Reviewer's report

Title: Access to regulatory data for systematic reviews: the times they are a-changing

Version: 1 Date: 5 August 2012

Reviewer: Erick Turner

Reviewer's report:

When assessing the work, please consider the following points:

1. Does it address an important or timely issue?
   Yes, this issue is both important and timely.

2. Is it well reasoned?
   The answer to this question depends on whether the authors are correct in their interpretation of the ICH E3 definition of a clinical study report (CSR). The authors state, "...making full CSRs available and providing raw data are two separate issues." But are they really? I had always thought of a study report as including the raw data.

   The authors state, "We define a CSR following the International Conference on Harmonisation (ICH) E3 guideline [13], according to which a CSR, in addition to containing the full protocol and summarized efficacy and safety data, also contains pseudonymized patient data listings, but not the full raw data set." I found this confusing. If indeed you have data listings at the level of individual patients, as opposed to summary data, is that not the raw data?

   I thought I would check the E3 guideline for myself. Searching the PDF, neither the phrase "raw data" nor the word "raw" appear anywhere. Using the terms "individual patient" and "patient data", I found the following. On page 9 of 55, ICH E3 states, "...and all individual patient data (archival listings requested only in the United States) should be provided in Appendix 16.4." (I admit that I do not understand the word "archival" in this context and thus whether it is relevant to this discussion.) On page 24 of 55, there is the header, "11.4 Efficacy Results and Tabulations of Individual Patient Data", suggesting that tables containing data for individual patients are a standard part of a study report. On page 55 of 55, at the very end of the E3 document, it states, "In the report of each controlled clinical study, there should be data listings (tabulations) of patient data utilized by the sponsor for statistical analyses and tables supporting conclusions and major findings. These data listings are necessary for the regulatory authority's statistical review, and the sponsor may be asked to supply these patient data listings in a computer-readable form." Are not such data from individual patients not the same as raw data?

   Or could it be that, when the authors distinguish "pseudonymized patient data
listings" from "full raw data set", the key word is "pseudonymized"? Do they define "full raw data" as including patient identifying information? I do not believe that most people would define it this way. Recalling earlier experiences as a site investigator in clinical trials, I do not believe that patient identifying data are transmitted from the trial site to the drug company. Not only would such transmission comprise a risk to patient confidentiality (Would this not, in the US, violate HIPAA regulations, and is it not prohibited by IRBs?), but there is no scientific need for such information.

But perhaps I am wrong. Although I could not find this question addressed in either the referenced article by Vickers or the one by Hutchon, I did stumble upon a recent article that seems to suggest that raw data can, in fact, contain patient identifying information.[1] (This article may need to be referenced. As an aside, glancing at this article, it states that HIPAA provides the most explicit guidance on identifying information.)

Having said that, I cannot understand why any researcher would insist that he or she needed patient identifying information and not be content with de-identified data.

The manuscript will need to be clarified with respect to the above questions.

3. Is it relatively balanced, or does it make plain where the author's opinions might not represent the field as a whole?

The authors do clearly state that others do not agree with their position. However, as a discretionary revision, they could elaborate on the main arguments of those others.

4. Is the standard of writing acceptable?

Some, but not a great deal, of copyediting will be necessary.

Beyond the copyediting issues, the authors could make their central argument somewhat more prominent. For example, in the concluding sentence, where they say, "...let's hope it is not unnecessarily delayed", they might add "...by an insistence that raw data be included in CSRs." (However, see above regarding the need for clarification on this point.)

Please divide your comments into the following categories:

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

There are two major issues with this paper in its current form. The first, elaborated on above, is that the definition of the term "clinical study report" with respect to the type of data included.

The other major issue with this paper is the fact that, regarding its scope, there is a mismatch between the title ("Access to regulatory data...") and the manuscript text. The title suggests that the paper will be about data from regulatory agencies in general. However, the paper focuses on only one regulatory agency, the EMA.
The FDA is not mentioned anywhere in the manuscript, and, if the paper is intended to cover the availability of data from regulatory agencies in general, the FDA should definitely be included. The FDA has been posting Drug Approval Packages on its website since 1997, and data from years prior to that have been and are accessible, through the US Freedom of Information Act. And, to borrow from the title and Bob Dylan, "the times they are a-changing" not only for the EMA, but for the FDA, as well. The FDA convened a Transparency Task Force a few years ago, recommendations have been made for broadening data access, and those recommendations are under considerations.[2] (However, in my opinion, this Transparency Initiative seems to have lost some momentum since the referenced article was written.)

So it seems there are two solutions to this scope issue: either the title should be revised from "Access to regulatory data..." to something like "Access to data from the European Medicines Agency..."., or the scope of the manuscript text should be broadened to fit the title. The simpler option would be, of course, to use the more narrow title. In this case, much less revision would be required, in my view. On the other hand, if the intent is to cover regulatory agencies in general, it seems that there is a great deal of material that should be added. If the scope of the manuscript is broadened so as to include the FDA, then it could me mentioned that FDA data have been used in publication bias studies conducted by myself[3,4] and other researchers.[5-7]

Within Europe, non-EMA regulatory agencies that could be mentioned, but again, only if the scope includes regulatory agencies besides the EMA. For example, Swedish [8] and Swiss [9] regulatory data have been used to examine questions related to the efficacy of antidepressant medications. However, the authors might wish to check with these authors to find out whether transparency initiatives are underway in those agencies, as well. My hunch is that, in Europe, the EMA is substantially in the lead with respect to transparency issues, but this would need to be verified.

Again, however, the suggestions immediately above can be ignored by simply making the title more narrow, so that it is clear to the reader that the scope is confined to the EMA.

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

Abstract, first sentence: "prevent" should be changed to "detect" or to "correct for". Access to regulatory data will not prevent publication bias.

Abstract, first paragraph, last sentence: "other researchers report" should be "other researchers have reported".

Abstract, third paragraph, second sentence: The ICH definition of a CSR is not provided until the Discussion section. I would provide a shortened version of it here (and also reference it), perhaps saying that, according to the ICH definition,
a CSR does not necessarily require the inclusion of raw data.

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

Please see above. If the scope of the manuscript it to be confined to the EMA and the title is modified accordingly, then the above comments regarding non-EMA drug regulatory agencies can be ignored.

References:


Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Declaration of competing interests:

I declare that I have no competing interests.