Draft advice to the European Medicines Agency from the clinical trial advisory group on Clinical trial data formats

The clinical trial advisory group at their TC on 4th February 2012 provided advice as follows:

1. The following definitions were agreed

1.1 This advice refers to all data recorded in a clinical trial (at a patient level or derived) that can be stored electronically and associated metadata (variable definition, terminology such as code lists or dictionaries) that are part of a submission for marketing authorisation to the Agency.

1.2 In this discussion proposal, data formats refer to the organisation of information according to pre-set specifications that facilitate the storage, exchange and archive of clinical data. It includes both the type of electronic files and the content of the files, as well as associated metadata.

The principles shall apply to clinical data submitted for regulatory submission throughout the life-cycle of medicinal products.

The data and metadata concerned by this policy are stored and submitted electronically, but not necessarily sourced via electronic tools.

2. There is a need to define data formats

Choice of formats should neither imply delays in the information to be made available nor impose unnecessary burden to the stakeholders.

Formats may be different depending on the type of information to be made publicly available and the intended use of it.

As there are not universally agreed standards or formats, in order to avoid errors, a minimum set of rules should be defined, including:

- Indexed list of all trials present in the submissions shall be provided so the data of overall clinical program is tracked
- Data shall be published in the format they have been submitted and evaluated
- Data should be readable and contain metadata to allow further analyses
- Consistency with agreed terms throughout the life cycle of the medicinal products shall be maintained
- Formats at high level should be readable with electronic non-proprietary software

3. What is to be included in data formats

There is an absolute need that formats agreed contribute to ensure privacy protection. Certain information such as CT scans, MRI and other imaging, interviews shall not be included in the formats

as they carry too many identifiers. Without appropriate guarantees public disclosure of clinical data might have a negative impact on recruitment.

Three levels of clinical data and corresponding formats shall be included

- Full clinical study reports: acceptable in PDF format for all approved medicinal products.
- Datasets and results used for the evaluation linked to the relevant protocols; full statistical analysis plan, details on methods and metadata are to be made always available to allow a meaningful re-assessment.
- Individual data such as CRF in PDF format are neither useful (as they will require substantial manpower for reloading in another usable format) nor appropriate as may contain subjects identifiers breaching privacy protection. Data from the annotated CRF are to be included in the format.

More detailed discussion is needed on what additional elements shall be provided along with the datasets.

4. Formats recommended

For the clinical study reports, the full documentation shall be made available according to the ICH E3 format.

To avoid delays any format shall be acceptable for products already authorised. The data shall be published in the format they are available at present then the format could move progressively to CDISC. However, CDISC provides a frame but for the data itself, there are no agreed standards: those shall be developed gradually applying the grandfather principle.

CDISC could be a useful format for datasets, but for metadata other formats might be more useful.

Harmonisation of formats such as CDISC SDTM and ADAM is of course desirable as this expands the usefulness of the data made available. This exercise shall be progressively implemented in a collaborative way to ensure consistency and versioning control.

Sustainability of a chosen standard might also require reducing the speed of versioning and ensuring availability of softwares adapted to the subsequent changes of the formats.

Whatever the format chosen, dataset formats in the long term are to be compatible.

For the datasets there is a need to:

- Harmonise a reference format worldwide
- Maintain versioning over time

A point to discuss further concerns mixed formats acceptability, e.g. for fixed combination of old and new active substances or hybrid mixed submission, when both clinical data from old studies and from new trials are included.

5. Who should adhere to the agreed formats

The formats agreed are to be adhered to by all stakeholders and also for locally run trials outside Europe. The Applicants should ensure correct implementation of the formats and should also consider implication of terms translations from different languages.

For trials owned in different measure by different partners (e.g. public-private partnerships), the above points should be taken into account from the beginning of the clinical studies.

6. Timelines for format implementation

The CTAG2 recommended the policy to be implemented from January 2014.

- Clinical data for products already approved to be published in the format available at the time of submission.
- Data for new marketing authorisation submissions to be made available in an open file format.
- Pro-active adoption of standard formats: as this has to be mandatory for the sake of fairness and clarity for all stakeholders, it was advised starting gradually to acquire experience and then mandate formats after 2-3 years of trial period.

7. International harmonisation across regulatory agencies

EMA is leading in terms of policy but global consultation of formats is recommended.

Global alignment for both the initial agreed formats and for the updates are necessary.

Under e-CTD, PDF, XML and other standards are allowed in MAA.

ISO, CEN and CDISC to define CSRs harmonised standards.

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
7	Comment: General	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
7	1) While the pharmaceutical industry strongly supports enhanced transparency of clinical research information, this increased transparency must be balanced with the legally required protection of commercially confidential information and intellectual property (as well as of personal data), so that the innovative research and development of new medicines continues to be supported and incentivised. 2) The implications of the release of patient level data on innovation and on individual patient protection and public health through re-evaluation of data by third parties needs careful consideration to identify the best solution to balancing the desire for transparency with the need to foster innovation. Furthermore, many EFPIA member companies already respond to requests for access to their clinical trial data on a case-by-case basis. 3) EFPIA would like to emphasize the need to establish consistency throughout the 5 groups on aspects of key importance for the establishment of a policy, such as patient confidentiality and the need for anonymization of data, scope and timing of application (e.g. only application for prospective submissions). In particular, key principles raised by EFPIA in the other groups need to be applied to the release of data (e.g. need for legitimate research purposes, minimise risk to participants' privacy and confidentiality, alignment with original informed consent, consultation with the MAH).	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland

Line number	Comment and Changes proposed	Name	Affiliation
7	4) EFPIA thinks a discussion on the IT solution concerning future access of patient level data is needed in terms of data submission and access process. Closely linked to that EFPIA would like to understand the resource and budget requirements as well as the process and timing for the establishment of such a system. In addition, clarification/ discussion would be needed whether the system will already be established by the envisaged date for entry into force of the policy. If it is unlikely that a system will be established by the date of entry into force aspects for a transition phase need to be discussed. 5) Term "format" - EFPIA would like to draw your attention to the fact that the meaning of the term is used variably throughout the document and would needs clarification (e.g. datasets formats, reference formats, metadata format, open file format). 6) EFPIA considers global alignment and harmonization are critical steps in the future process. While EFPIA understands that not all points raised below have been discussed in detail at the first meeting on 04 February, those points were still included since awareness on those aspects is important at an early stage. EFPIA calls for early consideration and inclusion of these aspects in the future discussion rounds.	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
8	Comment: Is there any provision for redaction of information? **Proposed change (if any)**: Suggest this is stated up front in the document.	Helen Spain	Vectura
8	Comment: Who will address any queries raised by those reading data and what is the process? **Proposed change (if any)**: To be clarified here or in one of the other workstreams.	Helen Spain	Vectura
10	Comment: Definition of meta data neeLine 52 connects but it may benecessary to list E3 itemsds to be clearer. I am uncertain whether it applies to items in E3 our outside E3.	Tom Jefferson	Cochrane ARI Group

Line number	Comment and Changes proposed	Name	Affiliation
10	Comment: There is currently no satisfying and agreed definition for "metadata", but vague ones (e.g., "data about data"). The definition suggested here is too restrictive, as you need much more than variable definition and codelists to have a proper metadata set (see for instance define.xml 2.0). Also terminology is one thing (a standard name for a given thing), code lists are other things (a set of choices for a question), although there is some overlap. We therefore propose the following change: **Proposed change (if any)**: … and associated metadata (data properties such as dataset keys, variable definition, terminology, code lists).	Thierry Lambert	AdClin
10	Comment: Clarification needed whether the listing of metadata (variable definition, terminology such as code lists or dictionaries) is a complete list or represents examples. This should be at least specified in more details somewhere later. **Proposed change (if any)**: No further suggestion at this stage.	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
10	Comment: Metadata should include also the context of the data interpretation, the rules chosen to code data, the hypothesis and the context of the study, the link between data and CRF and analysis. **Proposed change (if any)**: associated metadata(any data useful to interpret the clinical data: variable definition, terminology such as code lists or dictionaries, the context of the study and the data, the purpose of the analysis, etc.)	Patrick Lamplé	Institut de Recherches Internationales Servier, France
11	Comment: The scope requires further consideration in relation to the legal basis [8(3), 10, 10a, 10b,10c] of the MAA and in particular in the case of multiple applications. This should be clarified in developing the policy and associated guidance. In discussions to date the electronic data are associated with the MAA but the situation where the dataset may linked to more than one MAA has not been addressed.	Vicky Jones	Takeda, UK

Line number	Comment and Changes proposed	Name	Affiliation
11	Comment: Section 1.1 referrs to data that are "part of a submission for marketing authorisation to the Agency". It is not clear to me what this means. Would this also cover data submitted after marketing authorisation, e.g. data submitted later in the life cycle of a drug? This could be data from studies submitted as part of PSURs or other updates on the evidence on a drug. I think all data submitted to the Agency should be made publicly available. **Proposed change (if any)**: … that are submitted to the Agency.	Beate Wieseler	IQWiG, Germany
11	Comment: Clarification of scope of application: EFPIA would appreciate a clarification that the new rules apply only to studies as included in submissions as of January 1, 2014 and beyond. While EFPIA agrees with EMA' s summary that, as a matter of principle, all clinical trials should be under the scope of the future policy, there is an urgent need to discuss the obligation of marketing authorisation holders when it comes to the submission/ reporting on clinical trials for which the MAH was not the sponsor, e.g. purely academic trials. As a matter of fact, in their submission MAH reference to publications but do not have the ownership on the underlying data for those studies which were performed without the MAH' s sponsorship and support. In those cases, MAH cannot be made responsible for the submission of data in the format set by the future policy. **Proposed change (if any)**: "This advice refers to all data recorded in a clinical trial…..that a part of a submission for marketing authorization to the agency as of January 1, 2014 or beyond."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
12	Comment: need to exclude pdf formats. **Proposed change (if any)**: that facilitate the storage, exchange, analysis and archive of clinical data.	Alexis Clapin	a2m2, France
13	Comment: This document refers to clinical data but it is the intention to release computer programs as well. Where is the formatting of those to be considered?	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London

Line number	Comment and Changes proposed	Name	Affiliation
14	Comment: Data format should not refer to the content of the files, and should refer to organization of the data as noted in line 12. **Proposed change (if any)**: Remove reference to content.	Vicky Jones	Takeda, UK
14	Comment: What does "metadata" mean? Combination of data from multiple trials? **Proposed change (if any)**: Define "metadata".	Vicky Jones	Takeda, UK
15	Comment: Does the principle apply to all trial data or does it just apply to trial data submitted as part of the MAA? **Proposed change (if any)**: It should individually apply to all human trials irrespective of MA status or holding by EMA.	Tom Jefferson	Cochrane ARI Group
15	Comment: Again it is not clear to me, if "data submitted for regulatory submission" somehow restricts the data to be published. This should not be the case. **Proposed change (if any)**: The principles shall apply to clinical data submitted to the Agency throughout the life-cycle of medicinal products.	Beate Wieseler	IQWiG, Germany
16	Comment: Here the date/time point for the first release of data to a third party should be defined (e.g. date of decision of the EC; date of release of the EPAR …)	Thomas BRILL	ethris, GmbH (Munich, Germany)
17	Comment: Where will the data be stored post-submission? How will the data be stored? Any control of access? These are serious concerns to sponsors who have to provide data. **Proposed change (if any)**: Address these issues. If they have not been discussed, state so in the notes.	Vicky Jones	Takeda, UK

Line number	Comment and Changes proposed	Name	Affiliation
17	Comment: The meaning of "sourced via electronic tools" is unclear. The requirement must be that the data itself is machine readable, but that requirement may not exist for the metadata. **Proposed change (if any)**: The data and metadata concerned by this policy are stored and submitted electronically, but data must be machine readable but metadata may not need to be machine readable.	Vicky Jones	Takeda, UK
17	Comment: Even if not sourced via electronic tools, the data format must guarantee to link these documents to the data. **Proposed change (if any)**: submitted electronically and guarantee the coherence of the data and documents even if not sourced via electronic tools.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
18	Comment: It is unclear what is meant by the sentence "but not 17 necessarily sourced via electronic tools." Further detail or alternative wording should be added to clarify.	Catrin Tudur Smith	University of Liverpool, UK
18	Comment: Clarification needed on what is meant by "but not necessarily sourced via electronic tools". **Proposed change (if any)**: No further suggestion at this stage, need to understand.	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
20	Comment: We believe a statement like this, plus the time frame given (Jan 2014) binds you to existing standards: for data, it means whatever CDISC (prompted by the FDA) decides.	Thierry Lambert	AdClin
22	Comment: There is an implication here that different formats may be requested for different purposes or customers. We strongly recommend to keep to the grandfather principle and not convert legacy data. **Proposed change (if any)**: "Formats as used by the company for the analysis may be different from study to study. Data should be made available in this format irrespective of the type of information to made publicly available and the intended use of it. ("grandfather principle")."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland

Line number	Comment and Changes proposed	Name	Affiliation
26	Comment: Usually, companies are requested to provide the information already today as part of a submission. **Proposed change (if any)**: "Indexed list of all trials present in the submissions shall be provided so the data of overall clinical program is tracked as long as not available in the table of context of the submission."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
27	Comment: The list of trials should allow to track the studies also in other systems that present data on a trial, e.g. in clinicaltrials.gov. Therefore, the list should include an unique trial identifier. **Proposed change (if any)**: …clinical program is tracked and studies are identified by a unique study identifier.	Beate Wieseler	IQWiG, Germany
28	Comment: Clarification is needed that data need to be anonymized. **Proposed change (if any)**: "Anonymized data shall be published in the format they have been submitted and evaluated."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
28	Comment: In general, the dispossession of sponsors, who are the owner of the patient data, is questionable. An obligation to publish study data at a patient level is unacceptable without any further access protection mechanism like user identification and authorisation tools. A redaction system for anonymising patient data to protect personal rights is crucial.	Dr. Andreas Franken	AESGP, Germany
28	Comment: EMA may have received some approval file with all data in pdf files but the clinical studies have probably been analysed with adequat electronic files available by the firm. If the firm used a proprietary software, it should change the format to a format for non proprietary software (line 32 should apply). **Proposed change (if any)**: Data shall be published in the format they have been submitted or evaluated by the Marketing-authorisation holder.	Alexis Clapin	a2m2, France
29	Comment: Data should be readable - it is not clear by whom. Readability does not guarantee availabilty for analysis. **Proposed change (if any)**: Data should be presented as a structured	Vicky Jones	Takeda, UK

Line number	Comment and Changes proposed	Name	Affiliation
	database.		
29	Comment: This may be unnecessary because "readable" is addressed by line 32 and the meaning of "metadata" is unclear. Why containing metadata enables further analysis. **Proposed change (if any)**: Delete it or further clarify what the authors try to say.	Vicky Jones	Takeda, UK
29	Comment: Clarification needed through introduction of examples. Clarification is needed that data need to be anonymized. **Proposed change (if any)**: "Anonymized data should be readable and contain metadata to allow further analyses (e.g. SAS dataset format)."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
29	Comment: I advice to add the notion that the metadata will provide the context to interpret correctly the data. **Proposed change (if any)**: contain metadata that provide the context to interpret correctly the data and allow further analyses.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
29	Comment: readable could apply to pdf. Could you propose another term saying that the data could be analysed with a non proprietary software such as openoffice.org spreadsheets (free Excel). **Proposed change (if any)**: Data should be readable with a spreadsheet software such as openoffice.org one and contain metadata to allow further analyses .	Alexis Clapin	a2m2, France
30	Comment: Clarification needed what is meant by “agreed terms are”; Is this the agreement on the data format that was originally agreed? Does this mean all studies for one product should be in the same format (which may be difficult for long lasting projects)? What is meant by “consistency” in this context? **Proposed change (if any)**: No further suggestion at this stage.	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland

Line number	Comment and Changes proposed	Name	Affiliation
32	Comment: Formats at high level should be readable -not clear - data should be readable? What is meant by high level? **Proposed change (if any)**: Remove or combine with 29 (including the comments for 29).	Vicky Jones	Takeda, UK
32	Comment: Clarification needed through introduction of examples. **Proposed change (if any)**: "Formats at high level should be readable with electronic non-proprietary software. (e.g. reading SAS format)."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
32	Comment: The requirement to make the data readable with non-proprietary software contradicts the regulatory requirement, that for regulatorily relevant clinical studies only statistical software must be used which is validated and accepted.	Dr. Andreas Franken	AESGP, Germany
33	Comment: We should define a minimum standard which is realisable even in a small academic institution or a SME (e.g. scans of examination forms as PDF)	Thomas BRILL	ethris, GmbH (Munich, Germany)
35	Comment: As discussed at the meeting at EMA in November 2012, requesting absolute privacy protection might make publication of any data impossible. Standards set by the European data protection officer should be considered sufficient.	Beate Wieseler	IQWiG, Germany
35	Comment: Data Privacy should explicitly be mentioned to apply to genetic data. In addition, respecting data privacy goes beyond CT Scans, MRI or other imaging. There is a need for a reference to the EU data protection Directive 95/46/EC and to anonymization, in particular through de-identification, removal of free text, date of birth anonymization, obfuscation of subject study dates. **Proposed change (if any)**: “There is an absolute need that formats contribute to ensure data privacy protection through anonymization (reference to Directive 95/46/EC). Obviously, certain information such as CT scans, MRI and other imaging, interviews and genetic data shall not be included in the formats as they carry too many identifiers.”	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland

Line number	Comment and Changes proposed	Name	Affiliation
35	Comment: For imaging, interviews...: their analysis leads to a full set of data written in a specific CRF (volume, number, position of lesions, characteristics...). All rules that apply to the "clinical/biological" CRF should apply to the imaging,interview CRF.	Alexis Clapin	a2m2, France
36	Comment: CT scans, MRI, interviews etc. should not be excluded per se - this will jeopardize the whole CT Data Transparency idea. Many CT scans only show a small body region and if the name of the patient is replaced by the study ID than there will be no possibility to identify him the person. **Proposed change (if any)**: … and other imaging, interviews should be carefully checked so that they contanin no data which might be used to identify the patient.	Thomas BRILL	ethris, GmbH (Munich, Germany)
36	Comment: Patients' genetic/genomic data should not be included either. **Proposed change (if any)**: Add patients's genetic/genomic data to the exclusion list.	Vicky Jones	Takeda, UK
36	Comment: Whilst I agree that CT Scans, MRI and other information may compromise the data privacy,, I believe the results from these should be included and would not compromise the data privacy.	Niraj Ruparelia	Sascon Ltd
39	Comment: Items should listed, Line 52 connects but it may benecessary to list E3 items. **Proposed change (if any)**: Full CSR including protocol, amendments, dated SAP, CRFs, individual level data, certificates of analysis, list of IRB and IC, supplementary tables, informed consent forms, list of investigators, list of contributors to CSR.	Tom Jefferson	Cochrane ARI Group
39	Comment: Typo. **Proposed change (if any)**: Three levels of clinical data IN corresponding formats.	Vicky Jones	Takeda, UK

Line number	Comment and Changes proposed	Name	Affiliation
39	Comment: Three levels of study information, data and corresponding formats shall be included.Level 1: full list of trials of a given drug including a unique study identifier for each study; these lists should be fully searchable; the lists could be connected to the EPARs. Level 2: for each study: full clinical study report (CSR) according to ICH E3 including all appendices (this format according to ICH E3 among other things includes a full protocol with all amendments, a full statistical analysis plan and full summary tables and test outputs). A report according to ICH E3 also includes patient data listings. Measures needed to protect privacy to be discussed. Level 3: for each study: data sets (including individual patient data) and results used for the evaluation of the drug (including meta-data required to use the data set, like an annotated CRF, variable definitions, derived values etc.); including any test outputs. Proposed change (if any): Three levels of study information, data and corresponding formats shall be included. Level 1: full list of trials of a given drug including a unique study identifier for each study; these lists should be fully searchable; the lists could be connected to the EPARs Level 2: for each study: full clinical study report (CSR) according to ICH E3 including all appendices (this format according to ICH E3 among other things includes a full protocol with all amendments, a full statistical analysis plan and full summary tables and test outputs). A report according to ICH E3 also includes patient data listings. Measures needed to protect privacy to be discussed. Level 3: for each study: data sets (including individual patient data) and results used for the evaluation of the drug (including meta-data required to use the data set, like an annotated CRF, variable definitions, derived values etc.); including any test outputs.	Beate Wieseler	IQWiG, Germany

Line number	Comment and Changes proposed	Name	Affiliation
39	**Proposed change (if any)**: Three levels of study information, data and corresponding formats shall be included. Level 1: full list of trials of a given drug including a unique study identifier for each study; these lists should be fully searchable; the lists could be connected to the EPARs. Level 2: for each study: full clinical study report (CSR) according to ICH E3 including all appendices (this format according to ICH E3 among other things includes a full protocol with all amendments, a full statistical analysis plan and full summary tables and test outputs). A report according to ICH E3 also includes patient data listings. Measures needed to protect privacy to be discussed. Level 3: for each study: data sets (including individual patient data) and results used for the evaluation of the drug (including meta-data required to use the data set, like an annotated CRF, variable definitions, derived values etc.); including any test outputs.	Beate Wieseler	IQWiG, Germany
40	Comment: Need to also make full study protocols with all amendments dated available with CSRs, otherwise the reporting of appropriate outcome measures, statistical analysis plans, populations to be evaluated etc in CSRs cannot be assured. **Proposed change (if any)**: Full clinical study reports plus complete study protocols (including all amendments)…	John Abramson MD MS	Harvard Medical School
40	Comment: Here and at the beginning of line 41 and 44 there should be written "Level I data:" (41: "Level II data"; 44 "Level III data", resp.) - that will make discussion about these levels more precise.	Thomas BRILL	ethris, GmbH (Munich, Germany)
40	Comment: Why to narrow that on "all approved medicinal products" - this should also fit (if applicable) for other issues (e.g. MP under investigation; MP having not yet an approval).	Thomas BRILL	ethris, GmbH (Munich, Germany)

Line number	Comment and Changes proposed	Name	Affiliation
40	Comment: Full clinical study reports must not include patient level data unless anonymized. **Proposed change (if any)**: "Full clinical study reports (excluding individual patients level data)."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
40	Comment: See comment for line 28	Dr. Andreas Franken	AESGP, Germany
41	Comment: Datasets and ..... requires clarification. **Proposed change (if any)**: Patient level datasets and ……	Vicky Jones	Takeda, UK
41	Comment: Clarification: Results are already submitted with the study reports. Individual patient datasets need to be anonymized. **Proposed change (if any)**: "Anonymized individual patient data sets used for the evaluation linked to the relevant protocols; …."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland

Line number	Comment and Changes proposed	Name	Affiliation
41	Comment: full data should include all the data obtained on the CRF even if the data have not been analysed for the study report (as far as confidentiality is not engaged). For any specific subgroup of patients defined for analysis or as a consequence of an analysis (patients considered for per protocol analysis, responders vs non responders, patient with a specific characteristic or outcome) this should be clearly indicated in the dataset on a patient basis. For example a column should indicate if yes or no the patient should be considered in the group. **Proposed change (if any)**: Datasets and results used for the evaluation linked to the relevant protocols; full data included in the CFR (excep confidential information), full statistical analysis plan, details on methods and metadata are to be made always available to allow a meaningful re-assessment. For any subgroup of patients defined in the full clinical study report, dataset should include information on a patient basis on wether or not the patient is belonging to the group.	Alexis Clapin	a2m2, France

Line number	Comment and Changes proposed	Name	Affiliation
44	Comment: SAS files should be made available for IPD.	Tom Jefferson	Cochrane ARI Group
44	Comment: I disagree with the first sentence! PDF scans of printed out CRFs are by far ideal for reassessment of data but might be the minimal standard which is realisable even in a small academic institution or a SME. We should not forget that data transparency is the second step after generating the data in the hospital (first step) - and we should not set up unnecessary burden for financially weak "small sausage"holders (which cannot spend the money for a steak).	Thomas BRILL	ethris, GmbH (Munich, Germany)

Line number	Comment and Changes proposed	Name	Affiliation
44	Comment: Some old files probably contain datasets in pdf format. If yes and if the requester wishes, EMA should ask the marketing authorization holder to provide dataset in a format that can be used in spreedsheats (see line 28 comment). **Proposed change (if any)**: Individual data such as CRF in PDF format are neither useful (as they will require substantial manpower for reloading in another usable format) nor appropriate as may contain subjects identifiers breaching privacy protection. Data from the annotated CRF are to be included in the format. If the MA file contains data in pdf format, EMA should ask the MAH to provide the data in an adequate format readable in a non-proprietary spreedsheat format.	Alexis Clapin	a2m2, France
46	Comment: I do not catch the sense behind the second sentence "Data from annotated … in the format". Which annotations should be made onto a CRF? CRFs are "holy" raw data.	Thomas BRILL	ethris, GmbH (Munich, Germany)
46	Comment: Clarification: annotated CRFs never contain patient data. **Proposed change (if any)**: delete: "data from"; change into: "The clinical trial data should be accompanied by an annotated CRF."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
46	Comment: I strongly agree that the annotated CRF should be submitted. If CDISC SDTM Data is to be used as a pre-requisite or guide then the SDTM annotated CRF would also be very useful for the reviewer. SDTMs and ADaMs should both be submitted. Question is would the Raw CRF data be useful to EMA or is SDTM sufficient?	Niraj Ruparelia	Sascon Ltd

Line number	Comment and Changes proposed	Name	Affiliation
48	Comment: It would be good if such "additional elements" could also be harmonized, especially with FDA and eSUB requirements. **Proposed change (if any)**: "More detailed discussion is needed on what additional elements shall be provided along with the data. Harmonization with other agencies should not only be achieved with regard to data structures and formats but also with respect to such additional requirements."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
48	Comment: A general comment about unstructured data. Unstructured data have to be manage in order to enhanced their usability. The formats should include the links between structured data and unstructured data. This will help at every step, from analysis to review. **Proposed change (if any)**: All three levels of clinical data should be tighly linked in order to guarantee their readability and their usefulness.	Patrick Lamplé	Institut de Recherches Internationales Servier, France

Line number	Comment and Changes proposed	Name	Affiliation
49	Comment: There may be circumstances that would justify a different assessment of the confidentiality of the clinical trial data (this is being discussed in CTAG5 about legal aspects) and in such cases the level of clinical data and corresponding formats may need to be adjusted accordingly.	Vicky Jones	Takeda, UK
51	Comment: This comment concerns the format for the files to be shared. I believe that the rules of Good Clinical Practice (GCP) should be endorsed even for the (re)analysis of the shared data. With regard to this, I would like to draw the attention to the reflection paper (draft) published in 2007 by the inspector's working group (Doc. Ref. EMEA/505620/2007), although I am aware that this concerns the format of electronic source documents and is thus one step earlier that what is currently being discussed in the CTdataGroup2. However, the format chosen for the files should allow that any data analysis conducted afterwards can be followed in a very transparent way, i.e. an "audit trail" of some sort should be contained e.g. in the metadata. It should be ensured that transparency does not stop where EMA hands over data to the public but it needs to be ensured that the transparency continues.	Andrea Wohlsen	Federal Institute for Drugs and Medical Devices, BfArM, Germany
52	Comment: CSRs must be identical to the original document, signed and dated by the sponsor. **Proposed change (if any)**: For the clinical study reports, the full documentation in its original version signed and dated by the sponsor….	John Abramson MD MS	Harvard Medical School
52	Comment: Old clinical study reports may not fully comply with the current ICH E3 format. In these cases it should be acceptable to provide the clinical study report in the original format in which it was written. **Proposed change (if any)**: For the clinical study reports, the full documentation shall be made available according to the ICH E3 format, or the original format in which the report was written.	Vicky Jones	Takeda, UK

Line number	Comment and Changes proposed	Name	Affiliation
52	Comment: Clincial study reports should include appendices; possible measures with regard to privacy protection with regards to patient data listings to be discussed. **Proposed change (if any)**: For the clinical study reports, the full documentation shall be made available according to the ICH E3 guideline, including appendices.	Beate Wieseler	IQWiG, Germany
54	Comment: As stated in our discussion it might be very important to have access to the data for already authorised MPs - so I want to propose a timeframe within even for authorised MPs the data should be send to the agency - this could be five years after the final Data Transparency Rules have been implemented. At least this should be a duty for every MA holder who further wants to sell (and make money with) an authorised drug. We should not exclude the vast majority of already existing data (that are the data from authorised MPs) from transparency.	Thomas BRILL	ethris, GmbH (Munich, Germany)
54	Comment: In the meeting on 4th Feb 2013 EMA confirmed that the policy would be applied prospectively for new medicinal products after the implementation date therefore products already authorised should be out of scope. The wording should allow for release of old clinical data included in new MAAs. **Proposed change (if any)**: To avoid delays any format should be acceptable for active substances contained in authorised medicinal products.	Vicky Jones	Takeda, UK
54	Comment: Clarification. **Proposed change (if any)**: "To avoid delays any formats should be acceptable for those studies which have already been started at the point of entry into force of the new policy. The data shall be published …"	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
55	Comment: Datasets may already be available in CDISC formats. **Proposed change (if any)**: …….published in the format they are available at present, including CDISC, then the format could move progressively to CDISC as recommended.	Vicky Jones	Takeda, UK

Line number	Comment and Changes proposed	Name	Affiliation
56	Comment: CDISC does provide standards for the data to be collected. **Proposed change (if any)**: Remove the sentence 'However, CDISC provides a frame but for the data itself…….'.	Vicky Jones	Takeda, UK
57	Comment: Define grandfather principle, as it is unclear.	Tom Jefferson	Cochrane ARI Group
57	Comment: Could we clarify the "grandfather principle" - we don´t know that principle in Germany.	Thomas BRILL	ethris, GmbH (Munich, Germany)
57	Comment: If retained (see 56) the 'grandfather principle' should be clarified as the meaning of the sentence is unclear.	Vicky Jones	Takeda, UK
57	Comment: Can you be more explicit on the grandfather principle?	Patrick Lamplé	Institut de Recherches Internationales Servier, France
58	Comment: CDISC is not a format, but an organization. Please clarify what you are talking about here: ODM? SDTM? ADaM?	Thierry Lambert	AdClin
58	Comment: It seems odd to talk about "other formats" for metadata. CDISC has developed not only good and widely used standards for data, but also good and compatible standards for metadata, in the form of define.xml. As CDISC standards are already widely used in clinical research, it would be highly desirable to make use of them to the fullest extent possible and not to reinvent the wheel. **Proposed change (if any)**: Replace line with "CDISC have defined useful formats for both data, in the form of SDTM and ADaM standards, and metadata, in the form of define.xml. Use of these standards is strongly encouraged."	Adam Jacobs	Dianthus Medical Limited
58	Comment: CDISC standards for metadata are well defined and work well with CDISC formatted data. **Proposed change (if any)**: CDISC could be a useful format for datasets, but for metadata other formats may also be considered.	Vicky Jones	Takeda, UK
58	Comment: EMA should consider minimal requirements for metadata. **Proposed change (if any)**: "CDISC could be a useful format for datasets. EMA should define minimal requirements and standards for metadata."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland

Line number	Comment and Changes proposed	Name	Affiliation
58	Comment: Whilst CDISC formats provide a good guide to data formats, there remains much ambiguity over the Implementation Guides of CDISC with many Pharmaceuticals adopting their own interpretation. This would be a good opportunity to resolve this ambiguity and create clear and concise definitions.	Niraj Ruparelia	Sascon Ltd
58	Comment: I suggest to add a sentence to indicate the direction taken: multiple standards. Therefore, there is a need of a standard to link the differents standards. CDISC and HL7 propose BRIDG. **Proposed change (if any)**: Therefore, it seems that there is a need of a set of standards, for each type of data or exploitation.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
61	Comment: An explanation is needed what is meant with “in a collaborative way” and who would be included. **Proposed change (if any)**: No further suggestion at this stage.	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
62	Comment: This seems to mean "the CDISC is evolving SDTM way too fast, so the EMA should not follow this pace". If it is what is meant, please say it explicitely.	Thierry Lambert	AdClin
64	Comment: Please define "compatible": with what? This word has no meaning alone.	Thierry Lambert	AdClin
64	Comment: Clarification needed on what compatibility means in this context. **Proposed change (if any)**: “Whatever the format chosen, dataset formats in the long term are to be compatible in the standard format being used (e.g. like CDISC).”	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
68	Comment: EFPIA agrees to the need to accept long term studies with different formats attached when studies were finished at completely different time points. **Proposed change (if any)**: None at this stage.	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
68	Comment: Old non-formatted data in older studies may be an issue and could prove difficult to re-format to a newer version if required.	Niraj Ruparelia	Sascon Ltd

Line number	Comment and Changes proposed	Name	Affiliation
69	Comment: Old studies may include some conducted by agencies such as MRC, HTA, Wellcome, academic departments of medicine,and hospitals, going back 40 years. Data from these may no longer be around.	Anthony Johnson	UK Medical Research Council Clinical Trials Unit, London
73	Comment: Local studies outside Europe (e.g. local registration studies in Korea, China, Russia, Ethopia etc.), will only be expected to adhere to the agreed data formats, if they are part of a submission to the EMA. In any case, international harmonisation of data formats is needed before submission of trials in the future EU data formats can be required. **Proposed change (if any)**: "The formats agreed are to be adhered to by all stakeholders and also for trials run outside Europe if they become part of a submission to EMA."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
76	Comment: I would suggest to invite a representative of the European Organization for Research and Treatment of Cancer (EORTC), an academic organization of oncologic trialists which has implemented CDISC, to particpate in the future meetings of CTAG 2.	Christian Dittrich	ESMO
76	Comment: The scope of phase 4 clinical trials should be clarified as the MAH does not always have access to these data. We assume this applies only to studies conducted by the MAH which would be submitted to the MAA. **Proposed change (if any)**: Clarify intent of sentence and scope.	Vicky Jones	Takeda, UK
76	Comment: The situation regarding observational studies conducted by third parties requires further consideration and discussion. There are strict rules in place regarding industry use of third-party data. In these cases the Marketing Authorisation Holder is not permitted to share the data. Data from observational studies should be exempt from disclosure.	Vicky Jones	Takeda, UK

Line number	Comment and Changes proposed	Name	Affiliation
77	Comment: Additional comment on partnership programs. **Proposed change (if any)**: "This concerns a potential inconsistency between data submitted to the Agency for which the MAH takes accountability. In addition, it concerns publications which could fall under the remit of a public-private partner and which could use for example a different data cut. A clarification is needed on how agreements of public-private partners on secondary publications can be maintained when studies are being made available at the time of approval in agreement with the MAH."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
79	Comment: The recommendation to implement the policy from January 2014 includes publication of data already available at the Agency before January 2014 to be pro-actively published starting January 2014. Clinical study reports of all approved drugs available at the agency from submissions before January 2014 should be published by the Agency. These CSRs are required to assess drugs in current use beyond the assessment provided by the regulatory agencies for marketing authorisation. Examples for additional assessments which could be informed by these CSRs are questions of reimbursement or indirect comparisons required for comparative effectiveness research. **Proposed change (if any)**: The CTAG2 recommended the policy to be implemented from January 2014. It is furthermore recommended to pro-actively publish also those CSRs which are available at the Agency from submissions before January 2014. Publication of CSRs submitted before January 2014 should also start in January 2014.	Beate Wieseler	IQWiG, Germany
79	Comment: Clarification on applicability of the new policy. **Proposed change (if any)**: "The CTAG2 recommended the policy to be implemented in all submissions from January 2014 onwards."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
79	Comment: This part supposed that we act on this dates. During TC I understood it was a proposed estimation. **Proposed change (if any)**: Replace recommended by evaluate or suggest.	Patrick Lamplé	Institut de Recherches Internationales Servier, France

Line number	Comment and Changes proposed	Name	Affiliation
80	Comment: Should clarify that all CSRs previously submitted to EMA for approved drugs will be made available. **Proposed change (if any)**: Clinical data regarding all trials submitted to EMA for products already approved…	John Abramson MD MS	Harvard Medical School
80	Comment: See comment on "authorised/approved MPs" in line 54	Thomas BRILL	ethris, GmbH (Munich, Germany)
80	Comment: Clinical data for products already approved can be published in the format available at the time of submission. Clinical study reports of all approved drugs available at the agency from submissions before January 2014 should be published by the Agency. These CSRs are required to assess drugs in current use beyond the assessment provided by the regulatory agencies for marketing authorisation. Examples for additional assessments are questions of reimbursement or indirect comparisons required for comparative effectiveness research.	Beate Wieseler	IQWiG, Germany
80	Comment: Clarification of scope and timing for the future policy. **Proposed change (if any)**: "Anonymized clinical data from studies already started before implementation of this policy (and especially for those already analysed) should be published in the format available at the time of submission."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
82	Comment: Please clarify again what is meant with “open file format”, by adding an example like “e.g. SAS transport files”. **Proposed change (if any)**: No proposal at this stage.	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland

Line number	Comment and Changes proposed	Name	Affiliation
82	Comment: Sponsors and institutions for statistical anlyses are working with their own statistical software requiring sometimes standarized data formats, sometimes software specific data formats. If there is a mandatory data format, the software cannot work with, they will have to buy new software or extra migration tools. To guarantee a future readability of older files and formats, especially the huge pile of data collected before this discussion, sponsors and institutions have to keep the old software in parallel. Reformatting existing data into a new format is costly, complex and a source for errors. The group should keep in mind, that this also applies to academic researcher and it seems doubtful, they can afford or pay for it. Neverthless, if this group intends to discuss an approach to a common usable study data file format, it should observe a very similar approach of the US FDA. http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm248635.htm.	Dr. Andreas Franken	AESGP, Germany
82	**Proposed change (if any)**: The sponsor should not be forced to migrate his own data collecting tools and statiscal software environment to a new format as long as the electronic file used is a commonly used format or can be migrated by the third party user into such one. I suggest to use a varying statement at this timepoint according to the paragraf, I copied from the current FDA draft guidance on eStudy data format: A file format standard specifies a particular way that information is encoded in a computer file. Specifications for a format permit the file to be written according to a standard, opened for use or alteration, and written back to a storage medium for later access. Some file formats in widespread use are proprietary, others are open source. Examples of file format standards supported at FDA include Adobe Acrobat Portable Document (.pdf), SAS Transport File format (.xpt), text files (.txt), and Extensible Markup Language *(.xml). The use of a file format standard for study data exchange supports technical interoperability, but by itself is often insufficient for semantic interoperability.*	Dr. Andreas Franken	AESGP, Germany

Line number	Comment and Changes proposed	Name	Affiliation
83	Comment: While it seems reasonable to gain experience with formats of data sets and individual patient data, there is no need to have a test period for clinical study reports, because the format of the CSRs, i.e. ICH E3 is in effect since 1996. Therefore, the format for CSRs can be mandatory starting January 2014. **Proposed change (if any)**: … after 2-3 years of trial period. Since the format of clinical study reports is established since 1996, the CSR format (ICH E3) becomes mandatory in January 2014.	Beate Wieseler	IQWiG, Germany
85	Comment: More clarification needed what is meant with "…2-3 year of trial period". **Proposed change (if any)**: "…, it was advised starting gradually to acquire experience and then mandate formats after 2-3 years for all new studies."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland
86	Comment: Agree that harmonisation is required but also should implement what will be widely used in future to further standardise the process and prevent any re-formatting	Niraj Ruparelia	Sascon Ltd
87	Comment: I would like to propose a "probationary time period" during which the consequences of the strategy agreed upon could be re-assessed and in case some not expected consequences will turn out to be inadequate, adaptations could be performed before such complex new rules will be implemented.	Christian Dittrich	ESMO
87	Comment: The level for global consultation should be ICH (VICH resp.) **Proposed change (if any)**: … but global consultation of formats at the ICH (VICH resp.) level is recommended.	Thomas BRILL	ethris, GmbH (Munich, Germany)
87	Comment: Given the requirements from FDA on eSUBs, a recommendation on global consultation of formats is insufficient. Multiple formats will result in duplicative work and unsustainable burden on industry. **Proposed change (if any)**: "EMA is leading in terms of policy but global consultation of formats is essential to ensure that only one format is required to be produced by industry for regulators worldwide."	Sabine Atzor, Hans-Ulrich Burger	EFPIA, Switzerland

Line number	Comment and Changes proposed	Name	Affiliation
87	Comment: The international harmonisation is a critical point, as formating data is costly and, moreover, maintaining integrity of data in several formats is not a good practice that can lead to errors. EMA with FDA, SFDA and Japan PMDA, and any other public agency, should harmonise their recommandations on data formats and metadata requirements.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
N/A	Comment: Please specify when information where added out of the TC.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
N/A	Comment: Please add that we ask for inputs from other groups, and also to interact with them.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
N/A	Comment: Please clarify if a master data project is linked to this format group. Master data could be on, at least, the references of standards, the studies, the submitters, etc.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
N/A	Comment: Please clarify the impact of this policy with the guidance for EudraCT results-related submission, already planned for Q4 2013.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
N/A	Comment: Please clarify when this data must be submitted? It should be after the evaluation by EMA.	Patrick Lamplé	Institut de Recherches Internationales Servier, France
N/A	Comment: Last January 31st we published an article in the BMJ open on the increased atypical fracture risk in women taking bisphosphonates. We also published the full dataset of individual patients data. Please see the link below. If you scroll down you may read "Download: atypicalfractures database.xls " and if you click on it you get access to the database, http://datadryad.org/resource/doi:10.5061/dryad.h435m. If data are anonymized I can see no technical nor legal nor ethical reason to deny access to individual patient data of any trial.	Juan Erviti	ISDB secretary