DATA INNOVATION PROJECT

Issue Brief: What Data-Informed Organizations Need to Know About Protected Health Data

Working with protected health data must be done thoughtfully but it is not insurmountable.

Framing the Challenge

Non-profit organizations are increasingly encouraged to incorporate strategic decision-making that is grounded in robust data. When trying to access data regarding health outcomes, however, researchers and data stakeholders often come up against many challenges while navigating the requirements of the Health Insurance Portability and Accountability Act of 1996, commonly referred to as HIPAA. The goal of this issue brief is the help researchers and data stakeholders understand what HIPAA actually says about privacy, what data are actually protected, solutions for accessing protected information, and where to go for additional guidance.

What is HIPAA?

The Health Insurance Portability and Accountability Act of 1996 (HIPAA), Public Law 104-191, was passed by Congress in 1996 and required the United States Department of Health and Human Services (US DHHS) to adopt national standards for electronic health care transactions and code sets, unique health identifiers, and security. HIPAA was a broad law encompassing security, breach notification, compliance, enforcement and patient safety. Recognizing that the electronic aspects of the law could erode the privacy of health information, the act also contained federal privacy protections for individually identifiable health information.

The privacy rule established national standards to protect individually identifiable health information at ***covered entities***,[1] which are as follows:

1. Health care providers who conduct the standard health care transactions electronically (organization or individual who furnishes, bills, or is paid for health care[2] in the normal course of business);
2. Health care clearinghouses (public or private entity, including a billing service, repricing company, community health management information system or community health information system); and
3. Health plans (an individual or group plan that provides or pays the cost of medical care).

Also called ***protected health information (PHI)***, this refers to any information (electronic, paper or oral) created by a covered entity that relates to the following:

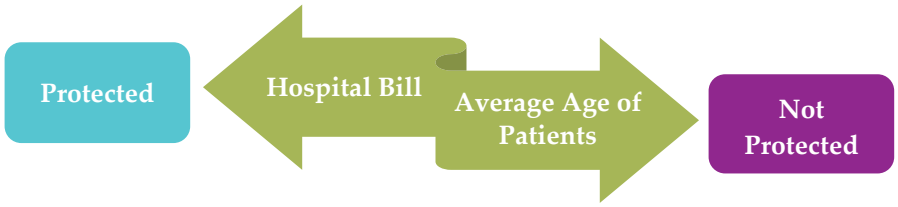
1. The past, present, or future physical or mental health or condition of an individual;
2. The provision of health care to an individual; or
3. The past, present, or future payment for the provision of health care to an individual that identifies the individual (or can be used to identify the individual).

[1] The rules also apply to ***business associates***: a person or entity who, on behalf of a covered entity, performs a function or activity involving the use or disclosure of individually identifiable health information.

[2] Care, services, or supplies related to the health of an individual, including (1) preventive, diagnostic, therapeutic, rehabilitative, maintenance, or palliative care, and counseling, service, assessment, or procedure with respect to the physical or mental condition, or functional status, of an individual that affects the structure or function of the body; and (2) sale or dispensing of a drug, device, equipment, or other item in accordance with a prescription.

Accessing Aggregate Reports

Protected health information exists when we can link an individual to their personal health information. Identifying information alone, such as personal names, residential addresses, or phone numbers, is not designated as PHI. Nor is a report that only contains the average age of health plan members; although developed by aggregating information from individual plan member records, the report does not identify any individual plan members and there is no reasonable basis to believe that it could be used to identify an individual.



To be considered "de-identified," **aggregate counts must be sufficiently large to ensure that individuals cannot be identified.** This becomes more difficult when researchers or data stakeholders request counts broken out by demographics or other characteristics. For example, pretend a researcher or data stakeholder asked a provider for a report of all newly diagnosed breast cancer patients of a certain age within a certain zip code. If the provider released a report with a total count of one, then the diagnosis date, zip code, and age of a cancer patient has been disclosed. This challenge can be addressed by suppressing information that does not meet a minimum reportable count threshold. While a uniform threshold has not been established by HIPAA, statisticians consider three to be the absolute minimum needed to prevent disclosure, although larger minimums (e.g., 5 or 10) are typical.[3]

Accessing Protected Health Information (PHI)

A covered entity is permitted to use or disclose PHI for research purposes if it obtains the individual's Authorization for the disclosure.[4] However, health service providers often use large, population-level databases containing thousands (or even millions) of records. As a result, contacting individual subjects to ask for their Authorization prior to examining health services across an entire population is not usually possible. In this case, a covered entity can still release PHI to a researcher or data stakeholder it the following conditions are met:

- Receives satisfactory documentation of an Institutional Review Board (IRB)[5] or Privacy Board waiver or alteration of the Authorization requirement;
- Uses or discloses PHI for research solely on decedents';
- Uses or discloses information that is **de-identified** in accordance with the set standards; or
- Provides a **limited data set** and enters into a data use agreement with the recipient of PHI.

[3] Disclosure Avoidance, Privacy Technical Assistance Center (PTAC). http://ptac.ed.gov/sites/default/files/FAQs_disclosure_avoidance.pdf

[4] For more information on securing authorized disclosure, please see http://www.hhs.gov/hipaa/for-professionals/faq/authorizations

[5] A board, committee, or other group formally designated by an institution to review research involving humans as subjects.

De-identified Data

De-identifying removes the identifiers protected health information. This process *removes the* privacy risks to individuals and allows a third-party to use of data for research and continuous quality improvement. Once de-identification is achieved, the Privacy Rule does not restrict the use or disclosure the health information, as it is no longer considered protected.

There are two ways to de-identify protected data. The first is to have a qualified expert review the requested information and go through a formal statistical process to determine the risks associated with releasing the information. The second is simply to remove 18 specific individual identifiers (see sidebar). In the second scenario, the covered entity cannot release the information if it has knowledge that the recipient has a method by which to identify the individual when information is received.[6]

Limited Data

A limited data set contains information that is not directly identifiable, but may contain more identifiers than data that has been stripped of the 18 identifiers, such as: all elements of dates, town or zip code, or unique identifiers.

Prior to releasing a limited data set, however, a **data-use agreement** must be established between the two parties. The data-use agreement must state who is permitted to use or receive the limited data set, for what purpose, and require the recipient to adhere to the following:

- not use or disclose the information other than as permitted by the agreement or as otherwise required by law;
- use appropriate safeguards to prevent uses or disclosures of the information that are inconsistent with the data-use agreement;
- report to the covered entity any use or disclosure of the information, in violation of the agreement, of which it becomes aware;
- ensure that any agents to whom it provides the limited data set agree to the same restrictions and conditions that apply to the limited data set recipient with respect to such information; and
- not attempt to re-identify the information or contact an individual.

HIPAA Identifiers

De-identification can be achieved by removing the following elements.

1. Names
2. Street address, city, county, precinct, ZIP code
3. All elements of dates (**except year**) for dates that are directly related to an individual; all ages over 89 unless aggregated into a single category of age 90 or older.
4. Telephone numbers
5. Vehicle identifiers and serial numbers, including license plate numbers
6. Fax numbers
7. Device identifiers and serial numbers
8. Email addresses
9. Web Universal Resource Locators (URLs)
10. Social security numbers
11. Internet Protocol (IP) addresses
12. Medical record numbers
13. Biometric identifiers, including finger and voice prints
14. Health plan beneficiary numbers
15. Full-face photographs and any comparable images
16. Account numbers
17. Any other unique identifying number, characteristic, or code
18. Certificate/license numbers

[6] For example, the covered entity knows that the researcher has obtained another data file with identifying information to which the protected health records will be matched.

Important Considerations

As a researcher or data stakeholder, it is important to weigh the costs and benefits of pursuing protected health information. Ask yourself whether you actually need that detailed level of information, and what it will provide to your research or quality improvement activities. How much effort will it take to satisfy the protections laid out in this issue brief? It is also important to explore whether there is another way to get meaningful information without requesting protected information; for example, through aggregate reports that already exist in the electronic system (or that might need moderate tweaks), or by accessing existing surveillance data (e.g., public health repository or survey). If you do choose to pursue protected health information, only request the information that you need to answer your questions.

Conclusion

This issue brief has described the various avenues by which researchers and data stakeholders can access health records according the rules laid out by HIPAA. Electronic health records from health insurance plans, authorization clearinghouses, and service providers contain a wealth of information about the overall well-being of a community population. Although carefully protected, using this information to inform research and data-driven quality improvement efforts is not insurmountable. The key takeaways around this issue are as follows:

- Aggregate data and data stripped of individual identifiers are not covered by the Privacy Rule and require no individual privacy protections.
- PHI may be used and disclosed for research with an individual's written permission in the form of an Authorization.
- PHI may be used and disclosed for research without an Authorization in limited circumstances.

However, it is worth noting that other federal, state or local privacy regulations may dictate additional expectations to covered entities about how they handle patient records. Researchers and data stakeholders should work with covered entities to understand all the privacy requirements to which they are beholden.

References and Additional Resources

U.S. Department of Health and Human Services.
http://www.hhs.gov/hipaa/for-professionals/special-topics/index.html

U.S Department of Health and Human Services. National Institutes of Health. Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule.
https://privacyruleandresearch.nih.gov/pr_02.asp

U.S Department of Health and Human Services. National Institutes of Health. Health Services Research and the HIPAA Privacy Rule
https://privacyruleandresearch.nih.gov/healthservicesprivacy.asp