health care systems research network

HCSRN SHARED PRINCIPLES: SMALL CELL ISSUES

Purpose

To document agreed upon principles for complying with HIPAA privacy requirements when sharing summary level data (e.g. counts or rates) between HCSRN sites.

Examples of issues that may arise include:

- When do such "small cells" constitute a re-identification risk?
- When is use of a data use agreement appropriate?
- Who decides when a small cell issue constitutes a re-identification risk?

Relevant Regulations and Guidance

The US Department of Health & Human Services, as well as various state and local governments, universities, and law firms, have written guidance and interpretation of HIPAA regulations relevant to the handling of small cell sizes by researchers. Please see references for exact citations.

- In general, **HIPAA regulations treat data aggregation as a form of de-identification**. Once data is de-identified, it is not considered 'protected health information' (PHI) and so HIPAA does not restrict its use or disclosure. That is, if data is not PHI, it falls outside HIPAA's purview.
- However, **HIPAA regulations place a burden of trust on covered entities** to determine that the steps they have taken to de-identify data have provided sufficient protection. Specifically, Section 164.514 (a) of the HIPAA regulation states "[h]ealth information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information."

  Section 164.514(b) states that "[a] covered entity may determine that health information is not individually identifiable health information" if all identifiers enumerated in Section 164.514 (b) (2)(i)(A-R) are removed and if a scientific/statistical expert renders the information not identifiable and documents that the risk of re-identification is "very small".

  Similarly, Section 164.514(b)(2)(ii) requires that the "covered entity does not 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."
- **HIPAA regulations do not generally state specific cell sizes as thresholds for de-identification**; instead, researchers are advised to make determinations based upon the specific situation. Note that some guidance is provided for zip codes and ages > 89 years; see Section 164.514 (b)(2)(i)(B) and (C).

The federal guidance cited below outlines several factors that impact the likelihood of re-identification, including replicability (does this value occur consistently with this individual?), data source availability (can parts of these data be found in public sources?), distinguishability (is data for an individual unique?), and characteristics and capabilities of recipients (small group of recipients vs. general public; recipients' access to linkable data; recipients' data re-identification expertise; etc.).

The point at which a data set contains information that could be "used alone or in combination with other information to identify an individual who is a subject of the information" depends on the situation (e.g., rarity of the event, publicized clinical events, knowledge of the data source, and so on).

- **Federal and state guidelines offer steps for assessing risk**. The federal guidelines cited below give examples of what "actual knowledge" means. The State of New Hampshire's guidance describes an easy-to-apply "Rule of Ones" test to determine re-identification risk. Washington State offers very thoughtful guidance on how to determine and mitigate de-identification risk.
- **Covered entities are not obligated to use data use agreements when sharing aggregate data** because HIPAA regulations do not apply to de-identified data. However, covered entities may voluntarily choose to ask recipients of de-identified information to enter into a data use agreement, and they may model the data use agreement on HIPAA-required DUA's. Points of agreement most relevant to sharing of aggregate data would be a prohibition from attempting to re-identify data and a time limit at which the data should be destroyed.
- **The risk of re-identification of a given dataset may change over time.** Although the data remains the same over time, the environment changes. For instance, the recipient may gain access to additional datasets over time, some of which could be linked to the data in question. Also, computational capability is likely to improve over time so that re-identification becomes easier. While the Privacy Rule does not require researchers to put expiration dates on data, they may want to consider asking recipients to destroy data after a certain time period.
- **State regulations and local organizational policies may affect sharing of aggregate data.** State or local regulations may be more restrictive than federal guidelines. Researchers should make sure that their approach is consistent with all applicable regulations.

Guiding Principles

HCSRN members agree upon the following guiding principles:

- **Each HCSRN member organization is responsible for ensuring its own staff are:**
    - **Adequately familiar with federal guidance** regarding methods for de-identification of protected health information (PHI) in accordance with the HIPAA privacy rule.
    - **Adhering to their local center's process** for determining if/when a data use agreement is needed.
    - The HCSRN Key Contacts directory lists DUA contacts and signatories at each site. These staff can advise on local processes, as needed.
- **The Principal Investigator at each local HCSRN site is responsible for ensuring that appropriate local processes are followed** relating to re-identification risk and the need for a data use agreement.
- **Each HCSRN site is responsible for documenting** the method and determination of re-identification risk assessment. The HCSRN has developed a checklist for documentation of the expert assessment method for sites to use, if desired.
    - Specific responsibility for "expert determination" of risk of re-identification varies across HCSRN research centers (e.g. formal consultation with a privacy office representative may or may not be required).
    - Each investigator is responsible for understanding and following those local requirements. Refer to the HCSRN Key Contacts Directory for DUA staff that can advise on local requirements, if needed.

Expert Determination Method

HCSRN sites commonly use the Expert Determination Method for de-identification of research data. HIPAA regulations define an expert as a "person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable". HIPAA guidance clarifies that there is "no specific professional degree or certification program for designating who is an expert", and that "relevant expertise may be gained through various routes of education and experience. Experts may be found in the statistical, mathematical, or other scientific domains."

- HIPAA Guidance: Expert Determination Method
- HCSRN Checklist for documentation of expert assessment method (optional)

References

US Department of Health & Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability & Accountability Act (HIPAA) Privacy Rule*. November 26, 2012: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html

US Department of Health & Human Services, *Summary of the HIPAA Privacy Rule*. May 2003: http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html

State of New Hampshire, Department of Health & Human Services, *Guidelines for the Public release of Public Health Data*. September 2008: http://www.dhhs.nh.gov/dphs/hsdm/documents/publichealthdata.pdf

Washington State Department of Health, Guidelines for Working with Small Numbers. October 2012: http://www.doh.wa.gov/Portals/1/Documents/5500/SmallNumbers.pdf