

# Acquisition, Analysis, and Sharing of Data in 2015 and Beyond: A Survey of the Landscape

## A Conference Report From the American Heart Association Data Summit 2015

Elliott M. Antman, MD, FAHA; Emelia J. Benjamin, MD, ScM, FAHA; Robert A. Harrington, MD, FAHA; Steven R. Houser, PhD, FAHA; Eric D. Peterson, MD, MPH, FAHA; Mary Ann Bauman, MD; Nancy Brown; Vincent Bufalino, MD, FAHA; Robert M. Califf, MD; Mark A. Creager, MD, FAHA; Alan Daugherty, PhD, DSc, FAHA; David L. Demets, PhD; Bernard P. Dennis; Shahram Ebadollahi, PhD; Mariell Jessup, MD, FAHA; Michael S. Lauer, MD, FAHA; Bernard Lo, MD; Calum A. MacRae, MD, PhD; Michael V. McConnell, MD, FAHA; Alexa T. McCray, PhD; Michelle M. Mello, JD, PhD; Eric Mueller; Jane W. Newburger, MD, MPH, FAHA; Sally Okun, RN, MMHS; Milton Packer, MD; Anthony Philippakis, MD, PhD; Peipei Ping, PhD, FAHA; Prad Prasoon; Véronique L. Roger, MD, MPH, FAHA; Steve Singer, PhD; Robert Temple, MD; Melanie B. Turner, MPH; Kevin Vigilante, MD, MPH; John Warner, MD; Patrick Wayte, MBA; on behalf of the American Heart Association Data Sharing Summit Attendees

**Background**—A 1.5-day interactive forum was convened to discuss critical issues in the acquisition, analysis, and sharing of data in the field of cardiovascular and stroke science. The discussion will serve as the foundation for the American Heart Association's (AHA's) near-term and future strategies in the Big Data area. The concepts evolving from this forum may also inform other fields of medicine and science.

**Methods and Results**—A total of 47 participants representing stakeholders from 7 domains (patients, basic scientists, clinical investigators, population researchers, clinicians and healthcare system administrators, industry, and regulatory authorities) participated in the conference. Presentation topics included updates on data as viewed from conventional medical and nonmedical sources, building and using Big Data repositories, articulation of the goals of data sharing, and principles of responsible data sharing. Facilitated breakout sessions were conducted to examine what each of the 7 stakeholder domains wants from Big Data under ideal circumstances and the possible roles that the AHA might play in meeting their needs. Important areas that are high priorities for further study regarding Big Data include a description of the methodology of how to acquire and analyze findings, validation of the veracity of discoveries from such research, and integration into investigative and clinical care aspects of future cardiovascular and stroke medicine. Potential roles that the AHA might consider include facilitating a standards discussion (eg, tools, methodology, and appropriate data use), providing education (eg, healthcare providers, patients, investigators), and helping build an interoperable digital ecosystem in cardiovascular and stroke science.

**Conclusion**—There was a consensus across stakeholder domains that Big Data holds great promise for revolutionizing the way cardiovascular and stroke research is conducted and clinical care is delivered; however, there is a clear need for the creation of a vision of how to use it to achieve the desired goals. Potential roles for the AHA center around facilitating a discussion of standards,

---

The American Heart Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

These Conference Proceedings were approved by the American Heart Association Science Advisory and Coordinating Committee on September 24, 2015. A copy of the document is available at <http://my.americanheart.org/statements> by selecting either the "By Topic" link or the "By Publication Date" link.

The Data Supplement is available with this article at <http://jaha.ahajournals.org/content/4/11/e002810/suppl/DC1>.

The American Heart Association requests that this document be cited as follows: Antman EM, Benjamin EJ, Harrington RA, Houser SR, Peterson ED, Bauman MA, Brown N, Bufalino V, Califf RM, Creager MA, Daugherty A, Demets DL, Dennis BP, Ebadollahi S, Jessup M, Lauer MS, Lo B, MacRae CA, McConnell MV, McCray AT, Mello MM, Mueller E, Newburger JW, Okun S, Packer M, Philippakis A, Ping P, Prasoon P, Roger VL, Singer S, Temple R, Turner MB, Vigilante K, Warner J, Wayte P; on behalf of the American Heart Association Data Sharing Summit Attendees. Acquisition, analysis, and sharing of data in 2015 and beyond: a survey of the landscape—a conference report from the American Heart Association Data Summit 2015. *J Am Heart Assoc*. 2015;4:e002810 doi: 10.1161/JAHA.115.002810. Expert peer review of AHA Scientific Statements is conducted by the AHA Office of Science Operations. For more on AHA statements and guidelines development, visit <http://my.americanheart.org/statements> and select the "Policies and Development" link.

**Correspondence to:** Elliott M. Antman, MD, FAHA, Associate Dean for Clinical/Translational Research, 75 Francis St., Boston, MA 02115. E-mail: [eantman@rics.bwh.harvard.edu](mailto:eantman@rics.bwh.harvard.edu)

© 2015 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley Blackwell. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

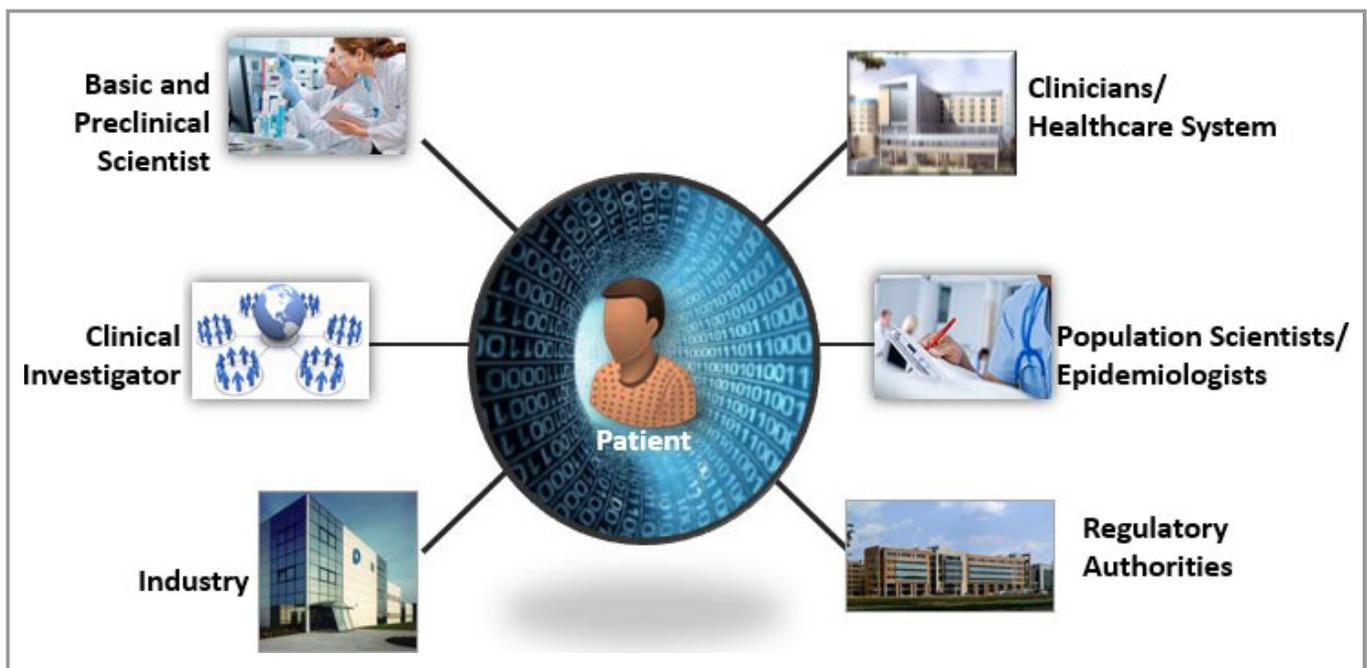
providing education, and helping establish a cardiovascular digital ecosystem. This ecosystem should be interoperable and needs to interface with the rapidly growing digital object environment of the modern-day healthcare system. (*J Am Heart Assoc.* 2015;4:e002810 doi: 10.1161/JAHA.115.002810)

**Key Words:** AHA Scientific Statements • clinical trials • data • epidemiology • ethics • mobile health • preclinical

The American Heart Association (AHA) is a global leader in the cardiovascular and stroke fields and has a long history of discovery, dissemination, and application of scientific knowledge. Recognizing the increasing importance of how modern approaches affect patients and the wide variety of persons working in the science and healthcare fields, the AHA convened a Data Summit on April 27–28, 2015, in Baltimore, Maryland. A central focus of the Data Summit was “Big Data,” which refers to large and complex data sets—including, for example in biomedicine, genomic, clinical, and environmental data—and entirely new approaches to data storage, management, integration, analysis, and visualization. A total of 47 participants representing stakeholders from 7 domains (patients, basic scientists, clinical investigators, population researchers, clinicians and healthcare system administrators, industry, and regulatory authorities) (Figure 1; Table 1) attended in person to debate and discuss critical issues in data management. Areas discussed included the acquisition, analysis, and sharing of data in the field of cardiovascular and stroke science. In addition to the writing

group members, a total of 28 persons (noted in the Acknowledgments) also provided input into the conference by providing strategic advice during the planning stage and/or commenting on drafts of this report.

The conference format included short presentations by groups of content experts followed by extensive panel discussions and audience interactions. Session topics on the first day included a survey of the Big Data landscape (traditional medical and nonmedical sources), a broad description of the current approaches to handling data and cutting-edge approaches using new tools and technologies, articulation of the goals of data sharing, and a description of the principles of responsible data sharing. An evening session focused on the AHA’s Cardiovascular Genome Phenome Study (CVGPS). The final day involved facilitated breakout sessions in which various stakeholders from the 7 domains developed scenarios of the future in response to 2 questions: (1) “What do stakeholders want from Big Data?” and (2) “What possible roles might the AHA play in each of these domains?” The attendees provided insights into additional topics related to



**Figure 1.** Patients were previously the passive source of data (ie, measurements were taken from patients by others), they are increasingly becoming active generators of data (eg, wireless sensors) and, in doing so, provide a vast new domain of data not previously available. In addition to the patient perspective, 6 other domains are shown. These lenses served as the organizing basis for the American Heart Association’s Data Summit.

**Table 1.** Stakeholder Domains Relevant to Acquisition, Analysis, and Sharing of Data in Cardiovascular and Stroke Science

Stakeholder	Interface With Data	
	Current	Future
Patients	<ol style="list-style-type: none"> <li>1. Passive source of data in healthcare environment</li> <li>2. Beneficiary of acquisition and analysis of data</li> </ol>	<ol style="list-style-type: none"> <li>1. Anticipated to assume progressively more active role in generating data (eg, quantified self)</li> <li>2. Participant in the acquisition and analysis of data (eg, PCORI)</li> <li>3. Beneficiary of acquisition and analysis of data</li> </ol>
Basic Scientists	<ol style="list-style-type: none"> <li>1. Focused on discovery science, usually in isolated units (ie, laboratories)</li> </ol>	<ol style="list-style-type: none"> <li>1. Sharing of data and experimental materials may facilitate replication of results beyond the original reporting laboratory</li> <li>2. Use of new methods for scientific discovery (eg, data mining, Big Data analytics)</li> </ol>
Clinical Investigators	<ol style="list-style-type: none"> <li>1. Focus on hypothesis-driven research</li> <li>2. Generally assume a common phenotype of trial participants for condition under study (eg, sample of larger universe of persons with the condition) but may analyze broad subgroup categories</li> <li>3. Case report form and trial database usually constructed for specific trial with limited or no plans for repeat use in future trials</li> </ol>	<ol style="list-style-type: none"> <li>1. Increasing use of Big Data approaches are anticipated to result in a shift from prespecified testable hypotheses to iteratively generated, data-driven hypotheses</li> <li>2. Increasing knowledge gained from deep genotyping and phenotyping will lead to more precise characterization of individual participant profiles and targeted treatments</li> <li>3. Trial documents and databases will be designed for repeated use</li> <li>4. Sharing of clinical trial data may facilitate planning of future trials (ie, avoid duplication of previous efforts) and permit validation of findings by external groups beyond original investigators</li> </ol>
Population Scientists/ Epidemiologists	<ol style="list-style-type: none"> <li>1. Each study collects data of interest, often without coordination across studies</li> <li>2. Participants interact at periodic intervals with the study</li> <li>3. Periodic surveillance of participants health encounters</li> </ol>	<ol style="list-style-type: none"> <li>1. Increasing coordination of “thick phenotyping” so that data can be harmonized, shared, and meta-analyzed across studies. Data collection protocols posted in real time at study Web sites</li> <li>2. Participants contribute data through mobile health technologies throughout course of year longitudinally</li> <li>3. Participants share healthcare encounter data in real time</li> </ol>
Clinicians/Healthcare System Researchers/ Administrators	<ol style="list-style-type: none"> <li>1. Navigating the role of the physician in the growing digital world</li> <li>2. Focus on balancing the validity of risk modeling from Big Data in making care decisions</li> </ol>	<ol style="list-style-type: none"> <li>1. Educating providers about the tools available to handle and find information</li> <li>2. Making the use of the data a natural part of the doctor–patient conversation</li> <li>3. Using data in areas in which there are physician shortages</li> </ol>
Industry	<ol style="list-style-type: none"> <li>1. Navigating the current fragmented environment that frequently results in cumbersome “linear” development pathways</li> <li>2. Limited integration of novel sources of data</li> <li>3. Increasing challenge of storing exponentially increasing amounts of data</li> </ol>	<ol style="list-style-type: none"> <li>1. Creating frameworks for collecting, storing, managing, curating, and analyzing data</li> <li>2. Navigating data that span various types of analytics including medical images, genomics, and natural language processing</li> <li>3. Capturing exogenous data to help inform clinical care</li> <li>4. Cloud computing to address capacity vs utilization issues</li> <li>5. New devices for monitoring patients with specific conditions</li> </ol>
Regulatory Authorities	<ol style="list-style-type: none"> <li>1. Current focus is on review of data from clinical trials usually conducted for a focused purpose</li> <li>2. Constrained by current system that limits ability to share and integrate data across multiple sources</li> </ol>	<ol style="list-style-type: none"> <li>1. Collecting data in standard formats for comparability and integration</li> <li>2. Ethical use of patient data, potential use of sensor data</li> <li>3. Greater use of data from public health sources</li> <li>4. Analyzing Big Data to accelerate understanding biological processes and heterogeneous response to treatments</li> <li>5. Policies and technical infrastructure to encourage data sharing</li> <li>6. Potential for modernizing the informed consent process</li> </ol>

PCORI indicates Patient-Centered Outcomes Research Institute.

Big Data that could be potential subjects of future conferences.

## Meeting Purpose

With the explosion of technological and computational advances, the amounts and types of data available to scientists have grown dramatically. Progress in many areas is expected to result from data sharing; however, data sharing needs to be approached with a clear understanding of the potential benefits and risks.<sup>1</sup> A thoughtful conversation is needed to provide the scientific community with a foundation from which to develop appropriate strategies for 2015 and beyond.<sup>2</sup> Although some efforts have been made in pooling and data sharing, what is needed now is a broader approach that will enhance the level of data and build on lessons learned in this and other fields to chart the course for cardiovascular disease (CVD) and stroke research and clinical care in the future.<sup>3-5</sup>

## Meeting Outcomes

The AHA Data Summit was convened to provide directional guidance to meet the needs of patients, researchers (basic, clinical, and population), clinicians, healthcare systems, industry, and regulatory authorities in the acquisition, analysis, and sharing of cardiovascular and stroke data. The structure of the summit was based on the planning committee's assessment of the field and was bolstered by new insights gathered during the conference. Areas of interest included a summary of the types of data to be considered, a broad description of how such data are acquired now and will be acquired in the future, methods for constructing data repositories, the goals of data sharing, and the principles of responsible data sharing. It is anticipated that the discussion at the conference will help formulate the AHA's approach and serve as the foundation for near-term and future AHA strategies in this area. The opinions and concepts evolving from this forum could also inform other fields of medicine and science.

## Data: Big and Small and Everything in Between, From Bench to Population

The presenters and attendees reviewed the landscape of data sources traditionally seen as relevant to and having an impact on decision making in cardiovascular and stroke science.

## Basic and Preclinical Data

Basic and preclinical cardiovascular data involve observations that are made from a small sample size but that hold great

potential from the perspective of informing and advancing the understanding of disease mechanisms to improve therapy.

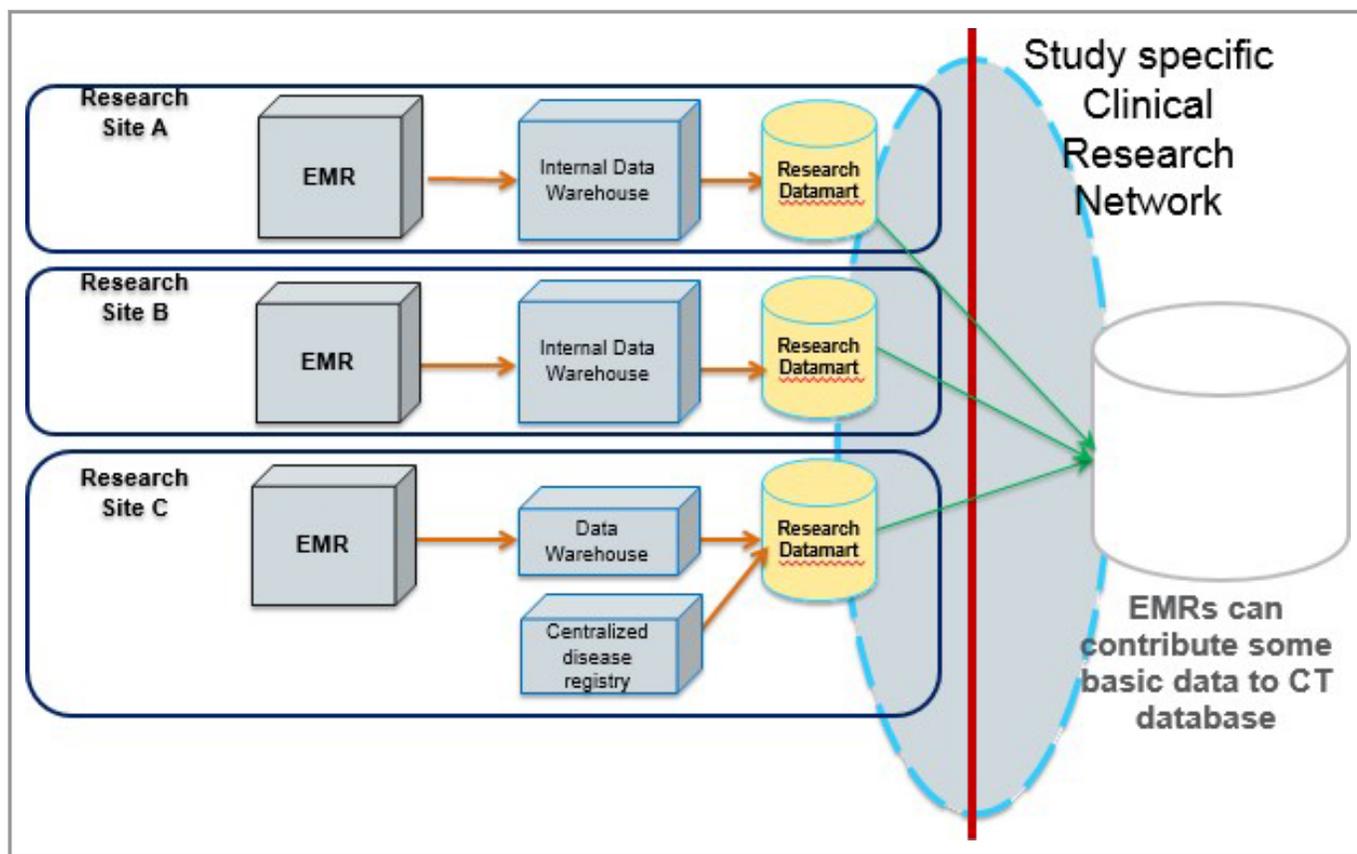
Basic cardiovascular research data sets are multidimensional in nature with a wide range of clinical and biomarker outcomes, for example, electrocardiogram, contractile function, molecular imaging, channel activities, genomics, proteomics, metabolomics, and phenotype characterizations. These data, however, are collected and presented in a variety of data formats, are variably distributed, and are both published and unpublished. The data are thus widely scattered and fragmented, making it difficult to extract knowledge either by individual laboratories or by organized scientific initiatives via team efforts. Consequently, there is limited ability to make well-informed decisions, in both bench and clinical settings; generally poor access to relevant information; and few opportunities to learn from and build on previous work.<sup>6</sup> Proper organization and management of data sets, rendering them accessible, complete, and analyzable, are important tasks for basic science investigators.

Revolutions in Big Data science have provided new digital technologies and informatics systems, and preclinical scientists can use them to address these challenges. These enabling platforms are designed to support integrated community efforts and are readily applicable in cardiovascular science. Big Data, by definition and concept, is rapidly evolving with respect to volume (bytes) but more so in terms of significance and relevance to scientific research.

## Patient Information and Clinical Care

The field of medicine is entering an unprecedented age of ubiquitous information.<sup>7</sup> Prior to the 1980s, clinical researchers were generally forced to review individual patients' paper charts to gather data; therefore, studies tended to be small, single-institution case series. During the 1990s, researchers began to tap into national medical claims data available from the large healthcare insurers including Medicare, Medicaid, the Veterans Health Administration, and private payers. These data sources provided very large patient sample sizes but were limited in their depth of detail, and the accuracy of the clinical information contained was suspect because the data were collected primarily for billing and not research.

With the enactment of the Patient Protection and Affordable Care Act of 2010, hospitals and clinics received a mandate to store their patients' clinical information in electronic medical records (EMRs). This digitization of patients' past histories and presenting complaints, treatments, and outcomes opened up a wealth of possibilities for clinical research (Figure 2). Although the initial vision of the EMR was full of promise, the lack of standardized data elements and definitions limited interoperability and presented challenges to widespread use by researchers. Over the



**Figure 2.** Evolving informatics for an EMR-based clinical research network. The model illustrates data transfers from individual-site EMR to storage in an internal data warehouse with data that can then be mapped to a research datamart (with standard data elements) and ultimately transferred to a CT database. CT indicates clinical trial; EMR, electronic medical records.

past few years, however, national standards have been promulgated, and EMRs are slowly mapping to these standards. In addition, large data infrastructure projects such as the National Institutes of Health (NIH) Collaboratory and the National Patient-Centered Clinical Research Network (PCORnet) (Figure 3) facilitated linking of EMR data across multiple large health systems, thus unlocking medical information on millions of patients to medical researchers. Such data are now being used to carry out large-scale postmarket surveillance studies and even to help recruit patients and collect information in practical clinical trials. The EMR is now being used increasingly not only to generate new evidence but also to incorporate quality improvement systems into the flow of clinical care. Within the past 30 years, the research community has seen remarkable growth and maturation of “big” clinical data resources. These resources offer the potential to allow clinical researchers to turn data into knowledge and drive knowledge into routine clinical practice.<sup>3</sup>

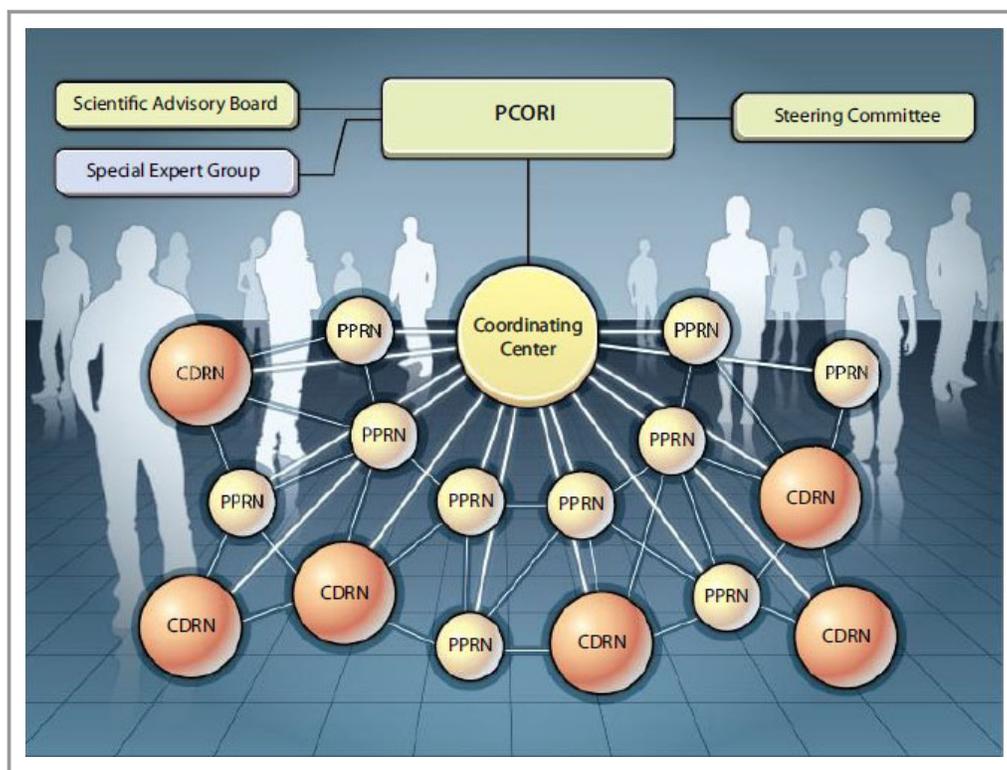
### “My Personal Health Information”

Patient-reported health information has long been part of the healthcare encounter (eg, chief complaint, quoted subjective

statements); however, by their very nature, these data are isolated within individual medical records. In our increasingly connected digital world, the ecosystem of “my” personal health information is ever changing and pushing the boundaries of its place alongside clinical data and within Big Data.

The myriad of opportunities to collect person-generated health data (PGHD) is expanding rapidly.<sup>9</sup> A person can track his or her health and fitness using wearable devices (eg, activity trackers). Data can be contributed proactively by participation in online patient communities and registries (eg, PCORnet). Sensors can be used to monitor activity level, daily weight, or other relevant markers of health (eg, smartphone “apps”). Patients can engage with health providers via telemedicine, email, or other electronic means and can even participate in clinical research via smartphones (eg, Apple ResearchKit, Health eHeart Study).

PGHD, when aggregated with traditional sources of health information, creates a complex tapestry of daily lives and healthcare experiences, with the potential to generate insights and new knowledge about living with and managing health and illness.<sup>1</sup> Making inferences regarding the population, however, is uncertain due to the nonrandom nature of



**Figure 3.** PCORnet: clinical research and patient engagement on a large scale. The proposed organization of the PCORnet is shown. Supported by a coordinating center, CDRNs, and PPRNs, a sustainable network of healthcare centers will be created with interoperable electronic medical records and active patient participation, all overseen by the PCORI staff, board of governors, and advisors. CDRNs include 8 networks with 1 million patients per network. PPRNs include 18 networks. Reprinted from Selby et al.<sup>8</sup> CDRNs indicates Clinical Data Research Networks; PCORI, Patient-Centered Outcomes Research Institute; PCORnet, National Patient-Centered Clinical Research Network; PPRNs, Patient Powered Research Networks.

self-selection in the use of these devices and online reporting portals.

Studies show that US patients, especially those living with chronic illness, are willing to share their personal information for the sake of improving care and treatment options for themselves and for others like them so long as appropriate protections are in place.<sup>10,11</sup> As people continue to participate in generating personal health information, these data are beginning to take on attributes commonly ascribed to Big Data: volume, velocity, and variety.<sup>12</sup>

The science of understanding PGHD, although continuing to emerge, remains nascent. The real-world PGHD needs to be subjected to standard analytical methods, and new methodologies are likely needed to effectively mine the data for insights to answer many questions that can now be asked when the various sources of data are woven together.

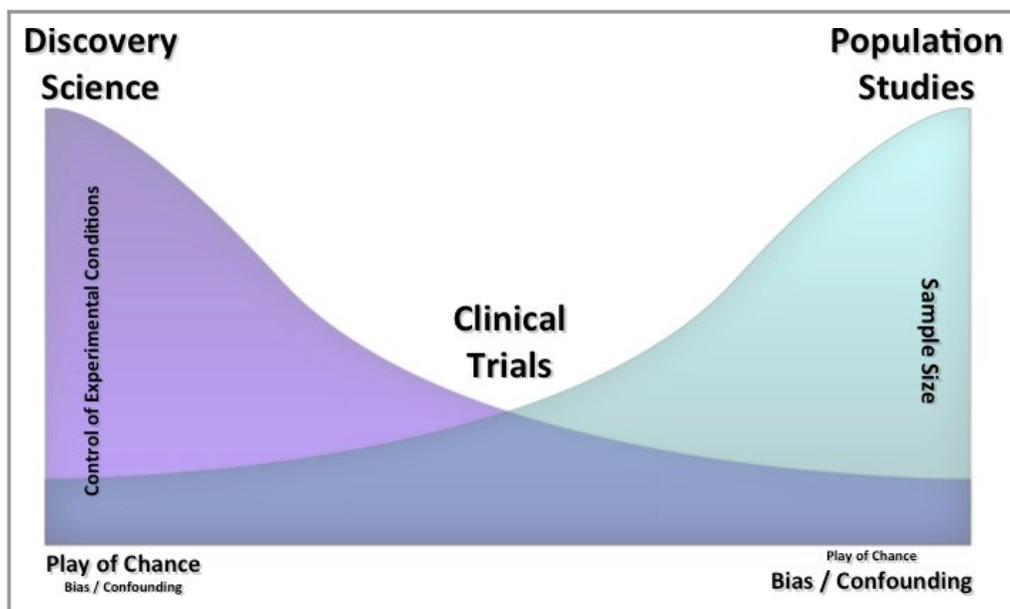
Collecting data from any source is one thing, but truly understanding what the data are saying is another. The proverbial needle just gets harder to find in the haystack, which keeps getting bigger and bigger.<sup>13</sup> Nevertheless, unlike PGHD, Big Data lacks context—the key holistic and interpre-

tive lens through which data are filtered and turned into real information. Further study will be needed to understand how personal health data can be optimally used to enrich what we know today as Big Data. With further study and understanding, it may be that personal health data are the key ingredient that is currently missing from Big Data.<sup>14</sup>

### Clinical Trial Data

The majority of clinical trials in cardiovascular medicine to date have been designed to assess the efficacy and safety of therapies administered to (eg, drug, biologic) or inserted into (eg, device) patients. Some clinical trials evaluate biomarker assays and imaging procedures, largely to assess whether these measurement tools can help determine whether a patient warrants treatment with a noted therapy or to assess the patient response to such treatments.

Three major potential threats to validity of research findings are “play of chance,” bias, and confounding (Figure 4). Play of chance is the predominant concern in discovery science research, with considerable control over the



**Figure 4.** Major potential threats to the validity of research findings. Discovery scientists work in an environment in which they have a high degree of control over the experimental conditions and use a small sample size. Population scientists operate in an environment in which there is less control over experimental conditions but a large sample size. Clinical trials fall between these 2 extremes and need to be interpreted with respect to internal validity and external generalizability. Other major threats to validity are shown (bottom left and right); the size of the font graphically illustrates the relative importance of the threats.

experimental conditions but a small sample size. Bias and confounding predominate in population studies, with loss of control over the experimental conditions but large sample sizes. Clinical trials fall between these 2 extremes, and investigators attempt to deal with threats to validity by minimizing type I and II errors during the design phase and using randomization during the implementation phase.

The following important developments currently influence clinical trials with increasing frequency:

1. Given the number of therapeutic options available to treat cardiovascular disorders, investigators are now testing many new agents against an active control and use clinical trials with noninferiority designs.<sup>15</sup>
2. Because emerging data either external to or observed from within trial can influence an ongoing trial, investigators may need to modify or adapt trials prior to their completion.<sup>16,17</sup>

Despite the rigor surrounding the clinical trial enterprise and the impact that clinical trials have on clinical practice guidelines, it is overly simplistic to frame the study hypothesis or interpret the results in terms of a treatment effect in a “population.” The cohort of patients enrolled in a trial may or may not be representative of the greater universe of patients who have the disease state of interest.<sup>4</sup> We typically operate under the assumption of a common phenotype in designing and analyzing clinical trials. Currently, we lack sufficient understanding of the genotypic and phenotypic aspects of

disease to adequately characterize the response to treatments in heterogeneous populations.

Sharing of clinical trial data has the potential to improve the design and efficiency of clinical trials and to enhance our understanding of the complexities of response to treatments. It is important to establish the principles of responsible sharing of clinical trial data and to describe the circumstances under which it is useful to move beyond sharing of the derived data sets and to provide individual participant-level data more broadly.<sup>18</sup>

### Observational and Epidemiological Data

More than a dozen governmental surveys provide snapshots of the health, risk factors, disease, and healthcare utilization status of American citizens. These surveys inform the public and policy makers, and they are incorporated into the AHA’s “Heart Disease and Stroke Statistics.”<sup>19</sup> The surveys include a variety of ascertainment strategies, such as phone surveys, in-person interviews, physical examinations, and surveillance of healthcare providers and institutions.

Complementary sources of data are epidemiological cohort studies funded by the NIH and foundations. The epidemiological cohorts track participants longitudinally, which enables study of the social, behavioral, biomarker, and genomic risk factors for subclinical and clinical CVD and stroke. The studies routinely collect interviews, physical examinations, subclinical

markers, biosamples, and outcomes that have been used to identify risk prediction algorithms, temporal trends, and healthcare disparities in CVD and other diseases. The oldest cohort study, the Framingham Heart Study, was founded in 1948, and the most recent study, The Hispanic Community Health Study/Study of Latinos, was initiated in 2008. There is variation in the number of sites, age at participant enrollment, ethnic and racial composition, and overall study focus. More recently, the epidemiological studies have been challenged to strategically transform to become more resource effective and to embrace new data-collection strategies. Suggestions include development of cross-cohort collaborations, electronic surveillance, and more accessible repositories for phenotype and genomic data and biospecimens; collaborations with clinical trialists also have been urged.<sup>20–23</sup>

## Big Data From the Real World

The presenters and attendees at the AHA Data Summit surveyed the landscape of data sources that are not traditionally considered by cardiovascular scientists but that emerging evidence suggests may inform or impact health. The attendees discussed efforts in the Big Data field and how they may inform or impact the study and management of cardiovascular disorders and medicine more generally.

## Computational Health Care in the Era of Big Data

From the viewpoint of health outcome determinants, almost 60% of data are exogenous (eg, behavioral, socioeconomic, environmental) and are rarely captured as part of EMR systems.<sup>24</sup> Inserting such data in the data flow and enabling the generation and/or capture of this exogenous data is crucial for emerging health ecosystems. Important aspects of these data are volume, velocity, and variety—the traditional Big Data characteristics. Another aspect is that all data are generated in uncontrolled environments (ie, no hospital or supply-side control), which create highly fragmented value chains that need a neutral entity that can collect, store, manage, curate, and analyze data for insights.

Computational health care is driven by a comprehensive set of technologies used to address the data deluge in medicine. Useful data span various types of analytics including medical images, genomics, and natural language processing. To implement behavior modification in clinical care, it will be important to study the biometrics, medication usage patterns, stress levels, sleep patterns, and social interactions of individual patients. Opportunities to improve disease management and treatment may exist through context-aware data acquisition, medication/dosage and comorbidity management, and patient education and engage-

ment.<sup>14</sup> In addition, behavior change and prevention can be addressed by using behavior models to develop recommendation services and by understanding habit-formation cycles to design new service models, incentives, and touch-point modifications.

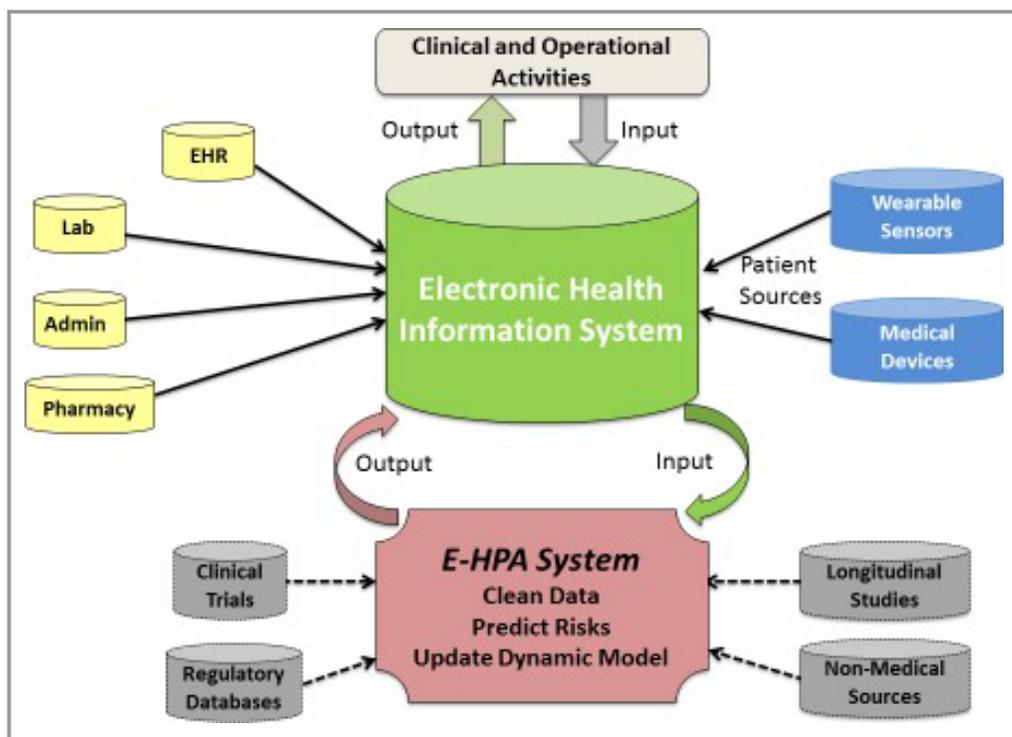
Rather than using data in traditional isolated analyses, a hybrid model of evaluating systems of insight, systems of record, and systems of engagement in a cloud environment may create a new future of health care.<sup>1,25</sup> All data types can be measured and analyzed to provide new decision-making models that allow providers to intervene at the right place and the right time for the most positive patient outcomes.

## Genomic Data and Digital Health

Advances in genetics, genomics, and proteomics over the past 20 years have catalyzed the capacity to address their experimental, translational, and clinical implications, as applied to cardiovascular health and disease.<sup>26–28</sup> The growth of these fields has been so exponential that the genomics community currently faces 4 computational barriers to transforming raw sequencing data into biomedical insights:

1. Processing massive sequence data sets requires costly computational infrastructure.<sup>29</sup> Few groups have the resources to meet this challenge, and those that do often end up duplicating each other's work.
2. The current generation of methods cannot scale to the petabytes of data already in existence, let alone the exabytes that will come.
3. Data are being collected and stored in silos, setting us on a trajectory toward a fragmented system analogous to what has been developed for EMRs.
4. Data copying is a prerequisite for data sharing, forcing the greater genomics community to shoulder the cost of storing multiple copies of massive data sets.

Cloud computing offers a potential solution to all of these challenges.<sup>30</sup> A frequent problem in planning technology infrastructure is capacity versus demand and the underlying expense. When planning data infrastructure requirements, organizations are forced to use peak load criteria when sizing their resource needs. Consequently, resources need to be provided for the highest possible load situation. On-site infrastructure to meet rising capacity needs is capital-intensive and can leave an organization either with excess capacity that is underutilized or with an excess in demand leading to service degradation. Cloud computing can more closely follow actual demand, and be scaled up and down readily along with actual demand.<sup>31</sup> Few areas are as well suited to cloud computing as genomics.<sup>29</sup> First, cloud computing allows the community of researchers to access a



**Figure 5.** The future of health information. New sources of patient data (in blue) are beginning to be merged with traditional healthcare data sources (in yellow) to better inform clinicians’ diagnosis, treatment and care decisions. The health care of the future can build on this model by incorporating additional existing data sources (in gray) to create an electronic healthcare predictive analytics systems (E-HPA) that could theoretically use data from any source to clean and analyze, run and/or update predictive models, and output risk estimates back into the health information system to trigger or monitor specific clinical and/or operational activity. Adapted from Amarasingham et al.<sup>32</sup> Admin indicates administration; EHR, electronic health record; Lab, laboratory.

shared pool of data in an environment equipped with extensive and elastic computing resources and a sophisticated model for access control. The cloud facilitates secure sharing of data at both technical and economic levels. Second, cloud computing allows researchers to rent a data center under a pay-as-you-go model, removing the significant capital expenditures associated with building a data center and staffing it with personnel. Finally, cloud computing is about not only a physical infrastructure for computers but also a paradigm for writing algorithms to enable massive parallelization, allowing for scalable on-demand “supercomputers.” Because genomic computations are easily parallelized by genomic locus, they are ideally suited to the cloud.

As we look to the future of genetic research in cardiology, it is important to consider the role of cloud computing as a powerful lever for advancing research.

### The Healthcare Environment of the Future

With the evolution of cloud technology, proactive medical treatments and protocols are now within reach of the

healthcare ecosystem at large. Data from the point of care and from devices and “wearables” driven by the “Internet of Things,” combined with environmental data (eg, social, financial), provide the bases of actionable advances in care delivery (Figure 5). Insights are gained by coupling Big Data with analytics and machine learning to create the foundation for a cloud-based interoperable ecosystem (Figure 5). Technology and treatment elasticity that can seamlessly follow the patient through the transitions of care are key to operationalizing the benefits of Big Data (Figure 5). By overlaying cloud Big Data foundations with evidence-based medicine, treatment advances can be shared quickly and globally. Technology should ultimately reduce the friction at the point of care by seamlessly integrating the healthcare consumer with the healthcare enterprise and allowing for a smooth interface with the healthcare ecosystem.

With the increased focus on EMR incentive programs in 2009 in the US healthcare system, the federal government put higher pressures on healthcare providers to leverage digital solutions for sharing of records and management of care.<sup>33</sup> With advances in technology and innovative

solutions, the role of the healthcare provider will change. We are on a verge of an explosion of a new era of mobile health and biosensor technologies that will make it easier to collect, track, and share real-time data and metrics among clinicians without extra processes or paperwork and that likely will be key patient engagement tools, offering alternatives to traditional in-person visits (Figure 5).<sup>34</sup> These new interactions will become richer and more prevalent as patients have greater access to highly sophisticated healthcare apps and diagnostic tools outside the traditional medical environment that allow them to take a more active role of the management of their care.

Because patients can be seen and treated faster, healthcare providers may be removed from certain low-risk routine clinical decisions, allowing more time to concentrate on higher risk aspects of a patient's care.<sup>32</sup> In the future healthcare ecosystem, it will be important to maintain critical healthcare provider engagement in decision making about individual patient care.

## Using Big Data

The presenters and attendees reviewed current approaches to handling the management of Big Data, which requires approaches that are different from classical approaches to data acquisition and analysis in cardiovascular science. Speakers discussed their experiences with acquiring large-scale data and data sets, the new methods that were required for managing and disseminating the data, and the policies that needed to be in place to ensure responsible use of the data. The speakers emphasized the importance of data sharing while acknowledging some of the inherent challenges.

## Building Big Data Repositories

The National Research Council's visionary report on precision medicine<sup>35-37</sup>; the commitment of the NIH, through its National Library of Medicine, to data science<sup>38</sup>; and the newly announced Precision Medicine Initiative<sup>39</sup> on creating a voluntary national research cohort of >1 million persons highlight the arrival of the era of Big Data in biomedicine. It will no longer be possible to view the collection and analysis of biomedical data as an activity that is conducted solely by a single investigator in a single laboratory for a single purpose.<sup>35,38,39</sup> New types of data, such as genomic and environmental data, which are themselves Big Data, are being correlated with current and longitudinal medical data drawn from EMRs. Big Data has the potential to lead to important discoveries, for example, diagnosing rare and undiagnosed diseases, discovering the etiology of highly heritable but poorly

understood conditions, and transforming the entire clinical trial enterprise.<sup>40-44</sup>

To fully effect this transformation, a number of technical, social, and policy changes will need to be addressed.<sup>1</sup> Data will need to be collected in standard formats using best practices so that systems that contain relevant data are technically interoperable, allowing for the comparability and integration of those data.<sup>45</sup> Research participants will need to be fully informed that their data will be shared for research purposes, and they will need to consent to that use. Researchers will need to ensure that the enormous benefits of data sharing outweigh the potential risks.<sup>44,46-48</sup> Policies that encourage or even mandate data sharing are beginning to be put in place,<sup>49,50</sup> and sharing data with research participants themselves has become an active area of research and discussion.<sup>51,52</sup>

## The dbGaP Database and BioLINCC

The National Heart, Lung, and Blood Institute (NHLBI), along with other institutes and centers at the NIH, has developed and supported platforms for sharing data generated by NIH-supported trials and observational studies. Examples include the NHLBI's Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC)<sup>53</sup> and the NIH National Center for Bioinformatics (NCBI) database of Genotypes and Phenotypes (dbGaP).<sup>54</sup>

The NHLBI's BioLINCC houses deidentified individual-level data from 88 clinical trials and 34 observational studies that altogether enrolled >600 000 separate participants. These data are available to qualified researchers, but a number of important constraints and caveats are designed to protect the interests of research participants, researchers, and the NHLBI. Data are deposited into the repository after study completion, allowing study investigators a 2-year time window to conduct analyses and publish their findings. The deposited data are deidentified and, at times, modified to ensure maximum protection of participant identity. Applicants who wish to use BioLINCC data or biological specimens must submit information about their analysis plans and must provide, through a data use agreement, assurance that they will not transfer data to third parties or conduct additional analyses without first contacting the NHLBI.

BioLINCC has been increasingly active since its inception in 2000. It now receives >300 data requests per year, about half for clinical trial data. Publication activity has also increased: In 2014, researchers published >80 papers based on BioLINCC downloads.

The NCBI's dbGaP database houses genomic and phenotypic data generated by >550 studies, with data derived from hundreds of thousands of samples. Similar to BioLINCC, investigators who wish to use dbGaP data must submit a brief formal application and must provide assurances regarding

research participants' consent and privacy. An analysis of dbGaP activity found increasing levels of interest. To date, dbGaP has provided data to >2000 investigators who have used the data to generate 924 publications.<sup>55</sup> As -omic and Big Data technologies evolve, the NIH is continuing to extend its genomic data-sharing policy to multiple types of data of interest to biomedical researchers.

### Integrating Sensor Data With Clinical Data

A major new source of Big Data is the emergence of wearable sensors and connected devices, which enable continuous health data acquisition (Figure 6). Although technology has been the initial driver of these data, particularly on the consumer side, there are powerful reasons to integrate sensor data with more traditional clinical and phenotypic data within clinical and population health studies. There are also important limitations for this form of PGHD, which requires careful consideration and further research.

The most powerful reason to incorporate sensor data is that a substantial portion of CVD and stroke events—across primordial, primary, and secondary prevention—are attributable to daily behaviors and modifiable risk factors, many of which largely go unmeasured. These behaviors are highlighted by the AHA's "Life's Simple 7" program, through which physical activity, food choices, cigarette smoking, blood pressure, weight, and glucose levels are all "trackable" on a daily or more frequent basis through wearable sensors, smartphone apps, and connected devices.<sup>57</sup> Real-world data should have advantages over, or at least complement, the

recall-based surveys and intermittent clinical testing in traditional studies.

Quantitative sensor data have already been incorporated in subgroups of several large population health studies, such as the use of wrist-based accelerometers in the Centers for Disease Control and Prevention National Health and Nutrition Examination Survey (NHANES)<sup>58</sup> and the UK Biobank.<sup>59</sup> The AHA-endorsed Health eHeart Study (a PCORnet Patient Powered Research Network) by the University of California San Francisco has developed an infrastructure to collect novel data from Web- and smartphone-based surveys, apps, sensors (eg electrocardiogram, physical activity, blood pressure, social networking, wearables), EMR data, and outcome measures to create a large, well-phenotyped cohort of participants from around the world who volunteer to provide data that are within a central academic institution and agnostic to device or sensor type.<sup>60</sup> Apple's release of open-source ResearchKit enables smartphone-based medical research, including the recent MyHeart Counts study by Stanford University in collaboration with the AHA,<sup>61</sup> which uses the phone's sensors to collect physical activity and fitness data plus wearable and connected health data linked through HealthKit. Extensive "physiome" data through wearable sensors are planned for a Baseline Study coordinated by Stanford, Duke University, and Google Inc,<sup>62</sup> with mobile health data also planned for the NIH's Precision Medicine Initiative cohort of 1 million US participants.<sup>35,36,63,64</sup>

Nevertheless, there are limitations to incorporating sensor data for clinical and population research. Few measurements from wearable sensors have been validated relative to existing metrics. For physical activity, some studies have validating



**Figure 6.** Vision for clinical research in the future. Examples of wearable sensors (top left). These communicate via Bluetooth to smartphones, which transmit data wirelessly to a research-grade database on the Internet. Mobile health enables enrollment of large numbers of diverse participants around the globe. Investigators can then more effectively study transitions from a state of ideal health to the development of risk factors and ultimately overt manifestations of disease. Reprinted from Antman.<sup>56</sup>

smartphones and/or consumer wearables (eg, step count,<sup>65</sup> 6-minute walk distance<sup>66</sup>), but reliable measures of moderate and vigorous activity—the basis of AHA and World Health Organization guidelines—may be more challenging. Mobile and connected health devices are also generating continuous ambulatory data that do not directly match the tests done in the clinic or laboratory. This warrants careful attention to having a common data schema for mobile health data and including relevant metadata (eg, device; software version; time; location; before, during, and after exercise), an effort spearheaded by the Open mHealth project.<sup>67</sup> These mobile health data are also typically not observed by study personnel, so data quality can be dependent on individual participants and their level of engagement. As is often the case for Big Data, researchers may be accepting trade-offs in precision for more frequent, scalable measures; in some cases, the more frequent data can be processed to account for small inaccuracies, but large inaccuracies may simply provide many inaccurate data points. As studies aim to reach larger populations, enabled by Web- and smartphone-based studies, there can be a selection bias from the participants who “opt in” and who have sufficient technological knowledge and access. Finally, mobile or connected health data, by their very nature, do not originate within the research site, so methods to ensure privacy and security of the data are critical, particularly because these data are shared and linked with other clinical data.

## Goals of Data Sharing

Many research funding organizations, including the AHA, have data sharing policies<sup>68,69</sup> that address the following common themes:

- Funding versus cost effectiveness of the data storage and subsequent management processes
- Types of data to be shared based on organizational goals and the impact of discovery versus the cost of sharing
- Timeliness of sharing data in relation to the end of a study
- Protection of intellectual property in relation to timeliness of data sharing
- Ethical use of data in terms of continued human subjects protection in data sharing and secondary analysis
- Required acknowledgement of the original data source in secondary analyses
- Length of time that data should be stored and used for secondary analysis
- Oversight and monitoring process of secondary analyses and resulting publications

It was agreed that when considering sharing data, it is important (1) to articulate the purpose of sharing the data and (2) to maximize the benefits while minimizing potential risks.

**Table 2.** Goals of Data Sharing

Facilitating discovery science: avoiding duplication, ensuring reproducibility
Increasing understanding of human disease
Improving the design, efficiency, and quality of clinical trials
Improving the quality of care in clinical settings
Increasing the effectiveness of prevention
Translation to public

Although much of the discussion to date has centered on research and data scientists, clinical investigators, and healthcare professionals, more emphasis needs to be placed on increasing public awareness of efforts to share personal information from healthcare encounters, biological specimens, and clinical trial participation (Table 2).

## Facilitating Discovery Science: Avoiding Duplication and Ensuring Reproducibility

Preclinical research has been the traditional venue for the identification and validation of molecular targets that have the potential to affect CVD<sup>70</sup>; however, there is growing concern that preclinical studies are not easily replicated beyond the reporting laboratory.<sup>6</sup> This has been assessed objectively by 2 reports that evaluated replication of studies prior to formalizing a pharmaceutical development program. Discouragingly, these studies demonstrated that only 11% to 22% of the preclinical studies could be replicated.<sup>71,72</sup> As a consequence of the concern about reproducibility, there is an NIH initiative to modify data acquisition and sharing in the preclinical arena. Much of this initiative focuses on enhancements of the quality of data acquired in preclinical studies. In this regard, the initiative details many parameters that are not commonly considered in preclinical studies including randomization, blinding, sample size estimations, and inclusion and exclusion criteria. The recommendations also call for full descriptions of statistical analyses. A pivotal part of the NIH recommendations is to enhance data and material sharing. This includes the need to make all primary data available during the manuscript review process. The plan also notes that all data should be available immediately on publication. Furthermore, data sets should be represented in a manner that facilitates reanalysis of data and incorporation into larger data sets. Although the recommendations have been endorsed by many leading journals, several unresolved impediments to implementation remain. These include standardization of data formats and funding sources that will support the central data repository. A confounding issue of shared data

is the potential bias due to more ready acceptance by journals of manuscripts describing positive data from preclinical studies. Balancing this bias requires development of a mechanism for acknowledging the importance of negative or neutral data.

## Increasing Our Understanding of Human Disease

In the clinical environment, robust sharing practices will be vital to realize the somewhat paradoxical goals of global data collection and personalization of care. The immense challenges inherent in maintaining this creative tension mandate transformative changes in the scale of each step in the translational cycle and the development of truly generalizable rules for biomedicine. Next-generation technologies have revealed unanticipated complexity in the genome of each person and have placed in stark relief the archaic redundancy of most clinical phenotypes.<sup>73</sup> We are compelled to reinvent the scope and scale of clinical phenotyping if we are to be able to deconvolute even the first-order information in a human genome.<sup>27,28</sup> Data sharing will bridge this “phenotypic” gap through the definition of new phenotypes, massive changes in the dimensionality of clinical assessment, and the consequent parsing of aggregate syndromes such as atherosclerosis and diabetes into their constituent etiologies.

Data sharing will be vital not only for traditional clinical data sets but also for diverse traits that are likely to be related in their basic characteristics: computability, linear dynamic range, and orthogonality to the phenotypes of the past. Meaningful data integration can be augmented by the collection of relevant metadata in the form of stimulus-response pairs, in which the organizing stimuli may be physiological, pharmacological, or environmental inputs. Carefully chosen stimuli will also facilitate vertical integration across species and from cells to organismal biology. In care delivery, the new incentives to share will align patients across the economic spectrum because, for many traits, the most informative patients are often those at the extremes. Finally, although wearable and other personal technologies will continue to proliferate, the need for rigorous and biologically relevant calibration will drive outpatient clinics to become a preferred venue for multidimensional data collection, in the process establishing a novel minimal clinical data set for the 21st century.

## Improving the Design and Efficiency of Clinical Trials

The clinical trials enterprise is vast and expensive. For a variety of historical reasons, despite the critical societal importance of clinical trials, most clinical trial data have not been shared broadly. The reasons include concern about maintaining the confidentiality and privacy of trial partici-

pants, protection of the intellectual property of companies and inventors, and academic control of ideas and results. As transparency efforts gain traction, it is timely to consider the potential benefits of sharing trial data to improve the clinical trials enterprise itself.

Clinical trials can be fundamentally divided into (1) mechanistic trials, which are intended to explore and understand the mechanisms governing human biology, and (2) pragmatic trials, intended to inform healthcare decision makers.<sup>74</sup> Because of the complexity of the types of questions in each of these categories, there is no single best approach to clinical trials. Instead, the principles of quality lead to efforts to optimize the value and efficiency of clinical trials by designing and operationalizing the trial based on the purpose of the trial.

ClinicalTrials.gov has provided a new ability to examine the clinical trials enterprise because the vast majority of trials have a legal requirement for registration prior to enrollment and results reporting within 1 year of ascertainment of the last primary end point.<sup>50</sup> Currently, >380 trials per week are registered, and initial analysis of the accumulating data reveals a large number of trials with major deficiencies that would make it unlikely that they are answering an important question. We now have the opportunity to treat the clinical trials enterprise as a learning ecosystem in which all previous trials can inform the planning and operationalization of the next trial.<sup>75</sup>

Several broad questions could be addressed:

- What types of trial designs are best for answering different types of questions?
- What are the characteristics of trials that answer the question asked by the trial or that fail to answer the question either because of design or conduct specifications?
- Over time, what are the characteristics of trials that answer key clinical questions?

More specific questions could also be addressed:

- What entry criteria for particular diseases or issues lead to the best recruitment?
- What biomarkers and putative surrogates actually work in the clinical trials setting?
- What are the clinical outcomes and natural history of particular disease states, and what event rates can be expected for given entry criteria?
- What operational approaches lead to the most efficient trial conduct?
- What types of data collection and which data items are most useful for different types of trials in different circumstances?
- What analytical methods lead to conclusions that are reliable over time and across different types of trials?

## Improving the Quality of Care in Clinical Settings

The AHA's work in quality improvement began in earnest in 2000 with the "Get With The Guidelines" program, which was implemented in the United States and in several other countries.<sup>76</sup> It has subsequently expanded beyond coronary artery disease to address stroke, atrial fibrillation, heart failure, and resuscitation, with >2000 hospitals and 6 million patients in the database. The most recent iteration is the outpatient clinical registry called The Guideline Advantage (TGA). This triagency program with the American Cancer Society and the American Diabetes Association provides feedback to clinicians regarding their compliance with performance measure-based care.<sup>77</sup> This program already has nearly 100 practices and >6 million medical records. It is vendor neutral so as to provide seamless data sharing to individual clinicians and practices and comparisons to a national data cohort. TGA provides performance tools to aid the providers in managing their patient care and to help with population health strategies. This database should help identify better measures and answer critical questions in the clinical research agenda.

Clinical quality improvement is an important initiative for the AHA to help the National Care Delivery System develop new approaches to better quality of care and provide tools to help clinicians manage care.<sup>78,79</sup>

## Increasing the Effectiveness of Prevention

The occurrence of CVD and of many chronic diseases is deeply shaped by behavior, health choices, and the context in which we live (ie, determinants of health outcomes<sup>80</sup>); these factors are responsible for a large proportion of CVD in the United States.<sup>19</sup> To understand how best to promote choices that are optimally conducive for health and disease management, it is important to recognize that patients with chronic diseases spend 3 to 5 hours per year with a care provider and 5000 waking hours elsewhere. Care providers can hardly influence choices that are critical determinants of disease occurrence and health and that are largely outside the realm of health care. Positive health choices are particularly challenging because people must confront, multiple times a day, powerful societal forces that have largely engineered physical activity out of everyday life and hinder healthy food selections.

Mobile health technology has the potential to deal with one of the most pressing problems in prevention: the failure of patients to comply with a prescribed regimen.<sup>81</sup> By incorporating simple reminder messages and alerts in the increasingly digital healthcare ecosystem, patients and healthcare providers are more likely to fully realize the benefits of already proven therapies when taken as prescribed. Mobile health devices and wireless technologies can provide a substantial

amount of "hovering" and appear conceptually well suited to facilitate healthy behaviors.<sup>81,82</sup> These tools generate high volumes of data that must be validated, analyzed, and interpreted, and methods such as predictive analytics and behavioral economics can be exploited to do so.<sup>83</sup>

At this juncture, the exact role of mobile health tools remains to be fully delineated, and key questions remain to be addressed. First, technology is necessary but not sufficient to induce health choices, and adherence to the use of mobile health technology is unclear. Second, new health-based high-volume data (Big Data) must be integrated into research and clinical practice approaches to prevention, and the impact on outcomes must be determined. Finally and importantly, the applicability and adoption of new approaches across diverse populations must be studied and understood. From the perspective of prevention, it is important to "reach people when they are not patients."

## Translation to the Public

The goals of data sharing must also include the patient and public perspectives. A primary concern of patients is that the confidentiality of their health information may be compromised. Concerns about confidentiality have a basis in reality. By statute,<sup>84</sup> the US Department of Health and Human Services Office for Civil Rights posts health data breaches that affect at least 500 patients. From January to May 1, 2015, there were 93 data breaches posted involving 92 350 555 patients. Since the onset of reporting in October 2009, there have been 1220 reports involving 133 257 322 patients from all 50 states. In addition to data breaches, deidentified health data potentially can be aggregated and identified by analyzing online forums, governmental records, and commercially available databases.<sup>85</sup>

Another public health concern is that Big Data may further exacerbate disparities in health outcomes. There is justified enthusiasm for harnessing Big Data from cell phones, geospatial location, and biological real-time monitoring of health conditions to improve health and disease management; however, access to smart phones and health literacy are unevenly distributed by age, race, socioeconomic status, and rurality.<sup>86,87</sup>

To address patient and public health concerns, we must pay attention to maintaining privacy and enhancing access. Failure to address the digital divide and to ensure privacy may enhance public distrust and exacerbate healthcare inequities.

## Principles of Responsible Data Sharing

The presenters and attendees discussed a wide variety of issues related to data sharing. The perspectives of multiple

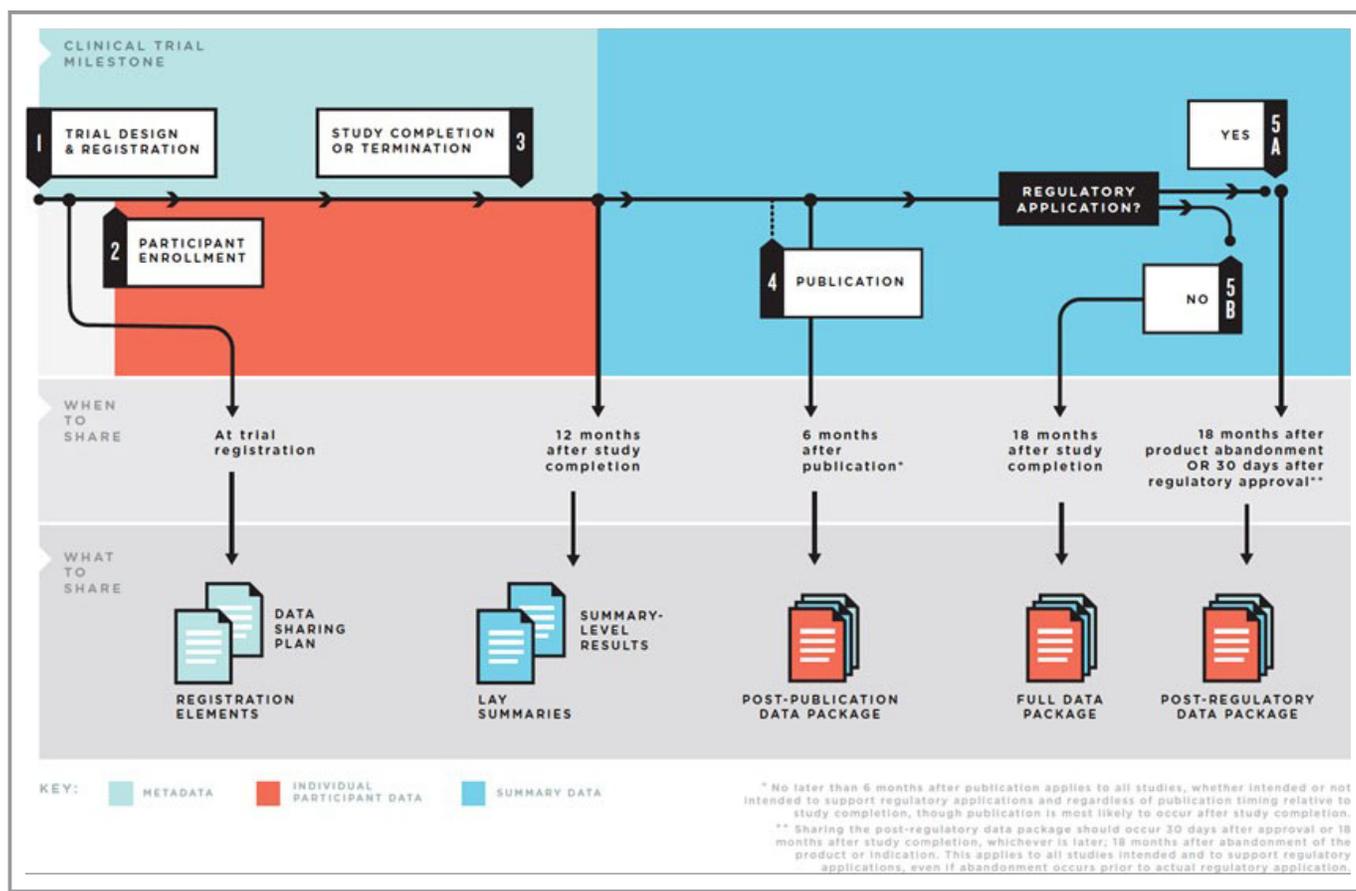
stakeholders were considered. The discussion focused on how principles for data sharing can help inform AHA research policies in the future and may serve as a precedent for other groups in medicine to address data sharing in the future.

### Ethical Considerations

As patients, as relatives, and as friends of patients, everyone wants to obtain the most complete and rigorous information possible about the effectiveness and safety of therapies. Responsible sharing of clinical trial data helps provide such information and thus is in the public interest.<sup>88</sup> The public good should be the guiding ethical concern regarding data sharing; however, the interests and concerns of stakeholders must be addressed and balanced (Figure 7).<sup>44</sup> For research participants, informed consent, privacy protections, and knowledge gained from their efforts are important.<sup>89</sup> Clinical trialists want time to publish their analyses and to get credit

for sharing data and analytic tools. Secondary investigators want to analyze unpublished data and reproduce published findings. Sponsors want their funding to yield new knowledge but do not want data sharing to compromise funding for new research projects. For-profit sponsors want to protect their intellectual property and commercially confidential information. A major challenge in data sharing is forging agreement on specific measures to address these stakeholder concerns.

The ethical principles of respecting research participants, maximizing benefits and minimizing harms, and acting justly or fairly should guide data sharing. To respect participants, we must address the issues of informed consent and privacy in the context of sharing data. For meaningful consent, research participants need to receive information about data sharing that is pertinent to their decision to participate. Although users of mobile apps and Web sites authorize data sharing by clicking “I agree” on data use agreements, such perfunctory permission does not meet the standards for consent for



**Figure 7.** Major stages of the clinical trial life cycle and recommendations for when to share specific data packages. Data are generated at nearly every stage of the clinical trial life cycle, including the initial protocol and statistical analysis plan prepared prior to registration, the collection of baseline participant data at participant enrollment, and the analysis of the analyzable data set. To help frame the discussion of what data should be shared at what times, the Institute of Medicine report on sharing clinical trial data described the clinical trial life cycle as consisting of 5 major stages with guiding principles and a practical framework for the responsible sharing of data, including the types of data available at different stages of a trial and the optimal times to share them. Reprinted from Institute of Medicine report on Sharing Clinical Trial Data.<sup>44</sup>

research. An unresolved issue is whether participants may consent to some types of data sharing but not others. For older completed clinical trials, consent for data sharing was inconsistent and often ambiguous or absent. Privacy is challenging in the Big Data era because data that are deidentifiable by Health Insurance Portability and Accountability Act (HIPAA) standards may become identified when additional data are brought to bear. To increase the benefits of data sharing and reduce the risks, organizations that share data should learn from their experience by collecting data on the outcomes of their model of data sharing, disseminating this information and the lessons learned, and continuously improving the data-sharing process. To reduce harms, data holders can require procedures that reduce the risk of invalid secondary analyses, such as data use agreements. To achieve fairness, an equitable and sustainable funding model for data sharing is needed. Those who benefit from data sharing should bear a fair share of the costs of sharing. Additional ethical considerations are sufficient access to shared data to achieve the benefits of data sharing, accountability of data generators and requestors, and practicability.

## Models of Responsible Data Sharing

A useful set of principles to govern the pursuit of responsible data sharing was laid out in an article in *The New England Journal of Medicine*.<sup>88</sup> First, the model should provide sufficiently broad access to data to achieve the sought-after benefits. It should apply to trials of all drugs, devices, and biologics approved in at least 1 country. Second, it should be designed to maximize protection of participants' privacy interest. Third, it should treat all qualified data requestors and trial sponsors evenhandedly. Fourth, it should ensure accountability by requiring data requestors to commit to protecting participants' privacy and conducting analyses that adhere to accepted scientific standards. If those who generate the data are allowed to influence when the data are released, they must commit to transparent, principled decision making. Finally, the system should be practicable. It must be able to render timely decisions and avoid undue burdens on data generators. To advance these principles, a data-sharing system should have the following specific features:

- A binding mechanism to ensure universal participation and compliance by data generators—a regulatory requirement is the most obvious mechanism, although trial sponsors would prefer a private alternative
- Minimum standards for what must be shared and how
- Equal application of any requirements to all trial sponsors
- Explicit decision criteria for data releases

- Public disclosure of the reasons for decisions
- Public disclosure of requestors' identities and analysis plans
- A mechanism to enforce conditions of data use, such as a data use agreement
- Provision of technical support to ensure that data requestors understand the data

In public discussions of data sharing, 4 models have emerged. In the first, an “open access model,” data sets and accompanying materials would be posted online for downloading. This model would serve the principles of broad access and transparency superbly but would not provide sufficient protection of the interests of participants and data generators.

In the second model, a “database query” model, the data generator would continue to hold the data and would run analyses on the data at the request of outside parties and send out the results. The data generator would be obliged to run any analysis that met the following 3 decision criteria:

1. Is there a reasonable scientific hypothesis, sound analytical plan, and adequate plan to disseminate findings?
2. Do the potential public health benefits of answering the proposed question outweigh the probable adverse effects on the data generator and risks to participants?
3. Does the requester have expertise sufficient to carry out the analyses?

The database query model serves sponsors' interests in controlling their data but otherwise is unappealing. It is resource intensive for data generators and could involve long delays in returning results. More important, it does not provide transparency. Requestors would not be able to verify that the analyses were conducted as envisioned.

The third model, a “sponsor review,” would release participant-level data sets but allow the trial sponsor to control decisions about releases, with some appeal mechanism. The sponsor would apply the decision criteria described above. This model allows sponsors to preclude what they deem inappropriate analyses but lacks the independence that is needed to cultivate public trust in the system. The potential for perceived inconsistencies and biased decisions is apparent. There would almost certainly need to be an appeal mechanism.

The fourth approach, the “learned intermediary” model, reposes responsibility for decisions in an independent board. The board would apply the decision criteria specified above, ensure that only the minimum data necessary to answer the study questions are released, and execute a data use agreement. Because of the intermediary's independence and weighing of the risks and benefits of each data request, this is the most desirable of the 4 models. Again, however, it seems essential to allow for appeal and to provide detailed

explanation for decisions, a vastly difficult and resource-intensive task.

## Institute of Medicine Report on Sharing Clinical Trials Data

In 2014, the Institute of Medicine (IOM) commissioned a committee to review the current practice of sharing of completed clinical trial data and to make recommendations regarding data sharing.<sup>44</sup> There are several compelling reasons for sharing data, including making it feasible for other investigators to reproduce initial published findings and to carry out additional analyses, strengthening and increasing scientific knowledge, maximizing the contributions of patients and sponsor investments, and stimulating new ideas for research. There are also some serious challenges to address, including the need to protect participant privacy and honor informed consent as outlined by HIPAA, to safeguard legitimate economic interests of sponsors (eg, intellectual property), to guard against invalid secondary analyses, to allow the academic investigators adequate time to conduct their secondary analyses and be recognized for their academic contributions, and to avoid unfunded mandates for data sharing.

In addition, there are several key stakeholders in the process of data sharing. These include the trial participants, the investigators who conducted the trial and who are typically involved in the design and analysis, the institutions at which the trial was conducted, funders and sponsors who provided the resources for the trial, regulatory agencies who may need to review the trial data for purposes of product approval, research ethics groups such as institutional review boards, medical journals that publish the results of such trials, professional societies that interpret the trial results and form practice guidelines, and patient advocacy groups.

In sharing clinical trial data, it is necessary to define the data to be shared, and there are many different types of data to be considered. First, there is the raw patient-level data, which may take the form of case report forms on which individual patient data are recorded manually or electronically, laboratory data such as x-ray films, magnetic resonance imaging data, ECG tracings and clinical laboratory data based on blood and urine specimens, quality-of-life questionnaires, and textual clinician notes. For data to have any meaning, there must be metadata, which are data about the data, describing the numerical fields, the methods used for collection, and other relevant descriptors. For data to be analyzable, it needs to be converted into numerical metrics at a participant level; however, clinical trials often contain much more analyzable data than are typically analyzed in publications and other summary reports. Consequently, there will be an analyzed participant-level data set that corresponds to each publication or report. Many trials have a complete

summary report, especially trials sponsored by industry, for which data are submitted for regulatory review. Finally, brief summaries of the analyzed data appear in publications in leading medical journals.

The IOM Report on Responsible Data Sharing contains 4 recommendations.<sup>44</sup> First, the stakeholders in a clinical trial should foster a culture in which data sharing is the expected norm and should be committed to a responsible strategy for this process.

Second, sponsors and investigators should share the various types of clinical trial data no later than the following timelines: Before the trial is initiated, the trial should be registered with ClinicalTrials.gov, or the equivalent in other countries, and should include the protocol, the data sharing plan, and the statistical analysis plan. Within 12 months of study completion in participant follow-up, (eg, last patient, last visit), a summary of results should be provided in ClinicalTrials.gov and in a lay or public-level presentation. Within 6 months of publication in a medical or scientific journal, the patient-level analyzed and deidentified data used in the paper should be made available along with the protocol, the statistical analysis plan, and the analytic code used in the analysis. Continuing on this timeline, within 18 months after trial completion (last patient, last visit), the full analyzable deidentified data set should be made available along with the protocol, the metadata for the full data set, the protocol, the statistical analysis plan, and all analytic code used in publications or summary reports. For trials that are submitted to regulatory agencies for product approval, the full analyzable deidentified data set, protocol, statistical analysis plan, and redacted complete summary report should be made available.

Third, holders of clinical trial data should use data use agreements, designate an independent review panel, and make access to clinical trial data transparent.

Fourth, sponsors and investigators must address the following specific issues:

- **Infrastructure:** There are currently insufficient platforms to store and manage data.
- **Technological:** The current platforms are not consistently discoverable, searchable, and interoperable.
- **Workforce:** Our current clinical trial research staff lacks skills and knowledge to manage operational and technical aspects of data sharing.
- **Sustainability:** The current model costs are borne by a small subset of sponsors, funders, and trialists and are unsustainable.

The IOM report includes the following key summary messages:

- Responsible clinical trial data sharing should become the norm.

- Stakeholders and institutions need to work together to agree on best practices, standards, and incentives.
- Evolution should be guided by empirical data, lessons learned, and best practices.

## Perspectives on Pediatric Research

Data sharing is vital in research performed in children, defined as persons who have not attained the legal age for consent to treatments or procedures involved in the research, under the applicable law of the jurisdiction in which the research is conducted (45 CFR 46.402).<sup>90</sup> Children differ from adults in organ physiology, development, and response to drugs and other interventions. They are a vulnerable group incapable of protecting their own interests and cannot take more than minimal risk when research has no benefit to the individual child. Many childhood diseases are rare and require multicenter trials, registries, and data repositories to achieve adequate sample sizes.

Legal consent is always obtained from parent or guardian proxies, whose values and viewpoints may differ from those of the children. Children who are old enough and capable of understanding must give their “assent” but may not fully understand the risks and benefits. A challenge of data sharing in pediatric research relates to the need to “reconsent” participants once they reach the age of 18 years, after which parental permission and child assent are no longer valid. Specifically, guidance issued by the Office for Human Research Protections advises that investigators seek and obtain a legally effective informed consent, as described in 45 CFR 46.116, for any ongoing human subjects research, including studies that involve the continued analysis of specimens or data for which the participant’s identity is readily identifiable to the investigators. Many years after the original study, now-adult participants may be difficult to locate. If appropriate, the institutional review board may consider a waiver under 45 CFR 46.116(d) of the requirements for obtaining informed consent for the patients to continue their participation in the research.

Waiver of consent may be obtained under circumstances that do not arise in adults. If the institutional review board determines, for example, that a research protocol is designed to study conditions in children or a subject population for which parental or guardian permission is not a reasonable requirement to protect the participants (eg, neglected or abused children), it may waive the parental permission requirements. Under these circumstances, potentially deleterious information could be revealed when neither the parent nor the child has consented.

In some circumstances, only the child has provided consent, and these situations are often in the most sensitive

areas. If research on a specific treatment involves treatments or procedures for which minors can give consent outside the research context under state and local law (eg, research on drug use, sexually transmitted diseases, or pregnancy), these patients do not meet the definition of children and thus parental consent is not required. This population may have a higher risk of inadequate cognitive or emotional maturity to anticipate the potential consequences of the later disclosure of this information.

Data sharing can also be challenging after the unexpected death of a child. The NHLBI and the National Institute of Neurological Diseases and Stroke are collaborating with the Centers for Disease Control and Prevention to create the Sudden Death in the Young (SDY) Case Registry to collect information in up to 10 states or jurisdictions on persons aged ≤19 years who experienced sudden unexpected death to create a registry of clinical information and DNA samples that can be used to investigate sudden death in the young.<sup>91</sup> The state public health agencies that conduct Child Death Reviews are grantees funded to collect data for the SDY Case Registry and differ in their opinions about the need for informed consent because use of a sample from a deceased subject is not considered human subject research by regulatory definitions. Because DNA will be linked with phenotypic data, albeit deidentified, the SDY Case Registry team decided that permission should be sought from the parents of a deceased child; however, to allow parents to make an informed decision, the steering committee favored waiting to require informed consent until after preliminary autopsy results had been discussed with the family. Because DNA sent soon after its collection has higher quality, it would be optimal to send blood samples to the biorepository immediately after autopsy. Nevertheless, state grantees differed in their opinions about whether blood samples could be collected and shipped to the biorepository before parents gave their permission. Ultimately, it was decided that specimens could be sent prior to family permission and could be accessed by medical examiners and coroners prior to consent but that the samples would only be made available for research once consents were signed.

Data sharing without consent in pediatric research has come under fire even when samples are deidentified. Specifically, the Newborn Screening Saves Lives Reauthorization Act of 2014 (Public Law No. 113-240) requires that all research using newborn dried blood spots be considered human subjects research regardless of whether the specimens are identifiable. This law eliminates the ability of institutional review boards to waive informed consent for such research. A Minnesota-based national organization dedicated to preserving patient-centered health care and protecting patient and privacy rights, the Citizens’ Council for Health Freedom, advocated for this amendment, which required that

parents give informed consent before their infant's DNA could be used in federally funded research. As a result, years of dried blood spots and associated data were destroyed in Minnesota.

Despite some of the challenges in pediatric research, data sharing provides the opportunity for extraordinary benefit to children with congenital and acquired forms of heart disease. Congenital heart disease affects almost 1% of live births and is the leading cause of mortality from birth defects.<sup>19</sup> Although common in aggregate, congenital and acquired pediatric heart diseases are rare and diverse disorders involving many possible causative genes, molecular pathways, and networks. Data sharing through multicenter consortia, such as the ongoing NHLBI's Pediatric Cardiac Genomics Consortium and, potentially in the future, the AHA's CVGPS, can improve our understanding of the genetic causes of congenital and pediatric heart disease and provide new avenues for investigating the relationship of genetic variants to clinical outcomes.

## Addressing Workforce Challenges With Continuing Education

The explosive growth of Big Data in health care will create constantly changing challenges and opportunities for the workforce of the future. Healthcare professionals will expect continuing education (CE) not only to enable them to meet licensing, credentialing, and other professional obligations but also to help them navigate and benefit from emerging technologies. With >1 million hours of instruction delivered annually,<sup>92</sup> accredited CE providers (including the AHA) have the reach, expertise, and resources to be strategic partners for health systems, institutions, care teams, and individual health professionals by facilitating the integration of new technologies into continuing professional development.

Educators will need to respond with nimbleness and flexibility to new educational needs and practice gaps that arise from advancing technology. The professional competencies, such as communication skills and patient-centered care, will serve as a framework for creating education that addresses the issues raised by Big Data,<sup>93</sup> for example, how to communicate with patients about data privacy issues. The workforce of the future will need practical education about how to integrate Big Data into practice and ongoing forums about the ethics involved. CE can provide opportunities for faculty and leadership development and peer interchange to support ongoing discourse and policy development about evolving ethics issues in the spheres of research, clinical care, and executive management.

Big Data will affect the entire cycle of workforce learning, from the identification of performance gaps and practice-based educational needs to outcomes measurement. Using

data that analyze cardiovascular health issues at the national, regional, local, and individual patient levels, educators and healthcare professionals can construct educational and systems interventions to improve health outcomes. Through collaborations and appropriate data sharing, CE programs can identify and respond to emerging public health priorities through workforce education and other interventions. Educators across the country can share information gleaned from Big Data to create relevant, effective learning solutions that can be distributed to healthcare professionals across the country where they work, live, and practice.

Numerous intervening variables exist between educational activities and patient outcomes, such as systems barriers and team performance. Big Data will enable the advancement of comparative effectiveness and implementation studies that link CE to performance improvement and patient care outcomes and will demonstrate how best to generate long-term retention and behavior change. With effective CE and support systems, the workforce of the future will have the data and expertise not only to improve their own practice but also to meaningfully address issues that affect population health in communities throughout the nation.

## Scenarios of the Future: What Do Stakeholders Want From Big Data?

The potential and promise of Big Data include opportunities to acquire, analyze, and share information for a multiplicity of goals. Summit attendees envisioned and reflected on the key objectives to which each stakeholder group might aspire in the utilization of the resultant data under ideal circumstances (Table 3). Some common themes emerged, including a passionate wish for accurate, secure, and complete data. Many groups detailed the urgent need for the development of tools that might distill data into visual depictions that could be used at the point of clinical care or for healthcare administration. An additional hope of attendees was that clinical and translational research could be facilitated and accelerated by the authorized sharing of patient data.

## What Do Patients Want From Big Data?

Patients expect excellence in health care. They are frustrated by the need to recount their personal history to every new provider, hospital, or service. They do not understand why our systems "can't just talk to each other."<sup>14</sup> They want to be partners in their care with the ability to access their personal information and to interface with their provider's office to make appointments, schedule tests, check results, and correct errors. Above all, they want their data to be secure and portable, whether they change doctors, systems, or

**Table 3.** Big Data and Stakeholders

Stakeholder	What Stakeholders Want From Big Data	The American Heart Association's Role
Patients	<ol style="list-style-type: none"> <li>1. Controlled access to portable secure medical information</li> <li>2. Access to best possible health outcomes at affordable cost</li> <li>3. Easy access to medical research/clinical trials</li> </ol>	<p>High priority</p> <ol style="list-style-type: none"> <li>1. Facilitate the use of Big Data by patients and professionals</li> </ol> <p>Medium priority</p> <ol style="list-style-type: none"> <li>1. Patient-centered advocacy on the use of Big Data</li> </ol> <p>Low priority</p> <ol style="list-style-type: none"> <li>1. Facilitate the evaluation of health technology devices</li> </ol>
Basic Scientists	<ol style="list-style-type: none"> <li>1. Develop and identify novel targets that otherwise would not be identified by traditional methods and query for CVD-related genes/proteins</li> <li>2. Use systems approach to identify multiple genes/proteins that collectively cause CVD</li> <li>3. Access and use of analytical partners to help advance research goals</li> </ol>	<p>High priority</p> <ol style="list-style-type: none"> <li>1. Create an AHA digital ecosystem to build an AHA knowledge base (AHA Commons) and enable AHA investment for sustainable and long-lasting impact</li> <li>2. Grant funding to establish training/learning, novel tools/products for analytic techniques</li> </ol> <p>Medium priority</p> <ol style="list-style-type: none"> <li>1. Broker partnerships to advance, collect and analyze data</li> </ol> <p>Low priority</p> <ol style="list-style-type: none"> <li>1. Establish AHA data science policies and guidelines (privacy, ethics, intellectual property)</li> </ol>
Clinical Investigators	<ol style="list-style-type: none"> <li>1. High-quality standardized clinical data for secondary use</li> <li>2. A standardized technology platform with interoperable, feasible, and federated access to a broad range of clinical data</li> <li>3. Better, novel, more rapid mechanisms of support for analysis of Big Data</li> <li>4. A mechanism for ongoing discussion of these topics that includes the clinical investigator community</li> </ol>	<p>High priority</p> <ol style="list-style-type: none"> <li>1. Training and funding a new generation of Big Data and users (clinicians to developers)</li> </ol> <p>Medium priority</p> <ol style="list-style-type: none"> <li>1. Bridge the gap between the theoretical promise of Big Data to potential use (eg, map EMR to "Get With The Guidelines"; provide leadership in data standard, quality, and validity)</li> </ol> <p>Low priority</p> <ol style="list-style-type: none"> <li>1. Be the match.com for data owners and data researchers</li> </ol>
Population Scientists/ Epidemiologists	<ol style="list-style-type: none"> <li>1. Resource effectively and integrate data from large numbers of diverse people, participants, patients longitudinally gathering across systems and populations</li> <li>2. Integration of data including, social determinants of health, exposure data, including mobile health as a source of data, physiological data, healthcare encounters/procedures; should be easily accessed</li> <li>3. Design and deploy the new prevention science using mobile health to empower the individual participant</li> </ol>	<p>High priority</p> <ol style="list-style-type: none"> <li>1. Provide resources for multidisciplinary investigators at all stages to broadly access science across the AHA portfolio including population science and epidemiology</li> <li>2. Create a data science fellowship whereby top tier data scientists can be funded and mentored by AHA multidisciplinary scientists</li> </ol> <p>Medium priority</p> <ol style="list-style-type: none"> <li>1. Commit resources to convene a forum focused on Big Data harmonization and validation inclusive of multiple stakeholders and career levels</li> <li>2. Help scientists along the clinical–translational continuum translate and educate the public, clinicians, and researchers about Big Data</li> </ol>

Continued

Table 3. Continued

Stakeholder	What Stakeholders Want From Big Data	The American Heart Association's Role
Clinicians/ Healthcare System Researchers/ Administrators	<ol style="list-style-type: none"> <li>1. Engagement across stakeholder domains</li> <li>2. Empowering patients using Big Data</li> <li>3. Identifying at-risk populations with Big Data decision support</li> <li>4. Assessing and equipping providers with tools for collection, distillation and visualization</li> <li>5. Leveraging Big Data to enrich the practice of medicine—more efficient and more enjoyable</li> <li>6. Teaching providers about Big Data</li> </ol>	<p>High priority</p> <ol style="list-style-type: none"> <li>1. Develop and disseminate accepted clinical standards and benchmarks</li> </ol> <p>Medium priority</p> <ol style="list-style-type: none"> <li>1. Sponsor the development of multidisciplinary tools for data analysis</li> </ol> <p>Low priority</p> <ol style="list-style-type: none"> <li>1. Convene all benchmarking communities and stakeholders</li> </ol>
Industry	<ol style="list-style-type: none"> <li>1. Understanding of the available data:                             <ol style="list-style-type: none"> <li>a. ability to access,</li> <li>b. appropriateness of use,</li> <li>c. context of collection,</li> <li>d. known quality issues,</li> <li>e. any inherent biases,</li> <li>f. clarity of business models</li> </ol> </li> <li>2. Robust analytical methods and scientifically accepted ways to deal with common data issues</li> <li>3. Rigor and disclosure of analytic methodologies when findings may affect healthcare policy decisions and clear communication of findings</li> <li>4. Clarity in how observational data may be used in support (or replacement) of RCT data in regulatory decision making</li> <li>5. Ensuring appropriate patient privacy and confidentiality</li> </ol>	<p>High priority</p> <ol style="list-style-type: none"> <li>1. Provide a mechanism to aggregate different databases allowing researchers to overcome limitation of any single data set (eg, statistical power, sample characteristics)</li> <li>2. Stimulate use of existing data sets through targeted research funding</li> </ol> <p>Medium priority</p> <ol style="list-style-type: none"> <li>1. Help researchers better understand the sampling methods and variables in existing data sets</li> </ol> <p>Low priority</p> <ol style="list-style-type: none"> <li>1. Facilitate appropriate availability of data to a larger community of researchers by serving as an “honest broker”</li> <li>2. Assist researchers in accessing data sets by clarifying data use requirements</li> </ol>
Regulatory Authorities	<ol style="list-style-type: none"> <li>1. Promote public health and innovation</li> <li>2. Careful attention to development of Big Data methodology—is it hypothesis generation only?</li> <li>3. Investigate heterogeneity of US population</li> <li>4. Rethink regulatory framework for informed consent</li> <li>5. Rethink regulatory framework for Big Data (eg, informed consent, OHRP, cluster randomized clinical trials)</li> </ol>	<p>High priority</p> <ol style="list-style-type: none"> <li>1. Develop analytic standards for                             <ol style="list-style-type: none"> <li>a. statistical methods,</li> <li>b. stated plans and protocols,</li> <li>c. data confirmation</li> </ol> </li> </ol> <p>Medium priority</p> <ol style="list-style-type: none"> <li>1. Develop common terminology and other data standards</li> <li>2. Preserve evidence standards for new uses and comparative effectiveness so Big Data can lead to changes in patient solutions and new hypotheses</li> </ol> <p>Low priority</p> <ol style="list-style-type: none"> <li>1. Registry of cardiovascular Big Data analyses</li> </ol>

AHA indicates American Heart Association; CVD, cardiovascular disease; EMR, electronic medical record; OHRP, Office for Human Research Protections; RCT, randomized controlled trial.

states, and they want smart-device interfaces so they have multiple options for access.

Patients assume their providers' recommendations will be based on the most up-to-date science and guidelines. They want to know all of the possible outcomes they can expect, and they want the best possible outcome at an affordable cost. Patients want to connect with others who share similar conditions and to use that community to better understand

their own symptoms or disease, to share concerns, and to gather information to become more effective partners in care.

Patients have shown a willingness to share their personal data to improve the health of themselves and/or others.<sup>94–96</sup> They want to know about the latest research, but often that information is not readily available in an understandable format. Patients are willing to become involved in medical research and clinical trials but need to be made aware of the

studies and then to have an easily accessible avenue for determining whether they or a loved one would be eligible to participate.

### What Do Basic Scientists Want From Big Data?

Most Big Data science approaches and methodologies such as data mining, machine learning algorithms, crowdsourcing annotation platforms, cloud computing infrastructure, and Bayesian network algorithms are new to the basic cardiovascular community. Regardless of the source, these approaches are feasible in the world of basic cardiovascular science. Many domain-specific databases, for example, are created in “standard” database technologies (eg, Oracle MySQL). These technologies are at the heart of most curated data repositories (eg, ICPSR, NCBI resources) and are very familiar to research teams.

Big Data concepts of integrating various types of data would allow basic scientists to potentially develop and identify novel targets that otherwise would not be identified by traditional methods. New data mining and analysis techniques would also allow researchers to query for genes and proteins linked to CVD and stroke. Using a systems approach would increase the possibility of identifying multiple genes and/or proteins that collectively cause CVD and stroke. Such advances could lead to new prognostic markers and, conceivably, therapeutic targets.

Last, access and use of analytical partners (eg, informatics, biostatistics, data to knowledge, architecture design of omics) is critical to advance basic research goals. To connect databases, linked data technologies may be used, connecting database fields through taxonomies or ontologies. Examples include Open PHACTS,<sup>97</sup> NIF,<sup>98</sup> and DataONE.<sup>99</sup> Internet search engines as both locators of information and immediate providers (via hyperlinks to the actual sources) have recast the notion of a catalog from a book or even a persistent database to a collection of dynamic, real-time computing functions. Web search technologies are also effective for creating data or tool indexes and repositories.

To fully appreciate the potential of basic science data in the evolving Big Data world, data should be Findable, Accessible, Interoperable, and Reusable (the FAIR model)<sup>100</sup> by:

- Organizing and building platforms rendering large-scale clinical cohorts findable
- Creating metadata standards to establish policies and guidelines on accessibility
- Building computational tools to define cardiovascular phenotypes, making them interoperable with other data platforms
- Transforming biomedical research culture to support advancement in precision medicine<sup>101</sup>

### What Do Clinical Investigators Want From Big Data?

Clinical investigators who are interested in evaluating the efficacy or safety of specific interventions in small- or large-scale trials generate substantial data that might benefit from merging with each other or with other data sets.<sup>17,102</sup> Using our current approach, however, we create the data platforms for characterizing patients and events de novo every time we design and carry out a clinical trial. Although informed by past experience, each clinical trialist develops a unique protocol and a unique case report form, and trialists are likely to use meaningfully different approaches to identifying patient phenotypes or clinical events. It has been difficult to develop standardized approaches to describing the symptoms, signs, severity, and other relevant features of common cardiac disorders.<sup>36</sup> Consequently, assuming that appropriate statistical and informatics methods were available to combine data sets, it is not clear whether the application of such methods to existing data sets would yield a useful synthesis. Efforts by the US Food and Drug Administration to standardize the definition and reporting of specific cardiac events (eg, myocardial infarction, stroke, or hospitalization for heart failure) represent an important advance toward this goal, but much work remains to be done. These efforts will be complicated by a simultaneous interest in streamlining the amount of information collected in clinical trials, based on the assumption that we already know what data to collect and how to collect them. Assuming that the data collected could be of higher quality and the merger of data sets could be made interoperable, clinical investigators not only would want to contribute to the creation of such data sets but also would be interested in posing questions that could be answered by such data sets. Nevertheless, it is currently very difficult to interact with the platforms available for merged data sets in a manner that is likely to reward clinical investigators for their time and effort. Regardless of the approach developed, clinical investigators wish strongly to be involved in the ongoing discussions with other stakeholders and trust that the financial resources will be available to enable the development and mining of large data sets to address important clinical questions.

### What Do Population Scientists and Epidemiologists Want From Big Data?

Population scientists recognize the potential of new tools to collect and analyze data for epidemiology research.<sup>37</sup> Digital data sources, which include mobile health electronic devices and clinical care databases, greatly expand the prospects for population science. The new tools can constitute powerful platforms to ascertain exposure, and the electronic access

to administrative and clinical data can be used to sample populations and capture outcomes. The steps to define and realize the promises of Big Data have been delineated; however, the best scientific approaches to optimally acquire, validate, standardize, and analyze high-volume data remain to be learned.<sup>20,21</sup> Of particular importance is the issue of training, which is critical for population scientists to use digital tools, and a robust data sciences curriculum must be integrated into population sciences education endeavors. Multidisciplinary team science that focuses on Big Data must be formally deployed through new funding mechanisms.

### What Do Clinicians and Healthcare Systems Want From Big Data?

Clinicians and health systems see significant value in Big Data, particularly if it is used to enhance engagement between patients and their healthcare providers.<sup>17</sup> If patients know that the data they generate will be used to improve their own health, they may be more likely to embrace the role of Big Data in their care. Clinicians will welcome novel information about their patients that will enable better care, but they will not be able to cope with vast amounts of unformatted data that require a great deal of their limited time to analyze and record in an EMR.

To derive clinical value from Big Data, tools must be developed to collect data efficiently in formats designed to allow its exchange, to distill data into information that facilitates clinical care and research, and to visualize data in ways that are accessible and actionable by both providers and patients.<sup>7,103</sup> With the appropriate tools, Big Data could be powerful in clinical care. It may be particularly useful in identifying at-risk populations, enhancing decision support, and performing predictive modeling to identify and continuously evaluate opportunities and strategies to improve population health.<sup>57,79</sup>

The use of Big Data has enormous potential to enrich the care of patients and to provide new ways for patients and providers to engage around health and wellness. If properly deployed, Big Data initiatives could make the practice of medicine more effective, more efficient, and more rewarding for providers and their patients.<sup>36</sup>

### What Does Industry Want From Big Data?

The biopharmaceutical industry sees significant potential to improve the lives of patients through the use of Big Data.<sup>17</sup> Big Data may help, for example, in the discovery, development, and delivery of medicines to patients. Although the use of Big Data to inform internal company processes is relatively clear, a number of questions need to be answered

to realize the full benefits for patients regarding the use of Big Data by all stakeholders (including sponsors, regulators, payors, academics, and providers) in informing healthcare policy decisions that may affect patients' ability to access medicines. At their core, the questions needing to be addressed relate to clarifying the nature and availability of high-quality data, elucidating the appropriate uses of those data in different contexts, ensuring quality and rigor in those uses, addressing healthcare policy decisions, and subsequently communicating findings. In addressing these issues, it will be essential to have constructive dialogue among patients, industry, regulators, providers, and academics because each party has an important stake and perspective. AHA could play a pivotal role by facilitating appropriate access to large data, providing detailed information about existing data sets, and stimulating the use of Big Data by the research community through targeted research funding.

### What Do Regulatory Authorities Want From Big Data?

Major goals of regulatory agencies are to actively promote public health and medical innovation. In addition, regulators are interested in better understanding heterogeneous treatment effects and the rare, but important, adverse safety problems produced by medical products. Big Data has significant potential for assisting regulators in all of these areas. Big Data, when used appropriately, may help develop key hypotheses or test important treatment strategies and thus accelerate understanding of fundamental biological processes.<sup>104</sup> Regulators are particularly interested in the use of Big Data to design and conduct randomized trials (including possibly cluster randomized trials).<sup>105</sup> These data can identify potential participants, facilitate invitations to participate, perhaps be used for the consent process, and decrease the amount of data that need to be newly collected, all of which could facilitate larger and more rapidly conducted trials for such studies as comparisons of multiple drugs. It will be necessary to see if critical features of good randomized trials (blinding) can be accommodated. As a first step, however, regulators need to be assured that the design and methodology used for Big Data studies and analyses are appropriate for the questions being asked<sup>38,106</sup>; otherwise, multiple unintended consequences may result. It is essential that regulators be assured that design and methodological considerations have been adequately addressed before fully integrating Big Data analytics into the regulatory science landscape.<sup>83</sup> Interestingly, some considerations such as the informed consent process may need to be rethought to take full advantage of Big Data while protecting patients.

## The American Heart Association's Role

The AHA has a track record of supporting cutting-edge science and already has initiated several innovative approaches to Big Data. These approaches are described below, followed by a summary of the final facilitated discussion on possible roles that the AHA might play in the future across stakeholder domains.

### AHA's Cardiovascular Genome Phenome Study

Despite extraordinary progress over the past 5 decades, heart disease remains the number 1 cause of death in the United States.<sup>19</sup> To accelerate the pace of research and transform cardiovascular care, the AHA has launched the CVGPS (Figure 8).<sup>56</sup> This effort will leverage the convening power of the AHA to develop a vision for the future of cardiovascular care and the research required to achieve that vision. A diversified portfolio of innovative grants will follow and include traditional multiyear research grants; smaller, rapid-turn, high-risk, high-potential innovation grants; prizes and challenges; data science grants; and community development grants that support informatics and related capability building. In addition, the AHA will drive the translation of discovery to care by asking grantees to present to a panel of industry evaluators; if there is translational promise, next steps will be suggested including potential industry partnerships or additional funding sources.

CVGPS will aggregate multiple cardiovascular studies to enhance data discoverability for a broader range of investigators. Using an online retail-like front end, researchers can search across multiple studies to assemble virtual cohorts, test new hypotheses, and engage in hypothesis-agnostic data mining. Finally, the AHA plans to enhance the

effort through analytic support. "Data navigators" will guide investigators through aggregated studies and data sources and provide methodological and analytic advice. In addition, robust analytic tools and services will be included for investigators who may have fewer resources. Together, these 5 CVGPS thrusts—a future vision for care, an innovative grant strategy, a translation engine, a cardiovascular Big Data repository, and analytic support—will enhance collaboration, democratize data accessibility, accelerate research, and advance cardiovascular care for the 21st century.

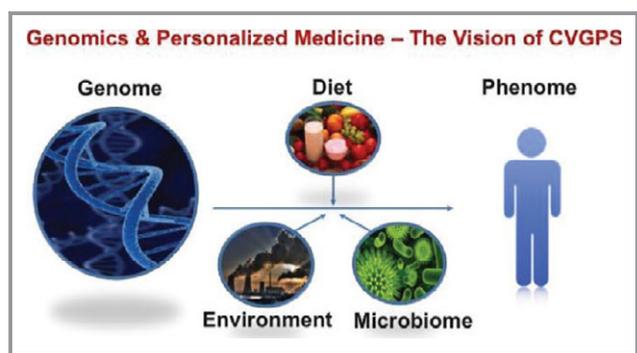
### The AHA's eHealth Initiatives

From a health education perspective, there is a strong belief in individual decision making and patient empowerment. The AHA has embraced the idea of functioning in this field as an information provider. Consequently, the goal is to deliver information at the right time and in the right way to patients, caregivers, and their families such that improved and informed health decisions can be made.

The AHA also understands that new health information technologies and increasing investments by government and private industry present enormous opportunities to reach patients, caregivers, and families in new ways.<sup>57,103</sup> As a result, it is actively pursuing relationships with a broad cross-section of established companies and startups to improve health outcomes.

Simply put, the AHA is striving to move squarely into the lives of healthcare consumers through a variety of different settings and mediums (Figure 9). To do this, the AHA will rely on a series of industry-based partnerships, both in terms of technology and sources of funding. The AHA will continuously evaluate its activities across the healthcare continuum, from primordial prevention, primary prevention, acute events, secondary prevention, and long-term care to the ultimate goal of creating seamless and effective offerings that support the evolving needs in the healthcare system—in other words, connecting health.

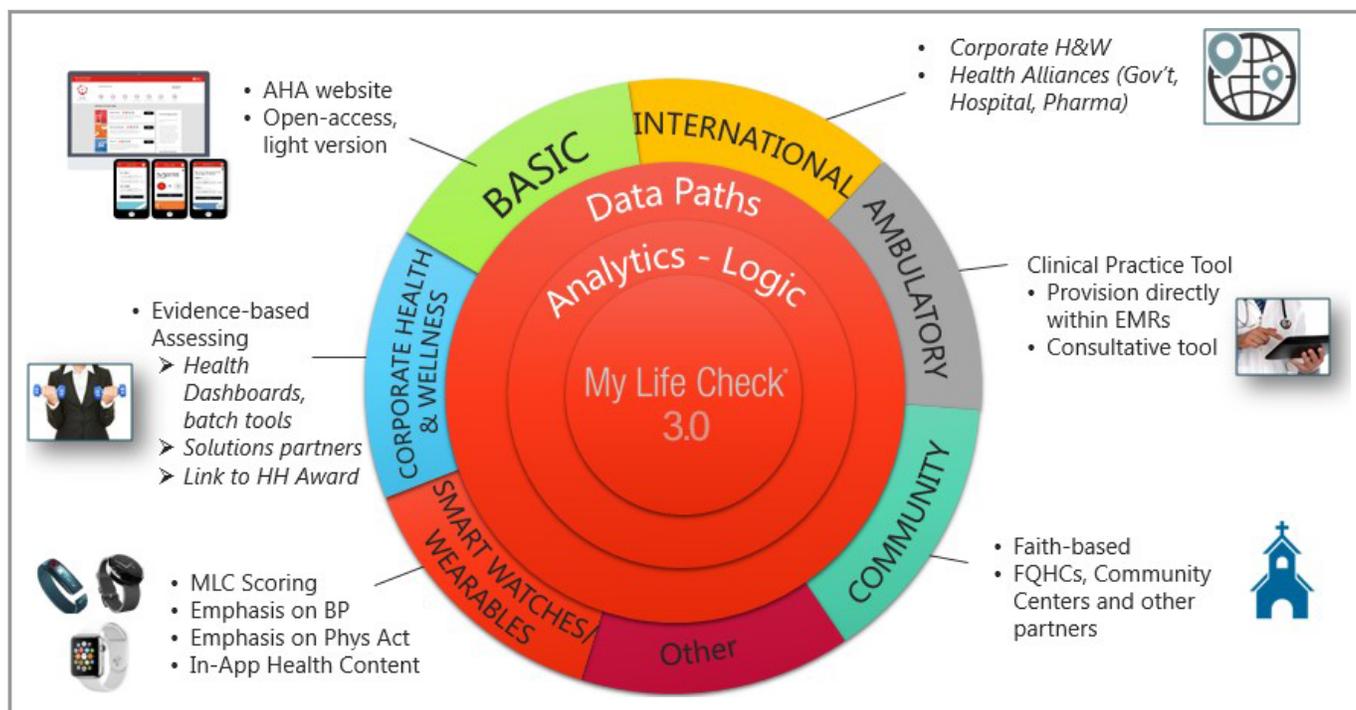
Based on the discussion of what stakeholders want or need from Big Data, the summit attendees discussed possible roles that the AHA might play across a range of domains (Table 3). The outcomes will be important input for the AHA to maximize the benefits of data sharing across a variety of activities in cardiovascular and stroke science.



**Figure 8.** Vision for the Cardiovascular Genome Phenome Study (CVGPS). Deep genotyping will provide assessment of the genetic and epigenetic determinants of disease. When data from diet, the environment, the microbiome, and deep phenotyping are combined, investigators will establish a 360° look at cardiovascular health. Reprinted from Antman.<sup>56</sup>

### What Possible Roles Might the AHA Play in the Patient Domain?

The AHA is uniquely positioned, from a patient perspective, to act as a key enabler in realizing the promise of



**Figure 9.** The AHA’s effort in modeling ideal cardiovascular health. There are many ways that ideal cardiovascular health can be modeled in the marketplace at the local community level and internationally. Within the model, there are different levels of engaging the public through Web-based and “app” technology; wearable devices; and corporate, clinic, and faith-based modules to create a cloud-based system of improving cardiovascular health. AHA indicates the American Heart Association; BP, blood pressure; EMR, electronic medical record; FQHC, Federally Qualified Health Center; Gov’t, government; H&W, health and wellness; HH, heart healthy; MLC, My Life Check; Phys Act, physical activity.

Big Data. The AHA can help ensure that Big Data is leveraged in the pursuit of world-class science and its clinical application. The AHA can provide a science-based framework for input into any patient health repository with regard to critical information on patients’ cardiovascular health.

Using Big Data capabilities, the AHA can assist in matching patients to clinical trials and can offer patients gateways to leverage their personal information in pursuit of scientific discovery. The AHA can leverage Big Data when presenting information to patients through a technical environment focused on patient support and improved patient information. The AHA can assist in the assessment of patient monitoring technologies (both devices and software) and promote the growth and linkage of new technologies and patient care.

In the area of scientific research, The AHA could not only promote the use of Big Data but also encourage research and advocacy focused specifically on leveraging knowledge about the capabilities afforded by a Big Data-rich environment. In summary, the AHA role as patient advocate needs to embrace the use of Big Data in the treatment of individual patients, in the advancement of science, and in the development of future technologies.

### What Possible Roles Might the AHA Play in the Basic Science Domain?

The AHA has been a leader in cardiovascular basic science research for the past 6 decades. Accordingly, the AHA is taking a leadership role to provide educational and how-to sessions to support AHA investigators in embracing the digital transformation of cardiovascular medicine as the next logical step to ensure that research investment is sustainable and will have long-lasting impact.

By creating an AHA community digital ecosystem and investing in digital technologies to engage AHA investigator communities, organizations, and writing groups, the AHA will continue on its path to a successful digital transformation that will ultimately support the AHA’s mission and its role as a global leader. By building a knowledge base connecting and aggregating AHA data sets, databases, Web services, and eventually relevant EMR information, secondary analyses can be conducted from basic research to clinical and population investigations. Establishing AHA data science policies and guidelines will help examine and address AHA-relevant data science issues with respect to ethics, privacy, intellectual property, and administration at an early stage to ensure success of AHA operations in the digital era.

## What Possible Roles Might the AHA Play in the Clinical Investigation Domain?

AHA should take the lead in bridging the theoretical promise of Big Data for clinical investigations and the practical applications of such data. This could include funding for demonstration projects that show how EMRs and other Big Data sources work in a variety of settings and environments.<sup>107</sup> Monitoring disease events and/or medical care in rural areas or in regions with poor resources, for example, may present significant practical barriers to getting high-quality data from these sources.

Clinical investigators using Big Data should question the quality and validity of the data they use and interpret. The AHA should provide leadership in efforts to demonstrate the quality and validity of these data sources.<sup>108</sup> Guidelines regarding avoidance of selection bias and enhancing data accuracy should be created and disseminated. The old rule of “garbage in, garbage out” still applies, even to Big Data.

The AHA should consider how it could assist the training of junior clinical and population scientists in a wide variety of uses of Big Data for clinical investigations.<sup>109</sup> This could include focused work on designing future Big Data sources and applications and perhaps consideration of student competitions and acknowledgment programs to facilitate the Big Data training initiative.

The AHA could consider expanding its efforts to support open access to clinical data sources. This could include search engines that allow owners of Big Data (eg, epidemiological data, trial data, biosamples) to list their data sources, elements, and other important metadata features so that researchers could easily find opportunities to use these sources.

## What Possible Roles Might the AHA Play in the Population Science and Epidemiology Domain?

Population scientists look to the AHA to pursue innovation, scientific rigor, and new technology while focusing on engaging patients and the public to increase public health.<sup>37</sup> The AHA has an important role to provide resources for multidisciplinary investigators at all career stages to broadly access science across the AHA portfolio.

An important step that would facilitate access would be the creation of a data scientist fellowship through which top-tier data scientists would be funded and mentored by AHA multidisciplinary scientists. Another major function of the AHA is to help scientists along the clinical–translational continuum to translate and educate the public, clinicians, government, and the media about Big Data. A more immediate task for the AHA will be to convene a forum focused on Big Data harmonization and validation that is inclusive of multiple stakeholders and career levels.

## What Possible Roles Might the AHA Play in the Clinicians and Healthcare Systems Domain?

Making Big Data useful in clinical care requires engagement of multiple stakeholders who come from different perspectives and who may not understand the needs and perspectives of others. The AHA has the opportunity to convene stakeholders with different backgrounds and expertise in workshops and conferences to overcome barriers to using Big Data in clinical care. New health applications, for example, can generate vast amounts of data on physiological parameters, but these data will not be useful in the healthcare system unless they can be readily shared with clinicians, summarized in useful formats, and stored in EMRs. An AHA-sponsored workshop could bring together the stakeholders to make these data useful in clinical care.

Large quantities of data of potential clinical value will soon be collected by different systems, each with its own format. These data cannot be readily shared if they are stored in different formats, but standards for data sharing could be developed. The AHA could help develop standard formats for cardiovascular and stroke data that could be used by different application developers, vendors, and testing companies to facilitate data being shared and stored.<sup>110</sup>

Large quantities of data need to be distilled into manageable amounts of useful information before they can be used in clinical care. Better tools for data analysis, reduction, and visualization would allow clinicians to apply new information at the point of care. The AHA could support research into development of tools that bridge the gap between research and clinical care.

## What Possible Roles Might the AHA Play in the Industry Domain?

The AHA could serve to help better understand the “digital divide” in terms of what it means for public health. The growing area of health technology and PGHD is an area of research opportunity that is just beginning to be explored. Providing mechanisms to aggregate different databases could allow researchers to overcome the limitations of a single data set (eg, statistical power, sample characteristics). Even the collection of passive data, such as environment and geospatial information, could be used to enhance clinical trials.

Ensuring that researchers better understand the sampling methods and variables in existing data sets is also critical. The AHA could facilitate appropriate data availability to a larger community of researchers by serving as an “honest broker.” This role could also provide researchers with assistance in accessing data sets by clarifying data use requirements. Last, as the largest nonprofit funder of CVD and stroke research in the United States, further research funding could be targeted to stimulate the use of existing data sets, as is being done with CVGPS.

## What Possible Roles Might the AHA Play in the Federal Regulatory Authority Domain?

Data quality and creating a fair playing field are essential regulatory elements of using and learning from Big Data. The AHA can act as a convener for standards for tools, methodology, terminology, and appropriate use of registry information for Big Data analysis. Developing common terminology and other data standards will preserve evidence standards for new uses and comparative effectiveness so Big Data can lead to changes in patient solutions and new hypotheses.

The AHA can also aid in addressing the current challenges with informed patient consent. Ensuring that patients fully understand the consent details is difficult, given the length and breadth of some consent forms. The possibility of broad consent for future research is perhaps an area that the AHA can aid in facilitating. Data that can now be collected can come to use in research before there is a clear vision of exactly how best to use it. Methodology and validation of data, integration into current system, ease of use and security and control by patients can all be tested; however, prevention of data misuse by lay users or researchers who cherry pick data to meet their needs is a potential future issue. Regulations for data use must ensure that users have a general research plan and questions to ensure the occurrence of data mining rather than fishing expeditions.

## Summary and Conclusion

There was a consensus across stakeholder domains that Big Data holds great promise for revolutionizing the way research is conducted and clinical care is delivered; however, there is a clear need for the creation of a vision of how to use Big Data to achieve the desired goals. Important areas that are high priorities for further study and discussion include description of the methodology of how to acquire and analyze findings from Big Data, how to validate the veracity of discoveries from Big Data research, and how to integrate Big Data into the investigative and clinical care aspects of the future of

cardiovascular and stroke medicine. Potential roles that the AHA might consider include facilitating discussion of standards (eg, tools, methodology, appropriate use of data generated by other investigators), providing education (eg, healthcare providers, patients, investigators), and helping foster the development of a digital ecosystem in cardiovascular and stroke science. This ecosystem should be interoperable and needs to interface with the rapidly growing digital object environment of the modern-day healthcare system.

The following suggestions were made for future conferences that might address some of the issues:

- A detailed technical discussion of Big Data that focuses on data types, data miners, data managers, and data analytics
- More discussion of the impact of EMRs on clinical care, healthcare systems, third-party payors, and multiple segments of our society with consideration of the ethics and safeguards against unintended consequences of the pervasive use of Big Data
- Consideration of how Big Data will provide new ecological measurements, new definitions of outcomes in clinical trials, and the implications for technology development and regulatory approval.

## Acknowledgments

The writing group wishes to thank the following for their contributions to the Summit planning, rich in-person discussion and review of this report: Michael H. Barbouche; Barbara A. Bowman, PhD; William J. Bryant, JD, LL.M.; Gregory L. Burke, MD, MSc, FAHA; Beth Croll, CMP; J. Michael Gaziano, MD, MPH; Meighan Girgus Vafa, MBA; Sarah M. Greene, MPH; Joseph A. Hill, MD, PhD, FAHA; Pat Hinton; Mark A. Hlatky, MD, FAHA; Yosef Khan, MD, MPH, PhD, MACE; Maria Kontaridis, PhD; Michael Levy, MSc, MBA; Joanne Odenkirchen, MPH; Jeffrey Olgin, MD; Sudip S. Parikh, PhD; Mitesh S. Patel, MD, MBA, MS; Rose Marie Robertson, MD, FAHA; Wayne D. Rosamond, PhD, MS, FAHA; John J. Ryan, M.D., FAHA; Jeffrey L. Saver, MD, FAHA; Laura M. Sol; Selena Smith; Jennifer E. Van Eyk, PhD, FAHA; Gayle R. Whitman, PhD, RN, FAHA; Bram Zuckerman, MD.

## Disclosures

### Writing Group Disclosures

Writing Group Member	Employment	Research Grant	Other Research Support	Speakers' Bureau/Honoraria	Expert Witness	Ownership Interest	Consultant/Advisory Board	Other
Elliott M. Antman	Brigham & Women's Hospital	None	None	None	None	None	None	None
Emelia J. Benjamin	Boston University School of Medicine	NIH/NHLBI <sup>†</sup>	None	None	None	None	Circulation Editorial Board <sup>†</sup>	None

Continued

Writing Group Disclosures Continued

Writing Group Member	Employment	Research Grant	Other Research Support	Speakers' Bureau/Honoraria	Expert Witness	Ownership Interest	Consultant/Advisory Board	Other
Mary Ann Bauman	INTEGRIS Health, Inc	None	None	None	None	None	None	None
Nancy Brown	American Heart Association	None	None	None	None	None	None	None
Vincent Bufalino	Advocate Healthcare Cardiology	None	None	None	None	None	None	None
Robert M. Califf	US FDA	None	None	None	None	None	None	None
Mark A. Creager	Brigham and Women's Hospital	None	None	None	None	None	None	None
Alan Daugherty	University of Kentucky	None	None	None	None	None	None	None
David L. Demets	University of Wisconsin	None	None	None	Marino Law Office <sup>†</sup>	None	Actelion Pharmaceutical*; Abbvie*; Duke University*; GSK*; Population Health Research Institute*; Teva Pharmaceutical*; Amgen*; Sanofi*; Dalgen*	None
Bernard P. Dennis	Dennis Associates, LLC	None	None	None	None	None	None	None
Shahram Ebadollahi	IBM Healthcare	None	None	None	None	None	None	None
Robert A. Harrington	Stanford University	None	None	None	None	None	None	None
Steven R. Houser	Temple University School of Medicine	None	None	None	None	None	None	None
Mariell Jessup	University of Pennsylvania	None	None	None	None	None	None	None
Michael S. Lauer	NHLBI	None	None	None	None	None	None	None
Bernard Lo	The Greenwall Foundation	None	None	None	None	None	None	None
Calum A. MacRae	Harvard Medical School	None	None	None	None	None	None	None
Michael V. McConnell	Stanford University Medical Center	GE Healthcare <sup>†</sup> ; Apple, Inc <sup>†</sup>	None	None	None	None	Google Life Sciences*; Open mHealth*	None
Alexa T. McCray	Harvard University	None	None	None	None	None	None	None
Michelle M. Mello	Stanford University Law School and School of Medicine	The Greenwall Foundation <sup>†</sup>	None	None	None	None	None	None

Continued

Writing Group Disclosures Continued

Writing Group Member	Employment	Research Grant	Other Research Support	Speakers' Bureau/Honoraria	Expert Witness	Ownership Interest	Consultant/Advisory Board	Other
Eric Mueller	Microsoft	None	None	None	None	None	None	None
Jane W. Newburger	Boston Children's Hospital/Harvard Medical School	Bristol Myer Squibb*; Pfizer*	None	None	None	None	Bristol Myer Squibb*; Merck*	None
Sally Okun	patientslikeme	None	None	None	None	None	None	None
Milton Packer	University of Texas Southwestern Medical Center	None	None	None	None	None	None	None
Eric D. Peterson	Duke Clinical Research Institute	None	None	None	None	None	None	None
Anthony Philippakis	Venture Partners	Intel <sup>†</sup>	None	None	None	Google Ventures <sup>†</sup> ; Google <sup>†</sup> ; Element Science <sup>†</sup>	None	Google Ventures <sup>†</sup> ; Fidelity <sup>†</sup>
Peipei Ping	UCLA School of Medicine	None	None	None	None	None	None	None
Prad Prasoon	American Heart Association	None	None	None	None	None	None	None
Véronique L. Roger	Mayo Clinic	None	None	None	None	None	None	None
Steve Singer	Accreditation Council for Continuing Medical Education (ACCME)	None	None	None	None	None	None	None
Robert Temple	FDA, Center for Drug Evaluation & Research	None	None	None	None	None	None	None
Melanie B. Turner	American Heart Association	None	None	None	None	None	None	American Heart Association <sup>†</sup>
Kevin Vigilante	Booz Allen Hamilton	None	None	None	None	None	None	None
John Warner	UT Southwestern Medical Center	None	None	None	None	None	None	None
Patrick Wayte	American Heart Association	None	None	None	None	None	None	None

This table represents the relationships of writing group members that may be perceived as actual or reasonably perceived conflicts of interest as reported on the Disclosure Questionnaire, which all members of the writing group are required to complete and submit. A relationship is considered to be "significant" if (1) the person receives \$10 000 or more during any 12-month period, or 5% or more of the person's gross income; or (2) the person owns 5% or more of the voting stock or share of the entity, or owns \$10 000 or more of the fair market value of the entity. A relationship is considered to be "modest" if it is less than "significant" under the preceding definition.

\*Modest.

<sup>†</sup>Significant.

## References

- Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA*. 2014;311:2479–2480.
- Jameson JL, Longo DL. Precision medicine—personalized, problematic, and promising. *N Engl J Med*. 2015;372:2229–2234.
- Mandl KD, Kohane IS. Federalist principles for healthcare data networks. *Nat Biotechnol*. 2015;33:360–363.
- Mega JL, Sabatine MS, Antman EM. Population and personalized medicine in the modern era. *JAMA*. 2014;312:1969–1970.
- Joyner MJ, Paneth N. Seven questions for personalized medicine. *JAMA*. 2015;314:999–1000.
- Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H, Finkelstein R, Fisher M, Gendelman HE, Golub RM, Goudreau JL, Gross RA, Gubitza AK, Hesterlee SE, Howells DW, Huguenard J, Kelner K, Koroshetz W, Krainc D, Lazic SE, Levine MS, Macleod MR, McCall JM, Moxley RT III, Narasimhan K, Noble LJ, Perrin S, Porter JD, Steward O, Unger E, Utz U, Silberberg SD. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012;490:187–191.
- The Robert Wood Johnson Foundation. Data for Health. Learning What Works. A Report From the Data for Health Advisory Committee. Available at: <http://www.rwjf.org/content/dam/farm/reports/reports/2015/rwjf418628>. Accessed April 6, 2015.
- Selby JV, Krumholz HM, Kuntz RE, Collins FS. Network news: powering clinical research. *Sci Transl Med*. 2013;5:182 fs113.
- Topol E. *The Patient Will See You Now: The Future of Medicine is in Your Hands*. New York, NY: The Perseus Books Group; 2015.
- Alston C, Paget L, Halvorson G, Novelli B, Guest J, McCabe P, Hoffman K, Koepke C, Simon M, Sutton S. *Communicating With Patients on Health Care Evidence*. Washington, DC: Institute of Medicine of the National Academies; 2012.
- Grajales F, Clifford D, Loupos P, Okun S, Quattrone S, Simon M, Wicks P, Henderson D. *Social networking sites and the continuously learning health system: a survey. Discussion Paper*. 2014. Institute of Medicine, Washington, DC. Available at: <http://nam.edu/wp-content/uploads/2015/06/SharingHealthData.pdf>. Accessed October 26, 2015.
- McAfee A, Brynjolfsson E. Big data: the management revolution. *Hav Bus Rev*. 2012;90:60–66, 68, 128.
- Brooks D. What Big Data Can't Do. New York Times Op-Ed February 18, 2013. Available at: <http://www.nytimes.com/2013/02/19/opinion/brooks-what-data-cant-do.html>. Accessed May 5, 2015.
- The Wall Street Journal. Tools to Track Your Health. Available at: <http://www.wsj.com/articles/tools-to-track-your-health-1432834250>. Accessed June 30, 2015.
- Stanley K. Design of randomized controlled trials. *Circulation*. 2007;115:1164–1169.
- Antman E, Weiss S, Loscalzo J. Systems pharmacology, pharmacogenetics, and clinical trial design in network medicine. *Wiley Interdiscip Rev Syst Biol Med*. 2012;4:367–383.
- Fordyce CB, Roe MT, Ahmad T, Libby P, Borer JS, Hiatt WR, Bristow MR, Packer M, Wasserman SM, Braunstein N, Pitt B, DeMets DL, Cooper-Arnold K, Armstrong PW, Berkowitz SD, Scott R, Prats J, Galis ZS, Stockbridge N, Peterson ED, Califf RM. Cardiovascular drug development: is it dead or just hibernating? *J Am Coll Cardiol*. 2015;65:1567–1582.
- Antman E. Data sharing in research: benefits and risks for clinicians. *BMJ*. 2014;348:g237.
- Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Despres JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner MB; on behalf of the American Heart Association Statistics Committee, Stroke Statistics Subcommittee. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation*. 2015;131:e29–e322.
- Roger VL, Boerwinkle E, Crapo JD, Douglas PS, Epstein JA, Granger CB, Greenland P, Kohane I, Psaty BM. Strategic transformation of population studies: recommendations of the working group on epidemiology and population sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. *Am J Epidemiol*. 2015;181:363–368.
- Roger VL, Boerwinkle E, Crapo JD, Douglas PS, Epstein JA, Granger CB, Greenland P, Kohane I, Psaty BM. Respond to “future of population studies”. *Am J Epidemiol*. 2015;181:372–373.
- Lauer MS. Time for a creative transformation of epidemiology in the United States. *JAMA*. 2012;308:1804–1805.
- Vasan RS, Folsom AR. Invited commentary: future of population studies—defining research priorities and processes. *Am J Epidemiol*. 2015;181:369–371.
- McGovern L, Miller G, Hughes-Cromwick P. Health Policy Brief: The Relative Contribution of Multiple Determinants to Health Outcomes. *Health Affairs*, August 21, 2014. Available at: [http://healthaffairs.org/healthpolicybriefs/brief\\_pdfs/healthpolicybrief\\_123.pdf](http://healthaffairs.org/healthpolicybriefs/brief_pdfs/healthpolicybrief_123.pdf). Accessed October 26, 2015.
- IBM. IBM Watson Health. Available at: <http://www.ibm.com/smarterplanet/us/en/ibmwatson/health/>. Accessed July 2, 2015.
- Lindsey ML, Mayr M, Gomes AV, Delles C, Arrell DK, Murphy AM, Lange RA, Costello CE, Jin Y-F, Laskowitz DT, Sam F, Terzic A, Van Eyk J, Srinivas PR; on behalf of the American Heart Association Council on Functional Genomics and Translational Biology, Council on Cardiovascular Disease in the Young, Council on Clinical Cardiology, Council on Cardiovascular and Stroke Nursing, Council on Hypertension, and Stroke Council. Transformative impact of proteomics on cardiovascular health and disease: a scientific statement from the American Heart Association. *Circulation*. 2015;132:852–872.
- Robinson PN, Kohler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008;83:610–615.
- Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, Pendergrass SA, Xu H, Hindorf LA, Li R, Manolio TA, Chute CG, Chisholm RL, Larson EB, Jarvik GP, Brilliant MH, McCarty CA, Kullo IJ, Haines JL, Crawford DC, Masys DR, Roden DM. Systematic comparison of phenotype-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31:1102–1110.
- Dean J, Ghemawat S. Mapreduce: Simplified data processing on large clusters. Presented at: OSDI '04, Sixth Symposium on Operating System Design and Implementation; December 6–8, 2004; San Francisco, CA.
- Mell P, Grance T. The NIST definition of cloud computing. 2011. Available at: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>. Accessed April 26, 2015.
- Rackspace Support Network. Moving your Infrastructure to the Cloud: How to Maximize Benefits and Avoid Pitfalls. Available at: [http://www.rackspace.com/knowledge\\_center/whitepaper/moving-your-infrastructure-to-the-cloud-how-to-maximize-benefits-and-avoid-pitfalls](http://www.rackspace.com/knowledge_center/whitepaper/moving-your-infrastructure-to-the-cloud-how-to-maximize-benefits-and-avoid-pitfalls). Accessed August 7, 2015.
- Amarasingham R, Patzer RE, Huesch M, Nguyen NQ, Xie B. Implementing electronic health care predictive analytics: considerations and challenges. *Health Aff (Millwood)*. 2014;33:1148–1154.
- Centers for Medicare & Medicaid Services. Data and Program Reports: 2014. Available at: <http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/DataAndReports.html>. Accessed August 12, 2015.
- Hawgood S, Hook-Barnard IG, O'Brien TC, Yamamoto KR. Precision medicine: beyond the inflection point. *Sci Transl Med*. 2015;7:300ps17.
- National Research Council. *Toward Precision Medicine: Building a Knowledge Network of Biomedical Research and a New Taxonomy of Disease*. Washington, DC: The National Academies Press; 2011.
- Ashley EA. The precision medicine initiative: a new national effort. *JAMA*. 2015;313:2119–2120.
- Khoury MJ, Evans JP. A public health perspective on a national precision medicine cohort: balancing long-term knowledge generation with early health benefit. *JAMA*. 2015;313:2117–2118.
- National Institutes of Health. National Library of Medicine (NLM) Working Group. Available at: <http://acd.od.nih.gov/reports/Report-NLM-06112015-ACD.pdf>. Accessed June 11, 2015.
- President Obama's Precision Medicine Initiative. Available at: <https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative>. Accessed May 12, 2015.
- Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, Carlson-Donohoe H, Gropman A, Pierson TM, Golas G, Wolfe L, Groden C, Godfrey R, Nehrebecky M, Wahl C, Landis DM, Yang S, Madeo A, Mullikin JC, Boerkoel CF, Tiffit CJ, Adams D. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet Med*. 2012;14:51–59.
- Gahl WA, Boerkoel CF, Boehm M. The NIH Undiagnosed Diseases Program: bonding scientists and clinicians. *Dis Model Mech*. 2012;5:3–5.
- McCray AT, Trevett P, Frost HR. Modeling the autism spectrum disorder phenotype. *Neuroinformatics*. 2014;12:291–305.

43. McCray AT. Better access to information about clinical trials. *Ann Intern Med*. 2000;133:609–614.
44. IOM (Institute of Medicine). *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. Washington, DC: The National Academies Press; 2015.
45. Chute CG, Ullman-Cullere M, Wood GM, Lin SM, He M, Pathak J. Some experiences and opportunities for big data in translational research. *Genet Med*. 2013;15:802–809.
46. Grande D, Mitra N, Shah A, Wan F, Asch DA. The importance of purpose: moving beyond consent in the societal use of personal health information. *Ann Intern Med*. 2014;161:855–862.
47. NIH policy supports broader sharing of genomic data, strengthens informed-consent rules: research participants must give consent for secondary sharing, even if data are de-identified. *Am J Med Genet A*. 2015;167a:viii–ix.
48. Drazen JM. Sharing individual patient data from clinical trials. *N Engl J Med*. 2015;372:201–202.
49. The AllTrials campaign. All Trials Registered, All Trials Reported. September 2013. Available at: <http://www.alltrials.net/wp-content/uploads/2013/09/What-does-all-trials-registered-and-reported-mean.pdf>. Accessed April 6, 2015.
50. Anderson ML, Chiswell K, Peterson ED, Tasneem A, Topping J, Califf RM. Compliance with results reporting at ClinicalTrials.gov. *N Engl J Med*. 2015;372:1031–1039.
51. NIH Data Sharing Policy. Available at: [http://grants.nih.gov/grants/policy/data\\_sharing/](http://grants.nih.gov/grants/policy/data_sharing/). Accessed May 12, 2015.
52. Bollinger JM, Scott J, Dvoskin R, Kaufman D. Public preferences regarding the return of individual genetic research results: findings from a qualitative focus group study. *Genet Med*. 2012;14:451–457.
53. National Institutes of Health. National Heart, Lung, and Blood Institute. Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). Available at: <https://biolincc.nhlbi.nih.gov/home/>. Accessed May 29, 2015.
54. National Institutes of Health. National Heart, Lung, and Blood Institute. Database of Genotypes and Phenotypes (dbGaP). Available at: <http://www.ncbi.nlm.nih.gov/gap>. Accessed May 29, 2015.
55. Paltoo DN, Rodriguez LL, Feolo M, Gillanders E, Ramos EM, Rutter JL, Sherry S, Wang VO, Bailey A, Baker R, Caulder M, Harris EL, Langlais K, Leeds H, Luetkemeier E, Paine T, Roomian T, Tryka K, Patterson A, Green ED. Data use under the NIH GWAS data sharing policy and future directions. *Nat Genet*. 2014;46:934–938.
56. Antman E. Saving and improving lives in the information age: presidential address at the American Heart Association 2014 Scientific Sessions. *Circulation*. 2015;131:2238–2242.
57. Neubeck L, Lowres N, Benjamin EJ, Freedman SB, Coorey G, Redfern J. The mobile revolution—using smartphone apps to prevent cardiovascular disease. *Nat Rev Cardiol*. 2015;12:350–360.
58. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40:181–188.
59. UK Biobank. Available at: <http://www.ukbiobank.ac.uk/physical-activity-monitor/>. Accessed May 29, 2015.
60. University of California, San Francisco. Health eHeart. Available at: <https://www.health-eheartstudy.org/>. Accessed May 29, 2015.
61. Stanford Medicine. MyHeart Counts. Available at: <http://myheartcounts.stanford.edu/>. Accessed May 29, 2015.
62. Wikipedia. Baseline Study. Available at: [http://en.wikipedia.org/wiki/Baseline\\_Study](http://en.wikipedia.org/wiki/Baseline_Study). Accessed May 29, 2015.
63. National Institutes of Health. Precision Medicine Initiative. Available at: <http://www.nih.gov/precisionmedicine/>. Accessed May 29, 2015.
64. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372:793–795.
65. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA*. 2015;313:625–626.
66. Brooks GC, Vittinghoff E, Iyer S, Marcus GM, Pletcher MJ, Olgin JE. Diagnostic accuracy of a smartphone based six-minute walk test. *Circulation*. 2014;130:A17496.
67. Open mHealth. Available at: <http://www.openmhealth.org/>. Accessed May 29, 2015.
68. Field D, Sansone SA, Collis A, Booth T, Dukes P, Gregurick SK, Kennedy K, Kolar P, Kolker E, Maxon M, Millard S, Mugabushaka AM, Perrin N, Remacle JE, Remington K, Rocca-Serra P, Taylor CF, Thorley M, Tiwari B, Wilbanks J. Megascience. ‘Omics data sharing. *Science*. 2009;326:234–236.
69. American Heart Association (AHA). Open Science Policy Statements for AHA Funded Research. Available at: [http://my.americanheart.org/professional/Research/FundingOpportunities/Open-Science-Policy-Statements-for-AHA-Funded-Research\\_UCM\\_461225\\_Article.jsp](http://my.americanheart.org/professional/Research/FundingOpportunities/Open-Science-Policy-Statements-for-AHA-Funded-Research_UCM_461225_Article.jsp). Accessed June 11, 2015.
70. National Institutes of Health. Principles and Guidelines for Reporting Preclinical Research. Available at: <http://www.nih.gov/about/reporting-preclinical-research.htm>. Accessed April 6, 2015.
71. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature*. 2012;483:531–533.
72. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10:712.
73. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, Plon SE, Ramos EM, Sherry ST, Watson MS. ClinGen—the Clinical Genome Resource. *N Engl J Med*. 2015;372:2235–2242.
74. Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical trials. *JAMA*. 2014;311:2381–2382.
75. Califf RM. Integrated efficacy to effectiveness trials. *Clin Pharmacol Ther*. 2014;95:131–133.
76. Bram JT, Warwick-Clark B, Obeysekare E, Mehta K. Utilization and monetization of healthcare data in developing countries. *Big Data*. June 2015;3:59–66.
77. Bufalino V, Bauman MA, Shubrook JH, Balch AJ, Boone C, Vennum K, Bradley S, Wender RC, Minners R, Arnett D. Evolution of “the guideline advantage”: lessons learned from the front lines of outpatient performance measurement. *Circ Cardiovasc Qual Outcomes*. 2014;7:493–498.
78. Bufalino VJ, Masoudi FA, Stranne SK, Horton K, Albert NM, Beam C, Bonow RO, Davenport RL, Girgus M, Fonarow GC, Krumholz HM, Legnini MW, Lewis WR, Nichol G, Peterson ED, Rumsfeld JS, Schwamm LH, Shahian DM, Spertus JA, Woodard PK, Yancy CW. The American Heart Association’s recommendations for expanding the applications of existing and future clinical registries: a policy statement from the American Heart Association. *Circulation*. 2011;123:2167–2179.
79. US Department of Health and Human Services. Office of the National Coordinator for Health Information Technology. Health IT Safety Center Roadmap—Collaborate on solutions, Informed by evidence. Available at: <http://www.healthitsafety.org/uploads/4/3/6/4/43647387/roadmap.pdf>. Accessed July 23, 2015.
80. Adler NE, Stead WW. Patients in context—EHR capture of social and behavioral determinants of health. *N Engl J Med*. 2015;372:698–701.
81. Burke LE, Ma J, Azar KMJ, Bennett GG, Peterson ED, Zheng Y, Riley W, Stephens J, Shah SH, Suffoletto B, Turan TN, Spring B, Steinberger J, Quinn CC; on behalf of the American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. *Circulation*. 2015;132:1157–1213.
82. Steinhilbl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? *JAMA*. 2013;310:2395–2396.
83. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. *N Engl J Med*. 2014;371:372–379.
84. U.S. Department of Health and Human Services. Health Information Technology for Economic and Clinical Health Act (HITECH). Section 13402 (e)(4)—Notification in the case of breach, Methods of Notice, Posting on HHS Public Website. Available at: <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveridentities/hitechact.pdf>. Accessed May 12, 2015.
85. Rothstein MA. Is deidentification sufficient to protect health privacy in research? *Am J Bioeth*. 2010;10:3–11.
86. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. *Inform Health Soc Care*. 2015;1–19. [Epub ahead of print].
87. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among Caucasian, Filipino, Korean, and Latino Americans: the digital link to health survey. *JMIR Mhealth Uhealth*. 2014;2:e43.
88. Mello MM, Franer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for responsible sharing of clinical trial data. *N Engl J Med*. 2013;369:1651–1658.

89. The White House. Precision Medicine Initiative: Proposed Privacy and Trust Principles. July 8, 2015. Available at: [https://www.whitehouse.gov/sites/default/files/docs/pmi\\_privacy\\_and\\_trust\\_principles\\_july\\_2015.pdf](https://www.whitehouse.gov/sites/default/files/docs/pmi_privacy_and_trust_principles_july_2015.pdf). Accessed July 10, 2015.
90. Department of Health and Human Services, Code of Federal Regulations, Protection of Human Subjects, 45 CFR 46. 2009.
91. Mitka M. US registry for sudden death in the young launched by the NIH and CDC. *JAMA*. 2013;310:2495.
92. Accreditation Council for Continuing Medical Education (ACCME®) 2014 Annual Report. 2015. Available at: [http://www.accme.org/sites/default/files/2014\\_Annual\\_Report\\_20150707\\_1.pdf](http://www.accme.org/sites/default/files/2014_Annual_Report_20150707_1.pdf). Accessed July 13, 2015.
93. Klass D. A performance-based conception of competence is changing the regulation of physicians' professional behavior. *Acad Med*. 2007;82:529–535.
94. Cho MK, Magnus D, Constantine M, Lee SS, Kelley M, Alessi S, Korngiebel D, James C, Kuwana E, Gallagher TH, Diekema D, Capron AM, Joffe S, Wilfond BS. Attitudes toward risk and informed consent for research on medical practices: a cross-sectional survey. *Ann Intern Med*. 2015;162:690–696.
95. PoliticoPro. Congress calls for re-examination of HIPAA. Available at: <https://www.politicopro.com/go/?wbid=56665>. Accessed July 9, 2015.
96. Lantos JD. U.S. research regulations: do they reflect the views of the people they claim to protect? *Ann Intern Med*. 2015;162:731–732.
97. Open PHACTS. Available at: <https://www.openphacts.org/> Accessed July 10, 2015.
98. Neuroscience Information Framework (NIF). NIF Standard Ontology. Available at: <http://bioportal.bioontology.org/ontologies/NIFSTD>. Accessed August 7, 2015.
99. Data Observation Network for Earth (DataONE). Available at: <https://www.dataone.org/> Accessed July 10, 2015.
100. Force11. Guiding Principles for Findable, Accessible, Interoperable and Reusable Data Publishing version b1.0. Available at: <https://www.force11.org/node/6062>. Accessed June 30, 2015.
101. Kohane IS. HEALTH CARE POLICY. Ten things we have to do to achieve precision medicine. *Science*. 2015;349:37–38.
102. Hicks KA, Tchong JE, Bozkurt B, Chaitman BR, Cutlip DE, Farb A, Fonarow GC, Jacobs JP, Jaff MR, Lichtman JH, Limacher MC, Mahaffey KW, Mehran R, Nissen SE, Smith EE, Targum SL. 2014 ACC/AHA key data elements and definitions for cardiovascular endpoint events in clinical trials: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Cardiovascular Endpoints Data Standards). *Circulation*. 2015;132:302–361.
103. Subramanian S, Dumont C, Dankert C, Wong A. Personalized technology will upend the doctor-patient relationship. *Harv Bus Rev*. June 19 2015. <https://hbr.org/2015/06/personalized-technology-will-upend-the-doctor-patient-relationship>. Accessed July 2, 2015.
104. U.S. Food and Drug Administration. Optimizing FDA's Regulatory Oversight of Next Generation Sequencing Diagnostic Tests—Preliminary Discussion Paper. Available at: <http://www.fda.gov/downloads/MedicalDevices/NewsEvents/WorkshopsConferences/UCM427869.pdf>. Accessed April 6, 2015.
105. Angus DC. Fusing randomized trials with big data: the key to self-learning health care systems? *JAMA*. 2015;314:767–768.
106. Evans BJ, Burke W, Jarvik GP. The FDA and genomic tests—getting regulation right. *N Engl J Med*. 2015;372:2258–2264.
107. Banchs JE, Scher DL. Emerging role of digital technology and remote monitoring in the care of cardiac patients. *Med Clin North Am*. 2015;99:877–896.
108. Fox CS, Hall JL, Arnett DK, Ashley EA, Delles C, Engler MB, Freeman MW, Johnson JA, Lanfear DE, Liggett SB, Lusic AJ, Loscalzo J, MacRae CA, Musunuru K, Newby LK, O'Donnell CJ, Rich SS, Terzic A. Future translational applications from the contemporary genomics era: a scientific statement from the American Heart Association. *Circulation*. 2015;131:1715–1736.
109. Musunuru K, Hickey KT, Al-Khatib SM, Delles C, Fornage M, Fox CS, Frazier L, Gelb BD, Herrington DM, Lanfear DE, Rosand J. Basic concepts and potential applications of genetics and genomics for cardiovascular and stroke clinicians: a scientific statement from the American Heart Association. *Circ Cardiovasc Genet*. 2015;8:216–242.
110. Hendel RC, Bozkurt B, Fonarow GC, Jacobs JP, Lichtman JH, Smith EE, Tchong JE, Wang TY, Weintraub WS. ACC/AHA 2013 methodology for developing clinical data standards: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards. *Circulation*. 2014;129:2346–2357.