BUILDING TRUST IN THE POWER OF “BIG DATA” RESEARCH TO SERVE THE PUBLIC GOOD

Eric B. Larson, MD, MPH

To improve health care in the United States, the Institute of Medicine (IOM) has advanced the “learning health care system”—a place where “each patient-care experience naturally reflects the best available evidence, and, in turn, adds seamlessly to learning what works best in different circumstances.”

With increasingly more data being digitally collected in every health care encounter, prospects improve for the integration of clinical care and research. Routinely collected data “provide great potential for extracting useful knowledge to achieve the ‘triple aim’ in health care—better care for individuals, better care for all, and greater value for dollars spent,” according to a recent report from the IOM’s Clinical Effectiveness Research Innovation Collaborative (CERIC).

Some have lamented that health care lags behind other industries in leveraging advances in information technology and analytical techniques. However, the recent doubling in use of electronic health records suggests that “big data” will inevitably be applied to health care, potentially improving quality and efficiency. With the Affordable Care Act encouraging the development of accountable care organizations, incentives are increasing for stakeholders (including clinicians, insurers, purchasers, and patients) to collect, analyze, and exchange health care information.

Although such developments hold great promise, continued progress cannot be ensured unless the cultural and ethical issues related to patient privacy and the need for individual institutions to maintain some degree of control over the data they collect are addressed. In addressing these issues, a social compact for serving the common good by advancing knowledge should be developed. Ideally all stakeholders, and especially the public, would converge on a consensus to allow more widespread use of data generated in the process of health care when that use is generating knowledge that serves the public good. We see evidence that this is already happening at the local level, when researchers and subjects develop trusting relationships. It is also happening in networks of institutions who gather, share, and analyze health care data.

A recent report addressed the need for patient privacy while calling for a new ethical framework for learning health care systems. The report highlights the challenge of applying strict features of research oversight—much of which evolved years ago in the wake of the Tuskegee Study and other ethics violations and long before the age of electronic medicine. Now, with large quantities of deidentified health data on individuals being amassed, the boundary between research (which requires the participant’s consent) and quality improvement (which happens constantly without consent) is becoming blurred. The change is causing some ethicists to ask: Are requirements for individual consent creating insurmountable barriers to life-saving innovation?

As the CERIC reports, many benefits of the use of big data are already evident. These include improved public health surveillance and response to incidents such as Escherichia coli outbreaks and influenza pandemics. Health systems can now do a better job of targeting services such as cancer screening or routine monitoring of chronic illness to specific populations. Pioneering work at Group Health Cooperative demonstrates how data from a breast cancer screening registry and program can lead to earlier diagnosis of breast cancer, resulting in lower rates of late-stage disease at the time of recognition. Clinicians and patients have increasingly better access to scientifically based information for clinical decision making. It is becoming easier for physicians and patients to isolate information about subpopulations so treatments can be based on the experience of similar patients. In addition, big data can be used to improve safety and prevent errors. The CERIC authors point to the case of rofecoxib—a heavily marketed pain reliever linked in 2005 to higher risk of acute myocardial infarction—as an example of harm that could now possibly be avoided. Retrospective analyses demonstrate that the adverse effects and mortality related to rofecoxib could have been detected much earlier by systematically monitoring large health care claims databases.

Increased availability of routine health care data is also accelerating research. Studies that once required decades of data collection from selected populations in experimental settings can now be accomplished in just months by mining large data sets, producing more generalizable results. Technological advances driving these developments are gain-

©2013 American Medical Association. All rights reserved.

Author Affiliation: Group Health Research Institute, Seattle, Washington.
Correspondence: Eric B. Larson, MD, MPH, Group Health Research Institute, 1730 Minor Ave, Ste 1600, Seattle, WA 98101 (larson.e@ghc.org).
ing momentum, spawning many ambitious research enterprises. For example, the Healthcare Systems Research Collaboratory of the National Institutes of Health is designing and rapidly conducting pragmatic clinical trials that exploit routinely collected data and patient-reported outcomes. In contrast to traditional randomized clinical trials seeking explanations of efficacy, these pragmatic trials aim to quickly demonstrate effectiveness in real-world care-delivery systems. The research collaboratory recently funded 7 new trials, including assessments of ways to prevent hospital-acquired infections, improve colon cancer screening, and prevent suicide attempts. Another example is the US Food and Drug Administration’s Mini-Sentinel, a rapid-response electronic safety-surveillance system established to monitor drugs, devices, and vaccines in a population of more than 120 million people.

These examples illustrate how the use of health data infrastructure might improve the safety, quality, and efficiency of care. But to benefit from these advances and others not yet imagined, overly burdensome oversight and consent rules for research processes must be avoided. The new framework the Hastings Center has proposed for research within learning health care systems is aimed at reducing such hurdles, so opportunities for improvement are not lost. The authors argue that moral consideration must be given to harm that can occur when physicians do not have information needed to determine which approaches work best. The authors point out, for example, that more than half of medical treatments are used without sufficient proof of their effectiveness. Faden et al describe an intervention preventing central line–associated bloodstream infections in intensive care units that was almost halted because of concerns about research ethics oversight. What about concern and oversight for the 3000 patients who “will die unnecessarily each year in the United States from this type of infection?” Faden et al ask.

Ultimately, oversight policies and practices must be developed that will eliminate barriers to sharing information and prevent losing precious opportunities to improve care and save lives. This work will likely require some fundamental reorientation and reformulation from opinion leaders in research ethics. Together, all stakeholders—including the general public—must better understand that allowing researchers reasonable access to routinely collected health data can improve health and health care for all.

At the same time, trusted partnerships must be developed among institutions that collect health care data so that they can share it in ways that best serve patients’ needs. To work, such partnerships must recognize and accommodate each entity’s obligation to protect its patients’ privacy as well as its own business interests. The HMO Research Network (HMORN) is one of many consortia learning to navigate this shared territory. As described in the CERIC report, the HMORN includes 18 disparate integrated health plans that download administrative and clinical data at scheduled intervals into a “virtual data warehouse.” This arrangement allows each member organization to keep its primary data secure while contributing to important public-interest, proprietary research. Typically working with academic partners at leading research universities, the HMORN institutions are conducting research that benefits their own patient populations while creating generalizable knowledge that serves the public good.

Through advances in health information technology, the needed tools are available to prevent future harm, eliminate waste, and learn with better certainty which treatments are most effective. This can be accomplished without risking patient privacy or proprietary business interests. By overcoming cultural impediments based on outdated ideas about the collection and use of everyday health care information, it will be possible to fulfill the IOM’s vision of an integrated, comprehensive health care system using routinely collected health data for continual learning that seamlessly serves individual patients and the public good.

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported receiving support from National Institutes of Health grant US54AT007748-01; travel support from the National Institute on Aging; and royalties from UpToDate.

REFERENCES