



KEEPING THE INTERNET
OPEN • INNOVATIVE • FREE

www.cdt.org

CENTER FOR DEMOCRACY
& TECHNOLOGY

1634 Eye Street, NW
Suite 1100
Washington, DC 20006

October 31, 2011

Donald Berwick
Administrator, Centers for Medicare & Medicaid Services
U.S. Department of Health and Human Services
Room 445-G, Hubert Humphrey Building
200 Independence Avenue, SW
Washington, DC 20201

Re: CDT Comments to CMS–9975–P

Dear Dr. Berwick:

The Center for Democracy and Technology (CDT), through its Health Privacy Project, promotes comprehensive, workable privacy and security policies to protect health data as it is exchanged using information technology. CDT is frequently relied on for sound policy advice regarding the challenges to health privacy and security presented by health information technology (health IT) initiatives. We have testified before Congress four times on the privacy and security issues raised by health IT, and we chair the privacy and security working group of the federal Health IT Policy Committee (called the “Tiger Team”). CDT submits these comments in response to the Department of Health and Human Services’ (HHS) Centers for Medicare and Medicaid Services’ (CMS) proposed rulemaking on *Standards Related to Reinsurance, Risk Corridors and Risk Adjustment*.¹

As required under the Patient Protection and Affordable Care Act of 2010, CMS’ proposed rule would compel every state (or HHS on the state’s behalf) to collect claims data from every payer in the individual and small group market to support a risk adjustment program.² As CMS’ proposed rule acknowledges, this approach raises patient privacy issues.³ Unfortunately, the proposed rule compounds existing privacy and security problems in two interconnected ways

- The proposed rule indicates CMS plans to collect copies of claims data into centralized databases at the state and federal level. The unnecessary duplication and aggregation of sensitive data worsens the risk and severity of data breaches.
- The proposed rule offers few specifics on the privacy protections CMS intends to apply to the collection and retention of claims data. The impression that the government will unnecessarily collect fully identifiable health records weakens

¹ 76 Fed. Reg. 41930.

² Id., at 41940-1.

³ Id., at 41941

public support for health IT and health reform.

The right mix of technical and policy solutions can address both issues while still achieving CMS' programmatic goals and preserving the accountability of plans for accurate data submissions. CDT therefore offers the following suggestions for the proposed rule. Taken together, these three suggestions can provide greater security against data breaches, address public concern over government collection of identifiable health data, and lower administrative costs and burdens on both plans and government agencies. CDT recommends that CMS

- Adopt a form of distributed network architecture, rather than the centralized approach proposed by CMS, as a more secure and privacy protective method of accessing and analyzing claims data. Specifically, HHS should require each plan to provide claims data – in a format determined by HHS – on plans' own secure edge servers; plans must then make their respective edge servers accessible to state or federal agencies for purposes of administering the risk adjustment program.⁴ Note, however, that we are not recommending the distributed approach described in the preamble of the proposed rule. A key feature of our recommended approach, the distributed "edge server" approach, is that HHS or states would have the ability to access and analyze claims-level data, rather than having to rely on a plan's response to queries. Similar distributed systems are already in use in health care and other sectors.
- Explicitly require the claims data that is submitted to state or federal agencies to be de-identified according to HIPAA standards using a methodology that can reveal a patient's identity only when necessary (such as in the event of an audit of the accuracy of a participating plan's data submissions).⁵
- Explicitly require states and participating plans to secure retained claims data "at rest" via cryptography (i.e., encryption or hashing).

These recommendations are described in more detail below.

I. CMS should adopt a distributed "edge server" approach

CMS should modify its proposed rule to require participating states to utilize a distributed edge server approach for the risk adjustment, risk corridor and reinsurance programs set forth in the rule. CDT's proposed distributed edge server approach would provide states and HHS with greater data access than the distributed approach outlined in the proposed rule. Specifically, HHS and states should require plans to upload standardized claims and encounter data (not aggregated or summarized data) into plans' own secure servers – to which state or federal agencies are granted access. State or federal agencies (not the plans) could then access each plan's server and run the necessary

⁴ At minimum, CMS should give states the option to adopt this distributed approach.

⁵ 45 CFR 145.514(b).

analyses on plans' data to calculate risk adjustment while leaving the physical possession of the claims data with the plans. Auditing and accountability controls should be incorporated to ensure accurate risk adjustment. Such an approach would allow HHS and the states to have access to the data they need to accomplish accurate risk adjustment without the privacy risks HHS acknowledges are present with the government centrally collects individual level data.

As written, the proposed rule would exacerbate a trend underway among states and other federal agencies: the large-scale collection and centralized retention of digital copies of individually identifiable health care claims data. Numerous states, as the proposed rule notes, have established "all-payer claims databases" (APCDs) to compile longitudinal digital claims data for broad public policy, law enforcement and research goals, including comparative effectiveness research.⁶ The federal Office of Personnel Management (OPM), as part of its management of the Federal Employee Health Benefits Program, is also in the process of building its "Health Claims Data Warehouse" for very similar purposes.⁷ Recently, the HHS Office of the Secretary announced plans to establish a "Multi-Payer Claims Database" (MPCD) that will access longitudinal claims data (and eventually information from electronic medical records) for comparative effectiveness research.⁸ Just as CMS appears to propose in its rule, OPM and state APCDs collect individually identifiable health data from health plans and aggregate that data into a centralized database in order to perform their analyses.⁹ Although we are not arguing against the goals and purposes for which these programs are collecting claims data, we believe these goals can be effectively accomplished without submitting raw copies of health claims data to government agencies.

Continually building huge repositories of medical data for new research or policy needs is risky, inefficient and a poor long-term strategy:

- **Data breaches:** As CMS is undoubtedly aware, maintaining copies of sensitive information in various locations for long periods of time sharply worsens the risk and severity of data breaches. Breaches of identifiable medical data are a growing – and extremely costly – problem for patients, health care companies

⁶ Id., at 41931, 41937.

⁷ See 76 Fed. Reg. 35050. CDT wrote a letter to OPM following the initial announcement of the Warehouse, urging OPM to consider alternatives to centralization and provide greater detail on the system's privacy protections. Center for Democracy & Technology, Letter to OPM Regarding the Health Claims Data Warehouse, October 27, 2010, http://cdt.org/files/pdfs/CDT_Letter_to_OPM_Re_Health_Claims_Data_Warehouse-102710.pdf.

⁸ 76 Fed. Reg. 59131.

⁹ Though it is early in the implementation process, the proposed MPCD architecture would centralize health information of lower sensitivity into a database, but leave sensitive and identifiable information with the health plans and made accessible via a distributed query system. See U.S. Dept. of Health and Human Services, Multi-Payer Claims Database (MPCD) for Comparative Effectiveness Research, Jun. 16, 2011, <http://www.ncvhs.hhs.gov/110616p1.pdf>.

and government agencies.¹⁰ Even if the data is de-identified here is still some risk – albeit much lower – of breach and misuse.¹¹

- **Public trust:** Unnecessarily funneling copies of patients' identifiable data to state and federal agencies inflames public perception of government snooping, eroding both trust in the confidentiality of medical records and support for health care reform.¹² As HHS has stated many times, public trust in the privacy of digital health records is fundamental to the evolution to a modern, information-driven health care system.¹³ Good health care depends on good information, but studies regularly show that patients who do not trust the confidentiality of their data are much less likely to be open with their care providers – sometimes going to great lengths to preserve their privacy.¹⁴
- **Inefficient, costly and burdensome:** Diverse entities at the state and federal level want access to health claims, sometimes for very similar purposes.¹⁵ It is burdensome and costly for plans to set up and secure multiple data feeds to different entities in various locations. This situation is especially inefficient when the entities are performing substantially similar analyses that could be fulfilled if

¹⁰ See, Ponemon Institute and ID Experts, *Benchmark Study on Patient Privacy and Data Security*, November 9, 2010, <http://www2.idexperts.com/press/healthcare-news/new-ponemon-institute-study-finds-data-breaches-cost-hospitals-6-billion>.

¹¹ See Center for Democracy & Technology, *Encouraging the Use of, and Rethinking Protections for De-Identified (and "Anonymized") Health Data*, June 2009, Pgs. 7-8, http://cdt.org/healthprivacy/20090625_deidentify.pdf.

¹² See Rep. Tim Huelskamp, *Obamacare HHS rule would give government everybody's health records*, September 23, 2011, <http://washingtonexaminer.com/opinion/op-eds/2011/09/obamacare-hhs-rule-would-give-government-everybody-s-health-records>. See also Rep. Denny Rehberg, *Chairman Rehberg Investigates Possible Violations of Private Health Care Information Under President Obama's Health Care Plan*, October 13, 2011, <http://pressrehberg.congressnewsletter.net/mail/util.cfm?gpiv=2100078808.1461.269&gen=1>.

¹³ David Blumenthal and Georgina Verdugo, *Building Trust in Health Information Exchange, Statement on Privacy and Security*, U.S. Dept. of Health and Human Services, http://healthit.hhs.gov/portal/server.pt?CommunityID=2994&spaceID=11&parentname=CommunityEditor&control=SetCommunity&parentid=9&in_hi_userid=11673&PageID=0&space=CommunityPage (last updated July 8, 2010).

¹⁴ See Markle Foundation, *Common Framework for Private and Secure Information Exchange, The Architecture for Privacy in a Networked Health Information Environment*, April 2006, Pgs. 3-4, http://www.markle.org/sites/default/files/P1_CFH_Architecture.pdf. In a recent study, more than a quarter of U.S. patients stated they would withhold information from clinicians and avoid treatment in order to preserve the confidentiality of their health data. New London Consulting and FairWarning, *UK: How Privacy Considerations Drive Patient Decisions and Impact Patient Care Outcomes*, Pg. 11, October 6, 2011, <http://www.fairwarningaudit.com/documents/2011-whitepaper-uk-patient-survey.pdf>.

¹⁵ For example, both OPM's Health Claims Data Warehouse and many state APCDs perform cost and quality comparisons across geography and demographics.

the entities had access to the same data set. In addition, it is costly for the government to establish and maintain centralized databases.

- **Scope creep:** When government possesses copies of claims, there is a risk that the government will incrementally expand its uses of the data beyond the limited set of purposes described in proposed rules and legislation – processes for which there is opportunity for public participation. It is more difficult for the public to understand, comment upon and vet new uses of data already in government possession.

Instead of outright requiring the creation of yet more centralized databases stocked by data feeds from health plans, CMS should require each plan to set aside a copy of structured claims and encounter data in a secure system, such as an edge server. HHS and states would have access to the data on the edge server to carry out the purposes HHS describes in the proposed rule, such as risk adjustment, enforcement validation of reinsurance and risk corridors. As described above, HHS or states could access and analyze data held in the system themselves, instead of requiring plans to respond to a variety of queries, which may burden some smaller plans, as the proposed rule notes.¹⁶ HHS and states could then retain the results of their analyses, rather than full copies of the claims data. Plans that exit the risk adjustment program may be required to maintain the secure edge server for a time period sufficient to enable CMS to complete any data analysis necessary to meet program needs. Similar distributed systems are already broadly deployed in the public and private sectors, support complex analyses of massive quantities of data from multiple external data sources, and have a high level of accountability.¹⁷

Leaving physical possession of the claims data with the plans can reduce the risk and severity of data breach, leverage existing infrastructure, minimize data transfer and cut down on redundant work. A distributed system would be less costly for state and federal government agencies to build than a centralized database. Uploading structured data to edge servers maintained by each plan would also likely be less costly and time-consuming for each plan than compiling and submitting regular reports to government agencies.

The distributed edge server system proposed above should incorporate policies and technical mechanisms that hold health plans accountable for the reliability of claims data held in the edge server. CMS should require the distributed edge server system to include immutable audit trails recording actions the plans perform on the claims data. CMS should require plans to certify the accuracy of the data they submit and subject plans to penalties for chronic or willful failure to meet accuracy standards. CMS should periodically audit the data submitted to the plans' edge servers for accuracy, comparing the plans' submissions to normative data and matching the submissions to the plans' internal records. The proposed rule cites concern that, under a distributed model

¹⁶ 76 Fed. Reg. 41940.

¹⁷ See Comments of Palantir Technologies to CMS proposed rulemaking on *Standards Related to Reinsurance, Risk Corridors and Risk Adjustment*, CMS–9975–P, 76 Fed. Reg. 41930.

whereby plans respond to queries, plans would make errors in calculating individual risk scores and plan averages; however, our proposed edge server model mitigates that concern because HHS itself would access the plans' data and perform the calculations.¹⁸

II. CMS should explicitly require data collected for risk adjustment to be de-identified

The proposed rule has drawn negative media attention in large part because of the perception that CMS intends to require government collection of confidential – i.e., not de-identified – medical records to carry out the risk adjustment program.¹⁹ Steve Larsen, HHS' Director of the Center for Consumer Information and Insurance Oversight, issued a blog post stating that these concerns were misplaced and CMS is not proposing or requiring states to collect patients' personal data, such as names or addresses, for the risk adjustment program.²⁰ However, Mr. Larsen's statements appear to conflict with the content of the proposed rule, as explained in more detail below. In its final rule, CMS should clarify this issue and explicitly require states and HHS to collect data that has been de-identified using hashing techniques that still allow for the creation of longitudinal records.

The proposed rule does not make clear that CMS will not collect (or require states to collect) individually identifiable data – in fact, the proposed section 153.340 suggests the opposite.²¹ The proposed section does not outright prohibit the collection of identifiable information. To the contrary, the proposed "minimum standards" indicate that CMS expects individually identifiable information to be collected – the minimum standards describe modest safeguards for the use, retention and disclosure of "individually identifiable information." The proposed section also includes an exception from the minimum standards for state APCDs, and the exception likewise does not restrict APCDs from collecting identifiable data. Finally, the proposed section would require states to use HIPAA standards ASC X12N 837 and ASC X12N 834 for risk adjustment data collection, yet health care providers applying these standards do appear to require collection of such individual identifiers as name and address.²²

¹⁸ 76 Fed. Reg. 41940.

¹⁹ See *supra*, fn. 11.

²⁰ Steve Larsen, *Risk Adjustment and Health Insurance*, U.S. Dept. of Health and Human Services, October 13, 2011, <http://www.healthcare.gov/blog/2011/10/riskadjust10132011.html>.

²¹ 76 Fed. Reg. 41954.

²² See, e.g., the following implementation guides requiring the collection of patient names and residential addresses. Blue Cross Blue Shield of South Carolina, ASC X12N 834 Benefit Enrollment and Maintenance Supplemental Implementation Guide, Pg. 7, August 15, 2011, <http://www.hipaacriticalcenter.com/UserFiles/hipaacritical/Documents/5010-834-rev5.pdf>. Blue Cross Blue Shield of South Carolina, ASC X12N 837I Benefit Enrollment and Maintenance Supplemental Implementation Guide, Pg. 5, August 15, 2011, <http://www.hipaacriticalcenter.com/UserFiles/hipaacritical/Documents/5010-837I-rev9.pdf>.

In the final rule, CMS should make clear that claims and encounter data retained for the risk adjustment program should be stripped of patient identifiers according to the HIPAA de-identification standard.²³ Under the distributed approach we recommend above, plans should store claims data in the edge server in de-identified form. To the extent that state or federal agencies need longitudinal records of individual patients, plans could use a one-way hash algorithm to mask patient identifiers while allowing agencies to track records belonging to the same patient.²⁴ At the outset of the program, plans could provide the government with the hash keys in the rare event that the government must know the identity of an individual patient, such as for audit purposes.

III. CMS should explicitly require retained claims data to be secured via encryption or hashing

CMS' proposed minimum standards require states to implement security standards that provide technical safeguards for identifiable information consistent with the security standards described at 45 CFR 164.312.²⁵ The HIPAA Security Rule requires covered entities and business associates to use security measures that are "reasonable" and "appropriate" to protect health data, including encryption. However, covered entities can determine that encryption is not appropriate to protect their health information – in which case they must document their decision and use an "appropriate" alternative protection to meet the Security Rule standards. Thus, while the Security Rule requirements of reasonable and appropriate safeguards are not optional, covered entities have some leeway with regard to the use of encryption. We believe such leeway is inappropriate for the risk adjustment program due to the fact that program will involve large scale databases of sensitive claims records.

CMS should modify the minimum standards in its proposed 45 USC 153.340(b) to explicitly require states to encrypt or hash claims and encounter data retained in connection with the risk adjustment program. Likewise, HHS should require plans participating in the program using the distributed model described above to secure the edge server and encrypt or hash the data held "at rest" on the server. HHS should also make clear that the cost of the encryption or hashing system can be funded through the Exchange Planning and Establishment Grants available under Section 1321 of the Patient Protection and Affordable Care Act.²⁶

²³ 45 CFR 145.514(b)

²⁴ A hash function takes a set of data and condenses it into a "representation" comprised of alphanumeric characters. Unlike encryption, hashing two identical sets of data will produce identical representations. This technique can therefore be used to match identical pieces of data without actually viewing the underlying data, hence supporting both de-identification and longitudinal records. See <http://csrc.nist.gov/groups/ST/hash/index.html>.

²⁵ 76 Fed. Reg. 41954.

²⁶ Pub.L. 111-148.

We thank CMS for the opportunity to submit comments. Please do not hesitate to contact us if we can be of any assistance.

A handwritten signature in black ink, enclosed in a thin black rectangular border. The signature reads "Deven McGraw" in a cursive, slightly slanted script.

Deven McGraw
Director, Health Privacy Project
Center for Democracy & Technology

A handwritten signature in black ink. The signature is highly stylized and cursive, appearing to read "Harley Geiger".

Harley Geiger
Policy Counsel
Center for Democracy & Technology