September 27, 2019

Possible Privacy, Cybersecurity, and Data Breach Issues in the Proposed National Medical Claims Database Under Section 303 of S. 1895

Peter Swire[1]

Introduction and Executive Summary

This White Paper analyzes possible privacy, cybersecurity, and data breach issues that may arise from a national medical claims database that is being considered by the Congress. Senators Lamar Alexander (R-TN) and Patty Murray (D-WA), the Chair and Ranking Member of the Senate Committee on Health, Education, Labor & Pensions (HELP) introduced "The Lower Health Care Costs Act," or Senate Bill 1895 ("S. 1895" or "the bill") on June 19, 2019. The bill, as amended, was approved by the HELP Committee on a bi-partisan vote on June 26. Titles I and II of the bill are entitled "Ending Surprise Medical Bills" and "Reducing the Prices of Prescription Drugs." Title III of the bill is entitled "Improving Transparency in Health Care," and includes Section 303, titled "Designation of a Nongovernmental, Nonprofit Transparency Organization to Lower Americans' Health Care Costs." The bill contains other provisions as well.

This White Paper solely discusses Section 303. The paper makes no observation about, or takes any position about, any other provision in S. 1895. Nor does the paper take any position about the overall advisability of the proposal in Section 303. Instead, the paper analyzes the current text of Section 303, and seeks to identify possible issues to inform deliberations on the proposed new medical claims database, to be administered under Section 303 by a nonprofit transparency organization ("the Non-Profit").

After introductory material, the White Paper in Figure 1 diagrams the four key stages of how data would flow in the proposed system:

1. Health insurance issuers and others who **supply data** to the Non-Profit:
    a. A first category of risk concerns what happens to individuals and their employers in the event of a data breach by the Non-Profit or a recipient of data from the Non-Profit.
    b. There are other risks that arise as the issuers are required to send claims information to the Non-Profit. For instance, the bill does not appear to authorize data use agreements to protect the data, and may not provide appropriate technical input on how to transfer comprehensive claims data to the Non-Profit.
2. **Processing data within the Non-Profit**:
    a. The Non-Profit would be subject to HIPAA privacy, security, and breach rules, under new rules by the Secretary of HHS ("the Secretary"). The scope of the Secretary's rulemaking authority is not clear, however, especially concerning

[1] Peter Swire is the Elizabeth and Tommy Holder Chair of Law and Ethics at the Scheller College of Business of the Georgia Institute of Technology, and Senior Counsel with Alston & Bird LLP. Support for this research was provided by the U.S. Chamber of Commerce. The views expressed here are solely those of the author, and should not be attributed to the U.S. Chamber of Commerce or any of its members, or to Alston & Bird LLP or any of its clients.

whether HIPAA protections would apply to other entities that receive claims data from the Non-Profit.

3. The Non-Profit exchanges data with **business associates**, who act on its behalf:
    a. The Secretary's rulemaking authority, on its face, does not appear to place the Non-Profit's business associates under HIPAA. The same was true under the original HIPAA rules, but Congress in 2009 ensured that business associates would be subject to HIPAA requirements.
4. Employers, researchers, and others who **receive data** from the Non-Profit:
    a. The bill authorizes a potentially large number of entities to access the claims database, including employers generally. As with business associates, it appears that employers and other authorized users would not be subject to the HIPAA Privacy and Security Rules, and HHS breach notice requirements.

For each stage, the White Paper sets forth the relevant provisions in the current version of S. 1895, and then analyzes possible privacy, cybersecurity, and data breach issues that may arise.

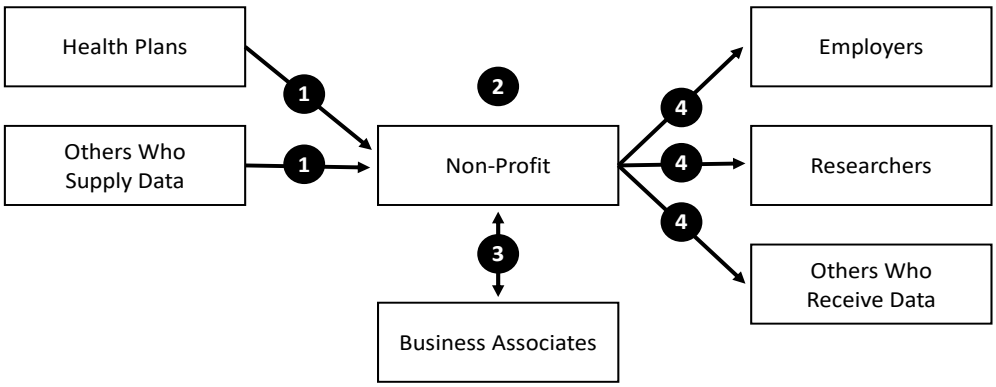
After discussing the stages of data flow, the White Paper turns to a topic already addressed in considerable detail in the bill, **the de-identification and possible re-identification of patients** when information about their claims is provided to the Non-Profit, subject to rulemaking by the Secretary. The White Paper summarizes risks of re-identification under the bill, and provides an Appendix to examine these issues in greater detail. The White Paper concludes with short observations on **miscellaneous provisions** in the current draft of the bill.

Biography of the Author

I am the Elizabeth and Tommy Holder Chair of Law and Ethics at the Scheller College of Business of the Georgia Institute of Technology, and Senior Counsel with Alston & Bird LLP. I have worked extensively on privacy and cyber-security for more than two decades. From 1999 to 2001 I served as Chief Counselor for Privacy in the U.S. Office of Management and Budget ("OMB"). Among my other activities in that role, I was White House coordinator for the 1999 proposed HIPAA Privacy Rule, and for the final Privacy Rule issued in 2000. I have continued to work on many aspects of medical privacy since that time.

Among my publications, I am lead author of the textbook published by the International Association of Privacy Professionals ("IAPP") for certification as a U.S. privacy professional. In 2013, I served as one of five members of President Obama's Review Group on Intelligence and Communications Technology, sometimes called the NSA Review Group. In 2015, the IAPP awarded me its Privacy Leadership Award. In 2018, I was named an Andrew Carnegie Fellow for research on human rights, national security, and cross-border data flows. In 2019, the Future of Privacy Forum honored me for Outstanding Academic Scholarship. Details of my work in the field are available at www.peterswire.net and https://www.alston.com/en/professionals/s/swire-peter.

Figure 1 - Data Flow Mapping



1. Supply data to Non-Profit
2. Process data within Non-Profit
3. Non-Profit exchanges data with Business Associates who act on its behalf
4. Employees, researchers, and others receive data from the Non-Profit

Data Flow Part 1: Health Insurance Issuers and Others Who Supply Data

Data flows under Section 303 begin with individuals, who are often covered under employer-provided health issuers. The health claim of these individuals goes to a health insurance issuer or other party that sends data to the Non-Profit. Figure 1, number 1 illustrates the path of health claims information through the proposed bill. All citations to the bill are to new Section 2796 of the Public Health Service Act (PHSA), to be added by the bill.

Relevant provisions in S. 1895: The Non-Profit shall "collect medical claims, prescription drug claims, and remittance data consistent with the [privacy and cybersecurity] provisions." Section 2796(a)(1)(B). The Secretary of Labor "shall ensure that the applicable self-insured group health plan, through its third-party administrator, pharmacy benefit manager, or other entity designated by the group health plan, as applicable, electronically submits all claims data with respect to the plan," including listed data elements. Section 2796(d)(1). In addition, the states can "require health insurance issuers and other payers to submit claims data to the database," provided that the data is provided to the Non-Profit as required under the bill. Section 2796(b)(4). Transfers from issuers and others to the Non-Profit come within the bill's general requirements for privacy and security, notably to "maintain effective security standards for transferring data or making data available to authorized users." Section 2796(b)(2)(D)(iii).

Analysis: There are **two main categories of risks** that arise from the initial stages of data flow, from individuals to issuers to the Non-Profit. **First are concerns about what happens to individuals and employers in the event of a data breach** at the Non-Profit or an entity that receives data from the Non-Profit. **Second are risks that arise as the issuers are required to send claims information to the Non-Profit.** Overall, the bill is less detailed on the relationship between those supplying the data and the Non-Profit than it is on the relationships between the Non-Profit and those receiving the data. For instance, recipients such as employers and researchers are expected to apply for access to the data, and to sign data use agreements. By contrast, there is no mention of data use agreements – contracts – between suppliers of data and the Non-Profit.

Concerns about what happens to individuals and employers in the event of a data breach:

Section 303 is designed to create a national, comprehensive health care database, containing data about all the health claims made in the United States each year. As we know from experience, data breaches have occurred repeatedly for large databases, both within and outside of the health care system. As of this September, the Privacy Rights Clearinghouse has documented **4,635 health care breaches made public since 2005,** involving 270 million records; breaches for all sectors of the economy during that period cover over 9,000 breaches and over 10 billion records.[2] This experience shows we should plan for the possibility of data breaches in the new system created by Section 303, such as by the Non-Profit or a recipient of data from the Non-Profit.

Our experience with **health care breaches** indicates that such breaches can be **more difficult to remedy** than other types of data breaches.[3] For instance, the breach of a credit card number can be fixed by the individual getting a new credit card within a few days, and with identity theft insurance provided as well. By contrast, once information about a person's

[2] Privacy Rights Clearinghouse, "Data Breaches," (visited Sept. 13, 2019), at https://www.privacyrights.org/data-breaches.

[3] See Tom Garrubba, "5 ways health care data breaches are far worse than financial ones," Healthcare IT News (Nov. 10, 2014), at https://www.healthcareitnews.com/news/5-ways-health-data-breaches-are-far-worse-financial-ones.

sensitive medical procedures become public, that information remains public. In addition, medical identity theft can lead to the fraudster's health information getting mixed with the medical records of the actual patient. The result can be serious medical harm, such as if the blood type is incorrect or a patient allergy to medication is omitted.

As drafted, it is not clear how notices about data breaches would be provided to **individuals whose data is affected.** As an initial point, the bill requires issuers to transfer claims data to the Non-Profit without any requirement to obtain the consent of employers or their employees to such transfer. If a **breach does occur at the Non-Profit**, for instance, it would appear to trigger HIPAA data breach notices to the affected individuals. In that event, individuals would receive notice from an entity they have likely never heard of – the Non-Profit – that information about their health insurance claims has been compromised. This sort of notice could be quite upsetting to individuals, who would learn about a breach of records from a hitherto-unknown entity. In addition, the **breach might occur by any employer, researcher, or other entity** who receives data from the Non-Profit. In that event, the questions from individuals could be even more difficult to answer; for instance, why would some employer that is not the individual's employer be gaining access to detailed health claims information? These sorts of notices could lead to a variety of problems, including loss of trust by individuals in the overall confidentiality of medical data. It is also unclear whether there would be any data breach notice responsibility on either **the individual's employer or a health plan selected by the employer**.

Breaches by the Non-Profit or downstream recipients of the claims data could also cause **difficulties for the individuals' employers**. Employees who receive health insurance through their employer will not have received notice of the transfer to the Non-Profit and downstream recipients, nor provided any consent to those transfers. Individuals may thus be upset about their confidential health information becoming exposed through a claims database of which they were likely unaware. As a legal matter, it is worth exploring whether the employer has any fiduciary duty or faces any legal liability from breaches that occur due to data supplied by the employer-provided insurance plan. More generally, breaches by the Non-Profit or recipients of the data from the Non-Profit raise the possibility of backlash against the claims database once it is created, which could undermine the goals of the database.

Other risks that arise as the issuers are required to send claims information to the Non-Profit.

Overall, the bill is less detailed on the relationship between those supplying the data and the Non-Profit than it is on the relationships between the Non-Profit and those receiving the data. For instance, recipients such as employers and researchers are expected to apply for access to the data, and to sign data use agreements. By contrast, there is no mention of data use agreements – contracts – between suppliers of data and the Non-Profit.

This lack of detail creates privacy and security risks for the issuers, and thus for patient data:

1. Data use agreements are standard practice in a wide range of settings where sensitive personal data is shared between organizations. For instance, under HIPAA, there must

be a written agreement of this sort between covered entities and business associates. Data use agreements are also used for medical research under HIPAA. The lack of data use agreements under the bill could put issuers at risk – they will not receive in writing assurances from the Non-Profit about how the latter will handle data sent to it.

2. The security and privacy expertise of issuers is not well reflected in the bill. The Advisory Committee to implement the database specifies that two members represent group health plan sponsors (one member representing an employer that sponsors a plan and one member representing an employee organization that sponsors a plan). . By contrast, there is no prescribed membership for an issuer administrator, yet the administrators are those with the technical expertise in how data would be transferred from the plan to the Non-Profit. Issuers are likely to have relevant knowledge about issues such as privacy, security, data quality, and proprietary financial data.
3. There are large technical challenges in how to transfer claims data from all of the issuers to the Non-Profit. Creation of the new database will involve very challenging architectural challenges, including the development of new application programming interfaces (APIs). The data should be encrypted, and there will be technical challenges to ensure inter-operability between the encryption used by the issuers and the ability of the Non-Profit to securely receive and decrypt the claims data. In practice, there will likely need to be extensive discussions and interaction between the issuers and the Non-Profit about how to create inter-operability in a secure manner. The bill provides no detail about how these interactions shall take place. Instead, the bill states that the issuers "shall" submit "all claims with respect to the plan." The bill does not supply the possibility for the issuer to decline to send information when the issuer learns about security vulnerabilities.
4. Issuers may be under legal obligations not to disclose the data to the Non-Profit. Issuers, similar to other entities that hold sensitive data, are often under contractual or other legal obligations not to disclose data to third parties. The bill does not address this possibility. This could put issuers into a double-bind: if the issuer transfers the data to the Non-Profit, then it may face liability under an existing contract or other law; if the issuer does not transfer to the Non-Profit, then it would be violating the bill's requirements.
    a. The bill states: "An entity required to submit data under this subsection may not place any restrictions on the use of such data by authorized users." Section 2796(d)(5). This bill language, however, may contradict legal obligations that already apply to the issuer.
5. The lack of safeguards applies to proprietary financial data as well. In many data use agreements, the recipient of the data agrees to follow safeguards, agrees to auditing by the entity sending the data, and prescribes consequences in the event of violation of the agreement. The bill appears to lack these protections for data supplied by issuers, both for patient/claims data and for proprietary financial data. Failure by the Non-Profit to protect proprietary financial data may harm issuers, such as by eroding their ability to negotiate effectively with providers.

Data Flow Part 2: Processing Data Within the Non-Profit

Once data is supplied by the issuers, the second stage for the data is processing within the Non-Profit. The biggest issue I see concerns the scope of privacy, cybersecurity, and breach protections that are authorized by the bill.

Relevant provisions in S. 1895: The bill instructs the Secretary of HHS to enter into an agreement with a Non-Profit within a year of the bill's passage, to manage the claims database. Section 2796(a). The bill sets forth requirements for the database, including discussion of de-identification requirements that are discussed further below in a separate section. The bill provides the Secretary the power "to issue regulations prescribing the extent to which, and the manner in which, the following rules (and any successors of such rules) shall apply to the activities of this section *of an entity* receiving a contract." Section 2796(b)(2)(A)(i) (emphasis added). The rules are the HIPAA Privacy and Security Rules, as well as the Breach Notification provisions passed as part of HITECH in 2009. In addition, the Secretary "may issue such supplemental regulations on the subjects of the rules listed under clause (i) as the Secretary determines appropriate to address differences between the activities described by this section and the activities covered by such rules." Section 2796(b)(2)(A)(ii).

Analysis: The provisions here provide the Secretary of HHS with significant power to apply the HIPAA Privacy and Security rules, and breach notification requirements, to the Non-Profit. The discussion here addresses three issues:

1. The Secretary's rulemaking power is limited in the first instance to the Non-Profit itself. The bill focuses on the Secretary's ability to extend privacy, cybersecurity, and breach rules to the Non-Profit itself. Those rules shall apply, according to the bill, "to the activities of this section of *an entity* receiving a contract." (emphasis added). That language would appear to exclude applying the rules to others who have access to the claims data. As discussed below, the rules under this authority would not apply to business associates of the Non-Profit, including for instance an IT company that provides back office and data processing services for the Non-Profit. Similarly, the privacy and other rules under this authority would not apply to employers or others who gain access to the database.
2. The scope of the supplementary authority is not clear as drafted. Section 2796(b)(2)(A)(ii) states that the Secretary may issue such supplemental regulations on the subjects of the rules listed under clause (i) as the Secretary determines appropriate to address differences between the activities described by this section and the activities covered by such rules." The scope of such supplemental regulations is not clear. Notably, would the Secretary have the authority to expand the universe of covered entities under the HIPAA Privacy and Security Rules, and the breach rules? One central feature of the HIPAA statute is that it differentiates between "covered entities," who must follow the rules' requirements, and other entities. Any expansion of the obligations to entirely new actors should occur only with clear notice in the law. Consider the employers who would gain access to the database in order to seek to reduce health care costs. Until now, employers have not been covered by the HIPAA Privacy and Security Rules; the universe of covered entities could expand dramatically if an employer became a covered entity in order to access the database. Such coverage might be justified in order to protect sensitive health claims data. My point is that the

scope of supplementary authority should be defined carefully to give notice as to what sorts of supplementary authority may exist.

3. The bill as drafted calls for careful attention to the issue of de-identification of data. I discuss separately below some challenging issues facing the Non-Profit concerning de-identification of patient data under the bill. As drafted, the bill requires both extensive de-identification, and the ability to link the data to an individually longitudinally, over time.

Data Flow Part 3: The Non-Profit Exchanges Data with Business Associates

The term "business associate" means an entity that performs work, on behalf of a covered entity, which involves access to protected health information. The term is defined in detail in 45 CFR § 160.103. As with covered entities under HIPAA, the Non-Profit may contract with business associates for a very wide range of functions, which in practice may include the back-office and IT infrastructure for the Non-Profit. Employees of a business associate may thus have access to a large fraction of the data available to the Non-Profit itself. As drafted, however, the bill appears to exclude business associates of the Non-Profit from the HIPAA Privacy and Security rule requirements.

Relevant provisions in S. 1895: The bill provides the Secretary the power "to issue regulations prescribing the extent to which, and the manner in which, the following rules (and any successors of such rules) shall apply to the activities of this section *of an entity* receiving a contract." Section 2796(b)(2)(A)(i) (emphasis added). In addition, the Secretary "may issue such supplemental regulations on the subjects of the rules listed under clause (i) as the Secretary determines appropriate to address differences between the activities described by this section and the activities covered by such rules." Section 2796(b)(2)(A)(ii).

Analysis:

1. The Secretary's rulemaking power on its face does not cover business associates. The bill focuses on the Secretary's ability to extend privacy, cybersecurity, and breach rules to the Non-Profit itself. Those rules shall apply, according to the bill, "to the activities of this section of *an entity* receiving a contract." (emphasis added). That language would not apply to business associates of the Non-Profit.
2. The scope of the supplementary authority does not appear to apply to business associates. Section 2796(b)(2)(A)(ii) states that the Secretary may issue supplemental regulations, "to address differences between the activities described by this section and the activities covered by such rules." Privacy and security protection by business associates, however, would appear to be fundamentally similar under HIPAA and for protection of the new database. Therefore, this statutory authority would appear to be a weak basis for extending HIPAA requirements and enforcement to business associates.
3. As originally promulgated, HIPAA similarly did not apply directly to business associates. When I worked on the proposed and final HIPAA Privacy Rule in 1999 and 2000, we discussed internally whether there was any statutory basis to regulate business associates directly, and concluded there was not. Enforcement at that point was only

indirect – the covered entity could be the subject of an enforcement action if it improperly supervised a business associate.

4. Congress addressed this problem in the HITECH Act of 2009, requiring privacy and security from business associates. The revised law is at 42 U.S.C. § 17934, applying HIPAA requirements and enforcement to the business associates of covered entities. Based on my experience, business associates took privacy and security requirements far more seriously after passage of the 2009 law.
5. Specifically, HIPAA protections would not apply to a business associate that performs de-identification of records on behalf of the Non-Profit. As discussed in the separate section on de-identification, technical and privacy advantages may result if de-identification is performed by a separate entity acting on behalf of the Non-Profit, rather than within the Non-Profit itself. Under the bill, however, de-identification by such a business associate would take place outside of HIPAA enforcement.

Data Flow Part 4: Employers, Researchers, and Others Who Receive Data from the Non-Profit

Part 4 of the Data Flow diagram occurs when data flows from the Non-Profit to "authorized users," who apply for access to the database. There are special provisions in the bill that apply to researchers, some of which are discussed below. The focus of the discussion here, though, is on employers and other non-researcher authorized users. It appears that this potentially broad group of users would not be covered by the HIPAA Privacy and Security rules, or the HHS data breach rule.

Relevant provisions in S. 1895: Those who can receive data from the Non-Profit are called "authorized users," defined as "including employers, employee organizations, providers, researchers, and policymakers," Section 2796(b)(1)(D)(ii). These authorized users are subject to application requirements in Section 2796(e)(2). For employers and other seeking access to the database "for the purpose of quality improvement or cost-containment," the applicant must provide the Non-Profit with "a description of the intended uses of such data." Section 2796(e)(2)(B)(ii). The Secretary shall through rulemaking "establish the form and manner and issue a regulation in which authorized users" may access data. Section 2796(e)(C)(ii).There is a limit on identifiability: "Data provided to such authorized users shall be provided in a form and manner such that users may not obtain individually identifiable price information *with respect to direct competitors*." *Id.* (emphasis added). Upon approval, such authorized user shall enter into a data use and confidentiality agreement with the entity. *Id.*

Analysis:

1. The range of authorized users is potentially quite large. The bill includes the many employers in the U.S. as potential authorized users. In addition, the bill language says authorized users "include" employers and the other listed categories, suggesting that other categories of authorized users may qualify as well.
2. As with business associates, it appears that employers and other authorized users would not be subject to the HIPAA Privacy and Security Rules, and HHS breach notice requirements. The bill states that these requirements apply only to the entity receiving the contract – the Non-Profit. The scope of the supplementary authority is vague. The

Secretary might nonetheless seek to use the supplementary authority to issue binding regulations on all employers and other users who access the database. If the Secretary were to do so, then the lack of clarity in the bill language could, in my opinion, lead to litigation by those challenging the Secretary's authority to so regulate.

3. There is doubt whether the data use and confidentiality agreements would empower the Secretary to bring HIPAA enforcement actions against employers and other authorized users. Assuming the analysis here is correct -- that there is no HHS rulemaking authority over employers and other authorized users -- then the Secretary might seek to bring enforcement actions based on the data use and confidentiality agreements required by the bill. As a matter of administrative procedure, however, there is a good chance that a contract of this sort would be insufficient to allow the Secretary to apply the enforcement procedures and penalties that apply to HIPAA covered entities. Similar contracts existed for business associates prior to the 2009 HITECH Act, and my understanding is that the contracts did not form the basis for a direct enforcement action by HHS against the business associate that signed such a contract.
4. The hybrid entity approach under HIPAA offers one possibility for applying privacy and security protections to authorized users. HIPAA offers one approach for covering the office within an employer that accesses the database, while avoiding the burden of HIPAA compliance for the rest of the employer's operations. In a "hybrid entity" under HIPAA, the company defines one or more components of its operation that access protected health information. 45 CFR § 164.105(a). Only those components need to comply with HIPAA, and there are restrictions on allowing PHI to flow from that component to the non-covered components of the operations. Presumably most employers (at least if they are not providers or other covered entities) would not want their entire operations to be subject to HIPAA. On the other hand, for specialized health-related personnel, it may be manageable for a defined component of the company to comply with HIPAA protections. In that way, the Secretary would retain enforcement authority, while avoiding burdens on most of an employer's operations.
5. There is a drafting quirk in authorized user access to identifiable claims data. For data provided from the Non-Profit to authorized users, the "users may not obtain individually identifiable price information *with respect to direct competitors*." The bill does not define "direct competitors." In addition, this prohibition on identifiable information "with respect to direct competitors" seems to imply that the authorized *can* receive identifiable price information for other claims, so long as they are not direct competitors. Perhaps there is some reason the bill drafters' included the "with respect to direct competitors" language, perhaps to avoid disclosure of proprietary health care information, but that reason is not apparent. If authorized users are receiving identifiable price information in this way, then the analysis here shows privacy and security risks, because HIPAA privacy and security protections do not appear to apply.

De-Identification and Re-Identification

The draft bill contains multiple provisions on the topic of de-identifying and re-identifying protected health information. The comments here seek to develop further the bill's goal of having effective de-identification of claims data, to protect the privacy and security of

individual patients and their claims information. **The Appendix provides greater technical detail on de-identification issues.**

Relevant provisions in S. 1895: One goal of the bill is to "improve transparency by using de-identified health data." Section 2796(b)(1)(A). The Non-Profit shall "establish a process under which data is de-identified consistent with" the HIPAA de-identification requirements, "while retaining the ability to link data longitudinally for the purposes of cost and quality, and the ability to complete risk adjustment and geographic analysis." Section 2796(d)(1)(C)(i). The bill contemplates that the Non-Profit may hire third-party contractors to perform de-identification, and such contractors shall "retain only the minimum necessary information to perform such a process, and adhere to effective security and encryption practices in data storage and transmission." Section 2796(d)(1)(C)(ii). The Non-Profit shall "store claims and other data collected under this subsection only" in compliance with HIPAA's de-identification requirements. Section 2796(d)(1)(C)(iii). The Secretary shall "take appropriate action to sanction [authorized] users who re-identify data." Section 2796(b)(5). The Advisory Committee shall provide advice on "best practices with respect to de-identification of data, as appropriate." Section 2796(b)(3)(C)(ii)(II). In addition, researchers who gain access to the database must sign a confidentiality and data use agreement that prohibits "attempts to re-identify and disclose individually identifiable health information and proprietary information." Section 2796(e)(2)(C)(i).

Analysis:

1. The bill drafters have already paid considerable attention to the important issue of de-identification and re-identification. The comments here seek to develop further the bill's goal of having effective de-identification of claims data, to protect the privacy and security of individual patients and their claims information.
2. The state of the art on re-identification has changed since the HIPAA Privacy Rule was drafted, and re-identification has become possible in a wider range of situations. During the process for drafting the proposed and final HIPAA Privacy Rule in 1999 and 2000, I worked extensively on de-identification provisions, as part of my role as Chief Counselor for Privacy at OMB. The de-identification provisions have stayed essentially the same since that time. There have been many technical advances during that period on the ability to re-identify data, however. These technical changes should be considered in designing the de-identification components of the bill.
3. These technical developments mean that effective protection increasingly relies on administrative safeguards, in addition to technical protections against re-identification. Due to the greater technical ability to re-identify data, it has become more important than previously to have effective administrative mechanisms to reduce the risk of re-identification. Examples of administrative safeguards are where only authorized users can gain access to the database, and there are mechanisms to reduce the risk of authorized users seeking to re-identify the data.[4] In seeking to assure effective de-

[4] Yianni Lagos & Jules Polonetsky, "Public vs. Nonpublic Data: The Benefits of Administrative Controls," 66 Stanf. Online L. Rev. 103 (2013), available at https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/66_StanLRevOnline_103_LagosPolonetsky.pdf.

identification of the data, it thus makes sense to consider both technical and administrative safeguards.

4. In conclusion on de-identification, the topic is specialized , complex, and important to the goals of the bill. The topic thus deserves careful attention.

Comments on Miscellaneous Provisions

The White Paper thus far has addressed the four components of data flows under the bill, and examined issues of de-identification and re-identification of claims data. For the Committee's consideration, here are some miscellaneous other comments that are related to addressing privacy and security concerns:

1. There are risks from the current bill's use of "Interim Final Procedures" rather than the usual notice-and-comment rulemaking process. The bill states that "The Secretary may make such initial set of regulations effective and final immediately upon issuance, on an interim basis, and provide for a period of public comments on such initial set of regulations after the date of publication." Section 2796(b)(2)(C)(ii). The discussion in this White Paper has raised a number of technical and complex issues for privacy and security, including the topic of how to meet the database's other goals while effectively protecting the de-identification of data.
    a. The HIPAA Privacy Rule benefited from numerous substantive public comments. When I was the White House coordinator for the HIPAA Privacy Rule, we received over 52,000 public comments. We learned a great deal from these comments and made numerous changes from the proposed to the final rule.
    b. This new national database for all medical claims data is similarly likely to receive insightful public comments. The bill proposes a database that will receive data from a wide array of health plans and other sources. The database will have new and unique operational requirements. The process to determine authorized users is similarly complex, due to the wide range of employers and others who may have access to the database. To go straight to final without public comment creates the risk of a badly-designed database and accompanying regulation, with potentially large risks of breach, privacy violation, or other security problems. In addition, if the database is initially created without thoughtful public comment, it may be difficult and expensive to retrofit a national system after the fact; for instance, entities that send claims information to the database might have to build their systems twice if public comments result in changes to the interim regulation. There is thus a strong basis to create such a large and complex architecture for sensitive health data only after the normal Administrative Procedure Act process with notice and comment.
    c. The bill's current timeline will make it difficult to incorporate expert input into the new system and database design. The bill as drafted would convene the Advisory Committee within 180 days and issue the interim final rule within twelve months. That provides little time to explore alternative approaches, receive expert input, and then make evidence-based and reasoned decisions on complex technical issues.

2. Privacy and security reviews are not included in the annual report. The bill requires an annual report. Section 2796(g). The bill specifies topics for that report, including "trends in the price, utilization, and total spending on health care services." The bill does not specify that the annual report discuss privacy, security, and related issues, including the quality of compliance with the privacy and security requirements.
3. The confidentiality of information, for litigation purposes, is not clearly defined. The bill states that "Data disclosed to authorized users shall not be subject to discovery or admission as public information, or evidence in judicial or administrative proceedings, without consent of the affected parties." Section 2796(i)(2). So far as I can determine, the bill does not define who counts as an "affected party." Would this include the Non-Profit? The health plans who supplied the claims data? The patient who may be identified from the claims data?
4. The "rule of construction" is not clearly defined. The bill states "Nothing in this section shall be construed to affect or modify enforcement of the privacy, security, or breach notification rules promulgated under" the HIPAA Privacy and Security Rules. Section 2796(k). The intent of this provision would seem to be, at a minimum, that entities already covered by HIPAA are subject to the same enforcement regime after the bill becomes law. On the other hand, the bill seems to extend enforcement at least to the Non-Profit. The discussion in this White Paper has also addressed the possibility that business associates of the Non-Profit, and perhaps others, would be subject to HIPAA enforcement. As currently drafted, the rule of construction could be read to indicate that there is not actually that expanded enforcement of HIPAA to the Non-Profit and others.

Conclusion

As stated in the introduction, the intent of this White Paper is to help inform deliberations about the proposed national medical claims database, as administered by the Non-Profit, and subject to rulemaking and enforcement by the Secretary of HHS. As with any detailed legislative proposal, it is possible that the author has mis-understood details of the legislative language. The overall intent of this White Paper has been to identify possible issues, to help enable any bill that passes to achieve its goals of reducing health care costs while also protecting patients' privacy.

APPENDIX ON DE-IDENTIFICATION AND RE-IDENTIFICATION

The draft bill contains multiple provisions on the topic of de-identifying and re-identifying protected health information. The main text of this paper provides summary discussion on this topic. This Appendix seeks to develop further the bill's goal of having effective de-identification of claims data, to protect the privacy and security of individual patients and their claims information.

Relevant provisions in S. 1895: One goal of the bill is to "improve transparency by using de-identified health data." Section 2796(b)(1)(A). The Non-Profit shall "establish a process under which data is de-identified consistent with" the HIPAA de-identification requirements, "while retaining the ability to link data longitudinally for the purposes of cost and quality, and the ability to complete risk adjustment and geographic analysis." Section 2796(d)(1)(C)(i). The bill contemplates that the Non-Profit may hire third-party contractors to perform de-identification, and such contractors shall "retain only the minimum necessary information to perform such a process, and adhere to effective security and encryption practices in data storage and transmission." Section 2796(d)(1)(C)(ii). The Non-Profit shall "store claims and other data collected under this subsection only" in compliance with HIPAA's de-identification requirements. Section 2796(d)(1)(C)(iii). The Secretary shall "take appropriate action to sanction [authorized] users who re-identify data." Section 2796(b)(5). The Advisory Committee shall provide advice on "best practices with respect to de-identification of data, as appropriate." Section 2796(b)(3)(C)(ii)(II). In addition, researchers who gain access to the database must sign a confidentiality and data use agreement that prohibits "attempts to re-identify and disclose individually identifiable health information and proprietary information." Section 2796(e)(2)(C)(i).

Analysis:

1. The bill drafters have already paid considerable attention to the important issue of de-identification and re-identification. The comments here seek to develop further the bill's goal of having effective de-identification of claims data, to protect the privacy and security of individual patients and their claims information.
2. There is a fundamental tension between two goals of the bill, to de-identify data and to assure "linkability." One goal of the bill is to "improve transparency by using de-identified health data." The bill also seeks to retain "the ability to link data longitudinally for the purposes of cost and quality, and the ability to complete risk adjustment and geographic analysis." It is technically difficult to ensure that records cannot be linked to an individual, and also that the database systematically retains the ability to link each claim to the individual.
    a. Under HIPAA, where de-identification is carried out by an expert, the expert must determine "that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information. CFR §164.514(b)(1)(i) (emphasis added).
    b. The bill's requirement of linkability means that the re-identification risk must remain "very small" even though the database has a pervasive ability to link each claim to an individual. As discussed further below, technical changes have made it easier to re-identify data, so it may be difficult to achieve the HIPAA requirement of "very small" risk of re-identification.
    c. HIPAA contemplates the possibility of linkability. The Privacy Rule states: "A covered entity may assign a code or other means of record identification to allow information de-identified under this section to be re-identified by the covered entity." 45 CFR §164.514(c). The rule provides protections, notably that the code

cannot be derived from the patient information, and the covered entity does not use or disclose the code for any other purpose. *Id.*

d. Under this HIPAA approach, the code that enables linking cannot be disclosed to other parties. For the proposed bill, that would mean not disclosing the code to authorized users, including employers and researchers. For the proposed bill, the drafters should consider whether this ban on disclosure will meet the goals of the bill. If the current HIPAA Rule does not achieve those goals, then the drafters may wish to make explicit that the bill will (or may) be different than the HIPAA approach.

e. There may be security advantages to having the linkability codes retained by a separate entity, likely a business associate of the Non-Profit. To reduce the possibility of breaches of the linkability codes, there may be advantages to having information about linkability held only by a separate entity. That entity would seemingly be a business associate of the Non-Profit, a possibility that the draft bill already contemplates. Designing and implementing the data flows will be an important technical task, both to ensure de-identification and to segregate the ability to link claims back to an individual patient.

3. The state of the art on re-identification has changed since the HIPAA Privacy Rule was drafted, and re-identification has become possible in a wider range of situations. During the process for drafting the proposed and final HIPAA Privacy Rule in 1999 and 2000, I worked extensively on de-identification provisions, as part of my role as Chief Counselor for Privacy at OMB. The de-identification provisions have stayed essentially the same since that time. I have also continued to work professionally on de-identification issues since then.

   a. In summary, two technical developments have made it substantially easier to re-identify data than was true two decades ago. First, the explosion of publicly-available data on the Internet means that previously-obscure facts about a person are often discoverable. Second, the quality of search engines has improved. The company Google was not founded until late 1998, the year before the proposed HIPAA Privacy Rule. Because we all have become so accustomed to instant and accurate search engines, it is difficult to remember how much harder it was previously to find facts about an individual. The combination of effective search, on massively greater amounts of data, means that we can link facts to individuals massively more often than previously.

   b. Two studies, among many, illustrate the change. Professor Latanya Sweeney, a pioneer in the field of re-identification, found that she was able to re-identify a large portion of hospital patients from a publicly-available data set in the state of Washington.[5] Sweeney found that other publicly-available information, notably local newspapers, enabled re-identification of 43% of the supposedly de-identified hospital records. Second, a recent paper in from *Nature Communications* found, even for heavily incomplete databases, that there is a

[5] Latanya Sweeney, "Matching Known Patients to Health Records in Washington State," *Computers & Society* (2013), available at arXiv:1307.1370 [cs.CY].

high likelihood of re-identifying individuals based on a small number of demographic characteristics.[6]

4. These technical developments mean that effective protection increasingly relies on administrative safeguards, in addition to technical protections against re-identification. As stated, the technical ability to re-identify data has become greater over time, due to better search engines and more publicly-available data. In order to protect patient identity, therefore, it becomes far more important than previously to have effective administrative mechanisms to reduce the risk of re-identification. Examples of administrative safeguards are where only authorized users can gain access to the database, and there are mechanisms to reduce the risk of authorized users seeking to re-identify the data.[7]
    a. HIPAA, by contrast, does not rely on administrative safeguards on de-identified data. The general rule under HIPAA is that data, once de-identified, goes outside of the protection of the Privacy Rule.
    b. Relying only on the HIPAA de-identification standard, therefore, is increasingly outdated in light of technical developments that assist re-identification.
5. The bill already includes an important administrative safeguard – a promise not to re-identify data – but the safeguard as drafted applies only to a sub-set of relevant categories.
    a. The Federal Trade Commission has included that promise in its test for de-identification. The FTC has stated: "data is not 'reasonably linkable' to the extent that a company: (1) takes reasonable measures to ensure that the data is de-identified; (2) publicly commits not to try to re-identify the data; and (3) contractually prohibits downstream recipients from trying to re-identify the data." [8]
    b. The bill includes that promise not to re-identify, but only for researchers. Section 2796(e)(2)(C)(i). The same promise is not clearly required for employers and other authorized users. Similarly, as discussed above, it appears that business associates are not covered by HIPAA, nor are they required to make the promise not to re-identify.
    c. The bill does not include the FTC's recommended practice of contractually prohibiting downstream recipients from trying to re-identify the data.
    d. There are advantages to requiring the public promise not to re-identify, rather than simply prohibiting in a statute or regulation the ability to re-identify. In my experience, some have argued that it would violate free speech protections of

[6] Luc Rocher, Julien M. Hendricx, and Yves-Alexandre de Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models," *Nature Communications* (2019), available at https://www.nature.com/articles/s41467-019-10933-3.

[7] Yianni Lagos & Jules Polonetsky, "Public vs. Nonpublic Data: The Benefits of Administrative Controls," 66 Stanf. Online L. Rev. 103 (2013), available at https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/66_StanLRevOnline_103_LagosPolonetsky.pdf.

[8] Federal Trade Commission, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers, (March 2012), at iv, available at https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

the First Amendment for the government to prohibit generally the possibility of re-identifying. The jurisprudence on the topic lacks clear precedent, in my view. Out of an abundance of caution, therefore, there is a stronger assurance of protection if a recipient voluntarily agrees not to re-identify, as a condition for receiving de-identified data.

6. In light of the importance of de-identification to the goals of the bill, there is reason to prescribe significant expertise on the topic in the Advisory Committee. The current make-up of the Advisory Committee, to the extent it prescribes expertise, addresses expertise in general data protection and security issues. In my experience, the topic of de-identification and re-identification is a complex sub-field of its own, and general knowledge of privacy and security does not necessarily include expertise in that sub-field.
7. In conclusion on de-identification, the topic is specialized , complex, and important to the goals of the bill. The topic thus deserves careful attention in drafting of the bill.