

Development of a National Distributed Research Network Data Infrastructure: Design of the PCORnet Common Data Model

Jeffrey S. Brown, PhD¹, Shelley A. Rusincovitch², Abel N. Kho, MD, MS³,
Keith Marsolo, PhD⁴, and Lesley H. Curtis, PhD⁵

¹Department of Population Medicine, Harvard Medical School and
Harvard Pilgrim Health Care Institute, Boston, MA; ²Duke Translational Medicine
Institute, Durham, NC; ³Northwestern University Departments of Medicine and Preventive
Medicine, Evanston, IL; ⁴Division of Biomedical Informatics, Cincinnati Children's
Hospital Medical Center, Cincinnati, Ohio; ⁵Duke University School of Medicine,
Department of Medicine, Durham, NC

Abstract

The PCORnet Common Data Model (CDM) is being developed to support the execution of common analytical programs across PCORnet, a national Distributed Research Network comprised of 29 networks located across the United States. The iterative development of the CDM is governed by a set of Guiding Principles and includes input from network stakeholders.

Introduction and Background

PCORnet is a national Distributed Research Network (DRN) sponsored by the Patient-Centered Outcomes Research Institute (PCORI). Formed in late 2013 with 29 networks located across the United States, the current phase of PCORnet will create a “network of networks” to form a highly representative national infrastructure for research. Using the PopMedNet™ software infrastructure, the architecture will allow queries and analytic programs to be disseminated to individual data stores. The PCORnet Common Data Model (CDM) has been developed to provide a standard organization and representation.

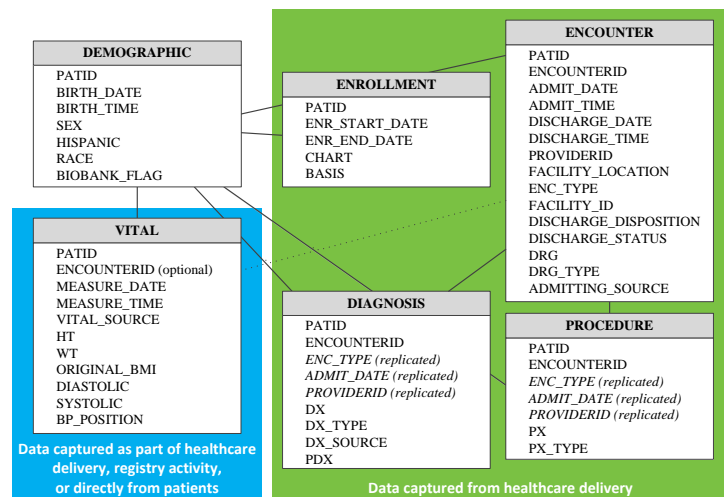


Figure 1. PCORnet CDM v1.0 entity-relationship diagram.

Methods

The development of the CDM began with the drafting of a set of Guiding Principles, which were included as part of the charter of the PCORnet Data Standards, Security, and Network Infrastructure (DSSNI) Task Force. As a result of the strategic direction set forth in the Guiding Principles, a decision was made to have the PCORnet CDM be a new stand-alone model instead of an existing standard. It is heavily derived from the Mini-Sentinel Common Data Model, however, and also informed by other DRN-related initiatives such as the Observational Medical Outcomes Partnership (OMOP) CDM and HMO Research Network Virtual Data Warehouse. The PCORnet CDM will be developed in incremental phases, allowing agility in deployment and for new data domains to be incorporated over the lifespan of the DRN. The modeling is optimized for analytic functionality.

Results

The PCORnet CDM v1.0 was released on May 30, 2014, after a review period in which more than 250 comments were received and incorporated. The first version of the PCORnet CDM is primarily based upon secondary data generated by healthcare delivery processes. Rather than developing standards for how these data should be collected and represented, the CDM takes a pragmatic approach, supporting the current state of these data.

Discussion

At present, the implementation of the PCORnet CDM v1.0 is in progress, and the next versions are under development. An area of future development is for those data domains that support multiple contexts: healthcare delivery, registries and clinical studies, and data generated and received directly from patients.