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

04 April 2013 - Version 9.0

This advisory group discussed the issues and questions listed below and offers the following views and positions for EMA's consideration:

1. Should the marketing authorisation holder be consulted before EMA discloses clinical trial data, in regards of commercial confidential information (CCI)? What elements of the clinical part of the dossier could be considered CCI?

No agreement was reached. The following positions were discussed:

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

Although the situations would be rare (perhaps when working with a new therapeutic class or a rare disease) it is possible that eCTDs and CSRs would contain competitively valuable information. The sorts of information (with historical examples that are no longer competitively relevant) are:

- Methods to pursue newly validated / devised endpoints that are persuasive to regulators:

e.g., the suite of validated measurements for assessing the effects of migraine on the whole body in support of the approval of a drug

- Identification of investigators that recruit well, especially for rare diseases / difficult patient populations

- A novel trial design, streamlining and making more economical the proof of efficacy for a novel compound:

Regarding the question "What elements of the clinical part of the dossier could be considered CCI after a marketing authorisation is granted?", the following examples were given:

information on the rationale or R&D strategy for the new medicine; new assay methodology for biomarkers; new validation methodology for a Patient Reported Outcomes; additional clinical results not included in the CSR but which are used to support the regulatory review (would be CCI until those results are released in a publication).

According to another position, what is CCI will need to be determined case-by-case, following consultation with the sponsor, as it will depend on factors such as the specific product, the way in which the documents have been written (will vary from sponsor to sponsor – some may have included information that may be CCI), and the timing of disclosure relative to the time of marketing authorisation.

- CSRs may contain information on bio-analytical product-characterization methods which are the intellectual property of the MAH - public disclosure could be an infringement of the MAH's IP rights. Furthermore, the use of some specific analytical tests described in the CSR can provide information indicative of the active product substance/molecule that can therefore be identified and used by competitor companies (e.g. tests on molecule-specific epitopes providing information allowing identification of the commercial confidential molecule).

Of particular concern with the proposed proactive broad disclosure of clinical trial data is the potential for inappropriate use of such data by third parties either to circumvent existing regulatory data protection (RDP) rules, or take advantage of the absence of such rules in the many countries which do not have robust systems of RDP equivalent to that in the EU. For instance, data exclusivity in Australia, China and Mexico is directly undermined by publication of the relevant data, anywhere in the world.

Industry contends that if data are obtained from EMA under its disclosure policy and used lawfully in a third country then the EU MAH would have no legal redress.

However, even if a CCI was defined, open access should be restricted ONLY for this sensitive part of the CSR. Moreover, EMA consultations to MAH should not imply long delays in releasing data.

b. EMA's consultation with the marketing authorisation holder (MAH) prior to disclosure will introduce delays that detract from the concept of "proactive" disclosure. Whether or not a particular material can be disclosed, and under what terms, should be decided prior to readying materials for disclosure.

With regards to the examples of CCI listed above: Some of the examples should nowadays not be legitimate examples of commercial sensitivity. At the time these drugs were being developed, they may have been thought to be legitimate examples simply because of the way drug development was done then. Today, these examples should be regarded as being examples that overall make clinical development more efficient and as such should be shared. Furthermore, if the new method, endpoint... is an argument for the approval, it should be made publicly available in the EPAR and properly described in any guideline applying to the evaluation of products in the indication.

It is emphasised that "competitively valuable information" is not necessarily CCI. For example, a negative study result is obviously competitively valuable information, but this should not make it CCI. CCI does not exist as far as clinical data are concerned.

Study methods and study results are never CCI. The information is essential for the interpretation of the study results and should be available for the public. EMA's policy will ensure that this will be done only after a decision about marketing authorisation has been made.

Third-party requestors may need some of this "competitively sensitive" information to carry out proper re-analysis and verification of results, such as trial protocols, but may not necessarily need all of them (e.g. identification of investigators that recruit well). Most of the information on 'good investigators' in CTD and CRS will also be available in publications.

Identity of investigators should always be public in order to make clear any conflicts of interest between MAH and professionals.

The questions listed below addressed the issue: what steps will a requester have to go through before being able to access clinical trial data from the EMA website? After accessing the dedicated domain of the EMA website:

2. Should requesters have to identify themselves?

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

1. Aggregate data: No agreement was reached. The following positions were discussed:
    a. There is no convincing rationale that identification of requesters could or should be required. Such data should be accessible freely (similar to EPAR information today). It is assumed that aggregate data contains no or few personal data (any personally identifiable information must be removed prior to release unless justified to remain). It is pointed out that the aim of transparency shouldn't be only to allow a potential reanalysis. For example, drug independent bulletins need full information of clinical trials not for research purposes but for education purposes in health areas. A watchdog activity is high useful to citizens and also for drug regulatory bodies. So in many cases there won't be a "legitimate scientific question" to be considered. Transparency goes beyond reanalysis purposes.
    b. In the interest of transparency, requesters should be identified, logged and their identity made public, primarily to ensure patient confidentiality is not compromised and to avoid the misuse of patient level data by third parties with commercial interests that are not related to healthcare research. It is technically possible to accurately identify requestors; one could perhaps use an ORCID ID to identify requestors.

    Note: It is pointed out that some ICH compliant clinical study reports might contain patient level data. Examples are patient narratives for serious adverse events, and sections in the report discussing these cases on an individual basis. For these parts of the study reports, considerations of personal data protection must be taken into account.

2. Patient-level data: No agreement was reached. The following positions were discussed:
    a. These data should be freely accessible <u>without</u> the need for identification. Arguments in favour of this position include (not in order of importance):
        i. Lowering the hurdle for patients who wish to access data related to their own disease. Asking requesters to publicly share their personal details, education and training before getting access would violate data protection regulations and induce a hurdle for non-professional user groups. Also, the rules of engagement should not include any pre-selection or pre-identification and publication of the requester name for a simple reason: a patient can ask for the data about a product he has to take for his/her disease. If specific qualifications are requested, one will easily know who are the requesters with a personal interest in the product (those without clear qualifications).

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

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

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

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

vi. Any suggestion that requestors of clinical trial data should also have sufficient qualifications and experience for any subsequent analysis of data is neither practical nor desirable for either aggregate data or patient-level data. It would entail subjective and arbitrary judgements about what qualifications and experience are "sufficient".

vii. The privacy of study participants is important and their privacy should be warranted. On the other hand, the privacy should also be warranted for study participants, patients or other (EU) citizens who like to access patient-level data for their own private use. Namely, publication of their name on the internet involves the risk of unintended use of the personal data of this person, especially if this information can be detected by search engines such as Google. For example, the information (name + type of medication) may be detected during a background search performed for a job application; the information can be used by insurance companies; or the information can be used for direct marketing for registered or falsified medicines, including spamming. This is an argument to carefully consider whether the benefits of publication of the names of private persons outweigh the risks of unintended use and breach of privacy of those who access data. Thus, benefits of publication of the names of those who access patient level data may not outweigh the risks, because publication of personal data in combination with (type of) medicines for which data have been accessed creates the possibility for unintended and undesirable use of personal data;

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

ix. Concerns about inappropriate analyses are misplaced, since the scientific community will or will not give their support to these analysis based on its scientific value; however, it was also discussed that the venues for inappropriate analysis and hyperbolic interpretations include the popular media, who often do not defer to the scientific community.

x. It is pointed out that the aim of transparency should not be only to allow a potential reanalysis.(see arguments above, para 2.1.a.)

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

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

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

aggregate data. Access to IPD restricted to being for the purpose of research in the interest of public heath - as demonstrated by provision of a protocol or research plan, disclosure of investigator name and affiliation and declaration of any potential conflict of interest (preferably at the point of release of data, but delayed if necessary);

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

iv. It is possible (and will be even easier in the future) to combine anonymised data sets with other data that is readily available publically to identify individuals. This is important for privacy particularly as the data contains health information that can be sensitive and assumed to be private by the clinical trial participant. For example please see : http://online.wsj.com/article/SB10001424127887323783704578247842499724794.html and the original article 'Identifying Personal Genomes by Surname Inference. Melissa Gymrek et al. Science 339:321, 2013'.

v. Requesters of patient-level clinical trial data should also have sufficient qualifications and experience for any subsequent analysis of data obtained from clinical trials, as aligned with ICH-E9 and 'statistical principles for clinical trials'. Also, in order for any analysis of data obtained from clinical trials, there should be a legitimate scientific question being proposed in order for the request for data access to be considered. Requesters should not only identify themselves, but they should also provide details of their qualifications and experience which supports they are sufficiently educated and trained to implement any subsequent analysis of the data being requested. This information should be made transparent by the requester at the time of seeking access to data.

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

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

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

ix. It is not clear how providing patients access to data relating to their own disease is aligned with the remit of access to data which is being able to independently re-analyse the benefit-risks. Anyone wishing to re-analyse data should have minimal qualifications and expertise and it should not be suggested that individuals who are not equipped with the relevant skills should attempt to re-analyse data.

x. When patients agree to participate in a clinical trial, they are doing so with the assurance that their data will be protected and appropriately used for clinical research. Another rationale for providing appropriate safeguards against access to patient level data is to ensure any requester for access to patient level data is going to respect the data that patients have agreed to be collected, and that the data remains protected if access is granted. Therfore, it is in the interest of the altruistic nature of patients participating in trials that such data will be used for further development of clinical research and healthcare and that their data would be protected;

xi. It should be recognised that clinical trial participants are providing sensitive health information while those who are accessing anonymised data would not be required to provide sensitive health information. For example they would only be required to provide their name, address and research institution. It is also difficult to understand why the name of a researcher/requester who accessed data for a particular disease would result in insurance or any other consequence. Merely accessing the data does not indicate or suggest that the individual has that disease or condition. In addition if an email address is not made public (and there is no reason to do so) there is little or no risk of spamming.

xii. There is also a risk of other unintended consequences: Some requesters may present out-of-context results that would lead to false impressions of drug safety issues and lead to unfounded health scares (e.g. http://www.biomedcentral.com/1471-2458/2/6). This risk is of high importance to the ultimate decision of whether patient level data should have open access and the long term consequences should be discussed. However, sometimes it's in fact the opposite. Some requesters use data from drug regulatory agencies to minimize unfounded health scares with potential harms in other senses: for example, the PPI-Clopidogrel interaction case: http://www.nature.com/ajg/journal/v106/n7/full/ajg2011126a.html

xiii. If a requestor uses data for an illegitimate use, is the EMA liable for failing to protect patient confidentiality? There is no secure path forward when granting control to anyone to secure patient confidentiality. Industry can do certain measures to ensure that data confidentiality is given within a dataset. But there is no measure available to secure this when a requester has access to the clinical trial data for the purpose to re-analyse it, as they would then have the potential to merge the clinical trial data with other available data. (Other available data will include those from other clinical trials. In chronic diseases (for example epilepsy) a patient may be entered into more than one trial so that the outcome measures from the first

become the baseline variables for the second even with a lapse of several years. Linkage of the trials therefore enables profiling patient histories over a long period such as a decade, and a greater risk of identification. Such linkage goes far beyond the remit of an individual trial.) The only way to secure patient confidentiality is to have a step that checks the request for access is scientific (good intent) and clear rules noting that data cannot be further disseminated. If the rules require the uploading of a protocol or analysis plan then this using a restrictive access approach increases the protection against unintended use of the data. The policy will need to clarify who is liable for any illegitimate use of data.

It is noted that verification of the identity of the requester does not necessarily imply that EMA should make public the names of requesters. In regards of publication, different positions were discussed:

aa. Although the identity of the requester indeed should be known to the database owner, it is not conclusive to request publication of these names and addresses.

bb. The name of the requestor should be public (with their consent). As mentioned above, verification of requestors is challenging. Hence this should be open to public scrutiny. This will also act as a deterrent to the mis-use of the data.

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

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

There is a need to modify the categories according to an optional user identification process, granting access to e.g. patient level after authorisation. If hierarchy for different user groups were finally considered, healthcare professionals should have access to the higher possible level of information. This would also allow for the processes discussed under topics 3, 4 and 6, setting reminders or making registered users aware of possible consequences after misuse.

Those specific trials should be identified where retroactive patient identification is a risk, and alternatives should be provided for these cases to harmonize patient and health professional rights. For example, access to data on clinical studies conducted in patients with rare diseases should be restricted and treated under different provisions, such as mandatory registration and identity verification of the

requestor, and contractual agreements covering the consequences of misuse and/or inadvertent identification.

Alternatively, open access could be granted for aggregate anonymised data and restricted access for patient level data where access is controlled by EMA.

Consider differentiating between requests for data to "independently re-analyse trial data" and requests for data to be used in "secondary analysis to address new clinical questions" and how this could determine the level of data access required. The complexity of taking patient level data and all the associated meta-data should be noted, and this complexity could lead to incorrect analyses being generated unless appropriate checks are put in place to deal with such situations.

Note whether it would be feasible for the EMA themselves to re-analyse patient-level trial data to address the "independent re-analysis" of trial data. If this approach was possible, this could lead to granting open access to aggregate anonymised data, and EMA and other nominated stakeholders considered "independent" to access to patient level data.

It is also noted that in order to allow for public access to patient-level data in the future, they would have to be a mandatory part of the clinical submission documents, and reflected in the relevant CHMP guideline documents such as CHMP/EWP/2998/03. Furthermore, the potential use of patient-level data outside of the clinical study scope should be covered in the study informed consent form such that the subject agrees to the future “secondary use” of patient-level data outside of the study scope.

3. Should requesters be required to ‘Agree’ to respect personal data protection?

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

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

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

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

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

If the identity of the requester has been verified:

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

Requesters have to be made aware of EU and local data protection regulations. Ticking a box implies a contractual relationship between the requester and the database owner/holder of the data. However, in that case both contractual parties need to be fully identifiable. A contractual but

not necessarily public "digital" agreement appears to be preferable compared to a purely anonymous process.

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

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

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

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

a) The requester should be required to sign a legally binding agreement affirming that the information and data will only be used for the agreed public health research purpose and not for any commercial use. Requests for patient level data from requesters to the EMA must be handled on a case-by-case basis, and follow consistent criteria to establish if and how the information provided will be used for valid scientific purposes and to benefit patients. (Please refer to discussion of CCI under Question 1)

b) It is unclear which situations we are talking about and "unintended commercial uses" may be used as a "killer argument". For example, if industry fears that one cannot exclude that a full CSR may be used for obtaining a marketing authorisation in a non-EU jurisdiction, this may prevent full transparency. The relationship between knowledge and profit-making is too complex to have it be contractually bound during the data release process; there is no simple distinction between using data for public health research and commercial use. The party suggesting a legally binding contract requiring the requestor to guarantee to use the data for public health purposes and not commercial purposes, should be clarified as to how commercial purposes and public health purposes will be defined and disentangled in practice.

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

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

a) It is emphasised that advising requesters of quality standards for additional secondary analyses should not and cannot impose any obligations on the requester. However, it would be appropriate to ask EMA to communicate their quality standards when a public statement is issued. (*Note: Reference is made to the work of CTAG4).*

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

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

It is pointed out that for some stakeholders, the aim of transparency goes beyond a potential reanalysis.(see arguments above, para 2.1.a.)

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

Again, it is pointed out that for some stakeholders, the aim of transparency goes beyond a potential reanalysis.(see arguments above, para 2.1.a.)

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

a) Immediately after uploading the protocol

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

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

d) Timing of publication decided by requester

Several comments/views along the following lines were expressed:

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

The process to be followed could be tailored to the remit for the request for access to data - independent re-analysis versus secondary analyses of existing data.

A protocol could be either uploaded or provided as link to a" trial register". An (ethics committee) review of the protocol should be provided by the requester.

Provision of a protocol demonstrating good research methods, fair use of data and the purpose to which it will be put seems an entirely reasonable exchange for access to data. There seems to be a danger of introducing double standards with requirement for access to clinical trial protocols and clinical trial data, but not to protocols for subsequent use. For IPD, make provision of a protocol (with delayed public access if necessary) a prerequisite for access to or release of data. A link to a formally published protocol would be acceptable. Protocols should be given a unique identifier, which is also quoted in each publication that arises from the analyses.

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

7. Should requesters be allowed to share accessed data?

It was agreed that this would become uncontrollable in case identification of the requester is not verifiable.

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

a) Should it be/have been possible to verify the identity of the requester, EMA may consider restricting data sharing. However, in such case any third party would have to be given access to the same data as the first requester directly from the EMA. If a collaboration between 2 requesters is necessary (e.g. Academia + industry or data management company), EMA should be informed and give approval. This can be anticipated in the analysis plan.

b) Requesters should not be allowed to share accessed data because that way the validity of the dataset cannot be controlled. Requestors will be responsible for the security of the data they gain access to. Without this accountability, the sharing of data could quickly become widespread; this can be avoided if requesters have restricted access to data sets in a controlled system. Requesters should need to explicitly confirm that they will not forward the downloaded original dataset to third parties. It is acknowledged that others must be able to repeat research findings; that is a basic principle of research. However, such groups would then have to identify themselves separately before accessing the same data.

c) The validity of the dataset cannot be controlled in any way; everybody can alter the original dataset once it is released by the drug agency. So the ban of sharing data is useless.

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

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

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

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

b) A staggered roll-out would be preferable as there are already many challenges to opening up access to aggregated data which need to be solved. Aligning with the roll-out of the EudraCT version 9 and access to results for many clinical trials could be an important step forward. Aggregated data, after consultation with the MAH for removal of CCI and protection of personal data, is more likely to have value to a wider audience and therefore should be of initial focus. A staggered roll-out should be done by running several pilots to evaluate potential issues.

c) A staggered approach would be pragmatic and could achieve much almost immediately. There are many issues around the release of individual patient data (IPD), particularly around open public access versus some model of conditional access. If this could be set aside for now with focus on release of aggregate data and results of all statistical analyses as set out in the trial protocol, rapid progress could be made. Access to IPD could follow

after sufficient time for discussion and enquiry. For example, potential impact of public release of IPD on participant consent needs to be investigated. Therefore, separate the issues of (1) release and access to trial information, results and aggregate data from (2) release and access to IPD, and move ahead immediately with 1. Do not delay implementation of 1 while 2 is addressed (it is much more complex and requires careful consideration). Extend the time period to allow proper consideration and investigation of issues pertaining to 2. However, the delay of the access to IPD should only be delayed for a short time - one year.

9. Should requesters be encouraged to provide feedback?

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

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

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

Several comments/views along the following lines were expressed:

- Just encouraging requesters to link their analyses back to the data accessed is not sufficient. Further discussion is needed on how any resulting publications arising from secondary analyses are linked back to data access requests. Principles should be included on minimal expectations of requesters and what should be fed back having been granted access to data. For example, should the requester have to summarise their key findings of their analyses as a minimum? Publishing has to be accepted not only in the form of articles in journals but also as other documents with open access from the internet.
- It is important that a third party who identifies a new potential safety issue liaises with the EMA and the MAH to verify the analysis and their conclusion to minimize the risk of unfounded health scares and to manage appropriate communication to patients and healthcare professionals.
- EMA should be committed to comment / answer in some way whatever new evidence brought up by requesters after its analysis.
- On the assumption that access to anonymised patient level data is granted for a defined research project, access to a secure area should be granted for a defined duration (the duration necessary to complete the project). An open-ended access (beyond the research project) would undermine the benefits of identification and declaration of research purposes. Requesters should be given a time frame within which they are obliged to publish/make public any outcomes and conclusions resulting from their analyses.
- Requestors should be required to make publications derived from this work open access either via a journal or via deposition in a publicly available repository within 12 months of the completion of the work and a copy of the work supplied to EMA.
- There should be no requirement for a time frame within which requestors are obliged to publish/make public the results of their analysis. However, if the EMA is constructing a database that will showcase the requests that have come in, also indicating which parties accessed what data, it would be nice to also include space for requestors to not only say

what outcomes have resulted from their analysis (e.g. publications) but also encourage requestors who did not publish any resulting analyses to explain the reasons for no publication.