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

20 February 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. All proposals are deliberately kept at a high level to enable discussion. It is expected that more detail will be added during the discussion process.

The draft proposal has been amended to reflect the comments and discussion (summarised in comment boxes) received during the first meeting of the clinical trial advisory group on protecting patient confidentiality held on 5 February 2013.

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 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 (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.

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.

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 persoonel as essential to be made public. In general, for clinical trials there is no great concern for revealing the names of investigators and study 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.

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, 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.

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 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.

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 PUBLICATION”, and the de-identified data clearly labelled as “SUITABLE FOR 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.

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 application of transformation methods to de-identify data may reduce the possibility of exact replication of 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.

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. 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:

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.

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. 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.

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.

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:

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

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

Comments:

Possibly, standardised formats should be developed to facilitate this.

3.9. 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.

Comments:

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.

3.10. The Agency may verify that the stated methodology conforms to standard transformation methods to de-identify the data. 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.

3.11. 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

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
2	EFPIA member companies recognise the potential benefit of providing scientists with information and data submitted to the EMA for further research. In this light, the pharmaceutical industry registers extensive information on their clinical trials at the time of initiation and publishes both positive and negative trial results through numerous channels (e.g., peer reviewed publications, EU CT register and clinicaltrials.gov). • However EFPIA contends that there are numerous essential policy, scientific, technological and legal issues to resolve prior to considering implementation of a new approach to access to clinical trial data.	Susanna del Signore	EFPIA
4	Five foundational principles underpin our responses to the questions below: ☐ Data and information on clinical trials that are not already publicly available should only be provided to other qualified scientists for legitimate research purposes on a case-by-case basis, directed by a scientifically sound hypothesis and research analysis plan ☐ The provision of data and information must be done in ways that minimise risks to research participants' privacy and commercial confidentiality ☐ Research use must align with permission provided by research participants through the informed consent obtained in the original clinical studies ☐ 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. ☐ Finally, but of chief importance, the MAH should always be consulted before release of information or data with the opportunity to comment and seek redactions	Susanna del Signore	EFPIA
10	EFPIA aknowledges that the current document reflects most of the new elements raised during the 5 February TC. Some EFPIA key comments are nevertheless summarised hereafter, specifically about protecting patient confidentiality: • EFPIA would recommend agreeing over a set of best practice rules aimed to effectively protect Patient Confidentiality, to restrict secondary research toward public health	Susanna del Signore	EFPIA

Line number	Comment and Changes proposed	Name	Affiliation
	benefit and to prevent the risks of 'bad science' or 'misuse" of various kind. • In the internet era, "key-coded" data on an individual in a clinical trial can not be considered anonymous, and should be handled as personal data falling under the personal data protection rules (EU and National). Should a broader access to such data be given, a special set-up is required to ensure proper protection of the individual and to ensure that the legal responsibility is clarified; mitigation strategies could be put in place to decrease the risk of re-identification; the risk of re-identification will not decrease to zero. We consider key-coded data as not appropriate for release. Although key-coded clinical trials data have had direct identifiers such as name and address removed, there may be other indirect identifiers included in a key-coded data set, such as patient initials, diagnosis, patient date of birth, and other dates related the patient's treatment (e.g., hospital admission/discharge dates). These indirect identifiers can sometimes be used in combination to re-identify an individual who is the subject of the data. If the same indirect identifier is present in several datasets it may even be enough with a single indirect identifier to re-identify a patient given that several datasets, containing the indirect identifier, are combined. 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. • In view of the above elements, EFPIA would favour 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. • Liability issues are also a reason of concern. In case of re-identification it is not clear where liability will stay Should the Applicant be liable for a process that is out of its control or for retrieval of personal data from documents shared in confidence and not originally intended for public disclosure? Will the Agency assume a role of gatekeeper to enable the respect of good rules of engagement?		

Line number	Comment and Changes proposed	Name	Affiliation
	• 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.		
16	Comment: Protection of personal data is mentioned in many places in this document. The issues raised are complicated and the debate is important but is there a risk that we are wasting our time as ultimately decisions may be made by lawyers? I understand that EMA has already taken a decision on its policy but it may be reassuring to hear that discussions with Information Commissioners and Ministers of Health throughout Europe have informed that policy.	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London
16	Is it really acceptable to cut all ways to identify a study subject retrospectively? Identification should be possible also after database closure to - prove that the patient really exists - in case of a Schadensfall for insurance purposes - in case of new medical and scientific knowledge a reevaluation of the data may be possible. This can result in new data and information relevant for the study subjects to know. Key-coding should be acceptable and the investigator is the owner of the patient identification list.	Dr. Uwe Gessner	Pfizer Pharma GmbH
20	Comment: EFSPI thinks that any advice on data formats and data anonymisation should distinguish between data from completed studies versus future studies and between submissions of applications (or subsequent submissions) as of 2014 versus submissions from before 2014. EFSPI is of the opinion feels that the grandfathering principle should be applied, meaning that legacy data can be submitted as analysed (after appropriate anonymisation).	Stefan Driessen	EFSPI
46	Comment: We query whether the 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. Proposed change (if any): lines 46-64 amended to reflect this point.	Grant Castle for Christiane Abouzeid	BioIndustry Association (BIA)
47	Comment: I suggest that this is almost impossible as data subjects will be able to identify themselves. Perhaps this is regarded as unimportant as I guess they will be able to request the information anyway. However in hospital clinics discussions with patients usually revolve around one or two	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London

Line number	Comment and Changes proposed	Name	Affiliation
	key markers of disease or disease progression, with remaining measurements / assays merely mentioned as unremarkable or normal. 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.		
52	Comment: Name, date of birth, address are obvious common identifiers but in clinical trials of chronic disease patients often record key targets such as BP, cholesterol, Hba1c, etc, or keep diaries of quality of life, seizures, etc. These are frequently discussed with other patients or relatives who sometimes accompany patients to clinical consultations. Doctors in hospital clinics often write to patients following clinic visits confirming values of key tests. Patients will easily identifiy themselves without need of the common identifiers.	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London
57	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(about the Informed Consent and the Ethical review) could change prospectively, however is not the case today 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. Moreover, about clinical data that are part of a submission to the EMA, as of today these are not formatted in order to undergo (secondary) statistical analyses. For the data format to be used a “grandfather principle” should apply, i.e. data should be made availabele in the format used by the company for the analysis irrespective of the type of information to be made publicly available and the intended use of it.” This kind of existing information, like narratives or lined data in tables should be carefully redacted in order to avoid disclosing details e.g. birth date, gender, rare disease status or name of the	Susanna del Signore	EFPIA

Line number	Comment and Changes proposed	Name	Affiliation
	hospital all could facilitate re-identification. It has been aknoledged that anomymisation and redaction these are very different operations. And that "key-coded" data are not anonymous, should be considered as personal and still falling under privacy data protection rules. Scope and definition should also separate the case of pro-active publication from third-party request based nominal release. Proposed change: 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. 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. o 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 o 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. o 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		

Line number	Comment and Changes proposed	Name	Affiliation
	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.		
57	What is meant by "original key-coded data"? Proposed change (if any): line 44; replace "investigator" by "investigator, and the data collected in the study for the analysis and reporting is key-coded data".	Stefan Driessen	EFSPI
64	In view of the above elements, EFPIA would favour 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. Proposed change: 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.	Susanna del Signore	EFPIA
66	Comment: Exactly who are the clinical trial personnel? In monitoring chronic disease in hospital clinics a doctor may, in consultation with the patient, make a decision that affects their trial treatment. That doctor is responsible for routine clinical care but is not part of clinical trial personnel.	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London
66	EFSPI disagrees with company's personnel personal data to become public for the sake of public interest or public health. 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, EFSPI is of the opinion that the same rules should apply to any requester of the data for the purpose of additional analyses as to the originating company that perfomred the intial analyses. In order to ensure good scientific practice and in the interest of public health, anyone wishing to analyse aggregate	Stefan Driessen	EFSPI

Line number	Comment and Changes proposed	Name	Affiliation
	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 acces clinical trial data. Proposed change: EFSPI favours Option 2 over Option 1 but would even opt for Option 3: no disclosure of personal data of industry personnel.		
74	EFPIA is in favour of option 2.	Susanna del Signore	EFPIA
80	Comment: This is not clear enough. There are several levels of "public". 1st level: providing (a relatively detailed level of) personal data to the sponsor for evaluation the qualification of the investigator to conduct this clinical trial. 2nd level: providing personal data to Regulatory Authorities and Ethics Committees for evaluation. 3rd level: a smaller dataset will be forwarded to the public audience. 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.	Dr. Uwe Gessner	Pfizer Pharma GmbH
89	Note that investigators are not company (study) personnel. The argument to disclose study personnel because the names of investigators are anyhow already widespread, does not hold.	Stefan Driessen	EFSPI
104	EFPIA is in agreement	Susanna del Signore	EFPIA
107	Comment: In Germany according to the Drug Law the principal investigator and and a delegate in his/her absence are responsible for conducting the clinical trial at the study site. So at least these two staff members have to offer their personal data. Local requirements in other EU countries should be checked. With regard of frugality of data the number of persons who should offer their personal data should be limited to truely responsible persons.	Dr. Uwe Gessner	Pfizer Pharma GmbH

Line number	Comment and Changes proposed	Name	Affiliation
108	Nevertheless the issue of alignmnet with the original informed consent still applies, a problem that could possibly be solved prospectively, but very bordersome for the existing documents. It also reinforces the need for a controlled case-by-case assessment approach that uses a risk-benefit approach to decision making.	Susanna del Signore	EFPIA
122	Comment: Why cannot a fully encrypted dataset be passed to a specific independent agency, perhaps a department within EMA, to undertake a re-analysis?	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London
122	According to EFSPI, this is one of the major issues to be solved. It is stated that there may be situations that it may be important to access patient data. EFSPI strongly feels that in the majority of cases for the purpose of re-constructing study results it is essential to have individual patient data. (For example, in case the primary analysis was consisting of a so-called Analysis of Covariance (ANCOVA) with as covariate "age", then in order to reconstitute the same results one absolutely needs the individual patient's age value.) And so in these cases a choice needs to be made; de-identification of the data might lead to the unability of any requester of the data to reproduce all results of the study. Unless, a completely different model is chosen, which could be called a server-solution; requesters access patient level data on EMA servers and analyse the data on the server and can only download summary statistics. Proposed change: EFSPI strongly feels that in the advice to EMA a clear statement should be included that full patient de-identification will in many cases be incompatible with re-production of study results by any third party requester of the study data.	Stefan Driessen	EFSPI
127	It should be defined what data should be made available and what data remains private.	Dr. Uwe Gessner	Pfizer Pharma GmbH
135	Comment: Not just patients but their carers as well. In trials where patients lack capacity to consent (children, dementia, mental illness, those who are unconscious, etc) assent to enter a trial has to be given by a third party. In these circumstances the issues around trial entry are difficult, distressing, and may be prolonged. Having to assent to release of data with no specified purpose at some undeclared time in the future will, I believe, gravely affect recruitment. Indeed clinical trials in these vital areas may become too difficult.	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London

Line number	Comment and Changes proposed	Name	Affiliation
135	Proposed change: Releasing data, even de-identified, may give rise to severe reactions, e.g., in patients suffering from psychosis, or facilitate discrimination, e.g. elderly in patients with dementia. Patients with rare diseases are particularly ar risk of re-identification. Further to the risk of accidental re-identification through a combination of indirect identifiers it subsists the risk of intentional re-identification, misuse of retrieved data for commercial purposes where selected indirect identifiers are recompiled and sold to interested parties "good and poor performing patients" per site and per country, insurance investigations, etc. "Discrimination", infectious diseases, dementia, patients with mental illnesses, etc., is as well a risk to be taken into consideration. These cases could bias future clinical trials and discourage effective participation to clinical research.	Susanna del Signore	EFPIA
138	In reference to an earlier comment, EFSPI feels that EMA should establish the rules for de-identification and not each individual company. The rules should also be such that it is to be expected that adherence will preculde patient de-identifcation even when applying all kind of linkages with other (social media) data carriers. In case, requesters of this de-identified data can not run their analyses as desired, then a procedure should be set in place to escalate the request to EMA and to align with the originating company the acceptablity and reasonability of the request as well the means for execution. This implicitly means that requesters of data should make themselves known (and be legitimate researchers and appropriately expereienced and qualified to run such analyses). Proposed change: EFSPI feels that it should be made clear in this section that EMA should set the rules for de-identifying study data.	Stefan Driessen	EFSPI
143	While from data protection view completely de-identified data is the goal, on the other hand only key-coded data offers the chance to re-evaluate the data. It should be clearly stated on which level of the study data generation a key-coding is acceptable and on which level or when the complete de-identification is needed. Example: it can happen that a drug is working only in a specific subgroup of patient population. To find out more about the specific biomarkers triggering the drug effect it can be necessary to initiate additional evaluation e.g. testing of tumor samples or blood samples. This would not be possible if the de-identification hinders the identification of	Dr. Uwe Gessner	Pfizer Pharma GmbH

Line number	Comment and Changes proposed	Name	Affiliation
	the study subjects.		
149	What is meant by "Publication" should be clarified: is: (i) to proactively make public; or (ii) only to respond to specific requests for information. From a Privacy Law perspective, this makes a big difference. As previously mentioned, the risk can not be assessed in absolute terms and is context- and time dependent. o Patient-level data should not be published and; o It is preferable that anonymised patient level data should not be 'released' or 'published'; but that access to anonymised patient level data for legitimate research purposes should be provided in a protected environment. We recommend to put in place gate-keeping principles and process, allowing a case-to-case assessment based on the actual use e.g. public interest justified metanalyses.	Susanna del Signore	EFPIA
150	It should be clearly defined which personal data from patients can be used. Example: meanwhile it is state of the art not to enter name initials or the date of birth into case report forms. On the other hand often initials and date of birth are entered on lab request forms or can be found on radiology exams or other examination reports. So these data are coming through the back door into the study database. This should definitely prevented and stopped by this guideline. To collect patient identifyers like phone numbers or registration numbers is obsolete. Also race attributes should be collected only, if there is a strong medical rationale for this (example: it can happen that the single datapoint "black american" can identify a subject if he/she is the only one in a certain country taking part in a clinical trial.)	Dr. Uwe Gessner	Pfizer Pharma GmbH
155	The general rule should be to retrieve as little as possible data and not as much as possible.	Dr. Uwe Gessner	Pfizer Pharma GmbH
167	EFPIA agrees with Option2	Susanna del Signore	EFPIA
167	EFSPI is favouring Option 2 over Option 1.	Stefan Driessen	EFSPI

Line number	Comment and Changes proposed	Name	Affiliation
184	EFSPI would prefer to see that any journal confronted with a re-analysis of data should sollicit comments from the originating company in the interest of transparency and good research abiding to hearing both sides.	Stefan Driessen	EFSPI
186	Comment: EFPIA agrees with Option2. Moreover, the 'level' of de-identification required to protect patient confidentiality needs to be assessed almost on a case-by-case basis. Proposed change: 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.	Susanna del Signore	EFPIA
186	EFSPI is favouring Option 2 over Option 1.	Stefan Driessen	EFSPI
195	Can not agree with option 1. It is not possible to control 2 datasets. Misuse is possible.	Dr. Uwe Gessner	Pfizer Pharma GmbH
214	Comment: Some explanation of divergent results will also be needed to patients entered in the trials. Indeed the controversy arising from disputed results may cause distress to patients wondering exactly what sort of research they have engaged in. The trial funder may have a duty of care to provide an explanation.	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London
218	According to EFSPI it is generally impossible to run meaningful additional analyses on the basis of only aggregate statistics from the trial data.	Stefan Driessen	EFSPI
225	EFSPI strongly feels that 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.	Stefan Driessen	EFSPI

Line number	Comment and Changes proposed	Name	Affiliation
227	Can one be more specific about how methods should be "individually tailored"? 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. 3.8. Applicant companies shall describe in general terms and justify for each document the anonymisation methods used. Is this agreed? We would rather suggest keeping flexibility and avoiding cumbersome processes	Susanna del Signore	EFPIA
235	Providing a de-identification description "in general" is reasonable, however, providing this on a document by document basis seems to be overly burdensome and not value-adding. Once again, EFPIA sees support for a case-by-case gatekeeper approach.	Susanna del Signore	EFPIA
238	Comment: In formats has to be specified which procedures, precautions and safeguards are followed	Hilje Lotenberg van der Grient	ELPA
238	EFSPI feels that EMA is to set these standards. It is unprecedented that individual patient data will be publicly availble 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.	Stefan Driessen	EFSPI
239	EFSPI strongly disagrees; see previous comment.	Stefan Driessen	EFSPI
240	EFPIA does not agree:We have already described the technical limits of de-identification operations.	Susanna del Signore	EFPIA
248	EFSPI feels that EMA should set the rules consistently and clearly thereby indicating what the consequences of the se rules can be in individual cases with respect to the analytical utility of the data. Proposed change: delete 3.8,3.9. 3.10 and change 3.11 into "The Agency will come forward with guidance to all involved companies regarding the methods of use for patient de-identification".	Stefan Driessen	EFSPI

Line number	Comment and Changes proposed	Name	Affiliation
n\a	Introductory observation The proposal assumes that only the EMA and the trial sponsor have the responsibility for protecting patient confidentiality. In my view the overriding ethical responsibility rests or should rest with the lead investigator of the trial, though the EMA and the sponsor must support him or her in fulfilling this responsibility. This argument derives from the experience of the last 50 years, in which adverse effects of medicines have been universally underreported and inadequately investigated, partly because patients have not been systematically followed up. Most are systematically lost to follow up, largely or partly because confidentiality rules have made it very difficult. The early detection and investigation of harmful effects is in the interests of patients and the community, and if patients understand that they will accept it and work together with professionals. When invited to take part in a trial all patients should be asked to agree to being followed up by the trial team or its successors (but not the trial sponsor or a body acting on its behalf). Follow up should be a separate part of the trial plan, for which the lead investigator should be responsible. S/he would therefore be the custodian of the patients' personal data, and so equipped to investigate later harms. This does not affect the scope and definitions in para 1 nor the trial personnel data (para2). Introductory observation The proposal assumes that only the EMA and the trial sponsor have the responsibility for protecting patient confidentiality. In my view the overriding ethical responsibility rests or should rest with the lead investigator of the trial, though the EMA and the sponsor must support him or her in fulfilling this responsibility. This argument derives from the experience of the last 50 years, in which adverse effects of medicines have been universally underreported and inadequately investigated, partly because patients have not been systematically followed up. Most are systematically lost to follow up, largely or partly because confidentiality rules have made it very difficult. The early detection and investigation of harmful effects is in the interests of patients and the community, and if patients understand that they will accept	Andrew Herxheimer	

Line number	Comment and Changes proposed	Name	Affiliation
	it and work together with professionals. When invited to take part in a trial all patients should be asked to agree to being followed up by the trial team or its successors (but not the trial sponsor or a body acting on its behalf). Follow up should be a separate part of the trial plan, for which the lead investigator should be responsible. S/he would therefore be the custodian of the patients' personal data, and so equipped to investigate later harms. This does not affect the scope and definitions in para 1 nor the trial personnel data (para2).		
n\a	Para 3.2, I would say no; 3.3 yes; 3.4 'sufficiently low' seems acceptable; 3.5 to 3.8 seem to need discussion with examples; 3.9 yes; 3.10 better at first to do it systematically for all and review this after a trial period; 3.11 yes	Andrew Herxheimer	
n\a	- Definitions used should be compliant with ICH GCP at the first place, as a common international standard. In this respect In addition to subjects / patients should be considered investigators, sponsors and ethics committees as well. So far no references have been made related to ethics committees In relation to this topic. - should also consider the trials not performed In EU sites but other nonEU countries as well, including USA – and the level of requirements related to personal data protection In all these countries, as well as national requirements In EU countries In relation with EU directive In data protection. - We are discussing past and ongoing clinical studies as well as new clinical studies – how this will practically impact all these different type of clinical studies? - Why full complete access would be required for non-medical/non-science people? Why is considered to be of help for a person of nonmedical/nonscientific background to have access to complete study report? - Could be considered a standard text to be included In all informed consent forms related to this aspects , once clearly defined what will be made public, at what extend, In what format, for whom, until when, etc - If for patient/volunteers the document to be considered is the informed consent form for investigators/institutions respective aspects should be included In study specific financial contracts, or for others In other type of contractual agreements / including employment agreements – Could be also for these contracts a standard text needed?	Cristina Oana Micsescu	

Line number	Comment and Changes proposed	Name	Affiliation
n\a	Consistent terminology The terminology is not consistent throughout the document, and this makes things confusing. For example, terms such as “anonymized” and “redacted” are used. A suggested, and simplest, approach is to talk about the “risk of re-identification” being “very small” or “sufficiently small”. This is consistent with current de-identification (sometimes also referred to as anonymization) guidelines from regulators. We would suggest that this be stated up front as the basis for dealing with the privacy issue, and then all subsequent points refer back to that. The concept of risk is mentioned in clauses 3.3 and 3.4, but that is not used consistently throughout. Also, the notion of “absent” risk is introduced. In general, it is not possible to have an absence of risk if any data will be disclosed. Therefore, the objective should be very/sufficiently small risk rather than an absence of risk. 2 Risk The definition of “very small” or “sufficiently small” has to be risk based. That means that it needs to take into account the context, such as the sensitivity of the data and any conditions that will be imposed on access to that data. This is consistent with the recent code of practice from the UK ICO and the guidance from the US HHS, to name a couple of recent examples. Because a risk based approach is the only defensible one, the cited article in BMJ/Trials should not be used as the basis for the de-identification method. That approach is not risk-based, does not use any metrics to evaluate the risk of re-identification (its stipulations, for example, shifting dates, but shifting dates cannot guarantee that the risk is very/sufficiently small because there is no requirement to measure the risk after such a transformation), and uses lists of fields to remove as the primary method of de-identification. That kind of approach has received considerable criticism and has little credibility in the disclosure control community, and does not withstand the test of time (the list of fields may change over time). It would be a poor standard to use for this purpose. One of its more serious problems is that it would allow data with a very high risk of reidentification	Khaled El Emam	University of Ottawa

Line number	Comment and Changes proposed	Name	Affiliation
	to be disclosed and can harm the whole initiative of making data available because it would result in a high risk of re-identifying patients. 3 Documentation It is important that the sponsor document the de-identification process. This is already stipulated in the document that was distributed. However, it would be useful to provide some suggestions or set some expectations about what should be included in that documentation. For example, all assumptions must be stated (e.g., about sampling fractions or adversary knowledge), reasoning for the definition of "very small", evidence that the actual risk in the data meets that threshold (eg, metrics used and their values), and the actual methods used to de-identify the data (this is critical for analysts to judge the impact of the transformations on data quality). 4 Mitigating Controls The binary distinction of published and not published needs to be thought about further (or at least clarified as the distinctions are not very clear). A sponsor may have a public data set, and a data set that is disclosed under certain conditions (such as an analyst/data user signing a data use agreement prohibiting re-identification attempts and requiring certain security practices to be in place). The former would have a more stringent definition of "sufficiently small" because there are no controls on the data, while the latter would have a more permissive definition of "sufficiently small" because of the added controls to manage the risk. A completely public data set may limit the kinds of analyses that can be performed. A suggestion would be stipulate that sponsors disclose at least a public data set, and then provide one or more versions of that data set where additional controls may be imposed. There may be more than two data sets (as suggested in clause 3.5). 5 Other Sources Another source that may be useful consider is this book (as per clause 3.7): http://www.amazon.com/Guide-De-Identification-Personal-Health-Information/dp/1466579064/ which is specific to the disclosure of health data, and describes a methodology for managing reidentification risk.		

Line number	Comment and Changes proposed	Name	Affiliation
n\a	I have reflected on the exchange of opinions around Item 3 of the draft paper and suggest that it would be useful to discuss - at the next virtual meeting - the concept of informed consent to limited disclosure by data subjects (trial participants) prior to recruitment to any clinical trial. Rebecca Li identified that it might be a logistical nightmare to obtain patient consent to their personal data being disclosed to third parties unless it was an established part of the entry criteria to any trial. Manfred Belent suggested that if such informed consent could be given then it might minimise - if not remove - the need to transform individual data items. Andrew Herxheimer also reminded us that it might be a good idea to ask the patient (participant) if they agreed to data being made available for related secondary research, to which Jose Drabwell - agreeing with this suggestion - proposed the use of a term such as "reputable medical researcher". Meeting participants will not need reminding that true informed consent requires an understanding of risk and benefit that eludes the majority of the population including many clinical practitioners. (as demonstrated by Gerd Giggerenzer among others). Yet while agreeing to the initial objective that data will be made accessible we were reminded by Anthony Brookes, Stefan Driessen and others that there is a trade off between the open sharing of data - allowing many eyes to review and if possible reproduce or repudiate (prove/disprove) trial results - and the likelihood of tracing data items back to individual participants. It was almost suggested that achieving 100% of one equated to 0% of the other. As Teresa and to a lesser extent Susanna del Signore reminded us it is important not to provide false pretexts for companies to withhold or change trial data and that it may well be necessary for the EMA to be in a position to wield sanctions in the event of it identifying such abuses as only in this way will there be a clear mechanism to enforce compliance to best practice standards. As an unaffiliated patient I do not want concerns about confidentiality to limit the potential benefit of participation in clinical trials. To do so will not fully reward the altruistic motives of many trial participants and may indeed contribute to long term harm at population level. Much like understanding that the safest PC is one that is locked in a safe and never switched on but will be of absolutely no use to anyone - I do believe individual participants (or those who are by law able to speak for	David Symes	

Line number	Comment and Changes proposed	Name	Affiliation
	them) will be able to understand the nature of the trade off between openness and confidentiality if properly described. Many decisions benefit from a few simple questions being addressed such as: (1) What do we want to achieve? (2) Why is it important? (3) What would happen if we did not do this? Regarding individual participation such simple questions might be: (1) What are we asking you to do? (2) Why are we asking you to be involved? (3) What will happen to the data we will collect and analyse? (4) What will we do to minimise the risk of your personal data being used for purposes other than the purpose of this trial? (5) Are you content to allow personal data being shared with other licensed researchers approved by EMA provided the EMA confirms such secondary use of your personal data is aligned with this informed consent? If others agree that it might be useful - rather than discuss these questions at the next meeting - perhaps the support team could see if meeting participants have examples of participant consent statements that would address the ability to obtain a truly informed consent that balances the risks and benefits of disclosure and confidentiality.		
n\a	I should be grateful if the following comments can be circulated to the group, as suggestions for principles to follow: Efforts to prevent the identification of trial participants should not damage the ability of researchers to attempt to replicate the original analyses or to conduct important, new analyses on the data. Confidentiality of trial participants can be protected by limiting access to the individual participant data so that recipients are highly unlikely to know enough about any participant to identify them from some of the data items. Anyone who is granted access to the individual participant data should be bound by appropriate regulations to respect the confidentiality of the participants and not to disseminate any identifying information if they become aware of it. Any processes that are adopted should be such that researchers in those trial are still able to reassure trial	Mike Clarke	

Line number	Comment and Changes proposed	Name	Affiliation
	participants that it will not be possible for anyone that they would not wish to see their data to identify them and access their data.		
n/a	"The following principles are recommended for organizations that conduct, sponsor, or regulate health research involving personally identifiable data. They can be transposed into professional guidelines, standard operating principles, regulations, or laws. Criteria and procedures should be established that are specific to the context. "Overall in health research, cultivate an atmosphere of respect for the privacy of the people whose health experience is being studied. Collect or use personally identifiable data only if the research is worthwhile and identifiability is required for scientific reasons. Urge Institutional Review Boards and other ethics review bodies to become fully engaged with the privacy, confidentiality, and security aspects of subject protection, in secondary research on data as well as in direct experimentation. Respect such standard fair-use practices as announcing the existence of data collections, allowing data-subjects to review data about themselves, and the like. If for scientific reasons exceptions have to be made to normal practice, this should be discussed as part of the informed consent process before the study starts. Attend sensitively to informing data-subjects and gaining informed consent. Safeguard personal identifiers as close to the point of original data collection as possible. Enforce a policy of "No access to personally identifiable information" as the default--then base exceptional access on need-to-know. Generally limit the cordon-of-access to personally identifiable data. Allow access for formally justified research uses and to appropriate researchers. Maintain and monitor access "audit trails." Remove data-subjects' personal identifiability as thoroughly as is compatible with research needs. If key-coding, aggregating, or otherwise removing personally identifying information, do so with adequate rigor. Maintain proper physical safeguards and cybersecurity measures. Periodically challenge them, to test their adequacy. Develop policies on seeking or allowing secondary use of personally identifiable data, and on the associated conditions and safeguards.	François Houÿez	

Line number	Comment and Changes proposed	Name	Affiliation
	Before either (a) transferring data to other researchers or organizations, or (b) using data for new purposes, make conscientious decisions as to whether to proceed and what the privacy protections should be. Then if proceeding, implement appropriate protections. Sensitize, train, and certify all personnel who handle personally identifiable data or supervise those who do. Make data stewardship responsibilities clear. Maintain internal and external accountability."		

272