# Detailed Problem Statements

### Analysis and discussions by the PSWG yielded high priority areas for discussion and solution development. This section outlines the key problem areas for focus and provides greater detail about the specific problem that needs to be addressed.

## Potential for Harmful or Discriminatory Practices

### During our hearings, among the most oft cited concerns about health “big data” is the potential for health data to be used in a way that harms individuals or groups. Discrimination is just one example of a harm that can result from certain uses of health big data. U.S. laws prohibit some discriminatory uses of health data – for example, use of health data to make decisions about health insurance coverage – but other discriminatory uses of health data are either not prohibited or are expressly permitted (for example, use of health information in life and disability insurance decisions).

### Beyond discrimination, some see other uses of health data as being “harmful” (for example, marketing and other “commercial” uses). However, there is a lack of consensus on which uses are “harmful,” as well as an inability to predict which future uses could be harmful, which creates challenges to enacting policies to prohibit or place additional constraints on such uses. During our hearings, some presenters expressed concern about the use of algorithms to make decisions about people, and the lack of “transparency” about the data used to inform these algorithms and about precisely how they are used. Poor transparency increases the potential for reinforcing bias that may lead to unfair practices.

### Failing to pay attention to these issues undermines trust in health big data, which could create obstacles to leveraging health big data to achieve gains in health and well-being.

## Two different domains of regulation (HIPAA and “Other”) Yields contradictions and unpredictability

## HIPAA covers many sources of health big data – but not all. Consequently, we lack comprehensive, FIPPS-based protections for health data (and data used for health purposes) in many domains, which is confusing for consumers and imperils trust in health big data.

### Access – Individuals often lack the ability to access, use, and share their own data, including for research and learning health system (LHS) activities. Even with respect to HIPAA covered entities, who are required to provide this right to individuals, the right is often difficult for individuals to exercise.

### Transparency – There is a lack of transparency regarding how holders of personal information use that information and how information is exchanged, especially in the Big Data ecosystem outside of traditional healthcare. This lack of transparency erodes trust and exacerbates the fear of discrimination.

### Research – when it is regulated, the rules do not necessarily regulate based on privacy risk and, as a result, create higher hurdles for uses of data for “research” purposes that intend to contribute to “generalizable knowledge” (i.e., the greater good).

## Protect Health Information by Improving Trust in De-identification Methodologies and Reducing the Risk of Re-identification

### De-identification is a useful tool for protecting privacy in big data research – but we over-rely on it as a matter of policy and do not have ways to hold people accountable for re-identifying data or negligently leaving data vulnerable to re-identification.

### HIPAA has standards – but there is no overarching standard(s) for de-identification of health data outside of HIPAA. HIPAA is often voluntarily used, but it is not required.

### Concerns have been raised about both methodologies currently used for de-identification under HIPAA – safe harbor and expert determination. The former may not be sufficiently protective in all contexts (particularly given increases in publicly available data) and there are no objective criteria governing the expert method.

### There is increased risk of re-identification when data sets are combined (the mosaic effect).

### In addition, de-identification – even under HIPAA – has never meant zero risk, but de-identified data is not subject to regulation (so the residual risk that remains is unregulated). We do not have consistent mechanisms for punishing people/entities who re-identify or who negligently leave datasets vulnerable to easy re-identification.

### De-identification is also not the panacea for all data use. Rendering data de-identified pursuant to HIPAA eliminates the potential for valuable uses of data and introduces burden on innovation and analytics.

## Security Threats and Gaps

### The lack of an end-to-end secure environment for health data was a problem mentioned by many who presented – but no entity (or federal agency) is responsible for assuring those end-to-end protections. Instead we have silos of protections (for example, HIPAA coverage applies in some places, FTC and FDA in others, Gramm-Leach-Bliley in financial contexts; state law may govern; some may be covered by multiple laws, and some may be covered by none). The lack of baseline security requirements was uniformly seen as a significant risk for deteriorating patient and consumer trust in the healthcare system and in entities involved inside and outside of healthcare. The call for such end-to-end security requirements was referenced as one of the highest priorities.

### The laws that do exist do not necessarily provide incentives for adopting privacy-enhancing technical architectures for big data analytics (for example, data enclaves).

### Congress is the only policy-making body equipped to authorize national security and/or cybersecurity requirements that would facilitate the requirement to provide a baseline level of security around health data, regardless of the entity that holds that data, in an end-to-end environment that is desirable for building trust.

# Solutions and Recommendations

## Addressing Harm, Including Discrimination Concerns

### Without a national consensus on what constitutes harm with regard to the use of big data, the Workgroup encourages ONC and other federal stakeholders to promote more public inquiry and projects to fully understand the scope of the problem and the potential for privacy harm, including discrimination. Such inquiry and projects should focus on identifying gaps in regulation and commonly identified harms.

### To address discriminatory practices: Policymakers should continue to monitor the use of health data to identify gaps in current legal protections and areas for further inquiry. Transparency about health data uses could illuminate which uses are harmful, and as a result advance a national consensus around harms

### To address distrust in big data algorithms: Improve trust through algorithmic transparency. Consider applying Fair Credit Reporting Act (FCRA) approaches to promote algorithmic transparency. The FCRA is a federal law that regulates consumer reporting agencies (CRAs) and empowers people by providing transparency about the use of consumer credit information. Regulations provide for how information is gathered, how it is used, and what CRAs must tell people. If information is used in a way that has an adverse impact on an individual, then certain disclosures must be made.

## Address Uneven Policy Environment

### To address a dearth of knowledge about uses of data and privacy laws: Leverage the most recent recommendations by the PSWG on better educating consumers about the current privacy and security laws, including how data is used within and outside of the HIPAA environment.

### We need comprehensive, FIPPs-based protections for health data not covered by HIPAA, which means not focusing on consent as the primary mechanism for such protections. Congress could address this through legislation, but voluntarily adopted codes of conduct can be enforced by the FTC for entities subject to their jurisdiction. A number of efforts are under way to develop such codes – those efforts should be encouraged and HHS, FTC and other relevant federal agencies should offer to review and provide suggestions for such efforts in order to more quickly establish dependable “rules of the road” that help build trust in health big data. [Note that we have already recommended that the Consumer workgroup consider an evaluation effort for consumer-facing apps.] Such codes of conduct should emphasize transparency (regarding data collection, transfer and use), individual access, accountability, and use limitations, at a minimum. They should also reward/promote the use of privacy enhancing architectures for big data analytics, such as data enclaves.

### We need to re-evaluate existing rules governing uses of data that could contribute to a learning health system to assure those rules promote the responsible re-use of data to contribute to generalizable knowledge. [Reference previous HITPC recommendations to treat certain types of research like operations when the covered entity remains in control of uses of the data]

### We need to re-evaluate the existing rules governing disclosure of data that could improve and make more efficient research efforts, especially when such disclosures are to entities or environments that afford appropriate protection of health data (for example, data enclaves or research entities agreeing to comply with HIPAA or FIPPs protections).

### We also need to strengthen existing rules giving individuals access rights to health information to bring them into the digital age, so that individuals can access, download, and transmit their health information as easily as they can access their financial information. [Do we want to say more here?]

## Protect Health Information by Improving Trust in De-identification Methodologies and Reducing the Risk of Re-identification

### We ask OCR to be a more active “steward” of HIPAA de-identification standards and conduct ongoing review of the methodologies to determine robustness and recommend updates to the methodologies and policies. The analysis could be performed by an outside expert, such as NIST, but would be vetted and ultimately endorsed by OCR.[[1]](#footnote-1)

### Consider the following additional recommendations that came out of the hearing testimony:

### - Limit use of safe harbor only to circumstances where data represent a random sample of a population [Look to Khaled El Emam’s recommendations]

### - Consider whether de-identification status of a dataset should be required to be re-evaluated when context changes (such as when data set is combined with other data).

### - Develop or encourage the development of programs to objectively evaluate statistical methodologies; consider granting safe harbor status to methodologies proven to be effective in particular contexts.

### - Congress should act to address accountability for re-identification or negligent anonymization/de-identification.

### Consideration should be given to risk-based de-identification requirements and re-identification risk when data is held by entities or in environments where re-identification risk remains low (for example, data enclaves or data repositories voluntarily adopting HIPAA security rules).

## Supporting Secure Use of Data for Learning

### To address the lack of a widely accepted security framework: Call on policy makers to enact security requirements for non-HIPAA covered entities. The FTC has previously recommended the enactment of strong, flexible, and technology-neutral legislation to strengthen the Commission’s existing data security enforcement tools.[[2]](#footnote-2)

### Federal policymakers should provide incentives for entities to use privacy-enhancing technologies and privacy-protecting technical architectures, such as secure data enclaves, secure distributed data systems, and distributed computation.

### [Point again to previous recommendations about HIPAA Security Rule and keeping it up to date with other security frameworks]

1. Reference NIST evaluation of efficacy of de-identification effort. [↑](#footnote-ref-1)
2. See FTC Internet of Things Report, January 2015, p. 49. [↑](#footnote-ref-2)