Clinical Chemistry 63:8
1321–1325 (2017)

Q&A

Patient Privacy and Clinical Laboratory Data

Moderator: Jason Y. Park[1*]
Experts: Toby C. Cornish,[2] Michael Hogarth,[3] Brian R. Jackson,[4] and Kristen B. Rosati[5]

We currently live in an age of massive personal data creation through activities such as social media, online banking, and healthcare. In healthcare, modern concepts of patient privacy were codified in the 1996 Health Insurance Portability and Accountability Act (HIPAA).[6] In 2009, the Health Information Technology for Economic and Clinical Health Act was passed to encourage the adoption of electronic medical records. These laws regulate the privacy of patients' clinical laboratory data.

With the generation and collection of large sets of clinical laboratory data, there is a privacy risk. However, this risk is balanced by the use of clinical laboratory data for biomarker discovery and the measurement of clinical outcomes. Recently proposed initiatives such as the Cancer Moonshot and Precision Medicine Initiative (now called "All of Us") will require the information of millions of patients, including both conventional laboratory and genomic data. In this Q&A, 4 experts explore the current state of data warehouses and health exchanges with a focus on the clinical laboratory.

Can you provide an example of a data warehouses or health information exchanges?

Toby C. Cornish: Health Data Compass, which went live in 2015, is the enterprise health data warehouse for the University of Colorado (UC) School of Medicine and UCHealth. It is a joint venture by the UC School of Medicine, University of Colorado Medicine, UCHealth, and Children's Hospital Colorado. Laboratory data were added to the data warehouse in July 2016 and initially included over 369 million patient laboratory test orders and results from UCHealth and over 124 million patient laboratory test orders and results from Children's Hospital Colorado. Health Data Compass does not yet contain genomic data, but it is expected to begin housing single-nucleotide polymorphism results from the Colorado Center for Personalized Medicine's Biobank project when that data become available. UCHealth participates in the Colorado Regional Health Information Organization, which contains records for over 4 million unique patients, 4000 providers, and 61 hospitals.

What is the clinical utility of data warehouses?

Kristen B. Rosati: A number of clients who have set up data warehouses report that data warehouses are key to being able to effectively curate and utilize data for internal quality improvement and other projects requiring large data sets. Patients benefit from rich data resources for proactive care management, improved quality improvement activities, and rich data resources for research.

Toby C. Cornish: Data warehouses are extremely important for coordinating disparate data sources, including clinical, administrative, billing, genomics, and other systems. Data warehouses integrate data from across the enterprise and are critical for clinical and translation research, cohort identification for clinical trials, healthcare quality improvement, and in reduction of overall healthcare costs. The primary clinical utility is in enabling data-driven processes in clinical research, quality improvement, and operations. Any benefit to individual patients

[1] Clinical Director, Advanced Diagnostics Laboratory, Department of Pathology and the Eugene McDermott Center for Human Growth and Development, Children's Medical Center Dallas and the UT Southwestern Medical Center, Dallas, TX; [2] Medical Director of Informatics, Department of Pathology, University of Colorado School of Medicine, Denver, CO; [3] Department of Pathology and Laboratory Medicine, UC Davis Medical Center, Sacramento, CA; [4] Vice President, Chief Medical Informatics Officer, ARUP Laboratories, Salt Lake City, UT; [5] Coppersmith Brockelman PLC, Phoenix, AZ.

* Address correspondence to this author at: Children's Medical Center Dallas, 1935 Medical District Dr, Dallas, TX 75235. Fax +214-456-4713; e-mail jaspar@childrens.com
Received April 20, 2017; accepted April 21, 2017.

[6] Nonstandard abbreviations: CDW, clinical data warehouse; EHR, electronic healthcare record; HIE, health information exchanges; HIPAA, Health Insurance Portability and Accountability Act.

is indirect. Although some data warehouses may return data to clinical information systems for clinical decision-making, at my institution, the data in Health Data Compass have not been certified for clinical decision-making.

Michael Hogarth: There have been discussions of the potential benefit of data warehouses. However, assuming the term "clinical utility" to be meant as "at the point of care, real-time, clinical utility," I would say the current utility of data warehouses is very limited. There are only a handful of examples of leveraging clinical data warehouses to provide "practice-based evidence" at the point of care. At my institution, patients may benefit from having their data in a clinical data warehouse (CDW) because their data are used to improve overall care delivery. For example, we had a sepsis registry that was used to reduce sepsis mortality by 25% in 24 months. The reasons for current limitations to clinical utility include: limited data quality; lack of essential key data elements being in structured form instead of semistructured narrative text; methods of validating decision support recommendations based on automated querying of a data warehouse; incorrect vital status on most of their deceased patients so the ultimate outcome, mortality, is grossly underestimated.

What is the clinical benefit of health information exchanges (HIE)?

Kristen B. Rosati: HIEs, which I define to mean exchange between legally separate entities, are key to getting the correct information about the correct patient at the correct time for treatment. HIEs have the potential to improve clinical outcomes and reduce the cost of care. The benefit to individual patients for their healthcare providers to participate in HIEs is substantial, especially if those individuals receive care from multiple legal entities. HIEs lift the burden from patients of having to ensure that each of their providers have a complete picture of the care they receive. HIEs are especially helpful in obtaining the information necessary to manage the care of a chronically ill family member.

Toby C. Cornish: The primary role of HIEs is to give regional providers participating in a patient's care access to that patient's electronic records generated by other providers, hospitals, and facilities. This can be especially useful for obtaining data generated by outside laboratories. This results in better and more coherent patient care, especially when patients transition between healthcare settings. HIEs also contain costs by reducing administrative and technical overhead and eliminating unnecessary duplication of laboratory testing.

Michael Hogarth: Patients may benefit from their data in HIEs by proxy when their providers have additional clinical information about them so that the clinical decision-making is optimized. I personally have experience with a HIE providing me with access to records about patients, which was very helpful. As an internist, I had a complex patient with dementia and unexplained abdominal pain who could not recall if he had been seen at another hospital. We proceeded with starting a workup and then used our local private connection with another institution for information about him. An unanticipated finding was that he had been recently diagnosed with biopsy-proven disseminated tuberculosis and had not followed up for treatment. This electronic information exchange allowed us to proceed to treat him without resorting to repeating everything, including a laparoscopic biopsy! So, I have personally experienced the great benefit to patients of being able to connect to other hospitals in the region and query for records. On an individual provider level, we should all be using our HIEs to manage patients. However, I have not yet seen a study that has done a good job of quantifying the benefit.

Brian R. Jackson: The most immediate and obvious use case for HIEs is the availability of data for emergency care. Although appealing, I see this as a pretty limited-use case. A second, larger one is availability of data for non-emergency care, for which the main benefit is reducing redundant testing (especially expensive imaging). But I think in the long run, the most compelling utility is being able to have more complete data sets available for analytic purposes, such as machine-learning approaches to decision support and quality measurement.

How is the quality controlled for data deposited in data warehouses or health information exchanges?

Michael Hogarth: The source of laboratory data in our clinical data warehouse is the electronic health record (EHR). The only "quality control" for this laboratory data is imposed in the laboratory information system (LIS)-to-EHR interface. We routinely inspect a random

10% of result messages that go between the LIS and EHR monthly. However, I am not aware of any quality control on the data warehouse side. We found that about 5% of our laboratory tests from certain clinical practices were manually entered test results without units of measure and no validation. This is a situation that might also exist elsewhere and can be a data quality problem (and care problem if the values are typed incorrectly).

Brian R. Jackson: The quality of data will be highly dependent on the setting. My concern, based on prior experience, is that most healthcare information technology people don't have enough laboratory expertise to adequately quality control the data. An example of this is data mapping, where some background knowledge is necessary to know whether 2 different test codes represent the same test or not.

Are there risks to patient privacy with the current practice of data warehouse and health information exchanges?

Michael Hogarth: Risks to privacy from data warehouses and health information exchanges are similar to risks from data in an EHR. The exchanges between an EHR and an HIE are all done in encrypted fashion, and organizations must follow standard agreements such as a Data Use and Reciprocal Support Agreement. However, there are limited opportunities for verification with compliance to the security requirements. As long as there are data in any database and exchange happening, there will always be a risk of privacy violations. It is not possible to have zero risk; however, it is possible to keep the risk low so that the benefits of sharing the data dramatically exceed the risks. It is also clear that a CDW does not need to have "identified" data for much of its benefit to be realized. If one relegates functions to the CDW that can be done without personal identifiers, I would then say that the risk of reidentification is quite low, assuming the CDW is not provided to an external entity in an uncontrolled fashion. A secure agreement would have restrictions on reidentification and thus be considered a limited data set. There should be a HIPAA business associates agreement that disallows reidentification by the recipient of the data. Ideally, one would enable patients to decide whether their data in a CDW is shared.

Toby C. Cornish: Each additional copy of identifiable patient health information adds risk to patient privacy. This additional risk cannot be eliminated, but every CDW and HIE must take steps that mitigate it. These steps include robust physical, technical, and administrative safeguards to ensure that protected health data remain private and secure. When considering the risks of CDWs and HIEs, it is very important to also consider the privacy risks of not having them. While centralization of data creates a large, high profile target, which magnifies the potential impact of a single breach, it also creates an opportunity to consolidate security and administrative efforts on a single source of data. When CDWs and HIEs are not employed, the same data will be stored in and accessed from a number of independent silos and disseminated in myriad ways. Ensuring that these data silos are all well managed and secured can be challenging, and the possibility of breaches or inappropriate use of data are very real. Likewise, when health data are transferred outside the context of an HIE, many means of communication may be used including printouts, faxes, email, and phone calls. In these cases, it is difficult to ensure that the transport mechanism itself is always secure.

Brian R. Jackson: Privacy is about controlling who can see your data, and this gets much harder in the era of interconnected data warehouses. Some patients don't want all of their healthcare providers to know all of their history, particularly history that is socially stigmatizing (e.g., drug use, sexually transmitted disease). These are legitimate concerns. Healthcare is a deeply personal domain and, though we don't like to admit it, some healthcare providers prejudge patients because of their social history. Some access by providers is direct, but increasingly it may include indirect inference. For example, what if a machine-learning algorithm labels a patient as being at high mortality risk, and the provider happens to know that the algorithm is partly based on obesity, unsafe sex, and illicit drug use? If employers can access healthcare data of applicants or employees, this would raise the stakes even further. I have a lot of concerns about many of the wellness programs in place at some large employers with their associated data acquisition. Many of these programs aren't really effective at promoting health but could be very effective at reducing the employer's healthcare costs by directly or indirectly reducing/avoiding employment of employees with expensive health conditions.

Kristen B. Rosati: Of course, any use of patient identifiable health information poses some risk to patient privacy. However, good data security and data governance processes can minimize those risks to the extent possible. Good HIE practices, for example, limit access to providers or plans that have an existing relationship documented in the HIE (such as through existing records showing the patient has been treated by a physician) or require attestation of that relationship before access. Accurate patient-matching algorithms are also important to ensure that the provider accesses the correct patient records, which is important both to privacy and preventing medical mistakes based on incorrect information. An increasing number of healthcare organizations are creating shared "big data" repositories involving multiple legal

entities. For those shared data resources, it is essential to establish rigorous data governance, rules of access, and downstream data use agreements to minimize potential privacy risks.

Are patients adequately informed of their current participation in data warehouses and health information exchanges?

Kristen B. Rosati: Most HIEs have good patient notification processes in place, which are in response to state laws or community policies. The specifics of patient notification vary greatly from one HIE to another, although most require actual written notice from healthcare providers that are participating in the HIE. On the other hand, most patients are not notified that their health information will be included in a data warehouse. That's because the creation of a data warehouse is a "healthcare operation" under HIPAA, under which the use of patient health information for internal business purposes is permitted without patient authorization or advance notice to the patient. State law also generally does not require patient consent for a healthcare organization's internal use of patient information, so notice to the patient would not carry any "actionable" information. Moreover, patient notice and the chance to object to the internal use of patient information in data warehouses would not be beneficial from a public policy perspective; this use of data is essential to clinical integration activities, rigorous quality improvement, and health services research.

Brian R. Jackson: I would say that patients are not adequately informed, particularly about potential harms. This is a really hard area, though. I am not sure how to create effective informed consents with opt-out possibilities given that healthcare data and the applications of the data are becoming so complex and interconnected. I just don't think that informed consent is a sufficient safeguard. I think we need other types of safeguards as well.

Michael Hogarth: Patients may be informed, but they may not understand where the data reside and how privacy protections are being managed. At my institution, we have a blanket consent for treatment that includes using the data for quality improvement. As you know, HIPAA does not require patient consent to access data about them in a remote system as long as you are accessing it to provide care. I do not believe our HIE participation in exchanging data with other institutions is subject to patient consent (i.e., it is opt-in by default).

Toby C. Cornish: Although the exact mechanism for informing patients varies, institutions are taking this issue very seriously. In my health system, all patients are notified that their data may be used for treatment, payment, healthcare operations, and research purposes. There is a waiver of informed consent and waiver of authorization from the University's Institutional Review Board (IRB) to make these data available for research projects that have IRB approval. Patients are also informed that their results will be sent to the HIE in the Notice of Privacy Practices that they are given when establishing care with a provider or facility. A patient can elect to opt-out of participating in the HIE at this time or at any time thereafter. If a patient opts out, their results will continue to flow from upstream information systems, but the data will not actually be searchable via the HIE. This mechanism allows a complete patient record to be maintained if the patient should choose to opt-in at a later date.

Can patients determine if their health information has been deposited in a data warehouse or is part of a health information exchange?

Kristen B. Rosati: Many HIEs have policies that allow patients to determine whether their information is held by the HIE. Some HIEs do this directly; others provide this information through a healthcare provider participating in the HIE. In contrast, the storage of patient information in data warehouses generally is not transparent to patients because it refers to the internal storage and use of patient information within an entity.

Michael Hogarth: I do not believe patients at our institution are told if their data reside in a data warehouse or whether it is part of a HIE. Patients are able to obtain an audit trail of the HIE exchanges involving their record, however. This is required by HIPAA.

Toby C. Cornish: Although the Notice of Privacy Practices as contributing entities may not specifically mention a CDW, any patient who has been treated at one of these entities will have data deposited in the data warehouse. In addition to privacy notifications, our HIE also publishes a list of participating providers and practices. Both our CDW and HIE can provide an audit of users that have accessed a patient's data upon request.

Brian R. Jackson: I strongly doubt that most patients know or can easily determine this.

What are some general guiding principles to balance patient privacy with societal use of clinical data warehouses?

Kristen B. Rosati: HIPAA provides a good "roadmap" for covered entities' internal use and external sharing of clinical data. For example, HIPAA has well-defined rules on how to deidentify patient information. However,

once information is deidentified, it is no longer subject to regulation by HIPAA or most state laws. With the rise of "big data" sets and the concomitant rise of patient privacy concerns, there has been a lot of discussion about allowing patients to control the use of their data—even if deidentified. However, the use of deidentified data is absolutely central to current efforts to improve the healthcare system and to discover new treatments and therapeutics. Requiring patient control (consent) to use deidentified information for these downstream purposes would seriously compromise these activities because of the expense of implementing a consent process and because of the consent bias that would be introduced into the use of large data sets. Moreover, there is substantial academic work showing that consent does not provide sufficient protection for people. That is especially true when the consent is "broad consent," which will not need to describe the particular type of activity or research that will be conducted with an individual's information. Rather, the answer is to prohibit reidentification of individuals in data used for research and quality improvement activities (with limited, thoughtful exceptions such as research conducted on reidentifiability). Federal law and the vast majority of state laws presently do not contain any prohibition against the reidentification of individuals unless the data are held by HIPAA covered entities. Prohibiting reidentification of individuals would address the acute need to use health information for research and at the same time honor individual rights.

Michael Hogarth: I believe HIPAA strikes a good balance between a degree of privacy protection (and substantial penalties for those who are intent on committing crime or harm) and benefit to the patient and society. HIPAA does not stipulate specific security controls, but does require a security officer to perform an annual evaluation of one's practices and that one must undertake any recommendations that are "common practice" at the time, which I think is a good way to go. We comply with HIPAA requirements and are working to become fully National Institute of Standards and Technology/Federal Information Security Modernization Act-compliant with protecting our clinical data (both identified and deidentified). A guiding principle we often use is that we always default to using deidentified data unless identification is required for the quality improvement measure or research being conducted using the CDW. Furthermore, we treat deidentified and identified data sets in the same fashion in terms of security controls. I would posit that deidentified data sets should be protected just as much as identified because if they are available to individuals bent on doing harm or fraud, patients can be reidentified fairly easily depending on other data sets that can be obtained (by purchase or publicly).

Toby C. Cornish: As a starting point, all CDWs must adhere to the HIPAA Privacy and Security Rules. One of the most important aspects of the HIPAA Privacy Rule is the "minimum necessary" standard. This standard directs that use or disclosure of protected health information should be limited to the minimum that is necessary for a particular purpose or function. To meet this standard, it is important that clinical data warehouses are capable of providing data in a variety of ways including fully deidentified data sets, limited data sets, and fully protected health information data sets. In each case, the provided data should be limited to the minimum data set needed to accomplish the intended purpose. Integration of an honest broker service with the data warehouse can frequently reduce the anxiety that researchers may have when working with deidentified data sets.

Brian R. Jackson: I believe that healthcare data warehouses and exchanges should have governance structures analogous to IRBs, with patient representation and transparency to the public. If you look at industries outside of healthcare, all sorts of consumer and financial data are warehoused and exploited for commercial activities. In some cases, consumers have been harmed without realizing it. It is incumbent on those of us in healthcare information technology to study these effects and develop ways to protect individuals from the risks associated with widespread use of big data.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: J.Y. Park, *Clinical Chemistry*, AACC; B. Jackson, ARUP Laboratories (nonprofit entity of University of Utah).
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: None declared.
Expert Testimony: None declared.
Patents: None declared.

Acknowledgment: We thank Sarah Davis, Principal Informatics Analyst, Health Data Compass; Cathy Boyd, Systems Architect Integration, UCHealth; and Connie Williamson, Director of Laboratory Information Systems, UCHealth.

Previously published online at DOI: 10.1373/clinchem.2016.266551