Author’s response to reviews

Title: Clinical trial data reuse – overcoming complexities in trial design and data sharing

Authors:

Toby Wilkinson (toby.wilkinson-2@postgrad.manchester.ac.uk)

Siddharth Sinha (siddharth.sinha@student.manchester.ac.uk)

Niels Peek (niels.peek@manchester.ac.uk)

Nophar Geifman (nophar.geifman@manchester.ac.uk)

Version: 1 Date: 08 Jul 2019

Author’s response to reviews:

Attached is our revised manuscript titled "Clinical trial data reuse – overcoming complexities in trial design and data sharing".

We have addressed all of the reviewer’s comments below and marked all the changes to the manuscript by using the “track changes” function.

Thank you for considering our submission and we look forward to a favorable outcome.

Reviewer reports:

Reviewer #1: This well-written manuscript explores a popular topic, clinical trial data sharing, from an atypical perspective, that of the data reuser. Most discussions of facilitating data sharing have focused on the needs and perspectives of data suppliers, so this commentary adds a valuable voice.

My suggestions for strengthening the manuscript are mostly organizational in nature. I found the substantive points interesting and important.

1. The introduction could be improved by providing a topic sentence and forecast that alerts readers to what the article is going to do and cover.

Thank you for this useful suggestion. We have added such a sentence to the introduction; page 3: “Here we describe a few of the challenges in the combining and pooling of shared clinical trial data for secondary analysis, from the perspective of a data re-user. Specifically, we focus on issues related to the quality of the data being shared and to differences in the design of the trials being reused.”
2. For the psoriasis example, I would have enjoyed hearing more about how the reported data problem affected the authors' work. It seems like a pretty small issue with a dataset that, apparently, was otherwise OK. Was it in fact significant? Is it indicative of some broader problem with datasets?

Thank you for this comment. We have added the following text to expand on how this affected our analysis on page 4: “This impacted our ability to include such information in our models and our analysis plan had to be adjusted accordingly, possibly resulting in less than optimal description of clinical response patterns.”

We have also added the following to the last paragraph of the Transparency and availability of information section on page 4: “The extent of such differences could vary, and as such so will the downstream effect on intended analyses.”

3. More generally, it might be helpful to more explicitly call out that there are 2 different kinds of problems that data reusers encounter that the authors are going to cover. First are problems with the datasets: they might have poor quality data, missing data/fields, or nonstandard ways of reporting key fields. Second is the fact that different trials have different designs and that can make it hard to pool data in a reasonable way. The first problem is theoretically fixable, the second really isn't. Both require that appropriate metadata be available to help reusers know what they are seeing and how different trials and dataset are or aren't comparable.

We would like to thank the reviewer once more for the useful suggestion. As mentioned in point 1, we have added a sentence to the introduction to be more explicit about the focus of this paper and the issues discussed. We have also added the following to the discussion on page 5: “While accurate description of trials and the data being shared is an issue that in principle can be tackled, handling complexity in trial design is more challenging”.

Reviewer #2: The topic discussed in the manuscript is certainly of major relevance for data sharing. Based upon own experiences with combining and pooling data from multiple different trials, potential hurdles, such as missing annotation, incomplete datasets and trial complexity are discussed and potential solutions addressed. Unfortunately, the manuscript cannot be accepted in its present form because it needs to be put into perspective with recent developments and the updated literature. The main points are:

1. The references about best practices for data sharing are not up to date and mis important documents (e.g. Ohmann et al., BMJ Open, 2017; Cancer Research UK with examples for good practice, https://www.cancerresearchuk.org/sites/default/files/hands_on_data_sharing_advice_-_clinical.pdf). The discussion is much further than suggested in the manuscript.

We would like to thank the reviewer for this comment. We have expanded on this section in the introduction; page 3 “A recent evaluation of issues related to the sharing of patient-level clinical trial data produced practical recommendations for such sharing from non-commercial trials. Other examples for good practice exist, specifically with regards to the sharing of non-trial research data.”
2. This is similar with the discussion on benefits and risks for granting free access to such data. Here, it was recently proposed to provide different modes of access (e.g. open, managed) to cope with the different requirements (e.g. Banzi et al., Trials, 2019, 15:169, Hopkins et al., J Clin Epidemiol 2016; 70:17; Sydes et al., Trials, 2015; 16:104).

Thank you for these suggested references. We have added the following to the second paragraph on page 3: While the benefits of open sharing of clinical trial data are recognised by many in the medical research and health-provider communities, the debate on benefits and risks for granting free access to such data is ongoing with different modes of access proposed as possible solutions to some of the related concerns.

3. The reviewer agrees that detailed information about the original studies is necessary when data are combined across trials. Here, activities trying to link clinical trials with other data objects characterising the trial (e.g. study protocol, DMP, SAP, CRF, interim analysis, reports) are of major importance and should be at least mentioned (e.g. Goldacre et al., Trials, 2016; 17:184).

We thank the reviewer for this suggestion. We have added the following text to page 5 of the paper: “Projects such as that described in Goldacre et al, look to link clinical trials with relevant trial documentation such as protocols, reports, and trial forms, as well as other literature of potential interest. This is a step in the right direction to provide researchers with the critical information needed when reusing trial datasets”.

4. The suggestion that secondary users should collaborate closely with primary clinical trialists has been challenged by others (Ohmann et al., BMJ Open, 2017). This should at least be mentioned.

Thank you for pointing this out; we have added the following to page 6 of the manuscript: “As suggested by Ohmann et al. involvement of data generators is not a necessity but primary data generators should have the option of being alerted about who requests access to their data and when”.

5. It can be agreed to the proposal to impose standards for data sharing but again there are already proposals in the literature, which need to be reflected.

We thank again the reviewer for this suggestion, in addition to the text added on page 3 (“A recent evaluation of issues related to the sharing of patient-level clinical trial data produced practical recommendations for such sharing, however these focus on non-commercial trials. Other examples for good practice exist, specifically with regards to the sharing of non-trial research data.”) we have also added the following to the discussion on page 6: “These have been proposed and lessons can be learned from efforts made to standardize electronic health records data, and recommendations made for non-commercial clinical trials”.