Clinical Research and Data: HIPAA, the Common Rule, the General Data Protection Regulation, and Data Repositories

Amy Jurevic Sokol, Associate General Counsel
The University of Kansas Medical Center

First, some context. My first computer was an Apple IIe. It had a 1.023MHz CPU, 64KB of RAM and booted off of floppy disks with a whopping 360KB capacity. For its time it was amazing; however, there was no such thing as "big data" thirty-four years ago. That computer, which I loved so much, would've failed miserably if given the task of crunching "big data" numbers. Fast forward to today. My iPhone has 128GB and my computer has a terabyte of storage. That terabyte is 8000GB, or to put that into context: more than 8000 times the storage capacity of that Apple IIe. In short, "big data"—and the machines required to process it—are now a reality.

The inexorable march of Moore's Law has resulted in changes in all areas of our lives, including how we do clinical research. Researchers and patients are more connected. We store, access, and manipulate data in different ways; we conduct studies in multiple countries sharing data and samples around the world; and cybersecurity and hacking are a reality. This article touches on different legal aspects arising at the intersection of technology, data, and clinical research—specifically HIPAA (the Health Insurance Portability and Accountability Act), human subjects research, the European data law (the General Data Protection Regulation), and data repositories. It attempts to explain how two different law-making bodies, the US and the EU, have tried to balance the necessity of using data for research purposes that benefit society with the privacy issues and risks of that same data.

Health Insurance Portability and Accountability Act (HIPAA) and Human Subjects Research

The US has a patchwork of federal and state laws that protect different types of data. Student records are protected by Family Educational Rights and Privacy Act (FERPA); financial data by the Gramm-Leach-Bliley Act (GLB); health data by the Health Insurance Portability and Accountability Act of 1996 (HIPAA); human subjects research by the Common Rule and FDA Regulations; and personal information, like driver's license number and social security number, by state law. There is no one overarching law that protects all data. Instead, it is a patchwork of laws that sometimes overlap and at other times have large holes of information that is not covered.[1]

The Federal Policy for the Protection of Human Subjects, or the Common Rule, outlines the basic provisions for Institutional Review Boards (IRBs), informed consent, and assurances of compliance.

Generally, it applies to research involving human subjects[2] conducted, supported or otherwise subject to regulation by one of eighteen different federal departments or agencies.[3] A different set of regulations apply to clinical investigations that are regulated by the Food and Drug Administration (FDA) or that support applications for research or marketing permits for products regulated by the FDA.[4] This includes sponsored trials of drugs and devices. Both the Common Rule and FDA Regulations require IRBs to, where appropriate, verify that there are adequate provisions in place to protect the privacy of subjects and to maintain the confidentiality of data.[5] The Common Rule requires that the informed consent form contain a statement describing the extent, if any, to which the confidentiality of records identifying the subject will be maintained.[6]

The Common Rule does not apply to public records or records in which the research subject cannot be identified directly or indirectly linked to the research subject.[7] So, if the information cannot be linked back to the subject, under the Common Rule it does not constitute human subjects research.[8] However, under the FDA regulations it may still be considered a clinical investigation.

HIPAA applies a much different standard than the Common Rule and FDA Regulations. HIPAA applies to "covered entities," which are defined as health care providers that transmit any information in an electronic form in association with standard transactions, health plans, and health care clearing houses.[9] It does not apply to all researchers; it applies to researchers that are covered entities and may apply, depending on the situation, to researchers who work for covered entities or obtain their data from a covered entity.[10] For instance, if a data repository is created by an academic medical center or health system, then HIPAA likely applies. However, if a group of individuals or a disease foundation create a data repository by submitting their own data, HIPAA likely does not apply.

Many researchers believe if they remove the patient's name and social security number they have de-identified data under HIPAA. These researchers would be incorrect. For information to be considered de-identified it has to meet the requirements of either the safe harbor or expert determination. The safe harbor requires removal of the following identifiers of the patient and the patient's relatives, employers, or household members:

- Names;
- All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census:
    - The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and
    - The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000;

- All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older;
- Telephone numbers;
- Vehicle identifiers and serial numbers, including license plate numbers;
- Fax numbers;
- Device identifiers and serial numbers;
- Email addresses;
- Web Universal Resource Locators (URLs);
- Social security numbers;
- Internet Protocol (IP) addresses;
- Medical record numbers;
- Biometric identifiers, including finger and voice prints;
- Health plan beneficiary numbers;
- Full-face photographs and any comparable images;
- Account numbers;
- Any other unique identifying number, characteristic, or code; and
- Certificate/license numbers.[11]

To fit within the safe harbor method all the identifiers above have to be removed, encoded, or randomized; no exceptions.[12] In addition, the researcher cannot have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.[13] It is important to note that the safe harbor method does not require removal of the physician or other health care provider's information, but only the patient's and family members' information.

Another option is the expert determination method. Under this method, a qualified statistician determines that the risk is very small that the information could be used alone or in combination with other reasonably available information by an anticipated recipient to re-identify an individual. In addition, the expert must document the methods and analysis that support this decision.[14] This method *could* allow certain identifiers to remain that would otherwise have to be removed under the safe harbor method of compliance.

If the data is de-identified under either the safe harbor or expert determination method, then HIPAA and its associated regulations do not apply to that data, and there are no limitations under HIPAA on its use or disclosure. For example, you may sell de-identified data (unless it is subject to a DUA or other agreement that prohibits it).

Researchers often need information that is not available in properly de-identified data sets. The most common request is for dates—often birth, death, admission, and discharge. In this instance, a researcher would use a *limited data set* (LDS), instead of fully identified information. A LDS is information from which the following identifiers of the individual or his or her relatives, employers or household members are removed:

- names;
- street addresses (other than town, city, state and zip code);
- telephone numbers;

- fax numbers;
- e-mail addresses;
- Social Security numbers;
- medical records numbers;
- health plan beneficiary numbers;
- account numbers;
- certificate license numbers;
- vehicle identifiers and serial numbers, including license plates;
- device identifiers and serial numbers;
- URLs;
- IP address numbers;
- biometric identifiers (including finger and voice prints); and
- full face photos (or comparable images).[15]

Examples of information that may remain and still be a LDS include:

- dates such as admission, discharge, service, birth, and death;
- city, state, five digit or more zip code; and
- ages in years, months, days, or hours.

An LDS is still considered protected health information even though there are fewer identifiers and less risk than fully identified protected health information.[16] The HIPPA Privacy Regulations require covered entities enter into a data use agreement with any recipient of a LDS. These agreements must include the following:

- a description of the permitted uses and disclosures of the limited data set;
- a list of who may use or receive the information;
- a requirement that the recipient will not use or further disclose the information, except as permitted by the agreement or as permitted by law;
- a requirement that the recipient use appropriate safeguards to prevent a use or disclosure that is not permitted by the agreement;
- a requirement that the recipient report to the covered entity any unauthorized use or disclosure of which it becomes aware;
- a requirement that the recipient ensure that any agents (including a subcontractor) to whom it provides the information will agree to the same restrictions as provided in the agreement; and
- a statement that the recipient will not re-identify the information or contact the individuals.[17]

A data use agreement has become a standardized agreement which is usually not a difficult agreement to negotiate. Increasingly, data use agreements are also being used for the disclosure of de-identified data, to prohibit the selling, re-disclosure, and noncompetitive use of de-identified data by the recipient.

Data Repositories

Creation of a Data Repository

A data repository is a collection or warehouse of data.[18] It can contain de-identified data, a limited data set, or fully identifiable information. There are two separate legal analyses that must occur when creating a data repository: first, does HIPAA apply; and second, is it considered human subject research under the Common Rule.

The HIPAA analysis starts with a seemingly simple concept: the use or disclosure of protected health information by a covered entity for research purposes requires that certain conditions under the HIPAA Privacy Rule be met. There is a lot of information packed in that one sentence. HIPAA applies to covered entities. So, if the researcher is not a covered entity, is not employed by a covered entity, and does not obtain the information from a covered entity then HIPAA and its associated regulations do not apply. Also, it only applies to protected health information. If the information is de-identified according to the HIPAA Privacy Rule (either by the safe harbor or expert determination method) then HIPAA no longer applies to the de-identified data. Finally, it must be for a research purpose. Research is defined as a "systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[19] If the use or disclosure is for health care operations of the covered entity then patient authorization is not required. Health care operations include conducting quality assessment and improvement activities, including outcomes evaluation and development of clinical guidelines, provided that obtaining generalizable knowledge is not the primary purpose; patient safety activities; and population-based activities relating to improving health or reducing health care costs, protocol development, case management and care coordination, and contacting of health care providers and patients with information about treatment alternatives.[20] Therefore, organizations could create a data repository for quality improvement activities without obtaining patient authorization or a waiver of authorization.

If a covered entity uses or discloses protected health information to create a database to conduct research, then the creation of the database itself is a research activity that must meet the requirements of the HIPAA Privacy Rule. If the data repository is created from a limited data set then one option is to use a data use agreement, which would enable the subsequent accessing of the data for research purposes to be achieved through a similar data use agreement.[21] Another option for creating the research repository is to obtain a HIPAA compliant authorization of each person's data that is contained in the repository. This may be a viable option for a small local database that is starting from scratch, but it is unlikely to work for anything but the smallest data repository. The more likely option is a waiver of the authorization requirement.

To create the research repository without obtaining each person's signed authorization, the researcher must get a written waiver of the authorization requirement from either an IRB or a privacy board that meets the Privacy Rule requirements.[22] The covered entity—prior to the use or disclosure—would obtain documentation of the following:

- Identification of the IRB or Privacy Board and the date on which the alteration or waiver of authorization was approved;
- A statement that the IRB or Privacy Board has determined that the alteration or waiver of authorization, in whole or in part, satisfies the three criteria in the Privacy Rule (listed below);

- A brief description of the protected health information for which use or access has been determined to be necessary by the IRB or Privacy Board;
- A statement that the alteration or waiver of authorization has been reviewed and approved under either normal or expedited review procedures; and
- The signature of the chair or other member, as designated by the chair, of the IRB or the Privacy Board, as applicable.[23]

The following three criteria must be satisfied for an IRB or Privacy Board to approve a waiver of authorization under the Privacy Rule:

- The use or disclosure of protected health information involves no more than a minimal risk to the privacy of individuals, based on, at least, the presence of the following elements:
    - an adequate plan to protect the identifiers from improper use and disclosure;
    - an adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law; and
    - adequate written assurances that the protected health information will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research project, or for other research for which the use or disclosure of protected health information would be permitted by this subpart;
- The research could not practicably be conducted without the waiver or alteration; and
- The research could not practicably be conducted without access to and use of the protected health information.[24]

At the same time as the waiver of authorization under HIPAA is being obtained, the IRB is also determining whether or not it considers the data repository to be "human subjects research" covered under the Common Rule.[25] The Common Rule does not apply if both of the following conditions are met: the data was not collected specifically for the currently proposed research project through an interaction or intervention with living individuals, and the investigator(s) cannot readily ascertain the identity of the individual(s) to whom the coded private information pertains.[26] Conversely, obtaining identifiable private information for research purposes constitutes human subjects research.[27]

If the IRB determines that the Common Rule applies, then: (1) either the researcher will get consent from each research subject; or (2) in addition to the HIPAA waiver of authorization, the researcher will also request waiver of informed consent under the Common Rule. To approve a waiver of the requirement to obtain informed consent or approve a

consent procedure which does not include, or which alters, some or all of the elements of informed consent requirements the IRB must determine and document the following:

- the research involves no more than minimal risk to the subjects;
- the waiver or alteration will not adversely affect the rights and welfare of the subjects;
- the research could not practicably be carried out without the waiver or alteration; and
- whenever appropriate, the subjects will be provided with additional pertinent information after participation. [28]

Several changes to the Common Rule go into effect July 19, 2018 that impact the informed consent process.[29] These revisions add a provision for secondary research uses of identifiable private information and identifiable biospecimens; [30] add a provision for broad consent for the storage, maintenance and secondary research use of identifiable private information and identifiable biospecimens; [31] modify the list of required elements of informed consent to include an additional statement if private identifiable information or identifiable biospecimens are collected[32] and additional language for the consent form if biospecimens (even if identifiers are removed) will be used for commercial profit[33] or if research of biospecimens will or might include whole genome sequencing; [34] and a provision approving research where a researcher obtains information or biospecimens without consent for the purpose of screening, recruiting, or determining eligibility of a prospective research subject if certain conditions are met. [35]

Once all of the approvals are obtained, the data repository may be populated with data. However, if new data sources are added, data elements collected are changed, or the protocol is revised, then the researcher would have to do the HIPAA analysis again. For instance, if the repository was originally considered to be de-identified or a limited data set, is that still the case after additional data is added? Also the IRB would want to review any modifications to the data repository or associated protocol prior to their implementation.

Accessing Data in a Data Repository

Each time protected health information (even a limited data set) is accessed for a research purpose, then the requirements for access must be met as well. Like the creation of a data repository, this is potentially a two-part analysis. The first part is the analysis under HIPAA; the second, an analysis under the Common Rule.

Under HIPAA, to access de-identified data no additional steps are required unless as part of the protocol that created the data repository the researcher stated that a data access or system access agreement would be signed by researchers accessing the data. In this case, the protocol must be followed. If a limited data set or fully identifiable protected health information is requested, then under the HIPAA Privacy Rule one of the following circumstances and conditions must be met:

- The request is a review preparatory to research and certain representations are obtained from the researcher;
- The research is solely on decedents' information and certain representations are obtained from the researcher;
- A HIPAA-compliant authorization was signed by each subject of the PHI, granting specific written permission for the access and use of the information;
- An IRB or Privacy Board has granted and documented the grant of a waiver or an alteration of the authorization requirement (if an alteration of the authorization is granted, a signed authorization is required from each individual);
- The PHI has been de-identified in accordance with the standards set by the Privacy Rule (in which case, the information is no longer PHI);
- The information is released in the form of a limited data set and a data use agreement between the researcher and the covered entity is signed;
- Informed consent of the individual to participate in the research, an IRB waiver of such informed consent, or other express legal permission to use or disclose the information for the research is grandfathered by the transition provisions. [36]

The Common Rule analysis is just as complicated as the HIPAA analysis, maybe even more so. Under the Common Rule, obtaining identifiable private information for research purposes constitutes human subjects research. This includes information that is already in the possession of the investigator. [37] Conversely, research is not considered human subject research if the research only involves coded private information of human subjects if both of the following conditions are met: (1) the private information was not collected specifically for the proposed research project through an interaction or intervention with living individuals; and (2) the investigator(s) cannot readily ascertain the identity of the individual(s) to whom the coded private information pertains. [38]

Information is considered identifiable when the information can be linked to specific individuals by the investigator(s) either directly or indirectly through coding systems. Additionally, an investigator is broadly defined to include anyone involved in conducting the research. [39] An investigator would not include an honest broker who solely provides de-identified or coded information as long as the honest broker does not collaborate on other activities related to the conduct of this research with the investigator(s) who receive(s) such information.[40] An example of the difference between these two scenarios in plain English is: First, a researcher accesses a database with patient information that identifies the patient, queries the information, and records the results in a way that is coded but maintains the key to decode the data. Or, second, an honest broker, who is not part of the research team, accesses the database, queries the data, codes the results so that the patients are not identifiable, and the honest broker maintains the linking code.

The first example is human subjects research, while the second is not.

Even if the accessing of data is considered human subjects research, it may be exempt under the Common Rule. According to the guidance published by the Office for Human Subjects Protections, the most relevant exemption is if the information obtained by the investigator is recorded in such a manner that the individuals cannot be identified directly or indirectly through identifiers linked to the individuals.[41] Using our previous examples, the researcher would access a database with patient information that identifies the patient, would query the information, and would record the results in a way that is coded with no way to decode the data to identify the individuals.

If the access to the research repository is considered human subjects research that is not exempt, then the investigator must submit a study submission as a new study or a modification to an existing study. In addition, the researcher is likely to request a waiver of the informed consent requirement as well.

The European Union and the General Data Protection Regulation

This final section is provided for illustrative purposes in order to highlight some of the issues that may arise when US researchers want to use data from other countries for their research. I have used the EU as an example. As this article has shown, the US has a complex maze of laws with limitations and exceptions which often makes researchers and their attorneys want to scream in frustration. The EU data protection laws are just as complex, but unlike the US, the EU has a single data protection regime that applies to all data. The previous law was the Data Protection Directive, [42] which is being replaced (effective May 25, 2018) by the General Data Protection Regulation (GDPR).[43] Like HIPAA, the GDPR has sanctions[44] and special rules for personal data breaches. [45] Unlike HIPAA, it also includes a right to be forgotten, or *erasure*, [46] and there are provisions for the portability of data. [47]

The GDPR was designed to harmonize the different data privacy laws across Europe, give all EU citizens control of how their data is used and protected, and to reshape the way organizations across the region approach data privacy. It is intended to reach beyond the territory of the EU to individuals and businesses that offer goods and services to or monitor the behavior of individuals in the EU. [48] This impacts US researchers because US researchers are increasingly wanting to use EU and other countries' data and are thus dragged into the quagmire that is international data protection law. This section provides an overview of the GDPR. What remains to be seen is what impact the GDPR will have on data repositories, big data research, and personal data in "the cloud" as it evolves over time, especially with the large potential sanctions. Since the GDPR is not yet in effect, it has yet to be interpreted by courts, researchers, and lawmakers throughout Europe.

The GDPR applies to "personal data," which is defines as:

> [A]ny information relating to an identified or identifiable natural person ("data subject"); an identifiable natural

> person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. [49]

This broad definition is likely a moving target. An identifiable person defined as someone who can be identified indirectly will change over time as more information becomes publicly available and as technology changes.[50]

Under the GDPR there is not a "de-identified data" safe harbor or expert determination, but instead data can be anonymized and pseudonymised. When data is anonymized, it is no longer personal data because the individual cannot be identified either directly or indirectly.[51] Pseudonymisation is defined by the GDPR as processing personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information. The GDPR not only requires the "additional information" be stored separately, but also requires various technical and organizational measures to ensure that the personal data cannot be attributed to an identified or identifiable natural person.[52] An example of pseudonymisation is encryption, which renders the original data unintelligible and the process cannot be reversed without access to the correct decryption key. The GDPR requires that this additional information (the decryption key) be kept separately from the pseudonymised data. Although the GDPR encourages the use of pseudonymisation to reduce risks to the data subjects, pseudonymised data is still considered personal data and, therefore, remains covered by the GDPR. [53]

The GDPR also recognizes special categories of personal data which are considered to be particularly sensitive. The processing (or use) of data related to these special categories is generally prohibited by the GDPR:

> Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited." [54]

There are two exceptions to this prohibition that are likely to apply to research involving the special categories of data, which includes health data or protected health information under HIPAA. The first exception requires the data subject give *explicit* consent to the processing of the personal data for one or more specified purposes. [55] The GDPR defines consent of the data subject as "any freely

given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." [56] If the consent is provided in a document that concerns other matters, the request for consent must be presented in a way that is clearly distinguishable from the other matters, in an intelligible and easily accessible form, and using clear and plain language. [57] In addition, consent should not be considered freely given if the data subject had no genuine or free choice or is unable to refuse or withdraw consent without penalty.[58] However, the GDPR also recognizes that it is often not possible to fully identify the purpose of personal data processing for research purposes at the time of data collection; therefore, data subjects should be allowed to provide their consent to certain areas of research or parts of research projects when in keeping with recognized ethical standards for scientific research.[59] If a researcher wants to reuse the data for research (e.g., secondary research) and does not have explicit consent for the secondary use, then the researcher would decide if the use was comparable to the previously consented use. The "compatibility test" looks at the following factors:

- any link between the purpose(s) for which the personal data was collected and the purpose(s) of the intended use;
- the context in which the personal data was collected, in particular the relationship between data subjects and the controller;
- the nature of the personal data, in particular are there special categories of personal data or personal data related to criminal convictions and offences;
- the possible consequences of the intended further processing for data subjects; and
- whether there are appropriate safeguards. [60]

The second exception is for research purposes that meet the requirements for applicable safeguards outlined in Article 89(1)[61] and based on EU or an EU country's law that is proportionate to the research aim pursued, respect the right to data protection, and provide for suitable and significant measures to safeguard the fundamental rights and interests of the data subject.[62]

For consent to be "waived" the research should have adequate safeguards in place to protect the data subject's information and have a valid research purpose. For secondary research the secondary use must meet the requirements for applicable safeguards outlined in Article 89(1) and be compatible with the initial purpose for collecting the data ("purposes limitation").[63] If personal data has not been provided by the data subject (e.g., secondary research), then unless the exception for research is met, the data subject should be provided with the following: the identity and the contact details of the controller and where applicable the data protection officer and the controller's representative; the purposes for processing the data and the legal basis for the processing; the categories of personal data concerned; the data recipients or categories of data recipients; and

where applicable, that the controller intends to transfer personal data to a recipient in a third country or international organization, whether there is an adequacy decision by the Commission, or if the transfer is made subject to appropriate safeguards.[64] The research exception for this requirement is if the provision of the information would be impossible or involve a disproportionate effort and likely render impossible or seriously impair the research. Then, subject to the conditions and safeguards referred to in Article 89(1), the researcher must take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.[65]

Conclusion

Research and technology are moving forward at an incredible pace. Technology has enabled researchers to store, manipulate and calculate data in new ways, which has created benefits and risks for researchers and data subjects. The US has a patchwork of laws to address the use of data in clinical research and data repositories, but there are some gaps. Further, as the world of research has gotten smaller and data is shared around the globe, "big data" and research has never been more complicated. US laws like HIPAA and the Common Rule complement and contradict other laws like the EU GDPR. So, researchers who use data from multiple countries must navigate not only their own country's laws but also international legal waters often without a clear path.

Endnotes

[1] An example of a data repository falling though a gap in the patchwork of laws is a commercial pharmaceutical company that has a data or biospecimen repository that uses the repository for internal non-funded research, which would not be subject to HIPAA (not a covered entity), the Common Rule (not funded or supported by one of the eighteen Federal agencies or department), or the FDA Regulations (not submitted to the FDA).

[2] A human subject is currently defined as "A living individual about whom a researcher: (1) Obtains data through intervention or interaction with the individual; or (2) Obtains identifiable private information." 45 CFR § 46.102(f). In the revised Common Rule a human subject is defined as "A living individual about whom a researcher: (1) Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (2) Obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens." Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7260 (Jan. 19, 2017) (to be codified at 45 CFR § 46.102(e)).

[3] 45 CFR § 46.101(a); *see also* Department of Health and Human Services, *Federal Policy for the Protection of Human Subjects ("Common Rule")* https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html.

[4] 21 CFR § 50.1(a).

[5] 21 CFR § 56.111(a)(7) and 45 CFR § 46.111(a)(7). One of the changes to the Common Rule was the statement that the Secretary of the Department of Health and Human Services after consulting with the Office of Management and Budget's privacy office and other Federal department and agencies that have adopted the Common Rule will issue guidance to assist IRBs in assessing adequate provisions to protect the privacy of research subjects and the confidentiality of data. Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7264 (Jan. 19, 2017) (to be codified at45 CFR § 46.111(a)(7)(i)).

[6] 45 CFR § 46.116(a)(5).

[7] 45 CFR § 46.101(b)(4).

[8] *Id.*

[9] 45 CFR § 160.103.

[10] Department of Health and Human Services, *Research Repositories, Databases, and the HIPAA Privacy Rule*, (January 2004), https://privacyruleandresearch.nih.gov/research_repositories.asp.

[11]45 CFR 164.514(b). *See also* Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, November 26, 2012, https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/#_edn1.

[12] Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, November 26, 2012, https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/#_edn1.

[13] 45 CFR § 164.514(b)(2)(ii).

[14] 45 C.F.R. § 164.514. *See also* Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/#_edn1.

[15] 45 CFR 164.514 § (e)(2).

[16] National Institutes of Health, *How Can Covered Entities Use and Disclose Protected Health Information for Research and Comply with the Privacy Rule?*, February 2, 2007, https://privacyruleandresearch.nih.gov/pr_08.asp.

[17] 45 CFR § 164.514(e)(4).

[18] For a description of types of data repositories, *see generally*, Amy Jurevic Sokol, *Big Issues in Big Data: Considerations for Research within Large Networks*, AHLA CONNECTIONS, August 2017, at 34.

[19] 45 CFR § 164.501.

[20] 45 CFR § 164.501

[21] Department of Health and Human Services, *Research Repositories, Databases, and the HIPAA Privacy Rule*, (January 2004), https://privacyruleandresearch.nih.gov/research_repositories.asp.
[22] 45 CFR § 164.512(i)(1)(i).
[23] 45 CFR § 164.512(i).
[24] 45 CFR § 164.512(i)(2)(ii).
[25] In addition, the IRB may determine in specific situations that the FDA Rules apply. Previously, the FDA did not have the statutory authority to permit an IRB to waive the informed consent requirements; however, an amendment to the Federal Food, Drug and Cosmetic Act has provided FDA with authority to permit an exception from informed consent for minimal risk clinical investigations when specific criteria are met. Food and Drug Administration, *IRB Waiver or Alteration of Informed Consent for Clinical Investigations Involving no More than Minimal Risk to Human Subjects, Guidance for Sponsors, Investigators, and institutional Review Boards*, July 2017, https://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM566948.pdf.
[26] 45 CFR § 46.102; *also see* Department of Health and Human Services, Office for Human Research Protections, *Guidance Coded Private Information or Specimens Use in Research*, October 16, 2008, https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html.
[27] 45 CFR § 46.102(f) (private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects); *see also* Department of Health and Human Services, Office for Human Research Protections, *Guidance Coded Private Information or Specimens Use in Research*, October 16, 2008, https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html.
[28] 45 CFR § 46.116(d).
[29] 45 C.F.R. § 46.101 (l)(4).
[30] Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149; 7266 (Jan. 19, 2017) (to be codified at 45 CFR § 46.104(d)(4)).
[31] Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149; 7266 (Jan. 19, 2017) (to be codified at 45 CFR § 46.116(d)). It is important to note that if an individual has declined to provide broad consent for the storage, maintenance, and secondary research use of identifiable private information or biospecimens, an IRB may not waive consent. *Id.*
[32] *Id.* One of the following must be included: "(i) A statement that identifiers might be removed from the identifiable private information or identifiable biospecimens and that, after such removal, the information or biospecimens could be used for future research studies or distributed to another investigator for future research studies without additional informed consent from the subject or the legally authorized representative, if this might be a possibility; or (ii) A statement that the subject's information or biospecimens collected as part of the research, even if identifiers are removed, will not be used or distributed for future research studies." *Id.*
[33] Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7266 (Jan. 19, 2017) (to be codified at 45 CFR § 46.116(c)(7)). The statement must also include whether the research subject will share in the commercial profit.
[34] Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7266 (Jan. 19, 2017) (to be codified at 45 CFR § 46.116(c)(9)).
[35] Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7267 (Jan. 19, 2017) (to be codified at 45 CFR § 46.116(g)). One of the following "(1) The investigator will obtain information through oral or written communication with the prospective subject or legally authorized representative, or (2) The investigator will obtain identifiable private information or identifiable biospecimens by accessing records or stored identifiable biospecimens." *Id.*
[36] National Institutes of Health, *Research Repositories, Databases, and the HIPAA Privacy Rule* (January 12, 2004), https://privacyruleandresearch.nih.gov/research_repositories.asp.
[37] Office for Human Research Protections, *Guidance on Research Involving Coded Private Information or Biological Specimens* (October 16, 2008), https://archive.hhs.gov/ohrp/humansubjects/guidance/cdebiol.htm.
[38] *Id.*

[39] *Id.*
[40] *Id.*
[41] 45 CFR § 46.101(b)(4).
[42] Council Regulation (EU) 95/46 of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281).
[43] Council Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the processing of Personal Data and on the Free Movement of Such Data, and repealing Directive 95/46/EC, (General Data Protection Regulation), 2016 O.J. (L 199) [hereinafter GDPR].
[44] The following sanctions may be imposed: (1) a warning in writing in cases of first and non-intentional non-compliance; (2) regular periodic data protection audits; (3) a fine up to 10,000,000 EUR or up to 2% of the annual worldwide turnover of the preceding financial year in case of an enterprise, whichever is greater; or (4) a fine up to 20,000,000 EUR or up to 4% of the annual worldwide turnover of the preceding financial year in case of an enterprise, whichever is greater. *Id.* at art. 83.
[45] A personal data breach means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed. *Id.* at art. 4(12). Under the GDPR, the Data Controller has a legal obligation to notify the Supervisory Authority without undue delay (within 72 hours) and the reporting requirement not subject to a de minims standard. I*d.* at art. 33. Individuals must be notified if adverse impact is determined, Individuals do not have to be notified if anonymized data is breached or personal data is protected by pseudonymisation techniques like encryption with adequate technical and organizational protection measures. *Id.* at art. 34.
[46] *Id.* at art. 17; there is a research exception to this right as well: "[t]he right to be forgotten will not apply to the extent it is likely to render impossible or seriously impair the achievement of the objectives of the research. The protections articulated in Article 89(1) must be met." *Id.* at art. 17(3(d).
[47] *Id.* at art. 20.
[48] *Id.* at art. 3(2).
[49] *Id.* at art. 4 (1).
[50] John Mark Michael Rumbold & Barbara Pierscionek, *The Effect of the General Data Protection Regulation on Medical Research*, 19 J. MED. INTERNET RES. (February 2017) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5346164/#ref21 *citing* Anonymisation: Managing Data Protection Risk Code of Practice, London: Information Commissioner's Office, (2012), https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/webcite. According to various studies the likelihood of re-identification is high if the researcher has the names, zip code and date of birth (Montreal, Canada almost 98% (Khaled El Emam et al., *The Re-identification of Canadians from Longitudinal Demographics*, BMC MED. INFORMATICS & DECISION MAKING (2011), https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-11-46); the Netherlands more than 99% (*id.*); and United States approximately 87 % (Latanya Sweeney, *Simple Demographics Often Identify People Uniquely*, Carnegie Mellon University, Data Privacy Working Paper 3 (2000), https://dataprivacylab.org/projects/identifiability/paper1.pdf)).
[51] GDPR, *supra* note 43, pmbl. 26.
[52] *Id.* at art. 4 (5).
[53] *Id.* at pmbl 26.
[54] *Id.* at art. 9 (1).
[55] *Id.* at art. 4 (2)(a) (unless the EU or the member law provides that the prohibition may not be consent to by the data subject). *Id.*
[56] *Id.* at art. 4(11).
[57] *Id.* at art. 7(2).
[58] *Id.* at pmbl. 42.
[59] *Id.* at pmbl. 33.
[60] *Id.* at pmbl. 50.
[61]"Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for

the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in L 119/84 EN Official Journal of the European Union 4.5.2016 order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner." *Id.* at art. 89(1).

[62] "[P]rocessing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject." *Id.* at art. 9(2)(j).

[63] *Id.* at art. 5(b).

[64] *Id.* at art. 14(1).

[65] *Id.* at art. 14 (5)(b).