MRCT Data Sharing and Informed Consent

Background: The European Medicines Agency has announced its final policy on publication of clinical trial data (October 2014)

Impact:

- Regulatory agencies worldwide are reviewing their current policies for clinical trial data sharing

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

2 October 2014
EMA/240810/2013

European Medicines Agency policy on publication of clinical data for medicinal products for human use

POLICY/0070
Status: Adopted
Effective date: 1 January 2015
Review date: No later than June 2016
Supersedes: Not applicable

1. Introduction and purpose

http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/general/general_content_000555.jsp

EMA Final Policy, Oct 2014 – Highlights

- All clinical study reports (CSRs) submitted after Jan 1, 2015 will be available for download by the public
- Sponsors will be allowed to redact commercially confidential information
- Informed consent will be respected
- Data will be shared with appropriate protections around “personal data” (anonymization)
  - The agency cites concerns on emerging data mining techniques and database linkages
  - A “terms of use” contract will govern access and usage

MRCT Informed Consent Subgroup

Informed Consent Subgroup

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

Objectives:

- Discuss the risk of re-identification and determine protective mechanisms

  1. Achieve clear definitions of identifiable data, anonymized data, anonymous data, de-identified data, and re-identified data.

- Provide guidance in two areas:

  2. Retrospective consent: Interpret the limitations imposed by language in existing informed consent forms (ICF's) upon data-sharing.

  3. Prospective consent: Develop ICF language for patient-level data-sharing acceptable to IRBs and patients groups.

Retrospective consent text

After review of forms, retrospective consent language regarding sharing of de-identified data can be divided into four categories :

(1) Explicitly permits data-sharing	(2) Specifically prohibits data-sharing	(3) Is silent on data-sharing	(4) Is inconsistent/ conflicting with regard to data-sharing

- Category 1: Assumes that the data have been sufficiently de-identified prior to sharing
- Category 2: Possibility of sharing data with other researchers, organizations, regulatory agencies, etc., is wholly forbidden in ICF
- Category 3: Silence regarding data-sharing creates ambiguity; data-sharing decisions left to the discretion of the researcher/sponsor; no requirement that data be de-identified
- Category 4: Statements made about maintaining strict confidentiality, but lists agencies with which de-identified data will be shared (e.g., regulatory agencies)

Review of "retrospective" ICF
Common consent language regarding data sharing

Topic	Text
Regulatory purposes • Auditing • Drug approvals • Legal obligation (e.g., post to ClinTrials.gov)	"Data may be seen by the IRB and with FDA, NIH, or other federal agencies, if applicable." "may share the information with regulatory agencies to approve new medicines." "share the information with people who check that the study is done properly (like the ethics committee or review boards)." "A description of this clinical trial will be available on http://www.ClinicalTrials.gov, as required by US law. This website will not include information that can identify you. At most, the website will include a summary of the results. You can search this on the website at any time."
Future use of data • Secondary use of data studies • Other studies related to disease condition or drug under study	"combine the information with results from other studies to learn more about the medicine and other medicines, and… other diseases and conditions." "share coded information with other companies, organizations, or universities to carry out research." "By signing this form, you are giving consent for any future studies of genes that we may perform in the laboratory."
Data ownership • Who owns data/samples after collection?	"Your DNA sample will remain the property of [company]." "[Company] will be the owner of the study results."
Privacy and Confidentiality • How data are stored	"study information will be labeled with a code number. It will not include your name or address." "your blood is stored and tested with an identifying number, and your name will not appear on the stored samples."

DRAFT – Suggested Definitions for ICFs - Definitions

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

Definitions

personally identifiable data	information that directly or indirectly identifies you (e.g. name, SSN, address, telephone number, email address)
coded data	study participant data and samples for which a code is given to replace a person's name or other specific information.
de-identified data	data modified to remove the presence of personally identifying information
your data	your personal and medical data collected during the study, which may be coded or de-identified

DRAFT - Suggested Prospective Text for ICFs - DOMAINS

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

Privacy & Confidentiality	• **How will my information be kept confidential?** • **Language describing the use of codes** • **Language describing modifying data for de-identification purposes**
Data Usage	• How will my data be used? • With whom my data will be shared ?
FDAAA, 42 U.S.C. 282(j)(1)(A), section 402(j)(1)(A) of the PHS Act Statement	"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."

HARVARD Global Health Institute

Our Draft ICF Language for Data Sharing

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

Key Sections

SEE HANDOUT

- What information about me will be used in the study?
- Who may see, use and share your personal and health information?
- How will my information be used?
- How may my data be used for additional research?
- What other information is shared?
- Do I have to participate?
- Can I change my mind?

HARVARD Global Health Institute

Open Issues

- Prospectively many institutions and sponsors will hopefully move to adopt more expansive data sharing language
- CSRs submitted after January 1, 2015 will include data from participants consented before this language was in place
- How to operate in compliance with the new regulations while maintaining respect for the original consent process?

Your feedback

Comments on draft language are welcome to Rebecca_Li@harvard.edu

Reference material

Agreeing on common definitions: "Identifiable"

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

Defining "Identifiable" Data: Statutes/Regulations

specified categories approach

Ex. 1: California Senate Bill 1386:
PI means an individual's first name or first initial and last name in combination with either (1) SSN, (2) driver's license# or California ID#, (3) account number, credit or debit card number, in combination with its pin/code

Ex. 2: HIPAA Privacy Rule
…information, including demographic data, that relates to:

- the individual's past, present or future physical or mental health or condition,
- the provision of health care to the individual, or
- the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe can be used to identify the individual…

"broader definition" approach

Ex. 1: US Department of Defense; Biometrics Identity Management Agency, Biometrics Glossary
…[i]information about an individual that identifies, links, relates, or is unique to, or describes him or her, e.g., a social security number; age; military rank; civilian grade; marital status; race; salary; home/office phone numbers; other demographic, biometric, personnel, medical, and financial information, etc.

Ex. 2: US Department of Labor Definition
…any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means.

Ex. 3: NIH Definition
the identity of the subject is or may be readily ascertained by the investigator or readily associated with the information

Ex. 4: Common Rule (45 C.F.R. § 46.102)
the identity of the subject is or may readily be ascertained by the investigator or associated with the information

Agreeing on common definitions: "Identifiable"

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

Defining "Identifiable" Data: EU Means-Based Approach

The EU Data Protection Directive (95/46/EC; Oct. 1995)

Article 2(a): "Identifiable" person is "one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."

…

Recital 26: "to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used … to identify the said person"

EMA Draft Policy on Publication and Access to clinical-trial Data (uses above definition)

"an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity"

EU Proposed General Data Protection Regulation (Jan. 25, 2012)

Article 4(2): "'personal data' means any information relating to a data subject"

…

Article 4(1) "'data subject' means an identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person"

Agreeing on common definitions: “Identifiable”

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

Defining “Identifiable” Data: Academic Input – Issues to Consider

"personally identifiable information" has no particular technical meaning. Algorithms that can identify a user in an anonymized dataset are agnostic to the semantics of the data elements. While some data elements may be uniquely identifying on their own, any element can be identifying in combination with others.

Arvind Narayanan & Vitaly Shmatikov, De-Anonymizing Social Networks, in Proc. 2009 30th IEEE Symp. on Security & Privacy 173-187.

“Information in the middle of the risk continuum relates to an identifiable individual when specific identification, while possible, is not a significantly probable event. In other words, an individual is identifiable when there is some non-remote possibility of future identification. The risk level for such information is low to moderate.

Schwartz, Paul M. and Solove, Daniel J., Reconciling Personal Information in the United States and European Union. 102 California Law Review (2014 Forthcoming)

…[L]aws differ in what actually qualifies for data protection. Some countries list specific examples of what can constitute personal data; others are satisfied with a more flexible—or ambiguous—definition. Although specific definitions may offer the benefit of greater certainty, they are subject to criticism as rigid and incapable of responding to new developments. Conversely, the flexible definitions do allow for future adaptability but can lead to uncertainty.

William B. Baker and Anthony Matyjaszewski, The changing meaning of "personal data," International Association of Privacy Professionals. https://www.privacyassociation.org/resource_center/the_changing_meaning_of_personal_data

Agreeing on common definitions: "De-identified vs. Anonymized"

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

De-Identified Data	Anonymized
Laws/Regulations	
HIPPA • removal of 18 types of identifiers or • expert determination using statistical or scientific principles that there is very small risk of identification **EMA Draft Policy** • minimum standard described in Hrynaszkiewicz (removal of all direct identifiers and ensuring that the data contains no more than 3-4 indirect identifiers)[1] • de-identification methods should be sufficient to prevent re-identification even when applying linkages with other data carriers (e.g. social media).	• No definition found
Academic /IRB Approach	
• data stripped of all subject identifiers, including all 18 HIPAA identifiers.[2] • removal or replacement of personal identifiers so that it would be difficult to reestablish a link between the individual and his or her data.[3] • personal identifiers in a record have been extracted and … it would be very difficult to re-establish any of the people mentioned in the original record.[4] • De-identification is different from anonymization as it doesn't protect individuals from inferences[5]	• data stripped of all subject identifiers and have no indirect links to subject identifiers. [2] • irreversible removal of the link between individual and his/her medical record data to the degree that it would be virtually impossible to reestablish the link[3] • all of the links between a person and the person's record have been irreversibly broken so that it would be virtually impossible to re-establish any of the people in the original record. [4] • (interpreting HIPPA privacy rule) -"Previously identifiable data that have been deidentified and for which a code or other link no longer exists. An investigator would not be able to link anonymized information back to a specific individual."[6]
sometimes used interchangeably • Stanford School of Medicine guidelines to "anonymizing" data direct removal of 18 HIPPA identifiers[7]	

Agreeing on common definitions: "De-identified vs. Anonymized"

MRCT Multi-Regional Clinical Trials
The MRCT Center at Harvard

Defining "Re-Identification"

Laws/Regulations

- **HIPPA §164.51499(c)**
  - "A covered entity may assign a code or other means of record identification to allow information de-identified under this section to be re-identified by the covered entity…"
- **Patient Safety Act 42 CFR § 3.212**
  - Under certain conditions, "a provider, PSO, or responsible person may assign a code or other means of record identification to allow information made nonidentifiable under this section to be re-identified by such provider, PSO, or responsible person…"

Academic Approach

- "The reverse of anonymization is reidentification or deanonymization. A person, known in the scientific literature as an adversary, reidentifies anonymized data by linking anonymized records to outside information, hoping to discover the true identity of the data subjects."
- Re-identification is the process by which anonymized personal data is matched with its true owner.[2]
- Re-identification of an individual or small group is essentially achieved by using "quasi-identifiers" to cross-reference specific elements that are included along with the genetic data, but also found in other databases; i.e., elements that can directly or indirectly narrow down the numbers of a given group or subgroup to recognize the subject.[3]
- **Summary:** not much disagreement on the definition - re-identification refers to the process of linking participant data to individual participant

HARVARD Global Health Institute