EMA/240810/2013

Submission of comments on 'Policy 0070 on publication and access to clinical-trial data'

EMA/293958/2005

Comments from:

Name and affiliation
European Federation of Statisticians in the pharmaceutical Industry (EFSPI)

Please note that these comments and the identity of the sender (not contact details) will be published unless a specific justified objection is received.

When completed, this form should be sent in Word format (not PDF) to: ctdatapolicy@ema.europa.eu

7 Westferry Circus • Canary Wharf • London E14 4HB • United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7418 8416
E-mail info@ema.europa.eu **Website** www.ema.europa.eu

Comments on text

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
General	EFSPI supports responsible data access. EFSPI believes access to clinical trial data should be implemented in a way which supports good research, avoids misuse of such data, lies within the scope of the original informed consent and fully protects patient confidentiality.	
General	The majority of EFSPI's comments relate to the provision of 'C' type data. EFPSI recognises the EMA's commitment to put in place appropriate standards, rules and procedures for de-identification of these data and to work with concerned parties towards this goal. EFSPI is committed to contributing to this work. EFSPI objective in providing these comments is to help ensure that the ultimate provision of 'C' type data leads to the best possible science within the constraints posed by the need to protect confidential information.	
General	EFSPI believes that there is a need for the policy to have a clear process mapped out with governance from submission of a research proposal up to and including publication of the additional post hoc analyses, including the consequences for not complying.	Include a process map from start to end, and describe all the steps in the process clearly.
General	A controlled system where the requestor can analyse the raw data but download only summary results is preferable with respect to patient confidentiality and the enforcement of any requirements for pre-specification of analysis plans. The controlled access system should allow the ability to combine data from multiple companies, e.g. to conduct patient level meta-analyses.	Include references to setting up a controlled system to manage access requests.
033-035	Allowing researchers to re-analyse and replicate primary analyses seems misaligned with the EMA current practice of not receiving the CT data themselves to re-analyse it before they make their decision to grant regulatory approval. Will the EMA analyse the CT data themselves?	Clarify if the EMA will begin to analyse CT data themselves as part of assessing a regulatory submission.
038-39	There is a reference to 'established ways and means to anonymise data and protect patients from retroactive identification.' References or details should be provided for these processes.	Add references of protecting patients from retroactive identification
047-48	The patient consent process for analyses outside the initial scope of the trial needs to be clarified. An Informed Consent template or, at a minimum, a list of minimum or essential elements that should be included in an informed consent should be specified in the policy.	Clarify the minimum elements, or provide an example of, an informed consent template that would be sufficient to prevent informed consent

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
		issues to grant access to data in line with the policy.
057	The use of the term "secondary" analyses is unfortunate and in relation to respected ICH guidelines, such as ICH E9, not appropriate. In E9 there is mentioning of secondary variables and secondary parameters, but this is then still in the context of pre-specified variables and parameters. This is, however, clearly not the setting of the analyses at stake in this draft policy, which are all post hoc, after the trial results have already been presented and in statistical language the "alpha" has been spent.	It is better to reference these analyses as "post hoc additional" or "replicate" analyses depending on their objectives.
059-61	EFSPI endorses EMA's plan to put measures in place to protect against claims resulting from inappropriate analyses. In our view, these measures should include: • Scientific rationale • Pre-specified statistical analysis plan • Qualified personnel • Independent review of the research proposal • Communication between the researcher and owner of the data • A governance process, including arbitration, in case of replicating analyses show results relevantly different form the original analyses	The measures to protect against claims resulting from inappropriate analyses should be stated. These measures should be mandatory and not optional as in the current draft policy.
070-72	It is stated that those conducting secondary analyses should be given a reasonable time to conduct their analyses without anyone being informed. We would feel it to be fair that the market authorization holder is informed about the identity of the requestor and the aims of the analysis, at the time when access to "C" data is granted. This would give the opportunity for researchers to communicate with the data owners on the proposed analyses. It also enables other researchers yet to request access to the data visibility to the proposed analyses, thus avoiding unnecessary duplication by other researchers to conduct similar proposed analyses. Is there a limit to the number of requesters who wish to re-analyse and replicate the primary analyses? Is one sufficient? If not how does more than one request support knowledge in the	Requests for data to be published when access to "C" data is granted. Clarify whether there is a limit to the number of requesters proposing to conduct a re-analysis of the primary analyses.

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
	interest of public health?	
083	There are major drawbacks with the policy only covering submission data, which is only a subset of trials that are available.	Note the limitation of the policy not covering all trials that could be available.
092	It is not true that "raw data" is customarily submitted to the EMA.	Delete this statement.
096-097	There will be situations where the sponsor will not have access to observational research data supporting a regulatory filing as they did not have direct access to the data but instead through a third party.	Clarify what is in scope for observational research methodologies.
120-121	CDISC differentiates between so-called SDTM and ADAM data sets, the first basically referring to "raw data", the latter to the "derived analysis data" underlying the statistical analysis and data presentation.	It would be helpful in case the policy more clearly indicates what is meant here in terms of required data and associated formats.
120-121	Annotated CRFs, variable definition, data specifications etc would better fall under another heading (and potential another process in terms of disclosure) than raw CT data – these are meta-data and don't have the confidentiality issues of the actual data.	Separate descriptions of meta-data from descriptions of actual data.
121	It is not clear what is meant by "test outputs". We would think of test output as being output that is created by a program prior to the program being peer-reviewed, validated and put in 'production' (its final read-only location). We see no purpose in storing test outputs or providing them to anyone. Perhaps "test output" means something different to the guidance authors?	We suggest removing this or define what is meant by test output, as it is not clear how it relates to raw data.
121-123	Statistical analysis software logs, test output of programs and SAS programs are mentioned as "raw CT data" here in the definition, but not later in the draft policy. These documents are not generally part of the CTD and CSR in Annex I and II. In addition, many data owners consider their SAS macros intellectual property. Requesting one-off SAS programs instead would be expensive.	We propose that logs, test outputs and programs are not made public unless they are contained in the CTD and CSR.
143	The term "adequately de-identified" should be defined. These definitions must be endorsed by European Data Protection authorities before the policy can be implemented.	Define and reference "adequately de-identified".
149	It is not clear whether the personal names of people involved in the conduct of the study need to stay in the report to be made available for "open access", or whether they can be deleted should the	Allow the sponsor to remove personal

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
	sponsor choose to do so (as part of the redaction of the study report that anyhow needs to take place).	names of people involved in the conduct of the study as part of the redaction process.
165-175	Guaranteeing confidentiality appears incompatible with making data available for replication of primary analyses or for secondary analyses or meta-analyses. There are certain data elements that are considered personal identifying information (PII), but would be necessary for performing the research/analyses. For example, Race is considered PII but can be very important to determine if there is a specific safety concern for a certain race of the population. How the level of de-identification in Category 3 differs from Category 2 is unclear. The way category 2 and category 3 are currently defined suggests overlap between the 2 categories depending upon what constitutes adequate de-identification.	Clarify and describe what constitutes "adequate de-identification". Clarify what is the difference between „de-identification" and "anonymization"? Clarify the definitions of category 2 and category 3 data and ensure there is no overlap between these 2 categories.
166-168	It is important that there is general acknowledgement that full transparency and full protection of data privacy (also for the long term future) is not feasible. Indeed there will be cases whereby data anonymisation will still leave the researchers with a data set that has a high level of utilization. But that is not the point. The point is that there will be cases whereby data anonymisation will simply mean that replication/reproduction of the original primary results will not be possible. And it is important that that limitation is a given and to be fully understood by everyone because it is key in understanding the intrinsic incompatibility of patient privacy and full transparency, whether we like it or not.	Add sentence acknowledging that "However, there will also be cases whereby data anonymisation (because of having to leave out parts of the raw data) will simply mean that full replication/reproduction of the original primary results will not be possible."
168	It is unclear to us whether EMA intends a full release of all data or a minimum release of only the data needed for the request's objective. In order to increase patient data confidentiality, a limited release is preferable as many of the data sets will contain tens if not hundreds of variables. Many of these will not be required for the intended purpose of the analysis.	In our view, a limited release requires a pre-specified analysis plan that specifies the variables to be analysed.

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
		Who will prepare the dataset with the limited data for each request?
168-171	It cannot be guaranteed that appropriately de-identified data sets will always preserve the ability to replicate the main analysis as it depends on the patient identifiers included in the main analysis and how much of the data requires de-identification to protect patient confidentiality.	Add *"if this is possible"* to the last sentence in this paragraph.
172	The reference [2] in the document regarding the de-identification requirements, while appropriate for the minimal data that usually appear in publications, would likely lead to problems when applied to the considerably greater amount of data that is collected in clinical trials supporting a regulatory submission.	Provide more details of how data should be de-identified and clarify what is expected to be submitted to describe how data was de-identified, or add a reference that this will be explained in a separate guidance document.
174-175	De-identified data will remain vulnerable to a persistent, intelligent match effort with access to databases of additional personal data such as medical records, insurance claims, vital statistics, and/or similar as well as social media. This is especially true in rare diseases. Data redacted to withstand a robust, sophisticated match effort would likely also lose much or all of its scientific and transparency value. It may be very difficult to implement the recommendation to de-identify data in such a way that "adherence will preclude subject de-identification, even when applying linkages with other data carriers (e.g. social media)." Even the cited reference (Hrynaszkiewicz and Norton, 2010) suggest some options that are difficult to implement such as "Consent for publication of appropriately anonymised raw data should ideally be sought from participants in clinical research" and that in some cases there should be a review by an ethics committee. As de-identification is very complex, it would be helpful to elaborate more on this topic rather than providing just one reference. We believe that a standard for de-identifying data needs to be developed that all can follow.	EMA should consider a closed environment for analyses that precludes the download of patient level data. Provide additional references on de-identification or note further guidance will be developed.

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
176-205	We would recommend adding expectations around appropriate storage of PPD data between downloading and destroying (e.g. Access, security – Physical/logical etc…). Ideally the data would stay in a "closed secure environment" that helps ensure appropriate protection of personal data. Data if accessed outside of a controlled system should only be destroyed after all the analyses have been completed, reported and published. If the data are destroyed after the analysis is completed but before the results are published, the researcher is unable to address any questions that may arise from the publication of the results.	Clarify expectations on appropriate storage, access and the destruction of data to researchers who are granted access to data. Confirm data should be destroyed once all the data analyses are completed, reported and published and there are no questions on the results. However, the statistical programs that generated the results of the post hoc analyses should be kept to allow for reproduction, if necessary (similar to the practice for primary results).
188-190	The policy does not make clear who will take responsibility in case of re-identification. What will the penalties be if patient confidentiality is breached? Who will be held liable?	Clarify who is responsible if data is retroactively identified and who is liable if patient confidentiality is breached.
191-192	It is unclear to us what is deemed "outside the boundaries of patient's informed consent".	Please clarify what is meant by "outside the boundaries".
198	The requester is required to 'have obtained ethics committee approval, as appropriate'. How would the requester know when this is required and to whom must they apply for approval?	Clarify how the requester obtains the necessary information on ethic committees to approach to see approval.

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
199	EMA's standards for good analysis practice should be made publically available and not just communicated to requestors.	EMA should publish the standards they expect for good analysis practice and this should be referenced in the policy.
203-204	All additional analyses conducted by the requester including all their supportive documents e.g. data derivation rules should be posted next to the request to ensure a similar transparency of the secondary analyses to the primary analysis.	Clarify the requestor has to post or publish all of their supporting documents for their additional analyses to promote full transparency. The publication should also mandatory have to indicate that it concerns a post hoc analysis after the trial results have already been published (and the acceptable error rate level (alpha) has already been spent).
203-204	Data owners should be notified/informed of the results prior to publication especially if there is discrepancy. If there are any deviations to the pre-specified plan, these should be identified appropriately and referenced in publications.	EMA should expect requesters to collaborate with the data owners if discrepancies arise in the re-analysis of primary results. For example, this might be to confirm the researcher hasn't used an inappropriate variable or misunderstood the data. EMA needs to put a governance process in place for publication of

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
		results that relevantly deviate from originator's analyses. Clarify that the requestor should include any deviations from their pre-specified analysis plan when publishing their results.
205	Destruction of accessed data: should not happen when the analysis is completed, but after publication. All data and statistical programs used to produce the secondary analysis should be archived for at least 5 years to facilitate further validation if needed. Certification of destruction is mandatory and should be enforced. How does the Agency intend to do so?	Data should not be destroyed for a period of time after additional results have been published. The same rules as for the original analyses should be applied cf. line 67ff.
210-215	The statistical analysis plan should be mandatory. The three issues 1. Replication of analysis / re-analysis (using different approaches / robustness of results) 2. Post hoc analysis, new questions 3. Meta analysis should be clearly separated. Regarding 1): use of the original analysis plan is needed as the additional analysis is a new sensitivity analysis. A dialogue between the researcher and the data owner should be encouraged. Regarding 2): there needs to be considerations to multiplicity as any additional analyses will be exploratory and not confirmatory. In addition the principles of ICH E9 should be followed: pre-specify population, endpoints, analysis model, handling missing data etc. Regarding 3): no additional requirements as there are plenty of existing guidance for conducting meta analyses. EMA could provide a SAP template that could ensure the above aspects are considered.	EMA consider working with industry and academic bodies to provide a template for a SAP for re-analysing data or for conducting secondary analyses. Access to 'C' type data should be contingent on the provision of an analysis plan.
214, 222-231	It is stated that the requester can decline to upload any documents, like an analysis plan, *at the time* of requesting access to 'C' data. The reference to the *time* of requesting access makes us wonder if	Clarify if there will be subsequent opportunities for a requester to

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
	there will be subsequent opportunities for the requestor to upload documents like an analysis plan. If the requestor does not submit it with the request for access, can they still submit it before receiving the data (assuming the request is approved)? Is there any time when the analysis plan is *required* to be submitted, for example at the time of disclosure of results?	upload documents and if there is a time when a SAP is required to be submitted.
216	To ensure scientific validity, the EMA should always judge the validity of the request and the competence of the requester.	EMA should review a request for access to data to confirm the scientific and statistical validity of the proposed analyses.
217	EFSPI believes that the same professional standards should be applied by EMA for secondary analyses as for the primary analysis of CT data.	Add the requirement of a qualified statistician as required by ICH E9 also for secondary analyses of CT data.
222-225	Is the requester of data required to share their computer code when information about the requestor is published by the agency (line 222-225)?	Clarify if the requestor should publish their computer code of their additional analyses.
222-231	The access to 'C' data should be fully transparent. The delayed publication of requests to access to 'C' documents/data and their aims may lead to duplicating research.	Requests including scientific rationale and statistical analysis plan should be published immediately, so that anyone (e.g. sponsor) can comment publically. In addition, immediate publication avoids other researchers developing duplicate requests for access to data.
242-244	Will somebody at some point during the process have to confirm that data have been appropriately de-identified when the data are provided to EMA? If the sponsor has performed an integrated analysis in the submission, the data set containing the integrated clinical trial data should not need to be resubmitted.	Clarify the process for who will be confirming data have been appropriately de-identified. Clarify that integrated data sets containing multiple clinical trial data

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
	There are no details provided on how CT data are to be submitted.	will not need to be submitted if the sponsor has conducted integrated analyses. Clarify how CT data is to be submitted and will the EMA put in place similar guidance to the FDA on data standards and how to submit compliant data sets?
251-252	The policy states that for a variation of a centralised marketing authorisation, CT data not previously submitted to the Agency would be in scope. For older studies the informed consent used previously may not permit the release of data to third parties. How will this conflict be resolved?	Clarify how CT data for a variation is allowed to be in scope of the policy with respect to informed consent in place when the study(s) were conducted.
251-252	The policy seems to only hold for centralised procedure submissions. It would be helpful to clarify that the policy does not hold for any submission as part of a decentralised and/or mutual recognition procedure even though it involves submission to an EU Member State.	Add statement that decentralized procedure and mutual recognition procedures are not within the scope of this policy.
260-261	Typically, a submission contains clinical trials that were conducted over a considerable time span. Do data and study reports e.g. from phase II studies need to be retrospectively adapted to the new rules?	Clarify the scope of the policy.
260-261	If a guidance document is made available 31Oct2014 then 1Jan2015 (2 months including the end of year holidays) could be a challenging timeline for a data owner to de-identify data as per the final guidance, especially if the regulatory filing includes many trials.	Could the final guidance document be available before the 31Oct2014 or the time between the final guidance being available and the effective date is more than 2 months apart?
279	It would be helpful to explain further what is meant by "key codes".	Clarify what is meant by 'key codes'

Line number (s) *(e.g. 20-23)*	Comment	Proposed changes, if any *(If changes to the wording are suggested, they should be highlighted using 'track changes')*
291-292	There are concerns about the legality of allowing access to the names and addresses of the personnel working on the trial. It is difficult to understand why this is in the public health interest.	Information concerning personnel involved in clinical trials should not be made public as the data is confidential.
Annex 2 page 15, Sections 14.3.1 – 14.3.3. compared with Annex 2, pg. 16, Section 16.2	We are wondering about the rationale for making listings of deaths, other serious and significant adverse events, narratives etc. with access "O, 1", while the access for patient listings of discontinued patients, adverse events are "C" category. They seem to be very similar in nature. Section 16.2 implies that all of these patient data listings will be available for every CSR. Whereas in reality, very few patient listings are now included in CSRs as the need to generate patient listings is substantially reduced.	All listings of patient data should be classified as 'C'. Clarify that 'C' access will only be granted where documents exist and the policy is not expecting that these listings be created for every CSR.

Please add more rows if needed.