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. J.; EFSPI; BIA; EFPIA)	6 March 2013	7 March 2013
CT data group 1 - Personal Data Protection - Revised After TC2	Revised document following second teleconference 15 March 2013 (Incl. comments from G.U. from Pfizer)	18 March 2013	25 March 2013
CT data group 1 - Personal Data Protection - Revised After TC2Comments	Further revisions following written comments (BIA; EFPIA; Parkinson's UK; I.H.; EFSPI; A.J.; G.U.; EURORDIS; M.B & D.P.)	17 April 2013	17 April 2013
CT data group 1 - Personal Data Protection – Draft Final Advice	Revised document following third teleconference 18 April 2013	18 April 2013	19 April 2013

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

1~~7~~8 April 2013

1. Summary

Different types of data pose different levels of risk of identifying patients. In most cases, the risk is considered sufficiently low in case of reports containing only aggregated data, such as the main body of the clinical study reports and appendixes (excluding line-listings). After de-identification of any indirect identifiers (e.g., case narratives, outliers, tables with sparse numbers), such reports could be considered for ~~proactive~~ publication and given unrestricted access. Applicant companies may use different transformation methods to de-identify the data. A recommended minimum standard for de-identifying data is described in Hrynaszkiewicz *et al.* (~~20~~101).

In case of the raw data and line listings, the risk is considered much higher due to the combined presence of many indirect identifiers. Further de-identifying key-coded raw data and line listings is a resource-intensive task and may (debatably, in some or most cases) compromise the analytical validity of the data, so that unrestricted publication of fully de-identified raw data and line listings may not (debatably, in some or most cases)~~(always)~~ be useful. ~~If unrestricted publication of de-identified raw data, or part of the data, is not considered feasible and useful, a~~Access to the ~~original key-coded~~ raw data and line listings (~~if necessary~~debatably, original key-coded or de-identified) should be allowed under similar rules to those applicable to processing of personal data by health care professionals subject to the obligation of professional secrecy. There should also be rules to ensure that additional use of raw data is within the scope of the informed consent/assent signed by trial subjects or on their behalf.

2. 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?

3. Scope and definitions

3.1. 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). When discussing the various options, a distinction is made between documents containing mainly aggregated data (i.e., the main body of the clinical study report and appendixes, excluding line-listings), and raw data and line listings, containing key-coded, patient-level data.

3.2. 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[1] (2).
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 mentioned in the submission (investigators, study site personnel, sponsor representatives, contracted workers, etc., hereinafter referred as “clinical trial personnel”).

3.3. De-identified data: Data that have been made anonymous in such a way that the data subject is no longer identifiable (directly or indirectly).

3.4. 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.

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[2] (3).

4. Clinical Trial Personnel’s Data

4.1. 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.

[1] Art. 2 (a) of the Directive 95/46/EC.
[2] Opinion 4/2007 on the concept of personal data of the Article 29 Data Protection Working Party.

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 of clinical trial personnel should be made public.

4.2. 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 for Option 1

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.

It is important that detailed information about the ethics committees is made available.

Comments for Options 2-3

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.

The assumption should be that such information should not be disclosed unless the relevant individual has consented, or unless the information becomes public as a result of publication of study data. Disclosing such information is contrary to the position taken by companies and the EMA when releasing information in other contexts. Releasing the names of such company employees can expose them to personal risks, particularly where the research involves technologies or techniques that some may find more controversial, such as stem cell or gene therapies.

5. Subjects' Data

5.1. Currently, subjects' clinical data are submitted as de-identified data (e.g., aggregated data in the form of tables within a clinical study report) or as key-coded data (e.g., using ~~a sub~~the ~~ject~~subject identification code instead of the subject's name as part of line listings).

5.2. 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 initials, date of birth, 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.).

5.3. For all the clinical trial data to be submitted or requested by 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. 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 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". If the risk is not considered sufficiently low, for study reports and, where applicable, for raw data (depending on the option chosen under 5.6), the company shall submit two sets of documents, the original documents clearly labelled as "NOT FOR PUBLICATION", and the documents containing de-identified data clearly labelled as "SUITABLE FOR PUBLICATION".

~~In most cases, aggregate statistics (frequencies, sums, etc., as found in the main body of the clinical study reports and appendixes excluding line listings) might be considered as sufficiently de-identified so as not to constitute personal data.~~

5.4. Metadata about the study reports and data sets should be provided so those wishing to seek data for reuse can easily see the nature of what is available. For example, datasets could be listed with: Name of the trial; intervention (drug/device) being studied; name(s) of investigator(s); number of subjects; file type(s) and format(s); sponsor of study; date(s) trial conducted; date of submission of data to the EMA; trial registration number (ISRCTN/NCT number).

~~5.3.~~ Study reports:

5.5. In most cases, aggregate statistics (frequencies, sums, etc., as found in the main body of the clinical study reports and appendixes excluding line listings) might be considered as sufficiently de-identified so as not to constitute personal data.

~~In case of the main body of the clinical study reports and appendixes (excluding line listings) if the risk cannot be considered to be absent or sufficiently low (e.g., inclusion of narratives including non-aggregated indirect identifiers, description of outliers, tables with sparse data), the applicant company shall submit two sets of documents, the original documents clearly labelled as "NOT FOR PROACTIVE PUBLICATION", and the documents containing de-identified data clearly labelled as "SUITABLE FOR PROACTIVE PUBLICATION".~~ Companies should be encouraged not to include promotional material in the study reports. Adequate disclaimers and visual prompts (for example, watermarking the pages of the study report) should be considered to avoid that any information made widely available could be misunderstood as representing Agency views.

~~5.4.~~

5.6. Raw data and line listings:

Option 1~~:~~

In case of raw data and line-listings, ~~making~~ adequately de-identified data ~~available, when it is~~

~~possible to protect patient privacy,~~ can be valuable and de-identifying the data does not necessarily compromise the analytical utility of the data (~~2~~4).

Adequately de-identified data should be made available for wide access. The data to be made available may include all the data sets or a relevant subset ~~of the data in case of~~ (e.g., the main analysis set, containing a limited number of indirect identifiers so that the risk of compromising subjects' identity in case of wide publication of those data is considered to be absent or sufficiently low whilst preserving the ability to replicate the main analysis). ~~Where this is possible, de-identified data clearly labelled as "SUITABLE FOR PROACTIVE PUBLICATION".~~

Option 2~~:~~

In case of raw data and line-listings, with few exceptions, available methods for de-identifying personal data cannot achieve sufficient de-identification while preserving sufficient analytical utility of the data, particularly for safety data and case narratives for adverse events.

Publication of such data would either fail to protect patient confidentiality or result in a burdensome yet futile exercise of no analytical use. ~~Option 1: Where de-identification of raw data and line-listings is not considered possible without compromising the analytic utility of the data, a~~Access to the original key-coded data should only be allowed under strict rules to ensure confidentiality and alignment of the purpose of access to the subjects' informed consent/assent. 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[3] (3).

According to one view, even when accessed under such strict rules in Option 2, the data may have to be also fully de-identified.

~~5.5. Option 2: Same as 5.5 except that even when accessed under strict rules such data shall be also fully de-identified.~~

~~5.6. Where datasets cannot be disclosed due to concerns about the privacy of subjects, metadata about the datasets should be provided so those wishing to seek data for reuse can easily see the nature of what is available. For example, datasets which cannot be published could be listed with: Name of the trial; intervention (drug/device) being studied; name(s) of investigator(s); number of subjects; file type(s) and format(s); sponsor of study; date(s) trial conducted; date of submission of data to the EMA; trial registration number (ISRCTN/NCT number).~~

5.7. Guidance should be provided in the format of standard templates for subjects' informed consent/assent to better inform subjects of possible further uses of the data in the interest of public health and under the chosen rules of engagement.

6. De-identification of personal data

6.1. 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

~~[3] Opinion 4/2007 on the concept of personal data of the Article 29 Data Protection Working Party.~~

identifying a subject does not exist or is negligible; such de-identified data are no longer considered as “personal data”.

6.2. A recommended 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 adherence will preclude subject de-identification even when applying linkages with other data carriers (e.g., social media).

6.3. 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 rare diseases.

6.4. Applicant companies shall describe in general terms and justify, if appropriate, for each document the de-identification methods used. 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.

6.5. The Agency should ~~consider~~ establish whether it wishes to systematically verify that the data submitted as de-identified data contain no personal data, or if this is considered the responsibility of the applicant company.

6.6. 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.

7. 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.

(2) Art. 2 (a) of the Directive 95/46/EC.

(3) Opinion 4/2007 on the concept of personal data of the Article 29 Data Protection Working Party.

(~~2~~4) Sandercock, P. A., M. Niewada, et al. (2011). "The International Stroke Trial database." Trials **12**: 101.

8. Summary of comments

Scope	Comment
Personal data protection	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.
Personal data protection	The proposal to de-identify data for raw data/line listings 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.
Personal data protection	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.
Rules of engagement	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.
Personal data protection	In some situations the minimum standard would provide sufficient de-identification of personal data. In other situations, this minimum standard

Scope	Comment
	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.
Personal data protection	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.
Personal data protection	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.
Rules of engagement	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.
Rules of engagement	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.
Rules of engagement	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.
Personal data protection	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

Scope	Comment
	research abiding to hearing both sides.
Personal data protection	Available methods for de-identifying personal data cannot achieve complete de-identification while preserving sufficient analytical utility of the data.
Personal data protection	Aggregate statistics (frequencies, sums, etc.) might be sufficient for many analyses purposes and provide sufficient reassurance about personal data protection.
Personal data protection	The entire context needs to be described to inform any statistics.
Rules of engagement	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.
Rules of engagement	If patients consent, no transformation is needed. In practice this can only be applied prospectively.
Legal	Regardless of the process followed, there should be clarity of where the responsibility lies in case of identification of subjects.
Personal data protection	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.
Personal data protection	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.
Personal data protection	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	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.
Personal data protection	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.
Rules of engagement	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

Scope	Comment
	clinical data being also linked to the actual use by third parties and this use can be monitored/restricted via adapted rules of engagement.
Personal data protection	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.
Rules of engagement	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 initial 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.
Personal data protection	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. Requiring applicant companies to submit two versions of every CSR is unnecessary and burdensome.
Rules of engagement	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.
Rules of engagement	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.
Rules of engagement	The MAH should always be consulted before release of information or data with the opportunity to comment and seek redactions.
Personal data protection	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.
Personal data protection	By signing a Personal Data Consent Form the investigator documents his/her willingness to share personal data with the sponsor and to allow the sponsor to use these data and to forward these data to involved parties - and to the public audience. A template of such a Personal Data Consent Form should be provided as an attachment of this guideline.
Legal	The Data Protection Directive 95/46/EC defines the "data controller" as the "natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data." Since the EMA has decided to operate a policy of granting broad access to study data it holds, it will

Scope	Comment
	be the data controller in respect of any information that it holds and that it intends to release for the purposes of its policy. It cannot delegate its legal responsibilities to a third party. The EMA will be the data controller in respect of any information that it holds and that it intends to release for the purposes of its policy on access to clinical trial data. While it may ask companies submitting data for an indication of whether the data may be suitable for disclosure by the Agency when implementing its policy on access to clinical trial data, this cannot negate the Agency's legal obligation to ensure that it complies fully with its obligations as data controller. Applicant companies may use different transformation methods to assist the EMA in ensuring that data the EMA proposes to pubish are de-identified. As data controller, the EMA remains responsible for ensuring that the privacy of subjects is adequately safeguarded in accordance with applicable data protection laws.
General	The implications of the release of patient level data on innovation and on individual patient protection and public health through re-evaluation of data by third parties needs careful consideration and discussion among regulators, patients, academia and industry to identify the best solution to balancing the desire for transparency with the need to foster innovation. Public health benefits include support for continued innovation and new drug development.
Legal	Redaction of commercially confidential information will also be needed and should be mentioned.
General	After de-identification of all indirect identifiers the risk for re-identification may be considered low. However this specific aspect does not solve other relevant issues e.g. the unsuitable format of a Clinical Study Report for "~~proactive~~ publication". We consider the CSR as a comprehensive technical document, aimed to provide very detailed scientific information to regulatory bodies and not tailored to the general public.
General	Unrestricted pro-active publication of raw data is not useful. Moreover we see no reason why access to key-coded data should be part of this proposal. There is no need to take this risk as access to de-identified data (compatible with the research request) would be best option.
Rules of engagement	Publication of the results of the additional analyses would seem very reasonable in terms of transparency.
General	It is necessary to understand the intended purpose for sharing the data and its intended benefit to science, public health and medicine. This then allows the risk to patient confidentiality to be assessed in the context of the informed consent and the necessary mitigation actions, i.e. level of de-identification, can be confirmed. This is part of the gatekeeper model.
~~General~~	~~Why does the EMA continually say "proactive publication"? What do they mean by "proactive"? Publication is publication – do we need the adjective?~~
Analysis	Making important datasets available for further scientific research should be implemented in a way which supports good research, avoids

Scope	Comment
	misuse of such data and fully protects patient confidentiality. Open access to data should only be made to aggregate level (summary) data and access to patient level data is only made available via a secure system controlled by the regulators or by the owner of the data in order to ensure patient confidentiality. The process of re-analysing data and drawing scientific valid conclusions from it is very complex, and in line with ICH E9, qualified and experienced individuals should be granted access to data to ensure quality research. It requires those wanting to re-analyse the individual patient data to submit upfront a research protocol or statistical analysis plan to verify the scientific integrity of the proposed analyses. The protocol, SAP and the results of the secondary analysis should be made public.
Personal data protection	Risk of Reidentification: The EMA should further consider the potential for re-identification of data that would appear to be "de-identified" under current standards. Because clinical trials often have very specific participation criteria, knowledge of such criteria can be used to re-identify participants. Orphan drugs and pediatric trials pose special concerns, because the population eligible to participate in such trials is extremely limited. Multi-year trials also increase the likelihood of re-identification, due to the increased specificity contained in multiple data points gathered at specific times over longitudinal periods. Furthermore, as technology advances it becomes increasingly likely that re-identification of subjects will be possible using genetic data that would not have allowed for re-identification previously. Given the reality that “de-identification” is becoming increasingly uncertain as a method of shielding research participant identities, the EMA policy must ensure that it does not rely solely on de-identification to protect participant privacy. Other alternatives would be, in addition to “de-identifying data,” establishing intermediaries to limit access to data to appropriate parties who have agreed to terms and conditions that include a pledge not to attempt to re-identify participants.
Personal data protection	Participant Consent: The EMA policy must address the issue of participant consent on two levels. First, the release of data from past studies requires analysis of the consent forms used in those studies to determine if the data sharing now required by the EMA was adequately explained to the participant, or was otherwise arguably accommodated or included in the consent terms. In cases in which data sharing was not contemplated by the language of the consent, the EMA must decide whether retroactive participant consent will be required and also consider the feasibility of obtaining such consent. Second, on a prospective basis, consent forms must be modified to inform participants of the new EMA data sharing policies. Ideally the consent form should make participants aware of what data will be shared, who will control access to the data and what restrictions will be placed on the use of the data.
Rules of engagement	Gatekeeping Function: The EMA policy should provide for a learned intermediary to control access to all data released. As part of the data release process, the data requester should be required to submit an appropriate and scientifically valid study protocol and to demonstrate experience in the statistical analyses needed to make proper use of the dataset. The learned intermediary can evaluate whether the proposed

Scope	Comment
	data use meets a true public health need or is an attempt to gain a commercial competitive advantage or otherwise harass or harm research sponsors, researchers and or participants. Furthermore, the learned intermediary should require that the data recipient sign a data use agreement restricting how the data can be shared with others and prohibiting re-identification of participants. • The EMA may wish to establish civil or criminal penalties for violation of the data use agreement so that violators face sanctions beyond breach of contract liability.

238
239

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

Line Number	Comment and Changes proposed	Name	Affiliation
1	Restrictions on Content of Clinical Study Reports: The EMA has indicated that it could consider restrictions on the content of clinical study reports (CSRs) submitted in support of a marketing authorization. For instance, the EMA suggested that pharmaceutical companies would likely be strongly discouraged from discussing additional indications for use of a drug in the reports. Such a policy is problematic because it infringes on the free speech rights of pharmaceutical companies to inform government about additional indications for use of a drug in the CSR. If the EMA is going to require CSRs to be made publicly available, it will need to live with the consequence that some of the information disclosed may be of a type that the EMA would have preferred to keep out of the public domain as direct representations from industry sponsors of clinical research.	Mark Barnes & David Peloquin	Harvard Law School and Ropes & Gray LLP
2	Challenges to EMA's Decision Making: If patient-level data are made publicly available, advocacy groups are likely to use such data to challenge particular EMA decisions regarding drug approval. If the EMA policy only requires disclosure of data pertaining to drugs for which a marketing authorization has been approved, the EMA should be prepared for disease-specific advocacy groups to reanalyze data to argue that the drug should be authorized for a broader indication, and for industry to make the same arguments. If, on the other hand, the EMA policy requires disclosure of data for all trials used to support a marketing authorization, regardless of whether the drug was in fact authorized, the EMA can expect advocacy groups to harness all available data to argue that the drug should have been approved. This is not necessarily an adverse outcome, but it is something the EMA should be prepared for.	Mark Barnes & David Peloquin	Harvard Law School and Ropes & Gray LLP
3	Disclosure of Ethics Committee Member Names: During a recent EMA call, it was suggested that the names of the members of ethics committees and IRBs that approved a given study be made public as part of the data disclosure process. Such a practice would be highly problematic because it would almost certainly deter people from serving on such research ethics committees. Members of ethics committees and IRBs who are involved in approving highly controversial studies will not want to face a backlash from persons, groups, or companies that may be opposed	Mark Barnes & David Peloquin	Harvard Law School and Ropes & Gray LLP

Line Number	Comment and Changes proposed	Name	Affiliation
	to a particular study. The only information regarding ethics committees and IRBs that should be made public is the fact that a given study was approved by an IRB or ethics committee; there is no need for the names of the members of such groups to be made public.		
5	Comment: Overall EFPIA acknowledges the clarifications included in the version circulated on April 18. However, as previously stated, the implications of the release of patient level data on innovation and on individual patient protection and public health through re-evaluation of data by third parties needs careful consideration and discussion among regulators, patients, academia and industry to identify the best solution to balancing the desire for transparency with the need to foster innovation. We think that public health benefits include support for continued innovation and new drug development. The EFPIA would welcome an invitation from the European medicines Agency to a face-to-face meeting about the final policy proposal developing and reconnecting the five subtopics in a consistent general picture. Finally, our present comments (dated April 26) should be considered in addition to our previous written comments dated February 28 and April 12. The three documents should rather be considered as a whole.	Susanna del Signore	EFPIA
7	Comment: Alter text to include the proposed text below Proposed change: Different types of data pose different levels of risk of identifying patients. The risk is non-existent in case of reports containing only aggregated data, such as the main body of the clinical study reports and appendixes (excluding non- key-coded line-listings). Such reports should be considered for proactive publication and given unrestricted access. In case of the raw data and line listings, the risk may be higher in a few limited cases such as rare diseases. Unrestricted publication of de-identified raw data and line listings is useful and that is the focus of this working group. Unrestricted publication of de-identified raw data, and original key-coded raw data and line listings does not infringe protection of privacy.	Teresa Leonardo Alves	Prescrire, International Society of Drug Bulletins
10	Comment: Please delete "and given unrestricted access". Personal information should be redacted/deleted from the CSR (not de-identified). Access to (anonymised) datasets should be via companies, or an independent data custodian. From a patient confidentiality perspective (remit of CTAG1), as long as personal information is either	Susanna del Signore	EFPIA

Line Number	Comment and Changes proposed	Name	Affiliation
	suitably redacted or deleted then such a 'reduced' document could be made available. This is made possible since such a 'reduced' document typically contains only aggregated information and other information that may already be in the public domain (e.g. study results) and so the risk to patient confidentiality is low. Whether the redacted documents are suitable for publication or controlled release then becomes a matter of debate on other points beyond patient confidentiality. This could include topics such as timing of release, the matter of CCI (whose criteria should be agreed and which can be managed through redaction) and the general suitability of the language and format if released in an unrestricted manner, etc. The last point needs to balance the issues of positively increasing transparency and trust as well as the benefits to medicine and science, against the issues of very technical and sophisticated terminology and language being potentially misunderstood and/or misinterpreted. Proposed change (if any): from the perspective of patient confidentiality protection such reports could be considered for publication provided that compliance criteria with chosen rules of engagement are met and both intellectual property rights and the protection of commercially confidential information are fully respected.		
20	EFPIA is not in agreement about giving access to raw data and line listings outside precise rules of engagement	Susanna del Signore	EFPIA
68	As previously mentioned EFPIA would favour option 2	Susanna del Signore	EFPIA
79	Comment: Both those that conduct clinical and non-clinical research are concerned Proposed change (if any): please add "clinical"	Susanna del Signore	EFPIA
98	Comment: EPIA agrees with the text from line 98 to 104 concerning need for restriction on disclosure of employees names Proposed change (if any): no change	Susanna del Signore	EFPIA
111	Comment: Even as an example it seems odd to include items like car registration numbers that are not part of what is collected in clinical research Proposed change (if any): please delete this example	Susanna del Signore	EFPIA
123	Comment: Anonymising data listings is resource intensive. This should not be required. Rather anonymised patient level data can be provided by the companies directly where there is a scientific request with a protocol and a commitment to publish or through the mechanism suggested through the rules of engagement group.	Susanna del Signore	EFPIA

Line Number	Comment and Changes proposed	Name	Affiliation
133	Comment: That is available on the result summary that is posted. This should be cross referenced not repeated	Susanna del Signore	EFPIA
149	We have been discussing this in length with your colleagues working on the EU register of clinical trials EUDRACT_R and we've been asking ourselves for such results to become public for a long time. There will be inappropriate information, and all parts of the public will be facing this: patients as well as healthcare professionals, media etc. There will be safeguards: disclaimers, warnings not to take the information for granted and validated, and we trust that with time, patients and their organisations will also learn to be cautious when reading not yet validated results. Industry as well will learn that it could be counter-productive to post inappropriate information, maybe with a short term positive impact, but negative over the long term if the liability of the results a company will post is regularly questioned. I think what the group is proposing clearly limits the risks: origin of the information (sponsor of the trial, be private or public), the fact that it has not yet been reviewed by experts / validated by EMA, and prevention of promotional claims in the text. Patients and their organisations will most probably look at the figures (subjects' disposition and outcomes) more than to the narrative texts. Even if a summary contains comments that overestimate the benefit or underestimate the risks, my experience with patients' representatives covering scientific conferences where the sponsors are presenting not yet validated results confirms that most of the time advocates listen to the results with a critical mind, and discuss them with their peers or with experts before publishing anything about them. Maybe another safeguard could be to invite the reader to contact a patient organisation or an expert (e.g. their own doctor) in case he or she needs to discuss the results. To this end we've proposed to mention a link to relevant patients' organisations with each trial results. This will take time as it is not easy to do. The EMA has posted a list of patients' organisations on the EUDRACT_R web site and this can be improved. so, regarding the trials' results: 1. patients may have more difficulties than healthcare professionals to judge for themselves: this will always be true, but still, it should not be a reason to prevent patients to access the information 2. we're are not worried about this risk: patients are responsible people, they know the importance of discussing with other patients or with their doctor before making any decision that could impact their health 3. the EMA should have a policy that discourages sponsors from including promotional contents and	François Houyez	EURORDIS

Line Number	Comment and Changes proposed	Name	Affiliation
	maybe even propose remedies/penalties in case of infringements 4. a disclaimer that these are the views of the sponsor and that the information has not yet been reviewed by experts is welcome 5. an invitation to contact a patients' organisation or to talk with a specialist or own doctor in case of any doubt or difficulty understanding the results would also be welcome 6. when a peer-reviewed article is published, or when the EMA publishes an opinion in relation with the results in question, a link to the article or to the EMA assessment should be provided on the same page as the page of the results. This will help the reader accessing the opinion of experts, even if there is a delay between the information as posted by the sponsor and the expert review.		
183	Comment: EFPIA agrees with the view that even when accessed under strict rules with the obligation of personal secrecy, nevertheless data should be de-identified.	Susanna del Signore	EFPIA
n\a	Two comments on the latest version: 1. There are a number of WHO-approved clinical trial registers (see http://www.who.int/ictrp/network/primary/en/index.html) but ISRCTN (isrctn.org) and NCT (clinicaltrials.gov) numbers are probably the most well known in Europe. It is up to you if you want to expand the information to reflect that in 5.4. 2. Thank you for noting my suggestion about the phrase "proactive publication". I have been receiving the documents from some of the other advisory groups and note that they still use this phrase. This will need to be made consistent across all the policy documents.	Iain Hrynaszkiewicz	
n\a	ON THE USE OF FEAR MONGERING ● Claims that the disclosure of clinical trial data would risk misinterpretation of data and to the dissemination of skewed information that would scare the public reflect outdated paternalism. ● Again, proportionality in ethics has to be taken into account. There is overwhelming evidence of drug-induced harm being routinely hidden by pharmaceutical companies in detriment of public health while there is no example of misinterpretation of data and misuse from the last 2.5 years during which the European Medicines Agency has released clinical data to researchers on request. ● There is no evidence of data manipulation from data	Teresa Leonardo Alves	Prescrire, International Society of Drug Bulletins

Line Number	Comment and Changes proposed	Name	Affiliation
	sharing/open data.		
n\a	GENERAL COMMENT: CAUTION NEEDED; THIS DOCUMENTS DOES NOT REFLECT THE VIEWS OF ALL STAKEHOLDERS! Public access to detailed and summary raw data is particularly important to protect public health as it allows independent analysis. The discussions of this working group seemed to revolve solely around exceptions, rather than reflecting on the implementation of overarching principles to facilitate a policy of access to data, as foreseen by the Agency. Due to the large over-representation of the pharmaceutical industry the final document does not reflect the spirit of the policy that the EMA has been implementing since 2010: which is a policy of transparency and access to information concerning clinical data submitted during the marketing authorization procedure. Having said that, I would urge the agency not to consider this document - in its present form -when reflecting about the implementation of the policy, since the views expressed here are not representative of all stakeholders and focus on exploring exceptions to hinder access to data.	Teresa Leonardo Alves	Prescrire, International Society of Drug Bulletins

244