Advice to the European Medicines Agency on rules of engagement for accessing clinical trial data

Draft – 27 February 2013 - Version 5.0 (versions 2.0 to 4.0 internal drafts)

Preliminary comment:

EMA should only disclose confidential commercial information from non-clinical and clinical study reports and patient level data when there is an overriding public interest reason for doing so, under conditions which serve that interest. The EMA should always consult with the marketing authorisation holder (MAH) prior to disclosure, to allow the MAH to take any necessary steps to protect against unfair competition and/ or prejudice to regulatory data protection, patent or other IP rights.

Note from EMA: stakeholders are invited to present at next CTAG3 meeting concrete (historic?) 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 what 'necessary steps' might be required. (See also comment under section3)

What steps will a requester have to go through before being able to access clinical trial data from the EMA website? After accessing the dedicated domain of the EMA website:

1. Should requesters have to identify themselves?

It is useful to distinguish between access to (1) aggregate data (e.g. lists of studies conducted, ICH compliant clinical study reports including the study protocol, statistical analysis plan and other appendices, but excluding patient level data) and (2) patient-level data (e.g. individual case record forms, SAS files with line listings).

1. Aggregate data: No agreement was reached. The following positions were discussed:

    a. There is no convincing rationale that identification of requesters could or should be required. Such data should be accessible freely (similar to EPAR information today). It is assumed that aggregate data contains no personal data.

    b. In the interest of transparency, requesters should be identified, logged and their identity made public, primarily to ensure patient confidentiality is not compromised and to avoid the misuse of patient level data by third parties with commercial interests that are not related to healthcare research. Requesters of clinical trial data should also have sufficient qualifications and experience for any subsequent analysis of data obtained from clinical trials, as aligned with ICH-E9 and 'statistical principles for clinical trials'. Also, in order for any analysis of data obtained from clinical trials, there should be a legitimate scientific question being proposed in order for the request for data access to be considered. Requesters should not only identify themselves, but they should also provide details of their qualifications and experience which supports they are sufficiently educated and trained to implement

any subsequent analysis of the data being requested. This information should be made transparent by the requester at the time of seeking access to data.

NOTE from EMA: such proposals may not be compatible with the legal framework under which EMA operates as a public body; to be discussed at upcoming CTAG3 meeting

2. Patient-level data: No agreement was reached. The following positions were discussed:

   a. These data should be freely accessible without the need for identification. Arguments in favour of this position include (not in order of importance):

      i. Lowering the hurdle for patients who wish to access data related to their own disease;

      ii. Proper verification of identity of the requester is near-impossible;

      iii. If the data are used for illegal actions such as illegitimate commercial use, there are legal actions which can be taken against the firm/country benefiting from the illegal action. Thus, this point should not be an argument to force requester-identification. Furthermore, if someone wishes the data for illegal action, he will surely and easily use a wrong identification or could only ask others to also request data in order to increase the number of suspects;

      iv. Any patient-level data that EMA makes available will be de-identified/anonymised, therefore the risk of retro-active patient identification is considered acceptably low, and the patient data protection is not an issue (it is argued that there is even no need to distinguish between aggregate data and patient level data). Therefore, there is no need to verify the identity of the requester (*Note: reference is made to CTAG1, which is discussing standards for de-identification/anonymisation to ensure patient data protection*);

      v. There are cases of harassment by pharmaceutical industry when a physician declared an adverse event to an agency (example: Dr Chiche in Marseilles about the Mediator story). If the name of the requesters is given to EMA, how will EMA make sure that the name of the requester will not be known by the Marketing Authorisation Holder? In case of harassment linked to a data request, what would be EMA's responsibility?

      vi. The privacy of study participants is important and their privacy should be warranted. On the other hand, the privacy should also be warranted for study participants, patients or other (EU) citizens who like to access patient-level data for their own private use. Namely, publication of their name on the internet involves the risk of unintended use of the personal data of this person, especially if this information can be detected by search engines such as Google. For example, the information (name + type of medication) may be detected during a background search performed for a job application; the information can be used by insurance companies; or the information can be used for direct marketing for registered or falsified medicines, including spamming. This is an argument to carefully consider whether the benefits of publication of the names of private persons

outweigh the risks of unintended use and breach of privacy of those who access data. Thus, benefits of publication of the names of those who access patient level data may not outweigh the risks, because publication of personal data in combination with (type of) medicines for which data have been accessed creates the possibility for unintended and undesirable use of personal data;

vii. As data would be anonymous there is no sensitive data. Retrospective patient identification cannot be prevented by verifying the identity of the requester, nor can any violator necessarily be identified through such knowledge as there will usually be no conclusive link between the violation and the requester. We should keep in mind article 6.1. b and c. in directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Pursuant to this article collection of data must be adequate, relevant and not excessive in relation to the purposes. Registering the requester is also processing of personal data and should only be done for legitimate reasons and should not be excessive in relation to the purpose.

b. These data should be freely accessible only after verification of the identity of the requester. Arguments in favour of this position include (not in order of importance):

i. Patient-level data is too sensitive to allow anonymous requesters to access because the risk of retrospective patient identification is never zero. The legal liability associated with the release of the patient data from a data privacy perspective needs to be considered. There is reference to the risk of retro-active patient identification being “acceptably low”, yet that still presents a risk to patient identification. Legal accountability needs to be addressed if a patient is in fact identified and this is used improperly against an individual patient;

ii. The level of de-identification required to render patient-level data suitable for open public access is likely to seriously compromise the utility of that data for the purpose of research in the interest of public health. Much of the value of analysis of patient-level data over aggregate data is the ability to link and take account of patient characteristics in analyses. For example, if age and gender were to be removed from the dataset, it would not be possible to investigate possible treatment interactions with these characteristics or with these in combination with other characteristics that remain in the dataset. If dates are removed this reduces scope for scrutiny and (unless replaced with a series of derived times from event to event) precludes time to event analyses. This would mean, for example, that survival analyses in cancer trials would not be possible. This is an important consideration for individual participant data systematic (IPD) reviews and meta-analyses. Re-consider whether tiered access is feasible. Open public access for all documentation including clinical study reports, results, and aggregate data. Access to IPD restricted to being for the purpose of research in the interest of public heath - as demonstrated by provision of a

protocol or research plan, disclosure of investigator name and affiliation and declaration of any potential conflict of interest (preferably at the point of release of data, but delayed if necessary);

iii. Strict assurances about the specific use of personal data are given as part of the consent process to trial entry; they do not include release except under strict rules. Release of individual patient data, even anonymised, contravenes the information provided as part of the consent process, and thereby infringes human rights.

iv. There is a risk of illegitimate commercial use of patient-level data (please refer to **point 3**). To mitigate this risk the identity of the requester must be verified;

v. The identity of the requester should be available and public. It is widely accepted in science that people have to disclose their financial interest. This principle should be applied here as well;

vi. The objective is clearly to restore trust in the system, not to create an all-purpose research tool. Patient data is not to be diverted to research purposes for which it was never intended or to "data mining", be it academic or commercial. Such misuse could otherwise lead to false claims of efficacy and safety of medicines. The EMA has previously stated the objective is to "(...) enable the independent re-analysis of the evidence used by the Agency's committees to determine their benefits and risks and is expected to lead to public-health benefits." The access process should be developed with this public health principle in mind;

c. For access, a hierarchy for different user groups should be foreseen with access to different types of data. For the EMA pharmacovigilance database, such an access policy already exists. (EMA/759287/2009 corr., EudraVigilance access policy for medicines for human use) This paper is adopted after consultation with the Patients' and Consumers' Working Party and consultation with the Health Care Professional Working Group. The paper defines 4 types of stakeholder groups:

- Medicines Regulatory Authorities, the European Commission and the Agency (hereafter referred to as Stakeholder Group I)
- Healthcare Professionals and the General Public (hereafter referred to as Stakeholder Group II)
- Marketing Authorisation Holders and Sponsors of Clinical Trials (hereafter referred to as Stakeholder Group III)
- Research Organisations (hereafter referred to as Stakeholder Group IV)

There is a need to modify the categories according to an optional user identification process, granting access to e.g. patient level after authorisation. This would also allow for the processes discussed under topics 3, 4 and 6, setting reminders or making registered users aware of possible consequences after misuse.

2. Should requesters be required to 'Agree' to respect personal data protection?

It is agreed that this point is only relevant for patient-level data.

It is agreed that any requirement for the requester to actively agree to respect personal data protection would depend on whether the identity of the requester can be/has been verified. (No agreement was reached on that point, see above)

If the identity of the requester has not been verified (two positions):

a) Without requester identification, such `agreement` to respect personal data protection is only for information, but cannot be legally binding. As far as CTAG1 rules for patient data anonymisation are applied and effective, respect of personal data protection mainly forbids linking the data obtained from EMA with other databases/information.

b) Even if the identity of a requester cannot be verified, a disclaimer about the need for personal data protection should be "read and accepted" by the requester.

If the identity of the requester has been verified:

Should it be/have been possible to verify the identity of the requester, and the requester actively agrees to respect personal data protection, any violation of this agreement should be legally enforceable.

Requesters have to be made aware of EU and local data protection regulations. Ticking a box implies a contractual relationship between the requester and the database owner/holder of the data. However, in that case both contractual parties need to be fully identifiable. A contractual but not necessarily public "digital" agreement appears to be preferable compared to a purely anonymous process.

Details of a contractual agreement should clarify that if any individuals are provided access to clinical trial data, then the holders of the data cannot be held accountable in any way for what the requesters subsequently do with the data; any re-analysis of the data is at the responsibility of the requester. If subsequent issues are found with respect to an incorrect re-analysis, misuse of the data for purposes outside of the research proposal originally specified, or any potential fraudulent behaviour, the original owner of the source data cannot be held accountable in any way.

3. Should the requester be required to 'Agree' to refrain from unintended commercial uses of information retrieved?

There is general agreement that EMA's policy on Access to clinical trial data should further the interest of public health, but should not abet usage of data for unintended commercial uses (e.g. obtaining a marketing authorisation in a third, non-EU, jurisdiction). EMA's policy should attempt to mitigate this risk without compromising transparency. The option of requiring anonymous data requesters to tick a 'read and accepted' tick box is considered ineffectual.

No agreement was reached on the following point (two positions):

a) The requester should be required to sign a legally binding agreement affirming that the information and data will only be used for the agreed public health research purpose and not for any commercial use. Requests for patient level data from requesters to the EMA must be handled on a case-by-case basis, and follow consistent criteria to establish if and how the information provided will be used for valid scientific purposes and to benefit patients.

b) It is unclear which situations we are talking about and "unintended commercial uses" may be used as a "killer argument". For example, if industry fears that one cannot exclude that a full CSR may be used for obtaining a marketing authorisation in a non-EU jurisdiction, this may prevent full transparency. Some real-life examples of "unintended commercial uses" should be given during the next CTAG3 session.

4. Should the requester be made aware of quality standards for additional / secondary analyses?

No agreement was reached on this point (two positions):

a) It is emphasised that advising requesters of quality standards for additional secondary analyses should not and cannot impose any obligations on the requester. (*Note: Reference is made to the work of CTAG4).*

The use of such advice is questioned. This may discourage non-professional users from downloading and using such data. There is no benefit from such advice but it may mean a subjective additional hurdle to lay groups/patients.

b) The requester should be advised of quality standards for additional secondary analyses. The same standards must be applied equally to the requester as would be applied to the MAH. It is emphasised that such advice should imply clear obligations on the requester.

5. Should the requester have to declare whether they wish to upload a protocol / analysis plan?

There is agreement that good scientific practise requires those who wish to engage in secondary data analysis to complete and submit a study protocol before accessing the data. Therefore, the opportunity (but not obligation) to upload a protocol on an EMA managed repository is welcomed. There was no consensus as to the time of publication of such uploaded protocols. Options discussed were:

a) Immediately after uploading the protocol

b) After a fixed time span (e.g. 1 month, 1 year?)

c) Around the time of publication of the results of secondary analysis

d) Timing of publication decided by requester

Several comments/views along the following lines were expressed:

A requester should have to submit a protocol or analysis plan before being granted access to the data as this enables full transparency of the purpose and intention for requesting access to the data and this helps to minimise any misuse by third parties. In order to ensure there is a legitimate research question(s) being proposed, pre-specifying the clinical hypotheses to be investigated ensures the scientific credibility of the research to be undertaken.

Provision of a protocol demonstrating good research methods, fair use of data and the purpose to which it will be put seems an entirely reasonable exchange for access to data. There seems to be a danger of introducing double standards with requirement for access to clinical trial protocols and clinical trial data, but not to protocols for subsequent use. For IPD, make provision of a protocol

(with delayed public access if necessary) a prerequisite for access to or release of data. A link to a formally published protocol would be acceptable.

Therefore the protocol must be reviewed before the patient level data is provided.

NOTE from EMA: such proposals may not be compatible with the legal framework under which EMA operates as a public body; to be discussed at upcoming CTAG3 meeting

6. Should requesters be allowed to share accessed data?

It was agreed that this is a moot point in case identification of the requester is not verifiable.

No agreement was reached on the following point of sharing data (two positions):

a) Should it be/have been possible to verify the identity of the requester, EMA may consider restricting data sharing. However, in such case any third party would have to be given access to the same data as the first requester directly from the EMA.

b) Requesters should not be allowed to share accessed data because that way the validity of the dataset cannot be controlled. Requesters should need to explicitly confirm that they will not forward the downloaded original dataset to third parties. It is acknowledged that others must be able to repeat research findings; that is a basic principle of research. However, such groups would then have to identify themselves separately before accessing the same data.

7. How should EMA's policy be rolled out (timelines)?

There was brief discussion as to whether the policy should be rolled out in a staggered way, starting with high-level (aggregated) data, followed by more granular (patient-level) data sets. No conclusion was reached (three positions).

a) If the name of the requester is not needed for aggregated data, then most points do not need further discussion. A staggered roll-out should not delay implementation of the rules to make data publicly available.

   There is no obvious benefit and no reason to use a staggered way other than limited capacity. Hence, there is no reason to postpone access to patient-level data

b) A staggered roll-out would be preferable as there are already many challenges to opening up access to aggregated data which need to be solved. Aligning with the roll-out of the EudraCT version 9 and access to results for many clinical trials could be an important step forward.

c) A staggered approach would be pragmatic and could achieve much almost immediately. There are many issues around the release of IPD, particularly around open public access versus some model of conditional access. If this could be set aside for now with focus on release of aggregate data and results of all statistical analyses as set out in the trial protocol, rapid progress could be made. Access to IPD could follow after sufficient time for discussion and enquiry. For example, potential impact of public release of IPD on participant consent needs to be investigated. Therefore, separate the issues of (1) release and access to trial information, results and aggregate data from (2) release and access to IPD, and move ahead immediately with 1. Do not delay implementation of 1 while 2 is

addressed (it is much more complex and requires careful consideration). Extend the time period to allow proper consideration and investigation of issues pertaining to 2.

8. Should requesters be encouraged to provide feedback?

There is agreement that users of data should be encouraged to link back the results of their analyses to the accessed data in order to ensure two-way transparency.

While a link back of results of individual analyses is desirable, it should be located on a separate database in order to not increase subjective hurdles to lay people. This database should/could be linked to the database of analysis plans/protocols.

It may also be useful to add a user/log-in concept to the repository to allow requesters to build project websites. These project websites would give requesters the opportunity to publish timelines, the protocol and the results of their project (or links to such documents).

Several comments/views along the following lines were expressed:

- Just encouraging requesters to link their analyses back to the data accessed is not sufficient. Further discussion is needed on how any resulting publications arising from secondary analyses are linked back to data access requests. Principles should be included on minimal expectations of requesters and what should be fed back having been granted access to data. For example, should the requester have to summarise their key findings of their analyses as a minimum?
- On the assumption that access to anonymised patient level data is granted for a defined research project, access to a secure area should be granted for a defined duration (the duration necessary to complete the project). An open-ended access (beyond the research project) would undermine the benefits of identification and declaration of research purposes. NOTE from EMA: such proposals may not be compatible with the legal framework under which EMA operates as a public body; to be discussed at upcoming CTAG3 meeting