019; **27**(2): 131–147.
PubMed Abstract | **Publisher Full Text**

53. Shimizu K, Ogura H, Hamasaki T, *et al.*: **Altered gut flora are associated with septic complications and death in critically ill patients with systemic inflammatory response syndrome.** *Dig Dis Sci.* 2011; **56**(4): 1171–1177.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

54. Rao K, Patel AR, Seekatz AM, *et al.*: **Gut microbiome features are associated with sepsis onset and outcomes.** *bioRxiv.* 2021; 2021.2001.2008.426011.
Publisher Full Text

55. Peled JU, Gomes ALC, Devlin SM, *et al.*: **Microbiota as Predictor of Mortality in Allogeneic Hematopoietic-Cell Transplantation.** *N Engl J Med.* 2020; **382**(9): 822–834.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

56. Bajaj JS, Vargas HE, Rajender Reddy K, *et al.*: **Association Between Intestinal Microbiota Collected at Hospital Admission and Outcomes of Patients With Cirrhosis.** *Clin Gastroenterol Hepatol.* 2019; **17**(4): 756–765.e753.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

57. Childhood Acute Illness and Nutrition Network: **Childhood Acute Illness and Nutrition (CHAIN) Network: a protocol for a multi-site prospective cohort study to identify modifiable risk factors for mortality among acutely ill children in Africa and Asia.** *BMJ Open.* 2019; **9**(5): e028454.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

58. Langholz B, Thomas DC: **Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison.** *Am J Epidemiol.* 1990; **131**(1): 169–176.
PubMed Abstract | **Publisher Full Text**

59. Kulathinal S, Karvanen J, Saarela O, *et al.*: **Case-cohort design in practice - experiences from the MORGAM Project.** *Epidemiol Perspect Innov.* 2007; **4**: 15.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

60. Wang HE, Szychowski JM, Griffin R, *et al.*: **Long-term mortality after community-acquired sepsis: a longitudinal population-based cohort study.** *BMJ Open.* 2014; **4**(1): e004283.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

61. Gold L, Ayers D, Bertino J, *et al.*: **Aptamer-based multiplexed proteomic technology for biomarker discovery.** *PLoS One.* 2010; **5**(12): e15004.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

62. Foroutan A, Guo AC, Vazquez-Fresno R, *et al.*: **Chemical Composition of Commercial Cow's Milk.** *J Agric Food Chem.* 2019; **67**(17): 4897–4914.
PubMed Abstract | **Publisher Full Text**

63. Foroutan A, Fitzsimmons C, Mandal R, *et al.*: **The bovine metabolome.** *Metabolites.* 2020; **10**(6): 233.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

64. Jenkins B, Ronis M, Koulman A: **LC-MS Lipidomics: Exploiting a Simple High-Throughput Method for the Comprehensive Extraction of Lipids in a Ruminant Fat Dose-Response Study.** *Metabolites.* 2020; **10**(7): 296.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

65. Kang M, Edmundson P, Araujo-Perez F, *et al.*: **Association of plasma endotoxin, inflammatory cytokines and risk of colorectal adenomas.** *BMC Cancer.* 2013; **13**: 91.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

66. Liu J, Gratz J, Amour C, *et al.*: **A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens.** *J Clin Microbiol.* 2013; **51**(2): 472–480.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

67. Saltzmann J, Xu Y, Gong YY, *et al.*: **Preliminary study on the relationship between aflatoxin-bovine serum albumin adducts in blood and aflatoxin M1 levels in milk of dairy cows.** *Mycotoxin Res.* 2020; **36**(2): 207–211.
     PubMed Abstract | **Publisher Full Text**

68. Wild CP, Hudson GJ, Sabbioni G, *et al.*: **Dietary intake of aflatoxins and the level of albumin-bound aflatoxin in peripheral blood in The Gambia, West Africa.** *Cancer Epidemiol Biomarkers Prev.* 1992; **1**(3): 229–234.
     PubMed Abstract

69. Routledge MN, Kimanya ME, Shirima CP, *et al.*: **Quantitative correlation of aflatoxin biomarker with dietary intake of aflatoxin in Tanzanian children.** *Biomarkers.* 2014; **19**(5): 430–435.
     PubMed Abstract | **Publisher Full Text**

70. Cotton RJ, Graumann J: **readat: An R package for reading and working with SomaLogic ADAT files.** *BMC Bioinformatics.* 2016; **17**(1): 201.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

71. Forster SC, Kumar N, Anonye BO, *et al.*: **A human gut bacterial genome and culture collection for improved metagenomic analyses.** *Nat Biotechnol.* 2019; **37**(2): 186–192.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

72. Liu J, Gratz J, Amour C, *et al.*: **Optimization of Quantitative PCR Methods for Enteropathogen Detection.** *PLoS One.* 2019; **11**(6): e0158199.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

73. Wild CP, Chapot B: **ELISA for quantification of Aflatoxin-albumin adducts and their application to human exposure assessment.** *Diagnostic Pathol.* 1991; **2**: 135–155.
     Reference Source

74. McCoy LF, Scholl PF, Sutcliffe AE, *et al.*: **Human aflatoxin albumin adducts quantitatively compared by ELISA, HPLC with fluorescence detection, and HPLC with isotope dilution mass spectrometry.** *Cancer Epidemiol Biomarkers Prev.* 2008; **17**(7): 1653–1657.
     PubMed Abstract | **Publisher Full Text**

75. Ghaemi MS, DiGiulio DB, Contrepois K, *et al.*: **Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy.** *Bioinformatics.* 2019; **35**(1): 95–103.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

76. Chen T, Guestrin C: **XGBoost: A Scalable Tree Boosting System.** In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016; 785–794.
     Publisher Full Text

77. Davis J, Goadrich M: **The relationship between Precision-Recall and ROC curves.** In *Proceedings of the 23rd international conference on Machine learning.* 2006; 233–240.
     Publisher Full Text

78. Ashburner M, Ball CA, Blake JA, *et al.*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet.* 2000; **25**(1): 25–29.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

79. Alexa A, Rahnenfuhrer J: **topGO: Enrichment Analysis for Gene Ontology.** *R package version 2.24.0.* 2016.

Open Peer Review

Current Peer Review Status: ✓ ✓

Version 2

Reviewer Report 09 November 2022

https://doi.org/10.21956/gatesopenres.15370.r32750

✓ **Quique Bassat** [iD]

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

I am satisfied with the responses to my comments, and happy for the paper to be published.

Competing Interests: No competing interests were disclosed.

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.

Version 1

Reviewer Report 29 July 2022

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

? **Quique Bassat** [iD]

Barcelona Institute for Global Health (ISGlobal), Hospital Clínic 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 etioogy), 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 Clínic 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.
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[2]

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[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.

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

1. CHAIN. Childhood Acute Illness and Nutrition (CHAIN) Network: a protocol for a multi-site prospective cohort study to identify modifiable risk factors for mortality among acutely ill children in Africa and Asia. *BMJ open* **9**, e028454-e028454, doi:10.1136/bmjopen-2018-028454 (2019).

2. CHAIN. Childhood mortality during and after acute illness in Africa and south Asia: a prospective cohort study. *The Lancet Global Health* **10**, e673-e684, doi:10.1016/S2214-109X(22)00118-8 (2022).

3. Taylor, A. W. *et al.* Initial findings from a novel population-based child mortality surveillance approach: a descriptive study. *The Lancet Global Health* **8**, e909-e919, doi:10.1016/S2214-109X(20)30205-9 (2020).

4. Wen, B. *et al.* Systemic inflammation and metabolic disturbances underlie inpatient mortality among ill children with severe malnutrition. *Science Advances* **8**, eabj6779, doi:doi:10.1126/sciadv.abj6779 (2022).

5. Njunge, J. M. *et al.* Biomarkers of post-discharge mortality among children with complicated severe acute malnutrition. *Scientific Reports* **9**, 5981, doi:10.1038/s41598-019-42436-y (2019).

6. Moisi, J. C. *et al.* Excess child mortality after discharge from hospital in Kilifi, Kenya: a retrospective cohort analysis. *Bulletin of the World Health Organization* **89**, 725-732, 732A, doi:10.2471/BLT.11.089235 (2011).

7. Wang, H. E. *et al.* Long-term mortality after community-acquired sepsis: a longitudinal population-based cohort study. *BMJ Open* **4**, e004283, doi:10.1136/bmjopen-2013-004283 (2014).

8. Winters, B. D. *et al.* Long-term mortality and quality of life in sepsis: a systematic review. *Crit Care Med* **38**, 1276-1283, doi:10.1097/CCM.0b013e3181d8cc1d (2010).

Competing Interests: None

Reviewer Report 28 July 2022

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

✔️ **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