# Iterative Design and Modeling for the PCORnet Common Data Model v3.0

Shelley A. Rusincovitch, Jeffrey S. Brown, PhD, Keith Marsolo, PhD, Jenny Ibarra, RN, MSN, Abel N. Kho, MD, MS, Daniella Meeker, PhD, Lesley H. Curtis, PhD

S36: Infrastructure for
Data-driven Translational Science
**Thursday, March 24, 2016**

AMIA
**2016 Joint Summits on Translational Science**
**March 21–24, San Francisco**

pcornet
The National Patient-Centered Clinical Research Network

# Disclosure

- Shelley Rusincovitch discloses that she has no financial relationships with commercial interests.

# Learning Objectives

- After participating in this activity, the learner should be better able to:
  - Assess the PCORnet Common Data Model iterative design process and mechanisms for soliciting and incorporating stakeholder input.
  - Evaluate the data-generating activity classifications and domain representation.
  - Understand the design drivers for development of the version 3.0 PCORnet Common Data Model release.

# The PCORnet Common Data Model

pcornet

AMIA

2016 Joint Summits on Translational Science
March 21–24, San Francisco

# *What is the CDM?*

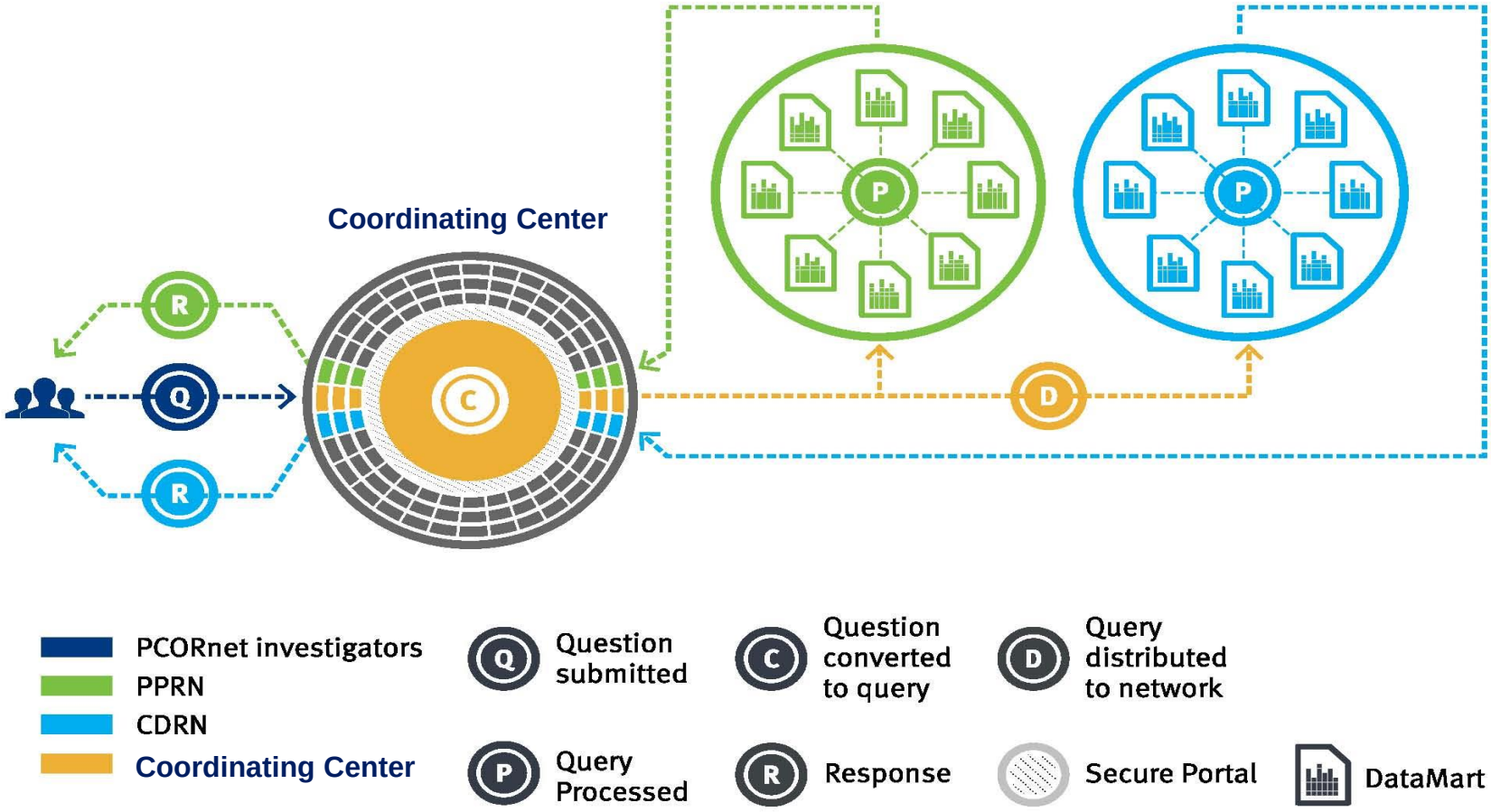
*The PCORnet Common Data Model (CDM) is a* ***specification*** *that defines a* ***standard organization*** *and* ***representation*** *of data for the PCORnet Distributed Research Network.*

pcornet

# The CDM within PCORnet

- PCORnet, the Patient-Centered Outcomes Network, is a distributed research network (DRN)
  - An initiative of the Patient-Centered Outcomes Research Institute (PCORI)
- A “network of networks” to form an infrastructure with participating institutions from across the nation:
  - Answering important scientific questions in a high-throughput model
  - Allowing data from multiple data partners to be queried securely

# The CDM is a key component of the PCORnet Distributed Research Network (DRN) infrastructure



# 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
- **Developed in incremental phases**, allowing agility in deployment and new data domains to be incorporated
- Modeling is **optimized for analytic functionality** and to be **intuitive to investigators**

pcornet

Brown JS, Rusincovitch SA, Kho AN, Marsolo K, Curtis LH. Development of a National Distributed Research Network Data Infrastructure: Design of the PCORnet Common Data Model. AMIA Jt Summits Transl Sci Proc 2015; 2015: 302.

AMIA
2016 Joint Summits on Translational Science
March 21–24, San Francisco

# Iterative Development of the CDM

pcornet

AMIA

2016 Joint Summits on Translational Science
March 21–24, San Francisco

# CDM in PCORnet Phase I

- Using an iterative design process, **we released 3 major versions** in a highly accelerated **16-month development period** (February 2014 through May 2015)
- Four feedback cycles with stakeholders and expert advisors
  - **More than 700 discreet comments** received during development and incorporated into model
- The CDM v3.0 is the **final version within PCORnet Phase I**

# Details of CDM Development

**PCORnet CDM v1.0**

- Released on **May 30, 2014**
- **276 discrete comments** received and incorporated in **1 feedback cycle**

**PCORnet CDM v2.0**

- Released on **February 27, 2015**
- **265 discrete comments** received and incorporated in **2 feedback cycles**
- Two stakeholder sessions with **79 and 60 attendees**, respectively

**PCORnet CDM v3.0**

- Released on **June 1, 2015**
- **236 discreet comments** received and incorporated in **1 feedback cycle**
- Two stakeholder sessions with **57 and 60 attendees**, respectively

*CDM Guiding Principle #4:*

*"The PCORnet CDM will be developed in a* ***modular****,* ***incremental****, and* ***extensible*** *fashion…"*

# PCORnet Common Data Model, v1.0

**DEMOGRAPHIC**

PATID
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG

**Fundamental basis**

**ENROLLMENT**

PATID
ENR_START_DATE
ENR_END_DATE
CHART
ENR_BASIS

**Data captured from processes associated with healthcare delivery**

**VITAL**

PATID
ENCOUNTERID (optional)
MEASURE_DATE
MEASURE_TIME
VITAL_SOURCE
HT
WT
DIASTOLIC
SYSTOLIC
ORIGINAL_BMI
BP_POSITION

**Data captured within multiple contexts: healthcare delivery, registry activity, or directly from patients**

**ENCOUNTER**

PATID
ENCOUNTERID
ADMIT_DATE
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
ENC_TYPE
FACILITYID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE

**DIAGNOSIS**

PATID
ENCOUNTERID
*ENC_TYPE (replicated)*
*ADMIT_DATE (replicated)*
*PROVIDERID (replicated)*
DX
DX_TYPE
DX_SOURCE
PDX

**PROCEDURE**

PATID
ENCOUNTERID
*ENC_TYPE (replicated)*
*ADMIT_DATE (replicated)*
*PROVIDERID (replicated)*
PX
PX_TYPE

**Data captured from healthcare delivery, direct encounter basis**

pcornet

# PCORnet Common Data Model, v2.0

New to v2.0

**DEMOGRAPHIC**
- PATID
- BIRTH_DATE
- BIRTH_TIME
- SEX
- HISPANIC
- RACE
- BIOBANK_FLAG

Fundamental basis

**ENROLLMENT**
- PATID
- ENR_START_DATE
- ENR_END_DATE
- CHART
- ENR_BASIS

**DISPENSING**
- PATID
- DISPENSE_DATE
- NDC
- DISPENSE_SUP
- DISPENSE_AMT

Data captured from processes associated with healthcare delivery

**VITAL**
- PATID
- ENCOUNTERID (optional)
- MEASURE_DATE
- MEASURE_TIME
- VITAL_SOURCE
- HT
- WT
- DIASTOLIC
- SYSTOLIC
- ORIGINAL_BMI
- BP_POSITION
- TOBACCO
- TOBACCO_TYPE

**CONDITION**
- PATID
- ENCOUNTERID (optional)
- REPORT_DATE
- RESOLVE_DATE
- CONDITION_STATUS
- CONDITION
- CONDITION_TYPE
- CONDITION_SOURCE

**PRO_CM**
- PATID
- ENCOUNTERID (optional)
- PRO_ITEM
- PRO_LOINC
- PRO_DATE
- PRO_TIME
- PRO_RESPONSE
- PRO_METHOD
- PRO_MODE
- PRO_CAT

Data captured within multiple contexts: healthcare delivery, registry activity, or directly from patients

**ENCOUNTER**
- PATID
- ENCOUNTERID
- ADMIT_DATE
- ADMIT_TIME
- DISCHARGE_DATE
- DISCHARGE_TIME
- PROVIDERID
- FACILITY_LOCATION
- ENC_TYPE
- FACILITYID
- DISCHARGE_DISPOSITION
- DISCHARGE_STATUS
- DRG
- DRG_TYPE
- ADMITTING_SOURCE

**DIAGNOSIS**
- PATID
- ENCOUNTERID
- *ENC_TYPE (replicated)*
- *ADMIT_DATE (replicated)*
- *PROVIDERID (replicated)*
- DX
- DX_TYPE
- DX_SOURCE
- PDX

**LAB_CM_RESULT**
- PATID
- ENCOUNTERID (optional)
- LAB_NAME
- SPECIMEN_SOURCE
- LAB_LOINC
- PRIORITY
- RESULT_LOC
- LAB_PX
- LAB_PX_TYPE
- LAB_ORDER_DATE
- SPECIMEN_DATE
- SPECIMEN_TIME
- RESULT_DATE
- RESULT_TIME
- RESULT_QUAL
- RESULT_NUM
- RESULT_MODIFIER
- RESULT_UNIT
- NORM_RANGE_LOW
- MODIFIER_LOW
- NORM_RANGE_HIGH
- MODIFIER_HIGH
- ABN_IND

**PROCEDURE**
- PATID
- ENCOUNTERID
- *ENC_TYPE (replicated)*
- *ADMIT_DATE (replicated)*
- *PROVIDERID (replicated)*
- PX_DATE
- PX
- PX_TYPE
- PX_SOURCE

Data captured from healthcare delivery, direct encounter basis

pcornet

2016 Joint Summits on Translational Science
March 21–24, San Francisco

# PCORnet Common Data Model v3.0

New to v3.0

## DEMOGRAPHIC
**PATID**
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG

Fundamental basis

## ENROLLMENT
**PATID**
**ENR_START_DATE**
ENR_END_DATE
CHART
**ENR_BASIS**

## DISPENSING
**DISPENSINGID**
**PATID**
PRESCRIBINGID (optional)
**DISPENSE_DATE**
**NDC**
DISPENSE_SUP
DISPENSE_AMT

## DEATH
**PATID**
**DEATH_DATE**
DEATH_DATE_IMPUTE
**DEATH_SOURCE**
DEATH_MATCH_CONFIDENCE

## DEATH_CAUSE
**PATID**
**DEATH_CAUSE**
**DEATH_CAUSE_CODE**
**DEATH_CAUSE_TYPE**
**DEATH_CAUSE_SOURCE**
DEATH_CAUSE_CONFIDENCE

Data captured from processes associated with healthcare delivery

## VITAL
**VITALID**
**PATID**
ENCOUNTERID (optional)
**MEASURE_DATE**
MEASURE_TIME
**VITAL_SOURCE**
HT
WT
DIASTOLIC
SYSTOLIC
ORIGINAL_BMI
BP_POSITION
SMOKING
TOBACCO
TOBACCO_TYPE

## CONDITION
**CONDITIONID**
**PATID**
ENCOUNTERID (optional)
REPORT_DATE
RESOLVE_DATE
ONSET_DATE
CONDITION_STATUS
**CONDITION**
**CONDITION_TYPE**
**CONDITION_SOURCE**

## PRO_CM
**PRO_CM_ID**
**PATID**
ENCOUNTERID (optional)
**PRO_ITEM**
PRO_LOINC
**PRO_DATE**
PRO_TIME
**PRO_RESPONSE**
PRO_METHOD
PRO_MODE
PRO_CAT

Data captured within multiple contexts: healthcare delivery, registry activity, or directly from patients

## ENCOUNTER
**ENCOUNTERID**
**PATID**
**ADMIT_DATE**
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
**ENC_TYPE**
FACILITYID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE

## DIAGNOSIS
**DIAGNOSISID**
**PATID**
**ENCOUNTERID**
*ENC_TYPE (replicated)*
*ADMIT_DATE (replicated)*
*PROVIDERID (replicated)*
**DX**
**DX_TYPE**
**DX_SOURCE**
PDX

## PROCEDURES
**PROCEDURESID**
**PATID**
**ENCOUNTERID**
*ENC_TYPE (replicated)*
*ADMIT_DATE (replicated)*
*PROVIDERID (replicated)*
PX_DATE
**PX**
**PX_TYPE**
PX_SOURCE

## LAB_RESULT_CM
**LAB_RESULT_CM_ID**
**PATID**
ENCOUNTERID (optional)
LAB_NAME
SPECIMEN_SOURCE
LAB_LOINC
PRIORITY
RESULT_LOC
LAB_PX
LAB_PX_TYPE
LAB_ORDER_DATE
SPECIMEN_DATE
SPECIMEN_TIME
**RESULT_DATE**
RESULT_TIME
RESULT_QUAL
RESULT_NUM
RESULT_MODIFIER
RESULT_UNIT
NORM_RANGE_LOW
NORM_MODIFIER_LOW
NORM_RANGE_HIGH
NORM_MODIFIER_HIGH
ABN_IND

## PRESCRIBING
**PRESCRIBINGID**
**PATID**
ENCOUNTERID (optional)
RX_PROVIDERID
RX_ORDER_DATE
RX_ORDER_TIME
RX_START_DATE
RX_END_DATE
RX_QUANTITY
RX_REFILLS
RX_DAYS_SUPPLY
RX_FREQUENCY
RX_BASIS
RXNORM_CUI

Data captured from healthcare delivery, direct encounter basis

## PCORNET_TRIAL
**PATID**
**TRIALID**
**PARTICIPANTID**
TRIAL_SITEID
TRIAL_ENROLL_DATE
TRIAL_END_DATE
TRIAL_WITHDRAW_DATE
TRIAL_INVITE_CODE

Associations with PCORnet clinical trials

## HARVEST
**NETWORKID**
NETWORK_NAME
**DATAMARTID**
DATAMART_NAME
DATAMART_PLATFORM
CDM_VERSION
DATAMART_CLAIMS
DATAMART_EHR
BIRTH_DATE_MGMT
ENR_START_DATE_MGMT
ENR_END_DATE_MGMT
ADMIT_DATE_MGMT
DISCHARGE_DATE_MGMT
PX_DATE_MGMT
RX_ORDER_DATE_MGMT
RX_START_DATE_MGMT
RX_END_DATE_MGMT
DISPENSE_DATE_MGMT
LAB_ORDER_DATE_MGMT
SPECIMEN_DATE_MGMT
RESULT_DATE_MGMT
MEASURE_DATE_MGMT
ONSET_DATE_MGMT
REPORT_DATE_MGMT
RESOLVE_DATE_MGMT
PRO_DATE_MGMT
REFRESH_DEMOGRAPHIC_DATE
REFRESH_ENROLLMENT_DATE
REFRESH_ENCOUNTER_DATE
REFRESH_DIAGNOSIS_DATE
REFRESH_PROCEDURES_DATE
REFRESH_VITAL_DATE
REFRESH_DISPENSING_DATE
REFRESH_LAB_RESULT_CM_DATE
REFRESH_CONDITION_DATE
REFRESH_PRO_CM_DATE
REFRESH_PRESCRIBING_DATE
REFRESH_PCORNET_TRIAL_DATE
REFRESH_DEATH_DATE
REFRESH_DEATH_CAUSE_DATE

Process-related data

**Bold font** indicates fields that cannot be null due to primary key definitions or record-level constraints.

pcornet

AMIA

2016 Joint Summits on Translational Science
March 21–24, San Francisco

# PCORnet CDM v3.0 Development

pcornet

AMIA

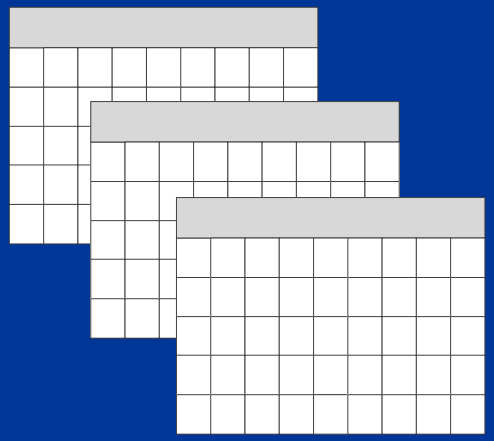
2016 Joint Summits on Translational Science
March 21–24, San Francisco

# Five primary design drivers for v3.0

1. Support of PCORnet pragmatic clinical trials
2. Support of PCORnet observational studies
3. Expanding the capabilities to represent the different workflows in medication data generation for either trial or observational purposes
4. Support of rapid querying capabilities
5. Incremental improvements based on feedback in newly-introduced feedback channels from users and implementers

# PCORNET_TRIAL serves as a connector and filter for CDM data within the parameters of a given trial protocol

**CDM Tables**
**(within a specific CDRN Datamart)**



Associate the CDM domains specified in the trial protocol

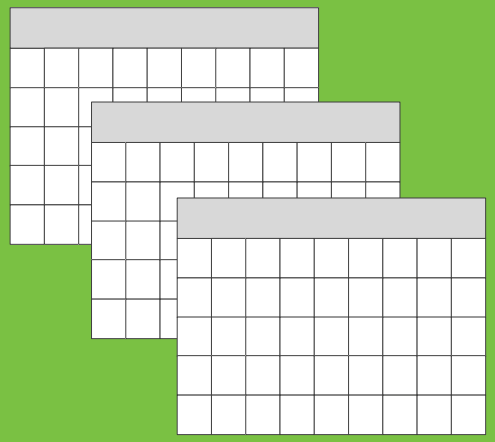
Work with CDM data in the correct timeframe

**PCORNET_TRIAL** v3.0

- PATID — Which patient?
- TRIALID — Which trial?
- PARTICIPANTID — Which person?
- TRIAL_SITEID — Which site? (may be trial-specific)
- TRIAL_ENROLL_DATE
- TRIAL_END_DATE
- TRIAL_WITHDRAW_DATE
- TRIAL_INVITE_CODE — If used by trial

Associate the study records

**PCORnet Trial Database**
For example: Adaptable

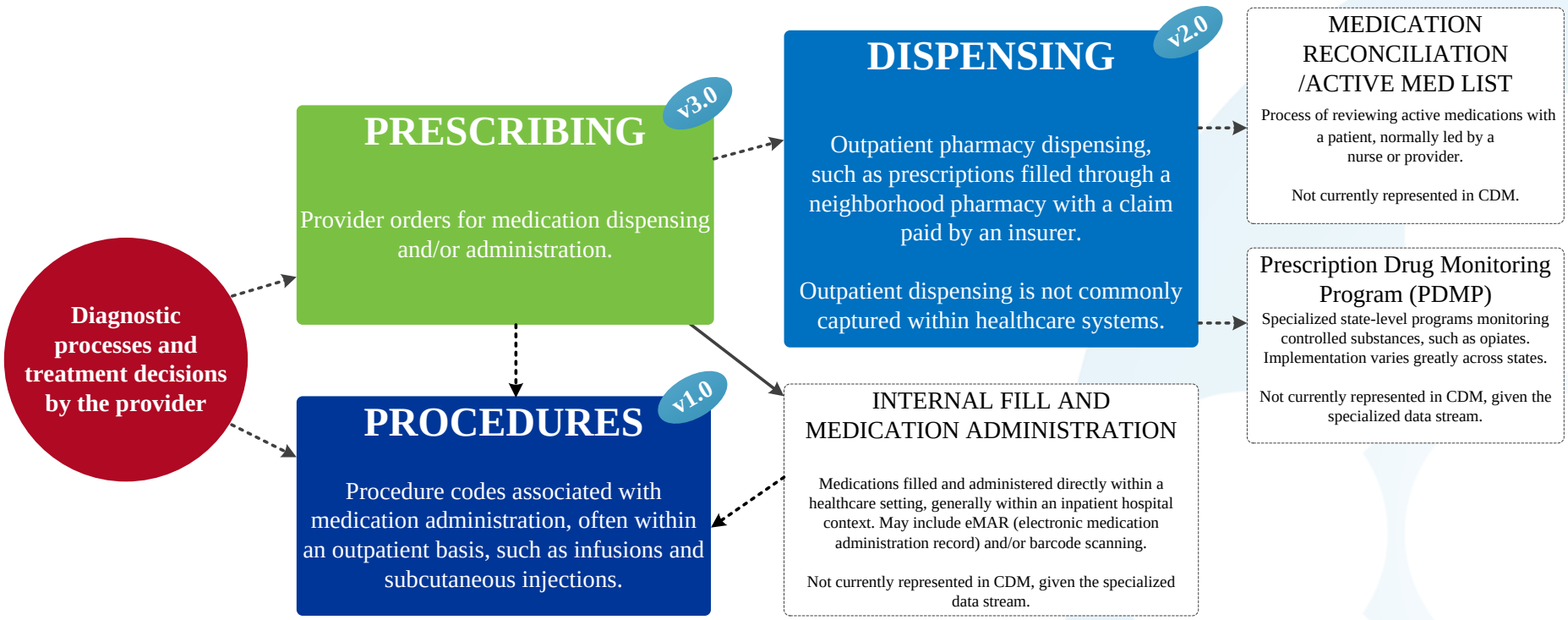


May contain:

- Consent module
- Randomization assignment
- Study-specific data collection
- Study-specific schedule of assessments

pcornet

# Medication domain modeling




pcornet

# v3.0 development process (part 1 of 2)

- **Expert guidance from the CDM Working Group**, which played an essential role in determining direction and feasibility
- Initial **presentations of design drivers** for PCORnet task force and principal investigator forums
- Draft CDM disseminated in **formal feedback cycle**; 236 discreet comments received from investigators, technical teams, and other stakeholders

# v3.0 development process (part 2 of 2)

- **Each comment was assessed and tagged** into 23 thematic categories, **then distilled** into 12 themes
- **Presented** during 2 stakeholder sessions with 57 and 60 attendees, respectively
- Final product, integrating feedback from all stakeholders, was **presented to and approved by the PCORnet Steering Committee**

# The 15 PCORnet CDM Domains, v3.0

## CONDITION v2.0

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 v3.0

Reported mortality information for patients.

## DEATH_CAUSE v3.0

The individual causes associated with a reported death.

## DEMOGRAPHIC v1.0

Demographics record the direct attributes of individual patients.

## DIAGNOSIS v1.0

Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery.

## DISPENSING v2.0

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

## ENROLLMENT v1.0

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 v1.0

Encounters are interactions between patients and providers within the context of healthcare delivery.

## HARVEST v3.0

Attributes associated with the specific PCORnet datamart implementation

## LAB_RESULT_CM v2.0

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 v3.0

Patients who are enrolled in PCORnet clinical trials.

## PRESCRIBING v3.0

Provider orders for medication dispensing and/or administration.

## PRO_CM v2.0

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 v1.0

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

## VITAL v1.0

Vital signs (such as height, weight, and blood pressure) directly measure an individual's current state of attributes.

pcornet

AMIA

2016 Joint Summits on Translational Science
March 21–24, San Francisco

# Acknowledgments

This work was funded through a the Patient-Centered Outcomes Research Institute (PCORI) Award (PCO-COORDCTR2013) for development of the National Patient-Centered Clinical Research Network, known as PCORnet.

The views, statements, and opinions presented in this work are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee or other participants in PCORnet.

pcornet

The PCORnet Common Data Model is a product of shared expertise, feedback, effort, and collaboration from many people. We gratefully acknowledge their work and individual contributions.

pcornet

AMIA

# Links and further reading

The Common Data Model specification, glossary, and lay guide: http://www.pcornet.org/pcornet-common-data-model/

CDM Forum and Errata on GitHub: https://github.com/CDMFORUM

Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. Journal of the American Medical Informatics Association: JAMIA. 2014;21(4):578-82. http://www.ncbi.nlm.nih.gov/pubmed/24821743

"Why PCORnet Exists": http://www.pcornet.org/why-pcornet-exists/

"PCORnet Phase II: What will success look like?" (video): https://www.youtube.com/watch?v=MBnUqomQ8vI

These slides are posted on GitHub in the CDM Guidance Repository:

https://github.com/CDMFORUM/CDM-GUIDANCE/wiki/CDM-related-Abstracts

# Contact Information


Shelley A. Rusincovitch

Project Leader in Applied Informatics and Architecture
Duke Translational Research Institute (DTRI)
shelley.rusincovitch@duke.edu

Twitter: https://twitter.com/rusincovitch
GitHub: https://github.com/rusincovitch
LinkedIn: https://www.linkedin.com/in/sarusincovitch


DukeHealth

AMIA
2016 Joint Summits on Translational Science
March 21–24, San Francisco

# Supplemental Slides

## Guiding Principles for the PCORnet Common Data Model (CDM)

PCORnet Data Standards, Security, and Networking Infrastructure Task Force, May 15, 2014

1. PCORnet member networks will develop a model for a physical or logical data set that will be implemented by each CDRN and PPRN. This common data model (CDM) will harmonize names, values, and definitions for data elements of general interest to PCORnet.

2. It is not expected that all CDRNs and PPRNs will be able to populate all parts of the PCORnet CDM. It is the responsibility of the CDRNs and PPRNs to communicate availability of each data domain and element.

3. The DSSNI Task Force, with guidance from other Task Forces as needed, will be responsible for development and maintenance of the PCORnet CDM. Each PPRN and CDRN will be expected to participate so that the CDM reflects local input.

4. The PCORnet CDM will be developed in a modular, incremental, and extensible fashion. New types of data will be needed, or newly available, during the life of PCORnet. Data domains and data elements will be added, revised, and deprecated throughout an iterative CDM lifecycle. Personnel from the CDRNs and PPRNs will work with the DSSNI Task Force (and other Task Forces as appropriate) to assist in these efforts.

5. Documentation will be clear and transparent so that its contents are understandable to all contributors. The CDM will be intuitive and easy for analysts and investigators to use. Investigators and analysts with prior experience using research data will not need additional skills or knowledge to use the CDM.

6. Other common data element and common data model initiatives exist. PCORnet will draw from the experience of others within and outside of PCORI, leveraging existing successful approaches and data model definitions wherever possible.

7. The CDM will reflect variables and values found in the local data. If some data are coded in a way that is unique to a site, mapping the data to a standardized format will be necessary. Values in the source data before mapping will also be included in the CDM. Derived variables should be avoided.

8. CDRNS and PPRNs may include additional domains and data elements in localized versions of the PCORnet CDM.

pcornet

# Distributed analysis methods that minimize the need to share patient-level data

- Only the minimum information necessary should be requested and shared
- Coordinating Center oversees minimum necessary policy implementation
- Many analyses can be completed without sharing any protected information
  - Risk sets
  - Propensity scores
  - Highly aggregated and summarized person-level information

Slide from DCRI Research Conference on September 22, 2015, “Moving Beyond the Blueprint: Doing Research in PCORnet," presented by Adrian Hernandez, MD, and Lesley Curtis, PhD.

Watch on YouTube at:
https://www.youtube.com/watch?v=dP8rJNV3nXI

pcornet

# Networks maintain physical and operational control over their data

- Distributed approach allows partners to maintain control of their data and all uses
- Partners have option to review requests before execution and review results before release
- No need to change local workflow related to release of information
- All activities secure and audited
- All queries are distributed through the PCORnet DRN Portal

Slide from DCRI Research Conference on September 22, 2015, "Moving Beyond the Blueprint: Doing Research in PCORnet," presented by Adrian Hernandez, MD, and Lesley Curtis, PhD.

Watch on YouTube at:
https://www.youtube.com/watch?v=dP8rJNV3nXI

pcornet

AMIA

2016 Joint Summits on Translational Science
March 21–24, San Francisco