Update: VINCI Observational Medical Outcomes Partnership (OMOP) Project

June 22, 2016

VINCI OMOP and OHDSI Services

Stephen Deppen, PhD

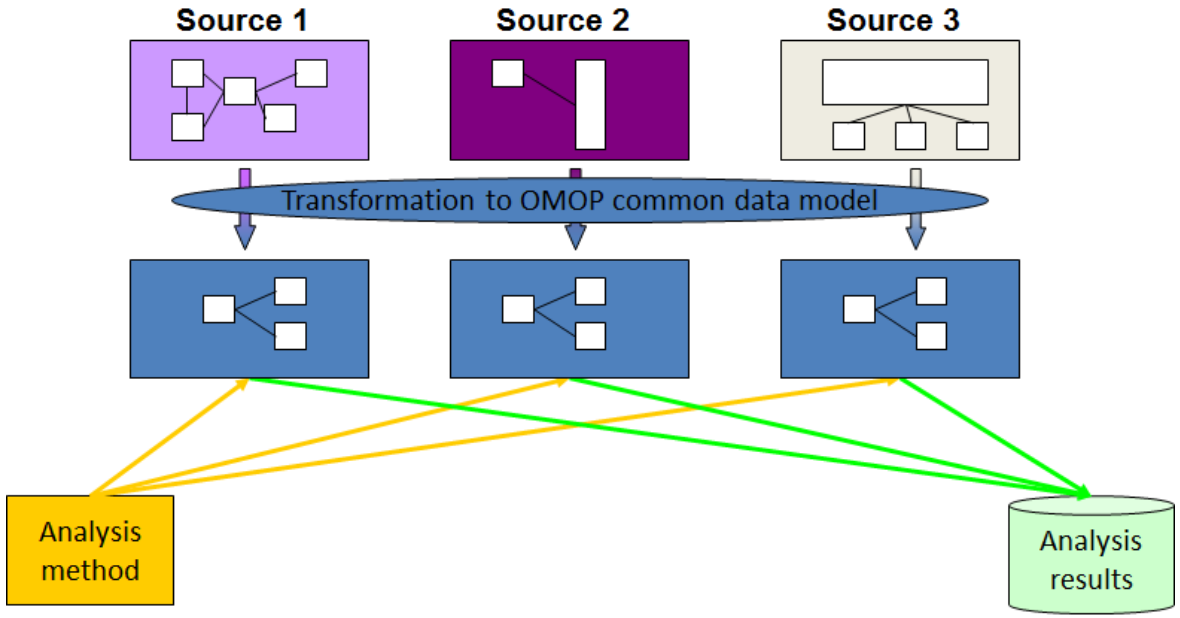
Goals

- Where are we with the VINCI - Observational Medical Outcomes Partnership (OMOP) project?
- Understanding some internal processes for VINCI-OMOP (Development, QA, Tools, Help Desk, etc)
- How to be ready for using VINCI-OMOP when it becomes available.

Outline

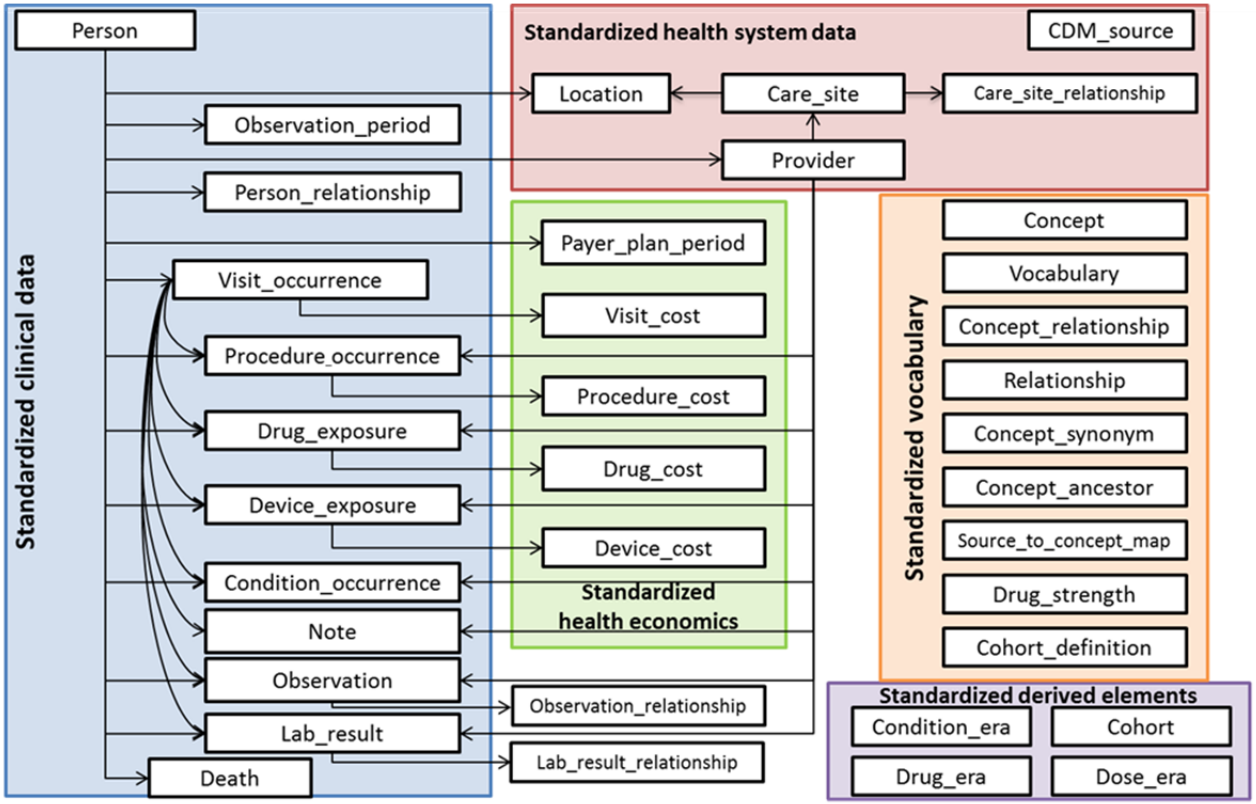
- Background
  - Brief intro to OMOP common data model
  - v5 OMOP and status
- QA process
- Timeline
  - Updates, architecture and process testing
- ODHSI – Achilles, Hermes, Cohort Design
- Getting Access being ready

Common Data Model is ?



- **Method for organizing data into a standard structure**
- **The standardized format allows for the systematic analysis of disparate observational databases**
- **The Observational Medical Outcomes Partnership (OMOP) CDM, now entering its fifth version.**
-

OMOP Version 5



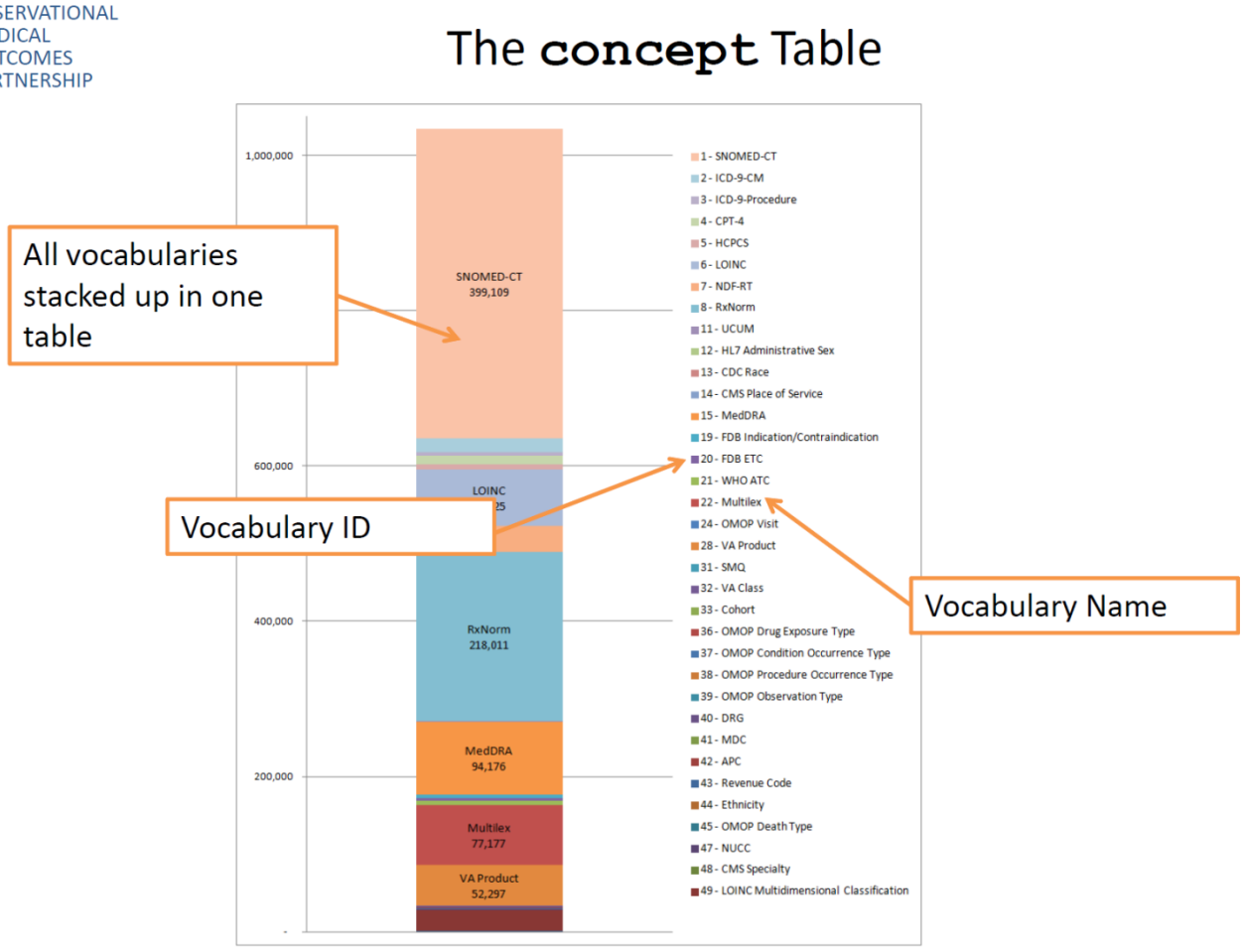
http://www.ohdsi.org/resources/publications/

OMOP v5

- Change of table structure to exploit the many to many mappings some concepts provide e.g some diagnoses also indicate procedures
- Necessary to use OHDSI suite of tools
- Applicability (and code) to transform to other common data models (PCORnet)
- Breaking out some of the larger tables smaller more manageable (observation into observation and measurement/lab results)
- ~14TB of data

Vocabularies in Concept Table

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP



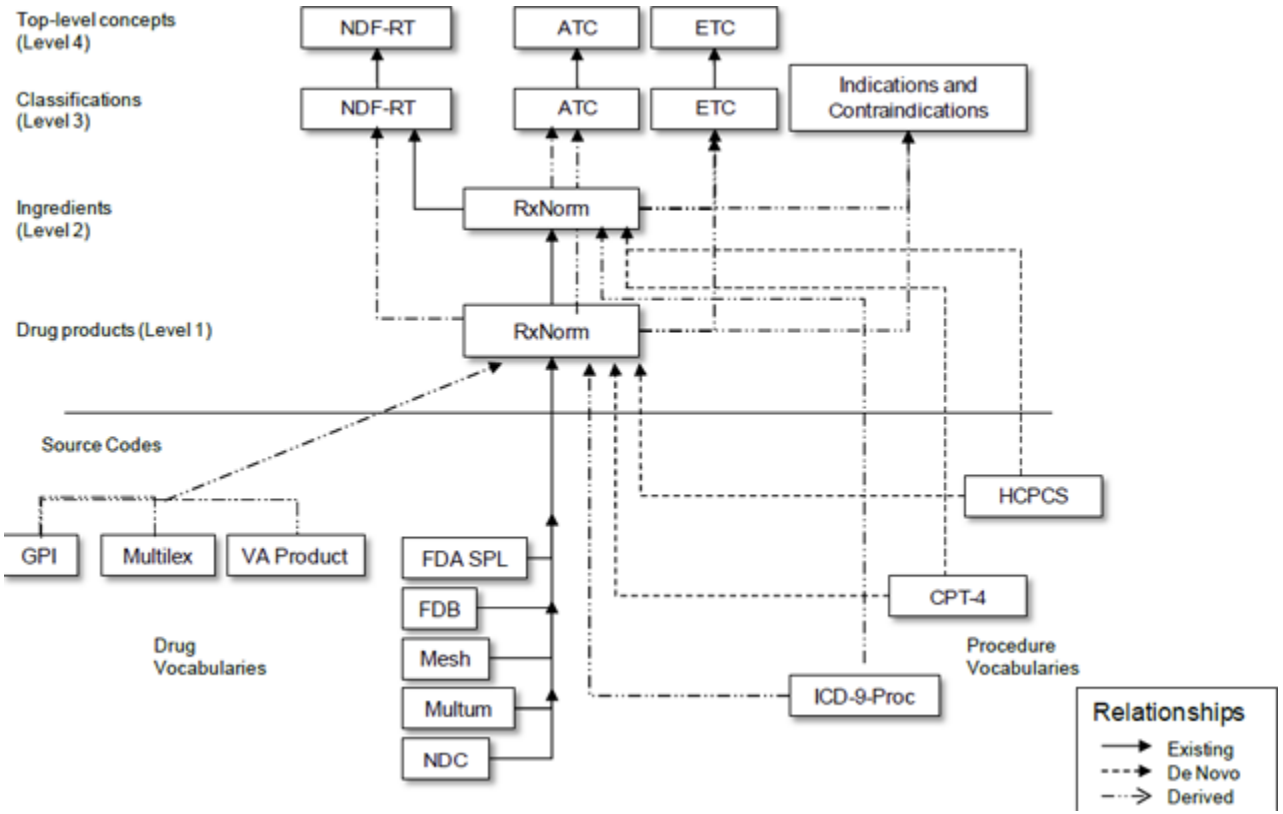
Standardized vocabulary

Standard Vocabulary	**Data type or source**
Snomed-CT	ICD-9/10, Pathology, HCPC
LOINC	Laboratory results
RXNorm,	Drugs, ingredients and
VA class, ETC and NDF-RT	Indications from RxOut, BCMA, HCPCS Drugs

Hierarchical Ladder for Drugs, Example

Concept ID	Concept Name	Concept Level	Concept Class	Vocabulary	Concept Code
19034886	Omeprazole 20 MG Enteric Coated Capsule [Prilosec]	1	Branded Drug	8	207212
923645	Omeprazole	2	Ingredient	8	7646
4319354	2-Pyridinylmethylsulfinylbenzimidazoles	3	Chemical Structure	7	N0000175098
4351005	Sulfoxides	3	Chemical Structure	7	N0000008055
4350914	Heterocyclic Compounds	3	Chemical Structure	7	N0000008095
4351444	Benzimidazoles	3	Chemical Structure	7	N0000007536
4340570	Infectious Diseases	3	Indication or Contraindication	7	N0000000007
4344424	Paraneoplastic Endocrine Syndromes	3	Indication or Contraindication	7	N0000002143
4342919	Esophagitis	3	Indication or Contraindication	7	N0000001165

Drug Domain Vocabularies



OMOP.org -standard vocabulary specification

Transformed CDW Domains

- SPatient
- BCMA
- Rxout
- Inpat
- Outpat
- Vitals/labs
- Chem
- Staff
- Vital Status
- Locations
- Institutions
- Dental

Further Education On OMOP

- Cyberseminar – Introduction to OMOP 2/4/2016

http://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=1112

- OHDSI.org website www.ohdsi.org
- VA Pulse – VINCI OMOP users group

https://www.vapulse.net/groups/vinci-omop-users-group

- Current documentation, FAQ, ETL and QA

Broad QA Process

- Issue Discovery
- Definition of Root Cause
  - Comparison to source
  - Examination of OMOP mapping or ETL
- Solution creation and testing
- Recommendation to ETL
- Documentation of issue, solution, addition to FAQ when necessary

QA Search Process

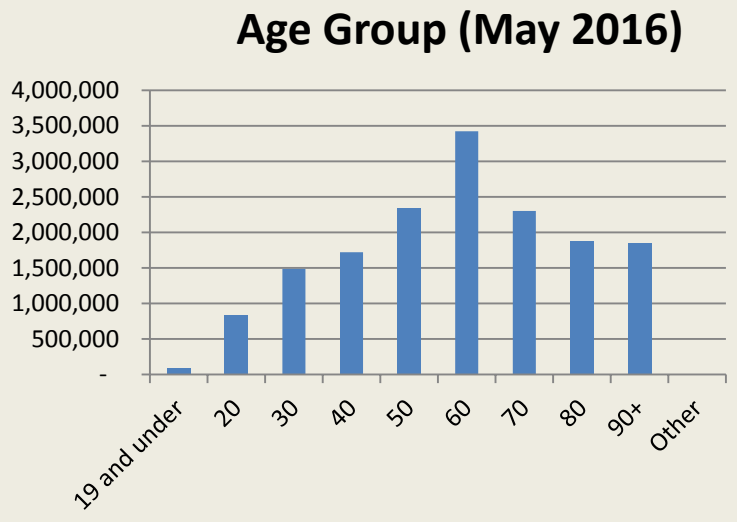
- Generic Checks for Concept Mapping by Table
- Null Value Searches
  - Source data without concept or higher level (RxNorm) mapping
- Volume/Activity Review
- Value-Range Review
- Hierarchy/Vocabulary Mapping Review
- Specific, High Interest Data Review
  - By data group, i.e. top 10 drugs by volume or class
  - Feasibility study Cardiac disease post-HIV diagnosis?
  - QI indicator: Narcan +/- 24hrs of opioid administration

QA Process

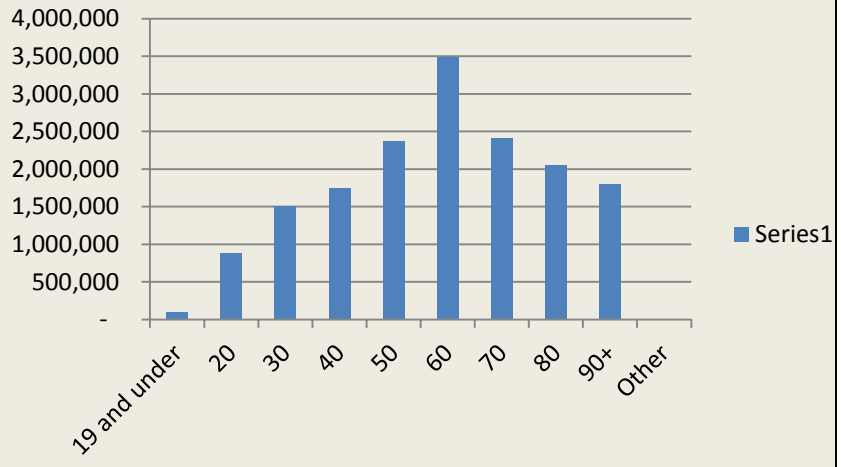
CONCEPT ID MAPPING

TABLE	TABLE_COUNT	TYPE	DISTINCT_CONCEPT_ID_MAPPED	MAPPED_VOLUME	PROPORTION_MAPPED_BY_VOLUME
CARE_SITE	1,071,307	PLACE_OF_SERVICE	22	866,334	81%
CONDITION_OCCURRENCE	2,315,173,933	CONDITION	19,038	2,184,238,751	94%
DRUG_EXPOSURE	2,305,990,943	DRUG	29,522	2,190,937,061	95%
OBSERVATION	3,734,118,158	OBSERVATION	195,832	3,733,962,487	100%
PERSON	22,736,834	GENDER	3	22,736,830	100%
PERSON	22,736,834	RACE	5	22,736,834	100%
PERSON	22,736,834	ETHNICITY	3	22,736,834	100%
PROCEDURE_OCCURRENCE	2,097,492,395	PROCEDURE	20,215	2,093,873,887	100%
PROVIDER	6,173,329	GENDER	3	3,206,189	52%
PROVIDER	6,173,329	SPECIALTY	680	1,088,504	18%
VISIT_OCCURRENCE	2,353,034,478	VISIT	1,246,926	2,352,813,115	100%

Value-Range Review



OMOP Person (excluding death patients)



CDW Spatient_Spatient (excluding death patients)

Specific Data Review

LocalDrugSID	Volume (OMOP)	Volume (CDW)	Difference (cdw-omop)	Diff proportion to CDW	Drugname	NDC (CDW)	Strength (CDW)	Dosage form (CDW)	Drug Unit (CDW)
1011846	402832	408602	5770	0.014121321	METFORMIN HCL 1000MG TAB METFORMIN HCL 1000MG TAB METFORMIN METFORMIN METFORMIN HCL 1000MG TAB	42291-0607-10	1000	TAB,ORAL	MG
	375326	388740	13414	0.034506354	METFORMIN HCL 1000MG TAB METFORMIN HCL 1000MG TAB METFORMIN METFORMIN (IMMEDIATE RELEASE) METFORMIN HCL 1000MG TAB	60429-0113-05	1000	TAB,ORAL	MG
1400008772	333339	341302	7963	0.023331243	METFORMIN HCL 500MG TAB METFORMIN HCL 500MG TAB METFORMIN METFORMIN METFORMIN HCL 500MG TAB	60429-0111-10	500	TAB,ORAL	MG
22516	311548	317220	5672	0.017880335	METFORMIN HCL 500MG TAB METFORMIN HCL 500MG TAB METFORMIN metFORMIN METFORMIN HCL 500MG TAB	60429-0111-10	500	TAB,ORAL	MG
408563	305866	310835	4969	0.015985973	METFORMIN HCL 500MG TAB METFORMIN HCL 500MG TAB METFORMIN METFORMIN METFORMIN HCL 500MG TAB	57664-0397-18	500	TAB,ORAL	MG

DRUG_SOURCE_VALUE	CONCEPT_NAME	CONCEPT_CODE(RxNorm)	x_NDC_Code	VOLUME
HYDROCHLOROTHIAZIDE 25MG TAB	Hydrochlorothiazide 25 MG Oral Tablet	310798	00603-3856-32	19,725,440
HYDROCHLOROTHIAZIDE 25MG TAB	Hydrochlorothiazide 25 MG Oral Tablet	310798	00172-2083-80	5,871,624
HYDROCHLOROTHIAZIDE 25MG TAB	Hydrochlorothiazide 25 MG Oral Tablet	310798	00603-3856-34	2,030,477
HYDROCHLOROTHIAZIDE 25MG TAB	Hydrochlorothiazide 25 MG Oral Tablet	310798	64125-0131-10	1,019,714
HYDROCHLOROTHIAZIDE 25MG TAB	Hydrochlorothiazide 25 MG Oral Tablet	310798	00172-2083-60	811,699
HYDROCHLOROTHIAZIDE 25MG TAB	Hydrochlorothiazide 25 MG Oral Tablet	310798	63739-0128-10	347,313
HYDROCHLOROTHIAZIDE 25MG TAB	Hydrochlorothiazide 25 MG Oral Tablet	310798	00603-3856-21	233,169
HYDROCHLOROTHIAZIDE 50MG TAB	Hydrochlorothiazide 25 MG Oral Tablet	310798	00603-3856-32	14,258
HYDROCHLOROTHIAZIDE 25MG TAB,UD	Hydrochlorothiazide 25 MG Oral Tablet	310798	00182-0556-89	412

Timeline

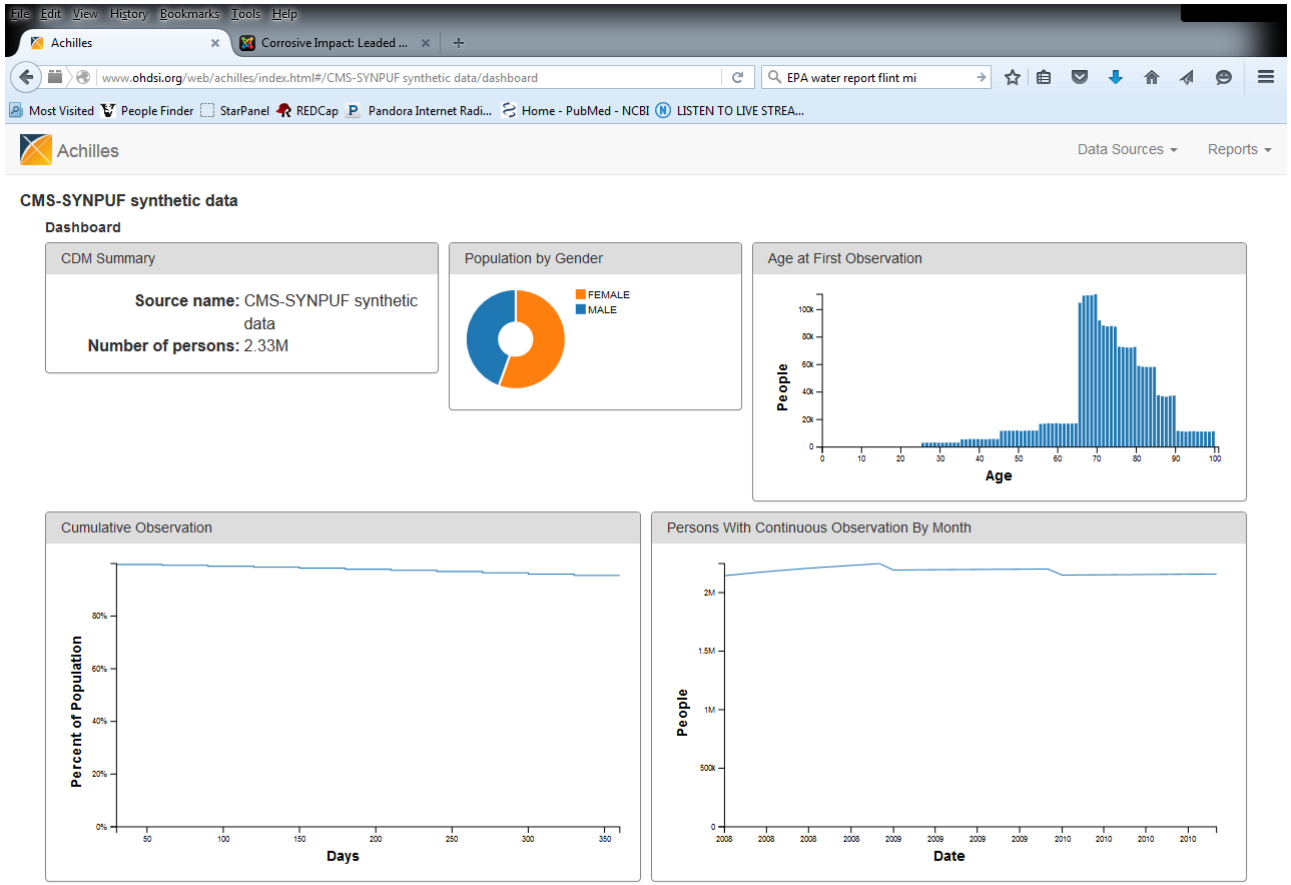
- June
  - Populating OMOP v5 Tables
  - QA Person, Death, Condition, Visit, Drug, Vitals and Observation Tables
- July
  - QA Location, Provider, Labs
  - Solutions and table reloads for June discoveries
  - Incremental load ETL for v5
  - External Group QA
  - ETL, QA and Documentation

Timeline (Cont)

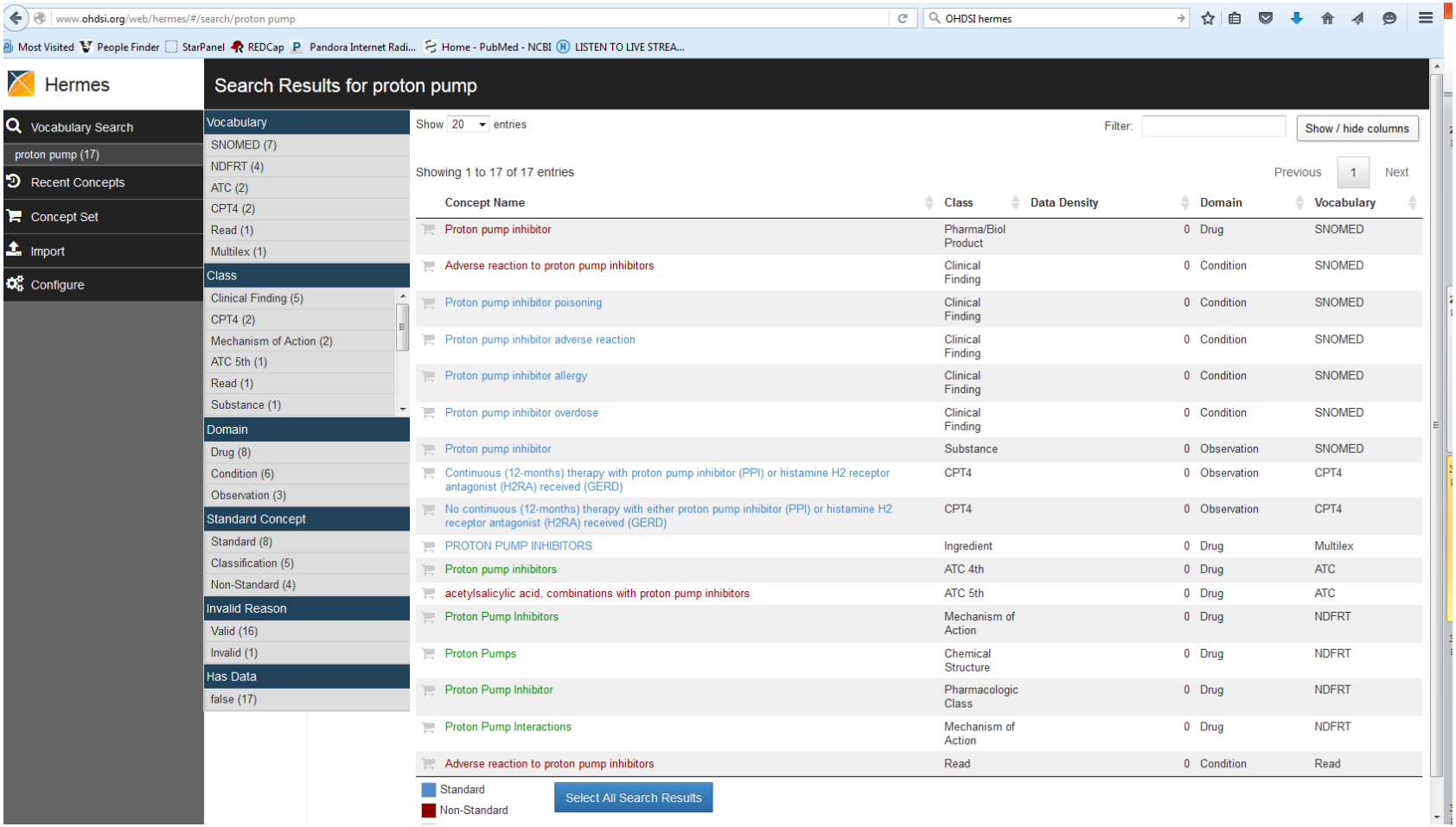
- September
  - COIN analysts training
  - Production open to QA groups
  - Load testing
  - Solutions and table reloads for July QA findings
  - OHDSI Tool testing
  - ETL Process
  - Architecture and performance improvement implementation
- October

Tools to help with data

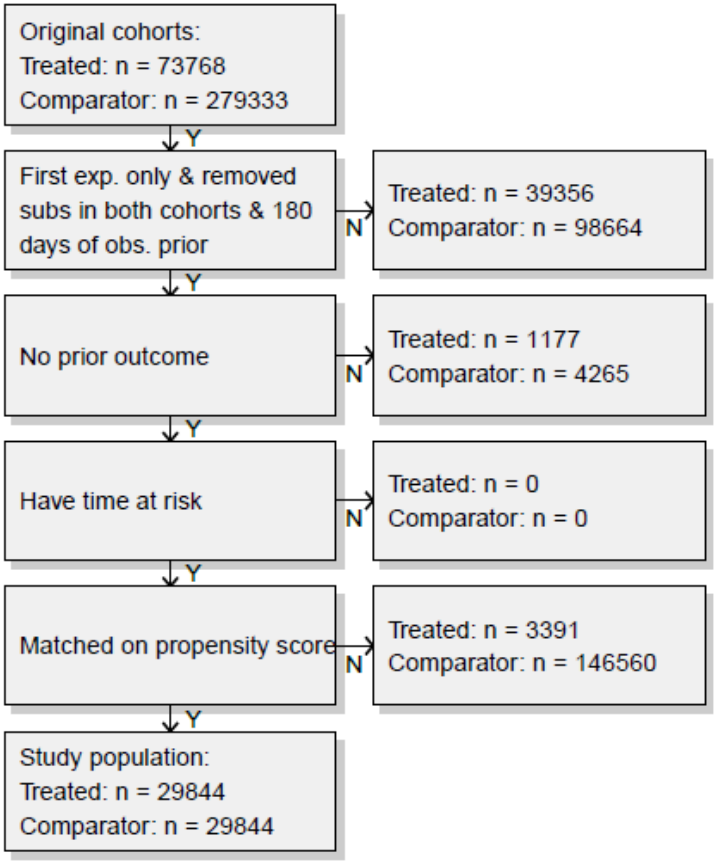
Visualizing data with Achilles



Finding concepts with HERMES

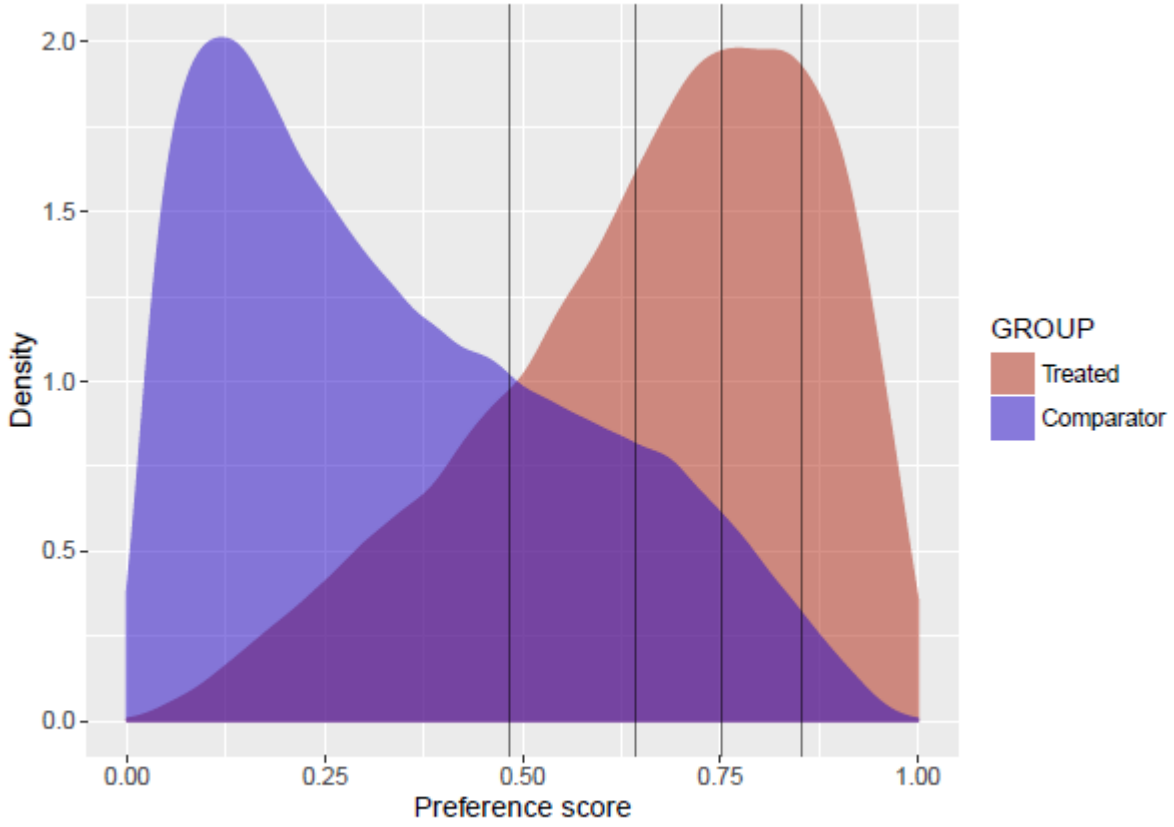


Cohort Method



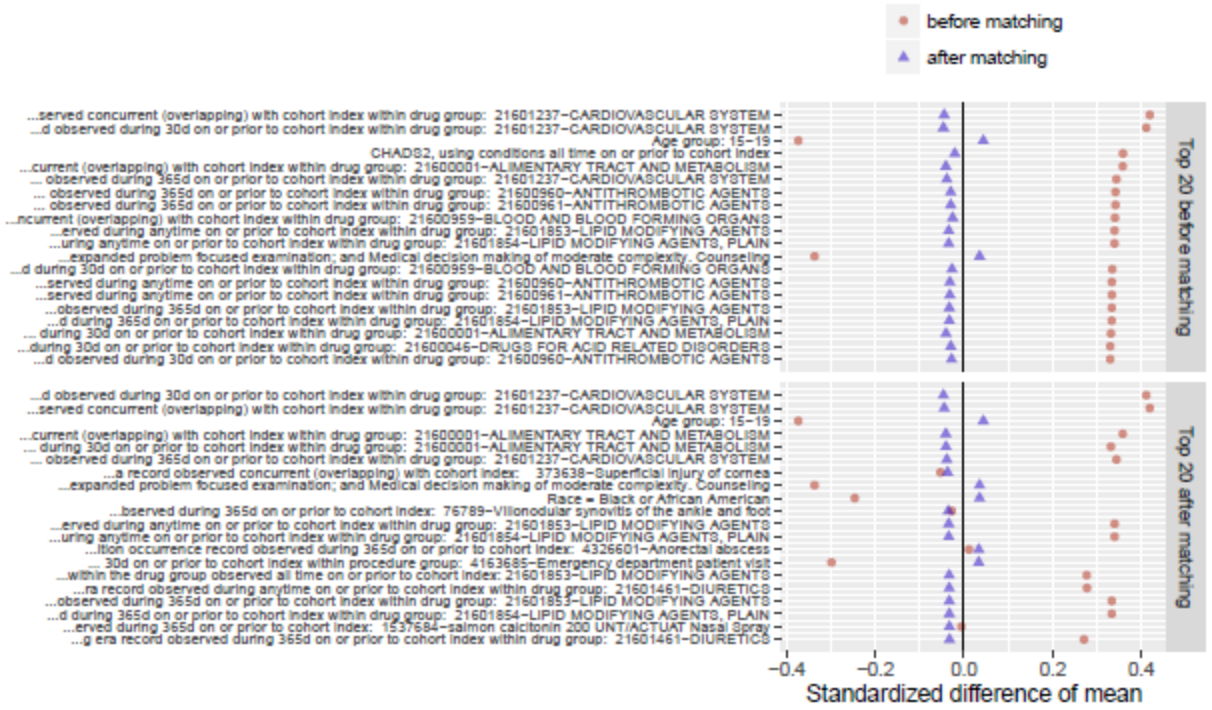
from OHDSI cohort method github

Propensity score matching



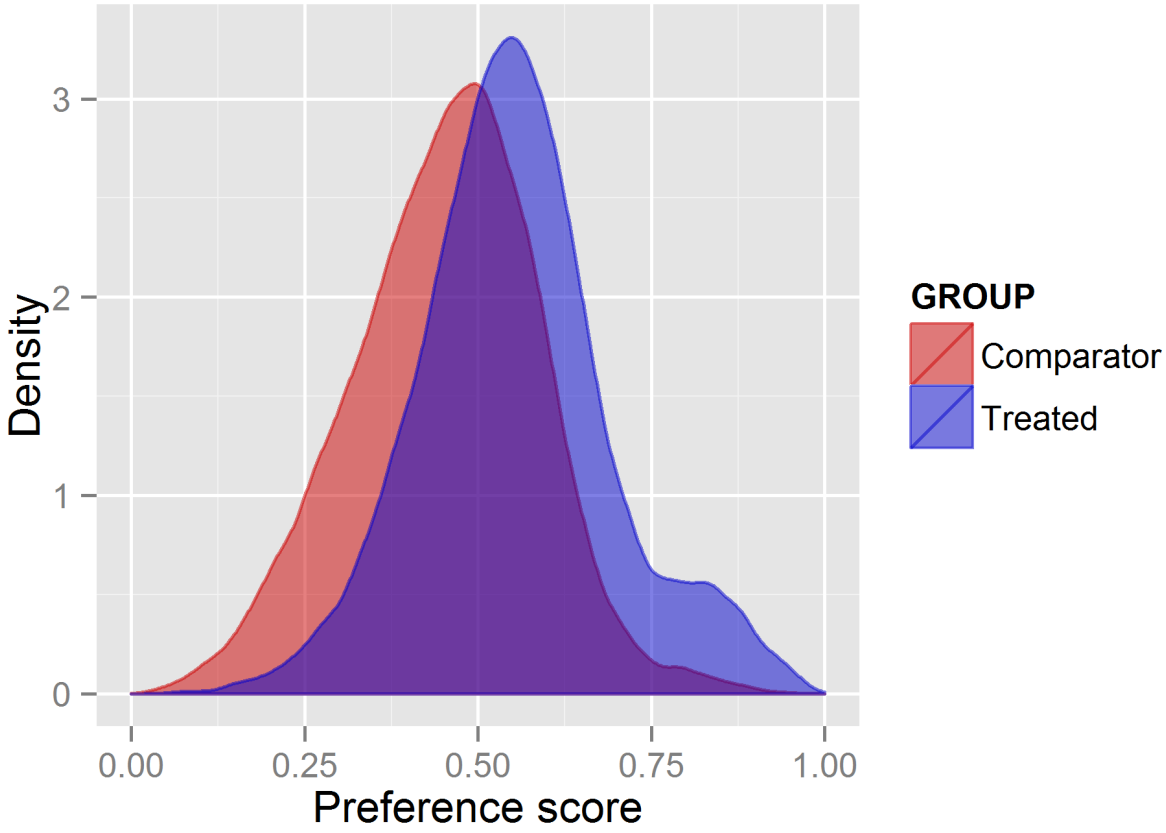
We can also match subjects based on propensity scores. In this example, we're using one-to-one matching:

Propensity score matching



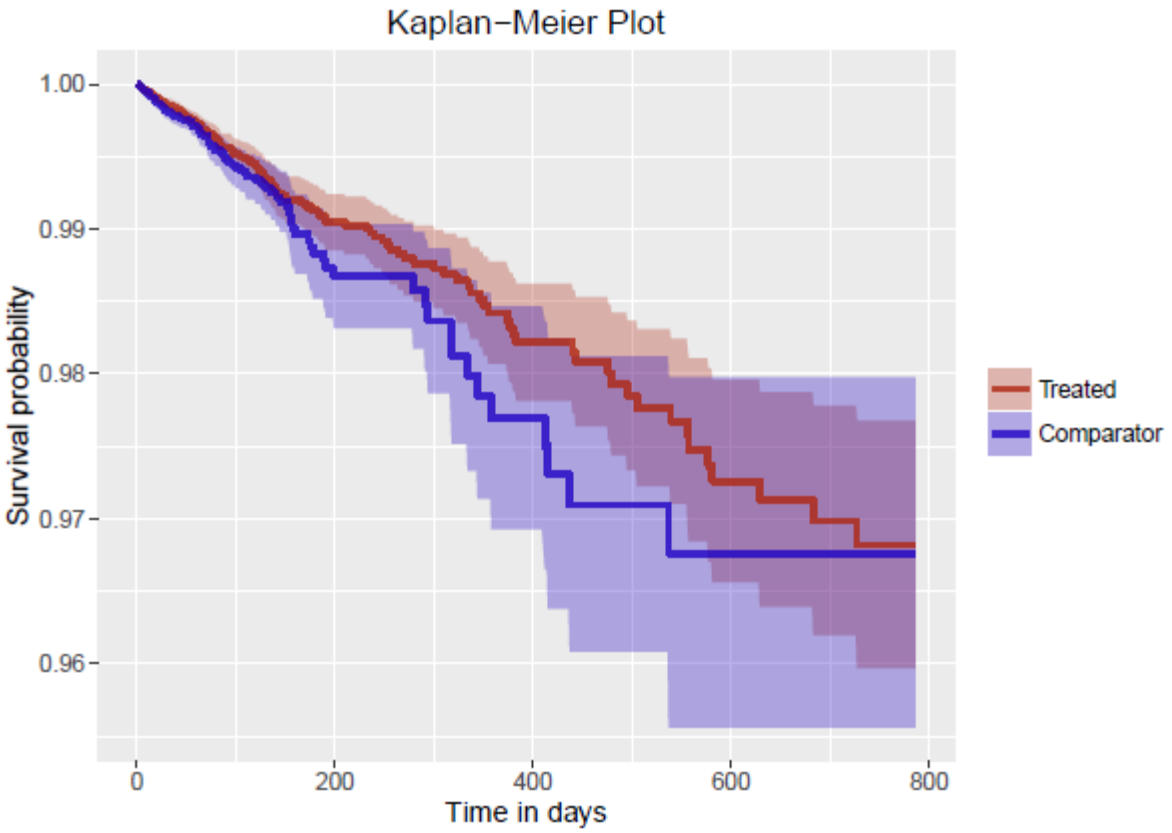
The 'before matching'population is the population as extracted by the `getDbCohortMethodData` function, so

Cohort Method

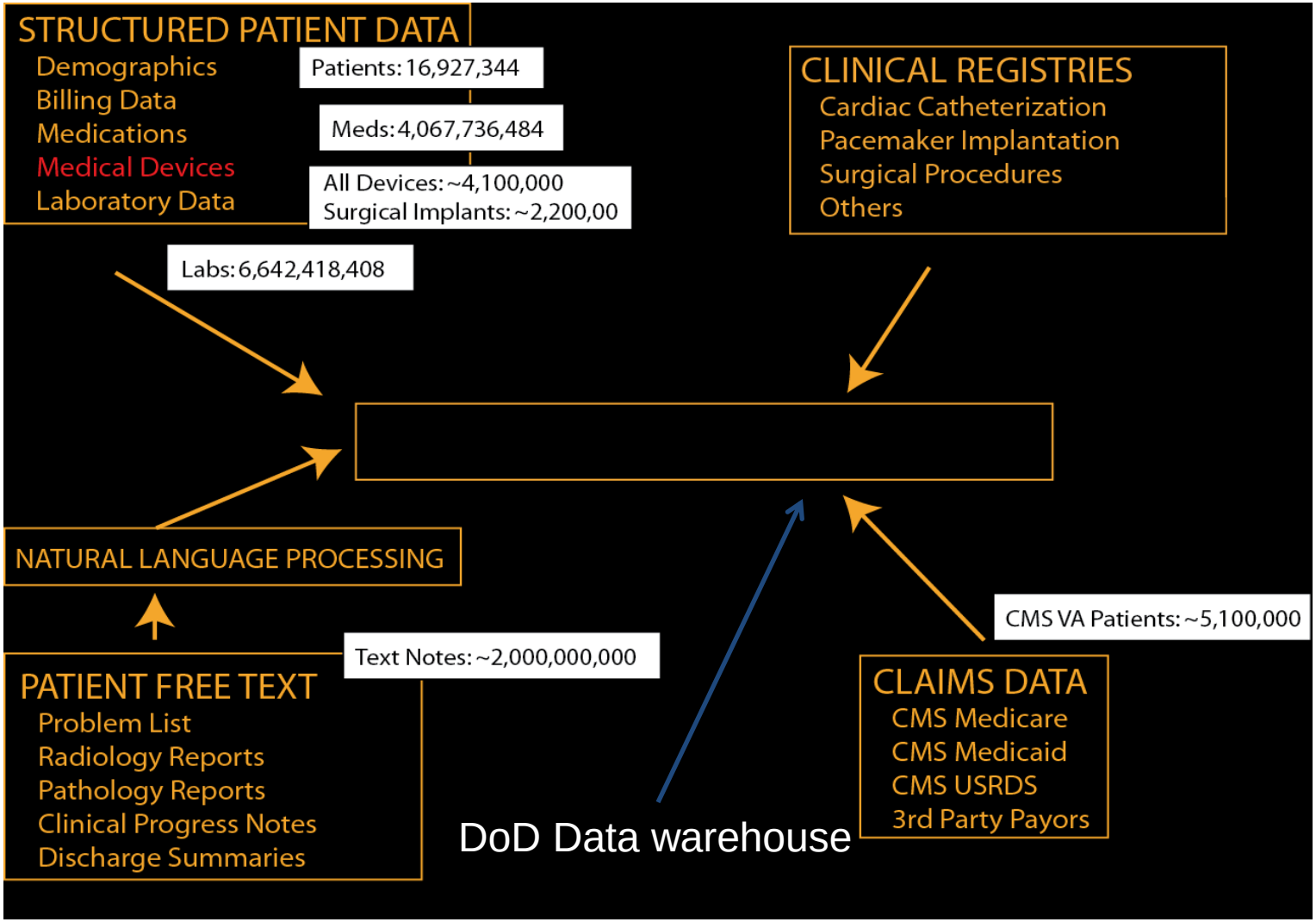


Propensity (preference score) distribution

Cohort Method



Future Plans



Future directions

- NPL defined conditions, occurrences and observations
  - Ejection Fraction – Scott Duvall
  - Acute kidney injury risk factors – Michael Matheny
  - Spinal cord injury risk factors – Steve Luther
- CMS and medicaid – VIREC, Denise Hynes
- DOD – DaVINCI – Bonnema and Duvall
- Cancer Registry -

Future directions

- Other CDW data – as requested
  - Microbiology - Makoto Jones
  - Clinical Assessment, Reporting, and Tracking (CART) for cardiac cath – Tom Maddox
- ODHSI tools – Achilles and Hermes 1st and Cohort Method, eventually Atlas (still in development at OHDSI)

Getting access to VINCI OMOP

- Researcher Access - Include VINCI OMOP as part of DART. OMOP tables and OMOP metadata tables will be added like other table views. Link and crosswalk using patient id and source table IDs.
  - A checkbox in the DART process
- Operational Access – As part of NDS request, include VINCI OMOP. Current CDW access grants access to OMOP tables as well.

OMOP Support

- VA Pulse VINCI OMOP group
  - Message list
  - Documentation
  - FAQ
  - Example code
  - Webnar videos
- VINCI help desk– include OMOP in your ticket descriptor.
  - VINCI@va.gov OR at VINCI Central website request help online

Questions

Acknowledgements

VINCI-TVHS

Development
Jesse Brannen
Fern Fitzhenry
Michael Matheny
QA-Help Desk
Stephen Deppen
Liz Hanchrow
Daniel Park
Jason Denton

VINCI-Salt Lake

Scott DuVall
Kevin Malohi
Brain Sauer
Jeff Scehnet
The many analysts
and content experts

Impetus Research with Common Data Model

- FDA Act of 2007-establish active surveillance systems using electronic health records data
- Medicare HER Incentive Programs
  - Clinicians must demonstrate meanin
    get payment adjustment.
  - Hospitals must do the same

SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.

(a) IN GENERAL.—Subsection (k) of section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

"(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—

"(A) DEFINITION.—In this paragraph, the term 'data' refers to information with respect to a drug approved under this section or under section 351 of the Public Health Service Act, including claims data, patient survey data, standardized analytic files that allow for the pooling and analysis of data from disparate data environments, and any other data deemed appropriate by the Secretary.

"(B) DEVELOPMENT OF POSTMARKET RISK IDENTIFICATION AND ANALYSIS METHODS.—The Secretary shall, not later than 2 years after the date of the enactment of the Food and Drug Administration Amendments Act of 2007, in collaboration with public, academic, and private entities—

"(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);

"(ii) develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

"(I) at least 25,000,000 patients by July 1, 2010; and

"(II) at least 100,000,000 patients by July 1, 2012; and

"(iii) convene a committee of experts, including individuals who are recognized in the field of protecting data privacy and security, to make recommendations to the Secretary on the development of tools and methods for the ethical and scientific uses for, and communication of, postmarketing data specified under subparagraph (C), including recommendations on the development of effective research methods for the study of drug safety questions.

"(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTIFICATION AND ANALYSIS SYSTEM.—

Why did we choose OMOP?

- Met the broadest needs for comparative effectiveness (1)
- Robust open source development teams for user tools to facilitate cohort creation and data visualization
- Applied across multiple databases with the least loss of data fidelity (2)

(1) Oganyemi et al, Med Care 2013; (2)Voss et al, JAMIA 2015

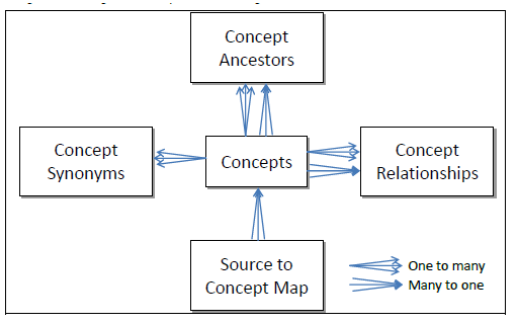
OMOP Vocabulary Tables

- Concept – Houses each individual concept from each standard vocabulary
- Concept Synonym – Alternate names for each concept
- Concept Relationship – Stores the relationship between each concept (i.e parent-child, is-a)
- Concept Ancestor – Stores ancestor-descendant relationships

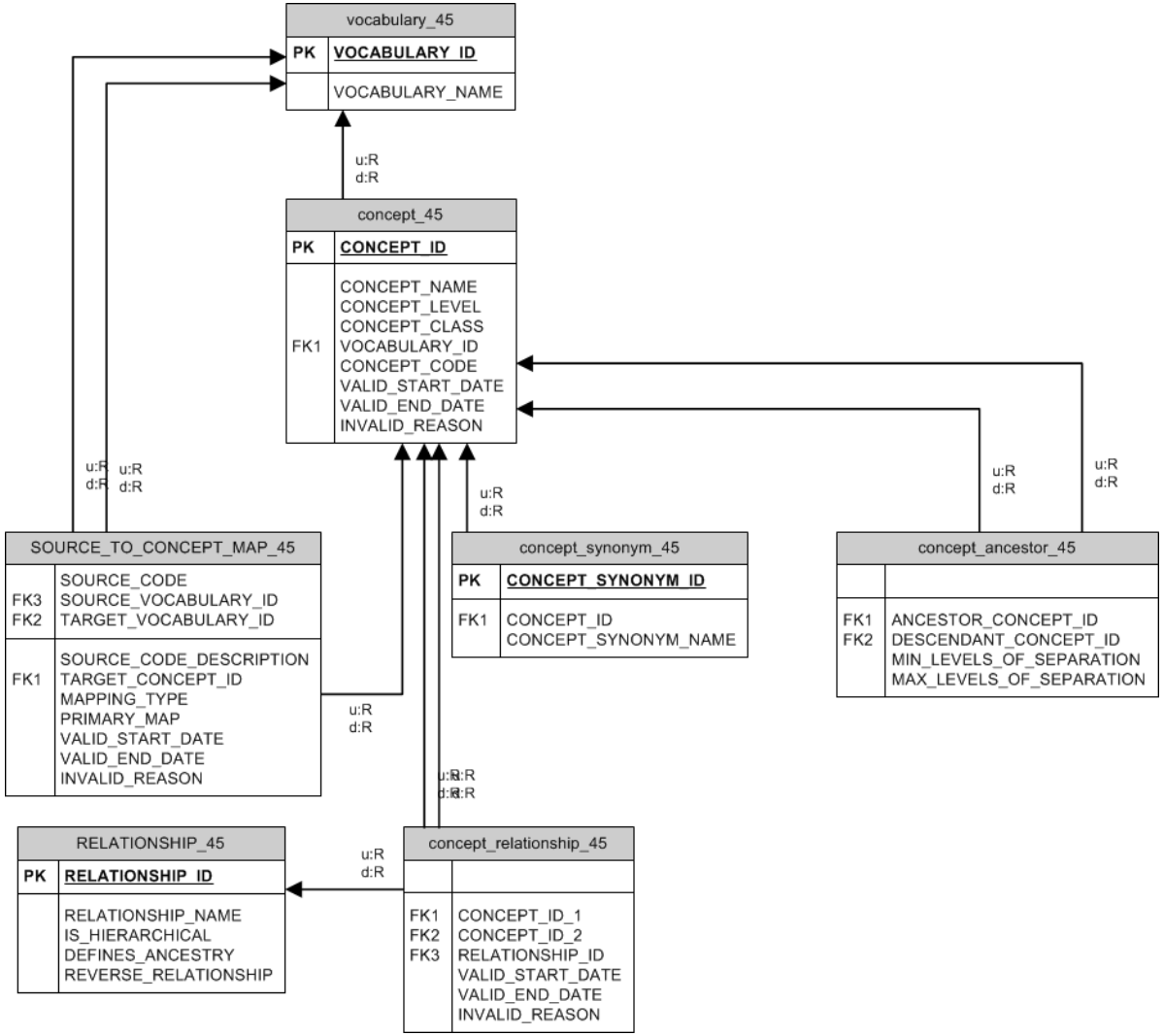
OMOP Vocabulary Tables cont.

- Vocabulary – A dim table of standard vocabularies in the Concept table
- Source-to-concept map – used to map one concept to an equivalent in another vocabulary
- Relationship – A dim table of relationships used in Concept Relationship table.

Concepts


Concept Ancestors
Concept Synonyms
Concepts
Concept Relationships
Source to Concept Map
One to many
Many to one

Vocabularies, concepts and hierarchies



Structure of Vocabulary Domains

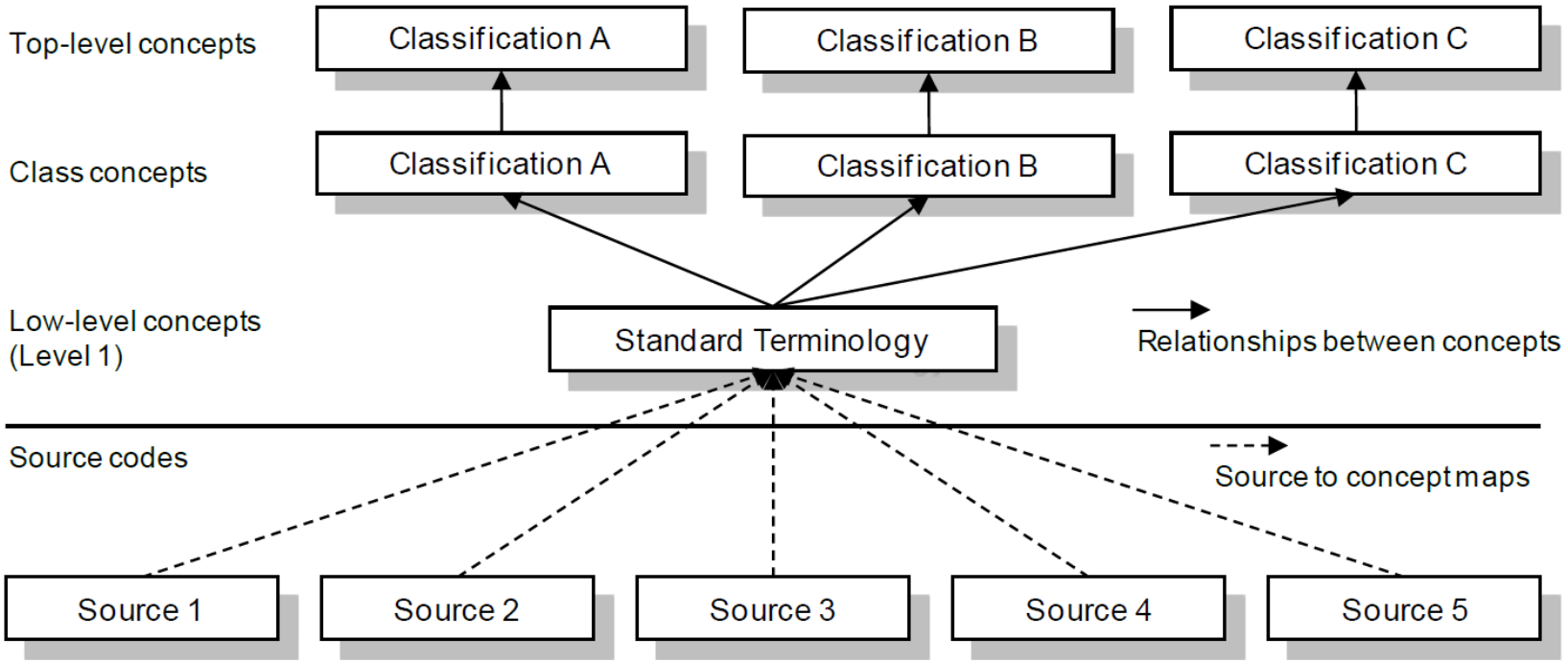


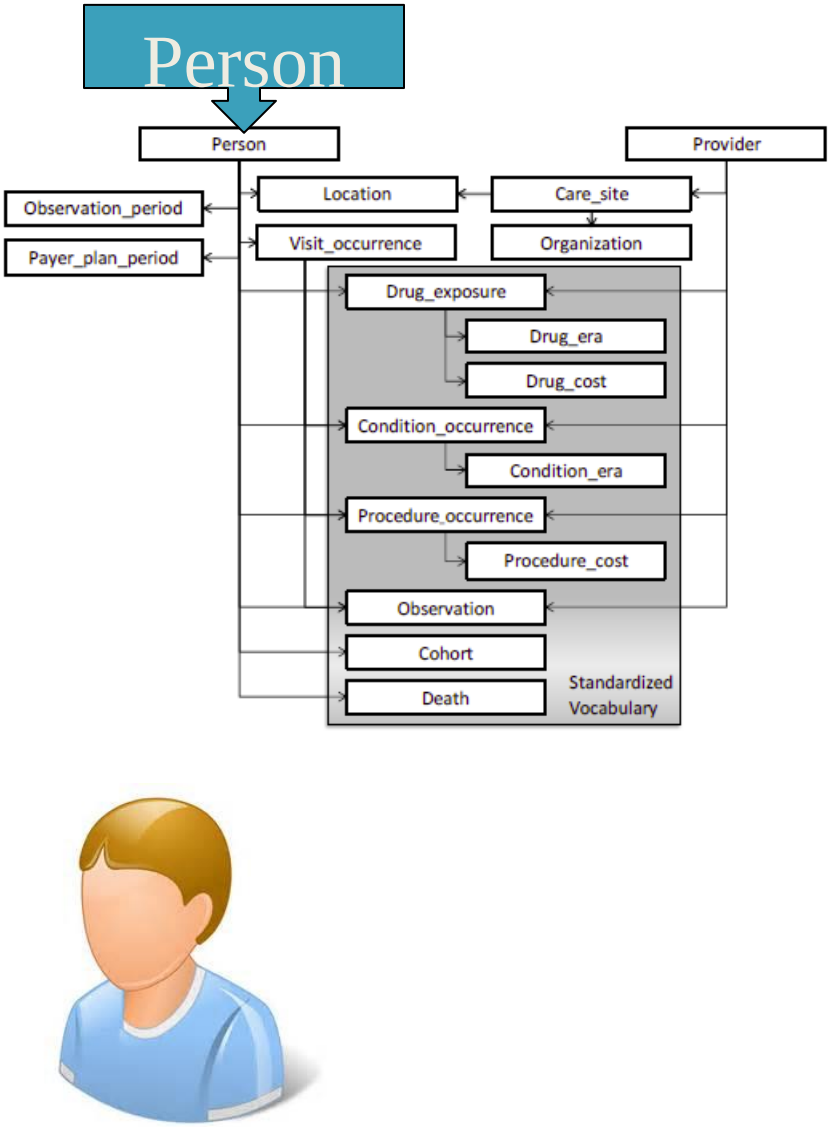
Figure 1: General Organization of Vocabulary Domains

OMOP.org -standard vocabulary specification

OMOP Data Tables Not Implemented

- Drug Cost
- Procedure Cost
- Cohort

Person



- Person table comes from person demographics data
- Data attributes are birthdate race, gender, and ethnicity
- References to other tables about the home address (location), primary care provider…

Person

- PatientICN is the PERSON_ID
- Patients are filtered to those ICNs that are veterans, not test patients, without multiple SSNs
- DOB determined by most frequent occurring not NULL, in a tie, DOB on first visit
- Gender determined by most frequent occurring not NULL, in a tie, most recent
- Race and ethnicity determined by most frequent , in a tie, first visit

PERSON	
PK	**PERSON_ID**
	GENDER_CONCEPT_ID YEAR_OF_BIRTH MONTH_OF_BIRTH DAY_OF_BIRTH RACE_CONCEPT_ID ETHNICITY_CONCEPT_ID LOCATION_ID PROVIDER_ID CARE_SITE_ID PERSON_SOURCE_VALUE GENDER_SOURCE_VALUE RACE_SOURCE_VALUE ETHNICITY_SOURCE_VALUE

Condition Era

- Because not all conditions have a true end date (i.e. Diabetes, HIV, Schizophrenia), the era end date is derived from the Condition Occurrence end date, and thus the discharge date in the record.
- Because there are duplicate diagnoses in the source data, a condition occurrence count field can have a high number despite a short era