

The PCORnet[®] Common Data Model & Patient-Reported Outcomes

Keith Marsolo, PhD

Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center

Co-Investigator – ImproveCareNow PPRN; PEDSnet CDRN

Co-Investigator – PCORnet Coordinating Center Distributed Research Network Operations Center

Member – PCORnet Data Committee



pcornet[®]


The National Patient-Centered
Clinical Research Network

The PCORnet Story

Our nation's clinical research system needs improvement.

- It doesn't answer all the questions that matter most to people,
- is too expensive,
- and too slow.

PCORI created PCORnet to help us improve it.




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

PCORnet enables more efficient, trustworthy clinical research to help people:

- Make **better-informed health decisions**,
- Achieve significant **cost savings**,
- and receive **faster clinical insights**

PCORnet's research community unites people, clinicians, and health systems with patient data from multiple sources: electronic health records, insurance claims data, data reported directly by people, and other data.



CDRNs
Clinical Data Research Networks
CDRNs originate in healthcare systems, such as hospitals, health plans, or practice-based networks




PPRNs
Patient-Powered Research Networks
PPRNs are operated and governed by groups of patients and their partners

PCORnet
a national infrastructure for people-centered clinical research

PCORnet represents:

~90 Million
patients who have had a medical encounter in the past 5 years
-some individuals may have visited more than one Network Partner and would be counted more than once
(as of July 15, 2016)

underpinned by a Common Data Model,

Sex 
Race 
Age 

Ask a question to hundreds of institutions and get back results you can trust.

to advance clinical research using PCORnet.

- Answer** pre-research questions faster.
- Collaborate** with our network partners.
- Benefit** from PCORnet study designation.

Think of all the different ways researchers can leverage PCORnet.

- pre-research**
- observational** studies
- interventional** studies

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

Why a Common Data Model?

Same information is represented differently at different institutions (e.g., Race)

SITE 1

Caucasian
African American
Asian
Multiple Race
Blank

SITE 2

101
201
300
401
500
600

SITE 3

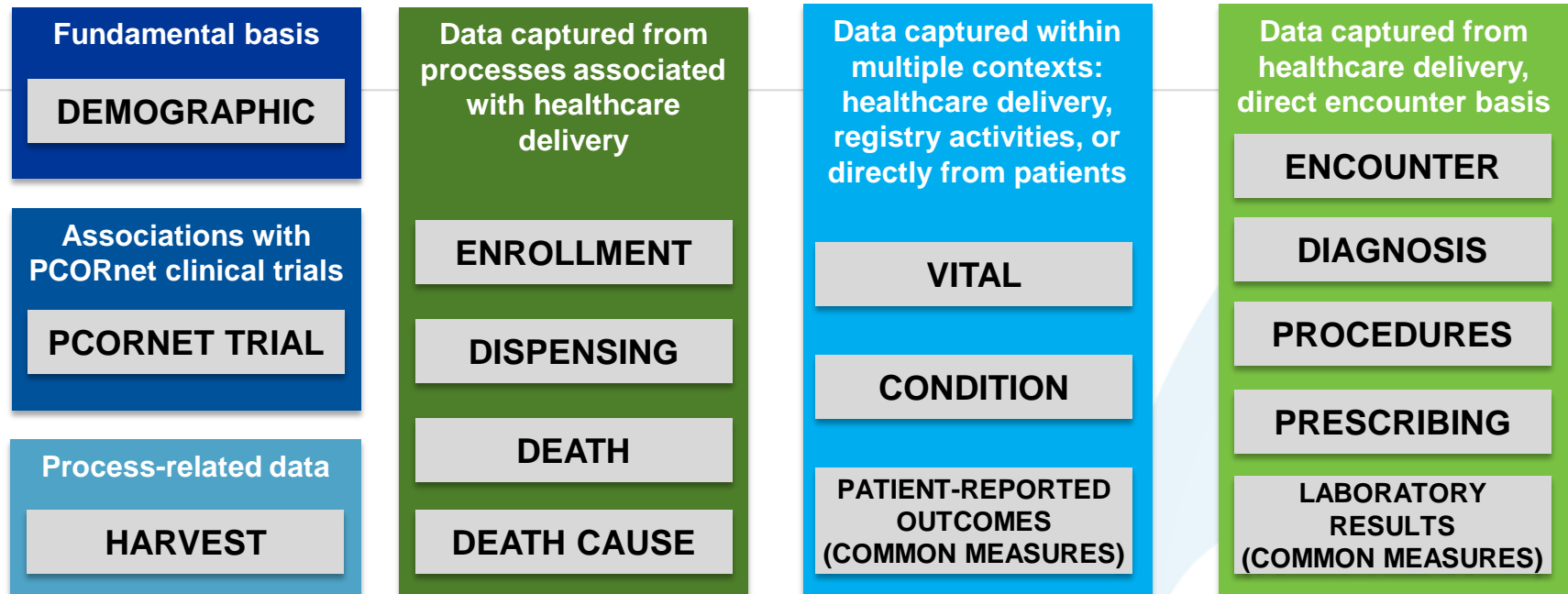
African American
American Indian
Asian American
White
Other
Unknown

Common Data Model Value Set

01 = American Indian or Alaska Native
02 = Asian
03 = Black or African American
04 = Native Hawaiian or Other Pacific Islander
05 = White
06 = Multiple Race
07 = Refuse to Answer
NI = No Information
UT = Unknown
OT = Other

*In order to be able to trust results of an analysis,
we need to have consistent representations*

PCORnet Common Data Model v3.1



- 4 versions released between May 30, 2014 & November 15, 2016

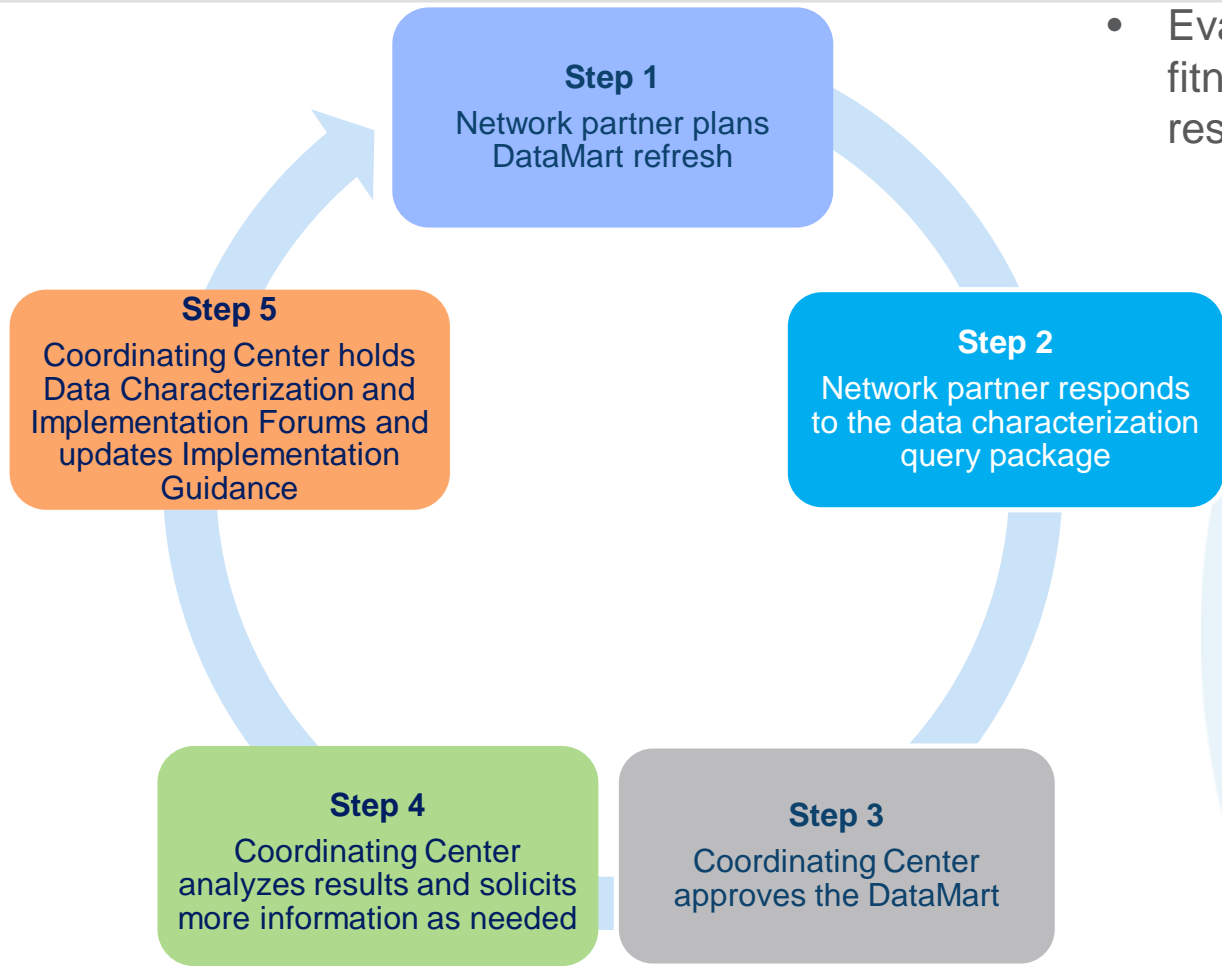
- 800+ discrete comments from stakeholders
- Multiple stakeholder sessions to review feedback

- Available at www.pcornet.org/pcornet-common-data-model/

Data within a distributed research network

- ❁ People interpret the CDM specification differently, resulting in variability in how CDM is populated
- ❁ Network composed of different health systems, with different EHRs, implemented at different times
- ❁ Clinical workflows differ across institutions & impact availability of data
- ❁ Understanding of EHR / claims data sources differs across institutions – may impact what gets loaded from source systems
- ❁ Important to assess data quality before running a query

Data curation



Data Curation assesses and improves global data quality

- Characterize the contents of the PCORnet CDM
- Evaluate global data quality and fitness-for-use across a broad research portfolio

Table IVc. Missing or Unknown Values
This table includes fields in the DEMOGRAPHIC, ENROLLMENT, ENCOUNTER, DIAGNOSIS, PROCEDURES and VITAL tables which are included in the query results and are not required to be populated (see Table IIc for required fields). VITAL measures are not included because the VITAL table structure (1 record per result) does not support measurement assessment. The table depicts the percentage of records with missing or unknown values and supports Data Checks 3.10 and 3.14.

		Records with missing, NUN, or OF values				Data Check		
Table	Field	Encounter Type	Count	Numerator	Denominator	%	Threshold	Source
DEMOGRAPHIC	BIRTH_DATE		0	605,817		3.03	>=5%	DEM_L3_AGEVROSTST1
DEMOGRAPHIC	SEX		160	605,817	0.0	3.03	>=5%	DEM_L3_SEXSTST1
DEMOGRAPHIC	HISPANIC		19,001	605,817	3.1	--	--	DEM_L3_HISPSTST1
DEMOGRAPHIC	RACE		55,221	605,817	9.1	3.04	>=15%	DEM_L3_RACESTST1
ENROLLMENT	ENR_END_DATE		0	605,817	--	--	--	ENR_L3_ENR_END
ENCOUNTER	DISCHARGE_DATE	IP, IS, EI	413	249,808	0.2	--	--	ENC_L3_ENCTYPE_IDDATE_YM
ENCOUNTER	ENC_TYPE		0	20,646,997	--	--	--	ENC_L3_ENCTYPE
ENCOUNTER	PROVIDERID		2,279,976	20,646,997	11.3	--	--	ENC_L3_N
ENCOUNTER	DISCHARGE_DISPOSITION	IP, IS, EI	73,716	249,808	29.5	3.04	>=15%	ENC_L3_ENCTYPE_DISDISP
ENCOUNTER	DISCHARGE_STATUS	IP, IS, EI	223,606	249,808	90.3	--	--	ENC_L3_ENCTYPE_DISSTAT
ENCOUNTER	DRG	IP, IS, EI	221,204	249,808	88.5	--	--	ENC_L3_ENCTYPE_DRG
ENCOUNTER	ADMITTING_SOURCE	IP, IS, EI	86,784	249,808	32.3	--	--	ENC_L3_ENCTYPE_ADMISC
DIAGNOSIS	DX_TYPE		0	25,811,129	--	3.03	>=5%	DXA_L3_DXTYPE_DISSOURCE
DIAGNOSIS	DX_SOURCE		0	25,811,129	--	--	--	DXA_L3_DXSOURCE
DIAGNOSIS	PDX	IP, IS, EI	0	1,734,843	0.0	3.04	>=15%	DXA_L3_PDX_ENCTYPE
PROCEDURES	PX_DATE		0	41,395,741	--	--	--	PRO_L3_PXDATE_Y
PROCEDURES	PX_TYPE		915,093	41,395,741	2.2	3.03	>=5%	PRO_L3_PXTYPE_ENCTYPE
PROCEDURES	PX_SOURCE		0	41,395,741	--	--	--	PRO_L3_PXSOURCE
VITAL	VITAL_SOURCE		0	38,787,908	0.0	3.03	>=5%	VIT_L3_VITAL_SOURCE

PCORnet Implementation Guidance – v1.0 for CDM v3.0

Table Name	Guidance
ENCOUNTER	<ul style="list-style-type: none"> The ENROLLMENT table provides an important analytic basis for identifying periods during which medical care should be observed, for calculating person-time, and for inferring the meaning of unobserved care (i.e., if care is not observed, it likely did not happen). Each ENCOUNTERID will generally reflect a unique combination of PATID, ADMIT_DATE, PROVIDERID and ENC_TYPE. Every diagnosis and procedure recorded during the encounter should have a separate record in the DIAGNOSIS or PROCEDURES Tables. Multiple visits to the same provider on the same day may be considered one encounter, especially if defined by a reimbursement basis; if so, the ENCOUNTER record should be associated with all diagnoses and procedures that were recorded during those visits. Visits to different providers on the same day, however, such as a physician appointment that leads to a hospitalization, would generally correspond to multiple encounters within the ENCOUNTER table. Rollback or voided transactions and other adjustments should be processed before populating this table. Although "Expired" is represented in both DISCHARGE_DISPOSITION and DISCHARGE_STATUS, this overlap represents the reality that both fields are captured in hospital data systems but with variation in how each field is populated. Do not include scheduled encounters. Partners should ensure that "administrative" encounters (e.g., e-mail, phone, documentation-only), are coded to the appropriate encounter type, which is typically "OA" for outpatient visits.
DIAGNOSIS	<ul style="list-style-type: none"> This table should capture all uniquely recorded diagnoses for all encounters. Diagnoses from problem lists will be captured in the CONDITION table. If a patient has multiple diagnoses associated with one encounter, then there would be one record in this table for each diagnosis. ENCOUNTERID should be populated for DIAGNOSIS and PROCEDURES. The definitions of the DIAGNOSIS and PROCEDURES tables are dependent upon a healthcare context; therefore, the encounter basis is necessary and the ENCOUNTERID, PROVIDERID, ENCOUNTER_TYPE, and ADMIT_DATE from the associated ENCOUNTER record should be included. While not desirable, a low percentage of orphan records is permissible to accommodate instances in which the associated ENCOUNTER details are missing from the source data. Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers, including technical/facility billing and professional billing and other data streams. Diagnoses are often only related to the treatment of the patient during the specific encounter. Chronic conditions that are not be pertinent to the treatment of a specific encounter, for example, would not be expected to be present.

PROs in the PCORnet CDM v3.1 (PRO_CM)

- 🌐 Focused on the PCORnet PRO Common Measures
- 🌐 Measures established by PCORnet PRO Task Force
 - Charge: develop recommendations for brief set (5-10 items) of common “core” questions, allowing comparisons across PCORnet
 - Resulted in 21 standardized measures (primarily PROMIS)
- 🌐 Table populated by about 10% of DataMarts (as of Feb. 2017)

PROs in future versions of the PCORnet CDM

- Currently developing draft specifications for a new version of CDM
- Considering expansion of PRO_CM table to include non-common measures
- Interested in comments on potential modifications
 - Support inclusion of domain scores
 - Allow for PROs from terminologies besides LOINC
 - Learn from existing models?
- Also preparing guidance/best practices for groups that wish to supplement the CDM with additional data domains (e.g., device data)

Questions?

- Send any thoughts/feedback on CDM modifications to keith.marsolo@cchmc.org