Health IT Policy Committee

A Public Advisory Body on Health Information Technology to the National Coordinator for Health IT

HIT Policy Committee
Privacy and Security Workgroup
FINAL Report of the December 8, 2014 Virtual Hearing on Health Big Data

Meeting Attendance: (see below)

Purpose of Hearing: The hearing continued from December 5. Workgroup Chairperson Deven McGraw repeated some of her introductory remarks and referred attendees to the meeting slide deck. Co-chairperson Stan Crosley introduced the presenters.

Current Law

HIPAA and the Common Rule: Melissa Bianchi, Hogan Lovells US LLP, referred briefly to research requirements under HIPAA, Human Subjects Research Requirements, Part 2 substance abuse regulations, and FERPA. She reminded the members of the permissible research mechanisms under HIPAA: an individual's written authorization, a waiver of authorization by an Institutional Review Board (IRB) or privacy board, a data use agreement regarding a limited data set, de-identified information, a certification for data reviews preparatory to research, or special provisions for research using decedent's information. Following the HITECH Rule, a HIPAA authorization may permit future research if the authorization adequately describes the future research such that it would be reasonable for the individual to expect that his/her PHI could be used or disclosed for that purpose. The Privacy Rule prohibits compound authorization for use or disclosure of PHI that authorizes research activities for which treatment is conditioned on signing the authorization and research activities for which treatment is not conditioned on signing the authorization. The Final Rule permits such compound authorizations provided: the authorization clearly differentiates between the conditioned and unconditioned research components; the authorization provides a clear opportunity for individuals to opt-in to the unconditioned component; and the research does not involve psychotherapy notes. The Human Subjects Research Regulations Common Rule applies to any entity conducting federally funded research and institutions obligated under a Federal Wide Assurance (FWA) to adhere to Common Rule restriction, regardless of funding source and any research involving a living individual about whom an investigator obtains data through intervention or interaction with the individual, or identifiable private information.

HIPAA "paradox": Deven McGraw, Manatt, Phelps & Phillips, LLP, pointed out that health care operation includes "conducting quality assessment and improvement activities, including outcomes evaluation and development of clinical guidelines, provided that the obtaining of generalizable knowledge is not the primary purpose of any studies resulting from such activities." Operation also includes population-based activities relating to improving health or reducing health care costs, and protocol development. The Common Rule has the same definition for research. Studies using data for quality improvement purposes using the same data points to address the same question or sets of questions, and are done by the same institution will be treated as operations if the results are only intended to be used internally and treated as research if a primary purpose is to share the results with others. She reminded the members that the HITPC had submitted comments to the Common Rule ANPRM that the use of clinical data to evaluate safety, quality and efficacy should be treated like operations, even if the intent is to share results for generalizable knowledge, as long as provider entity maintains oversight and control over data use decisions. Entities should follow the full complement of fair information practices in using PHI for these purposes. The recommendations provided some examples of activities with clinical data that should be treated as operations, but also acknowledged further work was needed to determine a new line for when analytics with EHR data should be treated under more robust rules. In terms of a fix for the paradox, McGraw suggested modification of HIPAA regulations for data reuse so that regulations more directly address privacy and confidentiality risks. The reuse of data that present greater risk to privacy, confidentiality and security should be regulated. She talked about several factors that may indicate greater risk and called for guidance with respect to the distinction between operations and research and/or when waivers can be granted. Perhaps HIPAA waivers could be granted in order to experiment with different models for protecting privacy in research.

Privacy Laws, including aspects of the Fair Credit Reporting Act: Kirk J. Nahra, Wiley Rein LLP, referred to the explosion of non-HIPAA-protected data. He reflected on an alternative to HIPAA. The FTC has broad authority in general to “prevent . . . unfair or deceptive acts or practices.” But there are no regulations in this area. The FTC has developed enforcement of data security standards (although these are under challenge). It has not to date undertaken broad privacy enforcement in the health care area. It clearly can take enforcement action against statements that are not true – e.g., privacy notices that misstate what is being done with information. He suggested that FDA may have broader authority. The FCRA regulates consumer reporting agencies (primarily) in connection with credit, employment and insurance. Consent is required to report medical information for these purposes (with some disclosure for medical debts). There are some prohibitions on using medical information, such as substance use treatment, for credit purposes (except for debt issues). He went on to present three options. Legislation could be enacted for non-HIPAA health care data. Something that covers all health care data (a general HIPAA) – either through HIPAA or otherwise could be designed, or a broader overall privacy law (with or without a HIPAA carve-out) could be enacted.

Q and A

Crosley asked about compound authorization. Bianchi said that HHS was attempting to ease requirements for research. Under the Privacy Rule, organizations could not compound authorizations. But now an exception for disclosure of PHI for research for the same or similar study is available. This can allow somewhat more leeway for researchers, for instance, to create a depository for research data. Tying authorization to receipt of treatment is prohibited. Nahra referred to tension in the policy debate to ease HIPAA restriction on research and let patients agree to more research.

Responding to a question about prevention of negative effects on disparities, Nahra referred to a model similar to the Fair Credit Reporting Act (FCRA) which would mandate disclosure. There is already a body of rules for insurance. The rules for using data could be better defined. There is currently no regulatory structure for doing so. According to Nahra, the FCRA is probably not expandable. HIPAA could be expanded beyond CEs. Currently, data are flowing via many entities. A Texas law defines anyone who touches health care as a CE. Bianchi indicated that she is opposed to expanding HIPAA per se, although something similar could be established. Baseline security requirements could give comfort. McGraw indicated that she is in favor of applying fair information practices. Data put to health care use should be allowed although certain practices should be held accountable. Someone said HIPAA does not address judgments such as are the data being used for good or bad. The merits are not taken into account. The HIPAA concept could be extended so that certain uses are normal and expected and consent is assumed.

Regarding the prohibition of discriminatory behavior, Nahra said that under HIPAA the use of data for good or bad is not addressed. To do so could be something new for a privacy law. Merit is dealt with under other laws. Bianchi suggested that existing laws be cataloged. Someone pointed out that under HITECH, the patient's right to hold certain information is tied to self-pay.

Gilad Kuperman asked McGraw how the HITPC's comments on HIPAA were received in 2011. McGraw said that the workgroup participants considered operations research as research. IBR approval is generally required for consideration for peer reviewed journals. According to Nahra, true independent researchers are not business associates. Kuperman said that ease of dissemination is important for a learning system. Nahra said that FTC has some authority for data safeguards, but this is being challenged in court. To date, FTC has not indicated an interest in building such principles for apps. During the time when the agency was building requirements for data security, on-going publicity about data breaches muted opposition.

Health Big Data Opportunities

Wearables, consumer generated health data, and genetic data research: Linda Avey, 23andME, Curios, Inc. (wearecurio.us), could not participate.

Big data analysis platform that uses cancer-related de-identified health information: Kald Abdallah, Project Data Sphere, reported on his organization's 3-year effort to address historical barriers to data sharing. It is an independent, voluntary, not-for-profit initiative that provides one place to broadly share, integrate, and analyze cancer trial data from academic and industry Phase III clinical trials using historical, comparator arm and raw anonymized patient level data, data dictionary, protocols and CRFs. SAS is a partner and provides the analytic tools. He said that the true power of the Project Data Sphere® initiative will come from an ever increasing volume of data and the engagement of a diverse global community focused on finding solutions for cancer patients. New members are welcome. For more information, visit www.ProjectDataSphere.org.

Data analytics to segment patient populations, define clinical pathways, and reduce payer and provider frictional costs, Ella Mihov, Ayasdi, presented graphic slides. She talked about her organization's mission. The emphasis is on novel approaches to data analysis. It seeks to identify and apply analytical approaches from other sectors. She opined that health care is catching up with other industries. Health organizations are building warehouses. They are attacking data complexity. She described looking at the shape of data and examining claims data to understand denied claims as well as fraud. An analytical shift is underway.

Q and A

McGraw asked for more information on the shape and structure of data. Mihov described applying thousands of algorithms to data to identify similarities and topological maps to matching patients. In one projects, a population of 20,000 diabetic patients was visualized, showing hot spots of groupings. Although type 2 is a continuum, within group differences are great. This is an example of starting with the data rather than with hypothesis. Abdallah went on to explain that this approach exposes the data to experts in math and other fields who traditionally have had no access to health data. In data driven inquiries, the findings have been unexpected. They are often generalizable.

David McCallie asked about legal mechanisms for acquiring data and wondered about the normalization of data sets. Mihov said that de-identified data were used. Getting data ready for analysis is difficult, but new sets are easily added. Upfront investment requirements are significant. Abdallah explained that in his experience the data providers do the de-identification in keeping with applicable law. Data sets are posted independently. Data from clinical trials are somewhat standardized and therefore less difficult to aggregate compared to data collected for clinical treatment.

Linda Kloss inquired about rules for secondary analysis. Abdallah said that the data are voluntarily shared by participating organization. Each data provider goes through its own required process. The studies are new ones; they are not reproducing research. The oncology community agrees on the importance of extracting all possible knowledge from available data. One example is a crowd sourcing challenge on prostate cancer.

Kuperman asked Mihov about any additional burdens for health data clients. She replied that they are similar to other clients. The unique challenge is interoperability for the continuum of care. The quality of data was often an issue as indicated by patients' editing of their existing data.

Regarding introduction of new types of data, Mihov reported that the challenge to understand EHR data and then combine them with claims is huge and is the current priority. The use of identified data possibly would yield information about subjective concerns and mental state. Abdallah agreed that since there is so much yet to learn from the de-identified data, little thought has gone into the possibility of identified data.

Learning Health System

Sharing information assets, technologies, knowledge tools and scientific expertise: Paul Wallace, Optum Labs, talked about his organization's key assets: very large linked medical claims and EHR data; forums to convene collaboration; translation partners; experts on staff, within partners and alongside Optum; and data visualization power tools. He referred to several policy Implications for the learning health care system. There are opportunities for better Informing and engaging IRBs, refining the relationship between quality improvement and research, and guidance in use of observational methods and de-identification. Data sets consist of more than 2,000 fields with medical and pharmacy claims, lab claims and results, health risk assessments, standardized costs of care, race, income, education level, language preference, geography, and mortality. Expanded with consumer data, additional data fields are available, such as the following: purchase behavior, income, assets, home value, marital status, occupation, home ownership, household composition, ethnicity, travel, various leisure activities, charitable giving, advocacy, volunteering, and community involvement. He said that Optum Labs employs certified de-identified data sets, together with a hashing methodology, to enable matching individuals from multiple sources, yet preserving statistical de-identification.

Perspective of a cloud-based service provider: Josh Gray, AthenaHealth, said that his clients all use the same version of the same software. As a result, real time data can be used to improve performance. This is a unique service. Researchers can easily access the data. Data flows relatively easily across these mostly community doctors who are representative of their national population. Dynamic guidelines are used for tracking patients for standard conditions. Clients can be shown where they can improve performances. Clients can send messages to their patients. Surveillance is 24 hours and weekly reports are used for predictions. Other entities, such as schools and health departments, can use the information. AthenaHealth is a partner with RWJF to examine the impact of the ACA on community doctors. Eventually EHR data will move to the cloud and the time gap from research to care will narrow.

Q and A

Crosley asked about regulation keeping pace with technology and the role of informal guidance. Gray said that only de-identified data are used. Informal guidance could be helpful. It takes time to deal with the variations in state laws. Wallace reported that his organization is conservative in staying within the bounds of the permissible. Because of the complexity of the operation, staff needs to know far in advance what may change.

McCallie asked about prohibitions. Wallace emphasized that only de-identified data are given to researchers. Staff works with researchers in advance to determine what is allowable. Removal of individual data is prohibited. The data never leave this environment. The data cannot be joined with local data. What is allowed is defined by a license. Whether an IRB is required depends solely on the policy of the researcher's organization. The owners of the data remove the direct identifiers before the data are obtained. They are sent an algorithm for the removal process. Since identifiers from the data sets are removed using the same algorithm, they can then be matched and integrated centrally. Gray indicated that he is eager to use a cloud-based infrastructure to identify doctors suitable for certain clinical trials and community trials. His organization is currently assessing the regulatory environment. Wallace said that in 2015 he expects to start bringing in other data. Depending on the problem, perhaps PGHD will be acquired. Learning occurs with each additional data set. One challenge is that de-identification experts are in limited supply. Gray pointed out that SES can reliably be inferred using residential address. He described a project to determine how much newly insured consumers are paying for care.

Regarding sustainability and funding, Wallace said that members have different ways to support projects, such as grants or cost recovery. Gray indicated that his funding sources need to be expanded. There is some funding from foundations and academic institutions.

In general, there is no systematic mechanism for going back and re-identifying a specific patient. That is not a use case. It would have to be an extreme finding. In which case, it could be referred back to the provider.

McGraw referred to keeping policy current with changes in technology and wondered about an option other than government action. What about trusted sector bodies? Gray confessed to a lack of knowledge about specific private sector bodies. According to Wallace, government must be flexible. HIPAA is an example of flexibility and is still a good model. Regarding the possibility of revisiting the concept of safe harbor, Wallace suggested that with more experience the expert method may be used more efficiently.

Health Big Data Concerns

Bioethics and public health law perspective: Melissa Goldstein, The George Washington University, reminded the members that medicine, unlike public health, is concerned with individuals. Public health ethics takes a more communistic emphasis. She referred to (and latter distributed) an article on public health ethics, saying that the fields of public health and ethics take similar approaches. The underlying principles and moral considerations of bioethics and public health ethics are quite similar. There are also similarities with FIPP. She went on to delineate the principles. The first is producing benefits, which is known as beneficence. The second is avoiding and removing harms. The third is producing the maximum benefits over harms. The fourth is distributing benefits and burdens fairly, which is called distributive justice, and ensuring public participation, including the participation of affected parties or procedural justice. Fifth is respecting autonomous choices and actions including liberty of action. The sixth is protecting privacy and confidentiality. Next is keeping promises and commitments or loyalty or fidelity. Number eight is disclosing information as well as speaking honestly and the last one is building and maintaining trust. She described a paper she co-authored on the future of mHealth, saying that the future is likely to be shaped by increasing concern about the need for a proper balance between data practices and innovation. She and her co-author took the position that the policy conversation on mHealth has been quite narrow. She urged the workgroup to go beyond simple notions of privacy and innovation to address the social consequences of health information collection, aggregation and use, and to delve deeper into actual business and employment practices that rely upon such information.

Big data for community health initiatives: Leslie Francis, University of Utah College of Law, reported on a project of the National Committee for Vital and Health Statistics (NCVHS) and the importance of avoiding surprises and suspicion in communities, especially minority communities. The Privacy, Confidentiality and Security Subcommittee designed a stewardship framework for community health data use and a Toolkit for Communities Using Health Data. A draft copy was provided to the workgroup. The Stewardship Kit consists of tools for accountability; openness, transparency, and choice; community and individual engagement and participation; purpose specification; data quality and integrity; data security and de-identification. The appendices include case studies and sample data use agreements. Francis emphasized that community members need to know where to go with concerns. They need information and tools and to understand the life cycle of data. Although there may be data use agreements for transfers, little is often known about what happens downstream. The toolkit is designed for use in communities. There are basic questions that have yet to be answered.

Q and A

McGraw inquired about data use agreements and Gellman's work on making them more enforceable. Francis said that data use agreements are only one component. There is no known instance of a third party bringing suit. Contract law does not allow punitive damages. There may be an absence of follow up with data transfers.

Kross talked about different definitions of community, such as a geographic community and a community of interest. There are many uses that fall outside of HIPAA. Regarding use of data, should one start with the data or the user? Can FFIP be applied? Additionally, there is a range of sophistication. Francis replied that HIPAA may not be a good model. Wherever data are, they should have the same protection. Consistent standards across federal and state boundaries are needed. Goldstein agreed, saying that FIPP is a good place to start.

McCallie asked about harm of exposed data downstream and the requirement for additional protection mechanisms. Frances agreed on the need for more protection; data use agreements are only one tool. FTC pertains to unfair or deceptive trade practice. The fundamental question is whether to work with existing laws or try to obtain new laws. Goldstein observed that society has identified items that cannot be used, such as existing conditions and genetic information, but these requirements do not apply to everything. GINA prohibits collecting certain information. Under employment discrimination law, the subject must first suffer from action and then seek legal recourse. It is much better to outright prohibit certain acts. Goldstein suggested that health data could be tagged so that they cannot be used in other spheres. Francis talked about having a downstream way of determine who has accessed and used data.

Kuperman referred to unprotected information such as an individual's Uber trips to a cancer center and asked about the urgency for change. Francis responded that some individuals think they have more privacy than they actually do. Waiting for a sense of public urgency is not the best approach.

Lucia Savage, ONC, asked about the extent to which the NCVHS toolkit has been used. Francis reminded her that it is draft status. It includes several examples of use. She reminded McGraw that a draft had been distributed to the workgroup in advance of the hearing.

McGraw talked about data protection varying by the holder. She wondered how sectorial law could be made specific. Francis talked about a basic framework of accountability, transparency, and use of data across lifecycle that, in addition, could be more specific.

Wrap-up

McGraw announced that several experts have been invited to present at the next meeting, which is scheduled for December 15.

Public Comment: None

Meeting Attendance				
Name	**12/08/14**	**12/05/14**	**11/24/14**	**11/10/14**
Adrienne Ficchi				
Bakul Patel				
Cora Tung Han	X	X		
David Kotz			X	X
David McCallie, Jr.	X	X	X	X
Deb Bass	X			
Deven McGraw	X	X	X	X
Donna Cryer	X	X	X	X
Gayle B. Harrell		X	X	X
Gilad Kuperman	X			X
Gwynne L. Jenkins				
Helen Caton-Peters	X	X		X
John Wilbanks				
Kathryn Marchesini	X	X	X	X
Kitt Winter	X	X	X	X
Kristen Anderson	X	X	X	X
Linda Kloss	X	X	X	X
Linda Sanches	X	X	X	X
Manuj Lal				
Mark Sugrue				X
Micky Tripathi		X	X	

Stanley Crosley	X	X	X	X
Stephania Griffin		X		
Taha A. Kass-Hout		X	X	
Total Attendees	**13**	**15**	**13**	**14**