Author's response to reviews

Title: Resource implications of preparing Individual Participant Data from a clinical trial to share with external researchers

Authors:

Catrin Tudur Smith (cat1@liv.ac.uk)
Sarah Nevitt (sjn16@liverpool.ac.uk)
Duncan Appelbe (dea@liv.ac.uk)
Richard Appleton (Richard.Appleton@alderhey.nhs.uk)
Pete Dixon (peted@liv.ac.uk)
Janet Harrison (Janet1h@liverpool.ac.uk)
Anthony Marson (a.g.marson@liv.ac.uk)
Paula Williamson (prw@liv.ac.uk)
Elizabeth Tremain (liz.tremain@nihr.ac.uk)

Version: 1 Date: 30 May 2017

Author’s response to reviews:

We thank the reviewers for their helpful suggestions to improve the manuscript. In response to specific comments:

Reviewer #1: Sharing of patient-level clinical trial data will, inevitably, become standard practice soon. The cost of, and steps required for, preparing datasets ready to be shared on some platform has not been studied, hence this paper fills a gap. The paper is nicely written.

The biggest limitation of this paper is that the cost estimates described, albeit very detailed, may be substantially different in other environments, whether publicly funded or not. In this respect, it is extremely unfortunate that the only figure coming from industry is from a personal communication stating that "an average of 7 days" is typically needed to prepare data for sharing, without any detailed breakdown between the different tasks involved. It is probably the case that some tasks vary enormously in length, especially (as pointed out by the authors) for legacy trials that were conducted a long time ago, and for which data access is far more time-consuming, when at all possible. While minor differences between the tasks and times required to prepare the data for the SANAD and MENDS trials are here described in exquisite detail, it would have been
far more useful to discuss widely different trials for which the tasks and times might have differed more substantially.

Along the same lines, this paper describes publicly funded trials carried out in the UK. It would have been nice, if at all possible, to quote figures (if available, even via personal communication) from other clinical trial data centers in Europe and the US.

- We agree with the reviewer that it would be ideal to increase the sample to include a greater number, and greater diversity, of trials. However, as a first step we believe that the two examples presented give a very reasonable indication of the type of activities and resources required that would be relevant to most ‘typical’ trials regardless of the setting. We have attempted to strengthen our note of caution that these empirical examples may not be representative of all trials and that the resources required for data preparation may differ in other settings. However, until researchers start to record these activities it is difficult to move forward. Our paper provides a starting point for these discussions.

- Unfortunately the experience from industry is only available as a personal communication and we do not have detailed information about the different tasks.

One interesting aspect of this paper is to provide guidance on the steps required to prepare a "data sharing pack" as well as to anonymize the data for sharing. It would be quite useful, in my view, to add a standard table of contents for the data sharing pack (in addition to the steps outlined in tables 1 and 2). Data sharing is a new paradigm, hence any thoughtful guidance on how to operationalize it is welcome at this stage.

- We have developed guidance (1, 2) that includes further detail on the suggested contents of a data sharing pack. We have added a table to the current manuscript (table 1) to highlight the recommendations from this guidance document.

The paper ends with a recommendation to include the costs of data sharing in a trial budget. It is comforting to see that these costs will generally be small compared with the overall costs of a trial, such that costs will not constitute a major obstacle to data sharing in the future. To put things in perspective, could the authors disclose the total costs of SANAD and MENDS?

- This information is publicly available and has been added to the manuscript (SANAD £1.35m; MENDS £1m)

Reviewer #2: Great work by authors to attack this issue. These seem like "typical " trials. Did you intend them to be?

- Yes, we chose these as typical trials co-ordinated by our trials unit but also to reflect the change in trial practice over time
Do you have plans to take a slightly larger sample? Valuable contribution and you now have the skills and process to look at this issue and most will not have the focus and patience to do what you have. Obviously large and more diverse trial set would give robustness to your findings.

- We are aware of a similar project being undertaken at another University. We are also planning to conduct a separate exercise to prospectively log the resource requirements for data sharing across a sample of trials units. We hope that this information will add further to the conclusions of our initial empirical study. We have added a further note of the limitation of this sample to the discussion.

Excellent analysis and to put in perspective the total cost is less than the cost of one patient in a typical clinical trial so the marginal cost is quite small and if made a standard part of the analysis plan as you imply would be a small marginal increase. Also, as you have said in your paper this will get more efficient over time reducing further. If you amortize the cost further over the number of reuses the costs seem quite reasonable.

- We thank the reviewer for these supportive comments

Did you think about the implications of using a common standard such as CDASH and SDTM? This would seem to go along with the efficiency gains you imply when talking about having a common platform.

- We do not currently use CDASH or SDTM at our trials unit and this is typical of other trials units in the UK. However, they are used routinely by the pharmaceutical industry and we are increasingly involved in discussions about the benefits of implementing these standards. We believe that this can only improve the efficiency of data preparation even further and so our estimates might be seen as worst case. We have added further discussion about standards.

One thing that does need attention is opportunity costs. If these skilled individuals are doing this work they are not working on new trials and that could be a cost to research unless we do as you say and shift the burden from the data analysts to the data scientists.

- We have added a sentence to the discussion to highlight that trial statisticians may be required to start work on new projects.
