Pragmatic Data Domain Selection for a National Distributed Research Network: The PCORnet Common Data Model Strategy

Shelley A. Rusincovitch[1], Abel N. Kho, MD, MS[2], Jon E. Puro, MPA:HA[3], Daniella Meeker, PhD[4], Pedro Rivera, MSCS[3], Aaron A. Sorensen, MA[5], Jeffrey S. Brown, PhD[6], and Lesley H. Curtis, PhD[7]

[1]Duke Translational Medicine Institute, Durham, NC; [2]Northwestern University Departments of Medicine and Preventive Medicine, Evanston, IL; [3]OCHIN, Inc., Portland, OR; [4]Department of Health, RAND Corporation, Santa Monica, CA; [5]Temple University School of Medicine, Philadelphia, PA; [6]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; [7]Duke University School of Medicine, Department of Medicine, Durham, NC

Abstract

The PCORnet Common Data Model (CDM) is the foundation for the PCORI national distributed research network. We describe our experiences in assessing potential data domains and making decisions for inclusion in the CDM, including modeling attributes, dimensions of assessment, and lessons learned.

Introduction and Background

The PCORnet Common Data Model (CDM) specifies the data foundation for the national distributed research network under development by the Patient-Centered Outcomes Research Institute (PCORI). The PCORnet CDM is developed with a phase-based approach, with each phase incorporating new concepts and data tables to support distributed clinical research (observational and interventional). The first version of the CDM established six tables reflecting key patient-level data captured routinely within healthcare delivery and billing systems. In order to establish priorities for subsequent CDM development, it was necessary to establish a method of assessing new concepts and making decisions for inclusion to serve the functional, pragmatic focus of the initiative.

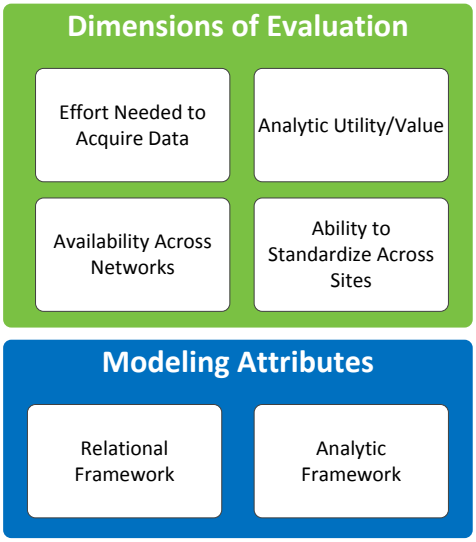

Figure 1. Overview of the data domain evaluation and modeling elements.

Methods

The assessment was organized by data domains; i.e., the high-level concepts of data organization based upon existing data sources, workflows, and processes. Our assessment included best practices established by existing data models and advice from external experts for specific topics. We chose four dimensions for assessment: Effort to acquire data; analytic value of data; ability to standardize data; and availability of data. Each of these dimensions was classified using a simple high, moderate, or low ranking. The CDM Working Group (CDM WG), initially convened in 3 meetings during the summer of 2014 to evaluate and prioritize new data domains for the CDM.

Results

During development and modeling of domains we paid close attention to PCORnet-specific requirements, such as capture of important analysis concepts and data availability. The assessment resulted in a recommendation for prioritization of data domains to be included in the CDM. The initiative has subsequently formalized a process for stakeholder review, facilitated discussion, and the approval process for adoption.

Discussion

A key lesson learned from this process has been the importance of identifying and articulating foundational strategic decisions, including interoperability within the analytic framework. One limitation of this evaluation approach is that dimensions were assessed by the CDM WG, rather than by a formal survey of all participants; however, the working group represented individuals with the deep expertise necessary to make informed recommendations. As the PCORnet initiative develops and evolves, a robust, pragmatic evaluation strategy will help to ensure that the data model strongly supports the mission of PCORI and the national distributed research network.