

Toward Reuse of Clinical Data for Research and Quality Improvement: The End of the Beginning?

Today's health information technology (HIT) landscape provides an unprecedented convergence of comprehensive electronic health records, robust computational processing power, data-sharing capabilities, and emerging financial incentives that favor the widespread adoption and meaningful use of such systems. Even as the primary value proposition for HIT adoption by clinical practices remains a matter of debate (1), 2 articles in this issue (2, 3) help to illustrate the potential for secondary benefits by discussing the value of distributed health data networks for improving research and health care quality.

The multiple goals of population health, pharmaceutical surveillance, comparative effectiveness research, and other major initiatives being advanced to address the needs of our health care and research enterprises have created a growing need for access to high-quality, patient-level health information. One approach to enabling such access is the creation of centralized repositories to which data can be transferred and then readily accessed to answer questions, an approach that certainly has its merits. Alternatively, the distributed health data networks described in this issue can be designed with appropriate methods, policies, and systems to enable access and use of data housed in their original, disparate locations. As Maro and colleagues point out (2), a distributed approach has potential advantages—given today's regulatory, economic, and cultural realities—and the technological capabilities to enable such an approach exist today. Why, then, do we not see more widespread use of such systems?

The history of HIT successes and failures indicates that the technological challenges involved often pale in comparison with the many socio-organizational issues that must be understood and addressed to enable HIT advances (4). Any visitor to a local technology superstore knows how easy and inexpensive it is to collect and store vast amounts of data—the equivalent of thousands of patient records—in a device that fits as easily in a shirt pocket today as a 1.44-megabyte floppy disk did in 1990. However, the ease of information dissemination this capability enables appropriately raises personal privacy and intellectual property concerns. Adequate funding is another important factor to the success of HIT initiatives; however, funding alone does not guarantee success. One need only look at the many recent examples of failed regional health information organizations—and their successful counterparts—to appreciate the importance of effective governance structures, regulatory policies, and properly aligned organizational incentives in establishing and sustaining an effective distributed health data network (5).

In addition to these prerequisites, developers and those who attempt to leverage HIT resources have recognized

even more fundamental issues inherent to reusing clinical data. Although systems to enable clinical research from large health information collections have been around for some time (6–9), they have often been criticized because the quality and comprehensiveness of the clinical data were not up to research standards or the analytical methods used to overcome these limitations were inadequate to overcome systematic biases inherent to data collected primarily for clinical care (10–12).

Has data quality improved since the time of these earlier systems? In recent years, increased data capture in electronic systems, improvements in the speed and standardization of data transfers between systems, and the ability to leverage data from multiple clinical sources (such as objective test results or therapeutic information) have reduced the oft-criticized reliance on administrative and billing data sources. However, no purely technical solution can overcome the capture of inaccurate information by the user of a clinical information system. As such, nontechnical innovations that help improve the accuracy of recorded information and incentivize consistently accurate data collection are critical to the success of research initiatives that rely on the presence of such data.

Is data capture more comprehensive now than in the earlier systems? Comprehensive data capture all relevant exposures and outcomes, features that are essential to meaningful comparative effectiveness assessments. Within a single health system, improved technology has facilitated the automated capture of more comprehensive data in electronic form, which minimizes the need for laborious manual data extraction from paper charts. However, many patients receive care across several health systems and alternative health care settings, such that a single, comprehensive, longitudinal record rarely exists for any given patient. Unfortunately, the goal of widespread information system interoperability—to enable integrated health information access that spans different health systems and vendor products—remains elusive. Achieving interoperability and ensuring adherence to common standards will be critical to the success of secondary-use initiatives (13). Without it, distributed data access could have limited value—or lead to misinterpretation of the prevalence of and relationships between exposures and outcomes.

Finally, have the methods for analyzing routinely collected clinical data improved? The process of clinical care introduces treatment bias, in which the statistical association between therapy and outcome is confounded by measured and unmeasured factors that influence both the choice of treatment and the likelihood of the outcome. Instrumental variable and propensity score analyses, which have been applied for years in the social sciences, and new

methods, such as prior-event rate ratios (14), are increasingly being applied in the medical domain to overcome treatment biases, with variable degrees of success and acceptance. Understanding the clinical circumstances and types of research questions for which these methods may yield valid results—perhaps even consistent with those of traditional clinical studies—requires simulations, sensitivity analyses, and validation against the findings of randomized trials.

The promises of our current HIT environment are clearly great and growing. The efforts of biomedical informaticians, health services researchers, biostatisticians, and others have significantly advanced our knowledge of how to collect, organize, retrieve, analyze, and apply health data to improve individual patient care, as well as for such additional purposes as population health and biomedical research. However, significant collaborative effort by many of the stakeholders involved, including health care institutions and clinicians, HIT vendors, researchers, informaticians, regulators and policymakers, payers, and patients, is required to realize the full promise of these resources. If history is any indication, fostering and nurturing this collaboration will be challenging and take some time. Nevertheless, as the 2 articles in this issue help illustrate, current technology and existing models of success have put us in a better position today than we have been in before to realize the promises of HIT to advance research and create a safer and more efficient health care enterprise. To quote Sir Winston Churchill, “This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.”

Mark G. Weiner, MD
University of Pennsylvania
Philadelphia, PA 19104

Peter J. Embi, MD, MS
University of Cincinnati
Cincinnati, OH 45267

Grant Support: Dr. Weiner’s contributions were supported in part by the Agency for Healthcare Quality and Research (5-R18-HS-017099-02, “Crossing the Quality Assessment Chasm: Aligning Measured and True Quality of Care”) and the National Center for Research Resources, National Institutes of Health (1-UL1-RR-024134-04, Institutional Clinical and Translational Science Award). Dr. Embi’s contributions were supported in part by the National Center for Research Resources, National Institutes of Health (1-UL1-RR-026314-01, Institutional Clinical and

Translational Science Award) and the National Library of Medicine, National Institutes of Health (1-R01-LM-009533-01A1, “Evaluating EHR-based, Point-of-Care Trial Recruitment Across Clinical Settings”).

Potential Financial Conflicts of Interest: *Grants received:* M.G. Weiner (National Heart, Lung, and Blood Institute).

Requests for Single Reprints: Mark G. Weiner, MD, University of Pennsylvania, 423 Guardian Drive, Room 1116, Philadelphia, PA 19104; e-mail, mweiner@mail.med.upenn.edu.

Current author addresses are available at www.annals.org.
This article was published at www.annals.org on 28 July 2009.

Ann Intern Med. 2009;151:359-360.

References

1. Blumenthal D. Stimulating the adoption of health information technology. *N Engl J Med.* 2009;360:1477-9. [PMID: 19321856]
2. Maro JC, Platt R, Holmes JH, Strom BL, Hennessy S, Lazarus R, et al. Design of a national distributed health data network. *Ann Intern Med.* 2009;151:341-4.
3. Pace WD, Cifuentes M, Valuck RJ, Staton EW, Brandt EC, West DR. An electronic practice-based network for observational comparative effectiveness research. *Ann Intern Med.* 2009;151:338-40.
4. Kaplan B, Harris-Salamone KD. Health IT success and failure: recommendations from literature and an AMIA workshop. *J Am Med Inform Assoc.* 2009;16:291-9. [PMID: 19261935]
5. Adler-Milstein J, McAfee AP, Bates DW, Jha AK. The state of regional health information organizations: current activities and financing. *Health Aff (Millwood).* 2008;27:w60-9. [PMID: 18073225]
6. Johnson SB, Hripcsak G, Chen J, Clayton P. Accessing the Columbia Clinical Repository. *Proc Annu Symp Comput Appl Med Care.* 1994:281-5. [PMID: 7949935]
7. Steib SA, Reichley RM, McMullin ST, Marrs KA, Bailey TC, Dunagan WC, et al. Supporting ad-hoc queries in an integrated clinical database. *Proc Annu Symp Comput Appl Med Care.* 1995:62-6. [PMID: 8563360]
8. Tierney WM, McDonald CJ. Practice databases and their uses in clinical research. *Stat Med.* 1991;10:541-57. [PMID: 2057654]
9. Safran C, Porter D, Lightfoot J, Rury CD, Underhill LH, Bleich HL, et al. ClinQuery: a system for online searching of data in a teaching hospital. *Ann Intern Med.* 1989;111:751-6. [PMID: 2802433]
10. Hlatky MA. Using databases to evaluate therapy. *Stat Med.* 1991;10:647-52. [PMID: 2057662]
11. Byar DP. Problems with using observational databases to compare treatments. *Stat Med.* 1991;10:663-6. [PMID: 2057663]
12. McDonald CJ, Hui SL. The analysis of humongous databases: problems and promises. *Stat Med.* 1991;10:511-8. [PMID: 2057652]
13. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc.* 2007;14:1-9. [PMID: 17077452]
14. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ.* 2009;338:b81. [PMID: 19174434]

Current Author Addresses: Dr. Weiner: University of Pennsylvania, 423 Guardian Drive, Room 1116, Philadelphia, PA 19104.
Dr. Embi: University of Cincinnati, 231 Albert Sabin Way, Box 670840, Cincinnati, OH 45267.