

Data Harmonization for Patient-Centered Clinical Research— A Workshop

In the context of work by the Patient Centered Outcomes Research Institute (PCORI) to foster establishment of a national patient-centered clinical research network (PCORnet), on October 31–November 1, 2013, the Institute of Medicine’s (IOM’s) Roundtable on Value & Science-Driven Health Care convened a workshop on issues and approaches to inter-network data harmonization. The workshop brought together experts in the areas of informatics, clinical research, patient data networks, and research data networks to reflect on past efforts and current state-of-the-art approaches to data harmonization; help identify barriers and challenges; share best practices; and identify options for progress in accelerating data interoperability among PCORnet awardees and between awardees and existing national data sharing activities. This workshop in brief summarizes the major themes of the workshop presentations and discussions.

The PCORnet program is intended to improve the nation’s capacity to conduct clinical research efficiently by creating a large, national, representative, patient-centered clinical research network—a network of networks—for conducting clinical outcomes research. The creation of such a broad network of networks could advance the nation’s evolution toward continuously learning health care and allow for large-scale research to be conducted with improved accuracy and efficiency. PCORnet will be composed of clinical data research networks (CDRNs), integrated research networks of two or more health care systems, and patient-powered research networks (PPRNs) (research networks of patients and individuals collecting patient-reported data, potentially including online patient communities). The goals of PCORnet are the development of a comprehensive, longitudinal data infrastructure; broader participation of patients, clinicians, health systems, and payers in the research process; and improvements in analytic methods for both observational and experimental clinical research. Additional components of PCORnet include a coordinating center to provide technical and logistical assistance to the component networks and support PCORnet-wide collaboration and communication, a steering committee made up of all awardees and key federal and private stakeholders, and several advisory and expert committees to inform the work of PCORnet.

In phase 1 of the program, beginning in January 2014, PCORI’s network awardees will have 18 months to partner with one another to build the initial research infrastructure. Success factors at the end of the 18-month period will include interoperable data sharing, the ability to conduct high-quality observational studies to support the use of the network of networks by external researchers and external data partners, and readiness for the conduct of large, embedded, pragmatic interventional trials. Given the expected diversity of the PCORI awardee networks with regard to data sources, types, and models of participating organizations, guidance will be needed on how to achieve interoperability quickly while preserving maximum flexibility in terms of future uses of the data.

The IOM workshop included four context-setting presentations from existing research networks and PCORI leadership; small-group discussions focusing on data models and terminology; data harmonization, implementations, and sustainability; strategies for dealing with expanded data sources; and moderated full-group discussions.

Individual workshop participants who spoke identified a few overarching themes and messages. The primary contributors to the points below are identified in parentheses. Statements, recommendations, and opinions expressed are those of individual presenters and participants and are not necessarily endorsed or verified by the Roundtable or the IOM, and they should not be construed as reflecting any group consensus.

Overarching Themes and Messages

- **Meet networks where they are.** Given the likely diversity across PCORnet participants, approaches to data harmonization that accommodate each grantee network's starting point, including entry point, pace, and definition of success, should be considered. The establishment of a common critical path with milestones to assess progress is one approach to ensuring that all grantee networks are moving toward a common goal, while allowing the flexibility to enter at the point and proceed at the pace most appropriate for them. (Jean Slutsky)
- **Capitalize on diversity of networks.** The diversity of the incoming PCORnet networks—from CDRNs comprised of established clinical research networks based in health care delivery organizations to PPRNs that are organic, online patient communities—presents a variety of complementary data types and sources and a rich opportunity to learn. Some networks will bring with them accounts of local successes and failures and refined approaches, while others will bring relatively blank slates, ripe for experimentation and paradigm-challenging ideas for how to proceed. Opportunities to learn from this diversity can be capitalized on to provide an engine for innovation and experimentation. (Charles Friedman, Sharon Terry)
- **Specify the “must-have narrow neck”; make everything else modular.** Drawing on the image of an hourglass, specification of standards focusing on the requirements at the “narrow neck”—the minimal set of components needed to facilitate exchange through a common pathway, while allowing for efficiency and innovation in the broader environment—such as the use of TCP/IP (Transmission Control Protocol/Internet Protocol) to enable the Internet, could be considered. This would include the maintenance of local data models that are optimized for individual network uses. Allow for all decisions to be reversible and no single decision to create irreparable harm by maximizing the use of modular approaches whenever possible. (Doug Fridsma, Charles Friedman)
- **Direct special attention to patient-generated data.** Patient-generated data from portals, social media, and mobile sources will be of increasing importance due to greater adoption of devices and the inclusion of PPRNs, for which the majority of data is expected to be of this type. A lack of standards to accommodate these data presents both a challenge (as an existing gap) and an opportunity (to begin fresh). This difference in likely data sources between CDRNs and PPRNs also suggests that they will face differing data harmonization challenges. (Sally Okun)
- **Maximize transparency.** Specify protocols and processes to ensure transparency for network members and stakeholders at every point in the data harmonization effort. Transparency can be critical to operational effectiveness and efficiency by building trust, allowing for learning from heterogeneous starting points and experiences, and fostering innovation. (Michael Kahn, Greg Simon, Wilson Pace)
- **Leverage existing standards and support institutional priorities.** Acknowledge the multiple competing informatics and financial institutional priorities, and align data harmonization approaches to leverage the efforts and investments made, such as the adoption of Meaningful Use standards. Adherence to widely adopted national standards when available, such as the use of LOINC (Logical Observation Identifiers Names and Codes) and SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms), can improve the consistency and quality of the data throughout PCORnet. (Michael McGinnis)
- **Ground the approach in iteration and validation.** Employ an approach that allows for simple, early wins, learning from successes and failures, and for gradual, increasing complexity as well as the adoption of data, standards, and tools that may be available in the future. Opportunities for assessment and validation of the quality and reliability of data and analytics can be built into every iterative step. (Michael Kahn, Doug Fridsma, Jeff Brown)

Challenges

Working toward the harmonization of data across PCORnet networks is challenging. Workshop discussions included the identification of a wide range of issues that the PCORnet coordinating center and the component CDRN and PPRN networks will face in their efforts to establish interoperable data sharing (with a focus on semantic rather than syntactic interoperability) as they seek to balance goals of allowing for near-term generation of new insights and planning for the long-term vision and capabilities of PCORnet.

Comments by Individual Participants

- **Balancing short- and long-term goals.** Success, Joe Selby from PCORI noted, will be measured by meeting the 18-month deadline for harmonization across phase 1 awardees while making decisions that will not shortchange the long-term potential of the network. Considerations stemming from these parallel goals will be to make decisions that are flexible and extensible enough not to preclude the addition of future PCORnet users, uses, and data sources. Associated challenges will include avoiding short-term and often expensive workarounds to meet near-term goals at the expense of long-term potential.
- **Network diversity.** Sharon Terry from Genetic Alliance noted that there is likely to be significant diversity among the grantee networks in terms of the data types they collect. Given the requirements of CDRNs to include health-delivery partners, they are likely to include clinical data and, potentially, administrative and claims data. PPRNs, on the other hand, have as part of their charge the obligation to advance the collection of patient reported outcomes, which are often collected from sources outside clinical settings. In both cases, data will likely be present in both structured and free-text formats. Diversity across PCORnet networks is also expected in terms of the kinds of staffing and technical expertise at their disposal, as well as the levels and richness of experience with previous data harmonization and data sharing efforts.
- **Industry trends.** PCORnet is designed to leverage institutional data for research to inform and improve health care decisions, Michael McGinnis from the IOM noted. As such, support from the highest leadership levels and integration into health delivery systems will be critical for its success. The U.S. health care delivery industry is currently in the midst of major changes regarding organization, payment, and informatics. Some of these will likely aid in the harmonization of data across networks, but others are likely to compete for attention and resources within delivery systems. Changes include consolidation of hospitals and practices that necessitate the integration of business and clinical data systems; the movement toward shared-risk payment contracts that involve the management of populations across sites; efforts toward better-integrated, team-based care; and the widespread implementation of Meaningful Use standards (see box) and conversion to the use of ICD-10 codes in health care delivery.
- **Data model.** Harmonization of data across PCORnet networks will likely require the selection of one or more common data or information model(s), noted Jeff Brown from Harvard Pilgrim Health Care. Given the diversity and different stages of maturity of the grantee networks, an approach that would allow individual networks to maintain their own local model and map to common PCORnet models is likely to be successful. Several decisions will need to be made in the selection of such common models. Among these are whether to adopt already-established models, adapt existing models, or create new ones. In addition, the kinds of models, including query, clinical, and

Meaningful Use

- Authorized as part of the Health Information Technology for Economic and Clinical Health provisions of the 2009 American Recovery and Reinvestment Act, the Meaningful Use program provides an estimated \$30 billion for staged incentives from the Centers for Medicare & Medicaid Services to eligible professionals and hospitals that demonstrate Meaningful Use of their electronic health records.
- Demonstration of Meaningful Use is determined by the ability to meet a set of objectives, including ability to record certain data types, implement clinical decision support, exchange digital information, and report on a series of clinical quality measures in a standardized way.
- Meaningful Use Stage 3, which is expected to begin in 2016, includes explicit requirements for data exchange using specific national and international terminology standards. The National Library of Medicine has been charged with creating standard value sets to be used in reporting clinical quality measures.
- As of 2013, more than 12 percent of eligible professionals⁹ and more than 42 percent of hospitals⁹ have successfully attested to Meaningful Use Stage 1.

⁹Wright, A., S. Henkin, J. Feblowitz, A. B. McCoy, D. W. Bates, and D. F. Sittig. 2013. Early results of the meaningful use program for electronic health records. *New England Journal of Medicine* 368(8):779-780.

⁹ DesRoches, C. M., D. Charles, M. F. Furukawa, M. S. Joshi, P. Kralovec, F. Mostashari, C. Worzala, and A. K. Jha. 2013. Adoption of electronic health records grows rapidly, but fewer than half of us hospitals had at least a basic system in 2012. *Health Affairs (Millwood)* 32(8):1478-1485.

SOURCE: Adapted from Meaningful Use (http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Meaningful_Use.html) (accessed January 8, 2014).

data typing models, and the level of specification of data, including macro-models and late-binding schema, will need to be decided.

- **Multi-use network.** Building a durable, national resource for large-scale, patient-centered clinical research requires the network to be able to support a wide range of research activities, Rachael Fleurence from PCORI highlighted. Such activities range from casual queries to data mining to generate hypotheses for future research to large-scale sophisticated analyses, including randomized trials. Balancing these diverse goals will be a challenge to the data harmonization effort. Approaches varying from the common-model and distributed-query approach that is employed in several existing networks to approaches that rely on machine learning and data mining, or a combination, could be considered.

Approach

Characterization of these challenges provided the conceptual backbone for the workshop discussion. Workshop participants spent time in small- and whole-group discussions exploring potential approaches to data harmonization across the PCORnet network of networks. This section briefly summarizes related comments by individual workshop participants, which should not be construed to reflect any group consensus. Highlights of these discussions can be found in the box below.

Data Harmonization Approach

- Outline a critical path and milestones for data harmonization (Slutsky)
- Partner with PCORnet participants on approach (Friedman)
- Employ use cases to shape and test harmonization approach (Kahn)
- Specify model and core data elements based on use cases and network goals (Ryan)
- Design data harmonization implementation for rapid-cycle improvement (Fridsma)
- Employ strategic accommodations to data sources and types (Setoguchi-Iwata, Okun)
- Establish the framework for data model governance and maintenance (Pace)
- Implement continuous improvement of data and insights (Simon)
- Develop and execute an evaluation framework (Holve, Humphreys)

Comments by Individual Participants

- **Outline a critical path and milestones for data harmonization.** Jean Slutsky from the Agency for Healthcare Research and Quality highlighted the need to identify a critical path with milestones that all networks must meet in order to be successful at the end of 18 months. Given the diversity of the incoming networks, individual workshop participants suggested that each one should be able to enter the path at the place appropriate for its situation, and the speed of its progress should accommodate its limitations within the 18-month window. Dividing CDRNs and PPRNs into tiers and tailoring strategies to navigate the critical path for each tier is one possible approach. Previously identified milestones can serve as assessment points for evaluating progress of individual networks and PCORnet as a whole.

TABLE Example Use Cases for PCORnet Data Harmonization Effort Discussed in the Workshop, in Increasing Order of Complexity

| Use Case Type | General Description | Specific Example |
|--|--|--|
| 1 Cohort counts across networks | Count of patients satisfying a defined set of inclusion criteria with each network, with counts exchanged to coordinating center | A. Count of unique individuals with diagnosis code(s) for obesity B. Percent of unique individuals with diagnosis code(s) for obesity in 1-year age cohorts to age 20 and 5-year age cohorts in adults |
| 2 Cohort summary statistics across networks | Summary statistics from running an against patient-level datasets at each network | Estimation of 12-month risk of cardiovascular event, adjusting for baseline cardiovascular risk factors (diabetes, hypertension, drug use, etc.) between patients taking one or more weight loss medications versus patients post bariatric surgery |
| 3 Cohort lists for re-identification and recruitment across networks | Lists of subject identifiers satisfying a defined set of inclusion criteria, with SUBJECT_ID values that can be used at network to re-identify patients for purposes of inclusion into future research study/additional data capture, etc. | Send outcomes survey to all unique individuals who had bariatric surgery in past 12 months with no documentation of elevated alanine transaminase (ALT) or aspartate aminotransferase (AST) 12 months prior to surgery followed by elevated ALT or AST documented at least once within the first 6 months post-surgery |
| 4 Expanded data sources for cohorts and outcomes | Comparable data from various “non- traditional” sources (e.g., eDiaries, apps, sensors, social media) that can be included in cross-network data acquisition | Compare functional status, quality of life, missed days of work, self-esteem, unintended consequences, adverse events in obese patients being treated with oral medications, and diet versus bariatric surgery |

SOURCE: Used with permission from Michael Kahn.

- Partner with PCORnet participants on approach.** Charles Friedman from the University of Michigan suggested that the coordinating center organize a multiday meeting within the first month of the award. Purposes for such a meeting could include level setting, team and collaboration building, and education across networks. The goal of the meeting would be discussion and selection of a specific but flexible target goal or set of goals tailored to different network capabilities, to be achieved in the first 6 months of the award, and consideration of subsequent 6-month goals.
- Employ use cases to shape and test harmonization approach.** Several individual workshop participants commented on the need to ground the data harmonization work in real use cases to ensure practicality by testing sample scenarios in the data models under consideration. Criteria discussed for the selection of candidate use cases included focus on a topic that engages the maximum number of networks, such as an obesity cohort, which is required of all CDRNs. Employing a sequence of use cases increasing in difficulty, such as beginning with generating cohort counts and increasing in complexity, was also suggested. Example use cases, introduced by Michael Kahn from University of Colorado, Denver, considered during the workshop discussions can be found in the table above.
- Specify model and core data elements based on use cases and network goals.** A strawman approach to an observational study based on use case number 2 from the table was presented by Patrick Ryan from Janssens Research and Development. By systematically reviewing the use case requirements, the data types needed and the expected sources and formats were identified. This information was used to pick an appropriate existing data model that supported the use case’s data needs, and the next steps were identified as selecting vocabulary

standards for each data domain and establishing common mapping from the sources to the data model standards. Additional considerations for data model selections noted by individual workshop participants included prioritizing the use of existing models and standards, maximizing extensibility, robust metadata and provenance capabilities, and abilities to support both quick-and-dirty queries and large-scale sophisticated analyses, such as randomized trials and prospective observational studies.

- Design data harmonization implementation for rapid-cycle improvement.** Several workshop participants, including Doug Fridsma from the Office of the National Coordinator for Health IT, suggested that the implementation of the data harmonization plan across the component networks should be set up to allow for fast, early failures that enable learning and subsequent iterations. Defining success at each stage of the implementation effort, continuously assessing progress against these goals, and establishing the expectation that failure is part of the normal process can facilitate this approach. Additionally, broad sharing of lessons from both successful and failed attempts across all network members can accelerate subsequent efforts.
- Employ strategic accommodations to data sources and types.** Characterizing current network data sources to inform data harmonization approaches that are tailored to their unique characteristics and plan for future sources was highlighted by individual workshop participants, including Soko Setoguchi-Iwata from Duke University, as an important component of the data harmonization approach. A starting list of current sources includes clinical data systems such as electronic health records and registries; administrative data systems, including claims systems; and data collected directly from patients, including through patient portals, from social media and mobile apps, and from mHealth sources such as remote devices and sensors, as was described by Sally Okun from PatientsLikeMe (see figure below). Data from these sources will likely be



FIGURE Results of a 2013 survey of PatientsLikeMe active members on their experience with insomnia.

SOURCE: Used with permission from PatientsLikeMe.

available in structured and free-text formats; extracting maximum utility from both formats will be important to a successful data harmonization effort.

- **Establish the framework for data model governance and maintenance.** The development of policies and processes to maintain the PCORnet data model(s), update them when needed, and address user issues (e.g., data access and incorporation of new standards and tools) was highlighted by several individual workshop participants, including Wilson Pace from the University of Colorado. The governance structure and processes, it was suggested, should be transparent and include representations from all data stakeholders.
- **Implement continuous improvement of data and insights.** The quality and reliability of the PCORnet data will be critical to the success of the program, because these efforts will reflect on the credibility of generated insights. Individual participants, including Greg Simon from GroupHealth, noted the need to establish processes and policies to continuously assess and improve data quality within individual CDRN and PPRN networks and with an eye to PCORnet-wide harmonization and interoperability. Approaches to this include making every network activity an opportunity to evaluate data quality, engaging local network experts in understanding the specifics of their data, encouraging transparent sharing of metadata on data quality and availability, and making data quality improvement a core tenant of the data harmonization governance.
- **Develop and execute an evaluation framework.** Several individual workshop participants who spoke, including Erin Holve from AcademyHealth, commented on the need to establish a PCORnet-wide evaluation framework and processes in order to ensure consistent interpretation of harmonization mappings and transformations. Betsy Humphreys from the National Institutes of Health suggested employing a prospective study use case as part of this strategy as a way to monitor progress and keep the harmonization effort targeted toward continuous improvement. Several workshop participants highlighted the importance of provenance as a key aspect of assessing semantic interoperability.

Additional considerations

Although the scope of the workshop was limited to providing pragmatic input to aid in data harmonization efforts across PCORnet, discussions often touched on issues that were related and potentially important to the PCORnet effort more widely. This section briefly summarizes these related comments by individual workshop participants, so that they might be considered by PCORI and the broader field of clinical research stakeholders.

Comments by Individual Participants

- **Leverage existing standards to collect data in standardized format at the source.** Several workshop participants, including Becky Kush from the Clinical Data Interchange Standards Consortium, noted that PCORnet could provide an opportunity to improve data consistency and quality by encouraging increased collection in standardized data formats from the point of collection. Mapping after the fact, she stated, often leads to lost information, and there are efforts, including those by the Food and Drug Administration, to encourage standardized data collection for research that should be tested within PCORnet and leveraged whenever possible.
- **Develop a continuously learning evidence-generation system.** Concepts of continuous, rapid assessment and improvement are applicable to all facets of a clinical research effort, including understanding the strengths and weaknesses of analytic methods and factoring this into the interpretation of results and their use for decision making. Martijn Schuemie from Janssens Research and Development highlighted the work that the Observational Medical Outcome Partnership has done to empirically evaluate observational methods and encouraged other research efforts to incorporate this approach into their work.

- **Harmonize analysis and output.** The need to harmonize the analysis and output processes, as well as the structure and content of data, was highlighted by individual workshop participants, including Dixie Baker from Martin, Blanck & Associates. Components of this effort include harmonizing the feasibility and quality assessments and modular, reusable approaches to analysis programs. Additionally, the need for network governance to ensure transparency in the sharing of summary results to allow for collective interpretation and establishment of overall conclusions was emphasized.
- **Consider an indexing approach.** Some workshop participants, including Nigam Shah from Stanford University, encouraged PCORnet stakeholders to think beyond the “traditional” approaches to clinical data networks and consider approaches more aligned with “big data” to better facilitate activities beyond hypothesis testing, such as discovery analytics, data mining, and the development of conditional predictions.

The PCORI PCORnet will be an important contribution to the development of a research infrastructure to power comparative-effectiveness research across diverse settings and continuously learning health care. It can accelerate progress toward reliable real-time knowledge generation from health data through interoperable sharing for clinical research among the component networks. Interoperable data sharing is an integral part of moving toward a clinical data utility to serve the knowledge-generation needs of a continuously learning health system. The discussions captured in this summary highlight the challenges and unique opportunities that the PCORnet coordinating center and component networks face in working toward the ambitious data harmonization goals for phase 1.

Terms and Acronyms

Administrative data: records of services, including patient registration information, billing codes and changes, and insurance transactions from hospitals, clinics, medical offices, pharmacies, and labs (Iezzoni, 1997).

CDRNs: Clinical Data Research Networks, one of two types of networks that comprise the PCORI National Patient-Centered Clinical Research Network, which are comprised of established or newly developed networks that involve two or more health care systems, with plans to function as integrated research network (Selby et al., 2013).

Common data model: the data content and process specifications to achieve, use, and refine functional data harmonization and interoperability among systems contributing data to defined activities. For example, the Virtual Data Warehouse model is the data model employed by the HMO Research Network to organize data from diverse data partners in a consistent way (Ogunyemi et al., 2013).

Continuously learning health system: a system in which science, informatics, incentives, and culture are aligned for continuous improvement and innovation—with best practices seamlessly embedded in the care process, patients and families active participants in all elements, and new knowledge captured as an integral byproduct of the care experience (IOM, 2012).

Data domain: the type of data included in a portion of a data model, such as “demographic” (Ogunyemi et al., 2013).

Data element: an atomic or indivisible unit of data, such as a blood pressure value (Ogunyemi et al., 2013).

Data exchange: protocol-determined transfer of data from one organization to another, such as the Direct Project messaging protocols created by the Office of the National Coordinator for Health IT to specify a simple, secure, scalable, standards-based way for participants to send authenticated, encrypted health information directly to known, trusted recipients over the Internet (ONC, 2014a).

Data harmonization: the process of standardizing data element definitions for core elements from heterogeneous sources that are most critical to the conduct of reliable clinical research and effective care and that can interface with related data from clinical diagnosis and treatment activities. This can require, for example, the mapping of data from individual sources to a common data model for patient care and evidence generation (Liu et al., 2010).

Data interoperability: the ability of different computer systems to communicate and digitally exchange data that can be seamlessly, automatically, and reliably used for research and to support individual and population health (Ogunyemi et al., 2013).

- **Syntactic interoperability:** the ability of two or more computer systems to communicate and exchange data without losing or distorting interpretive content (Ogunyemi et al., 2013).
- **Semantic interoperability:** the ability of two or more computer systems to automatically interpret the information exchanged meaningfully and accurately to produce reliable and useful results, as defined by the end users of both systems (Ogunyemi et al., 2013).

Data query: specially written software code sent to a database in order to get information back from the database. For example, querying a database for the patient IDs of all those having had bariatric surgery in the past 12 months (Ogunyemi et al., 2013).

Electronic health record: a digital version of a patient’s health care information that generally contains a patient’s medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiologic images, and laboratory and test results; it may also allow access to evidence-based tools that providers can use to make decisions about a patient’s care and streamline provider workflow (ONC, 2014c).

Experimental clinical research: clinical research that involves variation under the control of the investigator, conducted in accordance with research regulations (Gordis, 2009).

Governance model: a set of policies and procedures that stipulate the participants interacting around a common interest and the rules that govern the processes of their interaction (Deloitte, 2013).

HITECH: Health Information Technology for Economic and Clinical Health, a provision of the 2009 American Recovery and Reinvestment Act.

ICD-10: the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list developed and maintained by the World Health Organization. In health data harmonization, the ICD-10 codes are one way of categorizing diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or disease used primarily for insurance billing (WHO, 2014).

Meaningful use standards: the set of standards defined by the Centers for Medicare & Medicaid Services Incentive Programs that governs the use of electronic health records and allows eligible providers and hospitals to earn incentive payments by meeting specific criteria (Blumenthal and Tavenner, 2010).

Metadata: data about data, such as describing the context in which an element or file was collected, for example, whether a blood pressure reading was taken while the patient was sitting or standing (PCAST, 2010).

mHealth: an abbreviation for mobile health, a term used for the use of data from mobile devices—often personally held—which are increasingly evolving as a source of data for individuals, their clinicians, and the public health community (Health Affairs, 2013).

LOINC: Logical Observation Identifiers Names and Codes is a universal standard for identifying medical laboratory observations (Regenstrief, 2014).

“Narrow neck” standards: Used to refer to the minimal standards that facilitate a wide range of activities, such as the set of communication protocols, known as TCP/IP, that enable the Internet (NRC, 1994).

Observational clinical research: clinical research done based on observation of phenomena in their natural settings (Gordis, 2009).

OMOP: Observational Medical Outcomes Partnership, a public-private partnership between the Food and Drug Administration and pharmaceutical manufacturers to identify the most reliable methods for analyzing huge volumes of data drawn from heterogeneous sources (Overhage et al., 2012).

Online patient communities: communities of patients, often focused on a particular disease state, whose interactions are held primarily online, often through social networking platforms such as Diabetic Connect (Shaywitz, 2012).

Patient data networks: networks comprised primarily of patient data, such as PatientsLikeMe (Wicks et al., 2010).

Patient portals: portals, usually Web-based, through which patients can access their health data (ONC, 2014b).

PCORI: The Patient-Centered Outcomes Research Institute, a United States based nongovernmental institute created by clauses in the Patient Protection and Affordable Care Act .

PCORnet: PCORI's national patient-centered clinical research network, a large national network of health data representative of patients across the country (Selby et al., 2013).

PCORnet Coordinating Center: a center that will provide technical and logistical support to the individual components of PCORnet, such as technical assistance with inter-network data harmonization (Selby et al., 2013).

PPRNs: Patient-Powered Research Networks; one of two types of networks that comprise the PCORI National Patient-Centered Clinical Research Network, which are comprised primarily of patient-generated data (Selby et al., 2013).

Protocol transparency: openness about official procedures or systems of rules; in health, transparency is important on the rules that govern the sharing and use of patient-related data.

Registries: an organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons who have either a particular disease, a condition (e.g., a risk factor) that predisposes them to the occurrence of a health-related event, or prior exposure to substances (or circumstances) known or suspected to cause adverse health effects (AHRQ, 2007).

SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms is a collection of medical terms providing codes, terms, synonyms, and definitions used in clinical documentation and reporting (IHTSDO, 2014).

Social media apps: applications, often accessed through smartphones or tablets, that allow people to create, share, and/or exchange information and ideas in virtual communities and networks (Oxford Dictionaries, 2013).

Standardized data formats: consistent representation of data fields, such as the sex of all female patients being represented by an "F" (Ogunyemi et al., 2013).

TCP/IP: The Internet protocol suite including the Internet Protocol (IP), the principal communications protocol, and the Transmission Control Protocol (TCP), the protocol for sharing resources using packet-switching among the nodes (IETF, 1989a,b).

Use case: a narrative that describes real-world example uses of a technology and is used to test the functionality of a proposed approach to solving a problem (IBM, 2014).

References

- AHRQ (Agency for Healthcare Research and Quality). 2007. *Registries for evaluating patient outcomes: A user's guide*. Prepared by Outcome DEcIDE Center, Cambridge, MA.
- Blumenthal, D., and M. Tavenner. 2010. The “meaningful use” regulation for electronic health records. *New England Journal of Medicine* 363(6):501-504.
- Deloitte. 2013. *Governance operating model: A tool for more effective board oversight*. <http://deloitte.wsj.com/riskandcompliance/2013/06/11/governance-operating-model-a-tool-for-more-effective-board-oversight> (accessed January 14, 2014).
- Gordis, L. 2009. *Epidemiology*. 4th ed. Philadelphia: Elsevier/Saunders.
- Health Affairs. 2013. Mhealth and FDA guidance. http://healthaffairs.org/healthpolicybriefs/brief_pdfs/healthpolicybrief_104.pdf (accessed January 14, 2014).
- IBM. 2014. *Defining use cases*. http://pic.dhe.ibm.com/infocenter/rpcmpose/v2r0/index.jsp?topic=/com.ibm.rational.rrc.help.doc/topics/t_define_ucs.html (accessed January 14, 2014).
- IETF (Internet Engineering Task Force). 1989a. *Requirements for Internet hosts—application and support*. <http://tools.ietf.org/html/rfc1123> (accessed January 14, 2014).
- IETF. 1989b. *Requirements for Internet hosts—communication layers*. <http://tools.ietf.org/html/rfc1122> (accessed January 14, 2014).
- Iezzoni, L. I. 1997. Assessing quality using administrative data. *Annals of Internal Medicine* 127(8 Part 2):666-674.
- IHTSDO (International Health Terminology Standards Development Organisation). 2014. *SNOMED CT*. <http://www.ihtsdo.org/snomed-ct> (accessed January 14, 2014).
- IOM (Institute of Medicine). 2012. *Best care at lower cost: The path to continuously learning health care in America*. Washington, DC: National Academies Press.
- Liu, D., X. Wang, F. Pan, P. Yang, Y. Xu, X. Tang, J. Hu, and K. Rao. 2010. Harmonization of health data at national level: A pilot study in China. *International Journal of Medical Informatics* 79(6):450-458.
- NRC (National Research Council). 1994. *Realizing the information future : The Internet and beyond*. Washington, DC: National Academy Press.
- Ogunyemi, O. I., D. Meeker, H. E. Kim, N. Ashish, S. Farzaneh, and A. Boxwala. 2013. Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems. *Medical Care* 51(8 Suppl 3):S45-S52.
- ONC (Office of the National Coordinator for Health IT). 2014a. *Direct project*. <http://www.healthit.gov/policy-researchers-implementers/direct-project> (accessed January 13, 2014).
- ONC. 2014b. *What is a patient portal?* <http://www.healthit.gov/providers-professionals/faqs/what-patient-portal> (accessed January 14, 2014).
- ONC. 2014c. *What is an electronic medical record (EMR)?* <http://www.healthit.gov/providers-professionals/electronic-medical-records-emr> (accessed January 13, 2014).
- Overhage, J. M., P. B. Ryan, C. G. Reich, A. G. Hartzema, and P. E. Stang. 2012. Validation of a common data model for active safety surveillance research. *Journal of the American Medical Informatics Association* 19(1):54-60.
- Oxford Dictionaries. 2013. *Social media*. <http://www.oxforddictionaries.com/definition/english/social-media> (accessed January 13, 2014).
- PCAST (President's Council of Advisors on Science and Technology). 2010. *Report to the president: Realizing the full potential of health information technology to improve healthcare for Americans: The path forward*. Washington, DC: PCAST.
- Regenstrief. 2014. *Logical observation identifiers names and codes*. <http://loinc.org> (accessed January 14, 2014).
- Selby, J. V., H. M. Krumholz, R. E. Kuntz, and F. S. Collins. 2013. Network news: Powering clinical research. *Science Translational Medicine* 5(182):182fs113.
- Shaywitz, D. 2012. Getting better: Online communities elevate voice of the patient. *Forbes*. <http://www.forbes.com/sites/davidshaywitz/2012/02/07/getting-better-online-communities-elevate-voice-of-the-patient> (accessed January 14, 2014).
- WHO (World Health Organization). 2014. *International classification of diseases*. <http://www.who.int/classifications/icd/en> (accessed January 14, 2014).
- Wicks, P., M. Massagli, J. Frost, C. Brownstein, S. Okun, T. Vaughan, R. Bradley, and J. Heywood. 2010. Sharing health data for better outcomes on patientslikeme. *Journal of Medical Internet Research* 12(2):e19.

Roundtable on Value & Science-Driven Health Care

Mark B. McClellan (Chair)
Director
Engelberg Center
The Brookings Institution

David Blumenthal
President
The Commonwealth Fund

Raymond Baxter
SVP, Community Benefit
Research and Health Policy
Kaiser Permanente

Bruce G. Bodaken
Chairman, President, and
Chief Executive Officer
Blue Shield of California

Paul Chew
Chief Science Officer and
Chief Medical Officer
Sanofi U.S.

Carolyn M. Clancy
Director
Agency for Healthcare
Research and Quality*

Francis Collins
Director
National Institutes of Health*
(Designee: Susan Shurin)

Helen Darling
President
National Business Group on
Health

Susan D. Devore
Chief Executive Officer
Premier, Inc.

Judith Faulkner
Founder and CEO
Epic Health Systems

Thomas R. Frieden
Director
Centers for Disease Control
and Prevention*
(Designee: Gail R. Janes)

Patricia A. Gabow
Former CEO
Denver Health

Atul Gawande
General and Endocrine
Surgeon
Brigham and Women's Hospital

Gary L. Gottlieb
President and CEO
Partners HealthCare System

James A. Guest
President
Consumers Union

George C. Halvorson
Chairman and CEO
Kaiser Permanente

Margaret A. Hamburg
Commissioner
Food and Drug Administration*
(Designee: Peter Lurie)

James Heywood
Co-Founder and Chairman
PatientsLikeMe

Ralph I. Horwitz
Senior VP, Clinical Evaluation
Sciences
GlaxoSmithKline

Paul Hudson
EVP North America
AstraZeneca

Brent C. James
Chief Quality Officer and
Executive Director
Institute for Health Care
Delivery Research
Intermountain Healthcare

Craig Jones
Director
Vermont Blueprint for Health

Gary Kaplan
Chairman and CEO
Virginia Mason Health System

Darrell G. Kirch
President and CEO
Association of American
Medical Colleges

Richard C. Larson
Mitsui Professor
Massachusetts Institute of
Technology

James L. Madara
CEO
American Medical Association

Farzad Mostashari
National Coordinator
Office of the National
Coordinator for Health IT*

Mary D. Naylor
Professor and Director
Center for Transitions in Health
University of Pennsylvania

William D. Novelli
Former CEO, AARP
Professor, Georgetown
University

Sam R. Nussbaum
Executive VP, Clinical Health
Policy and Chief Medical
Officer
WellPoint, Inc.

Jonathan B. Perlin
Chief Medical Officer and
President, Clinical & Physician
Services, HCA, Inc.

Robert A. Petzel
Under Secretary for Health
Department of Veterans
Affairs*

Richard Platt
Professor and Chair
Population Medicine, Harvard
Medical School

John W. Rowe
Professor
Mailman School of Public
Health, Columbia University

Joe Selby
Executive Director
Patient-Centered Outcomes
Research Institute

Mark D. Smith
President and CEO
California HealthCare
Foundation

Glenn D. Steele
President and CEO
Geisinger Health System

Marilyn Tavenner
Administrator
Centers for Medicare &
Medicaid Services*
(Designee: Patrick Conway)

Reed V. Tuckson
Managing Director
Tuckson Health Connections

Mary Wakefield
Administrator
HRSA*

Debra B. Whitman
Executive VP
Policy and International
AARP

Jonathan Woodson
Assistant Secretary for Health
Department of Defense*

*Ex Officio

PLANNING COMMITTEE ON DATA HARMONIZATION FOR PATIENT-CENTERED RESEARCH**

Michael Kahn (*Chair*), Professor, Department of Pediatrics, University of Colorado, Denver; **Christopher G. Chute**, Professor of Medical Informatics, Mayo Clinic; **James J. Cimino**, Chief, Laboratory for Informatics Development, National Institutes of Health; **Lesley H. Curtis**, Professor, Duke Clinical Research Institute, Duke University; **Rachael Fleurence**, Acting Director, Accelerating PCOR Methods Program, Patient-Centered Outcomes Research Institute; **Doug Fridsma**, Director, Office of Standards and Interoperability, Office of the National Coordinator for Health IT; **Betsy L. Humphreys**, Deputy Director, National Library of Medicine, National Institutes of Health; **Rebecca D. Kush**, President and Chief Executive Officer, Clinical Data Interchange Standards Consortium; **Shawn N. Murphy**, Associate Director, Laboratory of Computer Science, Partners HealthCare System; **Patrick Ryan**, Head, Epidemiology Analytics, Janssen Research and Development; **Joe V. Selby**, Executive Director, Patient-Centered Outcomes Research Institute; and **Sharon Terry**, President and CEO, Genetic Alliance.

IOM STAFF

Claudia Grossmann, Senior Program Officer; **Katherine Burns**, Senior Program Assistant; **Diedra Henderson**, Program Officer; **Elizabeth Johnston**, Senior Program Assistant; **Elizabeth Robinson**, Research Associate; **Valerie Rohrbach**, Senior Program Assistant; **Julia Sanders**, Senior Program Assistant (through August 2013); **Robert Saunders**, Senior Program Officer; **Barret Zimmerman**, Senior Program Assistant (until October 2013); **J. Michael McGinnis**, Senior Scholar, Executive Director, Roundtable on Value & Science-Driven Health Care

DISCLAIMER: This workshop in brief has been prepared by Claudia Grossman as a factual summary of what occurred at the meeting. The statements made are those of the authors or individual meeting participants and do not necessarily represent the views of all meeting participants, the planning committee, or the National Academies.

**IOM planning committees are solely responsible for organizing the workshop, identifying topics, and choosing speakers. The responsibility for the published workshop in brief rests with the institution.

This workshop in brief was reviewed by **Greg Simon**, Group Health Research Institute, and **Dixie Baker**, Martin, Blanck & Associates, and coordinated by **Chelsea Frakes**, Institute of Medicine, to ensure that it meets institutional standards for quality and objectivity.

This workshop was sponsored by the Patient-Centered Outcomes Research Institute.

For additional information regarding the workshop, visit www.iom.edu/DataHarmonization.