By Joachim Roski, George W. Bo-Linn, and Timothy A. Andrews

DOI: 10.1377/hlthaff.2014.0147
HEALTH AFFAIRS 33,
NO. 7 (2014): 1115–1122

Creating Value In Health Care Through Big Data: Opportunities And Policy Implications

Joachim Roski (Roski_joachim@bah.com) is a principal at Booz Allen Hamilton in McLean, Virginia.

George W. Bo-Linn is a senior adviser in the Improvement Division, Alvarez and Marsal, in San Francisco, California.

Timothy A. Andrews is a vice president on the health team at Booz Allen Hamilton in McLean.

ABSTRACT **Big data has the potential to create significant value in health care by improving outcomes while lowering costs. Big data's defining features include the ability to handle massive data volume and variety at high velocity. New, flexible, and easily expandable information technology (IT) infrastructure, including so-called data lakes and cloud data storage and management solutions, make big-data analytics possible. However, most health IT systems still rely on data warehouse structures. Without the right IT infrastructure, analytic tools, visualization approaches, work flows, and interfaces, the insights provided by big data are likely to be limited. Big data's success in creating value in the health care sector may require changes in current polices to balance the potential societal benefits of big-data approaches and the protection of patients' confidentiality. Other policy implications of using big data are that many current practices and policies related to data use, access, sharing, privacy, and stewardship need to be revised.**

Decreases in the costs of computing power and storage, the proliferation of "smart" devices, and the growth of electronic communication are generating an explosion of health care data. In 2012 such data worldwide amounted to approximately 500 petabytes (one petabyte equals 10^{15} bytes of digital information), or the equivalent of the contents of 10 billion four-drawer file cabinets. By 2020 this amount will likely increase to 25,000 petabytes, or the equivalent of the contents of 500 billion file cabinets.[1]

A large integrated health system such as Kaiser Permanente, with approximately nine million members, may manage up to forty-four petabytes of data through its electronic health record (EHR) system alone. This is 4,400 times the equivalent of the data stored in the Library of Congress.[2]

However, this data explosion creates value only when it is used to make better and faster decisions. Big data—that is, the sophisticated and rapid analysis of massive amounts of diverse information—has enabled greater customer knowledge; customized outreach efforts in many business sectors; and significantly boosted productivity, sales, and overall economic benefit.[3–6]

Facilitated in part by the enormous investment of the health care sector in information technology (IT), big data could be among the catalysts driving improvements in health outcomes while lowering costs. Cleveland Clinic has identified big data as one of the biggest medical innovations in recent years.[7]

Given the imperative to improve performance in health care and big data's potential to generate value, why are big-data solutions not more common? Moving forward, what IT infrastructure, related technologies, and policies are required to support big data? This article examines these questions and the outlook for the integration of big data into the health care system.

What Is Big Data?

No single widely accepted definition of *big data* appears to be available. However, at least three defining features of it—the three Vs—seem to be generally accepted: volume, variety, and velocity.

Volume is a key characteristic of big data. Massive amounts of data strain the capacity and capability of traditional data storage, management, and retrieval systems such as data warehouses. Big data requires flexible and easily expandable data storage and management solutions.

The second characteristic is variety. Health care data today come in many formats, such as the structured and free-text data captured by EHRs, diagnostic images, and data streaming from social media and mobile applications. However, much of this information is not put to use to improve health or health care. For example, less than 15 percent of health data in EHRs might be entered in structured data fields that allow those data to be analyzed using traditional retrieval and analysis methods. Big-data approaches enable the efficient linking and analyses of disparately formatted data to answer particular operational, business, or research questions.[8]

The third characteristic is velocity. Most traditional health IT infrastructures are not able to process and analyze massive amounts of constantly refreshed, differently formatted data in real time. Big-data infrastructure makes it possible to manage data more flexibly and quickly than has been the case, as we explain below.

More identically formatted data are available for analyses than ever before. For example, the Centers for Medicare and Medicaid Services (CMS) is making Medicare claims data available to researchers and what the Affordable Care Act refers to as "qualified entities" for analysis. Analyzing these data together with claims data from private insurers can offer significant benefits. However, we would not characterize these efforts as leveraging big data's potential. Only if CMS data were combined with differently formatted data (for example, information from EHRs) and rapidly analyzed would the three Vs of big data be achieved and big data's full potential realized.

Big Data's Potential Value

Big data may have the potential to create approximately $300 billion annually in value in the health care sector, two-thirds of which would be generated by lowering health care expenditures.[9] Big data has already demonstrated its economic and clinical value on multiple occasions. First, the delivery of personalized medicine (individualized diagnoses and treatments based on a patient's detailed risk profile) has been demonstrated for care of patients with cancer or other conditions.[10,11] Second, the use of clinical decision support systems has been enhanced by the automated analysis of x-rays, computed tomography (CT) scan images, and magnetic resonance imaging (MRI) images and the mining (described below) of medical literature to tailor treatments to individual patients' risk profiles.

Third, reliance on patient-generated data has been demonstrated using mobile devices to tailor diagnostic and treatment decisions as well as educational messages to support desired patient behaviors.[12] For example, the Veterans Health Administration has launched a number of mobile health care initiatives that target specific patients and providers through the rapid collection and analysis of patient-generated data.[13]

Fourth, big data–driven population health analyses have revealed patterns that might have been missed had smaller batches of uniformly formatted data been analyzed instead. One example is the Durkheim Project, a collaboration between the Veterans Health Administration and Facebook, which is using real-time prediction software to analyze voluntary, opt-in data from veterans' social media accounts and mobile phones for suicide risk prevention.[14]

Finally, big-data tools for fraud detection and prevention, such as those used by CMS, have replaced earlier manual documentation processes. These tools have generated over $4 billion dollars in recovered costs in 2011 alone.[15]

IT Infrastructure Required For Big Data

Electronic health care data are beginning to be available in massive amounts. For example, 85 percent of US hospitals have already adopted an EHR system,[16] and most physicians' offices have started to digitize patients' records.[17,18]

However, for health organizations to rely on big data, enabling IT infrastructure has to be available. Installing such infrastructure and its components is becoming increasingly less expensive. Nonetheless, the installation still requires a significant investment of time (including time spent in training) and money, and it involves lost productivity during the transition process.

In addition, to be successful with big data, organizations need to develop processes and policies that accommodate new protocols, time requirements, risk factors, and mandates for managing data, especially in the area of privacy and security.

Most health IT systems rely on large data warehouses that have a static organization and a blueprint (or data schema) for how to construct a database. Sometimes long and complex processes collectively known as "extract, transform, and load" are required to transfer data from sources such as EHR systems into a data warehouse for analysis.

With big-data IT infrastructure, data can be rapidly ingested; tagged to indicate data properties, including origin; and stored indefinitely in a large, open information space called a "data lake." The data lake can accommodate differently formatted data elements, such as patients' names, laboratory values, documents (including progress notes, discharge instructions, and Clinical Document Architecture documents such as summaries of care that are encoded in Extensible Markup Language), and electrocardiograms and MRIs. New data and formats, such as streaming data from implanted pacemakers or Twitter data feeds, can be added without having to be transformed into uniform formats.

To answer a specific question with a particular analytic approach, analysts create a schema that links the appropriate items in the data lake. There can be as many schemas as there are questions, and schemas can be changed without affecting the raw data.

This approach takes advantage of technologies that are not available with data warehousing or the other traditional ways of managing databases that are still predominant in many health organizations. For example, current approaches to measuring the costs of care rely on payment claims records from data warehouses. Augmenting these analyses with clinical data from EHRs might make it possible to use episode grouping approaches that have higher validity and reliability than current approaches have. However, the static model of claims data in a traditional data warehouse is not conducive to this type of analysis, and adding clinical data often means having to unload the data in the warehouse and execute a new extract, transform, and load process.

In contrast, researchers conducting a cost-of-care analysis in a data lake could simply add new clinical records, queries, and algorithms (or mathematical formulas for solving the problems). The online Appendix provides additional information about health IT infrastructure elements needed to accommodate big data.[19]

Keeping Track Of Data Provenance

Big data offers many ways for organizations to tag and track the origin and use of data, as well as who is allowed to access the data. The ability to keep track of data's provenance (the history of the data's origin, ownership, use, and modification) can facilitate essential operations, such as meeting a state's legal requirements about how long data must be stored.[20]

In addition, the storage of tagged data in a data lake gives health care organizations the ability to use original data for repeated analyses to verify findings or identify problems. One example involves the analyses required to support new health care payment models. Organizations using these models may need to be able to adjust payments based on combinations of clinical quality and financial (that is, claims) data. If results from one set of data do not match results from another—say, the patients identified by diagnosis relying on laboratory values, such as blood sugar levels, in clinical records do not match those identified by *International Classification of Diseases* (ICD) codes in the claims records—health care organizations may need access to the original data to resolve the conflict.

$300 billion

In value

Big data may have the potential to create approximately $300 billion annually in value in the health care sector.

Protecting Data Security And Privacy

Security and privacy concerns related to big data and an IT infrastructure that is accessed through remote locations—for example, via a data cloud and services hosted in the cloud—have presented a significant barrier to the adoption of big-data approaches.[21–23] Some institutions have been experimenting with "private clouds," using Dropbox or other tools to share protected health information.[21] However, private clouds are limited in terms of elasticity and economies of scale.

An alternative that has arisen to address these limitations is the cloud service provider (CSP). Flexibility and elasticity are built into CSPs' business models. When more computing or storage capacity is needed, a CSP client can procure it almost instantaneously, pay for only what is needed, and then remove the extra capacity if it is no longer required.

For a number of reasons, CSPs may also be able to address data privacy and security concerns more effectively and efficiently than individual health organizations can. First, CSPs have already made extensive investments in security measures that could benefit additional organizations at no extra cost.[24] Economies of scale and scope enable CSPs to maintain defenses against cyber attacks or data hackers that may be more sophisticated than defenses that a single health care institution can afford.

Second, CSPs typically rely on up-to-date tools that support stringent security enforcement. For instance, the open-source application Accumulo,[25] originally developed to support national intelligence work, uses sophisticated cryptographic methods to place security tags on every piece of data to assign specific access rights to specific users.

Such applications could enable health care organizations to place different levels of security on different types of data, from demographic information to highly sensitive health data related to substance abuse and sexually transmitted diseases. This would provide the organizations with additional control over the ability to share information without compromising patients' privacy or releasing proprietary data.

Enabling Data Integration And Interpretation

Without the right IT infrastructure, analytical tools, visualization approaches, work flows, and interfaces, big-data insights are likely to be limited. The busy schedules of time-strapped health care providers prohibit the review of complex analyses unless the findings have an immediate impact on diagnosing or treating patients and the review is woven into current work flows.

Data in the patient record guides diagnostic and treatment decisions. Potentially useful information—for example, patient-generated data such as pedometer readings that are intended to provide a feedback mechanism and act as an incentive for a patient to maintain an exercise regimen—may not yet be integrated into health organizations' IT systems and clinicians' work flows. Thus, these data may not be used to improve diagnostic or therapeutic decision making during the patient-clinician interaction.[26]

To address potential liability concerns, health care organizations and providers must either devote more time and resources to evaluating new insights gained through big data or refuse to incorporate additional data and analysis results into the patient record.[27] Meanwhile, the patients being treated are often ill equipped to interpret complex analyses without the aid of sophisticated data visualization tools. Furthermore, these visualizations may still be incomplete because not all available and potentially relevant data (such as pedometer readings) have been included.

Policy Concerns

Current practices and policies related to data use, access, and protection may need to be examined and altered by policy makers and health care decision makers, to allow for creation of the greatest value.[28]

DATA SHARING AND COLLABORATION Big data's potential cannot be realized without amassing and analyzing vast amounts of diverse data. Moreover, the data involved may not all be owned or controlled by a single organization or institution. CMS and other government agencies are committed to greater data sharing. Nonetheless, market forces may be impeding wider data sharing and analysis in the private sector.

Market forces may be impeding wider data sharing and analysis in the private sector.

For example, competing health care organizations that treat overlapping patient populations in a community may be reluctant to share relevant data, typically because each organization fears that others could use its data for competitive advantage. This is the case even if sharing data might enable the organizations to identify and address suboptimal practice patterns and outcome gaps across their common coverage area, such as communitywide patterns of inappropriate use of antibiotic medications.

The Office of the National Coordinator for Health Information Technology has identified this phenomenon and the barrier it represents to harnessing big data's potential to improve care and population health.[29] Strategically pooling data—such as sharing data across organizations while controlling access to the source data—is one way to balance the competing demands of data protection and community benefit.[29] As described above, big data may make it possible to install effective and efficient data access controls while providing novel insights based on pooled data and sophisticated analyses.

REEXAMINING PRIVACY POLICIES High-profile data breaches and leaks may have led consumers to question whether health care organizations are doing enough to protect sensitive data.[30] At the same time, many patients and other consumers are sharing personally identifiable data—including photographs, videos, and messages—through social media such as Facebook and Twitter. Some organizations have used these voluntarily shared data in a deidentified state to achieve public health goals, including accurately predicting an influenza outbreak and surges in emergency department visits for flu-related symptoms.[8]

Yet recent developments have raised concerns about whether these deidentification efforts are enough to protect the privacy of patients in a big-data environment.[31] In a high-profile DNA study,

Policies guiding responsible data stewardship need to be drafted or adopted.

40 percent of the anonymous participants were reidentified through sophisticated analyses.[32]

CONSENT The concepts of consent, data ownership, and control may also need to be reexamined. In the course of a patient's treatment, data can be created with the potential to contribute to more effective therapies for everyone. Moreover, the creation of these data relies on providers' equipment and treatment decisions and is often compared against clinical reference data developed by private organizations and public institutions to aid in determining the appropriate diagnostic and treatment course for a given patient. How should data-use policies balance the need to preserve individual confidentiality and the potential to reap benefits for the entire community?

The concept of consent implies a person's right and ability to control information about him- or herself and to impose limitations on its use and reuse by entities such as researchers and health care organizations. Current consent policies imply that it is in society's best interest to prioritize individual control of data over societal benefits that could be achieved by sharing such data for multiple purposes. However, to harness big data's potential, consent may need to evolve from strict regulations for every potential use of data to a balance between personal control and informed sharing in the service of public health, environmental protection, or other goals.[33]

The Personal Genome Project has adopted consent procedures that allow for the use and reuse of data to support multiple research objectives.[34] The project has strict requirements to protect data. But it also educates patients about how their contributed identifiable data could benefit them and others when used and reused, even if the details about how the data will be used are not known when the project seeks patients' consent for future use. Its consent procedures also stipulate to what extent data can be protected and how participants can share their data with other scientific research projects to reap greater potential benefit.

In examining such an approach and the potential for its wider adoption, policy makers should balance the value of data uses and potential privacy risks, assess the practicability of obtaining true and informed consent, and understand the enforceability of restrictions on future data sharing and use.[35]

STEWARDSHIP Policies guiding responsible data stewardship—the knowledgeable and appropriate use of data—need to be drafted or adopted. These policies should address data collection, viewing, storage, exchange, aggregation, and analysis.

Efforts to develop policies in this area have already begun. Both the Federal Trade Commission[36] and the Organization for Economic Cooperation and Development[37] have identified principles of fair information use and guidelines for data privacy protection and data sharing and use. The two sets of principles address the issues of providing appropriate notice to consumers about the use of data, consumers' consent and options for controlling their data, consumers' access to and ability to modify collected data, data integrity and security, and enforcement options.

These principles may serve as guideposts for health care policy makers who seek to balance privacy protection with new ways of sharing data for societal benefit.[36,37] They also may provide guidance to health care organizations that are using big data and sharing data with other organizations.[38]

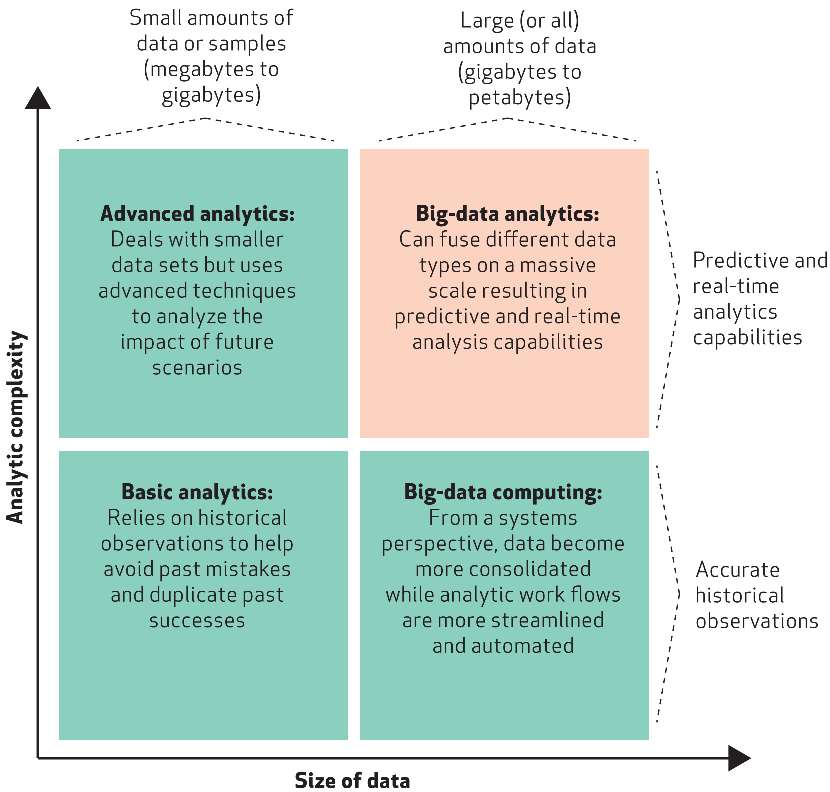
Big Data For Scientific Inquiry And Knowledge Gain

The emerging discipline of data science addresses the challenges of processing and analyzing big data. Data science relies heavily upon elements of signal processing, statistics, machine learning, text retrieval, and natural language processing to analyze data and interpret results. Effective data science teams working on health problems bring together three skill sets: domain expertise (understanding the health or health care context of the particular problem to be solved), mathematics (the theoretical space, or data models, in which data science problems are examined), and computer science (the environment in which data products are created).

Exhibit 1 illustrates different analytic approaches. Traditional research approaches in health are hypothesis driven and based on deductive reasoning, with analyses typically conducted on relatively small amounts of data that are collected in highly controlled circumstances, such as randomized clinical trials. Big data opens up additional possibilities for discovery in terms of scope, flexibility, and visualization.

EXHIBIT 1

Data Analytic Approaches, By Size Of Data And Analytic Complexity



SOURCE Booz Allen Hamilton. Cloud analytics playbook [Internet]. McLean (VA): Booz Allen Hamilton; 2012 [cited 2014 Apr 29]. Available from: http://www.boozallen.com/media/file/cloud-analytics-playbook.pdf. Adapted with permission.

Techniques such as data mining (the computational process of discovering patterns in large data sets) facilitate inductive reasoning and exploratory data analysis, allowing researchers to identify data patterns that are independent of specific hypotheses.

Take, for example, an analysis of progress notes and discharge instructions for a large population of patients with diabetes. In a big-data analysis, topic modeling techniques (statistical modeling used to discover the topics that occur in a collection of documents) could be used to present findings in the form of a graph[39] and yield insights that traditional analytical methods would not make obvious. Similarly, by relying on natural language processing to evaluate notes in EHRs in real time, deleterious outcomes such as severe sepsis can be predicted from the presence of clinical "red flags." This allows medical staff to initiate early treatment and decrease potential morbidity and mortality.[40]

Another example of big-data approaches to research is the analysis of logs of the search activities of populations of computer users to identify drug safety concerns. Research suggests that these logs can contribute to drug safety surveillance by uncovering previously unknown drug-drug interactions that would have been difficult to find through traditional clinical trials or other research methods.[41] Moreover, this method may be relatively inexpensive to implement, compared to analyzing data from sources such as EHRs.

By relying on an inductive approach such as data mining, big data has the potential to facilitate discovery across traditionally segregated scientific disciplines (such as health care, public health, economics, demographics, biology, and political science), explanatory models (for example, biomedical and socioeconomic models), and areas of expertise or influence. It may become possible to identify relationships among health phenomena that a single discipline alone could not uncover.

For example, health care providers in Camden, New Jersey, combined geospatial data on clinics' locations and patients' residences with health care organizations' utilization data. The providers were thus able to determine and address the complex reasons for why a subset of the community's population was using the emergency department to obtain care for routine medical needs and thus incurring unnecessary expenses.[42]

Conclusion

Health care organizations may be able to take their first steps toward implementing big-data solutions by defining the customers who will be using the information and analysis provided by big data; the solutions that will generate big data and enable analysis; and the value the organizations wish to generate, such as overall decreased costs and improved quality of care. Such an approach is likely to lead to focused, iteratively conducted efforts to move toward big-data solutions for specific problems instead of investing heavily in potentially ill-defined wide-ranging applications.

To realize big data's promise, health organizations and policy makers alike may need to set aside traditional mind-sets and embrace new approaches, overcoming barriers to promote data sharing with the appropriate protections, and collaboratively working toward the goal of delivering better outcomes at lower costs. ■

The authors thank Saniah Fatemi and Angela Bright for their editorial support in preparing this article.

NOTES

1 Pial S, Claps M. Bigger data for better healthcare. Framingham (MA): Intel; 2013 Sep.
2 Cusack CM, Hripcsak G, Bloomrosen M, Rosenbloom ST, Weaver CA, Wright A, et al. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. J Am Med Inform Assoc. 2013;20(1):134–40.
3 Bloom N, Sadun R, Van Reenen J. Americans do IT better: US multinationals and the productivity miracle. Am Econ Rev. 2012;102(1): 167–201.
4 McKinsey Global Institute. US productivity growth, 1995–2000. New York (NY): The Institute; 2001 Oct.
5 McKinsey Global Institute. How IT enables productivity growth. New York (NY): The Institute; 2002 Oct.
6 McKinsey Global Institute. The next frontier for innovation, competition, and productivity. New York (NY): The Institute; 2013.
7 Cleveland Clinic. Cleveland Clinic unveils top 10 medical innovations for 2012 [Internet]. Cleveland (OH): Cleveland Clinic; 2011 Oct 6 [cited 2014 Apr 28]. Available from: http://my.clevelandclinic.org/media_relations/library/2011/2011-10-6-cleveland-clinic-unveils-top-10-medical-innovations-for-2012.aspx
8 TechAmerica Foundation, Federal Big Data Commission. Demystifying big data: a practical guide to transforming the business of government [Internet]. Washington (DC): The Foundation; [cited 2014 Apr 28]. Available from: http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf
9 McKinsey Global Institute. The big-data revolution in US health care: accelerating value and innovation. New York (NY): The Institute; 2013.
10 McAfee A, Brynjolfsson E. Big data: the management revolution. Harv Bus Rev. 2012;90(10):61–68.
11 Terry K. Cancer researchers mine big data to individualize treatment. InformationWeek HealthCare [serial on the Internet]. 2013 Mar 28 [cited 2014 Apr 28]. Available from: http://www.informationweek.com/healthcare/electronic-health-records/cancer-researchers-mine-big-data-to-individualize-treatment/d/d-id/1109305
12 Institute for Health Technology Transformation. Population health management: a roadmap for provider-based automation in a new era of healthcare [Internet]. New York (NY): The Institute; [cited 2014 Apr 28]. Available from: http://ihealthtran.com/pdf/PHMReport.pdf
13 Slabodkin G. Veterans Health Administration launches mobile healthcare initiatives. FierceMobileHealthcare [serial on the Internet]. 2013 May 12 [cited 2014 May 7]. Available from: http://www.fiercemobilehealthcare.com/story/veterans-health-administration-launches-mobile-healthcare-initiatives/2013-05-12
14 Bertolucci J. Big data project analyzes veterans' suicide risk. InformationWeek [serial on the Internet]. 2013 Jul 9 [cited 2014 Apr 28]. Available from: http://www.informationweek.com/big-data/big-data-analytics/big-data-project-analyzes-veterans-suicide-risk/d/d-id/1110668
15 McNickle M. 5 basics of big data. Healthcare IT News [serial on the Internet]. 2012 Jun 13 [cited 2014 Apr 28]. Available from: http://www.healthcareitnews.com/news/5-basics-big-data
16 Charles D, Furukawa M, Hufstader M. Electronic health record systems and intent to attest to meaningful use among non-federal acute care hospitals in the United States: 2008–2011 [Internet]. Washington (DC): Department of Health and Human Services, Office of the National Coordinator for Health Information Technology; 2012 Feb [cited 2014 Apr 28]. (ONC Data Brief No. 1). Available from: http://www.healthit.gov/media/pdf/ONC_Data_Brief_AHA_2011.pdf
17 Moses H 3rd, Matheson DH, Dorsey ER, George BP, Sadoff D, Yoshimura S. The anatomy of health care in the United States. JAMA. 2013;310(18): 1947–63.
18 Wright A, Henkin S, Feblowitz J, McCoy AB, Bates DW, Sittig DF. Early results of the meaningful use program for electronic health records. N Engl J Med. 2013;368(8): 779–80.
19 To access the Appendix, click on the Appendix link in the box to the right of the article online.
20 Maryland Office of the Secretary of State. Regulation 10.01.16.04. 04 maintenance of medical records [Internet]. Annapolis (MD): The Office; [cited 2014 Apr 28]. Available from: http://www.dsd.state.md.us/comar/getfile.aspx?file=10.01.16.04.htm
21 HIStalk Advisory Panel. Advisory panel: PHI stored in the cloud by physicians and employees. HISTalk [serial on the Internet]. 2013 Aug 14 [cited 2014 Apr 28]. Available from: http://histalk2.com/2013/08/14/advisory-panel-phi-stored-in-the-cloud-by-clinicians-and-employees/
22 Halamka J. Cool technology of the week: HIPAA-compliant public cloud. Healthcare IT News [serial on the Internet]. 2012 Apr 13 [cited 2014 Apr 28]. Available from: http://www.healthcareitnews.com/blog/cool-technology-week-hipaa-compliant-public-cloud
23 Cox P. HIPAA in public cloud: the rules have been set. Cloud Management Blog [blog on the Internet]. 2013 Jul 3 [cited 2014 Apr 28]. Available from: http://www.rightscale.com/blog/enterprise-cloud-strategies/hipaa-public-cloud-rules-have-been-set
24 athenahealth. What cloud-based services can do for your medical practice or care organization [Internet]. Watertown (MA): athenahealth, Inc.; [cited 2014 Jun 15]. Available from: http://www.athenahealth.com/whitepapers/cloud-based-services
25 Henschen D. Defending NSA Prism's big data tools. InformationWeek [serial on the Internet]. 2013 Jun 11 [cited 2014 Apr 28]. Available from: http://www.informationweek.com/big-data/big-data-analytics/defending-nsa-prisms-big-data-tools/d/d-id/1110318?
26 Neff G. Why big data won't cure us. Big Data [serial on the Internet]. 2013 Sep [cited 2014 Apr 28]. Available from: http://ginaneff.com/wp-content/uploads/2013/09/Neff_Why-big-data-wont-cure-us.pdf
27 Stuckey M. Unsolicited data in EHRs posing workflow, legal challenges for some providers. iHealthBeat [serial on the Internet]. 2013 Oct 28 [cite4d 2014 Apr 28]. Available from: http://www.ihealthbeat.org/insight/2013/unsolicited-data-in-ehrs-posing-workflow-legal-challenges-for-some-providers
28 Porter ME. What is value in health care? N Engl J Med. 2010;363(26): 2477–81.
29 Hawai'i Island Beacon Community. Enabling health information ex-

change to support community goals [Internet]. Honolulu (HI): The Community; 2013 [cited 2014 Apr 28]. Available from: http://www.hibeacon.org/index.php/beacon_nation/package_details/enabling_health_information_exchange_to_support_community_goals

30 Bertolucci J. Privacy fears hit retailers' big data analytics plans. InformationWeek [serial on the Internet]. 2014 Jan 17 [cited 2014 Apr 28]. Available from: http://www.informationweek.com/big-data/big-data-analytics/privacy-fears-hit-retailers-big-data-analytics-plans/d/d-id/1113481

31 Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Rev. 2010;57;1701–77.

32 Tanner A. Harvard professor re-identifies anonymous volunteers in DNA study. Forbes [serial on the Internet]. 2013 Apr 25 [cited 2014 Apr 28]. Available from: http://www.forbes.com/sites/adamtanner/2013/04/25/harvard-professor-re-identifies-anonymous-volunteers-in-dna-study/

33 Davis K, with Patterson D. Ethics of big data. Sebastopol (CA): O'Reilly Media; 2012.

34 Lunshof JE, Chadwick R, Voraus DB, Church GM. From genetic privacy to open consent. Nat Rev Genet. 2008;9(5):406–11.

35 Tene O, Polonetsky J. Privacy in the age of big data: a time for big decisions. Stanford Law Review [serial on the Internet]. 2012 Feb 2 [cited 2014 Apr 28]. Available from: http://www.stanfordlawreview.org/online/privacy-paradox/big-data

36 Federal Trade Commission. Privacy online: fair information practices in the electronic marketplace: a report to Congress [Internet]. Washington (DC); FTC; 2000 May [cited 2014 Apr 28]. Available from: http://www.ftc.gov/sites/default/files/documents/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission-report/privacy2000.pdf

37 Organization for Economic Cooperation and Development. OECD work on privacy: the 2013 OECD privacy guidelines [Internet]. Paris: OECD; 2013 [cited 2014 Apr 28]. Available from: http://www.oecd.org/sti/ieconomy/privacy.htm

38 Fernandes L, O'Connor M, Weaver V. Big data, bigger outcomes: healthcare is embracing the big data movement, hoping to revolutionize HIM by distilling vast collection of data for specific analysis. J AHIMA. 2012;83(10):38–43; quiz 44.

39 Booz Allen Hamilton. The field guide to data science [Internet]. McLean (VA): Booz Allen Hamilton; 2013 [cited 2014 Apr 28]. Available from: http://www.boozallen.com/content/dam/boozallen/media/file/The-Field-Guide-to-Data-Science.pdf

40 Booz Allen Hamilton. Improving patient outcomes through advanced biomedical analytics: the Sisters of Mercy Health System [Internet]. McLean (VA): Booz Allen Hamilton; 2012 [cited 2014 Jun 16]. Available from: http://www.boozallen.com/content/dam/boozallen/media/file/leveraging-advanced-data-collection-cs.PDF

41 White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. J Am Med Inform Assoc. 2013;20(3):404–8.

42 Kitchenman A. Profile: Camden "hot-spotting" healthcare model captures national attention. NJ Spotlight [serial on the Internet]. 2013 Oct 9 [cited 2014 Apr 28]. Available from: http://www.njspotlight.com/stories/13/10/08/camden-hot-spotting-healthcare-model-captures-national-attention/