Developing a Common Data Model: The PCORnet® Experience

Darcy M. Louzao, Ph.D.
March 24, 2018
Questions to Answer Today

What is the goal of PCORnet?

What is PCORnet’s data strategy?

How has PCORnet’s CDM been developed? And under what key principles?

What are the major challenges and how do you navigate them?
PCORnet: the National Patient-Centered Clinical Research Network

PCORnet is a large, highly representative, national patient-centered clinical research network.

Our **vision** is to support a learning U.S. healthcare system and to enable **large-scale clinical research** conducted with **enhanced quality and efficiency**.

Our **mission** is to enable people to make informed healthcare decisions by efficiently conducting clinical research relevant to their needs.
PCORnet® embodies a “network of networks” that harnesses the power of partnerships
How do we do this?
The Data story
Enabling Research at a National Scale

How do you ask a research question at hundreds of institutions and get back results you can trust?

Option 1 — Write a description and have everyone create a local implementation to run on their data

Option 2 — Create an algorithm that can run against a single, common data model
Density and Heterogeneity of data partners
PCORnet Data Strategy

- Standardize data into a common data model
- Focus on data quality: data curation
- Operate a secure distributed query infrastructure
  - Develop re-usable tools to query the data
  - Send questions to the data and only return required information
- Learn by doing and repeat
CDM overview

- Strong basis of secondary data generated by healthcare delivery processes

- Heavily derived from the Mini-Sentinel Common Data Model and informed by other initiatives

- Development takes a pragmatic approach, supporting the current state of these data

- Modeling is optimized for analytic functionality and to be intuitive to investigators

CDM Guiding Principles (abridged)

- CDM will contain data of general interest to PCORnet
- Not all networks are expected to populate all parts of CDM
- CDM will change over time
- CDM will be intuitive & easy to understand
- Other CDMs exist – PCORnet will learn from them
- CDM will reflect values found in source data
- Networks are free to add data/domains to their local CDMs
Standardize to a common data model

**Fundamental basis**
- DEMOGRAPHIC

**Associations with PCORnet clinical trials**
- PCORNET TRIAL

**Process-related data**
- HARVEST

**Data captured from processes associated with healthcare delivery**
- ENROLLMENT
- DISPENSING
- DEATH
- DEATH CAUSE

**Data captured within multiple contexts: healthcare delivery, registry activities, or directly from patients**
- VITAL
- CONDITION
- PATIENT-REPORTED OUTCOMES (COMMON MEASURES)

**Data captured from healthcare delivery, direct encounter basis**
- ENCOUNTER
- DIAGNOSIS
- PROCEDURES
- PRESCRIBING
- LABORATORY RESULTS (COMMON MEASURES)
The 19 PCORnet Data Domains, v4.0 (to be implemented Summer 2018)

<table>
<thead>
<tr>
<th>PCORnet Common Data Model v4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEMOGRAPHIC</td>
</tr>
<tr>
<td>PATID</td>
</tr>
<tr>
<td>ETC</td>
</tr>
<tr>
<td>PAT_PREF_LANGUAGE</td>
</tr>
<tr>
<td>NET</td>
</tr>
<tr>
<td>VITAL</td>
</tr>
<tr>
<td>VITALID</td>
</tr>
<tr>
<td>PATID</td>
</tr>
<tr>
<td>MEASURE_DATE</td>
</tr>
<tr>
<td>VITAL_SOURCE</td>
</tr>
<tr>
<td>ETC</td>
</tr>
<tr>
<td>PRO</td>
</tr>
<tr>
<td>PRO_CM_ID</td>
</tr>
<tr>
<td>PATID</td>
</tr>
<tr>
<td>PRO_ITEM</td>
</tr>
<tr>
<td>PRO_DATETIME</td>
</tr>
<tr>
<td>PRO_TYPE</td>
</tr>
<tr>
<td>PRO_ITEM_TYPE</td>
</tr>
<tr>
<td>PRO_ITEM_XREF</td>
</tr>
<tr>
<td>PRO_ITEM_VALUE</td>
</tr>
<tr>
<td>PRO_MEASURE_VALUE</td>
</tr>
<tr>
<td>PRO_MEASURE_SEQ</td>
</tr>
<tr>
<td>PRO_MEASURE_SCORE</td>
</tr>
<tr>
<td>PRO_MEASURE_THETA</td>
</tr>
<tr>
<td>PRO_MEASURE_TSCORE</td>
</tr>
<tr>
<td>PROCEDURE</td>
</tr>
<tr>
<td>PROCEDURE_CODE</td>
</tr>
<tr>
<td>PROCEDURE_COUNT</td>
</tr>
<tr>
<td>PROCEDURE_DESCRIPTION</td>
</tr>
<tr>
<td>PROCEDURE_SOURCE</td>
</tr>
</tbody>
</table>

DANGER SLOW DOWN
CDM Development Timeline

- **V1.0**: Released May 30, 2014
  - 276 discrete comments, one feedback cycle
  - Two stakeholder sessions

- **V2.0**: Released Feb 27, 2015
  - 265 discrete comments, two feedback cycles
  - Two stakeholder sessions

- **V3.0**: Released June 1, 2015
  - 236 discrete comments, one feedback cycle
  - Two stakeholder sessions

- **V3.1**: Released Nov 15, 2016
  - 106 discrete comments, one feedback cycle
  - One stakeholder session

- **V4.0**: Released Jan 3, 2018
  - 365 discrete comments, two feedback cycles
  - Five stakeholder sessions

5 versions in 3.5 years
Each release of the CDM has a cascading effect on the Network

Coordinating Center

- Plan updates to data curation (DC) procedures, Implementation Guidance, analytic tools, etc.

Network Partners

- Refresh data using current CDM
- Provide updated DC results on current CDM
- Respond to queries using current CDM
- Develop extract-transform-load (ETL) procedures on new CDM
- Populate data into new CDM

First 3 versions of CDM occurred during period of minimal query activity
Evaluate data quality and fitness-for-use

- Data quality checks
- Annotated data dictionaries
- Implementation guidance
- Data curation and CDM implementation forums
- Analyses and reports
Develop reusable tools to query the data

Study Design:

Identify patients <AGE RANGE> with a <NEW / REFILL> dispensing of a <DRUG CLASS>. To be eligible, patients must have met the following criteria in the <LOOKBACK PERIOD> days before the index dispensing: (1) continuous enrollment in <TYPE(S) OF INSURANCE> benefits, (2) no prescription for <DRUG CLASS> or <DRUG CLASS>, and (3) no diagnosis of <DISEASE / CONDITION> in <INPATIENT / OUTPATIENT / ED> care setting.

The primary outcome of interest is <EVENT> identified with <DIAGNOSIS CODE> in <PRIMARY / SECONDARY / ANY> position during an <INPATIENT / OUTPATIENT / ED> encounter.
Learn by doing

Data Curation

Research & Analytic Tools

CDM Development
What are the challenges and how to navigate?

- Common data model paralysis
  - Establish guiding principles
  - Don’t start from scratch
  - Refine over time, but not too quickly

- Establishing the rules of the road
  - Who gets to send queries? How are they sent? Is everyone required to answer every queries? How are results returned? Can query results be shared publicly? Published?
  - Begin with a useful, relatively simple use case
  - Build trust
  - Refine over time
What are the challenges and how to navigate?

- Developing analysis code that runs in every environment is challenging
  - Develop reusable programs that are structured to answer common types of questions
  - Engage local experts in the network community

- Sustainability of infrastructure
  Extract, transform, and load (ETL) of data into a common data model is real work. When source data change, the ETL changes. When the CDM changes, ETL and tools have to change.
  - Be mindful of the cascade and proceed carefully
  - Develop useful products
Thank you!

Questions?