HEALTH BIG DATA RECOMMENDATIONS

HITPC Privacy and Security Workgroup
August 2015

Table of Contents

1 Introduction

This report addresses privacy and security concerns as well as recommendations related to the application of big data analytics in the healthcare space. The Privacy and Security Workgroup (PSWG) of the Health Information Technology Policy Committee (HITPC) is charged with investigating and providing recommendations to the National Coordinator for Health Information Technology at the U.S. Department of Health and Human Services (HHS) on privacy and security issues related to the electronic exchange of health information. The application of big data in healthcare impacts one of the PSWG's core values; specifically, that patients' needs and expectations should be considered, and that "patients should not be surprised about or harmed by collections, uses or disclosures of their information."[1]

The collection, analysis, and use of large volumes of electronic information will be a driver in the U.S. economy for the foreseeable future. Through the proliferation of software applications and mobile devices, the amount of health-related information is growing exponentially. As the volume, velocity, and variety of information continue to grow, so do the potential risks arising from unknown and inappropriate uses of protected health information (PHI).[2]

In response to a charge from the White House to consider the impacts of big data analyses, the PSWG invited relevant experts and interested stakeholders to testify on the opportunities and challenges of health big data, health big data concerns and harms, and the advantages and limits of current laws concerning the use of big data and emerging technologies. The PSWG held three (3) public hearings between December 2014 and February 2015 in which 21 individuals from across the healthcare spectrum were invited to speak.[3] The speakers are leading experts with a diverse perspective on issues related to big data in healthcare, and they represent a wide range of stakeholder groups, including consumer and privacy advocacy groups, consumer-facing enterprises, academia, big data analytics companies, and healthcare delivery systems.

1.1 Opportunities and Challenges

Many see the application of big data analytics in healthcare as an opportunity to improve the health of both individuals and their communities. These benefits include safer treatments, the ability to target communities and individuals with tailored interventions, and the ability to respond to the spread of diseases more rapidly.[4] Big data analytics can also support the growth of a learning health system (LHS), which is "an environment that links the care delivery system

[1] HITPC Transmittal Letter, September 1, 2010, p. 4, http://www.healthit.gov/sites/faca/files/hitpc_transmittal_p_s_tt_9_1_10_0.pdf.
[2] Protected Health Information is defined in 45 CFR § 160.103.
3 Speaker material and transcripts are available on www.HealthIT.gov for FACA hearing dates of December 5 and 8, 2014 and February 9, 2015.
[4] Public Hearing Responses of Richard Platt, p. 3, http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Richard_Platt_Reply_to_Questions_for_Panelists_2014-12-05.pdf [hereinafter "Richard Platt Responses"].

with communities and societal supports in 'closed loops' of electronic health information flow, at many different levels, to enable continuous learning and improved health."[5] ONC recently released a draft roadmap for the interoperability of clinical information to support research and big data analyses on the path to achieving a nationwide learning health system.[6]

Big data computing also poses challenges to privacy and security. Rapid growth in the volume of health-related information increases the risk of privacy violations,[7] particularly when data sets are combined.[8] Data anonymization tools such as de-identification are useful, but cannot eliminate risks to re-identification.[9] Additionally, the complex legal landscape around health privacy creates obstacles for individuals trying to access their personal information and hurdles for researchers attempting to grow the LHS. Much of the health-related information generated today is not regulated by the Health Insurance Portability and Accountability Act (HIPAA), which provides baseline privacy and security rules for covered entities (most healthcare providers, payers, and healthcare clearinghouses) and their business associates.[10] On the other hand, federal regulations applicable to human subjects research, although well intended, have posed challenges for researchers.[11]

1.2 Recommendations Summary

A high-level summary of the PSWG's health big data recommendations follows:

6.1 Address Harm, Including Discrimination Concerns

- ONC and other federal stakeholders should promote a better understanding by the public of the full scope of the problem – both harm to individuals and communities.
- Policymakers should continue to focus on identifying gaps in legal protections against what are likely to be an evolving set of harms from big data analytics.
- Policymakers should adopt measures that increase transparency about actual health information uses.

---

[5] Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap, Draft Version 1.0, p. 8, http://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf [hereinafter "Interoperability Roadmap"]; *see* also, Richard Platt Responses, p. 1 ("The term "Learning Health System" connotes a commitment to improve care, both by learning from all patients' experiences and by implementing the results of the learning activities.").

[6] Interoperability Roadmap, p. 35.

[7] Michelle De Mooy, Privacy and Security Workgroup Transcript, December 5, 2014, p. 30 [hereinafter "December 5"].

[8] Lucia Savage, December 5, p. 24.

[9] Michelle De Mooy, December 5, p. 30.

[10] For information on covered entities and business associates, see http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/; for definitions, see 45 CFR §160.103.

[11] See Deven McGraw & Alice Leiter, *Risk-Based Regulation of Clinical Health Data Analytics*, Colo. Tech. L. J., Vol. 12.2, p. 435.

- Policymakers should explore ways to increase transparency around use of the algorithms used in big health analytics, perhaps with an approach similar to that used in the Fair Credit Reporting Act (FCRA).

6.2 Address Uneven Policy Environment

- Promote Fair Information Practice Principles (FIPPs)-based protections for data outside of HIPAA:
    - Voluntarily adopt self-governance codes of conduct. In order to credibly meet the requirements of both protecting sensitive personal information and enabling its appropriate use. Codes must include transparency, individual access, accountability, and use limitations.
    - U.S. Department of Health and Human Services (HHS), Federal Trade Commission (FTC), and other relevant federal agencies should guide such efforts to more quickly establish dependable "rules of the road" and to ensure their enforceability in order to build trust in the use of health big data.
- Policymakers should evaluate existing laws, regulations, and policies (rules) governing uses of data that contribute to a learning health system to ensure that those rules promote responsible re-use of data to contribute to generalizable knowledge.
- Policymakers should modify rules around research uses of data to incentivize entities to use more privacy-protecting architectures, for example by providing safe harbors for certain behaviors and levels of security.
- To support individuals' rights to access their health information, create a "right of access" in entities not covered by HIPAA as part of the voluntary codes of conduct; also revise HIPAA over time to enable it to be effective at protecting health data in the digital age.
- Educate consumers, healthcare providers, technology vendors, and other stakeholders about the limits of current legal protection; reinforce previous PSWG recommendations.
    - Leverage most recent PSWG recommendations on better educating consumers about privacy and security laws and uses of personal information both within and outside of the HIPAA environment.

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

- The Office for Civil Rights (OCR) should be a more active "steward" of HIPAA de-identification standards.
    - Conduct ongoing review of methodologies to determine robustness and recommend updates to methodologies and policies.
    - Seek assistance from third-party experts, such as the National Institute of Standards and Technology (NIST).
- Programs should be developed to objectively evaluate statistical methodologies to vet their capacity for reducing risk of re-identification to "very low" in particular contexts.

- OCR should grant safe harbor status to methodologies that are proven to be effective at de-identification in certain contexts to encourage use of proven methodologies.
- OCR should establish risk-based de-identification requirements in circumstances where re-identification risk is very low.

6.4 Support Secure Use of Data for Learning

- Develop voluntary codes of conduct that also address robust security provisions.
- Policymakers should provide incentives for entities to use privacy-enhancing technologies and privacy-protecting technical architectures.
- Public and private sector organizations should educate stakeholders about cybersecurity risks and recommended precautions.
- Leverage recommendations made by the Privacy and Security Tiger Team and endorsed by the HITPC in 2011[12] with respect to the HIPAA Security Rule.

2 Background

2.1 Privacy and Security Workgroup Charge

In response to the White House report on big data and other complementary federal initiatives,[13] the PSWG was charged to investigate privacy and security issues related to big data in the healthcare space and recommend actions to address critical challenges. This section briefly summarizes the reports that shaped the PSWG's charge.

2.1.1 White House Report on Big Data and the President's Council on Advisors for Science and Technology Report

In May 2014, the White House released a report on big data that highlights the pressure on traditional privacy-protective measures (i.e., the Fair Information Practice Principles,[14] or FIPPs), such as de-identification, notice and consent.[15] The report recommends that

[12] HITPC Transmittal Letter, August 16, 2011, http://www.healthit.gov/sites/faca/files/HITPC_PSTT_Transmit_8162011.pdf.

[13] Big Data: Seizing Opportunities, Preserving Values, May 2014, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [hereinafter "White House Big Data Report"].

[14] There is no definitive version of the FIPPs, which are recognized worldwide as the foundational principles for data privacy. Appropriate sources include the OECD Guidelines on the Protection of Privacy, http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm, the Markle Connecting for Health Common Framework, http://www.markle.org/sites/default/files/CF-Consumers-Full.pdf, the White House's 2012 Consumer Bill of Rights, https://www.whitehouse.gov/sites/default/files/privacy-final.pdf, and the NIST National Strategy for Trusted Identities in Cyberspace, http://www.nist.gov/nstic/NSTIC-FIPPs.pdf.

[15] White House Big Data Report, p. 54, (stating that "re-identification is becoming more powerful than de-identification," and "focusing on controlling the collection and retention of personal data… may no longer be sufficient to protect personal privacy.").

government "lead a consultative process to access how HIPAA and other relevant federal laws and regulations can best accommodate the advances in medical science and cost reduction in healthcare delivery enabled by big data."[16] The report acknowledges the complexity of the current federal and state legal landscape regarding patient information and privacy, and suggested the "need to carve out special data use authorities for the healthcare industry if it is to realize the potential health gains and cost reductions that could come from big data analytics."[17] Finally, the report highlights that neither HIPAA nor other privacy laws regulate many organizations that collect health-related information, and that consumer privacy expectations may not be met in the current ecosystem.

The White House report is complemented by the President's Council of Advisors for Science & Technology (PCAST) report to the President,[18] released on the same day, which reinforces the pressure big data places on the FIPPs.

2.1.2 White House Open Government Partnership

In September 2014, as part of the U.S. Open Government Commitments through the Open Government Partnership (OGP),[19] the White House encouraged the use of big data to support greater openness and accountability, and highlighted the need to "ensure privacy protection for big data analysis in health."[20] Specifically, the White House recommended that to "ensure that privacy is protected while capitalizing on new technologies and data, the Administration, led by HHS, will: (1) consult with stakeholders to assess how Federal laws and regulations can best accommodate big data analyses that promise to advance medical science and reduce healthcare costs; and (2) develop recommendations for ways to promote and facilitate research through access to data while safeguarding patient privacy and autonomy."[21] As a result, ONC charged the PSWG to operationalize this commitment.

2.2 PSWG Plan of Action

Beginning in October 2014, the PSWG held several public meetings and hearings in which experts presented and discussed key issues. PSWG held two days of public hearings on

[16] White House Big Data Report, p. 62.
[17] White House Big Data Report, p. 23.
[18] Big Data and Privacy: A Technological Perspective, May 2014, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.
[19] The Open Government Partnership is a global initiative that began in 2011 to promote transparency and leverage new technologies, among other objectives. *See* Open Government Partnership, http://www.opengovpartnership.org.
[20] FACT SHEET: Announcing New U.S. Open Government Commitments on the Third Anniversary of the Open Government Partnership, https://www.whitehouse.gov/the-press-office/2014/09/24/fact-sheet-announcing-new-us-open-government-commitments-third-anniversa [hereinafter "Open Government Partnership"].
[21] Open Government Partnership.

December 5 and December 8, 2014.[22] The Workgroup invited panelists from industry, non-profit organizations, academia, and law to address the following issues as they relate to big data: (1) health big data opportunities, (2) health big data concerns, (3) the learning health system, (4) protections for consumers, and (5) current laws. Please see Appendix A for a list of public hearing topics and speakers.[23]

Following these hearings, the Workgroup analyzed the testimony and began drafting and refining its recommendations. In February 2015, the PSWG heard additional testimony on health big data security issues,[24] and in March and June 2015,[25] the PSWG updated the HITPC on the Workgroup's progress. The PSWG's public deliberations continued through July 2015.

3 Scope

In identifying specific issues to address within its charge, the PSWG remained mindful of the lessons and recommendations of other initiatives and activities.[26] Given the breadth of big data as a topic, the PSWG narrowed the scope of its discussions and recommendations to privacy and security concerns and potentially harmful uses of big data in healthcare. The PSWG also focused on prevailing legal frameworks and potential gaps in privacy and security protections, as well as the degree to which existing laws facilitate an environment that enables information to be "leveraged for good" while still protecting individual's privacy interests.

The PSWG identified several issues that were out of scope. These included matters related to data quality, data standards, and the non-representativeness of data (e.g., data that does not accurately reflect the composition of the population, which has the potential to ignore under-served communities). Where possible, the PSWG sought to avoid discussing issues that have been addressed by other projects and initiatives, as summarized above, though some topics and themes were complementary.

[22] *See* Policy: Privacy and Security Workgroup Virtual Hearing, December 5, 2014, http://www.healthit.gov/facas/calendar/2014/12/05/policy-privacy-security-workgroup-virtual-hearing and Policy: Privacy and Security Workgroup Virtual Hearing, December 8, 2014, http://www.healthit.gov/facas/calendar/2014/12/08/policy-privacy-security-workgroup-virtual-hearing.
[23] Please see Appendix A, Health Big Data Public Hearing Topics and Speakers.
[24] *See* Policy: Privacy and Security Workgroup, February 9, 2015, http://www.healthit.gov/facas/calendar/2015/02/09/policy-privacy-security-workgroup.
[25] *See* HITPC Meeting, March 10, 2015, http://www.healthit.gov/FACAS/calendar/2015/03/10/hit-policy-committee and Virtual HITPC Meeting, June 30, 2015, http://www.healthit.gov/FACAS/calendar/2015/06/30/virtual-hit-policy-committee.
[26] A brief summary of these activities is provided in Appendix C.

4 Public Testimony

This section provides a high-level summary of testimony from the Workgroup's public hearings and deliberations.[27] These hearings and meetings surfaced several key themes, which provide the following structure for this section: (1) concerns about tools commonly used to protect privacy; (2) preventing, limiting, and redressing privacy harms; and (3) the complex legal landscape, including issues of under- and over-regulation. A more detailed account of the supporting testimony is provided in Appendix B.

4.1 Concerns about Tools Commonly Used to Protect Privacy

Big data is blurring the lines between traditional health information (e.g., clinical or billing information) and other information (e.g., user-generated information about diet, steps, workouts, sleep, and mood).[28] Consequently, defining health information is becoming more difficult because almost all information has potential to, in some way, become health-related information, depending on how it is used.[29] Growth in the amount and availability of such information places additional pressure on core FIPPs.

4.1.1 De-identification

De-identification refers to the data anonymization methods that obfuscate health information to keep it confidential. HIPAA provides two methods of de-identification – safe harbor and expert determination – that are widely used to facilitate health research and are considered a powerful privacy protective tool.[30] Nevertheless, several presenters noted important weaknesses in the current HIPAA de-identification practices and offered specific solutions. Specifically, safe harbor poses a higher risk of re-identification and its use should be re-evaluated. With regard to expert determination, presenters cited the need to establish common, publicly scrutinized standards.[31] Several panelists urged that progress be made to prohibit re-identification.[32]

4.1.2 Patient consent

A patient's meaningful consent to authorize the use and sharing of personal health information is a valuable tool for protecting privacy and individual autonomy.[33] The Workgroup approached the consent issue by assessing how it works both within the HIPAA environment and outside the HIPAA environment, with a particular focus on consent for research and consent in the non-HIPAA Internet of Things environment. Panelists and PSWG members discussed the degree to

[27] Information about the hearing agendas and testifiers is provided in Appendix A.
[28] Stephen J. Downs, December 5, p. 8.
[29] Stephen Downs, December 5, p. 20; Mark Savage, December 5, p. 30; David McCallie, Jr., Privacy and Security Workgroup Transcript, December 8, 2014, p. 21 [hereinafter "December 8"].
[30] *See* Guidance Regarding Methods for De-Identification of Protected Health Information, http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html.
[31] See generally, Khaled El Emam, December 5, p. 48-49.
[32] Michelle De Mooy, December 5, p. 29; Mark Savage, December 5, p. 38, 41; Fred Cate, December 5, p. 63.
[33] Stanley Crosley, Privacy and Security Workgroup Transcript 2015-01-26, p. 17 [hereinafter "January 26"]. *See also*, www.healthit.gov/meaningfulconsent.

which people can and wish to control their health information, as well as the difficulty of obtaining meaningful consent when big data analytics is leveraged for various secondary uses of data.[34]

4.1.3 Data security

Security, including data breaches, was highlighted as one of the most pressing issues for big data in healthcare.[35] To build trust in a robust health big data ecosystem, panelists emphasized the need for a holistic approach to security, which includes embracing end-to-end security policy and technology frameworks that apply broadly, regardless of whether organizations are covered by HIPAA.[36] In any environment, information must be protected at the same time it is made available, and the methods of protection should not interfere with the responsible use of information.[37]

4.1.4 Transparency

Testimony provided that the "foundational principle of *openness and transparency* is perhaps the most important component of the FIPPs for all entities using big data."[38] Poor transparency engenders a lack of trust, and trust is essential for future learning and the beneficial application of big data in healthcare. Inadequate transparency extends to proprietary algorithms, which are used to make decisions about individuals and potentially shape their behavior.[39] To bolster public trust in health big data, participants suggested leveraging methods and lessons learned from transparency provisions in existing laws, such as the Fair Credit Reporting Act (FCRA).

4.1.5 Collection, use, and purpose limitation

Big data analytics and research begins with researchers examining trends and patterns in large data sets without first formulating a hypothesis.[40] As a result, the need to gather as much information as possible before identifying a research purpose conflicts with longstanding FIPPs that require defining the specific purpose(s) for which information is collected and limiting the amount of personal information to what is necessary to accomplish the specified purpose(s). Regardless of the challenge posed by big data, panelists and PSWG members agreed that organizations should examine their collection and retention practices and be mindful of over-collection.[41]

---

[34] Stanley Crosley, January 26, p. 17.
[35] Michelle De Mooy, December 5, p. 29.
[36] Michelle De Mooy, December 5, p. 30; Andrei Stoica, Privacy and Security Workgroup Transcript, February 9, 2015, p. 6 [hereinafter "February 9"].
[37] Fred Cate, December 5, p. 53.
[38] Testimony of CDT for the HIT Privacy and Security Workgroup Virtual Hearing, December 5, 2014, p. 3, http://www.healthit.gov/facas/sites/faca/files/PSWG_Testimony_Michelle_DeMooy_CDT_2014-12-05_0.pdf (emphasis added) [hereinafter "CDT Testimony"].
[39] Michelle De Mooy, December 5, p. 35-36.
[40] Ella Mihov, December 8, p. 30-31.
[41] Michelle De Mooy, December 5, p. 43.

4.2 Preventing, Limiting, and Redressing Privacy Harms

Defining privacy harm is very difficult, and panelists were not able to reach a consensus on which uses of information are "harmful" or "acceptable."[42] While defining the ends of the spectrum on appropriate use would be fairly straightforward (i.e., uses that are clearly good and uses that are clearly harmful), defining the middle of the spectrum would be very difficult. In the middle of the spectrum, what one community would define as a harmful or acceptable use of information could be different from how another community would define it.[43] Furthermore, the definition of harmful or acceptable use could change over time.[44] The combination of the lack of definition of harmful use and inability to predict what future uses could be harmful creates challenges in developing policies to prohibit harmful use of information.

4.3 The Complex Health Information Legal Landscape

Efforts to appropriately address health big data confront a complex legal landscape, which continues to confuse everyone: patients, providers, health IT developers and other stakeholders in the big data ecosystem, including mobile app developers. Traditional healthcare entities, and the information they collect and generate, are governed by the HHS Office for Civil Rights' (OCR) and State Attorneys General enforcement of the HIPAA Privacy and Security Rules. However, a great deal of health-related information is now generated and consumed outside of this HIPAA-regulated space.

Whereas covered entities and their business associates are bound by HIPAA's Privacy and Security Rules, non-covered entities are subject to different legal obligations, which include the FTC's consumer protection authority to combat unfair or deceptive trade practices under Section 5 of the FTC Act.[45] The exact same health-related information is regulated differently based on the entity processing the information. Additionally, information flowing between HIPAA and non-HIPAA environments may face both sets of laws and regulations.

Finally, state consumer protection laws based on similar principles of deception, enforced by State Attorneys General, as well as State HIPAA-like laws, add an additional layer of complexity. Consequently, privacy and security risks of non-compliance are difficult to understand without mapping the movement of information and a thorough knowledge of state and federal laws. This results in misapplication of existing rules, which has led to both lost opportunity and increased risk.

---

[42] Deven McGraw, December 8, p. 19; Khaled El Emam, December 5, p. 52.
[43] *See, e.g.,* The Journal of the National Center, Vol 2, No. 3, Spring 1989, regarding the Barrow, Alaska alcohol study, available at: http://www.ucdenver.edu/academics/colleges/PublicHealth/research/centers/CAIANH/journal/Pages/Volume2.aspx.
[44] Khaled El Emam, December 5, p. 52.
[45] HHS and the FTC share concurrent jurisdiction in some cases, such that the FTC's jurisdiction extends to certain HIPAA-covered entities, and the agencies coordinated where appropriate.

4.4 General Suggestions

In summary, the PSWG received many hours of helpful testimony over the course of several days of public hearings and meetings. In assessing the concerns raised about protecting privacy and security in health big data analytics, panelists offered several general suggestions, which follow:

- It is important to allow experimentation for technology and methods to improve. It is also important that organizations, that are initially slow to move, learn how best to take advantage of big data opportunities and realize potential benefits.
- The best approach for protecting privacy is to start with the FIPPs. The FIPPs are flexible yet structured, and can apply to the traditional healthcare sector as well as the emerging consumer applications market.[46]
- Finally, the PSWG might consider three options to address legal gaps:
    - Develop a specific set of principles applicable only to "non-HIPAA healthcare data" (with an obvious ambiguity about what "healthcare data" would mean);
    - Develop a set of principles (through an amendment to the scope of HIPAA or otherwise) that would apply to all healthcare information; or
    - Develop a broader general privacy law that would apply to all personal information (with or without a carve-out for data currently covered by the HIPAA rules.[47]

5 Detailed Problem Statements

As electronic health IT adoption has advanced, large data sets of health information have been amassed from across electronic health records (EHR) technology, health applications, and personal health records (PHRs). The application of big data analytics to these data sets offers promising opportunities for learning and improving patient outcomes, but big data analytics also creates potential problems for maintaining privacy and security. The PSWG's public hearings and meetings yielded high-priority issues that merit the HITPC's attention. This section outlines the key problem areas that the PSWG addresses in its recommendations.

5.1 Potential for Harmful or Discriminatory Practices

During the hearings, among the most often cited concerns about health "big data" is the potential for health information to be collected and used in a way that harms individuals or groups. Discrimination is just one example of a harm that can result from certain analytic uses of health big data. U.S. laws prohibit some discriminatory uses of health information – for example, use of health information to make decisions about health insurance coverage, which is now largely prohibited by the Affordable Care Act – but other discriminatory uses of health

[46] Michelle De Mooy, December 5, p. 28; Deven McGraw, December 8, p. 19.
[47] Kirk Nahra, Moving Toward a New Health Care Privacy Paradigm, Wiley Rein LLP, November 2014, p. 7, http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Kirk_Nahra_Health_Care_Privacy_Paradigm_2014-12-08.pdf.

information 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 information as being "harmful" (for example, marketing and other "commercial" uses). However, there is a lack of consensus on which uses are "harmful," particularly with respect to health big data analytics, as well as an inability to predict which future uses could be harmful and which beneficial, creating challenges to enacting policies to prohibit or place additional constraints on such uses. During the hearings, some presenters expressed concern about the use of algorithms to make decisions about people or communities, and the lack of "transparency" about both the information used to inform these algorithms and precisely how the algorithms are used. Certain practices resulting from the use of algorithms are more insidious when further obscured by a lack of transparency since the harm itself may be difficult if not impossible to detect. Often the existence of harmful bias or practices is only revealed when one understands the process used to arrive at a particular decision.

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

5.2 Two Different Domains of Regulation (HIPAA and "Other") Yields Contradictions and Unpredictability

HIPAA covers many sources of health big data – but not all. Consequently, the U.S. lacks comprehensive, FIPPS-based protections for health data analytics (and analytics leveraging information that on its face is not "health" but is used for health purposes or to infer a health status) in many domains, which is confusing for individuals and imperils trust in health big data. In addition, even when health data analytics is regulated, those rules may not have been written in a way that maximizes the healthcare industry's ability to learn from health information while still protecting it from risks to privacy, confidentiality and security. Three concerns in particular were surfaced by the hearings:

- Access – Individuals often lack the ability to electronically access their personal information from healthcare providers or plans, and therefore, their ability to use and share it with other providers, organizations, and researchers is limited. Even with respect to HIPAA covered entities, which are required to provide this right to individuals, the individual's right of access is often difficult for individuals to exercise.
- Transparency – There is a lack of transparency regarding how holders of personal information use and exchange that information, especially in the big data ecosystem outside of traditional healthcare. This lack of transparency erodes trust and exacerbates the fear of harm or 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 information for "research" purposes that intend to contribute to "generalizable knowledge" (i.e., the greater good).

5.3 Lack of Confidence in De-identification Methodologies and the Risk of Re-identification

De-identification is a useful tool for protecting privacy in big data research – but the healthcare industry over-relies on it as a matter of policy and do not have ways to hold people/organizations accountable for unauthorized re-identification of data or negligently failing to protect data that is vulnerable to re-identification. In addition, de-identification does not address the potential for harmful uses of health big data.

HIPAA has regulatory requirements for de-identification – but there are no such federal requirements for de-identification of health information outside of HIPAA. HIPAA standards for de-identification are often voluntarily used by non-HIPAA covered entities, but they are 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 information); the expert methodology is required to take "context" into account but there are no objective criteria governing it.

Furthermore, there is increased risk of re-identification when data sets are combined (the mosaic effect). A mosaic effect occurs when disparate threads can be pieced together in a way that yields information that is supposed to be private.[48]

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

Conversely, de-identification is also not the panacea for enabling valuable uses of information. Emphasizing (or favoring, through reduced regulatory requirements) data de-identified pursuant to HIPAA as the enabling mechanism for data use often significantly reduces the potential for valuable uses of information even where the risk associated with the use of more identifiable information is very low. In addition, de-identification using the expert methodology, which is generally believed to be both more effective at reducing re-identification risk (because it accounts for context) and more valuable for researchers (because it does not per se require the removal of certain data fields) is perceived by many research entities to be too expensive and time intensive.

[48] *See* Office of Management and Budget, M-13-13, Open Data Policy – Managing Information as an Asset, https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

5.4 Security Threats and Gaps

The lack of an end-to-end secure environment for health information was a problem mentioned by many who presented – but no entity (or federal agency) is responsible for assuring those end-to-end protections. Instead the U.S. has silos of information protections. For example, HIPAA Security Rule coverage applies in some places, the FTC and Food and Drug Administration (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 broadly seen as a significant risk for deteriorating patient and consumer trust in the healthcare system and in entities involved in health big data analytics both inside and outside of healthcare. The call for such end-to-end security requirements was referenced as one of the highest priorities.

In addition, existing laws do not necessarily provide incentives for adopting privacy-enhancing technical architectures for big data analytics (for example, data enclaves). In other words, the privacy rules governing analytic uses of data arguably are the same regardless of the technical architecture used to analyze the data.

Congress is the only policy-making body equipped to authorize national security and/or cybersecurity requirements that would facilitate the requirement to provide a consistent baseline level of security for health information, regardless of the entity that holds that information, in an end-to-end environment that is desirable for building trust. But the Workgroup did not recommend specifically directing Congress to address this issue at this time.

6 Solutions and Recommendations

6.1 Address Harm, Including Discrimination Concerns

To address discriminatory practices: without a national consensus on what constitutes harm with regard to health big data analytics, the Workgroup encourages **ONC and other federal stakeholders to conduct more public inquiry and pursue or promote initiatives or projects that could yield greater understanding of the scope of the problem and the potential for harm - both harm to individuals and harm to communities or subpopulations**.[49] While there are some voluntary efforts at creating trust frameworks across institutions for health

[49] Data for Health: Learning What Works, Robert Wood Johnson Foundation, April 2, 2015, p. 3, 29, http://www.rwjf.org/content/dam/farm/reports/reports/2015/rwjf418628.

information research and analysis, that we've shown in the footnotes,[50] there is some concern that purely voluntary efforts will be neither widespread enough nor rigorous enough to adequately protect against some of the potential harms; or that those voluntary efforts will lack the force of law. **Therefore, federal stakeholders should continue to focus on identifying gaps in legal protections against what are likely to be an evolving set of harms from big data analytics.**

Additionally, policymakers should adopt measures (for example, spending conditions, regulations, and guidance) that increase transparency about actual health information uses and convene multi-stakeholder work groups/hearings to help establish a consensus on privacy harms, particularly those that are likely in the era of Big Data. Greater education and knowledge about actual health information use and a convening of multi-stakeholder dialogues could help spur greater public dialogue about which uses are harmful, and as a result advance a national consensus around harms and the best ways to prevent or hold entities accountable for them, while also identifying uses of health data that are not likely to lead to these harms.

With respect to addressing distrust in big data algorithms, the Workgroup expressed a desire to have greater transparency about algorithms – for example, what data informs them, how the data are collected, how those data are weighted or used in the algorithm, and whether (and if so, how) the algorithm is evaluated with respect to the accuracy and fairness of its outcome. At the same time, the Workgroup recognizes that many algorithms are considered to be proprietary and frequently are machine-generated, so there is less than complete understanding of the inputs and the processes even among those using the algorithms. Additionally, the Workgroup recognized that detailing all of the data inputs for a given algorithm may, in many cases, be a near impossible task given the ephemeral nature of the data input and the volume of data utilized. **Nevertheless, the Workgroup recommends policymakers explore ways to increase transparency around use of algorithms, perhaps with an approach similar to that used in the FCRA**. The FCRA is a federal law that, among other things, regulates consumer reporting agencies (CRAs) and empowers people by providing transparency about the use of

[50] See, e.g., HealtheWay/Sequoia Project, http://sequoiaproject.org/ (one example of a multi-stakeholder trust community for health information exchange); Community Research Boards supported by the Community-Campus Partnership for Health, http://www.cph.info (a method for accounting for the impact on a community of research, not just the impact on the individual, and to thereby build trust); PCORnet, http://www.pcori.org/research-results/pcornet-national-patient-centered-clinical-research-network (an example of a multi-stakeholder trust community in support of research). See also Toolkit for Communities Using Health Data: How to collect, use, protect, and share data responsibly, NCVHS, May 2015, http://www.ncvhs.hhs.gov/wp-content/uploads/2013/12/Toolkit-for-Communities.pdf.

consumer credit information where the credit information is used in algorithms to create a credit score.[51]

Any such policymaker action regarding algorithms should aim to maximize transparency (to the extent possible), validity and fairness. Although this may be desirable for algorithms used in a range of contexts, it is particularly important where algorithms are used on/against health data to evaluate and/or make decisions that have an impact on individuals or communities.

6.2 Address Uneven Policy Environment

The Health IT Policy Committee has issued previous recommendations urging that holders of health information (and personal information being used for health purposes) implement protections based on the Fair Information Practice Principles (FIPPs) to protect the privacy, confidentiality and security of that information.[52] FIPPs are principles of responsible data stewardship and obligate data holders to adopt reasonable limits and safeguards regardless of whether an individual's consent is sought. FIPPs include provisions to enable individuals to make reasonable choices about the collection, use and disclosure of their health information – but the FIPPs do not focus just on consent as the primary mechanism. HIPAA and other privacy laws are based on FIPPs – but the U.S. healthcare industry lacks FIPPs-based protections for health information outside of the HIPAA environment.

Congress could address this through legislation, but the Workgroup believes such protections could better be achieved through voluntarily adopted codes of conduct, which can be enforced under Section 5 of the FTC Act by the FTC for entities subject to their jurisdiction. 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 ensure enforceability and to more quickly establish dependable "rules of the road" that help build trust in the use of health big data. (Of note: the Health IT Policy Committee has already asked the Consumer Empowerment Workgroup to consider an evaluation effort for consumer-facing health data tools like health

[51] The FCRA offers the consumers the following protections: must be told if his/her information has been used against them (e.g., to deny their application for credit, insurance, or employment); right to know what information is in their file; right to ask for a credit score, and CRAs must respond by disclosing the current or most recent score that has been computed for the consumer, the range of possible scores, all of the key factors (up to four of them) that adversely affected the score, the date it was created, and who provided the score or the file upon which it was created; right to dispute incomplete or inaccurate information in their credit file; CRAs must correct or delete inaccurate, incomplete, or unverifiable information; CRAs may not report outdated negative information and may not use such information in computing credit scores; access to and the use of a consumer's file must be limited; must give consent for reports to be provided to employers; may limit "prescreened" offers of credit and insurance based on information in their credit report; may seek damages from violators. *See* A Summary of Your Rights Under the Fair Credit Reporting Act, https://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf; see also 15 U.S.C. § 1681-1681x.

[52] *See* HITPC Transmittal Letter, September 1, 2010, https://www.healthit.gov/sites/faca/files/hitpc_transmittal_p_s_tt_9_1_10_0.pdf; *see also* HITPC Transmittal Letter, October 18, 2011, http://www.healthit.gov/sites/default/files/pdf/HITPC_Privacy_and_Security_Transmittal_Letter_10_18_11.pdf.

mobile apps.)[53]. In order to credibly meet the requirements of both protecting sensitive personal information and enabling its appropriate use, Codes, at a minimum, must include transparency, individual access, accountability, and use limitations.

They could also reward/promote the use of privacy enhancing architectures for big data analytics, such as data enclaves. A data enclave is a controlled, secure environment in which eligible researchers can perform analyses using restricted data resources.[54] Finally, the inclusion of accountability or risk review mechanisms such as through community or risk review boards should be considered.[55]

Policymakers also should evaluate existing laws, regulations, and policies (rules) governing uses of data that contribute to a LHS to ensure that those rules promote the responsible re-use of data to contribute to generalizable knowledge. The HITPC had previously recommended treating certain research uses of data conducted under the management and control of a HIPAA covered entity as operations (not requiring consent or IRB review), and the PSWG reiterates that recommendation.[56] Policymakers also should modify rules around research uses of data so that they provide incentives for entities to use more privacy-protecting architectures (for example, entities using secure data enclaves for research would not need to undertake as significant a level of de-identification).

Individuals should have strong access rights to their health information, sufficient to enable individuals to access, download, and transmit their health information as easily as they can access their financial information, for their own use or to facilitate research into conditions/diseases that impact them or in any area of learning that they seek to support. This will require **creating a "right of access" in entities not covered by HIPAA as part of the voluntary codes of conduct referred to earlier; it also will require strengthening HIPAA over time to bring it into the digital age**.

In the meantime, education of individuals, healthcare providers, technology vendors and other stakeholders about the limits of current legal protections, and about best practices to protect the privacy, confidentiality and security of health information is critical, particularly given the patchwork of regulation and the lack of comprehensively adopted, robust codes of conduct. The Health IT Policy Committee recently endorsed recommendations from this Workgroup with

[53] HITPC Meeting, May 22, 2015, http://www.healthit.gov/facas/sites/faca/files/HITPC_PSWG_Meeting_Slides_2015-05-22_Final.pdf.
[54] A data enclave is a controlled, secure environment in which eligible researchers can perform analyses using restricted data resources. *See* National Institute of Health. NIH Data Sharing Policy and Implementation Guidance, http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#enclave.
[55] For example, see the discussion by Al Richmond, MSW, of the Community-Campus Partnership for Health at the National Institutes of Health Precision Medicine Workshop on Patient Engagement, July 1-2, 2015, available at: http://www.nih.gov/precisionmedicine/workshop-20150701.htm.
[56] *See* HITPC Transmittal Letter, October 18, 2011, http://www.healthit.gov/sites/default/files/pdf/HITPC_Privacy_and_Security_Transmittal_Letter_10_18_11.pdf.

respect to providing guidance and educating stakeholders on these topics; we reinforce those recommendations again.[57]

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

OCR should 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. In particular, the Workgroup recommends the development of initiatives or programs to objectively evaluate statistical methodologies to vet their capacity for reducing the risk of re-identification to "very low" in particular contexts. OCR should also grant safe harbor status to those methodologies that are proven to be effective at de-identification in certain contexts, in order to encourage the use of proven methodologies.[58] Finally, OCR should establish risk-based de-identification requirements in circumstances where re-identification risk is very low (through mechanisms other than just treatment of the data), such as when access to data is limited through the use of secure data enclaves or "data havens," where those accessing or holding the data have low-to-no motivation for re-identifying and are prohibited from doing so in an environment where there are strong measures of accountability.

Establishing accountability for re-identification or negligent de-identification also was of interest to the Workgroup. This is another issue that Congress could address; however, the Workgroup does not believe specifically asking Congress to address this is advisable at this time.

6.4 Support Secure Use of Data for Learning

The PSWG seeks a widely-accepted security framework that assures accountability for security at all endpoints. Although the Workgroup believes Congress could address this issue,[59] consistent with its recommendations in Section 6.2 the PSWG instead urges the development of voluntary codes of conduct that also address robust security provisions. In addition, education of stakeholders about cybersecurity risks and recommended precautions is critical, and both the public and private sectors have a role to play in this effort.

---

[57] HITPC Transmittal Letter, August 16, 2011, http://www.healthit.gov/sites/faca/files/HITPC_PSTT_Transmit_8162011.pdf and HITPC Meeting, May 22, 2015, http://www.healthit.gov/facas/sites/faca/files/HITPC_PSWG_Meeting_Slides_2015-05-22_Final.pdf

[58] HITRUST Alliance recently released its views on methodologies for de-identification standards it suggests could be used for health information. https://hitrustalliance.net/de-identification/

[59] The FTC has recommended the enactment of strong, flexible, and technology-neutral legislation to strengthen the Commission's existing data security enforcement tools. *See* Internet of Things: Privacy & Security in a Connected World, FTC Staff Report, January 2015, p. 49, https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf [hereinafter "Internet of Things Report"].

Federal policymakers, through regulations, spending conditions, and guidance, 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.

The Workgroup also reiterates recommendations made by the Privacy and Security Tiger Team and endorsed by the HITPC in 2011[60] with respect to the HIPAA Security Rule. Specifically:

- Security policies for entities collecting, storing and sharing electronic health information needs to be responsive to innovation and changes in the marketplace.
- Security policies also need to be flexible and scalable to reflect differences in size and resources; at the same time a solid baseline of security policies needs to be established and consistently implemented across all entities.
- Providers will continue to need education and specific guidance on how to comply with the security rule.
- HHS should have a consistent and dynamic process for updating security policies and the rapid dissemination of new rules and guidance to all affected. As part of this process, HHS should look to other security frameworks to assure the Security Rule keeps up with the latest threats and innovations in security protections. NIST had previously issued guidance on HIPAA security compliance that many entities have found helpful; NIST should continue to update this guidance and keep it current and relevant for a changing risk environment.

[60] HITPC Transmittal Letter, August 16, 2011, http://www.healthit.gov/sites/faca/files/HITPC_PSTT_Transmit_8162011.pdf.

7 Bibliography

1. A Summary of Your Rights Under the Fair Credit Reporting Act, https://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf.
2. Big Data and Privacy: A Technological Perspective, May 2014, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf
3. Big Data: A Tool for Inclusion or Exclusion?, https://www.ftc.gov/news-events/events-calendar/2014/09/big-data-tool-inclusion-or-exclusion.
4. Big Data: Seizing Opportunities, Preserving Values, May 2014, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.
5. Community-Campus Partnership for Health, http; //www.ccph.info.
6. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap, Draft Version 1.0, http://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf.
7. Data for Health: Learning What Works, Robert Wood Johnson Foundation, April 2, 2015, http://www.rwjf.org/content/dam/farm/reports/reports/2015/rwjf418628
8. Deven McGraw & Alice Leiter, *Risk-Based Regulation of Clinical Health Data Analytics*, Colo. Tech. L. J., Vol. 12.2.
9. DRAFT NISTIR 8053, De-Identification of Personally Identifiable Information, April 2015, http://csrc.nist.gov/publications/drafts/nistir-8053/nistir_8053_draft.pdf.
10. FACT SHEET: Announcing New U.S. Open Government Commitments on the Third Anniversary of the Open Government Partnership, https://www.whitehouse.gov/the-press-office/2014/09/24/fact-sheet-announcing-new-us-open-government-commitments-third-anniversa.
11. For Covered Entities and Business Associates, U.S. Department of Health and Human Services,
12. Guidance Regarding Methods for De-Identification of Protected Health Information, http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html
13. http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/.
14. HITPC Meeting, March 10, 2015, http://www.healthit.gov/FACAS/calendar/2015/03/10/hit-policy-committee.
15. HITPC Meeting, May 22, 2015, http://www.healthit.gov/facas/sites/faca/files/HITPC_PSWG_Meeting_Slides_2015-05-22_Final.pdf.
16. HITPC Transmittal Letter, August 16, 2011, http://www.healthit.gov/sites/faca/files/HITPC_PSTT_Transmit_8162011.pdf.

17. HITPC Transmittal Letter, October 18, 2011, http://www.healthit.gov/sites/default/files/pdf/HITPC_Privacy_and_Security_Transmittal_Letter_10_18_11.pdf.
18. HITPC Transmittal Letter, September 1, 2010, http://www.healthit.gov/sites/faca/files/hitpc_transmittal_p_s_tt_9_1_10_0.pdf
19. HealtheWay/Sequoia Project, http://sequoiaproject.org/
20. Internet of Things: Privacy & Security in a Connected World, FTC Staff Report, January 2015, https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.
21. Kirk Nahra, Moving Toward a New Health Care Privacy Paradigm, Wiley Rein LLP, November 2014, http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Kirk_Nahra_Health_Care_Privacy_Paradigm_2014-12-08.pdf.
22. Markle Connecting for Health Common Framework, http://www.markle.org/sites/default/files/CF-Consumers-Full.pdf.
23. NIH Data Sharing Policy and Implementation Guidance, http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#enclave.
24. NIST National Strategy for Trusted Identities in Cyberspace, http://www.nist.gov/nstic/NSTIC-FIPPs.pdf.
25. OECD Guidelines on the Protection of Privacy, http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

- Office of Management and Budget, M-13-13, Open Data Policy – Managing Information as an Asset, https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf
- Open Government Partnership, http://www.opengovpartnership.org.

26. PCORnet, http://www.pcori.org/research-results/pcornet-national-patient-centered-clinical-research-network.
27. Policy: Privacy and Security Workgroup, February 9, 2015, http://www.healthit.gov/facas/calendar/2015/02/09/policy-privacy-security-workgroup.
28. Policy: Privacy and Security Workgroup Virtual Hearing, December 5, 2014, http://www.healthit.gov/facas/calendar/2014/12/05/policy-privacy-security-workgroup-virtual-hearing.
29. Policy: Privacy and Security Workgroup Virtual Hearing, December 8, 2014, http://www.healthit.gov/facas/calendar/2014/12/08/policy-privacy-security-workgroup-virtual-hearing.
30. Privacy and Security Workgroup Transcript, December 5, 2014, http://www.healthit.gov/facas/sites/faca/files/PSWG_Transcript_Final_2014-12-05.pdf.

31. Privacy and Security Workgroup Transcript, December 8, 2014, http://www.healthit.gov/facas/sites/faca/files/PSWG_Transcript_Final_2014-12-08.pdf.
32. Privacy and Security Workgroup Transcript, February 9, 2015, http://www.healthit.gov/facas/sites/faca/files/PSWG_Transcript_Final_2015-02-09.pdf.
33. Public Hearing Responses of Richard Platt, http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Richard_Platt_Reply_to_Questions_for_Panelists_2014-12-05.pdf.
34. Testimony of CDT for the HIT Privacy and Security Workgroup Virtual Hearing, December 5, 2014, http://www.healthit.gov/facas/sites/faca/files/PSWG_Testimony_Michelle_DeMooy_CDT_2014-12-05_0.pdf.
35. Toolkit for Communities Using Health Data: How to collect, use, protect, and share data responsibly, NCVHS, Draft 8: November 6, 2014, http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Leslie_Francis_NCVHS_Toolkit_2014-12-08.pdf.
36. Virtual HITPC Meeting, June 30, 2015, http://www.healthit.gov/FACAS/calendar/2015/06/30/virtual-hit-policy-committee.
37. White House's 2012 Internet Privacy Framework (Consumer Bill of Rights), https://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

8 Appendix A – Health Big Data Public Hearing Topics and Speakers

8.1 Health Big Data Public Hearings, December 5 and 8, 2014

Day 1 – Friday, December 5, 2014	Day 2 – Monday, December 8, 2014
Panel 1: Health Big Data Opportunities and the Learning Health System (LHS)	Panel 1: Current Law
• **Steve Downs, RWJF** • **Richard Platt, Harvard Pilgrim** • **Patricia Brennan, University of Wisconsin**	• Melissa Bianchi, Hogan Lovells • Kirk J. Nahra, Wiley Rein • Deven McGraw, Manatt, Phelps & Philips, LLC
Panel 2: Health Big Data Concerns	Panel 2: Health Big Data Opportunities
• **Michele DeMooy, CDT** • **Mark Savage, NPWF** • **Anna McCollister-Slipp, Galileo Analytics**	• Linda Avey, 23 and Me, Curios, Inc. (invited but could not attend) • Kald Abdallah, Project Data Sphere • Ella Mihov, Ayasdi
Panel 3: Protections for Consumers	Panel 3: Learning Health System
• **Khaled El Emam, University of Ottawa** • **Bob Gellman, Private Consultant** • **Fred Cate, Indiana University**	• Paul Wallace, Optum Labs • Josh Gray, athenahealth
—	Panel 4: Health Big Data Concerns
	• Leslie Francis, University of Utah • Melissa Goldstein, George Washington University

8.2 Data Security in Health Big Data Hearing, February 9, 2014

Panelist	Organization	Position
Andrei Stoica	IMS Health	VP of Global Systems Development and Security
Denise Anthony	Dartmouth College	Vice Provost for Academic Initiatives, Professor of Sociology; SHARPS contributor
Ryan Anderson	Milliman	Director of Software as a Service

9 Appendix B – Supporting Testimony

Topic: Big Data Opportunities
Testimony:

- Big data is expected to improve our understanding of the efficacy and safety of medical treatments and improve outcomes for the treatment of common diseases.[61]
- Additionally, big data can help us better understand the performance of medical devices like artificial joints.[62]
- Big data is anticipated to provide both personal and community-wide benefits. On an individual level, big data can advance personalized medicine by providing evidence for vaccine safety and providing insight into which treatments may work best for certain people.[63] On a community level, big data is expected to advance population health and improve community-wide care[64] by understanding how conditions like asthma, obesity, high blood pressure, and diabetes are concentrated in specific communities or grouped by age, gender, or other characteristics.[65]
- Big data will also continue to enable public officials to identify and track the spread of infectious diseases and respond in a timely manner.[66]

Topic: Big Data Challenges
Testimony:

- In a big data world, almost any kind of data can be health-related data.[67]
- Difficult to define privacy harm because the concept of harm is often subjective and dependent on context.[68]
- Big data introduces new risks of re-identification due to the volume of data and the broad variety of data sources in the big data ecosystem.[69]
- Data security is a crucial challenge. This is driven by the complexity in software and hardware that is used in the healthcare environment, which leads to greater vulnerabilities.[70]

---

[61] Public Hearing Responses of Richard Platt, p. 3, http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Richard_Platt_Reply_to_Questions_for_Panelists_2014-12-05.pdf [hereinafter "Richard Platt Responses"].
[62] Richard Platt Responses, p. 3.
[63] Richard Platt Responses, p. 3.
[64] Toolkit for Communities Using Health Data: How to collect, use, protect, and share data responsibly, NCVHS, Draft 8: November 6, 2014, http://www.healthit.gov/facas/sites/faca/files/PSWG_Presentation_Leslie_Francis_2014-12-08.pdf/
[65] Richard Platt Responses, p.3.
[66] Richard Platt Responses, p.3.
[67] Stephen Downs, December 5, p. 20; Mark Savage, December 5, p. 30; David McCallie, Jr., December 8, p. 21.
[68] Michelle De Mooy, December 5, p. 44.
[69] Michelle De Mooy, December 5, p. 30.
[70] Andrei Stoica, February 9, p. 6.

- Big data is not precisely defined, which makes it more difficult for legislators to enact laws that regulate big data collection, use, and analysis.[71]
- Additionally, algorithms that rely on big data and that are applied to make decisions about individuals, which may shape their behaviors and opportunities, are not well understood by the public. This lack of transparency, which was repeatedly cited as a concern, creates the potential for undetected discrimination to occur, which could reduce public trust in data exchange.[72]

Topic: Concerns About Tools Used to Protect Privacy
Testimony:

- All data can be health data, or data from which inferences about health are drawn or correlations with health are made.[73]
- Apps and wearables collect data that provide insights into peoples' day-to-day health. These data include "steps, workout, sleep, mood, pain, menstrual cycles and heart rate. Recent products also include hydration, stress and breathing rates and patterns. Still others are able to infer health experiences by analyzing data such as location, movement and social activity, not typically considered health data."[74]
- Some say the FIPPs are unsuited for the era of big data (e.g., analytical methods are putting pressure on traditional principles such as confidentiality, security, individual participation through meaningful patient consent, transparency and data minimization (including collection, use, and purpose limitation).[75]
- Nevertheless, presenters defended the FIPPs, stating they still provide "a strong, standardized structure that promotes responsible and efficient use of data while allowing for innovations in analytics and application."[76]

Topic: De-identification
Testimony:

- De-identification does not eliminate the risk of re-identification.[77] Nevertheless, if de-identification is done well, the risk of re-identification can be very low.[78]
- While some research organizations indicated their satisfaction with de-identified data sets,[79] others stated that sometimes it is necessary to use fully identified

---

[71] Robert Gellman, December 5, p. 51.
[72] Michelle De Mooy, December 5, p. 30.
[73] Stephen Downs, December 5, p. 20; Mark Savage, December 5, p. 31; D. McCallie, JR., December 8, p. 21.
[74] Stephen Downs, December 5, p. 8.
[75] CDT Testimony, p. 3.
[76] CDT Testimony, p. 3.
[77] Michelle De Mooy, December 5, p. 30.
[78] Khaled El Emam, December 5, p. 48.
[79] See testimony of Ella Mihov (Ayasdi), December 8, p. 32, 36 and Kald Abdallah (Project Data Sphere, LLC), December 8, p. 31, 36.

data (e.g. when electronic health data must be matched to an external source like the National Death Index).[80] Some stated that HIPAA's safe harbor de-identification method may not give researchers the data they need or want. Limited data sets are slightly more robust, but still may or may not be sufficient for research needs.[81]

- Workgroup agreed that when identifiable information is used, "it should be stored in highly protected locations like data enclaves."[82]
- Accumulating evidence exists that the safe harbor method has some important weaknesses that would allow data to be shared with a higher risk of re-identification. These risks include a reduction in the data utility and the consequence that under certain conditions, safe harbor allows data to be shared with a higher risk of re-identification. Safe harbor is being copied and used globally, so HHS should re-examine the value of such simple standards and provide additional guidance to limit situations when simple standards are applied.[83]
- No widely accepted standards for expert determination method and there is no homogeneity in how de-identification is actually done. Standards are needed to raise the bar in de-identification. Creating standards for the expert determination serves multiple purposes. These include (1) ensuring that methods are known, published, and scrutinized, and (2) creating a professional community of practice based on certification that could facilitate the development of more sophisticated methods and practices.[84]
- Participants echoed the rise in demand for standards, as a lack of guidance is inhibiting willingness to share data and IRBs are uncomfortable evaluating privacy issues in the face of conflicting advice.[85] The HITRUST alliance is already working on a general health standard for de-identification.[86]
- Additionally, as experience with expert determination grows, one could account for "the value of information in different settings" and balance whether to use experts to de-identify data or mimic or replace their processes with a degree of automation.[87] This works when experts can anonymize in a fixed space with known data elements, but the process may require change when external data elements are introduced (as mosaicking may increase re-identification risks).[88]
- De-identification can be enhanced by other controls. These include contractual controls (e.g., prohibiting the joining of data sets), privacy and security controls at

[80] Richard Platt, December 5, p. 10.
[81] Melissa Bianchi, December 8, p. 6.
[82] Richard Platt, December 5, p. 10.
[83] Khaled El Emam, December 5, p. 49.
[84] Khaled El Emam, December 5, p. 49.
[85] Khaled El Emam, December 5, p. 67 and Fred Cate, December 5, p. 67.
[86] Khaled El Emam, December 5, p. 66.
[87] Paul Wallace, December 8, p. 55.
[88] McCallie, Jr., December 8, p. 56.

recipient sites, and good governance mechanisms, such as ethics committees or data access committees, which determine acceptable uses of data.[89] Additionally, organizations can adopt privacy architectures, such as "safe havens" or data enclaves, and organizations can embrace distributed computation, which avoids risks associated with pooling data by performing analysis at the data sources.[90]

- Several presenters suggested the need for legal controls that prohibit and provide penalties for re-identification, especially since de-identification cannot be eliminate all risk of re-identification.[91] They thought that the Congress would need to address accountability for re-identification or negligent anonymization/de-identification.[92]

Topic: Patient Consent

Testimony:

- A patient's meaningful consent to authorize the use and sharing of personal health information is a valuable tool for protecting privacy and individual autonomy.[93] Individual control of data through informed consent has both advantages and disadvantages.[94] Consent empowers patients to control their information and take a more active role in their health, but consent also enables patients to withhold information, which can make data sets less valuable.[95]
- Presenters disagreed over the degree to which people want control over their health information.[96]
- Consent is a privacy and security issue for both providers in the HIPAA environment as well as for app developers and wearable device manufacturers outside the HIPAA space.[97] While HIPAA provides for certain expected uses of data that do not require consent (e.g. sharing for treatment, payment, and healthcare operations among covered entities), rules for consent outside the HIPAA space are less structured and rely on the FTC to protect consumers by working to prevent unfair or deceptive acts or practices.[98] In a big data analytics world, it is becoming more difficult to obtain meaningful consent because secondary uses of data may not be contemplated or anticipated, as the data itself can generate the hypotheses.[99]

[89] Khaled El Emam, December 5, p. 51-52.
[90] Khaled El Emam, December 5, p. 51-52.
[91] Michelle De Mooy, December 5, p. 29; Mark Savage, December 5, p. 38, 41; Fred Cate, December 5, p. 63.
[92] Fred Cate, December 5, p. 63.
[93] Stanley Crosley, PSWG Transcript 2015-01-26, p. 17.
[94] Richard Platt, December 5, p. 26.
[95] Mark Savage, December 5, p. 32.
[96] Fred Cate, December 5, p. 64; Robert Gellman, December 5, p. 64.
[97] Michelle De Mooy, December 5, p. 28.
[98] Kirk Nahra, December 8, p. 11.
[99] Stanley Crosley, January 26, p. 17.

- Presenters disagreed over the degree to which people want to and can control their health information. One presenter stated that it is not possible to obtain individual consent for all uses of an individual's data, and it may be impossible to notify every person about all the uses of their data.[100] Additionally, the length and complexity of privacy policies (which few people read) often makes consent meaningless.[101] Other presenters offered that current privacy laws are overly focused on individual control.[102] They urged that it is nearly impossible to expect that people will be able to control their own data.[103] Moreover, privacy is too valuable and important to expect individuals to shoulder the burden of policing themselves.[104] Nevertheless, others argued that new technologies can enable organizations to economically ask people for their consent, and more thought should be given to a person's ability to opt-out or opt-in to research.[105]
- Some argued that society has a collective right, expressed through law and regulation, to automatically include people in important research for the greater public good without asking for consent.[106] A study was cited, which revealed that people are often willing to contribute their data to research as long as their identity is protected.[107] Consequently, transparency may be a preferable strategy to engage individuals rather than consent.[108]

Topic: Data Security

Testimony:

- The security threat landscape changes constantly over time. These evolving security threats are driven by vulnerabilities that arise from designing and deploying highly complex software and hardware. Ultimately, there is no such thing as zero risk.[109]
- In response to this complexity, organizations should adopt a balanced, holistic approach to security that looks at operations end-to-end and applies a risk-based framework. This holistic approach should include considerations like physical security.[110]

[100] Richard Platt, December 5, p. 10 (Additionally, offering a universal opt-out may be undesirable because it would create unreliable answers from the data, which is a data quality concern).
[101] Fred Cate, December 5, p. 65.
[102] Fred Cate, December 5, p. 52.
[103] Fred Cate, December 5, p. 52.
[104] Fred Cate, December 5, p. 53 (Throughout Workgroup discussions, people commented that consent places a significant burden for privacy on the individual; see Stanley Crosley, January 26, at 17; see also Deven McGraw, February 9, at 26).
[105] Robert Gellman, December 5, p. 65.
[106] Robert Gellman, December 5, p. 66
[107] Stephen J. Downs, December 5, p. 15.
[108] See PSWG Meeting Slides, January 26, 2015, at 11, http://www.healthit.gov/facas/sites/faca/files/PSWG_Meeting_Slides_2015-01-26_v9.pptx.
[109] Andrei Stoica, February 9, p. 6.
[110] Andrei Stoica, February 9, p. 6.

- HIPAA defines high-level objectives, but panelists stated the need for a risk-based framework that defines very specific, contextual, and evolving controls that are applied to reduce risk to an acceptable level. "The only pragmatic way to secure data in healthcare and in any other domain is to consistently follow an industry developed risk-based framework."[111] HITRUST is an example of a common security framework that the healthcare community may consider applying.[112]
- HIPAA defines high-level objectives, but what is needed is a risk-based framework that will define very specific, contextual, and evolving controls that will be applied to reduce risk … to an acceptable level."[113] "The security objective should be based on outcomes, not the means, because the means (e.g., the hardware, the software, the attack mitigation) change constantly."[114]
- Moving to a common framework will be difficult for many organizations. Specifically, it will be challenging for organizations that do not have an IT department and rely on outsourcing, but it will be easier for organizations with sophisticated IT operations.[115] If an organization employs a good computer science approach, which involves backing up machines, firewalls, and antivirus software on desktops, then it should be a medium effort to achieve good security.[116]
- HIPAA compliance varies significantly across hospitals based on their levels of resources.[117] The resources go beyond IT sophistication to hospital infrastructure, and staffing.[118] Consequently, any regulatory incentive or effort must acknowledge that compliance varies across hospitals and providers.[119]
- Organizations can mitigate risk by adopting privacy architectures, such as "safe havens" or data enclaves. Even within these enclaves, data should be de-identified because the risk of re-identification is not eliminated.[120]
- Finally, distributed computation and distributed networks may augment good security practices. One participant testified that "[d]istributed data networks minimize the need to aggregate individual data[,] are increasingly powerful[,] and should be considered when they are appropriate. These methods move the analyses to the data systems that already possess the data and return results that can be combined across multiple sites. The Food and Drug Administration (FDA),

[111] Andrei Stoica, February 9, p. 6 (The view that specific controls are needed was not shared by everyone; in some cases, such specificity may be inconsistent with the FTC's approach).
[112] Andrei Stoica, February 9, p. 14.
[113] Andrei Stoica, February 9, p. 6.
[114] Andrei Stoica, February 9, p. 6.
[115] Andrei Stoica, February 9, p. 15, 21 (stating that "there is a huge impediment and huge cost differential for security, but as you go [down] a path to critical mass and you have a decent sized IT operation … then it becomes easier and easier….").
[116] Andrei Stoica, February 9, p. 15.
[117] Denise Anthony, February 9, p. 10.
[118] Denise Anthony, February 9, p. 16.
[119] Denise Anthony, February 9, p. 10.
[120] Khaled El Emam, December 5, p. 51-52.

the National Institutes of Health (NIH), and the PCORI have created distributed data networks to support some of their needs."[121] Others cautioned that some distributed networks do not have security evaluations or security proofs and it would be important to perform such proofs and evaluate security protocols before leveraging distributed computation systems.[122]

Topic: Transparency

Testimony:

- People are not fully aware of how their data are used and for what purpose; this extends to a common assumption that HIPAA covers all medical data when in reality it does not.[123]
- As a matter of ethical importance, transparency is crucial if data is used without a person's explicit consent or authorization.[124]
- One participant explained that entities should provide notice whenever individuals may think the usage or collection of data is unexpected or objectionable, and notice should be provided at a relevant time. Contextual (just-in-time) notice helps clarify consumer expectations. Such notice should explain what type of data is collected, when it is collected, what it is used for, the secondary uses contemplated, how long it will be retained, and what security measures are in place. That said, current notices are overly broad and vague, and they are drafted in highly technical language. Consequently, people do not read or understand notices, and they do not end up fostering transparency. Notices are drafted in highly technical language and are so vague that people do not read or understand them, so transparency is rarely achieved.[125]
- Without clear ground rules in the non-HIPAA space, organizations are less transparent about their data practices, and this extends to the use of algorithms, which are "crucial decision-making mechanisms." "Algorithmic transparency is crucial.... Many companies have entered the health data space and they consider their models proprietary and refuse to reveal them, which leaves a gaping hole where our understanding of these decision-making mechanisms should be."[126] Because sophisticated algorithms are proprietary intellectual property, it is very difficult to determine their inputs and outputs, and how they make decisions about people.[127] Moreover, "[a]lgorithms have become extremely sophisticated

[121] Written testimony of Richard Platt, p. 4, http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Richard_Platt_Reply_to_Questions_for_Panelists_2014-12-05.pdf.
[122] Khaled El Emam, December 5, p. 54.
[123] Michelle De Mooy, December 5, p. 37.
[124] Fred Cate, December 5, p. 72.
[125] CDT Testimony, p. 3; *see also*, Fred Cate, December 5, p. 69.
[126] Michelle De Mooy, December 5, p. 29-30.
[127] Michelle De Mooy, December 5, p. 36.

and nuanced to the point where they are [replacing the] human decision-making processes."[128]

- Another participant suggested that transparency and disclosure should extend to "what [data] informs the algorithms, how … cohorts are defined, and how individuals are separated. If that's opaque, … then nobody will ever trust the system."[129]
- Another participant drew parallels to transparency provisions in the Fair Credit Reporting Act (FCRA). When the FCRA was introduced, Credit Rating Agencies said that people were not asking for their credit information; nevertheless, access rights were put in the law. Today, 91% of people surveyed by the participant stated that it was important to find out to whom their personal information had been disclosed.[130] Although the FCRA is a statute that frames acceptable uses of data, it provides consumers with transparency if data has an adverse impact on them. Some cautioned that FCRA is tailored to particular circumstances and it may not scale well in the health arena.[131]

Topic: Collection, Use, and Purpose Limitation

Testimony:

- Organizations should ask themselves why they need the information they have collected, and they should avoid retaining data for some future, unnamed use simply because they think it might be valuable. Specifically with regard to health data, there should be a requirement to delimit the collection and use of data, and it is not acceptable to retain data for an unknown purpose.[132]
- Concerning use limitation, participants discussed the difficultly in clearly defining acceptable and unacceptable uses of health information in big data analytics.[133] Nevertheless, one panelist offered the following recommendations for processing electronic health data:
    - As a general principle, the minimum necessary amount of identifiable data should be used to answer a question;
    - There should be good processes for approval and oversight; and
    - The uses of data should be stated publicly and the number of individuals who have access to identifiable data should be minimized.[134]

Topic: Privacy Harms

Testimony:

- It is very difficult to define or put a frame around what is privacy harm.[135] It is similarly difficult to define acceptable uses of data because such an evaluation is

[128] Michelle De Mooy, December 5, p. 35.
[129] Anna McCollister-Slipp, December 5, p. 36.
[130] Denise Anthony, February 9, p. 13.
[131] Kirk Nahra, December 8, p. 17.
[132] Michelle De Mooy, December 5, p. 42.
[133] Khaled El Emam, December 5, p. 52.
[134] Richard Platt, December 5, p. 10.

naturally subjective, culturally specific, and will change over time.[136]

- Current rules, such as HIPAA and the Common Rule, cover permitted uses, but they do not enumerate "non-permitted abuses."[137] One participant stated that commercial use of personal information, without a clear disclosure, could be viewed as harmful. Additionally, any sort of discrimination or denial of opportunity, such as the loss of employment or insurance, or any public embarrassment would be classified as harmful.[138] Still others distinguished between focusing on harms and focusing on rights; thus, any collection, compilation, or sharing of data in violation of a person's rights could be harmful, even if the harm is not immediately visible.[139]
- In one survey, when people were asked about their greatest concern regarding the use of their health information, their top concern was that they would be contacted.[140]
- To arrive at a consensus around harms or non-permitted abuses, an effort could be made to identify laws that may already prohibit certain activities, identify gaps, and catalogue federal laws with non-discrimination provisions.[141] Additionally, the FTC can help can help identify boundaries through the cases it pursues under its ability to combat unfairness and deception.[142]

Topic: Complex Legal Landscape

Testimony:

- There continues to be a lack of clarity and understanding of privacy and security laws and rules.[143]
- Legal coverage is extensive and even contradictory in some areas (e.g., research under HIPAA and the Common Rule), while there are significant gaps in coverage other areas.[144]
- Many state laws add complexity, and these laws can be confusing, outdated, and seldom enforced.[145]

---

[135] Michelle De Mooy, December 5, p. 44.
[136] Khaled El Emam, December 5, p. 52.
[137] David McCallie, December 8, p. 21.
[138] Stephen Downs, December 5, p. 22.
[139] Robert Gellman, December 5, p. 66.
[140] Fred Cate, December 5, p. 55.
[141] Melissa Bianchi, December 8, p. 21.
[142] Michelle De Mooy, December 5, p. 48.
[143] See Anna McCollister-Slipp, December 5, p. 33; Michelle De Mooy, December 5, p. 37 ("… people think … HIPAA really covers all medical data and have … no idea or understanding that it doesn't in a lot of circumstances"); Michelle De Mooy, December 5, p. 37 (outreach to app developers is needed to ensure they understand their ethical responsibilities).
[144] Deven McGraw, HITPC Transcript, March 10, 2015, p. 19; see also, Moving Toward a New Health Care Privacy Paradigm, available at http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Kirk_Nahra_Health_Care_Privacy_Paradigm_2014-12-08.pdf.
[145] Kirk Nahra, December 8, p. 12.

- HIPAA applies only to covered entities (health plans, healthcare clearinghouses, healthcare providers) and business associates acting directly on their behalf. The bulk of health-related data being generated today falls outside of HIPAA regulation.[146]

Topic: Access to Information

Testimony:

- Generally, there is a need for greater data liquidity. On a personal level, patients want to access and combine their data in meaningful ways, but they face significant impediments. Privacy and security are seen as some of the biggest burdens/barriers.[147]
- Public access to data is also very important. Health data can be viewed as a social asset; there is a social responsibility to give back to the community by making data available to researchers and to patient groups.[148]
- Within HIPAA, patient consent is not required for the use and exchange of protected health information (PHI) for treatment, payment, or healthcare operations purposes.[149] One presenter cited this concept of normal, routine uses as one of HIPAA's greatest strengths.[150]
- Participants cited some progress regarding data liquidity. The recently finalized HIPAA Omnibus Rule introduced important changes, including (1) authorization to permit future research and (2) the ability to permit compound authorizations for research purposes. [151]

Topic: Under-Regulation

Testimony:

- [For the purposes of the PSWG's investigation, "under-regulation" refers to the gaps in law in which health-related data is not afforded the same privacy and security protections that exist under a regime like HIPAA.]
- A rapidly growing amount of health-related information is not regulated by the HIPAA. These include mobile applications, websites, and personal health records.[152]
- The FTC has become the default regulator of privacy and security over the past decade, but the FTC's Section 5 authority only extends to enforcement against organizations engaging in deceptive or unfair acts or practices. A company acts deceptively if it makes materially misleading statements or omissions about a matter, and such statements or omissions are likely to mislead reasonable consumers.[153] A company engages in unfair acts or practices if its practices cause or are likely to cause

[146] Kirk Nahra, December 8, p. 11 (noting the explosion of data created by mobile applications, websites, personal health records, and wellness programs that are not subject to HIPAA).
[147] Anna McCollister-Slipp, December 5, p. 33-34.
[148] Anna McCollister-Slipp, December 5, p. 35
[149] Melissa Bianchi, December 8, p. 5.
[150] Kirk Nahra, December 8, p. 19.
[151] Melissa Bianchi, December 8, p. 6.
[152] Kirk Nahra, December 8, p. 11.
[153] *See* Federal Trade Commission Policy Statement on Deception, appended to *Cliffdale Assocs., Inc.*, 103 F.T.C. 110, 174 (1984).

substantial injury to consumers that is neither reasonably avoidable by consumers nor outweighed by countervailing benefits to consumers or to competition.[154] The FTC has used its authority under Section 5 in cases where, for example, the FTC has reason to believe that a business made false or misleading claims about its privacy or data security procedures, or failed to employ reasonable security measures and, as a result, causes or is likely to cause substantial consumer injury. In the absence of a specific showing of deception or unfairness, however, the FTC cannot, as an enforcement matter, mandate certain basic privacy and data security protections.[155]

- The HIPAA model could be extended to define covered entities in a broader way. For example, in Texas, anyone who touches healthcare data is considered to be covered entity. This, however, alters the legal approach by shifting the analysis from which entities process personal information (i.e., whether the entity is a covered entities) to what kind of personal information is being processed (i.e., whether the data is health data). [156] This change would be difficult in a big data world in which health data is not clearly defined and information flows through many different people who don't necessarily have a direct relationship with the individual. [157]

Topic: Over-Regulation

Testimony:

- [The PSWG used the term "over-regulation" to refer to the multiplicity of laws addressing certain holders of health and health-related data and the extent to which those laws help leverage beneficial uses of health big data.]
- One panelist stated that the HIPAA Privacy Rule does not protect privacy as well as it should, and in fact, HIPAA impedes the use of data for important health research.[158] Others cited HIPAA's strengths, noting that it establishes common rules that apply uniformly; which serves to improve access to information.[159]
- Ideally, entities should not be penalized and disincentives should not be created when organizations contribute to the general knowledge base for healthcare.[160]
- There is an apparent paradox in HIPAA. While the definition of "research" is the same under both HIPAA and Common Rule,[161] different rules about patient consent are applied depending on whether the research results are shared for "generalizable knowledge" or are used for quality improvement purposes and kept within an

---

[154] *See* Federal Trade Commission Policy Statement on Unfairness, appended to *Int'l Harvester Co*., 104 F.T.C. 949, 1070 (1984).

[155] *See generally*, Kirk Nahra, December 8, p. 10-13.

[156] Kirk Nahra, December 8, p. 18.

[157] Kirk Nahra, December 8, p. 17, 18.

[158] Fred Cate, December 5, p. 53.

[159] Robert Gellman, December 5, p. 51.

[160] Deven McGraw, December 8, p. 9.

[161] Melissa Bianchi, December 8, p. 5; see also 45 CFR 164.501 (defining research as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."); see also http://www.hhs.gov/ocr/privacy/hipaa/understanding/special/research/.

organization (i.e., covered under "healthcare operations"). "Two studies that use data for quality improvement purposes using the same data points done to address the same question … by the same institution will be treated as operations if the results are *not* intended to contribute to generalizable knowledge, … but treated as research [requiring consent] if you intend to share the results with others so that learning may occur."[162]

- The question arose, how the learning health system (and data protection) can be advanced if a rule is based on what an entity intends to do with the results, not how data is safeguarded and treated in the underlying research project?[163] People should not be penalized and disincentives should not be created when organizations contribute to the general knowledge base for healthcare in the United States.[164]
- It was noted that the PSWG's predecessor, the Privacy and Security Tiger Team, provided recommendations [165] in the past on the topic of modernizing the Common Rule and creating more consistency with in HIPAA, and these were approved by the HITPC.[166]
- Acknowledging that the learning health system requires more widespread dissemination of information, the Tiger Team recommended that uses of EHR data to evaluate the safety, quality, and effectiveness of prevention and treatment activities should not require consent or IRB approval. Thus, such investigations should not be labeled as research – even if the results are used for generalizable knowledge – because doing so would pose an obstacle to learning. This exemption should be granted when the provider entity retains oversight and control over EHR data. [167]

[162] Deven McGraw, December 8, p. 8, 9; *see also*, Deven McGraw & Alice Leiter, *Risk-Based Regulation of Clinical Health Data Analytics*, Colo. Tech. L. J., Vol. 12.2, p. 435.
[163] Deven McGraw, December 8, p. 9.
[164] Deven McGraw, December 8, p. 9.
[165] HITPC Transmittal Letter, October 18, 2011, available at: http://www.healthit.gov/sites/default/files/pdf/HITPC_Privacy_and_Security_Transmittal_Letter_10_18_11.pdf [hereinafter "October 18 HITPC Recommendations"]
[166] Deven McGraw, December 8, p. 8; *see generally*, October 18 HITPC Recommendations.
[167] October 18 HITPC Recommendations.

10 Appendix C – Other Big Data Related Activities

The PSWG recognized the important big-data-related work that is being performed by both public and private stakeholders, often in partnership. The PSWG strove to complement these efforts and address gaps where they were identified. Below are brief descriptions of some of the efforts that are currently underway.

10.1 Federal Trade Commission Internet of Things Report and Big Data Workshop

In January 2015, the Federal Trade Commission (FTC) released a report on the Internet of Things (IoT).[168] The report, which summarized the discussions from a workshop held by the FTC in 2013, focused on FIPPs-related issues such as security, data minimization, and notice and consent. One of the report's recommendations was that Congress should enact general data security legislation.[169] The Commission also reaffirmed its commitment to strengthen data security enforcement tools, enforce existing privacy laws, educate consumers and businesses, participate in multi-stakeholder groups, and advocate for consumers. In September 2014, the FTC hosted a workshop[170] entitled "Big Data: A Tool for Inclusion or Exclusion?" to examine the potentially positive and negative effects of big data on low income and underserved populations. A report from that workshop will be forthcoming.

10.2 Precision Medicine Initiative

Launched in 2015, the Precision Medicine Initiative aims to "generate the scientific evidence needed to move the concept of precision medicine into clinical practice."[171] Among the Initiative's objectives is a commitment to protecting privacy. The White House intends to accomplish this via a "multi-stakeholder process with HHS and other Federal agencies to solicit input from patient groups, bioethicists, privacy, and civil liberties advocates, technologists, and other experts in order to identify and address any legal and technical issues related to the privacy and security of data in the context of precision medicine." The initiative also seeks to modernize regulations by evaluating what changes are needed to support new research and care models, including a privacy and trust principles framework.[172]

---

[168] Internet of Things Report, p. i.
[169] *Id*.
[170] Big Data: A Tool for Inclusion or Exclusion?, https://www.ftc.gov/news-events/events-calendar/2014/09/big-data-tool-inclusion-or-exclusion.
[171] About the Precision Medicine Initiative, National Institutes of Health, available at: http://www.nih.gov/precisionmedicine/.
[172] FACT SHEET: President Obama's Precision Medicine Initiative, https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative. *See also*, PMI: Proposed Privacy and Trust Principles, https://www.whitehouse.gov/sites/default/files/docs/pmi_privacy_and_trust_principles_july_2015.pdf; Advisory Committee to the NIH Director (ACD) PMI Working Group: Participant Engagement and Health Equity Workshop, Session 8: Interagency Proposed Privacy and Trust Framework for a PMI Cohort Day 2, http://www.nih.gov/precisionmedicine/workshop-20150701.htm

10.3 21st Century Cures

Launched in 2014, the 21st Century Cures initiative aims to "help accelerate the discovery, development, and delivery of promising new treatments and cures for patients and maintain the nation's standing as the biomedical innovation capital of the world."[173] The House Energy and Commerce Committee released an initial discussion document on January 27, 2015,[174] and on May 21, 2015, the Committee unanimously approved advancing the 21st Century Cure Act.[175] The 21st Century Cures Act aims to advance interoperability among patients, researchers, providers and innovators, modernize and personalize healthcare, while encouraging greater innovation and supporting research.[176]

10.4 Federal Health IT Strategic Plan and the Shared Nationwide Interoperability Roadmap

The 2015-2020 Federal Health IT Strategic Plan builds on ONC's previous strategy to advance the widespread adoption of health IT.[177] Under the plan, ONC's vision is that "health information is accessible when and where it is needed to improve and protect people health and well-being," and its mission is to "improve health, healthcare, and reduce costs through the use of information technology."[178] Objective 5B - Accelerate the development and commercialization of innovative technologies and solutions - references big data. For this objective, ONC plans to adopt a strategy to "fund organization learning and research, and promote innovation for new health IT products and solutions" that incorporate "advances in big data, computation and analytic methods, and other scientific discoverers that use health IT securely to help resolve challenging health problems."[179]

ONC's Shared Nationwide Interoperability Roadmap[180] leverages the second goal of the Federal Health IT Strategic Plan; which is to advance secure and interoperable health information.[181] This goal provides the foundation for achieving the balance of ONC's

---

[173] 21st Century Cures Discussion Document, http://energycommerce.house.gov/sites/republicans.energycommerce.house.gov/files/114/Analysis/Cures/20150127-Cures-Discussion-Document-One-Pager.pdf

[174] 21st Century Cures Discussion Document.

[175] Energy and Commerce Cures, http://energycommerce.house.gov/cures

[176] Id.

[177] Federal Health IT Strategic Plan: 2015-2020, http://www.healthit.gov/sites/default/files/federal-healthIT-strategic-plan-2014.pdf [hereinafter "Federal Health IT Strategic Plan 2015-2020"].

[178] Federal Health IT Strategic Plan: 2015-2020, p. 3.

[179] Federal Health IT Strategic Plan: 2015-2020, p. 26.

[180] See Interoperability Roadmap, *supra*.

[181] See Federal Health IT Strategic Plan: 2015-2020, p. 5.

goals.[182] Big data is referenced under the LHS requirement for shared policy and standards that enable interoperability across the health ecosystem. In the 2018-2020 timeframe, ONC plans to participate with stakeholders in a coordinated governance process to *define a policy framework* for the interoperability of clinical data that supports research and big data analyses. In the 2021-2014 timeframe, ONC and stakeholders will continue their coordinated governance process to *define criteria and implementation specifications* to support the interoperability of clinical data to support big data analysis nationwide.[183]

10.5 Patient-Centered Outcomes Research

The Patient-Centered Outcomes Research Institute (PCORI) is a nonprofit, nongovernmental organization established as part of the Patient Protection and Affordability Care Act of 2010.[184] PCORI's mandate is to "improve the quality and relevance of evidence available to help patients, caregivers, clinicians, employers, insurers, and policy makers make informed health decisions."[185] PCORI funds comparative clinical effectiveness research (CER) that will provide evidence to help patients and their caregivers make better-informed decisions. To facilitate more efficient CER that could significantly increase the amount of information available to healthcare decision makers, PCORI has created PCORnet. PCORnet is a national patient-centered research network that seeks to leverage the power of large amounts of data, including from EHR and patients to help draw information from real-world clinical settings to conduct critical CER and other types of studies.

10.6 Secretary's Advisory Committee on Human Research Protections (SACHRP)

The Secretary's Advisory Committee on Human Research Protections (SACHRP) provides expert advice and recommendations to the Secretary on issues and topics pertaining to the protection of human research subjects.[186] SACHRP recently provided recommendations regarding Human Subjects Research Implications of "Big Data" Studies.[187] Some of these recommendations called on the HHS Office for Human Research Protections (OHRP) to provide guidance on: consent waiver standards for research, proposed changes to rules to account for an exemption category for research involving big data, and asked OCR to clarify the extent to which HIPAA applies to big data research.[188]

---

[182] Goal 3: strengthen health care delivery; goal 4: advance the health and well-being of individuals and communities; and goal 5: advance research, scientific knowledge, and innovation. See Federal Health IT Strategic Plan: 2015-2020, p. 5.
[183] Interoperability Roadmap, p. 35.
[184] Patient Protection and Affordable Care Act, Pub. L. 111-148, Mar. 23, 2010, Sec. 6301(b), 124 Stat 727, http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf
[185] Patient-Centered Outcomes Research Institute, http://www.pcori.org/about-us.
[186] http://www.hhs.gov/ohrp/sachrp/
[187] Human Subjects Research Implications of "Big Data" Studies, US Department of Health and Human Services, http://www.hhs.gov/ohrp/sachrp/commsec/hsrimplicationsofbig_datastudies.html.
[188] Id.