

# Privacy in the Use of Big Data for Research

Applying Big Data to Address the Social Determinants of Health in Oncology  
A National Cancer Policy Forum Workshop  
October 28, 2019

Kristen B. Rosati  
Coppersmith Brockelman PLC  
[krosati@cblawyers.com](mailto:krosati@cblawyers.com)  
602-381-5464

# *A Complicated Web of Laws Regulating Privacy in Research*

- US federal law
  - HIPAA
  - Federal substance use disorder treatment regulations (the “Part 2 regulations”)
  - Common Rule
  - FDA regulations for clinical trials
  - NIH policies (the Clinical Trials Policy and regulations regarding Certificates of Confidentiality)
- US state laws
  - Consumer privacy protection laws (e.g., the California Consumer Protection Act)
  - State health information confidentiality laws
  - State licensure requirements
- EU General Data Protection Regulation – and individual countries’ laws throughout the world

# *HIPAA Compliance*

- HIPAA applies to “covered entities” and their “business associates”
- HIPAA applies to “protected health information” (PHI)
  - Name;
  - Street address, city, county, precinct, or zip code (unless only the first three digits of the zip code are used and the area has more than 20,000 residents);
  - The month and day of dates directly related to an individual, such as birth date, admission date, discharge date, dates of service, or date of death;
  - Age if over 89 (unless aggregated into a single category of age 90 and older);
  - Certain numbers related to an individual (telephone numbers; fax numbers; social security numbers; medical record numbers; health plan beneficiary numbers; account numbers; certificate/license numbers; vehicle identifiers, serial numbers, and license plate numbers; device identifiers and serial numbers);
  - Email addresses, Web Universal Resource Locators (URLs) and Internet Protocol (IP) addresses;
  - Biometric identifiers, such as fingerprints;
  - Full-face photographs and any comparable images; or
  - Any other unique identifying number, characteristic, or code

# *HIPAA Research Rules*

- The PHI is de-identified: either through removal of all HIPAA identifiers (the “safe harbor” method) or by certification of a statistical expert
- Only a “Limited Data Set” is used, subject to a “Data Use Agreement”
- The research participant or the research participant’s legally authorized representative signs a written HIPAA authorization
- An institutional review board (IRB) waives or alters the HIPAA authorization requirement
- The activities are only to prepare for research, and the investigator makes certain representations
- The activities are to recruit patients to participate in clinical research (or the patients of another health care provider under a business associate arrangement)
- The research involves the information of decedents only and the investigator makes certain representations
- The research is “grandfathered” under the HIPAA rules

# *HIPAA De-Identification*

- The PHI is de-identified: either through removal of all HIPAA identifiers (the “safe harbor” method) or by certification of a statistical expert
- Is genetic information Protected Health Information (PHI)?
  - Genetic information is “health information”
  - Health information is PHI if it is “individually identifiable information”: identifies the individual or “there is a reasonable basis to believe the information can be used to identify the individual”
  - Office for Civil Rights (OCR) has concluded that not all genetic information is “individually identifiable,” but has not provided guidance on when genetic information is individually identifiable
  - Common interpretation: genetic information is not PHI unless it is accompanied by HIPAA identifiers or unless you know recipient has the ability to link the genetic information to a person’s identity

# The Revised Common Rule

- Applies to federally-funded research in the US
- Effective date 1/19/19 (except for single IRB for collaborative research effective 1/20/20)
- Significant changes
  - Potential changes to “identifiability”
  - New HIPAA exemption
  - New requirements for informed consent
  - New exemption for research with “broad consent”
  - New exemption for publicly available information
  - New rule for preparing for research
  - New rule on single IRB for collaborative research

Federal Register/Vol. 32, No. 12/Thursday, January 16, 2017/Rules and Regulations 7149		
DEPARTMENT OF HOMELAND SECURITY	NATIONAL SCIENCE FOUNDATION	32. Definitions for Purposes of This Policy (S. .... 110)
6 CFR Part 85	45 CFR Part 980	IV. Learning to Compliance with This Policy (S. .... 110)
DEPARTMENT OF AGRICULTURE	DEPARTMENT OF TRANSPORTATION	V. Concept Research (S. .... 110)
7 CFR Part 1c	49 CFR Part 15	VI. Protection of Identifiable Private Information and Other Data (S. .... 110)
DEPARTMENT OF ENERGY	Federal Policy for the Protection of Human Subjects	VII. IRB Membership and Modification to Submissions to Vulnerability (S. .... 110)(a), (b), (c), (d), (e), and (f) (S. .... 110)(g)
10 CFR Part 785	AGENCY: Department of Homeland Security; Department of Agriculture; Department of Energy; National Aeronautics and Space Administration; Department of Commerce; Social Security Administration; Agency for International Development; Department of Housing and Urban Development; Department of Labor; Department of Defense; Department of Education; Department of Veterans Affairs; Environmental Protection Agency; Department of Health and Human Services; National Science Foundation; and Department of Transportation.	VIII. IRB Functions and Operations (S. .... 110)
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION	ACTION: Final rule.	IX. IRB Review of Research (S. .... 110)
14 CFR Part 1230		X. Expedited Review Procedures (S. .... 110)
DEPARTMENT OF COMMERCE		XI. Criteria for IRB Approval of Research (S. .... 111)
15 CFR Part 27		XII. Cooperative Research (S. .... 111)
SOCIAL SECURITY ADMINISTRATION		XIII. IRB Review (S. .... 111)
30 CFR Part 831		XIV. General Requirements for IRB Approval (S. .... 111)
AGENCY FOR INTERNATIONAL DEVELOPMENT		XV. Documentation of IRB Approval (S. .... 111)
22 CFR Part 226		XVI. Applications and Proposals Lacking Definite Plans for Implementation of Human Subjects (S. .... 111)
DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT		XVII. Research Under Review Without the Submission of Identifiable Human Subjects (S. .... 111)
36 CFR Part 68		XVIII. Conditions (S. .... 111)
DEPARTMENT OF LABOR		XIX. Regulatory Impact Analysis
38 CFR Part 21		XX. Environmental Impact
DEPARTMENT OF DEFENSE		XXI. Paperwork Reduction Analysis
32 CFR Part 219		XXII. Tribal Consultation, Statement
DEPARTMENT OF EDUCATION		Final Regulatory Text
36 CFR Part 97		Executive Summary
DEPARTMENT OF VETERANS AFFAIRS		Purpose of the Regulatory Action
38 CFR Part 16		Individuals who are the subjects of research may be asked to contribute their time and assume risk to advance the research enterprise, which benefits society at large. U.S. Federal regulations governing the protection of human subjects in research have been in existence for more than three decades. The Department of Health, Education, and Welfare first published regulations for the protection of human subjects in 1974, and the Department of Health and Human Services (HHS) revised them in the early 1980s. During the 1990s, HHS began a process that eventually led to the adoption of a revised version of the regulations by 15 U.S. Federal departments and agencies in 1999. The purpose of this effort was to promote uniformity, understanding, and compliance with human subjects protections as well as to create a uniform body of regulations across Federal departments and agencies (hereinafter “A” of 48 Code of Federal Regulations (CFR) part 46), often referred to as the “Common Rule” or “Protection of Human Subjects Regulations.” These regulations were last amended in 2009, and have remained unchanged until the issuance of this final rule.
ENVIRONMENTAL PROTECTION AGENCY		
40 CFR Part 26		
DEPARTMENT OF HEALTH AND HUMAN SERVICES		
45 CFR Part 46		
FOR DOST-1820		

# *“Identifiability” May Change over Time*

- Requires agencies to assess within one year of final rule whether there are technologies or techniques that should be considered to generate identifiable private information, even if not accompanied by traditional identifiers (such as whole genome analysis)
- May widen difference in interpretation of “non-identified” information under Common Rule (i.e., investigator cannot readily ascertain identify of research participants) and “de-identified” under HIPAA

# The New York Times

July 23, 2019, Gina Kolata

## Your Data Were ‘Anonymized’? These Scientists Can Still Identify You

Computer scientists have developed an algorithm that can pick out almost any American in databases supposedly stripped of personal information.




### ARTICLE

<https://doi.org/10.1038/s41467-019-10933-3>

OPEN

## Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher <sup>1,2,3</sup>, Julien M. Hendrickx<sup>1</sup> & Yves-Alexandre de Montjoye<sup>2,3</sup>



# *California Consumer Privacy Act*

- Cal. Civil Code 1798.100-1798.199
- New proposed regulations published October 11, 2019 at <https://www.oag.ca.gov/privacy/ccpa>
- Applies only to certain for-profit entities
- CCPA's definition of de-identified data is not harmonized with the HIPAA standards
- CCPA's existing clinical trial exemption does not apply non-interventional data-based research

# *European Union General Data Protection Regulation (GDPR)*

- Jurisdictional reach:
  - Applies to organizations “established” (with a physician location) within the European Economic Area (EEA)
  - Applies to organizations outside the EEA that offer goods or services to data subjects within the EEA (clinical trial recruitment) or monitor the behavior of data subjects within the EEA (collection of research data from research participants)
- Applies to EEA research collaborator transfer of “Personal Data” to the United States

# GDPR

- Any data that directly or indirectly identifies a living person (not just patients)
  - Name, identification number, location data, online identifiers, factors specific to the physical, psychological, genetic, mental, economic, cultural or social identity
- More sensitive data have special protection
  - Genetic data, biometric data for the purpose of creating unique identification, data concerning health, data regarding race, religion, politics, sex
- Treatment of de-identified data
  - No de-identification “safe harbor” – data is “anonymized” if under a “facts and circumstances” test, the data cannot be identified by any means “reasonably likely to be used ... either by the controller or by another person”
  - “Pseudonymised” (coded) still personal data

# GDPR

- GDPR requires a legal basis for “processing” data
  - Consent;
  - Necessary for compliance with a legal obligation of “controller”;
  - **Necessary for purposes of the “legitimate interests” of the controller (which includes research);** or
  - Other provisions not generally relevant in the healthcare setting
- GDPR requires additional legal basis for processing special categories of sensitive data
  - Explicit consent;
  - Necessary for preventive or occupational medicine, medical diagnosis, the provision of health or social care or treatment;
  - Necessary for public health;
  - **Necessary for scientific research;** or
  - Other provisions not generally relevant in the healthcare setting

# GDPR

- Requirements for transfer of personal data from the EEA to the US may apply to the sender:
  - Consent (and advising data subjects of the risks of transfer to the US);
  - Contract that contains model contractual clauses approved by the European Commission (which impose some GDPR requirements on receiving entity);
  - To US for-profit entities that have been certified under the EU-US “Privacy Shield”; or
  - Pursuant to codes of conduct by associations