NCVHS

National Committee on Vital and Health Statistics

Advising the HHS Secretary on National Health Information Policy

PSC Subcommittee Project: Privacy, Confidentiality and Security Considerations for Data Collection and Use During a Public Health Emergency

NCVHS Subcommittee on Privacy, Confidentiality and Security

June 18, 2020

Today’s agenda

1:45 - 2:00 p.m.: Reviewing yesterday

2:00 – 2:35 p.m.: Discussion; if time, more specific analysis of HIPAA and non-HIPAA covered entities.

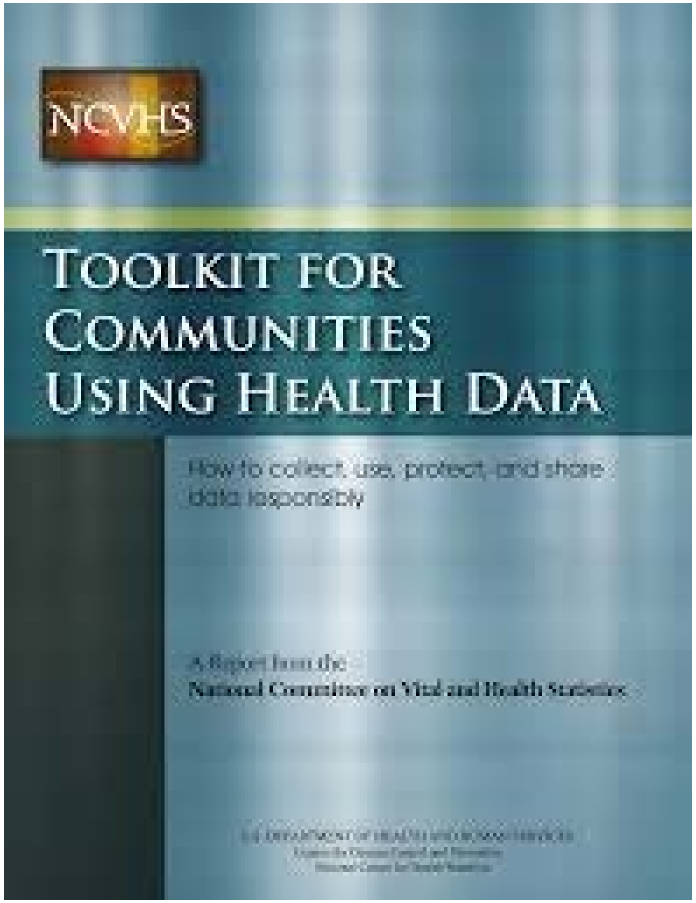
2:35 to 2:45 p.m.: Next Steps

Review of Potential Short Term Project

- What **should** happen with data in an emergency.
- What are fair information principles for a pandemic?
- What data should we be collecting?
- What rules are all right to override to advance public health, and what should remain in force, and perhaps inalienable?
- What level of identification of data is appropriate for which purposes?
  - When is there a need for identifiable data?
  - When is aggregate data more appropriate?
  - Is case-level data without identifiers an adequate compromise?
- How do our standards differ at the local / state / federal levels?

Review of Potential Short Term Project, continued

- Once collected, where may the data get disclosed?
- For what other purposes, if any, should it be used?
- How long can we keep it, and what guardrails to we put around it so it's not misused for law enforcement, immigration, or other purposes that would undermine trust in the public health system?



NCVHS

TOOLKIT FOR COMMUNITIES USING HEALTH DATA

How to collect, use, protect, and share data responsibly

A Report from the
National Committee on Vital and Health Statistics

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Proposed Focus Areas Based on Yesterday’s Discussion

- Updating extant toolkit
- Data use agreements
- Practical popularization of settled data protection principles
- Contact tracing
  - What is adequate training with respect to data protection?

Data Use Agreement Subjects

- De-identify Data When Appropriate
- Determine Retention Schedule
- Require Impact Assessment
- IRB/Institution Approval
  - Ensures shared data will be used responsibly
  - Purpose Specification; Ensure Minimum Necessary Standard
- Prohibit or Limit Third-Party Sharing

Accountability in Data Sharing

- Assign one point person
  - Accountable for data collection, transfer, and disclosure.
  - Identifying and responding to lapses in protocol.
- Enter Data Use Agreements (DUAs) with organizations requesting data.
  - Clarifies legal responsibilities in a legally enforceable document.

Security Given Sensitivity

- There can be a low baseline of data protection at many non-health entities
  - Ari Ezra Waldman, *Privacy as Trust*
- Compliance with HIPAA-mandated administrative, physical, and technical safeguards may be a step up.
- Continually evaluate and reduce security risks in transmitting COVID-19 patient data.
  - Woodrow Hartzog, *Privacy's Blueprint*

Non-Covered Entities

- Entities not covered by HIPAA may be accumulating data on COVID-19-related behavior
  - e.g. internet searches for symptoms on Google, self-reports of symptoms on Facebook, or purchase of cough drops on Amazon.
- What ethical guidelines should govern release of that data to

  a) private medical researchers?

  b) public health authorities?

Data Stewardship

- ***Health data stewardship*** refers to the responsibility of ensuring the appropriate use of personal health data.
- Failure to use good stewardship practices could harm individuals or communities, limit participation, and impede the use of data.
- Different types of data trigger different approaches to stewardship, with the burdens of stewardship and the balancing of interests changing from one type of data to another.
    - A data steward investigating the density of grocery stores in a neighborhood is not likely to encounter major concerns about privacy or confidentiality. But when that same data is repurposed to attribute clinical risks to individuals—e.g. risk of C19—it may trigger those concerns.

Deepening Commitments

- Accountability
- Openness, Transparency, and Choice
- Community and Individual Engagement and Participation
- Purpose Specification
- Quality and Integrity
- Security
- De-Identified Data

Purpose Specification

- **What question is the project designed to answer?** Data users should explicitly and carefully frame the question and be able to explain how the data will answer the question.
- **Repurposed data** are collected for one purpose and then used for another. Public health surveillance data collected by state health departments is repurposed when shared with communities or researchers to investigate a concern that the data may help explain. Laboratory tests performed to guide patient diagnosis and treatment are repurposed when combined with many other tests to show the prevalence of a condition in a subgroup of individuals.
- **When using repurposed data:**
  - consider concerns that may be raised by those whose data are being repurposed.
  - consider how changing the original purpose may trigger the need for additional notice or consent or if these changes are allowed under the DUA with the data steward.

Data Quality and Integrity

- Data quality refers to the accuracy, relevance, timeliness, completeness, validity, and reliability of the data.
- Data must be recorded or captured accurately, and it must represent what it is claimed to represent.
- Special Consideration for Merged Data Sets
  - Data users sometimes merge data from two or more sources to gain enriched data that is more useful than either data set alone. When two or more data sets are combined, users should ensure that a merger or aggregation is valid, and that the data retain integrity. Otherwise, the combined data may no longer accurately reflect the sources.
- It is seldom possible or necessary to have perfect data, but stewards should consider and make a judgment about whether data accurately and adequately measure what is being studied, and if the data can be trusted.
  - For more analysis on data stewardship, see Kristin Madison, "Health Regulators as Data Stewards," *North Carolina Law Review*, Vol. 92, pp. 1605-1636, 2014.

Principles for Sharing

- **Ensure that other researchers will respect ethical and privacy concerns.**
- If protective protocols were established for the data, new researchers who will use the data should agree to the same protocols.
- Data use agreements with enforceable consequences
- Auditing to detect unauthorized use and downstream sharing
- Fragile watermarks or other measures to heighten security?

De-Identification

- **De-identification** is the process of removing or obscuring any directly or indirectly identifying information from data in a way that minimizes the risk of unintended disclosure of individuals' identity and information. **Identity disclosure**—when an outside party can assign an identity to a record in a disclosed data set.
  - **Attribute disclosure**—allows an outside party to attribute characteristics to someone in a data set even if he or she has not been individually identified.
- De-identification combined with data management measures (such as data oversight boards, training and education of users, or penalties for misuse), and information technology solutions (such as encryption), are methods that may help to manage the risk of release while making relevant health care information more available to data users.

Privacy Rule De-identified Data Elements

Names

*Geographic subdivisions smaller than a state

*Dates

Telephone numbers

Fax numbers

E-mail addresses

Social Ssecurity numbers

Medical record numbers

Health plan beneficiary numbers

Account numbers

Certificate/license numbers

Vehicle identifiers and serial numbers, including license plate numbers

Device identifiers and serial numbers

Web universal resource locators (URLs)

Internet protocol (IP) address numbers

Biometric identifiers, including fingerprints and voiceprints

Full-face photographic images and any comparable images

Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification

*Identifiers marked with an asterisk may be included in a limited data set.

Quantifying and Evaluating the Risk of Re-identification

- **Merging data sets**, *in particular*, may increase the risk that individuals or small groups could be identified. Merged data sets raise concerns when people would not expect the data to be combined
  - E.g. correlations among prescriptions filled, food purchases, and method of payment for food that could be obtained from private supermarket data); when analysis of the combined data sets may have negative consequences for those whose data are used; or when merger raises the risk that private or confidential data may be disclosed.
- **Harms Modeling**: identify and take steps to mitigate
  - Risk of Injury
  - Denial of Consequential Services
  - Infringement on Human Rights
  - Erosion of Social and Democratic Structures

Resolving Ethical Tensions

- **Participatory Algorithm Design**
  - Researchers should include key stakeholders in the research process, including clinicians, social networks, and individuals who are the object of these predictions.
- **Developing Best Practices for Methods**
  - In published work, researchers should disclose study design and methods decisions to promote reproducibility, and the field should agree on what best practices are.
- **Beyond Ethics Boards**
  - Consider and discuss the implications of this research, outside of the normal considerations of ethics committees. Incorporate ethics as a key value in the research process from the beginning.

Stevie Chancellor et al., *A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media*, Conference on Fairness, Accountability, and Transparency, ASSOCIATION FOR COMPUTING MACHINERY (2019).

Proposed Next Steps

- **Scoping Document as Literature Review (summer/fall)**
  1) Updating extant toolkit
  2) Data use agreements best practices
  3) Practical popularization of settled data protection principles
  4) Contact tracing: What is adequate training with respect to data protection?

- **September Hearing** to canvass experts' latest views (not yet reflected in literature)