



OMOP – PCORnet Alignment

Rimma Belenkaya

26-April-2016

Rimma.Belenkaya@gmail.com



Outline

- History of OMOP adoption for CDRN
 - Development of crosswalk
 - ✓ Approach
 - ✓ Challenges
 - ✓ Outcomes
 - ✓ Remaining tasks
-



History of OMOP adoption for CDRN

- CDRN– Clinical Data Research Network
- PCORnet - the National Patient-Centered Clinical Research Network
 - ✓ Initiated in 2014
 - ✓ Funded by the Patient-Centered Outcomes Research Institute (PCORI)
 - ✓ Integrates data from 13 networks for CER
 - ✓ PCORnet CDM is based on the Mini-Sentinel CDM
- Out of initial 11 networks
 - ✓ 7 adopted i2b2
 - ✓ 4 adopted OMOP
 - ✓ 0 adopted PCORnet



PCORNet CDM V1

PCORnet Common Data Model, Original 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

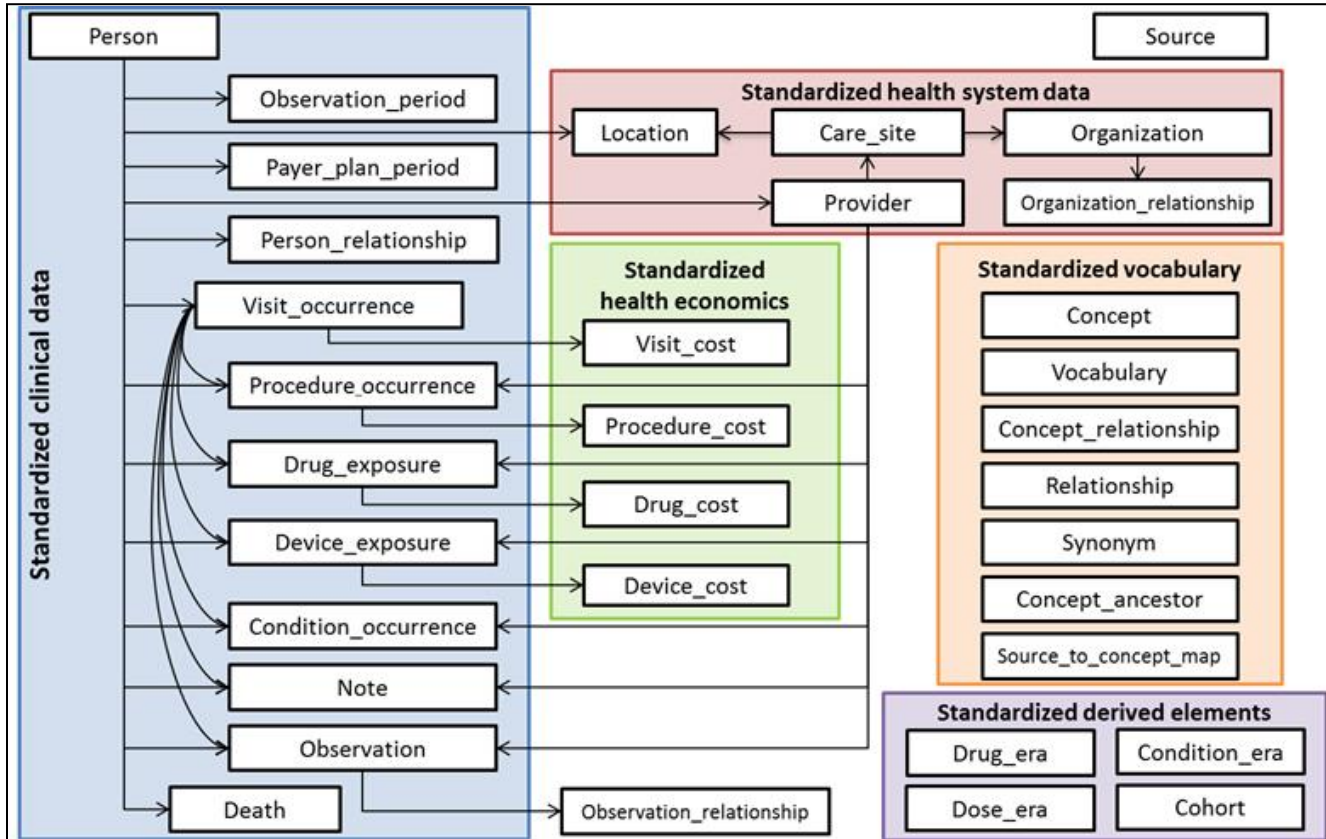
DIAGNOSIS
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
DX
DX_TYPE
DX_SOURCE
PDX

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

Data captured from healthcare delivery, direct encounter basis



OMOP CDM v4



Rationale for Choosing OMOP CDM

	C32 CCD	MiniSentinel	OMOP V5	PCORI CDM V1	SCILHS PCORI SHRINE	HARVARD SHRINE	NYC DMH QUERY HEALTH
Patient Demographics	X	X	X	X	X(HL7, ISO, CDC)	X(HL7, ISO, CDC)	X
Provider	X	X	X(NPI, DEA, Specialty(CDC))	X	O	O	X(NPI)
Encounters	X	X	X	X	X(in progress)	O	X(CPT, NQF)
Encounter cost	O	DRG	O	DRG	O	O	X
Locations	O	X	X	X	O	O	X
Care sites	O	X(3 CHAR ZIP)	X(CMS codes)	X(3 CHAR ZIP)			X
Organizations	O	X	X	O			X
Reason for encounter	O	O	X	O	O	O	X
Problem List(Condition)	X (SNOMED CT, ICD-*-CM)	O	X (SNOMED CT; MedDRA)	O	X(ICD-9-CM)	X(ICD-9-CM)	X(SNOMED CT, ICD9)
Encounter Diagnoses		X	X (SNOMED CT)	X			X
Principal Dx	O	X(ICD-9-CM; SNOMED CT; ICD-10-CM)	X (SNOMED CT)	X(ICD-9-CM; SNOMED CT; ICD-10-CM)			O
Cause of death	O	X(ICD-*-CM)	X (SNOMED CT)				O
Order diagnosis	O	O	X	O			O
	O		X (via relevant_condition_concept_id)	O			
Family History	PersonalRelationship:SNOMEDCT<<3030710014 Condition:SNOMEDCT<<ClinFinding	O	X	O	O	O	X
Social History	X	O	X (SNOMED CT)	O	O	O	X
Review of systems		O	O	O	O	O	X
Alerts/Allergies	X (SNOMED CT, RxNORM)	O	X (SNOMED CT)	O	O	O	X(SNOMED CT)
Vital Signs	X	X	X (LOINC, SNOMED CT)	X	X (some, in progress)	O	X(LOINC)
Physical findings	X (SNOMED CT, LOINC)	O	X	O			X
Lab Results	X (LOINC)	X (LOINC, CPT, Specimen)	X (LOINC, UCUM)	O	X(LOINC)	X(LOINC)	X(LOINC, NQF)
Radiology results					O	O	X
Medication history	X (RxNORM)	O	X (RxNORM; NDF-RT)	O	X(RxNORM, NDF-RT)	X(RxNORM, NDF-RT)	X(NDC, NF, RXNORM)
Medications prescribed(outpt)	X	O	X	O			X
Administered	X	O	X	O			X(NDC, NQF, RXNORM)
Rx dispensed	O	X (NDC)	X	O			X
Medication costs	O	O	X	O	O	O	X(NDC)



Rationale for Choosing OMOP CDM

- **Data model** - more reflective of clinical domain, granular, better structured
- **Vocabulary** - uniformly structured and well curated
- **Information Model** - formalized connection between data model and conceptual model (vocabulary)
- **Supportive Community** of developers and researches
- **Tools** for data integration, characterization, and analysis



Development of CDM Crosswalk



Step 1. Addressing PCORnet Requirements in OMOP CDM v5

Step 2. Establishing interoperability standards



PCORnet requirements in OMOP CDM v5

- ✓ New attributes and concepts added to Person, Visit, and Observation Period domains

OBSERVATION_PERIOD			
Field	Type	Required	Description
Observation_period_id	integer	Yes	A system-generate unique identifier for each observation period
person_id	integer	Yes	A foreign key identifier to the person who is experiencing the condition. The demographic details of that person are stored in the person table.
observation_period_start_date	date	Yes	The start date of the observation period for which data are available from the data source
observation_period_end_date	date	Yes	The end date of the observation period for which data are available from the source.
period_type_concept_id	integer	Yes	A foreign key identifier to the predefined concept in the Standardized Vocabularies reflecting the source of the observation period information
.....			

44814722	Period while enrolled in insurance'
44814723	Geography based period'
44814725	Period inferred by algorithm'
44814724	Period covering healthcare encounters'



PCORnet requirements in OMOP CDM v5

- ✓ Observations and fact relationships used as a workaround for unsupported attributes

VISIT_OCCURRENCE	
Field	Value
visit_occurrence_id	555555
person_id	804860
visit_start_date	10/10/2012
visit_start_time	23:00
visit_end_date	10/14/2012
visit_end_time	14:15
provider_id	666
.....	

OBSERVATION		
Field	Value	
observation_id	238923849	
visit_occurrence_id	555555	
person_id	804860	
observation_date	10/14/2012	
observation_time	14:15	
provider_id	666	
observation_concept_id	4137274	Discharge to establishment
value_as_concept_id	44814680	Residential Facility
.....		



PCORnet requirements in OMOP CDM v5

- ✓ PCORnet vocabulary created and mapped to standard vocabulary

CONCEPT_ID	CONCEPT_NAME	CONCEPT_LEVEL	CONCEPT_CLASS	VOCABULARY_ID	CONCEPT_CODE
44814649	Other	0	Generic	PCORNet	44814649-NI
44814650	No information	0	Generic	PCORNet	44814650-UN
44814653	Unknown	0	Generic	PCORNet	44814653-OT
44814647	Availbe in biobank	0	Biobank Flag	PCORNet	44814647-Y
44814648	Unavailable in biobank	0	Biobank Flag	PCORNet	44814648-N
44814651	Hispanic	0	Hispanic	PCORNet	44814651-Y
44814652	Not Hispanic	0	Hispanic	PCORNet	44814652-N
44814654	American Indian or Alaska Native	0	Race	PCORNet	44814654-01
44814655	Asian	0	Race	PCORNet	44814655-02
44814656	Black or African American	0	Race	PCORNet	44814656-03
44814657	Native Hawaiian or Other Pacific Islander	0	Race	PCORNet	44814657-04
44814658	White	0	Race	PCORNet	44814658-05
44814659	Multiple race	0	Race	PCORNet	44814659-06
44814660	Refuse to answer	0	Race	PCORNet	44814660-07
44814664	Ambiguous	0	Sex	PCORNet	44814664-A
44814665	Female	0	Sex	PCORNet	44814665-F
44814666	Male	0	Sex	PCORNet	44814666-M
44814670	Adult foster home	0	Admitting Source	PCORNet	44814670-AF
44814671	Assisted living facility	0	Admitting Source	PCORNet	44814671-AL
44814672	Ambulatory visit	0	Admitting Source	PCORNet	44814672-AV
44814673	Emergency department	0	Admitting Source	PCORNet	44814673-ED
44814674	Home health	0	Admitting Source	PCORNet	44814674-HH
44814675	Home / self care	0	Admitting Source	PCORNet	44814675-HO
44814676	Hospice	0	Admitting Source	PCORNet	44814676-HS
44814677	Other acute inpatient hospital	0	Admitting Source	PCORNet	44814677-IP



CDM Interoperability Collaborative

- Core members

- Rimma Belenkaya
- Rob Follett
- Ritu Khare
- Mark Khayter
- Toan Ong
- Lisa Schilling
- Don Torok



- Weekly meetings

- Deliverables: standard specifications

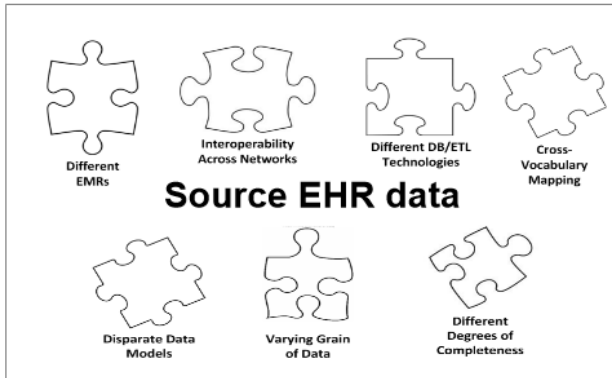


Approach

Crosswalk from source – to – OMOP CDM – to – PCORnet CDM

OMOP CDM Population Conventions

Patient->Person



OMOP Data Model



PCORnet CDM ETL Standards

Person -> Demographic

PCORnet Data Model



- ✓ Satisfy PCORnet-requirements
- ✓ Preserve granularity of source data
- ✓ Preserve integrity of OMOP CDM
- ✓ Enforce uniformity of workarounds

Convert data in the OMOP CDM to PCORnet CDM via a set of mechanistic transformation rules



Examples

✓ Extension of OMOP vocabulary to preserve null-values

NULL - values	Defintion
0	A data field is not present in the source system
No Information	A data field is present in the source system, but the source value is null or blank
Unknown	A data field is present in the source system, but the source value explicitly denotes an unknown value
Other	A data field is present in the source system, but the source value cannot be mapped to the CDM



Examples

- ✓ Convention for enforcing PCORnet referential integrity and cardinality
- ✓ Preserving integrity of OMOP CDM
- ✓ Enforcing uniformity of workarounds

VISIT_OCCURRENCE	
Field	Value
visit_occurrence_id	555555
person_id	804860
visit_start_date	10/10/2012
visit_start_time	23:00
visit_end_date	10/14/2012
visit_end_time	14:15
provider_id	666
.....	

OBSERVATION		
Field	Value	
observation_id	238923849	
visit_occurrence_id	555555	
person_id	804860	
observation_date	10/14/2012	
observation_time	14:15	
provider_id	666	
observation_concept_id	4137274	Discharge to establishment
value_as_concept_id	44814680	Residential Facility
.....		

Only one record with the same observation_concept_id per visit occurrence



Examples

✓ Convention for enforcing data completeness

VISIT_OCCURRENCE		
Field	Value	
visit_occurrence_id	555555	
person_id	804860	
visit_start_date	10/10/2012	
visit_start_time	23:00	
visit_end_date	10/14/2012	
visit_end_time	14:15	
provider_id	666	
Inpatient visit	visit_concep_id	9201
	

OBSERVATION		
Field	Value	
observation_id	238923849	
visit_occurrence_id	555555	
person_id	804860	
observation_date	10/14/2012	
observation_time	14:15	
provider_id	666	
observation_concept_id	4137274	Discharge to establishment
value_as_concept_id	44814680	Residential Facility
.....		

Diagram showing data mapping from VISIT_OCCURRENCE to OBSERVATION. Blue arrows point from the following fields in VISIT_OCCURRENCE to their corresponding fields in OBSERVATION: visit_occurrence_id, person_id, visit_end_date, visit_end_time, and provider_id.

Inpatient visits must have 'Discharge to establishment'



Examples

Granularity of source data preserved

- ✓ More granular Race representation
 - ✓ Multiple encounters to the same provider on the same date
 - ✓ Storing diagnosis date/time different from encounter date/time
-



Challenges

- **Missing attributes** - workarounds, overhead in ETL
 - **Missing concepts ('NULL' values)** - necessity to extend vocabulary
 - **Lab mappings** – PCORnet targets not defined
 - **PCORnet requirement to provide source values** – domain mismatch
 - **Different versions** of OMOP and PCORnet CDM, necessity to reconcile
 - **Time consuming** process of creating conventions
-



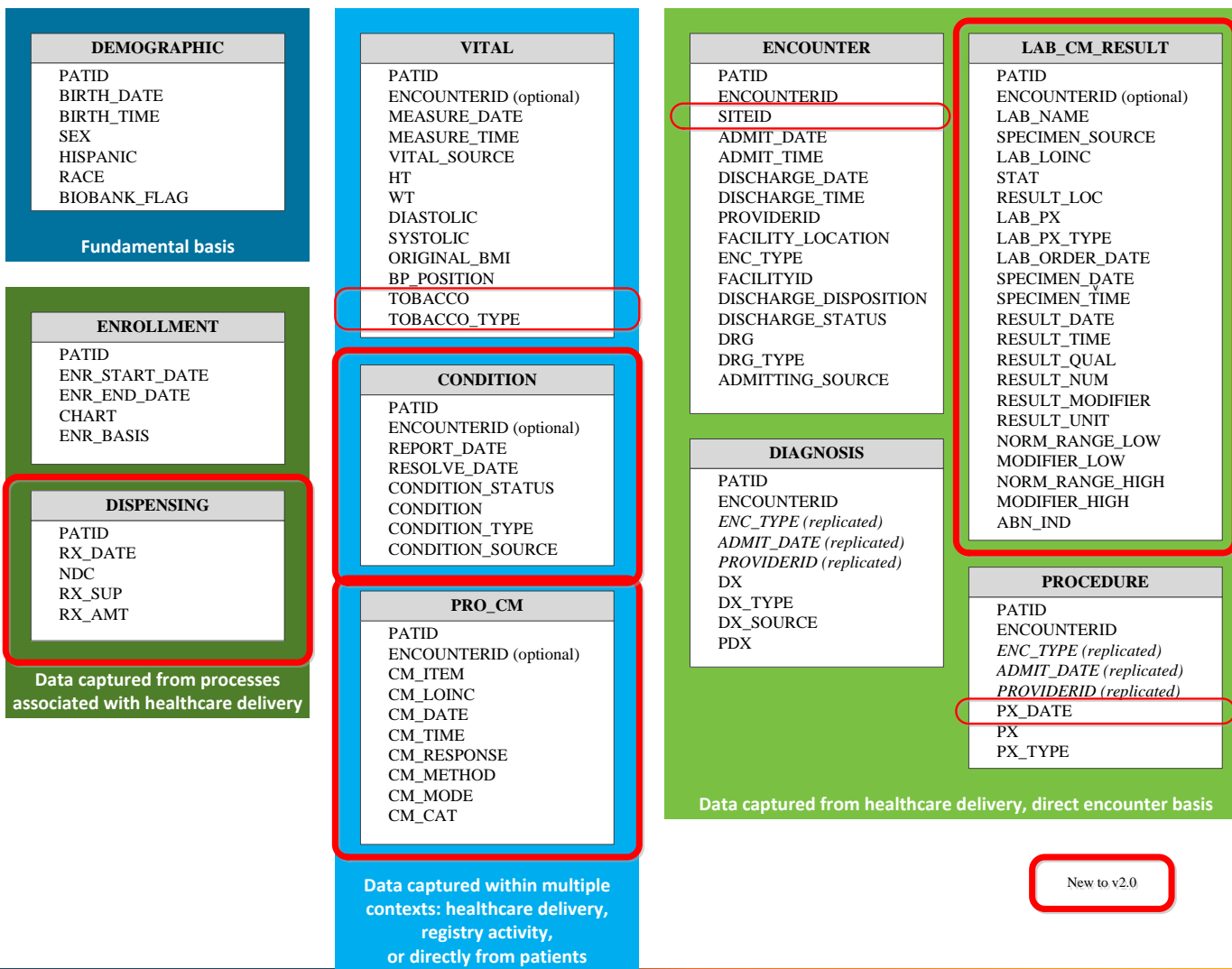
Outcomes

- **Interoperability standards** - from source (i.e. electronic health records) to OMOP CDM v4 (Specific to PCORI CDM) and from source to OMOP CDM v5 (specific to PCORI CDM)
- **A conventions document** - for populating the OMOP CDM, and an extract-transform-load specifications document for transforming to the PCORnet CDM v1, for both OMOP CDM v4 and OMOP CDM v5. The documents will be publicly available for the OHDSI community
- **OMOP CDM extended** – new important use cases
 - ✓ Added visit time
 - ✓ ADT – admission/discharge handling
 - ✓ Tobacco conventions – underway
- **OMOP CDM supports CDRN phenotypes** – ADAPTABLE study



Evolution of PCORnet CDMs

PCORnet Common Data Model, Draft v2.0 Modifications



Evolution of PCORnet CDMs

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_CONDITION
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

PCORnet Common Data Model v3.0

New to v3.0

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

Data captured from healthcare delivery, direct encounter basis

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

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



Remaining Tasks

- Complete OMOP v5 to PCORnet v3 CDM standards:
 - Death
 - Reconciliation between PCORnet v1, v2, and v3
- Make ETL code available in GitHub (pSCANNER)



Conclusions

- **Greater granularity of data representation** in OMOP CDM supports
 - Preservation of source data granularity
 - Straightforward transformation to less granular PCORnet
- **OMOP CDM is extendable** without breaking the CDM structure
 - Attributes that do not exist in OMOP can be represented without altering the standard table schema but require a set of documented conventions
 - OMOP CDM can support phenotype definitions not supported in PCORnet CDM
 - Adding non-standard/local concepts cover for missing vocabulary
- **OMOP CDM is a good intermediary data representation** when converting various healthcare datasets to PCORnet CDM



Great collaborative experience

NYC-CDRN

Parsa Mirhaji
Tom Champion
Rajan Chandras
Claudia Pulgarin
Ezra Fass
Karthik Natarajan
Uday Evani
Adler Perotte
Joan Leavey
Jim Singer



Patrick Ryan
Chris Reich



Mark Khayter
Don Torok
Lisa Schilling
Rob Follett
Daniella Meeker



Ritu Khare
Toan Ong
Michael Kahn



Shelley
Rusincovitch