Evidence-Based Care

New Care: Evidence-based Care

- In today's health system, it is widely believed that the best care for individuals is based on the training and experience of professionals. The new rule, on the other hand, could be stated: The best care results from the conscientious, explicit, and judicious use of current best evidence and knowledge of patient values by well-trained, experienced clinicians.
- Four steps: (1) formulation of a clear clinical question, (2) search for the relevant information from the best possible sources, (3) evaluation of the evidence for its validity and usefulness, and (4) implementation of those findings

Institute of Medicine (2001). Crossing the quality chasm: A new health system for the 21st century. Washington, DC: National Academies Press.

Improvement to reduce medical errors

- Access to the medical knowledge-base. Through use of the Web, it should be possible to help both providers and consumers gain better access to clinical evidence.
- Computer-aided decision support systems. Embedding knowledge in tools and training clinicians to use those tools to augment their own skills and experience can facilitate the consistent application of the expanding science base to patient care.
- Collection and sharing of clinical information. The automation of patient-specific clinical information is essential for many types of computer-aided decision support systems. Automation of clinical data offers the potential to improve coordination of care across clinicians and settings, which is critical to the effective management of chronic conditions.
- Reduction in errors. Information technology can contribute to a reduction in errors by standardizing and automating certain decisions and by aiding in the identification of possible errors, such as potential adverse drug interactions, before they occur.
- Enhanced patient and clinician communication.

Institute of Medicine (2001). Crossing the quality chasm: A new health system for the 21st century. Washington, DC: National Academies Press.

OHDSI/OMOP

OHDSI

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Mission: Improving health by empowering a community to collaboratively generate evidence that promotes better health decisions and better care

Vision: Creating a world in which observational research produces a comprehensive understanding of health and disease

Values: Innovation, Reproducibility, Community, Collaboration, Openness, Beneficence

Web: OHDSI.org

Discussion: forums.OHDSI.org

Code: github.com/OHDSI

Research Results: data.OHDSI.org

Twitter: @OHDSI

LinkedIn: LinkedIn.com/company/OHDSI

Contact Us: contact@ohdsi.org

OHDSI

- Observational Health Data Sciences and Informatics (OHDSI) is an international collaborative whose goal is to create and apply open-source data analytic solutions to a large network of health databases to improve human health and wellbeing.
- The OHDSI team comprises academics, industry scientists, health care providers, and regulators whose formal mission is to transform medical decision making by creating reliable scientific evidence about disease natural history, healthcare delivery, and the effects of medical interventions through large-scale analysis of observational health databases for population-level estimation and patient-level predictions.
- Over 90 participants from around the world have joined the collaborative with a vision to access a network of one billion patients to generate evidence about all aspects of healthcare, where patients, clinicians and all other decision-makers around the world use OHDSI tools and evidence every day.

OHDSI

Delivering Reliable Evidence

OHDSI strives to develop reliable real-world evidence through methodological research, open-source analytics development, and clinical evidence generation.

The OHDSI community has produced hundreds of peer-reviewed papers since its 2014 inception. The creation of the ATLAS open-source platform can be installed locally with any database that has been standardized to the OMOP CDM. Thus, research can be done in collaboration, and with a significant set of data.

https://www.ohdsi.org/

OHDSI: Evidence-based Care

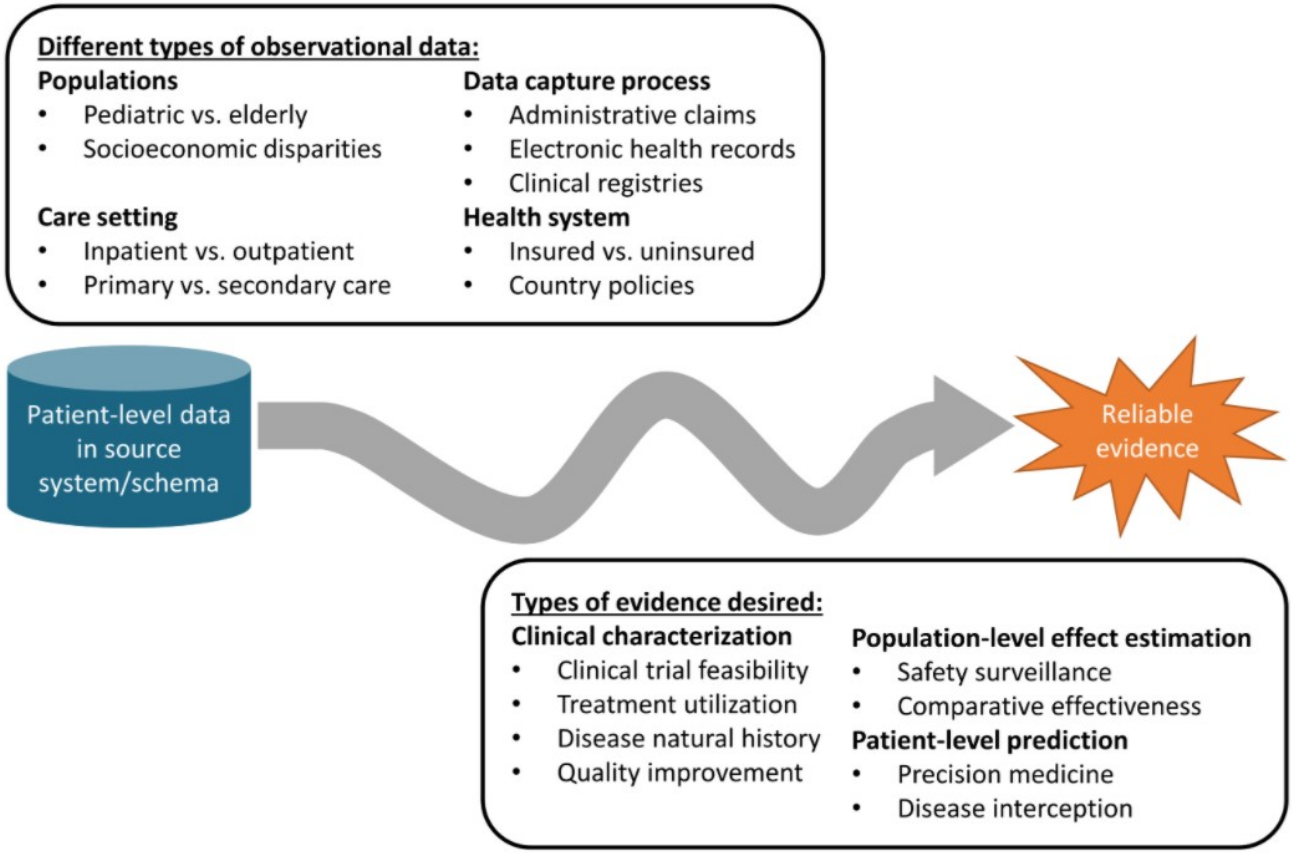


Figure 1.1: The journey from data to evidence

Mission:
To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

The OHDSI developer community has created a robust library of open-source analytics tools atop the OMOP CDM to support 3 use cases:

1) clinical characterization for disease natural history, treatment utilization, and quality improvement;
2) population-level effect estimation to apply causal inference methods for medical product safety surveillance and comparative effectiveness; and
3) patient-level prediction to apply machine learning algorithms for precision medicine and disease interception

Guiding principle

- Evidence-based, such that OHDSI's scientific research and development are driven by objective, empirical evidence to ensure accuracy and reliability;
- Practical, going beyond methodological research, but developing applied solutions and generating clinical evidence;
- Comprehensive, aiming to generate reliable scientific evidence for all interventions and all outcomes;
- Transparent, such that all work products within OHDSI are Open Source and publicly available, including source code, analysis results, and other evidence generated in all our activities;
- Inclusive, encouraging active participation from all stakeholders – patients, providers, payers, government, industry, academia – in all phases of research and development; and finally
- Secure, protecting patient privacy and respecting data

Open

- OHDSI is an open collaborative. Anyone who can give time, data, or funding is welcome, and participation in the operation of OHDSI is expected. Currently, participants come from around the world, including the United States, United Kingdom, Netherlands, Sweden, Italy, Korea, Taiwan, Hong Kong, and Australia.

OMOP Common Data Model (CDM)

v6..0

https://ohdsi.github.io/CommonDataModel/

OMOP

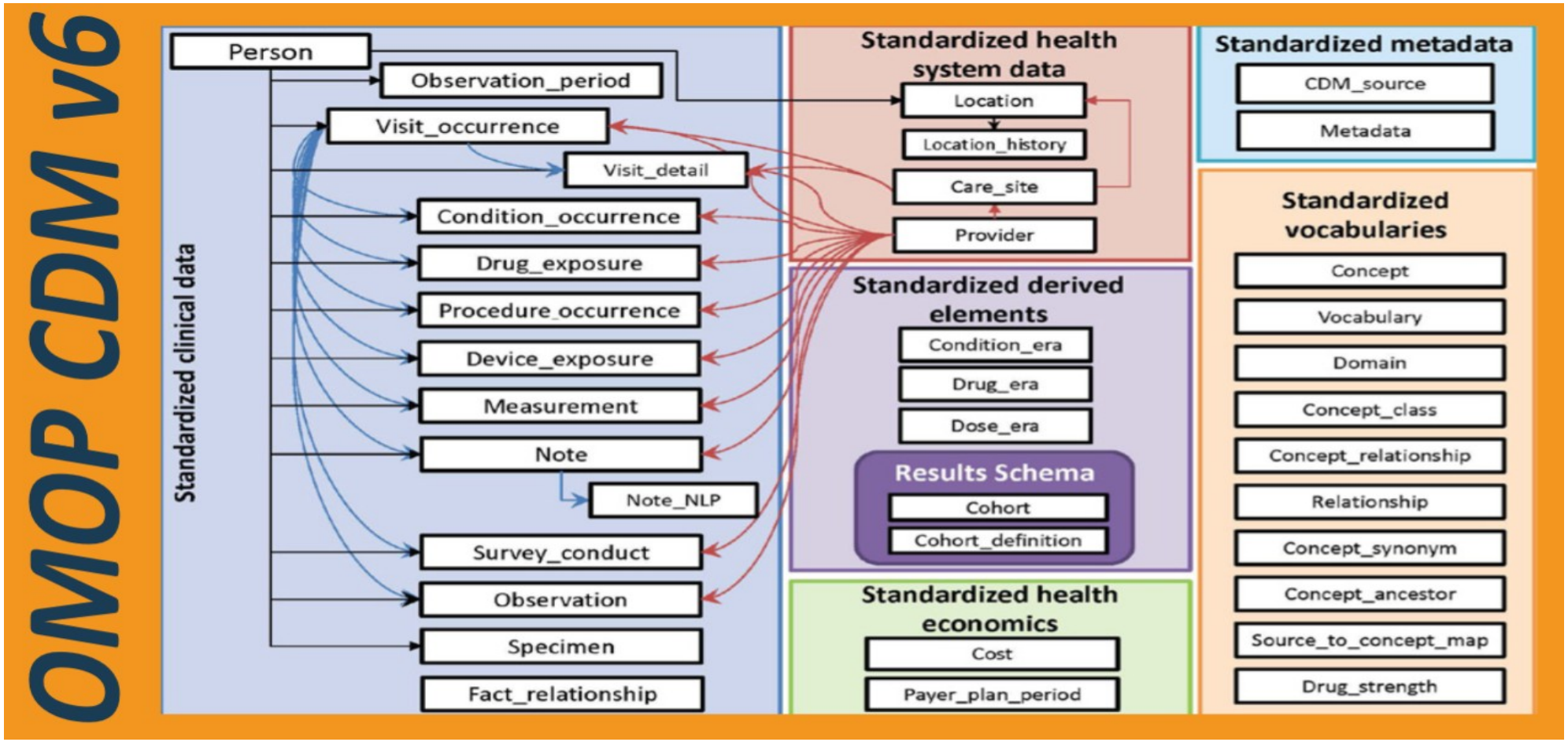
- Observational Medical Outcomes Partnership (OMOP)
- a public-private partnership established in the US to inform the appropriate use of observational healthcare databases for studying the effects of medical products.
- The five-year project developed new methods in observational research and established an observational research laboratory. At the conclusion of this five-year project, the OMOP research investigators initiated the OHDSI effort, and the research laboratory moved to the Reagan-Udall Foundation for the FDA under the Innovation in Medical Evidence Development and Surveillance (IMEDS) Program

OMOP CDM

- A centerpiece of the OMOP project was the development of the OMOP Common Data Model (CDM) which represents healthcare data from diverse sources in a consistent and standardized way.
- This CDM is a “strong” information model, in which the encoding and relationships among concepts are explicitly and formally specified.
- The OHDSI team has adopted and continued maintenance of this model and its associated vocabulary services. OHDSI’s overall approach is to create an open network of observational data holders, and require that they translate their data to the OMOP CDM.
- Each element in the participant’s database must be mapped to the approved CDM vocabulary and placed in the data schema. In return, this approach creates a unique opportunity of implementing a number of existing data exploration and evidence generation tools and participating in world-wide studies because any given query can be executed at any site without modification.
- This enables multicenter, global analyses to be executed rapidly and efficiently using applications or programs developed at a single site.

OHDSI Data Model (OMOP)

- Converting different health data into a common data model



OMOP

- The CDM is designed to support the conduct of research:
  - to identify and evaluate associations between interventions (drug exposure, procedures, healthcare policy changes etc.) and outcomes caused by these interventions (condition occurrences, procedures, drug exposure etc.).
  - Outcomes can be efficacious (benefit) or adverse (safety risk). Often times, specific patient cohorts (e.g., those taking a certain drug or suffering from a certain disease) may be defined for treatments or outcomes, using clinical events (diagnoses, observations, procedures, etc.) that occur in predefined temporal relationships to each other.
  - The CDM, combined with its standardized content (via the Standardized Vocabularies), will ensure that research methods can be systematically applied to produce meaningfully comparable and reproducible results.

Standardized Vocabularies

4 Standardized Vocabularies

CONCEPT
VOCABULARY
DOMAIN
CONCEPT_CLASS
CONCEPT_RELATIONSHIP
RELATIONSHIP
CONCEPT_SYNONYM
CONCEPT_ANCESTOR
SOURCE_TO_CONCEPT_MAP
DRUG_STRENGTH

Concepts

- Concepts are derived from vocabularies, which represent clinical information across a domain (e.g. conditions, drugs, procedures) through the use of codes and associated descriptions. Some Concepts are designated Standard Concepts, meaning these Concepts can be used as normative expressions of a clinical entity within the OMOP Common Data Model and within standardized analytics.
- Concepts can represent broad categories (like ‘Cardiovascular disease’), detailed clinical elements (‘Myocardial infarction of the anterolateral wall’) or modifying characteristics and attributes that define Concepts at various levels of detail (severity of a disease, associated morphology, etc.).
- Standardized Vocabularies tables are derived from national or international vocabularies such as SNOMED-CT, RxNorm, and LOINC, or custom Concepts defined to cover various aspects of observational data analysis.

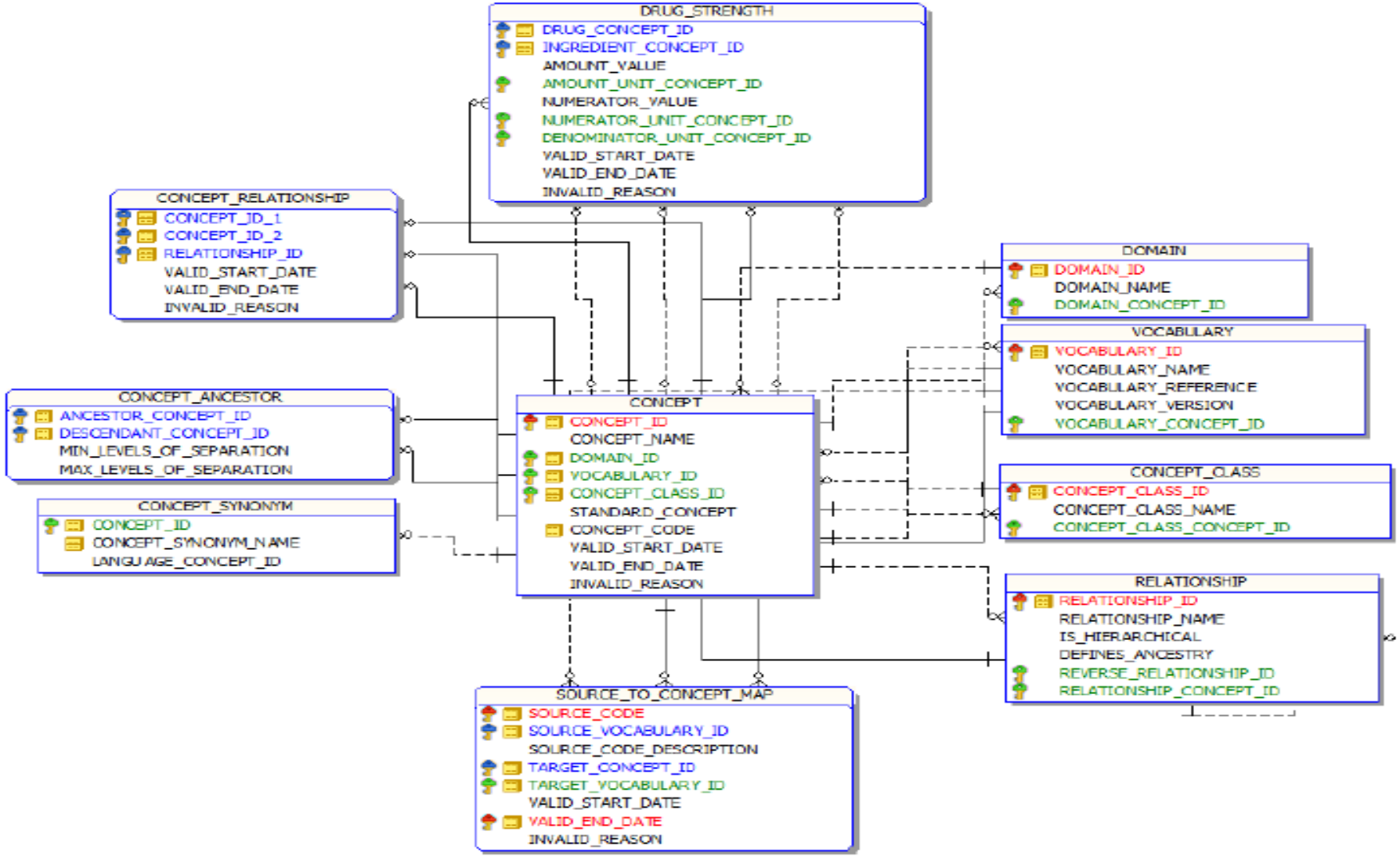


Figure 1: Vocabulary entity-relationship diagram

Athena search (**Pulmonary tuberculosis**)

https://athena.ohdsi.org/search-terms/terms/253954

DETAILS

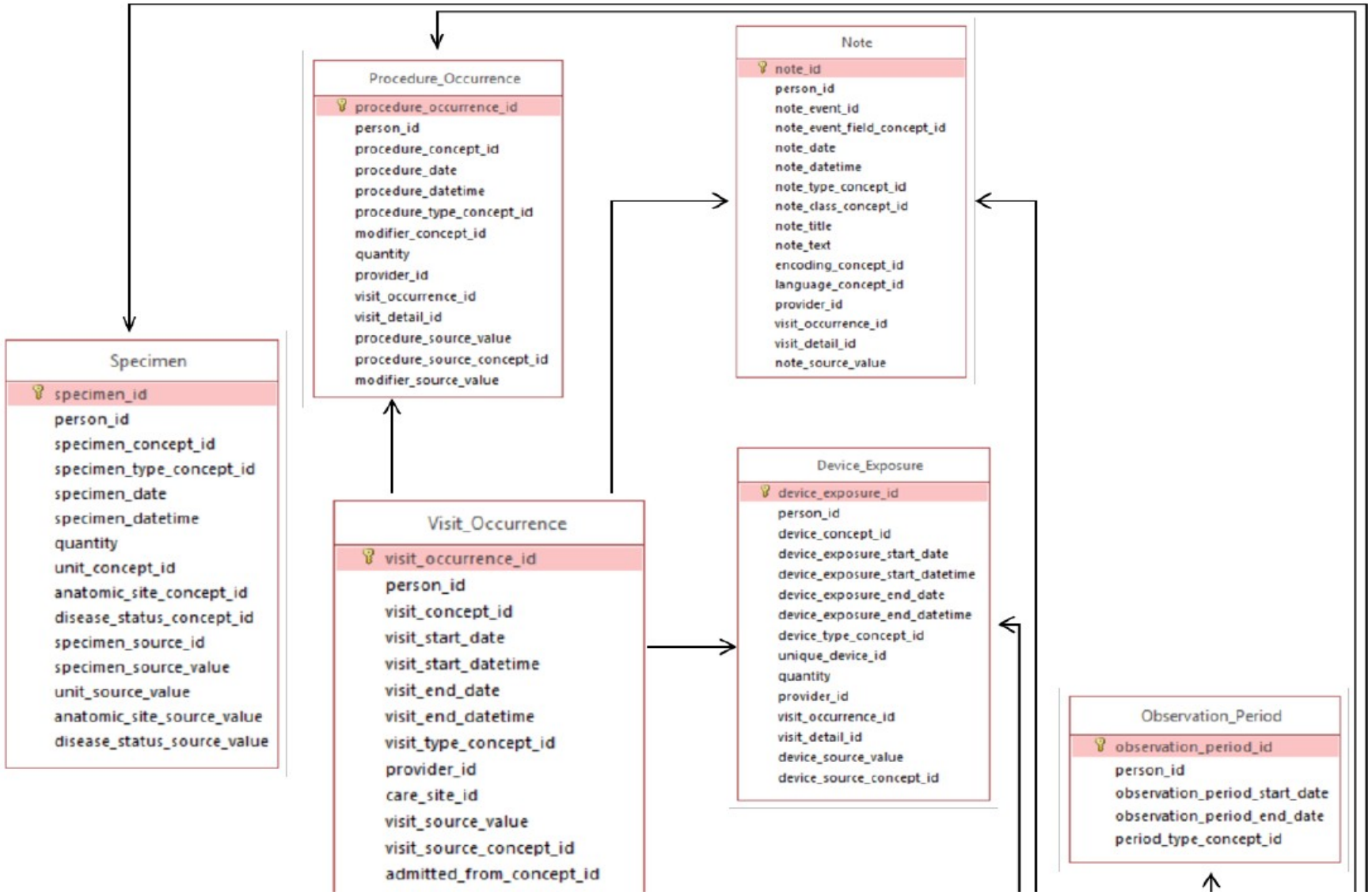
Domain ID	Condition
Concept Class ID	Clinical Finding
Vocabulary ID	SNOMED
Concept ID	253954
Concept code	154283005
Validity	Valid
Concept	Standard
Synonyms	Pulmonary tuberculosis (disorder) Pulmonary tuberculosis PTB - Pulmonary tuberculosis TB - Pulmonary tuberculosis
Valid start	01-Jan-1970
Valid end	31-Dec-2099

TERM CONNECTIONS (82)

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
ICD-9-CM to MedDRA (MSSO)	Pulmonary tuberculosis	36110777	MedDRA
Non-standard to Standard map (OMOP)	Pulmonary tuberculosis	253954	SNOMED
Subsumes	Other specified pulmonary tuberculosis	44830894	ICD9CM
	Other specified pulmonary tuberculosis, bacteriological or histological examination not done	44836741	ICD9CM
	Other specified pulmonary tuberculosis, bacteriological or histological examination unknown (at present)	44836742	ICD9CM
	Other specified pulmonary tuberculosis, tubercle bacilli found (in sputum) by microscopy	44821641	ICD9CM
	Other specified pulmonary tuberculosis, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture	44833188	ICD9CM

6 Standardized Clinical Data Tables

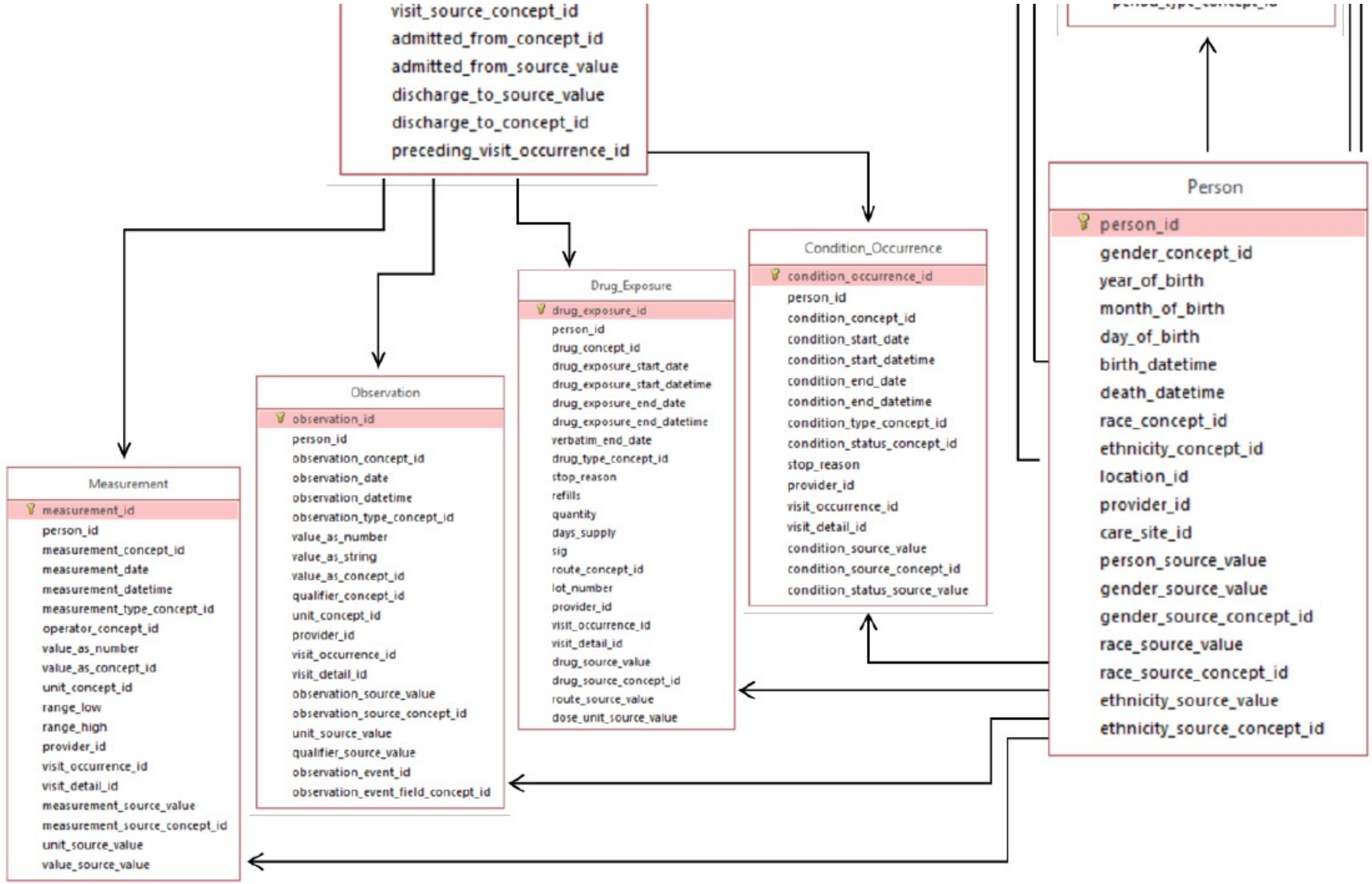
PERSON
OBSERVATION_PERIOD
DEATH
VISIT_OCCURRENCE
VISIT_DETAIL CONDITION_OCCURRENCE DRUG_EXPOSURE
PROCEDURE_OCCURRENCE
DEVICE_EXPOSURE
MEASUREMENT
NOTE
NOTE_NLP SURVEY_CONDUCT
OBSERVATION SPECIMEN
FACT_RELATIONSHIP

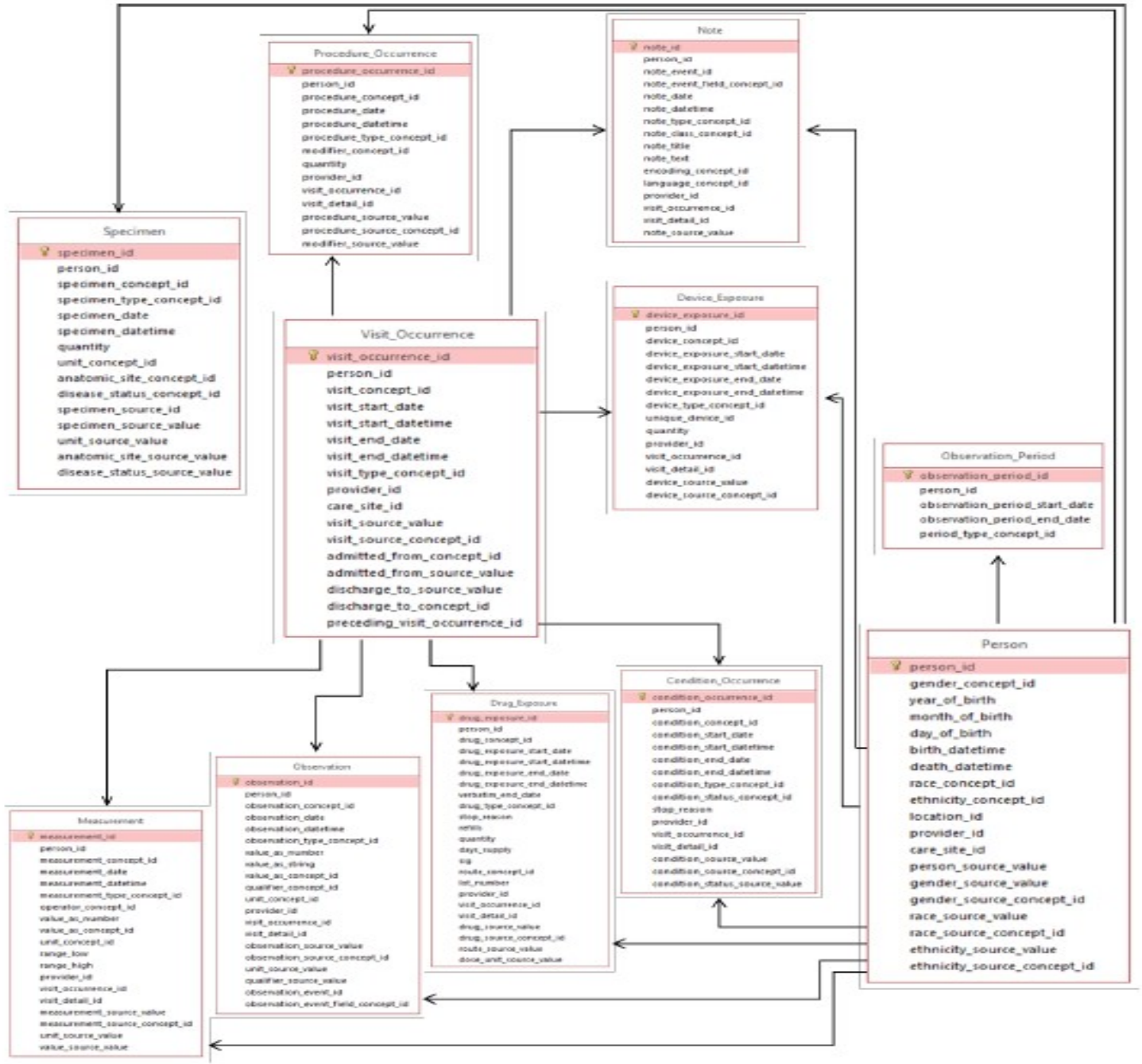


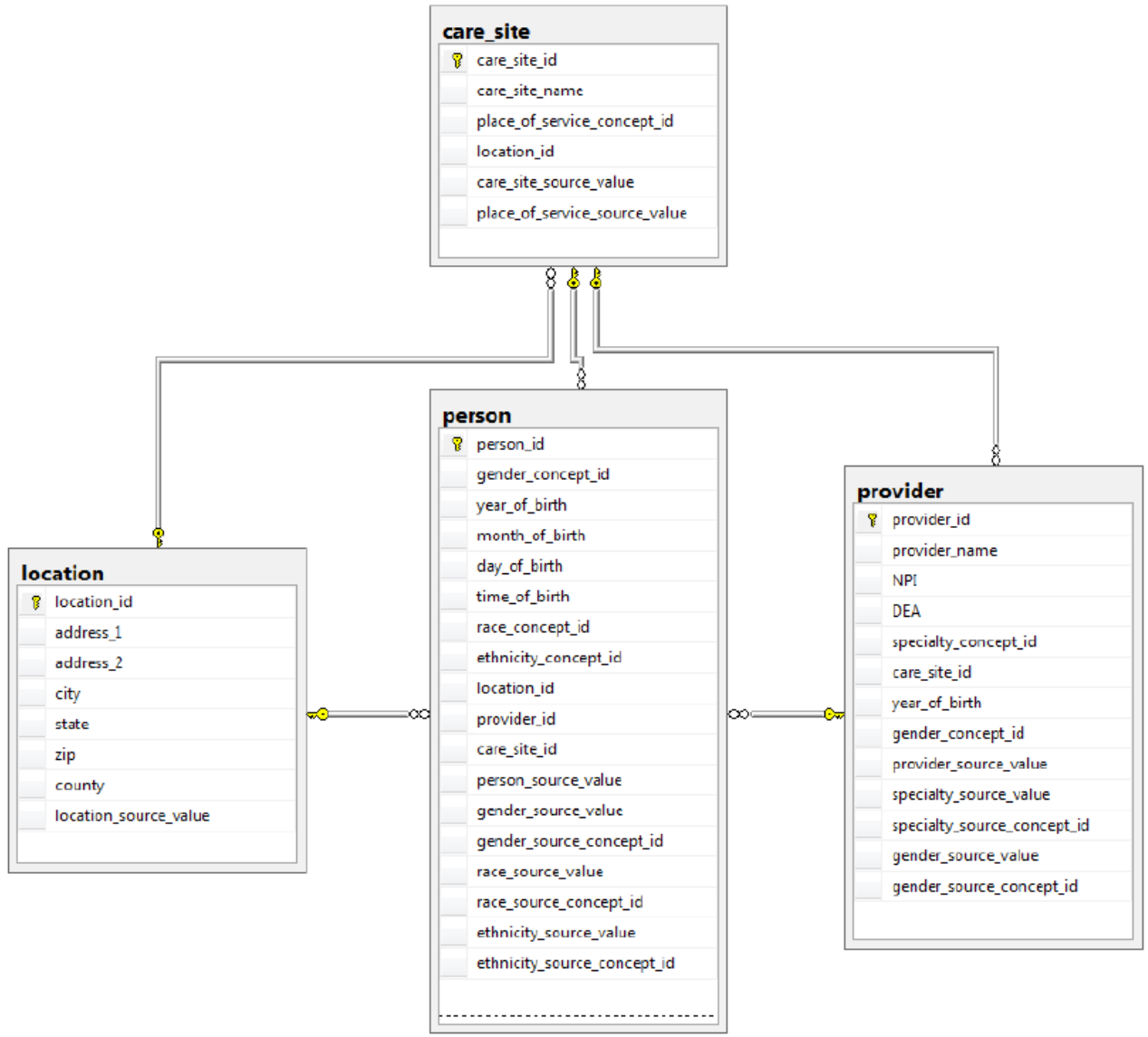
What are the 10 most common medical procedures?

Some of the most common surgical operations done i following:

- Appendectomy. ...
- Breast biopsy. ...
- Carotid endarterectomy. ...
- Cataract surgery. ...
- Cesarean section (also called a c-section). ...
- Cholecystectomy. ...
- Coronary artery bypass. ...
- Debridement of wound, burn, or infection.

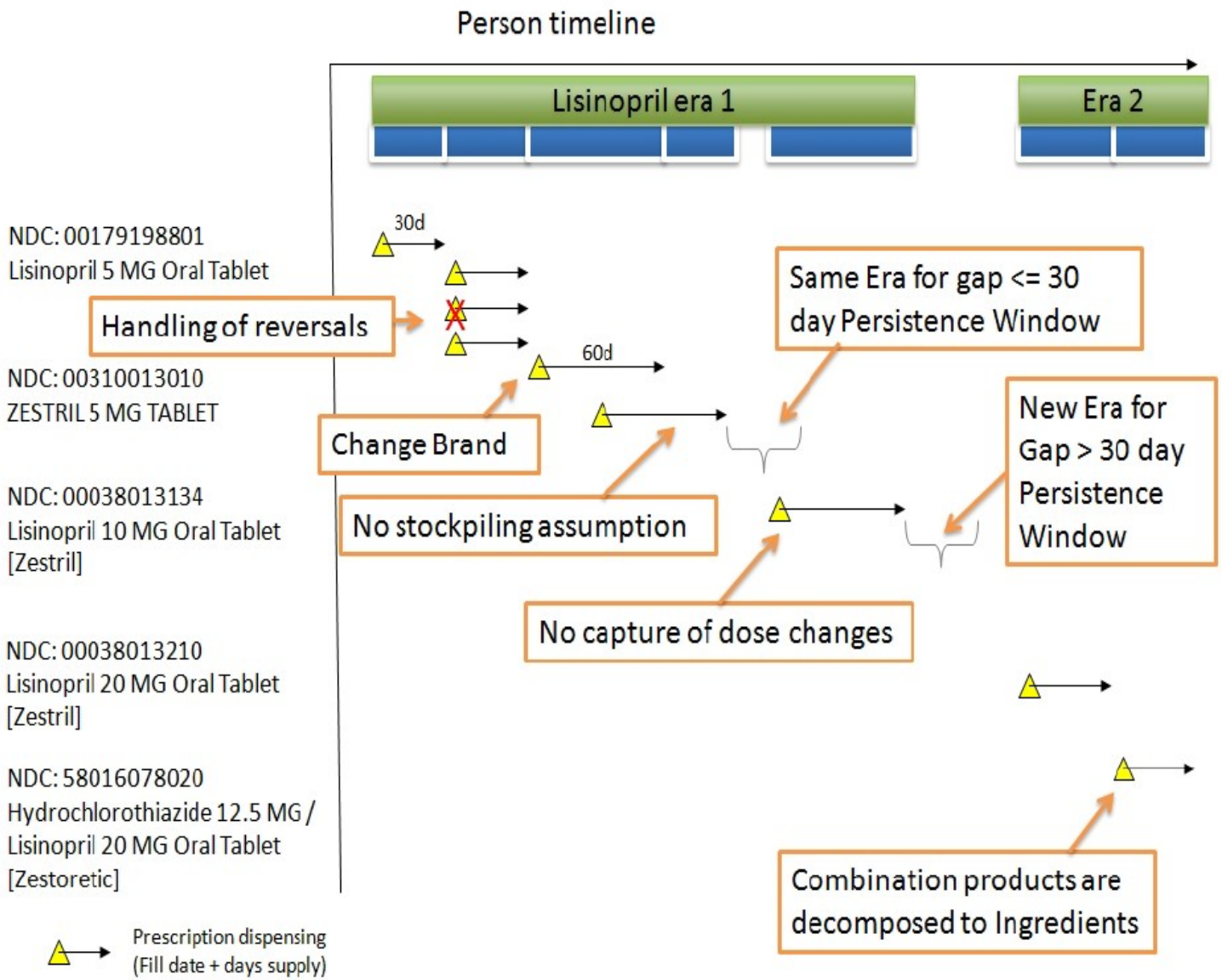

visit_source_concept_id
admitted_from_concept_id
admitted_from_source_value
discharge_to_source_value
discharge_to_concept_id
preceding_visit_occurrence_id
Person
person_id
gender_concept_id
year_of_birth
month_of_birth
day_of_birth
birth_datetime
death_datetime
race_concept_id
ethnicity_concept_id
location_id
provider_id
care_site_id
person_source_value
gender_source_value
gender_source_concept_id
race_source_value
race_source_concept_id
ethnicity_source_value
ethnicity_source_concept_id
Condition_Occurrence
condition_occurrence_id
person_id
condition_concept_id
condition_start_date
condition_start_datetime
condition_end_date
condition_end_datetime
condition_type_concept_id
condition_status_concept_id
stop_reason
provider_id
visit_occurrence_id
visit_detail_id
condition_source_value
condition_source_concept_id
condition_status_source_value
Drug_Exposure
drug_exposure_id
person_id
drug_concept_id
drug_exposure_start_date
drug_exposure_start_datetime
drug_exposure_end_date
drug_exposure_end_datetime
verbatim_end_date
drug_type_concept_id
stop_reason
refills
quantity
days_supply
sig
route_concept_id
lot_number
provider_id
visit_occurrence_id
visit_detail_id
drug_source_value
drug_source_concept_id
route_source_value
dose_unit_source_value
Observation
observation_id
person_id
observation_concept_id
observation_date
observation_datetime
observation_type_concept_id
value_as_number
value_as_string
value_as_concept_id
qualifier_concept_id
unit_concept_id
provider_id
visit_occurrence_id
visit_detail_id
observation_source_value
observation_source_concept_id
unit_source_value
qualifier_source_value
observation_event_id
observation_event_field_concept_id
Measurement
measurement_id
person_id
measurement_concept_id
measurement_date
measurement_datetime
measurement_type_concept_id
operator_concept_id
value_as_number
value_as_concept_id
unit_concept_id
range_low
range_high
provider_id
visit_occurrence_id
visit_detail_id
measurement_source_value
measurement_source_concept_id
unit_source_value
value_source_value


Procedure_Occurrence
procedure_occurrence_id
person_id
procedure_concept_id
procedure_date
procedure_datetime
procedure_type_concept_id
modifier_concept_id
quantity
provider_id
visit_occurrence_id
visit_detail_id
procedure_source_value
procedure_source_concept_id
modifier_source_value
Note
note_id
person_id
note_event_id
note_event_field_concept_id
note_date
note_datetime
note_type_concept_id
note_class_concept_id
note_title
note_text
encoding_concept_id
language_concept_id
provider_id
visit_occurrence_id
visit_detail_id
note_source_value
Specimen
specimen_id
person_id
specimen_concept_id
specimen_type_concept_id
specimen_date
specimen_datetime
quantity
unit_concept_id
anatomic_site_concept_id
disease_status_concept_id
specimen_source_id
specimen_source_value
unit_source_value
anatomic_site_source_value
disease_status_source_value
Visit_Occurrence
visit_occurrence_id
person_id
visit_concept_id
visit_start_date
visit_start_datetime
visit_end_date
visit_end_datetime
visit_type_concept_id
provider_id
care_site_id
visit_source_value
visit_source_concept_id
admitted_from_concept_id
admitted_from_source_value
discharge_to_source_value
discharge_to_concept_id
preceding_visit_occurrence_id
Device_Exposure
device_exposure_id
person_id
device_concept_id
device_exposure_start_date
device_exposure_start_datetime
device_exposure_end_date
device_exposure_end_datetime
device_type_concept_id
unique_device_id
quantity
provider_id
visit_occurrence_id
visit_detail_id
device_source_value
device_source_concept_id
Observation_Period
observation_period_id
person_id
observation_period_start_date
observation_period_end_date
period_type_concept_id
Person
person_id
gender_concept_id
year_of_birth
month_of_birth
day_of_birth
birth_datetime
death_datetime
race_concept_id
ethnicity_concept_id
location_id
provider_id
care_site_id
person_source_value
gender_source_value
gender_source_concept_id
race_source_value
race_source_concept_id
ethnicity_source_value
ethnicity_source_concept_id
Condition_Occurrence
condition_occurrence_id
person_id
condition_concept_id
condition_start_date
condition_start_datetime
condition_end_date
condition_end_datetime
condition_type_concept_id
condition_status_concept_id
stop_reason
provider_id
visit_occurrence_id
visit_detail_id
condition_source_value
condition_source_concept_id
condition_status_source_value
Drug_Exposure
drug_exposure_id
person_id
drug_concept_id
drug_exposure_start_date
drug_exposure_start_datetime
drug_exposure_end_date
drug_exposure_end_datetime
verbatim_end_date
drug_type_concept_id
stop_reason
refills
quantity
days_supply
sig
route_concept_id
lot_number
provider_id
visit_occurrence_id
visit_detail_id
drug_source_value
drug_source_concept_id
route_source_value
dose_unit_source_value
Observation
observation_id
person_id
observation_concept_id
observation_date
observation_datetime
observation_type_concept_id
value_as_number
value_as_string
value_as_concept_id
qualifier_concept_id
unit_concept_id
provider_id
visit_occurrence_id
visit_detail_id
observation_source_value
observation_source_concept_id
unit_source_value
qualifier_source_value
observation_event_id
observation_event_field_concept_id
Measurement
measurement_id
person_id
measurement_concept_id
measurement_date
measurement_datetime
measurement_type_concept_id
operator_concept_id
value_as_number
value_as_concept_id
unit_concept_id
range_low
range_high
provider_id
visit_occurrence_id
visit_detail_id
measurement_source_value
measurement_source_concept_id
unit_source_value
value_source_value


care_site
care_site_id
care_site_name
place_of_service_concept_id
location_id
care_site_source_value
place_of_service_source_value
person
person_id
gender_concept_id
year_of_birth
month_of_birth
day_of_birth
time_of_birth
race_concept_id
ethnicity_concept_id
location_id
provider_id
care_site_id
person_source_value
gender_source_value
gender_source_concept_id
race_source_value
race_source_concept_id
ethnicity_source_value
ethnicity_source_concept_id
location
location_id
address_1
address_2
city
state
zip
county
location_source_value
provider
provider_id
provider_name
NPI
DEA
specialty_concept_id
care_site_id
year_of_birth
gender_concept_id
provider_source_value
specialty_source_value
specialty_source_concept_id
gender_source_value
gender_source_concept_id

10 Standardized Derived Elements

DRUG_ERA
DOSE_ERA
CONDITION_ERA



Lisinopril

Common brands: Qbrelis, Prinivil, Zestril

ACE inhibitor

It can treat high blood pressure and heart failure. It can also reduce the risk of death after a heart attack.

Brands: Qbrelis, Prinivil, and Zestril

Availability: Prescription needed

Pregnancy: Consult a doctor

Alcohol: Interactions can occur

OMOP Tables

- The CDM contains 16 Clinical Event tables, 10 Vocabulary tables, 2 metadata tables, 4 health system data tables, 2 health economics data tables, 3 standardized derived elements, and 2 Results schema tables.
- https://ohdsi.github.io/CommonDataModel/cdm60.html

OMOP example about Lauren

- https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html

Section 4.3

More details

- 2019 Tutorials – OMOP Common Data Model and Standardized Vocabularies (6 hours long)
  - https://www.ohdsi.org/2019-tutorials-omop-common-data-model-and-standardized-vocabularies/
- Book of OHDSI
  - https://ohdsi.github.io/TheBookOfOhdsi/

Data Standardization

Data standardization is the critical process of bringing data into a common format that allows for collaborative research, large-scale analytics, and sharing of sophisticated tools and methodologies. Why is it so important?

Healthcare data can vary greatly from one organization to the next. Data are collected for different purposes, such as provider reimbursement, clinical research, and direct patient care. These data may be stored in different formats using different database systems and information models. And despite the growing use of standard terminologies in healthcare, the same concept (e.g., blood glucose) may be represented in a variety of ways from one setting to the next.

We at OHDSI are deeply involved in the evolution and adoption of a Common Data Model known as the OMOP Common Data Model. We provide resources to convert a wide variety of datasets into the CDM, as well as a plethora of tools to take advantage of your data once it is in CDM format.

Concept and Standard

CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	

Primary key

English description

Domain

Vocabulary

Class in vocabulary

Standard, Source of Classification

Code in vocabulary

Valid during time interval

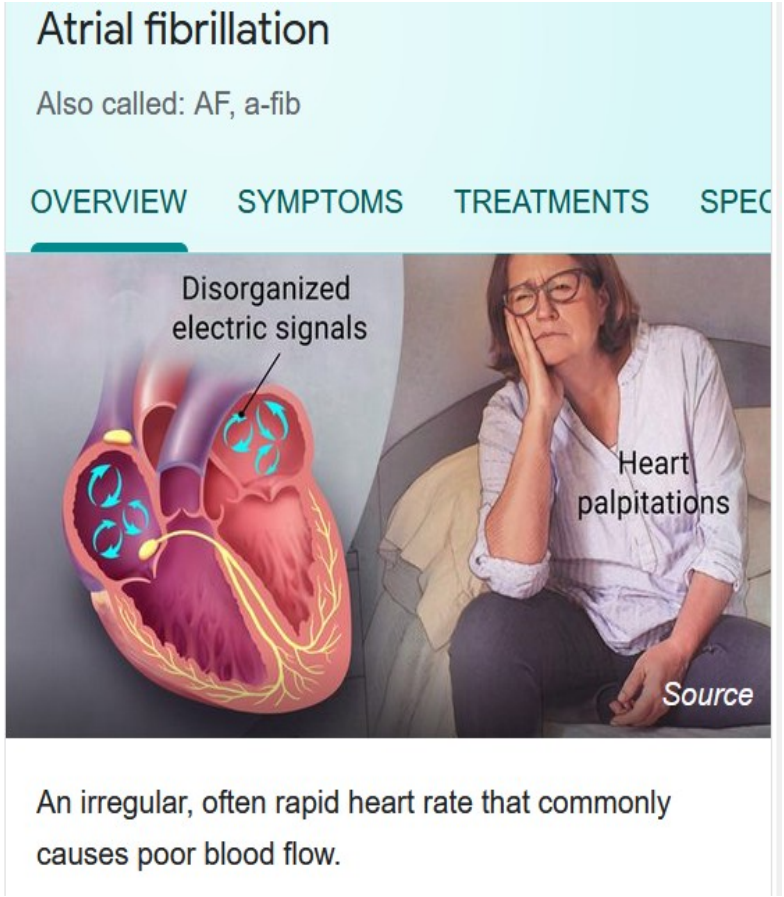


Figure 5.2: Standard representation of vocabulary concepts in the OMOP CDM. The example provided is the CONCEPT table record for the SNOMED code for Atrial Fibrillation.

Table 5.1: Vocabularies with or without horizontal and vertical sub-classification principles in concept class.

Concept class subdivision principle	Vocabulary
Horizontal	all drug vocabularies, ATC, CDT, Episode, HCPCS, HemOnc, ICDs, MedDRA, OSM, Census
Vertical	CIEL, HES Specialty, ICDO3, MeSH, NAACCR, NDFRT, OPCS4, PCORNET, Plan, PPI, Provider, SNOMED, SPL, UCUM
Mixed	CPT4, ISBT, LOINC
None	APC, all Type Concepts, Ethnicity, OXMIS, Race, Revenue Code, Sponsor, Supplier, UB04s, Visit

Horizontal concept classes allow you to determine a specific hierarchical level. For example, in the drug vocabulary RxNorm the concept class “Ingredient” defines the top level of the hierarchy. In the vertical model, members of a concept class can be of any hierarchical level from the top to the very bottom.

Standard and Hierarchies

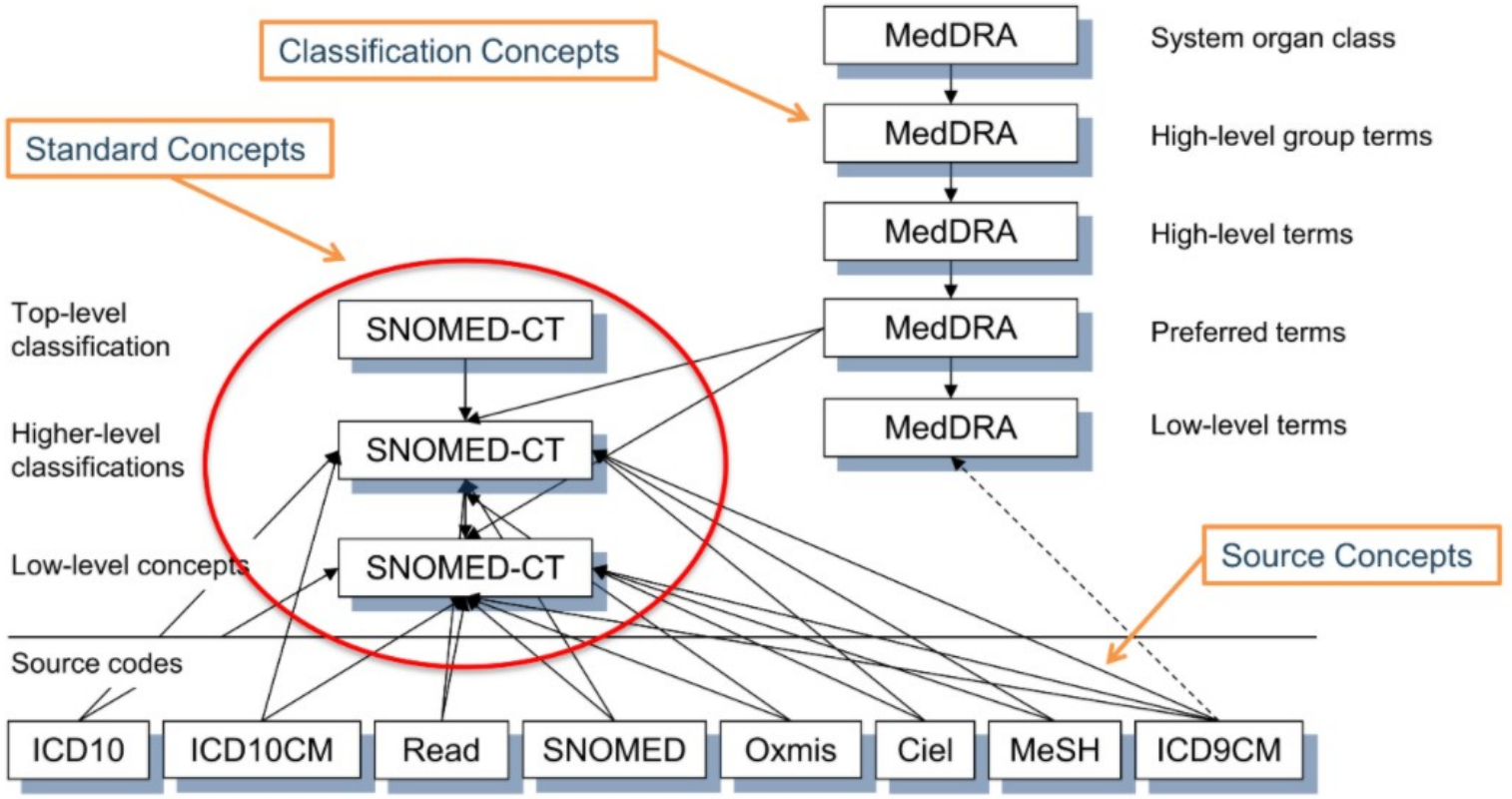


Figure 5.4: Standard, non-standard source and classification concepts and their hierarchical relationships in the condition domain. SNOMED is used for most standard condition concepts (with some oncology-related concepts derived from ICDO3), MedDRA concepts are used for hierarchical classification concepts, and all other vocabularies contain non-standard or source concepts, which do not participate in the hierarchy.

MedDRA

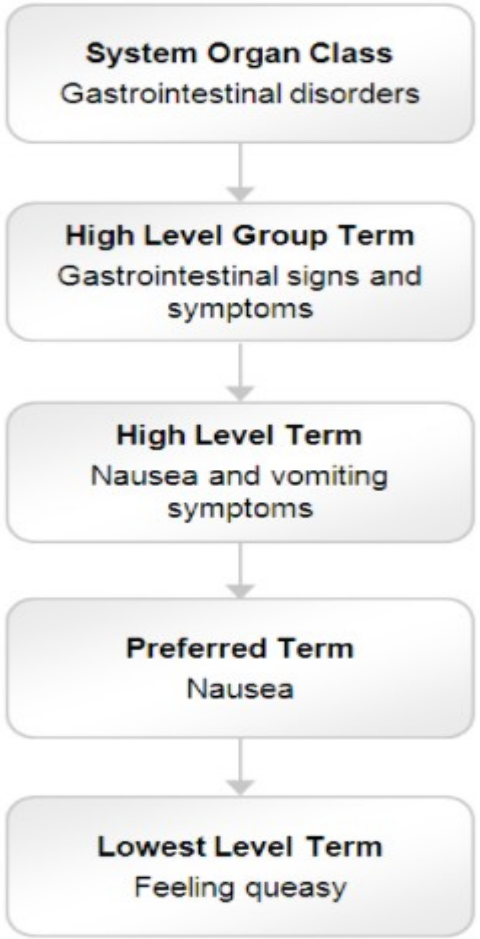
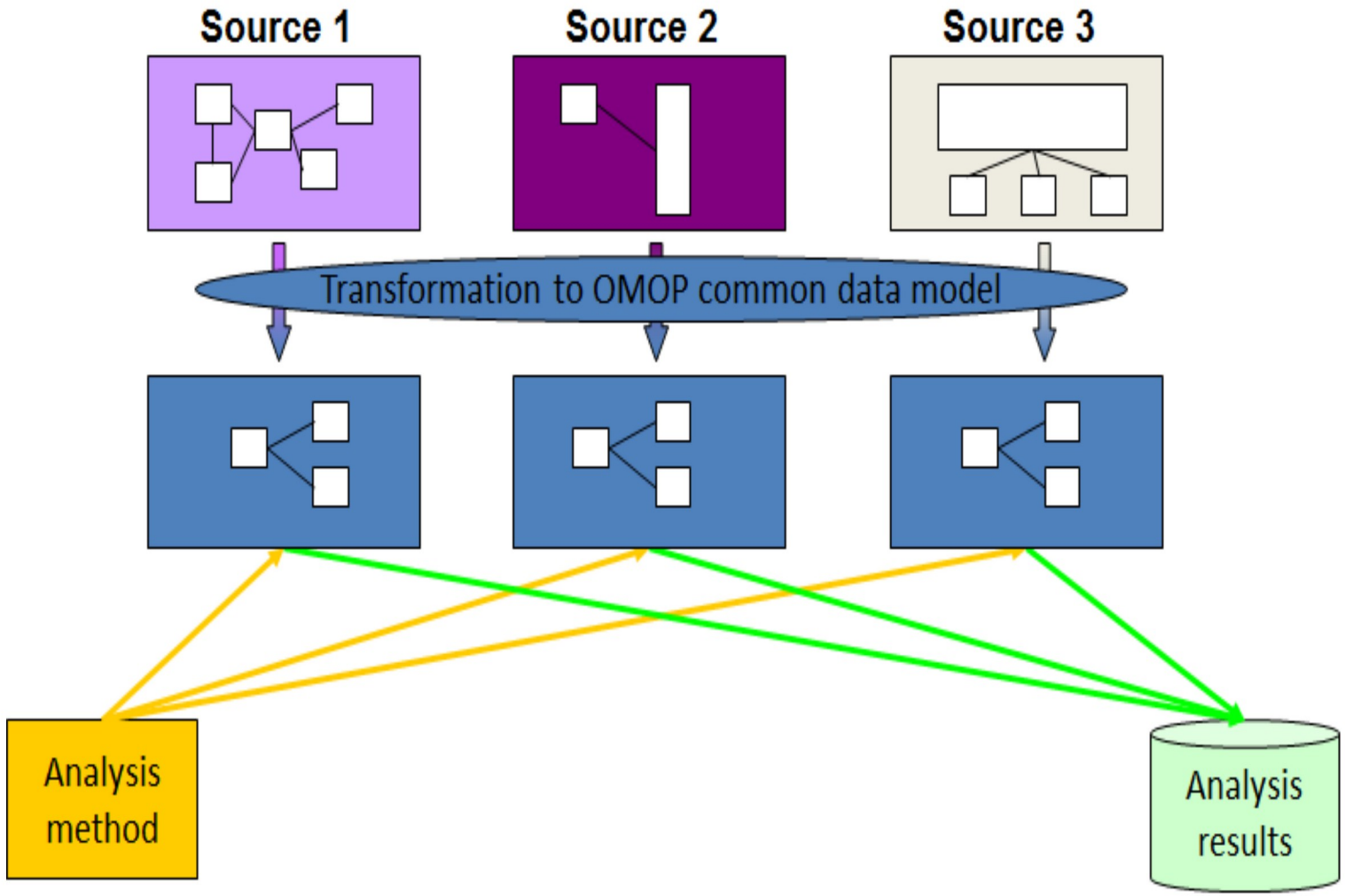


Table 5.2: List of vocabularies to utilize for Standard/non-standard/classification concept assignments.

Domain	for Standard Concepts	for source concepts	for classification concepts
Condition	SNOMED, ICDO3	SNOMED Veterinary	MedDRA
Procedure	SNOMED, CPT4, HCPCS, ICD10PCS, ICD9Proc, OPCS4	SNOMED Veterinary, HemOnc, NAACCR	None at this point
Measurement	SNOMED, LOINC	SNOMED Veterinary, NAACCR, CPT4, HCPCS, OPCS4, PPI	None at this point
Drug	RxNorm, RxNorm Extension, CVX	HCPCS, CPT4, HemOnc, NAAACCR	ATC
Device	SNOMED	Others, currently not normalized	None at this point
Observation	SNOMED	Others	None at this point
Visit	CMS Place of Service, ABMT, NUCC	SNOMED, HCPCS, CPT4, UB04	None at this point

OMOP Common Data Model



What is the OMOP Common Data Model (CDM)?

The OMOP Common Data Model allows for the systematic analysis of disparate observational databases. The concept behind this approach is to transform data contained within those databases into a common format (data model) as well as a common representation (terminologies, vocabularies, coding schemes), and then perform systematic analyses using a library of standard analytic routines that have been written based on the common format.

Why do we need a CDM?

Observational databases differ in both purpose and design. Electronic Medical Records (EMR) are aimed at supporting clinical practice at the point of care, while administrative claims data are built for the insurance reimbursement processes. Each has been collected for a different purpose, resulting in different logical organizations and physical formats, and the terminologies used to describe the medicinal products and clinical conditions vary from source to source.

The CDM can accommodate both administrative claims and EHR, allowing users to generate evidence from a wide variety of sources. It would also support collaborative research across data sources both within and outside the United States, in addition to being manageable for data owners and useful for data users.

Applications and Use Cases

- The OHDSI collaboration focuses on generating reliable evidence from real-world healthcare data, typically in the form of claims databases or electronic health record databases. The use cases that OHDSI focuses on fall into three major categories:
  - Characterization
  - Population-level estimation
  - Patient-level prediction

Characterization

- The data can provide answers to questions like:
  - For patients newly diagnosed with atrial fibrillation, how many receive a prescription for warfarin?
  - What is the average age of patients who undergo hip arthroplasty?
  - What is the incidence rate of pneumonia in patients over 65 years old?

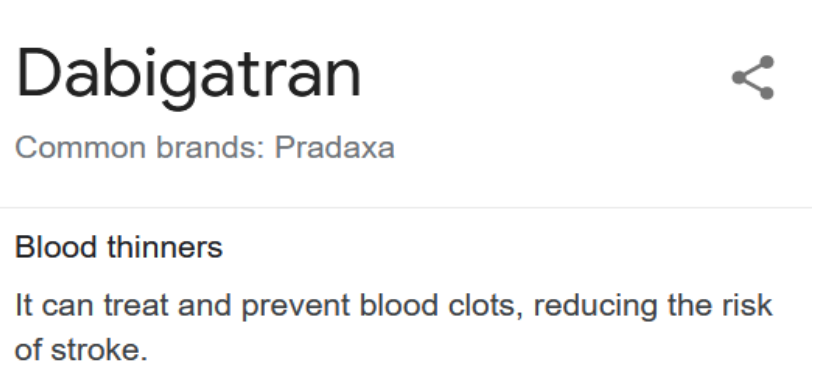
Typical characterization questions are formulated as:

- How many patients…?
- How often does…?
- What proportion of patients…?
- What is the distribution of values for lab…?
- What are the HbA1c levels for patients with…?
- What are the lab values for patients…?
- What is the median length of exposure for patients on….?
- What are the trends over time in…?
- What are other drugs that these patients are using?
- What are concomitant therapies?
- Do we have enough cases of…?
- Would it be feasible to study X…?
- What are the demographics of…?
- What are the risk factors of…?

And the desired output is:

- Count or percentage
- Averages
- Descriptive statistics
- Incidence rate
- Prevalence
- Cohort
- Rule-based phenotype
- Drug utilization
- Disease natural history
- Adherence
- Co-morbidity profile
- Treatment pathways
- Line of therapy

Population-level Estimatic



- The data can provide answers to questions like:
  - For patients newly diagnosed with atrial fibrillation, in the first year after therapy initiation, does warfarin cause more major bleeds than dabigatran?
  - Does the causal effect of metformin on diarrhea vary by age?

Typical population-level effect estimation questions are formulated as:

- What is the effect of…?
- What if I do intervention…?
- Which treatment works better?
- What is the risk of X on Y?
- What is the time-to-event of…?

And the desired output is:

- Relative risk
- Hazards ratio
- Odds ratio
- Average treatment effect
- Causal effect
- Association
- Correlation
- Safety surveillance
- Comparative effectiveness

Patient-Level Prediction

- The data can provide answers to questions like:
- For a specific patient newly diagnosed with major depressive disorder, what is the probability the patient will attempt suicide in the first year following diagnosis?
- For a specific patient newly diagnosed with atrial fibrillation, in the first year after therapy initiation with warfarin, what is the probability the patient suffers an ischemic stroke?

Typical patient-level prediction questions are formulated as:

- What is the chance that this patient will…?
- Who are candidates for…?

And the desired output is:

- Probability for an individual
- Prediction model
- High/low risk groups
- Probabilistic phenotype

OHDSI Analytics Tools

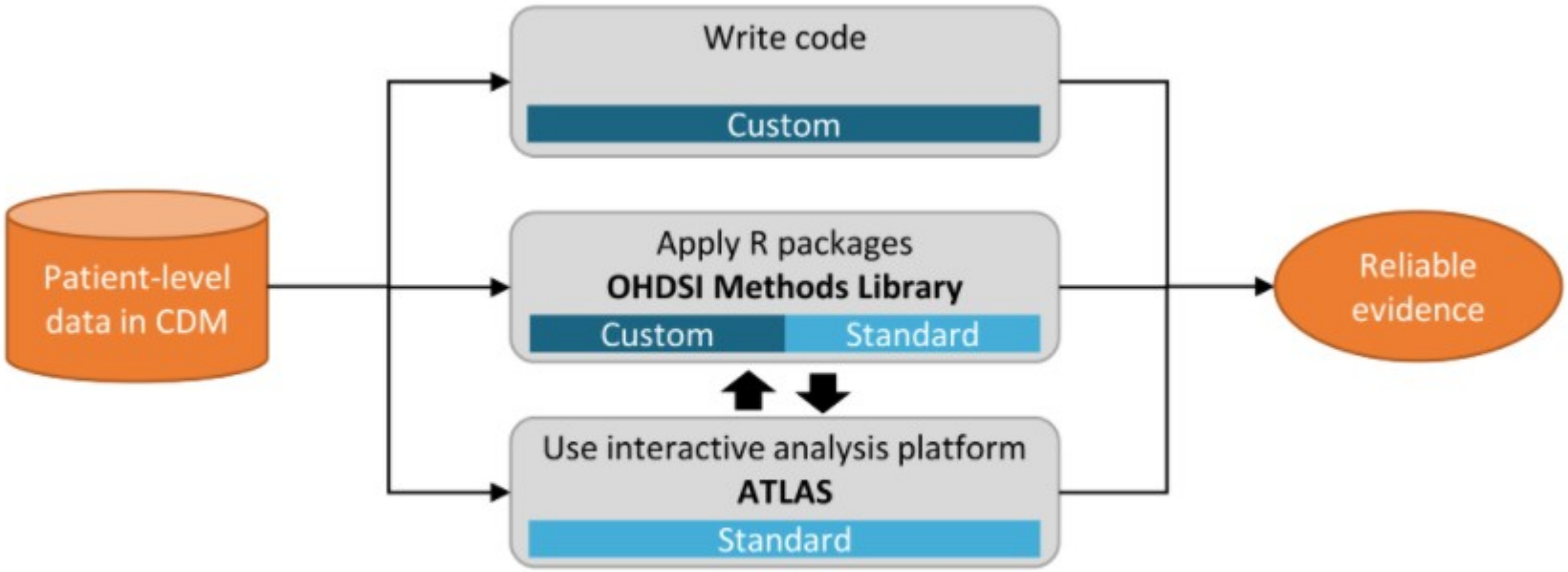


Figure 8.1: Different ways to implement an analysis against data in the CDM.

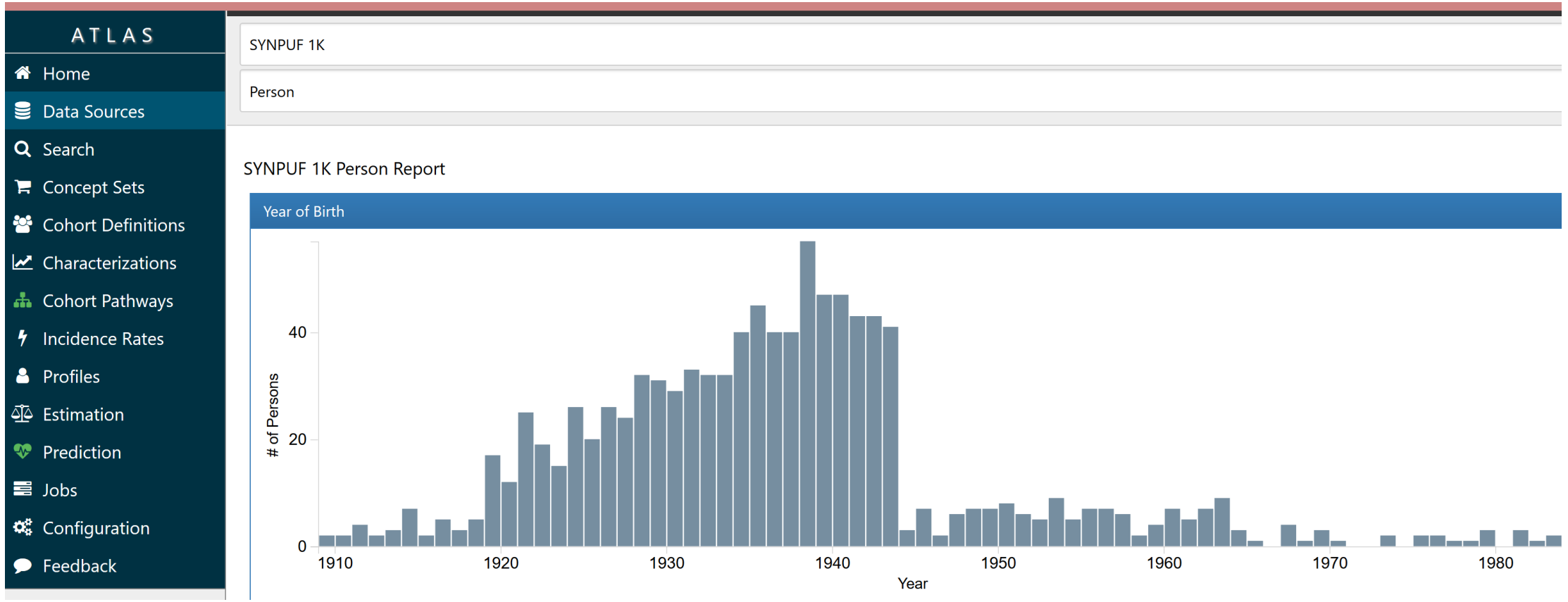
ATLAS

- ATLAS is an open source application developed as a part of OHDSI intended to provide a unified interface to patient level data and analytics.

- https://www.youtube.com/watch?v=dr9FhEkf04o&feature=youtu.be
- Or more ATLAS tutorial videos: https://www.youtube.com/user/OHDSIJoinTheJourney

ATLAS

https://www.ohdsi.org/software-tools/
https://atlas-demo.ohdsi.org/



More videos about ATLAS demo are here:
https://github.com/OHDSI/Atlas/wiki/Cohort-Pathways

Cohort Definition

- In OHDSI research, we define a cohort as a set of persons who satisfy one or more inclusion criteria for a duration of time.
- The term cohort is often interchanged with the term *phenotype*
- For instance, in a study aiming to predict the risk of angioedema in a group of people initiation ACE inhibitors, we define two cohorts: the outcome cohort (angioedema), and the target cohort (people initiating ACE inhibitors).

Figure 10.1: Building Blocks of Cohort definitions.

Cohort Definition

- When creating a cohort definition, you need to ask yourself the following questions:
  - *What initial event defines the time of cohort entry?*
  - *What inclusion criteria are applied to the initial events?*
  - *What defines the time of cohort exit?*

Cohort entry event: The cohort entry event (initial event) defines the time when people enter the cohort, called the **cohort index date**. A cohort entry event can be any event recorded in the CDM such as drug exposures, conditions, procedures, measurements and visits. Initial events are defined by the CDM domain where the data are stored (e.g. PROCEDURE_OCCURRENCE, DRUG_EXPOSURE, etc), the concept sets built to identify the clinical activity (e.g. SNOMED codes for conditions, RxNorm codes for drugs) as well as any other specific attributes (e.g. age at occurrence, first diagnosis/procedure/etc, specifying start and end date, specifying visit type or criteria, days supply, etc).

Inclusion criteria: Inclusion criteria are applied to the initial event cohort to further restrict the set of people. Each inclusion criterion is defined by the CDM domain(s) where the data are stored, concept set(s) representing the clinical activity, domain-specific attributes (e.g. days supply, visit type, etc), and the temporal logic relative to the cohort index date

Cohort exit criteria: The cohort exit event signifies when a person no longer qualifies for cohort membership. Cohort exit can be defined in multiple ways such as the end of the observation period, a fixed time interval relative to the initial entry event, the last event in a sequence of related observations (e.g. persistent drug exposure) or through other censoring of observation period.

Cohort Definition for Hypertension

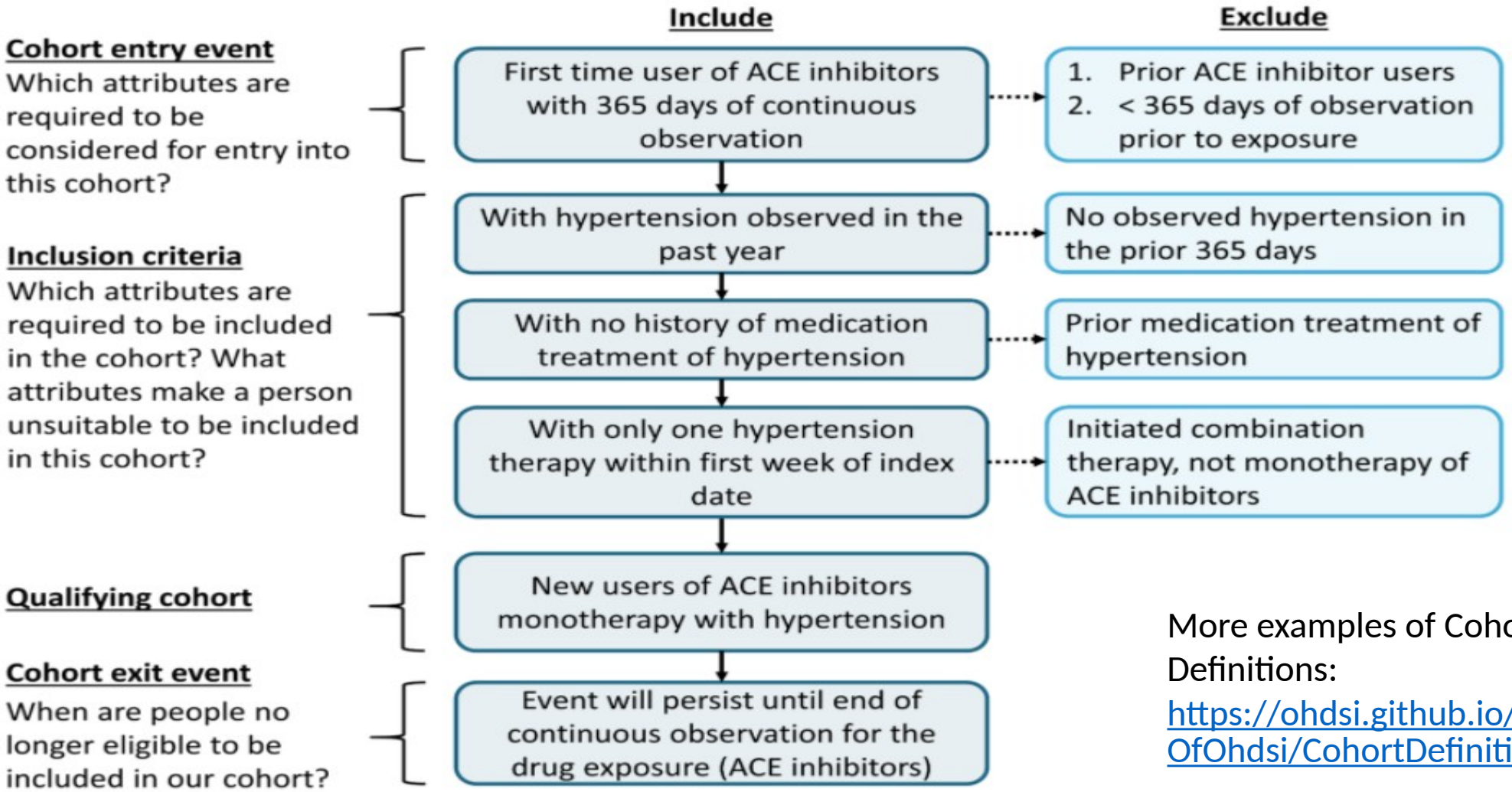


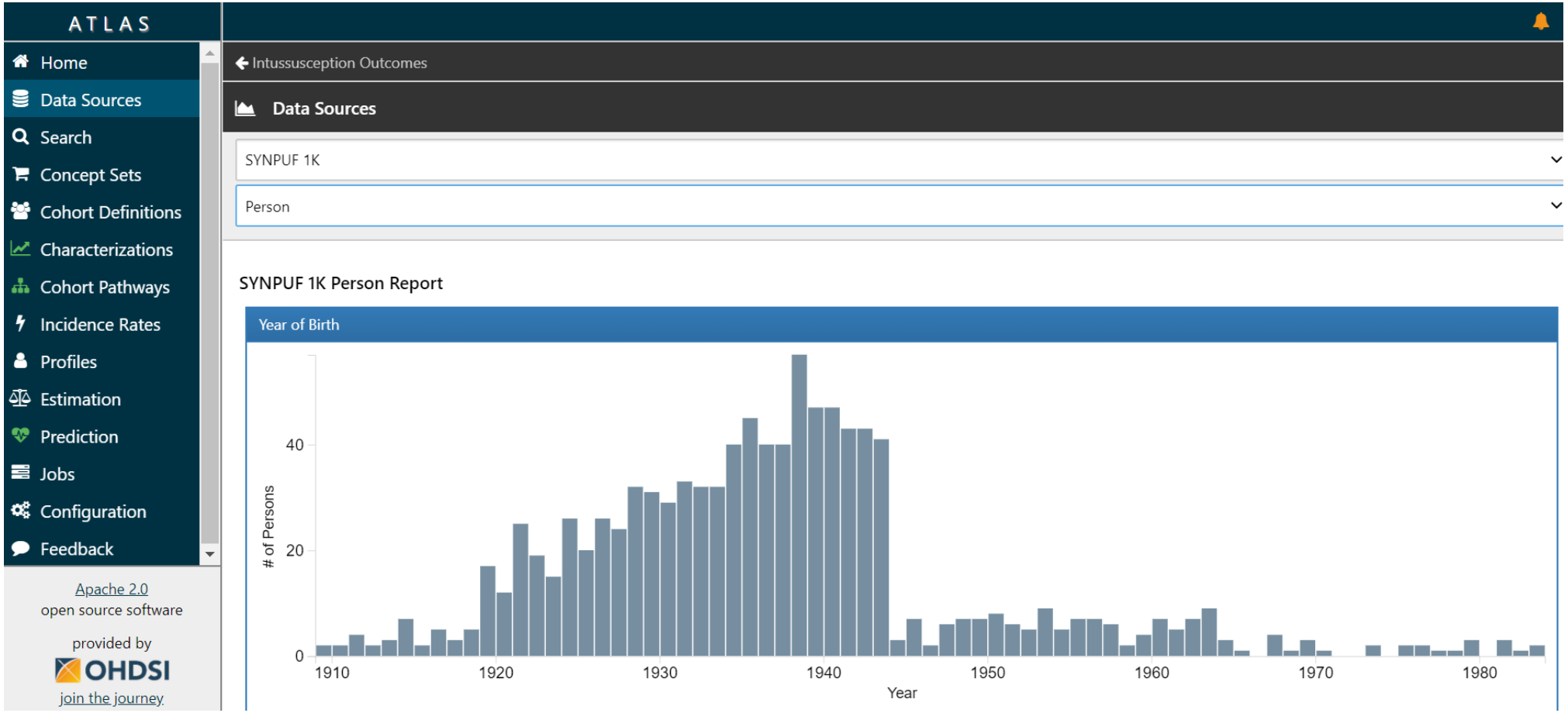
Figure 10.3: Logical Diagram of Intended Cohort

More examples of Cohort Definitions: https://ohdsi.github.io/TheBookOfOhdsi/CohortDefinitions.html

Characterization

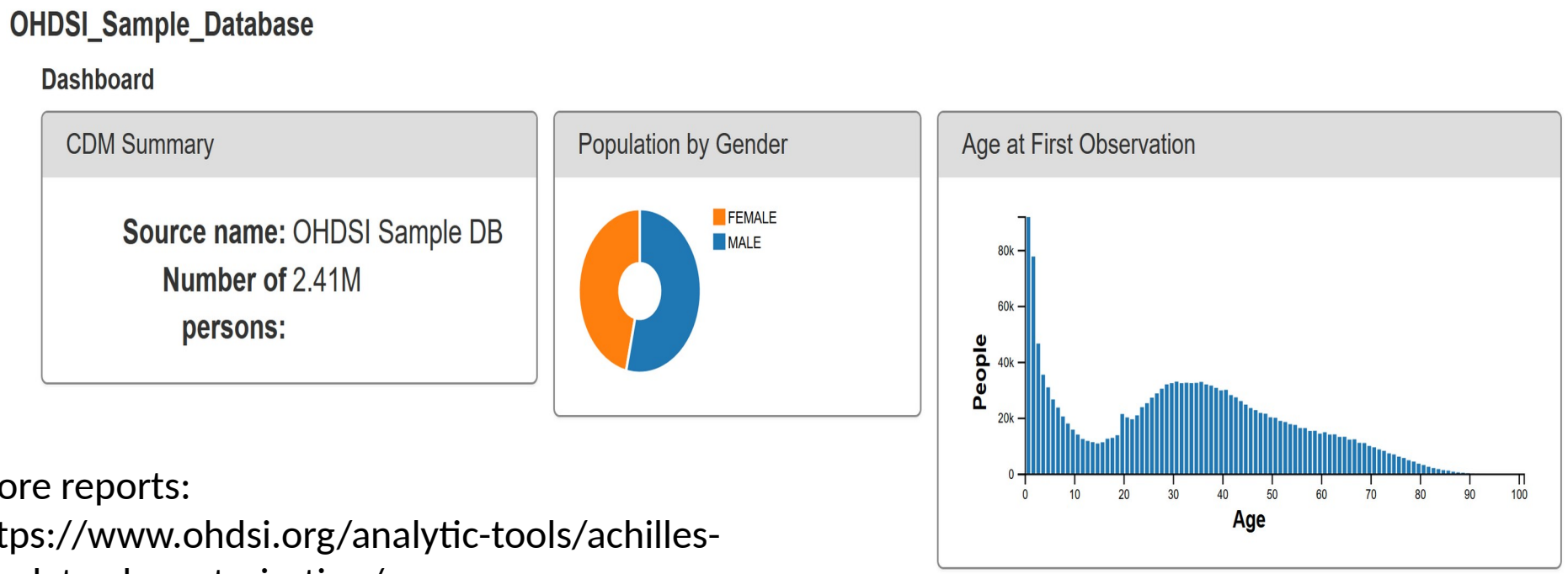
- **Database-level characterization**: provides a top-level set of summary statistics to understand the data profile of a database in its totality.
- **Cohort characterization**: describes a population in terms of its aggregate medical history.
- **Treatment pathways**: describes the sequence of interventions a person received for a duration of time.
- **Incidence**: measures the occurrence rate of an outcome in a population for a time at risk.

Database Characterization



ACHILLES

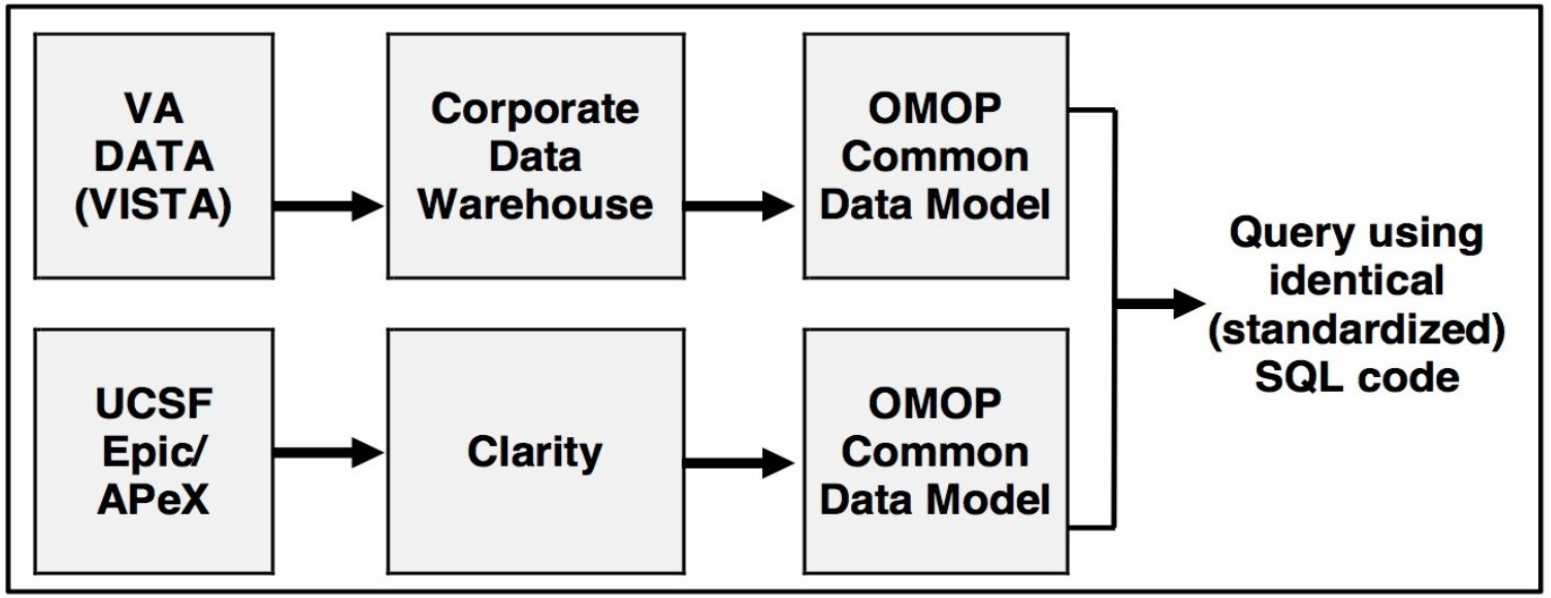
- Automated Characterization of Health Information at Large-scale Longitudinal Exploration System



More reports: https://www.ohdsi.org/analytic-tools/achilles-for-data-characterization/

OMOP Implementations

CHIME is supporting investigators who wish to use national OMOP data from the VA and/or local OMOP data from UCSF for clinical research projects. Through the pSCANNER project, data from all 5 University of California medical centers and the Veterans Health Administration (>150 medical centers) have been transformed into the OMOP common data model (version 4.0). Standardized structured query language (SQL) queries are shared in a common open-source repository and updated regularly. All data documentation is freely available online.



OHDSI Network of Networks

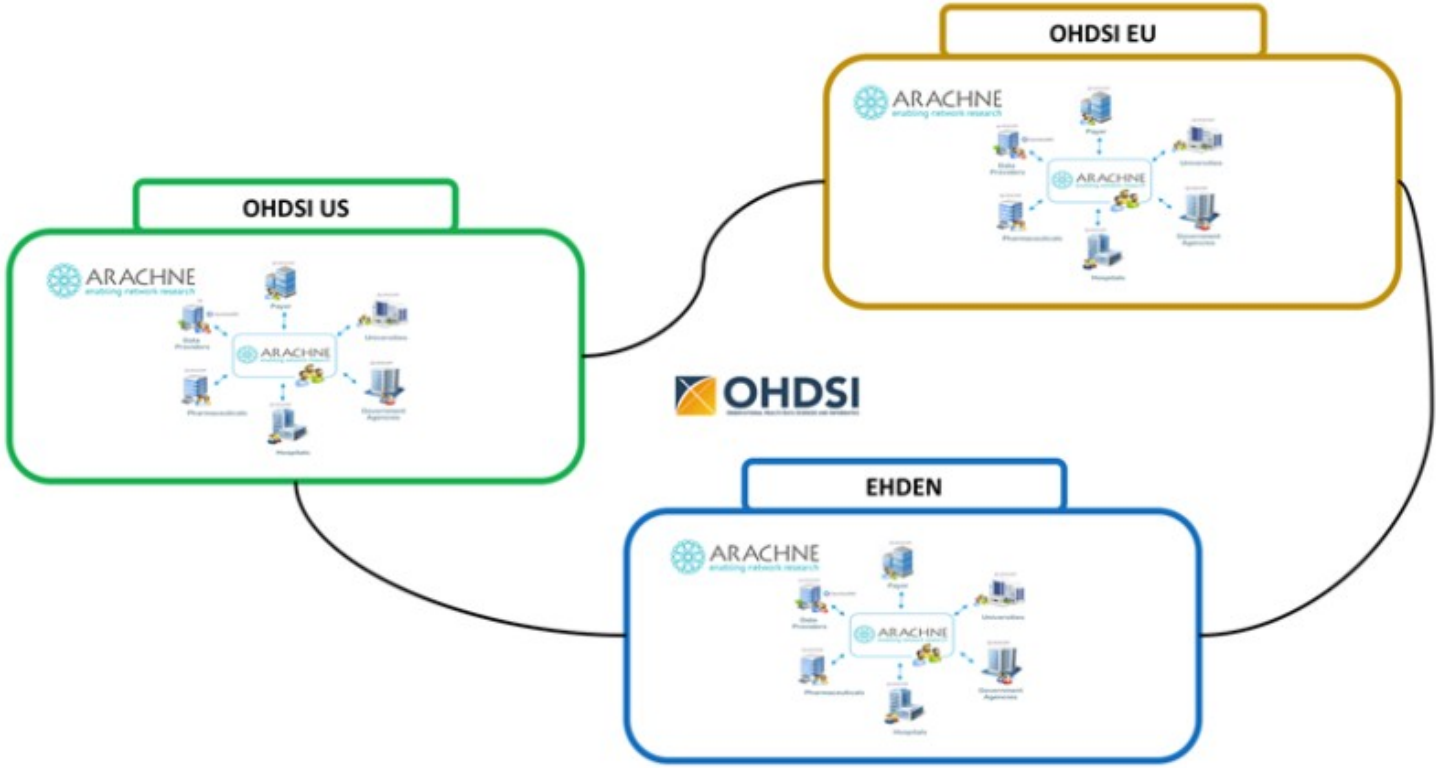


Figure 20.2: The ARACHNE Network of Networks.

Highly Recommended:
OHDSI Book:
https://ohdsi.github.io/TheBookOfOhdsi/

The European Health Data & Evidence Network (EHDEN)

MIMIC to OMOP

- https://physionet.org/content/mimic-iv-demo-omop/0.9/

Database Open Access

MIMIC-IV demo data in the OMOP Common Data Model

Michael Kallfelz , Anna Tsvetkova , Tom Pollard , Manlik Kwong , Gigi Lipori , Vojtech Huser , Jeffrey Osborn , Sicheng Hao , Andrew Williams

Published: June 21, 2021. Version: 0.9

MIMIC-IV demo available in the OMOP Common Data Model *(June 28, 2021, 12:36 p.m.)*

We are pleased to announce that a 100-patient demo of MIMIC-IV has been made available in the OMOP Common Data Model. The dataset is currently undergoing user testing and has known limitations (for example, the inputevents and outputevents tables are not yet incorporated). For more detail, please visit the project page on PhysioNet and the associated GitHub repository.

This work builds on previous efforts by Nicolas Paris, Adrien Parrot and colleagues on MIMIC-III. The project was in part supported by grants from Bill and Melinda Gates foundation and National Library of Medicine (NLM), National Institutes of Health.

i2b2

John Halamka

- Dr. John Halamka is he CIO of Beth Israel Deaconess Medical Center in Boston. In 2011, his wife, 49 year-old Korean woman, diagnose with breast cancer.
- Using big data available at Harvard hospitals, he can ask:
  - Of the last 10,000 Asian women near age 50 who were treated for the same tumor, what medications were used?
  - Was surgery or radiation necessary?
  - What are the outcomes?
- https://pt-br.facebook.com/SearchHealthIT/videos/john-halamka-uses-big-data-analytics-in-healthcare-to-fight-wi/1151468341554782/

i2b2

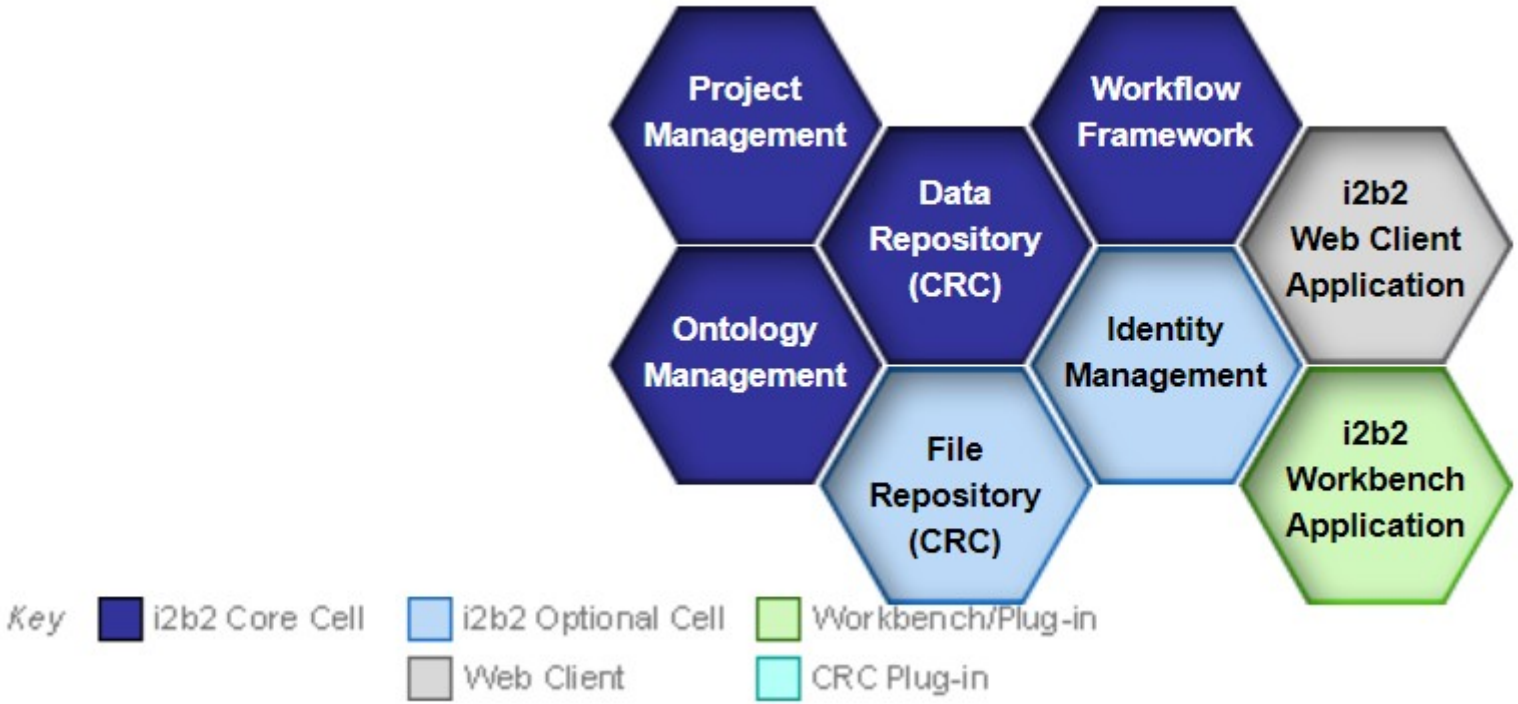
- i2b2 is an open-source clinical data warehousing and analytics research platform used at over 250 locations worldwide. i2b2 enables sharing, integration, standardization, and analysis of heterogeneous data from healthcare and research. https://community.i2b2.org/wiki/ : The i2b2 Community is a life-sciences-focused open-source, open-data community.

- The system was first developed within the Partner's HealthCare system in Boston at Massachusetts General Hospital (MGH). It served as the architecture for their Research Patient Data Registry (RPDR). Investigators at Harvard Medical School (Isaac Kohane, PI), in conjunction with investigators at MGH, applied to the NIH for funding to open-source the code and release it to the general research community. The grant was approved and the first version of the code was released in November 2007.

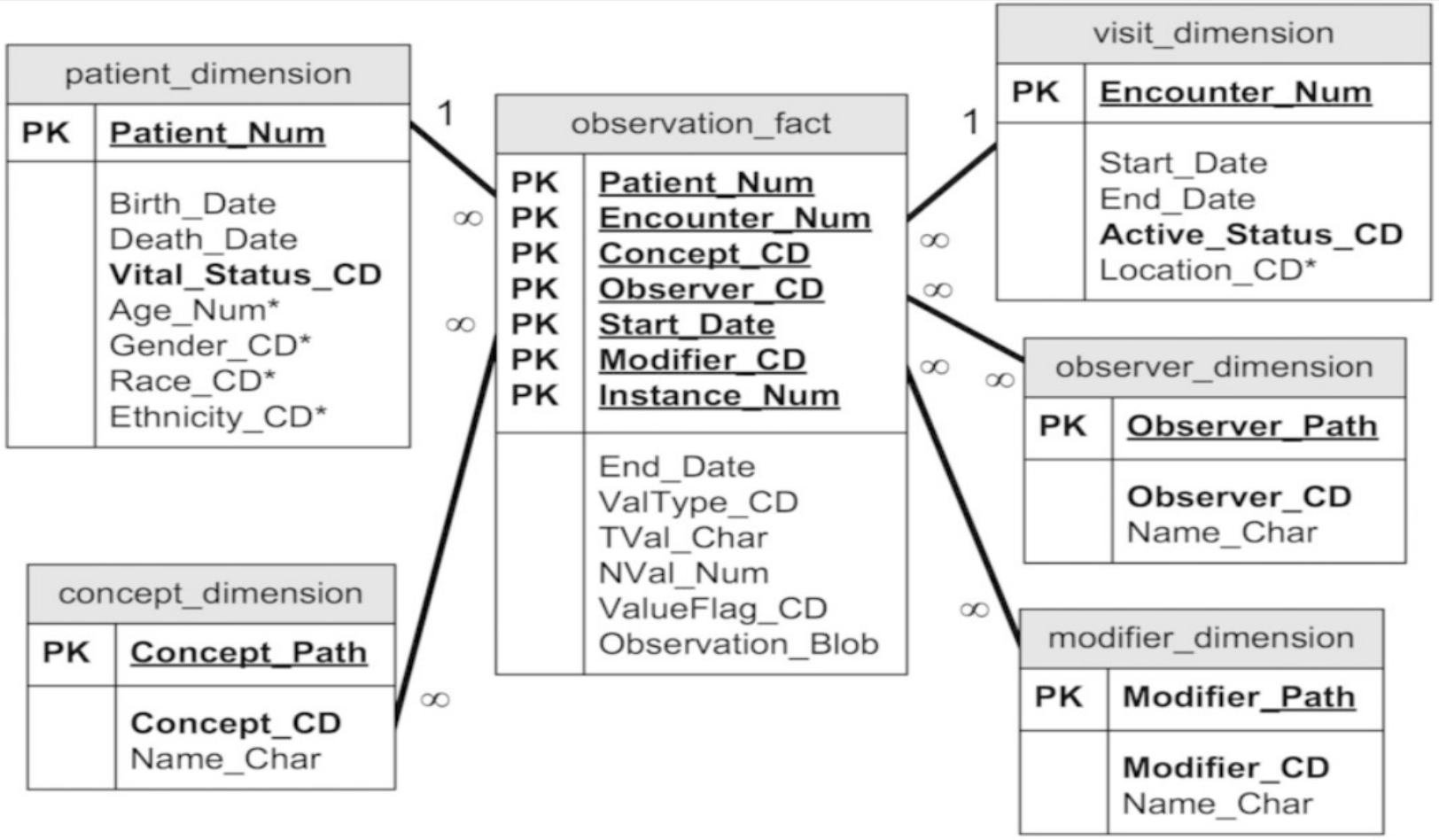
- i2b2 is part of the i2b2 tranSMART Foundation **(**https://i2b2transmart.org/**)**, which brings together an NIH-funded enterprise clinical research platform (i2b2) and pharma-developed software for translational research studies (tranSMART).

i2b2

- https://community.i2b2.org/wiki/ (interactive interpretation of each hive)



I2b2 data model



The i2b2 star schema.

Snowflake

Company

Snowflake Inc. is a cloud computing–based data cloud company based in Bozeman, Montana. It was founded in July 2012 and was publicly launched in October 2014 after two years in stealth mode. The firm offers a cloud-based data storage and analytics service, generally termed "data-as-a-service". Wikipedia

Stock price: SNOW (NYSE) $140.87 -1.28 (-0.90%)

Jan 13, 4:00 PM EST - Disclaimer

Founded: July 23, 2012, San Mateo, CA

Headquarters: Bozeman, MT

Number of employees: 3,992 (Jan 2022)

Revenue: 1.219 billion USD (2022)

Founders: Benoît Dageville, Thierry Cruanes

Subsidiaries: Streamlit, Inc., CryptoNumerics Inc., MORE

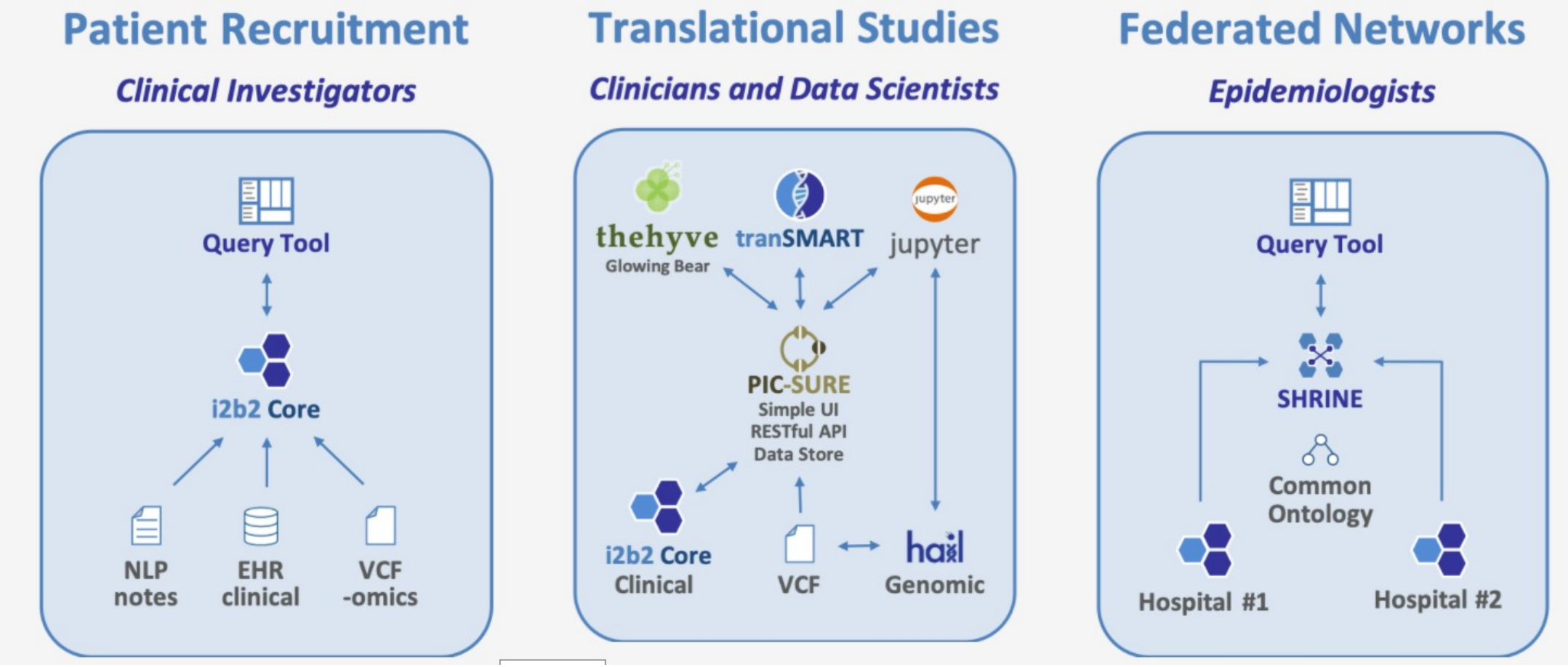
Total assets: 6.65 billion USD (2022)

I2b2 data model and OMOP

- The data models of i2b2 and OMOP have many similarities. We take advantage of these similarities to offer an evolution of the i2b2 software that adapts to the OMOP data model. This allows the query formulation in the i2b2 software that relies on the i2b2 Application Programming Interface (API) to be utilized on top of an OMOP data source. As a result, most of the functionality of i2b2 Software is preserved on an OMOP data source (with proper respect to the OHDSI ontologies). As a result a SHRINE tool or a SMART-on-FHIR tool that relies on the i2b2 API, over OHDSI ontologies, can run with an OMOP data source.
- The standard i2b2 data model is comprised of a central fact table (observation_fact) surrounded by multiple dimension tables arranged in a star schema. In the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), rather than a central fact table, we have a collection of them distinguished by domain: procedures, condition, drug, measurement, observation, etc. We run queries against multiple domain fact tables as dictated by the metadata.

i2b2 tranSMART tools

https://github.com/thehyve



i2b2 tranSMART tools

- **Core products**

  i2b2 Core – A common data model, application layer and APIs

  tranSMART – A suite of data exploration, visualization and ETL tools

- **Plugins and 3rd party products (examples) extend functionality**

  Query Tool: Advanced cohort builder

  PIC-SURE: RESTful API, data store, and simple user interface

  SHRINE: Federated data networks

  hail: High dimensional genomics data

  jupyter: Interactive web based computing

  thehyve: Tools and UIs from thehyve.nl

- **Libraries provide standards while allowing flexible configurations**

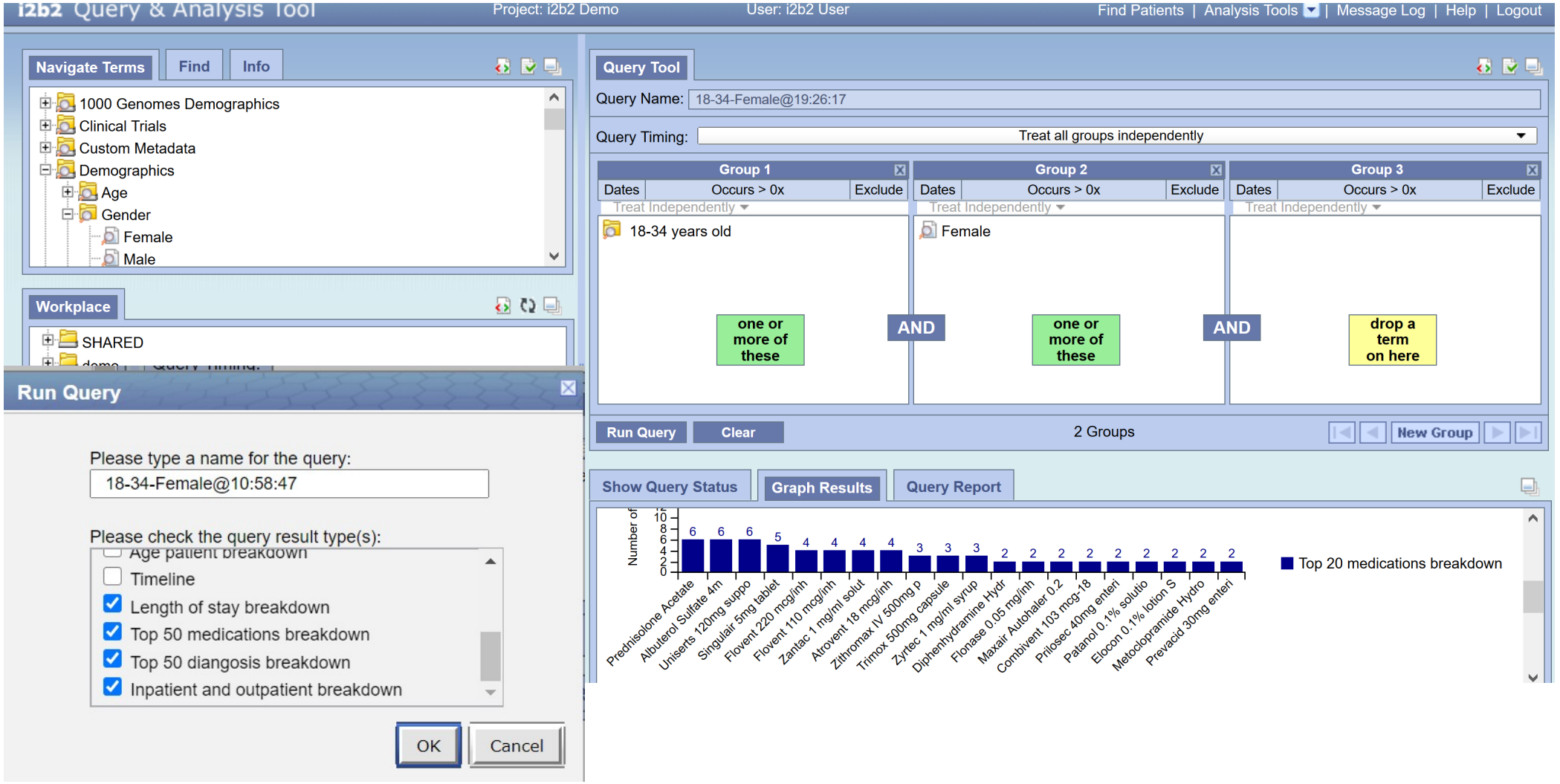
  Ontologies
  ICD9, ICD10, LOINC, RxNorm, ARCH, ACT, …

  Data / Networks
  demo data, NHANES, ARCH, ACT, …

  ETL Processes
  SQL Server, Oracle, Postgres, PopMedNet, …

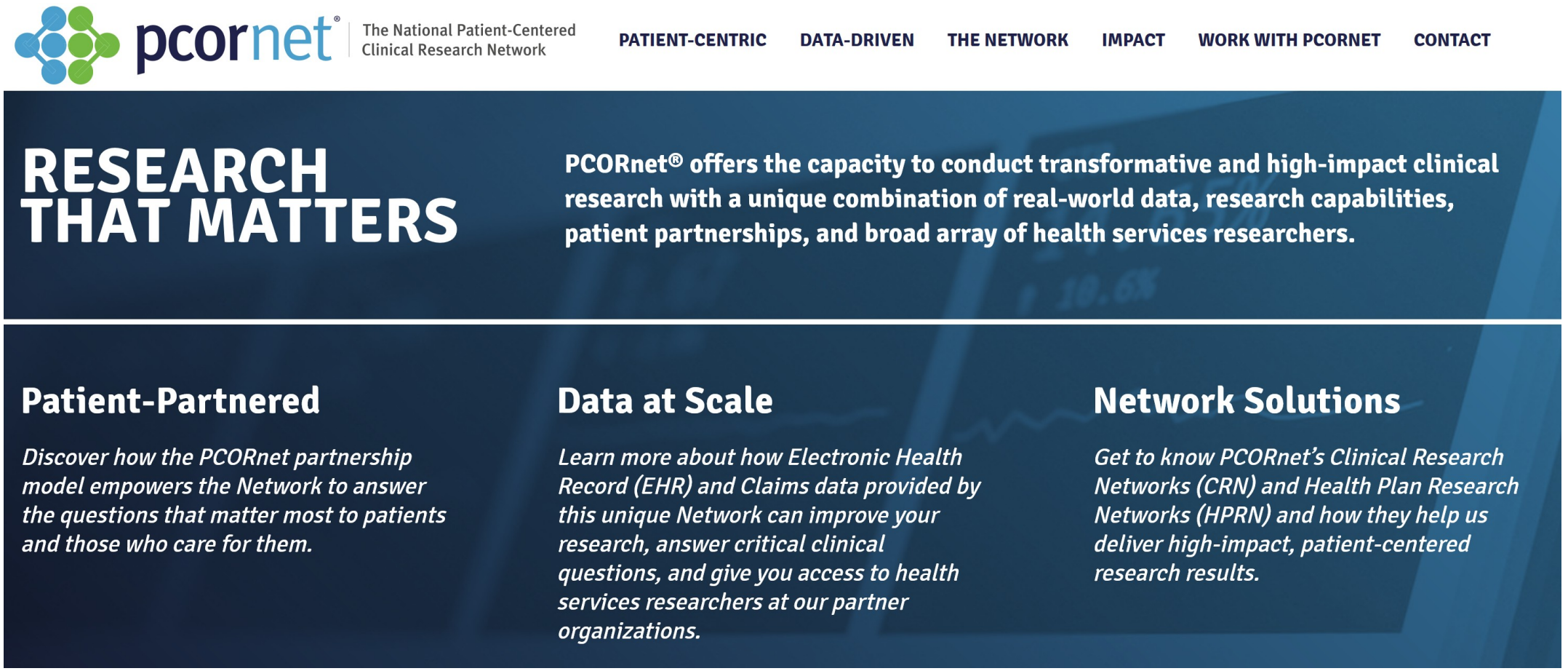
i2b2

https://www.i2b2.org/webclient/



PCORnet

PCOR net



https://pcornet.org/

Real-World Data

PCORnet's data draws from an unparalleled number of electronic health records (EHRs) with growing links to patient-reported and health plan data to create a powerful, standard data set that facilitates large-scale, multi-site research.

STRUCTURED CURATION

Data reside in PCORnet's Clinical Research Networks and are accessed through a rigorously curated query process that results in streamlined data for easier and lower-cost analysis than traditional methods.

HIGH-QUALITY DATA

Given the requirements of PCORnet's Common Data Model and data curation, the Network can identify and mitigate issues that are inherent with real-world data (linking EHRs, health plan data, and patient-reported outcomes).

BREADTH AND CONNECTEDNESS

Drawing from EHRs for millions of Americans along with health plan data, PCORnet is unique in scope. The ability to connect back to patients makes PCORnet an ideal platform for pragmatic trials and observational research.

CHALLENGE

Health systems collect data specific to each system resulting in large data challenges.

PCORnet® SOLUTION

Sourced from differing health systems, our process of querying the data in our Research networks allows it to be analyzed as if it were a single, actionable data set.

IMPACT

PCORnet® data can be used on its own or easily linked to other types of data (payor) to accelerate research.

Clinical Research Network Data - By the Numbers

66+
MILLION PATIENTS ELIGIBLE FOR AN OBSERVATIONAL STUDY

30+
MILLION PATIENTS ELIGIBLE FOR A CLINICAL TRIAL

547+
MILLION CLINICAL ENCOUNTERS IN THE PAST TWO YEARS

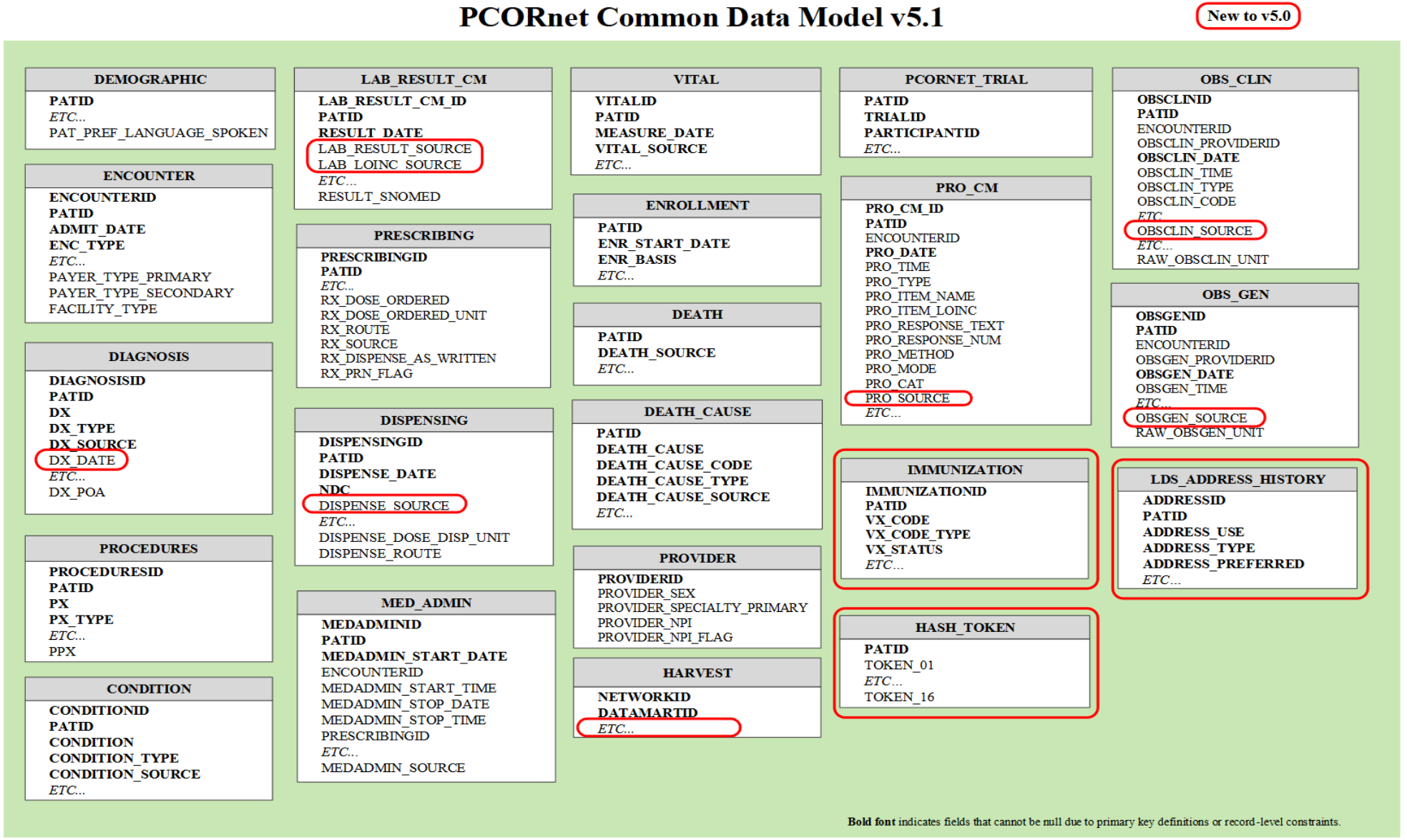
6+
BILLION LAB TESTS LINKED TO LOINC CODES

1+
BILLION MEDICATION ORDERS MAPPED TO A RXNORM CUI

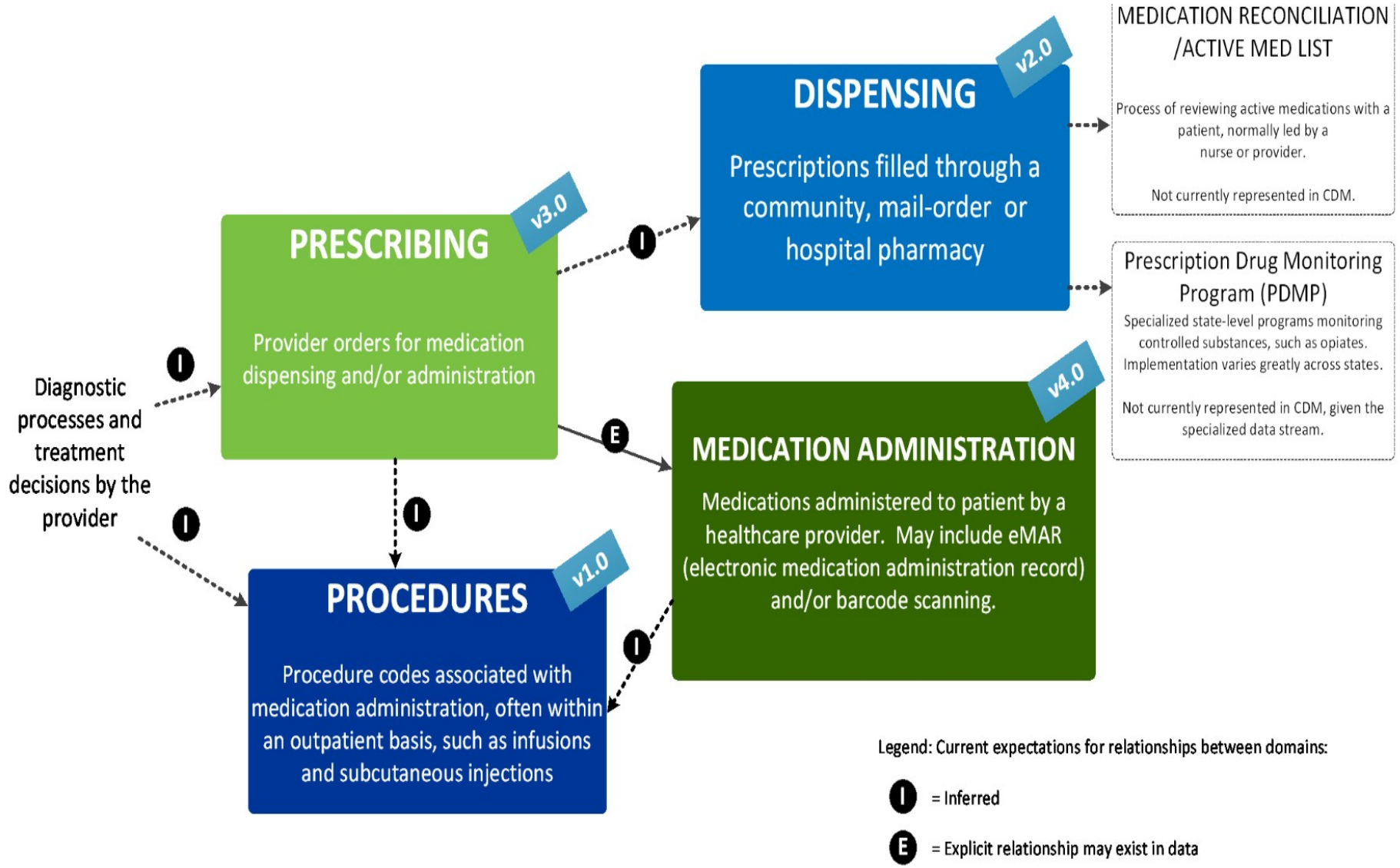
PCORnet Infrastructure is Well Suited for the Conduct of

- Real-world evidence studies
- Pragmatic clinical trials
- Population health research
- Health systems research
- Studies on how best to engage patients in research

PCOR Common Data Model (CDM)



https://cprn.org/data-submission/PCORnet-Common-Data-Model-v51-2019_09_12.pdf (208pages)


MEDICATION RECONCILIATION /ACTIVE MED LIST
Process of reviewing active medications with a patient, normally led by a nurse or provider.
Not currently represented in CDM.
DISPENSING
v2.0
Prescriptions filled through a community, mail-order or hospital pharmacy
Prescription Drug Monitoring Program (PDMP)
Specialized state-level programs monitoring controlled substances, such as opiates. Implementation varies greatly across states.
Not currently represented in CDM, given the specialized data stream.
PRESCRIBING
v3.0
Provider orders for medication dispensing and/or administration
MEDICATION ADMINISTRATION
v4.0
Medications administered to patient by a healthcare provider. May include eMAR (electronic medication administration record) and/or barcode scanning.
Diagnostic processes and treatment decisions by the provider
PROCEDURES
v1.0
Procedure codes associated with medication administration, often within an outpatient basis, such as infusions and subcutaneous injections
I
E
Legend: Current expectations for relationships between domains:
I = Inferred
E = Explicit relationship may exist in data

PCOR net data standardization

- Diagnosis codes would use ICD-9 before October 2015, and ICD-10 afterwards; procedure codes would be encoded in CPT/HCPCS or in ICD-9 before October 2015 and in CPT/HCPCS or ICD-10 afterwards;
- Problem list values would be encoded in SNOMED
- Unified Code for Units of Measure (UCUM) from LOINC
- Lab testing using LOINC
- Prescription medication: RxNorm

FDA Sentinel

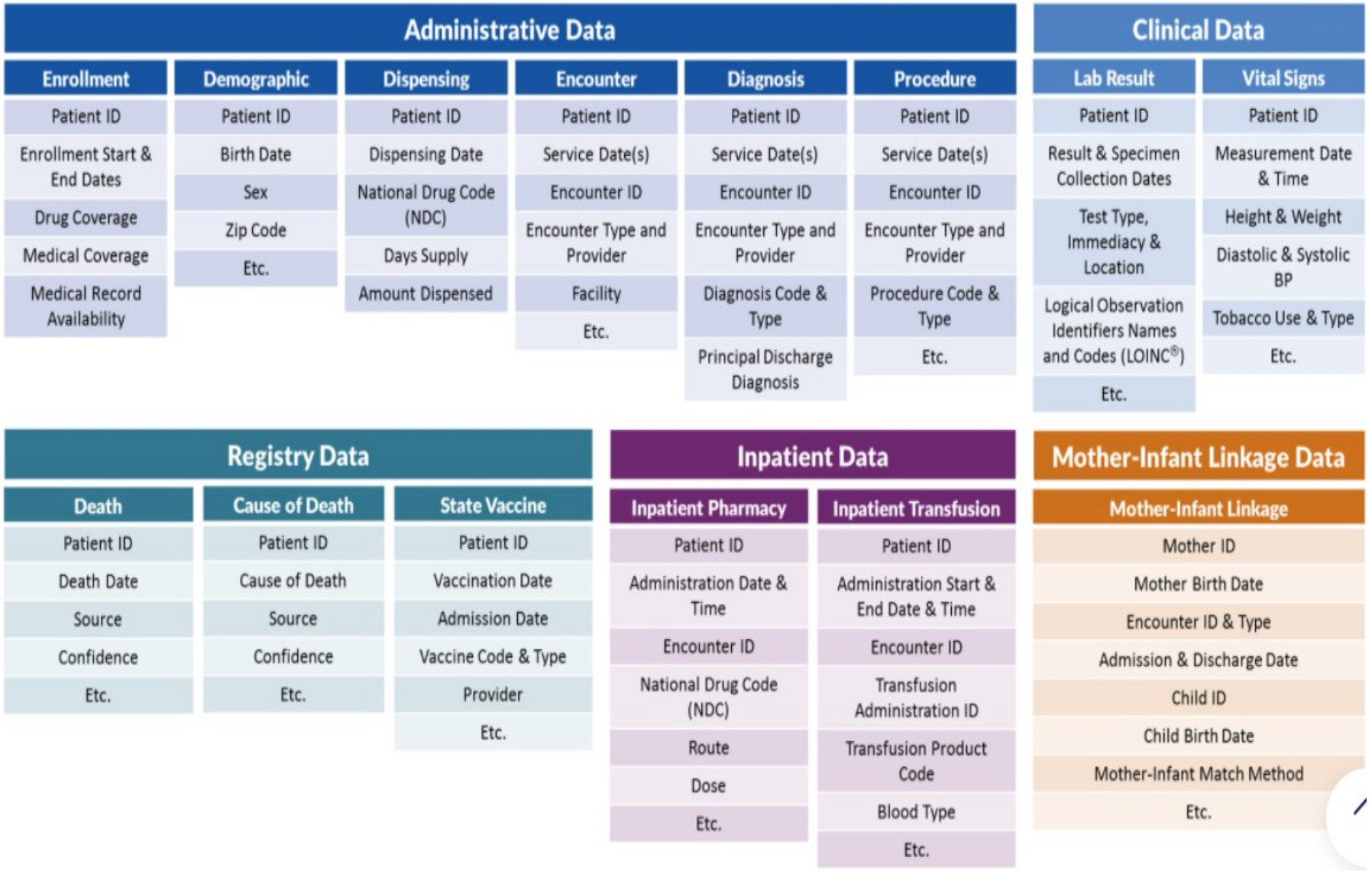
FDA Sentinel

- Sentinel is the FDA's national electronic system which has transformed the way researchers monitor the safety of FDA-regulated medical products, including drugs, vaccines, biologics, and medical devices

FDA Sentinel

- In September 2019, FDA announced Sentinel would expand to three distinct coordinating centers:
  - Sentinel Operations Center: continue to leverage organizational partnerships in the areas of epidemiology, clinical medicine, pharmacy, statistics, health informatics, data science (specifically, artificial intelligence (natural language processing, machine learning)), and network operations to support post-market safety analyses
  - Innovation Center: develop innovative methods to further advance Sentinel, including exploring novel ways to extract and structure information from electronic health records in the future
  - Community Building and Outreach Center: focus on communication and collaboration as well as deepening stakeholder involvement and broadening awareness, access, and use of Sentinel tools and data infrastructure.

Sentinel Common Data Model (SCDM)



Sentinel Common Data Model (SCDM)

- The Sentinel Operations Center (SOC) coordinates the network of Sentinel Data Partners and leads development of the Sentinel Common Data Model (SCDM), a standard data structure that allows Data Partners to quickly execute distributed programs against local data. The SOC Data Core manages creation of the Sentinel Distributed Database (SDD) using the SCDM, and maintains complete documentation of the implementation and characteristics of the SDD. The SDD refers to the data held and maintained by the Data Partners in the SCDM format.