Author’s response to reviews

Title: Data management and sharing policy: the first step towards promoting data sharing

Authors:

Naomi Waithira (naomi@tropmedres.ac)
Brian Mutinda (brian@tropmedres.ac)
Phaik Yeong Cheah (phaikyeong@tropmedres.ac)

Version: 1 Date: 02 Mar 2019

Author’s response to reviews:

1st March 2019

To the Editor
BMC Medicine

Dear Sir/Madam,

Resubmission of paper entitled “Data management and sharing policy: the first step towards promoting data sharing” BMED-D-19-00019

We thank the reviewers for the valuable comments. Below is a point by point response to their comments.

Reviewer #1: ◦ You observe that funding and journal requirements are currently the main vehicles for data sharing. Please explain the relationship between these and institutional policies: should institutional policies coexist with the other vehicles, and if so, why is this ideal? Should they supplant the other vehicles, and if so, why are they better? Why is it not enough to simply make the existing vehicles more robust?

Funders and journals set their requirements based on their specific goals and interests; most health research funders expect that research data should be made openly available with as few restrictions as possible, while journals require data to be made transparently available for purpose of validating the publication but not necessarily openly available. Subtle discrepancies in the requirements from funders and journals mean that neither is all-inclusive. Institutional policies on other hand should be specific to the research context and exist to promote the interest of the institution as well as safeguard the interests of their researchers and study participants. Institutional policies should not replace other policies but rather compliment them and provide direction for researchers when there are discrepancies between the funder and journal requirements.
Do institutions have sufficient incentives to adopt data sharing policies to make this approach viable, and if so, what are those incentives? The development process sounds resource intensive (you state that prerequisites include in-depth consultations with internal and external stakeholders, extensive literature reviews, etc.), so what, if anything, offsets this?

One of the key potential (not much empirical data is available) incentives of data sharing is generation of more research output such as publications from shared datasets. This is envisaged to translate to increased visibility, increased opportunities for collaboration and more funding opportunities for the research groups. More importantly, a policy that is aligned to institutional aims can serve as internal incentive.

You say that institutional policies are a "pre-requisite", the "first step", etc., to data-sharing, but data sharing seems to be possible (and occurring) even without such policies. Can you clarify? Is the idea that these policies could fill the gaps that still exist, and that you point to?

Data sharing is indeed occurring, however, there are gaps that exist such as inequity in data access and reuse. In addition, a lot sharing occurs without the basic data management standards such as sending datasets via unsecure channels, for example by email.

Institutional policies help address these gaps, offer protection of the intellectual property rights and security of research data generators while ensuring compliance with funder and journal requirements.

Please avoid repeatedly advising the reader to follow best practices without engaging with the content of those best practices (e.g. "in accordance with accepted ethical and quality standards", "meets standards and regulatory requirements", "in compliance with established ethical and scientific quality standards", "be aware of ... how datasets can be stored securely", have "knowledge of available data management tools and softwares", etc. etc.)

Noted with thanks. We have rewritten the ‘Data management for data sharing’ section.

However, we cannot be too specific as it is not the intention of this paper to discuss details of the specific standards and guidelines as these vary greatly depending on domain of research, local regulations and geographic location, where the research is conducted.

Is it really the case that "Every study should have a ... data sharing plan"? I'm not convinced the cited source supports this, and I'm not sure it is true, especially across all fields of study.

Thank you. We agree that not every study should have a data sharing plan.
The institutional data management and sharing policy is intended to provide overarching principles for all studies conducted in an institution while a study specific data sharing plan outlines detailed procedures that are specific to the study. For example, an institutional data sharing policy may dictate that data generated from any study should be deposited in an open-access data repository. A study data sharing plan will specify which repository the study data will be deposited. This detail will be different depending on the area of study - repositories for genetics data are different from repositories for clinical trial data. This paper focuses on studies conducted in biomedical research involving human participants where data sharing plans are valuable. We recognise that in other fields of study, individual studies may not need data sharing plans.

The text has been rephrased and a reference supporting this statement has been included (Research Data Oxford Funder requirements http://researchdata.ox.ac.uk/funder-requirements/)

- Page 3 Line 10: "who to share data with and what processes ..."

This sentence has been revised and we included a reference to substantiate this claim.

“Health researchers may not be well informed on what sharing options are available and what needs to be in place in order to share their data”

- P3L15: "... tend not to request data."

Text revised to the following. “Furthermore, very few researchers in LMICs have requested for data.”

- P3L19-20: "fear - that their failure to request data access means that data sharing benefits only HIC researchers, and leaves them worse off thereby exacerbating ...

We agree that this sentence is unclear. We have revised the sentence.

- P3L24: "This situation is ethically problematic ...

Sentence removed.

- P3L53: "aim to improve the treatment ...

Text updated to correct the grammatical error.

- P4L6: "... stakeholders are necessary. Scoping and cataloguing what type ...

Text updated as suggested

- P6L15: NB blanket consent is illegal: please remove text.

Text removed
Consider referencing GA4GH framework, policies or tools in the conclusion.

The GA4GH is a good framework and good to consider in sharing health related data.

We have included it as reference (Knoppers 2014) and included it in the conclusion as suggested.

Reviewer #2: The paper tries to motivate the need for a data management and data sharing policy in low-and-middle-income countries and also outlines some properties that this policy should have.

It is generally clear and easy to read.

Thank you for the positive comments!

My major question regarding the paper is what are the specific characteristics of low-and-middle-income countries that require a different policy compared to the ones in high-income countries. This should be discussed first thing in the background, in my opinion. The suggested policy seems to me as very similar to the ones in high-income counties (e.g., UK).

We think the fundamentals of a policy are the same but in this paper we will focus on LMICs. We have added text in Background, last paragraph.

Also, I would expect a brief review of technical means to protect data beyond policies, such as encryption, anonymization, and access control.

We agree this is important but it is not the intention of this paper to discuss the implementation of elements of data management such as data protection. However the text has been updated to provide a general overview. See Data management for data sharing

There are also the following concerns, questions and suggestions:

1. The main body part of the abstract does not mention the focus on low-and-middle-income countries, which is in contrast with the background section.

Added a sentence in the Abstract (last sentence)

2. The sentence in lines 54-61 on page 2 should be explained. Why are such datasets valuable?

We added a sentence in Paragraph 1 (last sentence).

3. The sentence in lines 32-34 on page 4 requires some explanation and references. Please correct the font size too.

This paragraph has been written in light of Reviewer #1’s comments.
4. The sentence in lines 57-59 on page 4 does not explain whether low-and-middle-income countries have actually established repositories.

This paragraph has been written in light of Reviewer #1’s comments.

5. The sentence in lines 4 to 6 on page 5 is not clear. Please describe what is blank case report forms CRfs).

A case report form refers to a questionnaire in clinical research studies. A blank case report form is an unfilled questionnaire and is provided alongside study data to give the secondary user an idea of what questions were asked at data collection. We have included a reference to explain it (Bellary et al, 2014)

6. The sentence in line 28 on page 5. Please explain what is DAC.

Done. The sentence has been revised. “Managed access via application to a Data Access Committee (DAC), a committee who have the responsibility of reviewing, assessing access requests and approving or disapproving request.”

In addition to the above, we have made some changes to improve the clarity of our arguments, corrected some grammatical errors and included additional references.

We hope that our paper is now suitable for publication in your journal.

Thank you.

Phaik Yeong Cheah, B.Pharm, MSc (Bioethics), PhD

Corresponding author