Brief Title: Improving diagnosis of sepsis using big-data

Global health systems data-science approach for precision diagnosis of sepsis in early life

Authors: Kenneth Iregbu MD[1], Angela Dramowski MBChB, PhD[2], Rebecca Milton MPH[3], Emmanuel Nsutebu MD[4], Stephen R. C. Howie MBChB, PhD[5], Mallinath Chakraborty MBChB, PhD[6], Pascal M Lavoie MD PhD[7], Ceire E Costelloe PhD[8] and Peter Ghazal PhD[9]

Author affiliations:
[1]Department of Medical Microbiology, National Hospital Abuja, Nigeria. [2] Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. [3] Centre for Trials Research, Cardiff University, Cardiff, CF14 4YS. [4] Infectious Disease Consultant and Chair of Infectious Diseases Division, Sheikh Shakhbout Medical City, Abu Dhabi, UAE. [5] Department of Paediatrics: Child and Youth Health, University of Auckland, Auckland, New Zealand. [6] Regional Neonatal Intensive Care Unit, University Hospital of Wales, Cardiff CF14 4XN. [7] Department of Pediatrics, University of British Columbia, Vancouver, Canada. [8] Global Digital health unit, School of Public health, Imperial College London, [9] Systems Immunity Research Institute, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN.

Corresponding author:
Peter Ghazal: FMedSci
Systems Immunity Research Institute,
School of Medicine,
Cardiff University,
Heath Park,
Cardiff,
CF14 4XN.
Email: GhazalP@cardiff.ac.uk

Abstract
Neonates and children in low-and-middle-income countries (LMICs) globally contribute to the highest number of sepsis-associated deaths. Interventions to prevent sepsis mortality are hampered by a lack of comprehensive epidemiological data and pathophysiological understanding of biological-pathways. Here, we discuss the challenges facing LMICs in diagnosing sepsis in this age group. We highlight a role for multi-omics and big eHealth-data to improve diagnostic accuracy of clinical algorithms, arguing that healthcare systems urgently need precision medicine to avoid pitfalls of missed-, mis- and over-diagnoses and attendant antimicrobial resistance. We discuss ethical, regulatory and systemic barriers related to the collection and use of bigdata in LMICs. Technologies such as cloud computing, artificial-intelligence and medical tricorders may help but require collaboration with local communities. Co-partnering could facilitate integration of these technologies as part of a future care-delivery system, offering a chance to transform globally the management and prevention of sepsis for neonates and children.

Box 1: Search strategy and selection criteria

We searched PubMed and African Journal Online to identify studies (original research, commentary, reviews) pertaining to neonatal and paediatric sepsis (including infants under one year of age for the latter), as well as clinical studies presenting data on mortality under 5 years of age, published over the past 15 years. Keywords used for searches included: neonatal sepsis, sepsis in children, risk factors for neonatal sepsis, risk factors for sepsis in children, epidemiology of sepsis in children, diagnosis of sepsis in children, biomarkers in neonatal sepsis. We also searched PubMed for AI, ML, EHR, cloud computing and medical tricorder technologies. Several national and regional studies were identified worldwide. We also reviewed PubMed and African Journal Online which looked at sepsis prevalence, aetiology, risk factors, diagnostics including biomarker (transcriptomic, proteomic, metabolomic, etc.) in paediatric or adult populations that included infants under 1 year of age. There were only few biomarker studies in developed countries, none specifically done in LMIC. Studies that did not separate data from infants or neonates were excluded as were cases with very low numbers of samples and where blood culture was not done were not considered.

Introduction
In 2017 an estimated 20.3 million children under-five years old developed sepsis, resulting in 2.9 million global deaths, the majority (85%) occurred in low- and middle-income countries (LMICs).[1-2] In sub-Saharan African countries almost half of neonatal deaths are accounted for by sepsis.[3,4] Global disparities in access to healthcare are well documented, especially for young children, pregnant women and the elderly.[1,5,6] Despite the major public health impact of sepsis in neonates and young children, there are limited data on the aetiology, and the biological causal pathways accounting for deaths in LMICs, significantly hampering development of effective interventions.[7]

Most LMICs follow the World Health Organisation (WHO) guidelines on Integrated Management of Childhood Illness when assessing a child with suspected sepsis.[8] This syndromic approach to managing these cases without laboratory tests has drawbacks, and hinders tailored treatment options based on aetiology. Sepsis due to bacterial infection requires urgent antimicrobial treatment to maximise survival[19–23]. Yet, overusing first-line antibiotics has led to the rapid expansion of antimicrobial resistance (AMR), particularly in LMICs.[9] The lack of access to effective therapy for AMR pathogens constitutes a global threat to effective sepsis management.

The last decade has seen an explosion of data-intensive health research underpinned by improved computational capabilities and molecular technologies to rapidly acquire and process complex biological and clinical data. Genomics, hastened by the Human Genome Project, and further catalysed by advances in high-throughput, technologies and bioinformatics methods, has enabled the acquisition of big multi-'omic' data, these include primarily transcriptomics, proteomics and metabolomics. Digital technology, engineering and wireless connectivity of mobile devices are now enabling the collection of these data on a large-scale, increasingly including remote areas.[10-11] These technological developments open the way for precision medicine, offering opportunities to improve health outcomes where the greatest burden of disease and mortality occurs and reduce the global health gap. The WHO highlighted the potential to achieve improved population health through precision in healthcare delivery and the use of data through collaboration.[12] The WHO SCORE (Survey, Count, Optimise, Review, Enable) global report urges countries to strengthen health-data systems, providing technical packages and support to improve data registries and collect better quality data.[13] This aims to provide an evidence-base for identifying low-cost context-specific recommendations, while opening possibilities for more advanced technological data-driven innovation. These near-term steps to strengthen acute healthcare delivery systems and the promotion of context-specific management guidelines have been comprehensively discussed.[14]

Practical problems remain in applying these data-driven approaches in LMICs, including ethical, regulatory and systems barriers related to the collection and use of sepsis data (Figure1). In comparison to other sectors, the health sector in LMICs has yet to harness the potential of large-volume digital clinical data and multi-'omic' systems-level patient data, to achieve a step-change in precision for sepsis diagnostics and therapeutics.[15] The debate on big healthcare data in LMICs has focussed on highlighting governance frameworks to protect individuals and to guide healthcare delivery systems to meet the target community needs.[16] The role advanced technology may play in improving health is gaining attention, yet the end user need in LMICs must be clearly identified and a collaborative approach taken.[17] Context-specific user requirements in LMICs must be well-understood through iterative user feedback from healthcare workers. While barriers to implementation should not be underestimated, there is an opportunity for LMICs to become early adopters, co-developing the next wave of diagnostic technologies.

We focus on the potential to save young lives from the impact of sepsis, highlighting clinical issues and limitations, ethical implications, and societal concerns related to current and future use of big-data (multi-omics and heath records) for enabling precision medicine in LMICs. We propose key technological stepping-stones LMICs should consider for harnessing, at earlier rather than later stages, the emerging point-of-care and medical tricorder technology, cloud computing and artificial intelligence (AI).

End-user need in LMICs: Sepsis aetiology and antimicrobial resistance

In young children, sepsis is often clinically defined as a systemic inflammatory response in the context of confirmed bacteraemia (definite sepsis) and/or abnormalities in relevant laboratory markers (suspected sepsis). Evidence of the sepsis-related Systemic Inflammatory Response Syndrome (SIRS) generally includes any two of: rectal temperature >38·5°C or <36°C; heart rate more than two standard deviations (SD) above the normal, or bradycardia in children older than one year of age (<10th percentile for age); respiratory rate more than two SD above normal (or pCO2 <32 mmHg); leukocyte count >12,000 cells/mm3, <4,000 cells/mm3, or >10% band forms.[24] This definition lacks sensitivity and specificity, failing to account for infection organism, sites of infection or early diagnosis. For neonates signs can be more subtle with closer overlap between infectious and non-infectious syndromes, and often only recognised by the mother as being "not right". Moreover, the definition is impractical in most resource-limited settings where monitoring technologies and diagnostic laboratories for routine blood culture or viral molecular testing and blood counts are mostly unavailable. Predominantly, the diagnosis (and empirical treatment) often rests on the classification of clinical signs by community-based healthcare workers.[25,26] Without early recognition and appropriate treatment sepsis can rapidly deteriorate into septic shock with its attendant high mortality. Hence the urgent need to develop screening tools and methods more suited to pre-hospital assessment and rapid early detection of infection.

Although aetiologic agents triggering sepsis include bacteria, viruses, fungi and parasites, the focus remains on bacterial infections due to effectiveness of antibiotic treatment, yet, access to the "right' antibiotic in many LMICs is problematical, increasing the risk of AMR.[22,27] A potential unintended consequence of early intervention, in the absence of a clear diagnosis is inappropriate antibiotic use. Resistance has emerged against commonly available and affordable antibiotics in many LMICs, especially in hospitals posing challenges to treatment, and where the use of carbapenems and the cephalosporins/beta-lactamase combinations are rarely available and/or unaffordable. A global meta-analysis in the paediatric population showed that resistance to ampicillin, nitrofurantoin, co-amoxiclav and ciprofloxacin antibiotics was higher in non-Organisation for Economic Co-operation and Development (OECD) than OECD countries, with resistance rates higher to first-line than to non-first-line antibiotics.[28,29] Evidence shows previous antibiotic exposure can more than double the risk of AMR for respiratory and urinary tract infections.[28] Highlighting the precarity between antibiotic use and their effectiveness, and the urgent need for improved diagnostics with improved specificity and negative predictive value, to inform a more judicious use of antibiotics.

In LMICs almost two-thirds of bacteraemia in children is caused by gram-negative bacteria (GNB), most notably *Klebsiella spp*, and *E coli*. Evidence suggests that causative organisms for neonatal sepsis are balanced between Gram-positive (mainly *S. aureus)* and GNB in community healthcare settings in LMICs, however, GNB dominate neonatal sepsis in home-born infants and hospital settings.[4,30] Considering increasing AMR to GNB, data to model optimum choices of antibiotics and identify emerging trends of resistance would transform management and outcomes in LMICs.[31] In very low birth weight premature neonates, compromised barrier integrity poses a further risk of sepsis due to commensals such as *coagulase-negative Staphylococcus aureu*s; and while treated as pathogens in neonates in high-income countries (HICs) are considered contaminants in LMICs, where few very premature neonates survive.[30]

Limitations of standard blood culture tests in LMICs

The Surviving Sepsis Campaign guidelines recommend obtaining blood cultures before antibiotic treatment in neonates and children with suspected sepsis, followed by empiric antibiotic administration within one hour of presentation.[32] Cultures of blood, urine, and cerebrospinal fluid, as well as other body specimen based on the patient's presentation, are critical for diagnosis and identifying the focus of infection. Blood cultures are the cornerstone of definitive aetiologic diagnosis and targeted antibiotic treatment in sepsis, but have very low pathogen positivity (5-10%) and are frequently contaminated by skin commensals owing to poor aseptic technique during culture collection.[33,34] Additionally, turnaround time for cultures exceeds time available before a decision needs to be made about initiating antibiotics, contributing to overuse of antibiotics in situations where bacterial sepsis is only one possibility. The lack of access to, and unaffordability of, laboratory tests in LMICs obscures our understanding of the epidemiology and aetiology of sepsis in these areas of the world.[7,35]

The fear of missing a true bacterial sepsis case leads clinicians in LMICs to give prolonged courses of broad-spectrum antibiotics in neonates and children. Automated blood culture systems that support the growth of a wider range of organisms and at lower inoculum levels than manual systems are not widely available in many LMICs. In addition, the presence of antibiotic removal devices such as resins which help to enhance microbial growth in the presence of antibiotics, as well as the continuous agitation of bottles by the system to encourage bacterial growth are scarce.[36-37] Newer diagnostic platforms present significant barriers for LMICs in terms of high cost, being bulky and requiring technical expertise and laboratory facilities (Figure 1). Even where microbiology laboratories are available many lack necessary resources and almost none provide updated systems (institutional or national) level antibiograms, and recommendations based on actual antibiotic sensitivity data to inform appropriate empiric antibiotics selection (Figure 1). Molecular methods offer advantages over blood cultures, including increased sensitivity and rapid diagnosis.[38,39,40] However, testing laboratories are not common outside research and there are few virology laboratories in LMICs, so the diagnosis of viral sepsis cases remains an underappreciated cause of sepsis further driving overuse of antibiotics.

Limitations of current host biomarker tests

Host biomarkers provides an important part of the clinical picture. These include Procalcitonin (PCT), available as a point-of-care assay[41] that has clinical usefulness in determining the termination point of antibiotic therapy rather than in establishing a diagnosis[41,45]. PCT, like most other single inflammatory biomarkers, is not specific for infection as it can be elevated in other disorders, especially following trauma,[42-43] and although its sensitivity in adults is 70-77%, it is only 49% shortly after birth. Its values are also influenced by day-of-life and gestational age amongst others.[44] CRP is an acute phase reactant that is produced in the liver in response to inflammation and is a useful low-cost biomarker. Serial measurements of CRP levels are required, as one-off tests lack sensitivity, therefore has limited usability for rapid assessment. Also, CRP values for neonates are influenced by events around birth such as premature rupture of membrane, maternal fever, meconium aspiration, fetal distress, gestational age..[40-46] Tests require skilled human resource and when available are often too costly for most families in LMICs, who mostly self-fund healthcare. The implementation of rapid, near patient systems for CRP/PCT that are cheap, easy to use and generate accurate results are becoming available in LMICs and may improve antimicrobial stewardship, but still lack the high specificity and negative predictive value needed to be confident about withholding antimicrobial treatment.

Research has continued to try and find the 'right' biomarker; leading to the investigation of approximately 200 different inflammatory biomarkers for sepsis, but none have yet to be used in diagnosis.[47-48] A key challenge is the ability to rapidly test blood samples directly without prior culture, to differentiate culture-negative sepsis from non-infectious SIRS, and to be universally suitable for point-of-care use. With the challenges and time required for microbial identification of pathogens, host biomarkers still have a strong potential role to play. A biomarker with high sensitivity, specificity, speed and accuracy would be revolutionary, as every hour of delay of treatment may increase the mortality of septic shock.[49] The use of a single biomarker for identification of sepsis, and especially inflammatory based biomarkers, is fraught with specificity and sensitivity problems. A single biomarker measured at the time of suspicion of sepsis ignores that the host responses are complex multi-system biological processes or pathways that evolve, with each pathway having specific roles temporally.

Future host molecular multi-omic biomarkers for diagnosis

For over 20 years profiling using omic-approaches to identify signatures unique for each pathogen has been discussed as a novel tool for diagnosis, prognosis, and clinical management of infectious diseases,[50-51] but remains to be embedded in clinical practice. Genomic signatures for infection type and sepsis have provided the necessary proof-of-principle studies in HICs.[52–53]These now need to be followed by large-scale multi-cohort validation studies, including in LMICs (Figure 2).

An overriding advantage of genomic and/or multi-omic signatures is the ability to measure multiple different pathways. Current standard diagnostic technology in HICs typically accommodates no more than a few biomarkers, and has mostly been optimised for single analytes.[54] Accordingly, small subsets of genomic markers that can be used with clinical algorithms to improve diagnostic accuracy have been sought. An example is work in critically ill septic paediatric patients where genome-wide expression profiling identified 12 candidate prognostic biomarkers; five of these marker genes were selected, based on predictive value; interleukin-8 (IL-8), C-C chemokine ligand 3 (CCL3), heat shock protein 70 kDa 1B (HSPA1B), granzyme B (GZMB), and matrix metalloproteinase 8 (MMP-8).[55] To increase accuracy a risk model incorporating other clinical parameters, such as platelet counts, was necessary for broad efficacy to estimate a specific patient's risk of mortality from sepsis.[56] This study provides evidence that linking genomic and clinical data improves accuracy as well as highlighting the need and challenges for using multi-parameter digital diagnostics. Global large-scale prospective studies are needed to demonstrate that these biomarkers can be implemented and can reduce unnecessary antibiotic use in healthcare delivery (Figure 2).

Genomics studies are largely focused in HICs and the question remains as to whether it is possible to develop genomic signatures with high accuracy, especially in LMICs.[54,57] To address this question and identify the optimal number of genomic biomarkers needed to accurately classify sepsis, a virtual clinical trial accounting for high variability was conducted using high-performance supercomputing simulation.[58] The virtual trial involved generating a digital population capturing data characteristics of neonatal whole blood (Figure 2), and revealed genomic signature are capable of accurately identifying sepsis. Modelling this digital population predicted 24 biomarkers as an optimal number for predicting neonatal sepsis.[58] The performance improved when considering pathways rather than random assortment of genes. Importantly, these digital populations can be adapted to data characteristics captured in LMICs using early-life multi-omic data that is beginning to emerge in LMICs.[54,59] In a subsequent prospective case-control investigation of neonatal sepsis 19 genomic biomarkers, comprising three pathways, were found to provide 0% misclassification.[60] A systems biology and AI analysis of the systemic host response in neonatal sepsis identified three principal pathogenic pathways, involving networks associated with elevated innate, suppressed adaptive and critically altered metabolic pathways that when combined generated a classifier showing high accuracy (>99%) of predicting bacterial sepsis in an independent population.[60-61] These studies need to be extended to LMICs where baseline levels and epidemiology of sepsis may differ from HICs (Figure 2).

This neonatal classifier for bacterial sepsis does not score positive for viral sepsis cases suggesting high specificity for discerning a bacterial sepsis aetiology. Further research is required to identify a pathway classifier for viral sepsis. A limitation of such studies is whether these makers associate with severity. Many published genomic signature studies of infection or sepsis are no more than proof-of-concept studies and require large-scale independent validation (Figure 2). Demonstrated in the meta-analysis of multiple tuberculosis studies where a small subset of genomic signature-biomarkers are shown to be reproducible between studies.[62]

Neonatal sepsis findings highlight the drawback of using exclusively inflammatory biomarkers and show the necessity for employing other pathophysiological pathways, most notably metabolism. In this connection, cellular bioenergetics drive innate immune responses and at the same time are closely linked to the maturation stage of myeloid cells during development.[63] While metabolic markers, such as lactate levels, have been empirically adopted for sepsis diagnosis, the neonatal studies provided the first evidence for the pathophysiological significance of integrating metabolic biomarkers in identifying sepsis at the time presentation.

Future digital diagnostics

AI and Machine Learning (ML) have vast potential to improve healthcare research[64] and in LMICs may help eradicate health inequalities and decrease the burden on healthcare systems (Figure 2). Widely available mobile phone-based applications that can monitor vital signs of patients or assist in the detection of early warning signs could greatly assist front-line workers who assess sick patients in remote areas with limited resources, this approach may help reduce the burden on clinicians and healthcare systems in LMICs. For instance, data on clinical signs described by mothers to community healthcare workers, reported on smartphone-based applications and used in dynamic classification algorithms, may aid in alerting awareness for prompt diagnosis. It is important to note that community-level involvement and co-development of such tools with healthcare workers is necessary for improvements in the management of neonatal sepsis and acute healthcare delivery systems in general is to be realised.[14]

AI and deep learning methods have been deployed for the diagnosis and treatment of neonatal sepsis in India; using a cloud-based data analytics platform.[65] Using ML and data analytics to power the platform allows standardisation of diagnosis across different regions of the country in a fast and cost-effective way (Figure 2).

Through inputting neonatal data points, the platform generates a predictive score to assist with diagnosis of sepsis. Deep learning is agnostic to a particular type of data, opening possibilities to "phenotype" sepsis cases beyond what can be extracted from Electronic Health Records (EHR), for example, using video-images. Generally much larger amounts of data are necessary to exploit the power of deep learning. These are currently limited to single institution studies and to-date a systematic EHR-based screening tool for children in HICs and LMICs has yet to be undertaken.[66]

Other issues that will need to be addressed for a successful health systems response to sepsis include: the need to better define at-risk populations; establishing the natural history of infection, including deterioration pathways, and mortality rate; identifying and characterising the causative organism; and, where appropriate modelling to suggest effective prevention and control measures.[67] Key health related information can be collected from national surveillance programs, individual healthcare systems electronic or paper-based health records, and existing literature. Collectively data can be used to train and prime an AI-application for its dedicated task (Figure 2). The key to this step is accessing and harmonising data, which may be highly granular, or sparse depending on the environment. A significant challenge to accelerating the use of AI-enabled tools in LMICs relates to the quality and quantity of available data (Figure 1). This will require standardisation, concerted efforts and addressing ethical and logistical barriers limiting data sharing across studies. While big-data is critical for underpinning the adoption of any AI driven approach, there is no current definition of acceptable performance standards, accuracy rates, and patient health outcomes against which to measure AI. Policy organisations such as The National Institute for Health and Care Excellence (NICE) have developed evidence standard frameworks which could be implemented in other countries, but perspectives vary widely on what standards AI tools should adhere to, to be used across LMICs, and for patients and doctors to have trust in them.[68] Any AI driven clinical decision support tools are reliant on high quality data and must be able to demonstrate patient benefit. Critically, this needs to be tailored to the context specific needs of LMICs.

Inclusive future research: Emerging medical tricorder technology for combining host biomarkers and digital diagnostics

Mobile smart health technology using digital diagnostics are actively being investigated as part of an increasing number of clinical trials in LMICs.[69-70] Additionally, the development of point-of-care diagnostics not requiring technical expertise or laboratory facilities further provides vital tools for pre-hospital screening of sepsis in the community.[71-72] Largely, this technology is being developed in HICs and full co-partnering with LMICs in terms of co-development and capacity building for such technology has not yet been fully implemented.

The direction of travel for diagnostic technologies is evolving toward point-of-care sensors capable of measuring multi-parameter molecular or physiological signs (Figure 3). Namely, mobile high acuity devices for capturing vital physiological signs are now seeing increasing use in both clinical and community settings.[73] Developing sensor technology for physiological recordings and molecular analyte detection are currently researched by separate fields of investigation due to primarily non-overlapping scientific disciplines. In this connection, neither host molecular-biomarkers nor digital diagnostics will achieve a truly 100% accuracy alone for identifying sepsis, however, if combined they have a far greater chance of achieving this goal.

We contend that early recognition and prediction of sepsis must involve integrating physiological signs and validated molecular biomarkers into a single read-out. An emerging health technology that meets this need for physically integrating sensors is the rapidly growing medical tricorder market.[74-75] Figure 3 shows how a medical tricorder device would work in terms of data-recording, integration and read-out and the necessary logic operations to predictively score sepsis. Currently, these are handheld devices that are limited in design to record physiological data from sensors, analyse and display results with digital alerts.[73] Future development should be inclusive of capturing LMIC user requirements, incorporating point-of-care testing of host-biomarkers, imaging and physiological parameters to enable precision diagnosis. While the emerging medical tricorder technology is still very much in its infancy and predominantly research based, it is perceived as potentially disruptive technology for the HICs diagnostics market but may provide an excellent opportunity for LMICs, particularly when appropriately linked to clinical management algorithms. Optimistically, the next decade is poised to see a step-change in the diagnosis of infectious disease and sepsis.

Emerging big-data in LMICs

Healthcare data is changing with many countries adopting EHR systems, enabling healthcare organisations to accumulate and reuse data to improve patient outcomes and support research alongside other benefits.[15,76] In 2015, the WHO reported a 46% global increase in the use of EHRs over five years. Despite this vast improvement only 51% of high-income, 65% of upper-middle, 35% lower-middle and 15% low-income countries have national EHR systems.[77] Reliable data for the incidence of neonatal sepsis from LMICs is patchy, which significantly increases

the challenges in managing the condition).[30] Hospital-based studies are limited by significant selection-bias and estimates from community-based studies suggest a wide range of incidence possibly due to inaccurate data.[30] Workflows, clinical protocols, terminology and EHRs vary greatly across and within LMICs, therefore, tools need to be developed to allow appropriate integration. The impactful yet small-scale neonatal care management tool used in Malawi and Zimbabwe (NeoTREE tool) (http://www.neotree.org/) is an example.

Frequently cited barriers to the implementation of EHRs include lack of funding, infrastructure, expertise and legal framework[78,79] (Figure 1). To work towards global health equality requires that LMICs to have investment, expertise, support and resources to remain ahead of the technological advancements, such as harnessing satellite technology and cloud data storage.[80] [83]Simple smartphone-based applications, designed for LMICs, which can run on low internet bandwidth, could improve data collection and dissemination of guidance for the management of sepsis. In time these data could feed into analytical tricorder systems, including ML models supported by cloud-based platforms. Big-data is also rapidly growing from post-genomics science.[81] While LMIC populations have participated in internationally led observational studies, future resourcing and capacity building for translational genomics is vital.[82]

Resourcing education in data-science is also essential for an effective strategy to close the gap, requiring support from both public and private organisations (e.g., https://www.goldenhelix.org). Keeping open access to these data not only accelerates advancement in identifying reliable biomarkers and understanding of disease but can also promote the translational and basic science base in low-resource settings.[84] With technology capable of rapidly collecting vast quantities of data via EHRs and genomics, the application of systems-science tools, such as AI for improving precision becomes essential. These systems data-science tools can enable better identification of health burdens, improve diagnoses and healthcare delivery at population and individual levels; the potential impact of these tools will be negated if local health systems lack the capacity and resources to implement population-level interventions or effectively treat the at-risk patient.

Ethical and Societal Barriers and Issues in LMICs

Limitations of using big-data in LMICs for sepsis diagnostics in early life include lack of population-level data.[11,85] Most LMICs lack high-quality civil registry systems with only 65% of births and 38% of all deaths registered worldwide, thus demonstrating a paucity of whole population representative data (Figure 4A).[86] The widespread availability and use of mobile phone technology to record births and deaths within communities, is an opportunity to partially address this problem. In addition to demographic data, clinical data such as vital signs and laboratory data may be measured and recorded to a variable degree due to a lack of training and equipment, staff shortages, limited laboratory resources and affordability (Figure 4B). Further, scaling up of big-data and health-related AI-enabled tools is complicated by differing and uncertain regulatory and policy environments, both intra- and inter-country.[65] For example, many LMIC governments lack the resources to create consistent policies on population health, such as disease burden analysis and monitoring and treatment protocols for use, across their various regions or states. Currently, the WHO's SCORE Global Report provides technical packages and support to improve death data registries and is an important step toward lowering the barrier for systems data-science tools for population health to scale at national level.

Data sharing enables independent replication of findings and maximises the translational use of data.[87]. In LMICs barriers to data-sharing are multi-faceted including cultural, linguistic, ethical, financial and technical.[88] Studies by foreign consortia using LMIC populations to collect data for their use without prioritizing local impact to change health outcomes or build capacity greatly undermines trust and future collaborations. In this regard, many LMICs already have regulations prohibiting private companies from taking health and other types of data outside their borders. Using big-data to answer research questions also requires research systems with ethical approval processes, expertise in research methodology and design, areas which are often under-developed in LMICs.[89] High-quality data is important for making valid, reliable, and safe decisions based on scientific evidence. Indeed, the value of the data is in its quality.[90,20]

Concluding Remarks

Neonates and children in LMICs bear an inequitable mortality and health burden, and experience the greatest gap between diagnostic need and provision. Mortality in this age group is mostly attributed to sepsis, but the rapidly growing threat from AMR requires that we tailor interventions more precisely, and urgently. The lack of aetiological and pathophysiological understanding of sepsis represents one of the greatest barriers to an effective deployment of large-scale preventative and intervention programmes. Thus, accurate diagnostics are essential to correctly inform the epidemiology of infectious diseases and they are also needed to ensure that the correct interventions are administered to the appropriate neonates and children, improving outcomes and avoiding AMR.

Systems-science investigations of multi-'omic' data are beginning to offer insight in host pathophysiological processes. These should extend and be inclusive of LMIC populations. This should include generating simulation models of digital neonates and children for predicting classification rules for biomarkers relevant to LMICs. Further, large-scale cross-country population cohort studies will be required to confirm the utility of these biomarkers to identify the aetiology of sepsis and improve the precision of healthcare delivery. EHRs and the solidification of mobile networks in LMICs can enable AI and digital diagnostics to reduce the burden on frontline healthcare workers and inform evidence-based health policies (Figure 3). It will be essential to support health technology education and training in data-science in LMICs to ensure that point-of-care innovations can be co-developed in LMICs, be affordable and adapted to local context, and effectively deployed.

We propose that big-data-approaches and multi-'omic'-derived biomarkers that converge on mobile health technologies, in particular in the emerging medical tricorder market, offer a strategy to transform the diagnostic ecosystem in LMICs. We cannot ignore that the building of capacity and resilience in already fragile LMIC health systems are needed for these technologies to fulfil their promise. Altogether, this will assure the future direction of travel for infectious disease diagnostics in LMICs, and it will be important to apply a technology foresight framework in considering what healthcare will or should look like in the long-term (2050).

Contributions
KI, AD, RM, EN, SRCH, MC, PML, CEC and PG contributed to the writing of this manuscript, data interpretation, data sourcing, literature search, and contributed to figures.
PG designed the structure of the manuscript and coordinated the work.
RM and PG collated comments, edited the manuscript and produced figures.
All authors have reviewed and commented on the final version of the manuscript.

Declaration of interests
We declare no competing interests.

Funding Acknowledgements
RM is supported by the Centre for Trials Research is funded by Health & Care Research Wales and Cancer Research UK. CEC is supported by a personal NIHR Career Development Fellowship (grant number NIHR-2016-090-015) and the Imperial NIHR BRC. PG and RM by The Global Challenges Research Fund (Grant number: JA1121RA58). PG by MRC (G0701289) and the European Regional Development Fund and Welsh Government (Ser Cymru programme – Project Sepsis). PML is funded by the Investigator Grant Award Program from the BC Children's Hospital Foundation (Vancouver, Canada). AD is supported by a NIH Fogarty Emerging Global Leader Award K43 TW010682.

The acknowledged funders for lead and co-authors had no role in manuscript writing; nor in the decision to submit the paper for publication. The corresponding author and lead author confirm that they had final responsibility for the decision to submit for publication.

References

1 Rudd KE, Johnson SC, Agesa KM, *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet* 2020; **395**: 200–11.

2 Liu L, Johnson HL, Cousens S, *et al.* Global, regional, and national causes of child mortality: An updated systematic analysis for 2010 with time trends since 2000. *Lancet* 2012; **379**: 2151–61.

3 Boeddha NP, Schlapbach LJ, Driessen GJ, *et al.* Mortality and morbidity in community-acquired sepsis in European pediatric intensive care units: A prospective cohort study from the European Childhood Life-threatening Infectious Disease Study (EUCLIDS). *Crit Care* 2018; **22**: 1–13.

4 Chaurasia S, Sivanandan S, Agarwal R, Ellis S, Sharland M, Sankar MJ. Neonatal sepsis in South Asia: Huge burden and spiralling antimicrobial resistance. *BMJ* 2019; **364**. DOI:10.1136/bmj.k5314.

5 WHO. World Conference on Social Determinants of Health. https://www.who.int/sdhconference/background/news/facts/en/ (accessed Jan 19, 2021).

6 Beckfield J, Olafsdottir S, Bakhtiari E. Health Inequalities in Global Context. *Am Behav Sci* 2013; **57**: 1014–39.

7 Popescu CR, Cavanagh MM, Tembo B, *et al.* Neonatal sepsis in low-income countries: epidemiology, diagnosis and prevention. *Expert Rev Anti Infect Ther* 2020; **18**: 443–52.

8 WHO. Management of Childhood Illness. https://www.who.int/teams/maternal-newborn-child-adolescent-health-and-ageing/child-health/integrated-management-of-childhood-illness (accessed Feb 27, 2021).

9 Fuchs A, Bielicki J, Mathur S, Sharland M, Van Den Anker JN. Reviewing the WHO guidelines for antibiotic use for sepsis in neonates and children. *Paediatr Int Child Health* 2018; **38**: S3–15.

10 Dickinson FM, McCauley M, Madaj B, Van Den Broek N. Using electronic tablets for data collection for healthcare service and maternal health assessments in low resource settings: Lessons learnt. *BMC Health Serv Res* 2019; **19**: 1–7.

11 Feroz A, Jabeen R, Saleem S. Using mobile phones to improve community health workers performance in low-and-middle-income countries. *BMC Public Health* 2020; **20**. DOI:10.1186/s12889-020-8173-3.

12 WHO. Big data in global health: improving health in low- and middle-income countries. https://www.who.int/bulletin/volumes/93/3/14-139022/en/ (accessed Jan 19, 2021).

13 WHO. WHO Score Global Report. 2021. https://www.who.int/news/item/01-02-2021-who-score-global-report-highlights-urgent-need-for-better-data-to-strengthen-pandemic-response-and-improve-health-outcomes (accessed April 14, 2021).

14 Rudd KE, Kissoon N, Limmathurotsakul Di, *et al.* The global burden of sepsis: Barriers and potential solutions. *Crit Care* 2018; **22**: 123305059.

15 Johnson SG, Pruinelli L, Hoff A, *et al.* A Framework for Visualizing Data Quality for Predictive Models and Clinical Quality Measures. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci* 2019; **2019**: 630–8.

16 Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA. Big data in global health: improving health in low- and middle-income countries. *Bull World Health Organ* 2015; **93**: 203–8.

17 DePasse J, Celi LA. Collaboration, capacity building and co-creation as a new mantra in global health. *Int J Qual Heal Care* 2016; **28**: 536–7.

18 Deutschman CS, Tracey KJ. Sepsis: Current dogma and new perspectives. *Immunity* 2014; **40**: 463–75.

19 WHO. WHO Sepsis. https://www.who.int/news-room/fact-sheets/detail/sepsis (accessed Feb 14, 2021).

20 Andre Kalil. Septic Shock Treatment & Management. 2020. https://emedicine.medscape.com/article/168402-treatment (accessed Feb 14, 2021).

21 Sands K, Carvalho MJ, Portal E, *et al.* Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries. *Nat Microbiol* 2021; **6**: 512–23.

22 Li G, Bielicki JA, Ahmed ASMNU, *et al.* Towards understanding global patterns of antimicrobial use and resistance in neonatal sepsis: Insights from the NeoAMR network. *Arch Dis Child* 2020; **105**: 26–31.

23 Hamer DH, Darmstadt GL, Carlin JB, *et al.* Etiology of bacteremia in young infants in six countries. *Pediatr Infect Dis J* 2015; **34**: e1–8.

24 Davis AL, Carcillo JA, Aneja RK, *et al.* American College of Critical Care Medicine clinical practice parameters for hemodynamic support of pediatric and neonatal septic shock. 2017 DOI:10.1097/CCM.0000000000002425.

25 Weber M, Klein N, Hartley J, Lock P, Malone M, Sebire N. Infection and sudden unexpected death in

infancy: a systematic retrospective case review. *Lancet* 2008; **371**: 1848–53.
26 Weber MW, Carlin JB, Gatchalian S, Lehmann D, Muhe L, Mulholland EK. Predictors of neonatal sepsis in developing countries. *Pediatr Infect Dis J* 2003; **22**: 711–6.
27 Okomo U, Akpalu ENK, Le Doare K, *et al.* Aetiology of invasive bacterial infection and antimicrobial resistance in neonates in sub-Saharan Africa: a systematic review and meta-analysis in line with the STROBE-NI reporting guidelines. *Lancet Infect Dis* 2019; **19**: 1219–34.
28 Costelloe C, Metcalfe C, Lovering A, Mant D, Hay AD. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: Systematic review and meta-analysis. *BMJ* 2010; **340**: 1120.
29 Bryce A, Hay AD, Lane IF, Thornton H V., Wootton M, Costelloe C. Global prevalence of antibiotic resistance in paediatric urinary tract infections caused by Escherichia coli and association with routine use of antibiotics in primary care: Systematic review and meta-analysis. *BMJ* 2016; **352**. DOI:10.1136/bmj.i939.
30 Zaidi AKM, Thaver D, Ali SA, Khan TA. Pathogens associated with sepsis in newborns and young infants in developing countries. *Pediatr Infect Dis J* 2009; **28**: 10–8.
31 Thaver D, Zaidi AKM. Burden of neonatal infections in developing countries: A review of evidence from community-based studies. *Pediatr Infect Dis J* 2009; **28**: 3–9.
32 Bloos F, Reinhart K. Rapid diagnosis of sepsis. *Virulence* 2014; **5**: 154–60.
33 Medugu, N., Iregbu K. Trends in profiles of bacteria causing neonatal sepsis in Central Nigeria Hospital. *African J Clin Exp Microbiol* 2017; **18**: 49–52.
34 Roche. Roche receives FDA clearance for its Procalcitonin (PCT) assay to help clinicians effectively assess sepsis risk and manage sepsis patients. .
35 Lavoie PM, Popescu CR, Molyneux EM, *et al.* Rethinking management of neonates at risk of sepsis. *Lancet* 2019; **394**: 279–81.
36 Arpi M, Larsen S. Comparison of detection speed and yield in agitated and non-agitated aerobic blood culture bottles. *Apmis* 1992; **100**: 1061–5.
37 Mancini N, Carletti S, Ghidoli N, Cichero P, Burioni R, Clementi M. The era of molecular and other non-culture-based methods in diagnosis of sepsis. *Clin Microbiol Rev* 2010; **23**: 235–51.
38 Holtzman C, Whitney D, Barlam T, Miller NS. Assessment of impact of peptide nucleic acid fluorescence in situ hybridization for rapid identification of coagulase-negative staphylococci in the absence of antimicrobial stewardship interventionδ. *J Clin Microbiol* 2011; **49**: 1581–2.
39 Hemmert A and GN. Rapid methods for pathogen detection in bloodstream infections. *MLO Med Lab Obs* 2013; **45**.
40 Wojewoda C. Pathology consultation on matrix-assisted laser desorption ionization-time of flight mass spectrometry for microbiology. *Am J Clin Pathol* 2013; **140**: 143–8.
41 de Jong E, van Oers JA, Beishuizen A, *et al.* Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: A randomised, controlled, open-label trial. *Lancet Infect Dis* 2016; **16**: 819–27.
42 Sridharan, P., Chamberlain R. The efficacy of procalcitonin as a biomarker in the management of sepsis: Slaying dragons or tilting at windmills? *Surg Infect (Larchmt)* 2013; **14**: 489–511.
43 Faix JD. Biomarkers of sepsis. *Crit Rev Clin Lab Sci* 2013; **50**: 23–36.
44 Chiesa C, Natale F, Pascone R, *et al.* C reactive protein and procalcitonin: Reference intervals for preterm and term newborns during the early neonatal period. *Clin Chim Acta* 2011; **412**: 1053–9.
45 Stocker M, van Herk W, el Helou S, *et al.* Procalcitonin-guided decision making for duration of antibiotic therapy in neonates with suspected early-onset sepsis: a multicentre, randomised controlled trial (NeoPIns). *Lancet* 2017; **390**: 871–81.
46 Hofa N, Müller W, Resch B. Non-infectious conditions and gestational age influence C-reactive protein values in newborns during the first 3 days of life. *Clin Chem Lab Med* 2011; **49**: 297–302.
47 Pierrakos C, Vincent JL. Sepsis biomarkers: A review. *Crit Care* 2010; **14**: 1–18.
48 Kim MH, Choi JH. An update on sepsis biomarkers. *Infect Chemother* 2020; **52**: 1–18.
49 Kumar A, Roberts D, Wood K. Dutation of hypertension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006; **34**: 1589–96.
50 Cummings CA, Relman DA. Using DNA microarrays to study host-microbe interactions. *Emerg Infect Dis* 2000; **6**: 513–25.
51 Campbell CJ, Ghazal P. Molecular signatures for diagnosis of infection: Application of microarray technology. *J Appl Microbiol* 2004; **96**: 18–23.
52 Cvijanovich N, Shanley TP, Lin R, *et al.* Validating the genomic signature of pediatric septic shock. *Physiol Genomics* 2008; **34**: 127–34.
53 Pennisi I, Rodriguez-Manzano J, Moniri A, *et al.* Translation of a host blood RNA Signature distinguishing bacterial from viral infection into a platform suitable for development as a point-of-care

test. *JAMA* 2021; **175**: 417–9.
54 Gliddon HD, Herberg JA, Levin M, Kaforou M. Genome-wide host RNA signatures of infectious diseases: discovery and clinical translation. *Immunology* 2018; **153**: 171–8.
55 Wong HR, Cvijanovich N, Allen G. Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum. *Crit Care Med* 2009; **37**: 1558–66.
56 Jacobs L, Berrens Z, Stenson EK, *et al.* The Pediatric Sepsis Biomarker Risk Model (PERSEVERE) Biomarkers Predict Clinical Deterioration and Mortality in Immunocompromised Children Evaluated for Infection. *Sci Rep* 2019; **9**: 1–8.
57 Lavoie PM, Popescu CR, Tembo B, *et al.* Whole blood genome-wide transcriptome profiling and metagenomics next-generation sequencing in young infants with suspected sepsis in a low-and middle-income country: A study protocol. *Gates Open Res* 2020; **4**: 1–32.
58 Khondoker MR, Bachmann TT, Mewissen M, *et al.* Multi-factorial analysis of class prediction error: Estimating optimal number of biomarkers for various classification rules. *J Bioinform Comput Biol* 2010; **8**: 945–65.
59 Lee AH, Shannon CP, Amenyogbe N, *et al.* Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. *Nat Commun* 2019; **10**. DOI:10.1038/s41467-019-08794-x.
60 Smith CL, Dickinson P, Forster T, *et al.* Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat Commun* 2014; **5**. DOI:10.1038/ncomms5649.
61 Dickinson P, Smith CL, Forster T, *et al.* Whole blood gene expression profiling of neonates with confirmed bacterial sepsis. *Genomics Data* 2015; **3**: 41–8.
62 Bah SY, Forster T, Dickinson P, Kampmann B, Ghazal P. Meta-analysis identification of highly robust and differential immune-metabolic signatures of systemic host response to acute and latent tuberculosis in children and adults. *Front Genet* 2018; **9**: 1–16.
63 Kan B, Michalski C, Fu H, *et al.* Cellular metabolism constrains innate immune responses in early human ontogeny. *Nat Commun* 2018; **9**.
64 Naseem M, Akhund R, Arshad H, Ibrahim MT. Exploring the Potential of Artificial Intelligence and Machine Learning to Combat COVID-19 and Existing Opportunities for LMIC: A Scoping Review. *J Prim Care Community Heal* 2020; **11**. DOI:10.1177/2150132720963634.
65 UNICEF. Avyantra: Using machine learning to facilitate early treatment to infants with neonatal sepsis. https://www.unicef.org/innovation/stories/avyantra-using-machine-learning-facilitate-early-treatment-infants-neonatal-sepsis (accessed Feb 15, 2021).
66 Weiss SL, Peters MJ, Alhazzani W, *et al.* Surviving sepsis campaign international guidelines for the management of septic shock and sepsis-associated organ dysfunction in children. *Pediatr Crit Care Med* 2020; : E52–106.
67 McBall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7129544/pdf/main.pdf (accessed Feb 14, 2021).
68 NICE. EVIDENCE STANDARDS FRAMEWORK FOR DIGITAL HEALTH TECHNOLOGIES. 2019. https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf (accessed Feb 15, 2021).
69 Lee V, Dunsmuir D, Businge S, *et al.* Evaluation of a digital triage platform in Uganda: A quality improvement initiative to reduce the time to antibiotic administration. *PLoS One* 2020; **15**: 1–15.
70 Paediatric Rapid Sepsis Trigger (PRST) Tool (PRST). https://clinicaltrials.gov/ct2/show/NCT04304235.
71 DeLuca A, Betz J, Bollinger R, Ray SC, Manabe YC. Impact of Point-of-Care Diagnostics on Clinical Decision-making in Low- and Middle-Income Countries. *J Appl Lab Med* 2018; **3**: 456–9.
72 O'Sullivan S, Ali Z, Jiang X, *et al.* Developments in transduction, connectivity and AI/machine learning for point-of-care testing. *Sensors (Switzerland)* 2019; **19**: 1–31.
73 Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. *Sensors (Switzerland)* 2019; **19**: 1–45.
74 Waters H. New 10 million X prize launched for tricorder-style medical device. *Nat Med* 2011; **17**: 754.
75 Chan J, Raju SC, Topol E. Towards a tricorder for diagnosing paediatric conditions. *Lancet* 2019; **394**: 907.
76 Muinga N, Magare S, Monda J, *et al.* Digital health Systems in Kenyan Public Hospitals: A mixed-methods survey. *BMC Med Inform Decis Mak* 2020; **20**: 1–14.
77 WHO. Global Health Observatory (GHO) data. https://www.who.int/gho/goe/electronic_health_records/en/ (accessed Jan 19, 2021).
78 Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv. Res. 2010; **10**. DOI:10.1186/1472-6963-10-231.
79 Ajami S, ArabChadegani R. Barriers to implement Electronic Health Records (EHRs). *Mater Socio*

Medica 2013; **25**: 213.
80 Drake TM, Knight SR, Harrison EM, Søreide K. Global inequities in precision medicine and molecular cancer research. *Front Oncol* 2018; **8**. DOI:10.3389/fonc.2018.00346.
81 Green ED, Gunter C, Biesecker LG, *et al.* Strategic vision for improving human health at The Forefront of Genomics. *Nature* 2020; **586**: 683–92.
82 Tekola-Ayele F, Rotimi CN. Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges. *Public Health Genomics* 2015; **18**: 242–7.
83 Mitropoulos K, Cooper DN, Mitropoulou C, *et al.* Genomic Medicine Without Borders: Which Strategies Should Developing Countries Employ to Invest in Precision Medicine? A New Fast-Second Winner Strategy. *Omi A J Integr Biol* 2017; **21**: 638–46.
84 Mangul S, Martin LS, Langmead B, *et al.* How bioinformatics and open data can boost basic science in countries and universities with limited resources. *Nat Biotechnol* 2019; **37**: 324–6.
85 Narayanan I, Nsungwa-Sabiti J, Lusyati S, *et al.* Facility readiness in low and middle-income countries to address care of high risk/ small and sick newborns. *Matern Heal Neonatol Perinatol* 2019; **5**: 1–14.
86 Garces AL, Mcclure EM, Pérez W, *et al.* The Global Network Neonatal Cause of Death algorithm for low-resource settings HHS Public Access Author manuscript. *Acta Paediatr* 2018; **2017**: 904–11.
87 Figueiredo AS. Data Sharing: Convert Challenges into Opportunities. *Front Public Heal* 2017; **5**: 1–6.
88 Alter GC, Vardigan M. Addressing Global Data Sharing Challenges. *J Empir Res Hum Res Ethics* 2015; **10**: 317–23.
89 Dornan L, Pinyopornpanish K, Jiraporncharoen W, Hashmi A, Dejkriengkraikul N, Angkurawaranon C. Utilisation of Electronic Health Records for Public Health in Asia: A Review of Success Factors and Potential Challenges. *Biomed Res Int* 2019; **2019**. DOI:10.1155/2019/7341841.
90 Liu C, Talaei-Khoei A, Zowghi D, Daniel J. Data Completeness in Healthcare: A Literature Survey. *Pacific Asia J Assoc Inf Syst* 2017; **9**: 75–100.

Figure Legend

Figure 1: Complexity of clinical, ethical and regulatory barriers to harnessing and harmonising neonatal and paediatric infection 'big' data from resource-limited settings

A mind-map visually demonstrating the complexity of known clinical, ethical and regulatory barriers surrounding the use of big-data in low- and middle-income countries (LMICs)

Figure 2 Shows workflow of how existing and future EHR big-data and multi-omics data (from HICs and LMICs) can be used to identify physiological signals and improve biomarker identification from multi-omics data.

The complexity and scale of these data require the use of simulation models (top panel) that generate digital populations for optimising classification rules for predicting and identifying infants at risk for sepsis. Data from HICs is biased towards hospital-acquired while LMICs are weighted for community acquired infections. Data-characteristics that account for these differences need be modelled for optimising parameters of prediction algorithms prior to real-world testing. For both HICs and LMICs large-scale cross-country multi-cohort independent clinical studies need to be performed. This emphasis the need to harmonise global policies and requirements to facilitate data sharing. It is notable that capabilities of clinical trial sites in LMICs can span the entire continuum of development, from discovery through to approval yet, e.g. only about 2% of trials come to Africa, and 70% of those trials are conducted in only three countries. https://www.ctc.africa/about. High quality curated databases will need to be established through global collaborative efforts involving expertise from both HICs and LMICs, will facilitate effective application of statistical inference, ML and AI.

Figure 3: A schematic of data-processing steps of a medical tricorder; providing an integrated digital solution enabling precision diagnosis of sepsis.

A schematic summarising how medical tricorder technology can be used for hospital or community acquired infections. These hand-held portable devices scan and integrate physiological and molecular signals from sensors, enabling precision diagnosis of sepsis lowering barriers for the need for specialised equipment or laboratory tests to diagnose. Left side of the panel shows a timeline for diagnostic information flow and outputs with results that can be transmitted to cloud healthcare systems. The digital early warning algorithm uses medical record inputs (marked by arrows) at presentation while the host early warning uses genomic/metabolic biomarker inputs from whole blood point-of-care testing. Systems integration involves using logic circuitry linking molecular pathophysiological biomarkers with physiological and clinical observations. Logic circuitry involves use of the Exclusive –NOR gate – this gives a low output (digital 0) if either, but not both of the two inputs are high (digital 1). With potential downstream pathogen detection, coupled to the final output by using an "AND" logic gate, giving a high output (1) if one or more of its inputs (shown in arrows) are high. Abbreviations: EHR:

Electronic/Medical Health Record of physiological clinical signs, POC: Point-of-care blood test. Logic gates: AND denotes an "AND" logic gate and "ENOR" denotes an "Exclusive" NOT-OR logic gate.

Figures 4a and 4b

Figure 4a. Percentages of population level data collected in LMICs

A bar chart representing the estimated percentages of recorded population level data in LMICs relating to births, deaths and healthcare deliveries. These data were sourced from: Garces AL, Mcclure EM, Pérez W, *et al.* The Global Network Neonatal Cause of Death algorithm for low-resource settings HHS Public Access Author manuscript. *Acta Paediatr* 2018; **2017**: 904–11.[83]

Figure 4b. A conceptual timeline on clinical, laboratory and research barriers

A conceptual timeline based on arbitrary data visually demonstrating the gaps between LMIC and high-income countries in terms of clinical, laboratory and research barriers.