Final advice to the European Medicines Agency from the clinical trial advisory group on good analysis practice

20 March 2013

Introduction

This is the fourth of five consultative groups related to the planned release of clinical trial data by EMA to third parties. The groups cover the following topics:

1. Protecting patient confidentiality
2. Clinical trials data formats
3. Rules of engagement
4. Good analysis practice
5. Legal aspects

The following report is made without individual attribution of opinions. It should be noted that the opinions contained in this report are those of members of the consultation group as listed at Annex 1. These opinions are not necessarily compatible but, at this stage of the consultation process, it was intended to reflect a full range of views. This document, together with those from the other four groups, will be used as a basis for a draft policy that will be subject to further public consultation.

Preliminary points

The following remarks apply to requests for data from groups involved in scientific research. Analysis of individual patient data requires specialist skills and biased conclusions may be drawn from selective inspection of events recorded in the data. It is recommended that an experienced epidemiologist or statistician be involved in all such analyses. Note that the involvement of a statistician in clinical trial analysis is a recommendation of ICH E9 and it is reasonable that similar expertise is required in other analyses of health data.

There are several potential uses of these data which may require access to somewhat different aspects of the data and metadata. This was discussed in a discursive way during the meeting but a more structured classification and commentary was offered after the meeting by an academic group. It may help to put the discussion into context and is reproduced in full below.

Roughly, the benefit of publication of clinical trial data at patient level can be classified in two types: (I) the opportunity for validation[1] of the main results and (II) reuse of clinical trial data for secondary research. The recommendation of analysis standards needs to be tailored to the respective objectives of the analyses:

[1] The word 'validation' is here, and elsewhere, used to refer specifically to the methods of analysis and not to checking the accuracy of the data.

I. The opportunity for validating the main results and for investigating their robustness. This would require access also to the clinical trial protocols (including amendments), statistical analysis plans together with software codes, data dictionaries and the study reports. For the validation purpose no prospective protocol seems to be necessary, as it will be guided by the original statistical analysis plan. However, the report on the validation should include all the necessary details to retrace the methodology applied. Standards should not be prohibitive and go beyond requirements for the original applicant. Still, in the first place this validation remains the responsibility of regulatory authorities in time during the assessment procedure.

II. Reuse of clinical trial data for secondary research. The scope of such research may range from quasi prospective research to full data mining, providing different levels of evidence.
   a. Higher levels of evidence will be provided by studies with a protocol written before access to the data. Meta-analysis based on individual patient data may be an example for such a higher level type of research. However, generally when planning secondary research projects, there will be study results already available, either from publications, reports or from other groups having access to the data, so that it will be difficult to exclude post hoc definitions of research objectives (e.g., resulting in hunting for significance). Early publication of protocols for secondary research, ideally before unblinded data of the phase III become available, would enhance credibility and persuasiveness of the planned secondary analyses. In any case, the protocol and resulting publications should clearly refer to time lines of data access and background knowledge available when formulating the research objectives - even though this will be difficult to verify independently.

   b. Full explorative discovery using data mining methodology provides a lower level of evidence but may reveal new and useful results. These results in general will have to be confirmed by further research and therefore such research projects have to be clearly identified as explorative when communicating results.

   The clinical trial protocol (including amendments) and the data dictionary should be accessible for all secondary research since it is hardly imaginable that anyone could otherwise perform and interpret analyses meaningfully. To avoid biased results due to incomplete access to data sets, administrative hurdles to get access to the data should be minimized.
   In general, it should be avoided that over-sophistication of protocol standards in secondary explorative research disguises its limited level of evidence. Therefore in all publications the nature of the research should be described clearly to assure an appropriate interpretation, for example by prominently indicating the source of the data and its secondary use.

Overall, the quality of both, validation and secondary research undertakings, will depend on the availability of data sets containing original measurements (in contrast to heavily pre-processed analysis data sets).

Protocols for new analyses

Need for protocols

There was consensus that a formal protocol is desirable. Although reservations were expressed that intentional misrepresentation of the data cannot be wholly avoided the use of a protocol was generally seen as a tool that can facilitate the interpretation of the large majority of research and provide some defence against erroneous conclusions related to multiple analyses. The results of a prespecified analysis should always be accorded greater confidence than an analysis that may be prompted by idiosyncratic features of the data unless it can be shown to be flawed.

An important technical point was that writing a protocol requires a detailed knowledge of the data fields that are available but should be independent of the actual values observed. Hence, if the potential data recipients wished to follow best practice in an overtly verifiable way, the original protocols and data

descriptions for the studies providing the datasets would have to be released to them prior to the datasets and a protocol finalised before the data were received.

The provision of a protocol template, including section headings corresponding to the recommended guidance documents, may help to ensure that the protocols follow the recommended guidance and provide sufficient detail. Similarly a template for statistical analysis plans might be provided.

The goal of having prospectively defined analysis methods prior to seeing the data cannot be wholly achieved since the author of the protocol will have seen the results of the studies submitted for regulatory approval. As such, the protocol should include any results from the company analysis relevant to the specified analysis to identify what was known about the data prior to specifying the analysis.

Public access to protocols

A majority of those who expressed a view said that protocols for secondary research should be publically accessible. This was seen both as a way of confirming the pre-specification of hypotheses and of inviting constructive dialogue on the study design. However, some strong reservations were expressed based on the potential use of litigation by companies to prevent legitimate research. Thus consideration should be given to providing a repository for protocols that would be made public upon completion of the study.

One researcher expressed the view that protocols for the primary research studies should also be made public before the studies were performed.

Review of protocols

The people with most detailed knowledge of the clinical trial data will be those involved in their production. It was thus suggested that review of protocols by pharmaceutical company statisticians should be invited. Opinions for and against this suggestion were robustly expressed. A more moderate view was also expressed by several people that review should be voluntary and that it would follow naturally if transparency was favoured.

Probably the strongest recommendation the EMA can make based on the views expressed is that researchers should seek opportunities to get informed review of their protocols. The point was made that transparency would be enhanced if all exchanges of views regarding the protocol were made public and suggestions for changes that were and were not adopted were recorded in an easily accessible format.

The point that central review of protocols (perhaps by EMA) might be desirable was made. It was acknowledged that this may not be possible with the current levels of resources available but it should be considered an aspiration.

A number of those present would have liked to see central review and approval as a prerequisite for data release. It was explained that this is not possible within the legal framework governing provision of the data and, moreover, could be interpreted as a form of censorship.

Guidelines for analysis

The group discussed whether available guidelines on good practice in analysis and checklists for quality of analyses should be recommended for data recipients.

It was noted that such guidelines are already an accepted part of research and that appropriate use of them is expected by peer reviewed journals. Hence formal endorsement by EMA is unnecessary.

The comparative dearth of guidelines in secondary research was noted and EMA suggested that the ENCePP Code of Conduct and Guide on Methodological Standards in Pharmacoepidemiology might be worth considering. These will be circulated.

CONSORT fulfils a well-established role in research reporting and was mentioned in written comments from two members.

EMA should note that researchers should be aware of relevant regulatory and scientific guidelines and apply them.

It was noted that analyses conducted to support marketing authorisation applications and regulatory pharmacovigilance processes are, in many cases, subject to regulatory guidelines designed to ensure that an acceptable standard of evidence is achieved. Analyses by recipients of the data intended to address the same hypotheses as the original research should implement the same guidelines. If regulatory guidelines are not followed in secondary analyses then this needs to be clearly stated in the protocol and the publication and a scientific justification needs to be provided.

Open access to codes and interim datasets

It was explained that the intention of this point was to promote full transparency and verifiability of the analyses by ensuring that all datasets generated from the data supplied by the EMA and all computer codes used to transform or analyse the data were made available. Given the appropriate computing environment, the whole analysis should be reproducible on the basis of these items.

There was strong support for this level of transparency and some additional points were made:

- The analysts might supplement the data from other sources. It would be important that these additional data were also revealed.

- Some analytical processes produce large datasets – for instance multiple imputation or bootstrapping. EMA opinion that all datasets required for replication of the analysis should be made available. (Note: Repeated runs of appropriately large resampling or simulation exercises should give similar results but only the datasets as used will allow exact replication.)

A caveat was raised with respect to giving public access to code. It was commented that the data might be used to develop commercial code that could not be made public. Assuming that no statements of immediate public health concern arose out of such work, it would be reasonable to waive the requirement for provision of code.

A further point made with respect to transparency was that it would be desirable to have a list of all requests for data and names of the organisations/persons making the request. This falls more obviously within the remit of Rules of Engagement and it was noted that this point should be added to the agenda for Group 3.

Providing access to codes and interim data should in no way alter the usual ethical requirement to publish the results of research. In this respect it was also noted that the agency could encourage good practice in several ways:

- Anonymous publication of results – as sometime happens on web pages – should be strongly discouraged.
- A forum should be provided for public review of these and other results.
- A document informing those requesting data of the expected standards of analysis and transparency should be prepared.

General point regarding this consultation

It was noted that language across the five consultation groups varied and that clarification of terms such as requestor, researcher, secondary research, new analyses, retrospective analyses, analysis plans (by requestor differentiated from by original data collector) and statistical analysis plan would be useful. The Agency should provide a glossary for the open policy consultation.