BMJ Global Health

Bridging research integrity and global health epidemiology (BRIDGE) guidelines: explanation and elaboration

Sandra Alba ,[1] Annick Lenglet,[2] Kristien Verdonck,[3] Johanna Roth,[4] Rutuja Patil ,[5] Walter Mendoza,[6] Sanjay Juvekar,[5] Susan F Rumisha[7,8]

To cite: Alba S, Lenglet A, Verdonck K, *et al*. Bridging research integrity and global health epidemiology (BRIDGE) guidelines: explanation and elaboration. *BMJ Global Health* 2020;**5**:e003237. doi:10.1136/bmjgh-2020-003237

Handling editor Seye Abimbola

Received 24 June 2020
Revised 30 August 2020
Accepted 1 September 2020

[1]Health, KIT Royal Tropical Insititute, Amsterdam, The Netherlands
[2]Médecins Sans Frontières, Amsterdam, North Holland, The Netherlands
[3]Tropical diseases, Institute of Tropical Medicine, Antwerp, Belgium
[4]European and Developing Countries Clinical Trials Partnership, The Hague, South Holland, The Netherlands
[5]Vadu Rural Health Program, KEM Hospital Research Centre, Pune, Maharashtra, India
[6]United Nations Population Fund, Lima, Peru
[7]National Institute for Medical Research, Dar es Salaam, United Republic of Tanzania
[8]Big Data Institute, University of Oxford, Oxford, UK

Correspondence to
Dr Sandra Alba; s.alba@kit.nl

ABSTRACT

Over the past decade, two movements have profoundly changed the environment in which global health epidemiologists work: research integrity and research fairness. Both ought to be equally nurtured by global health epidemiologists who aim to produce high quality impactful research. Yet bridging between these two aspirations can lead to practical and ethical dilemmas. In the light of these reflections we have proposed the BRIDGE guidelines for the conduct of fair global health epidemiology, targeted at stakeholders involved in the commissioning, conduct, appraisal and publication of global health research. The guidelines follow the conduct of a study chronologically from the early stages of study preparation until the dissemination and communication of findings. They can be used as a checklist by research teams, funders and other stakeholders to ensure that a study is conducted in line with both research integrity and research fairness principles. In this paper we offer a detailed explanation for each item of the BRIDGE guidelines. We have focused on practical implementation issues, making this document most of interest to those who are actually conducting the epidemiological work.

Summary box

- Over the past decade, two movements have profoundly changed the environment in which global health epidemiologists work: research integrity and research fairness.
- Both ought to be equally nurtured by global health epidemiologists who aim to produce high-quality impactful research, yet bridging between these two aspirations can lead to practical and ethical dilemmas.
- In the light of these reflections, we have proposed the BRIDGE guidelines for the conduct of fair global health epidemiology, targeted at stakeholders involved in the commissioning, conduct, appraisal and publication of global health research.
- The guidelines follow the conduct of a study chronologically from the early stages of study preparation until the dissemination and communication of findings.
- They can be used as a checklist by research teams, funders and other stakeholders to ensure that a study is conducted in line with both research integrity and research fairness principles.
- In this paper, we offer a detailed explanation for each item of the BRIDGE guidelines.
- We have focused on practical implementation issues, making this document most of interest to those who are actually conducting the epidemiological work.

INTRODUCTION

Over the past decade, two movements have profoundly changed the environment in which global health epidemiologists work: research integrity and research fairness. On one hand, questionable research practices may lead to spurious findings if studies are ill-designed, poorly implemented, inappropriately analysed or selectively reported. On the other hand, local communities, institutions and researchers are too often side-lined from the formulation of research questions, the design and implementation of studies and the dissemination of findings. Taking advantage of weak or inexistent ethics institutions, bypassing local expert knowledge, ignoring local context, failing to develop in-country capacity are some of the practices which de-value global health epidemiology.

The BRIDGE statement

As we have argued in the BRIDGE statement paper,[1] research integrity and research fairness need to be equally nurtured by global health epidemiologists who aim to produce high-quality impactful research. Yet bridging between these two aspirations can lead to practical and ethical dilemmas. In the light of these reflections, we have proposed guidelines for the conduct of fair global health epidemiology, targeted at stakeholders involved in the commissioning, conduct, appraisal and publication of global health research.

The BRIDGE guidelines were developed by a Delphi consensus with global health practitioners from over 20 countries in 5

continents. Our aim was to bring together existing principles in one overarching guideline, with a focus on practical implications for global health practitioners. The outcome consists of a set of 6 standards and 42 accompanying criteria covering the following steps of a study: (1) study preparation; (2) study protocol and ethical review; (3) data collection; (4) data management; (5) analysis; (6) dissemination and communication.

How to use this paper

This paper is linked to the BRIDGE statement paper [1] that introduced the BRIDGE guidelines and described the justification and methodology for their development. The guidelines follow the conduct of a study chronologically from the early stages of study preparation until the dissemination and communication of findings. They can be used as a checklist by research teams, funders and other stakeholders to ensure that a study is conducted in line with both research integrity and research fairness principles. In this paper, we offer a detailed explanation for each item of the BRIDGE guidelines. We have focused on practical implementation issues, making this document most of interest to those who are actually conducting the epidemiological work. This document is not necessarily meant to be read linearly from start to finish, but should rather serve as a source of further reading for readers interested in more in-depth discussion and justification for each item. A glossary can be found in online supplemental file 1 for all terms that are underlined.

THE ITEMS

Standard 1. Study preparation: carefully prepare the study, in partnership with local researchers, by taking into account existing knowledge and resources and engaging with key stakeholders

1.1. Plan and execute research in partnership with local researchers. When working in a setting where relevant epidemiological competences are limited or not available, consider what is in the study team's remit to strengthen local capacity

Global health research is rarely conducted by an organisation in isolation, but is the result of collaboration across different disciplines, expertise and countries. This often translates in research partnerships between institutes or organisations from high-income and low-income settings. These partnerships should always be established in a way that is highly advantageous for both parties.[2] Fair epidemiological research means that the local relevance of the research should be determined in collaboration with local partners.[2 3] Fair research partnerships also entail transparent and open communication between parties all throughout the research process from early planning stages to the communication of findings. On one hand, it is important that local human resources for health are not depleted to provide staff the research project (eg, nurses or laboratory staff).[3] On the other hand, lack of existing local capacity should not be viewed as a reason to forego such partnerships. Rather, when working in a setting where relevant epidemiological competences are limited or not available, epidemiologists from high-income countries should consider what is in the study team's remit to strengthen local capacity in order to meaningfully engage local researchers. This extent of this capacity strengthening should be commensurate to the scope of the research project and match the current professional needs and ambitions of people involved locally. Capacity strengthening activities may include, but are not limited to, establishing and/or strengthening ethics review committees, strengthening research capacity, developing relevant technologies, training of research or healthcare staff and education of the community involved in the research. These activities may be extended to more specialised domains of epidemiology.

1.2 Identify and engage key stakeholders throughout the study with approaches based on their needs, competences and expectations. Key stakeholders include representatives of affected populations and end-users of research

A stakeholder is anyone who has a 'stake' or an 'interest' in a particular initiative. In global health research, stakeholders may include include: the members of community where research was conducted (the affected populations), the community at large (at national or global level), local implementers (eg, local government or healthcare workers), national policymakers and policy implementers from governmental and non-governmental organisations, the scientific community who can benefit from the research, drivers of the international policy including bilateral and multilateral agencies. Key stakeholders in a global health research study include the representatives of the affected populations and the end-users of the research.

Stakeholders are increasingly claiming their right to be 'engaged'—that is, informed, consulted and involved—in the decision-making processes of research which affect them. Identifying the relevant stakeholders will therefore be the first step of a research project. Depending on the scale of the research, this can be done fairly quickly and informally, in groups using a participatory approach, or more rigorously using structured approaches for stakeholder mapping.[4] The method of engagement should be selected to best meet the needs, capacity and expectations of the relevant stakeholders as well as the strength of engagement sought, which can range from: (1) remain passive; (2) monitor; (3) advocate; (4) inform; (5) transact; (6) consult; (7) negotiate; (8) involve; (9) collaborate; (10) empower.[5]

1.3 Establish the knowledge gap by searching the literature (peer-reviewed publications and grey literature) as well as by consulting (local) experts, representatives of affected populations and end-users

A systematic literature review provides a complete, exhaustive summary and appraisal of current literature on a specific topic. While this is recommended whenever possible, there may not always be time and resources

available for such an exercise. In such cases, a literature review which thoroughly summarises the topic can suffice. A (systematic) literature may show that there is no knowledge gap to be filled and that the study is redundant. Alternatively, it may uncover useful sources of published information, which can form the basis of an analysis without the need for any new data collection—with all costs and burden to participants avoided. Even when new data collection remains necessary, experience in related studies may guide the design or indicate pitfalls to be avoided.

Depending on the complexity of the study and the amount of information already available on the topic to be studied, an exploratory needs assessment with key stakeholders may be warranted. Such an exercise can also improve the understanding of the research topic by different stakeholders (community, health facility staff, local administration, central ministry, other governmental bodies, donors, etc) and point towards the disciplines that need to be included in the study. Affected populations should also be consulted to ensure that their perspectives are fairly represented.[2 3 6]

1.4 Develop research questions and objectives in consultation with research partners and expected end-users

Research questions should be jointly formulated by all research partners involved.[2] The most relevant research questions are those which address the specific local issues. End-users should therefore be consulted early on in the study design to ensure that the research questions respond to their information needs. This can help ensure that the proposed research is in line with existing national research agenda or priorities.[7]

1.5 Select study design and research methods to best fulfil the study objectives and give due consideration to multidisciplinary approaches

Before embarking on any global health epidemiological study, researchers should consider whether they have incorporated the right disciplines to answer the proposed research questions. Global health research questions are often complex and multilayered as the issues at hand often involve many stakeholders. This requires the collaborations between disciplines beyond biomedical sciences.[8] As a result, in global health, epidemiological studies are, more often than not, conducted along-side or integrated with other quantitative (eg, economics and mathematical modelling, machine learning) or qualitative disciplines (eg, anthropology, sociology, political sciences). Multidisciplinary research is well suited to study multiple types of outcomes and provides a holistic understanding of causal pathways. While quantitative methods quantify change over time and associations (along with an estimate of the role of chance), qualitative methods are most suited to understand people's judgements, perceptions and preferences therefore providing insights into reasons behind changes or associations, or lack thereof.

1.6 Before embarking on primary data collection, assess whether existing data could be used, fully or partly, to fulfil the research objectives

During the planning stage, it is important to reflect on the need for primary data collection. What level of investment is permissible and justified and to what extent can (re-)analysis of existing quantitative data and an appropriate mix of qualitative and quantitative methods address the knowledge gap? These considerations need to consider how to make best use of funds available while not burdening people with unnecessary data collection. The past decade has seen a huge increase in the amount of publicly available data for research which offers tangible opportunities to forgo primary data collection in favour of secondary analyses of existing data. First, as discussed in criterion 6.6, open data sharing initiates have resulted in numerous repositories where data can be accessed for re-analyses. Second, many nationally representative health surveys are available for re-analysis, thanks to the efforts of organisations such as UNICEF and the United States Agency for International Development who have made many datasets from their Multiple Indicator Cluster Surveys (https://mics.unicef.org/) and Demographic and Health Surveys (https://dhsprogram.com/) openly accessible online. Lastly, health service data are also increasingly available to global health researchers, as health management and information systems data are digitalised through the DHIS2 platform (https://www.dhis2.org/) and other similar efforts.

1.7 Ensure data ownership and publication agreements have been agreed by all research partners

Agreements for data ownership, storage and access should be made during the preparatory phase of study. Data sharing agreements detail the understanding between the data provider and data receiver with regard to what data are shared and associated conditions of use. Within the frame of research, they should include provisions on the right to publish results. As mentioned in criterion 1.6, global health research often entails the re-analysis of health service data. In such cases, it is advisable that those who share the data also request researchers' compliance with the terms of a bespoke data-sharing agreement.

Much global health research is conducted in academia where peer-reviewed scientific publications remain the primary metric for career progression. This in itself is the source of much research unfairness, and the frequent (conscious or unconscious) bypassing of local researchers in the preparation of scientific publications. To counter that, it is important that fair agreements are made early on in the research process, considering the professional development of all partners involved equally. While it is very difficult to agree authorships before the work really gets going (because one rarely knows exactly how much each potential author will contribute), the principles of authorship

can still be agreed during preparatory stages. The International Committee of Medical Journal Editors provide guidance on authorship[9] and are endorsed by the vast majority of scientific journals. Yet they have a heavy emphasis on actual writing of the manuscript and it has been argued that this systematically disadvantages researchers from low-income and middle-income countries in global health research partnerships.[10]

1.8 Agree on work plans and governance structures with all study partners. Allocate adequate time, financial and human resources to all phases of the study

It is important that decision-making processes are clarified before the actual research starts. The roles and responsibilities of all parties involved should be transparently and fairly agreed in writing, so that study team members are on the same page in terms of expectations and contribution. This should including decision making in the event of disagreement. The RACI models are a useful way to do this, focusing on the four responsibilities most typically used: responsible, accountable, consulted and informed.[11] In larger multistakeholder studies, oversight bodies may be needed to advise and oversee the study conduct.

Clear study plans should be developed to ensure that adequate time, financial and human resources are available for all phases of the study. All team members should have a valid role and adequate resources in the project to fulfil that role. Trained and experienced local health professionals may possess the perfect skills mix for a given research position but their recruitment needs to be balanced with the potential health system weakening risk of depleting the local human resources for health. While it should go without saying that local research teams should be fairly remunerated for their contribution,[3] it is important to pay special attention to the working conditions of those in the lower echelons of the research hierarchy. This includes field staff (eg, interviewers, supervisors and field data editors). Mitigating the precarious nature of their freelance casual labour should be considered where possible, ideally with mid-term to long-term solutions such as long-term contracts and opportunities for career progression. This can take the form of online courses and qualifications that can be embedded into their roles within the research team, depending on their needs and aspirations. Long-term contracts for field staff should also consider employment benefits such as health insurance and pension planning (as appropriate for the context) and budget for these in the financial planning and budget plans for a study.

Precise budget estimates are not always part of a grant application process, but careful estimation of the different costs during the application stage is beneficial for the practical implementation of a research project. It is important to take currency fluctuations, inflation and uncertainty into account when preparing the budget, especially in fragile and conflicted affect settings.

Standard 2. Protocol development: prepare a detailed research protocol and ensure it has been approved by relevant ethical review boards if it includes research concerning human participants

2.1 Prepare a detailed research protocol in consultation with all research partners

The study protocol describes in detail all steps of a proposed study. The two primary purposes are funding acquisition or ethical approval. A number of templates are available to guide protocol writing.[12 13] While the protocol writing may be led by one party in research partnerships, it is important that all parties are engaged and given a fair opportunity to contribute. All parties should explicitly agree with all their roles and responsibilities in the protocol.

2.2 Write a clear and comprehensive analysis section

The study protocol may provide an overview of the planned analyses by describing the purpose of the study, the primary hypotheses, the design and the source population and a general description of the chosen analytical strategy. Depending on the complexity of the analyses, it may be advisable to write a stand-alone analysis plan. The purpose of a statistical analysis plan (SAP) is to ensure transparency and to minimise type I and type II errors resulting from the analysis strategy (eg, multiple testing, choice of confounders, etc) thus affecting inferential reproducibility. A minimum set of items to be included in a SAP for randomised clinical trials is available.[14] This is not yet the case for observational studies, but a recent paper has been published suggesting a modification of the recommended SAP format for clinical trials to fit observational studies.[15]

Broadly, the suggested items to cover in a statistical analysis plan include:

1. *Primary and secondary research questions and hypotheses*, as well as details of the primary and secondary outcome measures, and how these relate to the study objectives.
2. *Sampling procedure and recruitment/retention methods*, detailing the sampling method, the planned recruitment rate, the likely rate of loss to follow-up, interim analyses and stopping guidance (where applicable).
3. *Sample size justification*, including a description of the power and sample size calculations detailing the outcome measures on which these have been based, as well as any assumptions made underlying the power calculation and justification for these assumptions.
4. *Considerations about multiple testing*, explaining how false positive findings as a result of repeated subgroup analyses will be minimised.
5. *Potential confounders and effect modifiers* should be defined and approaches on how to address the effect of confounders and effect modifiers specified.
6. *Analysis strategy*, describing how results of this study will be analysed, including the use of statistical and/or mathematical models.

2.3 Consider studying the effect of locally relevant equity dimensions

With its focus on 'achieving equity for all people', global health acknowledges that social determinants have a major impact on health.[16–19] From an epidemiological perspective, this implies disaggregating analyses in order to reveal patterns that may be masked by aggregate data. Factors that may affect health opportunities and outcomes include place of residence, race, ethnicity, culture, language, occupation, gender/sex, religion, education, socioeconomic status and social capital—as described by the PROGRESS acronym and framework.[20] Sex/gender have been the subject of much attention[21–27] as gender is known to 'intersect'[28] with other social determinants, creating interdependent systems of disadvantage.[27] While intersectionality originates from gender studies, it is increasingly being proposed as a framework to study health equity in public health.[29–31]

There are a number of practical and statistical considerations related to studying equity in global health epidemiology. First, a thorough understanding of the local context is crucial to identify the relevant equity dimensions. Second, researchers need to be very mindful of the causal mechanisms they intend to study when using equity variables and cognisant of the potential for spurious results when using proxy variables.[32] Race and sex/gender are particularly challenging ones, with both biological (hereditary, genetic) and social (differentials in access to care) mechanisms. This further emphasises the importance of working within multidisciplinary frameworks (with either biological or sciences or both in these examples) to ensure a comprehensive understanding of the issues at hand (ref statement paper). Third, the choice of equity dimensions will have many practical implications for study design, ranging from sample size (more equity dimensions usually means more confounders and interactions[33 34] and therefore a larger sample size), sampling procedures (choice of sampling frame and definition of inclusion and exclusion criteria), research instruments and field procedures (which should be culturally appropriate and safe as described in criteria 3.2, 3.4 and 3.6).

2.4 When conducting multidisciplinary research, describe the purpose and strategies to integrate different analytical methods in the protocol

As described in criterion 1.5, addressing today's global health challenges frequently requires the involvement of different scientific disciplines, including but not limited to medicine, epidemiology, social sciences, economics and environmental sciences. Protocol writing is a team effort which requires the expertise of all disciplines involved.[35 36] Multidisciplinary research protocols should describe the purpose of combining different disciplines and also include strategies to integrate relevant qualitative and quantitative analytical methods.[37] Multidisciplinary research is typically conducted through several iterations of analyses—where analyses within one discipline are initially conducted independently *from* and then dependent *on* each other. To maximise the success of a multidisciplinary approach, study plans need to include regular moments of reflection with peers from the other disciplines throughout all study phases (especially in design, analysis and interpretation).[37 38]

2.5 Strive to make study protocols publicly available, either on a publicly accessible website or in appropriate study registers

Public availability of research protocols is one of the cornerstones of research integrity as it helps prevent post hoc revisions of study aims. Protocols can be either placed in a publicly accessible website or uploaded in an appropriate studies register.[39 40] An increasing number of journals now also offer the possibility to publish protocols, with the guarantee that study results will be published regardless of whether they show 'positive' or 'negative' results. This option is that it also enables a peer review of the research protocol.

2.6 For all data collection and data use concerning human participants, obtain ethical approval (or a waiver) ideally from all institutions and countries involved in the protocol. In case of multiple review and disagreement, the review of the country where the data are collected should take precedence

It is not always easy to determine whether a study needs ethical review as the boundary between research and public health practice can be blurred. Indeed, a recent review of ethical guidelines for epidemiology has shown that not all epidemiological or public health studies require an ethics review.[41] As a general rule, all studies involving primary data collection from human participants need to be reviewed ethically and scientifically by a competent and independent research ethics committees (REC) prior to the start of data collection.[42 43] Studies which perform secondary analysis of existing data may also require ethical review if the analyses fall outside the scope of the informed consent provided (or if no informed consent was provided).[43] Ethical review includes a thorough review of informed consent forms—ideally in the language in which they will be administered. Guidance for the formulation of informed consent forms can be found in the updated CIOMS 2017 guidelines.[43] While each REC may have their own templates, generic templates are also available.[12]

The latest CIOMS Ethical Guidelines for Health-related Research involving Humans ask for dual ethical approval for studies conducted by partnerships involving high-income and low-income and middle-income countries 'at the site of the sponsor as well as locally'.[43] The intention is to prevent 'ethics dumping'[3]—that is, the export of unethical research practices from high-income to low-income settings. However, it can also be argued that insisting on dual review perpetuates colonial notions that REC in low-income countries cannot be relied on. Certainly, 'researchers from high-income settings should show respect to host country REC'[3] and 'research projects should be approved by a REC in the host country, wherever this exists, even if ethics approval has already been obtained in the high-income setting'.[3] Difficulties can arise when ethical review is not possible at one site

(for lack of local capacity or willingness to review a study conducted in a foreign country) or if reviews conflict with each other. As a general rule, the review of the country where the research is conducted should take precedence.

The ethical review of research conducted in humanitarian emergencies deserves special attention here. In such settings, there is an intrinsic clash between ethical priorities: the research needs to be done swiftly, and participants are particularly vulnerable. A recent review suggests two useful strategies in such settings[44]: (1) pre-approved research protocol templates which can be quickly customised for use in individual emergencies[45] and (2) 'real-time responsiveness', which is an iterative strategy of constant dialogue between ethics reviewers and researchers while studies are being conducted.[46]

2.7 When working in a setting without ethical review boards or review boards with limited epidemiological capacity, consider what is in the study team's remit to strengthen their epidemiological capacity

Epidemiological studies may take place in countries with insufficient capacity to assess the ethical aspects and/or scientific quality of the research. Adequate capacity to conduct and review biomedical research does not automatically translate into the same for epidemiological and multidisciplinary projects, and this should therefore be regarded as a specific need. Taking advantage of this situation is one of the worst forms of 'ethics dumping'.[3] Instead, epidemiologists should consider what is in the study team's remit to strengthen their epidemiological capacity as part of broader capacity strengthen efforts (as described in criterion 1.1).

2.8 Explicitly state any open data access in the protocol submitted for ethical review and in the informed consent documents

Funding bodies and publishers increasingly encourage public data sharing to maximise the return of investment on research, to increase transparency and accountability, to reduce the cost of duplicating data collection and to promote potential new data uses.[47] Depending on the type of study and data collected, informed consent forms may include conditions of use and provisions for sharing with third parties. Any data sharing with third parties (whether fully open access or not) should be included in the protocol and informed consent documents. For collation of existing data to be used for secondary analyses, sharing with third parties should be agreed with data owners. The protocol should describe plans to publish data in online open access repositories (see criterion 6.6).

Standard 3. Data collection: use valid and reliable instruments and reproducible methods while ensuring culturally appropriate procedures

3.1. Use valid and reliable research instruments

Global health research relies on diverse types of data. Primary data are obtained by direct measurement using research instruments such as questionnaires, data extraction forms, interview guides, assessment by clinicians, laboratory and imaging techniques and global positioning system (GPS) and other devices. Studies can also rely on secondary analysis of existing data including health registries, routine operational information, weather and climate data, satellite information and census data.

Research instruments should be valid and reliable.[48] The development of research instruments requires skill. The design of, for example, a questionnaire is an iterative process in which the following steps can be distinguished: (1) definition and elaboration of the construct; (2) choice of measurement method; (3) selecting and formulating items; (4) scoring issues; (5) pilot testing; (6) field testing.[49] This process relies on scientific literature, theory, empirical evidence and statistical techniques. Before developing a new questionnaire, researchers should perform a review of existing instruments and their properties. If an instrument already exists, using it saves time and makes results comparable to other studies. The choice of research instruments remains a domain full of trade-offs, and reducing the risk of biases and error requires considerable efforts, which have to be delivered within time and budget constraints.[50]

Data collection modes have evolved over the past decades.[50] In the domain of surveys, for example, electronic methods are increasingly used, as a replacement of or in combination with face-to-face and telephone interviews. The advantages of computer-assisted methods include flexibility, reduced chance of error and possibly also of missing data, user-friendliness and time saving. This evolution also poses practical challenges (data capture design, data conversion, availability of internet, cost, training of field workers) as well as theoretical ones (unknown errors and biases resulting from new data collection modes).[50]

3.2. Ensure that research instruments are locally adapted and culturally appropriate

Global health epidemiologists often study a range different communities and countries. Researchers must be cognisant of local cultural sensitivities and should be careful not to violate customary practices with their data collection procedures.[3 51] In practice, time consuming, invasive or culturally insensitive data collection procedures can lead to non-response biases and measurement errors. It is well known that questions about sensitive topics, such as sexual practices, deaths or religious ideas can be difficult to handle for participants and data collectors.[52] It is less obvious but equally important to consider that apparently harmless topics (eg, questions about food consumption) may also embarrass or upset informants.[52] This further emphasises the importance of including (local) investigators with relevant skills who are experienced in dealing with such circumstances.

3.3. Provide concrete guidance for data collection in a document that is available to all data collection staff

Standardising data collection processes helps to ensure that instruments maintain their validity and reliability[48]

and contributes to methods reproducibility.[53] In general, quantitative measurements are easier to standardise than qualitative judgements. Standard operating procedures (SOPs) and job aids can help ensure uniformity for various procedures (inclusion and examination of study participants, collection and storage of specimens for the laboratory, laboratory assays, data management and quality assurance).[54] All guidance documents for data collection (field manual) should be developed with care so that they are legible, readable and comprehensible.[54] Generic templates are available for several types of SOPs.[54] All data collection guidance tools should be available whenever and wherever the people involved in data collection need them.

3.4. Select data collection staff according to technical as well as cultural criteria. Clarify the roles and responsibilities for each person involved and provide adequate training and support

In small studies, the lead researchers may be able to interview or examine all participants. But when there are many study participants or when there are sociocultural or linguistic barriers, field workers (intermediary research assistants) may be needed. Depending on the scale of the study, a hierarchy of field staff including interviewers, supervisors and field data editors may have to be recruited. Many global health research projects are highly dependent on field workers. These fieldworkers may be the only people who directly engage with the study participants, hence need to be well trained and oriented to understand the study objectives, ethical issues and the instruments used. Their influence on informed consent and data collection processes should not be underestimated.[52]

3.5. Pilot test, and if possible, field test all research instruments prior to the start of effective data collection

Pilot testing and field testing is recommended, regardless of the choice for an existing or a new research instrument. Pilot testing is intended to test the comprehensibility, relevance, acceptability and feasibility of the questionnaire in a small number of respondents, after which adaptations will follow. A pilot on the target population is crucial as only they can judge the comprehensibility and relevance of the questionnaire. In a pilot, after participants have answered all questions they should be asked about their experience in as much detail as necessary to enable changes.[49] When an instrument is considered to be satisfactory, it can be applied to a larger sample of the target population. Whereas pilot testing entails an intensive qualitative analysis of the formulation of questions and the layout of the questionnaire, field testing entails quantitative analyses. As such, all data management steps are also included in field testing. Possible analyses include: patterns of missing items (did respondents not understand the question? Do their answers not fit the response options?) and distribution of item responses (if some categories are seldom used, then can be combined with others).[49]

In practice, despite their clear usefulness, pilot and filed tests remains problematic in epidemiology. Ideally, pilot and field testing should be integrated in grant applications and study timelines, but that is not always be possible, as many research funders do not support these financially (in terms of budget lines) and logistically (in terms of the time investment). Unfortunately, researchers may even find themselves in stalemate situation when preparing funding applications for large-scale studies, as funders (and particularly also external reviewers of these funders) may request to see pilot data before granting their funding.

3.6.Collect data in a respectful and safe manner, in an environment which safeguards the confidentiality of respondents

When data collection is prepared and field workers are selected and trained, it is important to focus exclusively on technical aspects of using the research instruments and to reflect on how study participants and field workers can be protected from harm due to the study. Fieldwork is sometimes conducted in dangerous settings and associated with considerable risks.[55] The gender of data collection teams is an important factor to consider. In many contexts, women can feel uncomfortable if they are interviewed by men or in the presence of their husbands and partners. In such settings, gender-segregated interviews are important part of ensuring a respectful and safe environment for participants. One trick for achieving this is to carry out women's and men's interviews simultaneously to keep men occupied while women participate in the study.[56] Beyond gender, other sociodemographic characteristics (eg, socioeconomic or ethnic or religious backgrounds, etc) may lead to cultural hierarchies which make it difficult for people to relate to each other.[56] A good understanding of the local context is necessary to ensure that the data collection can be as culturally sensitive as possible.

3.7. Put in place quality assurance and quality control mechanisms to ensure data accuracy, completeness and coherence

Data accuracy refers to the degree to which data correctly estimate or describe the quantities or characteristics they are designed to measure.[57] In this respect, data fabrication is a common concern in global health epidemiology and it appears to be widespread and very difficult detect.[52] The chief concern is that field workers do not visit the sampled locations and fabricate data. There are a number of quality control activities that can be put in place to ensure accurate data. The use of electronic data collection offers a number of opportunities to check that the sampled locations were visited, including geopositioning, attachment of photographs and monitoring the start and end date of the interview. Spot-checks and re-collecting data in a random sample (eg, 10%) of sampled units (eg, households or facilities) is another commonly used approach to ensure that data were correctly collected. However, the reality is often more nuanced than total data fabrication, with field workers deviating from the verbatim use of the questionnaire. This can be done for very valid reasons, for example, when field workers prefer using local terms and language, or

exercise their own judgement when asking sensitive questions. Efforts to foster a safe an open dialogue with field workers, combined very good understanding of the local context and a willingness to adapt research processes (as advocated by the slow research movement[51]) are key for quality assurance.

Data completeness is usually described as the amount of available data in a database compared with the amount that was expected to be obtained. Prompt review of research instruments by a field supervisor is important to ensure that missing data can be re-collected in time. Distinguishing between different types of missing information on the research instruments is a good way to ensure data completeness during data collection (eg, (1) the question could not be asked; (2) the respondent did not reply; (3) the respondent replied 'do not know').[58 59]

Coherence refers to the degree to which data are logically connected and mutually consistent.[57] During surveys one way to ensure coherence is include cross-checks within number of questions which should be internally consistent. Electronic data collection offers the possibility of programmed consistency checks which notify (and can even prevent) data collectors from entering inconsistent values.

Standard 4. Data management: manage data with reproducible procedures and ensure compliance with relevant data protection rules

4.1 Put in place data management procedures before effective start of data collection and provide concrete guidance in a document available to all data management staff

A data management plan is essential to ensure the planning around data collection, storage and sharing are adequately planned for at the start of the research. Broadly, the suggested items to cover in a data management plan include:

1. *Data management overview*: a description of the system(s) used, the data flow, the data management roles and responsibilities, the system for unique identification of individuals (or entities) and if relevant, the hierarchy and links between datasets and a codebook (c.f. criterion 4.3).
2. *Creation of database:* description of data entry application (which in the case of electronic data collection will coincide with the data collection application), quality assurance and quality control mechanisms (c.f. criterion 4.4), database lock and statistical file creation.
3. *Data safety and security:* relevant national/supranational legal framework(s); methods for back-ups, storage and archiving; data security protocol including access rights to ensure the anonymisation and privacy of data collected and processes for data sharing; procedures used to ensure national and international frameworks of data protection are adhered to.

In addition, depending on the complexity of the study and the data management procedures, SOPs and job aids may be useful for data management staff.

4.2 Create and pretest a data entry application prior to effective start of data collection

From the moment of effective data collection, it is important that the data management system is up and running adequately. To ensure this, it is important to test the system ahead of time. This testing may coincide with field testing (c.f. criterion 3.5) of data collection instruments.

4.3 Describe all variables in a codebook and consider preparing additional metadata documentation

Metadata are a set of data which describe the data collected through research. Metadata serve as a reference for the team members involved in study and is essential to ensure the re-usability of data for future analyses. A codebook is the primary metadata document to link the questionnaire to the study database and includes information on all the variables in the database, which question (or other source) they were obtained from, codes and valid ranges, format of notation as well as variable definitions, especially for derived (calculated) variables. Another useful document is the annotated data collection form. It is best prepared before data entry and used during data entry. It is essentially a copy of the last or latest version of the data collection form with text boxes next to every entry indicating the variable name annotated data collection form. This should ideally not replace a codebook as it does not include the same level of detail, but can be an additional useful aid. There are numerous international efforts to harmonise metadata collected as part of research for multicentre studies.[60–62]

4.4 Put in place quality control mechanisms to ensure data accuracy, completeness and coherence

Data accuracy refers to the degree to which data correctly estimate or describe the quantities or characteristics they are designed to capture.[57] The most common method to ensure accuracy with paper-based data collection is double data entry. Alternative methods include partial data checks, which can be implemented in a number of ways. One option is to select a random proportion of data points (eg, 10%) from the database and to check them visually against the completed questionnaires. A less time-consuming variant is to randomly choose a number of respondents to check (rather than a number of data points) and to check all data for those respondents against the questionnaire. For electronically collected data, accuracy can be ensured by programming the database with precoded answer options, logical ranges for continuous data and skip logic.

Data completeness is usually described as the amount of available data in a database compared with the amount that was expected to be obtained. Methods to check completeness includes tabulating the data in

the database against the sampling list to ensure that all expected data are included. However, even when all sampled elements are included, certain variables may have missing entries. This can be checked by tabulating selected 'critical' variables (eg, those most important for analysis or most likely to be missing) to ensure that there are no systematic and patterns of missingness. Ideally, this should be done at regular time points throughout the implementation of the study to ensure that mid-course corrective measures can be put in place.

Coherence refers to the degree to which data are logically connected and mutually consistent.[57] Coherence has four important subdimensions: (1) within a dataset, (2) across datasets, (3) over time, (4) across countries. Data coherence within a dataset can be ensured by cross-checking variables, which ought to be perfectly correlated. One important element of coherence across datasets is ability to merge datasets, for which a good system of assigning unique identifiers is crucial. Standardised procedures and good guidance for data collection (c.f. criterion 3.3) and data management (c.f. criterion 4.1) can help ensure coherence over time and across countries.

4.5 Annotate all data cleaning and processing steps and strive for reproducibility by means of stored programming code

Programming facilitates the documentation of study analyses and thus enables external parties to verify study results and claims and reproduce these. Most statistical software packages offer the possibility of doing data management using dropdown menus. Although this may be useful as a first step to explore, the data programming should be preferred to ensure methods reproducibility and results reproducibility. Most statistical programmes also have functionalities to store programmed code and annotate the data cleaning and analysis in a structured format (ie, R software scripts and markdowns, STATA do-files, SPSP syntax and SAS programs). Furthermore, when data are made available at different stages, programming makes it possible to progress on both data management and statistical analyses before the full database is ready. If analyses are not done by means of statistical software packages (eg, spreadsheets or qualitative analysis tools), it is important that they are nevertheless well documented and annotated to ensure results reproducibility.

4.6 For each data file define levels of anonymisation and privacy protection as well as corresponding access rights in line with national and international frameworks

Data security measures should be made explicitly clear for each stage of the research process in line with national and international frameworks—such as the General Data Protection Regulation in the EU. Personal data, and especially sensitive personal data should be treated with extreme caution.[63] Personal identifiers can be either direct or indirect.[64] Although none of the indirect identifiers on its own would point to an individual, several indirect identifiers might do. The appropriately anonymised data have: (1) no single direct identifier or less than three indirect identifiers and (2) if dates are necessary for certain analyses, methods should be used to preserve anonymity without compromising statistical analyses, such as adding or subtracting a small, randomly chosen number of days to all dates.

It should be clear at the start of the research, which research team members will have access to which data and how access will be managed (different team members might have different access rights). There are numerous ways to protect sensitive personal data. One method is by saving the personal identification data in a dataset that is separate from the bulk of the study data and only providing the ability to certain research team members to link the two datasets. This can be done by providing different passwords for different datasets or encrypting electronic database files. In the event the data are collected in a paper format, securely stored data forms (in locked cabinets with password access-locks or key-locks to which only specific research members have access) is the best way to keep the data secure.

4.7 At the beginning of the study, prepare an electronic secured study file to store all study documentation and outputs. Regularly update this file and archive it the end of the study

Maintaining a secure electronic study file helps to ensure that the most up to date versions of all study materials are stored in a single location. An electronic study file should include protocols, data analysis plans, data management plans, ethical review submissions and responses, informed consent forms, data collection tools, anonymised datasets and transcripts, metadata, data management programmes, analysis programmes, statistical outputs, reports and publications. To ensure secure storage, the study file should resists on two physically separate regularly synchronised storage mediums, for example, on a local laptop hard disk and remote backup server. When setting up the storage system, it is important to think about risks to data integrity, externally (eg, fire, flooding), and internally (disgruntled staff member, ransomware, virus attacks, etc) and how to mitigate those.

The choice of where to store and especially where to archive the data may be straightforward if researchers have an established data management facility in their institution. Cost-free remote data repositories may be a useful alternative when these are not available. There are three important considerations when choosing an online repository for data storage and archiving (as opposed to data sharing, which is discussed in criterion 6.6) : 1) Does it offer closed access and protected against unauthorised access; 2) Is it hosted by a trusted institution with a vision and capacity to provide long-term secure storage (eg, at least 10 years). Zenodo (zenodo.org) is a general purpose repository hosted by CERN which fits both criteria, while both can be problematic with public could storage services (such as DropBox, Google Drive, etc).

4.8 Retain source data safely, in their original form, preserving data confidentiality for as long as has been described in the protocol

Source data refers to materials collected as part of the research at the primary source of data collection (study participants, household respondents, etc). Thus, source data includes: signed informed consent forms, filled in data collection forms, audios files, videos and photos and biological samples (data, images, photos of slides, but not sample itself). The study protocol should specify how long source data will be stored and under which conditions (and security guidance).

Standard 5. Data analyses: analyse data according to the protocol and integrate statistical analyses with approaches from other disciplines in the study

5.1 Only work with personal identifiers that are necessary to answer the research questions

During the process of data analysis, the person analysing the data should work on an appropriately anonymised or pseudo-anonymised dataset. Respondent-specific identifier number should be used to identify individual respondents in the data and the key between the identifiers and the personal confidential information of the respondents must be hosted by an independent person. As described in criterion 4.6, identifiers can be direct or indirect, and a combination of indirect identifiers may be sufficient to identify a person.[64] Therefore, it is important to realise that there are limits to the extent to which this criterion can be met, especially if a number of indirect identifiers are relevant to analyses (eg, nationality/ethnicity, sex and age). Even if clearly personal information (such as name, address, telephone number, ie, direct identifiers) are removed from a dataset, it is usually still possible to identify individuals though combinations of indirect identifiers (such as disease status, sex, age and ethnic background). Such indirect identifiers are often relevant for the analysis. It is therefore important to realise that in practice, a dataset from a global health research project is rarely anonymised. However, pseudo-anonymisation may well be achieved.

5.2 Conduct statistical analyses in accordance with the protocol and distinguish preplanned from exploratory analyses

One of the cornerstones of research integrity in epidemiology is ensuring that analyses do not deviate from the plan. As discussed in criterion 2.2, it is important to think about analyses before conducting a study because of dangers associated with performing multiple statistical tests. Most GEP guidelines recommend that any deviations from the statistical analysis plan are justified and documented. As discussed in criterion 5.4 below, such requirements may be difficult to fulfil in multidisciplinary studies where qualitative research informs the quantitative research (exploratory model) or vice versa (explanatory model).[65] The prespecification of all analyses goes against the iterative nature of qualitative research. Certainly, analyses which were preplanned in the protocol and for which the study is powered should be distinguished from other exploratory analyses. Furthermore, all analyses should clearly relate to the research questions the study was set out to answer.

5.3 Fully annotate all analysis steps and strive for reproducibility by providing programming code

All analyses steps need to be replicable to ensure results reproducibility and inferential reproducibility. As discussed in criterion 4.5, this can be facilitated by means of stored and annotated programmes or plain language instructions in a spreadsheet or word processing document. When data are made available at different stages, programming makes it possible to progress on both data management and statistical analyses before the full database is ready. Ideally, programming code should be organised in a way that enables results to be reproduced from the 'clean raw database' at the click of a button.

5.4 In multidisciplinary studies, integrate statistical analyses with analyses from other study disciplines in an iterative process to coherently address the research objectives

As discussed in criterion 2.4, global health promotes multidisciplinary collaboration. In order to maximise the success of a multidisciplinary approach, study plans need to include regular moments of reflection with peers across all involved disciplines, throughout all study phases, but especially in design, analysis and interpretation of findings.[38]

At the analysis stage, one of the defining features of multidisciplinary research is the iterative cycles through which information from the various disciplines are integrated in order to coherently address the research questions. Multidisciplinary research involving disciplines with both quantitative and qualitative research traditions are especially challenging as it requires researchers to overcome and compromise on at times deep epistemological divergences. In our experience, the following iterative approach can help to ensure that the quantitative data are coherently mixed with the other qualitative disciplines: (1) start qualitative data analysis early on during data collection to ensure that all emerging themes are being explored; (2) conduct preliminary descriptive analyses of both quantitative and qualitative data as soon as data are available for analyses; (3) convene with peers from other research disciplines to discuss further statistical analysis of quantitative data (descriptive and inferential) and synthesis of the qualitative data (key themes); (4) combine analyses from the various disciplines to answer the research questions comprehensively; (5) define further higher-level analyses (either qualitative or quantitative) where gaps persist; (6) take note of elements which still need to be explored with new data and new research.

5.5 Put in place quality control mechanisms to ensure that data have been correctly analysed

The most robust method to prevent erroneous analyses from being disseminated is having results (or a purposeful selection thereof) reproduced by a qualified person who

was not previously involved in the analyses. Inconsistencies in the results should be discussed and a consensus reached between the two analysts. However, these types of approaches are often not possible in research settings as they are costly and time-consuming. Furthermore, there may not be another qualified person in the team capable of performing of an independent analysis. In such cases, one option is to ensure that the research team meets frequently, at different phases of results generation process, to review results and assess their validity, in order to spot any errors and mistakes at an early stage.

Standard 6. Dissemination and communication: report and disseminate results, preferably in the public domain, with means of communication which appropriately target key stakeholders

6.1 Develop user-specific dissemination and communication plans in consultation with key stakeholders (representatives of the affected populations and end-users)

Dissemination usually refers to making results known to research peers, policy-makers and other professional organisations to enable them to use the results in their own work.[66] Communication refers to the promotion of results to communities and societies as a whole and possibly engaging in a two-way exchange.[66] Publication of papers in peer-reviewed journals is often epidemiologists' preferred mode of dissemination. Yet, it primarily targets the scientific community and international agencies while in order to have an impact (eg, change policies, practices or behaviour), global health research findings needs to be disseminated and communicated more broadly, in ways that will enable end-users to understand and find them.[2] Research findings must be translated into different 'formats and languages' appropriate to the respective target audience, and should be delivered through effective communication channels.[2] Dissemination materials may include policy briefs and white papers summaries for pamphlets and websites. Communication material can take the form of news articles and social media posts; community meetings, newspaper articles, videos or short films, documentaries, podcasts, infographics, etc. Art-based approaches, such as theatre, music, visual arts, storytelling and film[67] are especially useful to reach and engage large numbers of people. Study findings need to be communicated neutrally and impartially, and where necessary conflicts of interest need to be clarified/declared.

6.2 Report data reporting in a non-stigmatising, non-discriminatory, culturally sensitive and non-identifying manner

The information included in the dissemination and communication material must not stigmatise, discriminate or identify the study participant. Country-specific regulations must be followed during the dissemination of epidemiology study results. However, less stringent data protection standards in low-income countries can never be an excuse for researchers from high-income countries to condone potential privacy breaches.[3] Special attention must be paid to ensure the protection of research participants who are at risk of stigmatisation, discrimination or incrimination.[3] More specifically, epidemiologists should bear in mind that presenting data from small groups in tables or maps may make individuals easily identifiable and thus break confidentiality. If any participants are quoted with names or in picture, due consent for publicising their information must be obtained, paying particular attention to the protection of minors, elderly and other vulnerable populations.

6.3 Conform to reporting guidelines for the given study design and methods in academic publications

Reporting guidelines are structured tools to guide researchers in the preparation of their scientific manuscripts. A reporting guideline provides a minimum list of information needed to ensure a methods and/or results can be understood by a reader, reproduced by a researcher, used by a practitioner and included in a systematic review.[68] The Enhancingthe QUAlity and Transparency Of Health Research network is an online platform that promotes and disseminates reporting guidelines for health research, which can be consulted to identify relevant guidelines.[68] Guidelines relevant for epidemiological study reporting include the Strengthening the Reporting of Observational Studies in Epidemiology,[69 70] RECORD,[71] Consolidated Standards of Reporting Trials,[72 73] STARD[74] and Preferred Reporting Items for Systematic Reviews and Meta-Analyses[75] guidelines, and Standards for Reporting Qualitative Research[76] for qualitative research.

6.4 Put in place quality assurance and quality control mechanisms to ensure complete, accurate, accessible and interpretable data reporting

Complete and accurate reporting in scientific publications is key to research integrity. Previous items have described approaches to guaranteeing prepublication of the protocol (criterion 2.5) and use of reporting guidelines (criterion 6.3). Accessibility of results on the other hand is the primary driver behind open access of publications and is discussed in following sections (criteria 6.5 and 6.6).

Interpretability reflects the ease with which users may understand and properly use data products.[57] As discussed in criterion 6.1, this is very important in global health, as research findings need to be adequately communicated to end-users in order to have an effect on behaviour, decision making or policies—and an ultimate impact on health. Participatory approaches which engage users in the compilation of dissemination findings are especially useful to ensure that messages speak to the needs and concerns of users, are delivered through the most effective channels, and are understood as intended.

6.5 Consider indexed open access journals for scientific publications

Open access to scientific publications is one of the cornerstones of efforts to foster research integrity and transparency. There are two main routes to open access[77]: (1) self-archiving ('green' open access) where researchers archive the published article or the final peer-reviewed manuscript in an online repository before,

at the same time as, or after publication; (2) open access publishing ('gold' open access) where an article is immediately published in open access mode. With open access publishing publication costs (referred to as article processing charges (APCs)) are borne by the authors instead of readers. The charges of journals with high impact factors can be expensive and need to be considered when budgeting for the research. Many journals do offer discounted or waived rates for researchers from low-income and middle-income countries (and further discounts for students).

Although the international status and impact factors of journals is an important aspect to establish the credibility of the research, sometimes, local or national-level journals can better reach targeted audiences and demonstrate a commitment to address local research questions and policy issues.[78] These may or may not be open access. Where possible it is good to favour indexed journals, which can be found by search engine databases (eg, PubMed). A journal's membership of the COPE network (https://publicationethics.org/) also indicates a commitment to ethical publishing practices In this regard, it is important to be aware of predatory publishing, an exploitive academic publishing business model that involves charging APCs to authors without providing editorial services, peer review or indexation. Young and inexperienced researchers from low-income and middle-income countries are most likely to publish in these journals.[79] The line between 'serious and reputable' journals and predatory journals is blurred and unfortunately, a number of national journals in low-income and middle-income countries are deemed predatory.[68]

6.6 On study completion, consider publication of the archive in an openly accessible online repository. Consult key stakeholders and research partners to identify strategies within the study team's remit to encourage as much as possible re-analyses by local researchers

Open access data sharing is increasingly being encouraged and at times a condition for funding and publication. It is considered necessary to maximise the return on investment in research, with benefits ranging from the generation of novel findings as researchers re-examine the data applying different hypotheses, the possibility combine data sets from multiple studies, and the development of new research collaborations.[77 80] There are many online repositories which support open access data sharing. The considerations to choose a repository for data sharing are slightly different that those discussed for data storage and archiving (criterion 4.7), as the main aim here is to maximise the ease with which peers will be able to find and access the data. The FAIR guidelines aim to improve the findability, accessibility, interoperability and reuse of digital data by both humans and machines.[81] On one hand, researchers may want to privilege repositories which comply with these guidelines. On the other hand, there are difficulties in interpreting and putting these principles into practice and many repositories are still not able to comply, especially those in social sciences.[82] Therefore, domain-specific open access repositories may be the most effective route to implement open access data sharing for global health epidemiologists, regardless of their compliance with FAIR. The Registry of Research Data Repositories (re3data.org) offers an overview of existing international repositories for research data.

However, epidemiologists should also be aware that there is also a less noble side to data sharing in global health. It can end up being a lot more advantageous for scientists in high-income countries with higher analytical capacities than those in low-income countries where the data have been collected. While scientists in high-income countries may be highly trained to perform analyses, they have neither shared the legwork in collecting the data (including intellectual design and practical troubleshooting) with scientists in the low-income countries where the data were collected .[83] In order to ensure that data sharing is mutually advantageous to all parties, the principle of 'as open as possible, as closed as necessary'[77] should therefore be followed. Embargoes are a useful short-term strategy to afford more time to local researchers, but fair data sharing should be considered within the frame of comprehensive long-term approaches to knowledge sharing—that is, epidemiological capacity building of researchers and more general investments in research infrastructure in low-income countries. As discussed in criterion 1.1, the extent of this capacity strengthening should be commensurate to the scope of the research.

CONCLUSION

None of what is described here will be new to experienced global health epidemiologists and researchers. Yet we know from first-hand experience that it is not easy to navigate the competing demands on a researcher's loyalty in the complex multistakeholder environment in which we operate. With the benefit of hindsight there are certainly many things we would now do differently. By giving a name and a space to recurring challenges, by stimulating a reflection on routine practice and common assumptions, by offering arguments and background references, we intend to support those who are trying to stand up against questionable research practices and research unfairness.

The notes of caution and invitations to reflect on research integrity and research fairness issues jointly can be valuable for teaching purposes for young epidemiologists and researchers embarking on the field of global health epidemiology. The exposure of these notions early on in their educational and professional development can ensure that the new generation of global health epidemiologists is more aware of the intricacies and challenges of our field, so that they do not unknowingly repeat known mistakes and reinforce unfair patterns of research behaviour. Yet we also hope that more experienced researchers will also be open to reflect on some deeply

engrained practices and assumptions in global health epidemiology. Ultimately, we are aware that dissemination of these guidelines to a broad audience—including commissioners, funders, reviewers and publishers of research—is key to have to have a tangible impact.

GLOSSARY

Affected populations: individuals and communities that are affected by the data collection process. This may be the people on whom data collection was actually done, but also their families and the wider community which may be directly or indirectly affected by it.

End-users: individuals, communities or organisations external to those who conducted the research, who will directly use or directly benefit from the output, outcome or results of the research. Examples of end-users include researchers, policy-makers from governmental and non-governmental organisations, the service providers, communities and community organisations.

Job aids: instructions, lists or quick reference materials derived from the main SOP. Job aids can be used when the full procedure is not needed at the time the task is performed.[54]

Multidisciplinary research: research which combines and, in some cases, integrates concepts, methods and theories drawn from two or more disciplines. Others may refer to this as 'mixed methods',[37] 'cross-disciplinary'[8] or 'multiple discipline' research.[35]

Quality attributes: the formulation of quality assurance and quality control activities revolves around goals for quality attributes. Quality attributes in epidemiology include data quality dimensions such as relevance; accuracy; credibility; timeliness; completeness; accessibility; interpretability and coherence.[57] These can either be attributes of the system that produced the data (ie, the process) or of the data itself (data output/outcome).[82]

Quality assurance: set of activities to ensure quality in the processes by which products are developed. Quality assurance aims to prevent defects with a focus on the process used to make the product. It is a proactive and ongoing quality process. Quality assurance includes quality control activities.

Quality control: set of activities for ensuring quality in products. The activities focus on identifying defects in the actual products produced. Quality control aims to identify (and correct) defects in the finished product. Quality control, therefore, is a reactive process. Quality control activities a part of a broader quality assurance.

Reliability: the degree to which a measurement is free from error, or more extensively, the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions[48]:

- *internal consistency*: using different sets of items from the same multi-item measurement instrument;
- *over time*: test–retest;
- *inter-rater*: by different persons on the same occasion;
- *intra-rater*: by the same raters/responders on different occasions.

Parameter/methods to measure reliability include: the SE of measurement, intra-class correlation coefficient, coefficient of variation, Cohen's kappa, Cronbach's alpha and Bland-Altman plots.[48]

Research instrument: set of questions or items used to collect information about research participants. Examples of research instruments: questionnaires for primary data collection with respondents, data extraction forms for collection of existing data records, case report forms to collect clinical data or interview guides for qualitative data collection, laboratory and imaging techniques, global position system and other devices. Synonym: research tool

Reproducibility[53]: an overall term which refers to

- *Methods reproducibility*: provision of enough detail about the procedures of a study so that these study procedures can be repeated exactly.
- *Results reproducibility:* ability of an independent study with closely matched procedures to give the same results as the original study.
- *Inferential reproducibility*: an independent replication of a study or a reanalysis of a study lead to qualitatively similar conclusions as the original study.

Standard operating procedures (SOPs): written step-by-step instructions on how to carry out procedures correctly. SOPs are meant to ensure consistency, accuracy and quality of data. They help ensure compliance to the study protocol, regulations and international standards. SOPs can also be used as training tools.[54]

Validity: the degree to which an instrument truly measures what it purports to measure. Parameters/methods to measure validity include: specificity and sensitivity, receiver operating characteristic curves, weighed kappa, Spearman's or Person's correlation coefficients and Bland-Altman limits of agreement, factor analysis. Three different types of validity can be distinguished[48]:

- *Content validity*: does the content of the instrument correspond with what one intends to measure, with regard to relevance and comprehensiveness?
- *Criterion validity*: in situations where there is a gold standard for the measurement, how well do the scores of the measurement instrument agree with the scores on the gold standard?
- *Construct validity*: when there is no gold standard, does the instrument provide expected scores, based on knowledge on what it is trying to measure?

Twitter Rutuja Patil @Ru2ja

Contributors JR drafted the section on protocol development, KV on data collection, AL on data management, RP, WM and SJ on dissemination and communication. SA compiled all contributions and finalised the document. SFR reviewed and complemented the first draft. All authors reviewed and approved the final version of this manuscript.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iDs
Sandra Alba http://orcid.org/0000-0003-2435-624X
Rutuja Patil http://orcid.org/0000-0001-6390-4363

REFERENCES

1 Alba S, Verdonck K, Lenglet A. Bridging research integrity and global health epidemiology (bridge) statement: guidelines for good epidemiological practice. *BMJ Global Health* 2020:e003236.
2 Swiss Commission for Research Partnerships with Developing Countries (KFPE). A Guide for Transboundary Research Partnerships - 11 Principles, 2018. Available: https://naturalsciences.ch/service/publications/9505-a-guide-for-transboundary-research-partnerships-3rd-edition-2018-
3 TRUST. Global code of conduct for research in resource poor settings. Available: https://www.globalcodeofconduct.org/wp-content/uploads/2018/05/Global-Code-of-Conduct-Brochure.pdf
4 Varvasovszky Z, Brugha R. A stakeholder analysis. *Health Policy Plan* 2000;15:338–45.
5 Accountability. AA1000 Stakeholder engagement standards. Available: http://www.accountability.org/standards/ [Accessed 26 Oct 2017].
6 The research Fairness initiative. RFI reporting guide, 2018. Available: http://rfi.cohred.org/wp-content/uploads/2018/04/RFI_ReportingGuide_EN_v2.0_20180315.pdf
7 Cáceres CF, Mendoza W. Globalized research and "national science": the case of Peru. *Am J Public Health* 2009;99:1792–8.
8 Ding Y, Pulford J, Bates I. Practical actions for fostering cross-disciplinary global health research: lessons from a narrative literature review. *BMJ Glob Health* 2020;5:e002293.
9 International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals, 2019. Available: http://www.icmje.org/icmje-recommendations.pdf
10 Smith E, Hunt M, Master Z, *et al*. Authorship ethics in global health research partnerships between researchers from low or middle income countries and high income countries. *BMC Med Ethics* 2014;15:42.
11 Blokdijk G. *The Service Level Agreement SLA Guide - SLA book, Templates for Service Level Management and Service Level Agreement Forms. Fast and Easy Way to Write your SLA. Lightning Source*, 2008.
12 WHO. Recommended format for a 'research protocol'. Available: https://www.who.int/ethics/review-committee/format-research-protocol/en/ [Accessed 9 May 2020].
13 Chan A-W, Tetzlaff JM, Altman DG, *et al*. Spirit 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;158:200–7.
14 Gamble C, Krishan A, Stocken D, *et al*. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA* 2017;318:2337–43.
15 Hiemstra B, Keus F, Wetterslev J, *et al*. DEBATE-statistical analysis plans for observational studies. *BMC Med Res Methodol* 2019;19:233.
16 Koplan JP, Bond TC, Merson MH, *et al*. Towards a common definition of global health. *Lancet* 2009;373:1993–5.
17 A History of Global Health. Johns Hopkins University press books. Available: https://jhupbooks.press.jhu.edu/title/history-global-health [Accessed 29 Apr 2020].
18 Victora CG, Wagstaff A, Schellenberg JA, *et al*. Applying an equity lens to child health and mortality: more of the same is not enough. *Lancet* 2003;362:233–41.
19 GHO. Health equity monitor. Available: https://apps.who.int/gho/data/node.main.HE-1540 [Accessed 11 May 2020].
20 O'Neill J, Tabish H, Welch V, *et al*. Applying an equity lens to interventions: using progress ensures consideration of socially stratifying factors to illuminate inequities in health. *J Clin Epidemiol* 2014;67:56–64.
21 García AM, Bartley M, Alvarez-Dardet C. Engendering epidemiology. *J Epidemiol Community Health* 2007;61:ii1–2.
22 Krieger N. Genders, sexes, and health: what are the connections--and why does it matter? *Int J Epidemiol* 2003;32:652–7.
23 White A, Richardson N. Gendered epidemiology: making men's health visible in epidemiological research. *Public Health* 2011;125:407–10.
24 Ruiz-Cantero MT, Vives-Cases C, Artazcoz L, *et al*. A framework to analyse gender bias in epidemiological research. *J Epidemiol Community Health* 2007;61:ii46–53.
25 Jahn I, Börnhorst C, Günther F, *et al*. Examples of sex/gender sensitivity in epidemiological research: results of an evaluation of original articles published in JECH 2006-2014. *Health Res Policy Syst* 2017;15:11.
26 Morgan R, George A, Ssali S, *et al*. How to do (or not to do)… gender analysis in health systems research. *Health Policy Plan* 2016;31:1069–78.
27 Alba S, Wong F, Bråten Y. Gender matters in household surveys. *Signif* 2019;16:38–41.
28 Crenshaw K. Demarginalising the intersection of race and sex: a black feminist critique of anti-discrimination doctrine, feminist theory, and anti-racist politics. *Univ Chic Leg Forum* 2011;140:25–42.
29 Hankivsky O, Christoffersen A. Intersectionality and the determinants of health: a Canadian perspective. *Crit Public Health* 2008;18:271–83.
30 Bauer GR. Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity. *Soc Sci Med* 2014;110:10–17.
31 López N, Gadsden VL, University of Pennsylvania. Health inequities, social determinants, and intersectionality. *NAM Perspect* 2016;6.
32 Kareem Carr on Twitter: 'Saying "Race is a risk factor" is problematic. Statisticians and social scientists need to be more careful about this kind of language and avoid it where possible. Available: https://twitter.com/kareem_carr/status/1270756754201219074 [Accessed 19 Jun 2020].
33 Mena E, Bolte G, *et al*, ADVANCE GENDER Study Group. Intersectionality-based quantitative health research and sex/gender sensitivity: a scoping review. *Int J Equity Health* 2019;18:199.
34 Scott NA, Siltanen J. Intersectionality and quantitative methods: assessing regression from a feminist perspective. *Int J Soc Res Methodol* 2017;20:373–85.
35 Choi BCK, Pak AWP. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. definitions, objectives, and evidence of effectiveness. *Clin Invest Med* 2006;29:351–64.
36 Tobi H, Kampen JK. Research design: the methodology for interdisciplinary research framework. *Qual Quant* 2018;52:1209–25.
37 O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. *J Health Serv Res Policy* 2008;13:92–8.
38 BMJ GH Blogs. Not everything that counts can be counted: mixed methods impact evaluations in global health, 2020. Available: https://blogs.bmj.com/bmjgh/2020/05/11/not-everything-that-counts-can-be-counted-mixed-methods-impact-evaluations-in-global-health/ [Accessed 11 May 2020].
39 Home - ClinicalTrials.gov. Available: https://www.clinicaltrials.gov/ [Accessed 10 May 2020].
40 WHO. Welcome to the who ICTRP. Available: http://www.who.int/ictrp/en/ [Accessed 10 May 2020].
41 Piasecki J, Waligora M, Dranseika V. What do ethical guidelines for epidemiology say about an ethics review? A qualitative systematic review. *Sci Eng Ethics* 2017;23:743–68.
42 Council for International Organizations of Medical Sciences, World Health Organization. *International ethical guidelines for epidemiological studies*. Geneva: CIOMS, 2009.
43 Council for International Organizations of Medical Sciences. *International ethical guidelines for health-related research involving humans*. Geneva: CIOMS, 2017.
44 Bruno W, Haar RJ. A systematic literature review of the ethics of conducting research in the humanitarian setting. *Confl Health* 2020;14:27.
45 Hunt M, Tansey CM, Anderson J, *et al*. The challenge of timely, responsive and rigorous ethics review of disaster research: views of research ethics Committee members. *PLoS One* 2016;11:e0157142.
46 Eckenwiler L, Pringle J, Boulanger R, *et al*. Real-Time responsiveness for ethics oversight during disaster research. *Bioethics* 2015;29:653–61.
47 Alba S, Mergenthaler C. Lies, damned lies and epidemiology: why global health needs good epidemiological practice guidelines. *BMJ Glob Health* 2018;3:e001019.
48 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and

definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
49 de VHCW, Terwee CB, Mokkink LB. *Measurement in medicine: a practical guide*, 2011.
50 de Leeuw ED. To mix or not to mix data collection modes in surveys. *J Official Stat* 2004;21.
51 Adams V, Burke NJ, Whitmarsh I. Slow research: thoughts for a movement in global health. *Med Anthropol* 2014;33:179–97.
52 Kingori P, Gerrets R. The masking and making of fieldworkers and data in postcolonial global health research contexts. *Crit Public Health* 2019;29:494–507.
53 Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Sci Transl Med* 2016;8:341ps12.
54 Barbé B, Verdonck K, Mukendi D, *et al*. The art of writing and implementing standard operating procedures (SOPs) for laboratories in low-resource settings: review of guidelines and best practices. *PLoS Negl Trop Dis* 2016;10:e0005053.
55 Hall-Clifford R, Addiss DG, Cook-Deegan R, *et al*. Global health fieldwork ethics: mapping the challenges. *Health Hum Rights* 2019;21:1–5.
56 Elias M. *Practical tips for conducting gender-responsive data collection*, 2013: 2.
57 OECD. Quality framework for OECD statistical activities. Available: http://www.oecd.org/sdd/qualityframeworkforoecdstatisticalactivities.htm [Accessed 2 Oct 2019].
58 Juul S. *Take good care of your data*, 2004: 57.
59 Epidemiology for the uninitiated. Available: https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated [Accessed 11 May 2020].
60 ddi3. Data documentation initiative. Available: https://ddialliance.org/category/tags/ddi3 [Accessed 11 May 2020].
61 Finnie TJR, South A, Bento A, *et al*. EpiJSON: a unified data-format for epidemiology. *Epidemics* 2016;15:20–6.
62 Welcome - Humanitarian Data Exchange. Available: https://data.humdata.org/ [Accessed 11 May 2020].
63 What personal data is considered sensitive? Eur. Comm. - Eur. Comm. Available: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en [Accessed 11 May 2020].
64 Hrynaszkiewicz I, Norton ML, Vickers AJ, *et al*. Preparing RAW clinical data for publication: guidance for Journal editors, authors, and peer reviewers. *BMJ* 2010;340:c181.
65 Designing and conducting mixed methods research, 2017. Available: https://wordery.com/designing-and-conducting-mixed-methods-research-john-w-creswell-9781506386621 [Accessed 24 Apr 2020].
66 Funding & tenders. Available: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq/933 [Accessed 19 Jun 2020].
67 Bunn C, Kalinga C, Mtema O, *et al*. Arts-based approaches to promoting health in sub-Saharan Africa: a scoping review. *BMJ Glob Health* 2020;5:e001987.
68 The EQUATOR Network. Enhancing the quality and transparency of health research. Available: https://www.equator-network.org/ [Accessed 11 May 2020].
69 von Elm E, Altman DG, Egger M, *et al*. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology* 2007;18:800–4.
70 Vandenbroucke JP, von Elm E, Altman DG, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Epidemiology* 2007;18:805–35.
71 Benchimol EI, Smeeth L, Guttmann A, *et al*. The reporting of studies conducted using observational Routinely-collected health data (record) statement. *PLoS Med* 2015;12:e1001885.
72 Schulz KF, Altman DG, Moher D, *et al*. Consort 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
73 Consort - Welcome to the CONSORT Website. Available: http://www.consort-statement.org/ [Accessed 25 May 2020].
74 Cohen JF, Korevaar DA, Altman DG, *et al*. Stard 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.
75 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
76 O'Brien BC, Harris IB, Beckman TJ, *et al*. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014;89:1245–51.
77 European Comission. Guidelines to the rules on open access to scientific publications and open access to research data in horizon 2020, 2017. Available: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
78 Abimbola S. The foreign gaze: authorship in academic global health. *BMJ Glob Health* 2019;4:e002068.
79 Xia J, Harmon JL, Connolly KG, *et al*. Who publishes in "predatory" journals? *J Assn Inf Sci Tec* 2015;66:1406–17.
80 Parker M, Bull S. Sharing public health research data: toward the development of ethical Data-Sharing practice in low- and middle-income settings. *J Empir Res Hum Res Ethics* 2015;10:217–24.
81 Wilkinson MD, Dumontier M, Aalbersberg IJJ, Ijj A, *et al*. The fair guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:1–9.
82 Dunning A, De Smaele M, Böhmer J. Are the fair data principles fair? *Int J Digit Curation* 2017;12:177–95.
83 Tangcharoensathien V, Boonperm J, Jongudomsuk P. Sharing health data: developing country perspectives. *Bull World Health Organ* 2010;88:468–9.