Author's response to reviews

Title: Access to regulatory data from the European Medicines Agency: the times they are a-changing

Authors:

Beate Wieseler (beate.wieseler@iqwig.de)
Natalie McGauran (n.mcgauran@iqwig.de)
Michaela F Kerekes (florina.kerekes@iqwig.de)
Thomas Kaiser (thomas.kaiser@iqwig.de)

Version: 2 Date: 10 September 2012

Author's response to reviews: see over
Dear Systematic Reviews Editorial Team,

Attached please find the revised version of our manuscript MS: 1007539543741308 (Amended Title: Access to regulatory data from the European Medicines Agency: the times they are a-changing)

We have considered all of the reviewer’s comments (please see point by point response below as well as the “compare documents” version of the manuscript). We have also considered the comments by the Editorial Team (we have included an Acknowledgements section and corrected the formatting of the references).

Thank you for considering our manuscript for publication.

Best regards
Beate Wieseler

Dr. Beate Wieseler
Head of Department
Dept. Drug Assessment

Tel.: +49-221/35685-254
Fax: +49-221/35685-1
E-Mail: beate.wieseler@iqwig.de

Institute for Quality and Efficiency in Health Care (IQWiG)
Im Mediapark 8
D-50670 Köln

Germany

Director: Jürgen Windeler

Internet: www.iqwig.de
Health Information: www.informedhealthonline.org
<table>
<thead>
<tr>
<th></th>
<th>Reviewer's report</th>
<th>Authors’ response</th>
</tr>
</thead>
</table>
| **1** | **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** |  
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. | We agree that our distinction between CSRs following ICH E3 and CSRs supplemented by raw data was unclear. We hope to clarify the issue by the changes suggested below. |
| **3** |  
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 | Our article refers to the publication by Eichler et al. who provide a view “on the call for openness from a European Union drug regulatory perspective”. Our definition of raw data is in line with the statements by Eichler et al.:  
When addressing “important benefits from public disclosure of raw trial data” Eichler et al describe “large, information-rich datasets needed to support the computer science and artificial intelligence research” required to develop applications needed, for example to allow better individualized therapeutic decisions based on a comparison of a patient’s health record with trial data. This would clearly require an electronic database of trial data from individual patients. Such a format allows a computer-based analysis of the data. In contrast to this format of an electronic database, a CSR is in principle a “paper-based” format. Even if the CSR is provided as a PDF file, including not only text and |
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?

summary data but also tabulations (or listings) of individual patient data, this format could not be used to perform computer-based analyses of these individual data. We therefore would not refer to these listings of individual patient data in the CSR as “raw data”. (They could theoretically be converted to raw data by manually re-entering the data from the CSR into a database. However, due to the vast amount of resources required, in practice this is not very realistic.)

Our understanding of the format and use of listings of individual patient data from CSRs derives from our own experience in the writing of CSRs for submission to regulatory authorities and the use of CSRs as a data source for systematic reviews. Furthermore, the ICH E3 guideline describes the format and intention of the CSR in the “INTRODUCTION TO THE GUIDELINE (page 1 of the document)” as follows:

“The clinical study report described in this guideline is an "integrated" full report of an individual study of any therapeutic, prophylactic or diagnostic agent (referred to herein as drug or treatment) conducted in patients, in which the clinical and statistical description, presentations, and analyses are integrated into a single report, incorporating tables and figures into the main text of the report, or at the end of the text, and with appendices containing the protocol, sample case report forms, investigator related information, information related to the test drugs/ investigational products including active control/comparators, technical statistical documentation, related publications, patient data listings, and technical statistical details such as derivations, computations, analyses, and computer output etc.”

From our point of view, this description underlines the intended paper-based format of a CSR.

In addition, as noted by Erick Turner, at the very end of the ICH E3 guideline it states:

“Format and Specifications for Submission of Data Requested by Regulatory Authority’s Statistical Reviewers: In the report of each controlled clinical study, there should be data listings (tabulations) of
patient data utilised 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.”

This section of the guidelines clarifies that patient data listings in a computer-readable form (i.e. as an electronic database) are not included as a standard but are provided on request only. Thus, the computer-readable format is a potential add-on to the CSR, not an integral part of it.

The availability of individual patient data in the format of an electronic dataset as an add-on to CSRs seems to vary between different regulatory agencies. In contrast to the FDA, EMA does not accept SAS data sets* and thus does not allow for submission of an electronic format of individual patient data that could be used for computer-based analysis.


The understanding of raw data being a dataset in an electronic format that can be used for computer-based analysis is also supported by other authors. In their paper on “Preparing raw clinical data for publication...” Hrynaszkiewicz et al. (2010) also define raw data as an electronic format suitable for computer-based analysis when they describe the requirements for making raw data available (on page 2 of 5):

“File preparation ...Authors should provide a clean, well annotated dataset in a suitable format so that statistical analyses could be conducted. By “clean,” we mean reviewed systematically for duplicates,
errors, and missing data; by “well annotated,” we mean that sufficient information is given about each variable to allow replication of the originally published results. For example, the dataset included as supplementary material by Vickers [5] includes a brief description of the study and data and a detailed explanation of each variable on the dataset. It is recommended that file formats be as general as possible. Microsoft Excel is widely used and delimited text format is universally convertible, so these formats are preferable to files saved in formats specific to statistical software such as SAS or STATA. If a dataset may be updated in the future—for example, in cancer studies where follow-up is continued over many years—it could be given a version number or date.”

For clarification we have amended the abstract and discussion as follows:

Abstract
We define a CSR following the International Conference on Harmonisation (ICH) E3 guideline, which does not necessarily require the inclusion of the full raw data set, i.e. an electronic database allowing for computer-based analysis.

Discussion (p. 7)
We define a CSR following the International Conference on Harmonisation (ICH) E3 guideline [13]. According to ICH E3, a CSR, in addition to containing the full protocol and summarized efficacy and safety data on all outcomes, also contains individual patient data in the format of tabulations or listings (usually provided in PDF files), but not the full raw data set, i.e. an electronic database allowing for computer-based analysis. (This type of format is only intended to be made available by sponsors on request but is not a regular component of the CSR [13]). However, the above regulators extend this definition of a
| 6 | **CSR to include the full raw data set, which on the basis of the suggested applications would have to be provided as an electronic database.**  
We would prefer not to go into more detail about electronic formats in the commentary and, for example, choose not to address on SAS formats and differences between regulatory agencies, as we feel that this would go beyond the scope of the paper. |
|---|---|
| 4 | 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.  
We do not consider full raw data as data including patient-identifying information and agree that there is no scientific need to transmit such information. We hope the issue is clarified by our definition of raw data above “…the full raw data set, i.e. an electronic database allowing for computer-based analysis.”  
We used the adjective “pseudonymized” as a technical term to describe the fact that the individual patient data in a clinical trial are not fully anonymized and that in principle a patient could be identified using additional information from the trial site. However, the term “pseudonymized” is not relevant in the context of our manuscript and we have therefore deleted it. |
| 5 | 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 suggests 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.  
Hrynaskiewicz et al. address the problem that patients could be identified from a set of raw data by combination of indirect identifiers such as gender, age and trial site (e.g. in trials of rare diseases).  
However, the issue of patient confidentiality is not specific to raw data, but also potentially applies to (paper-based) patient data listings and has also been addressed by pharmaceutical companies in our discussions on making CSRs following ICH E3 available.  
We have clarified this point in the text. |
| Discussion (p.8) | **Patient confidentiality has been a major concern in the discussion about making extended information (including individual patient data) from clinical trials available. Since CSRs also include individual patient data** |
listings, in principle, this issue not only applies to raw data but also to CSRs. However, as stated by EMA, “current European legislation requires patient information to be included in non-identifiable form in the marketing authorization application submitted to competent authorities” [12]. Given these requirements, it seems unrealistic that data listings in a CSR would contain patient-identifying information, although the risk might be higher for raw data sets, where patient characteristics can be re-arranged and combined electronically. As Eichler et al. note, there might be exceptional cases (for example, trials in “ultra-rare” diseases) where it could be difficult to ensure patient confidentiality. In such cases, a simple preliminary solution would be to split off individual patient data listings before the release of CSRs. This would allow discussion of measures to ensure patient confidentiality without delaying the release of large parts of the information on these trials.

### 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.

As Erick Turner notes, we clearly state that there are differing views on the issue of the format of CSRs to be made publicly available.

In addition, we have included the following sentence:

**Discussion (p. 7)**

We acknowledge the potential advantages of analysing raw data and that making both CSRs and raw data publicly available in the near future would be a huge step forward. We thus understand the rationale for this proposal.

### 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.)

In addition to incorporating Erick Turner’s suggestions, we have made a number of editorial changes (all indicated in the “compare documents” version)

We have also amended the concluding sentence:

“As a result of EMA’s policy change, a milestone for data transparency in clinical research is within our reach; let’s hope it is not unnecessarily delayed by an insistence that raw data be released together with CSRs.”
| 8 | 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. |
| 9 | 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 necessary. |
| | Please see out comments above and the corresponding changes to the text. |
| | Our commentary specifically focuses on EMA and the article by Eichler et al. Inclusion of the FDA would go beyond the scope of the manuscript. We agree that the title is inaccurate and have amended the title, running head and keywords. (We have abbreviated the title elsewhere as otherwise it would be too long.). |
| | **Amended title**
Access to regulatory data from the European Medicines Agency: the times they are a-changing
Access to regulatory data from EMA

| Keywords: “European Medicines Agency” added |
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 be 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.

<table>
<thead>
<tr>
<th>10</th>
<th>Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abstract, first sentence: &quot;prevent&quot; should be changed to &quot;detect&quot; or to &quot;correct for&quot;. Access to regulatory data will not prevent publication bias.</td>
</tr>
<tr>
<td></td>
<td>Abstract, first paragraph, last sentence: &quot;other researchers report&quot; should be &quot;other researchers have reported&quot;.</td>
</tr>
<tr>
<td></td>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11</th>
<th>Discretionary Revisions (which are recommendations for</th>
</tr>
</thead>
</table>
|    | We have considered all of Erick Turner’s suggestions in our changes to the abstract (except for the inclusion of a reference; according to the authors’ guidelines for Systematic Reviews, references are not allowed in abstracts).
|    | Changes to abstract                                                                                                               |
|    | ...to correct for publication bias                                                                                                  |
|    | Other researchers have reported                                                                                                     |
|    | We define a CSR following the International Conference on Harmonisation (ICH) E3 guideline, which does not necessarily require the inclusion of the full raw data set, i.e. an electronic database allowing for computer-based analysis. |
|    | The extended abstract exceeds the word limit and as the submission                                                              |
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.

<table>
<thead>
<tr>
<th>11</th>
<th>References:</th>
</tr>
</thead>
</table>

- We have added the publication by Hrynaszkiewicz et al. References 2-9 refer to the FDA and the Swedish regulatory agency. (Please see Point 9).
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td><strong>Level of interest:</strong> An article of importance in its field. <strong>Quality of written English:</strong> Needs some language corrections before being published. <strong>Declaration of competing interests:</strong> I declare that I have no competing interests.</td>
</tr>
<tr>
<td>13</td>
<td><strong>Additional comment by authors:</strong> EMA has announced that it will be holding a workshop on access to clinical trial data and transparency in November 2012. We have added a sentence to the conclusion and included the citation. <em>The debate between EMA, its stakeholder groups and interested parties on access to CSRs, including the issue of the level and format of data to be provided, is still ongoing [22].</em></td>
</tr>
</tbody>
</table>