Special Report

Report of the National Heart, Lung, and Blood Institute Working Group
An Integrated Network for Congenital Heart Disease Research

Sara K. Pasquali, MD, MHS; Jeffrey P. Jacobs, MD; Gregory K. Farber, PhD; David Bertoch, MHA; Elizabeth D. Blume, MD; Kristin M. Burns, MD; Robert Campbell, MD; Anthony C. Chang, MD; Wendy K. Chung, MD, PhD; Tiffany Riehle-Colarusso, MD, MPH; Lesley H. Curtis, PhD; Christopher B. Forrest, MD, PhD; William J. Gaynor, MD; Michael G. Gaies, MD, MPH; Alan S. Go, MD; Paul Henchey, MS; Gerard R. Martin, MD; Gail Pearson, MD, ScD; Victoria L. Pemberton, RN, MS; Steven M. Schwartz, MD; Robert Vincent, MD; Jonathan R. Kaltman, MD

Abstract—The National Heart, Lung, and Blood Institute convened a working group in January 2015 to explore issues related to an integrated data network for congenital heart disease research. The overall goal was to develop a common vision for how the rapidly increasing volumes of data captured across numerous sources can be managed, integrated, and analyzed to improve care and outcomes. This report summarizes the current landscape of congenital heart disease data, data integration methodologies used across other fields, key considerations for data integration models in congenital heart disease, and the short- and long-term vision and recommendations made by the working group. (Circulation. 2016;133:1410-1418. DOI: 10.1161/CIRCULATIONAHA.115.019506.)

Key Words: database [publication type] heart diseases outcome assessment, health care

As medicine moves into the era of big data, it is important to develop a common vision for how the rapidly increasing volume of data will be managed, integrated, and analyzed to improve care and outcomes. This holds true across a variety of different disciplines and specialties, including the field of congenital heart disease (CHD). Only through coordinated efforts will the CHD community be able to fully leverage available and emerging data sources to support important investigations and to conduct research most efficiently. To facilitate this process, the National Heart, Lung, and Blood Institute (NHLBI) convened a working group meeting in January 2015 in Bethesda, MD, to explore issues related to CHD data integration. The goals of the working group were to develop a vision for an integrated data network to support CHD research and to identify critical elements and potential barriers. The working group consisted of experts in pediatric and adult cardiology, cardiothoracic surgery, health services and outcomes research, epidemiology, informatics, and statistics.

The Era of Big Data
The past several years have been characterized by what has been called an era of big data. During this time, the volume, velocity, and variety of data captured across numerous sources have increased exponentially, outpacing traditional techniques for managing and analyzing data. Newer data platforms, computing capabilities, and analytic techniques have been developed to better manage, integrate, analyze, and provide more real-time feedback to various industries about their data, with the goal of optimizing performance and outcomes.

For example, the automotive industry is capturing data generated...
Medicine initiatives. Both of these programs involve recent launching of Big Data to Knowledge and Precision as further evidence of this, the National Institutes of Health value of leveraging the increasing volume of available data. Trends have led to a greater recognition in medicine of the importance of data captured electronically in health care have increased exponentially in recent years, including data captured in the electronic health record (EHR), clinical registries, research data sets, monitoring systems, and other sources. With this has come the recognition that analyzing and integrating these data sets can expand the range of questions that can be answered. For example, early results suggest that integrating continuous data streams generated by various monitoring systems with clinical outcomes data may enable better prediction and treatment of adverse events in intensive care settings. Second, along with this increase in availability of data, there has been a simultaneous decrease in funding to support biomedical research. This has led to further interest in improving our understanding of how to leverage available data to power research more efficiently, for example, using existing registries as platforms to support clinical trials with the goal of reducing time and costs associated with data collection. Finally, the current national emphasis in health care on improving quality and optimizing healthcare value has necessitated the analysis and integration of healthcare quality and cost data across a variety of sources to understand the landscape of care delivery and outcomes, to investigate relationships between quality and cost, and to develop strategies for improvement. These and other recent trends have led to a greater recognition in medicine of the value of leveraging the increasing volume of available data. As further evidence of this, the National Institutes of Health recently launched the Big Data to Knowledge and Precision Medicine initiatives. Both of these programs involve efforts to integrate information across a variety of sources to conduct research more efficiently and to improve care.

**Data in Medicine**

Historically in medicine and in the hierarchy of medical research, the value of databases, registries, and other data sources in the cycle of scientific discovery has not always been recognized. Mining data sets and database research have often been seen as lesser pursuits compared with basic science research or clinical trials. However, several developments have begun to change the way data in medicine are viewed. First, similar to other fields, the volume and granularity of data captured electronically in health care have increased exponentially in recent years, including data captured in the electronic health record (EHR), clinical registries, research data sets, monitoring systems, and other sources. With this has come the recognition that analyzing and integrating these data sets can expand the range of questions that can be answered. For example, early results suggest that integrating continuous data streams generated by various monitoring systems with clinical outcomes data may enable better prediction and treatment of adverse events in intensive care settings. Second, along with this increase in availability of data, there has been a simultaneous decrease in funding to support biomedical research. This has led to further interest in improving our understanding of how to leverage available data to power research more efficiently, for example, using existing registries as platforms to support clinical trials with the goal of reducing time and costs associated with data collection. Finally, the current national emphasis in health care on improving quality and optimizing healthcare value has necessitated the analysis and integration of healthcare quality and cost data across a variety of sources to understand the landscape of care delivery and outcomes, to investigate relationships between quality and cost, and to develop strategies for improvement. These and other recent trends have led to a greater recognition in medicine of the value of leveraging the increasing volume of available data. As further evidence of this, the National Institutes of Health recently launched the Big Data to Knowledge and Precision Medicine initiatives. Both of these programs involve efforts to integrate information across a variety of sources to conduct research more efficiently and to improve care.

**Current Landscape of CHD Data**

**Data Sources, Infrastructure, and Collaboration**

The current CHD data environment has many assets (Table 1). Numerous existing clinical registries, administrative/billing databases, public health surveillance databases, research data sets, and other sources contain a wealth of important information that can be used to facilitate research, surveillance, and quality improvement activities in the field. In addition, data are increasingly being captured via a variety of newer modalities, including the EHR, data generated from medical monitors and devices, and genetic and biomarker data. Some centers are also beginning to capture longer-term outcomes data such as quality of life and neurodevelopmental outcomes, as described in more detail in subsequent sections. Most congenital heart programs across the United States have existing local infrastructure to support data collection for various registries and other data sets, and infrastructure to integrate data across centers exists as a part of several national registries, multicenter quality improvement activities, and research efforts. Finally, there is an environment of collaboration among investigators and many congenital heart programs related to participation in these efforts, including the Pediatric Heart Network, National Pediatric Cardiology Quality Improvement Collaborative, and Pediatric Cardiac Critical Care Consortium (PC4), among many others. Annual meetings of the Multi-societal Database Committee for Pediatric and Congenital Heart Disease further aid in facilitating sharing of ideas and collaboration across different registries and databases.

**Standardized Nomenclature**

Another particularly important aspect of the CHD data landscape (Table 1) has been the major effort over the past 2 decades to develop a standardized nomenclature system. In the 1990s, both the European Association for Cardio-Thoracic Surgery and the Society of Thoracic Surgeons (STS) created databases to assess congenital heart surgery outcomes and established the International Congenital Heart Surgery Nomenclature and Database Project. By 2000, a common nomenclature and common core minimal data set were adopted by both the European Association for Cardio-Thoracic Surgery and STS Congenital Heart Surgery databases. Subsequently, the International Society for Nomenclature of Pediatric and Congenital Heart Disease was formed, and its Nomenclature Working Group cross-mapped the nomenclature developed by the European Association for Cardio-Thoracic Surgery and STS with the European Pediatric Cardiac Code of the Association for European Pediatric Cardiology, thereby creating the International Pediatric and Congenital Cardiac Code (IPCCC), which is available for free download at http://www.IPCCC.NET. The IPCCC is currently used by multiple databases that span the spectrum of pediatric and congenital cardiac care, including the following:

- Cardiac surgery (STS Congenital Heart Surgery Database, European Association for Cardio-Thoracic Surgery Congenital Heart Surgery Database, Japan Congenital Cardiovascular Surgery Database, and United Kingdom Central Cardiac Audit Registry)
- Cardiology (The IMProving Pediatric and Adult Congenital Treatment [IMPACT] Registry of the National Cardiovascular Data Registry of the American College of Cardiology)
- Anesthesia (the Joint Congenital Cardiac Anesthesia Society–STS Congenital Cardiac Anesthesia Database)
- Critical care (PC4 and Virtual Pediatric Intensive Care Unit System Databases)
Investigators have previously linked data from outpatient pediatric registries. However, data are not collected in many databases as a result of a variety of factors. Although linkages based on direct or unique identifiers are the easiest way to merge data sets, these identifiers are often limited. This has led to the development of indirect identifiers.23

**Linking on Indirect Identifiers**

Indirect identifiers include variables such as date of birth, date of admission, date of discharge, sex, and center where hospitalized. It has been shown that nearly all records at a given congenital heart center can be uniquely identified with a combination of these indirect identifiers and that a crosswalk can then be created between 2 data sets, linking patients on the values of center where hospitalized and the indirect identifiers.23

This methodology has been used in the pediatric cardiovascular population to merge information from a large clinical registry (The STS Congenital Heart Surgery Database) with a pediatric administrative data set (Pediatric Health Information System Database).1,13,21-26 Linking these 2 data sets allows the use of the detailed operative and outcomes data from the clinical registry and the resource use data from the administrative data set. The current linked data set includes records from >60,000 children undergoing congenital heart surgery at 33 different hospitals from 2004 to 2010, with expansion and updating of the data set underway. Several comparative-effectiveness studies and analyses of healthcare costs have been successfully conducted with this data set, leveraging variables from both data sources to facilitate analyses not otherwise possible with either data set independently.24-26 Similar methodology has also been used to merge clinical trial data from the Pediatric Heart Network Single Ventricle Reconstruction Trial with data from the Children’s Hospital Association Case Mix data set to perform integrated analyses of clinical outcomes and costs.27

Combinations of indirect and direct identifier linkage methodologies have also been used. For example, data sets may use indirect methods to link to a common data set and then use the common data set to identify unique individuals. Using these methods, surveillance data from the Metropolitan Atlanta Congenital Defects Program were indirectly linked to vital records, as were data from the Special Education Database of Metropolitan Atlanta. Then deterministic (direct) linkage was done between the 2 data sets to evaluate the use of special education services among children with CHD.28

**Center-Level Linkages**

Linking registry data to other independently collected center-level data through matching on center can be easily accomplished (Table 2). For example, independently collected survey data on intensive care unit care models and nursing education and staffing levels have been successfully linked to the STS Congenital Heart Surgery Database.1,13,29 These linkages enable evaluation of the variables collected in the survey in relation to outcomes data collected in the registry.

<table>
<thead>
<tr>
<th>Method</th>
<th>Examples of Linked Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linking on unique identifiers</td>
<td>Clinical data + survival data</td>
</tr>
<tr>
<td>Linking on indirect identifiers</td>
<td>Registry + administrative/cost data</td>
</tr>
<tr>
<td>Center-level linkages</td>
<td>Clinical trial + administrative/cost data</td>
</tr>
<tr>
<td>Collaboration/partnering between databases</td>
<td>Hospital survey data + registry data</td>
</tr>
<tr>
<td>Supplementary data modules to main registry</td>
<td>Registries with shared platforms, variables, definitions</td>
</tr>
</tbody>
</table>

Table 2. CHD Data Integration Activities to Date

CHD indicates congenital heart disease.
central data coordinating center, along with surgical data for participating centers. This approach is likely more time- and cost-efficient than creating a separate anesthesia database in which many of the fields on patient characteristics and the operative procedure would have been duplicated. Related efforts are underway to incorporate electrophysiology data within the IMPACT Registry. An alternative method involves a more distributed approach with sharing of data definitions and variables between data sets and organizations, information technology solutions allowing single entry of data at the local level, and subsequent submission and distribution of both shared and unique data variables to applicable national data sets and data coordinating centers. An example of this is the shared IPCCC data definitions for certain variables between the STS, PC4, and IMPACT registries.

**Supplementary Data Modules**

Methods have also been developed to create data modules enabling efficient collection of supplemental data points to the primary data source (Table 2). The modules are generally Web based and can be quickly created and deployed to allow real-time collection of additional data needed to answer important research questions that may arise. This methodology has been recently successfully used by PC4 to collect supplemental data to the main registry to study the relationship between vasoactive-inotropic score and outcome after infant cardiac surgery. A module allowing capture of additional data related to inotrope use was created, deployed, and linked to the main registry data for each patient. This facilitated efficient data collection with 391 infants prospectively enrolled across 4 centers in just 5 months.

**Powering Clinical Trials Through Registries**

In recent years, it has become increasingly recognized that many variables of interest for prospective investigation, including clinical trials, are already being captured within clinical registries across an engaged group of sites on a routine basis. It has been proposed that leveraging these existing registry data (and linkage with modules containing additional study-specific data when necessary) may be a more efficient way to power research, avoiding duplicate data collection and minimizing study start-up timelines. To date, these methods have been used to facilitate 2 clinical trials in adult cardiovascular medicine: the Study of Access Site for Enhancement of PCI for Women (SAFE-PCI) and Thrombus Aspiration During ST-Segment Elevation Myocardial Infarction (TASTE) trials, both of which leveraged existing information in 2 different cardiac catheterization registries. In the CHD community, work is currently underway within the Pediatric Heart Network to assess the completeness and accuracy of local clinical registry data at sites and to determine whether prospective studies and clinical trials may leverage these data to minimize duplicate data collection and to promote greater efficiency.

**Current CHD Data Limitations**

Although a great deal of progress has been made over the past several years to better integrate and leverage available CHD data sources to more efficiently conduct research, many limitations remain Table 1). As described in the preceding sections, most current methods to integrate data have involved 1:1 linkages of a certain data set to another to answer a specific question. Regulatory and contracting issues have generally prevented the use of these integrated data sets to answer additional research questions after the primary study is completed. Broader and more comprehensive strategies for data integration across numerous sources are lacking, and the landscape still consists primarily of individual data silos. There is relatively limited capability from an information technology perspective to broadly share information across data sets, and data governance and collaboration models have yet to be developed across congenital heart centers and national organizations to facilitate such data sharing. Methodology to allow the use of existing data for more efficient prospective studies and clinical trials is just beginning to be developed in the field. Issues concerning consent and confidentiality also require further evaluation. Finally, important longitudinal outcomes information is lacking, and simply linking existing data sets together will not address this issue.

**Data Integration Models Across Other Fields**

Several models supporting more comprehensive data integration exist across other fields and may be useful examples for the CHD community to consider.

**National Institute of Mental Health and National Database of Autism Research**

To link information across studies enrolling patients with autism, the National Institute of Mental Health uses a global unique identifier (GUID). The GUID is a unique code generated in a manner to protect confidentiality. To generate a GUID, a combination of patient-identifying information (typically name at birth, sex, city of birth, and date of birth) is entered into a software program at the local site. Through 1-way encryption, these elements are translated into hash codes, which cannot be traced back to the patient. The hash codes are then sent to a central data server, which generates a GUID. The GUID is returned to the site, where it is entered into the data set for that patient. The GUID protects privacy in 2 ways: The identifiable information never leaves the local site, and the GUID cannot be traced back to the patient. Another important feature of the GUID is that the patient or family does not need to remember it. Patients and families need to remember only the individual data elements described above, and from these, the same GUID is generated each time, regardless of location or timing of enrollment in a study. Thus, the GUID functions to uniquely identify the patient, is composed of information easily known to the patient that is invariant over a lifetime, and maintains patient privacy. GUIDs are currently being piloted in the field of CHD for patients enrolled in studies conducted by the Pediatric Cardiac Genomic Consortium.

**PEDSnet**

PEDSnet, funded by the Patient Centered Outcomes Research Institute, is a clinical data research network consisting of 8 of the nation’s largest pediatric academic medical centers, 2 existing pediatric consortia/quality improvement collaboratives,
and 2 national data partners (Express Scripts, a national pharmacy benefits management company, and IMS Health, a data aggregator of multipayer claims data). The goal of PEDSNet is to create a learning health system that integrates research into routine care settings and supports structured data capture and quality improvement processes to rapidly implement advances in new knowledge. To date, PEDSNet has harmonized data from 4.5 million patients captured across the EHR systems of member site using a common terminology and uses open-source software to support data submission and aggregation. Analyses are done primarily at a centralized data coordinating center, although distributed queries across sites may also be possible (as discussed in further detail in subsequent sections) for certain research questions.

The Cardiovascular Research Network

The Cardiovascular Research Network consists of 15 geographically distributed healthcare delivery systems caring for >10 million members that was established through initial funding by the NHLBI to conduct large-scale adult cardiovascular research more efficiently, including epidemiological studies, outcomes research, comparative-effectiveness studies, and clinical trials. Within the Cardiovascular Research Network, data captured through the EHR of each healthcare system and multiple other electronic databases are linked at the site level with the use of medical record numbers. These data include clinical and resource use data across inpatient, emergency department, and outpatient settings; procedures; diagnoses; inpatient and outpatient pharmacy data; and laboratory test results. Data capture and architecture are standardized across the virtual data warehouse at each site with common data elements, naming conventions, and definitions to facilitate combining information in aggregate analyses. Recent Cardiovascular Research Network efforts within the CHD population have involved developing natural language text processing algorithms to attempt to identify patients with CHD from unstructured EHR data across multiple health systems.

Mini-Sentinel Program of the US Food and Drug Administration

The goal of this program is to facilitate active surveillance and monitoring of the safety of medical products across the United States. Mini-Sentinel uses a distributed data approach in which participating data partners maintain physical and operational control over their own data. A common data model was designed to meet the needs of the program, and each participating organization developed a process to extract, transform, and load its source data, applying the common data model, to create the distributed database. These data are then analyzed with programs developed centrally and executed locally by participating organizations.

Potential Data Integration Models in CHD

Several different integration models may be considered to support CHD research that build on the existing models used across other fields. Two specific models that have received the most attention to date and their strengths and weaknesses in terms of the needs of the CHD community are discussed below.

Creation of a CHD GUID and Data Linkages at the National Level

One option to support data integration in CHD may build on the work done by the autism research community described in the preceding sections. This could involve creation of a CHD GUID and collaboration among researchers, professional societies, and other groups to share and merge data sets containing these identifiers at the national level. Potential advantages of this approach include its previous success within the autism research community, the ability for multiple linkages, and ability to maintain privacy in that direct identifiers are not sent outside of local sites. However, there are also certain disadvantages to consider. Some of the data elements needed to generate a GUID in its current form are not necessarily found in the medical record and require direct patient contact. Although this may be feasible for certain research studies, there would be several difficulties to consider within the current data collection infrastructure of most large registries and data sets for which there generally is not direct patient interaction to capture data elements, so additional local personnel or modification of data collection workflow would be required. There may also be issues related to consent to consider. In addition, to enable linkages, not only must the GUID be generated and incorporated into individual data sets, but professional societies and other organizations must agree to collaborate and share their data sets for linkage and analysis. Negotiating the various data sharing and governance policies of multiple professional societies and different organizations, which often have a focus primarily on adult cardiology and cardiac surgery, and current policies in some cases prohibiting sharing of data outside of central data coordinating centers may prove to be challenging.

Supporting Local Linkages and a Distributed Data Network Model

An alternative option involves building on the experience of the Cardiovascular Research Network, PEDSNet, and the Mini-Sentinel program to support data linkages at the local level and sharing of these integrated data across heart centers, creating a distributed data network in CHD. Local data linkages are feasible because most often research and registry data reside locally at the institution of each participant site, in addition to being aggregated into larger multicenter data sets. Local linkages are relatively easy to perform because direct or unique identifiers are readily available in these data sets and in the EHR. Merged local data sets can then be deidentified, and groups of institutions or heart centers can collaborate to share and aggregate information at a central site for analysis. Alternatively, data may be kept at each site, and standard algorithms can be developed to query and analyze the data locally, with results aggregated and combined across sites, similar to the Mini-Sentinel approach. This model addresses some of the limitations identified with the use of GUIDs and makes linked information available for both local purposes.
and aggregate research. Data linkages within congenital heart center data warehouses are already taking place at several centers. For example, at 1 center participating in the working group, local data linkages have supported improved accuracy in determining surgical-site infection rates (through merging infection data with CHD clinical registry data), which has in turn aided in efforts to reduce these infections. Furthermore, as discussed in previous sections, sharing data across heart centers for multicenter research is already a common practice, but the methodology could leverage these existing collaborative relationships.

**Additional Considerations**

In addition to strategies for integrating existing data, there are several other related areas of consideration. As described in previous sections, data on longer-term outcomes remain very limited, and efforts to promote efficient collection of these data have just recently begun. Preliminary work suggests that engaging with patients and families directly can allow successful capture of critical longitudinal outcomes data such as survival, rehospitalizations (particularly those that occur at institutions other than the surgical center), and important aspects of quality of life and burden of disease. For example, standardized patient-reported outcomes data have been successfully captured across 2 heart centers participating in the working group on >2000 patients to date. Methods are being developed to further the use of Web portals, mobile technology, and social media to allow more efficient and widespread capture of patient-reported data.

In addition, the Cardiac Neurodevelopmental Outcomes Consortium is working toward developing methods to capture standardized information obtained during neurodevelopment follow-up clinic visits. It will be important to incorporate these emerging longer-term outcomes data into the overall strategy developed for data integration.

A second area of consideration relates to the EHR. It has been hypothesized that leveraging EHR data can provide improved efficiency and reduce data collection burden for various registries or research data sets. However, these efforts will require additional work to improve the quality and standardization of data currently contained in the EHR. Although certain types of structured data may be efficiently captured through the EHR (eg, laboratory values, medications), other data critical to CHD research may be more difficult to capture due to the lack of granularity in the EHR and associated coding schemes (eg, detailed information on anatomic diagnoses and procedures) and the lack of standardized definitions (eg, for preoperative comorbidities or postoperative complications).

Finally, with the expansion in the number and types of data sets and opportunities for linkages, it remains important to consider several key factors concerning data collection and analyses in general to ensure that research conducted with these data sets is meaningful. These include issues related to accuracy and completeness of data, standardization (or lack thereof) of data elements and definitions, and availability of variables within the data set to perform appropriate risk adjustment or adjustment for differences in case mix across hospitals. The availability and use of linked or integrated data sources do not lessen the importance of these critical factors.

**Short- and Long-Term Vision and Recommendations**

The working group outlined a vision for how data might be integrated in the CHD community and developed several recommendations to achieve that vision over the short and long term.

**Future Vision**

The working group acknowledged that our current conceptualization of data and data management is largely outdated. In addition, the volume and granularity of available data will continue to increase through more widespread capture of genomic and biomarker data and real-time physiological data, for example. Our current databases, data structure, and analytics will be insufficient for the task of managing and understanding these data. The working group recommended that further collaboration and consultation with data scientists and experts from diverse fields and industries outside of medicine will be important in understanding and incorporating modern data storage, manipulation, and analytic techniques into short- and long-term data solutions for the CHD community. For example, 3-dimensional graphing techniques of large volumes of data have been shown recently in other industries to identify patterns that would otherwise not be apparent; however, these and other novel techniques have had limited application in medical research to date. The importance of considering these techniques within the context of medical decision making was also discussed because simple associations found in the data do not necessarily indicate cause and effect, and ensuring accuracy, reliability, and integrity of the data will continue to be important concepts regardless of the data storage and integration techniques used.

**Short-Term Goals**

The working group recommended that the CHD community take steps in the near term to facilitate more comprehensive integration of information across currently available CHD data sources. The group felt that this could support further research that could not be conducted with individual data sets alone and could help to promote efficiency in research. The group recommended that, of the data integration strategies discussed in previous sections, methodology supporting local data linkages and a distributed data network across collaborating sites seemed to meet more needs of the CHD community compared with alternative approaches. The group acknowledged work in this area already taking place across several congenital heart center data warehouses, as described in previous sections.

The working group recommended further efforts toward developing such a data network in CHD through engaging interested sites; developing a common data model, strategies for integration, and data governance policies; identifying potential funding sources; and conducting pilot studies to better understand the value of data integration and to demonstrate proof of concept. In addition, the working group recognized the lack of important longitudinal outcomes data in the field and acknowledged that this issue cannot be addressed with linkages of current data sources alone. Further development and testing of strategies to support efficient capture of
longer-term outcomes data that may be merged with existing data were recommended, and the working group recognized ongoing efforts in this area described in the preceding sections. Exploring potential funding for data integration efforts was also recommended, including funding through the Pediatric Heart Network, other NHLBI funding opportunities, the Patient Centered Outcomes Research Institute, and foundation and philanthropic opportunities.

Data Standardization
The working group acknowledged the importance of data standardization to facilitate data pooling and analysis and to decrease data entry burden at sites. The working group recognized the IPCC terminology as the standard nomenclature within the field and recommended that IPCC terminology and definitions be incorporated into all relevant data sources when possible, including CHD registries, clinical data and the EHR, and research data sets, and recognized ongoing work to incorporate IPCC terms into the International Classification of Diseases coding system. The working group recommended that consensus across sites and other stakeholders should be developed to standardize data collection in other areas such as the collection of longitudinal follow-up information and neurodevelopmental outcomes and recognized recent work in this area by the Cardiac Neurodevelopmental Outcomes Consortium and others, as described in the preceding sections.

Align Goals With Stakeholder Interests
Moving toward more fully integrated data systems will require engagement with multiple stakeholders, including hospital systems, researchers, national organizations and professional societies, and patients and families. To make a case for change, the field will need to discuss and identify the value of data integration from the perspectives of multiple stakeholders. At the hospital level, one such value point is the optimization of strategic investments already made in the EHR and clinical registries. Data integration will provide a more complete picture of the care provided and thus may enable improved quality, reduction in errors, and increased value. Patients and families are also critical stakeholders, and it will be important for data integration activities to seek to align with their interests, which may include improving care and outcomes, participating in longitudinal patient-reported data collection activities, and engaging in the process of determining quality improvement and research priorities.

Conclusions
Several concurrent trends have provided the opportunity to recalibrate our approach to data collection, integration, and analytics in biomedical research. In the CHD community, several advances in recent years, including nomenclature standardization, development of rich clinical data sets, an environment of multicenter collaboration, and implementation of several data integration techniques, have provided a strong foundation for future work. There is now a need for further integration and collaboration to meet present and future challenges and to develop a more efficient and comprehensive research enterprise to improve the care and outcomes of patients with CHD.

Acknowledgments
Working group members: Sara K. Pasquali, MD, MHS (University of Michigan C.S. Mott Children’s Hospital), Jeffrey P. Jacobs, MD (Johns Hopkins University, All Children’s Heart Institute), Gregory K. Farber, PhD (National Institute of Mental Health), David Bertoch, MHA (Children’s Hospital Association), Elizabeth D. Blume, MD (Boston Children’s Hospital), Kristin M. Burns, MD (NHLBI), Robert Campbell, MD (Emory University), Anthony C. Chang, MD, MBA, MPH (Children’s Hospital of Orange County), Wendy K. Chung, MD, PhD (Columbia University), Tiffany Riehle-Colarusso, MD, MPH (Centers for Disease Control and Prevention), Lesley H. Curtis, PhD (Duke Clinical Research Institute), Sherry Farr, PhD (Centers for Disease Control and Prevention), Christopher B. Forrest, MD, PhD (Children’s Hospital of Philadelphia), William J. Gaynor, MD (Children’s Hospital of Philadelphia), Michael G. Gaies, MD, MPH (University of Michigan C.S. Mott Children’s Hospital), Alan S. Go, MD (Kaiser Permanente Northern California), Paul Henchey, MS (ArborMetrix, Inc), Gerard R. Martin, MD (Children’s National Medical Center), Gail Pearson, MD, ScD (NHLBI), Victoria L. Pemberton, RN, MS (NHLBI), Steven M. Schwartz, MD (The Hospital for Sick Children), Mario Stylianou, PhD (NHLBI), Robert Vincent, MD (Emory University), and Jonathan R. Kaltman, MD (NHLBI).

Sources of Funding
This working group was funded by the NHLBI. The views and conclusions expressed in this report are those of the authors and do not necessarily represent the official position of the NHLBI, the National Institutes of Health, or the Centers for Disease Control and Prevention.

Disclosures
Dr Pasquali receives funding from the NHLBI and the Janette Ferrantino Professorship related to CHD research. Dr Pasquali is a member of the STS Congenital Heart Surgery Database Taskforce, directs the PC*- Data Coordinating Center, and leads the Integrated CARDiac Data and Outcomes Collaborative within the NHLBI-sponsored Pediatric Heart Network. Dr Jacobs is chair of the STS National Database Workforce and Congenital Heart Surgeons Society Committee on Quality Improvement and Outcomes; PC*- executive committee member; and American College of Cardiology Improving Pediatric and Adult Congenital Treatment Registry Steering Committee Member. Dr Farber manages the National Institute of Mental Health data archives. D. Bertoch is Vice President of Comparative Data and Informatics, Children’s Hospital Association. Dr Blume is the co-principal investigator for NHLBI contract HHSN268200548198C for the Interagency Registry for Mechanically Assisted Circulatory Support. Dr Chung receives funding from the NHLBI related to CHD research and is a member of the Pediatric Cardiac Genomics Consortium and Pediatric Cardiomyopathy Registry. Dr Curtis receives funding from the US Food and Drug Administration as the data core lead for the Mini-Sentinel Program. Dr Forrest is the principal investigator of Pedsnet (National Pediatric Learning Health System), funded by the Patient-Centered Outcomes Research Institute; chair of the PCORnet Research Committee; and chair of the PEPR Steering Committee (Validation of PROMIS Pediatric Measures in Chronic Disease Consortium). Dr Gaynor is a member of the STS Congenital Heart Surgery Database Taskforce and Public Reporting Committee and the PC*- Executive Committee. Dr Gaies receives funding from the NHLBI related to congenital heart disease research and is the executive director of the PC*. Dr Go receives research funding from the Patient-Centered Outcomes Research Institute and is the principal investigator of the PORTAL Network Congenital Heart Defect Cohort. P. Henchey is an employee of ArborMetrix, Inc. Dr Schwartz is a member of the PC*- Executive.
Committee and the Pediatric Heart Network Executive Committee. Dr Vincent is the Steering Committee chair, American College of Cardiology Improving Pediatric and Adult Congenital Treatment Registry, and PC¹ Executive Committee member. The other authors report no conflicts.

References


2. Why “big data” is a big deal. http://harvardmagazine.com/2014/03/why-


S1047951118002813.


pediatrics.2015-0259.


Report of the National Heart, Lung, and Blood Institute Working Group: An Integrated Network for Congenital Heart Disease Research


_Circulation_. 2016;133:1410-1418
doi: 10.1161/CIRCULATIONAHA.115.019506

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/133/14/1410

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/