Journal of Biomedical Informatics 71 (2017) 143–146

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Commentary

Ethics and subsequent use of electronic health record data

Lisa M. Lee

Presidential Commission for the Study of Bioethical Issues, 330 C St SW, L001, Washington DC 20201, United States

ARTICLE INFO

Article history:
Received 26 March 2017
Revised 28 May 2017
Accepted 29 May 2017
Available online 31 May 2017

Keywords:
Ethics
Electronic health records
Privacy
Research
Learning health system

ABSTRACT

The digital health landscape in the United States is evolving and electronic health record data hold great promise for improving health and health equity. Like many scientific and technological advances in health and medicine, there exists an exciting narrative about what we *can* do with the new technology, as well as reflection about what we *should* do with it based on what we value. Ethical reflections about the use of EHR data for research and quality improvement have considered the important issues of privacy and informed consent for subsequent use of data. Additional ethical aspects are important in the conversation, including data validity, patient obligation to participate in the learning health system, and ethics integration into training for all personnel who interact with personal health data. Attention to these ethical issues is paramount to our realizing the benefits of electronic health data.

Published by Elsevier Inc.

1. Introduction

It is an exciting time in digital health in the United States; the landscape is evolving and data hold great promise for improving health and fostering health equity. Digital health has the potential to improve patient care and to provide an important source of data for health services and epidemiologic research as well as quality improvement and cost analyses. Determining whether and how to use these data for purposes other than point of care patient service requires us to consider the ethical dimensions of the collection and use of health data.

The backbone of digital health is the electronic health record (EHR). In the United States in 2015, after 6 years and nearly $30 billion of federal investment, 87% of physicians and 95% of hospitals are solely or partially using electronic health records for recording and storing patient encounter data [1]. The promise of immediate provider access to a patient's record has been partially realized, at least within the same health care organization or practice system. A truly mobile medical record—one that 'travels' to the computer screen of any health care provider a patient happens to see, whether in New York or New Mexico—is not yet a reality for all Americans, but promises to be in the near future. A fully implemented mobile EHR promises more efficient, accurate, timely, and useful information [2]. Anticipated benefits include a reduction in medical errors, faster reimbursement processing time, improved health outcomes for individuals and communities, and estimated cost savings of $81 billion [3]. These benefits will be realized not only through effective use of EHR data by clinicians and administrators, but also through use by researchers.

As is the case with many scientific or technological advances in health and medicine, there are exciting narratives about what we *can* do with the new technology, as well as reflection about what we *should* do, based on what we value. One important question about EHR technology is whether we should use patient data collected for direct point of care services for activities intended for the collective good—activities such as epidemiologic research and health system quality improvement? Practically speaking, decisions about what we should do often entail an examination of what we are willing to risk for the expected gain. This risk benefit ratio is a fundamental concern that institutional review boards (IRBs) must consider when evaluating whether research participants are adequately protected. And it is something that we must think about when reflecting on what we ought to do with the growing stores of data flowing into EHRs.

Most ethical reflections about using EHR data for research have considered the important issues of privacy and data security, informed consent for data uses, and ownership of patient data. The basic bioethical principle of respect for persons drives the major concern about data privacy and security. EHR policies and systems employ numerous techniques, including legal requirements, encryption, access limits, and audit logs, to protect data privacy. In the United States, federal and state policies protecting EHR data have existed for over a decade and a great deal of informatics and ethics literature has addressed these issues. Such national policies as the HIPAA Privacy Rule [4], the HITECH Act of 2009 [5], and the Federal Policy for the Protection of Human Subjects ('Common Rule') [6] outline legal and policy protections for health data in

E-mail address: Lisa.Lee@bioethics.gov

http://dx.doi.org/10.1016/j.jbi.2017.05.022
1532-0464/Published by Elsevier Inc.

practice and research settings. While reflections about data ownership and privacy strike a chord with American individualism, there are additional ethical issues such as non-maleficence, reciprocity, and other professional duties that play an important role in the conversation.

2. Ethical considerations and potential benefits

The potential public benefits of using EHR data for health services and epidemiologic research are enormous. Creating a learning health system—a system that makes ethical use of the vast stores of patient data to yield effective and efficient health interventions in real time for all persons—is a goal that nearly all of us agree is worth pursuing [7,8]. In order to achieve projected benefits, EHR data must be valid and useful. Generally speaking, there is greater tolerance for lower quality data in administrative data compared with clinical trial data; most researchers are aware of the limitations of data collected for one purpose and used for another. EHR data, however, must be of high quality for patient care. This higher quality could benefit researchers as well, especially compared with the low quality of many other types of administrative data that researchers use. Still, there is little agreement or consistency on how to assess the quality of EHR data [9]. Validity is a multidimensional construct that represents both the accuracy of the data—how correct it is relative to the truth, as well as the reliability of the data—how likely it is that the measure will be the same when assessed again and again. For example, if a patient with a particular risk factor denies the risk factor every time the question is posed, the measure would be inaccurate (patient has the risk factor, but responds 'no') but reliable (patient provides the same response every time the question is posed).

Data validity depends on a number of inputs including coding of diagnoses and procedures, self-reported measures of risks and outcomes, accurate data entry, and, when more than one database is used to construct a story about a patient over time, when matching algorithms incorrectly match or fail to match an event to a particular record. There is a large literature indicating that errors can be (and are) introduced at any and all of these inputs. Once validity-compromising errors are introduced into an EHR, there are serious challenges to correcting the information. Although the HIPAA Privacy Rule provides for a patient to request an amendment to their EHR data, there are few effective processes in place to do so in medical centers. In addition, with the increased sharing of EHR data (especially de-identified datasets) to promote health care efficiency and effectiveness, errors might not be discovered until after datasets are in the hands of health services researchers and epidemiologists.

In addition to being valid, EHR data must be useful. Given the broad coverage across many population subgroups, EHR data can provide large sample sizes with adequate power to look at small subgroups, which is especially useful for health equity research. There is much hope that EHR data can be used to inform health policy. To be useful in informing policy, policy makers must trust that the data and the findings that result from studies using the data are valid and complete, and that the recommendations will prove beneficial and cost effective. The kind of data that best guides policy are data that lead to unbiased information, whether about cost or clinical practice. Unbiased findings from research using EHR data result from data that are complete and statistically representative. Weiskopf et al. have shown that completeness of EHR varies in important ways [10] and this incompleteness can affect the validity of studies conducted with EHR. In addition to problems with overall completeness, differential completeness can reduce validity. EHR of ill patients contain more complete data than EHR of healthy patients, which can lead to incorrect or biased—and less useful—results [11].

The typical designs of health services research and epidemiological studies that are supported by EHR are observational, whether prospective or retrospective, and case-control. While observational and case-control studies provide some evidence for action, they are considered weaker than the gold standard, randomized controlled trials (RCTs). The observational studies that can be conducted using EHRs hold promise for a number of important research questions, especially for research questions that are not ethically amenable to an RCT. A well-conducted observational trial is also useful for hypothesizing about causal factors, though in many circumstances, findings must be confirmed with RCT results. Researchers must communicate clearly about the strength of evidence resulting from observational studies using EHR data, as methods and findings will have to be strong enough to gain trust of policy makers and regulators who are in the position to use the results to affect change.

An additional potential benefit of large-scale EHR use is the opportunity for all patients to contribute to public beneficence by participating in improving the health system. Involvement in the learning health system has been considered an obligation by some, and might contribute to an increase in a sense of solidarity in communities and the country [12,13]. It might provide the rugged individualist with the opportunity to see how working together for the good of *all* of us is good for *each* of us.

Knowing whether and the degree to which we can realize the potential benefits of EHR is essential for our ability to judge whether the benefits outweigh the risks. These are empirical questions, many of which still need answers. We are not there yet, but given that our enormous investment in technology is not risk-free, it is incumbent upon us to accurately measure and communicate the benefits of EHRs.

3. Ethical considerations and potential hazards

In addition to numerous potential benefits of EHR, there are risks. The primary risk of EHRs is that data will be used in ways for which it was not intended and that such uses will bring harm to individuals or communities. Harm, if it comes, could result from disclosure of accurate or inaccurate information and could range from a minor annoyance such as predictive marketing to more serious harms such as discrimination and stigma, job loss, denial of certain types of insurance coverage, or worse. Unauthorized disclosures of private health data are common and can be particularly harmful [14]. Unlike victims of breaches of financial data, to whom reparations can be made, victims of breaches of private health data cannot be 'made whole'; information cannot be 'taken back'. The perpetrator of unauthorized health data disclosures might be an identifiable 'bad actor' as occurred in the 2011 case of a contract employee posting a photo and medical information of a patient at Providence Holy Cross Medical Center in Mission Hills, California [15], or an unidentifiable hacker like the one who recently disclosed health information of 2016 Olympic athletes [16]. Data security is a necessary condition for the ethical optimization of electronic health information.

Another risk that has received less attention is the risk that EHR data will be biased in unexpected or uncharacterized ways that will result in erroneous conclusions that lead to harmful or costly policy changes. While EHR data are relatively new, erroneous or incomplete conclusions from observational research have resulted less than optimal policies and practices in some cases [17]. Information gleaned from the research use of EHR data are subject to these biases and errors as well, and we must remain aware and vig-