Reviewer's report

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

Version: Date: 12 February 2006

Reviewer: Scott Halpern

Reviewer's report:

General

This paper addresses an important and timely issue in the conduct of clinical trials, and is therefore appropriate for a commentary in the journal Trials. The general sentiment, perhaps best surmised by the author's statement that "there appeared little recognition that producers of data are part of a community with reciprocal obligations," is one that needs to be better recognized by trialists.

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

none

-------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

none

-------------------------------------------------------------------------------

Discretionary Revisions (which the author can choose to ignore)

1. 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; to the extent that this happens, the impact of the authors' astute arguments will be regrettably limited. I suggest that the author reserve adjectives like "unscientific" and "unethical" for the paper's conclusion, rather than presenting them in the introduction before the arguments supporting these assertions are made.

2. In discussing the ownership of data, on p. 7, the author might consider mentioning the practical difficulties of actually granting ownership to trial participants. Each patient, for example, could not (practically) be contacted prior to the release of data for each specific purpose. This in no way changes the validity of the author's argument - I merely suggest that the focus should be on recognizing the sentiment that trialists are not the true owners of data, despite the practical requirement that they serve as gatekeepers to data (at least until the data are published in an accessible format, as the author suggests).

3. 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. For example, the process of cleaning the data for sharing might uncover errors in data entry or previous analyses that have led - or could lead - to spurious results and conclusions. Thus, the expectation that data will be shared might have a salutary effect of augmenting the validity of published reports.

4. 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. This person might be the original PI, the editor of the journal in which the original analyses are published, or the funding agency who supported the work. The author might discuss his views on whether such a policing is warranted, and if so, who should be charged with doing it. 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. From a public health standpoint, this would be a regrettable outcome. Although I agree with the author that researchers should recognize their duty to serve the interests of public health and of the patients who participate in their trials, I would point out that the more selfish interests of investigators in promoting their own careers does, at present, motivate the much-needed collection of new data.

**What next?:** Accept after discretionary revisions

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable