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

27 February 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:

Protecting patient confidentiality

Clinical trials data formats

Rules of engagement

Good analysis practice

Legal aspects

The following report is made without individual attribution of opinions.

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.

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 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:

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.

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.

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.

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

1.1 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.

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.

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.

1.2 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, a strong reservation was 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.

1.3 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.

The opinion was expressed that EMA should note that researchers should be expected to be aware of relevant guidelines and apply them.

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.

1. ANNEX 1 List of attendees

European Medicines Agency
Hans-Georg Eichler – Medical director
Frank Petavy - Statistician
Jim Slattery – Chair
Urszula Piotrowska - Support

Other organisations

Roberto D'Amico	Academia
Kay Dickersin	Academia
Peter Doshi	Academia
Anthony Johnson	Academia
Kunal Merchant	Academia
Martin Posch	Academia
Karen Robinson	Academia
Jana Skoupa	Academia
Leslie Huson	Consultant
Paul Smith	Consultant
Alexis Clapin	Healthcare professional
Javier Garjón	Healthcare professional
Eugene Milne	Healthcare professional
Gisela Schott	Healthcare professional
Carla Souza	Healthcare professional
Aart Van der Molen	Healthcare professional
Roland Gordon-Beresford	Healthcare professionals' organisation
Christiane Abouzeid	Industry
Manfred Beleut	Industry
Helga Blasius	Industry
João Duarte	Industry
Eric Genevois-Marlin	Industry
Christoph Gerlinger	Industry
Merete Joergensen	Industry
Sören Kristiansen	Industry
Hans-Juergen Lomp	Industry
Duncan McHale	Industry
Toby Lasserson	NGO
Adam Jacobs	Other/Unknown
Kieran Breen	Patients' organisation
Laila Abdel-Kader Martin	Payer / HTA
Ralf Bender	Payer / HTA
Mirjam Knol	Public health organisation
Regine Lehnert	Regulator
Mary Ann Slack	Regulator

Peter C Gøtzsche	Research institute
David Carroll	Student

Annex I - Comments from participants below may or may not have been made on behalf of the organisation they are affiliated with.

Comment form

Line Number	Comment and Changes proposed	Name	Affiliation
N/A	Comment: I believe one important point is missing (although it was not discussed at our t-con): ICH E9 (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002928.pdf) in its section 1.2 stipulates that the statistical analysis of a clinical trial should be carried out by a qualified statistician. On behalf of the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) I suggest that the same requirement should be applied to secondary analyses of the clinical trial data. EMA reply: Document changed to reflect need for statistician/epidemiologist	Christoph Gerlinger	European Federation of Statisticians in the Pharmaceutical Industry(EFSPI)
N/A	Comment: Just a comment re : your document you might want to consider, as I see it a time bomb. It is written "There are several potential uses of these data which may require access to somewhat different aspects of the data and metadata." […] "Roughly, the benefit of publication of clinical trial data at patient level can be classified in two types: (I) the opportunity for validation of the main results and (II) reuse of clinical trial data for secondary research." "opportunity for validation of the main results" are quite strong words directly pointing at the immediate roles and responsibilities of EMA and other health authorities…which is not the intent	Roland Gordon-Beresford	Cardio3 BioSciences S.A

Line Number	Comment and Changes proposed	Name	Affiliation
	May I suggest something like "the opportunity for further evaluating clinical data in the light of accumulating information" EMA reply: I am not sure that I understand your point here. It is, of course, true that one function of ourselves and national regulatory authorities is to ensure that every assertion that goes into the dossier for a MAA is evaluated with some care. However, in that process we actively invite the views and expertise of a much wider audience. We do not believe we are infallible and we strongly believe that promoting a lively and wide ranging debate of the issues in the scientific and wider community is the best way of approaching a good understanding. Hence I think the text as it stands is not incompatible with our views. Having said that, it did not arise from us but is a verbatim quote from an academic group involved in the consultation. Hence if you feel it needs some qualification I would be happy to put in some separate words to that effect. Since the initial request for consultation, the coordinator has kindly supplied me with a form for feedback which I attach in the hope it may help to focus the process. Comment: Sure, sorry for not being clear. Let me try to explain better. (this is also summarized in the comments form attached) The word 'validation' is reserved in clinical data management. For example, "ICH GCP 5.5.3 When using electronic trial data handling and/or remote electronic trial data systems, the sponsor should: (a) Ensure and document that the electronic data processing system(s) conforms to the sponsor's established requirements for completeness, accuracy, reliability, and consistent intended performance (i.e. validation)."		

Line Number	Comment and Changes proposed	Name	Affiliation
	In general terms 'validation' is about performing the required test(s) to support the reliability of the results (i.e. it allows discriminating unreliable results). I don't think anyone suggested that clinical results obtained in the EU (where Dir 2001/20 applies) carry a systemic risk of not being GCP compliant, thereby requiring external validation in addition to that of Sponsors and Competent Authorities. In my opinion, the sentence I was referring to carries the potential for being interpreted such as to claim that, in EMA's own words, Good Analysis Practice is not compliance to GCPs, Dir 2001/20/EC and the many other normative references, and compliance with these texts IS NOT sufficient for claiming 'validated' results (fraud or misconduct being excluded). In my opinion what is meant to be said is that validated (i.e. 'reliable') results could become inaccurately interpreted. That is the point. Are we validating the results or their interpretation? One need not repeat the calculations with the original raw data being provided, these numbers ARE reliable, they ARE validated. What is interesting though, is to understand HOW these results were obtained and calculated (the sciences of statistics) and HOW they are interpreted at a given point in time and knowledge (the science of medicine). In that respect, allowing new calculations being carried in the light of accumulated evidence is the very essence of science that we all support. But here again, what are we really meaning? If it is about re-assessing a supplemented dataset of a single molecule, then most likely this activity falls under the realm of pharmacovigilance and other post-marketing commitments/PSUR/other activities. We should not suggest this is not sufficient and that 'external validation' is somehow actively or passively, suggested. This is a benefit/risk judgment call for a specific product that Competent Authorities only can make reliably, as per their mandate to protect public health. If it is about merging datasets for data-mining in the search for new evidences, then it is not a validation of existing results, it is scientific research. But then (and I appreciate this goes beyond my initial mail) the new results of this re-assessment should also be validated like the pre-existing results, per the GCP. There is no good reason why re-assessment activities should be exempted from the ICH E6 introductory paragraph we all abide to "Good Clinical Practice (GCP) is an international ethical and scientific quality standard for designing, conducting, recording and reporting trials that involve the participation of human subjects. Compliance with this standard provides public assurance that the rights, safety and well-being of trial subjects are protected, consistent with the principles that have their origin in the Declaration of Helsinki, and that the clinical trial data are credible." (emphasis added). This includes all clinical trial data reporting; it is not limited to the original Sponsor. Again, I believe this was the intention.		

Line Number	Comment and Changes proposed	Name	Affiliation
	EMA reply: That is much clearer. I am sure the authors did not mean to imply validation of the data. However, they probably did mean validation of the analysis. They are all statisticians and will have had in mind is examining the assumptions of the analysis by using different models in which to test the hypotheses. For example using multiple imputation rather than last observation carried forward to handle missing data. This is not an implied criticism of the original analysis. There are generally a fair number of possible modelling assumptions and we generally hope that sufficient data will make the conclusions robust to sensible variations in them. However, examination of the fit of the model may sometimes lead you to the conclusion that a particular model is less appropriate than another and that may be worth knowing. In short, I take your point and agree that some clarification is required that it is not the adequacy of the data that is the primary target of the re-analysis. However, neither is it necessarily re-evaluation in the light of accumulating evidence. What they are really discussing is robustness of the results to modelling assumptions. Can I insert a note to that effect? Comment: Being not a statistician, I have no objection. Being a regulatory person, I can only caution about robustness being raised as grail of clinical excellence. Lack of statistical robustness (i.e. fragility) means more results are desirable to build statistical confidence in the conclusion drawn from tested hypotheses, not that results or the conclusions are wrong or wrongly interpreted. Statisticians and we, the well-versed persons of goodwill know this very well. Once in public domain, this document will, and is meant to, escape this wise community. Surely, you will find the right balance.		
N/A	Comment received; awaiting permission to publish	Peter Gøtzsche	Nordic Cochrane Centre

Line Number	Comment and Changes proposed	Name	Affiliation
N/A	Comment: In CTdataGroup4 (promoting good analysis) and CTdatagroup3 (rules on engagement), there is always two global thoughts : 1. requesters are to satisfy various rules or requirements 2. it would be better if requesters could satisfy various rules or required (but without obligation). This thought includes : no obligation at all to get the data. A long list of rules and requirements has been proposed (mainly : give names, have qualifications, give a protocol, get validation for the protocol…) It seems, considering some remarks from EMA on the legal framework, that it will be difficult to only give data to a certain kind of people fulfilling specific characteristics (a patient can ask for the data just for information and probably will not publish). If it is the case, a large part of our discussion is useless. May be should we wait for this information on the legal framework before any other meeting. If the legal framework allows to put mandatory rules, then, another meeting will probably be necessary. If nothing mandatory can be done, discussion will be probably shortened and perhaps unnecessary. I would thus propose to wait for this information before any new meeting for the CTdatgroup4. For CTdatagroup3, examples and case scenarios how confidential commercial information from CSRs could be used for unfair competition and/ or prejudice to regulatory data protection, patent or other IP rights and some real-life examples of "unintended commercial uses" should be given during the next CTAG3 session. If examples are given, opportunity to discuss these example should be given to group through another meeting. If no examples are given, the discussions on other points could be shortened if we have the information on the legal framework. EMA reply: See discussion with Eric Genevois below	Alexis Clapin	Neurologist, Paris
N/A	Comment received; awaiting permission to publish	Eric Genevois	Sanofi
73	Comment received; awaiting permission to publish	Sören Kristiansen	Takeda Pharmaceuticals International

Line Number	Comment and Changes proposed	Name	Affiliation
			GmbH
142	Comment received; awaiting permission to publish	Sören Kristiansen	Takeda Pharmaceuticals International GmbH
150	Comment received; awaiting permission to publish	Sören Kristiansen	Takeda Pharmaceuticals International GmbH
N/A	Comment received; awaiting permission to publish	Sören Kristiansen	Takeda Pharmaceuticals International GmbH
170	Comment received; awaiting permission to publish	Sören Kristiansen	Takeda Pharmaceuticals International GmbH
N/A	Comment received; awaiting permission to publish	Hans-Jürgen Lomp	Boehringer Ingelheim Pharma GmbH & Co. KG