Document History

Filename	Description	Dated	Circulated on
CT data group 1 - Personal Data Protection - Discussion Document_1	Initial discussion document with questions	10 January 2013	10 January 2013
CT data group 1 - Personal Data Protection - Revised After TC1	Revised document following first teleconference 5 February 2013	10 January 2013	20 February 2013
CT data group 1 - Personal Data Protection - Revised After TC1Comments	Further revisions following written comments (ELPA; A. Johnson; EFSPI; BIA; EFPIA.)	6 March 2013	7 March 2013

Draft advice to the European Medicines Agency from the clinical trial advisory group on Protecting Patient Confidentiality

07 March 2013

Introductory note

This is a draft proposal intended to stimulate and structure the upcoming discussion among members of the advisory group on protecting patient confidentiality, which is set up to inform the upcoming EMA policy on clinical trial data transparency. The draft document is not intended to pre-empt the content of the policy the agency will ultimately adopt. The draft proposal has been amended to reflect the comments and discussion (summarised in comment boxes) received during the meetings of the clinical trial advisory group on protecting patient confidentiality.

Problem statement

How can EMA ensure through its policy that patient and other personal information will be adequately protected i.e., that patients cannot be retroactively identified when clinical trial data are released, and that applicable legislation, standards, and rules regarding personal data protection will be respected?

Discussion proposal

1. Scope and definitions

1.1. This advice refers to any information containing clinical data (e.g., raw data, clinical study reports) that are submitted to the Agency as part of a marketing authorisation application, or subsequent submission (e.g., in the context of clinical variations of the marketing authorisation, submission of results of post-authorisation safety studies).

Comments:

Clarify that the scope refers to initial approval and subsequent changes.

1.2. Personal data: Any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. In this document, a distinction is made between persons included in clinical trials (e.g., patients or healthy volunteers and their legal representatives, hereinafter referred to as “subjects”), and any other person

(investigators, study site personnel, sponsor representatives, contracted workers, etc., hereinafter referred as “clinical trial personnel”).

Comments:

The basis for the definition of personal data should be the definition provided in Art. 2 (a) of the EU Data Protection Directive (Directive 95/46/EC), namely that 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

1.3. De-identified data: Data that have been made anonymous in such a way that the data subject is no longer identifiable (directly or indirectly). A similar term is “anonymised data”.

1.4. Key-coded data: These data refer to information that relates to individuals that are assigned a code, while the key making the correspondence between the code and the common identifiers of the individuals (like name, date of birth, address) is kept separately. In clinical trials, the key is typically held by the investigators. Information to the pharmaceutical company or other parties involved is provided only in this coded form.

Comments:

This refers to the activity of rendering data anonymous in such a way that the data subject is no longer identifiable. The preferred term “de-identified” should be used consistently throughout the document. The term “data redaction” should not be used as a synonym of de-identification.

Key-coded data refers to information that relates to individuals that are assigned a code, while the key making the correspondence between the code and the common identifiers of the individuals (like name, date of birth, address) is kept separately. In clinical trials, the key is typically held by the investigator, and the data collected in the study for the analysis and reporting is key-coded. Information to the pharmaceutical company or other parties involved is provided only in this coded form.

Key-coded data constitutes information relating to identifiable natural persons for all parties that might be involved in the possible identification and should be subject to the rules of data protection legislation (see Opinion 4/2007 on the concept of personal data of the Article 29 Data Protection Working Party).

The original key-coded data were never conceived to be published. If such personal data were to be shared by the Agency, a special set of rules would be required, similar to those applicable to processing of personal data for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy.

Data management/data access control should be defined. This can be obtained through the establishment of a governance function/structure that will assume gate-keeper responsibilities controlling the good implementation of rules of engagement and processes necessary for MA data disclosure. The risk of re-identification of submitted personal clinical data being also linked to the actual use by third parties and this use can be monitored/restricted via adapted rules of engagement.

With free access patients and their relatives will be able to view all their clinical data in detail outside of clinical consultation; that may not be wise as such information may be mis-interpreted. These disclosures may affect the patient-doctor relationship.

Legal Aspects: It should be verified if release of study participant data is possible under the EU Data Protection Directive, given that study participants will not have contemplated this, or consented to it.

2. Clinical Trial Personnel's Data

2.1. Option 1: Personal data of clinical trial personnel (name, CV, affiliation, etc.) are considered as professional information that is essential to be made public. Clinical trial personnel have legally defined responsibilities and roles with respect to aspects of the marketing authorisation dossier and the clinical trials that are part of the dossier. Assessment of the qualifications of the researchers and other clinical trial personnel is an important public interest in the area of public health protection and scientific research. Companies are advised that non-essential information (e.g. personal address, personal phone number) should not be included in the dossier.
Option 2: Personal data relating to the principal investigator and the experts who sign the clinical study report are considered as professional information that is essential to be made public. This is justified by grounds of important public interest in the area of public health protection and scientific research. For any other clinical trial personnel there is no presumption of important public interest why such data should be made public.
Option 3: There is no presumption of important public interest why any personal data should be made public.

2.2. There should be sufficient protection for the privacy of pharmaceutical company employees and researchers that perform non-clinical research. Similar considerations should apply to personnel participating in research that could be considered to be sensitive or controversial. In such cases, companies should be allowed to justify de-identification of data related to clinical trial personnel.

Comments:

One view was to agree with the approach to consider personal data related to clinical trial personnel as essential to be made public. In general, for clinical trials there is no great concern for revealing the names of investigators and study/company personnel, as shown by the ample information generally in the public domain about the investigators involved (e.g., as listed as authors or investigators in publications of medical journals, including their affiliations, contact details and emails). In multinational studies it is also important to know who the investigator in charge in that country is.

A divergent view was that except for a few people (the principal investigator, the persons responsible for the study or its interpretation, the experts who sign the report), there is no public health interest for disclosing such information about any other clinical trial personnel or persons whose names may appear in the dossier. Data related to such persons should be considered as personal data, not to be released without adequate de-identification. There is also a concern that publishing all investigators' names may add to the risk of identifying the clinical trial subjects.

In line with GCP and ICH E9 the company needs to ensure that appropriately experienced and qualified personnel, including trial statistician, is available to design, conduct, analyse and report the trial and their results. EMA (or any other regulatory authority) is able to check on this through Inspections. In any case, however, the same rules should apply to any requester of the data for the purpose of additional analyses as to the originating company that performed the intial analyses. In order to ensure good scientific practice and in the interest of public health, anyone wishing to analyse aggregate data should be sufficiently qualified and trained otherwise the requester is not sufficiently able to implement legitimate scientific research. Given statisticians who are involved in the design and analysis of clinical trials must

be appropriately qualified and trained as per ICH-E9, surely these minimum standards should be expected of any requester wanting to access clinical trial data.

There should be sufficient protection for the privacy of pharmaceutical company employees that perform non-clinical research. Similar considerations would apply to investigators and researchers participating in research that could be considered to be controversial, e.g., stem cell research. In such cases, companies should be allowed to justify de-identification of data related to investigators.

It would be useful to describe in more detail what data would normally be included here.

3. Subjects' Data

3.1. Currently, subjects' clinical data are submitted as key-coded data (e.g., using a subject identification code instead of the subject's name). Key-coded data constitute information that might be involved in possible identification and should be subject to the rules of data protection legislation. Key-coding is generally insufficient for de-identifying data.

3.2. Key-coded data that are not sufficiently de-identified should only be used for public health-related purposes. A special set of rules would be required for providing access to these data. Such rules should be similar to those applicable to processing of personal data for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy (see Opinion 4/2007 on the concept of personal data of the Article 29 Data Protection Working Party).

Comments:

Key-coded data is the standard practice. Key-coding is generally insufficient for de-identifying data (see also 1.2). Key-coded data should only be used for specific needs, e.g., for certain public health-related purposes by health care professionals or other persons subject to a legal obligation of professional secrecy.

There may be situations (e.g., unusual reaction, adverse effects), when individual data may be important. A balance would have to be struck between personal and public health interest. There need to be ways to allow analysing such data.

3.3. Apart from direct identification, there is a risk that clinical trial data may allow identifying the subjects indirectly, through a combination of potential indirect identifiers. For instance, a person may be identified indirectly by a telephone number, a car registration number, a social security number, a passport number or by a combination of significant criteria which allows him to be recognized by narrowing down the group to which he belongs (age, occupation, place of residence, etc.).

Comments:

Clearly, all requirements of EU data protection legislation and any applicable national laws need to be complied with.

Clarify what is meant by "combination of potential indirect identifiers".

Releasing data, even de-identified, may give rise to severe reactions, e.g., in patients with psychosis or elderly in patients with dementia, or their carers.

Real risk of discrimination; rare diseases

3.4. For all the clinical trial data to be submitted to the Agency (e.g., study report, data set), including any subsequent revisions, the applicant company shall assess the risk of compromising subjects' identity in case of wide publication of those data. In most cases, aggregate statistics (frequencies, sums, etc.) might be considered as sufficiently de-identified so as not to constitute personal data.
Assessment of the risk should take into particular consideration data that could be considered to be sensitive or controversial and that might lead to discrimination if the subject can be identified, as well as situations with an intrinsic higher risk of identification such as very rare diseases.
If for any data the risk of compromising subjects' identity in case of wide publication of those data is considered to be absent or sufficiently low, the applicant company shall clearly label the data as "SUITABLE FOR PROACTIVE PUBLICATION".

Comments:

Need to consider all documents not just individually.

Need to clarify what is meant by publication: controlled or wide access?

If wide access is to be given, industry considers risk to be context dependent. Context may change over time and one cannot predict future. Gate-keeping principle and case by case approach should be applied.

Set the default to have anonymised data publicly available – applicant to state why not possible. If impossible use a gate-keeping approach. Require on application that data set has been anonymised and reviewed by ethics committee. Show the process they followed so that no unacceptable residual risk.

There is a risk of abuse under false pretext of protecting patient confidentiality. Need to ensure data are those needed to enable further research. Develop guidance. EMA should make the risk assessment.

Ask the patient if they agree their identity to be disclosed for specific purposes, e.g., research for confirmations, for further investigation. Eventually should go into informed consent but not unlimited public disclosure but sufficient for research. Only reputable medical investigators should be allowed to conduct the research.

Defining upfront what is "suitable for publication" in the case of very detailed documents and data sets, remains purpose and context related, and necessitates a reliable process to be in place. The MAH should contribute in the preparation of motivated research access and in the monitoring of agreed, scientifically planned and performed secondary analyses, in the interest of the public health.

3.5. Option 1: If for any data the risk cannot be considered to be absent or sufficiently low, the applicant company shall submit two sets of data, the original data clearly labelled as "NOT FOR PROACTIVE PUBLICATION", and the de-identified data clearly labelled as "SUITABLE FOR PROACTIVE PUBLICATION".
Option 2: If for any data the risk cannot be considered to be absent or sufficiently low, the data shall not be widely released. Such data may only be made available in well-justified cases, based on best practice rules to ensure patient confidentiality (to be developed), restricting the purpose of the use of the data towards public health benefits, and preventing the risk of misuse of the data compared to what has been agreed in the informed consent.

Comments:

The proposal (Option 1) is quite complex even from a process point of view. A second set cannot be provided by default but only when justified.

Applicants should not be required to provide additional documents, when necessary, beyond the internationally agree Common Technical Document format. This will de facto void the huge

benefits achieved through ICH regarding harmonized application dossiers and clinical study reports, which have largely contributed to increasingly simultaneous submissions and subsequently accelerated patient access to innovative medicines.

3.6. Option 1: Applicant companies may use different transformation methods to de-identify the data. Generally, using such methods, it is possible to adequately de-identify data in such a way that, taking into account all the means likely reasonably to be used to identify subjects, the risk of identifying a subject does not exist or is negligible; such de-identified data are no longer considered as “personal data”.
A minimum standard for de-identifying data is described in Hrynaszkiewicz et al. (1) In some situations, this minimum standard should be supplemented by additional de-identification methods (e.g., statistical).
The methods of de-identification should also be such that it is to be expected that adherence will preclude patient de-identification even when applying linkages with other data carriers (e.g., social media).
The application of transformation methods to de-identify data may reduce the possibility to conduct certain types of analysis or to replicate exactly certain analyses. This aspect should be considered and adequately communicated when interpreting or publishing results from analyses based on de-identified data compared to those based on key-coded data. If access to the untransformed data is required, this should follow the rules as for key-coded data (see section 3.2).

Option 2: Available methods for de-identifying personal data cannot achieve complete de-identification while preserving sufficient analytical utility of the data. Thus, clinical trial data should not be published unless this is done under strict conditions of access and confidentiality, for public-health purposes only (see also 3.2). Best practice rules should be developed to ensure patient confidentiality. Risk of re-identification should be assessed on a case-by-case basis. The purpose of the use of the data should be exclusively for the benefit of public health and should be in agreement with the informed consent.

Comments:

Some personal clinical data that are part of a submission to the EMA, like narratives or lined data in tables should be carefully redacted in order to avoid disclosing details e.g. birth date, height, gender, rare disease, status or name of the hospital all could facilitate re-identification. This also applies to information such as CT scans, MRT and other imaging, interviews and genetic data. Patient level data in line listings and datasets should not be publically released. Identifiable data in the main body of study reports can be relatively easily redacted. This is not the same as anonymisation of datasets. Access to anonymised trial data should be provided in a secure environment with controls in place to prevent the data and documents from being downloaded or distributed beyond the scope of the approved use of the data. The requestor should be required to sign a legally binding agreement affirming that that they will not seek to re-identify individuals.

A major reason of concern is the alignment of secondary use of Clinical Trial data and the initial Informed Consent. Patients/healthy volunteers participating to a clinical trial gave their informed consent in the frame of the planned use of their clinical data, as described in the information received before to accept participating. Overall secondary use and disclosure of data should be aligned with the original informed consent. Most of the time secondary use for novel/secondary research was not within the scope of the original informed consent, neither the intention to have patient level data published in the public domain, with risk of re-identification. Ethical review boards were not informed of this step either. These provisions (with respect to Informed Consent and Ethical Board review) could change prospectively, however is not the case for the great majority of current submitted clinical data in MAs. It is not pragmatic nor feasible to envisage amendment of past ICFs nor the ECs in each relevant country.

In some situations the minimum standard would provide sufficient de-identification of personal data. In other situations, this minimum standard would have to be supplemented by additional methods (e.g., statistical). The current standards are in the format of a non-technical report that provides general rules. More sophisticated techniques using computer software to assess the risk have been proposed. Common electronic format could present challenges. The merits of different standards could be evaluated with this respect. Alternative methods of assessing adequacy of standards can be applied.

Generally, using such methods, it is possible to adequately de-identify data in such a way that taking into account all the means likely reasonably to be used to identify subjects do not exist or are negligible, and the information would not be considered as “personal data”. Even using additional methods, generally, sufficient analytical utility of the data can be preserved. It is understood that in the case of very small data sets for very rare conditions, the transformation methods used to de-identify personal data may be such that for many types of analyses, the analytical utility would be reduced.

It is difficult to agree on a single standard, the risk can change based on the dataset or type of research. Standard practice is difficult to recommend, there is a need for a case-by-case approach. Best practice rules should be developed to ensure patient confidentiality, to restrict the purpose of the use of the data towards public health benefits and to prevent the risk of misuse following uses not aligned with the initial informed consent. The secondary use of the data has to be in line with the informed consent.

Controlled access to data whereby recipients must agree not to attempt to re-identify data subjects, to protect the confidentiality of the data, and to use the data only for certain specified purposes, is far more privacy-protective than public release.

A governance function/structure should be established that will assume gate-keeper responsibilities controlling the good implementation of rules of engagement and processes necessary for MA data disclosure. The risk of re-identification of submitted personal clinical data being also linked to the actual use by third parties and this use can be monitored/restricted via adapted rules of engagement.

In general, the application of transformation methods will reduce the analytical utility of the data due to the loss of information. In addition, exact replication of analyses and results may not be possible using de-identified data. The controversy arising from disputed results may cause distress to patients wondering exactly what sort of research they have engaged in. This likelihood has to be borne in mind when interpreting the results of analyses done based on de-identified data. Complete de-identification is incompatible with exact reproducibility of all analyses. It needs to be clarified whose responsibility it is to explain divergent results due to data transformations. Any journal confronted with a re-analysis of data should solicit comments from the originating company in the interest of transparency and good research abiding to hearing both sides.

Available methods for de-identifying personal data cannot achieve complete de-identification while preserving sufficient analytical utility of the data.

Aggregate statistics (frequencies, sums, etc.) might be sufficient for many analyses purposes and provide sufficient reassurance about personal data protection.

The entire context needs to be described to inform any statistics.

There are practical issues with informed consent if some subject were allowed to agree or disagree within one study. If this was an entry criterion it may be more workable. But there are concerns about additional burden on sponsors or incomplete data sets. The solution needs to be practical.

If patients consent, no transformation is needed. In practice this can only be applied prospectively.

Regardless of the process followed, there should be clarity of where the responsibility lies in case of identification of subjects.

EMA should set the rules for all parties involved for patient de-identification because these rules will determine the analytical utility of the data as a result and are a direct consequence of it, which will enable a far better communication to public health and to the public in general.

EMA is to set these standards. It is unprecedented that individual patient data will be publicly available and sufficient safeguards should be put in place to prevent misuse from happening, including patient identification, while trying to reach transparency about the data underlying health claims.

3.7. De-identification methods shall be individually tailored to the specific dataset and situation to ensure that a maximum of information is available while at the same time ensuring sufficient personal data protection. Methods and extent of de-identification should be adapted to sensitive or controversial situations that might lead to discrimination if the subject can be identified, as well as situations with an intrinsic higher risk of identification such as very rare diseases.

Comments:

How methods should be “individually tailored” is to be discussed on case-by-case basis; according to the specific context of the secondary research it may be appropriate to keep some indirect identifiers and not others in order to adapt to the disclosure context while preserving scientific validity of the sample. However these data fall under the scope of EU Data Privacy Directive and may raise issues of liabilities in case of subsequent misuse.

Methods and extent of de-identification should be adapted to sensitive situations.

3.8. Option 1. Applicant companies shall describe in general terms and justify for each document the de-identification methods used.
Option 2. Applicant companies shall describe in general terms the de-identification methods used.

Comments:

Suggest keeping flexibility and avoiding cumbersome processes.

Possibly, standardised formats should be developed to facilitate this, detailing the procedures, precautions and safeguards that have been followed.

Providing a justification on a document-by-document basis seems to be overly burdensome and not value-adding. A case-by-case gatekeeper approach is recommended.

3.9. Option 1. The Agency will not systematically verify that the data submitted as de-identified data contain no personal data – this is considered the responsibility of the applicant company.
Option 2. The Agency should systematically verify that the data submitted as de-identified data follow Agency standards and contain no personal data.

Comments:

Sufficient safeguards should be put in place to prevent misuse from happening, including patient identification, while trying to reach transparency about the data underlying health claims.

EMA’ s mission and legal role necessitates its active involvement in the assessment of data held by EMA which is to be made available and necessitates an effective oversight of the process.

The MAH should always be consulted before release of information or data with the opportunity to comment and seek redactions.

This would only work if not abused (excessive anonymisation of data). The Agency should refuse applications where invalid methods have been used or if an abuse may be identifiable.

If the de-identification methods are deemed insufficient or excessive, the Agency shall ask the applicant company to further justify and if necessary modify the de-identification method.

3.10. The Agency shall produce further guidance on the standards and methods for de-identifying data. Upon request, the Agency shall provide advice to applicant companies, (where necessary involving relevant patient groups and members of the public), on the adequacy of the methods for de-identifying data.

4. References

(1) Hrynaszkiewicz, I., M. L. Norton, et al. (2010). "Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers." BMJ **340**: c181.

Additional points for discussion:

It may be worthwhile discussing this issue with the European Commission and the Article 29 Data Protection Working Party.

A point is raised about commercial (mis)uses of data.

Face-to-face meeting recommended for the end of the work of this advisory group.

Revised proposal: 22 February

Second teleconference: around 12 March

Final proposal: End of March

Last teleconference: 19 of April