PCORnet Common Data Model 3.0 User Guide

Introduction

This document aims to explain the PCORnet Common Data Model (CDM) 3.0. You can find the full CDM guide online at: [http://pcornet.org/pcornet-common-data-model](http://pcornet.org/pcornet-common-data-model)

**What is the Common Data Model?**

PCORnet created the Common Data Model for networks in PCORnet. The CDM makes it easier for networks to share information by setting common definitions and organizing data so that:

1) PCORnet can analyze data more quickly.
2) PCORnet CDM can support the different platforms that the networks use to organize their data.
3) Organizations can better compare their data. Comparison is easier and more efficient because everyone can now use the same organization and definitions of their data.

PCORnet developed the CDM for wide distribution. The CDM is licensed under Creative Commons and easy for anyone to access, use, and share. It also works for various stakeholders because it is based upon an understanding of the data commonly generated by organizations that deliver healthcare, such as hospitals, doctor’s offices, and insurance plans. It is based on work done by many other groups and is designed to promote multi-site, patient-centered research. This includes observational research (studies where researchers observe, rather than test a drug or procedure) and clinical trials (studies that answer specific questions about a drug, procedure, device, etc.).

Like any work in progress, the CDM will continue to be flexible and change to meet stakeholder needs. The CDM’s standards for how data is organized and defined will make it easier to do collaborative studies.

**How does the PCORnet Common Data Model help my network?**

The CDM allows PCORnet networks to compare their data with each other, which is powerful and essential when doing research. Not all networks will be able to collect the same data elements. Using the CDM, organizations can contribute to overall knowledge with their available data sources, even when they are different than other networks.
The CDM provides guidance on how to build and manage your PCORnet analysis database. Your organization’s database programmers can access these guides, including standardized keys (codes). The CDM provides a way to map between the format in which you entered data and the CDM representation. This preserves the values and variables within your data sets while allowing for data comparison across organizations.

Data handling can be costly and often takes a lot of time and effort. The CDM is based upon standardized codes that are routinely used in healthcare settings. This means that data can be compared more reliably.

For example, date of birth is a common piece of patient information. Most studies require it, but organizations use different codes for it. This can quickly become confusing and hard to manage, unless there are standards like those in the CDM. A clinic might code birthdate as “Date_of_Birth,” a health system might call it “Birth_DT,” and a health registry might use the code “DOB.” In this example, the CDM offers a universal variable: “BIRTH_DATE.” Each organization’s system can map its own birth date, no matter how it is labeled, to this variable. By using the CDM, your organization keeps control over your members’ data, but you can also collaborate with other groups or researchers to run more effective observational research and clinical trials.

**What is the history of the Common Data Model?**
The CDM has already released several versions, and continues to evolve.

- Version 1.0 (released in May 2014) focused on a foundation of data elements that tend to be readily available.
- Version 2.0 (released February 2015) added data elements in new domains of data.
- Version 3.0 (released in June 2015) continues to expand domains, including key information to support pragmatic clinical trial execution.

Implementation of each new version incorporates lessons learned during the phases before it. Each organization that uses the CDM must document which version it uses and how it is used. This helps PCORnet track progress and continue to improve the CDM.

**What should I know about version 3.0?**
The CDM version 3.0 includes several upgrades that make it easier to organize data and manage the database:

1. **Additional data elements.** Ten new fields in existing tables and five new tables have been added to expand the amount and types of information being represented.
2. **Modifications to relational integrity specifications and to date formatting practices.** This means that data tables have more rules around null values (when no information is entered) and foreign keys (relationships between tables).
3. **New specifications to allow the CDM to run with SAS, a computer program used for statistical analyses.**
4. **A new method for representing dates consistently across organizations.**
5. Additional PCORnet trial tables to connect and filter CDM data inside of a clinical trial’s protocol.
6. Written guidance with additional instructions and descriptions.

What influenced the development of the CDM?
The PCORnet CDM is based on the Mini-Sentinel Common Data Model (MSCDM), an FDA surveillance platform that incorporates EHR and health plan data. The MSCDM was built to support rapid data analysis.

What are the technical specifications of the CDM?
The PCORnet Common Data Model:

- Protects patient confidentiality by using the “minimum necessary” of patient data. The CDM also specifies that each network does not use medical record numbers (MRNs) to identify patients, but instead creates an arbitrary identifier number.
- Has been influenced by many projects and databases. These include the Health Care Systems Research Network, the Vaccine Safety Datalink, various AHRQ Distributed Research Network projects, and the ONC Standards & Interoperability Framework Query Health Initiative.
- References standard terminologies such as ICD, SNOMED, CPT, HCPSC, and LOINC. To learn more about these terms, visit the PCORI website.
- Is covered under a Creative Common license, so anyone interested can freely obtain and use it.

Overview Diagram
The below overview diagram shows the 15 new fields added in CDM version 3.0.

The 15 PCORnet CDM Domains, v3.0

- CONDITION +2.0
  - A condition represents a patient’s diagnosed and self-reported health conditions and diseases. The patient’s medical history and current state may both be represented.

- DEATH +2.0
  - Reported mortality information for patients.

- DEATH_CAUSE +3.0
  - The individual causes associated with a reported death.

- DEMOGRAPHIC +1.0
  - Demographics record the direct attributes of individual patients.

- DIAGNOSIS +1.0
  - Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery.

- DISPENSING +2.0
  - Outpatient pharmacy dispensing, such as prescriptions filled through a neighborhood pharmacy with a claim paid by an insuror. Outpatient dispensing is not commonly captured within healthcare systems.

- ENROLLMENT +3.0
  - Enrollment is a concept that defines a period of time during which all medically-attended events are expected to be observed. This concept is often insurance-based, but other methods of defining enrollment are possible.

- ENCOUNTER +2.0
  - Encounters are interactions between patients and providers within the context of healthcare delivery.

- HARVEST +3.0
  - Attributes associated with the specific PCORnet datamart implementation

- LAB_RESULT_CM +2.0
  - Laboratory result Common Measures (CM) use specific types of quantitative and qualitative measurements from blood and other body specimens. These standardized measures are defined in the same way across all PCORnet networks.

- PCORNET_TRIAL +3.0
  - Patients who are enrolled in PCORnet clinical trials.

- PRESCRIBING +1.0
  - Provider orders for medication dispensing and/or administration.

- PRO_CM +2.0
  - Patient-Reported Outcome (PRO) Common Measures (CM) are standardized measures that are defined in the same way across all PCORnet networks. Each measure is recorded at the individual item level: an individual question/statement, paired with its standardized response options.

- PROCEDURES +1.0
  - Procedure codes indicate the discreet medical interventions and diagnostic testing, such as surgical procedures, administered within healthcare delivery.

- VITAL +1.0
  - Vital signs (such as height, weight, and blood pressure) directly measure an individual’s current state of attributes.