EVALUATION OF THE CHNRI PROCESS FOR SETTING HEALTH RESEARCH PRIORITIES

SACHIYO YOSHIDA

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

University of London

May 2019

Department of Infectious Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Supervisors: Professor Simon Cousens of the London School of Hygiene & Tropical Medicine and Professor Igor Rudan of the University of Edinburgh

No funding received

Declaration

I, Sachiyo Yoshida, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

29 May 2019

In loving memory of my Grandmother, Toshiko Yamashita (1927–2017)

Acknowledgements

First of all, I would like to express my special appreciation, and most sincere thanks, to my supervisors, Professor Simon Cousens and Professor Igor Rudan. Without your supervision, this work would simply not have been possible. Simon, I admire your crystal-clear thinking and mental sharpness. Each time I have read your comments, I have learned something new from the way you think. Igor, I admire your creativity, agility and flexibility in helping me to find the solutions to the problems and challenges that I have encountered over the past five years. You have both shared your warmth and kindness with me. It has been an honour, pleasure and privilege to have completed my PhD under your supervision.

I would like to thank my supervisors at work, Dr Rajiv Bahl and Dr Jose Martines for encouraging me to pursue my PhD studies, and for allowing me to grow and develop in research. It has already been eight years since I started to work with you. You have always been most supportive and caring, and eager to encourage me to learn more. Rajiv and Jose, I hope that you know just how grateful I am to you.

I would like to express my most sincere gratitude to my family for their love and sacrifices. It was not easy to pursue a PhD with a baby who is now two and a half years old. Without my family, this PhD thesis would never have been completed. I dedicate my thesis to the loving memory of my Grandmother, Toshiko Yamashita, with whom I spent my childhood and whose influence on my life is inestimable. The last few words of acknowledgment I have saved for my dear Husband, Yoshinobu Nagamine. This journey would not have been possible without you. I am grateful to my Parents, who supported me emotionally and even flew to Geneva to enable me to finish the final sprint. I am also thankful to my Parents-in-law who always believed in me and wanted the best for me. Maguy and Mario, my son's day care "grand-parents", thank you so much for taking care of my son as if he were your own child. Knowing that Kei has been with you has made me feel so much less guilty about not being able to play with him as much as he, and I, would have liked.

Kei, I am sorry for not always being able to be with you when you wanted me most. Nothing made me sadder to see you crying. You may not remember that I used to breastfeed you before you fell asleep in one arm while I was typing this thesis with my other hand. One day you will know.

To my friends, thanks for listening, and for offering me advice. I reserve special thanks for Professor David Ross, my lunch mate, for his very thoughtful suggestions, as well as for Professor Suman Rao for babysitting Kei, to Neil Livingstone for your generosity in editing my PhD thesis, to Theadora Swift Koller and Sapna Manglani for being such brave and caring friends. Your supports meant a lot to me. Finally, I would like to thank all the participants in the CHNRI exercises for their input.

Abstract

Background: The Child Health Nutrition Research Initiative (CHNRI) method is a tool used in health research prioritisation (RP). Though the method is widely used, inherent challenges remain: the method may be affected by on-going research where self-selected participants have personal interests; no evidence-based guidelines exist on the sample size of the expert group; and little evaluation has been conducted on the quality and impact of the method. This PhD presents an example of the application of the method, studies these inherent challenges, and assesses the quality and impact of previously conducted CHNRI exercises.

Methods: The methods include a comprehensive review of RP approaches published between 2000 and 2014; coordination of two global RP exercises; statistical analyses of previously conducted CHNRI exercises; and assessment of the quality of method's application using an evaluation framework and a survey.

Results: Approximately one in four RP exercises used the CHNRI method between 2000 and 2014. In the previously conducted CHRNI exercises, substantial potential for self-selection bias was noted. Statistical analyses identifying a minimum sample size of experts yielded varied results across different CHNRI exercises. The evaluation of the quality of the process identified that the CHNRI exercises met most of the requirements to be qualified as good practice.

Conclusion: To my knowledge, this is a first attempt to evaluate some key components of the CHNRI method. The varied results in the sample size analyses prevented any recommendation being made at this point. Many RP exercises end once the priorities are identified, without assessing whether the RP exercise is effective in mobilizing funds for identified priorities. Future RP exercises should add a follow up at a later time point to assess whether funding has been allocated and if so, how much funding has been allocated to priority areas.

Table of Contents

List of Tables

List of Figures

List of Appendices

General introduction

Motivation to pursue the PhD

I work in the Department of Maternal, Newborn, Child and Adolescent Health at the World Health Organization (MCA/WHO). I have chosen my PhD topic on basis of the role I play in the MCA/WHO. I have two key responsibilities in the Department. Firstly, I support the overall project management of several multi-country studies in newborn health, for which the Department provides the overall technical coordination and supervision. The other equally important responsibility directly linked to this PhD is to coordinate the research prioritisation exercises that MCA/WHO coordinates globally, and to provide methodological guidance on the process of prioritisation in the area of maternal, newborn, child and adolescent health. I have chosen my PhD topic based on this latter role in the Department, i.e. to coordinate research prioritisation exercises in maternal, newborn, child and adolescent health, with the intention of strengthening methodological expertise.

In WHO, setting global health research priorities is part of the organization-wide agenda. In 2010, the WHO strategy on research for health was adopted at the sixty-third World Health Assembly.[1] The strategy aims to achieve five visions:

1) Strengthening the research environment across the organization.

2) Setting research priorities (RP) based on the important health needs.

3) Strengthening national capacity to conduct health research.

4) Promoting good practice in health research.

5) Reinforcing the linkage between research outcomes and their translation to policy and programming.

Clearly, research prioritisation is the key organization-wide priority in WHO, not only because evidence obtained through research supports the development of health interventions to solve health problems but also this reinforces the role and responsibility of WHO in global health architecture.

The MCA Department is the WHO's leading department in health research prioritisation; the MCA Department produced the largest number of information products based on research prioritisation exercises in the area of child and adolescent health.[2] Since 2007, our department has coordinated ten global research prioritisation exercises using the Child Health Nutrition

Research Initiative (CHNRI) methodology in the area of maternal, newborn, child and adolescent health. The department has also led several workshops in various countries to develop research proposals in the priority areas identified by the exercises. Providing methodological guidance on how to conduct research prioritisation, as well as coordinating global health research prioritisation exercises, are obviously key roles of the department, which address one of the six core functions of WHO, namely "shape the research agenda and simulate the generation, translation and dissemination of valuable knowledge".[3]

The CHNRI method is a method widely used in health research prioritisation exercises. Comprehensive review revealed that more than one in four RP exercises used this method[4], reaching over 50 health RP exercises at global, national and sub-national level till date).[5] Though widely used, the method has not been evaluated for its quality of process and impact of prioritisation exercises. The CHNRI method uses the collective opinion of "crowds" of experts to identify the health research agenda. A CHNRI exercise produces a ranking of many research ideas based on the collective opinion of the expert group. However, it is yet to be demonstrated that the collective opinion of an expert group should be regarded as more useful than the opinion of individual experts in the group, and under what circumstances collective opinion outperforms individual opinion. Secondly, the method is not free of potential bias. An example of a potential bias is self-selection bias by which scoring may be affected by on-going research in which self-selected participants have personal interests. Thirdly, current guidelines recommend a large and diverse group of experts to be invited into the process, however how large is optimal for the CHNRI method? The guideline does not provide any indication on optimal sample size of experts to be invited in the process. Fourthly, there has not been evaluation of quality of process and impact of previously conducted research prioritisation exercises. The overarching aim of my PhD thesis is to examine the steps in CHNRI method and identify recommendations for further refinement of the process. Furthermore, this PhD assesses the quality of the method's application and impact of previously conducted CHNRI exercises.

Contribution statement

I led the work described in Chapter 1, Chapter 2 and Chapter 4. The initial idea of Chapter 3 came from Professor Igor Rudan, and I have conducted all the data analyses. Professor Simon Cousens provided substantial guidance on data plan and results.

Introduction to the Child Health Nutrition Research Initiative (CHNRI) methodology

History of CHNRI

The Child Health Nutrition Research Initiative (CHNRI) methodology was developed by the Child Health and Nutrition Initiative of the Global Forum for Health Research to assist policy makers and research funding agencies to make investment decisions in health research[6] and to address investment gaps relating to health problems affecting poor and vulnerable populations.[7]

Fifteen interdisciplinary experts developed the CHNRI methodology between 2005 and 2007. Their expertise included international health, health policy, paediatrics and child health, economics and management science, political science, law and ethics, and included researchers, programme leaders from low- and middle-income countries and members of international organisations. The expert panel was coordinated by Professors Igor Rudan (University of Edinburgh, UK), Shams El Arifeen (ICDDR, B in Dhaka, Bangladesh) and Robert Black (Johns Hopkins University, Baltimore, USA). The panel identified common challenges and problems by reviewing methodologies previously used in setting research priorities (**Table 1**). In addressing a few key challenges, the group concluded that there were no methods or tools available that could enable a systematic, transparent, legitimate and fair, scientifically rigorous and replicable process of priority setting. The absence of such tools became the driving force behind the development of the CHNRI method.

Table 1. Twenty "universal challenges" in setting priorities in health research investments, according to the CHNRI expert group[6]

1. Deciding who should be involved in the process of setting research priorities
2. Defining what constitutes a health research investment option opportunity
3. Defining what constitutes the expected "return" on the investment
4. Defining what constitutes a potential "risk" in the investment
5. Finding a way of dealing with uncertainty of health research outcomes
6. Defining health research, its boundaries, and its levels of "depth"
7. Systematic listing of many competing research investment options
8. Defining what is meant by "priority setting" in the context of health research
9. Defining criteria relevant to priority setting in health research investments
10. Comparing different domains of health research using the same criteria
11. Development of a simple quantitative way to rank competing research options
12. Limiting the potential of personal biases to substantially influence the outcome
13. Ensuring that priority setting process is fully transparent
14. Ensuring that it can be repeated and validated
15. Ensuring that it is flexible and adjustable to all contexts and levels of application
16. Ensuring that it is iterative with a feedback loop, instead of a one-way process
17. Ensuring that it is perceived by the users as legitimate and fair
18. Ensuring that it is simple and intuitive enough to become popular among the users
19. Linking quantitative ranks of research options with specific investment decisions
20. Involving stakeholders from the wider community into the process

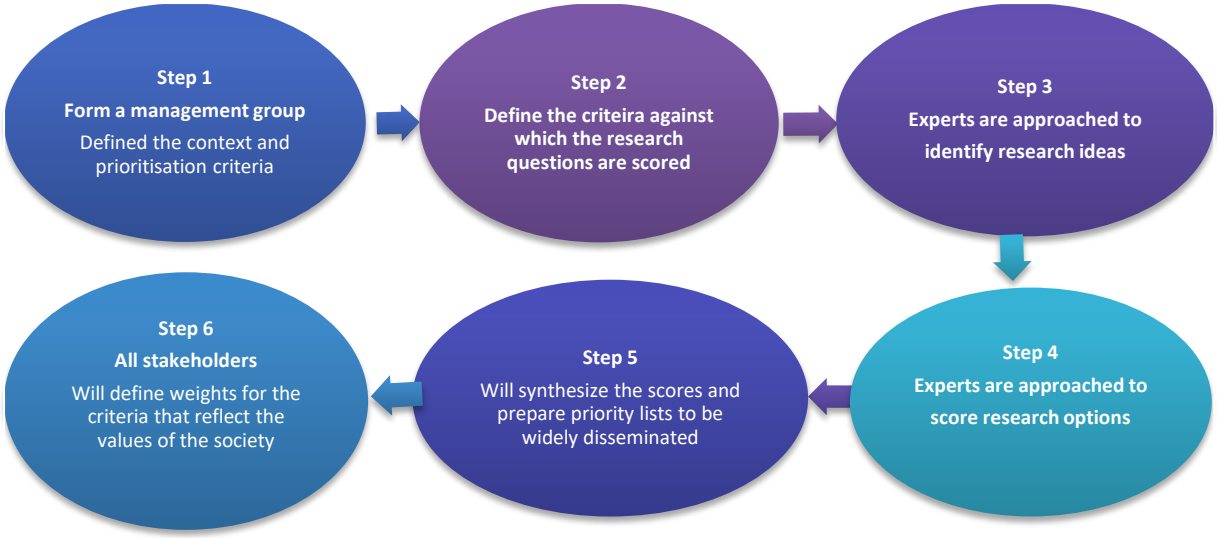
The CHNRI methodology

Outline of the process

The CHNRI method offers comprehensive guidance to the process of research prioritisations RPs (**Figure 1**). The CHNRI process involves several steps, and involves a multidisciplinary group of individuals, which may include policy makers, government officials, health professionals, researchers, programme managers and donors. A management team consisting of technical and methodological experts leads and coordinates the process. It selects a group of experts with relevant technical knowledge and experience in the research area. The roles of experts are two-fold n: to identify research ideas; and to independently score research options. Stakeholders are selected based on their professional background and the type of interests they have. They are selected by the management team and are usually involved in the final stage of the method. The profile of stakeholders is specific to each RP. They can be technical experts such as researchers, policy makers, programme managers, as well as a non-technical crowd such as consumer groups, patient groups, and care providers. Their role in the process is to set the thresholds and weights for each criterion to reflect their value system.

In CHNRI exercises four domains are used to classify the type of health research. “Descriptive research includes epidemiological research questions that assess the burden of health problems in a given population. “Delivery” research includes operational research to improve the health status of the population by improving the delivery of existing effective interventions. “Development” research includes research questions relating to modifications to existing interventions or to the ways that interventions are delivered, to be more effective, affordable or sustainable. “Discovery” research includes biomedical research questions that would lead to innovation and discovery in basic science.

Figure 1. Steps in the CHNRI methodology



Step 1

Usually, the process is run by a relatively small number of people who coordinate the entire process. The first step of the method is, therefore, to select the members of the management group who will coordinate the process. A management group typically includes five to six people including funders, methodological experts and experts in the technical area for which the research priorities are set. In the globally lead research prioritisation exercises, the management group usually includes members from international organisations such as the WHO and UNICEF, as well as technical and methodological experts from academia. In most cases, the instigator of the research prioritisation process is a member of the management group. Though not all CHNRI exercises included funders in the management group, it is helpful to involve them because one of the aims of the CHNRI priority setting method is to help funders learn about the potential risks associated with each investment option.

Subsequently, the management group decides the context and scope of the research area for prioritisation. There are four elements to be discussed in defining the context of the research prioritisation exercise. Firstly, the management group needs to define target populations that would benefit from the research output by reducing the disease burden and improving health. For example, the target population could be those living in low- and middle-income countries rather than those in high-income countries. The population could also be an age-specific group.

Secondly, the health problem needs to be clearly defined in terms of existing knowledge about disease burden, disability and existing research gaps.

Thirdly, it is important to pre-define the timeline within which the results of the research are expected. The expected timeline may depend on the nature of the research, given that basic biomedical research requires longer time spans for the development of new interventions or

tools, while implementation research requires less time to adapt existing interventions to the local needs.

Lastly, the nature of the investment strategy needs to be discussed considering funding support. Some funders like high-risk and high-return research such as studying the mechanisms of Alzheimer's disease for new therapeutic strategies, while others prefer combined research in which varied patterns of risk and return exist.

Step 2

The subsequent step is about discussing and deciding the criteria against which the research questions are scored, hence the criteria should be specific to the context of the exercise. There are 18 criteria proposed in the CHNRI method. Most of previously conducted CHNRI exercises used following five standard criteria: answerability, efficacy, deliverability and acceptability, potential impact and equity (**Table 2**). These five criteria were derived based on a framework in which research options would initially generate new knowledge, and then the new knowledge will be translated into either an improved health intervention or a new intervention. Implementation of such a health intervention is expected to lead to disease burden reduction. The standard criteria were expected to assess the likelihood of the progress in this framework.

The CHNRI method provides other 13 criteria (Table 3). The management group is expected to choose criteria depending on the focus of prioritisation exercise or can introduce new criteria that are more relevant to the context of prioritisation exercise, therefore number of criteria is not fixed to five and more than five criteria could be used. For example, Kennedy et al introduced a new criterion "community involvement and sustainability" since the objective of the research prioritisation exercise included community engagement in the design and conduct of research as well as community ownership of potential research result.[8] Tomlinson et al also introduced a new criterion "applicability and sensitivity" as focus of the exercise was on patient-centred care and improving access among people with disability. [9] In all CHNRI exercises, research options will be judged according to the selected set of criteria. The selected criteria will help assess the likelihood that each research option is worth investing in within the specified context.

Table 2. CHNRI criteria that are most frequently used (adapted from Rudan et al, 2008 and Rudan 2016)[10,11]

Answerability Guiding question: Based on the elements below, can a study be designed to answer the research question in ethical way? • Clarity of the research question and its proposed endpoints • Level of research capacity required to conduct the proposed research • Size of the gap from current level of knowledge to the proposed endpoints • Proposed research obtaining ethical approval without major concern
Efficacy Guiding question: Is the proposed research likely to result in an intervention or programme that is efficacious? • The study is likely to find a beneficial effect on proposed outcomes under research conditions based on the best current evidence and knowledge available in the public domain
Deliverability and acceptability Guiding question: Based on the elements below, can the proposed research lead to deliverable interventions that are accepted by users (or improve their deliverability)? • The infrastructure and resources required to deliver the intervention or programme (human resources, health facilities, communication and transport infrastructure, etc.) • The need for change in demand, beliefs and attitudes of users to ensure that they accept the intervention
Potential impact Guiding question: Is it likely that the study will, directly or indirectly, help to reduce the global burden of mortality and severe long-term disability by at least 5%to 10%?
Effect on equity Guiding question: Based on the elements below, is a study likely to help reduce the burden of mortality or severe long-term disability in the most vulnerable socioeconomic strata and therefore improve equity? • The present distribution of the burden mortality • The evidence of the long-term effects on equity of the existing health interventions

Table 3. CHNRI other optional criteria (adapted from Rudan 2016)[12]

Criteria	Description
Attractiveness	Research idea is likely to result in publication in high-impact journals
Novelty	Research idea is likely to result in knowledge that is novel and does not exist at present
Potential for translation	Research idea is likely to result in the translation of knowledge into health intervention

Affordability	Research idea in which translation of knowledge leads to intervention that is affordable for the beneficiaries of the research
Public opinion	Research idea that is justified and acceptable to the public
Ethical aspect	Research idea is likely to result in raising ethical concerns
Community involvement	Research idea is likely to result in community mobilization and empowerment through community involvement
Relevance	Research idea is likely to be relevant to the context in which the research will be conducted
Fills key gap	Research idea is likely to fill the key gap in knowledge
Cost	Research idea is likely to result in requiring more funding than others
Fundability	Research idea is more likely to result in receiving funding support within the defined context than others.
Alignment with political priorities	Research idea is likely to be aligned with government priorities
Likelihood of generating patens/lucrative products?	Research idea is likely to result in generating patents or potentially marketable product that some financial return on investments are expected

Step 3

Having agreed which criteria are to be used, technical experts are contacted to take part in two steps of the process: the generation of research ideas; and the scoring of research options. Technical experts are those who have in-depth knowledge of the area the exercise relates to. These technical experts are usually researchers and programme experts who, once identified by the management team, are asked to provide research ideas within the pre-defined context and the scope of prioritisation exercise. The method of selecting and involving technical experts is presented in Chapter C. Usually, research ideas are submitted to the management team through a web-based platform on an anonymous basis. When all the research ideas have been received by the management team, they are reviewed, to identify any similarity between research ideas, either by merging similar ideas or by improving the clarity of the research ideas. This laborious review process helps to transform the original submitted research ideas into a list of research options in which the target population, intervention, and research outcomes are clearly indicated.

Step 4

Experts are then contacted again to score each research option. Technical experts are asked to systematically score each research option against the pre-defined criteria on a three-point scale: 1 point (Yes, I agree that the research option meets the criterion), 0 point (No, I do not agree), 0.5 point (Informed but undecided answer) and missing (Insufficiently informed to answer).

Although participants are given the choice of three responses, a yes or no is usually encouraged except in cases where scorers feel that they may not be able to state yes or no, in which case they are asked to use the midpoint score. The scoring is conducted by each technical expert independently. The scoring sheet, when completed, is submitted online to the management team by the technical expert.

Step 5

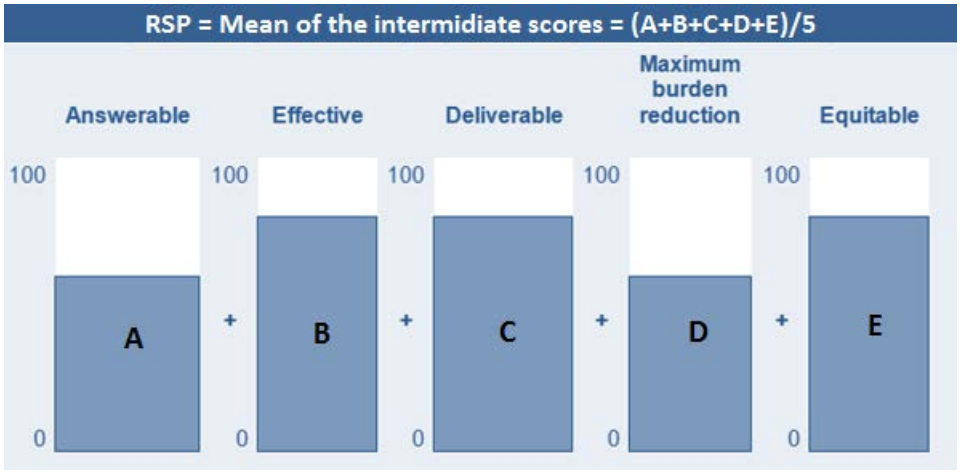
When the scores have been received, the first step is to calculate the intermediate scores. The intermediate scores are the mean of all scores received by a given research option for each criterion, i.e., sum of scores divided by the number of respondents who provided a non-missing response including 1 point, 0 point or 0.5 point. Missing inputs (Blank) are neither part of the numerator nor part of the denominator. All research options have assigned values ranging between 0% and 100% for each criterion. The overall research priority score (RPS) is then computed as the mean of all intermediate priority scores (Figure 2). "Weights" provided by the stakeholder group are used to weight each criterion, and a "threshold" is used to determine a minimum acceptable score against each criterion that would be required for a research option to be considered a research priority. RPS represents the technical experts' perceived likelihood that each of the proposed research options would satisfy the priority setting criteria. Research options can be sorted in descending order of RPS to develop the final list of priority research options. Further details of the calculation are provided in Chapter B.

Average Expert Agreement (AEA) scores were computed as the average proportion of scorers who provided the most frequent response. AEA indicates the extent of agreement or controversy reported for each research options. The computation of AEA is as follows:

$$\text{AEA (average expert agreement)} = \frac{1}{5} \times \sum_{q=1}^{5} \frac{\text{N (scorers who provided most frequent response)}}{\text{N (scorers who provided any response)}}$$

(where q is a criterion that experts are being asked to evaluate research options, ranging from 1 to 5 for example).

Figure 2 Research priority score (RPS)



Application of CHNRI methodology

As explained in the previous section, the CHNRI method provides an approach for soliciting many research ideas and ranking them in a systematic, transparent and replicable way. The method is widely used in setting research priorities. It has been used by research institutes and academia at the national level, as well as by international NGOs and United Nations Agencies [11 13 14 15 16] as evidenced in a recently conducted independent review of the field.[14]

The CHNRI method has been used in most research prioritisation exercises coordinated by the World Health Organization in areas such as newborn and child health,[16–20] sexual and reproductive health, adolescent health,[21,22] and quality of care from pre-conception to perinatal health.[8] A recent review highlighted an even wider use of the method: of the 165 identified research prioritisation exercises, more than one in four used the CHNRI method, followed by the Delphi method (24%). Interestingly, an increasing number of publications reporting applications of the CHNRI method have been observed in PubMed indexed journals for the past five years. Clearly, there has been an increasing demand for prioritisation of the health research agenda and the CHNRI method has played a central role in providing guidance.

The Child Health Nutrition Research Initiative (CHNRI) method is a tool used in health research prioritisation (RP) exercises. Since its advent in 2007, it has been widely used at the global, national and sub-national level. However, inherent challenges remain. Firstly, there is a concern that scoring may be affected by on-going research in which self-selected participants have personal interests. Secondly, there are no evidence-based guidelines for the sample size of the expert group. Thirdly, there has been little evaluation of the impact of the method on research conducted. This PhD presents an example of the application of the method and studies these inherent challenges. Furthermore, this PhD assesses the quality of the method's application and the impact of previously conducted CHNRI exercises. Given the absence of any assessment of the method to date, it is of both professional and academic interest to evaluate the method, to provide better technical guidance on global research prioritisation. In this thesis I aim to address the aim and objectives mentioned below.

Aim and objectives

The aim of the thesis is to evaluate the process and impact of the CHNRI methodology. The specific objectives of this PhD thesis are to address the research questions mentioned below:

Research Question 1: What are the available tools and approaches used in setting health research priorities in the 21st century?

Research Question 2: What are the lessons to be learned from the research prioritisation exercises using the CHNRI method?

Research Question 3: A CHNRI exercise uses collective opinion of a group of experts to produce a ranking of research ideas. Whether collective opinion of experts outperforms individual expert's opinion is to be examined however the difficulties related to validating personal opinions is that there is no right answer against which to validate the opinion. On the contrary, personal knowledge can be validated against the right answer to a factual question. The accuracy of personal knowledge is an underlying basis of the individual's opinion. Therefore, in this thesis I aim to address following research question: Is collective knowledge better than individual knowledge in predicting correct answer.

Research Question 4: What is the optimal sample size of the experts in a CHNRI exercise?

Research Question 5: What recommendations for involving participants can be drawn from previous experience with CHNRI exercises?

Research Question 6: What are the quality of process and impact of the previously conducted CHNRI exercises?

Introduction of the chapters

This PhD thesis consists of six chapters. There are eight published papers included in this thesis and they are presented in Chapter 1 to Chapter 3.

Chapter 1 addresses research question 1 by presenting review of tools and approaches used in health research prioritisation. It provides the results of a comprehensive review of the available tools and approaches that have been used in health RP exercises between 2000 and 2014. It discusses the strengths and limitations of the different methods and approaches and shows who

has used these tools. Moreover, a paper points to the need for the further refinement and evaluation of the CHNRI method (**Paper 1**).

Chapter 2 addresses research question 2 by examining two examples of the research prioritisation exercises using the CHNRI method. This chapter describes the process and presents the results of two research prioritisation exercises for which I provided methodological expertise. One, in newborn health and birth outcomes, was conducted between 2012 and 2013; for this, more than 600 programme managers and researchers working in newborn and perinatal health were approached. This exercise yielded research priorities in the areas of delivery, development and discovery research respectively (**Paper 2.1 and Paper 2.2**). In 2015, a new research prioritisation exercise was initiated around quality maternal and newborn care that focused on the continuum of care from pre-pregnancy, through birth, postpartum and the early weeks of life. This exercise focused particularly on different models of care that respond better to the needs of women, infants and families. More than a thousand experts were approached, including government officials, researchers, programme officers, members of civil society, and NGOs (**Paper 2.3**).

Chapter 3 addresses research questions 3, 4 and 5 through an assessment of various aspects of the CHNRI methodology. This chapter critically reviews some of the underlying assumptions of the CHNRI method and addresses one of the most frequently asked questions regarding the method.

The CHNRI method uses the collective opinion of "crowds" of experts to identify the health research agenda. More precisely, a CHNRI exercise produces a ranking of many research ideas based on the collective opinion of the expert group. It is yet to be demonstrated that the collective opinion of an expert group should be regarded as more useful than the opinion of individual experts in the group. However, the difficulties related to validating personal opinions do not apply to the validation of personal knowledge, and the accuracy of personal knowledge is an important component underlying an individual's opinion. Because of this, we should expect some parallels between the quantitative properties of human collective knowledge and human collective opinion. **Paper 3.1** examines the accuracy of collective knowledge compared to individual knowledge and whether the there is a variation in the accuracy in the situation when the knowledge is obtained from experts versus when the knowledge is elicited from non-experts.

One of the most frequent questions asked by the users of the method is whether there is an "optimal" sample size of researchers to be invited to participate in the research prioritisation exercise, and if and how the composition of the experts affects the outcome. **Paper 3.2**

addresses the question of sample size; using data from 4 exercises and examines how the composition of the group affects the result.

Chapter C also presents the findings of reviews on the involvement of three groups of participants in the health research prioritisation exercise, researchers, stakeholders and funders in the previously conducted exercises, and makes recommendations for the optimal use of each group (**Paper 4.1, 4.2 and 4.3**).

Chapter 4 addresses research question 6 by evaluating previously conducted health research prioritisation exercises. This chapter aims to assess the quality of the process and the impact on resource allocation resulting from the previously conducted CHNRI exercise.

Chapter 5 provides a discussion of the overall findings from review/analysis/experiments conducted to answer each research question and explicates the implication of findings to future CHRNI exercises.

Chapter 6 provides a conclusion of the study

Chapter 1. Review of tools and approaches used in health research prioritisation

Chapter 1 addresses research question 1 by presenting the landscape of tools and approaches used in health research prioritisation exercises between 2000 and 2014. Chapter 1 summarises overall process, selection of participants, scoring criteria, and discusses advantages and disadvantages of the six most frequently used methods in health research prioritisation exercises.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? Journal of Global Health

1.2. When was the work published? June 2016

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes** / **No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I am the sole author of this publication

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE [redacted] Date 25.07.2018

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

journal of global health

Approaches, tools and methods used for setting priorities in health research in the 21st century

Sachiyo Yoshida

Department of Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland

Background Health research is difficult to prioritize, because the number of possible competing ideas for research is large, the outcome of research is inherently uncertain, and the impact of research is difficult to predict and measure. A systematic and transparent process to assist policy makers and research funding agencies in making investment decisions is a permanent need.

Methods To obtain a better understanding of the landscape of approaches, tools and methods used to prioritize health research, I conducted a methodical review using the PubMed database for the period 2001–2014.

Results A total of 165 relevant studies were identified, in which health research prioritization was conducted. They most frequently used the CHNRI method (26%), followed by the Delphi method (24%), James Lind Alliance method (8%), the Combined Approach Matrix (CAM) method (2%) and the Essential National Health Research method (<1%). About 3% of studies reported no clear process and provided very little information on how priorities were set. A further 19% used a combination of expert panel interview and focus group discussion ("consultation process") but provided few details, while a further 2% used approaches that were clearly described, but not established as a replicable method. Online surveys that were not accompanied by face–to–face meetings were used in 8% of studies, while 9% used a combination of literature review and questionnaire to scrutinise the research options for prioritization among the participating experts.

Conclusion The number of priority setting exercises in health research published in PubMed–indexed journals is increasing, especially since 2010. These exercises are being conducted at a variety of levels, ranging from the global level to the level of an individual hospital. With the development of new tools and methods which have a well–defined structure – such as the CHNRI method, James Lind Alliance Method and Combined Approach Matrix – it is likely that the Delphi method and non–replicable consultation processes will gradually be replaced by these emerging tools, which offer more transparency and replicability. It is too early to say whether any single method can address the needs of most exercises conducted at different levels, or if better results may perhaps be achieved through combination of components of several methods.

Correspondence to:
Sachiyo Yoshida
Department of Maternal, Newborn, Child and Adolescent Health
World Health Organization, Geneva, Switzerland
Avenue Appia CH–27
Geneva, Switzerland
yoshidas@who.int

Apart from the continuing need to prioritize investments in health systems and health interventions, there is also a need to prioritize health research. Health research is difficult to prioritize, because the number of possible competing ideas for research is large, the outcome of research is inherently uncertain, and the impact of research is difficult to predict and measure [1]. A systematic and transparent process to assist policy makers and research funding agencies in making investment decisions is a permanent need.

At national level several methods have been tried: some of the best examples are the Council on Health Research for Development's approach (COHRED) in Brazil, Cameroon, Peru and Philippines; the Essential National Health Research (ENHR) approach in Cameroon and South Africa; and the Combined Approach Matrix (CAM) in Malaysia, Pakistan and Argentina [2,3]. COHRED, ENHR and CAM were all developed by committees set up by international agencies, such as the World Health Organization (WHO) or the Global Forum for Health Research (GFHR). These methods are useful for organizing the available information so that the research prioritization can take place.

To obtain a better understanding of the landscape of approaches, tools and methods used to prioritize health research I conducted a methodical review of the PubMed database covering the period 2001–2014. My primary aim was not to perform an exhaustive review of the field, which would include searching all available scientific databases and grey literature. Instead, I was interested in identifying the methods and tools that are being commonly used in the papers that are most readily accessible through databases in the public domain such as PubMed, and to assess their relative importance and applicability. The review of PubMed for the period between 2001 and 2014 achieves this aim, because this limits the search of priority–setting tools to health topics only, which is the main interest of this analysis, while drawing on a very large database which is publically available and which should contain the vast majority of relevant studies.

METHODS

My search terms included "research priorit* OR priorit* research". These terms were chosen as the most informative combination of search terms after experimenting with several versions of search terms. The search terms identified 343 publications, 138 of which were excluded from the analysis because their contents were irrelevant to health research priority setting. A further 40 studies were excluded because they were review articles which did not attempt to set priorities. In total, 165 relevant studies were identified and retained for the analysis. **Figure 1** shows a flowchart of the review on all research priority setting exercises conducted between 2001 and 2014.

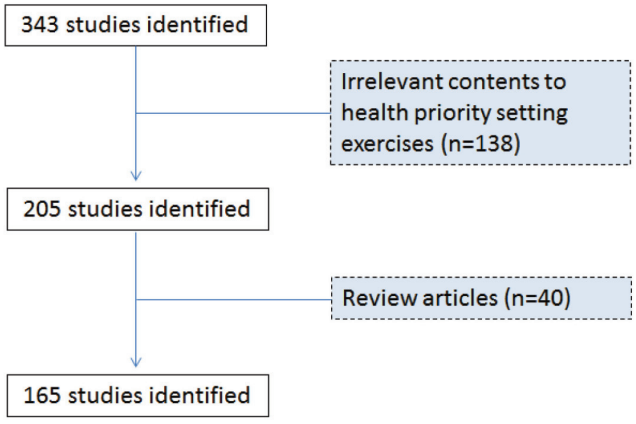


Figure 1. Flowchart of the review on all priority–setting exercises for health research conducted between 2001 to 2014.

RESULTS

Approximately 12 exercises were initiated each year between 2001 and the end of 2014. Since 2012, there has been a steady increase in the number of exercises published with the peak in 2014 with 34 exercises published (**Figure 2**). Of the 165 publications identified, the most frequently used was the CHNRI method (26%), followed by the Delphi method (24%), James Lind Alliance method (8%), the Combined Approach Matrix (CAM) method (2%) and the Essential National Health Research method (<1%). COHRED method, although frequently mentioned and clearly described in the historic context of national–level research priority setting, was not underlying any specific priority–setting process in the time period which I studied. Online surveys that were not accompanied by face–to–face meetings were used in 8% of studies, while 9% used a combination of literature review and questionnaire to scrutinise the research options for prioritization among the participating experts. About 3% of studies reported no clear process and provided very little information on how priorities were set. A further 19% used a combination of expert panel interview and focus group discussion ("consultation process") but provided few details, while a further 2% used approaches that were clearly described, but not established as a replicable method (**Figure 3**). At this point, I would like to clarify that "replicable" refers to the method's description in sufficient detail, so that all other users could apply it in the same way. It does not refer to method's property to yield the same results when repeated, which is a different meaning of the term "replicable" when assigned to a method.

Tables 1 to **6** provide a brief description of the approaches and processes used by the specific methods mentioned in **Figure 3**. The methods range from those that are not described at all, through vaguely described processes of

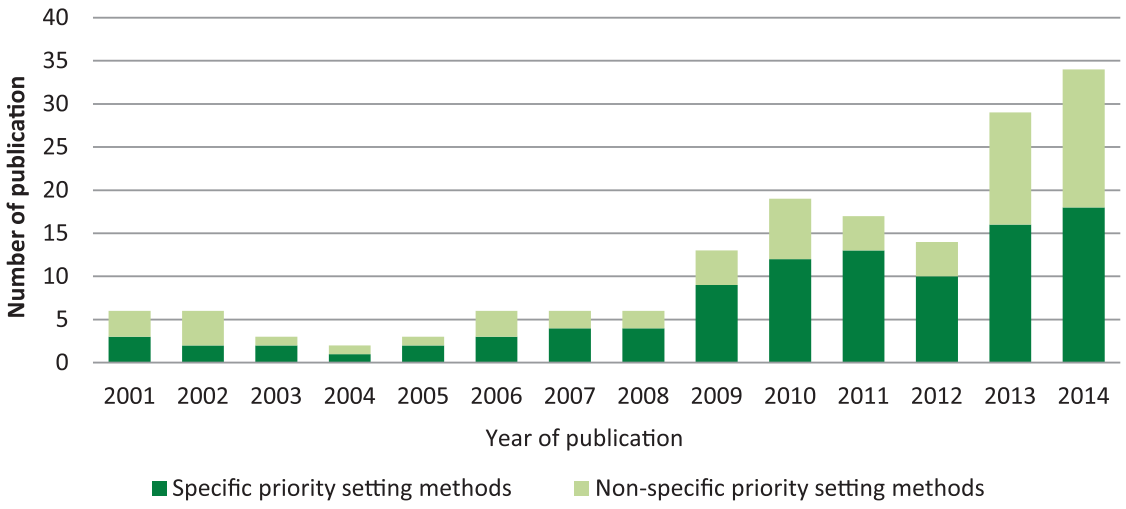


Figure 2. Total number of publication by year (source: PubMed, 2001 to 2014).

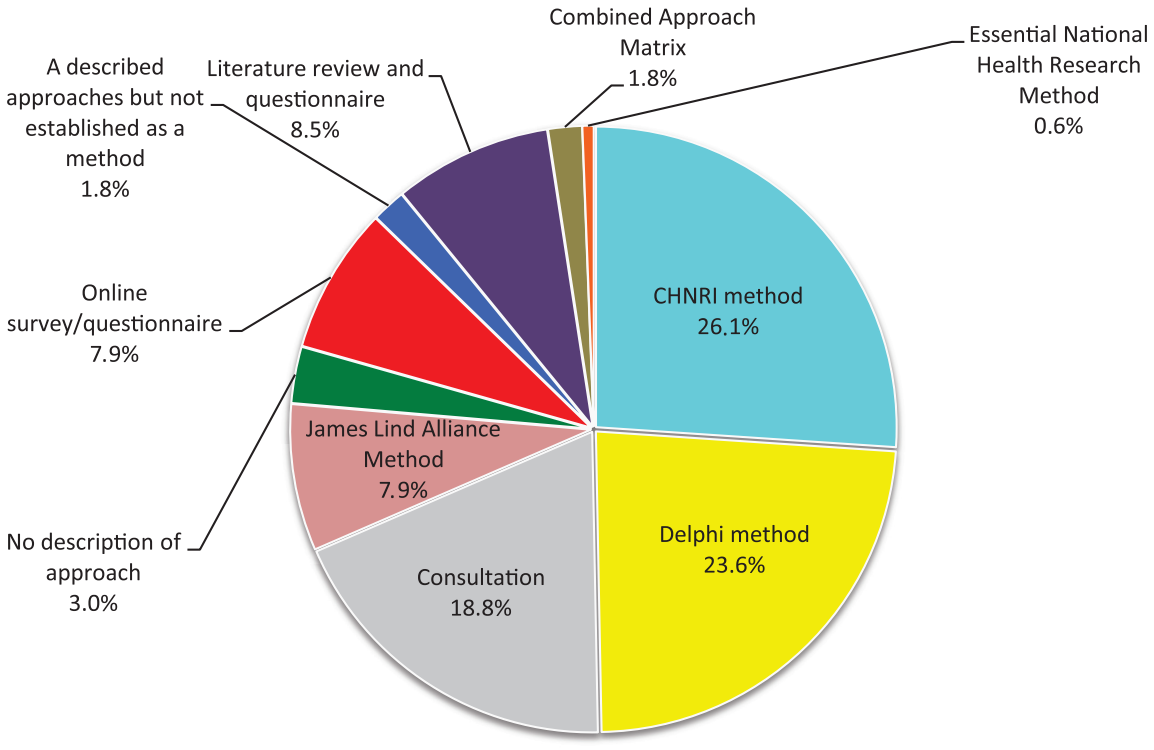


Figure 3. Methods, tools and approaches used for setting health research priorities (source: PubMed, 2001 to 2014).

group decision making, to those that follow a certain structure/process and use transparent criteria. Their output is typically quite general, ie, pointing to broad research areas in which more research activity is needed. As described above, COHRED, ENHR and CAM are used in assembling the evidence that can be used for the consultation but not for the ranking of priorities. Nevertheless, the use of any method, regardless of its limitations, is preferable to the alternative of having no clearly defined approach at all [3].

Among the 165 identified studies that set health research priorities, 21% were conducted at global level, 50% of the exercises were focused on High Income Countries (HICs)

Table 1. Brief explanation of the Essential National Health Research (ENHR) [4–6]

Overall process	ENHR was developed by Commission on Health Research for Development in 1990. It is a step by step guide for national research priority setting, focused on equity in health and development. Strategy focused on inclusiveness in participation, broad–based consultations at different levels, both quantitative and qualitative information used, and stewardship by small working group.
How are participants identified?	Participants are involved through a small representative working group which can facilitate the process, through various consultations. These stakeholders have a major stake in the goal of equity in health and development. The four major categories of participants include: researchers, decision makers, health service providers and communities.
How are research ideas identified	Stakeholders suggest priority areas, via evidence based situation analysis (such as looking at health status, health care system, health research system). Research ideas are gathered from a nomination process from different stakeholders. Consensus building using methods such as brainstorming, multi–voting, nominal group technique, round–table is then used to select research ideas.
Scoring criteria	Criteria is selected as to be: – Appropriate to the level of the action of the priority setting *i.e.* global, national, district; – Detailed in definition; – Independent of each other; – Contain information base; – Reflect equity promotion and development; – Manageable number; – Expressed in a common language. Criteria are agreed on by brainstorming of large collection of possible criteria, eliminating duplicates and clearly defining the meaning of each criterion from stakeholders. Criteria will then be put into representative categories and finally selected depending on purpose and level of action of priority–setting exercise.
Scoring options	Each criteria is scored: Point score to each criteria OR Number of score choices to each criteria
Advantages	– Broad based inclusion and participation of different stakeholders. – Multidisciplinary and cross–sectoral approach – Partnership development – Transparent process – Systematic analyses of health needs
Disadvantages	– Vague criteria and lack of transparency in individual process used by countries – Few countries had guidelines on how to develop nor apply criteria – Needs stronger representation of groups such as private sector, parliamentarians, donors, international agencies– Does not provide methodology for identifying participants

Table 2. Brief explanation of the Combined Approach Matrix (CAM) [7,8]

Overall process	Developed by the Global Forum for Health Research, CAM was to bring together economic and institutional dimensions into an analytical tool with the actors and factors that play a key role in health status of a population. It also aims to organise and present a large body of information that enters the priority setting process. This will help decision makers make rational choices in investment to produce greatest reduction in burden of disease.
How are participants identified?	Institutional approach involving: individual, household and community; health ministry and other health institutions; other sectors apart from health; and macroeconomic level actors.
How are research ideas identified	Five step process including measuring the disease burden, analysing determinants, getting present level of knowledge, evaluating cost and effectiveness, and present resource flows. For each main disease and risk factor, institutions and stakeholders with particular knowledge are brought together to provide information via workshops and brainstorming. Each institution will feed into matrix the information at disposal, regarding a specific disease or factor; the matrix will reveal how little information is available in some areas which can then be candidates for research. Each participant determined the priority research topics based on CAM evidence, then grouping the topics and cutting down to establish the top priorities.
Scoring criteria	Criteria based on questions of what is a research priority in the context, and what is not known but should be.
Scoring options	N/A
Advantages	– Creates framework of information – Identifies gaps in knowledge – Facilitates comparisons between sectors – Broad inclusion of actors – 3D–CAM includes equity
Disadvantages	– Difficult and time–consuming as involves multi–stage discussion – Does not provide algorithm to establish and score research priorities therefore is not repeatable nor systematic – Does not provide methodology for identifying participants

Table 3. Brief explanation of the James Lind Alliance Method [9]

Overall process	Focuses on bringing patients, carers and health professionals in order to identify treatment uncertainties which will become research questions. The method uses a mixture of data gathering, quantitative and qualitative analysis to create research priorities in areas of treatment uncertainty.
How are participants identified?	Participants are identified through Priority Setting Partnerships which brings patients, carers and clinicians equally together and agree through consensus priorities.
How are research ideas identified	Treatment uncertainties are defined as no up to date, reliable systematic reviews addressing treatment uncertainty, or systematic review that shows such uncertainty exists. Step 1: Recommendations by PSPs, or through looking at existing literature, creates a list of uncertainties. Step 2: These are then verified through systematic reviews of databases to verify they are research gaps using Cochrane, DARE, NICE, Sign. An uncertainty is deemed genuine when a reported confidence interval in a systematic review does not cross the line of effect or line of unity. A virtual interim priority ranking, and a final priority setting workshop takes place to agree upon 10 prioritised uncertainties through consensus building.
Scoring criteria	No clear criteria are identified with which to use.
Scoring options	Ranked AND Qualitative consensus
Advantages	– Takes into account underrepresented groups – Applicable to small scale prioritisation (eg, hospital) – Mixture of methods
Disadvantages	– Time consuming to identify and verify treatment uncertainties – Selection of criteria not clear – Not suitable for global level, nor specific disease domains – Very clinically orientated – Disproportionate mix of participants may skew information base

Table 4. Brief explanation of the Council on Health Research for Development (COHRED) [10]

Overall process	COHRED uses a management process for national level exercises to show important steps for priority setting processes.
How are participants identified?	Participants are identified through the chosen methods outlined in the steps of the COHRED guide.
How are research ideas identified	Identification of priority issues much choose method best suited to local context and needs either through compound approaches (ENHR, CAM, Burden of Disease) or foresighting techniques (Visioning, Delphi). Consider using more than one method to optimize usefulness of results.
Scoring criteria	COHRED presents ranking techniques that can be used to rank priority issues including direct and indirect valuation techniques.
Scoring options	Ranked
Advantages	– Overview approach providing steps – Discusses wide range of options – Flexible to contexts and needs
Disadvantages	– Too general and unspecific – Lack of criteria transparency

and 28% were focused on Low and Middle Income Countries (LMIC). At the national level, the countries where research priority exercises were most frequently initiated were the UK (27%), USA (16%), Australia (15%), and Canada (11%) (**Table 7**).

Topic areas for which research priorities were identified included non–communicable diseases (18%), followed by child and adolescent health (17%), mental health (10%), nursing/midwifery (8%) and infectious disease (8%). The remaining exercises (39%) covered a wide variety of topics, including policy and health system, occupational health /therapy, reproductive health/women's health, emergency care, environmental health, occupational health, forensic science and injury prevention (**Table 7**).

DISCUSSION

The number of priority setting exercises in health research published in PubMed–indexed journals is increasing, especially since 2010. These exercises are being conducted at a variety of different levels, ranging from the global level to the level of an individual hospital. With the development of new tools and methods which have a well–defined structure – such as the CHNRI method, James Lind Alliance Method and Combined Approach Matrix – it is likely that the Delphi method and non–replicable consultation processes (see the definition of "replicable" earlier in the text) will gradually be replaced by these emerging tools, which offer more transparency and replicability. This is a process that should be endorsed, as a natural progression of the

Table 5. Brief explanation of the Delphi Process [11]

Overall process	Delphi, mainly developed in the 1950s, is a systematic, interactive forecasting method which relies on a panel of experts and questionnaires.
How are participants identified?	Participants are eligible to be invited if they have related backgrounds and experiences concerning the target issue, are capable of contributing, and are willing to revise their initial judgements in order to reach consensus. Participants are considered and selected through investigators, ideally through a nomination process, or selection from potential leaders or authors through publication. It is suggested that the three groups are used: top management decision makers who will utilise outcomes of Delphi study; professional staff members and their support team; respondents to the Delphi questionnaire. It is recommended to use the minimally sufficient number to generate representative pooling of judgements – however no consensus yet as to optimal number of subjects.
How are research ideas identified	In the first round an open–ended questionnaire is sent to solicit information about a content area from Delphi participants. Investigators will then turn the responses into a well–structured questionnaire to be used as survey for data collection. Through four rounds experts answer questionnaires; the facilitator summarises anonymously the forecast after the first round and the experts are then asked to revise their earlier answer thereby decreasing the range of answers and converging towards the correct answer. Up to four iterations can be used.
Scoring criteria	N/A
Scoring options	Rate or ranking AND Consensus building
Advantages	– Multiple iterations and feedback process – Flexible to change – Anonymity of respondents
Disadvantages	– Does not provide methodology for identifying participants – Lack of criteria transparency – Potential for low response rate due to multiple iterations – Time–consuming – Potential for investigators and facilitators to bias opinions

Table 6. Brief explanation of the CHNRI process [12–15]

CHNRI method Child Health Nutrition Research Initiative	
Overall process	The CHNRI methodology was introduced in 2007 by the Child Health and Nutrition Research Initiative of the Global Forum for Health research. The methodology was developed to address gaps in the existing research priority methods. The CHNRI method is developed to assist decision making and consensus development. The method include soliciting ideas from different carder of participants on the given health topic and use independent ranking system against the pre–defined criteria to prioritise the research ideas.
How are participants identified?	Participants are identified by management team based on their expertise (eg, number of publications, experience in implementation research and programmes etc). Participants includes stakeholders who might not have the technical expertise but have view on the health topic of concern.
How are research ideas identified?	Research ideas are generated by participants or by management team based on the current evidence. If former, usually each participant is asked to provide maximum of three research questions against the predefined domain of health research (eg, descriptive research, development research, discovery research and delivery research). The ideas are usually submitted via online survey and consolidated by the management team.
Scoring criteria	Five standard criteria are usually used: – Answerability – Equity – Impact on burden – Deliverability – Effectiveness. Though the five standard criteria are used in more than 70% of the research priority setting exercises, the method offers optional criteria to be used to replace the standard criteria depending on the needs and context of the exercises. For example, criteria such as low cost, sustainability, acceptability, feasibility, innovation and originality are used to replace or in addition to the standard criteria.
Scoring options	Each criteria is scored: Point score to each criteria in the scale of 0, 0.5 and 1 or in the scale of 0 to 100.
Advantages	– Simple, inclusive and replicable and thus systematic and transparent process. – Independent ranking of experts (avoid having the situation where one strongly minded individual affecting the group decision) – Less costly
	– Potentially represent collective opinion of the limited group of people who were included in the process. – Scoring affected by currently on–going research

Table 7. Distribution of identified studies by geographic context and countries where the research priority setting exercises have been initiated and research priority areas addressed

Geographical area	Number	%	Technical areas	Number	%
Global	35	21	Non–communicable disease	29	18
High income countries	82	50	Child and adolescent health	28	17
Low middle income countries	47	28	Mental health	16	10
Humanitarian settings	1	<1	Infectious disease	14	8
TOTAL	**165**	**100**	Nursing/Midwifery	13	8
National level			Public health in general	10	6
Australia	15	15	Policy and health system	8	5
Brazil	1	1	Occupational health/therapy	6	4
Canada	11	11	Reproductive health/women's health	6	4
Colombia	1	1	Skin disease	5	3
Chile	1	1	Emergency care	3	2
Cuba	1	1	Environmental health	3	2
Hong Kong	2	2	Disability	3	2
India	1	1	Child development potential	2	1
Iran	2	2	Injury prevention	2	1
Ireland	3	3	Maternal and perinatal health	2	1
Italy	1	1	Pharmaceuticals	2	1
Malaysia	1	1	Microbial Forensics	2	1
Nepal	1	1	Behavioural science	1	1
The Netherlands	1	1	Diagnostic accuracy	1	1
Nigeria	1	1	Tuberculosis	1	1
Peru	1	1	Medical science	1	1
Portugal	2	2	Neurological	1	1
South Africa	3	3	Nutrition	1	1
Saudi Arabia	1	1	Surgical	1	1
Spain	3	3	Surveillance system	1	1
United Republic of Tanzania	2	2	Water and sanitation	1	1
United Kingdom	26	27	Primary health care–related disease	1	1
United States of America	16	16	Others	1	1
TOTAL	**97**	**100**	**TOTAL**	**165**	**100**

priority–setting field from the period in which hardly any structured processes existed to fill a need, to the new era which will be increasingly dominated by structured and well–defined tools.

This review is not the first attempt to assess approaches, tools and methods to set health research priorities. Searching the literature, I identified five earlier attempts to review and discuss priority–setting processes. The first review was published by Rudan and colleagues in 2007 in an attempt to develop an evidence base for the development of conceptual framework and guidelines for implementation of the CHNRI methodology [1]. This paper identified ambitious attempts by several large organizations at the international level to define health research priorities for either the whole developing world, large world regions or nationally. These attempts date back to the year 1990, with the "... Commission on Health Research for Development usually being referred to as the first truly significant international initiative aimed toward systematic approach to setting priorities in global health research." Other initiatives that followed were the "Ad Hoc Committee (AHC) on Health Research Relating to Future Intervention Options" (in 1994), the "Global Forum for Health Research" (in 1998), the "Council on Health Research and Development (COHRED)" (in 2000), "The Grand Challenges" proposed at the World Economic Forum in Davos, Switzerland (in 2003) and the "Combined Approach Matrix" as the first specific priority–setting tool for health research (in 2004). The paper concluded that the processes, initiatives and tools fell short of being informative on what the specific research priorities should be and how exactly are they derived [1].

In 2010, Viergever et al. [16] reviewed the articles that set health research priorities and they specifically reviewed exercises coordinated by World Health Organization Headquarters since 2005. This resulted in the total of 230 documents or reports, many of them unpublished (hence, not included in my review). The authors concluded that, at that point in time, there was no "gold standard" approach for health research prioritisation. This was not surprising, given the heterogeneity in the context of research prioritization exercises and different levels at which they were being conducted. Nevertheless, the authors observed several

common themes of "good practice" and proposed a generic framework – in the form of "checklist", like a form of "guidelines" – which also suggested various options for each step of the process. Nine themes were identified through a review of the previously conducted priority–setting processes. They were categorized as the "themes during the preparatory work" (defining context, use of comprehensive approach, ensure inclusiveness of participants, information gathering, planning for implementation), followed by the steps in the process of deciding on the priorities (defining the criteria, methods for deciding on priorities), and two steps in the last phase after the priorities have been set (plan the timing of evaluation in terms of how the research priorities are being used, and write the clear report of the methodology used to ensure the transparency in the process). The authors proposed that the provision of the framework should be of assistance to policy makers and researchers. It could have a dual role: it could not only assist priority–setting process, but also planning the follow up and implementation of the priorities [16].

In the same year, in 2010, the World Health Organization's Department for Research Policy and Cooperation held a consultation between methodology–developing experts to identify optimal characteristics of priority–setting methods that could be applicable at the national level. The aim was to empower low and middle–income countries to take more ownership of their own health research agenda. Tomlinson reviewed the progress made at this meeting and published the main conclusions in 2011 [2]. Three methods emerged as applicable at the national level: the Combined Approach Matrix (CAM), the Council on Health Research and Development (COHRED) and the Child Health and Nutrition Research Initiative (CHNRI). The authors presented and discussed strengths and weaknesses of each method [2]. They also noted that, across the countries surveyed, genuine engagement of stakeholders was difficult to achieve and was typically missing. Countries also varied in the extent to which they would document priority–setting processes, with not a single country having an appeal process for outlined priorities. Another problem was that the identified priorities usually outlined broad disease categories, rather than more specific research questions [2]. The authors concluded that priority–setting processes should aim to include mechanisms for publicizing results, effective procedures to translate and implement decisions and processes to ensure that the revision of priorities eventually does occur.

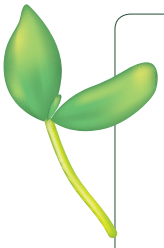
In a more recent report, an independent team from the Kirby Institute in Sydney, Australia, systematically reviewed all studies undertaken in low– and middle–income country (LMIC) settings that attempted to set research priorities over the period from 1966 to 2014. The studies included were not reported but they found 91 studies, including 16 which used the CHNRI method [17]. The authors concluded that almost half of these processes took place at the global level (46%). For regional or national initiatives, a half focused on Sub Saharan Africa (49%), followed by East Asia and the Pacific (20%) and Latin America and the Caribbean (18%). Most commonly, studies were initiated by an international organization or collaboration (46%). Researchers and governments were the most commonly represented stakeholders. The most frequently used process was a conference or workshop to determine priorities (24%), followed by the CHNRI method (18%) [17]. The review revealed inconsistent use of existing methods and approaches in health research prioritization processes. It also showed that while there was strong involvement of government and researchers, participation of other key stakeholders was limited. The authors argued that many processes, regardless of the method used, lacked an implementation strategy to translate the result of the process into implementation of research projects. Finally, the authors concluded that research prioritization exercises would often remain "one–time exercises", given the lack of follow up and implementation strategies involving the funders, researchers and government officials.

Finally, in 2014, as a part of the Lancet series on increasing value and reducing waste in health research generally, one paper of the series (by Chalmers et al. [18]) explored how to increase value and reduce waste when research priorities are set. The group of authors argued that many basic research endeavours do not lead to knowledge that is useful to the end user of the research results. By using various examples, the authors reiterate the same argument: if research does not meet the needs of the users of research, evidence will have little impact on public health and clinical practice. The authors argue that many research studies that fall in the area of basic (fundamental) research were duplicative. Although a replication of positive findings is a welcome process, an excessive repetition of conducting similar research can be prevented by either: (i) conducting systematic reviews and also involving the end user of the research as well as clinicians in the process (where they used the example of hospital based research priority setting exercise using the James Lind Alliance method); and (ii) mapping research portfolios of major agencies, that could help to prevent duplication in the nature of supported research. The main message of the article is, therefore, a need for better co–ordination among the researchers and the funders over the research that is being conducted and increased focus on the translational value of the information that is being generated through research [18].

It is evident from my own methodical review, and from the systematic review undertaken by the researchers from the Kirby Institute, that there is a need for a transparent, replicable, systematic and structured approach to research pri-

ority setting, because the large majority of the previous exercises were not based on processes meeting all of these criteria. The review by McGregor et al. [17] shows how, although a very recent addition to the set of tools, the CHNRI method is set to become the most widely used approach.

The results of my review broadly confirmed the observations of all previous reviews, with an additional insight into time trend – showing an increase in the number of exercises conducted over time, and gradual replacement of poorly defined processes with those that use particular methods and tools, as shown in **Figure 2**. The next step in the field of health research priority setting should therefore involve monitoring whether any single method may address the need for most exercises conducted at different levels, or if better results may perhaps be achieved through combination of strengths of several methods.



Acknowledgements: The author thanks James Gao, an intern at the World Health Organization, for his assistance to the development of Tables 1–6. She also acknowledges helpful input from her mentor, Professor Simon Cousens from the London School of Hygiene and Tropical Medicine.

Disclaimer: SY is an employee of the World Health Organization. The views expressed in this paper are the responsibility of the author and do not necessarily represent the views of the World Health Organization.

Funding: None.

Ethical approval: Not required.

Authorship declaration: SY wrote the paper as a part of her PhD thesis.

Declaration of interest: The author completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author). The author reports no conflicting financial or other interest related to the work detailed in this manuscript.

REFERENCES

1 Rudan I, Gibson J, Kapiriri L, Lansang MA, Hyder AA, Lawn J, et al; Child Health and Nutrition Research Initiative (CHNRI). Setting priorities in global child health research investments: Assessment of principles and practice. Croat Med J. 2007;48:595-604. Medline:17948946

2 Tomlinson M, Chopra M, Hoosain N, Rudan I. A review of selected research priority setting processes at national level in low and middle income countries: towards fair and legitimate priority setting. Health Res Policy Syst. 2011;9:19. Medline:21575144 doi:10.1186/1478-4505-9-19

3 Rudan I, Kapiriri L, Tomlinson M, Balliet M, Cohen B, Chopra M. Evidence–based priority setting for health care and research: tools to support policy in maternal, neonatal, and child health in Africa. PLoS Med. 2010;7:e1000308. Medline:20644640 doi:10.1371/journal.pmed.1000308

4 Council on Health Research for Development (COHRED). Essential national health research and priority setting: Lessons learned; 1997. Available: http://www.cohred.org/downloads/586.pdf. Accessed: 21 August 2015.

5 Okello D, Chongtrakul P; The COHRED Working Group on Priority Setting. A manual for research priority setting using the ENHR Strategy; 2000. http://www.cohred.org/downloads/578.pdf. Accessed: 21 August 2015.

6 The Working Group on Priority Setting. Priority setting for health research: lessons from developing countries. Health Policy Plan. 2000;15:130-6. Medline:10837035 doi:10.1093/heapol/15.2.130

7 Ghaffar A, Collins T, Matlin SA, Olifson S. The 3D combined approach matrix: An improved tool for setting priorities in research for health. The Global Forum for Health Research. 2009. Available: http://www.webcitation.org/query.php?url=http://www.globalforumhealth.org/content/download/7860/50203/file/CAM_3D_GB.pdf&refdoi=10.1186/1478–4505–8–36. Accessed: 21 August 2015.

8 Ghaffar A, de Francisco A, Matlin S. The Combined Approach Matrix: a priority–setting tool for health research. Geneva: Global Forum for Health Research; 2004. Available: http://www.webcitation.org/query.php?url=http://www.globalforumhealth.org/Media–Publications/Publications/The–Combined–Approach–Matrix–A–priority–setting–tool–for–health–research&refdoi=10.1186/1478–4505–8–36. Accessed: 21 August 2015.

9 Cowan K, Oliver S. The James Lind Alliance Guidebook (version5). The James Lind Alliance; 2013. Available: http://www.jlaguidebook.org/pdfguidebook/guidebook.pdf. Accessed: 21 August 2015.

10 Montorzi G, Haan S, IJsselmuiden C. Priority setting for research for health: A Management process for countries. Council on Health Research for Development (COHRED); 2010. Available: http://www.cohred.org/downloads/Priority_Setting_COHRED_approach_August_2010.pdf. Accessed: 21 August 2015.

11 Hsu C, Sandford B. The Delphi technique: Making sense of consensus. Pract Assess, Res Eval. 2007;12:1-8.

12 Kapiriri L, Tomlinson M, Gibson J, Chopra M, El Arifeen S, Black RE, et al. Child Health and Nutrition Research Initiative (CHNRI): Setting priorities in global child health research investments: Addressing the values of the stakeholders. Croat Med J. 2007;48:618-27. Medline:17948948

REFERENCES

13 Rudan I, Chopra M, Kapiriri L, Gibson J, Lansang MA, Carneiro I, et al. Setting priorities in global child health research investments: universal challenges and conceptual framework. Croat Med J. 2008;49:307-17. Medline:18581609 doi:10.3325/cmj.2008.3.307

14 Rudan I, Gibson JL, Ameratunga S, El Arifeen S, Bhutta ZA, Black M, et al; Child Health and Nutrition Research Initiative (CHNRI). Setting Priorities in Global Child Health Research Investments: Guidelines for implementation of the CHNRI Method. Croat Med J. 2008;49:720-33. Medline:19090596 doi:10.3325/cmj.2008.49.720

15 Rudan I, El Arifeen S, Black RE, Campbell H. Childhood pneumonia and diarrhoea: Setting our priorities right. Lancet Infect Dis. 2007;7:56-61. Medline:17182344 doi:10.1016/S1473-3099(06)70687-9

16 Viergever RF, Olifson S, Ghaffar A, Terry RF. A checklist for health research priority setting: nine common themes of good practice. Health Res Policy Syst. 2010;8:36. Medline:21159163

17 McGregor S, Henderson KJ, Kaldor JM. How are health research priorities set in low and middle income countries? A systematic review of published reports. PLoS ONE. 2014;9:e108787. Medline:25275315 doi:10.1371/journal.pone.0108787

18 Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet. 2014;383:156-65. Medline:24411644 doi:10.1016/S0140-6736(13)62229-1

Chapter 2. Example of the application of the CHNRI method

Chapter A provided an overview of various methods and approaches used in health research prioritisation between 2001 and 2014. The number of health research prioritisation exercises listed in PubMed has been increasing since 2010. The 165 health research prioritisation exercises identified in the review were conducted at global, regional, country or individual hospital level. Approximately one in four exercises used the CHNRI method, followed by consensus building methods including the Delphi method and consultations.

The comparative advantage of the CHNRI method is that the method is simple, inclusive and replicable. Independent scoring by experts is expected to avoid situations in which one strongly minded individual dominates the group decision. It is also less costly compared to consultation where large number of experts are physically invited to a meeting to reach group consensus.

Having reviewed various health research prioritisation methods and learned about the CHNRI process, I co-ordinated two global health research prioritisation exercises to gain hands on experience with the CHNRI method. In both research prioritisation exercises, I provided methodological supports in the process. In this chapter, I will present two global research prioritisation exercises in which I used the CHNRI methodology. This chapter also reflects on strengths and limitation of the CHNRI method based on this experience.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed for each 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? Journal of Global Health

1.2. When was the work published? 2016

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes** **No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I have coordinated the research prioritization process, conducted data analysis, and wrote the first draft of the paper.

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE Sachiyo Yoshida Date 25 July 2018

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

Newborn health research priorities beyond 2015

Published Online
May 20, 2014
http://dx.doi.org/10.1016/
S0140-6736(14)60263-4

In 2012, an estimated 2·9 million newborn babies died[1] and 2·6 million were stillborn in 2009.[2] An even greater number have long-term impairment associated with preterm birth, intrauterine growth restriction, congenital anomalies, and intrapartum or infectious insults. Despite the increasing proportion of child deaths that are neonatal—estimated at 44% at present—programme and research funding is modest.[3] In view of the Millennium Development Goal (MDG) deadline in 2015 and the shift to a new framework targeting the unfinished survival agenda and beyond, including healthy development, growth, and human capital, there is increased attention to birth outcomes as highlighted in the *Lancet* Every Newborn Series[3-7] and the upcoming Every Newborn Action Plan. Research priorities are required for this wider agenda and longer timeframe.

In 2007–08, WHO held a series of exercises to set global research priorities to reduce mortality among newborn babies and children until 2015.[8-12] In 2013, a new priority setting process was initiated for the post-MDG era, initially to 2025, regarding maternal, newborn, child, and adolescent health. As part of this initiative, the global exercise to set research priorities for newborn health was coordinated by WHO and Saving Newborn Lives/Save the Children, with support from the Bill & Melinda Gates Foundation.

We adapted and used the Child Health and Nutrition Research Initiative (CHNRI) method.[13] The CHNRI process is transparent, replicable, and feasible for online application and has been used for many exercises varying from mental health to primary care.[14] We identified and approached 200 of the most productive researchers in the field in the past 5 years and 400 programme experts, and 132 of them submitted their three best research ideas online. Ideas were collated into a set of 205 research questions, and sent for scoring to the 600 experts originally approached. The 205 research questions were scored against five predefined criteria (answerability, efficacy, deliverability, impact, and equity) by 91 responding experts. Research priority scores were then computed as the mean of the aggregated scores to identify priorities in the three domains of research: delivery, development, and discovery.

Nine of the ten top-ranked priorities were in the domain of delivery (table), exploring how to take effective interventions to every mother and every newborn baby. Research priority scores ranged from 79% to 90%, and the interscorer variability analyses showed a high level of agreement (65–77%). The top delivery research priorities included identifying approaches to scale up simplified newborn resuscitation at lower levels of the health system, identification and management of newborn infection at community level, addressing barriers in the scaling up of exclusive breastfeeding

	Research priorities	Score
Delivery domain		
1	Can a simplified neonatal resuscitation programme delivered by trained health workers reduce neonatal deaths due to perinatal asphyxia?	90
2	How can health workers' skills in preventing and managing asphyxia be scaled up?	88
3	Can simple clinical algorithms used by community health workers identify and refer neonates with signs of infection and consequently reduce newborn mortality?	86
4	How can exclusive breastfeeding in low-resource contexts be promoted to reduce neonatal infections and mortality?	85
5	Can training of community health workers in basic newborn resuscitation reduce morbidity and mortality due to perinatal asphyxia?	83
6	How can the administration of injectable antibiotics at home and first-level facilities to newborns with signs of sepsis be scaled up to reduce neonatal mortality?	82
7	How can facility-based initiation of kangaroo mother care or continuous skin-to-skin contact be scaled up?	80
8	How can chlorhexidine application to the cord be scaled up in facility births and in low neonatal mortality rate settings to reduce neonatal infections and neonatal mortality?	80
9	How can quality of care during labour and birth be improved to reduce intrapartum stillbirths, neonatal mortality, and disability?	79
10	Can community-based extra care for preterm/low birthweight babies delivered by community health workers reduce neonatal morbidity and mortality in settings with poor access to facility care?	79
Development domain		
1	Can community-based initiation of kangaroo mother care reduce neonatal mortality of clinically stable preterm and low birthweight babies?	82
2	How can the accuracy of community health workers in detecting key most important high-risk conditions or danger signs in pregnant women be improved?	77
3	Can perinatal audits improve quality of care in health facilities and improve fetal and neonatal outcomes?	74
4	Can intrapartum monitoring to enhance timely referral improve fetal and neonatal outcomes?	74
5	Can training community health workers to recognise and treat neonatal sepsis at home with oral antibiotics when referral is not possible reduce neonatal mortality?	74
Discovery domain		
1	Can stable surfactant with simpler novel modes of administration increase the use and availability of surfactant for preterm babies at risk of respiratory distress syndrome?	71
2	Can the method to diagnose fetal distress in labour be made more accurate and affordable?	66
3	Can strategies for prevention and treatment of intrauterine growth restriction be developed?	64
4	Can novel tocolytic agents to delay or stop preterm labour be developed in order to reduce neonatal mortality and morbidity?	63
5	Can major causal pathways and risk factors for antepartum stillbirth be identified?	61

Overall and criterion specific scores ranged from 0% to 100%.

***Table*: Research priorities for improving newborn health and birth outcomes by 2025 as ranked by 91 experts**

and facility-based kangaroo mother care, evaluating chlorhexidine cord cleansing for neonates born in health facilities, and developing strategies to improve the quality of facility-based care during labour and childbirth.

In the domain of development to improve existing interventions, the overall research priority scores ranged from 74% to 82%, with moderate to high agreement between scorers (57–64%). The top ranked priorities included evaluating the impact and safety of kangaroo mother care initiated at the community level, early detection of high-risk women in pregnancy and labour, improved and simplified intrapartum monitoring, evaluation of appropriate oral antibiotics for treatment of neonatal sepsis, and the role of perinatal audits in improving quality of care during labour and childbirth.

Discovery research priorities emphasised the need to invest in science and technology to expand the arsenal of effective interventions. Overall research priority scores ranged from 61% to 71% and agreement scores from 43% to 49%. The highest priorities in this domain were to discover causal pathways of preterm labour, new tocolytics to delay preterm birth, stable surfactant with easier mode of delivery, and to discover more accurate and affordable ways to detect fetal distress. These research priorities align with solution pathways for understanding the biological basis of preterm birth and devising new methods of prevention.[15]

Large inequities exist in present research funding for newborn health as compared with other diseases globally, and also between different neonatal disorders themselves. Disorders that affect newborn babies in high-income countries receive more funding and attention than those affecting newborn babies in low-income countries. For instance, research on care for preterm babies in neonatal intensive care units has received substantially more funding[16] in comparison with intrapartum-related birth outcomes.

In coming years, the newborn health research agenda should be placed at the forefront of efforts to reduce global under-5 child mortality and improve human capital. The results described here will assist both donors and researchers in setting evidence-based priorities to address the key gaps in knowledge that could make the most difference in saving newborn lives, preventing stillbirth, and other birth outcomes.

We challenge the many partners linked to the Every Newborn Action Plan, including governments, non-governmental organisations, research institutes, and donors, to ensure that the top ranked priorities are evaluated and inform accelerated progress around the world for every women, every newborn baby, and every child.

See **Series** pages 174 and 189

See **Online/Series**
http://dx.doi.org/10.1016/S0140-6736(14)60792-3,
http://dx.doi.org/10.1016/S0140-6736(14)60582-1, and
http://dx.doi.org/10.1016/S0140-6736(14)60750-9

*Sachiyo Yoshida, Igor Rudan, Joy E Lawn, Stephen Wall, João Paulo Souza, José Martines, *Rajiv Bahl, and members of the neonatal health research priority setting group*
Department of Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva 1211, Switzerland (SY, RB); Centre for Population Health Sciences and Global Health Academy, University of Edinburgh Medical School, Edinburgh, UK (IR); MARCH (Maternal Reproductive & Child Health), London School of Hygiene & Tropical Medicine, London, UK (JEL); Saving Newborn Lives, Save the Children, Washington, DC, USA (JEL, SW); Department of Social Medicine, Ribeirão Preto School of Medicine, University of São Paulo, Brazil (JPS); and Centre for Intervention Science in Maternal and Child Health, Centre for International Health, University of Bergen, Norway (JM)
bahlr@who.int

We declare no competing interests. SY and RB are employees of WHO; the views expressed in this paper are the responsibility of the authors and do not necessarily represent the views of WHO. See appendix for full list of members of the neonatal health research priority setting group.

1 You D, Bastian P, Wu J, Wardlaw T. Level and trends in child mortality. Estimates developed by the UN Inter-agency Group for Child Mortality Estimation. Report 2013. New York: UNICEF, WHO, The World Bank, and United Nations, 2013. http://www.childmortality.org/ (accessed Dec 18, 2013).

2 Cousens S, Blencowe H, Stanton C, et al. National, regional, and worldwide estimates of stillbirth rates in 2009 with trends since 1995: a systematic analysis. *Lancet* 2011; **377:** 1319–30.

3 Darmstadt GL, Kinney MV, Chopra M, et al, for The Lancet Every Newborn Study Group. Who has been caring for the baby? *Lancet* 2014; published online May 20. http://dx.doi.org/10.1016/S0140-6736(14)60458-X.

4 Lawn JE, Blencowe H, Oza S, et al, for The Lancet Every Newborn Study Group. Progress, priorities, and potential beyond survival. *Lancet* 2014; published online May 20. http://dx.doi.org/10.1016/S0140-6736(14)60496-7.

5 Bhutta ZA, Das JK, Bahl R, et al, for The Lancet Newborn Interventions Review Group and The Lancet Every Newborn Study Group. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? *Lancet* 2014; published online May 20. http://dx.doi.org/10.1016/S0140-6736(14)60792-3.

6 Dickson KE, Simen-Kapeu A, Kinney MV, et al, for The Lancet Every Newborn Study Group. Health-systems bottlenecks and strategies to accelerate scale-up in countries. *Lancet* 2014; published online May 20. http://dx.doi.org/10.1016/S0140-6736(14)60582-1.

7 Mason E, McDougall L, Lawn JE, et al, for The Lancet Every Newborn Study Group, on behalf of the Every Newborn Steering Committee. From evidence to action to deliver a healthy start for the next generation. *Lancet* 2014; published online May 20. http://dx.doi.org/10.1016/S0140-6736(14)60750-9.

8 Fontaine O, Kosek M, Bhatnagar S, et al. Setting research priorities to reduce global mortality from childhood diarrhoea by 2015. *PLoS Med* 2009; **6:** e41.

9 Rudan I, El Arifeen S, Bhutta ZA, et al. Setting research priorities to reduce global mortality from childhood pneumonia by 2015. *PLoS Med* 2011; **8:** e1001099.

10 Lawn JE, Bahl R, Bergstrom S, et al. Setting research priorities to reduce almost one million deaths from birth asphyxia by 2015. *PLoS Med* 2011; **8:** e1000389.

11 Bahl R, Martines J, Ali N, et al. Research priorities to reduce global mortality from newborn infections by 2015. *Pediatr Infect Dis J* 2009; **28:** S43–S48.

12 Bahl R, Martines J, Bhandari N, et al. Setting research priorities to reduce global mortality from preterm birth and low birth weight by 2015. *J Glob Health* 2012; **2:** 010403.

13 Rudan I, Gibson JL, Ameratunga S, et al. Setting priorities in global child health research investments: guidelines for implementation of the CHNRI method. *Croat Med J* 2008; **49:** 720–33.

14 Rudan I. Global health research priorities: mobilizing the developing world. *Public Health* 2012; **126:** 237–40.

15 Lackritz E, Wilson CB, Guttmacher AE, et al. A solution pathway for preterm birth: accelerating a priority research agenda. *Lancet Glob Health* 2013; **1:** e328–30.

16 Howson CP, Kinney MV, Lawn JE. Born too soon: the global action report on preterm birth. Geneva: March of Dimes, PMNCH, Save the Children, World Health Organization, 2012.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? The Lancet

1.2. When was the work published? May 2014

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I led data analysis and wrote first draft of the method and result section.

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE Sachiyo Yoshida Date 25 July 2018

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

Electronic supplementary material:
The online version of this article contains supplementary material.

journal of global health

PAPERS

Setting research priorities to improve global newborn health and prevent stillbirths by 2025

Sachiyo Yoshida[1], José Martines[2], Joy E Lawn[3,4], Stephen Wall[4], João Paulo Souza[5], Igor Rudan[6], Simon Cousens[3]; The neonatal health research priority setting group; Peter Aaby[7], Ishag Adam[8], Ramesh Kant Adhikari[9], Namasivayam Ambalavanan[10], Shams El Arifeen[11], Dhana Raj Aryal[12], Sk Asiruddin[13], Abdullah Baqui[14], Aluisio JD Barros[15], Christine S Benn[16], Vineet Bhandari[17], Shinjini Bhatnagar[18], Sohinee Bhattacharya[19], Zulfiqar A Bhutta[20], Robert E Black[21], Hannah Blencowe[22], Carl Bose[23], Justin Brown[24], Christoph Bührer[25], Wally Carlo[26], Jose Guilherme Cecatti[27], Po–Yin Cheung[28], Robert Clark[29], Tim Colbourn[30], Agustin Conde–Agudelo[31], Erica Corbett[32], Andrew E Czeizel[33], Abhik Das[34], Louise Tina Day[35], Carolyn Deal[36], Ashok Deorari[37], Uğur Dilmen[38], Mike English[39], Cyril Engmann[40], Fabian Esamai[41], Caroline Fall[42], Donna M Ferriero[43], Peter Gisore[44], Tabish Hazir[45], Rosemary D Higgins[46], Caroline SE Homer[47], DE Hoque[48], Lorentz Irgens[49], MT Islam[50], Joseph de Graft–Johnson[51], Martias Alice Joshua[52], William Keenan[53], Soofia Khatoon[54], Helle Kieler[55], Michael S Kramer[56], Eve M Lackritz[57], Tina Lavender[58], Laurensia Lawintono[59], Richard Luhanga[60], David Marsh[61], Douglas McMillan[62], Patrick J McNamara[63], Ben Willem J Mol[64], Elizabeth Molyneux[65], G. K Mukasa[66], Miriam Mutabazi[67], Luis Carlos Nacul[68], Margaret Nakakeeto[69], Indira Narayanan[70], Bolajoko Olusanya[71], David Osrin[72], Vinod Paul[73], Christian Poets[74], Uma M Reddy[75], Mathuram Santosham[76], Rubayet Sayed[77], Natalia E Schlabritz–Loutsevitch[78], Nalini Singhal[79], Mary Alice Smith[80], Peter G Smith[81], Sajid Soofi[82], Catherine Y Spong[83], Shahin Sultana[84], Antoinette Tshefu[85], Frank van Bel[86], Lauren Vestewig Gray[87], Peter Waiswa[88], Wei Wang[89], Sarah LA Williams[90], Linda Wright[91], Anita Zaidi[92], Yanfeng Zhang[93], Nanbert Zhong[94], Isabel Zuniga[95], Rajiv Bahl[1]

Affiliations of authors are listed at the end of the manuscript.

Correspondence to:
Dr Rajiv Bahl
Department of Maternal, Newborn,
Child and Adolescent Health
World Health Organization
Geneva, Switzerland
20 Avenue Appia 1211
Geneva–27, Switzerland
bahlr@who.int

Background In 2013, an estimated 2.8 million newborns died and 2.7 million were stillborn. A much greater number suffer from long term impairment associated with preterm birth, intrauterine growth restriction, congenital anomalies, and perinatal or infectious causes. With the approaching deadline for the achievement of the Millennium Development Goals (MDGs) in 2015, there was a need to set the new research priorities on newborns and stillbirth with a focus not only on survival but also on health, growth and development. We therefore carried out a systematic exercise to set newborn health research priorities for 2013–2025.

Methods We used adapted Child Health and Nutrition Research Initiative (CHNRI) methods for this prioritization exercise. We identified and approached the 200 most productive researchers and 400 program experts, and 132 of them submitted research questions online. These were collated into a set of 205 research questions, sent for scoring to the 600 identified experts, and were assessed and scored by 91 experts.

Results Nine out of top ten identified priorities were in the domain of research on improving delivery of known interventions, with simplified neonatal resuscitation program and clinical algorithms and improved skills of community health workers leading the list. The top 10 priorities in the domain of development were led by ideas on improved Kangaroo Mother Care at community level, how to improve the accuracy of diagnosis by community health workers, and perinatal audits. The 10 leading priorities for discovery research focused on stable surfactant with novel modes of administration for preterm babies, ability to diagnose fetal distress and novel tocolytic agents to delay or stop preterm labour.

Conclusion These findings will assist both donors and researchers in supporting and conducting research to close the knowledge gaps for reducing neonatal mortality, morbidity and long term impairment. WHO, SNL and other partners will work to generate interest among key national stakeholders, governments, NGOs, and research institutes in these priorities, while encouraging research funders to support them. We will track research funding, relevant requests for proposals and trial registers to monitor if the priorities identified by this exercise are being addressed

About 2.9 million newborns died in 2011, accounting for 44% of the world's under-5 child deaths [1]. The proportion of neonatal mortality continues to increase because the neonatal mortality rate is declining at a slower rate than the mortality rates for older children [1]. Moreover, 2.7 million stillbirths occur each year, at least 40% of which occur during labour [2]. The leading killers of newborns are preterm birth complications, intrapartum–related events and neonatal infections such as pneumonia, sepsis or meningitis [3]. A high proportion of stillbirths, neonatal and also maternal deaths happen at birth and during the first days after birth – a total of over 3 million deaths [4]. This is also a critical time window to address acute morbidity and long–term impairment associated with preterm birth, intrauterine growth restriction (IUGR), congenital abnormalities, and perinatal or infectious insults [5,6].

With the approaching deadline for the achievement of the Millennium Development Goals (MDGs) in 2015, and the creation of new framework for development goals [7], there is an increasing need to guide the limited research capacity and funding to obtain the maximum impact on maternal and child health. Hence the World Health Organization (WHO) has initiated a set of global research priority–setting exercises in 2007–2008 for improving health of mothers, newborns, children and adolescents [8–12]. The five–year evaluation of that exercise from the perspective of donors, policy–makers and researchers is currently under way and it is showing an increased focus on identified research priorities from all three groups of stakeholders – in terms of investments by the donors [13,14], initiatives launched by policy–makers [15–19] and publication output from researchers [2,20–23], respectively. As part of this initiative, the Department of Maternal, Newborn, Child and Adolescent Health undertook this exercise for setting research priorities in newborn health and stillbirth, in collaboration with Saving Newborn Lives (SNL), a program of Save The Children. The time frame for the expected impact of the research extends to 2025 to allow for medium term and long–term research investments to also be considered. Alongside the persisting urgency of reducing mortality and the findings from previous research priority exercises the group believed that the research should also address morbidity, development, and long–term sequelae of preterm birth, small for gestational age as well as other hypoxic or infectious insults in the neonatal period (**Box 1**). In the exercise, we focused on intrapartum stillbirth as a high proportion of stillbirths occurs during the labour.

Box 1 The purpose and remit of this research priority setting exercise

Population of interest:
Newborns and stillbirths, survival and health, preterm birth, growth and impairment–free development

Time frame:
2013–2025, reaching beyond the timeframe of the Millennium Development Goals

Research domains:
DISCOVERY (new interventions)
DEVELOPMENT (improved interventions)
DELIVERY (implementation of existing interventions)
(note: not including description eg, epidemiology)

Audience (stakeholders):
Governments, researchers in low and middle–income countries, international donors

METHODS

A working group that managed the agenda–setting process consisted of staff responsible for newborn health in WHO and Saving Newborn Lives. The group defined the scope of the priority setting exercise (**Box 1**). Methodology developed by the Child Health and Nutrition Research Initiative (CHNRI) was adapted and used for this priority setting exercise, to enable systematic listing and transparent scoring of many competing research questions [24–26]. This methodology had been used in the previous priority setting exercises by the WHO on five major causes of child deaths: pneumonia, diarrhea, preterm birth and low birth weight, neonatal infections, and birth asphyxia [8–12]. The previous exercise coordinated by the WHO was sharply focused on short–term gains, ie, within the MDG4 target of the year 2015. In addition, the CHNRI methodology has been used by many other subject groups and multiple organizations [27–33]. **Box 2** shows the steps we followed during this priority setting process.

A large group of researchers and program experts were identified and asked to submit three ideas for improving newborn health outcomes by 2025 (**Box 2**). Two hundred of the most productive researchers, representing a broad range of technical expertise and regional diversity, identified through Web of Science® ranking tools, were invited by email to propose research questions on newborn health and birth outcomes. A further 400 program experts in newborn health programmes were also invited to propose research questions.

The proposed research questions and scoring criteria were refined by a small group of 14 experts who were invited by the WHO to participate in a two–day workshop. Each question was assigned to a domain and a technical area. The first of the three domains was "discovery", which included research aimed at finding new solutions such as new medicines, vaccines or other preventive interventions, or new diagnostics. The second domain was "development", which included research questions aimed at improving existing interventions, reducing their costs or mak-

Box 2. Adapted Child Health and Nutrition Research Initiative's (CHNRI) methodology applied to set newborn research priorities

1. Selection of individuals to submit ideas and to score questions:

Individuals representing a wide range of technical expertise in the area of newborn health and birth outcomes were selected by including

- Top 100 most productive researchers in the previous 5 years (2008–2012), according to the Web of Science®, in any research that involved neonates anywhere in the world, including (but not limited to) fundamental research, obstetrics and gynaecology, social science, and other fields;
- Top 50 most productive researchers in the previous 5 years (see above) in research specifically involving neonates in low and middle income countries (LMICs);
- Top 50 most productive researchers in the previous 5 years (see above) in any research involving stillbirths;
- 400 program experts in newborn health, who were contacted through the Healthy Newborn Network Database, representing mainly national–level health programme managers in LMICs.

2. Identification of questions to be scored:

All the identified individuals were approached and asked to submit their three most promising ideas for improving newborn health outcomes by 2025. An expert group meeting was convened to review the 396 questions received from 132 experts. After removing or merging seemingly duplicate ideas, the submissions were consolidated into a set of 205 research questions and clarity of the questions was improved.

3. Scoring of research questions:

A set of 5 criteria to assess the proposed 205 research questions was agreed on.

The scoring criteria were based on CHNRI methodology [8–12]

i. Likelihood of answering the question in an ethical way
ii. Likelihood of efficacy
iii. Likelihood of deliverability and acceptability
iv. Likelihood for an important disease burden reduction
v. Predicted effect on equity

During the preliminary meeting, 14 experts invited from the larger pool of responders completed their scoring to test the methodology. The remaining experts were asked independently to answer a set of questions via an online survey on all the chosen criteria for all listed research options. Scores from a total of 91 experts were received.

4. Computation of scores for competing research options and ranking:

The intermediate scores were computed for each of the five criteria and they could potentially range between 0–100%. Those scores indicate the "collective optimism" of the group of scorers that a given research question would fulfil each given criterion. The overall research priority score for each research question was then computed as the mean of the intermediate scores. The average expert agreement scores were also calculated (**Online Supplementary Document**).

ing them simpler to deliver. The third domain was "delivery", which included research questions that would help deliver existing interventions to more mothers and newborns with high quality. The five separate technical areas included: (i) preterm birth; (ii) intrapartum–related events including intrapartum stillbirths; (iii) newborn infections; (iv) congenital malformations and other specific conditions; and (v) integrated care including the care for mothers and neonates;

The final list of research questions and scoring criteria were sent to the original group of 600 experts with an invitation to score them. Each research question was assessed by the expert and received a score of 1.0, 0.5 or 0 for five preset criteria, with the option of not assigning any score in case the expert did not feel confident to decide on that criterion. Scoring took place over eight weeks and was conducted and returned to the coordinators at the WHO by 91 experts.

Intermediate scores for each research question against the 5 criteria were computed as the sum of the scores for that particular criterion divided by the total number of scorers. This resulted in a number between 0–100% that captured the "collective optimism" of the group of 91 scorers that a given research question would fulfill each given criterion. The overall research priority score (RPS) for each research question was then computed as the mean of the intermediate scores calculated for each of the five criteria: RPS = [(Criterion 1 score %) +(Criterion 2 score %)+(Criterion 3 score %)+(Criterion 4 score %)+(Criterion 5 score %)]/5. The confidence interval was calculated using the bootstrapping methods in STATA version 11.2.

RESULTS

In total, 132 of the 600 invited experts proposed a total of 396 research questions, which were then checked for similarity and consolidated in a final list of 205 questions to be scored. The characteristics of respondents are summarized in **Figure 1**. The 205 research questions were then scored by 91 experts. About 40% of the scorers were based in low and middle income countries (LMICs) in Africa, Asia, and South America. About two–thirds (65%) worked in academic or research institutions and the remainder was divided between program managers (16%), clinicians (7%), donor representatives (7%) and policy makers (5%) (**Figure 1**).

The overall research priority scores given to the 205 proposed questions ranged from 90% (high) to 47% (low; full list of scored questions is presented in the **Online Supplementary Document**). The level of agreement between the 91 experts ranged from 77% (high) to 34% (low), suggesting that on average, for each research question of interest, between three–quarters and one–third of the scorers were in agreement in their responses to each criterion.

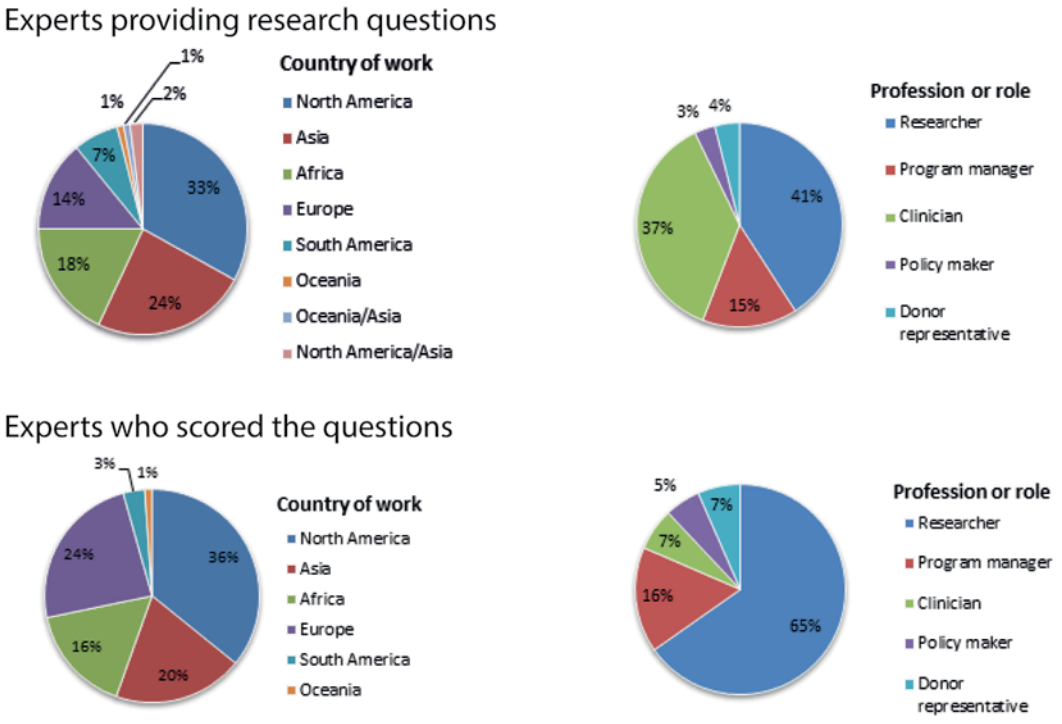


Figure 1. Background characteristics of 132 experts who provided questions and 91 experts who scored the questions.

The overall scores for the highest priority questions ranged from 79% to 90% (**Table 1**). Agreement scores indicated that more than two thirds of the experts had a common view towards the list of research priorities. Nine of the ten top priorities were in the domain of "delivery", with simplified neonatal resuscitation programs and clinical algorithms and improved skills of community health workers leading the list. Among the 11 priorities shown in this table, three addressed preterm birth, four addressed intrapartum–related events and four addressed newborn infections.

In the domain of "development", the top 10 priorities (**Table 2**) were ranked between 8th and 50th on the list of all research questions (displayed in full in **Online Supplementary Document**). They were led by ideas on improved Kangaroo Mother Care, improve accuracy of diagnosis by community health workers, and perinatal audits. Two priorities among the leading ten in this domain were identified in each of the areas of preterm birth, intrapartum related events and newborn infections, while the remaining 4 priorities related to integrated care.

The 10 leading priorities for discovery research (**Table 3**) ranked between 55th and 129th on the list of all research questions (see **Online Supplementary Document**) and they focused on stable surfactant with novel modes of administration, ability to diagnose fetal distress and novel tocolytic agents. Agreement scores for the ten leading questions ranged from 42% to 49%. Three priorities were identified in each of the areas of preterm birth and newborn infections, two on preventing intrauterine growth restriction and one each on intrapartum–related events and antepartum stillbirths.

There was a remarkable similarity in the scoring pattern between experts from a research background and those from a program background for the top 10 ranked priorities (**Table 4**). The programme experts had a tendency to assign somewhat higher overall scores to "delivery" questions, which was mediated through their higher scoring of maximum potential impact and equity criteria. Among "development" questions, the scorers with a background in research gave higher scores for efficacy and deliverability, while programme experts gave higher scores for impact and equity criteria. Surprisingly, the scoring pattern of both groups of experts for "discovery" questions was very similar, both for overall score and for each of the 5 criteria.

DISCUSSION

In this paper, we present global research priorities that have the potential to impact mortality, morbidity, child development, and long–term health outcomes among neonates in the period between 2013–2025. Despite the broad focus on these outcomes and a 12–year timeline, "delivery" questions received highest scores, followed by "development" and "discovery" questions, as was the case in previous exercises with shorter time lines focusing only on reducing mortality [8–12].

Table 1. Top ten research priorities for improving newborn health and birth outcomes by 2025 as ranked by 91 experts

Rank	Research questions	Domain	Total score (confidence interval)	Agreement between scorers	Answerable?	Efficacy?	Deliverability?	Impact?	Equity?
1	Can simplified neonatal resuscitation program delivered by trained health workers reduce neonatal deaths due to perinatal asphyxia?	Delivery	90 (85–91)	77	96	91	94	77	92
2	How can the health worker's skills in preventing and managing asphyxia be scaled up?	Delivery	88 (83–89)	74	96	91	89	75	86
3	Can simple clinical algorithms used by CHW identify and refer neonates with signs of infection and consequently reduce newborn mortality?	Delivery	86 (83–89)	72	92	92	92	66	88
4	How can exclusive breastfeeding in low–resource contexts be promoted to reduce neonatal infections and mortality?	Delivery	85 (79–89)	72	94	89	86	69	86
5	Can the training of CHWs in basic newborn resuscitation reduce morbidity and mortality due to perinatal asphyxia?	Delivery	83 (78–86)	67	94	84	84	64	88
6	How can the administration of injectable antibiotics at home and first level facilities to newborn with signs of sepsis be scaled up to reduce neonatal mortality?	Delivery	82 (78–86)	64	89	88	88	59	84
7	Can community–based initiation of Kangaroo Mother Care reduce neonatal mortality of clinically stable preterm and low birth weight babies?	Development	80 (74–84)	66	86	87	81	69	77
8	How can facility based initiation of Kangaroo Mother Care or continuous skin–to–skin contact be scaled up?	Delivery	80 (71–84)	62	90	82	84	62	81
9	How can chlorhexidine application to the cord be scaled up in facility births and in low NMR setting to reduce neonatal infections and neonatal mortality?	Delivery	80 (70–83)	67	91	85	89	52	81
10	How can quality of care during labour and birth be improved to reduce intrapartum stillbirths, neonatal mortality and disability?	Delivery	79 (71–82)	65	83	84	82	72	75
11*	Can community based "extra care" for preterm/LBW babies delivered by CHWs reduce neonatal morbidity and mortality in settings with poor accessibility to facility care?	Delivery	79 (70–82)	63	87	87	81	62	81

*The overall and criterion specific scores ranged from 0% to 100%.The 11th question added to complete the list of top 10 priorities in the domain of "delivery". The question originally ranked 5th was omitted from this table because it was a variant of question that already received a higher overall score.

Table 2. Top ten development research priorities for improving newborn health and birth outcomes by 2025 as ranked by 91 experts

Rank	Research questions	Total score (confidence interval)	Agreement between scorers
8*	Can community–based initiation of Kangaroo Mother Care reduce neonatal mortality of clinically stable preterm and low birth weight babies?	82 (78–86)	64
26	How can the accuracy of community health workers in detecting key most important high risk conditions or danger signs in pregnant women be improved?	77 (70–80)	61
35	Can perinatal audits improve quality of care in health facilities and improve fetal and neonatal outcomes?	74 (67–79)	58
37	Can intrapartum monitoring to enhance timely referral improve fetal and neonatal outcomes?	74 (67–79)	57
38	Can training community health workers to recognize and treat neonatal sepsis at home with oral antibiotics when referral is not possible reduce neonatal mortality?	74 (62–78)	57
40	Can oral amoxicillin at home for treatment of neonatal pneumonia reduce neonatal mortality?	73 (64–78)	58
43	Can models for strengthening capacity of health Professionals in caring for neonates in peripheral hospitals improve neonatal outcomes?	73 (63–77)	54
44	Can intervention package for CHWs to prevent and manage perinatal asphyxia be delivered by community health workers?	72 (64–77)	55
47	Can low–cost devices for facility care of newborns be developed and tested for the effectiveness at various levels of the health system (eg, CPAP devices, syringe drivers, IV giving sets, phototherapy units, oxygen concentrators, oxygen saturation monitors incubators, ventilators, therapeutic hypothermia technology) ?	72 (65–76)	53
50	Can surfactant reduce preterm morbidity and mortality in low and middle income countries?	72 (65–78)	56

*Also in the overall top 10 priorities.

Table 3. Top ten discovery research priorities in discovery for improving newborn health and birth outcomes by 2025 as ranked by 91 experts

Rank	Research questions	Total score (confidence interval)	Agreement between scorers
55	Can stable surfactant with simpler novel modes of administration increase the use and availability of surfactant for preterm babies at risk of respiratory distress syndrome?	71 (62–73)	49
71	Can the method to diagnose fetal distress in labour be more accurate and affordable?	66 (57–71)	49
97	Can strategies for prevention and treatment of intrauterine growth restriction be developed?	64 (51–68)	46
105	Can novel tocolytic agents to delay or stop preterm labour be developed in order to reduce neonatal mortality and morbidity?	63 (54–68)	42
116	Can major causal pathways and risk factors for antepartum stillbirth be identified?	61 (52–66)	43
118	Can novel point of care diagnostics for congenital syphilis be identified in low resource setting to improve management?	60 (53–64)	49
120	Can novel antibiotic or other biological agents be identified?	60 (51–65)	40
121	Can the new method identify intrauterine growth restriction at the early stage (including biomarkers) and predict abnormal postnatal growth and body composition?	60 (52–63)	43
125	Can novel vaccines for maternal immunization be developed and evaluated to prevent newborn infections (eg, GBS, Klebsiella, *E coli*, Staph)?	60 (51–64)	41
129	Can preterm birth be delayed or averted with antioxidant and/or nutrient supplementation (eg, Vitamin D, omega–3 fatty acids)?	58 (48–63)	42

GBS – group B streptococcus, Staph – staphylococcus

Table 4. Overall scoring pattern by profile of experts

	Median (IQR)		
	All scorers (n = 91)	Researchers (n = 61)	Programme experts (n = 30)
TOTAL SCORE			
Delivery	82 (80–86)	83 (78–86)	86 (81–87)
Development	74 (72–74)	75 (71–76)	75 (68–79)
Discovery	61 (59–64)	62 (60–62)	63 (58–65)
AGREEMENT			
Delivery	67 (65–72)	68 (64–73)	70 (65–75)
Development	57 (55–58)	58 (56–60)	55 (54–62)
Discovery	43 (42–49)	45 (42–47)	44 (39–49)
ANSWERABLE?			
Delivery	92 (87–94)	92 (88–95)	91 (90–94)
Development	84 (82–89)	87 (81–90)	84 (78–89)
Discovery	76 (73–78)	76 (74–79)	76 (70–79)
EFFICACY?			
Delivery	87 (84–91)	87 (83–91)	88 (84–90)
Development	81 (77–83)	84 (79–84)	78 (76–81)
Discovery	68 (64–70)	68 (65–72)	69 (59–72)
DELIVERABILITY?			
Delivery	85 (82–89)	86 (82–91)	87 (82–89)
Development	77 (75–80)	79 (77–81)	74 (70–84)
Discovery	68 (66–72)	69 (64–72)	70 (64–72)
IMPACT?			
Delivery	68 (62–72)	65 (58–70)	73 (69–80)
Development	56 (53–57)	53 (52–58)	62 (52–65)
Discovery	46 (39–50)	46 (38–48)	44 (36–54)
EQUITY?			
Delivery	84 (81–88)	84 (76–89)	87 (79–88)
Development	74 (66–77)	71 (65–76)	76 (75–80)
Discovery	54 (50–59)	52 (50–58)	53 (50–65)

The major emerging themes in the domain of "delivery" included simplifying intervention delivery to implementation at lower levels of the health system, evaluating delivery of interventions by community health workers, developing strategies to improve quality of care during labour and childbirth, and addressing barriers in the scaling up of high impact interventions. It is interesting to note that 5 of the questions were related to neonatal resuscitation. This could be related to neonatal resuscitation being the most dramatic intervention in newborn care. The major themes in the domain of "development" were adapting known interventions to make them deliverable at the community level, adapting effective interventions to increase deliverability in health facilities in low and middle income countries, and approaches such as perinatal audits to improve quality of care to mothers and newborns. The themes in the domain of "discovery" included new, more effective and less expensive medicines for preventing preterm birth and treating sepsis, point of care diagnostics for infections, maternal vaccines to prevent newborn infections, and basic science work on causal pathways for identifying intervention targets and biomarkers for preterm birth, IUGR, and antepartum stillbirths. It is noteworthy that preterm prevention was not ranked highly, even though it may have the largest impact. This appears to be the result of these questions being scored low in answerability.

The relatively lower scores for the "development" and "discovery" groups of research questions may have several possible explanations. First, more than 95% of the neonatal deaths occur in low and middle–income countries (LMICs). Therefore, research addressing neonatal health issues that are relatively more important in wealthy countries may be

perceived to contribute less to global reduction in mortality and morbidity, explaining some of the lower scores received by potentially promising research on novel interventions based on high technologies. Second, "discovery" research often takes longer to be translated into measurable benefits in terms of mortality burden reduction, and by definition the link to reduction in mortality and inequity is less direct. One specific example is research on prevention of preterm birth – while it was likely to have high impact, it was ranked only 129th among the 205 questions. Thereby, respondents sent a message that this research question would likely be difficult to answer given the current stage of knowledge. Third, the process of delivery of novel interventions usually requires specific funding mechanisms, such as PEPFAR or Advance Market Commitment (AMC), which require time for a political agreement [34,35].

The CHNRI process we followed for setting priorities has several strengths. The methodology is transparent, replicable, and feasible to apply via e–mail [8–12, 27–33]. The output is intuitive and easily understood, and it has been refined and improved through many exercises over the past several years [36]. In this particular exercise, further improvements have been introduced to the process. We chose a large number of experts based on their productivity in the previous five years using Web of Science®, thus transparently identifying the group that was most likely to understand the field and its present research challenges and gaps. A very wide global network of programme experts in the Saving Newborn Lives' Network was also invited. Moreover, we used online data collection tools, such as Survey Monkey® and Google Analytics®, which allowed monitoring of the progress of the exercise in real time, ensured adequate representation of experts by their background and region, and increased the efficiency of data management. Finally, 132 experts proposed research questions and 91 scored all the questions in this exercise; this is considerably more than in previous priority setting exercises using CHNRI methodology, where we typically involved fewer scorers, research ideas, and criteria scored by each expert.

There may be concern that the results derived from the CHNRI approach might represent only the collective opinion of the limited group of people who were included in the process. However, we were able to obtain questions and scores from a large number of experts worldwide, who were selected in a transparent and replicable manner, based on their research productivity in the field. The large number of participants and the protection against potential bias provided by the CHNRI approach make our results more credible, although it remains apparent that the highest scored questions may still be biased towards those that researchers are most familiar with and so may bias reflect research already in progress. This issue may be particularly relevant in view that only about a quarter of originally invited researchers, policy makers and programme experts eventually contributed to generating research questions, and only about one in six completed the scoring process, making response bias an important potential concern. Second, even though the list of proposed questions was reviewed and refined before sending for scoring, there were still overlaps in some research questions, possibly creating confusion in scoring such questions. Those and other possible strengths and limitations of CHNRI methodology are described and discussed in greater detail in **Online Supplementary Document**.

A recent analysis of funding committed globally to improving neonatal health and birth outcomes has shown that donor mention of the "newborn" has increased quite sharply since 2005. However, given a total of only 10% of all donor aid to RMNCH mentioning the word "newborn", and only 0.01% referring to interventions expected to reduce newborn deaths, it still seems unlikely that donor aid is commensurate with the large burden of 3.0 million newborn deaths each year, or with the burden of morbidity, developmental and long–term health outcomes [37]. The word "stillbirth" occurred only twice in the OECD database between 2002 and 2010, suggesting even lower attention for the world's 2.7 million stillbirths.

Large inequities in current research funding support exist not only in the amounts invested in newborn health in comparison to other diseases globally, but also between different neonatal conditions themselves. Conditions that affect newborns in high–income countries receive more funding and attention than conditions that largely affect newborns in low–income countries. For instance, the research on care of preterm babies in neonatal intensive care units has received considerably more funding over the past several years in comparison to intrapartum–related birth outcomes or newborn sepsis [38].

The results presented in this paper will assist both the donors and the researchers in setting evidence based priorities to address the key gaps in knowledge, that could make the most difference in saving newborn lives and preventing stillbirth. In addition, attention to many of these questions could also improve maternal and child health outcomes. Likewise, research priorities to address other related areas such as maternal, child and adolescent health and health system issues may have substantial effect on newborn health. Complementary exercises are under way to identify research priorities in these areas. Using the identified research priorities, WHO, SNL and other partners, that are linked to the Every Newborn action plan launched in 2014 [39], will work to generate research interests among key national stakeholders, governments, NGOs, and research institutes, while encouraging research funders to support these priorities. We will track research funding, relevant request for proposals and trial registers to monitor if the priorities identified by this exercise are being addressed, and highlight those that are not being addressed.

AFFILIATIONS OF AUTHORS

[1] Department of Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland
[2] Centre for Intervention Science in Maternal and Child Health, Centre for International Health, University of Bergen, Norway
[3] London School of Hygiene and Tropical Medicine, London, UK
[4] Saving Newborn Lives, Save the Children, Washington, USA
[5] Department of Social Medicine, Ribeirão Preto School of Medicine, University of São Paulo, Brazil
[6] Centre for Population Health Sciences and Global Health Academy, The University of Edinburgh Medical School, Scotland, UK
[7] Bandim Health Project, Indepth Network, Guinea–Bissau
[8] Faculty of Medicine, University of Khartoum, Sudan
[9] Kathmandu Medical College, Nepal
[10] Department of Pediatrics, University of Alabama at Birmingham, Birmingham, AL, USA
[11] Centre for Child and Adolescent Health, International Centre for Diarrhoeal Disease Research, Bangladesh
[12] Department of Neonatology Paropakar Maternity and women's Hospital, Nepal
[13] TRAction Bangladesh Project, University Research Co., LLC
[14] Johns Hopkins Bloomberg School of Public Health, USA
[15] Centro de Pesquisas Epidemiológicas, Universidade Federal de Pelotas, Brazil
[16] Research Center for Vitamins and Vaccines, Statens Serum Institut, and University of Southern Denmark/Odense University Hospital, Denmark
[17] Program in Perinatal Research, Yale University School of Medicine, USA
[18] Pediatric Biology Centre, Translational Health Science and Technology Institute, India
[19] University of Aberdeen, UK
[20] Center of Excellence in Women and Child Health, the Aga Khan University, Karachi, Pakistan
[21] Institute of International Programs, Johns Hopkins Bloomberg School of Public Health, USA
[22] The London School of Hygiene & Tropical Medicine, UK
[23] The University of North Carolina at Chapel Hill School of Medicine, USA
[24] Thrasher Research Fund, USA
[25] Department of Neonatology, Charité University Medical Center, Germany
[26] University of Alabama at Birmingham, USA
[27] Department of Obstetrics and Gynaecology, School of Medical Sciences, University of Campinas, Brazil
[28] Departments of Pediatrics, Pharmacology & Surgery, University of Alberta, Canada
[29] Charities, USA
[30] University College London Institute for Global Health, UK
[31] Perinatology Research Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development/National Institutes of Health/Department of Health and Human Services, Bethesda, Maryland and Detroit, Michigan, USA
[32] Independent consultant maternal health research, Rwanda
[33] Foundation for the Community Control of Hereditary Diseases, Hungary
[34] Biostatistics and Epidemiology, RTI International, USA
[35] LAMB Integrated Rural Health & Development, Bangladesh
[36] Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases National Institute of Health, USA
[37] All India Institute of Medical Sciences, India
[38] Pediatrics and Neonatology, Yıldırım Beyazıt University Medical Faculty, Turkey
[39] Nuffield Department of Medicine & Department of Paediatrics, University of Oxford, UK and KEMRi–Wellcome Trust Research Programme, Nairobi, Kenya
[40] Newborn Health, Family Health Division, The Bill & Melinda Gates Foundation and the University of North Carolina Schools of Medicine and Public Health, USA
[41] Moi University, School of Medicine, Kenya
[42] International Paediatric Epidemiology; Affiliations: Medical Research Council Lifecourse Epidemiology Unit, University of Southampton, UK
[43] Department of Pediatrics, UCSF Benioff Children's Hospital, USA
[44] School of Medicine, Child Health and Pediatrics, Moi University, Kenya
[45] Children's Hospital, Pakistan Institute of Medical Sciences, Pakistan
[46] Eunice Kennedy Shriver NICHD Neonatal Research Network, Pregnancy and Perinatology, Branch, National Institute of Health, USA
[47] Centre for Midwifery, Child and Family Health, University of Technology, Sydney, Australia
[48] Centre for Child and Adolescent Health, International Centre for Diarrhoeal Disease Research, Bangladesh
[49] University of Bergen and Norwegian Institute of Public Health, Norway
[50] Japan International Cooperation Agency (JICA), Bangladesh
[51] Save the Children, USA
[52] Zomba Central Hospital, Ministry of Health, Malawi
[53] Saint Louis University, USA
[54] Paediatrics and Head of Department Shaheed Suhrawardy Medical College, Bangladesh
[55] Centre for Pharmacoepidemiology, Karolinska Institute, Sweden
[56] Departments of Pediatrics and of Epidemiology, Biostatistics and Occupational Health, McGill University Faculty of Medicine, Montreal, Quebec, Canada
[57] Global Alliance to Prevent Prematurity and Stillbirth (GAPPS), USA
[58] University of Manchester School of Nursing Midwifery & Social Work, University of Manchester, UK
[59] Indonesian Midwives Association, Indonesia
[60] Save the Children, Malawi
[61] Save the Children, USA
[62] Department of Paediatrics, Dalhousie University, Canada
[63] Departments of Paediatrics & Physiology, University of Toronto; Physiology & Experimental Medicine program, Hospital for Sick Children, Toronto, Canada
[64] Department of Obstetrics and Gynaecology, Academic Medical Centre Amsterdam, the Netherlands
[65] Paediatrics and Child Health, College of Medicine Malawi
[66] International Baby Food Action Network, Uganda
[67] STRIDES for Family Health, Management Sciences for Health, Uganda
[68] Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, UK
[69] Kampala Children's Hospital Limited and Childhealth Advocacy International, Uganda
[70] United States Agency for International Development /Maternal and Child Health Integrated Program, USA.
[71] Centre for Healthy Start Initiative, Nigeria
[72] Wellcome Trust Senior Research Fellow in Clinical Science, Institute for Global Health, University College London, UK
[73] All India Institute of Medical Sciences, India
[74] University of Tubingen, Germany
[75] Eunice Kennedy Shriver National Institute of Child Health and Human Development National Institutes of Health, USA
[76] Center for American Indian Health, Johns Hopkins Bloomberg School of Public Health, USA
[77] Save the Children, Bangladesh
[78] Department of Obstetrics and Gynecology, University of Tennessee Health Science Center, USA
[79] University of Calgary, Canada
[80] Environmental Health Science Department, University of Georgia, USA
[81] Tropical Epidemiology Group, London School of Hygiene & Tropical Medicine, London, UK
[82] Department of Pediatrics & Child Health, Women & Child Health Division, Aga Khan University, Pakistan
[83] Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, USA
[84] National Institute of Population Research and Training (NIPORT), Ministry of Health and Family Welfare, Bangladesh
[85] Kinshasa School of Public Health, School of Medicine, University of Kinshasa, Democratic Republic of Congo
[86] Department of Neonatology, University of Utrecht, the Netherlands
[87] Institute for Global Health Technologies Rice University, USA
[88] Division of Global Health, Karolinska Institutet, Sweden
[89] School of Medical Sciences, Edith Cowan University, Australia and School of Public Health, Capital Medical University, China
[90] Save the Children UK
[91] Eunice Kennedy Shriver National Institute of Child Health and Human Development National Institutes of Health, USA
[92] Aga Khan University, Pakistan
[93] Department of Integrated Early Childhood Development, Capital Institute of Paediatrics, China
[94] Developmental Genetics Laboratory, New York State Institute for Basic Research in Developmental Disabilities, USA
[95] Médecins sans Frontières, Belgium

Acknowledgements: The authors thank the expert group for their time and effort in this priority–setting exercise, and the contribution of anonymous reviewers to the final version of this report. We acknowledge inputs from Dr Diane Morof from Centers for Disease Control and Prevention, USA.

Disclaimer: SY, RB, JPS, JM are employees of the World Health Organization. The views expressed in this paper are the responsibility of the authors and do not necessarily represent the views of the World Health Organization.

Funding: This work was supported by Save the Children.

Ethical approval: Not required.

Authorship declaration: SY and IR conducted the analysis and wrote the first draft of the paper. RB, JM, JL, SW, JPS and SC revised the draft and finalized the paper. All the named authors had a major role in design, implementation or analysis of the priority setting exercise. The Newborn Health Priority setting group proposed research questions, helped finalize the list of questions and provided scores that served as the dataset for all quantitative analyses

REFERENCES

1 UNICEF, World Health Organization, The World Bank and United Nations. Level & Trends in Child Mortality, Estimates developed by the UN inter–agency group for child mortality estimation. New York: UNICEF, 2014.

2 Cousens S, Blencowe H, Stanton C, Chou D, Ahmed S, Steinhardt L, et al. National, regional, and worldwide estimates of stillbirth rates in 2009 with trends since 1995: a systematic analysis. Lancet. 2011;377:1319-30. Medline:21496917 doi:10.1016/S0140-6736(10)62310-0

3 Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, 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. Medline:22579125 doi:10.1016/S0140-6736(12)60560-1

4 Oza S, Cousens SN, Lawn JE. Estimation of daily risk of neonatal death, including the day of birth, in 186 countries in 2013: a vital–registration and modelling–based study. Lancet Glob Health. 2014;2:e635-44. Medline:25442688 doi:10.1016/S2214-109X(14)70309-2

5 Mwaniki MK, Atieno M, Lawn JE, Newton CRJC. Long–term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. Lancet. 2012;379:445-52. Medline:22244654 doi:10.1016/S0140-6736(11)61577-8

6 Horton R. GBD 2010: understanding disease, injury, and risk. Lancet. 2012;380:2053-4. Medline:23245595 doi:10.1016/S0140-6736(12)62133-3

7 United Nations. A new global partnership: Eradicate poverty and transform economies through sustainable development. The report of the high–level Panel of Eminent Persons on the Post–2015 Development Agenda. New York: United Nations, 2013.

8 Fontaine O, Kosek M, Bhatnagar S, Boschi–Pinto C, Chan KY, Duggan C, et al. Setting research priorities to reduce global mortality from childhood diarrhoea by 2015. PLoS Med. 2009;6:e41. Medline:19278292 doi:10.1371/journal.pmed.1000041

9 Rudan I, El Arifeen S, Bhutta ZA, Black RE, Brooks A, Chan KY, et al. Setting research priorities to reduce global mortality from childhood pneumonia by 2015. PLoS Med. 2011;8:e1001099. Medline:21980266 doi:10.1371/journal.pmed.1001099

10 Lawn JE, Bahl R, Bergstrom S, Bhutta ZA, Darmstadt GL, Ellis M, et al. Setting research priorities to reduce almost one million deaths from birth asphyxia by 2015. PLoS Med. 2011;8:e1000389. Medline:21305038 doi:10.1371/journal.pmed.1000389

11 Bahl R, Martines J, Ali N, Bhan MK, Carlo W, Chan KY, et al. Research priorities to reduce global mortality from newborn infections by 2015. Pediatr Infect Dis J. 2009;28:S43-8. Medline:19106763 doi:10.1097/INF.0b013e31819588d7

12 Bahl R, Martines J, Bhandari N, Biloglav Z, Edmond K, Iyengar S, et al. Setting research priorities to reduce global mortality from preterm birth and low birth weight by 2015. J Glob Health. 2012;2:010403. Medline:23198132 doi:10.7189/jogh.01.010403

13 Pitt C, Lawn JE, Ranganathan M, Mills A, Hanson K. Donor funding for newborn survival: an analysis of donor–reported data, 2002–2010. PLoS Med. 2012;9:e1001332. Medline:23118619 doi:10.1371/journal.pmed.1001332

14 Sridhar D. Who sets the global health research agenda? The challenge of multi–bi financing. PLoS Med. 2012;9:e1001312. Medline:23049485 doi:10.1371/journal.pmed.1001312

15 Greenwood B. A global action plan for the prevention and control of pneumonia. Bull World Health Organ. 2008;86:322-2A. Medline:18545727 doi:10.2471/BLT.08.053348

16 Zipursky A, Wazny K, Black R, Keenan W, Duggan C, Olness K, et al. Global action plan for childhood diarrhoea. Developing research priorities: Report from a Workshop of the Programme for Global Paediatric Research. J Glob Health. 2013;3:010406. Medline:23826510 doi:10.7189/jogh.03.010406

REFERENCES

17 Bhutta ZA, Chopra M, Axelson H, Berman P, Boerma T, Bryce J, et al. Countdown to 2015 decade report (2000–10): taking stock of maternal, newborn, and child survival. Lancet. 2010;375:2032-44. Medline:20569843 doi:10.1016/S0140-6736(10)60678-2

18 Lawn JE, Gravett MG, Nunes TM, Rubens CE, Stanton C; Global Action Plan for Prematurity and Stillbirths (GAPPS) Review Group. Global report on preterm birth and stillbirth (1 of 7): definitions, description of the burden and opportunities to improve data. BMC Pregnancy Childbirth. 2010;10 Suppl 1:S1. Medline:20233382

19 World Pneumonia Day. Available: http://worldpneumoniaday.org. Accessed: 1 August 2013.

20 Walker CL, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta ZA, et al. Global burden of childhood pneumonia and diarrhoea. Lancet. 2013;381:1405-16. Medline:23582727 doi:10.1016/S0140-6736(13)60222-6

21 Seale AC, Mwaniki M, Newton CR, Berkley JA. Maternal and early onset neonatal bacterial sepsis: burden and strategies for prevention in sub–Saharan Africa. Lancet Infect Dis. 2009;9:428-38. Medline:19555902 doi:10.1016/S1473-3099(09)70172-0

22 Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, Narwal R, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet. 2012;379:2162-72. Medline:22682464 doi:10.1016/S0140-6736(12)60820-4

23 Lawn JE, Cousens S, Zupan J; Lancet Neonatal Survival Steering Team. 4 million neonatal deaths: When? Where? Why? Lancet. 2005;365:891-900. Medline:15752534 doi:10.1016/S0140-6736(05)71048-5

24 Rudan I, El Arifeen S, Black RE, Campbell H. Childhood pneumonia and diarrhoea: setting our priorities right. Lancet Infect Dis. 2007;7:56-61. Medline:17182344 doi:10.1016/S1473-3099(06)70687-9

25 Rudan I, Gibson JL, Ameratunga S, El Arifeen S, Bhutta ZA, Black M, et al. Setting priorities in global child health research investments: Guidelines for implementation of the CHNRI method. Croat Med J. 2008;49:720-33. Medline:19090596 doi:10.3325/cmj.2008.49.720

26 Rudan I, Chopra M, Kapiriri L, Gibson J, Lansang MA, Carneiro I, et al. Setting priorities in global child health research investments: Universal challenges and conceptual framework. Croat Med J. 2008;49:307-17. Medline:18581609 doi:10.3325/cmj.2008.3.307

27 Tomlinson M, Swartz L, Officer A, Chan KY, Rudan I, Saxena S. Research priorities for health of people with disabilities: an expert opinion exercise. Lancet. 2009;374:1857-62. Medline:19944866 doi:10.1016/S0140-6736(09)61910-3

28 Brown KH, Hess SY, Boy E, Gibson RS, Horton S, Osendarp SJ, et al. Setting priorities for zinc–related health research to reduce children's disease burden worldwide: an application of the Child Health and Nutrition Research Initiative's research priority–setting method. Public Health Nutr. 2009;12:389-96. Medline:18426636

29 Wazny K, Zipursky A, Black R, Curtis V, Duggan C, Guerrant R, et al. Setting research priorities to reduce mortality and morbidity of childhood diarrhoeal disease in the next 15 years. PLoS Med. 2013;10:e1001446. Medline:23690756 doi:10.1371/journal.pmed.1001446

30 Tomlinson M, Rudan I, Saxena S, Swartz L, Tsai AC, Patel V. Setting priorities for global mental health research. Bull World Health Organ. 2009;87:438-46. Medline:19565122 doi:10.2471/BLT.08.054353

31 Souza JP, Widmer M, Gülmezoglu AM, Lawrie TA, Adejuyigbe EA, Carroli G. Maternal and perinatal health research priorities beyond 2015: an international survey and prioritization exercise. Reprod Health. 2014;11:61. Medline:25100034 doi:10.1186/1742-4755-11-61

32 Tomlinson M, Swartz L, Officer A, Chan KY, Rudan I, Saxena S. Research priorities for health of people with disabilities: an expert opinion exercise. Lancet. 2009;374:1857-62. Medline:19944866 doi:10.1016/S0140-6736(09)61910-3

33 Hindin MJ, Christiansen CS, Ferguson BJ. Setting research priorities for adolescent sexual and reproductive health in low– and middle–income countries. Bull World Health Organ. 2013;91:10-8. Medline:23397346 doi:10.2471/BLT.12.107565

34 Goosby E, Von Zinkernagel D, Holmes C, Haroz D, Walsh T. Raising the bar: PEPFAR and new paradigms for global health. J Acquir Immune Defic Syndr. 2012;60 Suppl 3:S158-62. Medline:22797738 doi:10.1097/QAI.0b013e31825d057c

35 Zandonella C. If you build it, they will pay. A novel incentive called and Advance Market Commitment could help spur private sector investment in AIDS vaccine research and development. IAVI Rep. 2005;9:13-21. Medline:16276579

36 Rudan I. Global health research priorities: mobilizing the developing world. Public Health. 2012;126:237-40. Medline:22325672 doi:10.1016/j.puhe.2011.12.001

37 Pitt C, Lawn JE, Ranganathan M, Mills A, Hanson K. Donor funding for newborn survival: an analysis of donor–reported data, 2002–2010. PLoS Med. 2012;9:e1001332. Medline:23118619 doi:10.1371/journal.pmed.1001332

38 Howson CP, Kinney MV, Lawn JE, editors. March of Dimes; PMNCH; Save the Children; World Health Organization. Born too soon: The global action report on preterm birth. Geneva: World Health Organization, 2012.

39 Every newborn action plan – an action plan to end preventable deaths. Available: http://www.globalnewbornaction.org/about/. Accessed: 31 July, 2014.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? The Lancet Global Health

1.2. When was the work published? September 2016

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I have designed the project, analysed the result and wrote the first draft of the method section.

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE [redacted] Date 25 July 2018

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

Asking different questions: research priorities to improve the quality of care for every woman, every child

Lancet Glob Health 2016
Published Online
September 20, 2016
http://dx.doi.org/10.1016/S2214-109X(16)30183-8

Unacceptably high rates of adverse outcomes persist for childbearing women and infants, including maternal and newborn mortality, stillbirth, and short-term and long-term morbidity.[1] In light of the challenges to achieve the UN Sustainable Development Goals, it is timely to reconsider priorities for research in maternal and newborn health. Are we asking the right questions?[2] Recent evidence indicates the importance of seeking knowledge beyond the treatment of complications, to inform better ways of providing sustainable, high quality care, including preventing problems before they occur.[3]

The 2014 publication of *The Lancet*'s Series on Midwifery presented a unique opportunity to generate future areas of inquiry by drawing on the most extensive examination to date of evidence on the care that all women and newborn infants need across the continuum from pre-pregnancy, birth, postpartum, and the early weeks of life.[4–6] The series summarised the evidence base for quality maternal and newborn care in a new framework that focuses on the needs of women, infants, and families and differentiates between *what* care is provided, *how* it is provided, and by *whom*.[4] These are concepts that are often confused or ignored in existing studies. Midwifery was identified as a cost-effective and fundamentally important element of quality care, with the potential to improve over 50 different maternal and newborn outcomes including mortality and morbidity. However, there are substantive barriers to proper implementation and integration of midwifery into health systems.[1]

We adapted the Child Health and Nutrition Research Initiative (CHNRI) methodology to score competing future research topics on quality maternal and newborn care and the contribution of midwifery to that care.[7] This method has been used to set health research priorities for infant and childhood conditions,[8,9] reduction of maternal and perinatal mortality,[2] and preterm birth and stillbirths.[10]

A team representing expertise in maternal and newborn health research, including authors from *The Lancet*'s Series on Midwifery, contributors from WHO, UNFPA, the International Confederation of Midwives, and a representative of or advocate for service users conducted the work. The team identified

	Research priorities	Research priority score
1	Evaluate the effectiveness of midwifery care across the continuum in increasing access to and acceptability of family planning services for women	90·4
2	Evaluate the effectiveness of midwife-led care when compared to other models of care across various settings, particularly on rates of fetal and infant death, preterm birth, and low birthweight	89·8
3	Determine which indicators are most valuable in assessing quality maternal and newborn care	89·7
4	Identify and describe aspects of care that optimise, and those that disturb, the biological/physiological processes for healthy childbearing women and fetus/newborn infants and those who experience complications	89·3
5	Evaluate the effectiveness of midwifery care in providing culturally appropriate information, education, and health promotion (eg, nutrition, substance use, domestic violence, and mental health)	89·1
6	Identify and describe enabling factors from examples of successful implementation of evidence-based maternal and newborn care across a variety of settings	89
7	Describe and evaluate the effectiveness of midwives working with others (such as health professionals, community health workers, and traditional birth attendants) in achieving quality maternal and newborn care including, but not limited to: Timely transfer of women to appropriate level/site of care Management of emergency situations Maximal use of skills and competencies Shared decision-making and accountability	89
8	Assess the views and preferences of women and families across a variety of settings about their experiences of maternal and newborn care including, but not limited to, care providers and sites of care (eg, place of birth, antenatal care)	88·8
9	Develop setting-specific benchmarks to assess measurable progress on implementation of quality maternal and newborn care	88·3
10	Identify and describe aspects of maternal and newborn care that strengthen or weaken women's psychosocial wellbeing and mental health	88·0
11	Assess whether new measures of morbidity are needed to more effectively evaluate outcomes of maternal and newborn care	88·0

Table: **Ranking of research topics by overall research priority score**

30 research topics based on an analysis of gaps in the evidence presented in the 2014 *Lancet* Series on Midwifery. Stakeholders were asked to consider the potential research topics in terms of their relevance, significance, and potential future implementation based on five criteria: answerability, community involvement, sustainability, equity, and maximal impact.[7] The 30 research topics and scoring criteria were distributed in English, French, and Spanish online surveys to 1191 stakeholders, including constituents of the global Partnership for Maternal, Newborn, and Child Health (PMNCH) and representatives from all WHO regions. Stakeholders were asked to score each of the 30 research topics as 1·0 (yes), 0·5 (informed but undecided answer), or 0 (no) on whether they met each of the five criteria. It was possible to omit a score if a respondent did not feel confident to decide on a criterion; these were regarded as missing data and not part of the denominator. Summary scores for each criterion and an overall score were then computed as the sum of the scores divided by the number of actual scorers.

Responses were received from all WHO regions, with a total response rate of 23% (n=271). Most (83%) responses were submitted in English, 13% in French, and 4% in Spanish. The highest proportion (24%) of those who provided demographic information came from the Western Pacific Region and the lowest (2·6%) from southeast Asia. Over a quarter (26%) came from the academic, research, or training institution sector of the PMNCH constituents. Of the 199 respondents who identified themselves as health professionals, 168 (84%) were in roles associated with maternal and child health.

Our goal was to identify the top 10 priorities; however, two scored equally, and so the top 11 are presented in the table. The stakeholders prioritised research that would increase knowledge about ways to prevent complications and reduce unnecessary interventions, strengthen women's own capabilities, and optimise biological, social, and cultural processes. They also identified the importance of examining the role of midwifery in providing quality care for all women and infants. Stakeholders also identified research to improve skilled, knowledgeable, and compassionate care provided by an appropriate workforce that ensures timely referral when complications arise. The top two priorities indicate the fundamental importance of effective family planning services and of quality care to reduce rates of preterm birth, low birthweight, stillbirth, and perinatal mortality. Evidence indicates that midwifery care can be a key intervention to improve these outcomes, but more research is urgently needed to determine clinically and cost-effective models of care in diverse settings, especially in low-resource areas.[11] A focus on new measures and indicators of care components that have not traditionally been well examined will enable new benchmarks to be set for developing systems of care that meet the needs of all women and newborns.

The priorities identified reveal broad knowledge domains rather than individual research questions. Research funding in the past has often targeted management of critical situations that contribute to high mortality, such as haemorrhage, hypertensive disorders, obstructed labour, preterm birth, and sepsis.[12] The priorities identified in this study do not eschew the importance of complication management, but potentially restore balance by moving towards a focus on prevention. Studying ways of providing such care has the potential to improve the provision of quality care for all, enhance women's and infants' own capabilities, and maximise the health promotion potential of midwives. The Global Strategy for Women's, Children's and Adolescents' Health 2016–2030 is designed to help women, children, and adolescents survive, thrive, and transform.[13] The concepts of thriving and transforming particularly resonate with the research priorities identified in this exercise. Importantly, this new knowledge could contribute to achieving Sustainable Development Goal 3, for healthy lives and wellbeing for all people. Investment in these innovative priorities has the potential to enable the rights of women and children to life and to health, and help women, infants, and families to survive and thrive. It would be transformative for families, communities, and science.

*Holly Powell Kennedy, Sashiyo Yoshida, Anthony Costello, Eugene Declercq, Marcos A Dias, Elizabeth Duff, Atf Gherissi, Karyn Kaufman, Frances McConville, Alison McFadden, Michaela Michel-Schuldt, Nester T Moyo, Kerri Schuiling, Anna M Speciale, Mary J Renfrew

Yale University, West Haven, CT 06516, USA (HPK); World Health Organization, Geneva, Switzerland (SY, AC, FM); Boston University, Boston, MA, USA (ED); Instituto Fernandes Figueira—FIOCRUZ, Rio de Janeiro, Brazil (MAD); National Childbirth Trust, London, UK (ED);

Université Tunis-El Manar, Tunis, Tunisia (AG); McMaster University, Hamilton, ON, Canada (KK); University of Dundee, Dundee, UK (AM, MJR); UNFPA, New York, USA (MM-S); International Confederation of Midwives, The Hague, Netherlands (NTM, KS); American College of Nurse-Midwives, Silver Spring, MD, USA (AMS)
holly.kennedy@yale.edu

We declare no competing interests. Funding from the Bill & Melinda Gates Foundation and NORAD was used to support author meetings, data collection and analysis, and editing.

1 International Confederation of Midwives, WHO, UNFPA. State of the World's Midwifery: a universal pathway, a woman's right to health. http://www.unfpa.org/sowmy (accessed Sept 8, 2016).
2 Chalmers I. The perinatal research agenda: whose priorities? *Birth* 1991; **18:** 137–41.
3 Souza JP, Widmer M, Gulmezoglu AM, et al. Maternal and perinatal health research priorities beyond 2015: an international survey and prioritization exercise. *Reprod Health* 2014; **11:** 61.
4 Renfrew MJ, McFadden A, Bastos MH, et al. Midwifery and quality care: findings from a new evidence-informed framework for maternal and newborn care. *Lancet* 2014; **384:** 1129–45.
5 Homer CS, Friberg IK, Dias MA, et al. The projected effect of scaling up midwifery. *Lancet* 2014; **384:** 1146–57.
6 Van Lerberghe W, Matthews Z, Achadi E, et al. Country experience with strengthening of health systems and deployment of midwives in countries with high maternal mortality. *Lancet* 2014; **384:** 1215–25.
7 Rudan I, Gibson JL, Ameratunga S, et al. Setting priorities in global child health research investments: guidelines for implementation of CHNRI method. *Croat Med J* 2008; **49:** 720–33.
8 Yoshida S, Rudan I, Lawn JE, et al. Newborn health research priorities beyond 2015. *Lancet* 2014; **384:** e27–29.
9 Lawn JE, Bahl R, Bergstrom S, et al. Setting research priorities to reduce almost one million deaths from birth asphyxia by 2015. *PLoS Med* 2011; **8:** e1000389.
10 Bahl R, Martines J, Bhandari N, et al. Setting research priorities to reduce global mortality from preterm birth and low birth weight by 2015. *J Glob Health* 2012; **2:** 010403.
11 Sandall J, Soltani H, Gates S, Shennan A, Devane D. Midwife-led continuity models versus other models of care for childbearing women. *Cochrane Database Syst Rev* 2015; **9:** CD004667.
12 Tuncalp, Were WM, MacLennan C, et al. Quality of care for pregnant women and newborns-the WHO vision. *BJOG* 2015; **122:** 1045–49.
13 World Health Organization. The Global Strategy for Women's, Children's, and Adolescent's Health (2016-2030). Geneva: World Health Organization: 2015.

Lessons learned from the application of the CHNRI method in two RP exercises

These two RP exercises used the CHNRI method to identify research priorities in their respective health areas. The first research prioritisation exercise focused on interventions to improve newborn health and birth outcomes, and 132 experts and 91 scorers participated. The second RP exercise focused on the improvement of the quality of maternal and newborn care through midwifery care. The exercise broadly followed the CHNRI method but modified the way in which research ideas were identified by using the evidence resulting from the analyses in the 2014 *Lancet* series on midwifery. This RP exercise involved 270 scorers.

Lessons learned from the RP exercise on newborn and birth outcomes were fed back into the subsequent RP exercise and how these changes made difference in the subsequent RP exercise are described below.

1. We improved the way of identifying participants in the latter RP exercise, regarding programme experts and health professionals whose participation was lower in the previous exercise. We used only one source of information to obtain list of program experts which also contained list of professional roles other than program experts. We engaged the Partnership for Maternal, Newborn, and Child Health (PMNCH) to ensure that a large and diverse group of experts representing all WHO regions was approached. The PMNCH constituents including non-governmental organizations [NGOs], service user/advocate groups, and healthcare professional associations; academics, researchers and training institutes, donors and foundations, multilaterals [UN], and partner countries were invited to participate in the survey. The list had 1191 experts in total. With this new method, we observed more balanced representation of the different sectors of professional and civil society among the respondents than in the first RP prioritisation exercise.

2. We translated the research ideas into French and Spanish to increase the inclusion of the participants in the regions where English is not the first language. Surprisingly, most RP exercises at global level have been conducted in English language only, including the first RP exercise. We consider it important to make research ideas available in French and Spanish since 24 out of 54 African countries (44%) are francophone, and Spanish is spoken in most Latin American countries. This resulted in increased proportion of response from both Francophone and Spanish speaking countries compared to the previous RP exercise. The language barrier should never be a reason for not participating in the RP exercise.

3. We randomised the order in which research questions were presented to individual experts, to rule out any potential bias in scoring the questions if the order of research questions were

uniform. In the newborn health exercise, all the participants received the questions in the same order for scoring.

4. We used a systematic approach to select criteria. We asked 30 members of the management team to rank 15 CHNRI criteria used in other CHNRI exercises with written rationale as to which were the most important for this exercise. We selected the five most highly ranked criteria that were considered relevant for the exercise. This approach was different to the former exercise in which criteria were selected based on consensus through discussions rather than systematic listing of independently provided scores.

Two exercises leading to a publication in the high impact journals reflect the potential acceptance of the prioritisation exercise, as well as the credibility of the process involved in the exercises. Common challenges and successes in these RP exercises are discussed in more detail in Chapter E.

Chapter 3. Assessment of some key assumptions of the CHNRI methodology

Chapter B presented the application of the CHNRI method in two globally led research prioritisation exercises. The same chapter also reflected on potential limitations of the method and discussed potential methodological improvements to address some limitations noted during the implementation of the method.

Chapter C goes deeper to examine the underlying assumptions of the method. The method is based on crowdsourcing to identify and rank research priorities. It uses the collective opinion of a group of experts to generate and prioritize between many competing research ideas. Though the method uses collective opinion this is impossible to validate since there is no right answer against which to validate "opinions". This chapter therefore investigates collective "knowledge" given that "knowledge" is a critical underlying component of "opinion" and it can be validated for factual questions. Chapter C will examine the accuracy of collective knowledge as compared to individual knowledge, and whether the benefit of collective versus individual knowledge is different in situations in which the knowledge is obtained from experts versus situations in which the knowledge is elicited from non-experts.

The CHNRI method recommends involving a large and diverse group of participants. But how large is optimal and is there any minimum sample size of experts? The CHNRI method relies on purposively selected samples of experts in a certain domain of research as opposed to probability-driven samples. In the absence of clear guidance on the appropriate sample size for purposive samples, most purposive sample sizes are decided upon by those who conduct the research.[23,24] The CHNRI method follows the same logic. The guideline of the method suggests that a "large and diverse" groups of participants are more appropriate for priority setting for health research".[10] However, a question often raised during the planning stage is "how many experts do we need in a CHNRI exercise"? Chapter C investigates the sample size required to obtain stable results using data from four previously conducted CHNRI exercises to provide practical guidance to the future CHNRI users.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? The Journal of Global Health

1.2. When was the work published? June 2016

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes** / **No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I have supported the first author in data compilation and analysis, and provided inputs to the paper drafted by the first author.

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE [redacted] Date 25 July 2018

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

journal of global health

Setting health research priorities using the CHNRI method: V. Quantitative properties of human collective knowledge

Igor Rudan[1], Sachiyo Yoshida[2], Kerri Wazny[1], Kit Yee Chan[1,3], Simon Cousens[4]

[1] Centre for Global Health Research, The Usher Institute for Population Health Sciences and Informatics, the University of Edinburgh, Edinburgh, Scotland, UK

[2] Department for Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland

[3] Nossal Institute for Global Health, University of Melbourne, Melbourne, Victoria, Australia

[4] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

Introduction The CHNRI method for setting health research priorities has crowdsourcing as the major component. It uses the collective opinion of a group of experts to generate, assess and prioritize between many competing health research ideas. It is difficult to compare the accuracy of human individual and collective opinions in predicting uncertain future outcomes before the outcomes are known. However, this limitation does not apply to existing knowledge, which is an important component underlying opinion. In this paper, we report several experiments to explore the quantitative properties of human collective knowledge and discuss their relevance to the CHNRI method.

Methods We conducted a series of experiments in groups of about 160 (range: 122–175) undergraduate Year 2 medical students to compare their collective knowledge to their individual knowledge. We asked them to answer 10 questions on each of the following: (i) an area in which they have a degree of expertise (undergraduate Year 1 medical curriculum); (ii) an area in which they likely have some knowledge (general knowledge); and (iii) an area in which they are not expected to have any knowledge (astronomy). We also presented them with 20 pairs of well–known celebrities and asked them to identify the older person of the pair. In all these experiments our goal was to examine how the collective answer compares to the distribution of students' individual answers.

Results When answering the questions in their own area of expertise, the collective answer (the median) was in the top 20.83% of the most accurate individual responses; in general knowledge, it was in the top 11.93%; and in an area with no expertise, the group answer was in the top 7.02%. However, the collective answer based on mean values fared much worse, ranging from top 75.60% to top 95.91%. Also, when confronted with guessing the older of the two celebrities, the collective response was correct in 18/20 cases (90%), while the 8 most successful individuals among the students had 19/20 correct answers (95%). However, when the system in which the students who were not sure of the correct answer were allowed to either choose an award of half of the point in all such instances, or withdraw from responding, in order to improve the score of the collective, the collective was correct in 19/20 cases (95%), while the 3 most successful individuals were correct in 17/20 cases (85%).

Conclusions Our experiments showed that the collective knowledge of a group with expertise in the subject should always be very close to the true value. In most cases and under most assumption, the collective knowledge will be more accurate than the knowledge of an "average" individual, but there always seems to be a small group of individuals who manage to out–perform the collective. The accuracy of collective prediction may be enhanced by allowing the individuals with low confidence in their answer to withdraw from answering.

Correspondence to:
Professor Igor Rudan
Centre for Global Health Research
The Usher Institute for Population Health Sciences and Informatics
The University of Edinburgh
Teviot Place
Edinburgh EH8 9AG
Scotland, UK
igor.rudan@ed.ac.uk

In 1906, Galton suggested that a group of individuals make better predictions as a collective than any individual expert [1]. Since then, our understanding of the "Wisdom of Crowds" has grown: in recent years, a widely appreciated example of this phenomenon has been evident to the audience of the quiz show "Who Wants To Be A Millionaire?" In this quiz show, a contestant needs to answer a series of increasingly difficult questions by picking from one of four possible responses, only one of which is correct – so that the probability that a random response is correct is 25%. In this show, an "*Ask the audience*" joker is available, whereby 100 persons in studio audience get to submit electronically their opinion on what the correct answer is, and the distribution of their individual opinions is then shown to the contestant. As an alternative, a "*Phone a friend*" joker allows contestants to phone one friend whom they consider the most knowledgeable, and then ask for his/her individual answer. Comparative analyses of the performance of the two jokers showed that the relative majority of the audience chose the correct answer about 91% of the time, while the most knowledgeable friend was right about 65% of the time. There are methodological concerns over the direct comparison between these two percentages, because these success rates were based on different questions, but the difference is still quite striking [1].

Crowdsourcing has become an increasingly popular human tool to address many problems–from government elections in democracies [2], formation of stock market prices [3], to modern online platforms such as TripAdvisor (to advise on the best hotels and restaurants) [4] or Internet Movie Database (IMDb) (to advise on the best movies, TV shows, etc.), all of which are based on the personal opinions of many hundreds or thousands of participants [5]. When crowdsourcing is used for gathering information, or in decision–making processes, there is probably a need to distinguish between at least three different scenarios in which collective knowledge might be used. The first is getting the right answer to a factual question, which we may consider "objective knowledge" and it represents the simplest case. The second is predicting the outcome of some future event, which can subsequently be verified with certainty and within a reasonable time frame. An example is betting on an outcome, eg, of football games or horse races. This is different from stock market predictions, where those who participate in predictions (investors) can also influence the outcomes through their actions. Finally, crowdsourcing could be used to gather information on subjective opinion on something that cannot be easily verified. This last scenario is the closest to how crowdsourcing is used in the CHNRI method (the acronym for: Child Health and Nutrition Research Initiative) [6,7], which seeks to gauge collective optimism with respect to different health research ideas and the benefits they might lead to at some point in the future.

The CHNRI method for setting health research priorities uses "crowds" of experts in global health – researchers, policy makers and programme implementers – to generate, assess and prioritize between many competing ideas in global health research. A CHNRI exercise produces a ranking of many research ideas according to the collective opinion of the expert group, but it is not possible to verify objectively how "valid" that ranking may be, not least because low ranked ideas are unlikely to be funded and therefore no outcomes are available for them. It is yet to be demonstrated that the collective opinion of an expert group should be regarded as more useful than the opinion of individual experts in the group [1,8]. However, the difficulties related to validating personal opinions do not apply to the validation of personal knowledge, and the accuracy of personal knowledge is an important component underlying the individual's opinion. Because of this, we should expect some parallels between the quantitative properties of human collective knowledge and human collective opinion. In this paper, we report several experiments to explore the quantitative properties of human collective knowledge and discuss their possible relevance to the validity of the CHNRI method. The aim of this paper is to examine the accuracy of collective compared to individual knowledge, using different approaches of assessment.

METHODS

We conducted a series of experiments among a group of undergraduate medical students. The number of participating students ranged from 122 to 175 in each exercise. Students who completed the second year lectures in Epidemiology and Statistics, as part of a practical application of epidemiological and statistical concepts were asked to answer 10 questions on each of the following: (i) an area in which they have a degree of expertise (subjects related to the medical curriculum for the first year undergraduate); (ii) an area in which they have some knowledge but do not have expertise (general knowledge); and (iii) an area in which they are not expected to have any knowledge (astronomy). The content of the lecture was entirely unrelated to the questions that were asked from the students. The ethics approval was obtained from a relevant research centre (Centre for Population Health Sciences at the University of Edinburgh).

The questions were chosen so that the answer to each question was numerical (an integer), and so that the answers ranged from a 1–digit number to a 10–digit number over the course of 10 questions in random order, with students unaware of this element of the design. This element was included to allow us to assess whether the students' answers were more accurate when the correct answer was a smaller or larger number (see **Online Supplementary Document**).

Table 1 shows the questions that were asked in each of the three areas, and the correct answers. The questions were asked at the end of 3 consecutive lectures spanning 10 days. Students were given 30 seconds to answer each question. The students were asked to record an answer for every question. For questions for which they were unsure of the answer they were asked to write down their best guess.

In addition, students were shown 20 pairs of well–known celebrities and asked them identify which was the older of the two. **Table 2** shows the pairs of celebrities in the order that the questions were asked. The questions were phrased as: "*Would you say that Celebrity X is older than Celebrity Y?*", and the possible answers were either "*Yes*" or "*No*", where they had to choose one of those two options. However, they were also given an option next to each answer to choose their "secondary" answer as either "*Not sure*" (when they were familiar of both celebrities, but it was too difficult to judge), or leaving the answer "*Blank*" deliberately, when not knowing one or both celebrities. Those two options would indicate their low confidence in their "Yes"/"No" answer. By adding "*Not sure*" (which would be coded with half a point) or "*Blank*" (which would remove them from the sample, leaving the others with more confidence in their answers), they could prevent a wrong answer and increase the chance of the collective answer to be close to the correct answer. This latter type of "scoring" is also used by the CHNRI method. In this way, the same group of students provided two different data sets with scores: one, where they all needed to provide a binary ("Yes"/"No") answer to each question, regardless of their confidence in answering the question correctly; and the other one, where they were able to use the answer "*Not sure*", or leave the answer blank, when they were not confident in their answer. Their input was then turned into a data sheet that was analogous to those produced in the CHNRI exercise, where "*Yes*" was

Table 1. Questions posed to a group of undergraduate Year 2 medical students*

Questions in an area of students' high expertise (undergraduate Year 1 medical curriculum)	
1. How many valence electrons does carbon have?	(4)
2. How many pairs of cranial nerves are there?	(12)
3. How many bones in the adult human body?	(206)
4. In which year did Freud publish "The interpretation of dreams"?	(1900)
5. How many genes does a human have?	(23 000)
6. What is an average salary of a GP in the UK?	(104 000)
7. How many erythrocytes in 1 mL of blood?	(5 000 000)
8. How many refugees are there in the world?	(15 400 000)
9. How many people in the world have diabetes?	(347 000 000)
10. How many bases (A, T, C or G letters) are in the haploid human genome?	(3 000 000 000)
Questions in an area of students' moderate expertise (general knowledge)	
1. How many marriages did Elizabeth Taylor have?	(8)
2. How old was Mozart when he died?	(35)
3. How many minutes does the movie "Casablanca" last?	(102)
4. In which year was Hamlet first published?	(1603)
5. How many diseases in ICD–10?	(14 400)
6. What is the average house price in the UK (in GBP)?	(238 976)
7. How many people live in Cape Town?	(3 740 000)
8. How much was Van Gogh's "sunflowers" painting sold for (in US$)?	(39 700 000)
9. What is the population size of Indonesia?	(246 900 000)
10. How many views did Psy's "Gangham Style" video have to date?	(1 764 039 000)
Questions in an area of student's low expertise (astronomy)	
1. How many light years from our Sun is Sirius?	(9)
2. How many moons does Saturn have?	(62)
3. How many times is Jupiter heavier than Earth?	(318)
4. In which year was Uranus first discovered?	(1781)
5. Distance between our Sun and the centre of Milky Way galaxy (in light–years)?	(27 000)
6. How many times is the Sun heavier than Earth?	(332 900)
7. What is the speed of the solar wind (in Km/h)?	(1 440 000)
8. How many years ago did the comet impact killed off dinosaurs?	(65 000 000)
9. Distance between the Sun and the Jupiter (in kilometres)?	(780 000 000)
10. How many years ago was our Solar System formed?	(4 568 000 000)

*The group was about 170 (range: 167–175) undergraduate Year 2 medical students from: (i) an area of their high expertise (ie, undergraduate Year 1 medical curriculum); (ii) an area where they have some expertise (general knowledge); and (iii) an area where they should have no expertise (astronomy). Correct answers are shown in brackets.

Table 2. Questions posed to a group of 122 undergraduate medical students to guess which well–known celebrity is older than the other*

Pair 1: Justin Bieber vs Miley Cyrus (19 vs 20)
Pair 2: George Clooney vs Brad Pitt (52 vs 49)
Pair 3: Madonna vs Susan Boyle (55 vs 52)
Pair 4: Beyonce vs Shakira (32 vs 36)
Pair 5: Dustin Hoffman vs Robert de Niro (76 vs 70)
Pair 6: Katy Perry vs Rihanna (28 vs 25)
Pair 7: Mick Jagger vs Paul McCartney (70 vs 71)
Pair 8: Lewis Hamilton vs Tiger Woods (28 vs 37)
Pair 9: Angela Merkel vs J. K. Rowling (59 vs 48)
Pair 10: Tony Blair vs George W. Bush (60 vs 67)
Pair 11: David Cameron vs Barack Obama (47 vs 52)
Pair 12: Ashton Kutcher vs Ben Affleck (35 vs 41)
Pair 13: Tom Cruise vs Nicole Kidman (51 vs 46)
Pair 14: Paris Hilton vs Jennifer Anniston (32 vs 44)
Pair 15: Jennifer Lopez vs Britney Spears (44 vs 31)
Pair 16: Eminem vs Jay–Z (40 vs 43)
Pair 17: Kim Kardashian vs Adele (33 vs 25)
Pair 18: Roger Federer vs Andy Murray (32 vs 26)
Pair 19: David Beckham vs Prince Harry (38 vs 29)
Pair 20: Elvis Presley vs Michael Jackson (42 vs 50)

*Correct answers (expressed in years of their age at the time of this exercise) are shown in brackets. The indicated age of individuals is relevant to October 17, 2013. For the last pair, the age at the time of death was being compared. The question was posed as: "Would you say that celebrity X is older than celebrity Y?" and possible answers were "Yes", "No", "Not sure" or "Blank" (see details in the text).

coded as "1", "*No*" as "0", "*Not sure*" as "0.5" and "*Blank*" responses were simply left as blank cells in the data sheet.

This design was carefully developed to allow us to study two questions: (i) how the students' collective opinion performs in comparison to that of individuals when the answers are no longer in a quantitative, but rather in a categorical format; and (ii) whether the *format* of categorical answer (with or without allowing for "*Not sure*" when students' confidence in their answer is low, or "*Blank*" when they simply don't have any knowledge on the question) altered the performance of the students' collective answer. Our hypothesis was that allowing students to answer "*Not sure*" or "*Blank*" would give better results, because it allows the participants within a team who are not sure of the correct answer to "withdraw" from providing their (possibly inaccurate) input, which would give more weight to the responses from students who were more confident in their individual knowledge.

Thus, four different experiments were conducted over the course of four consecutive lectures, which we label "Medical knowledge–quantitative" (MKQ), "General knowledge–quantitative" (GKQ), "Astronomy knowledge–quantitative" (AKQ) and "Celebrity knowledge–categorical" (CKC). In the MKQ, GKQ and AKQ exercises, we conducted the analyses in the following way: (i) we determined the median and the mean response for each of the 10 questions, based on all answers collected from the students (sample sizes were N = 167, N = 175 and N = 170, respectively); (ii) we also developed a parameter that we called "error size", to quantify the extent to which each student deviated from the correct answers over a series of 10 questions, and then we also applied it to the collective median and mean. Given that the responses could both over– or under–estimate the true value, we were interested in the ratio between the larger and the smaller of the two (ie, the correct answer and the answer provided by the student). As an example, this means that, if the correct answer was "10", and one student provided the answer "2" and the other "50", they would be making errors of the "same size": in our evaluation, it was equally wrong to over– or underestimate some value 5–fold. This also means that if the correct answer was provided for each question, then all the ratios contributing to "error size" parameter would be "1". Any deviation from the correct answer in either direction would increase the parameter from this theoretical minimum. (Note that this differs from other possible approaches, such a proportionally expressed increase or decrease, because the latter system would favour under–estimation as a smaller error than over–estimation, and under–estimation would be limited to 100% while overestimation would not be limited in any way). Once the individual errors, expressed as the ratio of the greater vs the smaller of the two values, was determined for each answer to each question, they were summarized for each individual student across all 10 questions and their sum was called "error size". In this way, each student was assigned his/her own "error size" in each of the three exercises (GKQ, MKQ and AKQ), and the students were then ranked by the error size parameter, from the smallest to the largest error made. This was then repeated for the entry of a collective (both using medians and means), and median and mean value rank within the entire student sample was then determined.

In the fourth exercise (CKC), which we designed as a series of 20 "*Yes* or *No*" questions, the task for the students was changed. In the first instance, the collective answer was taken to be the answer given by the majority of students–either "*Yes*" or "*No*". Then, there was an additional methodological caveat. First, those who were not confident about their answer could change some of their answers into the "*Not sure*" option, the effect of which contributed a certain 0.5 points to a total score, and minimised the risk of dropping a whole point for the collective for an incorrect answer. Second, those who had no knowledge of the question (eg, not recognising the names of celebrities) were allowed to change some of their responses to "*Blank*". This would have the effect of reducing the sample size of the collective, leaving all those with no knowledge out, and reducing the overall threshold of correct answers required from other students that the collective would need to answer correctly. Clearly, for those who are confident of their knowledge, this system would mean that they should answer "*Yes*" or "*No*" to all questions and not use either "*Not sure*" or "*Blank*" options at all.

The *correct* answer was then coded as "1", "*not sure*" as "0.5", the *incorrect* answer as "0", and "*blanks*" were excluded from the analysis, thus reducing sample size. The points assigned as described above were added ("1" for correct, "0.5" for "*not sure*", and "0" for incorrect) and then divided by the total number of "*non–blank*" responses received. The result was expressed as "*the percentage of correctness*" of the collective answer, and any value greater than 50% was considered a correct collective answer. This produced two data sheets–CKC1 (where everyone was required to submit either a *Yes* or a *No* answer) and CKC2 (with a *Yes–No–Not sure– Don't know scoring system*). The comparison between the two exercises was expected to reveal if "self–removal" through the use of "*Not sure*" or "*Blank*" improves the score of the collective considerably.

RESULTS

Students' collective answers (median and mean) to the 10 questions in three areas: (i) an area of their expertise, ie, Year 1 medical curriculum; (ii) the area of general knowledge; and (iii) the area outside of their expertise, ie, astronomy are shown in **Tables 3 to 5** (a total of 167, 175 and

Table 3. Year 2 undergraduate medical students' collective answers to the 10 questions in the area of their knowledge*

Question	Correct answer	Students' collective answer–median	Students' collective answer–mean
1. Valence electrons in carbon?	4	4	6
2. Number of cranial nerve pairs?	12	12	13
3. Number of bones in human body?	206	206	210
4. Freud's "Interpretation of dreams" published?	1900	1901	1890
5. Number of human genes?	23 000	38 000†	1 124 128 437
6. Average GP's salary in the UK?	104 100	76 001	85 568
7. Erythrocytes in 1 mL of blood?	5 000 000	8679	12 124 582
8. Number of refugees in the world?	15 400 000	80 000 000	394 267 469
9. Number of people with diabetes?	347 000 000	100 000 000	444 785 232
10. Number of ATCGs in human genome?	3 000 000 000	23 500 327	178 090 845 668

*Number of responses N = 167.

†Question 5 was problematic because the number of human genes was revised down from about 40 000 to 23 000 only recently, ie, after the students learned of the former number; therefore, the median response from students was, in fact, very close to what they were likely to have learnt earlier in the course of their education).

Table 4. Year 2 undergraduate medical students' collective answers to the 10 questions in the area of general knowledge*

Question	Correct answer	Students'collective answer (median)	Students' collective answer–mean
1. Number of marriages of Elizabeth Taylor?	8	4	4
2. How old was Mozart when he died?	35	38	40
3. Minutes duration of "Casablanca"?	102	120	122
4. Year when "Hamlet" was published?	1603	1642	1637
5. Number of diseases in ICD–10?	14 400	48 132	76 480 054
6. Average house price in the UK?	238 976	193 271	369 819
7. Population size of Cape Town?	3 740 196	3 000 000	19 384 089
8. Price of van Gogh's "Sunflowers"?	39 700 000	15 000 000	3 875 825 789
9. Population size of Indonesia?	246 900 000	20 000 000	682 312 629
10. Number of views of "Gangnam Style"?	1 764 039 000	278 000 000	1 610 122 583

*Number of responses N = 175.

Table 5. Year 2 undergraduate medical students' collective answers to the 10 questions in the area outside of their expertise (astronomy)

Question	Correct answer	Students' collective answer (median)	Students' collective answer (mean)
1. Distance Earth–Sirius (in light–years)?	9	6900	5 800 659 084
2. Number of Saturn's moons?	62	12	20
3. How many times Jupiter heavier than Earth?	318	811	5 681 716 865
4. When was Uranus first discovered?	1781	1807	1720
5. Distance Sun–Milky Way Centre (in ly)?	27 000	5 000 000	22 584 267 640
6. How much Sun heavier than Earth?	332 900	8000	8 561 716 703
7. Speed of Solar Wind (in km/h)?	1 440 000	43 027	7 948 573 823
8. Years since comet killed off dinosaurs?	65 000 000	24 564 456	1 396 252 256
9. Kilometres from Sun to Jupiter?	780 000 000	8 728 001	1 239 338 648 469
10. Years since solar system created?	4 568 000 000	7 119 851 052	721 049 090 361

*Number of responses N = 170.

171 responses received, respectively). **Table 6** shows the summary result of the three exercises, presenting both the rank and the percentile of the collective answer (based on either median or mode) among all individual answers provided by the students in three consecutive exercises where students had a decreasing level of expert knowledge. When answering the questions in their own area of expertise, the collective numerical median answer was 35/168 (21st centile) of the most accurate answers; in general knowledge, it was 21/176 (12th centile) most accurate answers; and in an area with no expertise, the group answer was the 12/171 (7th centile). However, the mean value of the collective didn't rank highly in any of the three exercises–in fact, it ranked near the bottom: 127/168 (76th centile) in Year 1 medical knowledge, 164/176 (93rd centile) in general knowledge and 164/171 (96th centile) in astronomy.

Table 7 shows the results of the exercise in recognizing the older of the two celebrities, based on the sample of 122 participating students. The age indicated in the table was relevant to October 17, 2013. All 20 questions were phrased as: "*Would you say that Celebrity X is older than Celebrity Y*?" The possible answers in the first round were

Table 6. The rank and the percentile of the collective answer (based on either median or mean) among all individual answers provided by the students in three consecutive exercises where students had a decreasing level of expert knowledge*

	Collective answer–median			Collective answer–mean		
Exercises on collective knowledge	Rank	Percentile (% top answers)	"Error size" parameter	Rank	Percentile (% top answers)	"Error size" parameter
Medical (Year 1) knowledge	35/168	20.83%	725	127/168	75.60%	48975
General knowledge	21/176	11.93%	38	164/176	93.18%	5430
Astronomy knowledge	12/171	7.02%	1132	164/171	95.91%	663265715

*Addition of the collective answer increased the total number of received answers by one, resulting in 168, 176 and 171 responses being ranked in each exercise, respectively; percentile of eg, 20.83% means that the collective response ranked among the 20.83% most accurate individual responses).

Table 7. Results of the exercise in recognizing the older of the two celebrities (N = 122)*

Older celebrity	Younger celebrity	Difference (years)	% correct (2–category system: yes/no)	% correct (4–category system: yes/no/ns/b)
Roger Federer (32)	Andy Murray (26)	6	97%	97%
George Clooney (52)	Brad Pitt (49)	3	95%	96%
David Beckham (38)	Prince Harry (29)	11	96%	96%
Tiger Woods (37)	Lewis Hamilton (28)	11	93%	95%
Jennifer Aniston (44)	Paris Hilton (32)	12	97%	94%
Miley Cyrus (20)	Justin Bieber (19)	1	93%	92%
Ben Affleck (41)	Ashton Kutcher (35)	6	85%	85%
George W. Bush (67)	Tony Blair (60)	7	85%	80%
Kim Kardashian (33)	Adele (25)	8	82%	79%
Jennifer Lopez (44)	Britney Spears (31)	13	83%	78%
Angela Merkel (59)	JK Rowling (48)	11	71%	73%
Michael Jackson (50)	Elvis Presley (42)	8	75%	67%
Barack Obama (52)	David Cameron (47)	5	66%	62%
Tom Cruise (51)	Nicole Kidman (46)	5	64%	60%
Katy Perry (28)	Rihanna (25)	3	63%	59%
Jay–Z (43)	Eminem (40)	3	56%	57%
Dustin Hoffman (76)	Robert de Niro (70)	6	44%	52%
Paul McCartney (71)	Mick Jagger (70)	1	59%	52%
Madonna (55)	Susan Boyle (52)	3	55%	51%
Shakira (36)	Beyonce (32)	4	43%	43%

*The questions were phrased as: "Would you say that Celebrity X is older than Celebrity Y?". The possible answers in the first round were "Yes" or "No" (2–category system); and in the second round the students were also allowed "Not sure" (when they were familiar of both celebrities, but it was too difficult to judge) and leaving the answer "Blank" deliberately (when not knowing one or both celebrities), in order to increase the chance of the entire collective of students to answer correctly. The latter type of "scoring" is used in the CHNRI method.

"Yes" or "No" (2–category system); and in the second round the students were also allowed "Not sure" (when they were familiar of both celebrities, but it was too difficult to judge) and leaving the answer "Blank" deliberately (when not knowing one or both celebrities), in order to increase the chance of the entire collective of students to answer correctly. The latter type of "scoring" is used in the CHNRI method.

The results show that, when everyone needed to provide a "Yes" or "No" answer, regardless of their confidence in their own answer, the collective was correct in 18/20 cases (90%), with 8 students outperforming the results of the collective–all of them with 19/20 correct answers (95%). This means that the collective answer based on this type of response ranked in the top 7.3% of individual answers. However, when the students were allowed to use the system of responses in which those who were not confident of their answer were allowed to ask for half a point, or withdraw from responding entirely, in order to improve the scores of the collective, the results changed somewhat. Looking at all specific celebrity pairs, they were not clearly better than when everyone gave an answer regardless of their confidence in being correct. However, with this type of scoring the collective was correct in 19/20 cases (95%), while the 3 most successful individuals among the 122 students now had 17/20 correct guesses (85%). This clearly shows that many students opted to only receive half a point, or withdrew, because the small group among them who gave best individual answers did not repeat the level of success from the first round of scoring in this second round–although they did manage to further improve the collective answer. A subsequent analysis showed that the median frequency of choosing the "*Not sure*" answer when this was possible was 44 (range: 3–59), or about one third of students, with very wide range–depending on the level

of difficulty of the question. The option "blank" was used much less frequently, with a median of 7 (range: 0–35).

The **Online Supplementary Document** presents several additional analyses. Figures S1–S3 show that the number of digits of the correct answer does not seem to be related to the likelihood that the group will identify the correct answer–this only seemed to possibly be the case in the exercise where students had expertise (ie, Year 1 medical curriculum questions), but was not replicated in the other two exercises. Figure S4, related to the fourth exercise, shows that the proportion of those guessing correctly in the group was associated with the age difference between the two celebrities, as might be expected.

DISCUSSION

The analyses conducted in this study tried to provide insights into quantitative properties of human collective knowledge, many of which are relevant to better understanding of the properties of the CHNRI method as originally proposed. First, the CHNRI method relies on the opinion of experts that is based on their knowledge of a specific subject, and asks them to express their optimism about research ideas through scores. Through this series of exercises we wanted to explore if this approach is likely to result in better predictions than if persons with limited knowledge of the subject are also invited to prioritize health research, or if persons with no knowledge at all are invited. In the student exercise in their own area of expertise (Year 1 medical curriculum, **Table 3**), the first 5 answers given by the students as a collective median value were all exactly right or extremely close (taking into account that the number of genes in the human genome was indeed close to 40 000 in their earlier textbooks, and it was only revised down to about 23 000 more recently). This level of precision was not observed in their responses to general knowledge questions (**Table 4**), or questions on astronomy (**Table 5**).

However, there are worrying signs that, when the majority of students don't know the correct answer to a question that should be covered by their expert knowledge, the collective median can be very wrong. The examples are the case of the number of erythrocytes in 1 mL of blood (where the collective median was 3 orders of magnitude smaller than the correct value) or the number of nucleotides in the human genome (where the underestimate was by 2 orders of magnitude) (**Table 3**). Because of those two questions, where most of the students didn't even know the right order of magnitude, the parameter "error size" of the collective median was even greater for the exercise on Year 1 medical knowledge, than it was for the exercise in general knowledge (**Table 6**). Although this may seem surprising at first, it can be easily explained. The parameter "error size" is very sensitive to the size of the departure from each of the 10 correct answers. In general knowledge questions, collective median answers were always reasonably close to the correct answers in terms of students' being able to guess the correct order of magnitude for the answer, as all the questions were related to topics in which they had at least some knowledge. However, a specific question in their own area of expertise in which they had no knowledge could quickly lead to very large departures from the correct answer. It would be difficult, given a small sample size, to reach a definite conclusion that there are some experts who do better than the crowd–"*the superforecasters*" [8], although this remains a possibility.

The exercise in the knowledge of astronomy (**Table 5**) was interesting because it clearly showed that humans do not possess a "cryptic" ability to collectively predict values on which they do not have any knowledge as individuals with any precision. This suggests that "wisdom of crowds" only works when the majority of participants in the group have at least some private knowledge of the quantity that is being predicted. As an example, the students had some intuition on the possible year when Uranus could have been discovered, the number of Saturn's moons, or even the number of years since the Solar system was created–they got the order of magnitude correct in those three questions. However, when asked about quantities of which they knew nothing, nor had any intuition, they were typically wrong by several orders of magnitude when their collective medians were compared to the correct answers.

Collective medians typically performed well across all three exercises: the collective median was among the 20.83% of the most accurate responses in the medical knowledge, 11.93% in the general knowledge, and 7.02% in the astronomy knowledge. We propose that the collective median is actually not among the top 10% scores in the area of expertise, because there is a smaller group of students among the entire cohort with excellent knowledge, and who would be seen as the top of their class. These students know the correct answers and the rest of the class simply dilutes their accuracy and moves the collective median away from the perfectly accurate response. We believe that this explains why the collective median in the area of expertise was only at the 21st percentile of the most accurate answers. However, as the collective moves towards answering the questions outside of the area of their expertise, the collective median begins to move up the ranks. Once there are no longer individuals who could easily answer all 10 questions with high accuracy, the collective median progresses to the 12th percentile (in the general knowledge exercise) and 8th percentile (in astronomy exercise).

We propose a mathematical explanation for this, which is relevant to the relationship between the correct answer and

the distribution of all responses in a series of questions. After each question, the collective median will be exactly at the 50th percentile of answers. When the distribution of answers is compared to the correct answer, the error size of the median will either be at the 50th percentile of the group or smaller. For individual students who don't have any knowledge on the subject and are simply guessing, they can expect to alternate between a position above and below the 50th percentile randomly, and occasionally making gross mistakes. After enough time and many iterations, the collective median of a group who are guessing entirely unknown quantities will always be either at the 50th percentile, or above, while the rest of individual answers will be above or below the 50th percentile half of the time. After a sufficient number of questions, this should ensure that the collective median acquires Rank 1, because median can sometimes be very close to a correct answer, but never worse than 50th percentile of all group's guesses. This protects it from gross errors that all other students will eventually experience over a large number of guesses. This may be a general mechanism that explains why collective median eventually outperforms individuals in a long time series of predictions of entirely unknown quantities.

All of the above is relevant to collective medians. Turning our attention to collective means, they did not fare well at all. They were at the 76th percentile of ranks in the area of medical knowledge, 94th in the area of general knowledge, and 96th in the area of astronomy. We found the explanation to this poor performance in a number of extremely wrong predictions made by several individuals, who made mistakes of such magnitude that they completely dominated the collective mean. Because of this, we suggest that – when the answers are being predicted in a quantitative form – medians will be more reliable than the means. One question that could be raised here is whether the entire cohort of medical students can be trusted to take this sort of exercise seriously, because if a small group deliberately put down extreme responses, this would certainly have an effect of skewing the mean.

The exercise in "guessing the older of the two celebrities" allowed us to establish that, in an area of "relative" expertise (because it has become difficult to avoid information on the celebrities that were chosen). There is considerable accuracy in collective prediction when "Yes"/"No" answers are allowed and the answer given by the majority is chosen as the correct one. The collective was correct in 90% of cases, and this translated to the rank 9/123 (8th percentile in the ranks), with 8 individuals who recorded 95% of correct answers and out–performed the collective. This exercise was analogous to a large extent to the "Ask the audience" joker that is used in the quiz show "Who wants to be a millionaire?", as mentioned earlier, and the accuracy of 90% is very similar to the one of about 91% observed in the quiz show.

The key question in this exercise was whether the collective response could be further improved by allowing some individuals, who were not confident in their answers, to minimise the "damage" to the collective by choosing "not sure" (which still gives them a guaranteed 50% of available points) or to drop out from the sample. When this option was given, the accuracy of the collective answer increased to 95%, while the three best individual answers only achieved 85%. A question–by–question comparison of 20 individual answers between the two types of scoring doesn't indicate that the collective answer with the 2nd type of scoring (4 options) is consistently better than the binary "Yes"/"No" type of scoring, so we cannot be sure that this finding is generalizable, rather than a chance effect, and we should continue to explore this with more questions and using larger sample sizes to confirm it.

We will now consider how the findings of this study are relevant to "validation" of the CHNRI method. This study shows that the collective knowledge in an area of expertise is likely to lead to more accurate responses than the collective knowledge in an area outside of the expertise. Moreover, the exercise shows that it may be better to only invite a reasonably small, highly selected group of experts and rely on their collective prediction, rather than trying to seek expertise from a large group, which may lead to deviations from the optimal collective prediction. This justifies the strategy that has been used in many early CHNRI exercises, where as few as 10–15 leading experts in a narrow research field were invited to conduct the exercise on setting research priorities in their field. Moreover, the type of response used in CHNRI exercises ("Yes" – "No" – "Not sure" – "Blank") seems to slightly improve the collective prediction in comparison to the alternative, where all scorers are forced to choose between only two binary options. However, the difference between the two types of scoring resulted in predictions that could be considered surprisingly similar, so further experiments will need to resolve whether there is a real difference between the two approaches or not. If there is no difference, then perhaps the "Yes"/"No" answer could be preferred as simpler and more discriminative in the process of prioritisation, because too many "not sure" answers lead to scores that show regression to the mean and the discriminatory power of the scoring process is gradually lost. This, therefore, remains an unresolved question that warrants further investigation.

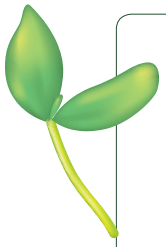
Applications of "crowdsourcing" are finding ways into many areas of human activity. In parallel, many interesting scientific experiments are being performed to improve our understanding of the principles underlying and governing crowdsourcing. Recent studies showed that sharing the in-

formation on confidence in their answers between the individuals in the group can substantially improve the prediction of the group, as we could see in our study (**Table 7**), but if those most confident are wrong, then it can also lead the collective opinion to dramatically wrong decisions [8,9]. Independence of the provided opinion, such as in the CHNRI exercise, is very important because studies have convincingly shown that interactions between participants in the group and social influence may both improve and undermine the "wisdom of crowds" effect [10,11]. We should also mention that this research was conducted in "artificial", well–controlled conditions, but in the real world every group will have its own unique dynamics. In many contexts, collective knowledge, opinion or intelligence may not be the main factor influencing the decisions, which is a limitation of this type of research and of its applications in complex real–world scenarios.

There seems to be agreement between researchers that select groups of "best–performing" experts can reach an optimal collective result with sample sizes as small as five, which cannot be easily improved by increasing sample size [12,13]. This observation has a potential practical application in the field of medical diagnostics [13]. However, it has also been shown that a well–designed mathematical or statistical model would still outperform any collective human opinion [13]. Two further interesting applications of crowdsourcing in the fields of medicine and health research have been proposed recently. One study proposed that, in the absence of clear guidelines on indications, stabilization of the prevalence of use of certain drugs–such as antidepressants–at the level of the whole population might indicate the optimal usage. This is because the stabilized frequency at the population level is likely to reflect hundreds of thousands of decisions on continued usage, made by treated individuals based on their personal experiences [14]. Finally, it has been proposed that complex, expensive and bureaucratic processes of research evaluations, such as the Research Excellence Framework (REF) that takes place every 6 years in the UK, could be replaced by crowd–sourced "prediction markets" [15]. Prediction markets enable individuals to trade "bets" on whether a specific outcome would occur or not, and they have been shown to be successful at predicting outcomes in different areas of human activity, such as sport, entertainment and politics. Given that they are based on expert judgements, which also form the basis of REF in the UK, there is no reason why prediction market could not theoretically offer an alternative to the REF that could be updated annually, or even track the performance in real time [15].

CONCLUSION

Our experiments showed that the collective knowledge of a group with expertise in the subject should always be very close to the true value. In most cases and under most assumptions, the collective knowledge will be more accurate than the knowledge of an "average" individual, but there always seems to be a small group of individuals who manage to out–perform the collective. The accuracy of collective prediction may be enhanced by allowing the individuals with low confidence in their answer to withdraw from answering. This study showed that the CHNRI method is based on the premises and designs that are likely to maximise the predictive value of the group: experts are being invited to score proposed research ideas (instead of persons with limited knowledge, or lay persons); experts are providing their answers independently (to protect the end result from social influences); and they are using the scoring system that is expected to maximise the accuracy of the collective answer over the individual ones.



Acknowledgements: The authors thank the Year 2 medical students of the University of Edinburgh for their participation in this study.

Funding: This work was supported by the CHNRI Foundation award to the University of Edinburgh.

Ethical approval: Obtained from the Centre for Population Health Sciences, the University of Edinburgh.

Authorship declaration: IR developed the idea for the paper. IR and KYC conducted the student exercise. IR and SY co–wrote the paper. IR, SY and SC designed and performed the analyses. SC, SY, KYC and KW provided important intellectual content.

Declaration of interest: IR is an editor–in–chief of the Journal of Global Health. To ensure that any possible conflict of interest relevant to the journal has been addressed, this article was reviewed according to best practice guidelines of international editorial organizations. The authors completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author). Authors declare no conflicting financial or other interest related to the work detailed in this manuscript.

REFERENCES

1 Surowiecki J. The wisdom of crowds. London: Abacus, 2013.
2 Surowiecki J. Democracy: dreams of the common good. In: Surowiecki J. The wisdom of crowds. London: Abacus, 2013.
3 Graham B. The intelligent investor. New York: Harper Collins, 2003.
4 Tripadvisor. Available at: https://www.tripadvisor.co.uk/. Accessed: 15 April 2016.</eref>
5 Internet Movie Database. Available at: http://www.imdb.com/. Accessed: 15 April 2016.
6 Rudan I, Chopra M, Kapiriri L, Gibson J, Lansang MA, Carneiro I, et al. Setting priorities in global child health research investments: universal challenges and conceptual framework. Croat Med J. 2008;49:307-17. Medline:18581609 doi:10.3325/cmj.2008.3.307
7 Rudan I, Gibson JL, Ameratunga S, El Arifeen S, Bhutta ZA, Black M, et al. Setting priorities in Global child Health Research Investments: guidelines for implementation of the CHNRI Method. Croat Med J. 2008;49:720-33. Medline:19090596 doi:10.3325/cmj.2008.49.720
8 Tetlock PE, Gardner D. Superforecasting: the art and science of prediction. London: Random House Books 2015.
9 Koriat A. When are two heads better than one and why? Science. 2012;336:360-2. Medline:22517862 doi:10.1126/science.1216549
10 King AJ, Cheng L, Starke SD, Myatt JP. Is the true "wisdom of the crowd" to copy successful individuals? Biol Lett. 2012;8:197-200. Medline:21920956 doi:10.1098/rsbl.2011.0795
11 Lorenz J, Rauhut H, Schweitzer F, Helbing D. How social influence can undermine the wisdom of crowd effect. Proc Natl Acad Sci U S A. 2011;108:9020-5. Medline:21576485 doi:10.1073/pnas.1008636108
12 Mannes AE, Soll JB, Larrick RP. The wisdom of select crowds. J Pers Soc Psychol. 2014;107:276-99. Medline:25090129 doi:10.1037/a0036677
13 Kattan MW, O'Rourke C, Yu C, Chagin K. The wisdom of crowds of doctors: their average predictions outperform their individual ones. Med Decis Making. 2016;36:536-40. Medline:25878196 doi:10.1177/0272989X15581615
14 Patten SB. The wisdom of crowds (vox populi) and antidepressant use. Clin Pract Epidemiol Ment Health. 2015;11:1-3. Medline:25674154 doi:10.2174/1745017901511010001
15 Munafo MR, Pfeiffer T, Altmejd A, Heikensten E, Almenberg J, Bird A, et al. Using prediction markets to forecast research evaluations. R Soc Open Sci. 2015;2:150287. Medline:26587243 doi:10.1098/rsos.150287

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed for each 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? The Journal of Global Health

1.2. When was the work published? June 2016

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes** **No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I have collected the data, conducted the data analysis and wrote the first draft of introduction, method and result section.

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE [redacted] Date 25 July 2018

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

journal of global health

Setting health research priorities using the CHNRI method: VI. Quantitative properties of human collective opinion

Sachiyo Yoshida[1], Igor Rudan[2], Simon Cousens[3]

[1] Department for Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland

[2] Centre for Global Health Research, The Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, Scotland, UK

[3] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

Introduction Crowdsourcing has become an increasingly important tool to address many problems – from government elections in democracies, stock market prices, to modern online tools such as TripAdvisor or Internet Movie Database (IMDB). The CHNRI method (the acronym for the Child Health and Nutrition Research Initiative) for setting health research priorities has crowdsourcing as the major component, which it uses to generate, assess and prioritize between many competing health research ideas.

Methods We conducted a series of analyses using data from a group of 91 scorers to explore the quantitative properties of their collective opinion. We were interested in the stability of their collective opinion as the sample size increases from 15 to 90. From a pool of 91 scorers who took part in a previous CHNRI exercise, we used sampling with replacement to generate multiple random samples of different size. First, for each sample generated, we identified the top 20 ranked research ideas, among 205 that were proposed and scored, and calculated the concordance with the ranking generated by the 91 original scorers. Second, we used rank correlation coefficients to compare the ranks assigned to all 205 proposed research ideas when samples of different size are used. We also analysed the original pool of 91 scorers to to look for evidence of scoring variations based on scorers' characteristics.

Results The sample sizes investigated ranged from 15 to 90. The concordance for the top 20 scored research ideas increased with sample sizes up to about 55 experts. At this point, the median level of concordance stabilized at 15/20 top ranked questions (75%), with the interquartile range also generally stable (14–16). There was little further increase in overlap when the sample size increased from 55 to 90. When analysing the ranking of all 205 ideas, the rank correlation coefficient increased as the sample size increased, with a median correlation of 0.95 reached at the sample size of 45 experts (median of the rank correlation coefficient = 0.95; IQR 0.94–0.96).

Conclusions Our analyses suggest that the collective opinion of an expert group on a large number of research ideas, expressed through categorical variables (Yes/No/Not Sure/Don't know), stabilises relatively quickly in terms of identifying the ideas that have most support. In the exercise we found a high degree of reproducibility of the identified research priorities was achieved with as few as 45–55 experts.

Correspondence to:
Sachiyo Yoshida
Department for Maternal, Newborn, Child and Adolescent Health
World Health Organization
Avenue Appia 20
CH–1211 Geneva 27
Switzerland
yoshidas@who.int

In 1906, Galton suggested that a group of individuals tend to make better predictions as a collective than any individual. Since then, our understanding of collective decision–making, termed by some as the "Wisdom of Crowds", has grown considerably [1]. Crowd–sourcing has become an increasingly important human tool to address many problems – from government elections in democracies [2], formation of stock market prices [3], to modern online platforms such as TripAdvisor (to advise on the best hotels and restaurants) [4] or Internet Movie Database (to advise on the best movies, TV shows, etc.) [5], all of which are based on personal opinions of many hundreds or thousands of participants. The CHNRI method (the acronym for: Child Health and Nutrition Research Initiative) also uses crowdsourcing as the major component of the process to set priorities among many competing health research ideas [6,7]. It relies on large groups of scientists who are invited to participate in each exercise. Within the CHNRI process, several dozens (or even hundreds) of scientists are typically invited first to generate, and then to assess many competing health research ideas using a pre–defined set of priority–setting criteria. Their collective optimism towards each research idea with respect to specific criteria is measured and the research ideas are then ranked according to the scores they achieve across all criteria.

However, researchers typically question several concepts in relation to the "validity" of the CHNRI exercises. The first question is fundamental to the entire process, asking the developers of the method to demonstrate convincingly that the opinion of a large expert group is more reliable and trustworthy than the opinion of only one, or a very small number of experts. This question has been addressed in a previous paper in this series [8], which demonstrated that the collective knowledge of a group (rather than opinion) generally outperforms the knowledge of any single individual. While for factual knowledge there is a "gold standard" against which we can compare the response of the collective to that of individuals, for opinions about future outcomes there is no such "gold standard". Nevertheless, given that individual knowledge, or lack of it, underlies a significant part of individual opinion, and that the same governing principles that make the collective knowledge superior to individual knowledge (described in our previous paper [8]) should also apply to opinion, we consider this question largely addressed. The substantial literature on so–called "prediction markets" provides further evidence of the reliability and effectiveness of collective opinion in comparison to individual opinion in predicting future events [9,10].

The second question concerns the "optimal" sample size of researchers to be invited to conduct a CHNRI exercise. Here, "optimal" refers to a minimum number of experts needed from a larger, global "pool" of experts, in order to reduce the cost and complexity of conducting the exercise while obtaining a replicable collective opinion. The question of the "sufficient" sample size can be investigated by exploring at which point addition of further experts from the larger, global "pool" of experts ceases to influence the outcomes of the CHNRI process. The third question is related to the composition of the sample of experts, and how this composition can potentially affect the final scores. Do the background characteristics of the experts invited to participate affect their collective opinion in such a way that one subgroup of experts would provide systematically different scores from another subgroup?

In this article, we address the latter two questions by exploring some of the quantitative properties of human collective opinion. We study the special case where the collective opinion is based on a set of individual opinions, all of which are expressed in the form of simple categorical variables. These variables relate to the optimism expressed by each participating expert regarding the extent to which each proposed research idea meets the different priority–setting criteria [6,7]. The opinion provided by the participating experts can be expressed as "Yes" (equals 1), "No" (equals 0), "Not sure" (equals 0.5) and "I don't know" (equals blank input), which is the typical input required in the CHNRI method. This special case is of particular interest, because in our previous paper [8] we demonstrated the effectiveness of this method of expressing individual opinion in comparison to other types. Finally, one of the concerns about this way of collecting opinion from groups of experts is the impact of low response rates and subsequent self–selection bias. We will mention this concern here because we find it potentially very important, although it will be difficult to study and we will not attempt to address it in this paper.

METHODS

In order to answer the latter two questions posed in the introduction, we conducted statistical analyses of the inputs provided by the group of experts who took part in a previous CHNRI exercise. These analyses focused on identifying whether there was a point of "saturation" in collective opinion. "Saturation" here refers to the idea that beyond a certain sample size of experts, adding further experts' opinions does not significantly change the results of the process. To study this, we used the data set with quantitative input from the experts who took part in a CHNRI exercise on newborn health in this series [11], which is freely available as a supplementary online material to the article in question [11]. All input was provided in the form of a simple categorical variable (ie, optimism towards each idea expressed as "Yes" (equals 1), "No" (equals 0), "Not sure" (equals 0.5) and "I don't know" (equals blank input)).

Our analysis strategy involved drawing many random sub–samples, with replacement, from the full sample of 91 expert participants in the CHNRI exercise on newborn health. The experts scored a set of 205 proposed research ideas [11]. Our aim was to identify the minimum sample size of experts required to produce stable results. We used two metrics to assess stability. First, we compared the 20 most highly ranked ideas for each resampled data set with the 20 most highly ranked ideas in the whole data set (ie, all 91 experts) and calculated how many ideas appeared in both top 20 lists. If all the opinions were assigned entirely at random, then we would only expect about 2 research ideas on average (out of the total of 205) to be in common across two samples. Given this reasonably low expected agreement by chance, we arbitrarily defined results as being stable when 15 (or more) of the 20 highest ranked ideas were concordant with those based on the opinion of the full sample of 91 experts. We believe that such an occurrence indicates a high level of stability/replicability compared with the 2 expected purely by chance.

Previous studies into the point of saturation in collective opinion

The question of the sample size at which the "saturation" of information occurs has been vigorously discussed over many years in relation to qualitative research, where interviews conducted with the participants are recorded and analysed to obtain insights into a wide variety of research topics. In qualitative research, saturation is typically described in the context of obtaining the "appropriate" sample size at which no new ideas, findings, or problems are found. Determining the "appropriate" sample size is critical, because a sample that is larger than needed would result in inefficient use of research funds, resources and time. On the other hand, too small a sample size may result in limited validity of the research findings.

The idea of "saturation" was first introduced in the late 1960s [12] through the notion that, though every research participant can have diverse ideas in principle, the majority of qualitative studies will inevitably reach a point of saturation. Since the work by Glaser and Strauss [12], researchers have attempted to provide sample size guidance for various research disciplines. Proposed sample sizes have ranged from fifteen in all qualitative research disciplines [13] to sixty [14] in the area of ethnographic interviews. These proposed sample sizes were rarely accompanied by a clear justification or description of how they were derived.

However, the idea of saturation does not necessarily translate to CHNRI exercises, where opinions are submitted in a form of quantitative categorical variables. This gives us perhaps a rare opportunity to perform an assessment of the quantitative properties of human collective opinion by analysing a data set underlying a typical CHNRI exercise. We found one study that attempted to analyse the stability of responses of the 23 health care and patient safety experts who participated in a Delphi survey using a categorical rating scale [15], which is the most similar case to the CHNRI process that we were able to find in the literature. In that study [15], the responses to each item were scored on a rating scale from 1 to 4, with "1" being unimportant to "4" being very important. The responses obtained in the first round of the survey were processed using sampling with replacement to produce hypothetical samples of 1000 and 2000 participants, from the initial sample size of 23 subjects. Then, means and 95% confidence intervals for the scores of the original 23 participants were compared with the hypothetical samples. Substantial similarity of inferential statistics between the actual and hypothetical samples was observed, from which the authors concluded that the "stability" of results was already achieved with only 23 actual study participants [15]. Clearly, this interpretation was limited by having an original sample as small as 23 individuals to generate large bootstrapping samples, and the result needs to be replicated using a larger initial sample of individuals to generate bootstrapping samples. In our study, the key improvement will be drawing sub–samples smaller than the original sample, while in the approach described in this study samples were created that were much larger than the original sample – which is an approach with major limitations.

Defining "saturation" in our study

In our study, we defined "saturation" in two ways. First, we defined it as the point where we observed replicability in the collective rankings of top 20 research ideas (among a total of 205 assessed) between two randomly generated sub–samples of a given sample size. In other words, involving further experts would no longer be expected to make any important difference to the 20 most highly ranked priorities. Given that randomness inherent to the process of sampling makes it unrealistic to expect all 20 priorities to always replicate at a certain sample size, and taking into account low "a priori" probability of replication (only 2 among the 20 most highly ranked research ideas would be expected to replicate by chance alone), we needed to define "saturation" arbitrarily. We considered the specific sample size as "saturation–reaching" when the same 15 (or more) research ideas in any two randomly generated samples of a specific size were expected to be found among the 20 most highly ranked research ideas in both samples.

Second, we used Spearman's correlation coefficient to compare the ranks assigned to all 205 proposed research ideas by the randomly generated sub–samples with the ranks derived from the full sample. We considered "saturation" to be achieved when the median rank correlation coeffi-

cient reached or exceeded 0.95 (which is an extremely high rank correlation coefficient). We believe that both definitions of saturation are stringent and conservative from the statistical point of view.

Database used in this analysis

We used anonymised raw scores provided by the participants in the CHNRI exercise on newborn health [11]. The database included all individual scores from 91 participating experts that were assigned to all 205 proposed research ideas using 5 pre–defined criteria. The criteria used in the exercise are summarized in **Box 1**, and they were posed in the form of simple "yes/no" questions. The requested input was provided in the form of numbers: 0 (meaning "no"), 0.5 ("informed, but undecided answer"), 1 ("yes"), and blank ("insufficiently informed"). "Blank" was used whenever the participants did not feel that they possessed enough technical knowledge to be able to answer, which is different from an "informed, but undecided" answer, where the expert could neither agree nor disagree although they felt that they had enough knowledge on the topic.

Statistical analysis

We used resampling with replacement, sometimes referred to as "bootstrapping", to simulate the diversity of samples drawn from a larger global pool of experts. All analyses were performed using the statistical program STATA 13.0 (www.stata.com). To study how the rankings assigned to proposed research ideas change and converge with increasing sample sizes of experts, we generated samples ranging in size from minimum 15 to a maximum of 90. For each selected sample size, 1000 random bootstrap samples were drawn.

Two statistical analyses were then performed to examine how the ranking list of research ideas changed as the number of experts contributing to the CHNRI exercise increased. In the first analysis, we examined the concordance in the top 20 research ideas between 1000 randomly generated subsamples of the same size that were developed using the bootstrap method. In the second analysis, we used Spearman's rank correlation coefficient to examine the concordance in the ranking order of all 205 research ideas between 1000 randomly generated subsamples of the same size that were developed using the bootstrap method.

Box 1. The five criteria used in the exercise.

Criterion 1. Answerability: Can the research question be answered ethically?

Criteria 2. Efficacy/Effectiveness: Can the new knowledge lead to an efficacious intervention or programme?

Criteria 3. Deliverability and acceptability: Is the proposed intervention or programme deliverable and acceptable?

Criteria 4. Maximum potential for disease burden reduction: Can the intervention or program improve newborn health substantially?

Criteria 5. Effect on equity: Can the interventions on program reach the most vulnerable groups?

Analysis of subgroups within the full sample

Research priority scores (RPS) were recalculated for each research question in sub–samples of scorers that were defined by participants' self–classified background and the country in which they were based. Participants originally classified themselves as researchers, policy makers, donor representatives, program managers or health practitioners (multiple choices were not allowed), and this information is available in the original paper [11]. In this exercise, we had combined all categories other than researcher into one category as "non–researcher", as the numbers of participants falling into each of the non–researcher categories were small. The country where the scorer was based was classified by the level of income as either a "high–income country" (HIC) or a "low– or middle–income country" (LMIC), using the World Bank's categorization [16]. We explored: (i) the differences in median scores that different sub–groups of scorers (ie, researchers vs non–researchers; and HIC–based vs LMIC–based) assigned to different criteria; the median scores were determined across all 205 research ideas to investigate whether subgroups of scorers systematically scored particular criteria differently; (ii) the overlap between the top 20 research ideas identified by different sub–groups of scorers (ie, researchers vs non–researchers; and HIC–based vs LMIC–based).

RESULTS

Figure 1 shows the how concordance with respect to the top 20 priorities increased as the number of sampled scorers increased. Note that when resampling 90 scorers with replacement, concordance with the top 20 priorities based on the original sample of 91 experts would not be expected to reach 100%. This reflects the fact scores derived from the original sample of 91 experts are themselves subject to sampling variation. The median concordance (across the 1000 sub–samples drawn for each sample size) increases from 12/20 (60%) with a sample size of 15 to 15/20 (75%) with a sample size of 55 experts. Thereafter there is no clear improvement in concordance with increasing sample size. The interquartile range for concordance with a sample size of 55 is 14/20 to 16/20 (70% to 80%) and this also appeared relatively stable as sample sizes were increased further. At a sample size of 90, the median concordance was 16/20 (85%) (IQR 15–16). Given that this gives an indication of the variability of the sample size we had available

to us for analysis, it appears that relatively stable results can be achieved with sample of 50 experts (median 14, IQR 13.5–15). There is little further increase in achieved overlap by increasing the pool of experts from 50 to 90 (**Figure 1**).

Figure 2 shows the relationship between the sample size of the scorers within the CHNRI newborn health exercise [11] and the median, IQR and range of Spearman's rank correlation for the ranks of all 205 proposed research ideas. As expected, the rank correlation coefficient increases as sample size becomes larger and a median correlation of 0.95 was reached at a sample size of 45 experts (median of the rank correlation coefficient = 0.95; IQR 0.94–0.96).

Among the 91 scorers in the newborn health exercise, 61 self–classified as "researchers" and 30 as "non–researchers"; 53 participants were based in HIC and 38 in LMIC. **Table 1** shows the differences in median scores (with inter–quartile range, IQR) that different subgroups of scorers (ie, researchers vs "non–researchers"; and high–income country (HIC)–based vs low– or middle–income country (LMIC)–based) assigned to different criteria. The differences between researchers and non–researchers were small, with non–researchers being slightly more optimistic about maximum potential impact, but all differences were well within the limits predicted by inter–quartile ranges. Larger differences were observed between HIC–based and LMIC–based researchers, with the latter tending to provide more optimistic scores, ranging from a 7 to a 24 point–difference on a scale from 0 to 100. The smallest difference was noted for answerability, followed by effectiveness and

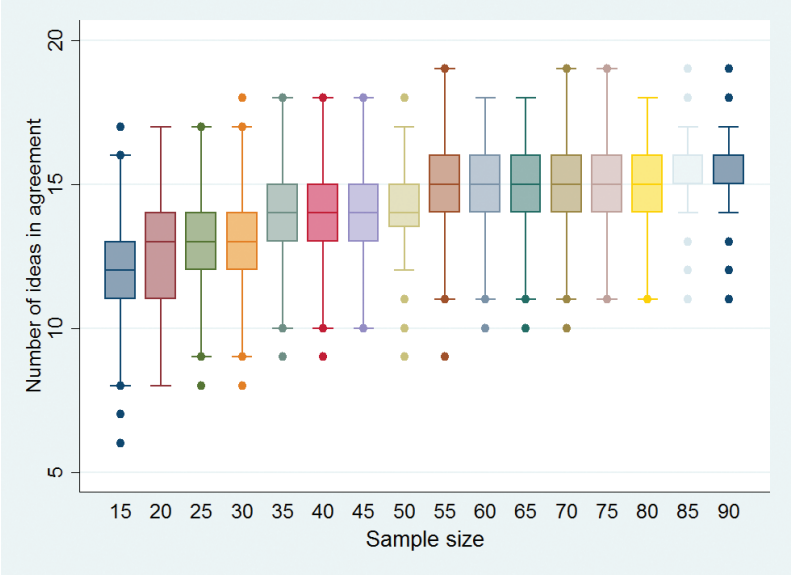


Figure 1. Level of overlap among the top 20 ranked research ideas (Y–axis) by the size of the sample of randomly selected experts (X–axis) from a total pool of 91 experts using a bootstrap method (simulation 1000 times with replacement of already selected experts, using bsampling function). The size of randomly generated samples ranged from 15 to 90 and it was based on the CHNRI exercise on newborn health research priorities [11].

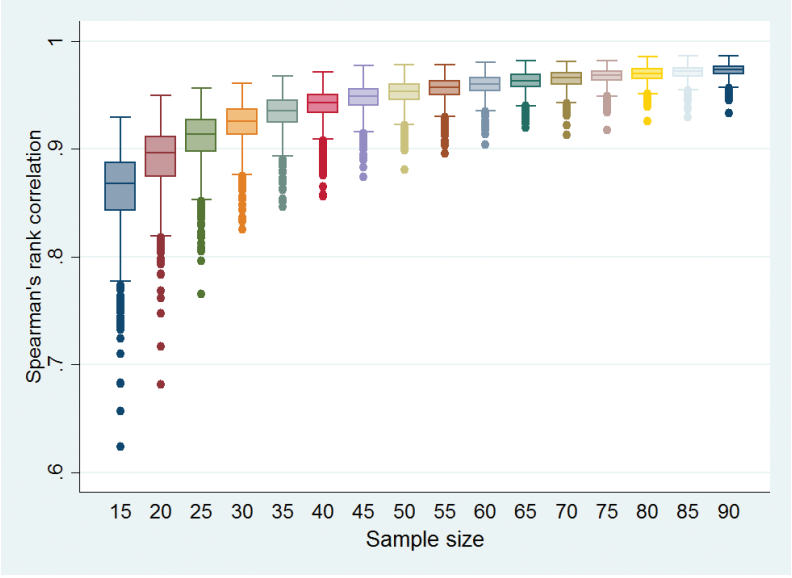


Figure 2. Spearman's rank correlation among all 205 ranked research ideas (Y–axis) by the size of the sample of randomly selected experts (X–axis) from a total pool of 91 experts using a bootstrap method (simulation 1000 times with replacement of already selected experts, using bsampling function). The size of randomly generated samples ranged from 15 to 90 and it was based on the CHNRI exercise on newborn health research priorities [11].

Table 1. The differences in median scores (with inter–quartile range, IQR) that different sub–groups of scorers (ie, researchers vs "non–researchers"; and high–income country (HIC)–based vs low– or middle–income country (LMIC–based) assigned to different criteria*

	All scorers (median, IQR) (N = 91)	Researchers (median, IQR) (N = 61)	"Non–researchers" (N = 30)	HIC–based (median, IQR) (N = 53)	LMIC–based (median, IQR) (N = 38)
Total score	63 (54–71)	62(54–70)	64 (53–73)	57 (47–66)	72 (61–80)
Answerability	76 (68–83)	76 (68–84)	77 (67–85)	74 (63–81)	81 (73–89)
Effectiveness	70 (61–77)	69 (61–78)	68 (59–78)	66 (54–74)	76 (66–84)
Deliverability	69 (58–77)	69 (59–78)	67 (57–78)	65 (54–72)	77 (65–84)
Maximum impact	42 (32–52)	39 (32–50)	44 (32–55)	32 (23–41)	54 (44–66)
Equity	57 (47–70)	57 (46–66)	60 (46–75)	48 (37–61)	72 (60–81)

IQR – interquartile range, HIC – high–income, LMIC – low– and middle–income

*The median scores were determined across all 205 research ideas in order to investigate if any sub–group of scorers deviated in their scoring of any particular criterion.

deliverability, while the largest differences were noted for maximum potential impact and equity.

Table 2 shows the overlap between the top 20 research ideas (RI–) identified by different sub–groups of scorers (ie, researchers vs "non–researchers"; and HIC–based vs LMIC–based). There was an overlap between researchers and "non–researchers" for 10 out of top 20 research ideas (50%). For HIC–based vs LMIC–based researchers, 8 of top 20 research ideas (40%) overlapped. We could judge this level of overlap against the expectation provided by the bootstrap analysis for comparable sample sizes. There is likely to be an effect of sub–stratification, which is smaller for the "researchers vs. non–researchers" comparison, but more considerable for the "HIC–based vs. LMIC–based" comparison.

Table 2. The overlap between the top 20 research ideas (RI–) identified by different sub–groups of scorers (ie, researchers vs "non–researchers"; and HIC–based vs LMIC–based)*

Rank	All scorers (N = 91)	Researchers (N = 60)	"Non–researchers" (N = 31)	HIC–based (N = 53)	LMIC–based (N = 38)
1	RI–30	**RI–30**	**RI–30**	**RI–30**	**RI–30**
2	RI–28	**RI–28**	**RI–28**	**RI–28**	**RI–23**
3	RI–15	**RI–15**	**RI–15**	RI–29	**RI–15**
4	RI–23	RI–29	RI–5	**RI–15**	RI–47
5	RI–33	**RI–23**	**RI–33**	**RI–33**	**RI–28**
6	RI–29	RI–36	RI–79	RI–7	RI–44
7	RI–149	RI–7	**RI–23**	RI–13	RI–18
8	RI–37	RI–13	RI–52	**RI–23**	RI–12
9	RI–5	**RI–33**	**RI–149**	RI–149	**RI–33**
10	RI–13	RI–58	RI–46	RI–36	RI–86
11	RI–79	**RI–149**	RI–47	RI–5	RI–58
12	RI–78	**RI–37**	RI–44	RI–37	**RI–46**
13	RI–36	RI–67	**RI–8**	RI–21	RI–60
14	RI–46	RI–75	**RI–78**	RI–55	RI–11
15	RI–8	**RI–78**	RI–129	**RI–79**	RI–8
16	RI–55	RI–86	RI–11	RI–22	RI–35
17	RI–52	**RI–55**	**RI–37**	RI–52	RI–67
18	RI–75	RI–12	**RI–55**	**RI–78**	RI–10
19	RI–58	**RI–8**	RI–127	RI–75	**RI–79**
20	RI–67	RI–158	RI–138	**RI–46**	**RI–78**

HIC – high–income, LMIC – low– and middle–income

*The research ideas that overlap between researchers vs "non–researchers", and HIC–based vs LMIC–based sub–samples, respectively, are in bold for easier recognition. Note: eg, RI–30 indicates research idea number 30 in the list of 205 ideas.

DISCUSSION

In this paper, we addressed two important questions relating to the quantitative properties of human collective opinion: (i) whether there is a point of "saturation" in the sample size, after which no significant changes in the collective opinion should be expected when more experts are brought into the exercise; and (ii) whether there is evidence that opinions differ between subgroups of experts defined by their professional background or their geographic location. We addressed both questions using data from a previous CHNRI exercise [11]. The data set based on the CHNRI exercise was useful in this regard, because it quantified a large number of expert opinions about 205 competing research ideas in a systematic and structured way, based on five pre–defined criteria, using simple categorical responses. We did not attempt to demonstrate that the collective would give more "useful" predictions than individual experts would, since this is examined in another paper on collective knowledge [8]. Perhaps the best support for the view that the opinion of a collective will prove more useful over time than that of individuals is provided in the literature on stock markets and prediction markets [3,9,10]. Over long periods of time, following the collective wisdom seems to be the most successful strategy. There are some important differences, though, because stock markets to a degree involve betting individual opinions against those of others, where investors are trying to identify stocks and shares that are undervalued by the collective opinion. Together, our previous paper from this series [8] and the large experience with stock markets and prediction markets [3,9,10] make a compelling case for collective decision–making.

Our analyses indicate that, in bootstrap samples that ranged in size from only 15 to 90, the level of overlap

among the top 20 scored research ideas increased with sample size up to about 50–55 experts. At this point, the median level of concordance stabilized at 15/20 top ranked questions (75%), with the interquartile range also generally stable (14–16). There was little further increase in overlap when the bootstrap sample of experts increased from 55 to 90. However, it should be noted that the overlap of 12/20 top ranked research ideas was achieved with sample sizes as small as 15 experts, as opposed to only 2 research ideas that would have been expected by chance. The conclusion from this analysis is that human collective opinion, when expressed in simple quantitative terms, tends to converge towards a similar outcome and saturate quickly. A sample size of 15 persons already shows an appreciable level of reproducibility, but with 50–55 experts the level of replicability becomes nearly equal to to that which is achievable with a sample size of 90.

It is important to note that the total sample of 91 experts, which is the maximum that we had available, represents only a sub–sample of a much larger global pool of experts. Therefore, it also carries a certain inherent random variation relative to the "total expert population". Sampling with replacement enables us to examine how variable the results for a given sample size will be, assuming that are full sample of 91 experts is representative of the diversity of the wider global pool. Thus two bootstrapped samples of size 91 participants would not be expected to have the top 20 research ideas fully replicated (although this is the entire original sample!). We used sampling with replacement to overcome, at least partly, the concern that the 91 experts are still only a reasonably small sample of the larger population and to produce a conservative estimate of the minimum sample size that produces replicable results in this particular CHNRI exercise.

We also tested the relationship between the sample size of the scorers and Spearman's rank correlation coefficient for the ranks of all 205 proposed research ideas. As expected, the rank correlation coefficient increased as the bootstrap sub–samples became larger. A median correlation of 0.95 was reached at the sample size of 45 experts (median of the rank correlation coefficient = 0.95; IQR 0.94–0.96), which again points to high reproducibility and relatively quick saturation.

Studying quantitative properties of human collective opinion, as opposed to collective knowledge verifiable against accepted facts, has the limitation that no gold standard is available against which the "accuracy" of the opinion can be judged. We therefore focused on the questions of saturation, reproducibility and subgroup stratification. Another limitation of this preliminary analysis is that it was based on a single data set from a previous CHNRI exercise. An analysis of multiple data sets with large numbers of experts and different numbers of research ideas being scored may offer further interesting insights into a nature of human collective opinion and results that are more generalizable than those based on the analysis of a single data set. Ideally, an analysis should involve as many experts as possible, because testing on exercises that only included reasonably small groups of experts will not be very useful. At this point, we should also declare that we can't predict the effects of low response rate and self–selection bias on the level of saturation achieved. The issue of missing responses of the experts who do not choose to participate should be explored separately and it remains an unresolved uncertainty related to the validity of the approach used in the CHNRI method.

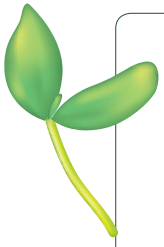
Any future work in this area could plan to acquire more data sets and replicate the analyses from this study. One emerging question that it would be interesting to answer is to examine the main determinants of the observed level of concordance in ranking lists. Examples of possible determinants are the composition and the nature of the proposed research ideas, the composition and sample size of scorers, and the criteria used for discrimination. Answering this question would require a study into how an increasing number of experts participating in the CHNRI exercise introduces variation in the data set across different exercises; then, how does the number of research questions in the data set introduce variation; how does the substance (ie, content, plausibility) of research ideas introduce variation; and how does the level of agreement between all experts participating in the CHNRI exercise introduce further variation. It would be important to understand whether the key determinant of variation in the data set is the number of experts, the diversity of experts, the number of research ideas, or the content and diversity of research ideas. This could be understood if the number of research ideas and the number of experts are standardized (ie, made equal) across several different CHNRI exercises and then the rank correlation analysis and a comparison of the concordance of the top 20 research priorities are repeated using the methodology in this paper.

An important question is whether by increasing the sample size of scorers we would obtain a wider spectrum of opinions, and therefore greater variation between responses, or whether we would simply continue to observe the same level of variation. One way of addressing this would be to look at a CHNRI exercise where we could separate those who responded to the initial request and those who only responded after reminders, and study whether there was evidence that the late responders differed from the early responders in their opinions.

A search for the presence of sub–stratification in this study could only examine the two characteristics that were known

for each invited scorer: a background in research vs "non–researchers", and affiliation to HIC vs LMIC. When the analysis of concordance was conducted, a reduced level of agreement was detectable when HIC–based vs LMIC–based samples were compared. This observation lends support to the recommendation that an inclusive approach to the sample selection in the CHNRI method should be preferred, so that the result of the exercise reflects the opinion of a wide group of experts. This should help to prevent any particular sub–group among the scorers, with particular views, having undue influence on the results. An analysis of a much larger set of data set from the CHNRI exercises might help to suggest how best to manage the problem of sub–stratification within the sample of invited experts and whether there were examples of exercises in which this concern was reduced to a minimum, or even avoided [17].

Finally, it is of interest to the field of qualitative research to draw analogies between the observations on "saturation" of quantitatively expressed human collective opinion, which we observed in this study, and the long–term notion of quick saturation of information content obtained through interviews with human subjects. Researchers studying the question of the "saturation of ideas" in qualitative research often conclude that 15 interviews may be all it takes to reach a very high degree of "saturation", with 20–30 interviews being sufficient [18]. The numbers as small as those proposed are often counter–intuitive to researchers who conduct quantitative research in the fields such as epidemiology, public health and/or clinical trials, where new information is still discovered even after hundreds or thousands of participants have been enrolled, and having larger sample sizes often leads to a better study with more statistical power to demonstrate convincing results. We conclude that the results of our study seem to support the notion that human collective opinion tends to saturate surprisingly quickly and there does seem to be a point at which adding further experts is unlikely to significantly affect the results that were derived from the initial 45–55 experts. This interesting finding warrants further exploration to understand why this seems to be the case and whether there is a wider significance of this finding, or perhaps any immediate opportunities to implement it in solving practical problems in different areas of human activity.



Acknowledgements: The authors thank the participants of the CHNRI exercise on newborn health for providing scores that were used as a basis for analyses in this study. This work was supported by the CHNRI Foundation award to the University of Edinburgh.

Funding: None.

Ethical approval: Not required.

Authorship declaration: SY led the work on database development and co–wrote the paper. IR participated in study design, co-wrote the paper and provided important intellectual content. SC led the design of the analyses, wrote the script for analyses, mentored SY in running the analyses and provided important intellectual content.

Declaration of interest: IR is an editor–in–chief of the Journal of Global Health. To ensure that any possible conflict of interest relevant to the journal has been addressed, this article was reviewed according to best practice guidelines of international editorial organizations. The authors completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author). Authors declare no conflicting financial or other interest related to the work detailed in this manuscript.

REFERENCES

1 Surowiecki J. The wisdom of crowds. London: Abacus, 2013.
2 Surowiecki J. Democracy: dreams of the common good. In: Surowiecki J. The wisdom of crowds. London: Abacus, 2013.
3 Graham B. The intelligent investor. New York: Harper Collins, 2003.
4 Tripadvisor. Available at: https://www.tripadvisor.co.uk/. Accessed: 15 April 2016.
5 Internet Movie Database. Available at: http://www.imdb.com/. Accessed: 15 April 2016.
6 Rudan I, Chopra M, Kapiriri L, Gibson J, Lansang MA, Carneiro I, et al. Setting priorities in global child health research investments: universal challenges and conceptual framework. Croat Med J. 2008;49:307-17. Medline:18581609 doi:10.3325/cmj.2008.3.307
7 Rudan I, Gibson JL, Ameratunga S, El Arifeen S, Bhutta ZA, Black M, et al. Setting priorities in global child health research investments: guidelines for implementation of the CHNRI Method. Croat Med J. 2008;49:720-33. Medline:19090596 doi:10.3325/cmj.2008.49.720
8 Rudan I, Yoshida S, Wazny K, Chan KY, Cousens S. Setting health research priorities using the CHNRI method: V. Quantitative properties of human collective knowledge. J Glob Health. 2016;6:010502 doi:10.7189/jogh.06.010502.

REFERENCES

9 Prediction market. Available: https://en.wikipedia.org/wiki/Prediction_market. Accessed: 15 April 2016.

10 Pennock DM, Lawrence S, Giles CL, Nielsen FA. The real power of artificial markets. Science. 2001;291:987-8. Medline:11232583 doi:10.1126/science.291.5506.987

11 Yoshida S, Martines J, Lawn JE, Wall S, Souza JP, Rudan I, et al. Setting research priorities to improve global newborn health and prevent stillbirths by 2025. J Glob Health. 2016;6:010508. Medline:26401272 doi:10.7189/jogh.06.010508

12 Glaser BG, Strauss AL. The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine Publishing Co, 1967.

13 Bertaux D. From the life-history approach to the transformation of sociological practice. In: Bertaux D, ed.: Biography and society: The life history approach in the social sciences. London: Sage, 1981.

14 Bernard HR. Social research methods. Thousand Oaks: Sage, 2000.

15 Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. BMC Med Res Methodol. 2005;5:37. Medline:16321161 doi:10.1186/1471-2288-5-37

16 World Bank country data. Available: http://data.worldbank.org/country. Accessed: 15 April 2016.

17 Rudan I. Global health research priorities: mobilizing the developing world. Public Health. 2012;126:237-40. Medline:22325672 doi:10.1016/j.puhe.2011.12.001

18 Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods. 2006;18:59-82. doi:10.1177/1525822X05279903

Supplementary analysis of three CHNRI exercises

To deepen our understanding of the point at which stabilisation of expert opinion is achieved, similar statistical analyses to those in Paper 3.2 were performed on three additional datasets. The primary purpose of repeating the analysis on different datasets was to obtain more robust findings in which there is more confidence that the results are likely to be generalizable to other CHNRI exercises, and to discuss the results considering the findings from the previously conducted analysis.

Three datasets were used in this analysis (**Table 4**). Each dataset includes the scores given by anonymous individuals who participated in the CHNRI exercises in health-related domains: a) maternal and perinatal health,[25] b) quality of care for every woman and every newborn,[8] c) disability and health access.[9] In the RP exercise on maternal and perinatal health, 650 experts were approached, 140 of whom provided scoring for 190 research ideas. Most research questions were specific to interventions relating to specific diseases or conditions. The RP on quality of care for every woman and every newborn, on the other hand, conducted the literature review and pre-selected 30 research questions prior to the exercise, and only contacted the participants for scoring (Paper 2.3). This exercise approached over 1000 experts, achieving a 27% response rate. In this exercise, the identified RPs included was more like knowledge domains than specific research questions. The RP exercise on disability and health access did not report the number of experts approached. Fifty experts provided scoring for 83 research questions focusing on how to best address the barriers to accessing different levels of services provided to people with disabilities.

Table 4. Summary table of previous CHNRI exercises used in the supplementary data analysis

Title	Scope	Year	Time frame for assessment of ideas	Areas of ideas sought	Experts approached (N)	Experts generated ideas (N)	Research ideas generated(N)	Experts participated in the scoring exercise (N)	Questions ranked (N)	Criteria used (N)
RP exercise on maternal and perinatal health	Maternal morbidity, health system antenatal and perinatal care	2013-2014	2015-2025	Delivery, discovery and development ideas around labour and delivery, obstetric haemorrhage, hypertensive disorder or pregnancy, abortion, antenatal care, health system, neonatal care, and other	650	339	980	130	190	5
RP exercise on disability and health access	Improving health and functioning of people with disability	2008	Not specified	Delivery, discovery, descriptive and development research on the health of people with disability (improve availability and accessibility of health services, opportunity for education)	Not mentioned	70	348	50	83	5
RPs to improve the quality of care for every woman, every child	RP exercise on midwifery-led maternal and newborn healthcare	2014-2015	2016-2030	Delivery, discovery and development research on maternal and newborn healthcare and the contribution of midwifery care.	1000	30	30	270	30	5

While the RP exercise on maternal and perinatal health exercise used the standard CHNRI criteria, the other two used slightly modified criteria. Kennedy et al chose community involvement and sustainability instead of effectiveness and deliverability, and Tomlinson introduced patient-centred criteria such as applicability, sensitivity and support within the context. They all used five criteria. The usual scale of 0, 0.5, 1 or blank was used for scoring for all the exercises.

Two statistical analyses were performed on each of the three datasets. The first analysis examined the concordance of the ranking of the top-ranked questions. The second statistical analysis examined the degree of correlation in the rank assigned to all research questions. Both analyses compared the results from the bootstrap samples with those of the original scorers. For each sample size, 1000 bootstrap samples were drawn with replacement for each simulation. We used arbitrary cut-off points of 75% concordance and a rank correlation coefficient of 0.95 to define stable results.

Result

Concordance of ranks on top ranked questions

The median concordance and interquartile range (IQR) are presented in the Box Whisker plots (**Figure 3, Figure 4, Figure 5**). In this analysis, we considered that stability is achieved when a median concordance of 75% is achieved (i.e. agreement on 15 out of the top 20 research options).

Figure 3 shows the median concordance, IQR and range for the number of overlapping questions found within the top 20 research priorities identified by the priority setting exercise on maternal and perinatal health. The sample sizes examined ranged from 15 to 130. The result indicates that the agreement on 15 of the research options is only observed at a sample size of 130 (median 15, IQR 14-15). However, it could also be argued that having a sample of 65 experts, which is almost half of the reference group, already, achieves a median overlap of 13.

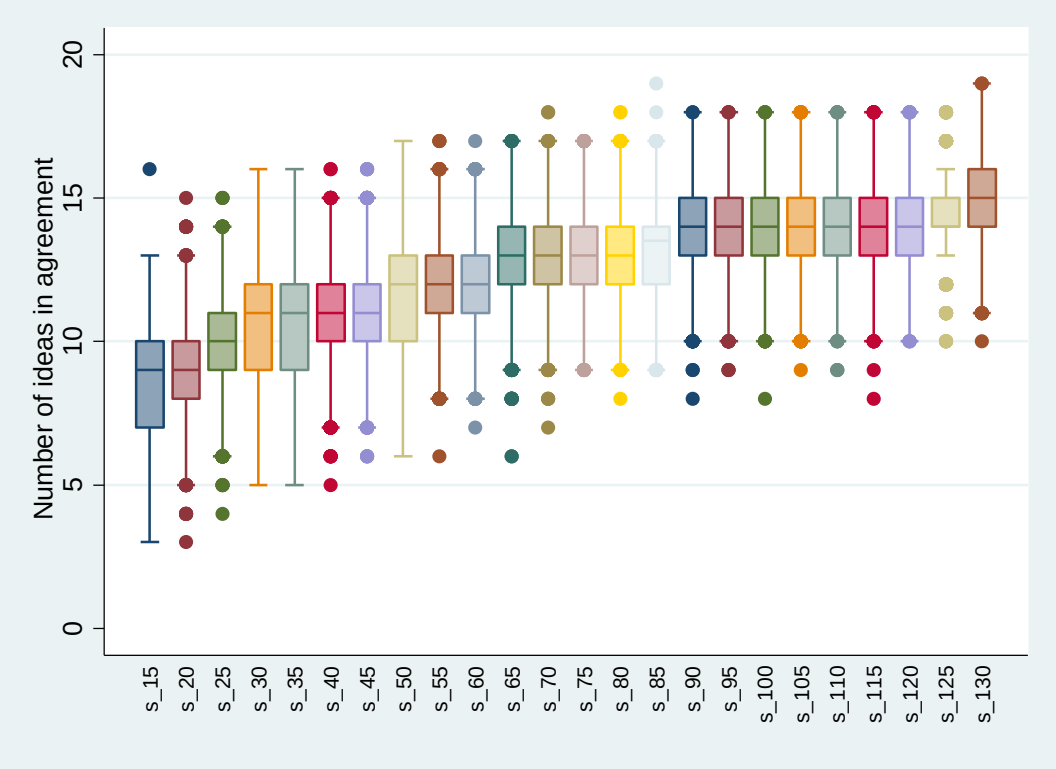


Figure 3. Level of overlap among the top 20 ranked research ideas (Y–axis) by the size of the sample of randomly selected experts (X–axis) from a total pool of 130 experts using a bootstrap method (simulation 1000 times with replacement of already selected experts, using bsampling function)

Figure 4. shows the analysis performed on the priority setting exercise on disability with varying number of subsamples, ranging from 15 to 50. A median concordance of 16 (IQR 15-17) is obtained at the subsample size of 30 and does not increase further with increasing sample size. The full sample size of 50 demonstrated the same level of agreement as the 30 experts.

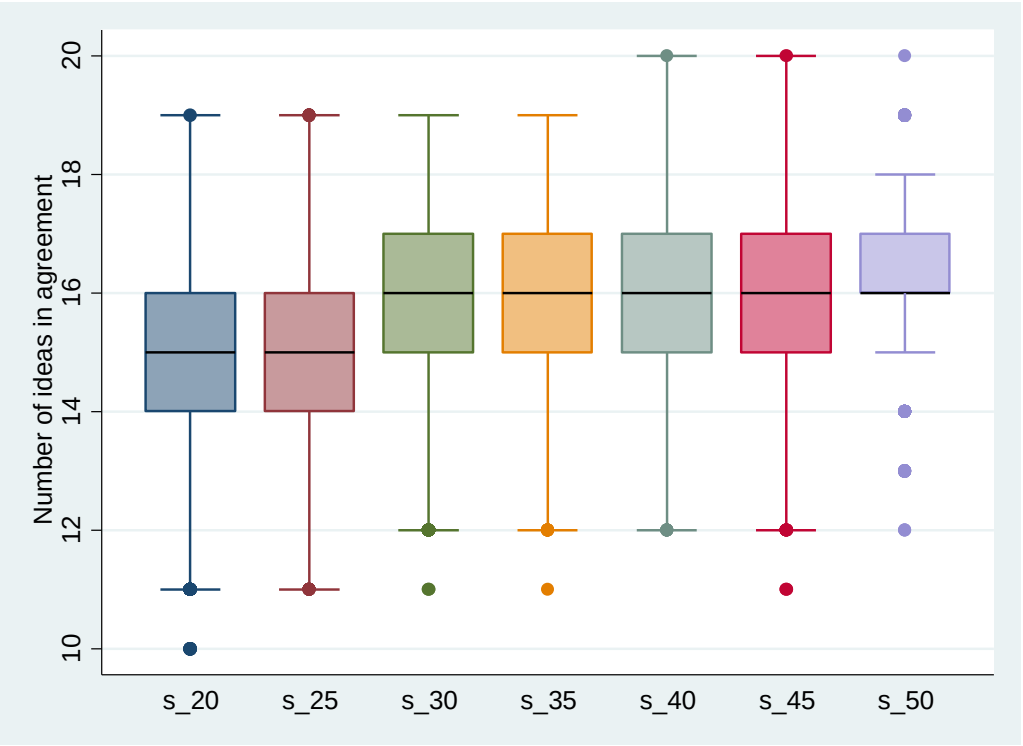


Figure 4. Level of overlap among the top 20 ranked research ideas (Y–axis) by the size of the sample of randomly selected experts (X–axis) from a total pool of 83 experts using a bootstrap method (simulation 1000 times with replacement of already selected experts, using bsampling function)

Figure 5 shows the same analysis conducted on the dataset obtained from the CHNRI exercise on the contributions of midwifery care for maternal and newborn health. This dataset analysis was performed with varying number of subsamples, ranging from 5 to 270. A median concordance of 15 (IQR 14–16) is obtained at the subsample size of 5 and does increase further with increasing sample size. The full sample size of 270 demonstrated the same level of median agreement as the 210 experts.

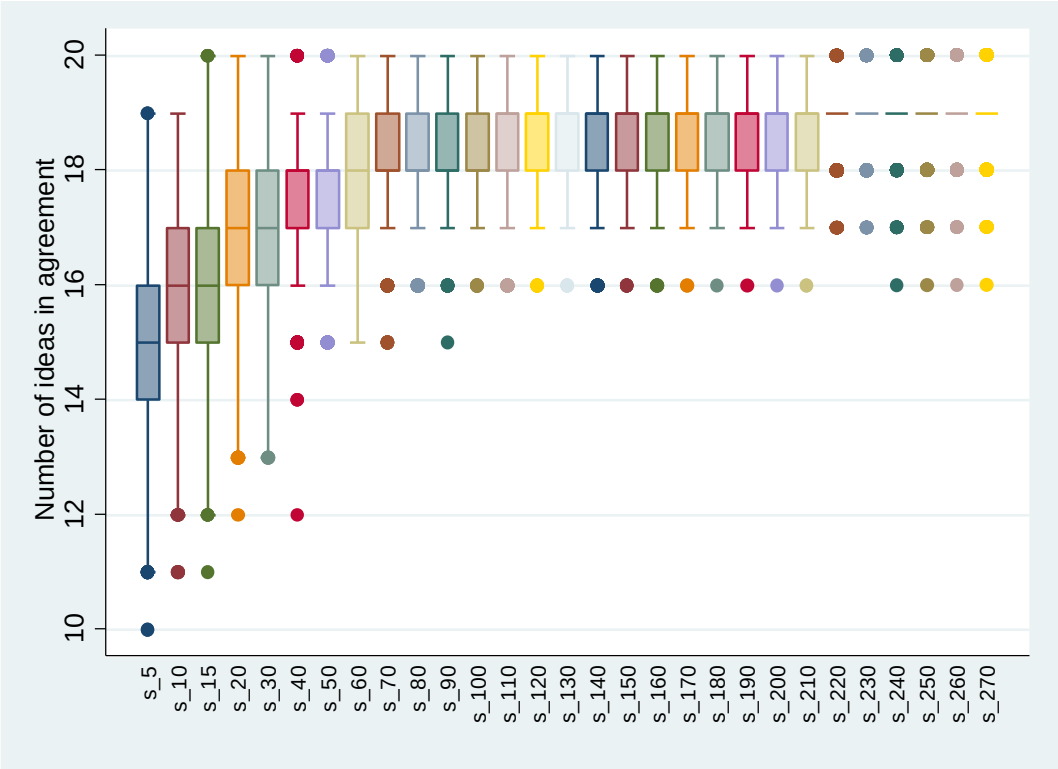


Figure 5. Level of overlap among the top 20 ranked research ideas (Y–axis) by the size of the sample of randomly selected experts (X–axis) from a total pool of 270 experts using a bootstrap method (simulation 1000 times with replacement of already selected experts, using bsampling function)

Rank correlation of all the research questions

In the second analysis, the rank correlation coefficient gradually increases as the bootstrapped sample size increases in all three datasets. In this analysis, we considered a correlation coefficient of 95% to be the optimal achievable correlation between the ranks obtained from bootstrapped subsamples and the original sample of experts. The median correlation and IQR are presented in the Box Whisker plots (**Figure 6, Figure 7, Figure 8**).

Figure 6 shows the association between the sample size of the bootstrap sub-samples (X-axis) and the Spearman's rank correlation (Y-axis) in maternal and perinatal health CHNRI research prioritisation exercise. A median correlation coefficient of 0.95 (95%) is achieved with the sample size of 110 experts (median 0.95, IQR 0.94-0.95).

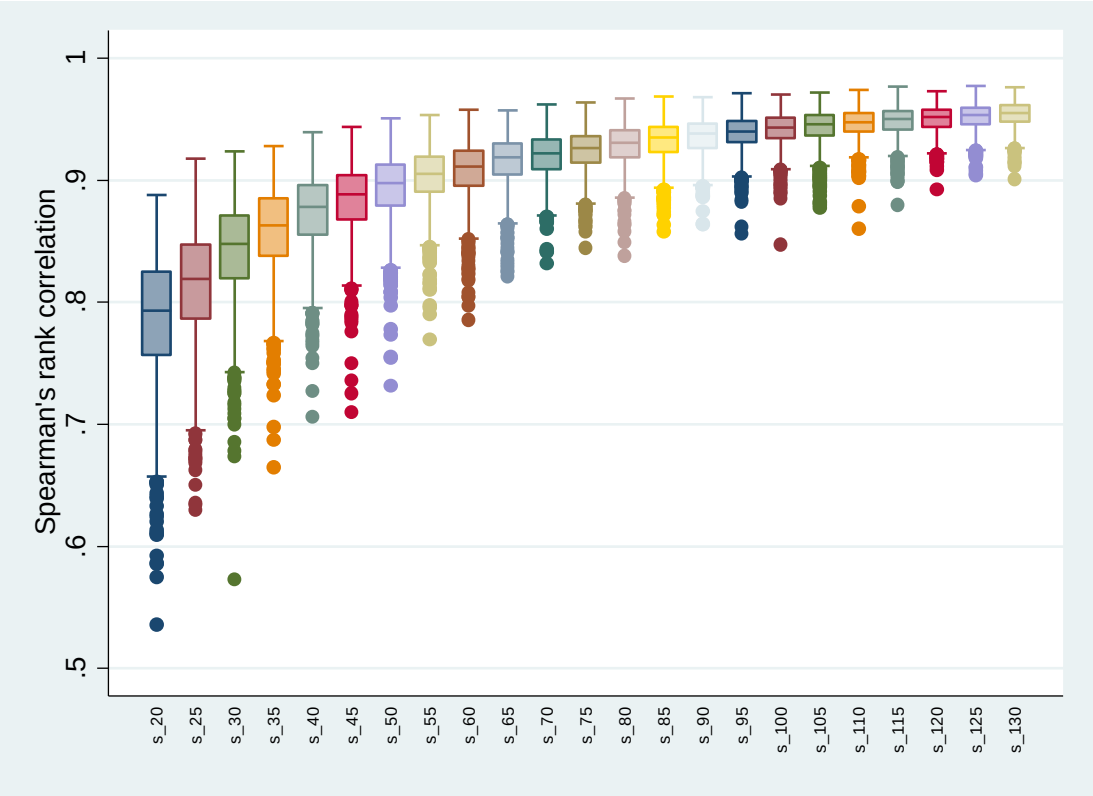


Figure 6. Spearman's rank correlation for all 190 research questions (Y-axis) by the size of the sample of randomly selected experts (X-axis) from a total pool of 130 experts using a bootstrap method (simulation 1000 times with replacement of already-selected experts, using bsampling function)

Figure 7 shows the association between the sample size (X-axis) and Spearman's rank correlation (Y-axis) in the research priority setting exercise on disability. The correlation coefficient increases with the increasing sub-sample size however a median correlation coefficient of 0.95 is not quite reached with a bootstrap sample size equal to the total sample size for the experts (50: median correlation coefficient 0.94, IQR 0.93-0.95)

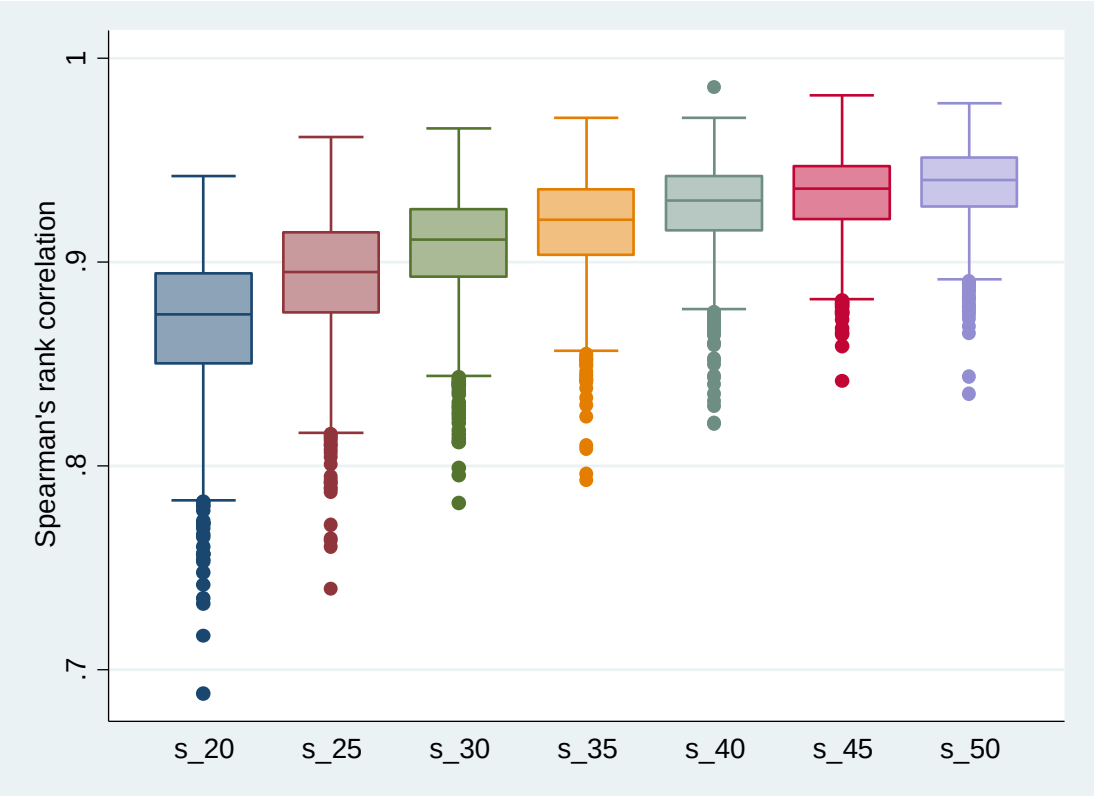


Figure 7. Spearman's rank correlation for all 83 research questions (Y-axis) by the size of the sample of randomly selected experts (X-axis) from a total pool of 50 experts using a bootstrap method (simulation 1000 times with replacement of already selected experts, using bsampling function).

Figure 8 shows the association between the sample size of bootstrap subsamples (X-axis) and the Spearman's rank correlation (Y-axis) in the research prioritisation exercise on the midwifery model of care. The rank correlation appears stable with a bootstrap sample size of 150 with correlation coefficient of 0.95 (IQR 0.93, 0.97).

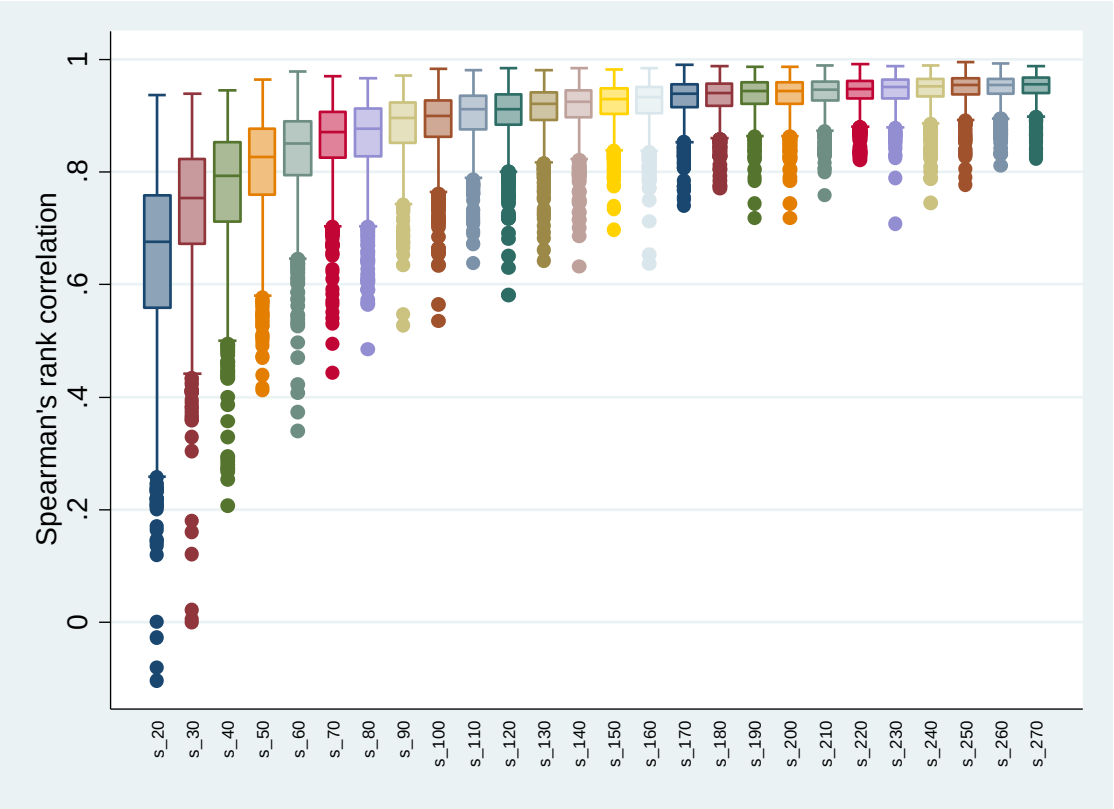


Figure 8 Spearman's rank correlation for all 30 research questions (Y-axis) by the size of the sample of randomly selected experts (X-axis) from a total pool of 270 experts using a bootstrap method (simulation 1000 times with replacement of already selected experts, using bsampling function)

Further analyses were conducted to examine if the number of experts needed to achieve the stability of results correlates with varying the number of research questions. In other words, if an RP exercise with a greater number of research questions would require more experts to achieve stable results. We repeated the statistical analysis using bootstrap samples this time varying the number of research questions from 40 randomly selected samples to 50, 60, and finally 70 randomly selected samples. The analysis included three RP exercises including data from the RP on newborn health and birth outcome (Paper 3.2) and did not include RP on quality of care because the number of research questions was much smaller than in the other RP exercises. Table 5 and Table 6 show the sample size at which three levels of concordance and rank correlations are achieved at varying number of research questions for each RP exercise, respectively. Table 7 presents mean and standard deviation (SD), median (IQR), and coefficient of variation of the RPS for each exercise.

Table 5. Summary of sample size at which three level of concordance is achieved in each exercise by varying the number of questions

	Sample size at which 85% concordance is achieved			
RP Exercise	*40 RQs*	*50 RQs*	*60 RQs*	*70 RQs*
Newborn health and birth outcome	20	20	25	40
Maternal and perinatal health	35	35	45	65
Disability and health access	<20	<20	<20	>50
RP Exercise	**Sample size at which 80% concordance is achieved**			
Newborn health and birth outcome	15	15	15	15
Maternal and perinatal health	15	15	15	45
Disability and health access	<20	<20	<20	>50
RP Exercise	**Sample size at which 75% concordance is achieved**			
Newborn health and birth outcome	<15	<15	<15	<15
Maternal and perinatal health	<15	<15	<15	25
Disability and health access	<20	<20	<20	45

Table 6. Summary of sample size at which three levels of rank correlation are achieved in each exercise by varying the number of questions

	Sample size at which 95% rank correlation is achieved			
RP Exercise	*40 RQs*	*50 RQs*	*60 RQs*	*70 RQs*
Newborn health and birth outcome	60	43	40	40
Maternal and perinatal health	105	100	100	>130
Disability and health access	50	50	>50	50
	Sample size at which 90% rank correlation is achieved			
RP Exercise	*40 RQs*	*50 RQs*	*60 RQs*	*70 RQs*
Newborn health and birth outcome	18	18	18	20
Maternal and perinatal health	50	45	45	70
Disability and health access	35	35	35	25
	Sample size at which 85% rank correlation is achieved			
RP Exercise	*40 RQs*	*50 RQs*	*60 RQs*	*70 RQs*
Newborn health and birth outcome	15	<15	<15	<15
Maternal and perinatal health	35	28	28	40
Disability and health access	20	20	20	<20

Table 7. Variability of research priority score in each RP exercise

RP Exercise	Research Priority Score
Newborn health and birth outcome	
Mean (SD)	62 (11.8)
Median (IQR)	63.0 (53.9-71.4)
Coefficient of variation	19.1
Maternal and perinatal health	
Mean (SD)	55 (25.5)
Median (IQR)	59.0 (42.3-76.0)
Coefficient of variation	46.3
Disability and health access	
Mean (SD)	67 (11.7)
Median (IQR)	68.6 (61.1- 75.2)
Coefficient of variation	17.5
Quality of care for every mother and every child	
Mean(SD)	85 (4.1)
Median (IQR)	86.0 (83.4-88.8)
Coefficient of variation	4.9

Discussion

In this Chapter, I replicated previously conducted analyses on different CHNRI datasets, to make a further study of an aspect of sample size in human collective opinion. It is important to note that with the bootstrap method it is not possible to achieve 100% concordance because of replacement that occurs during sampling for each simulation. In four analyses, including the original study, we set the threshold to define stability of results at 75% concordance for the top 20 research questions, and at a rank correlation coefficient of 0.95 for all research questions. The number of experts, the number of scored research questions, and the sample sizes, at which these thresholds are achieved, are presented in the summary table (**Table 8**). The experiment showed that the concordance among the top 20 research questions reached 75% to 95% at maximum, and these maximum concordances were obtained at 40% to 60% of the original sample size in three out of the four experiments. The only outlier was the maternal and perinatal health exercise for which maximum concordance was reached at 95% of the original sample size.

The sample size at which 75% agreement is observed for the top 20 research questions ranged widely. It ranged from five experts in the RP exercise on quality of care for every woman and every child to 125 experts in maternal and perinatal health exercise. The wide range is possibly affected by number of total research options scored in these RP exercise. Sample sizes at which 95% rank correlation was achieved varied from 45 experts in the RP exercise on newborn health to 150 experts in the RP exercise on quality of care for every woman and every child.

Table 8. Summary of results

Title of research prioritisation exercise	Number of experts	Number of research options	Sample size at which 75% agreement is observed in top 20 research questions	Sample size at which maximum agreement is observed (% of max agreement)	Sample size at which 95% rank correlation is observed in all research questions
Newborn health and birth outcomes	91	205	55	55 (75%)	45
Maternal and perinatal health	130	190	125	130 (75%)	110
Disability and health access	50	83	20	30 (80%)	50
Quality of care for every woman, every child	270	30	5	110 (95%)	150

In contrast to exercises reaching the stability point in the sample size of experts, we could not locate a point of stability for one of the priority setting exercises (maternal and perinatal health). This may be due to the broad scope of this exercise covering health system, preconception, abortion, maternal and perinatal care as well as newborn care. This RP exercise reported higher variations in the scores represented by wide SD and IQR than other exercises [Mean, SD: 55 (26); Median (IQR) 59.0, (42.3-76.0)]. The RP exercise on quality of care for every mother and every child had very similar scores among the top priorities. This was seen by the low variability of scores with narrower SD and IQR in the exercise [Mean (SD): 85 (4.1) Median (IQR) 86.0 (83.4-88.8)]. However, for this exercise, concordance for the top 20 research priorities is achieved more quickly due to the smaller number of research options as compared with other RP exercises

We conducted further analysis to examine if there is any correlation between the number of research questions and the number of experts required to achieve stability of results for three exercises. Overall, not surprisingly, concordance on the top 20 questions is more easily achieved when the total number of questions is fewer. For example, to achieve 75% concordance of top 20 research priorities would require 55 experts when there are 205 research questions while it would require less than 15 experts for 40 to 70 research questions for the RP exercise on newborn health. Similarly, 120 experts were required to achieve 75% concordance for 190 research questions while it would only require less than 25 experts for 40 to 70 research questions in the RP exercise on maternal and perinatal health. The RP exercise on disability and health access showed that 75% concordance is achieved with less sample size of experts if number of questions were smaller than 70. Similar pattern was observed both at 80% and 85% concordance. A sudden increase in the number of experts at 70 research questions points to the need to investigate whether there is any increase in the number of experts when there are more than 70 research questions.

In the analysis that looked at stability by rank correlation, the finding did not show any correlation between increasing the number of questions and stability of results. This was not surprising that the rank correlation to be less susceptible to the number of questions than concordance of top 20 research questions. The number of experts, at which 95% rank correlation was observed for each number of question ranging from 40 to 80, were similar to those required for achieving stability for all research questions.
Although we defined an *a priori* threshold at 95% and used this cut-off point to compare the stability across exercises, 95% is a very high correlation. We therefore looked for stability point at 90% and 85% correlation while varying the number of research questions. It showed that to obtain 90% correlation, the newborn health RP exercise would need 18 to 20, to have stable results, 45 to 70 experts for maternal and perinatal health RP exercise, and 25 to 35 for the RP exercise on disability and health access. Therefore, we could argue that these three RP exercises would have only needed a minimum of 18 to a maximum of 70 experts to achieve stability of results. Similarly, 85% rank correlation is achieved up to15 experts in the newborn health RP exercise, while it is slightly more in the maternal and perinatal health RP exercise, ranging from 28 to 40, and the RP exercises on disability and health access required up to 20 experts for the same range of the number of questions. Thus, for 85% rank correlation, these three exercises would have needed approximately 15 to 40 experts.

There are two suggestions for further analyses. Firstly, it would be interesting to perform a further analysis to examine if there is a stability point at which one can be reasonably confident that adding further experts would not change the result. In other word, does the stability point differ if varying the number of experts? This analysis could be performed by comparing the point at which stability is achieved by experts in the exact order of the scores received i.e., first 20 experts who provided scores, first 30 experts etc. I plan to conduct this analysis in the future.

Secondly, it would be interesting to examine how many experts are needed to achieve the "saturation" of ideas. In our experience, we observed that initially there were as many new ideas as experts submitting the ideas. However, we noted that after a time the ideas being submitted were similar to those already received. This is not unexpected i.e., we might expect there to be a saturation point beyond which further experts do not contribute different ideas. An analysis could be conducted by taking 5 experts at a time, for example, to look at the cumulative number of distinct ideas submitted as the number of experts increases.

In this PhD thesis, we did not investigate thematic-saturation, rather we defined saturation as the stability point at which by adding more experts would not change the level of concordance or ranking of the research questions. Analysis on thematic-saturation may be able to provide users of the method with useful guidance on optimal sample size of experts and how many mutually exclusive ideas one could expect from the sample size of expert.

Saturation can only be ascertained after the data collection has been made. Hammersley et al consider this as general problem of qualitative research. So is the stabilization of result. In our analysis, about 40% to 60% of the original sample size was sufficient to obtain reasonably stable results except in one exercise. One of the factors likely to influence how quickly stabilization occurs is the homogeneity of the sample, which may not be representative of the population of experts since stabilization of response occurs relatively faster among such groups.[23] Although we sought to approach "diverse" groups of experts ranging from programme managers, to researchers to policy makers in previously conducted CHNRI exercises, the fact remains that they were purposive samples. They were selected based on their interests and expertise in a given health area; therefore, it could be argued that our expert samples are homogeneous with respect to their professional interests.

Presenting research questions in the same order in the scoring sheet to experts could have influenced rank stabilization. Two exercises, the RP exercises on quality of care and maternal and perinatal health, randomised the order of research questions, in order to rule out the potential bias in scoring the questions if the order of research questions were uniform.[8,25] The remaining two used a uniform order. If there is an "order effect" it could have influenced top ranked priorities but not overall.

Strengths and limitations

This exercise was a unique opportunity to explore some properties of human collective opinion with datasets obtained from four global CHNRI exercises.

As one of the study limitations, it is worth mentioning that some variation in the CHNRI method might have contributed to variations in scoring. While, the usual scale of 0, 0.5, 1 or blank was used for scoring for all the exercises, different criteria were used in all exercises. While Souza et al used the standard CHNRI criteria; the other two had used slightly modified criteria. Kennedy et al chose community involvement and sustainability instead of effectiveness and deliverability, and Tomlinson introduced patient-centred criteria such as applicability, sensitivity, and support within the context. However, the same number of criteria was used across all the RP exercises.

Finally, the purpose of repeating the analysis on different datasets was to obtain results that are likely to be generalizable to other CHNRI exercises. Our analysis provided a broad indication of the range in the number of experts at which stability of results was observed at the three levels of rank correlations. It is important to note that there is not enough evidence to provide a clear guidance on the number of experts required for future CHNRI exercises. To provide useful guidance, there is a need to repeat the same analysis on more data on CHNRI exercises, to increase the generalizability of the findings.

Involving participants (Paper 4.1, Paper 4.2, Paper 4.3)

The experiments to assess accuracy of the underlying assumption of the CHNRI method suggested that collective knowledge outperforms individual knowledge. The same exercise confirmed that experts' collective knowledge in an area of their expertise is better than their collective knowledge in an area outside of their expertise (Paper 3.1).

Statistical analyses were also undertaken to examine whether it was possible to identify a minimum sample size of experts required for scoring the results of these analyses varied across different CHNRI exercises preventing any recommendation being made at this point. Further analyses using data from more CHNRI exercises might help to identify whether there are context-specific characteristics of different exercises which have implications for sample size (Paper 3.2).

In this chapter, I will shift the focus from quantitative to qualitative aspect of the process. Three papers (paper 4.1, paper 4.2 and paper 4.3) introduce results of qualitative review on how participants are approached, involved and in what way their inputs were reflected in the process. Participants in a CHNRI exercise include researchers, stakeholders, and funders. The three papers provide useful guidance and examples of best practices in involving these participants in future CHNRI exercises.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? The Journal of Global Health

1.2. When was the work published? June 2016

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes** / **No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
I conducted literature review, analysed and synthesized available information and wrote first draft of introduction, method and result section of the paper.

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE [redacted] Date 20.09.2016

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

Setting health research priorities using the CHNRI method: III. Involving stakeholders

Sachiyo Yoshida[1*], Kerri Wazny[2*], Simon Cousens[3], Kit Yee Chan[2,4]

[1] Department for Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland
[2] Centre for Global Health Research, The Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, Scotland, UK
[3] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[4] Nossal Institute for Global Health, University of Melbourne, Melbourne, Australia
[*] These authors contributed equally to the work.

Setting health research priorities is a complex and value–driven process. The introduction of the Child Health and Nutrition Research Initiative (CHNRI) method has made the process of setting research priorities more transparent and inclusive, but much of the process remains in the hands of funders and researchers, as described in the previous two papers in this series [1,2]. However, the value systems of numerous other important stakeholders, particularly those on the receiving end of health research products, are very rarely addressed in any process of priority setting. Inclusion of a larger and more diverse group of stakeholders in the process would result in a better reflection of the system of values of the broader community, resulting in recommendations that are more legitimate and acceptable.

The CHNRI method, as originally proposed, took into account the importance of stakeholders and made provisions for their participation in the process. Although the involvement of a large and diverse group of stakeholders is desirable, they were not expected to propose research ideas, or score them against the set of pre–defined criteria. Because of this, the original CHNRI method proposed that stakeholders should be allowed to "weigh" pre–defined criteria and set "thresholds" for a minimum acceptable score against each criterion that would be required for a research idea to be considered a "research priority". In choosing the stakeholders, the context of each exercise will be very important and the goals of the specific exercise should be defined before choosing an appropriate "stakeholder group". Among stakeholders, we would expect to see those affected by the disease of interest and their family members, their carers and health workers, members of general public, media representatives interested in the topic, community leaders, representatives of the consumer groups and industry, but also potentially researchers and funders themselves. Although the latter two groups – researchers and funders – already have a different role assigned in the CHNRI process, this does not exclude them from also being stakeholders in the process [1,2]. In this paper, we aim to review and analyse the experiences in stakeholder involvement across the 50 CHNRI exercises published in the 10–year period between 2007 and 2016, the proposed approaches to involving stakeholders and their effects on the outcome of the prioritization process.

One paper in the original CHNRI method series focused on involving stakeholders [3]. That paper presented practical experiences from three separate attempts to involve stakeholders that took place in 2006. The three groups approached were: (i) members of the global research priority setting network; (ii) a diverse group of national–level stakeholders from South Africa; and (iii) participants at a con-

Setting health research priorities is a complex and value–driven process. The introduction of the CHNRI method has made the process of setting research priorities more transparent and inclusive, but much of the process still remains in the hands of funders and researchers. However, the value systems of numerous other important stakeholders, particularly those on the receiving end of health research products, are very rarely addressed in any process of priority setting. Including a larger group of stakeholders in the process would result in a better reflection of the system of values of the broader community, resulting in recommendations that are more legitimate and acceptable.

ference related to international child health held in Washington, DC, USA. Each group was asked to complete a short questionnaire to assess the relative importance of the five original CHNRI criteria. Different versions of the questionnaire were used with each group [3]. The results of this exercise indicated that groups of stakeholders vary in the weights they assigned to the 5 criteria, reflecting divergence in the "value" placed on each criterion by each stakeholder group.

The diverse group of respondents within the priority–setting network placed the greatest weight on the criterion of "maximum potential for disease burden reduction" and the most stringent threshold on "answerability in an ethical way". Among the attendees at the international conference on child health, the criterion of "deliverability, answerability and sustainability" was identified as the most important. Finally, in South Africa, where inequity has been a national problem historically, the greatest weight was placed on the "predicted impact on equity" criterion.

This comparative analysis by Kapiriri et al. [3] effectively demonstrated that involving a wide range of stakeholders is an important goal for any research priority setting exercise. The criteria that may be of importance to funders, scientists and other technical experts involved in the process of planning and conducting the exercise may not be well aligned with the values of those who should eventually benefit from health research, or with the sentiments of wider society as a whole [3]. This is an important observation, because if the CHNRI process is conducted without regard for the broader social value or research then it is unrealistic to expect it to fulfil its purpose of being accepted as a fair, transparent and legitimate process for setting investment priorities for health research.

THE CONCEPTS OF THRESHOLDS AND WEIGHTS IN THE CHNRI METHOD

These concepts were introduced as a part of the initial CHNRI method description [4,5]. The multi–disciplinary working group that developed the CHNRI method recognised the need to find a practical way to involve a much larger group of stakeholders in the priority–setting process. An agreement was reached that, at least in principle, most members of the public would not be expected to generate research ideas or score them, because they do not possess the knowledge that would enable them to discriminate among the proposed research ideas. Instead, it was agreed that their contribution to the process and the final results of the exercise would be in the assignment of "weights" to the criteria that reflect their collective preferences and beliefs. Over the years of CHNRI implementation, it has been shown that stakeholders originating from funding institutions or political organizations prefer the criterion of maximum potential for disease burden reduction, because their targets are usually set around this criterion; programme managers are typically more focused on the deliverability and sustainability criterion; stakeholders from the industry tend to prefer knowing the likelihood of effectiveness of resulting interventions; while members of the general public often emphasize equity and ethics as their preferred criteria [6].

In addition to placing more "weight" on some criteria than others, which could affect the final rankings of *all* research ideas as a result of stakeholders' input into the CHNRI process, the stakeholders can also disqualify *some* research ideas using the system of "thresholds". This means they may agree a priori that a research idea will not be considered a priority unless it reaches a certain minimum score against a particular priority–setting criterion. This can be important in a specific context; eg, in the aforementioned example of South Africa, where equity was a very important concern for all stakeholders, they could have insisted that a research idea must have a minimum score of 80% on the "equity" criterion to qualify as a priority. In practice, this means that a research idea with scores 50–70% on all other criteria, but 90% on "equity", could be considered a research priority. However, another idea with scores of 80–90% on all other criteria, but 60% on "equity" would be disqualified from the exercise – or at least delayed, until it addresses the recognized issues with equity. Common examples of the latter are the new, high technology–based interventions that would likely first be utilised by the wealthy. In this way, research ideas with lower overall

The original CHNRI method proposed that large and diverse groups of stakeholders should "weigh" different criteria according to their perceived value and importance for society as a whole. They were asked to set "thresholds" for minimum acceptable scores for each of the pre–defined criteria. In this paper, we aim to review and analyse the experiences with stakeholder involvement across the 50 CHNRI exercises published in the 10–year period between 2007 and 2016, the proposed approaches to involving stakeholders and their effects on the outcome of the prioritization process.

scores could be seen as greater priorities if they pass all the pre–defined "thresholds" [3,4].

Although the interdisciplinary group that developed the CHNRI method considered this approach as practical and inclusive, the question remained of how best to select the stakeholders and ensure their representativeness to the entire community of interest. Possibly the best solution to this problem to date has been achieved by Kapiriri et al. [3] who aimed to develop a "global" group of stakeholders by conducting an internet–based survey of the affiliates to the "Global research priority setting network", which had been assembled in the years prior to the development of the CHNRI method by the staff from the University of Toronto, Canada. Between March and May 2006 a large number of affiliates to the "Global research priority setting network" agreed to participate in a pilot on the condition of anonymity. They agreed to provide stakeholder input to five forthcoming exercises that aimed to set research priorities to address the five major causes of global child mortality. Respondents included a very diverse mix of researchers, policymakers and health practitioners with an interest in priority setting in health care from high–, middle– and low–income countries. Participants were given a simple version of the questionnaire, and were asked to rank the five "standard" CHNRI criteria from 1st to 5th in the order of their perceived importance of the criteria. They were also asked to set a threshold for each of the five criteria. The respondents placed the greatest weight (1.75) on potential for disease burden reduction, while the weights for the remaining four criteria were similar to each other, and ranged between 0.86 to 0.96. The highest threshold was placed on the criterion of answerability in an ethical way (0.54), while the lowest was placed on potential for disease burden reduction (0.39).

CASE STUDIES OF STAKEHOLDER INVOLVEMENT IN CHNRI EXERCISES

We identified 50 research prioritization exercises using the CHNRI method that were published between 2007 and 2016. Of the 50 exercises, 38 (76%) did not seek inputs from stakeholders and 12 (24%) involved stakeholders as their larger reference group. This already shows how it may be remarkably difficult in most cases to identify and involve an appropriate group of stakeholders that would be representative of the wider community of interest – whether this is a global, regional, national or local population. It seems that, in the absence of simple solutions, most authors who conducted the CHNRI exercises preferred not to include stakeholders in the process, rather than including an ill–defined and non–representative group and then having to adjust the final ranks based on their input. By not including input from stakeholders, the CHNRI exercises simply remained "unfinished" to an extent, though weights and thresholds could still be applied *post–hoc* should an appropriate group of stakeholders be identified at some later stage – unless the context changes substantially in the meanwhile.

Among the 12 CHNRI exercises that involved stakeholders and took their input into account, 5 were papers that belonged to the series of exercises related to addressing research priorities for the five major causes of child mortality globally – eg, pneumonia, diarrhoea, neonatal infections, preterm birth/low birth weight, and birth asphyxia [7–11]. All of these papers were co–ordinated by the World Health Organization (WHO) and they used the weights and thresholds defined above by Kapiriri et al. [3]. However, the remaining seven exercises made their own individual attempts, using guidelines for implementation of the CHNRI method, to identify appropriate stakeholders within their own contexts and involve them in the process. This section explores the experiences and results from these seven studies. **Table 1** summarizes the approaches to involving stakeholders in these seven exercises.

Two exercises were carried out at the global level. They were focused on mental health research and acute malnutrition in infants less than six months, respectively [12,13]. The remaining five exercises were conducted at the national level and focused on research in child health in South Africa [14], zoonotic disease in India [15], health policy and maternal and child health in China [16,17], and Prevention of Mother–to–Child Transmission of HIV (PMTCT) in Malawi, Nigeria and Zimbabwe [18]. Given that the large majority (over 80%) of the 50 CHNRI exercises were focused on either the global context, or on all low– and middle–income countries (LMIC), the high representation of national–level exercises among those CHNRI studies that

Table 1. Summary tables on the involvement of stakeholders

Reference	Profiles and mode of identification	Number of stakeholders	Responsibility	Criteria	Weights and thresholds applied to the criteria	Impact of stakeholders' involvement on the final scores
[12]	Psychiatrists (9), psychologists (4), social workers (2), government employees (3), non–governmental organization representatives (6), researchers (6), users of mental health services (6) and members of the public service (7), including those from low–and middle–income countries; No indication as to how they were identified and selected	43	They were asked to rank the five pre–defined criteria with range of 1 to 5 (1–highest rank to 5–lowest rank)	5 standard CHNRI criteria used [4]	Weights were assigned based on ranking: effectiveness (+21%), maximum potential for burden reduction (+17%), deliverability (+0%), equity (–9%), answerability (–19%); Thresholds not applied	There was no description whether the ranks significantly differed between non–weighted and weighted scores
[13]	Mostly researchers and policy makers; also included technical experts, senior practitioners in the area of nutrition and child health (including 9 members of "MAMI" groups: Management of Acute Malnutrition for Infant less than six month reference group). Above profiles included all the participants and there was no clear description of the profile of stakeholders. Identified from the participants at meetings, symposia related to the technical area of concern	64	They were asked to score the research questions against the pre–defined criteria, rather than place weights on the criteria	5 standard CHNRI criteria (two composite criteria split into two – 7 in total) [4]	Weights and thresholds not applied	See main text: the stakeholder group was used for scoring, rather than weighting
[14]	Researchers, academics, clinicians, government officials, clinical psychologists, and member of the public. Identified based on their availability and accessibility with an attempt to ensure diversity of the group	30	Same as reference [12]	5 standard CHNRI criteria used [4]	Weights were defined using the rank given to the 5 pre–defined criteria: equity (+30%), efficacy and effectiveness (+9%), deliverability, affordability and sustainability (+2%), maximum potential for disease burden reduction (–9%), answerability and ethics (–19%); Thresholds not applied	The paper presented both the weighted and non–weighted scores. The stakeholders' inputs changed the ranking of the research options somewhat, but the top 20 research options remained the same in both cases
[15]	Scientists, students and lay people. Identified from staff members of the Public Health Foundation of India (PHFI) and those identified through personal networks of authors	Not mentioned	They are asked to rank the pre–defined five criteria from most important (ranked 1) to least important (ranked 5) within the national context	5 standard CHNRI criteria used [4]	Weights were defined using the rank given to five pre–defined criteria: deliverability, affordability (+18%), maximum potential for disease burden reduction (+18%), efficacy and effectiveness (+13%), equity (–17%) and answerability and ethics (–18%); thresholds not applied	The final outcome was not affected by the stakeholders' inputs on the criteria in that the top 15 research options remained the same across weighted and non–weighted scores
[16]	Managers from medical institutions, doctors, patients, and representatives of public (5 representatives of each group). Method of identification not mentioned	20	They were asked to rank the and also provide the thresholds on the pre–defined five criteria. However it was unclear whether or not other participants also provided the ranking to the criteria	5 criteria used: potential to affect change, maximum potential for disease burden reduction, deliverability, economic feasibility and equity	Weights: Potential to affect change (0.1925), maximum potential for disease burden reduction (0.1925), deliverability (0.2160), economic feasibility (0.1890) and equity (0.2050); Thresholds: Potential to affect change (33.5%), maximum potential for disease burden reduction (29.7%), deliverability (27.0%), economic feasibility (28.0%) and equity (27.8%).	It was unclear whether any major differences in the ranks were observed after applying the weights and thresholds

Table 1. Continued

Reference	Profiles and mode of identification	Number of stakeholders	Responsibility	Criteria	Weights and thresholds applied to the criteria	Impact of stakeholders' involvement on the final scores
[17]	Obstetricians, gynaecologists, paediatricians, representatives of patients group, industry and international organizations; mode of identification was not mentioned	19	They were asked to rank the and also provide the thresholds on the pre–defined ten criteria	10 criteria used: answerability and ethics, efficacy and effectiveness, deliverability, maximum potential for disease burden reduction, equity, acceptability, sustainability, translation to policy, and economic feasibility and equity	Weights: answerability (0.11), efficacy and effectiveness (0.09), deliverability (0.10), maximum potential for disease burden reduction (0.14), equity (0.11) acceptability (0.07), sustainability (0.11), translation to policy (0.10), economic feasibility (0.10) and equity (0.07). Thresholds: answerability (33%), efficacy and effectiveness (38%), deliverability (28%), maximum potential for disease burden reduction (29%), equity (29%), acceptability (41%), sustainability (33%), translation to policy (33%), economic feasibility (40%) and equity (38%)	It was unclear whether any major differences in the ranks were observed after applying the weights and thresholds
[18]	The article addressed three country–led research prioritization exercises. In each country, stakeholders were researchers, academics, policy makers, district health workers, frontline health workers, implementing partners, people living with HIV/AIDS; mode of identification was not mentioned	40 to 70 participants each in Malawi, Nigeria and Zimbabwe	Stakeholders participated in the entire process ie, generation of research ideas and the scoring of research ideas. The weighting of scores was not applied in the exercise, because all stakeholders participated in the entire process.	6 criteria were used: answerability and ethics; potential maximum disease burden reduction on paediatric HIV infections; addresses main barriers to scaling–up; innovation and originality; equity; and likely value to policy makers	Weights and thresholds not applied	This exercise included diverse group of stakeholders. In this regard the relevance of the research ideas identified in the respective exercise to the national context was high.

used stakeholders input (5/12) is likely a reflection of the fact that it is much easier to involve stakeholders at the national or sub–national level than it is on a regional or global level.

In all exercises, the stakeholders involved were first given an induction course about the CHNRI process. Then, an opportunity for asking and sharing questions and concerns with respect to the CHNRI process was provided. In five of the seven exercises, stakeholders were asked to rank the relative importance of the pre–defined criteria from most important one ("1") to the least important ("5"), while considering the context of the research prioritization. The average score was calculated for each criterion and was then used to calculate the relative weights by dividing the average expected score of 3.0 (ie, the average expected rank if all criteria were valued the same) by the mean assigned rank. For example, a mean assigned rank for "answerability" criterion of 2.47 translates a relative weight of 1.21 (ie, 3.00/2.47 = 1.21). In this way, "answerability" will receive 21% greater weight than if all the criteria were weighted equally.

The concept of thresholds was very rarely used. Even when it was applied, it was clear that it wasn't properly explained to participating stakeholders. This is not surprising, because the thresholds really refer to a measure of "collective optimism" of the scorers, rather than a real computation of likelihood or probability that is rooted in any real–world parameters. It is very difficult to estimate what this measure of "collective optimism" could amount to for different criteria. This is why such attempts to set thresholds typically resulted in them being set at 25%–30%, much too low to have any discriminatory power and disqualify many research ideas, so that almost all research ideas passed all the thresholds.

In the remaining two exercises, the nature of stakeholder involvement was modified radically from that which was originally envisaged in the CHNRI exercises with reasonable justification [13,18]. Instead of using the group of stakeholders only to adjust the ranks that were derived from an expert–driven scoring process, the authors involved a broad range of stakeholders in the generation of research ideas [18] and/or scoring the research ideas [13,18]. We will now reflect on these experiences in a critical way, identify some lessons learnt and make recommendations for future exercises.

CRITICAL ASSESSMENT OF STAKEHOLDER INVOLVEMENT IN CHNRI EXERCISES

In the 7 studies that tried to develop a larger reference group of stakeholders that would be appropriate to their respective contexts, the number of stakeholders involved was disappointingly small: it ranged from 20 to 70. Although attempts were clearly made to ensure diversity of the stakeholders involved, such small sample sizes can hardly be considered sufficiently inclusive of many different groups of stakeholders and their representativeness. Although good representativeness of stakeholders can be ensured without necessarily requiring a very large number of participants – such as, eg, in many examples of national parliaments in democratic societies, who represent all the people of the nation through a relatively small number of their elected members – we still feel that bigger numbers would ensure more legitimacy to the process, or more relevance of the outcomes to the context of the exercise.

It would be difficult to consider the examples in the reviewed exercises as truly representative of the wider communities, let alone the nation or the world. This shows that despite the authors' best intentions to fully adhere to the guidelines and complete the CHNRI process, they didn't really manage to find a satisfactory solution to involving large and diverse group of stakeholders. In these papers, the profile of stakeholders often included researchers, who would have been better reserved for the scoring process. Other stakeholders included clinicians, government officials, and representatives of academia and professional organizations, which again are rare in the society and hardly representative of the wider community. The examples of the profiles of persons who we would expect included in the larger reference group are also laypersons, frontline health workers and direct beneficiaries of health services, such as patients who contracted disease of concern. We encourage the authors of the future CHNRI exercises to try to get as much feedback as possible from those groups, because they have their own specialised knowledge (including lived experience), which would not be captured by other participating groups in the process. They also have "stake", or interest, in the outcome of the exercise.

The small sample sizes and differences in approaches to ensure diversity and representativeness of the stakeholders led to large variations in stakeholders' input [12–18]. In the global exercise, the greatest relative importance was assigned to effectiveness, and the lowest to answerability, though these results should not be generalized. Stakeholders at the national level varied in their preferences, alternately supporting the criteria equity, deliverability (with affordability and sustainability), or the maximum potential for disease burden reduction (**Table 1**). Clearly, small sample sizes used in these exercises limit the generalizability of such preferences even within their local context, let alone more broadly.

It is also important to note that in all exercises that applied the "weights", this procedure didn't really have dramatic effects on the final rankings of the research ideas. Although a research idea might move a few places up or down the list following the weighting procedure, these shifts did not profoundly affect the non–weighted ranking order that was determined by the researchers and experts. Perhaps this is one of the additional reasons why so many groups conduct-

Photo: Meeting with a group of stakeholders at the maternity health clinic in Ghana (Courtesy of Dr Alice Graham, personal collection)

ing the CHNRI exercise did not place sufficient importance on involving stakeholders. From the exercises that involved stakeholders, one might conclude that the process of expert scoring is sufficient and the outcome of the exercise will not be greatly altered by the involvement of stakeholders. We believe that such a view is premature and would like to see more examples of the involvement of the stakeholders in the CHNRI process before such judgements could be made.

In two exercises that actively involved stakeholders, their involvement wasn't limited to weights or thresholds, but rather they were also involved in research idea generation and scoring [13,18]. In the exercise on PMTCT in three African countries [18], about 40–70 people took part in respective countries, and all participants contributed to all stages of the CHNRI process. This included academics/researchers, district health workers and implementing partners such as UN agencies, people living with HIV/AIDS, frontline health workers and policy makers. The authors' justification for including these diverse groups in all stages of the CHNRI process was to avoid discriminating within this diverse range of groups, but to truly engage the groups according to their technical expertise and to enhance inclusiveness and participation in similar priority–setting exercises across the nation. Eventually, the stakeholders' weighting of the scores was not even applied, possibly due to an assumption that it was no longer needed. This example represented a rather interesting deviation from the original CHNRI conceptual framework, but we can see a rationale for this modification, which makes it an illuminating exception.

The other exercise, on the management of acute malnutrition in infants in low– and middle–income countries, involved stakeholders only in the scoring process [13]. The stakeholder group included participants at meetings and symposia related to the topic area (**Table 1**). In this exercise, the core group of researchers ("management team") developed the list of research questions based on the review of the literature in this field that preceded the CHNRI exercise as the preparatory step. The final list of questions was then circulated for scoring to both researchers invited to the CHNRI process and also the conference participants, who were considered stakeholders. Equal weighting was given to all criteria. The management team justified this on the grounds that malnutrition was a new area of research in infants younger than 6 months and they therefore believed that unweighted estimates would be more suitable and interpretable by their intended policy–maker audience. However, the authors stated that the lack of weighting of criteria might have resulted in limited reflection of the values in the broader community. In this case, we can conclude that the borderline between the invited researchers and the "stakeholders" (who were likely to include unrelated researchers and any other people of similar profile who could be expected to attend an international conference in this topic), was blurred and not really clear. It is likely that this deviation from the suggested approach didn't really invalidate the conceptual framework, because all the scorers would still be expected to possess knowledge on the topic of interest. It would perhaps be more appropriate not to call the second group "stakeholders", but rather an additional, "convenience" sample of scorers that increased the number of scorers considerably.

PROPOSED SOLUTIONS AND WAY FORWARD

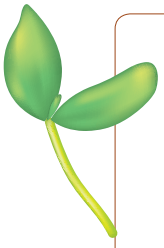
So far, there hasn't really been a good example of stakeholder involvement as originally envisaged by the CHNRI across the first 50 implementations, apart from perhaps the Kapiriri's priority–setting network involvement that was used in 5 child mortality papers [3,7–11]. This is certainly a shortcoming of all the previously conducted processes. This finding may also reinforce the initial concern that involving stakeholders in research priority setting processes is very challenging and that the solutions proposed in the original CHNRI method were quite difficult to implement as envisaged.

This is not to suggest that the results of previous CHNRI exercises are not useful, and the thresholds and weights can be applied later, if a good solution to obtain them can be found within the time scale during which the context described to scorers would still remain largely unchanged. The efforts conducted to date to perform the CHNRI exercises were not wasted and their results can be used. However, it must be acknowledged that most CHNRI exercises to date are, in fact, incomplete at least with respect to the original vision for them. To bridge this gap better definition is needed of who are the stakeholders at different levels (ie, global, regional, national and local) and how best to represent them.

For global exercises, we'll inevitably need a very large and inclusive crowd–sourcing exercise of many stakeholder representatives, who would place weights and thresholds on all 25 priority–setting criteria that were used to date across all 50 CHNRI exercises (5 "standard" and 20 new). The sample of stakeholders will need to be truly large, because we may later need several sub–samples that could provide us with region–specific stakeholders, or allow selecting specific groups of stakeholders and leaving others out of the exercise. In this way, the large "global" sample of stakeholders would also serve as a base for the regional samples of stakeholders. A major concern relating to this suggested approach would be how to avoid a strong urban bias in low–income settings and be inclusive of un-

developed and/or rural areas. In terms of national–level or local–level exercises, it is likely that highly targeted samples that aimed to include 500–1000 stakeholders would already be sufficient and representative of national or local context. The "targeting" component of the sampling strategy would define the profile of the stakeholders that would be most appropriate to the exercise, and then a person could be found in the community to fit each such profile.

How could these large sample sizes be achieved technically? How could we engage thousands of people globally, or hundreds nationally? With further attention to the development of the area of "crowd–sourcing" in the age of the internet and social networks (such as Facebook, Twitter, etc.), we should be able to do lot more in the future with respect to truly engaging the stakeholders in the process of setting priorities in health research investments at different levels of the human population.



Acknowledgements: We sincerely thank the representatives of the stakeholders for taking part in the previous CHNRI exercises and providing useful information. SY is the employee of the World Health Organization. Her views expressed here do not necessarily represent the views of the Organization.

Funding: None.

Authorship declaration: SY and KW wrote the article. KYC and SC revised the article and made useful comments.

Competing interests: All authors have completed the Unified Competing Interest form at www. icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no conflict of interest.

REFERENCES

1 Rudan I, Yoshida S, Chan KY, Cousens S, Sridhar D, Bahl R, et al. Setting health research priorities using the CHNRI method: I. Involving funders. J Glob Health. 2016;6:010301. doi:10.7189/jogh.06.010301

2 Yoshida S, Cousens S, Wazny K, Chan KY. Setting health research priorities using the CHNRI method: II. Involving researchers. J Glob Health. 2016;6:010302. doi:10.7189/jogh.06.010302

3 Kapiriri L, Tomlinson M, Gibson J, Chopra M, El Arifeen S, Black RE, et al. Child Health and Nutrition Research Initiative (CHNRI): Setting priorities in global child health research investments: Addressing the values of the stakeholders. Croat Med J. 2007;48:618-27. Medline:17948948

4 Rudan I, Gibson JL, Ameratunga S, El Arifeen S, Bhutta ZA, Black M, et al. Child Health and Nutrition Research Initiative (CHNRI). Setting Priorities in Global Child Health Research Investments: Guidelines for implementation of the CHNRI Method. Croat Med J. 2008;49:720-33. Medline:19090596 doi:10.3325/cmj.2008.49.720

5 Rudan I, Gibson J, Kapiriri L, Lansang MA, Hyder AA, Lawn J, et al. Child Health and Nutrition Research Initiative (CHNRI). Setting priorities in global child health research investments: Assessment of principles and practice. Croat Med J. 2007;48:595-604. Medline:17948946

6 Rudan I. The complex challenge of setting priorities in health research investments. Indian J Med Res. 2009;129:351-3. Medline:19535827

7 Fontaine O, Kosek M, Bhatnagar S, Boschi–Pinto C, Chan KY, Duggan C, et al. Setting research priorities to reduce global mortality from childhood diarrhoea by 2015. PLoS Med. 2009;6:e1000041. Medline:19278292 doi:10.1371/journal.pmed.1000041

8 Rudan I, El Arifeen S, Bhutta ZA, Black RE, Brooks A, Chan KY, et al. Setting research priorities to reduce global mortality from childhood pneumonia by 2015. PLoS Med. 2011;8:e1001099. Medline:21980266 doi:10.1371/journal.pmed.1001099

9 Lawn JE, Bahl R, Bergstrom S, Bhutta ZA, Darmstadt GL, Ellis M, et al. Setting research priorities to reduce almost one million deaths from birth asphyxia by 2015. PLoS Med. 2011;8:e1000389. Medline:21305038 doi:10.1371/journal.pmed.1000389

10 Bahl R, Martines J, Ali N, Bhan MK, Carlo W, Chan KY, et al. Research priorities to reduce global mortality from newborn infections by 2015. Pediatr Infect Dis J. 2009;28:S43-8. Medline:19106763 doi:10.1097/INF.0b013e31819588d7

11 Bahl R, Martines J, Bhandari N, Biloglav Z, Edmond K, Iyengar S, et al. Setting research priorities to reduce global mortality from preterm birth and low birth weight by 2015. J Glob Health. 2012;2:010403. Medline:23198132 doi:10.7189/jogh.01.010403

12 Tomlinson M, Rudan I, Saxena S, Swartz L, Tsai AC, Patel V. Setting priorities for global mental health research. Bull World Health Organ. 2009;87:438-46. Medline:19565122 doi:10.2471/BLT.08.054353

REFERENCES

13 Angood C, McGrath M, Mehta S, Mwangome M, Lung'aho M, Roberfroid D, et al. Research Priorities to Improve the Management of Acute Malnutrition in Infants Aged Less Than Six Months (MAMI). PLoS Med. 2015;12:e1001812. Medline:25898252 doi:10.1371/journal.pmed.1001812

14 Tomlinson M, Chopra M, Sanders D, Bradshaw D, Hendricks M, Greenfield D, et al. Setting priorities in child health research investments for South Africa. PLoS Med. 2007;4:e259. Medline:17760497 doi:10.1371/journal.pmed.0040259

15 Sekar N, Shah NK, Abbas SS, Kakkar M; Roadmap to Combat Zoonoses in India (RCZI) Initiative. Research options for controlling zoonotic disease in India, 2010–2015. PLoS ONE. 2011;6:e17120. Medline:21364879 doi:10.1371/journal.pone.0017120

16 Yang M. Zhao K, Qi X, Li X. Determining the priority issues of health policy based on CHNRI. Chinese Journal of Health Policy. 2015;8:74-9.

17 Li X, Zhao Z, Qi X, Xiao Y, Guo W. Determining the priority issues of maternal and child health system in Western China based on CHNRI. Chinese Health Service Management. 2014;4:244-7.

18 Rollins N, Chanza H, Chimbwandira F, Eliya M, Nyasulu I, Thom E, et al. Prioritizing the PMTCT Implementation Research Agenda in 3 African Countries: INtegrating and Scaling up PMTCT Through Implementation REsearch (INSPIRE). J Acquir Immune Defic Syndr. 2014;67 Suppl 2:S108-13. Medline:25310115 doi:10.1097/QAI.0000000000000358

Correspondence to:
Sachiyo Yoshida
Department for Maternal, Newborn, Child and Adolescent Health
World Health Organization
Geneva, Switzerland
yoshidas@who.int

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? The Journal of Global Health

1.2. When was the work published? June 2016

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes** / **No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I conducted review of literature and wrote one of the examples of previously conducted CHNRI exercises.
I provided inputs on introdcution and conclusion of the paper.

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE [redacted] Date 25 July 2018

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

Setting health research priorities using the CHNRI method: II. Involving researchers

Sachiyo Yoshida[1], Simon Cousens[2], Kerri Wazny[3], Kit Yee Chan[3,4]

[1] Department for Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland
[2] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[3] Centre for Global Health Research, the Usher Institute for Population Health Sciences and Informatics, the University of Edinburgh, Edinburgh, Scotland, UK
[4] Nossal Institute for Global Health, University of Melbourne, Melbourne, Australia

Large groups of researchers who agree to offer their research ideas and then score them against pre–defined criteria are at the heart of each CHNRI priority–setting exercise. Although the roles of funders and other stakeholders are also very important, much of the exercise is focused on selecting and engaging a large group of researchers, obtaining their input and analysing it to derive the initial results of the process. In a sense, a CHNRI exercise serves to "visualise" the collective knowledge and opinions of many leading researchers on the status of their own research field. Through a simple "crowdsourcing" process conducted within the relevant research community, the CHNRI approach is able to collate a wide spectrum of research ideas and options, and come to a judgement on their strengths and weaknesses, based on the collective knowledge and opinions of many members of the research community. In doing so, it provides valuable information to funders, stakeholders and researchers themselves, which is obtained at low cost and with little time necessary to conduct the exercise.

Success in involving researchers within each research community, and ensuring their voluntary participation and engagement, is therefore essential to the successful completion of a CHNRI exercise. Over the past few years, we have been involved in assembling groups of researchers to participate in several CHNRI research priority–setting exercises. In this paper, we share our experience of what works well and what works less well and try to answer the most frequently asked questions when it comes to engaging researchers in the CHNRI exercises.

Figure 1 shows where within the CHNRI process researchers should be involved –which is after the funders have provided their input, and before other stakeholders are approached and asked to contribute.

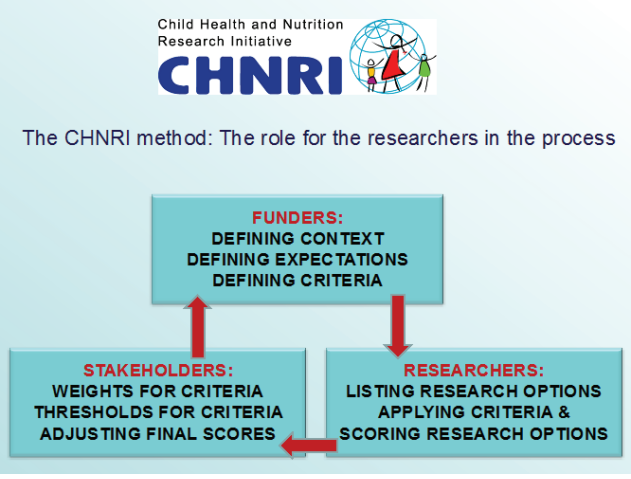


Figure 1. The role of researchers shown within the broader CHNRI process.

WHY DO RESEARCHERS NEED TO BE INVOLVED IN THE CHNRI EXERCISE?

Following input from funders, as described in a previous paper of this series [1], the managers of the CHNRI process then need to involve a sufficiently large sample of researchers. We discuss the considerations relevant to the optimal size of this sample of researchers in another paper of this series [2]. Researchers have two important roles in the CHNRI process: (i) providing the managers with a broad spectrum of research ideas, which usually span the spectrum of "description", "delivery", "development" and "discovery" research; and (ii) providing their own judgement on the likelihood that each submitted research idea will meet a set of pre–defined criteria. These judgements allow the ranking of a large number of submitted research ideas.

At this point, we should explain why CHNRI uses only researchers to provide research ideas, and not other groups of people–eg, funders, programme leaders and managers, other stakeholders, or simply members of the public. This is typically justified on the grounds that researchers are expected to possess far more knowledge and understanding of the state of their research field and the questions that have real potential to generate new knowledge. Importantly, their judgement of each research idea against the priority–setting criteria will also be based on an understanding of the realities of the research process and the success rate in their field. Including participants without this prior knowledge would likely introduce "random noise" into the exercise, resulting in most or all of the ideas receiving similar scores. Thus, restricting participation in these steps to researchers is expected to improve discrimination between the competing research ideas by using the collective knowledge and opinion of a small group of very knowledgeable people.

There is also a practical reason for this: by selecting the most productive, or highly cited researchers over the several preceding years, we are targeting the very group of people who will be most competitive for the research grant calls and likely be awarded the majority of the grants in the immediate future. We should also stress that this is, potentially, a "double edged sword", because researchers may not be entirely objective in their scoring and may tend to score highly their own preferred areas. This is why the chosen group always needs to be large enough, to prevent anyone's individual input having a substantial effect on the overall scores. Therefore, the leading researchers are given power through this method to influence the priorities and shape the topics for the future grants, ie, influence the subjects of the calls that are advertised by the funders, rather than simply responding to them. This could also be helpful to the funders, who do not have an easy access to a collective opinion of their research field.

It is worth bearing in mind that an important characteristic of the CHNRI method is its flexibility. Suggestions provided in the guidelines are not prescriptive, and each exercise can be tailored to meet the specific needs of the exercise. For example, some exercises may be mainly focused on implementation ("delivery") or fundamental ("discovery") research, particularly if the exercise is related to a specific intervention or geographic context. There have been several examples of such exercises, eg, the implementation of zinc interventions [3], implementation research for maternal and newborn health [4], emerging (discovery–based) interventions for childhood pneumonia and diarrhoea [5,6] and others. In such cases, there is scope for involving further groups of people whose knowledge and experience can provide informative input, particularly if this input is limited to the priority–setting criteria where the researchers would be unlikely to possess any first–hand knowledge. For example, many programme managers contributed to the scoring of questions on the newborn research agenda in relation to its deliverability, affordability and sustainability [7]. Our analyses of previous exercises have shown that the researchers tend to be less optimistic than programme managers on the criterion of answerability, while they tend to be more optimistic on the criterion of deliverability, affordability, sustainability and maximum potential for burden of disease reduction; similarly, programme managers tend to prioritise implementation research questions, whereas researchers prioritised technology–driven research [2,8]. Clearly, a good understanding of the complexities and challenges involved tends to make the experts–whoever they are–more cautious about the prospects of the suggested research ideas.

HOW TO INVOLVE RESEARCHERS IN THE CHNRI EXERCISE?

In planning the involvement of the group of researchers, the minimum target sample size needs to be decided early in the process. The optimal number will be derived based on the analyses conducted by Yoshida et al. [2], as mentioned previously. Yoshida's analyses suggest that the ranking of proposed research ideas, relative to each other, stabilises at surprisingly small sample sizes–ie, once that 30–50 people with private knowledge on the topic are involved, it is unlikely that the ranking of proposed research ideas will change markedly with the addition of further researchers and their opinions. Given this finding, targeting sample sizes of 50 or greater should result in a replicable CHNRI priority–setting exercise [2].

However, in planning the number of scorers needed, an important issue needs to be considered, which can reduce not only the actually achieved sample size quite substantially, but also introduce potential bias that can invalidate

the entire exercise. This is the issue of *(self–)selection bias*. The nature of CHNRI process means that researchers are usually invited (using e–mail or other means) by the management team to take part in the exercise. Their participation is needed in two consecutive steps of the process: (i) providing research ideas that they think would stand a good chance against all other ideas, given the pre–defined priority–setting criteria; and (ii) scoring a long list of research ideas against the pre–defined criteria. While the first step, providing research ideas, is not very time–consuming for researchers, the second step is a lot more time consuming and it may require several hours of input.

In an analysis of the first 50 CHNRI exercises, in which more than 5000 scorers were approached, Rudan et al. reported that the initial response rate (ie, submitting research ideas) was about 60%, with each expert submitting an average of about 3 research ideas. However, when all the initially invited experts were approached again to score the "consolidated" list of research ideas, the response rate dropped to only about 35%. Thus 40% of potential scorers are lost at the first stage, and further 25% of the total number are lost at the second stage (Rudan I, personal communication). The reason for re–contacting everyone who was initially invited to participate, even if they didn't offer any research ideas, is that there may be experts who are not keen giving away their ideas, but would be prepared to score ideas generated by others. This may help to preserve the initial sample that was contacted to the maximum extent possible.

Non–response has two important implications for an exercise. First, it reduces the actual sample size. This can be accounted for–eg, if the desired sample is 100 scorers, then about 300 probably need to be invited to participate in the exercise. Second, and more worrying, is the potential for bias in the results if responders and non–responders differ in their opinions. Results based on inputs from only about one third of the initial pool of researchers contacted may suffer from self–selection bias. For example, if individuals are more likely to respond to an invitation from the management group if they know the members of that group well, they may also be more likely to share similar views with the management group members. Others, who may disagree with those views and may, in fact, be in a majority in that particular research community, would not have their opinions recorded, or would be underrepresented. The high proportion of non–responders in many CHNRI exercise is therefore an important issue and we plan to conduct further work to explore non–response in previous exercises by comparing the characteristics of responders vs non–responders. The important thing to realise in relation to this self–selection bias is that it cannot be attenuated or controlled by further increasing sample size with new invitees because, no matter how large the sample size, they may still be based on the opinions of an unrepresentative subset of research community. In summary, increasing the achieved sample size can be done by inviting more people to participate, or by improving the response rate. The former approach will not attenuate possible self–selection bias, while the latter would tend to reduce the scope for bias and should be preferred. Several reminders are, therefore, usually sent to all invited participants to maximise the response rate.

SELECTING AND APPROACHING THE RESEARCHERS

The approach to identifying whom to invite to participate in the exercise can be very flexible, but must be credible to both the reviewers of the resulting publication, and also to any researchers who are left out of the exercise (ie, don't get an invitation). We present three examples of previous CHNRI exercises to examine how different strategies may work in different specific situations.

EXAMPLE OF THE CHNRI EXERCISE ON RESEARCH PRIORITIES FOR CHILDHOOD PNEUMONIA MORTALITY REDUCTION

This exercise [9], published in 2011, involved a small community of researchers working on childhood pneumonia in the low– and middle–income (LMIC) settings. A search for publications on childhood pneumonia in low–resource settings over the previous 5 years listed by the Web of Science identified only a few hundred publications in total. Ranking the authors of these publications ranked by the number of those papers that they had co–authored, revealed that the 100 most productive names were associated with a large majority of papers, and that those authors who were not among the most productive 100 had each contributed 3 papers or fewer over the previous 5 years. The decision was therefore taken to invite the most productive 200 researchers on the basis that this would cover almost the entire research community on this topic, regardless of the nature or importance of their discoveries.

It was agreed that an official approach through the World Health Organization (WHO), that agreed to serve as the hosting hub for the management group, would be most likely to persuade invited researchers to participate in the exercise. Moreover, mentioning that they were selected based on their placement among the 200 most productive researchers in this field would help to make them feel appreciated and that their work is valued. Nevertheless, even with these measures taken, the final response rate in terms of scoring in this small research community was 45/200 (22.5%).

Initially, the researchers were contacted through individual e–mails sent from the WHO, which explained the aim

of the exercise, acknowledged the contribution of each researcher to the field, and explained the type of the research idea that was sought – ie, neither too broad, nor too specific (this was further explained in the guidelines for implementation of the CHNRI method) [10]. They were also asked to consider different instruments of health research, ie, "description", "delivery", "development" and "discovery" and they were given an example of a "valid" research idea from each of those four types of research. They were initially given up to one month to submit as many research ideas as they wished, and two further reminders were sent at two weekly intervals following the initial deadline before the total number of submitted ideas reached 500. At that point, reminders were stopped and the management group studied the potential bias introduced because some researchers submitted many more ideas than others. At that point, a "consolidation" of the list of research ideas was conducted to ensure that the retained questions were evenly distributed across different research instruments and main research avenues and cover them all reasonably well. In this phase, all duplicate ideas were removed, while similar ideas were compressed into a single research question. This resulted in the reduction of the number of research ideas considered for scoring from 500 to 158, thus also making the scoring process more manageable.

Depending on the number of research ideas and the anticipated time required for scoring, one option is to offer the scorers the option of only scoring the criteria that they feel most comfortable with scoring – another flexibility in the CHNRI method. It is important that each scorer scores all research ideas on the same criterion, rather than scoring some but not all ideas for all criteria. This ensures that each research idea is scored by the same set of scorers, avoiding any personal preferences towards some ideas and keeping the process transparent and fair.

Given that scoring is time consuming, it was considered reasonable to allow the scorers about a month to reply, with two further reminders sent at monthly intervals after the deadline. After 3 months, the scoring process should typically be considered completed, the drop–out rate recorded, and the analyses can begin. The process of analysis of the scores is described in great detail in another paper [10].

EXAMPLE FROM THE CHNRI EXERCISE ON RESEARCH PRIORITIES FOR NEWBORN HEALTH

This study has been published in its extended form in this theme issue [7]. Although the field of newborn health in low–income settings is very recent and the research community is still quite small, and although the process of involving researchers followed many steps that were in common to the exercise on pneumonia 5 years earlier, several important innovations were introduced.

Similarly to the pneumonia exercise, the management group selected the 200 most productive researchers, based on the number of co–authored publications in peer–reviewed journals in the previous 5 years. However, the composition of those 200 researchers was more targeted in this case: in addition to inviting the 100 most productive researchers on newborn health globally, the 50 most productive researchers affiliated to institutions in low and middle–income countries (LMIC) were also invited. The final 50 invitations were reserved for the most productive researchers in the area of stillbirth research globally. The purpose of this approach to sampling was to avoid under–representation of researchers from LMIC and the small number of researchers who worked on the increasingly important issue of stillbirths. This was a carefully thought–through approach and is another example of the flexibility allowed in the CHNRI process. It is important to "design" the sampling process in a way that captures researchers who could be most informative for the specific exercise, which is likely to be more important for exercises that are very broad in scope and less important for those which are very narrow.

Photo: Researchers in Bangladesh working in their laboratory (Courtesy of Dr Ozren Polašek, personal collection)

Another innovation in this newborn health exercise was the inclusion of

programme managers, identified through the Healthy Newborn Network database. This was a suggestion made by several members of the management board in light of broad agreement that "description" research was no longer a priority and that the new focus should be on implementation. Therefore, the group recognised the need to include experts with first–hand understanding of the challenges with delivery, cost and sustainability of newborn health and stillbirth prevention programmes in LMIC settings. This resulted in about 600 potential scorers being invited to participate in the exercise, of which the majority (400) were program managers familiar with the challenges in low–resource settings. Eventually, 132 persons participated in the generation of ideas and 91 in scoring, bringing the final response rate to about 15%.

Another innovation in this exercise was the use of "Survey Monkey", which allowed the management group to keep track of the age, gender, geographic area, background and affiliation of each participating researcher/programme manager in real time. This innovation was seen as very useful, because it allowed more intense reminders that were being sent to specific groups of invitees who were falling behind and becoming under–represented.

To improve the response rate, the management team sent four and five reminders to the invitees for both submitting the ideas and the scores. The team met in Geneva for a week to consolidate the initial list of research ideas they had received from about 400 down to about 200 that were eventually scored. In summary, this exercise stands out in three ways: (i) the targeted sampling of researchers; (ii) the inclusion of programme managers as the majority of invited scorers, to better reflect the community with useful knowledge on the criteria, which is not necessarily reflected in academic articles; and (iii) the tracking of score responses in real time using survey monkey [7].

EXAMPLE FROM THE CHNRI EXERCISE ON RESEARCH PRIORITIES FOR DEMENTIA

The examples on childhood pneumonia and newborn health are both relevant to research fields with relatively small research communities. In both exercises, the CHNRI method was used primarily as a way to galvanise the community and define the strategy for the development of the field. The small number of productive researchers in both fields meant that nearly everyone who had contributed to the research field over the previous 5 years was invited to participate in the exercise. However, how should we select researchers when the research field is very large and has tens of thousands of actively participating researchers? One such recent example is the CHNRI exercise on dementia and Alzheimer disease, a field in which tens of thousands of researchers are active. This exercise represents a good example of the strategies that can be used to solicit input from researchers in such circumstances.

The management group numbered 15–20 members at various stages of the process and included representatives of the World Health Organization, several international societies and funders interested in this topic (eg, Alzheimer Disease International, USA–based Alzheimer Association, UK's National Institute for Health Research, Canadian Institute for Health Research and USA–based National Institute of Aging), together with leading researchers and opinion–leaders in the field who were based in academic institutions (Rudan I, personal communication). This diverse group needed to devise a plan for recruiting a large number of researchers to provide research ideas and scores for the vast multi–disciplinary field of dementia and Alzheimer disease research. They held several meetings and teleconferences during which they discussed the best strategy to address this difficult task.

Their discussions soon focused on finding the proper justification for inviting some researchers, while leaving many thousands of others outside of the exercise. The group started to look for an appropriate response to a likely *post–hoc* question "*Why wasn't I invited to participate, and other colleagues were?*" that would eventually be acceptable to all those who might ask this question. The group eventually agreed that a justification that was likely to be accepted by researchers in this area should have the following format: "*You were not invited because: (i) you were not among the most productive 500 researchers (in terms of the number of publications) in this field in the past 5 years; (ii) you were neither the lead, nor the senior author on any of the 50 most cited papers in each of the past 5 years; and (iii) you don't belong to any of the groups of researchers specifically targeted for inclusion (even if they do not fall into the first two categories); this mainly relates to the few researchers from low– and middle–income countries (LMICs)*".

Given that the line of whom to invite needs to be drawn somewhere, the CHNRI management group agreed that the justification provided above would have a good chance for being accepted by the entire research community. Indeed, if a researcher isn't among the 500 most productive in the field in the previous 5 years, they cannot easily take an issue over those 500 more productive researchers being invited. Moreover, if a researcher hasn't led the research on a paper that was later ranked among the 50 most cited papers on the topic in each of the 5 previous years, then they cannot easily take an issue over the invitation of those 500 further authors who were in this position (5 years × 50 papers × (1 lead +1 corresponding author) = 500 authors). This rule implied that up to 1000 researchers would be invited to participate – some based mainly on their productivity in this field, and others mainly on high impact of their work, with some overlap expected between the two groups.

Finally, given that the exercise was global in terms of geographical scope, and that the vast majority of the most productive and/or cited authors were based in wealthy countries, the group concluded that every effort should be invested to identify the third group to invite – composed of an unrestricted, but likely quite small number of prominent published researchers based in low– and middle–income countries, which would be sought for through a separate effort.

The productive authors for the first group were identified through a search of Web of Sciences' "Core Collection", which ranked all researchers in the world in the field of dementia or Alzheimer disease by the number of publications, limited to the output in the preceding 5 years (2009–2013). This allowed the CHNRI management group to identify the 500 most productive researchers. The group also needed to check and merge results for the same author who published with different initials (ie, interchangeably using only one or both initials in their papers). The contact details were then successfully obtained from their publications for a sizeable subset, although not for all. This potentially introduced a bias related to dropping those who couldn't be contacted from further stages of the process.

The group then used Web of Science's "Core Collection" to rank the papers published in each of the years 2009–2013 by the number of citations that each paper received by the end of 2014. For the 50 most cited papers in each year, the group identified the lead and the corresponding author (ie, the first and last listed). After removing duplicate entries – because some authors would be found on several such papers, and then also on the previous list of the most productive authors – the identified authors would be invited to participate in the exercise wherever their contact details could be found. All duplicates were removed, but the "new free places" were not filled with further scientists, because the justifications for inclusions were pre–set and it was not clear whether to keep filling the places based on productivity, citations, or some other criterion. This meant that the final number of invited researchers would decrease from 1000 to a smaller number. Due to the overlap, the described process yielded 672 researchers to be contacted.

In addition, Chinese databases were systematically searched. The papers published in those databases didn't have many citations (as checked through Google Scholar), so the ranking of papers by citations received could not have been used as a selection criterion in a truly meaningful way. The group therefore invited the most productive 50 authors from the Chinese literature over the preceding 5–year period (2009–2013). To identify the few researchers from other low– and middle–income countries, the Alzheimer Association, Alzheimer Disease International (ADI, which is the global umbrella organization of all national Alzheimer associations) and 10/66 dementia research group (broad network of researchers from low and middle income countries) were actively involved in identifying and contacting the experts in LMIC. In the end, about 800 researchers were identified for contact, and the contact details were successfully obtained for 69% of them, each of whom was asked to submit 3–5 research ideas. Then, a total of 201 experts responded and submitted 863 research ideas. Those ideas pertained to prevention, diagnosis, treatment or care for dementia and represented "basic", "clinical–translational" or "implementation" research, as categorized by the management group. The management group then decided that this number was too large to score, so they convened a meeting to review all received research ideas. They consolidated the list to 59 representative "research avenues/themes", which were broader than specific research ideas/questions. These broader avenues/themes were then scored using a slightly modified set of the 5 standard CHNRI criteria. Thus, this exercise developed not only an approach to the sampling of experts when a very large number of experts exists in the world, but also developed an approach to deal with an unmanageable number of specific research ideas/questions received from such a large expert group. It is possible that, in the final version of the published paper (which is now still under review), some minor practical modifications from this protocol will be observed (Rudan I, personal communication).

ETHICAL AND OTHER CONSIDERATIONS

Given that the CHNRI method essentially relies on input from human subjects (who are researchers in this case), we consider here the ethical aspects of conducting CHNRI exercises. The CHNRI exercises are a form of research that uses various measures of collective opinion as an output – eg, the level of collective support for a particular research idea, the extent of agreement within the collective, the variance in all expressed opinions, the average level of support across several criteria, and possibly others. Nevertheless, the input is based on individual opinions received from individual participants.

The method itself, as initially proposed [10], underwent ethical scrutiny at the institution where it was conceived – at the Croatian Centre for Global Health at the Faculty of Medicine of the University of Split, Croatia. The following recommendations were made:

(i) It is important to let all participants know, at the stage of inviting them to participate in the CHNRI exercise, that by responding to the invitation through submitting their ideas, and then their numerical scores, they acknowledge their voluntary participation in the exercise; this will deal

with the ethical concern over whether their participation is voluntary, and they would not need to sign a special informed consent;

(ii) Although the input received from the participants is encoded as a sequence of numbers (the scores), if it is presented in the supplementary material of the resulting papers under the scorers' personal names or surnames, and aligned against the research ideas that were scored, this can still be used to reconstruct their personal opinions on a wide range of research topics; this may make the participants (ie, scorers) uncomfortable. Therefore, unless specific approval is obtained at the individual or a group level to disclose all individual scores in the interest of transparency of the CHNRI exercise (which is a motivation that can be seen as being in conflict with ethics concerns in this case), we recommend that all scores disclosed in the public domain through publications should be anonymized. If the scores received from the scorers are anonymized in a proper way, and only the opinion of the entire collective is studied and interpreted, there should not be any ethical concerns related to the CHNRI exercise.

(iii) We see another theoretical ethical concern that should potentially be carefully managed; namely, if all participants and their scores are disclosed in the public domain, and the participants haven't been anonymised at their own request (ie, in the interest of transparency and legitimacy of the CHNRI exercise), then the participants should still be warned that further statistical analyses could potentially be performed on the data set that involves their names. Those analyses could focus on participants themselves as subjects, and "ranking" and comparisons among the participants, rather than research ideas. Therefore, everyone's input could be statistically compared to that of one or more other participants. Although this is never the intention or a focus of the CHNRI exercise, it is a theoretical possibility and it could identify some scorers as "outliers" in terms of scoring with respect to their colleagues, which may cause them an unforeseen concern.

If these theoretical concerns are appropriately addressed and managed, which can most easily be achieved through informing the participants of the scope of the exercise, explaining that by self–selecting themselves for the exercise they are acknowledging their voluntary participation, and anonymising their scores once they are received, the CHNRI method should be considered free from ethics concerns.

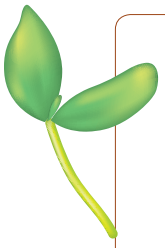
The managers of CHNRI exercises often ask whether the results of the exercise should be returned to all participants. We endorse this practice, because we can see no reason why this should not happen. It is in everyone's interest to inform them of the collective optimism/pessimism towards various research ideas within each research community, especially when the participants have freely offered their ideas and time for scoring.

This brings us to another frequent question, which is how to thank the participants for their contributions in terms of suggesting research ideas and dedicating their time to scoring? In the vast majority of the previously conducted CHNRI exercises, this was done through involving the participants in the resulting publication. This involvement could either take the form of equal co–authorship, or listing under the group co–authorship, or simply acknowledging their contribution in the acknowledgement at the end of the paper. The decision as to which of these three options to employ typically depends on the number of participants, the realistic prospects in involving them in other stages of writing of a resulting CHNRI publication (beyond purely providing the scores), and the preferences, restrictions, or authorship criteria of the journals to which the papers have been submitted. It is also possible to motivate the participants to participate in the CHNRI exercise by organising a meeting in a convenient location and supporting participants' travel and accommodation expenses, and then conduct the entire exercise over a few days in a location of preference or convenience. In some cases, this has been done to expedite the scoring process when speed is important as exercises can take quite a long time when conducted via e–mail [4–8].

CONCLUSIONS

To date, we have gained considerable experience with involving researchers as participants who provide research ideas and scores for the CHNRI exercises. We have tried to summarise some informative examples in this paper, irrespective of whether the chosen examples were necessarily the most successfully conducted CHNRI exercises. Indeed, it is difficult to judge whether the CHNRI exercise has been "successful", and what criteria should be used to do so. Clearly, a high participation rate should limit the scope for response bias (through self–selection), which is a major concern with CHNRI exercises. Then, a large and broadly inclusive spectrum of research ideas provided by participants and made available for scoring would certainly signal a success in conducting the exercise, although it is difficult to quantify this inclusiveness. Moreover, it would reflect researchers' willingness to share their ideas freely and take part in the process. Large differences in the final research priority scores (RPSs) received by various research ideas indicate that the criteria used are able to discriminate between ideas. If an exercise results in only small differences in RPSs then any ranking of research ideas based on the scores is unlikely to be very robust, and the exercise will have largely failed to meet its own objectives.

Finally, if the exercise is conducted reasonably quickly (typical time is about 3–6 months) and at low cost (typical direct financial costs are up to US$ 15000, unless the costs of organizing one or more meetings are envisaged), and all participants accept the results and co–author a resulting publication, then the exercise has served its purpose. This will be even more so whenever there is a vision of a follow–up to the exercise, in which a workshop is organised to arrange research proposal writing, or a special meeting with the funders is agreed to ensure that the priorities have been properly communicated. Dissemination of the results and an appropriate follow–up at national, regional and global levels are important parts of the CHNRI process, to increase the likelihood that the research on identified priorities is conducted in the near future. Evaluating whether CHNRI exercises have had an impact on those who invest in health research and influenced investment decisions is challenging and is will be addressed in future papers on the CHNRI method.



Acknowledgements: We sincerely thank the representatives of the researchers for taking part in the previous CHNRI exercises and providing useful information. SY is the employee of the World Health Organization. Her views expressed here do not necessarily represent the views of the Organization.

Funding: None.

Authorship declaration: SY wrote the article under supervision by her PhD mentor, Professor Igor Rudan. SC, KW, KYC revised the article, made many useful comments and provided important intellectual content.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no conflict of interest.

REFERENCES

1 Rudan I, Yoshida S, Chan KY, Cousens S, Sridhar D, Bahl R, et al. Setting health research priorities using the CHNRI method: I. Involving funders. J Glob Health. 2016;6:010301. Medline:26401269 doi:10.7189/jogh.06.010301

2 Yoshida S, Rudan I, Cousens S. Setting health research priorities using the CHNRI method: VI. Quantitative properties of human collective opinion. J Glob Health. 2016;6:010503. doi:10.7189/jogh.06.010503

3 Brown KH, Hess SY, Boy E, Gibson RS, Horton S, Osendarp SJ, et al. Setting priorities for zinc–related health research to reduce children's disease burden worldwide: an application of the Child Health and Nutrition Research Initiative's research priority–setting method. Public Health Nutr. 2009;12:389-96. Medline:18426636

4 GAPPS Expert Group on Community Based Strategies and Constraints, George A, Young M, Bang A, Chan KY, Rudan I, et al. Setting implementation research priorities to reduce preterm births and stillbirths at the community level. PLoS Med. 2011;8:e1000380. Medline:21245907 doi:10.1371/journal.pmed.1000380

5 Rudan I, Theodoratou E, Zgaga L, Nair H, Chan KY, Tomlinson M, et al. Setting priorities for development of emerging interventions against childhood pneumonia, meningitis and influenza. J Glob Health. 2012;2:010304. Medline:23198129 doi:10.7189/jogh.01.010304

6 Bhutta ZA, Zipursky A, Wazny K, Levine MM, Black RE, Bassani DG, et al. Setting priorities for development of emerging interventions against childhood diarrhoea. J Glob Health. 2013;3:010302. Medline:23826502 doi:10.7189/jogh.03.010302

7 Yoshida S, Martines J, Lawn JE, Wall S, Souza JP, Rudan I. Setting research priorities to improve global newborn health and prevent stillbirths by 2025. J Glob Health. 2016;6:010508. Medline:26401272 doi:10.7189/jogh.06.010508

8 Wazny K, Sadruddin S, Zipursky A, Hamer DH, Jacobs T, Källander K, et al. Setting global research priorities for integrated community case management (iCCM): Results from a CHNRI (Child Health and Nutrition Research Initiative) exercise. J Glob Health. 2014;4:020413. Medline:25520803 doi:10.7189/jogh.04.020413

9 Rudan I, El Arifeen S, Bhutta ZA, Black RE, Brooks A, Chan KY, et al. Setting research priorities to reduce global mortality from childhood pneumonia by 2015. PLoS Med. 2011;8:e1001099. Medline:21980266 doi:10.1371/journal.pmed.1001099

10 Rudan I, Gibson JL, Ameratunga S, El Arifeen S, Bhutta ZA, Black M, et al. Setting priorities in global child health research investments: guidelines for implementation of CHNRI method. Croat Med J. 2008;49:720-33. Medline:19090596

Correspondence to:
Dr Sachiyo Yoshida
Department for Maternal, Newborn, Child and Adolescent Health
World Health Organization
Avenue Appia 20, CH–1211
Geneva 27
Switzerland
yoshidas@who.int

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? The Journal of Global Health

1.2. When was the work published? June 2016

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

Not applicable

1.3. Was the work subject to academic peer review? Yes, the article went through peer review.

1.4. Have you retained the copyright for the work? **Yes** / **No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I have co-wrote the paper along with other co-authors. I did not contribute to the design and conceptualization of the exercise since exercise already took place in 2008.

NAME IN FULL (Block Capitals) SACHIYO YOSHIDA

STUDENT ID NO: 380010

CANDIDATE'S SIGNATURE [redacted] Date 25 July 2018

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) [redacted]

Setting health research priorities using the CHNRI method: I. Involving funders

Igor Rudan[1], Sachiyo Yoshida[2], Kit Yee Chan[1], Simon Cousens[3], Devi Sridhar[1], Rajiv Bahl[2], Jose Martines[2]

[1] Centre for Global Health Research, The Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, Scotland, UK

[2] Department of Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland

[3] London School of Hygiene and Tropical Medicine, London, UK

In 2007 and 2008, the World Health Organization's Department for Child and Adolescent Health and Development (later renamed as WHO MNCAH – Maternal, Newborn, Child and Adolescent Health) commissioned five large exercises to define research priorities related to the five major causes of child deaths for the period up to the year 2015. The exercises were based on the CHNRI (Child Health and Nutrition Research Initiative) method, which was just being introduced at the time [1,2]. The selected causes were childhood pneumonia, diarrhoea, birth asphyxia, neonatal infections and preterm birth/low birth weight [3–7]. The context for those exercises was clearly defined: to identify research that could help reduce mortality in children under 5 years of age in low and middle income countries by the year 2015. The criteria used in all five exercises were the "standard" CHNRI criteria: (i) answerability of the research question; (ii) likelihood of the effectiveness of the resulting intervention; (iii) deliverability (with affordability and sustainability); (iv) potential to reduce disease burden; and (v) effect on equity [3–7].

The five criteria used by the scorers were intuitive as they followed the path from generating new knowledge to having an impact on the cause of death. They were chosen with a view to identifying research questions that were most likely to contribute to finding effective solutions to the problems. However, after the five exercises – all of which were published in respected international journals [3–7] – the WHO officers were left with an additional question: how "fundable" were the identified priorities, ie, how attractive were they to research funders? More specifically, should another criterion be added to the CHNRI exercises, which would evaluate the likelihood of obtaining funding support for specific research questions?

To answer these questions, coordinators of the CHNRI exercises at the WHO agreed that it would be useful to invite a number of representatives from large funding organizations interested in child health research to take part in a consultation process at the WHO. The process aimed to explore funders' perspective in prioritization of health research. The funders would be presented with the leading research priorities identified through the CHNRI exercises and asked to discuss any potential variation in their likelihood of being funding. If all the leading priorities were equally attractive to funders and likely to attract funding support, this would indicate that the "standard" CHNRI criteria were sufficient for the process of prioritization. However, if there were large differences in attractiveness of the identified research priorities to funders, then adding another criterion to the exercise – "likelihood of obtaining funding support", or simply "fundability" – would be a useful addition to the standard CHNRI framework.

In 2007 and 2008, the World Health Organization's Department for Child and Adolescent Health and Development commissioned five large research priority setting exercises using the CHNRI (Child Health and Nutrition Research Initiative) method. The aim was to define research priorities related to the five major causes of child deaths for the period up to the year 2015. The selected causes were childhood pneumonia, diarrhoea, birth asphyxia, neonatal infections and preterm birth/low birth weight. The criteria used for prioritization in all five exercises were the "standard" CHNRI criteria: answerability, effectiveness, deliverability, potential for mortality burden reduction and the effect on equity. Having completed the exercises, the WHO officers were left with another question: how "fundable" were the identified priorities, i.e. how attractive were they to research funders?

THE MEETING WITH THE FUNDERS (GENEVA, 27–29 MARCH 2009)

In March 2009, MNCAH invited 40 representatives from funding organizations, including the Bill and Melinda Gates Foundation, the Wellcome Trust, National Institutes of Health USA, Department for International Development UK, Save the Children, INCLEN, EPICENTRE, UNICEF, USAID, PATH, Ministry of Science and Technology of India, Ministries of Health of Zambia, Pakistan and Brazil, Global Forum for Health Research, Trinity Global Support Foundation, Children's Investment Fund Foundation, Osaka Research Institute for Maternal and Child Health. Eventually, 16 representatives of funding agencies agreed to take part in the exercise under the condition of anonymity. Moreover, it was understood that their input would not necessarily be the official position of their respective funding agencies, nor would it create any form of funding obligation.

Having explained the aims of the consultation meeting to the representatives of funding agencies, the 16 participants were presented with a list of the top 10 research priorities for each of the five major causes of child deaths: pneumonia, diarrhea, birth asphyxia, neonatal infections and preterm birth/low birth weight [3–7]. This set of 50 research priorities represented roughly the top 5% of all the research ideas submitted for scoring during the CHNRI exercises.

The WHO coordinators (RB and JM) explained each of the 50 leading research priorities to the 16 donor representatives. Then, the 16 donor representatives were provided with the list of research priorities and asked to individually identify those that were most likely to receive funding support from their respective organizations.

Funding attractiveness was measured in two ways. First, funder representatives were asked to rank the identified research priorities according to their likelihood to receive funding support under an organization's current investment policies and practices. Second, funding attractiveness was measured by asking funder representatives to distribute a theoretical US$ 100 among the research priorities that seem most fundable. Results were used to facilitate discussion on what makes a research question attractive (or unattractive) for funding support. The scoring sheet that was given to meeting participants is shown in **Figure 1**. While they did not need to provide their name or organization, they were asked to assign ranks 1–10 to the ten research priorities identified for each of the five causes of death (column 1), and also to distribute a hypothetical US$ 100 to different research priorities in concordance to the likely funding support that they may obtain.

Sixteen participants scored the identified research priorities according to the instructions (**Figure 1**). The average ranks across the 16 participants (1 = most likely to be funded; 10 = least likely to be funded) assigned to the 50 research priorities ranged from 3.7 to 7.2. The average US$ amount assigned to research priorities ranged from US$ 20.1 to US$ 2.5. There was general consistency between ranks and the US$ assigned to research priorities.

CAUSE OF DEATH (1 / 5) - e.g. PNEUMONIA

NAME:
ORGANIZATION:

	RANK?	US$?
RESEARCH QUESTION 1		
RESEARCH QUESTION 2		
RESEARCH QUESTION 3		
RESEARCH QUESTION 4		
RESEARCH QUESTION 5		
RESEARCH QUESTION 6		
RESEARCH QUESTION 7		
RESEARCH QUESTION 8		
RESEARCH QUESTION 9		
RESEARCH QUESTION 10		

Figure 1. A questionnaire that was given to 16 funder representatives at the meeting to obtain information useful to understanding funding attractiveness of different research priorities.

Importantly, the analysis of the collective input based on the 2nd column (ie, assigned US$), presented in **Figure 2**, clearly shows that there was a rather substantial departure of the assigned funds from that expected at random: if all research priorities were equally likely to obtain support from the funders, then all the bars would be extending only to the line that represents an investment of US$ 10.0. Furthermore, 4 research priorities (8%) clearly stood out from the rest [8]. It was agreed that they might provide a starting point from which MNCAH Department could concentrate its efforts. These 4 research priorities are shown in **Table 1**.

Table 1. The 4 research priorities (8%) that were identified as positive outliers in terms of their likelihood to obtain funding support

Evaluate the quality of community workers to adequately assess, recognize danger signs, refer and treat acute respiratory infections (ARI) in different contexts and settings.
What are the barriers against appropriate use of oral rehydration therapy?
What are the feasibility, effectiveness and cost of different approaches to promote the following home care practices (breastfeeding, cord/skin, care seeking, handwashing)?
What are the feasibility, effectiveness and cost of a scheme of routine home visits for initiation of supportive practices, detection of illness and newborn survival?

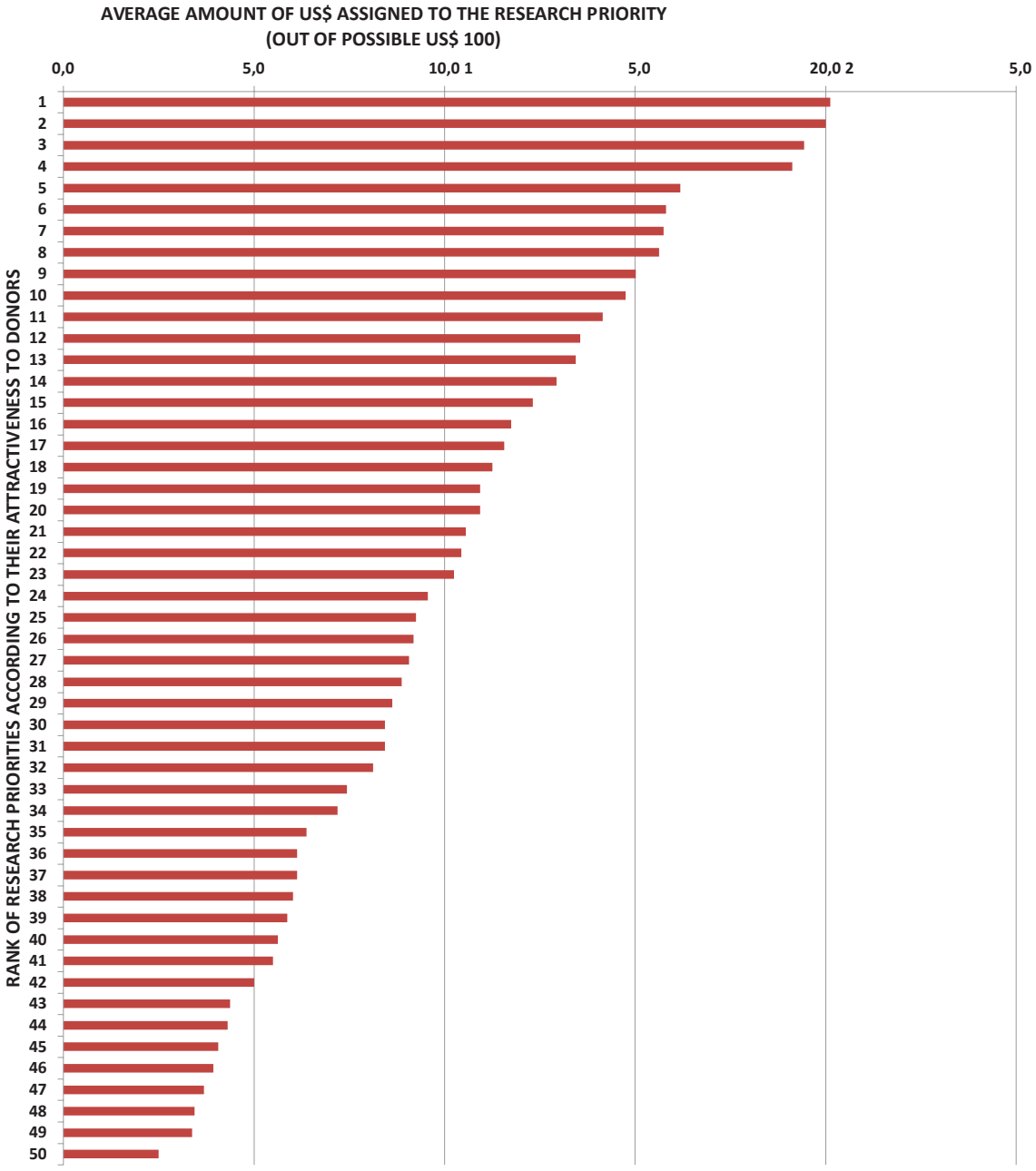


Figure 2. The results of the collective input from 16 funder representatives, showing large differences in funding attractiveness between 50 research priorities. No substantial differences in funding attractiveness would be indicated by equality of the scores on the horizontal axis at the US$ 10.0 line.

AN ANALYSIS OF THE EXERCISE WITH FUNDER REPRESENTATIVES

The results were analysed after the first day of the meeting and presented to donor representatives at the beginning of the second day of the meeting. An open discussion was held with participants to understand and interpret the results of their collective input. Participants agreed that the most important criteria for research prioritisation differed between researchers and funders. Researchers tended to value answerability, effectiveness, deliverability, impact on the burden and equity. Funders were also interested in the clarity and specificity of research ideas, value for money, novelty, international competitiveness of the groups proposing the research, linkages to broader societal issues, and complementarity with other long–term strategic investments that were already made. An important point in the discussion was that researchers and research funders, especially those in the private sector, often speak quite different languages. Researchers need to be clear on what their goals are and communicate these in more readily understood terms. This point is particularly important because it implies that the CHNRI exercises' research priorities that were identified as most likely to generate useful new knowledge may not be considered equally relevant by the funders.

In March 2009, WHO officers invited 40 representatives from organizations that provide substantial funding support for global child health research to take part in a consultation process at the WHO. The process aimed to explore funder's perspective in prioritization of health research. Eventually, 16 funders' representatives agreed to take part in the exercise under the condition of anonymity. Participants agreed that the most relevant criteria for prioritisation differed between researchers and funders. Funders are interested in clarity and specificity of research ideas, value for money, novelty, international competitiveness of the groups proposing the research, links to broader societal issues, and complementarity with other long–term strategic investments that they have already made. Some may be particularly interested in the potential for forming partnerships between researchers and industry to improve the translation of findings and their application

This should certainly be taken into account when presenting and discussing the results of the CHNRI exercises.

Moreover, there seem to be important differences between the categories of funders in the criteria that they use to decide on research priorities. Generally, all investors in health research are concerned with answerability of the proposed research ideas in an ethical way, feasibility and value for money. However, some may be particularly interested in potential for forming partnerships between researchers and industry to increase the translation of findings and their application. Ministries and international organizations appeared more interested in deliverability, affordability and sustainability of the resulting interventions, local and national research capacities to carry out the proposed research ideas, and whether a research question is linked to an ongoing public debate or an important societal issue. Industrial donors may be primarily motivated to generate patents and translate research results into commercial products. Finally, society as a whole may be more concerned with issues of safety and equity issues and ask whether implementation of research results would widen the existing socio–economic gaps..

Transparency of research priority setting processes must, therefore, begin with those who invest. Perceived returns on investments in health research should be clearly stated at the beginning of the process. They may be defined as reduction in disease burden wherever public money is being invested. Investors from industries may see patentable products as their preferred returns. Non–profit organizations may be primarily interested in increased media attention for their agenda. The context in which investment prioritization takes place is thus primarily defined by expected returns of the funders. Moreover, their investment styles may be balanced and responsible (suggested for those investing public funds), risk–averting (which may be preferred among some industrial partners) or risk–seeking and biased towards high risk – high profit avenues of health research (which may be typical for some industry and not–for–profit organizations).

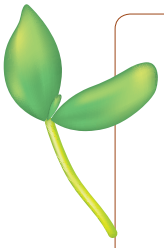
Apart from funders' perceived returns and their investment styles, the population, geographic area and disease burden of interest, the time frame in which returns are expected is an important defining component of the overall context. Priorities can differ substantially if the overall context is one of great urgency to tackle a problem, or whether decisions are made on very long–term, strategic investments.

CONCLUSIONS

The meeting with research funders organized by the WHO MNCAH department in March 2009 was exceptionally useful in understanding that funders certainly have their own views on what represents an attractive funding option.

Those views are not generalizable and may differ between categories of funders. Moreover, funders' perspectives are often quite different from those of researchers, or wider stakeholder groups. It is important to involve funders early in the process of setting research priorities, such as the CHNRI process, to encourage their ownership of the results. Funder–supported criteria must be taken into account, in addition to those preferred by the researchers and wider stakeholders. Otherwise, the outcomes of research prioritization exercises may have very limited impact on funders' decision making.

The key value of the CHNRI method to funders lies in its ability to transparently lay out the potential risks and benefits associated with investing in many competing research ideas, drawing on collective knowledge of the broad research community. Results of the CHNRI process represent an attempt on the part of researchers to communicate their views and opinions to funders in a way that is easily understood, transparent, replicable and intuitive. It provides useful additional information that funders may, or may not take into account when deciding on their own research agenda. From a methodological perspective, finding appropriate and effective ways of involving funders in future CHNRI exercises, communicating the outcomes clearly, and securing their commitment to acknowledge the results of the CHNRI process remain considerable challenges. An even greater challenge in future years will be to develop tools that can detect and evaluate the impact of CHNRI exercises on funder decision making and any change in funding priorities as a direct result of the CHNRI process. This should be particularly relevant to those who make decisions about investing public funds, whose primary agenda should be improving public health in the most cost–effective way – a target that CHNRI exercises should serve quite well.



Acknowledgements: We sincerely thank the representatives of the funding agencies for taking part in the consultation process and providing useful information under the condition of anonymity. SY, RB and JM are the employees of the World Health Organization. Their views expressed here do not necessarily represent the views of the Organization.

Funding: None.

Authorship declaration: JM and RB organized the consultation at the World Health Organization and IR moderated the consultation process. IR, SY and KYC wrote the article. SC and DS revised the article and made useful comments.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no conflict of interest.

REFERENCES

1 Rudan I, Chopra M, Kapiriri L, Gibson J, Lansang MA, Carneiro I, et al. Setting priorities in global child health research investments: universal challenges and conceptual framework. Croat Med J. 2008;49:307-17. Medline:18581609 doi:10.3325/cmj.2008.3.307

2 Rudan I, Gibson JL, Ameratunga S, El Arifeen S, Bhutta ZA, Black M, et al. Setting priorities in global child health research investments: guidelines for implementation of the CHNRI Method. Croat Med J. 2008;49:720-33. Medline:19090596 doi:10.3325/cmj.2008.49.720

3 Rudan I, El Arifeen S, Bhutta ZA, Black RE, Brooks A, Chan KY, et al. Setting research priorities to reduce global mortality from childhood pneumonia by 2015. PLoS Med. 2011;8:e1001099. Medline:21980266 doi:10.1371/journal.pmed.1001099

4 Fontaine O, Kosek M, Bhatnagar S, Boschi–Pinto C, Chan KY, Duggan C, et al. Setting research priorities to reduce global mortality from childhood diarrhoea by 2015. PLoS Med. 2009;6:e41. Medline:19278292 doi:10.1371/journal.pmed.1000041

5 Bahl R, Martines J, Ali N, Bhan MK, Carlo W, Chan KY, et al. Research priorities to reduce global mortality from newborn infections by 2015. Pediatr Infect Dis J. 2009;28 Suppl:S43-8. Medline:19106763 doi:10.1097/INF.0b013e31819588d7

REFERENCES

6 Lawn JE, Bahl R, Bergstrom S, Bhutta ZA, Darmstadt GL, Ellis M, et al. Setting research priorities to reduce almost one million deaths from birth asphyxia by 2015. PLoS Med. 2011;8:e1000389. Medline:21305038 doi:10.1371/journal.pmed.1000389

7 Bahl R, Bhandari N, Bhutta ZA, Biloglav Z, Edmond K, Iyengar S, et al. Setting research priorities to reduce global mortality from low birth weight by 2015. J Glob Health. 2012;2:010403. Medline:23198132 doi:10.7189/jogh.01.010403

8 World Health Organization. Consultation Proceedings: Identifying priorities for child health research to achieve Millennium Development Goal 4. DRAFT April 30, 2009. Geneva: WHO, 2009, pp. 1–29.

Correspondence to:
Professor Igor Rudan
Centre for Global Health Research
The Usher Institute for Population Health Sciences and Informatics
The University of Edinburgh Medical School
Teviot Place
Edinburgh EH8 9AG
Scotland, UK
igor.rudan@ed.ac.uk

Summary of key findings

Involving stakeholders

Many CHNRI exercises did not involve stakeholders in research priority setting processes. When stakeholders were involved, their profile and level of involvement varied between exercises. In some CHNRI exercises, stakeholders were involved in research idea generation, in other exercises they weighted the criteria or provided thresholds for the criteria. The fact that most previously conducted CHNRI exercise did not place great importance on involving stakeholders may be due to the fact that the profile and the quantity of stakeholders are not well defined. Moreover, even when they are involved, their inputs did not change the ranking of the top research options. To bridge this gap recommendation of profile and quantity of stakeholder is needed to guide the users of the CHNRI method.

Involving researchers

To date, CHNRI exercises attracted participation from more researchers than other groups of experts. Thus, considerable experience was gained in involving researchers in providing research ideas and scores for the CHNRI exercises. The paper presented two CHNRI exercises conducted in newborn health and birth outcomes, and dementia. Paper 4.2 argued that *"High participation rate by researcher would reflect researchers' willingness to share their ideas freely and take part in the process. Large differences in the final research priority scores (RPSs) received by various research ideas indicate that the criteria used can discriminate between ideas. If an exercise results in only small differences in RPSs then any ranking of research ideas based on the scores is unlikely to be very robust, and the exercise will have largely failed to meet its own objectives".*[26] However, I believe that it is difficult to judge whether these previously conducted CHNRI exercises were successful without an evaluation based on some framework. To fill this gap Chapter 4 evaluates the quality of the process in previously conducted CHNRI exercises using a pre-defined evaluation framework.

Involving funders

Involving funders in the CHNRI process is another aspect of the process that has been less explored. Paper 4.3 presented the outcome of an exploratory exercise with donors to understand the donors' perspective. Analysis of the explanatory exercise showed that the funders' perspective differs from that of the researchers or stakeholders in that funders were interested in value for money, novelty, international competitiveness of the groups proposing the research, linkages to broader societal issues, and whether the new priorities are aligned to funders' prior investment. On the other hand, researchers place greater importance on values such as effectiveness, deliverability, impact on the burden reduction i.e., standard CHNRI

criteria. To incorporate their view, the paper pointed to a need for funders to be involved at an early stage to help ensure that the outcome of the exercise will be easily understood by the funders. However, approaches to involve funders in a CHNRI process may vary by the context and scope of the exercises therefore the best way to involve them is yet to be determined.

Chapter 4. Evaluation of previously conducted health research prioritisation exercises

Impact of the research prioritisation exercises

In the previous chapters, I was able to assess components of the CHNRI method. Firstly, I explored fundamental principle of the method, the notion of the "wisdom of crowds". Secondly, I reviewed and showcased examples of how participants were involved in previously conducted CHNRI exercises. Thirdly, I conducted statistical analyses to try to identify optimal number of experts required in a CHNRI exercise.

Chapter D is a synthesis of previous chapters. This chapter provides thorough evaluation of the previously conducted research prioritisation exercises using CHNRI method to assess quality of the process and impact of such research prioritisation exercises.

Till date, many papers have been published on the strengths and weaknesses of RP methods and their implementation. However, evaluations of RP exercises are rarely reported, regardless of the method or approach used.[7,29] There is indeed scant guidance on how evaluations should be conducted, and which aspects should be assessed. Firstly, this is because of the heterogeneity of contexts and the different approaches being used at different levels. Secondly, most research prioritisation exercises are one-off exercises and there is rarely an *a priori* plan put in place before the RPs to translate the result into implementation.[13,30] Viergever et al advocate having a pre-determined strategic plan for the translation of priorities into actual research;[31] however, many research prioritisation exercises do not have such strategy.

Viergever et al proposed a checklist to assess the quality and impact of an RP exercise.[31] This checklist provided overall guidance for nine themes:

1. Context
2. Inclusiveness
3. Information gathering
4. Planning for implementation
5. Use of a comprehensive approach
6. Criteria
7. Method for deciding on priorities
8. Transparency
9. Evaluation

These were suggested as themes for which RP exercises should have clear processes. However, the checklist did not provide specific guidance as to what to check and the means against which to verify how each theme should be assessed. To fill this gap, Mador et al provide examples of

indicators and suggested modifications to the existing checklist including examples of indicators and data sources for each theme. They argued that the checklist is not a validated tool for evaluation but is a process guide to good RP practice.[2] Kapiriri et al propose a framework for successful priority setting in LMICs based on a literature review and interviews.[32] They provided indicators for immediate and delayed measures of both quality and impact of RP exercise; some parameters overlap with themes identified in the checklist such as increased use of evidence (information gathering in the checklist), a fairer priority setting process (criteria, inclusiveness, transparency in the checklist). On the "evaluation" theme in the checklist, three parameters were proposed: reflection on public values; increased public awareness of priority setting; and increased public confidence and acceptance. To the best of my knowledge, this framework developed by Kapiriri and Martin has not been used to evaluate RP exercises using the CHNRI method.

This Chapter proposes a modified evaluation framework adapted from various sources.[2,10,31–33] It also applies this framework to assess the previously conducted CHNRI exercises and to examine the feasibility of applying it.[34]

Method

The following sources have provided the basis for the conceptual framework to evaluate the quality and impact of RP exercises: *Guidelines for Implementation of CHNRI Method* by Rudan et al;[10] the framework for evaluation developed by Kapiriri and Martin;[32] the checklist to assess the RP process by Viergever et al;[31] further modifications of the checklist suggested by Mador et al,[2] and the synthesis of desired features presented by Kapiriri et al.[33]

A modified framework for evaluation of quality and impact of RP exercise was developed based on four sources and lessons learned from my experience in coordinating one of the CHNRI RP exercises. The proposed framework is presented in **Table 9**. Themes were adopted from the checklist of nine common themes by Viergever. Evaluation questions on the theme of "context" were adapted from guidelines written by Rudan et al. The remainder of the evaluation questions were adapted from an evaluation conducted by Mador et al, a review of tools by Kapiriri et al, and feedback from participants in the previous CHNRI exercises. Examples of objectively verifiable indicators were adapted from the framework for evaluation suggested by Kapiriri and Martin.

The "evaluation" section consists of two types of evaluation: process evaluation and impact evaluation. Process evaluation assesses the quality of the process in RP exercises and could be conducted immediately after the completion of RP exercises. Impact evaluation assesses the

impact of RP exercises on the alignment of funding allocation to priorities identified by the exercise, and the effect on health research institutions. Impact evaluation can be initiated sometime after the dissemination of the results, while the assessment of effect on research institutions will require a much longer time since priorities should first be translated into research, which usually takes about 5 years.

Table 9. The proposed framework for evaluation adapted from various sources

Theme	Description as outlined in the checklist.	Evaluation question as adapted from various sources.	Examples of objectively verifiable indicators to assess CHNRI exercises, if applicable.	Means of verification.
Context	Articulating the contextual factors that underpin the process.	i. What is the population to which health research should contribute in reducing disease burden and in improving health?	Non-quantifiable, i.e., definition of population the end-product of research should contribute to.	
		ii. What is the time-span of the research itself?	Number of years in which the result of research is expected.	Consultation with management team.
		iii. What is the overall motivation of research prioritisation?	Existing documentation on the process of selection.	
		iv. What nature of research is this prioritisation focused on?	Domains of research such as 4Ds (Discovery, Delivery, Development and Description research).	
Inclusiveness	Deciding who should be involved in setting research priorities.	i. How were the participants selected?	Existing documentation on the process of selection.	Observation at meetings, meeting minutes, publication.
			Number (%) of respondents who provided research ideas.	List of participants.
		ii. Was there a balanced representation?	Number (%) of professional background of respondents.	List of participants with baseline characteristics.
		iii. Was there a scope of self-selection bias?	Number (%) of professional background of respondents vs. non- respondents.	Analysis of non-respondents.
		iv. Were stakeholders involved in the process?	Explanation on how and when they are involved.	Observation at meetings, meeting minutes, publication.
		v. What was their profile?	Number (%) of professional background of stakeholders.	List of participants.
Information gathering	Choosing what information should be gathered to inform the process.	i. What sources and types of evidence/information/ data required were identified?	All-cause or cause-specific mortality rate due in a given population group. Morbidity in a population group due to disease (disability-adjusted life years – DALYs).	Consultation with management team.
		ii. Did the process consider available priorities?	Existing report on review of available priorities.	Consultation with management team, existing report documenting

		iii. How was the decision made about the level of RP (global level RP, regional level, national level, sub-national level)?	Existing report about how the decision was made	the process of RP.
Planning for implementation	Establishing plans for translating research priorities into action.	i. Was there a pre-defined plan to translate research priorities into implementation?	Existing plan describing the process.	Report.
Criteria	Selecting relevant criteria to focus discussion.	How were the criteria decided?	Number of decision appealed, number of decision revised.	Report documenting the process of RP.
Method for deciding on priorities	Choosing a method for deciding on priorities.	How was the method chosen?	Existing report documenting the selection of method.	Report documenting the process of RP.
Use of comprehensive approach	Assessing whether a comprehensive approach is necessary or if a tailored process and methods are required.	i. Were other comprehensive approaches considered?	Existing report discussing the comparative advantage of the method.	Report documenting the process of RP.
		ii. Was the process of consolidating, refining the research questions documented?	Existing report on how original research questions were modified by management team to address duplication and improve clarification	
Transparency	Communicating the approach that was used to set priorities.	i. Was the process of RP documented?	Existing report or publication.	Report documenting the process of RP.
		ii. Was information material sent to the participants involved in the PS process?	Existing proof of information shared.	Proof of email sent
Process evaluation	Defining when and how evaluation of process and outcome will occur.	i. Was the evaluation of RP process conducted with participants directly involved in the process?	Virtual meetings, face-to-face meetings with participants.	Note of the meeting(s).
		ii. Was there an increased awareness of priority setting?	Number (%) public aware of existing priority setting process. Number of complaints from the public.	Awareness survey.
		iii. Were the results of the RP exercise accepted by users?	Number of citing articles, Number (%) of institution, country, and donor affiliated with authors of the citing articles.	
Impact evaluation		iv. Degree of alignment of resources allocated to priorities	Amount of funds allocated to the priorities.	Financial report (if accessible).

	v. Did RP exercise lead to building capacity of health research system?	Number of local health research institutes in conducting research.	Technical report provided by health research institute participating in research under priority area.

Data analysis and interpretation

Data from reports and publications, and informal feedback from participants of the previously conducted CHNRI exercises were consolidated and reviewed. The RP process on newborn health and birth outcomes was evaluated for all themes except indicators *ii* and *iii* in the theme of "evaluation". The evaluation question *ii* was assessed for the past four CHNRI exercises which were led by the MCA/WHO in the area of pneumonia,[16] diarrhoea,[17] preterm birth,[19] and neonatal infections.[35] These four exercises were strategically chosen to ensure the time span of five years after the CHNRI exercises, with an assumption that after five years results of research would be would be made available in publications. Consequently, we did not include the CHNRI exercise on preterm birth and low birth weight babies, which was published in 2012.

Parameter *iii* was assessed for the above four and for one more CHNRI exercise on the theme on preterm birth and low birth weight babies.[18] The primary goal of the analysis is to understand the acceptance of results from the previously conducted CHNRI exercises prior to 2012 by users of research priorities including research community. Consequently, I have included all five CHNRI exercises coordinated by MCA/WHO. The two RP exercises I have coordinated were not included in this analysis because they were published in 2016.

To assess if there was a likelihood of self-selection bias in the process, a web-based search on Google and Google Scholar was carried out to identify their professional profile and affiliation of the non-respondents and respondents of the CHNRI exercise. In this exercise, a total of 578 experts were approached by the CHNRI management team. These experts consisted of researchers, identified based on publications between 2007 and 2011 (n=200), and professionals in the Global Newborn Network (n=378). In this analysis, the non-respondents refer to those who were approached but did not provide scores. This information for the respondents was obtained from the first part of the online survey in which they had provided research ideas. Professional background was available for 501 non-respondents. The information was classified into each professional category: the response rate is presented for each category and the proportion in each professional category was compared between the respondents and non-respondents.

To assess whether there was an increased awareness of the RP exercises, an awareness survey was conducted. The survey targeted individuals in communities in which research prioritisation exercises would be expected to have had some impact. To identify them, I conducted a literature review aimed at identifying publications relating to the top five research priorities identified in the four previously conducted CHNRI exercises between 2009 and 2011. Peer reviewed articles published within five years of the publication of CHNRI exercises were examined. For example,

for the CHNRI research prioritisation exercise on childhood diarrhoea published in 2009, I looked for literature published between 1 January 2010 and 31 December 2014. PubMed was used to identify publications on the priority areas. Medical subject headings (MeSH terms) were used to allow for multiple terms with identical meanings within one search. The abstracts of all search results were screened manually, to assess whether the publication related to the RPs. If it did, the paper was reviewed to see if there was any mention of the CHNRI exercises in the main body of the text. For papers in which there was no mention of the CHNRI exercises, I obtained contact details of the corresponding authors from the article and got in touch with them to ask if they were aware of the CHNRI prioritisation exercise and if so whether it had influenced them in submitting their proposal or in planning for the submission of the proposal.

As an indirect measure to assess acceptance of the results of the previously conducted RP exercises, a web-based search was conducted to analyse articles citing the past five CHNRI RP exercises. A total of 248 citing articles were found using Web of Science® citation analysis as at 2 June 2018. Information such as author-institutions, country of the institutions, and funding sources were extracted from these articles.

Results

Context

Context was well described in the report and the publications reviewed for the RP exercise on newborn health and birth outcomes. Context included the target population, the motivation, the domains of research, and the time frame. Context was defined by the management team. The time frame for the expected impact of the research extended to 2025 to allow for medium term and long–term research investments.

Inclusiveness

i. How were participants selected?

The selection of the participants for the RP exercise on newborn health and birth outcome was described in the reports and the publications. In the absence of clear guidance on optimal sample size of participants in a CHNRI exercise, the management group chose to contact as many experts as possible.

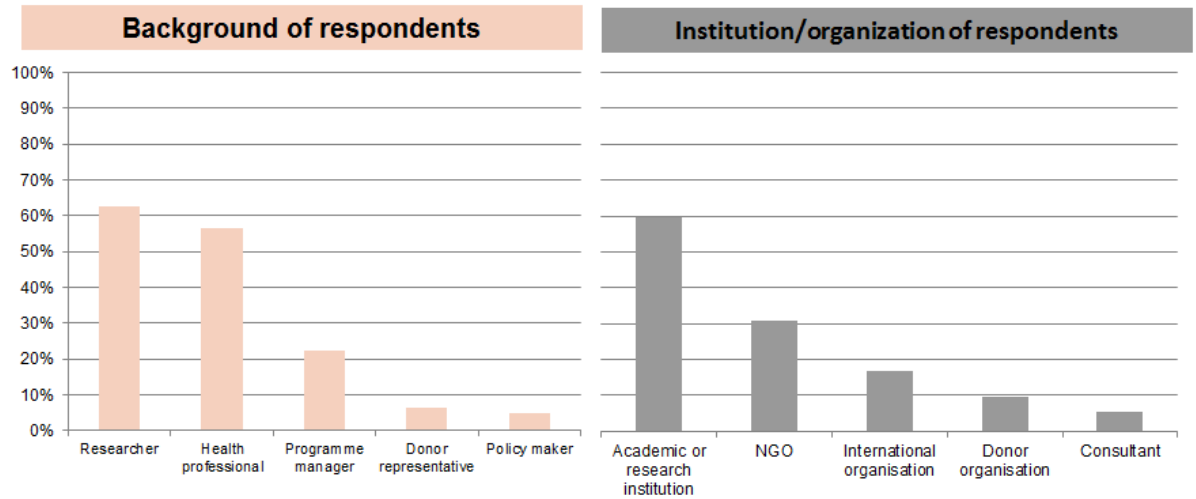
In summary, 578 people were identified in total, of whom 100 were identified based on academic achievement by number of publications in peer reviewed journals. Web of Science® was used with the following restrictions: neonatal OR newborn* OR neonate* AND document type (Article) AND year published (2008-2012). In total, 39,377 names were found as at March 2012 and the

top 100 were selected. The same criteria were used with the addition of the entire list of countries categorised under LMIC, e.g. Iraq OR Iran OR Jamaica, etc. In total, 3775 names were identified and the top 50 were selected. Similarly, we conducted a search using the following criteria: (stillbirth*) AND document type (Article) AND year published (2008-2012). Of the 1494 names identified in our search, the top 50 were selected. Programme experts were identified through databases maintained by Save the Children and Saving Newborn Lives (SNL) which also contained government officials, donors and UN staff. A total of 378 programme experts were chosen. Each expert was included as either the most productive researcher or the programme expert.

iii. Is there a balanced presentation?

Of the 578 approached, 132 (23%) responded with ideas. Half of the respondents were based in LMICs in Africa, Asia and South America, and the rest in HICs (**Figure 9**). According to the self-reported professional background (multiple choices allowed), approximately 60% classified themselves as a researcher. More than half were affiliated to academic or research institutions. The analysis showed that, although 378 programme managers were invited, their representation remained low (20%). The low representation of program managers despite twice as many program experts as researchers were approached, implies much lower response rate than that of researchers

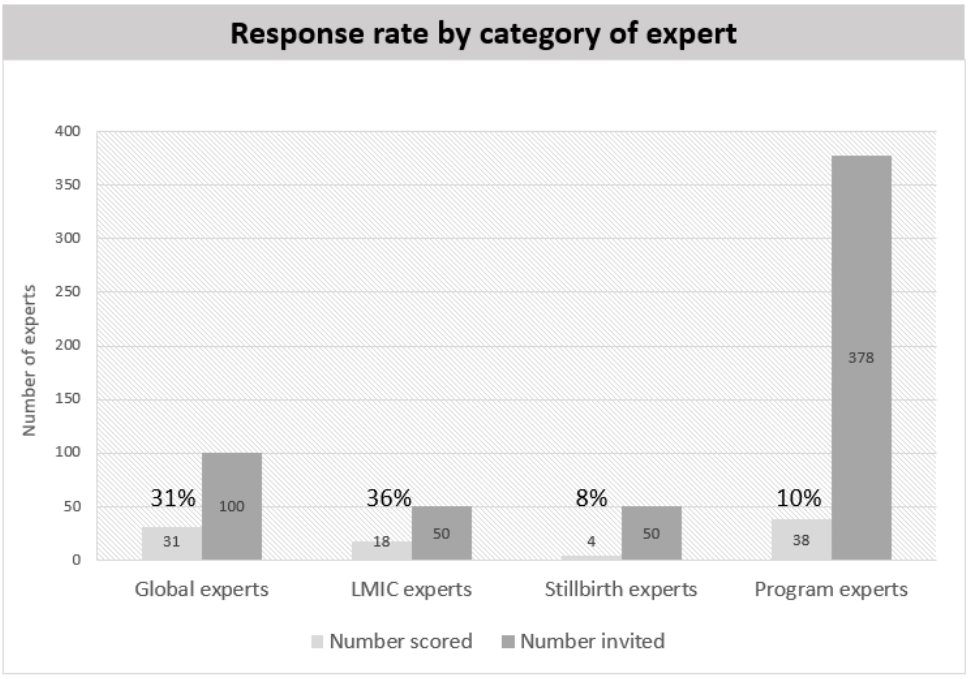
Figure 9. Professional background of the participants who provided research ideas (n=132)



Of the 578 contacted, 91 (16%) provided scoring. Of the 91 scorers, 31 (34%) were from the group of 100 whom we identified based on academic achievement expressed as the number of publications in peer-reviewed journals (global experts), 18 (20%) were from the group of 50 experts based in LMICs (LMIC experts), 4 (4%) were from the group of experts we identified by limiting the search to those who published in the area of stillbirth research (stillbirth experts),

and 38 (42%) scorers were from the group of 378 program experts (program experts). When response rate was calculated by originally selected experts, the highest response was given by LMICs experts (36%), followed by global experts (31%). The lowest response was given by group of stillbirth experts. Only one in 10 program experts provided scoring (**Figure 10**)

Figure 10. Response rate by category of expert



iii. Was there a likelihood of self-selection bias?

The CHNRI process is not free of potential bias. A potential criticism of the CHNRI method is that the process obtains the collective opinions of the limited group of people who self-select to participate in the exercise. In other words, the scoring may be affected by on-going research in which self-selected participants have relevant interests. An analysis was conducted to explore if there are any identifiable differences between self-selected respondents (scorers, n=91) and non-respondents (non-scorers, n=592). The analysis shows the response rate by professional category. The response rate was very low for most categories, ranging from 9% to 17% (**Table 10**). Higher response rates in donor representatives (44%), donor organizations (75%) and consultants (82%) were based on small numbers of professionals in these categories. The low overall response rate, though spread across all categories, indicates substantial potential for self-selection bias. The analysis of scorers and non-scorers by professional category shows that most of the scorers were researchers (**Figure 11**). This contrasts with our strategy, according to which we approached almost twice as many programme experts as researchers. Further analysis was conducted to examine if there is any differential pattern in the scores between

researchers and programme managers among the top 10 priorities. Median (IQR) of scores given by researchers and programme experts for total and by criteria are presented **(Figure 12**). The results show that the total score was similar; both had the same tendency to value more operational research over development or discovery research. However, if analysed by criteria, different patterns were observed; programme managers valued delivery and development research more than researchers through "impact in reducing mortality". Another notable difference was observed in development research in that score on "equity" was higher among programme experts than among researchers.[27] Scoring patterns on "discovery" research options were similar in both groups, overall and for each of the criteria. The similarity in the overall scores between the two groups implies that the finding was not be affected by self-selection bias

Despite our efforts to reach out to programme managers, we seem to have identified experts with roles other than programme manager. Of the "programme experts" contacted, only 42% were categorised as a programme manager by the web-based search among non-respondents. The rest were classified as researchers (35%), health professionals (12%) policy makers (12%), and donors (2%) **(Figure 13**). We only used the database managed by Save the Children-SNL which had an extensive list of experts working around programmes on maternal and newborn health. Of the scorers whom we identified in the list of programme experts, nearly half self-classified as researchers, and less than one in three as programme managers (**Figure 14**). It would have been helpful to look for other sources of information to cross-check the experts' background and to ensure better coverage of experts. Examples of such sources include the list of programme experts maintained by Partnership for Maternal, Newborn and Child Health (PMNCH).

Table 10. Response rate of scorers

	Total number of experts contacted (scorers+non-scorers)	Number of scorers	Response rate (%)
By professional background			
Researcher	338	59	17%
Program manager	155	15	9%
Health professional	45	6	14%
Policy maker	40	5	12%
Donor representative	14	6	44%
Total	592	91	15%
By institution/Organization			
Academic or research institution	305	62	21%
NGO	165	12	6%
Government	68	5	13%
Donor organization	8	5	75%
International organisation	42	5	0%
Consultant	4	2	82%
Total	592	91	15%

Figure 11 Professional background/institution of the scorers (n=91) and non-scorers (n=501)

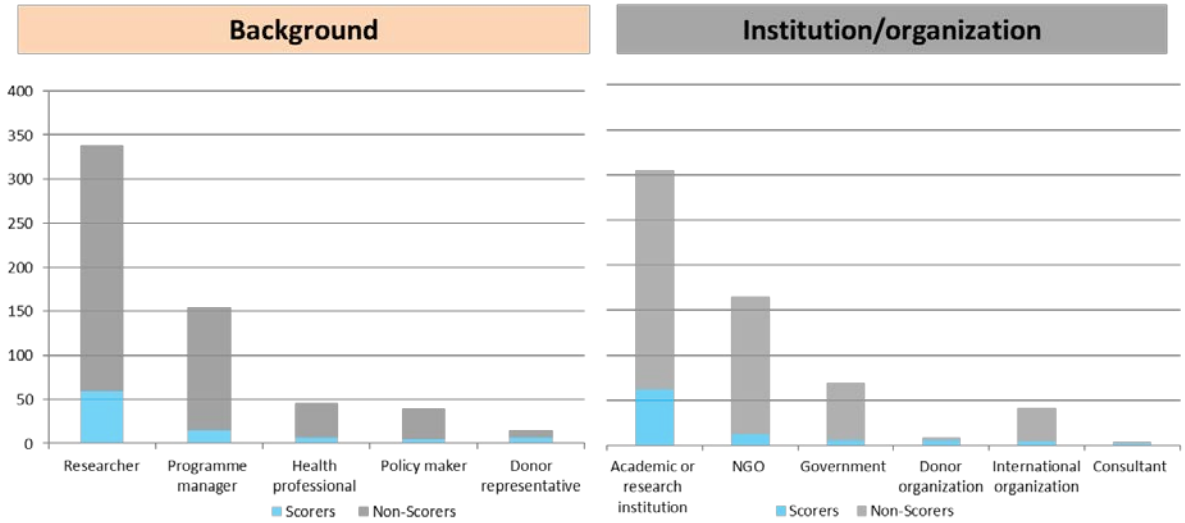


Figure 12 Overall scoring patterns by profile of experts (adopted from Yoshida et al 2016)

	Median (IQR)		
	All scorers (n = 91)	Researchers (n = 61)	Programme experts (n = 30)
TOTAL SCORE			
Delivery	82 (80–86)	83 (78–86)	86 (81–87)
Development	74 (72–74)	75 (71–76)	75 (68–79)
Discovery	61 (59–64)	62 (60–62)	63 (58–65)
AGREEMENT			
Delivery	67 (65–72)	68 (64–73)	70 (65–75)
Development	57 (55–58)	58 (56–60)	55 (54–62)
Discovery	43 (42–49)	45 (42–47)	44 (39–49)
ANSWERABLE?			
Delivery	92 (87–94)	92 (88–95)	91 (90–94)
Development	84 (82–89)	87 (81–90)	84 (78–89)
Discovery	76 (73–78)	76 (74–79)	76 (70–79)
EFFICACY?			
Delivery	87 (84–91)	87 (83–91)	88 (84–90)
Development	81 (77–83)	84 (79–84)	78 (76–81)
Discovery	68 (64–70)	68 (65–72)	69 (59–72)
DELIVERABILITY?			
Delivery	85 (82–89)	86 (82–91)	87 (82–89)
Development	77 (75–80)	79 (77–81)	74 (70–84)
Discovery	68 (66–72)	69 (64–72)	70 (64–72)
IMPACT?			
Delivery	68 (62–72)	65 (58–70)	73 (69–80)
Development	56 (53–57)	53 (52–58)	62 (52–65)
Discovery	46 (39–50)	46 (38–48)	44 (36–54)
EQUITY?			
Delivery	84 (81–88)	84 (76–89)	87 (79–88)
Development	74 (66–77)	71 (65–76)	76 (75–80)
Discovery	54 (50–59)	52 (50–58)	53 (50–65)

Figure 13. Background of programme managers based on the web-search of non-scorers

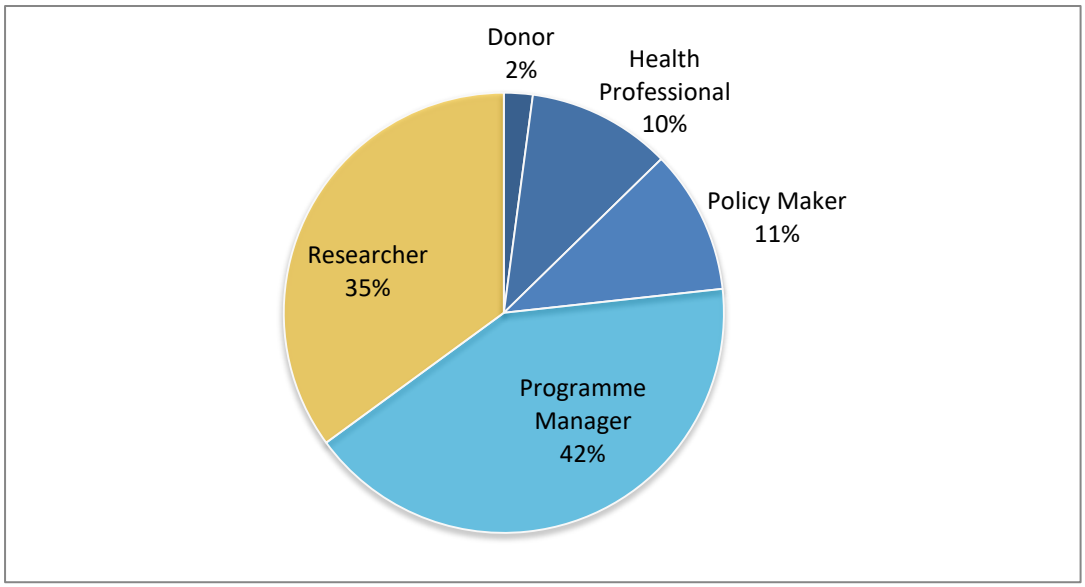
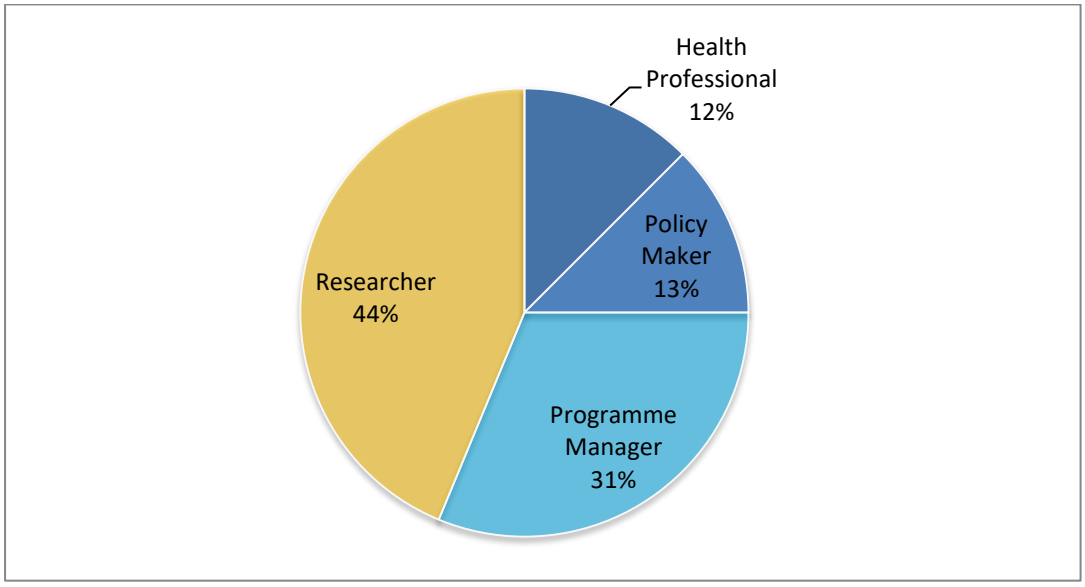


Figure 14 Self-categorization of "programme managers"



iv. Were stakeholders involved in the process?

Stakeholders are distinguished from participants since their role is to provide CHNRI exercises with broader societal perspectives and values, instead of generating or scoring research ideas.

Stakeholders were not involved in this exercise. Profile of stakeholders may include experts such as researchers and program experts depending on the RP exercises. Stakeholders and experts both have a stake in the result of exercise. However, stakeholders are distinguished from these experts by their role. Stakeholders' role is mainly to provide CHNRI exercises with broader societal perspectives and values while experts' role is to generate or scoring research ideas. Nevertheless, the involvement of such reference groups was discussed but. The management team assumed that approaching nearly 600 experts, identified in a systemic way, would lead to inclusiveness and diversity of participants, and global representation of priorities. In addition, the urgency to propose post-MDG global research priorities well before the end of 2015 drove the decision not to include them. From a methodological point of view the inclusion of stakeholders would have added strength to this RP exercise.

v. What was their profile?

Stakeholders, other than technical experts, were not involved in this exercise.

Information gathering

i. What sources and types of evidence/information/ data required were identified?

Global estimates on rates, causes and timing of neonatal deaths and stillbirth were reviewed by the management team.[36,37] Evidence on the causes of severe disability associated with neonatal and congenital conditions was reviewed in the light of child development.[38] The management team identified five major causes leading to severe disability: preterm birth, intrapartum related events, congenital conditions, neonatal infections and neonatal jaundice. These categories were used to classify the research ideas at the time of consolidation.

ii. Did the process consider existing priorities?
The CHNRI exercise on newborn health and birth outcome considered the previously conducted RP exercises on five major causes of child death. Previous priorities were reviewed and were taken into consideration in defining the scope and vision of the RP exercise on newborn health and birth outcomes. For example, the newborn RP focused on child development with longer time-span set at 10 years post MDGs as the previous RPs focused on child survival and research that are short-term within the MDG4 time frame.

iii. How was the decision on the level of the RP made?
At the time of conceptualization, there were no global research agendas that were identified as priorities in the global community beyond 2015. The MCA/WHO identified the need shape global research agenda for newborn health; therefore, the RP exercise was conducted at global level.

Planning for implementation

Is there a pre-defined plan to translate research priorities into proposals?
At the time of designing the exercise, there was no follow-up plan documented for the translation of research priorities into actual conduct of research. However, two members of the management team were donors and are very strong advocates for newborn health, including prevention of stillbirths. Having them in the management team was expected to promote linkage with donors, as well as ownership by donors, which would eventually lead to calls for proposals. Face to face discussions on how to promote the top 10 priority areas were held on various occasions.

Criteria

i. How were the criteria decided and on what basis?

We used a consensus-based approach to select the criteria. Firstly, the core management team reviewed the definitions and sub-criteria for all criteria used in past CHNRI exercises. Secondly, the team examined these criteria based on the context and focus of the exercise and selected

five standard CHNRI criteria. In doing so, there were a few intensive discussions on whether the guidance for each criterion was clear enough for the participants. However, there is neither a meeting report nor a note for the record available that indicated the number of revisions made.

Finally, we asked the 14 members of the working group meeting to provide preliminary scores using the five criteria. Comments from the working group provided feedback to develop guidance material for each criterion.

It is worth noting the feedback from wider participants that answerability and deliverability criteria were considered as interlinked and were not mutually exclusive in judging research options by the wider expert group. Future CHNRI exercises should consider this feedback, to improve the clarity of the guidance material.

Methods for deciding on priorities

How was the method chosen?

The CHNRI method was chosen for all research prioritisation exercises coordinated by MCA/WHO firstly because of the strengths of the method and its ability to systematically solicit ideas from a large pool of participants. Secondly, three management team members had led the previous CHNRI RP exercises, thus they were familiar with the entire process. Thirdly, the management team felt that it was important to keep the consistency of method of all RP exercises coordinated by the MCA/WHO.

Use of comprehensive approach

i. Were other approaches considered?
There was no consideration given to other method for the reasons specified in the response above.

ii. Was the process of consolidating, refining the research questions documented?
Approximately 400 research ideas were submitted and reviewed. These research questions were trimmed down to 205 research questions in the review process. During the process, duplicate questions were excluded, similar questions were merged and retained, and questions were clarified if they were not clear. The entire process was documented with track changes. We considered it important to document the process, given the possibility that some decisions could have been influenced by the perceptions of the management members involved in the review process. Originally, the submitted research ideas were used as reference material in case agreement on rephrasing or merging the research questions was not achieved by the core the management members. In such cases, we referred the matter to another member of

management team for a second opinion. Once the review process was completed for all 205 research questions, we convened a small working group meeting in which 14 participants were invited to review scoring criteria and the list of revised research questions.

Transparency

i. Was the process of RP documented?
Each step of the RP process was documented and published. Participants who provided scores were asked if they would like to be part of the group authorship. Those who consented were acknowledged in the commentary published and were part of the group authorship in the second publication.

ii. Was information material sent to the participants involved in the scoring process?
There were two stages at which guidance materials were sent to participants: the generation of research ideas and the scoring of research options. The guidance material for the former stage included the description on the context, relevant epidemiological data, and guidance to help formulate research ideas. The guidance material for the scoring exercise included descriptions on each criterion and an explanation of how to proceed with scoring.

Evaluation

i. Was the evaluation of the RP process conducted with participants directly involved in the process?

Informal feedback received from participants and management team was used to evaluate the process and to discuss potential solutions to problems. However, it was a passive exercise rather than an active exercise in that it was not conducted systematically and so we did not reach out to all the participants to ask for input.

ii. Was there an increased awareness of priority setting?

The awareness survey was administered to authors of the articles published in the areas of 4 previously conducted CHNRI research priority exercises. To identify the authors, a literature search using MeSH terms (**Appendix 1**) was performed and a total of 131 papers were identified; these had been published between 1 January 2010 and 31 December 2015 (**Table 11**). Of the 131 papers, 33 were related to the research priorities on childhood diarrhoea, 54 to childhood pneumonia, 25 to intrapartum related neonatal deaths and 19 to neonatal infection (**Appendix 2**). Eleven publications had seven co-authors who were participants in at least one of the previous CHNRI exercises. Five of them were either a founding parent or an adoptive parent of the method, the people who initiated four previous CHNRI exercises. We assumed that

they were aware of the exercises and may have been influenced by the RP exercises to some extent, and thus they were excluded from the analysis. Of the remaining 120 papers, 108 (82%) authors did not mention the CHNRI exercises in the article, the remaining 12 (18%) were either or were written in a language other than English or access to the full article was not granted. In these 12 articles it was not sure whether the CHNRI exercise was mentioned in the article. Therefore, the authors of these 120 papers were contacted by email and asked whether the CHNRI exercise made any influence on their work.

Of the 120 first authors contacted, 57 (48%) responded to the survey. Nine researchers (16% of those who responded) were aware of the method and responded that the outcome of the research prioritisation exercise had had some influence on the choice of the research theme. Another five people (9%) mentioned that they were aware of it but stated that it had not had any influence. Three quarters of those who responded were not aware of the CHNRI exercises at the conception of the research studies.

Of the 14 researchers who mentioned being aware of the previously conducted CHNRI exercises, about half of their work was funded through grants from the Bill & Melinda Gates Foundation (BMGF) or the Rockefeller Foundation. Others were funded by multilateral agencies and government agencies. BMGF also provided grant to the CHNRI exercise on preterm birth.

Table 11. Numbers of papers relating to research priorities identified 4 CHNRI exercises

CHNRI Exercise Topic (Year)	RP	Number of Post-CHRNI Publications	Dates
Childhood Diarrhoea (2009)	1: What is the acceptability and effectiveness of the new reduced osmolality ORS in clinic, as well as in the community?	2	01/01/2010-31/12/2014
	2: What is the effectiveness of zinc supplementation on the outcome and incidence of diarrhoea in the community?	9	01/01/2010-31/12/2014
	3: What are the barriers against appropriate use of ORT?	6	01/01/2010-31/12/2014
	4: Design locally adapted training programmes to orient health workers on IMCI.	12	01/01/2010-31/12/2014
	5: What is the impact of IMCI in different population groups on timely identification and treatment of acute diarrhoea?	4	01/01/2010-31/12/2014
Childhood Pneumonia (2011)	1: Study the main barriers to healthcare seeking and healthcare access for children with pneumonia in different contexts and settings in developing countries.	10	01/01/2012-31/12/2015
	2: Identify the key risk factors predisposing to the development of severe pneumonia and identify children who require hospitalisation.	28	01/01/2012-31/12/2015
	3: Study the main barriers to increasing coverage by available vaccines - Hib vaccine and pneumococcal vaccine - in different contexts and settings.	4	01/01/2012-31/12/2015
	4: Study whether the coverage by antibiotic treatment can be greatly expanded in safe and effective ways if it was administered by community health workers.	5	01/01/2012-31/12/2015
	5: Study the main barriers to increasing demand for/compliance with vaccination with available vaccines in	7	01/01/2012-31/12/2015

	different contexts and settings - for measles and pertussis vaccines, Hib vaccine, and pneumococcal vaccine.		
Intrapartum Related Neonatal Death (2011)	1, 3, and 4: Can community cadres of workers identify a limited number of high-risk conditions/danger signs (e.g. multiple pregnancy, breech, short maternal stature, etc.) and successfully refer women for facility birth? What is the predicative value and cost effectiveness? // Behavioural/community participation package to improve recognition and acting for simplified danger signs for mother in labour, including transport and phone/radio communication ("emergency preparedness")? // Effectiveness of community cadre roles, e.g. social support, bringing to facility when woman is in labour, danger recognition/referral?	5	01/01/2012-31/12/2015
	2: What strategies are effective in increasing demand for, and use of, skilled attendance (e.g. conditional cash transfers)?	17	01/01/2012-31/12/2015
	5: Does regular use of perinatal audit reduce the incidence of adverse outcomes related to acute intrapartum events?	3	01/01/2012-31/12/2015
Neonatal Infection (2009)	1: What are the feasibility, effectiveness and cost of different approaches to promote the following home care practices: early initiation and exclusivity of breastfeeding, hygienic cord and skin care, prompt care seeking for illness from an appropriate provider?	1	01/01/2010-31/12/2014
	2: What is the role of local application of disinfectants in the prevention of umbilical infections and sepsis?	6	01/01/2010-31/12/2014
	3: What are the feasibility, effectiveness and cost of approaches to increase coverage of clean delivery practices in facilities and in homes?	5	01/01/2010-31/12/2014
	4: What are the feasibility, cost and effectiveness of setting up newborn care corners in first referral units and district hospitals?	3	01/01/2010-31/12/2014
	5: What is the feasibility effectiveness and cost of a scheme of routine home visits for initiation of supportive practises, detection of illness and newborn survival?	4	01/01/2010-31/12/2014

iii. Were the results of the RP exercise accepted by users?

First, there was no purpose-made platform to which participants could appeal or complain about the result of the RP exercise.

As an indirect measure to assess acceptance of the result of the previously conducted RP exercises, analysis on the citing articles of the past RP exercises was conducted to assess the acceptance of the priorities identified in the past five CHNRI exercises. In total, there were 248 citing articles. When institutions affiliated to the CHNRI method was excluded, 188 citations were found (**Table 12**). The top 10 institutions to which authors of the citing articles were affiliated are presented in **Table 13**. The University of Edinburgh, University College London and WHO were the top 3 institutions, probably because they are part of the global community who use and promote the CHNRI method. Institutions in LMICs such as Pakistan, Bangladesh and India accounted for 18% of the citing articles. The distribution of the top 10 funding agencies from which citing papers received funding is presented in **Figure 15**. When multiple sources of funding are reported, all funding sources are considered. About 30% of citing papers

received funding from the Bill and Melinda Gates Foundation, and 17% from WHO; the rest were mostly from UK-based agencies or foundations. The geographical distribution of the top 10 countries indicated that more than half were based in the UK and USA. This is similar to the distribution of institutions that authors of the citing articles are affiliated with (**Figure 16**).

Table 12. **Numbers of citations of four CHRNI exercises**

Title	**Author**	**Year of Publication**	**Number of Citations**	**Number of Citations (CHNRI affiliates removed)**
Setting Research Priorities to Reduce Global Mortality from Childhood Diarrhoea by 2015.	Fontaine et al.	2009	52	30
Research Priorities to Reduce Global Mortality from Newborn Infections by 2015.	Bahl et al.	2009	50	36
Setting Research Priorities to Reduce Global Mortality from Childhood Pneumonia by 2015.	Rudan et al.	2011	54	40
Setting Research Priorities to Reduce Almost One Million Deaths from Birth Asphyxia by 2015.	Lawn et al.	2011	46	41
Setting Research Priorities to Reduce Global Mortality from Preterm Birth and Low Birth Weight by 2015.	Bahl et al.	2012	46	41
TOTAL			248	188

Table 13. Institutions to which the authors of citing articles were affiliated

Institution	**Count**
University of Edinburgh, UK.	70
University College London, UK.	48
World Health Organization, Switzerland.	48
Johns Hopkins University, USA.	39
London School of Hygiene Tropical Medicine, UK.	39
Johns Hopkins Bloomberg School of Public Health, USA.	36
University of Split, Croatia.	25
Aga Khan University, Pakistan.	22
International Centre for Diarrhoeal Disease Research, Bangladesh.	22
Public Health Foundation of India, India.	16

Figure 15. Funding agency from which publications of citing articles received funding

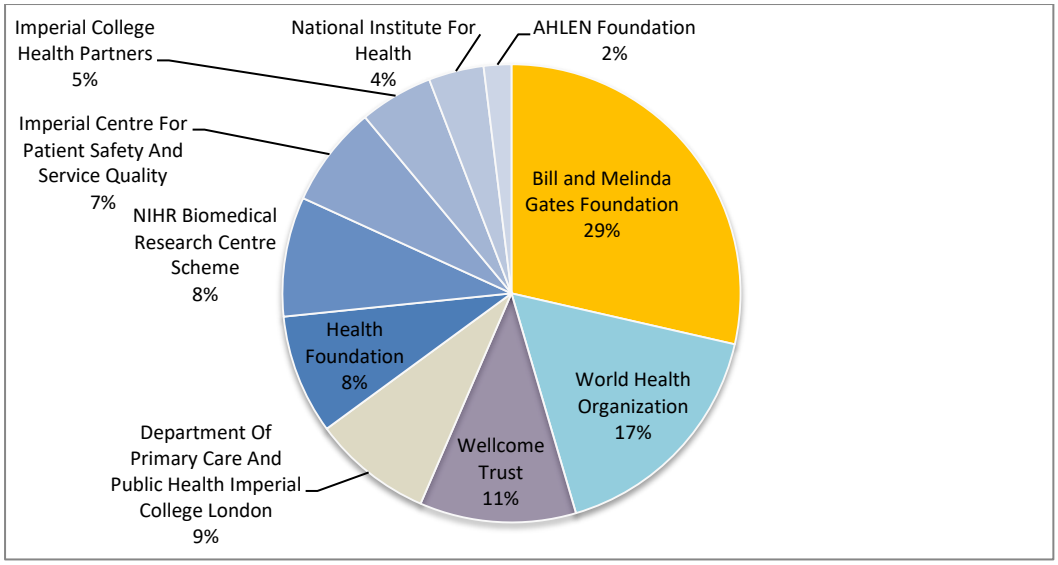
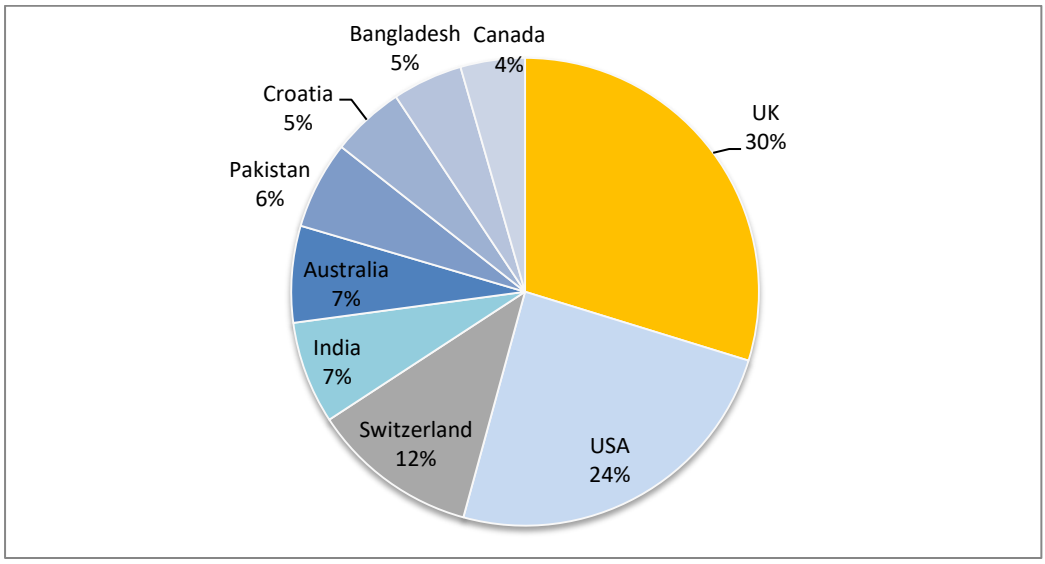


Figure 16 Country where authors of citing articles were based



iv. Degree of alignment of resources allocation and agreed upon priorities

Six priorities identified in the RP exercise on newborn health and birth outcome were funded by major funding agencies through a grant to WHO: scale-up research to identify newborns with signs of infection by community health workers, community-initiated KMC, simplified antibiotics regimens for neonatal sepsis, immediate initiation of KMC,[39] KMC implementation research for accelerating scale-up,[40] and a study to improve the quality of the services for pregnant women and newborns.[41] For KMC-related research priorities, BMGF convened two meetings to discuss research priorities in this area, and a proposal writing workshop was organized and donors, including BMGF, were invited. The research studies are all being implemented by national health research institutes or universities. Overall coordination is

provided by MCA/WHO. More than 20 million USD were allocated to the implementation of the six studies.

v. Did the RP exercise lead to building capacity of the health research system in the country?

In MCA/WHO experience, research priority setting and fund raising are managed by the same organization. In this operational model, we aim to build capacity of researchers in LMICs through technical guidance on the implementation of the study and the analysis of data. There are eighteen research institutions and universities in ten countries involved in the studies mentioned in the previous section. They are all invited to meetings to share knowledge and challenges. Once the study is completed, we conduct data analysis and manuscript writing workshop to which not only the Principal Investigators (PIs) but also the co-PIs and data managers are invited to analyse their own data.

Discussion

Reflections on applying the framework

This experience in applying the framework suggested that the framework is useful to examine key components of the RP exercises. Some elements were incorporated into the framework based on personal experience such as information sharing with participants since limited accessibility to up-to-date evidence is an issue in many LMICs.[42,43] National or sub-national policy makers may not be as informed as the global health community about the available evidence on which to base their decisions. These bottlenecks pose questions about the validity of the outcome of research prioritisation in countries.

Assessment of self-selection bias was also added to examine the possibility of response bias introduced through self-selection which may lead to limited generalizability of priorities to a wider context.

Applying the framework to the previously conducted CHNRI exercises revealed that "flexibility" was an underlying element. For the "inclusiveness" theme, flexibility in the timing to approach stakeholders was a key. To illustrate, the decision on whether to include stakeholders is usually made before the initiation of the exercise. However, it would be more helpful to revisit the decision based on whether there is a balanced representation of participants. Based on personal experience in the previous CHRNI exercises I believe a good balance of representation would be as follows: researchers (of which some may also be health professionals) and programme experts (≈35% each), donors (20%), and policy makers (10%). In our experience, we identified that experts working in the programmes were underrepresented when we

received research ideas. At this point, it would have been helpful to reconsider the involvement of the stakeholders including that of programme experts in determining weights for the different criteria. In general, reflecting "voices" from diversified participants would reduce potential response bias relating to experts preferring their particular research field.[13]

Resource allocation to the priority areas indicated the impact of the RP exercise with implementation despite not having a pre-defined plan. It would have been helpful to document how some of the top 10 priorities were successfully translated into proposals that were eventually funded through a grant to WHO. This would provide some guidance in visualising the steps towards bridging priorities and implementation. Moreover, the documentation of such processes will help to develop an *a priori* action plan for future research prioritisation exercises.

The awareness survey conducted among specific group of experts indicated that one in four (16% who are aware of the exercise and admitted that the exercise had had some impact plus 9% who are aware of the exercise and admitted that exercise had no impact plus 9%) was aware of the RP exercises. Of them, 16% reported that the exercise had had some influence in their work. The impact of RP exercises can also include other factors such as knowledge sharing, meetings with stakeholders and donors, and convening prioritisation exercises that do not necessarily result in publication. The low response rate was one of the limitations of the exercise. In addition, it would have been useful to conduct a more comprehensive survey or an interview to get to know the details of influence for future exercises. I only contacted a targeted group of researchers, but in future surveys it would be helpful to contact a much wider audience of the CHNRI exercises including stakeholders. A low response rate (48%) was also a limitation of this survey.

Limitations

The application of the framework to evaluate the CHNRI exercises has some limitations. Firstly, I developed my own evaluation questions and my own objectively verifiable indicators based on various sources, but none of these tools were tested as an evaluation tool of RP exercises.

Secondly, the comprehensive evaluation could not be conducted on one CHNRI exercise since there are two evaluations of the RP exercises. The first evaluation is an immediate review of the process to highlight issues and to propose solutions immediately after the exercise. The second evaluation is a review of impact that can be conducted after five years of the RP exercise since timing to assess any influence on authors of publication related to top research priorities would require a 5-year period. Evaluation of process was conducted on newborn health and birth

outcomes published in 2016 while evaluation of impact was conducted on the four CHNRI exercises conducted between 2009 and 2011. It was not possible to conduct a comprehensive evaluation of the earlier CHNRI exercises because I did not have access to reports and materials to be able to assess the quality of the process.

Finally, for the theme on "evaluation", a different way in categorising the professional role between scorers and non-scorers limits the comparison of results by professional category. I used analyses of citing articles of past CHNRI exercise, and the awareness survey, to obtain information about the awareness and the influence of past PR exercises. These provided some information but did not provide a comprehensive view on what impact the past CHNRI exercises had. Analysis of citing articles does not to assess how many articles disagreed with the results of the RP exercises, therefore presumably do not cite the RP exercises. Furthermore, it is not sufficient to measure the impact of research prioritisation exercises by using scientific articles as a proxy outcome of success. The number of publications does not necessarily indicate the quality of research conducted. For example, one large-scale study with sound methodology in a given area could provide definitive evidence meaning that no further studies are needed on the same topic. Though the ultimate success of RP exercises should be measured through success in resources mobilised for priority research areas, there is also a need to consider some process indicators to measure the impact. Examples of such indicators include the number of research items incorporated in the national health strategies and plans, the number of meetings conducted to discuss the implementation of priorities, and the number of proposal-writing workshops conducted.

Recommendations

To my knowledge, this is the first attempt to apply a framework to guide the evaluation of the CHNRI exercises. The application of the evaluation framework identified four recommendations to improve components of the CHNRI priority setting process:

1. *A priori* decision on the involvement of stakeholders to provide weights or thresholds should be revisited based on the representation of participants.
2. There should be a pre-defined plan on how to take forward the results of the RP exercise
3. Systematic evaluation of the CHNRI exercises with respect to both process and impact should be an integral part of the process rather than an option. We recommend that a review of the process is conducted by someone involved in the organisation of the RP exercises since most information required to evaluate the process is only available internally. A review of the RP process should be considered as an integral part of the RP exercise rather than a separate exercise. On the other hand, evaluation on impact of the

RP exercise could be conducted internally or externally provided that all materials are available publicly.

Conclusion

While many reports identified the importance of linking research with action, I could not find any reports about whether the identified research priorities were funded after the prioritisation. Many RP exercises end once the RPs have been identified, with no subsequent follow-up to assess the extent to which the RP exercise was effective in mobilizing funds for identified priorities. Despite the absence of written plan to promote the research priorities, it was always the intention of the instigator of the process to consider prioritisation process as a means to promote and support the priority studies rather than the end in the newborn health RP exercise. Moreover, the instigator of the process worked closely with donors to facilitate the transition from questions to implementation. Publication in the high-impact journals might have helped to increase the credibility of results.

The process of evaluation confirmed that CHNRI exercises that were assessed met most of the requirements to be qualified as good practices. This was the first attempt to evaluate the past CHNRI exercises, which demonstrated an example of how the assessment of the CHRNI method can be conducted at each step. In the absence of an existing practical evaluation framework, and having tested this new framework, I believe that the impact and quality of the CHNRI exercise should be assessed based on structural guidance and, therefore, I believe that the modified framework can be further used to assess future RP exercises including CHNRI exercises.

In future it will be important to address how each parameter and theme, which has yes/no answers or semi-quantitative response, can be combined to evaluate the overall performance of RP exercises. In other words, on what basis can we make a difference between “good” exercise and a very “good” exercise? I am aware that this is a remaining challenge and I plan to address this in the future.

Chapter 5. Overall discussion

Review of tools and approaches used in health research prioritisation

The first aim of this thesis has been to fully understand the landscape of all approaches, tools and methods used in research prioritisation exercises in the 21st century. To achieve this, I reviewed 165 health prioritisation exercises. The review scrutinised six major approaches with respect to the overall process, how participants and ideas are identified, how scoring is done, and highlighted weaknesses and comparative advantages of the different approaches. The review clearly showed that CHNRI is the most frequently used method to date in the 21st century.[34] Since its publication in 2016, this review has been used as an introduction to various health research prioritisation tools and has already been cited by twenty-two papers. This is an early indication of the usefulness of the review in providing a critical overview of health research prioritisation methods. A recently published review of 50 CHNRI exercises[5] helped our understanding of the evolution of the method, in terms of areas of application, criteria modification in number and actual criteria, the use of context. The review of funders, researchers and stakeholders' involvement helped us to elicit challenges that were dealt with in a satisfactory way, but which still need to be improved in various ways. The review also showed us how this could be achieved.

The consultation held by the World Health Organization's Department for Research Policy and Cooperation in 2010, and the independent review conducted by Skye et al, identified two major shortcomings in the CHNRI method. Both are related to the end product of the prioritisation exercises. [13] They argued that identified research priorities should be broad enough to fit with the interests of donors but also sufficiently specific to be translated into a study proposal. Finding the proper balance to fit with both perspectives is challenging. Many users of the CHNRI method have faced the dilemma of how specific individual research priorities should be. Research ideas submitted by participants vary in their specificity and clarity. Some such as PICO (population, intervention, control, outcome) clearly include the conventional characteristics of the research idea, while others merely specify a broad area of interest, e.g. maternal depression, quality of care. The latter example was seen in many CHNRI exercises and has been the source of headaches for the management group when it comes to consolidation. On the other hand, we observed that very specific research questions were likely to be scored less highly than similar but less specific research questions.[26] In this case, very specific research questions only catch the attention of the few people with the same interest. The best guidance in forming the research question is similar to forming the research hypothesis. "A hypothesis can be defined as a tentative explanation of the research problem, a possible outcome of the research, or an

educated guess about the research outcome."[44] and is neither too specific nor too general. The research ideas should be declarative and contain a clear statement of what is intended to be studied based on the current knowledge. Listed below is an example of how a broad research question can be formulated to a specific research question:

Broad research question: evaluate delivery strategies to reach the poor and marginalized.

The above research question is not specific in that it does not mention what intervention to be delivered is and who are the poor and marginalized population. Hence a more specific research question would be:

Specific research question: evaluate home visits to deliver postnatal care as a strategy to reach the poor and marginalised mothers and newborns.

Similarly, research questions that are too specific can be formulated to be less specific. Listed below is an example:

Too specific question: what are the effects of intervention programs in the elementary schools on the rate of childhood obesity among 3rd - 6th grade students in Eldoret, Kenya?

The above research question has very specific geographical and population focus: generalizability of the research question is limited. Hence less specific research question would be:

Less specific research question: evaluate effects of school-based intervention programs to reduce childhood obesity in low and middle-income countries?

In past CHNRI exercises, I have created guidance materials to help the participants to formulate the research ideas. To my knowledge, such supplementary material is not currently available online and thus it will be important to revise the current supporting materials to be more user-friendly and more easily available. Such material will be instrumental in saving much time in revising the research questions as this is the most time-consuming component of the CHNRI process.

Most CHNRI research prioritisation exercises ended up as a one-off exercise, meaning that most RP exercises do not go beyond the identification of priorities.[13,30] Viergever et al advocate having in their nine checklist items a pre-determined strategic plan for the translation of priorities into actual research.[31] Many research prioritisation exercises, however, do not have such a pre-defined strategy for implementation and this is true of the CHNRI process. Mostly research priorities are published and disseminated in conferences and there is little, if any,

follow up and action. In addition to the lack of focus on how to take research priorities forward there has been very little assessment of the quality and impact of RP exercises. The impact of the RP should be assessed by the success in the allocation of funds to priority areas. Without an assessment of the process and impact, it is hard to demonstrate the success of RP exercises. In retrospect, the absence of a useful evaluation framework to date might have also contributed to this cause. In the same Chapter, I presented a new framework and applied the framework to assess previously conducted CHNRI exercises in detail. This was a very helpful exercise, not only to reflect on what could have been done better but also to reflect on the ultimate goal that RP exercises should try to achieve at the global, national and sub-national level.

Application of the CHNRI method and lessons learned

The second aim was to conduct a research prioritisation using the CHNRI method to gain a better understanding of the process and use of the method, and to generate the database for my PhD analysis. I coordinated the global RP exercise in an area of newborn health and birth outcomes using the CHNRI method, [20,34]which generated the dataset that was studied for further research questions on the CHNRI methodology. In this RP exercise, three broad areas were identified as priorities: care of preterm and low birth weight babies; prevention and management of infections; and prevention of asphyxia and intrapartum stillbirth.

Similarly, I provided methodological expertise to another exercise in the area of midwifery in which the collective opinions of 270 participants were compiled in setting the global research agenda for the midwifery model of care.[8] This was the first attempt in the history of midwifery to set a global research agenda. In this exercise, a modified version of the CHNRI methodology was applied to identify research gaps. Normally, in the CHNRI method the identified participants are asked to provide a maximum of three research ideas; however, in this exercise, instead of asking the participants to generate the research questions, a systematic review was conducted by experts and the refined result of this systematic review was sent to participants for scoring. This exercise was published in *The Lancet* midwifery series in 2016.

The results of both exercises, published in the high-impact journals, reflect the potential acceptance of the prioritisation exercise, as well as the potential for further discussion and advancement of the research among the scientific community.

Limitations and recommendations

I would like to share a few lessons learned and corresponding recommendations for future RP exercises:

I noted that in RP exercise on quality of care; all questions received relatively similar scores [Mean (SD) 85 (4.1); Median (IQR) 86.0 (83.4-88.8)]. Similarity of scores in all research questions could be affected by three possibilities: research options are not distinctive enough; the criteria were perceived to be similar; a combination of both; the research questions were equally important. Although similarity of scores in some research questions is expected, all research questions receiving similar scores, as in this exercise, is indicative of some sort of a problem in the method. On the other hand, some questions receiving similar scores is not indicative of a problem. In principle, research options or scoring criteria or combination of both should be mutually exclusive enough for the respondent to provide an accurate judgement in the RP exercise. Informal feedback from participants included difficulties in scoring because of the similarity in the definition of criteria. They reported that answerability and deliverability criteria were interlinked, and they were less distinctive in judging research options. Future CHNRI exercises should improve the clarity of the guiding material on both the criteria and on research questions.

We noted that the process to compile and combine research questions may introduce another risk of bias. The way questions are phrased, or how broadly or narrowly they are framed, could possibly influence the judgement of the respondent. In both exercises, compiling and combining the research questions, either from a list of ideas generated by the respondents or by evidence review, required intensive discussion as to how best to clarify the original research questions and knowledge gap without deviating from the original questions. We found that two approaches help in this process. Firstly, we found that it is important to document the process in 'track changes' so that any deviation from source is trackable and original and final formulation of research question is compared once review of all research questions are completed Secondly, a pilot test for a scoring exercise by smaller group of experts other than the management team can provide an objective view.

Future RP exercises need to consider using different criteria to score discovery research such as biomedical research or, alternatively, use the same criteria but present the results separately by type of research for sake of fairness.

In the two exercises in which I was involved, implementation research (delivery research) was likely to be scored higher than the development or discovery-type research because of a higher certainty in the outcome and the fact that the outcome is expected to be achieved within a shorter time frame. This raises the possibility that someone who is aware of this might classify his or her preferred research options as development or discovery research with intention to increase the chance of having them listed as top priorities in these research domains.

Though the definitions of delivery and development research are clearly stated [10], it is not always clear when it comes to application of these definitions to some research options. For example, research options on community-based initiation of KMC is categorised as development research because community-based KMC is considered as a new intervention, while research to scale up KMC that is based in health facility is regarded as delivery research. Further improvements need to be made on how to apply the research domain definitions in the classification research options in the future. However, in the research prioritisation exercises I was involved, research options were discussed and reviewed multiple times by a smaller group of meeting participants as well as by the management team independently. Two layers of review process should be able to prevent intentional misclassification.

In the same research prioritisation exercise, we did not ask the participants to generate epidemiological research ideas. This is because the management team felt that major priorities in descriptive research on maternal and newborn health were already identified through a set of previously conducted global research prioritisation exercises in child health, conducted between 2009 and 2012. This exercise rather focused on medium-term and long-term research options on newborn growth and development through better delivery of existing intervention or through new or modified interventions. However, it may be true that there may be emerging epidemiological research priorities which might have been missed in the previous CHNRI exercises therefore this could be one of the limitations of the exercise.

Common successes

Both exercises achieved more than just the identification of research priorities. The RP exercise on newborn health and birth outcomes led to resources being allocated to six out of top 10 research priorities. Among these, one of the priorities that ranked among the top 10 related to formative and implementation research on Kangaroo Mother Care (KMC). The MCA/WHO has been coordinating three projects on KMC: community initiated KMC in India; scale-up of facility based KMC in three states in India and Ethiopia; and a project to evaluate the efficacy and safety of KMC initiated immediately after birth for babies weighing between 1.0kg to 1.8kg compared to providing intermitted KMC (routine care in the tertiary hospital) in Ghana, India, Malawi, Nigeria and Tanzania. Moreover, implementation research to scale up outpatient treatment of possible serious bacterial infection when referral is not possible has been conducted in two countries.

Why has there been so much success in KMC-related research areas? The key was to create interest among the donors with respect to the research priorities. Our research prioritisation exercise caught the attention of BMGF, which has been supportive of interventions provided by

community health workers and KMC. Following the publication of the RP exercise on newborn health, BMGF organized a meeting with researchers and WHO to discuss how to take the KMC research agenda forward into actual research. Eventually, proposal writing workshops were organized by WHO, and interested donors were invited. The RP was published in *The Lancet* and the *Journal of Global Health*[20,34] and cited by 49 articles, thus indicating a high level of interest in this area.

The RP exercise on the improvement of quality through midwifery care led to several substantive outcomes, including establishment of a research alliance on midwifery, and a few international meetings to bring donors and stakeholders together to explore the funding of priority areas. However, funding has not been allocated to priorities areas yet.

Two key ingredients in the success of RPs are identified. Firstly, having at least one key person in the core management team who is well connected to the donors is very important, to link demand and the needs of the stakeholders and donors on the research priorities. Secondly, many global-level research prioritisation exercises in the late '90s seem to have predominantly targeted researchers and there was less involvement of programme experts (personal communication). In our exercise, we attempted to approach as many programme experts as possible. Programme experts could help shape the questions, point to their priorities from the programmatic angle, endorse the need for answering specific questions and support the implementation of studies by facilitating access to sites and covering variable portions of the costs of intervention implementation. From my first CHNRI exercise, I believe that the way the programme experts were identified could have been much improved. Thus, in the subsequent RP exercise, I used a more comprehensive approach to list programme experts in our area of interest for future RP exercises, which has resulted in better coverage of programme experts.

Human collective knowledge

Having reviewed and applied the CHNRI methodology, I then explored some of underlying assumptions of the CHNRI methodology. I first explored some properties of human collective knowledge. The aim of the experiments was to investigate the circumstances under which a collective knowledge is better than individual knowledge. This exploration was performed through a series of experiments in a group of about 160 (range 122 to 175) undergraduate Year 2 medical students. The experiments suggested that collective knowledge outperforms individual knowledge. [45] Although not a surprise, this exercise did confirm that the collective knowledge of experts in an area of their expertise is likely to produce more accurate responses than the collective knowledge in an area outside of their expertise. It might be argued that an

individual expert may perform as well as a group of experts where there is generally agreed certainty about things such as "facts" rather than "opinion". Where collective wisdom is particularly useful is when there is uncertainty either due to a lack of "expertise" or due to inherent uncertainty in the subject area.

Independent ranking is the distinctive feature of the CHNRI, which based on the notion behind the "wisdom of crowd" method. Although the students' exercise was conducted in a rather controlled environment, in that each student answered the questions independently without any discussion or consultation, there may be factors other than the collective knowledge, opinion or intelligence that influenced the responses, possibly including group dynamics in the classroom. This is, therefore, is a limitation of this type of exercise.

Implication of the findings from this experiment

This experiment supports two underlying assumptions of the CHNRI method. Firstly, it showed that collective wisdom exceeds individual performance, and second it demonstrated that experts' collective wisdom outperforms non-expert's collective wisdom. In addition to confirming the two baseline concepts of the method, it also showed that when the experts did not answer the questions in areas where they did not have any technical knowledge, this seems to have improved the performance of the collective prediction. Interestingly, having the "not sure" option did not make much difference over just having either "Yes" or "No" answer options. The above two findings are useful in the application of the CHNRI methodology. It implies that for future CHNRI exercises there could be three options: "Yes", "No" and "Blank" and no "Not sure" in scoring since having "Not sure" did not make any difference in the collective expert performance.

Human collective opinion

Following the exploration of human collective knowledge, we conducted an exploration of some quantitative aspects of human collective opinion. Studying quantitative properties of human collective opinion is challenging given that that collective opinions are non-verifiable and, therefore, no gold standard exists. In the absence of any means of validating human collective opinion, we examined the reproducibility of the human collective opinion.

The CHNRI method uses purposively selected samples (e.g. experts in a certain domain of research) as opposed to probability-driven samples for quantitative research. Most purposive sample sizes are determined by the concept of "saturation".[46,47] "Saturation" is a guiding principal defined as the point of diminishing returns at which the collection of new data does not shed any further light on the issue under investigation.[48] The definition is subjective in a

way to operationalise in practice. Indeed, more than two decades have passed since this problem was highlighted[46] and yet there are no clear guidelines on how to assess "saturation". In the absence of clear guidance, most purposive sample sizes are *a posteriori* or "rule of thumb" decided upon by those who conduct the research[23,24] and this is the case in the CHNRI exercises.

In this Chapter, I attempted to provide a response to the question most frequently asked since I used CHNRI method for the first time: "How many participants are enough in a CHNRI research prioritisation exercise?" Despite the interest, to the best of my knowledge nobody has studied the appropriate sample size for a CHNRI exercise. To date, there is no clear guidance in the published literature, thus leaving the decision to those who instigate the exercises. The aim of the analysis was to explore the point at which collective opinions tend to stabilise. Stabilization of collective opinions implies that there is a point, at which adding more participants brings about very little change in the ranking of research questions, i.e. the minimum sample size of experts to obtain stable result.

To explore the minimum sample size of the CHNRI exercise, we conducted two statistical analyses on four different databases consisting of scores received in four previously conducted CHNRI exercises. Each dataset includes the scores given by anonymous individuals who participated in CHNRI exercises in health-related domains: newborn health and birth outcomes;[20] maternal and perinatal health;[25] neonatal and maternal health through the contribution of midwifery care;[8] disability and health access.[9] The rank of research priorities is generated on the basis of the total scores for each research question. Unlike qualitative research studies, we could not conduct content analysis since inputs from participants were dichotomous and numerically coded. We used scores given by all participants as a reference point and compared these against the ranking of research questions from bootstrap samples of increasing size. Hence, we defined stabilization as a point at which by adding more participants there was very little change in the ranking of research questions, i.e. *rank stabilization*. We first compared the concordance of the top 20 research priorities between the increasing bootstrapping sample sizes and the ranking by the reference group. In the second analysis, the rank correlation coefficient was calculated for all research questions. In our analysis, concordance of 75% and a rank correlation coefficient reaching 95% were set as arbitrary *a priori* thresholds to define rank stabilization.

We conducted further analysis to examine if the number of research questions has any influence on the point at which rank stabilization occurs. The findings did indicate that less number of questions required less number of experts to achieve stabilization of results expressed by concordance, however there was no consistent pattern of positive correlation between

increasing number of research questions and increasing number of experts to achieve stability. Similarly, there was no correlation between increasing the number of questions and the stability of results regarding rank correlation. What this additional analysis indicated was that sample sizes at which rank stabilization occurred were varied across four CHNRI exercises, and this was not solely due to varying the number of questions. Other factors that are likely to influence the stability point are the total number of research questions in an exercise; the composition and diversity of scorers; the number of criteria used; variations in RPS (i.e., similarity of RPSs in all research questions).

To conclude, at present there is insufficient evidence to provide any recommendation of the optimal number of experts that are likely to be required in future CHNRI exercises. This was the first attempt to explore the optimal sample size of CHNRI exercises. I believe this was an important attempt in this, yet, undeveloped scientific field. This study provided a detailed method and process on which subsequent researchers can expand. We encourage further methodological research using more data on CHNRI exercises, to clarify what factors in CHNRI RP exercises are likely to influence the minimum sample size of the CHNRI exercises.

Involving stakeholders

Who are the stakeholders? In health research, profiles of stakeholders include policy makers at different levels (central government agencies, ministries of health, local government, etc.), health governing boards, national and local health organizations (non-governmental & non-profit organizations), health governing boards, unions, suppliers, international health organizations, representatives of health professionals organizations and healthcare institutions, and representatives of the public.[49,50]

Stakeholders in the CHNRI exercise are defined as a reference group, selected based on their professional background and the types of interest they have, and who are usually involved at one stage of the method. They can be technical experts such as researchers, policy makers or programme managers, as well as a non-technical crowd such as consumer groups, patient groups or care providers. Unlike participants who provide scores on individual research options, stakeholders are involved in providing thresholds or weighting the criteria, during which they represent societal values at large for those who receive the end product of research in the respective CHNRI exercises.

Why is involving stakeholders so important? Reflecting the value of stakeholders in the process is usually recommended to ensure the transparency, legitimacy and fairness of the process.[27] Firstly, stakeholders' involvement in health research prioritisation promotes the successful implementation of research activities after the prioritisation exercises are conducted. Sufficient

evidence is documented in the area of healthcare decision making[51] and health technology assessment.[52] Successfully conducted research is reported when not only are the "voices" of stakeholders reflected, but the stakeholders are also actively involved in the implementation of the research. A study in Afghanistan showed that having well recognised stakeholders in an advisory capacity in the study facilitated the dissemination of information and increased the demand to use maternal health services in a post-conflict area.[50,53] The same study mentioned that the successful factor for the study was to get stakeholder buy-in by convincing them to be a supporter rather than an obstructer within their local context. In research prioritisation, country-led research prioritisation exercises in Malawi, Nigeria and Zimbabwe actively involved patients living with HIV/AIDS in the entire process along with programme managers and health professionals working with HIV patients.[54] Similarly, a multi-level research prioritisation exercise in India conducted at the national and sub-national level provided a successful example of bringing together a group of key actors from within the nation and involving them in the process. They later became an advocate of the research prioritisation exercise in increasing the participation rate at national level. In the exercise, stakeholders consisted of policy makers, politicians, senior researchers, programme managers and funders. Each plays a crucial role in Indian national health research activities.[55]

Secondly, why is involving stakeholders at the national or sub-national level so critical? Stakeholders' involvement in the process ensures legitimacy and fosters the integration of research priorities into the current health system planning cycle and infrastructure in countries.[33,56,57] Most prioritisation exercises have been initiated by researchers from outside the health system planning cycle[33,58] resulting in the generation of priorities parallel to existing government priorities.[33,59] It goes without saying that the research prioritisation exercise should not be instigated at the expense of national health priorities.[60] Therefore, it is even more crucial to have stakeholders, including policy makers, on board at the sub-national level research prioritisation since decisions on the integration of programmes on cross-cutting health areas, resources and strategies to deliver intervention are taken mostly at the sub-national level, while the focus of national-level research prioritisation is usually disease or intervention-specific in a decentralised system.[61] Strategic approaches and plans for engaging well-defined stakeholders promote their ownership of the process, and the integration of priorities into activities provided by national health research systems.

Thirdly, how do global priorities feed into to national priorities? And what are the potential implications of a global level prioritisation exercise which result in globally identified priorities that do not necessarily reflect national challenges with regard to financial alignment and disease burden.[57,62] In most of previously conducted global research prioritisation exercises, priorities

were generated by individuals and institutions rather than from beneficiaries at national level. Usually globally-led exercises pick up universal challenges and concerns at broader level, some but not all of which may be applicable to a specific country's needs. In a few cases, this has led to a nationally led workshop or in a re-prioritisation exercise following the global exercise in which specific health themes or interventions applicable to a country's priorities are solicited. One of the good example is a global research prioritisation exercise coordinated by the International AIDS Society and international partners. They identified 20 priority research questions, among which was a weight-based prescribing range, which was not a national priority in some countries. Following the global prioritisation exercise, nationally-led research prioritisation exercises were conducted in three African countries to take into account national and district needs.[54] Similarly, a globally- led research prioritisation exercise for maternal, newborn, child and adolescent health in humanitarian settings identified 25 priorities (exercise completed 18 April, 2019). Included in the top priorities was provision of organised and inclusive nurturing care for early childhood development through health services during protracted emergencies. Early childhood development is a global priority for child health including for protracted emergencies setting however this topic may not be applicable for acute humanitarian crisis settings. On the other hand, epidemiological research to collect mortality rate and cause of death data, including stillbirth, was not seen as one of the global priorities in this exercise. However, this may be the most critical and applicable first research in acute humanitarian settings since in many of these settings health workers do not collect such information in a systematic way. Though it was a first attempt to set global research priorities on humanitarian setting, diversity of context, distribution of disease burden and nature of population at risk posed great challenges to such an exercise. It requires interpretation through the country's own lens to see the situation within the country since adaptability and applicability differ by country. Participants felt a need to conduct a workshop, including countries of similar context, to further scrutinise the priorities that are suitable to their own context.

What are the remaining challenges to be addressed? CHNRI exercises provide, potentially, an opportunity to reflect stakeholders' views in the outcome. However, stakeholders were only involved in 20% of all previously conducted CHNRI exercises[5] because of various challenges documented earlier.[27,63] Stakeholders' profiles and number of stakeholders that is appropriate to involve are unexplored areas of interest given the limited involvement of stakeholders in the previous CHNRI exercises. Paper 4.1 presented two potential roles of stakeholders: have them determine how to weight the various criteria or have them set thresholds. It also showcased various ways of involving the stakeholders, starting from the identification, number,

responsibility, weights and thresholds applied to the criteria, and the impact of stakeholders' involvement in the final scores. It concluded that, to date, a good definition of stakeholders does not exist and pointed to the need to have a large group of stakeholder representatives. A recent review by Wazny highlighted the potential option of using crowdsourcing as a complementary source of knowledge participation from low- and middle-income countries given that they are primarily the beneficiaries of the outcomes of CHNRI exercises.[64] Wazny suggested conducting a large-scale crowdsourcing-based exercise to collate inputs from wide range of people globally to weight the CHNRI criteria, out of which any sub-group analysis would be possible to retrieve input from specific groups such as fields of expertise, interests or any geographic area for the globally led exercise. A pool of stakeholders' views will be particularly useful in global research prioritisation exercises. However, for country-led exercises it would be suitable to have more targeted sample of stakeholders who are familiar with the local context and preferably a smaller number of stakeholders compared with the global exercise.[58] I believe that this exercise provides an option to respond to the earlier questions with regard to the type and quantity of stakeholders.

Involving researchers

Paper 3.1 confirmed that the collective wisdom of experts is better than the collective wisdom of non-experts. In CHNRI exercises, the collective wisdom of experts was largely represented by researchers in past CHNRI exercises since the exercises have been more successful in attracting interest from people from the research community, and less successful among other experts such as programme managers and policy makers. This is clearly seen in the high numbers of researchers participating in many of the previous CHNRI exercises.[17–19,35,65] In CHNRI exercises, one of the major concerns is about the response bias resulting from having a large representation of researchers and less in the other group. Response bias can be reduced to an extent by having a larger number of participants, but when most participants are self-selected researchers, as in most previously conducted CHNRI exercises, a larger number of participants would only address part of the problem. One of the analyses in Chapter D indicated the potential for self-selection bias not only because the overall response rate for most professional categories was very low (9-17%) but also because the majority of those who scored were researchers. A similar tendency was reported in other studies in which relatively few policy makers and programme managers were found to have participated in the exercise, so limiting the representativeness of the full range of experts or their views, unless the theme and concept of the RP exercise is very focused on issues where input from a specific group of experts are relevant.

I believe that the CHNRI method should continue to involve researchers while increasing interest among experts in other professional categories. Involving researchers in the CHNRI process is likely to promote the successful implementation of research activities after the prioritisation exercises have been conducted and is likely to contribute to one of the long-term goals of RP exercises, i.e. building capacity among researchers in national health research systems.

Involving funders

Involving funders in the CHNRI process is another aspect of the process that has been less explored.[27] Paper 4.3 provided examples of when and how to involve funders. The MCA/WHO conducted an exploratory exercise with donors to understand the donors' perspective. The exercise showed that the funders' perspective differs from that of the researchers or stakeholders.[28] Therefore, it is important to involve the funders at an early stage, including the planning stage, so that their views are reflected in the conception and focus of the prioritisation exercise. This will help to ensure that the outcome of the exercise will be easily understood by the funders. Having said that however, from a methodological perspective, determining the best way to involve funders in the CHNRI process is challenging. Not only their involvement but also the clear communication the outcomes and securing their commitment to acknowledge the results of the CHNRI process constitute the challenges that the CHNRI users have been facing to date.

In this exploratory exercise, some 40 representatives of funding agencies were invited to provide their views and perspectives on the research priorities identified. The paper highlighted some important areas of agreement from the meeting. Firstly, funders should be involved early in the process of research prioritisation, to increase their ownership of the results. Secondly, it was recommended that the criteria include the funders' collective preference in the process. The meeting clearly identified important differences in funders' preferred criteria. Quite a few funders were interested in the potential for partnership between academic researchers and industry, to translate the findings into the production of tools or devices. Incorporating donors' preferred criteria would enhance the relevance of an exercise to a donor, potentially increasing the likelihood that the exercise will influence donors' subsequent funding decisions. The exercise recommended involving donors in the conceptualisation of the process given that the "perceived returns on investments in health research should be clearly stated at the beginning of the process". Having them in the conceptualisation of the process is likely to help secure funding commitment to move the identified priorities to the next level, in which these are funded as future research projects.

Remarkable progress has been seen in the reduction of child mortality since 1990, although the reduction of neonatal mortality proceeds at a slower pace. The newborn health research prioritisation exercise in 2014 highlighted the lack of funding and attention given to the area of newborn health and life-saving newborn interventions.[20]

The more recent analysis, published after the RP on newborn health, showed a substantial increase in the funding benefitting prenatal and neonatal health from 2003 to 2013 both in absolute and in relative terms to funding given to maternal newborn and child health. Most funding given to prenatal and newborn health was from the USA ($2848 million, 40%), Canada ($1198 million, 17%) and the International Development Association (World Bank, $585 million, 8%).[66] And yet perinatal health and newborn health were rarely mentioned in funding descriptions. The paper pointed out the need for more resource mobilisation for effective interventions in saving newborn lives, in addition to high-quality and timely monitoring of the activities. Research prioritisation processes should consider these needs, to make sure that the donors' perspectives are considered in identifying the priorities, and that more and more research prioritisation exercises should encourage donors' participation.

In this thesis, we identified how the donors' perspective differs from other group such as experts and stakeholders, as well as how perspectives could also differ between the funders. The paper showcased examples of how they could be involved in the CHNRI exercise and recognised the challenges in targeting one or a few funders with similar perspective. We have assessed the impact of CHNRI exercises to some extent by looking at funding allocated to priorities. What I could not do is explore how funding organizations make priorities on what they fund. (With hindsight, it would have been better if I had been able to do this.) A recent study highlighted the vast diversity in mechanisms by which funding organizations decide on what to fund.[67] What is noted from an initial web-based search and personal communications was that the funders' priority-setting processes were generally not well documented and even when documented, the description is not clear enough to understand who was involved, how many people were involved, whether the priorities identified were based on comprehensive approaches or by discussion based consensus. The result of the initial web-search is presented in **Appendix 3.** In general, most funders seem to use varied approaches and I found it difficult to analyse based on the available information. In the future, I would like to conduct interviews with donors to deepen my understanding of how funders identify their research priorities to better comprehend their perspectives and needs.

Chapter 6. Conclusion

I am confident that I have been able to achieve aim of the PhD thesis, namely to evaluate some aspects of the CHNRI method and evaluated several specific CHNRI exercises. Firstly, I reviewed the tools and approaches used in health RPs, to deepen my understanding of the available tools. Second, I conducted two RP exercises using the method, to learn more about the methodology. Thirdly, I revisited some of the underlying assumptions of CHNRI methodology and assessed these assumptions to provide useful guidance to the user of the method. Finally, I have developed a new evaluation framework to assess quality of the RP exercises in general based on available materials,[2,10,32,33] and applied the framework on the previously conducted CHNRI exercises. The application of the framework led to methodological recommendations for future CHNRI exercises to consider.

To my surprise, the evaluation of research prioritisation exercises is a poorly documented area. Many research prioritisation exercises put a greater effort into the identification of research questions and a lesser effort beyond that point, which is exactly why RP exercises were criticised as one-off events. Evaluation of the process and quality of the CHNRI exercises should be an integral part of future CHNRI exercises. The first evaluation is an immediate review of the process to assess transparency, inclusiveness and relevance of prioritisation exercises and the second evaluation is a review of impact that can be conducted after some years. It provides credibility for such exercises in translating the identified priorities into funding allocation.

"CHNRI method is a lasting legacy of the Child Health Nutrition Initiative that has lasted even after the initiative was resolved in 2015."[68] In the next few years, CHNRI is likely to be the most frequently used RP method in setting the health research agenda. It is highly recommended that an evaluation is conducted in future CHNRI exercises. But first, the development of a framework for the evaluation of the process and impact of RP exercises is required. I propose that the new framework can be used to assess future CHNRI exercises.

References

1. World Health Assembly resolution 63.21. full-text. *World Heal Organ*. 2010;A63/VR/8.

2. Mador RL, Kornas K, Simard A, Haroun V. Using the Nine Common Themes of Good Practice checklist as a tool for evaluating the research priority setting process of a provincial research and program evaluation program. *Heal Res Policy Syst*. 2016;14(1):1-9. doi:10.1186/s12961-016-0092-5

3. WHO. a47df47068a5cd15cafa92723c99fcbf4405f8d6 @ www.who.int. Working for health.

4. Yoshida S. Approaches, tools and methods used for setting priorities in health research in the 21st century. *J Glob Health*. 2015;6(1). doi:10.7189/jogh.06.010507

5. Rudan I, Yoshida S, Chan KY, et al. Setting health research priorities using the CHNRI method: VII. A review of the first 50 applications of the CHNRI method. *J Glob Health*. 2017;7(1). doi:10.7189/jogh.07.011004

6. Rudan I. Global health research priorities: mobilizing the developing world. *Public Health*. 2012;126(3):237-240. doi:10.1016/j.puhe.2011.12.001

7. Rudan I, Gibson J, Kapiriri L, et al. Setting priorities in global child health research investments: assessment of principles and practice. *Croat Med J*. 2007;48(5):595-604. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2205967&tool=pmcentrez&rendertype=abstract.

8. Kennedy HP, Yoshida S, Costello A, et al. Asking different questions: research priorities to improve the quality of care for every woman, every child. *Lancet Glob Heal*. 2016;4(11):e777-e779. doi:10.1016/S2214-109X(16)30183-8

9. Tomlinson M, Swartz L, Officer A, Chan KY, Rudan I, Saxena S. Research priorities for health of people with disabilities: an expert opinion exercise. *Lancet*. 2009;374(9704):1857-1862. doi:10.1016/S0140-6736(09)61910-3

10. Rudan I, Gibson JL, Ameratunga S, et al. Setting Priorities in Global Child Health Research Investments: Guidelines for Implementation of the CHNRI Method. *Croat Med J*. 2008;49(6):720-733. doi:10.3325/cmj.2008.49.720

11. Rudan I, Chopra M, Kapiriri L, et al. Setting Priorities in Global Child Health Research Investments: Universal Challenges and Conceptual Framework. *Croat Med J*. 2008;49(3):307-317. doi:10.3325/cmj.2008.3.307

12. Rudan I. Setting health research priorities using the CHNRI method: IV. Key conceptual advances. *J Glob Health*. 2016;6(1):1-8. doi:10.7189/jogh.06.010501

13. McGregor S, Henderson KJ, Kaldor JM. How are health research priorities set in low and middle income countries? A systematic review of published reports. *PLoS One*. 2014;9(10). doi:10.1371/journal.pone.0108787

14. Rudan I, El Arifeen S, Black RE, Campbell H. Childhood pneumonia and diarrhoea: setting our priorities right. *Lancet Infect Dis*. 2007;7(1):56-61. doi:10.1016/S1473-3099(06)70687-9

15. George A, Young M, Bang A, et al. Setting Implementation Research Priorities to Reduce Preterm Births and Stillbirths at the Community Level. *PLoS Med*. 2011;8(1):e1000380. doi:10.1371/journal.pmed.1000380

16. Rudan I, El Arifeen S, Bhutta Z a, et al. Setting research priorities to reduce global mortality from childhood pneumonia by 2015. *PLoS Med*. 2011;8(9):e1001099. doi:10.1371/journal.pmed.1001099

17. Fontaine O, Kosek M, Bhatnagar S, et al. Setting research priorities to reduce global mortality from childhood diarrhoea by 2015. *PLoS Med*. 2009;6(3):e41. doi:10.1371/journal.pmed.1000041

18. Bahl, R., Martines, J., Bhandari, N., Biloglav, Z, Edmond, K., Iyengar, S., Kramer, M., Lawn, J.E, Manandhar, D.S., Mori, R., Rasmussen, K.M., Sachdev, H. P. S., Singhal, N., Tomlinson, M., Victora, C., Williams, A.F., Yee Chan, K., Rudan I. Setting research priorities to reduce global mortality from preterm birth and low birth weight by 2015. *J Glob Health*. 2012;2(June 2012).

19. Lawn JE, Bahl R, Bergstrom S, et al. Setting research priorities to reduce almost one million deaths from birth

asphyxia by 2015. *PLoS Med*. 2011;8(1):e1000389. doi:10.1371/journal.pmed.1000389

20. Yoshida S, Rudan I, Lawn JE, et al. Newborn health research priorities beyond 2015. *Lancet*. 2014;384(9938):e27-9. doi:10.1016/S0140-6736(14)60263-4

21. Hindin MJ, Sigurdson Christiansen C, Ferguson J. Setting research priorities for adolescent sexual and reproductive health in low-and middle-income countries. *Bull World Heal Organ*. 2013;9112(10). doi:10.2471/BLT.12.107565

22. Nagata JM, Ferguson BJ, Ross DA. Research Priorities for Eight Areas of Adolescent Health in Low- and Middle-Income Countries. *J Adolesc Heal*. 2016;59(1):50-60. doi:10.1016/j.jadohealth.2016.03.016

23. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? *Field methods*. 2006;18(1):59-82. doi:10.1177/1525822X05279903

24. Mason M. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum Qual Sozialforsch / Forum Qual Soc Res Vol 11, No 3 Methods Qual Manag Res Context Soc Syst Think*. 2010. doi:10.17169/fqs-11.3.1428

25. Souza JP, Widmer M, Gülmezoglu AM, et al. Maternal and perinatal health research priorities beyond 2015: An international survey and prioritization exercise. *Reprod Health*. 2014;11(1):1-9. doi:10.1186/1742-4755-11-61

26. Yoshida S, Cousens S, Wazny K, Chan KY. Setting health research priorities using the CHNRI method: II. Involving researchers. *J Glob Health*. 2016;6(1). doi:10.7189/jogh.06.010302

27. Tomlinson M, Chopra M, Hoosain N, Rudan I. A review of selected research priority setting processes at national level in low and middle income countries: towards fair and legitimate priority setting. *Heal Res Policy Syst*. 2011;9:19. doi:10.1186/1478-4505-9-19

28. Rudan I, Yoshida S, Chan KY, et al. Setting health research priorities using the CHNRI method: I. Involving funders. *J Glob Health*. 2016;6(1). doi:10.7189/jogh.06.010301

29. J. B, R. S-F, J. W. Health research priority setting in selected high income countries: A narrative review of methods used and recommendations for future practice. *Cost Eff Resour Alloc*. 2014;12(1):no pagination.

30. Chanda-Kapata P, Ngosa W, Hamainza B, Kapiriri L. Health research priority setting in Zambia: A stock taking of approaches conducted from 1998 to 2015. *Heal Res Policy Syst*. 2016;14(1). doi:10.1186/s12961-016-0142-z

31. Viergever RF, Olifson S, Ghaffar A, Terry RF. A checklist for health research priority setting: nine common themes of good practice. *Heal Res policy Syst*. 2010;8:36. doi:10.1186/1478-4505-8-36

32. Kapiriri L, Martin DK. Successful priority setting in low and middle income countries: A framework for evaluation. *Heal Care Anal*. 2010;18(2):129-147. doi:10.1007/s10728-009-0115-2

33. Kapiriri L, Chanda-Kapata P. The quest for a framework for sustainable and institutionalised priority-setting for health research in a low-resource setting: the case of Zambia. *Heal Res Policy Syst*. 2018;16:11. doi:10.1186/s12961-017-0268-7

34. Yoshida S, Martines J, Lawn JE, et al. Setting research priorities to improve global newborn health and prevent stillbirths by 2025. *J Glob Health*. 2016;6(1):010508. doi:10.7189/jogh.06.010508

35. Bahl R, Martines J, Ali N, et al. Research priorities to reduce global mortality from newborn infections by 2015. *Pediatr Infect Dis J*. 2009;28(SUPPL. 1):43-48. doi:10.1097/INF.0b013e31819588d7

36. Liu L, Johnson HL, Cousens S, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*. 2012;379(9832):2151-2161. doi:10.1016/S0140-6736(12)60560-1

37. Lawn JE, Blencowe H, Pattinson R, et al. Stillbirths: Where? When? Why? How to make the data count? *Lancet*. 2011;377(9775):1448-1463. doi:10.1016/S0140-6736(10)62187-3

38. Mwaniki MK, Atieno M, Lawn JE, Newton CRJC. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. *Lancet*. 2012;379(9814):445-452. doi:10.1016/S0140-6736(11)61577-8

39. Immediate Kangaroo Mother Care (i-KMC). 2019;(November):2019.

40. Scale-up A. Kangaroo Mother Care Implementation Research for. 2018.

41. Improving Quality of Care for Mothers and Newborns at District Hospitals and FRUs of Three Districts of Haryana. 2018.

42. Kok MO, Rodrigues A, Silva AP, de Haan S. The emergence and current performance of a health research system: lessons from Guinea Bissau. *Heal Res Policy Syst*. 2012;10(1):5. doi:10.1186/1478-4505-10-5

43. AbouZahr C, Boerma T, Byass P. Bridging the data gaps: do we have the right balance between country data and global estimates? *Glob Health Action*. 2017;10(sup1):1299978. doi:10.1080/16549716.2017.1299978

44. Sarantakos S. *Social Research*. Macmillan Education UK; 2012.

45. Rudan I, Yoshida S, Wazny K, Chan KY, Cousens S. Setting health research priorities using the CHNRI method: V. Quantitative properties of human collective knowledge. *J Glob Health*. 2016;6(1). doi:10.7189/jogh.06.010502

46. Morse JM. The Significance of Saturation. *Qual Health Res*. 1995;5(2):147-149. doi:10.1177/104973239500500201

47. Morse JM. Analytic Strategies and Sample Size. *Qual Health Res*. 2015;25(10):1317-1318. doi:10.1177/1049732315602867

48. Glaser BG, Strauss AL, Strauss AL. *Discovery of Grounded Theory*. Routledge; 2017. doi:10.4324/9780203793206

49. Hoekstra D, Mütsch M, Kien C, Gerhardus A, Lhachimi SK. Identifying and prioritising systematic review topics with public health stakeholders: A protocol for a modified Delphi study in Switzerland to inform future research agendas. *BMJ Open*. 2017;7(8):1-8. doi:10.1136/bmjopen-2016-015500

50. Hyder A, Syed S, Puvanachandra P, et al. Stakeholder analysis for health research: Case studies from low- and middle-income countries. *Public Health*. 2010;124(3):159-166. doi:10.1016/j.puhe.2009.12.006

51. Berger ML, Teutsch S. Cost-effectiveness analysis: from science to application. *Med Care*. 2005;43(7 Suppl):49-53.

52. Kazanjian A. Reflections on the social epidemiologic dimension of health technology assessment. *Int J Technol Assess Health Care*. 2004;20(2):167-173. doi:10.1017/S0266462304000947

53. Research — Future Health Systems.

54. Rollins N, Chanza H, Chimbwandira F, et al. Prioritizing the PMTCT Implementation Research Agenda in 3 African Countries. *JAIDS J Acquir Immune Defic Syndr*. 2014;67:S108-S113. doi:10.1097/QAI.0000000000000358

55. Arora NK, Mohapatra A, Gopalan HS, et al. Setting research priorities for maternal, newborn, child health and nutrition in India by engaging experts from 256 indigenous institutions contributing over 4000 research ideas: a CHNRI exercise by ICMR and INCLEN. *J Glob Health*. 2017;7(1). doi:10.7189/jogh.07.011003

56. Lavis JN, Lomas J, Hamid M, Sewankambo NK. Assessing country-level efforts to link research to action. *Bull World Health Organ*. 2006;84(8):620-628. doi:10.2471/BLT.06.030312

57. Conalogue DM, Kinn S, Mulligan JA, McNeil M. International consultation on long-term global health research priorities, research capacity and research uptake in developing countries. *Heal Res Policy Syst.* 2017;15(1):1-6. doi:10.1186/s12961-017-0181-0

58. El-Jardali F, Makhoul J, Jamal D, Ranson MK, Kronfol NM, Tchaghchagian V. Eliciting policymakers' and stakeholders' opinions to help shape health system research priorities in the Middle East and North Africa region. *Health Policy Plan*. 2010;25(1):15-27. doi:10.1093/heapol/czp059

59. Pollock A, George BS, Fenton M, Crowe S, Firkins L. Development of a new model to engage patients and clinicians in setting research priorities. *J Health Serv Res Policy*. 2013;19(1):12-18. doi:10.1177/1355819613500665

60. Sridhar D. Who Sets the Global Health Research Agenda? The Challenge of Multi-Bi Financing. *PLOS Med*.

2012;9(9):e1001312.

61. Hipgrave DB, Alderman KB, Anderson I, Soto EJ. Health sector priority setting at meso-level in lower and middle income countries: Lessons learned, available options and suggested steps. *Soc Sci Med*. 2014;102:190-200. doi:10.1016/j.socscimed.2013.11.056

62. Nuyens Y. Setting priorities for health research: lessons from low- and middle-income countries. *Bull World Health Organ*. 2007;85(4):319-321. doi:10.2471/BLT.06.032375

63. Kapiriri L, Tomlinson M, Chopra M, et al. Setting priorities in global child health research investments: addressing values of stakeholders. *Croat Med J*. 2007;48(5):618-627.

64. Wazny K. Crowdsourcing's ten years in: A review. *J Glob Health*. 2017;7(2):1-13. doi:10.7189/jogh.07.020601

65. Schouten J a, Hulscher MEJL, Natsch S, et al. Potential interventions for the prevention of childhood pneumonia in developing countries: A systematic review. *Bull World Health Organ*. 2007;126(1):1636-1650. doi:10.1016/S1473-3099(06)70687-9

66. Pitt C, Grollman C, Martínez-Álvarez M, Arregoces L, Lawn JE, Borghi J. Countdown to 2015: an analysis of donor funding for prenatal and neonatal health, 2003–2013. *BMJ Glob Heal*. 2017;2(2).

67. Viergever RF, Hendriks TCC. The 10 largest public and philanthropic funders of health research in the world: What they fund and how they distribute their funds. *Heal Res Policy Syst*. 2016;14(1). doi:10.1186/s12961-015-0074-z

68. Black RE. The legacy of the Child Health and Nutrition Research Initiative (CHNRI). *J Glob Health*. 2016;6(1):010101. doi:10.7189/jogh-06-010101

Appendices

Appendix 1 List of search terms used in the literature search

CHNRI Exercise Topic (Year)	Research Priority	Search terms
Childhood Diarrhea (2009)	1: What is the acceptability and effectiveness of the new reduced osmolarity ORS in clinic as well as in the community?	(("Fluid Therapy"[Mesh]) OR "World Health Organization oral rehydration solution" [Supplementary Concept]) AND (“Diarrhea, Infantile"[Mesh] OR "Diarrhea"[Mesh]) AND ("Patient Acceptance of Health Care"[Mesh] OR "Evaluation Studies as Topic"[Mesh])
	2: What is the effectiveness of zinc supplementation on the outcome and incidence of diarrhoea in the community?	((((("Dietary Supplements"[Mesh]OR "supplementation"[All Fields] OR “Supplement"[All Fields]) AND "Zinc"[Mesh]) AND ("Diarrhea"[Mesh] OR "Diarrhea, Infantile"[Mesh]) AND ("Residence Characteristics"[Mesh]OR "household" OR "home") OR “Incidence”)
	3: What are the barriers against appropriate use of ORT?	(("Fluid Therapy"[Mesh] OR "World Health Organization oral rehydration solution" [Supplementary Concept]) AND (“Diarrhea"[Mesh] OR "Diarrhea, Infantile"[Mesh])) AND (“Evaluation Studies as Topic"[Mesh] OR "Evaluation Studies" [Publication Type] OR "Program Evaluation"[Mesh] OR “Appropriate”)
	4: Design locally adapted training programmes to orient health workers on IMCI.	("Integrated Management of Childhood Illness" OR "IMCI") AND ("Evaluation Studies" [Publication Type] OR "Program Evaluation"[Mesh]OR "Workers" OR "Teaching"[Mesh] OR "Staff Development"[Mesh])
	5: What is the impact of IMCI in different population groups on timely identification and treatment of acute diarrhoea?	("Program Evaluation"[Mesh] OR "Evaluation Studies as Topic"[Mesh] OR "Evaluation Studies" [Publication Type]) OR (“Quality Assurance, Health Care"[Mesh] OR "Quality Improvement"[Mesh] OR "Quality of Health Care"[Mesh]) AND ("IMCI" OR "Integrated Management of Childhood Illness")
Childhood Pneumonia (2011)	1: Study the main barriers to health care seeking and health care access for children with pneumonia in different contexts and settings in developing countries.	"Pneumonia"[Mesh] AND ("Health Services Accessibility"[Mesh] OR “Patient Acceptance of Health Care"[Mesh])
	2: Identify the key risk factors predisposing to the development of severe pneumonia and identify children who require hospitalization.	("Risk Factors"[Mesh]) AND "Pneumonia"[Mesh] AND ("Severe" OR " Hospitalization" OR "hospitalization" OR "serious" OR "invasive")
	3: Study the main barriers to increasing coverage by available vaccines - Hib vaccine and pneumococcal vaccine - in different contexts and settings.	(("Pneumococcal Vaccines"[Mesh] OR "Haemophilus influenzae type b polysaccharide vaccine" [Supplementary Concept]) AND "Pneumonia"[Mesh]) AND (“Evaluation Studies as Topic"[Mesh] OR "Evaluation Studies" [Publication Type] OR "barriers")
	4: Study whether the coverage by antibiotic treatment can be greatly expanded in safe and effective ways if it was administered by community health workers.	(("Pneumonia"[Mesh]) AND (“Anti-Bacterial Agents"[Mesh] OR "Anti-Bacterial Agents" [Pharmacological Action])) AND "Community Health Workers"[Mesh]
	5: Study the main barriers to increasing demand for/compliance with vaccination with available vaccines in different contexts and settings - for measles and pertussis vaccines, Hib vaccine, and pneumococcal vaccine.	("Pneumococcal Vaccines"[Mesh] Rheophiles influenzae type b polysaccharide vaccine" [Supplementary Concept] OR “Measles-Mumps-Rubella Vaccine"[Mesh] OR “Pertussis Vaccine"[Mesh] OR “Measles Vaccine"[Mesh]) AND ("Patient Compliance"[Mesh] OR "demand”) NOT "adult" NOT

		"elderly" AND ("Evaluation Studies as Topic"[Mesh] OR "Evaluation Studies" [Publication Type] OR "barriers")
Intrapartum Related Neonatal Death (2011)	1,3,4: Can community cadres of workers identify a limited number of high-risk conditions/danger sings (e.g. multiple pregnancy, breech, short maternal stature, etc.) and successfully refer women for facility birth? What is the predicative value and cost effectiveness? // Behavioral/community participation package to improve recognition and acting for simplified danger signs for mother in labour, including transport and phone/radio communication ("emergency preparedness")? // Effectiveness of community cadre roles, e.g., social support, bringing to facility when woman is in labour, danger recognition/referral?	("residence characteristics"[MeSH Terms] OR ("residence"[All Fields] AND "characteristics"[All Fields]) OR "residence characteristics"[All Fields] OR "community"[All Fields]) AND ((("mothers"[MeSH Terms] OR "mothers"[All Fields] OR "maternal"[All Fields]) AND "danger sign"[All Fields]) OR "facility delivery"[All Fields] OR (("mothers"[MeSH Terms] OR "mothers"[All Fields] OR "maternal"[All Fields]) AND ("referral and consultation"[MeSH Terms] OR ("referral"[All Fields] AND "consultation"[All Fields]) OR "referral and consultation"[All Fields] OR "referral"[All Fields]))))
	2: What strategies are effective in increasing demand for, and use of, skilled attendance (e.g. conditional cash transfers)?	("Demand"[All Fields] OR ("Utilization"[Subheading] OR "Utilization"[All Fields] OR "Use"[All Fields]) OR ("Utilization"[Subheading] OR "Utilization"[All Fields])) AND (("Skilled"[All Fields] AND ("Parturition"[MeSH Terms] OR "Parturition"[All Fields] OR "Birth"[All Fields]) AND (attend*) OR "Facility delivery"[All Fields])
	5: Does regular use of perinatal audit reduce the incidence of adverse outcomes related to acute intrapartum events?	("Perinatal Care/organization and administration"[MeSH]) AND ("Infant Mortality/trends"[MAJR] OR audit*)
Neonatal Infection (2009)	1: What are the feasibility, effectiveness and cost of different approaches to promote the following home care practices: early initiation and exclusivity of breastfeeding, hygienic cord and skin care, prompt care seeking for illness from an appropriate provider?	(("infant, newborn"[MeSH Terms] OR ("infant"[All Fields] AND "newborn"[All Fields]) OR "newborn infant"[All Fields] OR "newborn"[All Fields]) OR ("infant, newborn"[MeSH Terms] OR ("infant"[All Fields] AND "newborn"[All Fields]) OR "newborn infant"[All Fields] OR "neonate"[All Fields])) AND promotion[All Fields] AND home[All Fields] AND care[All Fields]
	2: What is the role of local application of disinfectants in the prevention of umbilical infections and sepsis?	"umbilical infection"[All Fields] OR omphalitis[All Fields] OR ((neonatal[All Fields] OR ("infant, newborn"[MeSH Terms] OR ("infant"[All Fields] AND "newborn"[All Fields]) OR "newborn infant"[All Fields] OR "newborn"[All Fields])) AND ("sepsis"[MeSH Terms] OR "sepsis"[All Fields])) AND (local[All Fields] AND ("disinfectants"[Pharmacological Action] OR "disinfectants"[MeSH Terms] OR "disinfectants"[All Fields] OR "disinfectant"[All Fields]))
	3: What are the feasibility, effectiveness and cost of approaches to increase coverage of clean delivery practices in facilities and in homes?	("Neonatal"[All Fields] OR ("Infant, newborn"[MeSH Terms] OR ("Infant"[All Fields] AND "Newborn"[All Fields]) OR "Newborn infant"[All Fields] OR "Newborn"[All Fields])) AND ("Infection"[MeSH Terms] OR "Infection"[All Fields]) AND ("Clean"[All Fields] OR "Sterile"[All Fields]) OR ("Hygiene"[MeSH Terms] OR "Hygiene"[All Fields]) OR "Hygienic"[All Fields]) AND ("Delivery, obstetric"[MeSH Terms] OR ("Delivery"[All Fields] AND "Obstetric"[All Fields]) OR "Obstetric delivery"[All Fields] OR "Delivery"[All Fields])
	4: What are the feasibility, cost and effectiveness of setting up newborn care corners in first referral units and district hospitals?	(("Delivery Rooms"[Mesh]) AND ("Evaluation Studies as Topic"[Mesh] AND ("Hospitals, District"[Mesh] OR "Community Health Centers"[Mesh])
	5: What is the feasibility effectiveness and cost of a scheme of routine home visits for initiation of supportive	("Neonatal"[All Fields] OR ("Infant, newborn"[MeSH Terms] OR ("Infant"[All Fields] AND "Newborn"[All Fields]) OR "Newborn infant"[All Fields] OR "Newborn"[All Fields])) AND "Home visit"[All Fields

	practices, detection of illness and newborn survival?	

Appendix 2 List of articles identified

Question	Publication	Year	CHNRI in Article	Funders	Type of Funder
1.1	Predictors of Oral Rehydration Therapy use among under-five children with diarrhoea in Eastern Ethiopia: a community-based case control study	2012	No	College of Health Science of Harmala University for the financial support. Kersa district health office also deserves acknowledgment for its material assistance.	University/Institution
1.1	Knowledge and practices regarding oral rehydration therapy among mothers in rural area of Vasind, India	2013	No	Not included within paper	N/A
1.2	Oral rehydration salts, zinc supplement and rota virus vaccine in the management of childhood acute diarrhoea	2010	No	Source of Support: Nil	None
1.2	The added benefit of zinc supplementation after zinc treatment of acute childhood diarrhoea: a randomized, double-blind field trial	2010	No	Funded through a grant from the Bill & Melinda Gates Foundation.	Private Foundation
1.2	Zinc for the treatment of diarrhoea: effect on diarrhoea morbidity, mortality and incidence of future episodes	2010	No	Supported in part by a grant to the US Fund for UNICEF from the Bill & Melinda Gates Foundation (grant 43386) to "Promote evidence-based decision making in designing maternal, neonatal and child health interventions in low- and middle-income countries".	Private Foundation
1.2	Role of zinc in Pediatric diarrhoea	2011	No	Source of Support: Nil.	None
1.2	Preventive zinc supplementation in developing countries: impact on mortality and morbidity due to diarrhoea, pneumonia and malaria	2011	No	Supported in part by a grant to the US Fund for UNICEF from the Bill & Melinda Gates Foundation (grant 43386) to "Promote evidence-based decision making in designing maternal, neonatal and child health interventions in low- and middle-income countries".	Private Foundation
1.2	Short-course prophylactic zinc supplementation for diarrhoea morbidity in infants of 6 to 11 months	2013	No	Supported by the Indian Council of Medical Research, Department of Health Research (Ministry of Health and Family Welfare), and Government of India. Reference No. 3/2/2011/PG-thesis-MPD-10.	Ministry

1.2	Zinc and other micronutrients supplementation using sprinkles: impact on the occurrence of diarrhoea and respiratory infections in institutionalized children	2013	No	Donated by Emory University, Atlanta, GA, USA.	University/Institution
1.2	Therapeutic effects of oral zinc supplementation on acute watery diarrhoea with moderate dehydration: a double-blind randomized clinical trial	2013	No	Office of Vice Chancellor for Research of Urmia University of Medical Sciences for financial support of this study	University/Institution
1.3	Practice and attitudes regarding the management of childhood diarrhoea among pharmacies in Thailand	2010	No	Supported financially by Prince of Songkla University (grant number PHA5122020048S).	University/Institution
1.3	Evaluation of a social marketing intervention promoting oral rehydration salts in Burundi	2011	No	Funding for this study was provided by USAID	Government Agency
1.3	Community case management of childhood diarrhoea in a setting with declining use of oral rehydration therapy: findings from cross-sectional studies among primary household caregivers, Kenya, 2007	2011	No	Funded by the United States Agency for International Development (USAID) and the Centres for Disease Control and Prevention (CDC).	Government Agency
1.3	Examining the use of oral rehydration salts and other oral rehydration therapy for childhood diarrhoea in Kenya	2011	No	Supported by the United States Agency for International Development.	Government Agency
1.3	A study to evaluate the acceptability, feasibility and impact of packaged interventions ("Diarrhoea Pack") for prevention and treatment of childhood diarrhoea in rural Pakistan	2013	No	Not found online	N/A
1.3	Protocol for the economic evaluation of the diarrhoea alleviation through zinc and oral rehydration salt therapy at scale through private and public providers in rural Gujarat and Uttar Pradesh, India	2014	No	The DAZT Program is in partnership between the Micronutrients Initiative, Family Health International-360, UNICEF, Clinton Health Access Initiative (CHAI), the US Fund, and Johns Hopkins Bloomberg School of Public Health that is made possible only through the generous	Private Foundation

				support of the Bill and Melinda Gate's Foundation (BMGF).	
1.3	Household Management of Childhood Diarrhoea: A Population-based Study in Nicaragua	2014	No	Supported by 5K01TW008401-04 from the Fogarty International Center at the US National Institutes of Health.	University/Institution
1.4	The rise and fall of supervision in a project designed to strengthen supervision of Integrated Management of Childhood Illness in Benin	2010	No	Funded by the United States Agency for International Development's Africa Integrated Malaria Initiative (project number 936-3100).	Government Agency
1.4	Assessment of implementation of integrated management of neonatal and childhood illness in India	2011	No	Not mentioned in paper	N/A
1.4	Integrated management of childhood illness in Lahej, Yemen: a qualitative analysis from the perspective of health providers	2011	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
1.4	Global challenges with scale-up of the integrated management of childhood illness strategy: results of a multi-country survey	2011	No	Not mentioned in paper	N/A
1.4	The challenges of achieving high training coverage for IMCI: case studies from Kenya and Tanzania	2011	No	Supported by the Consortium for Research on Equitable Health Systems (CREHS) which is funded by the United Kingdom's Department for International Development (DFID). CG and FW are members of the KEMRI/Wellcome Trust Research Programme, which is supported by a grant from the Wellcome Trust (#077092).	Private Foundation
1.4	Evaluating health worker performance in Benin using the simulated client method with real children	2012	No	Supported by the United States Agency for International Development's Africa Integrated Malaria Initiative (project number 936–3100). The funding agency did not influence any aspect of the study or the decision to submit it for publication	Government Agency
1.4	Trends in health worker performance after implementing the Integrated Management of Childhood Illness strategy in Benin	2012	No	Funded by the United States Agency for International Development's Africa Integrated Malaria Initiative.	Government Agency

1.4	Teaching of the Integrated Management of Childhood Illness strategy in undergraduate nursing programs	2012	No	Supported by Conselho Nacional de Pesquisa e Desenvolvimento Científico e Technologic (CNPq), process # 479475/2010-5	Ministry
1.4	Clinical mentorship to improve Pediatric quality of care at the health centres in rural Rwanda: a qualitative study of perceptions and acceptability of health care workers	2014	No	Supported by funds from the African Health Initiative of the Doris Duke Charitable Foundation	Private Foundation
1.4	Design of an interactive medical guideline application for community health workers	2014	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
1.4	Factors influencing the implementation of integrated management of childhood illness (IMCI) by healthcare workers at public health centres and dispensaries in Mwanza, Tanzania	2014	No	Thank the Community Medicine department at Catholic University of Health and Allied Sciences (CUHAS) and CUHAS in general for their guidance and technical assistance to ensure this study is a success.	University/Institution
1.4	Improving and sustaining quality of child health care through IMCI training and supervision: experience from rural Bangladesh	2014	No	Arranged, coordinated and funded by the Department of Child and Adolescent Health and Development of the World Health Organization, and with the financial support of the Bill and Melinda Gates Foundation and the US Agency for International Development.	Private Foundation and Government Agency
1.5	Clinical signs predicting severe illness in young infants (<60 days) in Bolivia	2010	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
1.5	Evaluation of short term integrated management of childhood illness training on the clinical competency of village doctors in Yunnan, China	2012	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
1.5	Using partnership approach to reduce mortality and morbidity among children under five in Limpopo province, South Africa	2012	No	Makhuduthamaga Child Survival Project team for their contribution in implementing the project and AMREF UK for securing funding for the project from	Private Foundation

				UBS-Optimus Foundation and AMREF South Africa management team for their technical support in managing the project	
1.5	Integrated management of mother and child health in the region of Monastir (Tunisia)]	2014	CHNRI not mentioned in abstract but whole article in French	Not mentioned in paper or missed because article in French	N/A
2.1	[Social inequalities in spatial distribution of hospital admissions due to respiratory diseases]	2013	CHNRI not mentioned in abstract but whole article in Portuguese	Not mentioned in paper or missed because article in Portuguese	N/A
2.1	Understanding care seeking for child illness in sub-Saharan Africa: a systematic review and conceptual framework based on qualitative research of household recognition and response to child diarrhoea, pneumonia and malaria	2013	No	Co-authors (AG and JK) representing the funding source (UNICEF) helped interpret the review findings and implications, advised on writing up the findings, and contributed to revising drafts of the manuscript	Multilateral Agency
2.1	Healthcare seeking for diarrhoea, malaria and pneumonia among children in four poor rural districts in Sierra Leone in the context of free health care: results of a cross-sectional survey	2013	No	Funds from the Canadian International Development Association and is supported by the United Nations Children Emergency Fund (UNICEF).	Multilateral Agency
2.1	Increased access to care and appropriateness of treatment at private sector drug shops with integrated management of malaria, pneumonia and diarrhoea: a quasi-experimental study in Uganda	2014	No	Einhorn Family Foundation – Sweden for funding this study. The Einhorn Family Foundation had no role in study design, data collection, data analysis, interpretation or writing the report.	Private Foundation
2.1	Observational follow-up study on a cohort of children with severe pneumonia after discharge from a day-care clinic in Dhaka, Bangladesh	2014	No	Funded by the Swiss Agency for Development and Cooperation (SDC), Bern; the Gastrointestinal Research Foundation, Liestal; and the University of Basel, Switzerland	Government Agency and University/Institution
2.1	Local barriers and solutions to improve care-seeking for childhood pneumonia, diarrhoea and malaria in Kenya, Nigeria and Niger: a qualitative study	2014	No	Grants from the Bill and Melinda Gates Foundation and the Rockefeller Foundation	Private Foundation

2.1	Household health care-seeking costs: experiences from a randomized, controlled trial of community-based malaria and pneumonia treatment among under-fives in eastern Uganda	2014	No	This study received financial support from Sida/SAREC and UNICEF/UNDP/ World Bank/WHO Special Programme for Research and Training in Tropical Diseases. FM received financial support from the Malaria Capacity Development Consortium (MCDC) at the London School of Hygiene and Tropical Medicine under the PDP Programme.	Multilateral Agency
2.1	Oxygen and pulse oximetry in childhood pneumonia: surveys of clinicians and student clinicians in Cambodia	2014	No	National Immunisation Programme within the Ministry of Health in Cambodia for supporting the distribution of the questionnaires	Ministry
2.1	Determinants of care seeking for children with pneumonia and diarrhoea in Guatemala: implications for intervention strategies	2014	No	Funding for the study was provided by the UBS Optimus Foundation	Private Foundation
2.1	Care seeking behaviour for children with suspected pneumonia in countries in sub-Saharan Africa with high pneumonia mortality	2015	No	ACN is an external PhD candidate at the University of Maastricht; she received no funding for this work. LCV, ABS and MY are employed by the United Nations Children's Fund (UNICEF), they received no specific funding for this work. JWLC is employed by the University of Maastricht and is supported by a Veni-grant (91614078) of the Netherlands Organization for Health Research and Development (ZonMw). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.	University/Institution
2.2	Prenatal and postnatal risk factors for infantile pneumonia in a representative birth cohort	2012	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
2.2	Evaluation of risk factors for severe pneumonia in children: the Pneumonia Etiology Research for Child Health study	2012	No	Supported by grant 48968 from The Bill & Melinda Gates Foundation to the International Vaccine Access Center, Department of International Health,	Private Foundation

				Johns Hopkins Bloomberg School of Public Health.	
2.2	Risk factors of severe pneumonia among children aged 2-59 months in western Kenya: a case control study	2012	No	Funded by the Centers for Disease Control and Prevention (CDC) Kenya and the Ministry of Public Health and Sanitation	Government Agency and Ministry
2.2	Factors associated with severe disease from malaria, pneumonia and diarrhoea among children in rural Tanzania - a hospital-based cross-sectional study	2012	No	Funded by the Norwegian government through Quota programme and was part of PhD training	Government Agency
2.2	Risk Factors for Mortality in Community –Acquired Pneumonia Among Children Aged 1-59 Months Admitted in a Referral Hospital	2012	No	Funding: Nil	None
2.2	[Risk factors for influenza A (H1N1)-associated pneumonia on hospitalized people less than 18 years old in China, 2009-2010]	2012	Not in abstract. Full article in Chinese	Not mentioned in paper or missed because article in Chinese	N/A
2.2	Lactate as a predictor of mortality in Malawian children with WHO-defined pneumonia	2012	No	SMG was recipient of core funding from the Wellcome Trust, UK	Private Foundation
2.2	Risk factors for community-acquired pneumonia in pre-school-aged children	2012	No access to full article but CHNRI not mentioned in abstract or keywords	Funded by research grants from the Health Research Council of New Zealand, the Auckland Medical Research Foundation and Curekids.	Private Foundation and Ministry
2.2	High rates of pneumonia in children under two years of age in a South East Asian refugee population	2013	No	CT and FN are supported by the Wellcome Trust of Great Britain (Grant No. 077166/Z/05). PT is also supported by the Wellcome Trust (Grant 083735). SMRU is part of the Mahidol-Oxford University Tropical Medicine Research Programme and is supported by the Wellcome Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript	Private Foundation
2.2	Factors associated with mortality in Pediatric patients requiring extracorporeal life support for severe pneumonia	2013	No	Not mentioned in abstract	N/A

2.2	Bed-sharing and risk of hospitalisation due to pneumonia and diarrhoea in infancy: the 2004 Pelotas Birth Cohort	2013	No	Supported by the Wellcome Trust, UK, through grant number 086974/Z/08/Z	Private Foundation
2.2	Clinical profile of recurrent community-acquired pneumonia in children	2013	No	Supported by grants from the Italian Ministry of Health (Bando Giovani Ricercatori 2007).	Ministry
2.2	Childhood Anemia at High Altitude: Risk Factors for Poor Outcomes in Severe Pneumonia	2013	No	Dr Moschovis has received funding through the American Medical Association Foundation and the Harvard Global Health Institute for research in childhood pneumonia. Dr Hibberd has received funding through National Institutes of Health grant 7U01 HD058322 for research in childhood pneumonia. Dr Qazi is a staff member of the World Health Organization. The expressed views and opinions do not necessarily express the policies of the World Health Organization. Dr Saha has received funding from the Department of Microbiology, Bangladesh Institute of Child Health. The other authors have indicated they have no financial relationships relevant to this article to disclose	Private Foundation and University/Institute
2.2	A population-based analysis of children with pneumonia among intensive care units in Taiwan	2015	No	The authors have no financial or nonfinancial conflicts of interest related to the subject matter or materials discussed in the manuscript	None
2.2	[The risk factors of ventilator-associated pneumonia in newborn and the changes of isolated pathogens]	2013	Only abstract read because rest in Chinese (NO)	Not mentioned in paper or missed because article in Chinese	N/A
2.2	Clinical risk factors of death from pneumonia in children with severe acute malnutrition in an urban critical care ward of Bangladesh	2013	No	Funded by the Dhaka Hospital of International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B; grant no Gr- 00233) and its donors, which provide unrestricted support to ICDDR, B for its operations and research. Current donors providing unrestricted support include: Australian Agency for	Government Agency

				International Development, Government of the People's Republic of Bangladesh, Canadian International Development Agency, Swedish International Development Cooperation Agency, and the Department for International Development, United Kingdom. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.	
2.2	First report on prevalence and risk factors of severe atypical pneumonia in Vietnamese children aged 1-15 years	2014	No	Supported by The National Foundation for Science and Technology Development (NAFOSTED), grant no. 106.03-2010.36 from The Ministry of Science and Technology, Vietnam.	Ministry
2.2	Risk Factors for Ventilator-Associated Pneumonia in Infants and Children: a Cross-sectional Cohort Study	2014	No	Funded in part by grant 04-13361-2 from Fundação de Amparo a Pesquisa do Estado de São Paulo	Government Agency
2.2	Risk factors for a poor outcome among children admitted with clinically severe pneumonia to a university hospital in Rabat, Morocco	2014	No	Funding sources: Spanish Agency of International Cooperation for Development (AECID) through grant 07-CO1-021 awarded to Fundació Clínic per a la Recerca Biomèdica. JR has a fellowship from the programme I3, of the ISCIII (grant number CES11/012). QB has a fellowship from the programme Miguel Servet of the ISCIII (grant number CP11/00269).	Government Agency
2.2	Impact of air pollution on respiratory diseases in children with recurrent wheezing or asthma	2014	No	Supported by a grant from the Italian Ministry of Health (Bando Giovani Ricercatori 2009).	Ministry
2.2	Risk factors for ventilator-associated pneumonia in neonatal intensive care unit patients	2014	No	Not mentioned in paper	N/A
2.2	Pneumocystis pneumonia in South African children diagnosed by molecular methods.	2014	No	Supported by an NHLS Research Trust grant; the National Research Foundation, South Africa; ASTRA-Zeneca Respiratory Award from the South African Thoracic Society and the Medical Research Council of Southern Africa	Government Agency

2.2	Increased risk for respiratory syncytial virus-associated, community-acquired alveolar pneumonia in infants born at 31-36 weeks of gestation	2014	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
2.2	Incidence and severity of childhood pneumonia in the first year of life in a South African birth cohort: the Drakenstein Child Health Study	2015	No	Funding Bill & Melinda Gates Foundation, South African Thoracic Society, Federation of Infectious Diseases Societies of South Africa, and University of Cape Town.	Private Foundation
2.2	Risk factors of progressive community-acquired pneumonia in hospitalized children: a prospective study	2015	No	Supported by the Taiwan National Health Research Institutes. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript	University/Institution
2.2	Incidence and Risk Factors of Childhood Pneumonia-Like Episodes in Biliran Island, Philippines--A Community-Based Study	2015	No	Supported by Japan Science and Technology Agency and Japan International Cooperation Agency, Science and Technology Research Partnership for Sustainable Development, research name Comprehensive Etiological and Epidemiological Study on Acute Respiratory Infections in Children: Providing Evidence for the Prevention and Control of Childhood Pneumonia in the Philippines. Also supported by Japan Society for the Promotion of Science, Grant-in-Aid for Young Scientists (B), Grant Number 21790570. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.	Government Agency
2.2	Stunting is associated with poor outcomes in childhood pneumonia	2015	No access to full article but CHNRI not mentioned in abstract or keywords	Funded by •Department of Child and Adolescent Health and Development, WHO •Center for International Health and Development, Boston University •Johns Hopkins Bloomberg School of Public Health, Baltimore •National Institutes of Health. Grant	Mulitlateral Agency and University/Institute

				Number: F32HL124951 •Harvard Catalyst | The Harvard Clinical and Translational Science Center •Harvard University	
2.2	Concurrent Pneumonia in Children Under 5 Years of Age Presenting to a Diarrheal Hospital in Dhaka, Bangladesh	2015	No access to full article but CHNRI not mentioned in abstract or keywords	Supported by grants from the National Institutes of Health, including National Institute of Allergy and Infectious Diseases grants AI058935, AI100023, AI106878, AI077883 (to Edward T. Ryan), AI100923 (to Daniel T. Leung), a Thrasher Research Fund Early Career Award, and a Postdoctoral Fellowship in Tropical Infectious Diseases from the American Society of Tropical Medicine and Hygiene/Burroughs Wellcome Fund	University/Institution
2.3	Cost-Effectiveness Analysis of Pneumococcal Conjugate Vaccine in Taiwan: A Transmission Dynamic Modelling Approach	2012	No	Supported by a grant from the National Science Council (NSC95-2320-B-182-022-MY2).	Ministry
2.3	A qualitative study on knowledge, perceptions, and attitudes of mothers and health care providers toward pneumococcal conjugate vaccine in Bandung, West Java, Indonesia	2013	No	Not mentioned in paper	N/A
2.3	The effect of distance on observed mortality, childhood pneumonia and vaccine efficacy in rural Gambia	2014	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
2.3	Distance to health services affects local-level vaccine efficacy for pneumococcal conjugate vaccine (PCV) among rural Filipino children	2014	No	Not mentioned in paper	N/A
2.4	Community case management of fever due to malaria and pneumonia in children under five in Zambia: a cluster randomized controlled trial	2010	No	Funded by United States Agency for International Development through Child and Family Applied Research project Cooperative Agreement GHSA-00-00020-00 with Boston University and the President's Malaria Initiative. The funders had no role in study design, data collection and analysis, decision to	Government Agency and University/Institution

				publish, or preparation of the manuscript.	
2.4	Factors affecting availability of essential medicines among community health workers in Ethiopia, Malawi, and Rwanda: solving the last mile puzzle	2012	No	We thank the Ministry of Health leaders in all three countries for their support of the baseline studies	Ministry
2.4	Community case management of severe pneumonia with oral amoxicillin in children aged 2–59 months in Haripur district, Pakistan: a cluster randomised trial	2012	No	Funding: United States Agency for International Development	Government Agency
2.4	Measuring coverage in MNCH: a prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia	2013	No	Supported by a sub-grant "MA13 project: careseeking and coverage of child health interventions" from the Child Health Epidemiology Reference Group grant from The Bill & Melinda Gates Foundation grant no. 50140. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.	Private Foundation
2.4	Measuring coverage in MNCH: challenges in monitoring the proportion of young children with pneumonia who receive antibiotic treatment	2013	No	Conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, or preparation of the manuscript. The funders supported the decision to publish.	Private Foundation
2.5	Impact of MMRV combination vaccine on childhood vaccination compliance	2012	No	Funding Source: Kaiser Permanente Southern California received funding from Merck to conduct a phase IV post licensure safety study (completed 2009)	Private Foundation
2.5	Coverage for the entire population: tackling immunization rates and disparities in Saskatoon Health Region	2012	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A

2.5	Local discrepancies in measles vaccination opportunities: results of population-based surveys in Sub-Saharan Africa	2014	No	Not mentioned in paper	N/A
2.5	Reaching hard-to-reach individuals: Nonselective versus targeted outbreak response vaccination for measles	2014	No	Not mentioned in paper	N/A
2.5	Low measles vaccination coverage among medical residents in Marseille, France: reasons for non-vaccination, March 2013	2015	No	Funding was provided by the Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security and the Fogarty International Center, National Institutes of Health	University/Institution
2.5	Comparing the health and social protection effects of measles vaccination strategies in Ethiopia: An extended cost-effectiveness analysis	2015	No	The Bill & Melinda Gates Foundation for funding through the Disease Control Priorities Network grant to the University of Washington.	Private Foundation
2.5	Paid maternity leave and childhood vaccination uptake: Longitudinal evidence from 20 low-and-middle-income countries	2015	No	Funding for this research provided by the Canadian Institutes of Health Research (CIHR) fellowship award programme. All authors acknowledge funding from the Canadian Institutes of Health Research Operating Grant, “Examining the impact of social policies on health equity"	University/Institution
3.1,3,4	Contribution of community-based newborn health promotion to reducing inequities in healthy newborn care practices and knowledge: evidence of improvement from a three-district pilot programme in Malawi	2013	No	TG, DS, RL, EC, and EZ are employees of Save the Children. FK is employed by the Malawi Ministry of Health.	Ministry
3.1,3,4	Can community health officer-midwives effectively integrate skilled birth attendance in the community-based health planning and services programme in rural Ghana?	2014	No	The Navrongo Health Research Centre, the Ghana Health Services and the Boston University School of Public Health for their support in the research work.	University/Institution

3.1,3,4	Can she make it? Transportation barriers to accessing maternal and child health care services in rural Ghana	2015	No	Not mentioned in paper	N/A
3.1,3,4	Use of mobile phone consultations during home visits by Community Health Workers for maternal and newborn care: community experiences from Masindi and Kiryandongo districts, Uganda	2015	No	Financial support from the Institute of Tropical Medicine-Antwerp and partial funding from the African Doctoral Dissertation Fellowship (ADDRF).	University/Institution
3.1,3,4	Risk indicators for referral during labour from community midwife to gynaecologist: a prospective cohort study	2015	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
3.2	What influences the decision to undergo institutional delivery by skilled birth attendants? A cohort study in rural Andhra Pradesh, India	2012	No	Funded by the Public Health Foundation of India (PHFI), New Delhi, India.	Private foundation
3.2	Influence of birth preparedness, decision-making on location of birth and assistance by skilled birth attendants among women in south-western Uganda	2012	No	Funding for this study was made possible by Grants from Swedish International Development Cooperation Agency/SAREC and Global Health Research Initiative. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.	Government Agency
3.2	A qualitative evaluation of the choice of traditional birth attendants for maternity care in 2008 Sierra Leone: implications for universal skilled attendance at delivery	2013	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
3.2	Determinants of skilled attendance for delivery in Northwest Ethiopia: a community based nested case control study	2013	No	University of Gondar and World Health Organization for financial support	Multilateral Agency and University/Institute
3.2	Head of household education level as a factor influencing whether delivery takes place in the presence of a skilled birth attendant in Busia,	2013	No	Support of Irish Aid and the people of Ireland for funding this work through the “Access to Infant and Maternal Health Programme” (AIM-Health), a	Government Agency

	Uganda: a cross-sectional household study			collaborative initiative with World Vision Ireland and the Centre for Global Health, Trinity College Dublin	
3.2	Effectiveness of health sector reforms in reducing disparities in utilization of skilled birth attendants in Tanzania	2013	No	Funded by USAID	Government Agency
3.2	Barriers to using skilled birth attendants' services in mid- and far-western Nepal: a cross-sectional study	2013	No	Funded by a grant from WHO in Geneva. University of Gothenburg, Sweden, provided travel grants through a "Global University" grant (A11 0524/09).	Multilateral Agency
3.2	The effect of community maternal and newborn health family meetings on type of birth attendant and completeness of maternal and newborn care received during birth and the early postnatal period in rural Ethiopia	2014	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
3.2	A cluster randomized implementation trial to measure the effectiveness of an intervention package aiming to increase the utilization of skilled birth attendants by women for childbirth: study protocol	2014	No	Funded by the World Health Organization (WHO), Switzerland.	Multilateral Agency
3.2	Using the community-based health planning and services programme to promote skilled delivery in rural Ghana: socio-demographic factors that influence women utilization of skilled attendants at birth in northern Ghana	2014	No	Navrongo Health Research Centre (NHRC), the Ghana Health Services and the Boston University School of Public Health for research support.	University/Institution
3.2	Barriers to skilled birth attendance: a survey among mothers in rural Gambia	2014	No	Norwegian Research Council's Econpop Programme for funding this project.	Government Agency
3.2	Spatial analysis of skilled birth attendant utilization in Ghana	2014	No	Not mentioned in paper	N/A
3.2	Conditional cash transfer schemes in Nigeria: potential gains for maternal and child health service uptake in a national pilot programme	2014	No	CIFF provided financial and technical assistance to the CCT programme, including organizing a study tour to Mexico for CCT staff to observe and	Private Foundation

				discuss the established Oportunidades social assistance programme	
3.2	Practices and determinants of delivery by skilled birth attendants in Bangladesh	2014	No	Japan International Cooperation Agency (JICA) provided financial and technical support to the whole project.	Government Agency
3.2	Utilisation of skilled birth attendance in Northern Nigeria: a cross-sectional survey	2014	No	Not mentioned in paper	N/A
3.2	Rate of Utilization of Skilled Birth Attendant and the Influencing Factors in an Urban Myanmar Population	2015	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
3.2	Determinants of use of skilled birth attendant at delivery in Makueni, Kenya: a cross sectional study	2015	No	We wish to acknowledge the Amref Health Africa Mama na Mtoto wa Afrika MNCH project, funded by Comic Relief of UK, for granting access to the project's baseline survey conducted in August of 2012.	Private Foundation
3.5	Perinatal death audits in a peri-urban hospital in Kampala, Uganda	2012	No	Not mentioned in paper	N/A
3.5	A multi-country study of the "intrapartum stillbirth and early neonatal death indicator" in hospitals in low-resource settings	2013	No	Support for the study design meeting and study were provided by the Bill & Melinda Gates Foundation-funded Maternal Health Task Force at EngenderHealth, GAPPS and by the Eunice Kennedy Shriver NICHD Global Network for Women's and Children's Health Research grants (U01 HD042372, U01 HD040607, U01 HD040636, U01 HD058322, U01 HD058326). GAPPS (C.E.R.), MHTF (A.L., A.B.)	Private Foundation
3.5	Clinical audit to enhance safe practice of skilled birth attendants for the fetus with nuchal cord: evidence from a refugee and migrant cohort.	2014	No	Supported by the Wellcome Trust of Great Britain (Major Overseas Programme–Thailand Unit Core Grant). The Shoklo Malaria Research Unit is part of the Wellcome Trust Mahidol University Oxford Tropical Medicine Research Programme.	Private Foundation
4.1	The Baby-Friendly Hospital Initiative shows positive effects on breastfeeding indicators in Brazil	2012	No	supported by Brazilian Ministry of Health	Ministry

4.2	Impact of 4.0% chlorhexidine cord cleansing on the bacteriologic profile of the newborn umbilical stump in rural Sylhet District, Bangladesh: a community-based, cluster-randomized trial	2012	No	Funding for the Projahnmo Project is provided by the United States Agency for International Development, Office of Health, Infectious Diseases, and Nutrition, Global Health Bureau and the Dhaka Mission through the Global Research Activity Cooperative Agreement (GHS-A-00-03-00019-00), and the Saving Newborn Lives initiative of Save the Children Federation - USA through a grant from the Bill and Melinda Gates Foundation.	Government Agency and Private Foundation
4.2	Topical application of chlorhexidine to neonatal umbilical cords for prevention of omphalitis and neonatal mortality in a rural district of Pakistan: a community-based, cluster-randomised trial	2012	No	Funded by PAIMAN (Pakistan Initiative for Mothers and Newborns) and John Snow Inc, via a grant from by the US Agency for International Development	Government Agency
4.2	The effect of cord cleansing with chlorhexidine on neonatal mortality in rural Bangladesh: a community-based, cluster-randomised trial	2012	No	United States Agency for International Development and Save the Children's Saving Newborn Lives program, through a grant from the Bill & Melinda Gates Foundation.	Private Foundation
4.2	Umbilical cord antiseptics for preventing sepsis and death among newborns	2013	No	The National Institute for Health Research (NIHR) is the largest single funder of the Cochrane Pregnancy and Childbirth Group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR, NHS or the Department of Health	University/Institution
4.2	Effect of topical application of chlorhexidine for umbilical cord care in comparison with conventional dry cord care on the risk of neonatal sepsis: a randomized controlled trial	2013	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
4.2	Impact of chlorhexidine cleansing of the umbilical cord on cord separation time and neonatal mortality in comparison to dry cord	2014	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A

	care - a nursery-based randomized controlled trial.				
4.3	Improving hygiene in home deliveries in rural Ghana: how to build on current attitudes and practices	2010	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
4.3	The influence of distance and level of care on delivery place in rural Zambia: a study of linked national data in a geographic information system	2011	No	This work was done using existing data without particular funding. Funders thus had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.	None
4.3	Association between clean delivery kit use, clean delivery practices, and neonatal survival: pooled analysis of data from three sites in South Asia	2012	No	Funded primarily by the Wellcome Trust under a Strategic Award. Partner sites have received funding from the Health Foundation (UK), Women and Children First (UK), the UK Big Lottery Fund, Saving Newborn Lives, the UK Department for International Development, the United Nations Children's Fund, and the United Nations Fund for Population Activities. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.	Private Foundation
4.3	Clean delivery practices in rural northern Ghana: a qualitative study of community and provider knowledge, attitudes, and beliefs	2012	No	Not mentioned in paper	N/A
4.3	Clean home-delivery in rural Southern Tanzania: barriers, influencers, and facilitators	2013	No	Supported by a grant from the Bill & Melinda Gates Foundation through Save the Children-USA	Private Foundation
4.4	Assessing Indian public health standards for community health centers: a case study with special reference to essential newborn care services	2011	No	WHO India Country Office for providing financial support in conducting the study	Multilateral Agency
4.4	Assessment of Special Care Newborn Units in India	2011	No	Supported by funding from the UNICEF	Multilateral Agency

4.4	Assessment of Essential Newborn Care Services in Secondary-level Facilities from Two Districts of India	2014	No	Supported by Ministry of Health and Family Welfare, Government of India.	Ministry
4.5	Evaluating fidelity in home-visiting programs a qualitative analysis of 1058 home visit case notes from 105 families	2012	No	Financed with a grant from the National Ministry of Health Hospital Clinical Research Programme and the National Institute for Promotion and Health Education	Ministry
4.5	Do home visiting services received during pregnancy improve birth outcomes? Findings from Virginia PRAMS 2007-2008	2014	No access to full article but CHNRI not mentioned in abstract or keywords	Not mentioned in abstract	N/A
4.5	Effect of home visit training programme on growth and development of preterm infants: a double blind randomized controlled trial	2015	No	Financially supported by Shiraz University of Medical Sciences, Shiraz, Iran (ICR-87-4275).	University/Institution
4.5	Effects of home visiting and maternal mental health on use of the emergency department among late preterm infants	2015	No	No relevant financial relationship	None

Appendix 3 List of donors and the method to identify research priority

List of donors and the method for research prioritisation	***Terminologies***: The first threes columns are the funding body/organization, the location of the organization and the type. ***Methodology*** means whether there is a method described for research priority setting (and therefore funding allocation). If there is none detailed, 'NA' is entered. ***Transparency*** means that the method is clearly described in a relatively detailed manner on the website. ***Criteria*** means what criteria the body clearly specifies that it uses to judge said allocation. ***Funding allocation*** means whether there's a list of the amount of funding allocated to individual projects. ***The last column*** is a summary of what the website says. ***Sources*** are most of the webpages which held relevant information.						
Institution	**Location**	**Type**	**Methodology**	**Transparency**	**Listed Criteria**	**Information**	**Sources**
Gates Foundation	USA	Private foundation	Consultation	NO	strategic importance	broad strategies determined internally 'concepts developed' of areas of focus. "use a variety of ways to explore and refine concepts, with the help of organizations in the field". Funding allocated according to strategy by 3 means: Direct solicitation (to a ""well-suited" organization), Discussion (consult with 1 or more organization and invite a proposal for funding), Request for proposal (public or private notes of available funding sent out for submission of proposals).	http://www.gatesfoundation.org/How-We-Work
Wellcome Trust	UK	Charitable foundation	Peer review	NO	Novelty/strategic importance/individual's track record	Process dependent on type of grant (more systematic process for larger grants). Procedure 1: "written peer review by external expert referees... interview by the Collaborative Awards Committee" awarding decisions taken by interviewers. Procedure 2: review and decisions by Expert Review Groups	http://www.wellcome.ac.uk/Funding/Biomedical-science/Funding-schemes/Science-collabourative-awards/index.htm http://www.wellcome.ac.uk/Funding/Biomedical-science/Funding-schemes/Seed-Awards/index.htm http://www.wellcome.ac.uk/Funding/Biomedical-science/Funding-schemes/Strategic-awards-and-initiatives/WTD018098.htm
US CDC	USA	Federal agency	NA	NO	scientific merit/ability to meet programme needs	selection of successful funding on an individual case basis according to Funding Opportunity Announcements (FOA). Mention of reviewers ("i.e. Use language that can be easily understood by peer reviewers, scientists, and the public") but no explanation of process	http://www.cdc.gov/grants/interestedinapplying/applicationprocess.html#process http://www.cdc.gov/grants/interestedinapplying/eligibility.html
National Institutes of Health	USA	Federal agency	Peer review	YES (rating)	Overall Impact, Scored	First review by Scientific Review Group (SRG) of non-federal scientists, 2nd review by Institute and Center	http://grants.nih.gov/grants/peer_review_process.htm#Initial

					Review Criteria, Significance, Investigator(s), Innovation, Approach, Environment + 'additional criteria'	(IC) National Advisory Councils or Boards (scientific and public representatives)	
GAPPS	USA	Public Private	*Ongoing CHNRI exercise according to website, but no record of how current priorities set*	NO	NA	setting priorities' page talks about CHNRI and cites CHNRI document but exercise is at stage of "currently reviewing data and preparing to publish...findings". No details on how current/previous priorities have been set	http://gapps.org/research/our_approach/setting_priorities/
Sanfo-Aventis	France	Pharmaceutical Company	NA	NO	NA	"We concentrate our research efforts where the most pressing medical needs and public health issues are"… but no attempt to detail how they work out what these are	http://en.sanofi.com/rd/research_area/research_area.aspx
European Commission	Belgium	International Agency	NA	NO	in keeping with EU funding programme sections	Processes appear to be specific to funding programme. Horizon 2020: unclear how overall priorities are set but funding allocated in some form of peer review process "Experts, as peer reviewers, assist in the: evaluation of proposals". Third health: unclear	http://ec.europa.eu/programmes/horizon2020/en/experts
Roche	Switzerland	Pharmaceutical Company	NA	NO	first/best in class drugs	No details given for how R+D priorities set	
Australian NHMRC	Austrailia	Federal agency	Peer review	NO	best'/ strategy specific	Stage 1: 2 independent peer review processes ('mail review' and external review panel) , stage 2: international joint Peer Review Panel (PRP)	https://www.nhmrc.gov.au/print/book/export/html/42239
USAID	USA	Federal agency	NA	NO	NA	Global development lab' of USAID supports research but no details of how it chooses which projects to fund. Its focus areas are: Food Security and Nutrition Modernizing Food Assistance; Ending Preventable Child and Maternal Deaths; Energy Access; Water Solutions: Child Literacy; Financial Inclusion; Human Rights, Participation, and Accountability; and Humanitarian Response, but again no details on priority setting.	https://scms.usaid.gov/sites/default/files/documents/15396/PEER%20One%20Pager.pdf http://www.usaid.gov/sites/default/files/documents/1869/FY2016DevelopmentBudget_FactSheet.pdf http://www.usaid.gov/GlobalDevLab/about
Wyeth						*NO DATA:* Wyeth now Pfizer at global level (since 2009)	http://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=250312

Merck	USA	Pharmaceutical Company	NA	NO	NA	No details given for how R+D priorities set	http://www.merck.com/research/discovery-and-development/home.html
National Natural Science Foundation of China	China	Federal agency	NA	NO	NA	Programmes appear to be run individually and there is no indication of the process by which funding is allocated	http://www.nsfc.gov.cn/Portals/0/fj/english/fj/pdf/2014/021.pdf
Ministry of Health of China	China	Federal agency	*ACCESS ISSUE*- google translated page hard to understand				http://www.nhfpc.gov.cn/zhuzhan/
Burroughs Wellcome Fund	USA	Private foundation	NA	NO	young scientists/ under developed research areas	Advisory Committees of external experts review submitted proposals, interview finalists, and make recommendations for approval by the foundation Board of Directors. Same process for each type of award but with award specific experts	http://www.bwfund.org/grant-programs/biomedical-sciences/career-awards-medical-scientists/advisory-committee http://www.bwfund.org/grant-programs/biomedical-sciences/career-awards-medical-scientists/advisory-committee
Edward Mallinckrodt Jr Foundation	USA	Private foundation	NA	NO	NA	Funding awarded to individual people so person as well as project influences decision. Allocation decisions made by the board of directors at the foundation	http://www.emallinckrodtfoundation.org/Submission.html
Research Council of Norway	Norway	Federal agency	Peer review	YES (scoring)	NA	"applications will be reviewed by external expert referees, individually or in referee panels, before the final decisions regarding grant awards are taken. Allocation decisions are usually taken by the programme boards, expert committees or research board of the relevant Research Council division" criteria each award is marked against are specified in each call for proposals	http://www.forskningsradet.no/en/Application_processing/1138882215874
EFIC	Belgium	Professional organization	Panel review	NO	NA	EFIC Scientific Research Committee' reviewed proposals for grant funding- 13 people international group, unclear whether individuals on the panel are internal or external	http://www.efic.org/index.asp?sub=eSW7DhmX26HBOD http://www.efic.org/index.asp?sub=gSX7EjoX36ICOE
ETC (Netherlands)	Netherlands	INGO	NA	NO	NA	No info available on priority setting	http://www.etc-international.org/projects/health-projects/
Danish Council of Developmental research	Denmark	Federal agency				does not appear to exist anymore- various bodies represented on web page	
MRC UK	UK	Federal agency	Peer review/internal panel	YES (scoring)	Importance, potential impact,	Stage 1 – External peer review (or "triage") by specialist referees (UK and international). Stage 2- board/panel decision	http://www.mrc.ac.uk/funding/guidance-for-applicants/how-we-assess/

			assessment		Resources requested		
Premup	France	Public Private	NA	NO	NA	No info available on priority setting only disclose which particular lab they are supporting	https://www.premup.org/eng/research
Foundation Grace de Monaco	Monaco	Private foundation	NA	NO	NA	No info available on priority setting	http://www.fondation-psse-grace.mc/en/
Swedish Government	Sweden	Federal agency	Peer review	YES (rating)	novelty/scientific quality/individual merit/feasibility	800 active Swedish researchers (+ some international experts) in 90 panels review relevant proposals	http://www.vr.se/inenglish/researchfunding/assessment/peerreview.4.aad30e310abcb9735780007646.html
French Ministry of Research	France	Federal agency	Consultation	NO		Regardless of looking at French ministry of research and education or national institute of health and medical research, there is no information about priority setting	http://www.enseignementsup-recherche.gouv.fr/cid56095/biologie-et-sante.html http://english.inserm.fr/
Batchworth Trust	UK	Private foundation	NA	NO	NA	no website available for the trust	/
Belgian Directorate General of Development Cooperation	Belgium	Federal agency	NA	NO	NA	No info available on priority setting	
DFID	UK	Federal agency	NA	NO	NA	Clearly list of agency priorities for development work but no clear strategy or criteria for setting priorities for research: "We fund research that can lead to new technologies and better ways of helping the poorest people in the world"	https://www.gov.uk/government/organisations/department-for-international-development/about#priorities https://www.gov.uk/government/organisations/department-for-international-development/about/research#research-we-fund
GAVI	Switzerland	Private public	NA	NO	vaccine-development focus	no details on setting health *research* priorities. Overall organization priorities decided by small internal panel, but no specific reference to health research	file:///C:/Users/mcphersonk/Downloads/09-Strategy%20development%20process.pdf
Laerdal Foundation for Acute Medicine	Norway	Company	NA	NO	NA	No info available on priority setting	
Canadian Institutes of Health Research	Canada	Federal agency	Peer review	YES (rating)	NA	Process described- Step 1: Recruit Peer Reviewers (external experts) Step 2: Assign Applications	http://www.cihr-irsc.gc.ca/e/48958.html

						Step 3: Evaluate Applications Remotely (based on grant-specific criteria) Step 4: Finalize Evaluations and Ratings Step 5: Make Funding Decisions- final recommendations made by Chief Scientific Officer and Chief Financial Officer to be given final approval by the Science Council	http://www.cihr-irsc.gc.ca/e/39410.html
French Development Agency	France	DFI	NA	NO	NA	No info available on priority setting, again, far more information available about intervention than research aspect	http://www.afd.fr/lang/en/home/recherche
Swedish Research Council	Sweden	Federal agency	Peer review	YES (weighted scale)	Novelty, originality, scientific quality, merits	500 Swedish and foreign researchers as part of expert evaluation panel which peer review grant applications. Not clear who makes final decision	http://www.vr.se/inenglish/researchfunding/assessment.4.45a6e939122880e7d8e80002028.html
Meltzer Foundation in Bergen	Norway	University	Peer review	NO	NA	No info available on priority setting	http://meltzerfondet.no/hovedsiden/english/

Appendix 4 Ethical approval

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Miss Sachiyo Yoshida
LSHTM

26 February 2018

Dear Sachiyo

Study Title: CHNRI- an assessment of collective opinion

LSHTM Ethics Ref: 11807

Thank you for responding to the Observational Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered by the Committee.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Investigator CV	CV_Sachiyo_Yoshida_short	30/05/2016	1.0
Protocol / Proposal	Protocol_Assessment_Collective_Opinion_SC	21/07/2016	v1
Investigator CV	CV0214	21/07/2016	1.0
Investigator CV	CV-Igor-eng17-21052015	21/07/2016	1.0
Covering Letter	Cover_letter_response_for_questions_for_clarification_1	19/09/2017	v1
Covering Letter	Cover_letter_response_for_questions_for_clarification_4	12/02/2018	4
Covering Letter	Invitation_letter	12/02/2018	1
Covering Letter	Paper_2.1_NB_lancet_commentary	12/02/2018	NA

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: https://leo.lshtm.ac.uk

Additional information is available at: www.lshtm.ac.uk/ethics

Yours sincerely,

[redacted]

Professor John DH Porter
Chair

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

University of Edinburgh,
Centre for Population Health Sciences
RESEARCH ETHICS SUBGROUP

Self-Audit Checklist for Level 1 Ethical Review for Staff/research fellow projects

To be completed by the project PI

See **Intra** website for further information: http://www.cphs.mvm.ed.ac.uk/intra/research/ethicalReview.php

Proposed Project *(State research question and topic area, and briefly describe method/ data. Specify also countries in which data will be collected.):*

Project: "Comparison of collective knowledge versus individual knowledge in responding to general knowledge questions of increasing degree of difficulty among Year 2 medical students"

Topic area: Decision science; Crowd-sourcing.

During a series of 6 lectures in Year 2 "Epidemiology and statistics" at the University of Edinburgh, a few minutes after each lecture will be used to run a pre-prepared power-point presentation. Each slide will be asking the entire group of 200-250 students general knowledge questions of increasing level of difficulty, where a correct response will always be an integer (a number). The students will be asked to simply enter their best guess of the number that they think is as close as possible to a correct answer. An example is: "*How many marriages did the actress Elizabeth Taylor have?*"; *"How old was the composer W. A. Mozart when he died (in years)?*"; or "*How many minutes does the movie "Casablanca" last?*" The students willing to take part will stay on and write their responses on a blank piece of paper as they watch the presentation.

The purpose of the exercise is to analyse their numerical responses and investigate whether the mean or a median of the whole group's response outperforms individual responses/guesses, and under what conditions. Students will write their responses on a blank sheet of paper, and they will not, in any way, be individually identified, or asked to write anything else other than their best guesses of a number that is a correct answer. The project is being conducted to examine validity of "crowd-sourcing" approaches to decision-making. Students will be explained that it is their right not to participate in the exercise, and that they should only stay on after the lecture if they are doing it willingly and with full consent, in order to support research.

1. Bringing the University into disrepute

Is there any aspect of the proposed research which might bring the University into disrepute? YES/ **NO**

2. Data protection and consent

Are there any issues of DATA PROTECTION or CONSENT which are NOT adequately dealt with via established procedures? YES/ **NO**

These include well-established sets of undertakings. For example, a 'No' answer is justified *only if*:

(a) There is compliance with the University of Edinburgh's Data Protection procedures (see www.recordsmanagement.ed.ac.uk);

(b) Respondents give consent regarding the collection, storage and, if appropriate, archiving and destruction of data;

(c) Identifying information (eg consent forms) is held separately from data;

(d) There is Caldicott Guardian approval for (or approval will be obtained prior to) obtaining/ analysing NHS patient-data.

(e) There are no other special issues arising about confidentiality/consent.

3. Study participants

a) Will a study researcher be in direct contact with participants to collect data, whether face-to-face, or by telephone, electronic means or post, or by observation? (eg interviews, focus groups, questionnaires, assessments) YES/ **NO**

b) **Answer this only if qu. 3 above = 'YES':**
*In ethical terms, could any participants in the research be considered to be '**vulnerable**'?*
e.g. children & young people under age of 16, people who are in custody or care (incl. school), a marginalised/stigmatised group
Please tick one: 'vulnerable' ☐ not 'vulnerable' ☐

4. Moral issues and Researcher/Institutional Conflicts of Interest

Are there any SPECIAL MORAL ISSUES/CONFLICTS OF INTEREST? YES/ **NO**

(a) An example of conflict of interest for a researcher would be a financial or non-financial benefit for him/herself or for a relative of friend.

(b) Particular moral issues or concerns could arise, for example where the purposes of research are concealed, where respondents are unable to provide informed consent, or where research findings could impinge negatively/ differentially upon the interests of participants.

(c) Where there is a dual relationship between researcher and participant (eg where research is undertaken by practitioners so that the participant might be unclear as to the distinction between 'care' and research)

5. Protection of research subject confidentiality

Are there any issues of CONFIDENTIALITY which are NOT adequately handled by normal tenets of confidentiality for academic research? YES/ **NO**

These include well-established sets of undertakings that should be agreed with collaborating and participating individuals/organisations. For example, a 'No' answer is justified *only if*:

(a) There will be no attribution of individual responses;

(b) Individuals (and, where appropriate, organisations) are anonymised in stored data, publications and presentation;

(c) There has been specific agreement with respondents regarding feedback to collaborators and publication.

6. Potential physical or psychological harm, discomfort or stress

(a) Is there a FORSEEABLE POTENTIAL for PSYCHOLOGICAL HARM or STRESS for participants? YES/ **NO**

(b) Is there a FORSEEABLE POTENTIAL for PHYSICAL HARM or DISCOMFORT for participants? YES/ **NO**

(c) Is there a FORSEEABLE RISK to the researcher? YES/ **NO**

Examples of issues/ topics that have the potential to cause psychological harm, discomfort or distress and should lead you to answer 'yes' to this question include, but are not limited to:

relationship breakdown; bullying; bereavement; mental health difficulties; trauma / PTSD; violence or sexual violence; physical, sexual or emotional abuse in either children or adults.

7. Duty to disseminate research findings

Are there issues which will prevent all relevant stakeholders having access to a clear, understandable and accurate summary of the research findings if they wish?* YES/ **NO**

** If, and only if, you answered 'yes' to 3 above, 'stakeholders' includes the participants in the research*

Overall assessment

- If every answer above is a definite NO, the self-audit has been conducted and confirms the ***ABSENCE OF REASONABLY FORESEEABLE ETHICAL RISKS*** – *please tick box* ☐

  *This means that regarding this study, as currently self-audited, no further ethical review actions are required within CPHS. However, if in the coming weeks/months there is any change to the research plan envisaged now (and outlined above), the study should be **re-audited** against a Level 1 form, because it may be that the change made negates the absence of ethical risks signed off here.*

- If one or more answers are YES, then risks have been identified and prior to commencing any data collection **formal ethical review** is required - either:
  - by NHS REC (NB copy of ethics application and decision letter to be sent to CPHS Ethics); or
  - if not to be formally reviewed by NHS REC, then CPHS level 2/3 ethical review required.

  [If either 4 is 'yes' or 3b is 'vulnerable' then it is possible level 3 review is required.]

The completed and signed form should be returned to the CPHS Ethics administrator.

Professor Igor Rudan
PI name

[redacted]
PI Signature *

*** NOTE:** *The CPHS Ethics Subgroup cannot check validity of responses made on this form (the light touch Level 1 form means we have insufficient detail to do so). By countersigning this check-list as truly warranting all 'No' answers, **you** are taking responsibility, on behalf of CPHS and UoE, that the research proposed truly poses no ethical risks. Therefore, if there is any doubt on any issue, it would be a wise precaution to mark it as 'uncertain' and contact the Ethics Subgroup as to whether a level 2 form might be required as well. (See **Intra web-site at** URL given above.)*

25 March 2014