Precision Medicine and Sharing Medical Data in Real Time: Opportunities and Barriers

Y. Tony Yang, ScD, and Brian Chen, PhD, JD

ABSTRACT

Sharing massive amounts of medical data is critical to precision medicine. The California Department of Public Health recently started to partner with certain hospitals in the state to better understand cancer trends by collecting and securely sending standardized cancer data directly to the California Cancer Registry. This initiative is the first of its kind in the United States. This has afforded the cancer registry the opportunity to perform real-time surveillance on data reported via participating hospitals, and researchers can use advanced methods to analyze these data. Other states are likely to follow California's lead. However, there are barriers to increased data-sharing efforts. How these barriers can be addressed to facilitate data sharing while protecting individual privacy, reducing the risk of data misuse, and enhancing public trust becomes critical as precision medicine moves forward.

Am J Manag Care. 2018;24(8):356-358

The growing use of big data and the ability to share data across entities and fields have been called the "biggest medical innovation in recent years" and hold great promise for the future of population health and precision medicine.[1] Big data are characterized by "volume, variety, and velocity," or the ability to store massive amounts of data, to analyze differently formatted data structures, and to do so quickly and efficiently.[1] Big data can be used to deliver personalized medicine, help identify population health patterns in communities, and predict and track disease.[2] California recently introduced a real-time cancer tracker, which allows pathologists and oncologists to upload current patient data into a central database.[3] Researchers and providers will be able to use the database to recommend treatments for patients based on real-time information and outcomes.[3] This model is promising, but there are barriers that may impede the system's adoption elsewhere.

The California Model

In 2014, the California Department of Public Health (CDPH) and the College of American Pathologists (CAP) began a pilot project to track cancer data in real time in a central database that will be accessible to providers across the state.[3] One of the reasons California's model is so novel is because of how the information is uploaded. The CDPH and CAP worked together to develop a standardized checklist into which physicians and providers enter patient data.[3] Currently, pathologists write brief paragraphs about their observations and outcomes and send them to the California Cancer Registry. A researcher then reads the paragraphs and enters the information into the system.[3]

A real-time registry offers many benefits. If enough hospitals adopt the same format, physicians will be able to use all of the standardized data to help determine treatment plans for their patients, including recommending clinical trials. The new format will also allow cancer centers to better conduct internal evaluations to detect problems with their laboratories or pathology departments. Currently, only a few of the hundreds of hospitals in California are utilizing the new format.[3] As of today, there remain many barriers that impede speedy adoption.

TAKEAWAY POINTS

- The California Department of Public Health recently started to partner with hospitals to better understand cancer trends by collecting and sending standardized cancer data directly to the California Cancer Registry.
- This has afforded the registry the opportunity to perform real-time surveillance on data reported via participating hospitals, and researchers can use advanced methods to analyze these massive amounts of data.
- There are barriers to increased data-sharing efforts. How these barriers can be addressed to facilitate data sharing while protecting individual privacy, reducing the risk of data misuse, and enhancing public trust becomes critical as precision medicine moves forward.

Barriers

Technical. Most hospitals use an electronic health record (EHR) system, which, in theory, should give researchers immediate access to big data, but barriers exist. First, these systems vary and may not be compatible with other systems.[2] Second, clinical trial and other research data are not standardized. Researchers work around these limitations by extracting, transforming, and loading data from EHRs into a "data lake," which is a massive, easily accessible, centralized repository of large volumes of structured and unstructured data.[1] Researchers can then create 1 large data schema, but this can be very expensive in time and resources, requiring intensive and customized data coding.[4]

Also, some consumers may feel uncomfortable with their data being stored and accessed by multiple users, even when deidentified.[1] There are also questions about whether anonymization truly deidentifies data. One study found, for example, that 40% of the participants in a deidentified DNA study were reidentifiable.[1]

Legal. Data sharing is regulated by federal and state laws. The content of the data, their identifiability, and the context of use determine how these laws apply.[5] The Public Health Service Act limits the use of identifiable data and requires the secretary of HHS to ensure that data are protected from inappropriate use.[6] The Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulates what types of identifiable data can be disclosed by healthcare plans and providers.[7] The HIPAA Privacy Rule, which applies to identifiable health information only, protects such data from being improperly disclosed by health plans, providers, clearinghouses, and business associates.[8] The HIPAA Security Rule requires covered entities to establish structures to protect data from being breached or inappropriately disclosed.[8] In addition, tax considerations restrict nonprofit organizations from entering into certain types of data-sharing arrangements.[5] Many governmental agencies also have their own privacy and data-sharing policies.[9]

The ability to collect public health data, as well as to disclose and share such data, varies by state.[10] Some states have privacy protections that are stronger than those required under HIPAA[10] or specific protections for certain diseases, such as HIV.[5] Other state laws are comprehensive and apply generally to system participants who collect, acquire, use, or disclose information within the state.[5] Federal and state antitrust laws may also regulate contractual information exchange and sharing arrangements to promote competition.[5] Providers may have to consider state licensing requirements.[5] Some states create legal obligations and requirements on government entities regarding government data, and others utilize broader privacy laws.[10] Mandatory data-reporting laws also vary across states.[10] Such variations in state laws may present barriers to information sharing.

Institutionalized. Institutionalized barriers must also be overcome. Of particular significance are patients' privacy concerns and reluctance to share data.[1] Many consumers are concerned that their data may be used by third parties, even though they frequently and voluntarily upload identifiable data on unsecure websites and social media.[2] Consumers also fear that their data will be used in unethical or harmful ways,[2] such as to discriminate against minorities.[5] Patients want data to be shared if it benefits them, but they want to protect their data and keep them to themselves if they do not have a vested interest in sharing the data.[5] In order to bring consumers on board, the government, health systems, and research institutions will have to prove that they can be trusted with sensitive health information and explain the exact purpose of data collection.[5]

In addition, health systems are worried that their competitors will be able to use their data against them when competing for customers.[1] Providers fear that if their health statistics are publicly available, they will lose patients or be sanctioned if they do not measure up in performance.[2] The cost of implementation of data management and sharing software is also a barrier. Companies are reluctant to implement large data systems requiring high front-end costs if they do not see a clear financial benefit.[4]

A Path Forward

Overcoming technical barriers. The government has access to an enormous amount of healthcare data. In 2013, the government accounted for 64% of total healthcare spending in the United States, and this figure is expected to rise to over 67% by 2024.[11] This presents a major opportunity for the federal government to lead the way toward innovative data collaboration. CMS already shares Medicare claims data with "qualified entities" in order to evaluate provider performance.[12] Value-based purchasing and performance-based payment also encourage the study of claims data to detect patterns and reward providers of high-quality care.

Possible legal solutions. Legislative or regulatory fixes may help assure the public that their health data will not be compromised. Policy makers should swiftly punish the organizations and people responsible for data breaches to foster an environment of accountability.[2] Safeguards should ensure that healthcare data can be used only by entities with sufficient technical capabilities to maintain security.[2] The Affordable Care Act has also enacted standards for the collection of certain kinds of especially sensitive health data, such as race and ethnicity.[10]

The government can clarify restrictions around data sharing without requiring legislation. Agencies can issue guidance and clarify language around existing statutes and regulations, including issuing guidance that coordinates agencies in order to clarify the ways in which all of the potential problem areas interact.[5]

States can also remove barriers to data sharing by providing incentives to share data by working to connect public health agencies to providers, mandating the reporting of data and minimizing barriers that could limit reporting or data sharing, and sharing best practices so that other states can apply lessons learned to their own systems.[10]

Changing institutionalized barriers. One way to address privacy concerns is to create data stewardship guidelines.[1] At the federal level, the Federal Trade Commission and the Organisation for Economic Co-operation and Development have created guidelines for researchers to use data in a fair and secure way.[1] The Markle Foundation has created the Connecting for Health Common Framework for Private and Secure Health Information Exchange (Common Framework) that institutions can use as technical guidance when creating sharable data systems.[13] The Common Framework is based on US Fair Information Practice Principles, which stress transparency, individual participation, purpose specification, use limitation, data security, and institutional accountability.[14]

Policy makers should engage the patient community and explain the benefits of data collection in a concrete and tangible way.[2] To such end, Jan Liphardt, PhD, of Stanford University, has proposed a patient-driven cancer database.[15] The site will respect patient privacy by anonymizing data and following patient directives on what the data can be used for.[15] The initial phase will only ask patients to answer 5 basic questions, such as "what is your diagnosis?" and "did your cancer metastasize?" Eventually, however, the team hopes to synthesize the data so that patients can help chart their own treatment plans by looking at what the data set shows for similar patients.[15] A model like this represents a path forward for patient engagement and trust, the foundation upon which the future of big data must be built.[15] ■

Author Affiliations: Center for Health Policy and Media Engagement, George Washington University School of Nursing, and Department of Health Policy and Management, George Washington University Milken Institute School of Public Health (YTY), Washington, DC; Department of Health Services Policy and Management, University of South Carolina (BC), Columbia, SC.

Source of Funding: None.

Author Disclosures: The authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (YTY); drafting of the manuscript (YTY); critical revision of the manuscript for important intellectual content (BC); and supervision (YTY).

Address Correspondence to: Y. Tony Yang, ScD, 1919 Pennsylvania Ave NW, Ste 500, Washington, DC 20006. Email: ytyang@gwu.edu.

REFERENCES

1. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. *Health Aff (Millwood)*. 2014;33(7):1115-1122. doi: 10.1377/hlthaff.2014.0147.
2. Heitmueller A, Henderson S, Warburton W, Elmagarmid A, Pentland AS, Darzi A. Developing public policy to advance the use of big data in health care. *Health Aff (Millwood)*. 2014;33(9):1523-1530. doi: 10.1377/hlthaff.2014.0771.
3. California Department of Public Health (CDPH) Partners for breakthrough for sharing cancer data [news release]. Sacramento, CA: California Department of Public Health; July 27, 2015. cdph.ca.gov/Programs/OPA/Pages/NR15-051.aspx. Accessed December 28, 2017.
4. Bernstein AB, Sweeney MH; Centers for Disease Control and Prevention. Public health surveillance data: legal, policy, ethical, regulatory, and practical issues. *MMWR Suppl*. 2012;61(3):30-34.
5. Rosenbaum SJ, Painter MW. Assessing legal implications of using health data to improve health care quality and eliminate health care disparities. George Washington University website. hsrc.himmelfarb.gwu.edu/sphhs_policy_facpubs/253/. Published 2005. Accessed December 28, 2017.
6. Public Health Service Act, 42 USC §300kk(e).
7. Health Insurance Portability and Accountability Act, PL 104-191 (1996).
8. Summary of the HIPAA Privacy Rule. HHS website. hhs.gov/hipaa/for-professionals/privacy/laws-regulations. Accessed December 28, 2017.
9. CDC/ATSDR policy on releasing and sharing data. CDC website. cdc.gov/maso/policy/releasingdata.pdf. Published April 16, 2003. Updated September 7, 2005. Accessed December 28, 2017.
10. Partnership for Public Health Law. Legal issues related to sharing of clinical health data with public health agencies. Association of State and Territorial Health Officials website. astho.org/Public-Policy/Public-Health-Law/Legal-Issues-Related-to-Sharing-Clinical-Health-Data-with-Public-Health-Agencies. Published April 2016. Accessed December 28, 2017.
11. Himmelstein DU, Woolhandler S. The current and projected taxpayer shares of US health costs. *Am J Public Health*. 2016;106(3):449-452. doi: 10.2105/AJPH.2015.302997.
12. Qualified entity program. CMS website. cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/QEMedicareData/index.html?redirect=/QEMedicareData. Updated October 13, 2017. Accessed December 28, 2017.
13. Markle Common Framework. Markle website. markle.org/markle-common-framework-connecting-professionals. Accessed December 28, 2017.
14. National strategy for trusted identities in cyberspace. National Institute of Standards and Technology website. nist.gov/sites/default/files/documents/2016/12/08/nsticstrategy.pdf. Published April 2011. Accessed December 28, 2017.
15. Huber J. Introducing CancerBase: a way to share personal medical data to help cancer research. Scope website. scopeblog.stanford.edu/2016/08/01/introducing-cancerbase-a-way-to-share-personal-medical-data-to-help-cancer-research. Published August 1, 2016. Accessed December 28, 2017.