EMA/240810/2013

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

Comments from:

Name and affiliation
Multi-Regional Clinical Trials Center (MRCT) at Harvard University

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

Introduction

Harvard University's *Multi-Regional Clinical Trials Center* (MRCT) respectfully submits the following comments on the EMA's proposal, 0700 (June 24, 2013), entitled Publication and Access to Clinical Trial Data" (EMA/240810/2013**),** a proposal that will complement the existing 'Policy on access to documents (related to medicinal products for human and veterinary use)' (POLICY/0043) (EMA/110196/2006). The current draft seeks to make clinical trial (CT), detailed and high-quality data available for analysis by a wider scientific community enabling third party analysis and providing opportunities to broaden our understanding of human biology.

MRCT strongly supports EMA's objective to increase sharing of participant-level clinical trials data to ensure regulatory integrity and to further public health. That said, MRCT is aware of the complexities and the need to balance personal autonomy and privacy of the clinical study participant on the one hand, and wider sharing in the service of public health and scientific innovation, on the other. We should state at the outset that any approach to data-sharing generally, including the specific considerations of sharing participant-level data, should apply equally to all public and private study sponsors, whether government, foundation, academic, industry or other, and all data-generators and data-users.

Formed in 2011, MRCT at Harvard is a partnership of academia, government, non-profits and industry dedicated to improving the design and conduct of multi-regional clinical trials, especially those involving sites in the developing world. Of note, MRCT does not fund, plan, conduct, or monitor clinical trials, but rather studies their regulatory, practical and ethical aspects, in order to improve design and conduct of clinical trials. MRCT has convened a number of working groups to study ways to facilitate greater access to participant-level clinical trials data while maintaining the security of private health information. One MRCT work group explored models for sharing access to participant-level data, favouring an 'independent, learned intermediary' to ensure appropriate access.[1] A second MRCT work group has studied the issues impacting informed consent, including participant privacy, confidentiality and identifiability. The comments submitted herein by MRCT focus on the highly nuanced informed consent issues offering a multi-stakeholder perspective, one which MRCT may be uniquely positioned to represent.

Fundamental Comments

1. Protecting Personal Data

The EMA has defined three categories of data: Category 1, containing elements of commercially confidential information (CCI) that will not be made available to the public, Category 2, without personal data concerns that would be released immediately, and Category 3, containing personal data that would be made available (1) after adequate de-identification and (2) through a controlled access process.

Personal data includes any information relating to an identified—either directly or indirectly—or identifiable person. MRCT at Harvard agrees with the Agency's acknowledgement of the limitations of current technologies that purport to protect participants from later identification. Given

[1] The models were reviewed in May and presented to a large multi-stakeholder group including experts from academia, industry, government and patient groups for feedback. The majority of the participants favored the use of an independent, learned intermediary (independent of the trial sponsor) that would review and approve individual research requests to ensure access by qualified researchers. The conference materials, including slides and a summary of the discussions from the conference are available on our website, http://mrct.globalhealth.harvard.edu

advances in data mining technologies, availability of databases and the potential for linkage, and the ever-expanding information available in social media, the ability to guarantee individual participant privacy and confidentiality is limited, especially where access is more open and information can be downloaded and combined with other publically available datasets. Though participant re-identification and misuse of data may be unlikely overall, the consequences of one incident of re-identification occurring may severely threaten public trust in the scientific investigative process with potential detrimental impact on participation, volunteerism, and confidence in the clinical trial system. We therefore respectfully submit that maximal clarity with regard to ownership of personal data be provided together with the steps necessary to achieve adequate data de-identification given the current informatics environment to inform policy on the use of participant-level data.

2. Defining "de-identified data"

The risk of identifying study participants varies depending on the underlying type of data, as different methods of de-identification accord different levels of protection. Particularly, in clinical trials, there exists an important analytical and practical distinction between key-coded, de-identified and anonymized data. In key-coding, every participant is assigned a code and the link between the coded data and the individuals, such as name, date of birth, and address, is kept separately. Investigators are the only party with access to the key, while sponsors of the clinical trial and other parties involved receive data in the coded form only, it is important to note also that sponsors apply additional controls to assure confidentiality of the key coded data related to access and storage such that the combination of the coding and the controls affords adequate levels of protection. Of note, the data has not been de-identified completely at this point, for example, some attributes such as age, dates of birth may be present. General sharing of key coded data without the aforementioned controls does not provide the most advanced level of protection for participants' data nor does mere de-identification even if completed according to generally accepted standards and therefore has an associated risk of re-identification. On the other end of the spectrum is complete anonymization of participants' data through such techniques as removal of indirect identifiers[2] and a second coding of this de-identified data or through aggregation of data. Individual participants are no longer identifiable and the data cannot reasonably be traced back to them. As a result, the risks posed by sharing such anonymized data are considered to be lower. MRCT is conscious of the different levels of protection afforded by key-coded, de-identified and anonymized data and in referring to de-identified data through this comment, MRCT is referring to data than has been fully de-identified and anonymized.

In contrast, the EMA draft Policy does not elaborate on what is meant by "de-identified" data and which de-identification standard would EMA believe provides an adequate level of protection? Is de-identification according to a particular standard acceptable or are more advanced levels of protection, such as removal of indirect identifiers and verbatim text, and a second anonymization with discarding of key code, required? Significant clarity could be gained by defining the meaning of "de-identified data" within the policy. EMA should evaluate potential ethical and practical considerations raised by methods and from a practical perspective, how these methods will be impacted by the degree to which data can be combined with other available datasets. Very lenient de-identification and accessibility increases the risk of participant privacy breach, which may in some jurisdictions be a breach of Good Clinical Practice, and could serve to discourage future clinical trials participation. On the other hand, a more rigorous approach may place undue burdens on data-generators, as well as render anonymized data useless for future researchers by stripping it of relevant identifiers.

[2] Final advice to the European Medicines Agency from the clinical trial advisory group on Protecting Patient Confidentiality. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2013/04/WC500142853.pdf, accessed 26.9.2013.

3. Recognizing the Role of Informed Consent Documents in Protecting Personal Data

Mitigating against instances of re-identification and resulting erosion of public trust is the education and information given to the participant during the informed consent process, embodied in the informed consent document ("ICD"). There is no question that the EMA acknowledges the importance of the informed consent. The EMA states in the draft policy that it takes "*a guarded approach to the sharing of patient level data*" and further indicates that "*respect for the boundaries of patients' informed consent*" is required. However, the statement that follows in the text of the policy (reproduced below) raises some ambiguity when applied to the content of individual study site ICD and therefore has the potential to be interpreted in a number of ways:

> "*Patients participate in clinical drug trials in the hope that their data will support the development and assessment of a particular medicine that is useful for the treatment of their disease, and will benefit the advancement of science and public health. The Agency takes the view that any other use of patient data oversteps the boundaries of patients informed consent and shall not be enabled by the policy*"

MRCT finds the clause "…and will benefit the advancement of science and public health" a potential point of confusion and broad interpretation.[3] Further elaboration would be of tremendous benefit to the research community. For example, does the EMA suggest that participants "hope that their data…will benefit the advancement of science and public health" independent of application to or treatment of their disease, or only in the more limited circumstance of application to their specific disease? Does the language above indicate the EMA only advocates data access to clinical trial data for research as related to one particular medicine and only in the narrow and specific disease area under study, and that future use of the data by third parties for any other research oversteps boundaries of the ICD? Or, does the EMA advocate that data users must only demonstrate a commitment to advance science and public health to have broad access to data, regardless of the specifics of the ICD? MRCT also notes that the EMA refers to third party researchers respecting the "spirit of the informed consent," which may be inferred to mean that researchers must not use data for purposes which the informed consent language does not contemplate.

4. Aligning Data Sharing with Informed Consent Documents

The informed consent is a legal document, the terms protected by contract law provisions. However, there are also strong ethical concerns relating to the degree to which the study participant is informed and truly understands that which is conveyed and in turn consented to by signing the document. Patients participate in clinical trials with the knowledge and expectations as conveyed by the language of the ICD and as explained by the principal investigator. The template ICD provided by the data generator and/or sponsor of the clinical trial may well contain broad language that explains how the data will be used (e.g. that the data from the trial may be used for the current research, in future research regarding the same medicine or disease area or by third party parties in related or unrelated future research). MRCT believes that there is an ethical responsibility on the part of sponsors and data-generators, principal investigators and individuals, who have attained informed consent. Likewise, secondary data-users also have an ethical responsibility to respect the specific language and the intent as represented in the informed consent document. Thus, researchers have an obligation to act in concert with these agreed upon original commitments, whether agreed to

[3] Actually, most confusing is the comma that precedes the "and will benefit…" Deletion of the comma would more affirmatively lead the reader to the conclusion that both conditions (development and assessment … useful for the treatment of their disease and will benefit the advancement…) must be met.

explicitly or implicitly.[4] However, determining what acting in concert with the original commitments entails is not a simple determination. MRCT considers the position of the research participant: on the basis of the ICD, what would the participant expect? The ICD is the understanding, albeit with varying levels of clarity, with which individuals agree to participate in clinical trials.

The analysis of how to proceed depends upon whether one is discussing a study in which informed consent has already been obtained from the participant (retrospective/current studies) or prospective studies, in which the ICD will elucidate, in advance, the intent to share participant data broadly. We will discuss each in turn.

a. Analysing Informed Consent Documents in Prospective Studies

It is more straight-forward to analyse use-cases relating to future, prospective studies. In this case, consistent with the current regulatory requirements, the ICD should disclose to any potential participant that his or her data will be shared, how it will be de-identified and anonymized, and that every attempt will be made to protect the individual from secondary re-identification. It should also include provisions that while the risk is small, there will always be the possibility that a research participant's identity will be revealed with the potential for compromising their privacy. Thus, any participant who has been properly informed and agrees to participate in a clinical trial will be made aware of the risk of re-identification as technology advances, even though the data may be de-identified according to the requirements of applicable regulations at the time of obtaining the consent. On the other hand, participation in the research study is then conditional on the agreement to allow future use of data for the greater good of science and public health, a framing that, absent due care, might be interpreted by a potential subject as compromising their autonomy, or even coercive. Further, such a condition—even if approved by the ethics committee or institutional review board—may potentially introduce bias to the analyses made on the subject cohort,[5] compromising the ability to generalize an interpretation from the dataset. An additional nuance and ethical consideration is raised by an analysis of when data is removed from association to a particular participant such that it is no longer "their data" and consideration of broader balancing with public interest allows for sharing of this data for further research.

b. Analysing Retrospective Informed Consent Documents.

1. Permissibility Language in Retrospective Informed Consent Documents

A potentially more challenging scenario describes one in which a study has already been initiated, with the proper consent documents obtained, but the data not yet submitted to the EMA. MRCT believes that the intent and language of the specific ICD (such that an analysis of what the study participant would likely have understood and consented to) must be respected, thus a retrospective review of all ICDs will be necessary prior to any decisions about sharing the clinical trial data. After reviewing many informed consent documents, MRCT can categorize the documents into five broad categories:

[4] A similar conclusion pertaining to biospecimens has been reached by the Health and Human Services Secretary's Advisory Committee on Human Research Protections (SACHRP) and can be equally applied to data: "In the case where secondary use of tissue samples is not compatible with the original consent for tissues that are de-identified, coded, or anonymized and are not readily identifiable, the samples are no longer subject to human subject regulations. Thus, there is no regulatory violation. Nevertheless, the original investigator and his/her institution have made an agreement with the subjects about use of their specimens, and have an obligation to honor that agreement." http://www.hhs.gov/ohrp/sachrp/20110124attachmentatosecletter.html, accessed 14.9.2013.

[5] It is well appreciated that trust in the "medical establishment" varies among different ethnic, racial and socioeconomic groups.

1. ICD is *explicitly prohibitive*, stating clearly that 'the identifiable data will not be shared.' Example language includes, "All information collected throughout this study will remain strictly confidential." Wherein the ICD says "Records of your participation in this study will be held confidential except as disclosure is required by law," the participant fully anticipates that legal disclosure of their data will be for specific and pre-defined research purpose of the particular clinical trial and not generally released by a change in the law.

2. ICD is *explicitly permissive*, stating affirmatively that the 'identifiable data will be shared' with researchers for 'future research.'

3. ICD is *explicitly but selectively permissive*, stating affirmatively to whom the identifiable data may be released, and identifying a restricted and defined population of individuals or situations that may receive data. This type of phrasing is used with increasing frequency amongst recent ICDs. For instance, the ICD may read "your study information will be shared with the study sponsor and its representatives including companies that it works with, the study team and researchers at other sites, government health agencies (such as the FDA and the EMA). The study sponsor will use and disclose your information only for research or regulatory purposes or to prepare publications." In specifying the groups to whom identifiable data *will* be released, the participant will infer that it will *not* be released to others not mentioned.

4. ICD is silent, with no provisions or descriptions on how a participant's data will or could be shared for future research.

5. ICD is contradictory, in which varying segments of the ICD are discrepant.

In addition, all ICDs for applicable clinical trials initiated after March 7, 2012 comply with 21 CFR 50.25(C) and include the FDA mandated language: "A description of this clinical trial will be available on *http://www.ClinicalTrials.gov*, as required by U.S. Law. This Web site will not include information that can identify you. At most, the Web site will include a summary of the results. You can search this Web site at any time." This statement informs participants of the availability of their summary information on the public website. MRCT recommends that, if the EMA will allow data availability according to 150-154 ("data will be available as downloads from the Agency's website, at the time of publication of the European Public Assessment Report (EPAR)…"), then comparable mandatory language should be specified by the EMA prior to posting.

Envisioning potential interpretations by participants, what the participant believes to be "his/her data" may vary[6] as does the desire to protect this data for fear of re-identification. To the extent that data is aggregated, de-identified and/or anonymized, participants may no longer view the data as his or her "own." Moreover, given the interest study participants have in advancing science and public health, whether in relation the original medicine or disease or an interest in contributing more broadly, participants may be keenly interested having data shared for use in further research when the data are no longer reasonably tied them. Conversely, as the risk of re-identification increases, study participants may be more likely to view data as their "own" in effort to maintain control and protect their privacy. While these issues can be more clearly communicated and agreed

[6] Arguably many would assume that 'your' data refers to personally identifiable information, not aggregate or de-identified data.

upon in ICD moving forward, clarity of the original consent communication and therefore the intent of the study participant is difficult to ascertain, retrospectively. A policy decision as to approach to use in analyzing these issues may be very helpful from a practical perspective.

2. Interpreting Informed Consent Documents Language on Data-Sharing

The five categories of ICDs merit different approaches to data sharing and raise a number of considerations. In general, if there is explicit language prohibiting release of the data, the agreement within the ICD should be honoured and identifiable data should not be shared (Category 1). However, it is not clear whether the prohibition on data release should also apply to anonymized data. Arguably, since such data can no longer be connected to individual study participants, it no longer belongs to that individual. In such a circumstance there are potentially significant benefits that can be gained from future research uses of fully anonymized data, whether for further research within the specific disease area of original study or for broader research purposes. It is also unclear whether regulatory obligations of data-generators override any express prohibition on data-sharing. In situations where regulations impose a legal obligation on data-generators to share data, is sharing permitted and where does the legal responsibility to protect participant privacy fall?

Similarly, the Category 2 explicitly permissive provisions create a number of challenges. The express language of ICDs affirming future use suggests that there should be no prohibition on releasing patient-derived study data.[7] In practice, however, the analysis can likely be more complicated as relevant laws, regulations and ethics committees' guidelines, in multiple countries where the clinical trial has been conducted, may place significant limitations on data sharing, overriding express consent. Furthermore, there may be ethical reasons to reconsent a prior participant who required a consent by a legal representative at the time of clinical trial enrolment. For example, in pediatric trials, the parent/guardian may have consented on behalf of the research participant, but the datasets that are under consideration for sharing may now refer to an individual who has reached the age of consent. Lastly, the extent of such expressly permitted data sharing is ambiguous; it is unclear whether by agreeing to data use for future research, the patient believes his or her data would be shared within his or her specific disease context or can be used for other, unrelated research.

In Category 3, the specific use of and audience to receive the data has been defined and identified in the ICD. Considering the perspective of individual participants, such language can be seen to suggest that data will only be shared to the circumscribed list included in the ICD. While such interpretation can be debated, MRCT believes that for retrospective studies, a more conservative approach strikes a better balance between participants' privacy and the public's interest in furthering scientific research and innovation. Thus, in such circumstances, data-sharing outside of the intended audience will be problematic. If the ICD has committed to sharing the data in only very specific settings, deidentifying data does not eliminate the responsibility to the participant. Is there then a point at which data is no longer reasonably associated with a participant such that it is no longer "their data" and would this allow data falling in this category to be shared? A broad policy decision may be necessary to facilitate this interpretation by stakeholders otherwise a conservative approach is more likely.

The setting in Category 4 would perhaps allow one to de-identify and share potentially identifiable data provided such data was de-identified to the maximum extent which may involve considerable

[7] MRCT recommends that data that is to be shared generally be de-identified to the extent consistent with the intended use. So long as the ICD did not promise confidentiality, future use is envisioned.

resource in the case of multi-regional trials;. In the event that the ICD is contradictory (Category 5), the more conservative commitment (i.e. most restrictive access) will most likely be honoured. In both Category 4 and 5, ethics committees and institutional review boards can provide case-by-case guidance on data-sharing where data is identifiable.

3. The Role of Ethics Committees and Institutional Review Boards

MRCT suggests that ethics committees (ECs) and institutional review boards (IRBs) can play an important role in ensuring that data sharing is conducted "in line with the spirit" of pertinent ICDs. In each of the settings that raise some concern as to the permissiveness of the ICD, should ECs/IRBs review the ICD and the data tables or study documents proposed for release or posting? These committees are charged with respecting and protecting the "rights and welfare" of the participants and can assess the specific trial, the ICDs, and the potential risk of sharing data more widely. Their deliberation may consider such elements as the sensitivity of the data (e.g. disease states including mental health status; social/behavioural, demographic and reputational harms associated with potential re-identification, for instance sexually transmitted diseases, high risk activities and illegality), among other issues. The EC/IRB can determine whether, depending on the specific situations, the planned action is consistent with the ICD, the participants should reconsent to the planned use (impractical in many studies), or if a waiver of consent is permissible. MRCT is conscious of the administrative burden imposed upon the EC/IRB for which they may be neither prepared nor resourced.

4. Addressing privacy concerns through Informed Consent Documents

While data privacy considerations are linked to adherence to the terms of the informed consent, the two are separable. An informed consent should articulate in clear and understandable language what will and what will not happen to personally identifiable information and data that has the potential to be easily re-identified, together with the protections in place by law and additional protections put in place by sponsors and investigators. Though participants may understand that their data will be submitted to regulatory authorities, either because of explicit or implicit language, they do not likely contemplate the sharing of this data by regulators for use by external researchers. At this foundational step, an approach to data sharing that fails to adequately protect patient privacy is violative of the "spirit" their informed consent. Furthermore, participants may have heightened concerns should there be increased risk of re-identification given a particular approach to data sharing such as open access or even through a controlled access approach that is lacking true control from a practical perspective or as implemented. Control of release of data may require greater oversight of data release such as to whom and whether the data can be downloaded for future use and combined with other available data. However, when ICDs do clearly articulate data sharing for further research and data are de-identified and/or anonymized so as to adequately protect the participants' identity and then shared via an approach that adequately protects the participant from re-identification, these privacy concerns lessen and benefits to public health can be realized. Therefore, any data-sharing decision based on retrospective ICDs should carefully balance privacy obligations, International Conference on Harmonisation (ICH) of GCP ethical guidelines, and the language of ICDs in determining whether data-sharing is appropriate and expected.

5. Concluding comments and suggestions for further consideration

How best to proceed with prospective and retrospective ICDs involves consideration of various complexities. Significant difficulties are also presented by situations calling for individualized determination of ICDs permissibility. Such situations can arise in a number of circumstances. For example, some authorities propose that participants be given the option as to whether, and how,

they permit their data to be shared—that participants may "opt out". To complicate this situation further, even if the data generators and sponsors intend the ICD to contain broad consent for future use, some local ECs/IRBs will modify the local consent or give participants the choice of whether or not to participate in data-sharing. Broad consent language is often modified by ethics committees at the national or site level, sometimes in order to comply with local laws, resulting in restrictions to language limiting or prohibiting future research or analysis of the data. In some instances the ICD may not include such language from the outset or be silent on the subject of future research.

These situations present a statistical problem for data integrity. One cannot post data tables in which subsets of subjects are deleted or not represented as the statistical precision of the interpretation may be compromised and such exclusion of data would potentially yield biased results. This could also affect the secondary analyses of such data, which lacks the statistical power needed to validate the original claims made by the sponsor and/or EMA, potentially increasing the burden for the regulatory agency in defending its (appropriate) initial conclusion. If this scenario were to be realized, how would the EMA intend to address access to studies where some portion of the participants opted out of sharing their data with further research and thus the data may lack both integrity and usability?

Similarly, EMA proposed approach to situations where there is an affirmative obligation to protect personal data requires further consideration. Though, appropriately, de-identification is proposed, the specifics are not delineated. As a starting point, the "controlled access" the EMA proposes is dependent on defining "de-identified" data and delineating the scope of data-sharing – whether such sharing should be permitted only in the narrow disease area or more broadly, for unrelated studies. EMA states that its "control access" analysis will be "in line with the spirit of informed consent." How will the EMA ensure the concordance of use and ICD? Presumably, EMA will review the ICD since the data generators, those most familiar with the study, will not receive the data request directly from the external researchers. One would assume that the requester will provide a protocol or study plan so that alignment with consent can be confirmed. Will EMA have legal authority to review and deny access if there is lack of alignment? Will EMA have the resources to review each ICD for concurrence to data request?

Alternatively, will EMA be informed by the data-generators as to the appropriateness of sharing contributed data? Should data-generators determine whether clinical trial data can be released in compliance with national and global data privacy laws, ethical norms and the principles of both the Declaration of Helsinki and ICH GCP; they will also consider adherence to the specific ICD. Would EMA then retain authority to agree or disagree with the analysis of the data-generator? Will then EMA/data generators then interpret the ICDs, the available data, and the qualifications of and specific studies proposed by the data requestors. If EMA shares data despite a recommendation *not* to share the data by the data-generator, is the EMA liable for subsequent use or misuse of the data, for data and privacy breaches? If EMA takes action against the data-generators, will the EMA indemnify the data generator for any third party claims in circumstances where data is released contrary to the terms of the ICD?

***

MRCT strongly encourages the development and publication of practical guidance clarifying how the language of the ICD, and the requirement of the data-generator and data-user to adhere to such language. In addition to concerns about privacy and confidentiality, and risk of re-identification, the choice of the individual study participant regarding sharing of his/her data for use in further research must be balanced with the greater good of using data to speed scientific discovery, increase efficiency of clinical trials (possibility of fewer patients exposed to potential harms and more expedient trials) and the public interest of bringing beneficial medicines to patients who need them.

Thoughtful and respectful approaches to situations described above are possible though far from straight-forward, as EMA contemplates moving forward from a practical standpoint.

MRCT thanks the Agency for the opportunity to comment on this draft proposal, and stands ready to work the EMA on these important issues

Sincerely,

Barbara E. Bierer, M.D.
Co-Chair Multi-Regional Clinical Trials Center, Harvard University

Rebecca Li, Ph.D.
Executive Director Multi-Regional Clinical Trials Center, Harvard University

Mark Barnes, J.D.
Co-Chair Multi-Regional Clinical Trials Center, Harvard University

Jessica S. Scott, MD, JD
Co-Chair of MRCT Center Data Sharing Implementation Workgroup
Director, North America Medical Advocacy and Policy
North America Medical Affairs, GlaxoSmithKline

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:	MRCT strongly encourages the development and publication of practical guidance clarifying how the language of the ICD, and the requirement of the data generator to adhere to such language, influences and impacts the implementation of the EMA policy.	MRCT strongly encourages the development and publication of practical guidance clarifying how the language of the ICD, and the requirement of the data generator to adhere to such language, should inform the implementation of, and data generators' compliance with, the EMA policy.
44-48:	The language of the draft Policy is open to significant interpretational differences. In stating that patients "hope that their data…will benefit the advancement of science and public health," it is unclear whether the EMA believes that patients consent to a broader access to their data, independent of application to or treatment of their disease, or to the more limited access for research in the narrow disease area. Does the language indicate the EMA only advocates data access to clinical trial data for research in the narrow disease area of the clinical disease area and that future use of the data by third parties for any other research oversteps boundaries of the ICD? Or does the EMA advocate that data users must only demonstrate a commitment to advance science and public health to have broad access to data, regardless of the specifics of the ICD?	Respect for the boundaries of patients' informed consent: Patients participate in clinical drug trials in the hope that their data will support the development and assessment of a particular medicine that is useful for the treatment of their disease, and will benefit the advancement of science and public health. The Agency takes the view that any use or disclosure of patient data which oversteps the boundaries of patients' informed consent, shall not be enabled by the policy.
57-61:	The Agency properly recognizes the difficulties of guaranteeing that all secondary data analysis will be conducted and reported to the highest possible scientific standard. The draft Policy states that measures will be put in place to safeguard the public from inappropriate analyses. However, the specifics of such measures are not delineated. What ICD information will the third party researcher receive to ensure compatibility with the proposed project? The third party should exercise caution	

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')*
	if the data set has been reformatted because of "opt out" clauses for use of the data beyond the protocol. Similarly, will the EMA require researchers to gain approval of local or national ethics committees to conduct studies using the data?	
64-65:	The responsibility for ethical and legal compliance rests with the party releasing the data and such responsibility must be executed in good faith. If the regulatory agency contravenes the recommendation of the party releasing the data, it has responsibility for follow-on consequences and will indemnify the data generator from subsequent action.	
172-175:	Common standards, rules and procedures for deidentification and protection of patient privacy are needed. For meaningful data analysis, data can be de-identified but anonymity is, at best, challenging. Some data (e.g. genetic data, orphan diseases) may need to be excluded even if appropriate de-identification has occurred. The acceptable processes and procedures should be delineated.	
183:	Allowing controlled access in line with "the spirit of informed consent" raises a number of challenges. Since, at a minimum, ICDs present an ethical contract with a study participant, the release of clinical trial data must be concordant with the ICD. The ethical and legal release should tract the language in the ICD. However, the draft Policy does not address how future data releases plan to comply with existing ICDs. Is there a specific mechanism that the EMA proposes to use to ensure compatibility of the ICD with data sharing? Has the EMA considered to whom and to what extent the sharing of key coded study data is permitted under the informed consent documents for each study?	
244-247:	Standard data formats for existing and completed studies are imperative if aggregation and compilation of multiple datasets is desired.	