National Committee on Vital and Health Statistics

Advising the Secretary of Health and Human Services on health information policy since 1949.

# Privacy, Confidentiality and Security Subcommittee

## June 15, 2016

A. Summary of Hearing on De-Identification and the Health Insurance Portability and Accountability Act (HIPAA)" May 24-25, 2016

B. Preview of Hearing on Minimum Necessary and HIPAA, June 16

C. 18 month priorities for the PCS Subcommittee

## *De-identification Hearing Objectives:*

1. Increase awareness of current and anticipated practices such as the sale of information to data brokers and other data-mining companies for marketing and/or risk mitigation activities;
2. Understand HIPAA's de-identification requirement in light of these practices, and
3. Identify areas where outreach, education, technical assistance, a policy change, or guidance may be useful.

# *Overarching Themes*

1. There is a "privacy-data collision"
2. Few agreed upon processes for de-identification and managing de-id data
3. Divergent views of problem/issues
4. Expanding data science research
   - Raise the sophistication of current practice
   - Translation and spread into practice
5. Value of use cases for education
6. Further guidance or a change to regulation
   - Adding genomics
   - Suite/tier of methodologies
   - Lifecycle/Data Stewardship
   - Controls

# *Take Aways -I*

- Every data set presents different de-identification challenges
- De-identified data does not stay de-identified
- Need for physical, technical and administrative solutions
  - Depends on data, recipient, context
- Suggestion for a controlled study of the efficacy of Safe Harbor
- Policy incentives to improve application of de-identification techniques and/or tools

# *Take Aways - II*

- Need a workable definition of re-identification
- Different rules depending on recipient
- Current methods depending on recipient
  - For direct access = licensing and security
  - For dissemination = de-identification
  - Query-based = now limited to differential privacy
- Laws are based sector-specific, data in different environments treated differently
- De-identification worthwhile given changes in technology; useful, but not sufficient, may need other restrictions

# *More Take Aways - III*

- Options for mitigating risk of re-identification
  - Synthetic data sets
  - Enclaves (Semi-trusted analytic 'sandbox')
  - Secure multi-party computing
- Desire to address provenance of data
- Need to limit burden on research
- Limited use v. public use data
- More de-identification v. more utility
- Data Use Agreements move enforcement from regulatory realm to contract enforcement
- De-identified PHI has no restriction on re-identification
- Too focused on de-identification alone v. spectrum of disclosure limitations, techniques and tools

# *Panel 1: Policy Interpretations*

## Panel Observations

- Complexity of de-identifying narrative data v. unstructured data
- Risk means something different to everyone
- Who is responsible for certifications when multiple data sets integrated? Which expert controls?
- Formalists and pragmatists don't meet/talk
- No standardization of cell suppression
- No standard for minimizing risk

# *Panel 1: Policy Interpretations*

## Panel Suggestions

- No oversight of de-identified PHI --"ludicrous"
- Time-limited certification of a data set under safe harbor or expert analysis
- Education of IRBs in de-identification science
- Agreement on what constitutes a "small risk" under HIPAA
- Clearinghouse for best practices in de-identification practice
- Process for minimizing risk similar to data security policy
- Tiered access levels of data
- Economic incentives for adopting policy
- Adding a provision to Data Use Agreements prohibiting re-identification

# *Panel 2: De-Identification Challenges*

## Panel Observations

- Citizen scientists, not covered
- Individuals want > control → access
- Need for more robust risk assessment tools, processes
- Risk of propagating existing biases in data
  - Drawing inaccurate inferences
- Governance and ethics discussion
- We're bad at talking about risk, societal values
  - Risk of harm v. statistical risk of re-id v. benefits
  - Reputational v. economic v. bodily v. other?
- Technology alone is not the solution

# *Panel 2: De-Identification Challenges, continued*

## Panel Suggestions

- Best practices, e.g. life cycle management, audits
- FTC report on Internet of Things
  - Reasonable steps to de-identification
  - Commitment to re-identification
  - Need for enforceable contracts
  - Prohibiting downstream re-identification
- Consideration of "context" of collection, opt=in/opt-out
  - Under what circumstances should individuals be notified of a new use?
- Real v. perceived harm, re-identification is not necessarily actual harm
- Multi-step process for consent

# *Panel 3: Approaches for De-Identification and Re-Identification*

## Panel Observations

- Separating data set from potential uses
- Machine learning, algorithms – respect limits
- Policy chases technology, need for evolutionary policy process
- De-identification conflicts with data exchange regulations
- Lack of resources to audit de-identification
- Vendors want to keep de-identified data to use later
- Look for benefit to patient, then allow re-identification - No guidance for going back to data source
- Using pseudo-identifiers common privacy architecture
- Improve linkages for mortality and outcome data, including longitudinal
- Self-reported genomic data can be used to re-identify

# *Panel 3: Approaches for De-Identification and Re-Identification, continued*

## Panel Suggestions

- Use cases, e.g., pop health, precision medicine, would be helpful
- Rules need room for innovation
- Consider how partners address privacy especially when linking data
- Need guidance for publishing reports, standards for how to report
- Revisit rules for genetic information regularly (dynamic)
- Policy should anticipate more ways to identify data at later dates
- Improve education on implications of genomic data, other kinds
- Aligning best practices in administration with IRB practices

# *Panel 4: Models for Privacy-Preserving and Use of Private Information*

## Panel Observations

- Consumer expectations = legal, just, fair
- Increased capture of observational data, e.g., cameras, activity trackers, apps, behavior-related

# *Panel 4: Models for Privacy-Preserving and Use of Private Information, continued*

## Panel Suggestions

- Catalogue of privacy controls + calibration to different levels of risk, updating over time such as FISMA controls
- Appropriate controls for each stage of life cycle
- Crosswalk to different standards, e.g., HITRUST framework
- Where no experts on staff -- Catalogue of controls, licensing agreements, expert panels, develop guidance, design for interoperability
- Governance and accountability against a set of social values

# *Research Impediments*

- Expense of obtaining an expert
- Administratively cumbersome to get consents
- Use of safe harbor reduces utility
- Fear of re-identification due to change in technology
- Cumbersome to prepare data collected for one purpose into data useful for a secondary purpose

# *Minimum Necessary Hearing Objectives—June 16*

1. Understand current industry policies and practices involving minimum necessary practices;
2. Understand challenges and potential areas for clarification in light of these practices, new and emerging technology developments, and new and evolving policy directions since the Privacy Rule became effective, and
3. Identify areas where outreach, education, technical assistance, or guidance may be useful.

## 2016 Subcommittee Workplan