Author’s response to reviews

Title: Whose data set is it anyway? Sharing raw data from randomized trials

Authors:

   Andrew J Vickers (vickersa@mskcc.org)

Version: 2 Date: 23 March 2006

Author’s response to reviews: see over
Dear Editor

We would like to thank the reviewers for their useful comments. Our response to each comment, detailing where we have made changes, is given below.

Yours truly,

Andrew Vickers

Reviewer: Hutchon.

Comment: No reference was made to Hutchon DJR Publishing raw data and real time statistical analysis on e-journals. BMJ 2001;322:530
Response: Rereading my paper, I see how that I may give the impression that this is all my idea, whereas I am merely building on much prior work, including that of Hutchon. I have now amended the manuscript to cite Hutchon’s paper and some others.

Reviewer: Halpern

Comment: I agree wholeheartedly with Dr. Vickers’ arguments. I am a bit concerned, however, that the tone of his prose may be slightly off-putting to the very people he seems committed to influencing. The "Introduction" paragraph, most prominently, may alienate the casual reader.
Response: I have toned down the initial paragraph, removing pejorative terms.

Comment: In discussing the ownership of data, on p. 7, the author might consider mentioning the practical difficulties of actually granting ownership to trial participants.
Response: I agree that ownership of data by trial participants is a moral point not a practical one. I have added a clarification that trialists might be seen as “custodians” of the data.

Comment: In discussing how to share data, the author might wish to make an additional point that knowledge that data will ultimately be shared might spur greater attention to the cleanliness of a dataset.
Response: This is an excellent point and I have added it to table 1 and the abstract.

Comment: There is one obvious concern with the proposal to make all raw data widely accessible - that this will lead to data dredging, inappropriate reanalyses or subgroup analyses, and a net public health setback as we get mired in paper after paper looking at the same data in new ways. For the author's proposal to be as great a public health boon as he predicts, someone must be charged with policing the uses of the data.
Response: I think there are two separate points here, one concerning the extent of additional analyses and the second concerning their appropriateness. With respect to the first point, I don’t find it particularly plausible that many trial data sets will lead to unreasonably large numbers of new analyses. This is partly because statistical analysis is
difficult, and not that many people can do it, and partly because few trials lend
themselves to very large numbers of different analyses. The second point is of greater
concern: what if an investigator downloaded the trial data, did an incorrect analysis, and
then published results claiming to have refuted the authors’ conclusions? I deal with this
point in a new section entitled Protecting the interests of trialists.

Comment: A related point is that if raw data became so widely available that talented
researchers could build a career on analyzing data produced by others, this might create
an unfortunate disincentive to go through the arduous process of collecting new data.
Response: This is not an important problem in my view. Researchers currently can and
do “build … career[s] on analyzing data produced by others”: meta-analysis is an
important field and it is good that some researchers are dedicated to it. Any researcher
will be judged by the quality and importance of the work they produce.

Reviewer: Jonas Ellenberg

Comment: [The paper involves an] essential dismissal of the sensitivity
of individual investigators to their appropriate yield from their work, [this is] an affront
to fair play.
Response: I accept that the paper did downplay this point. I have now added a section
entitled Protecting the interests of trialists.

Comment: Author needs to confront the workload and expense associated with the
creation of a public domain data set. This will include not only the initial publication of
the data set in user friendly form and ready distribution, but also the continuing costs of
potential updates, responding to inquiries, etc. It is a non trivial exercise to create a data
set that will yield its fruits to investigators having no relation to the original data
collection. Commentary is necessary on whether federal agencies have added dollars to
grant and contract awards for the creation of public data files (including their long term
maintenance and responsibilities for outside inquiries) or whether this has become a new
requirement without compensatory increases in funds.
Response: I am not convinced that creation of a public domain data set is a particularly
expensive or onerous exercise. Good practice would involve the creation of a well-
annotated data set, hence the only additional work would be in de-identification, which
involves changing or removing only a few variables. In particular, I don’t think there is
any necessity for authors to update data (the data set made public should be that analyzed
for the associated publication). If data are sufficiently well-annotated I don’t think there
will be important additional burden to investigators over and above the current
responsibility that they have to respond to inquiries from other investigators requesting
information. Note also that the clinical trial publication should contain most of the
information on experimental methods that any investigator might require to understand a
data set. I have added some comments along these lines to the text.
Comment: Author needs to provide a practical plan for investigators that insure that their investment along with that of the funding agency is fairly compensated. To imply that anything less than complete forfeiture of publication rights would be an offense against the patient participants is neither realistic nor fair. Publications and intellectual productivity are the currency with which investigators survive, for a substantial part of the research community. The author should provide the reader with some broad mechanism that will define a reasonable time and scope of work for exclusivity. This is not an easy task nor will it necessarily yield a universal algorithm.

Response: I have added a section entitled “Protecting the interests of trialists”. Note that I have only recommended data sharing or publication of data sets for which the trialists have published a paper. In my view, therefore, the issue of exclusivity does not arise: the trialists already have an exclusive.

Comment: Author needs to pay some attention to the moral, legal, and other obligations of the provider. This would include liability for errors.

Response: I discuss these obligations on page 8: I don’t feel that “liability for errors” is any different for publication of raw data than for publication of study results.

Comment: Author should address the distinction between the argument presented for clinical trials and the case for observational studies.

Response: We only deal with clinical trials here. If I am missing something, please get in touch and I'll amend!!