OMOP Common Data Model and Standardized Vocabularies

24-March-2018

Christian Reich, MD, PhD
Mui van Zandt
Erica A. Voss, MPH, PMP

After the Tutorials, you will know…

1. What's OMOP, OHDSI?
2. How does the Standardized Vocabulary work?
3. How do find codes and Concepts?
4. How do I navigate the hierarchy?
5. What is the OMOP CDM?
6. How to use the OMOP CDM

Agenda

Section	Speaker	Time	Item(s)
Registration	-	8:00- 9:00 (1 hour)	
Introduction	Christian	9:00 - 10:00 (1 hour)	Introductions and Ground Rules Foundational • History of OMOP • Why and How • Birth of OHDSI Introduction to OMOP Common Data Model OHDSI Community Example of Remote Study VM Overview
Vocabulary – Part 1	Mui	10:00 – 10:30 (30 min)	Basic Relationship
Break	-	10:30 - 10:45 (15 min)	
Vocabulary – Part 2	Mui	10:45- 12:30 (1 hour & 45 min)	Ancestors & Descendants How does it work for Drugs SQL Examples

Agenda (cont.)

Section	Speaker	Time	Item(s)
Lunch	-	12:30 - 13:30 (1 hour)	-
Vocabulary – Part 3	Mui	13:30 - 14:00 (30 min)	Continued
Common Data Model	Erica	14:00 - 15:35 (1 hour & 30 min)	History of the model In depth discussion of model Era discussion Real World Scenario ETL Pitfalls
Break	-	15:30 - 15:45 (15 min)	-
CDM Examples	Erica / Mui	15:45 - 17:00 (1 hour & 16 min)	Leveraging OHDSI Tools (GitHub/Forums/Working Group) Exercises OHDSI Community Conclusion Game

Instructors

Christian Reich, MD, PhD IQVIA	Mui van Zandt IQVIA	Erica A. Voss, MPH, PMP Janssen R&D

Ground Rules

- Watch your step on the sides of the room!

- We are recording today's session.

- We may table some questions if they are too specific.

- If we cannot get the remote desktop working on your machine let's try to buddy you up. Do not worry the presentation will still walk you through the content.
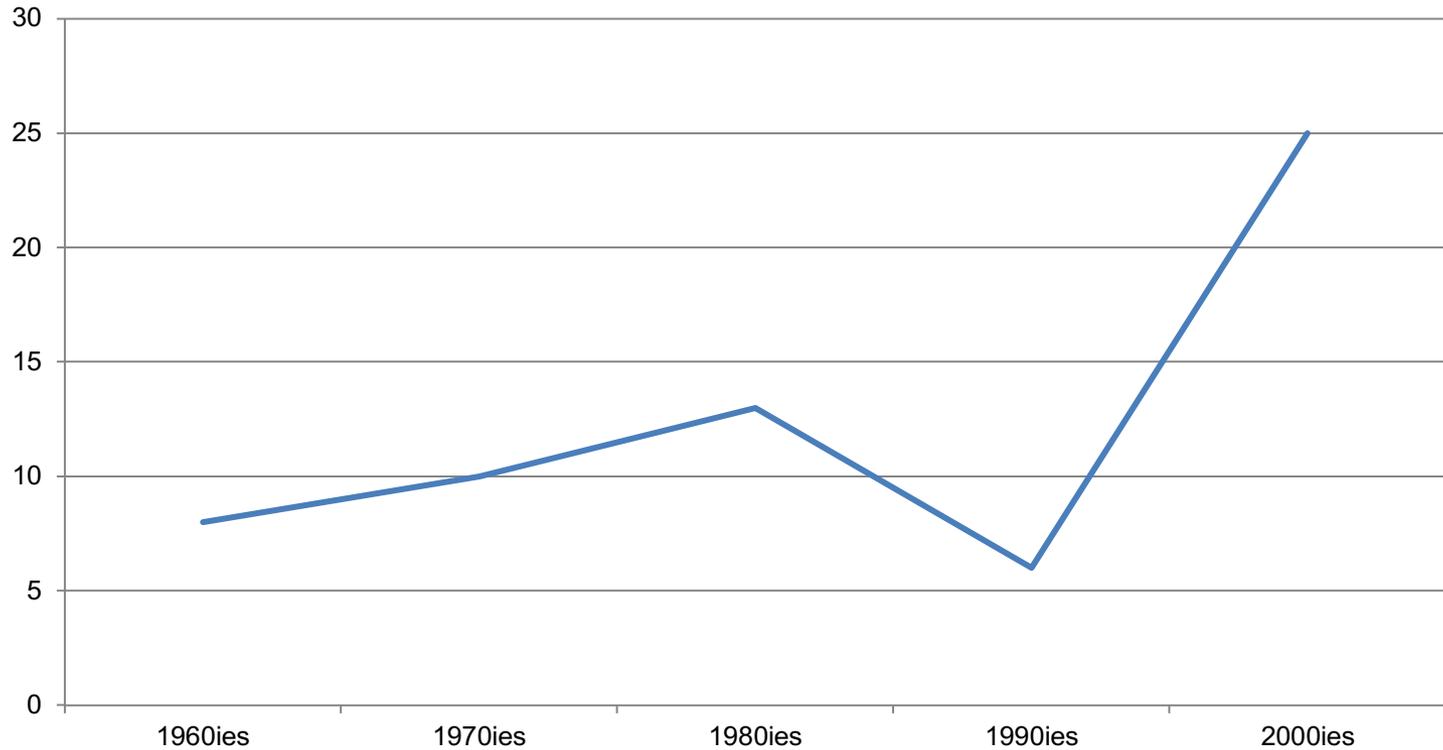
FDA Regulatory Action over Time

Number of FDA-caused Withdrawals

FDAAA calls for establishing Risk Identification and Analysis System

Observational Medical Outcomes Partnership

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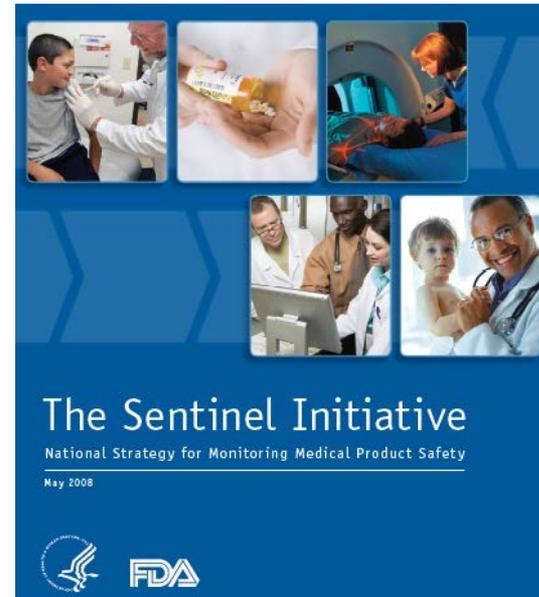
"(ii) develop validated methods for the establish-ment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

"(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.
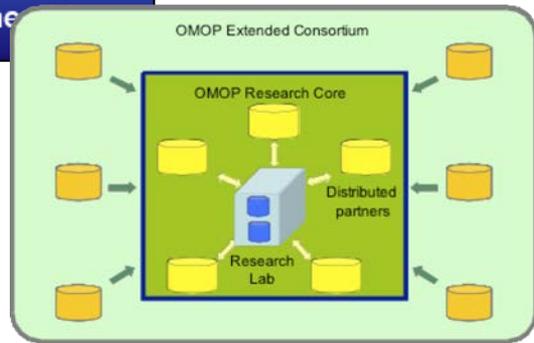
"(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTI-FICATION AND ANALYSIS SYSTEM.—

The Sentinel Initiative
National Strategy for Monitoring Medical Product Safety
May 2008
FDA
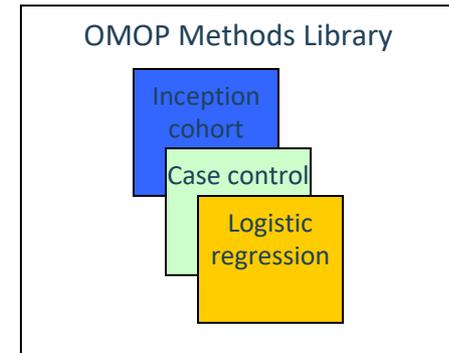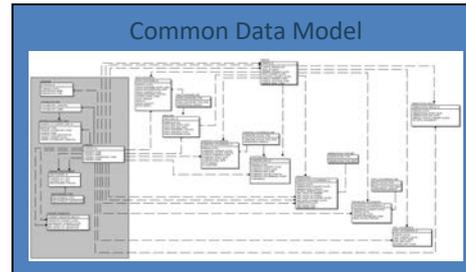
Risk Identification and Analysis System:

a systematic and reproducible process to efficiently generate evidence to support the characterization of the potential effects of medical products from across a network of disparate observational healthcare data sources

OMOP Experiment 1 (2009-2010)

Observational Medical Outcomes Partnership

OMOP Extended Consortium
OMOP Research Core
Distributed partners
Research Lab

- 10 data sources
- Claims and EHRs
- 200M+ lives

- Open-source
- Standards-based

Common Data Model

OMOP Methods Library
Inception cohort
Case control
Logistic regression

- 14 methods
- Epidemiology designs
- Statistical approaches adapted for longitudinal data

Drug

Outcome	ACE Inhibitors	Amphotericin B	Antibiotics: erythromycins, sulfonamides, tetracyclines	Antiepileptics: carbamazepine, phenytoin	Benzodiazepines	Beta blockers	Bisphosphonates: alendronate	Tricyclic antidepressants	Typical antipsychotics	Warfarin
Angioedema	red	blue		blue	blue	blue				blue
Aplastic Anemia	blue	blue	blue	red	blue	blue	blue	blue		blue
Acute Liver Injury		blue	red		blue	blue	blue	blue		
Bleeding			blue		blue		blue	blue		red
Hip Fracture	blue	blue		red	blue			blue		blue
Hospitalization	green									
Myocardial Infarction			blue		blue		blue	red	red	
Mortality after MI		blue		blue	blue	green				blue
Renal Failure		red	blue	blue	blue		blue	blue		blue
GI Ulcer Hospitalization	blue		blue		blue		red		blue	

OMOP Experiment 2 (2011-2012)

Observational
Medical
Outcomes
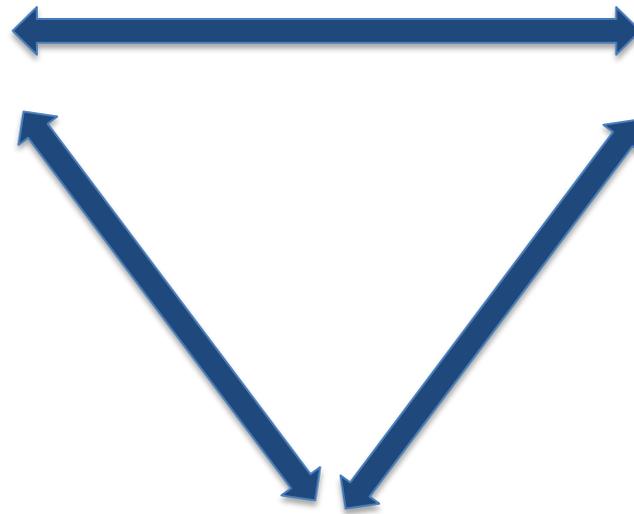**Partnership**

Observational data

4 claims databases

1 ambulatory EMR

Methods
- Case-Control
- New User Cohort
- Disproportionality methods
- ICTPD
- LGPS
- Self-Controlled Cohort
- SCCS

Drug-outcome pairs

	Positives	Negatives
Total	165	234
Myocardial Infarction	36	66
Upper GI Bleed	24	67
Acute Liver Injury	81	37
Acute Renal Failure	24	64

European OMOP Experiment

Observational data

Aarhus   Pedianet

ARS   IPCI

HS   PHARMO

eu-adr

Methods

- Case-Control
- New User Cohort
- Disproportionality methods
- ICTPD
- LGPS
- Self-Controlled Cohort
- SCCS

Drug-outcome pairs

	Positives	Negatives
Total	165	234
Myocardial Infarction	36	66
Upper GI Bleed	24	67
Acute Liver Injury	81	37
Acute Renal Failure	24	64

Observational Medical Outcomes Partnership
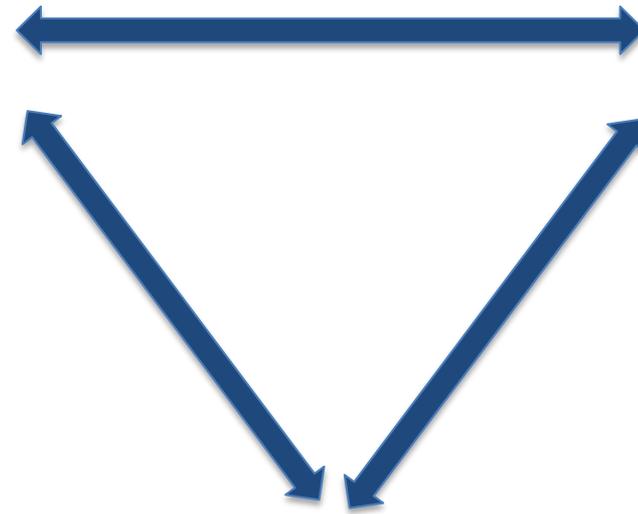
Ground Truth for OMOP Experiment

Observational Medical Outcomes Partnership

isoniazid

fluticasone

indomethacin

clindamycin

	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

ibuprofen

loratadine

sertraline

pioglitazone

Criteria for positive controls:
- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label
- Drug listed as 'causative agent' in Tisdale et al, 2010: Drug-Induced Diseases
- Literature review identified no powered studies with refuting evidence of effect

Criteria for negative controls:
- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as 'causative agent' in Tisdale et al, 2010: Drug-Induced Diseases
- Literature review identified no powered studies with evidence of potential positive association

13

Data Used in European Experiment

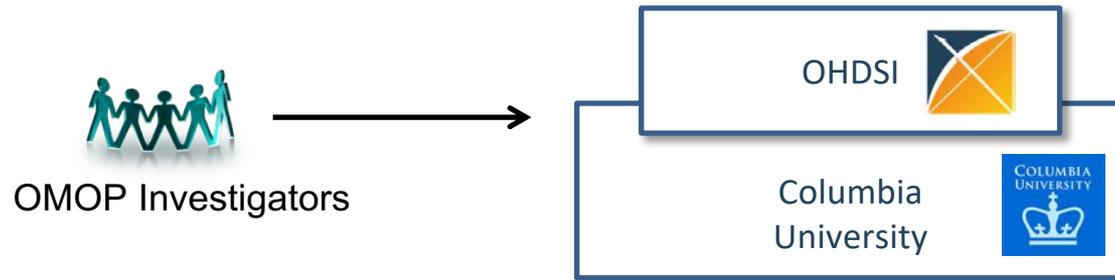| Name	Description	Population
Aarhus	Danish national health registry, covering the Aarhus region. Includes inhabitant registry, drug dispensations, hospital claims, lab values, and death registry.	2 M
ARS	Italian record linkage system covering the Tuscany region, including inhabitant registry, drug dispensations, hospital claims, and death registry	4 M
Health-Search	Italian general practice database (no children)	1 M
IPCI	Dutch general practice database	0.75 M
Pedianet	Italian general practice pediatric database	0.14 M
PHARMO	Dutch record linkage system. Includes inhabitant registry, drug dispensations, hospital claims, and lab values.	1.28 M

Results

Main findings in OMOP experiment

- Heterogeneity in estimates due to choice of database

- Heterogeneity in estimates due to analysis choices

- Except little heterogeneity due to outcome definitions

- Good performance (AUC > 0.7) in distinguishing positive from negative controls for optimal methods when stratifying by outcome and restricting to powered test cases

- Self controlled methods perform best for all outcomes

Fate of OMOP - OHDSI

OMOP Investigators → OHDSI / Columbia University

- The Observational Health Data Sciences and Informatics (OHDSI) program is a **multi-stakeholder, interdisciplinary collaborative** to create **open-source** solutions that bring out the value of observational health data through large-scale analytics

- OHDSI has established **an international network of researchers and observational health databases** with a central coordinating center housed at Columbia University
  - Public, Open
  - Not Pharma-funded
  - International

http://ohdsi.org

OHDSI's vision

OHDSI collaborators access a network of 1 billion patients to generate evidence about all aspects of healthcare. Patients and clinicians and other decision-makers around the world use OHDSI tools and evidence every day.

Join us on the journey

http://ohdsi.org

OHDSI: a global community

OHDSI Collaborators:
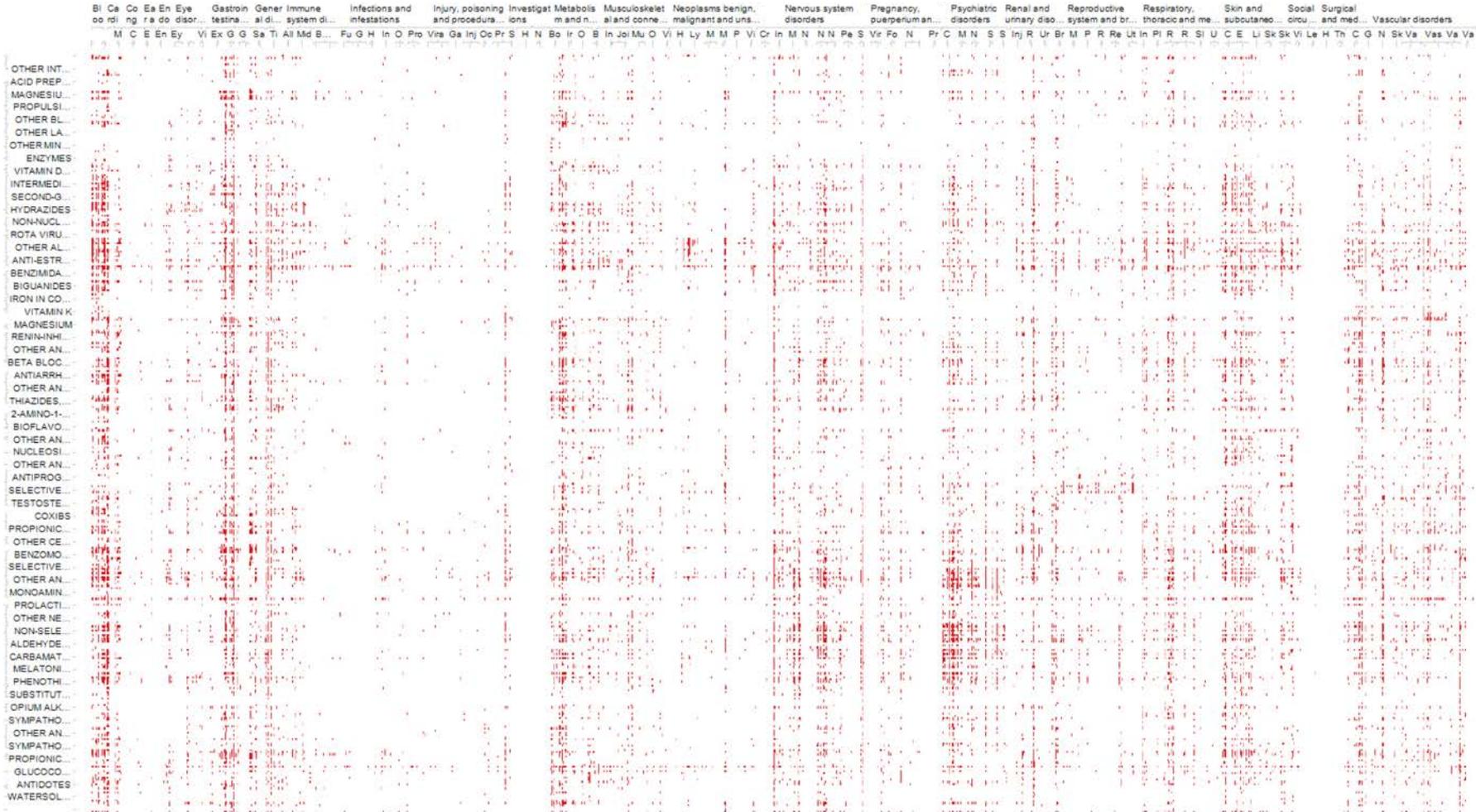- >200 researchers in academia, industry and government
- >17 countries

OHDSI Data Network:
- >82 databases from 17 countries
- 1.2 billion patients records (duplicates)
- ~115 million non-US patients

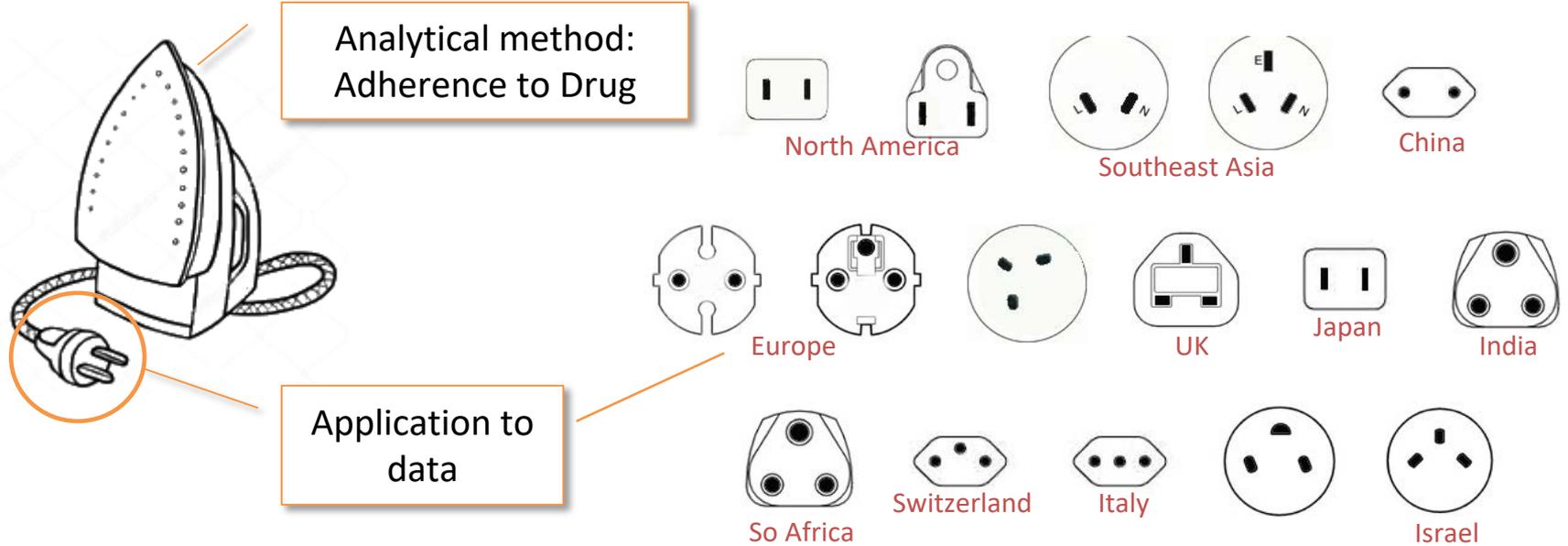# Current pace of evidence generation in healthcare

All health outcomes of interest

Current Approach: "One Study – One Script"

"What's the adherence to my drug in the data assets I own?"

Analytical method:
Adherence to Drug

Application to data

North America

Southeast Asia

China

Europe

UK

Japan

India

So Africa

Switzerland

Italy

Israel

Current solution:

One SAS or R script for each study
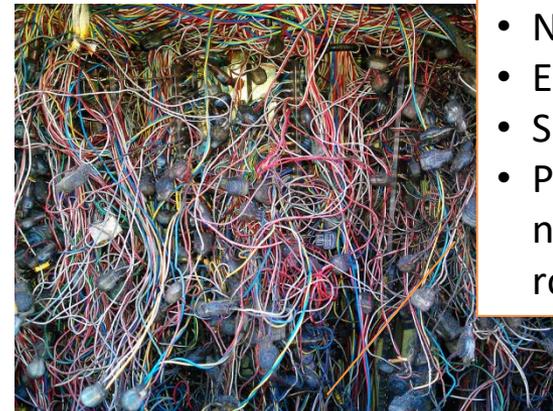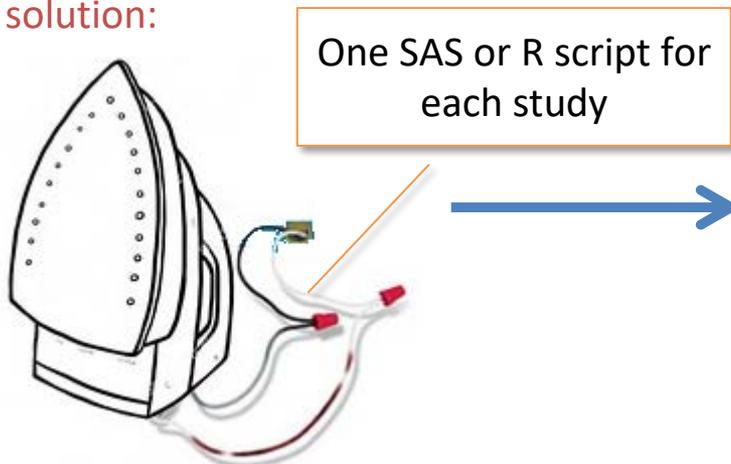
- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use

Solution: Data Standardization Enables Systematic Research

Adherence

Mortality

Source of Business

Safety Signals

North America   Southeast Asia   China

Europe   UK   Japan   India

So Africa   Switzerland   Italy   Israel

Standardized data
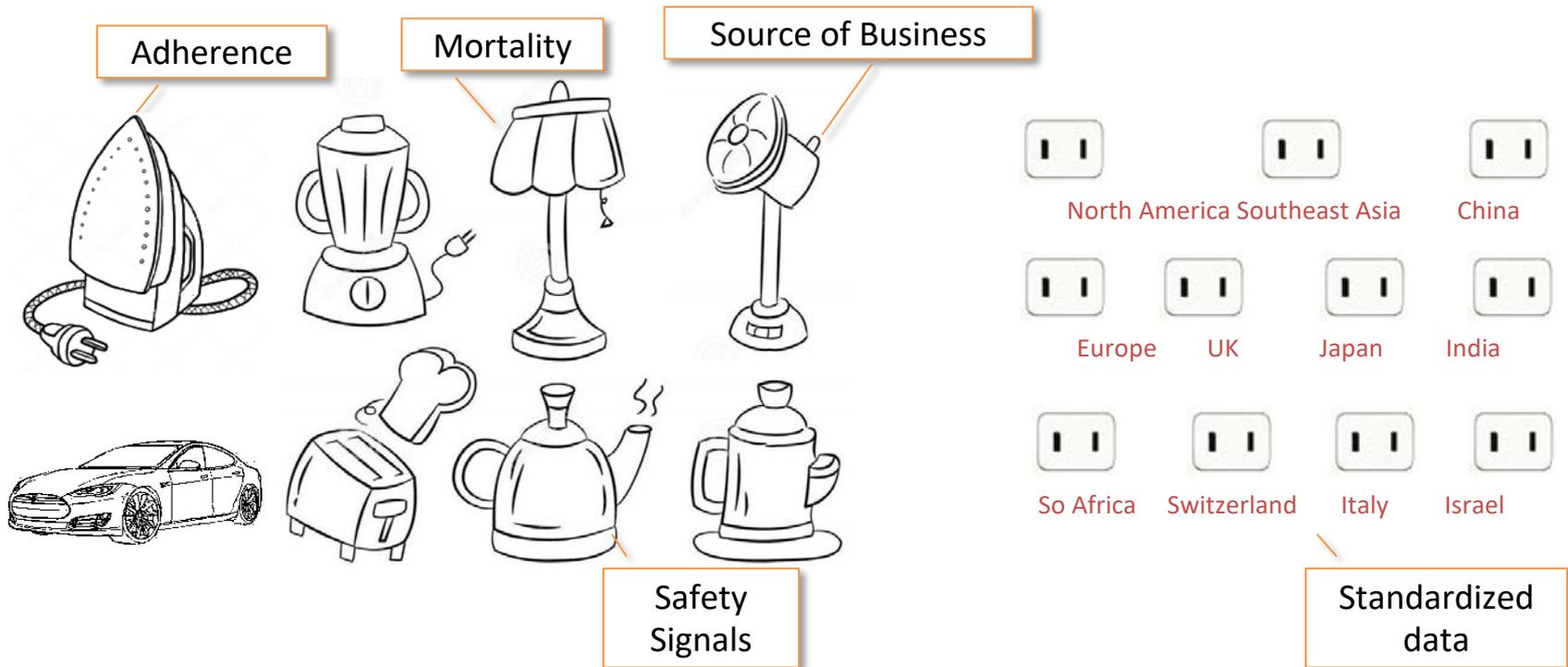
OHDSI Tools

OMOP CDM

Analytics can be remote
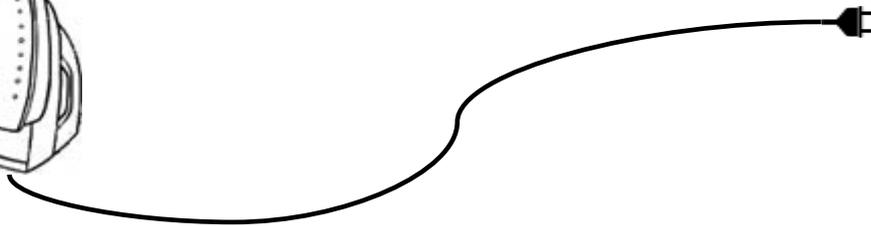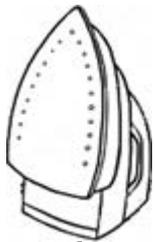
Analytics can be behind firewall
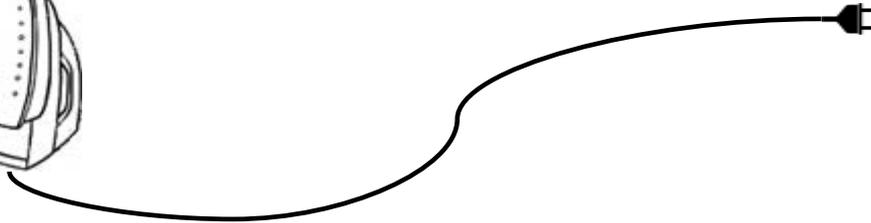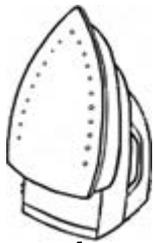
North America   Southeast Asia   China

Europe   UK   Japan   India

So Africa   Switzerland   Italy   Israel

Network Studies Networks of networks

ISDN

EMR — University Medical Center

EMR — Inpatient Hospital

EMR — Outpatient Hospital

Another Network

Claims Asset

Claims Asset   EMR Asset

Coordinating Center

Network

Coordinating Center

Claims Asset   Claims Asset   Claims Asset

EMR Asset   EMR Asset   EMR Asset   EMR Asset
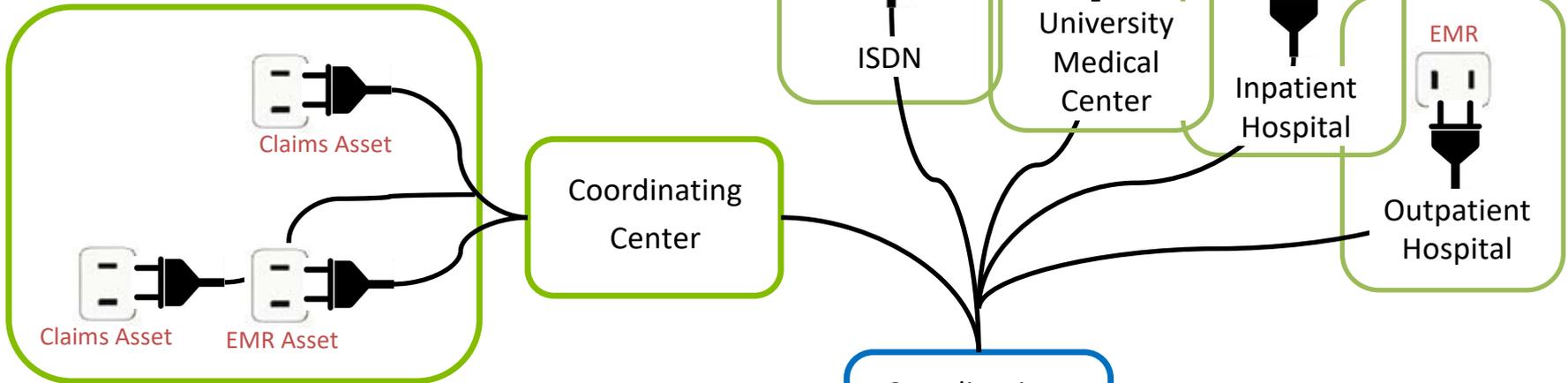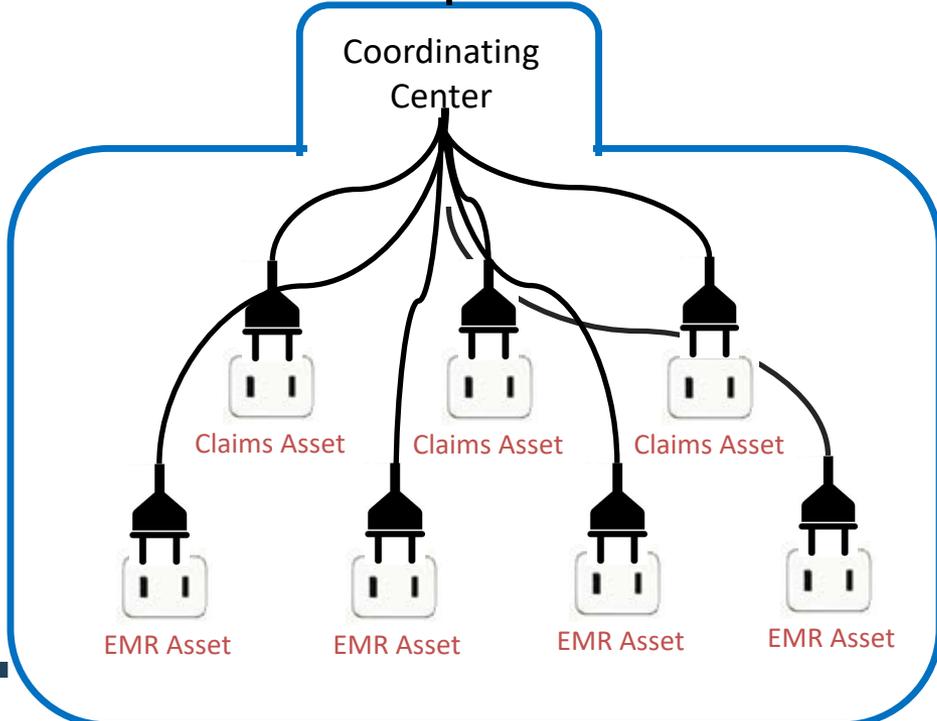
CDM Version 5 Key Domains
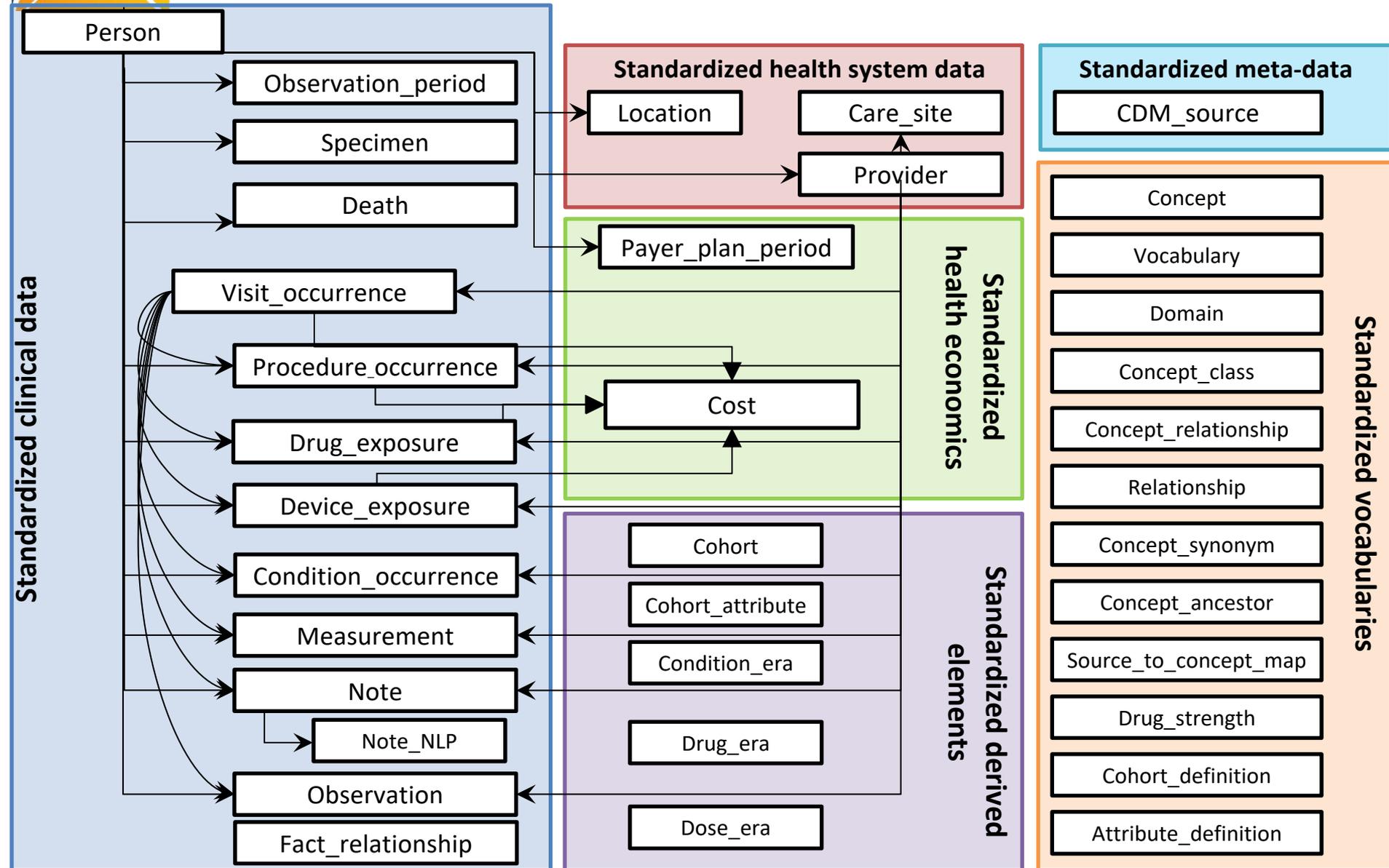
Standard Variable Name Conventions

Field name	Purpose	Example
<entity>**_concept_id**	Foreign key into the Standard Vocabulary for **Standard Concept**	condition_concept_id 313217 (SNOMED "Atrial Fibrillation")
<entity>**_source_concept_id**	Foreign key into the Standard Vocabulary for **Source Concept**	condition_source_concept_id 44821957 (ICD9CM "Atrial Fibrillation")
<entity>**_source_value**	Verbatim information from the source data, **not to be used** by any standard analytics	condition_source_value 427.31 (ICD9CM "Atrial Fibrillation")
<entity>**_type_concept_id**	Foreign key into the Vocabulary for the **origin of the information**	condition_type_concept_id 38000199 ("Inpatient header – primary")
<entity>**_id**	Unique identifiers for **entities** (row numbers, or IDs imported from source)	person_id 1234567 visit_occurrence_id 7654321 could be a person identifier or an autogenerated number by the CDM builder

Single Concept Reference Table

All vocabularies stacked up in one table

Vocabulary ID

SNOMED
ICD9CM
ICD9Proc
CPT4
HCPCS
LOINC
NDFRT
RxNorm
NDC
GPI
UCUM
Race
Place of Service
MedDRA
Multum
Read
OXMIS
Indication
ETC
ATC
Multilex
VA Product
SMQ
VA Class
Cohort
ICD10
ICD10PCS
Drug Type
Condition Type
Procedure Type
Observation Type
DRG
MDC
APC
Revenue Code
Death Type
MeSH
NUCC
Specialty
SPL
GCN_SEQNO
OPCS4
Gemscript
HES Specialty
Domain
PCORNet
Currency
Relationship
Vocabulary
Concept Class
ICD10CM
ABMS
CIEL
DA_France
DPD

What's in a Concept

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	

For use in CDM

English description

Domain

Vocabulary

Class in SNOMED

Concept in data

Code in SNOMED

Valid during time interval: always

OMOP CDM Standard Domain Features

Feature	Description and purpose	Field name convention	Example
Patient centric	Every domain table has **patient identifier**. Patient data can be retrieved independently from other domains.	person_id	person_id 123
Unique domain identifier	Every domain table has a unique primary key to identify domain **entities**	<entity>_id	condition_occurrence_id 470985
Standard concept from a respective vocabulary domain	Integration with the vocabulary. Foreign key into the Standard Vocabulary for **Standard Concept**	<entity>_concept_id	condition_concept_id 313217 (SNOMED "Atrial Fibrillation")
Source concept from a respective vocabulary domain	Provenance. Foreign key into the Standard Vocabulary for **Source Concept**	<entity>_source_concept_id	condition_source_concept_id 44821957 (ICD9CM "Atrial Fibrillation")
Source value	Provenance. Verbatim information from the source data, **not to be used** by any standard analytics	<entity>_source_value	condition_source_value 427.31 (ICD9CM "Atrial Fibrillation")
Source type	Provenance. Foreign key into the Vocabulary for the **origin of the**	<entity>_type_concept_id	condition_type_concept_id 38000199 ("Inpatient header – primary")

Integration of CDM and Vocabulary

CONCEPT
concept_id:	44821957
concept_name:	'Atrial fibrillation'
vocabulary_id:	'ICD9CM'
concept_code:	'427.31'
primary_domain:	condition
standard_concept:	N

CONCEPT
concept_id:	312327
concept_name:	'Atrial fibrillation'
vocabulary_id:	'SNOMED'
concept_code:	49436004
primary_domain:	condition
standard_concept:	Y

CONDITION_OCCURRENCE
person_id:	123
condition_concept_id:	312327
condition_start_date:	14Feb2013
condition_source_value:	'427.31'
condition_source_concept_id:	44821957

Remote Desktop

Logging into Erasmus PCs

- Username: `medinformatica`
- Password: `2013ebm`

OHDSI in a Box
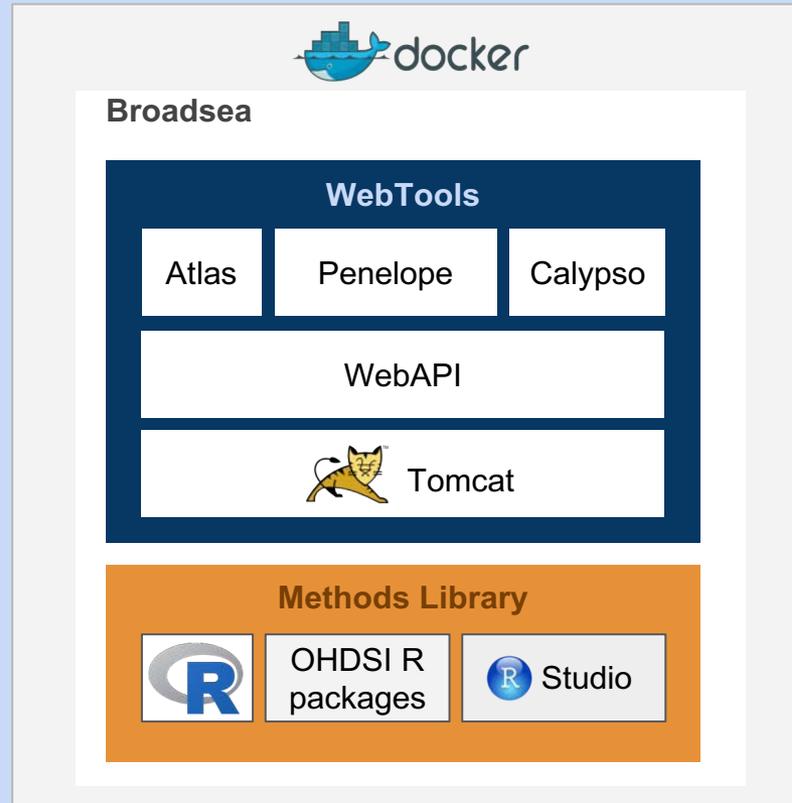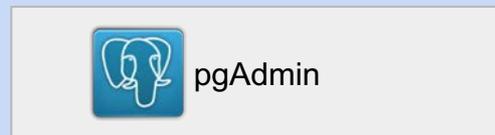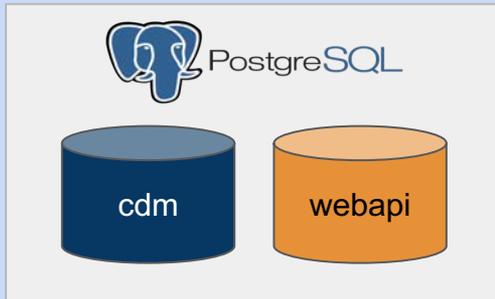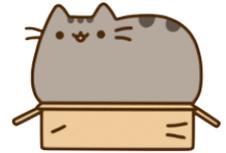
How to Sign into the Remote Desktop

From your command prompt, type %systemroot%/system32/**mstsc**.exe to launch Remote Desktop

How to Sign into the Remote Desktop

- Use the shortcut on the desktop named "Remote Desktop"

goo.gl/aXKY9e

- Pick one of the rows and put your name on the second column

How to Sign into the Remote Desktop

- Take Column A from spreadsheet and copy into the "Computer" field
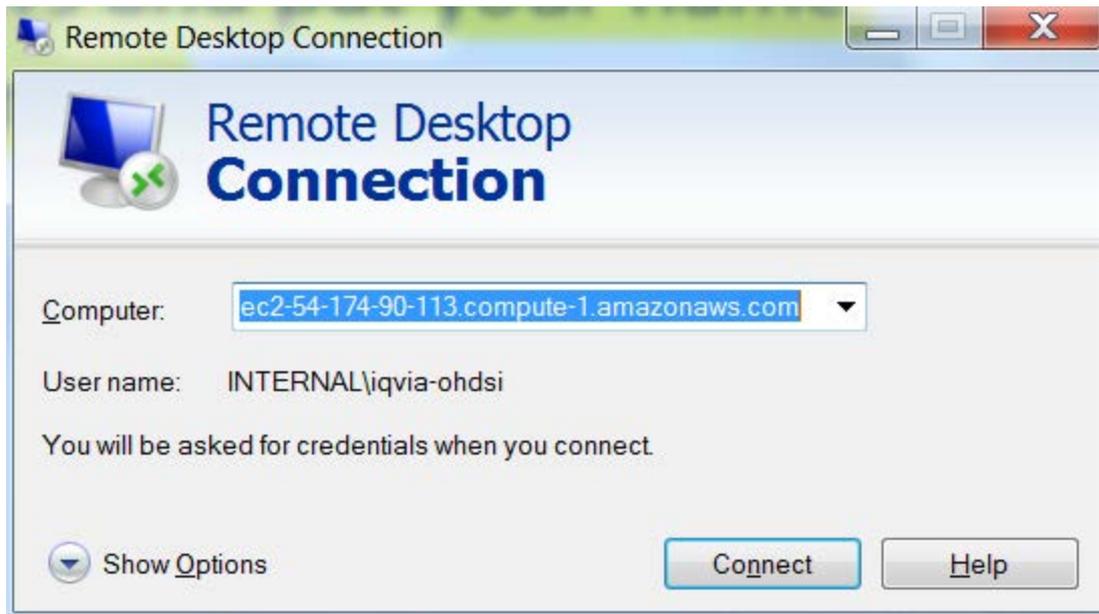
How to Sign into the Remote Desktop

- Pick 'Use Another Account'

- Copy username from Column C

- Copy password from Column D

A	B	C	D
RDP URL	**Name**	**Username**	**Password**
ec2-34-226-245-112.compute-1.amazonaws.com	Erica Voss	iqvia-ohdsi	I!QViAOH@DSI18
ec2-52-87-207-197.compute-1.amazonaws.com	Mui Van Zandt	iqvia-ohdsi	I!QViAOH@DSI18

How to Sign into the Remote Desktop

- If you get this page, select "Yes"

OHDSI in a Box – Ready

CDM Database:
pgAdmin III New Server

- Click on PGAdmin

CDM Database: Connect

- Password: `ohdsi`

CDM Database: Open SQL Sheet

CDM Database:
Ready

```
set search_path to 'public', 'ohdsi';
```

Vocabulary

Basic Relationship, Ancestors, & Descendants
How does it work for Drugs
SQL Examples

OMOP Common Vocabulary Model

What it is

- **Standardized structure** to house existing vocabularies used in the public domain
- **Compiled standards** from disparate public and private sources and some OMOP-grown concepts

What it's not

- **Static dataset** – the vocabulary updates regularly to keep up with the continual evolution of the sources
- **Finished product** – vocabulary maintenance and improvement is ongoing activity that requires community participation and support

OMOP Vocabulary CDM
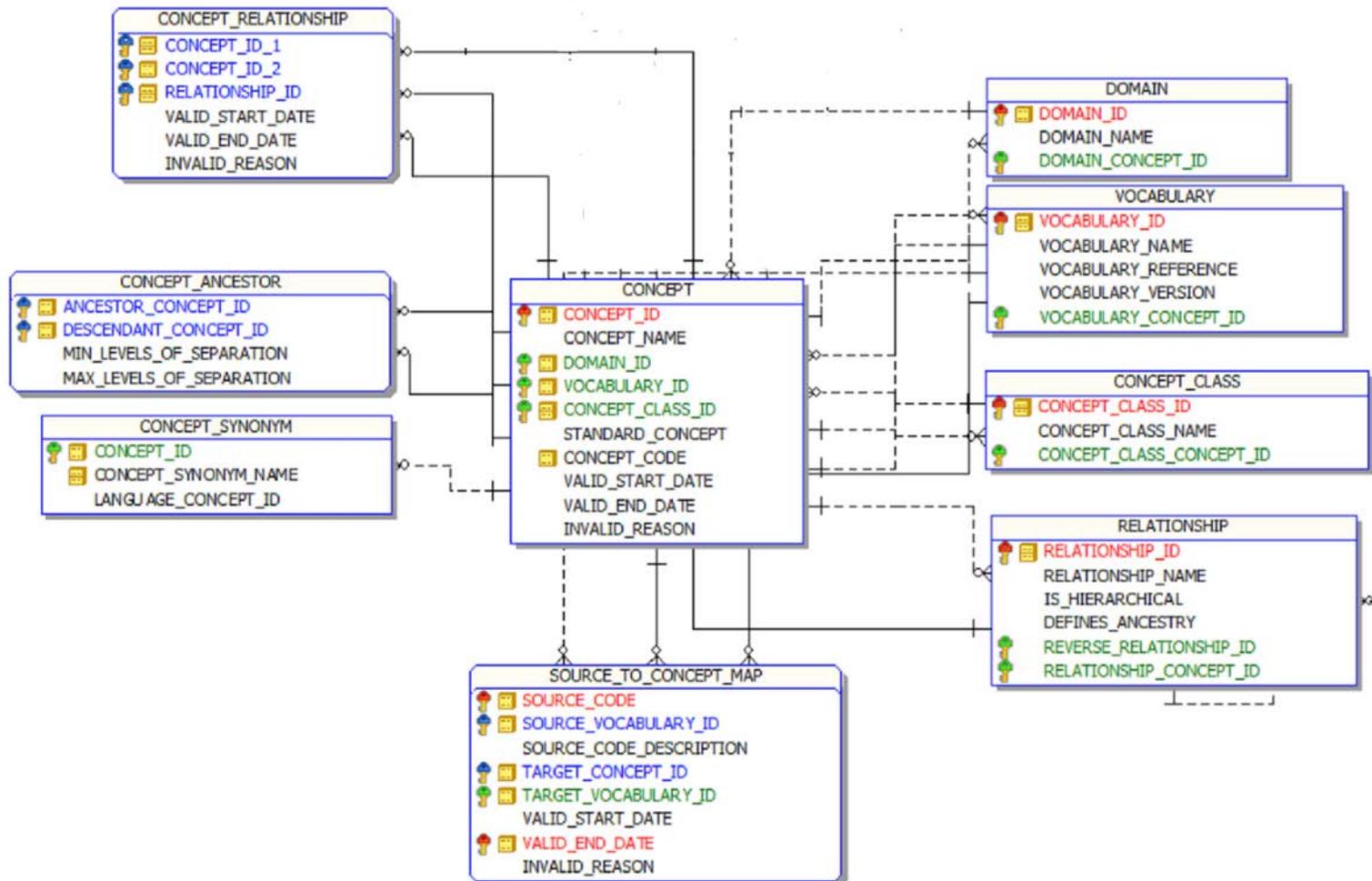
1. All content: concepts in **concept** table
2. Direct relationships between concepts listed in **concept_relationship**
3. Multi-step hierarchical relationships pre-processed in **concept_ancestor**

MiniSentinel in use: Dabigatran and bleeding

Dabigatran and Postmarketing Reports of Bleeding

Questions we are regularly asked:

1) What does it take to do an analysis like this?

2) How can this be done against the OMOP CDM?

	Patients	Events	100,000 days at risk)	Patients	Events	dence events/ 100,000 days at risk)
Gastrointestinal hemorrhage						
Analysis with required diagnosis of atrial fibrillation	10,599	16	1.6	43,541	160	3.5
Sensitivity analysis without required diagnosis of atrial fibrillation	12,195	19	1.6	119,940	338	3.1
Intracranial hemorrhage						
Analysis with required diagnosis of atrial fibrillation	10,587	8	0.8	43,594	109	2.4
Sensitivity analysis without required diagnosis of atrial fibrillation	12,182	10	0.9	120,020	204	1.9

All Content in CDM is Coded as Concepts

- Concepts are referred to by concept_id
- All details are in the **CONCEPT** table:

```sql
SELECT *

FROM concept

WHERE concept_id = 313217
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL

Dozens of schemes, formats, rules

LOINC_248_MULTI-AXIAL_HIERARCHY.CSV

PATH_TO_ROOT			SEQUENCI	IMMEDIATE_PARENT	CODE	CODE_TEXT
			1		LP31755-9	Microbiology
LP317			1	LP31755-9	LP14559-6	Microorganism
LP317	loinc.csv		1	LP14559-6	LP98185-9	Bacteria
LP317			1	LP98185-9	LP14082-9	Bacteria

LOINC_NL	COMPONENT	PROPERTY	TIME_ASPCT	SYSTEM	SCALE_TYP	METHOD_TYP	CLASS	SOURCE	DATE_LAST_CH	CHNG_TYI	COMMEN	STATUS
10454-7	Xylose^2H post 25 g xylose PO	MCnc	Pt	Ser/Plas	Qn		CHAL	SH	19961220	ADD		ACTIVE
10455-4	Xylose^30M post 25 g xylose PO	MCnc	Pt	Ser/Plas	Qn		CHAL	SH	19961220	ADD		ACTIVE
10456-2	Xylose^post 6H CFst	MCnc	Pt	Ser/Plas	Qn		CHAL	SH	19961220	ADD		ACTIVE
10457-0	Actin Ag	ACnc	Pt	Tiss	Ord	Immune stain	PATH	SH;DL-M	20060706	MIN		ACTIVE
10458-8					Ord	Immune stain	PATH	DL-M	20060706	MIN		ACTIVE
10459-6					Ord	Immune stain	PATH	DL-M	20060706	MIN		ACTIVE
10460-4									20060706	MIN		ACTIVE
10461-2									20060706	MIN		ACTIVE
1046-2									20130529	MAJ		ACTIVE
10462-0									20060706	MIN		ACTIVE
10463-8									20060706	MIN		ACTIVE
10464-6									20060706	MIN		ACTIVE
10465-3									20060706	MIN		ACTIVE
10466-1									19961220	ADD	[(NA+)]-[(	ACTIVE

CMS32_DESC_LONG_SHORT_DX.xlsx

DIAGNOSIS CODE	LONG DESCRIPTION	SHORT DESCRIPTION
0010	Cholera due to vibrio cholerae	Cholera d/t vib cholerae
0011	Cholera due to vibrio cholerae el tor	Cholera d/t vib el tor
0019	Cholera, unspecified	Cholera NOS
0020	Typhoid fever	Typhoid fever
0021	Paratyphoid fever A	Paratyphoid fever a
0022	Paratyphoid fever B	Paratyphoid fever b
0023	Paratyphoid fever C	Paratyphoid fever c
0029	Paratyphoid fever, unspecified	Paratyphoid fever NOS
0030	Salmonella gastroenteritis	Salmonella enteritis
0031	Salmonella septicemia	Salmonella septicemia
00320	Localized salmonella infection, unspecified	Local salmonella inf NOS
00321	Salmonella meningitis	Salmonella meningitis
00322	Salmonella pneumonia	Salmonella pneumonia
00323	Salmonella arthritis	Salmonella arthritis
00324	Salmonella osteomyelitis	Salmonella osteomyelitis
00329	Other localized salmonella infections	Local salmonella inf NEC

Condition Concepts
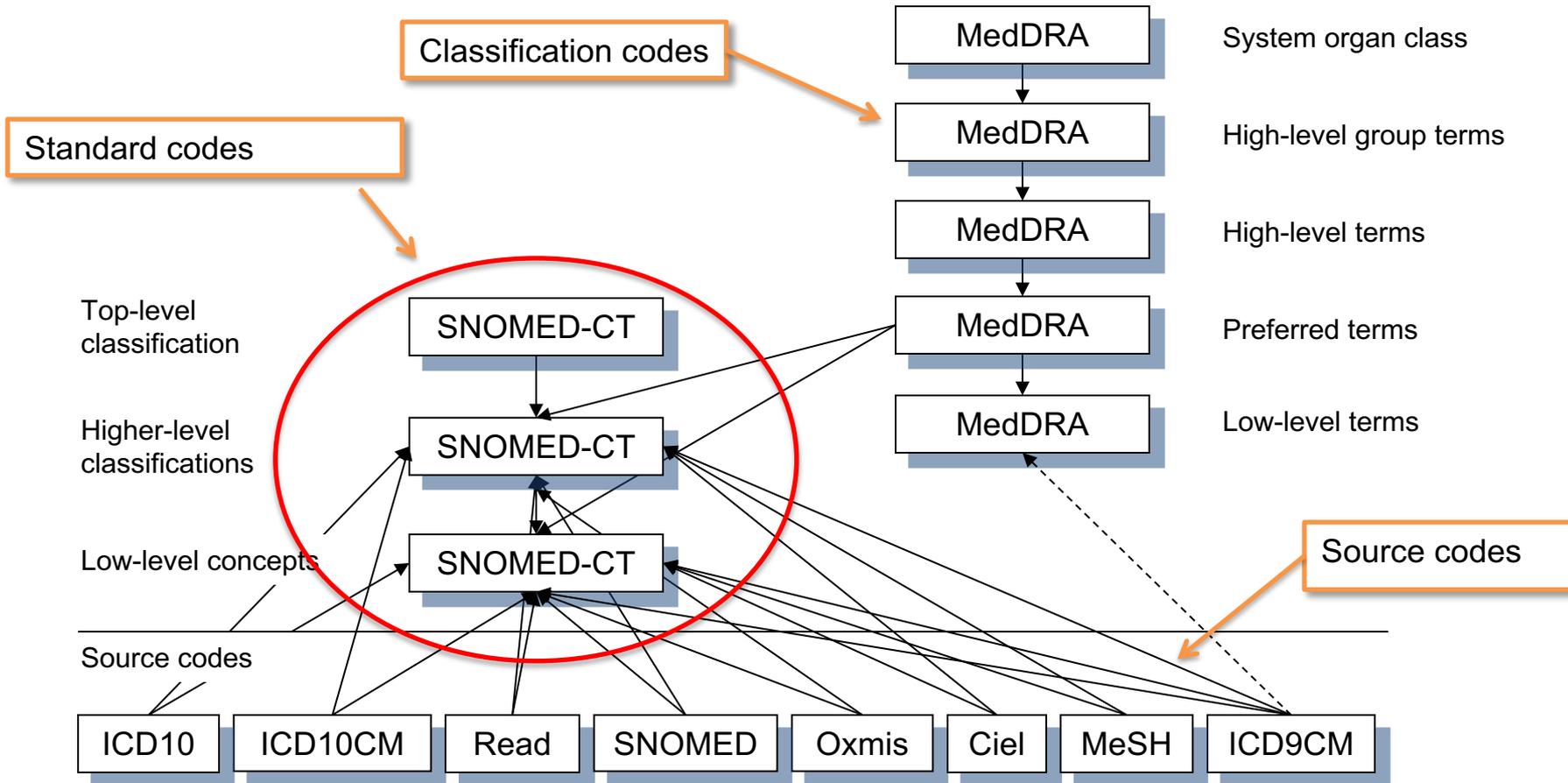
Finding the Right Concept: #1

1. ..if I know the **ID**

```sql
SELECT * FROM concept WHERE concept_id = 313217;
```

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULA RY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	01-Jan-1970	31-Dec-2099	

> SNOMED code

2. ..if I know the **code**

```sql
SELECT * FROM concept WHERE concept_code = '49436004';
```

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULA RY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	01-Jan-1970	31-Dec-2099	

Concept code 49436004 in SNOMED Browser

Concept ID versus Concept Code

```
SELECT *
FROM concept
WHERE concept_code = '1001';
```

Same code
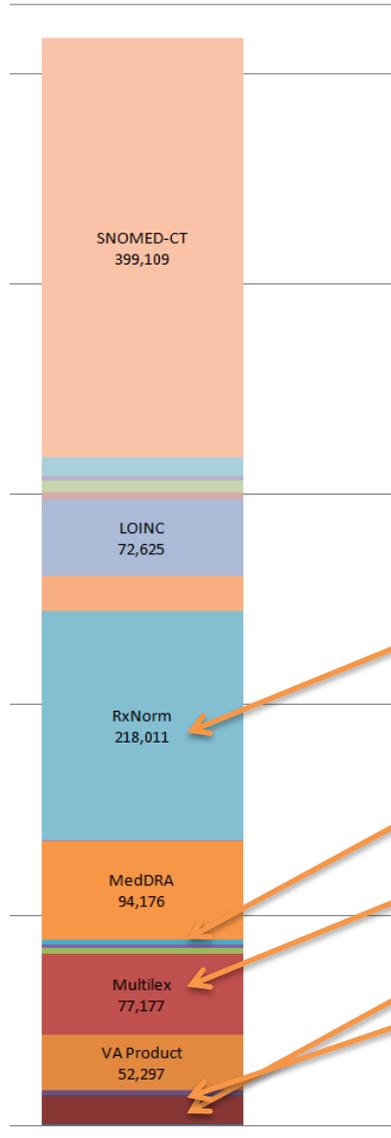
Concept_Name	Concept Class	Vocabulary_ID	Concept_Code
Antipyrine	Ingredient	RxNorm	1001
Aceprometazine maleate	Ingredient	BDPM	1001
Serum	Specimen	CIEL	1001
methixene hydrochloride	Ingredient	Multilex	1001
Brompheniramine Maleate, 10 mg/mL injectable solution	Multum	Multum	1001
ABBOTT COLD SORE BALM 4%/0.06% W/	Drug Product	LPD_Australia	1001
Residential Treatment - Psychiatric	Revenue Code	Revenue Code	1001

SNOMED-CT
399,109

LOINC
72,625

RxNorm
218,011

MedDRA
94,176

Multilex
77,177

VA Product
52,297

Finding the Right Concept: #2

3. ..if I know the **name**

```
SELECT * FROM concept WHERE concept_name = 'Atrial fibrillation';
```

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31
35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658
45500085	Atrial fibrillation	Condition	Read	Read		G573000
45883018	Atrial fibrillation	Meas Value	LOINC	Answer	S	LA17084-7

Finding the Right Concept: #3

1. if don't know any of this, but I know the code in another vocabulary

ICD-9 is not a Standard Concept

`SELECT * FROM concept WHERE concept_code = '427.31';`

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31

`SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957;`

Mapping to different vocabularies

Kind of relationship

CONCEPT _ID_1	CONCEPT ID_2	RELATIONSHIP _ID	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
44821957	21001551	ICD9CM - FDB Ind	01-Oct-13	31-Dec-2099	
44821957	35204953	ICD9CM - MedDRA	01-Jan-70	31-Dec-2099	
44821957	44824248	Is a	01-Oct-14	31-Dec-2099	
44821957	44834731	Is a	01-Oct-14	31-Dec-2099	
44821957	313217	Maps to	01-Jan-70	31-Dec-2099	

Why are we mapping?

LANGUAGES
Supporting language learning and linguistic diversity

European Commission > Languages > Policy > Linguistic diversity

Official languages of the EU

What is it?

The European Union has 24 official and working languages. They are:

Bulgarian	French	Maltese
Croatian	German	Polish
Czech	Greek	Portuguese
Danish	Hungarian	Romanian
Dutch	Irish	Slovak
English	Italian	Slovenian
Estonian	Latvian	Spanish
Finnish	Lithuanian	Swedish

What is the Commission doing?

With a permanent staff of 1,750 linguists and 600 support staff, the Commission has one of the largest translation services in the world, bolstered by a further 600 full-time and 3,000 freelance interpreters.

How many different ways do you express one meaning?

Gëzuar

Наздраве

Salut

Živjeli

Na zdravi

Skål

Proost

Terviseks

Skál

Santé

Salud

На здравје

Kippis

Υγεια

Zum Wohl

Fenékig

Sláinte

Noroc

Salute

į sveikatą

Saúde

Priekā

Na zdrowie

На здоровье

Cheers

Mapping = Translating

Step 1. Find the Source Concept

`SELECT * FROM concept WHERE concept_code = '427.31';`

CONCEPT_ID	CONCEPT_ NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31

Step 2. Translate

`SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957 AND relationship_id = 'Maps to';`

CONCEPT_ID_1	CONCEPT_ID_2	RELATIONSHIP _ID	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
44821957	313217	Maps to	01-Jan-1970	31-Dec-2099	

Step 3. Check out the translated Concept

`SELECT * FROM concept WHERE concept_id = 313217;`

```
ICD-9: '427.31' : 313217
Read: 'G573000' : 313217
ICD-10: 'I48.0' : 4154290 'Paroxysmal Atrial Fibrillation'
```

Concept

Source codes

| ICD-9-CM 427.31 Atrial Fibrillation | Read G573000 Atrial Fibrillation | ICD-10-CM I48.0 Atrial Fibrillation |
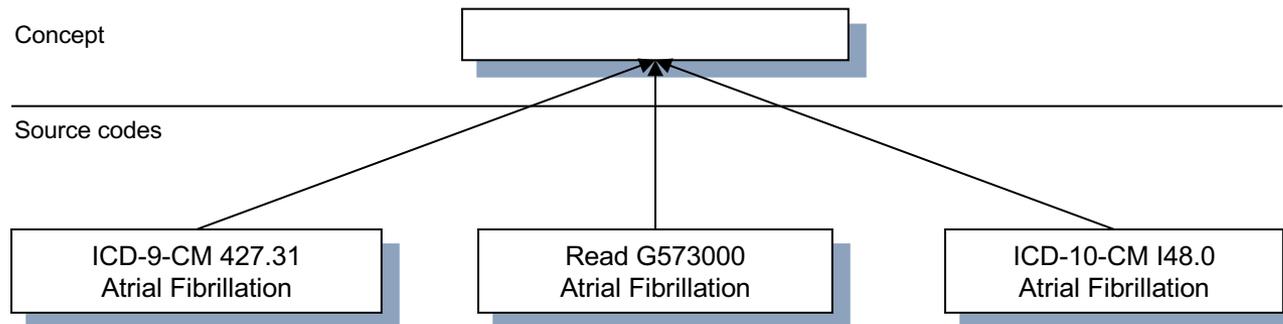
Break

Please return in 15 minutes

Reason #2: Disease Hierarchy
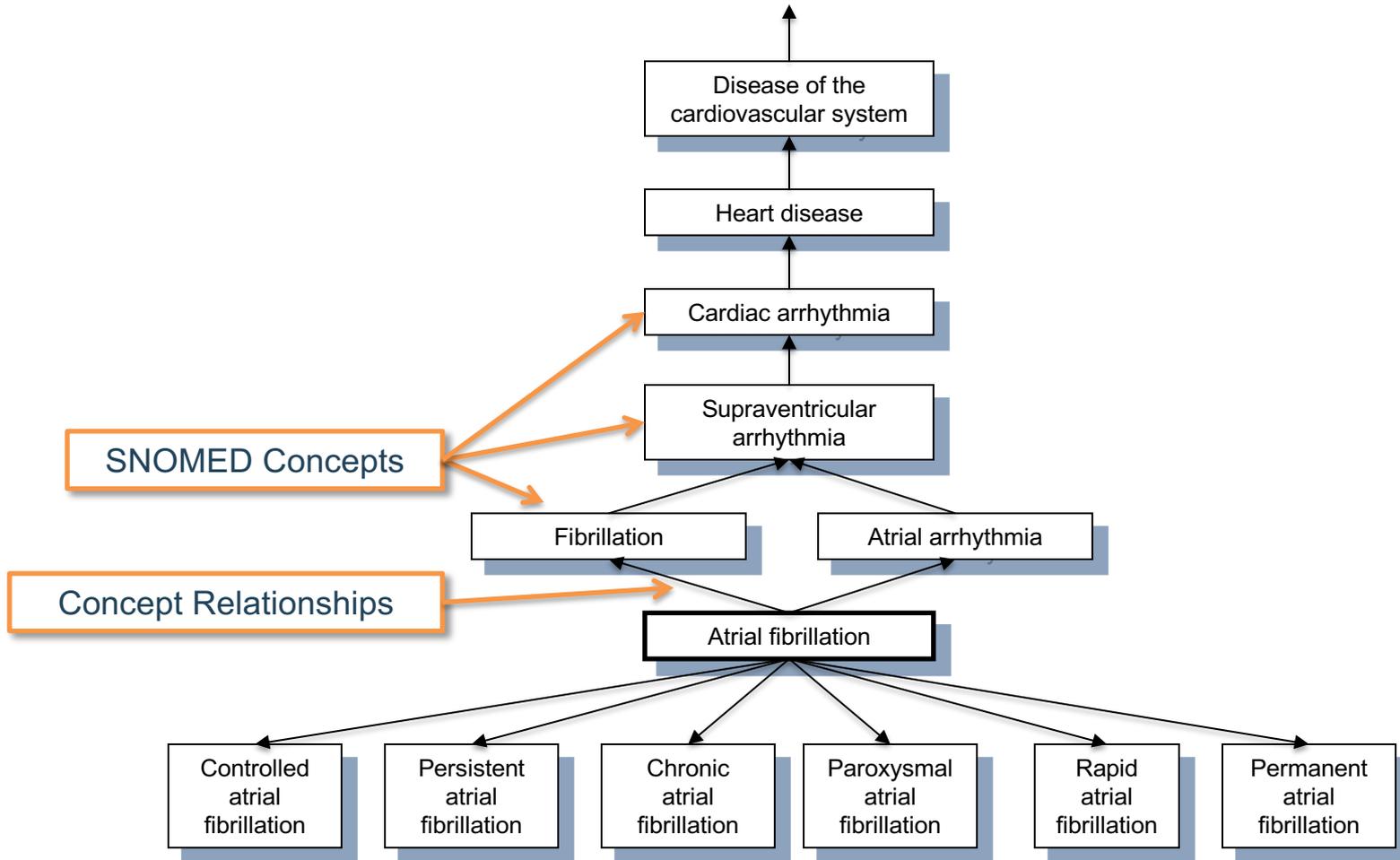
Exploring Relationships

```sql
SELECT
  *
FROM
  concept_relationship
WHERE
  concept_id_1 = 313217;
```

Related Concepts

Relationship ID

CONCEPT _ID_1	CONCEPT _ID_2	RELATIONSHIP _ID
313217	4232697	Subsumes
313217	4181800	Focus of
313217	35204953	SNOMED - MedDRA eq
313217	4203375	Asso finding of
313217	4141360	Subsumes
313217	4119601	Subsumes
313217	4117112	Subsumes
313217	4232691	Subsumes
313217	4139517	Due to of
313217	4194288	Asso finding of
313217	44782442	Subsumes
313217	44783731	Focus of
313217	21003018	SNOMED - ind/CI
313217	40248987	SNOMED - ind/CI
313217	21001551	SNOMED - ind/CI
313217	21001540	SNOMED - ind/CI
313217	45576876	Mapped from
313217	44807374	Asso finding of
313217	21013834	SNOMED - ind/CI
313217	21001572	SNOMED - ind/CI
313217	21001606	SNOMED - ind/CI
313217	21003176	SNOMED - ind/CI
313217	4226399	Is a
313217	500001801	SNOMED - HOI
313217	500002401	SNOMED - HOI
313217	4119602	Subsumes
313217	40631039	Subsumes
313217	4108832	Subsumes
313217	21013671	SNOMED - ind/CI
313217	21013390	SNOMED - ind/CI
313217	313217	Maps to
313217	44821957	Mapped from
313217	2617597	Mapped from
313217	45500085	Mapped from
313217	313217	Mapped from
313217	45951191	Mapped from
313217	21013856	SNOMED - ind/CI
313217	21001575	SNOMED - ind/CI
313217	21001594	SNOMED - ind/CI

Exploring Relationships #2

```sql
SELECT cr.relationship_id, c.*
FROM concept_relationship cr
JOIN concept c ON cr.concept_id_2 = c.concept_id
WHERE cr.concept_id_1 = 313217;
```

Find out related concept

Ancestor concepts

| relationship_id	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
Asso finding of	4194288		Observation	SNOMED	Context-dependent	S	312442005	1/1/1970 0:00	12/31/2099 0:00	NULL
Asso finding of	4203375		Observation	SNOMED	Context-dependent	S	433276002	1/31/2009 0:00	12/31/2099 0:00	NULL
Asso finding of	42689685	Atrial fibrillation not detected	Observation	SNOMED	Context-dependent	S	1.06706E+15	4/1/2017 0:00	12/31/2099 0:00	NULL
Asso finding of	44807374	Atrial fibrillation excluded	Observation	SNOMED	Context-dependent	S	8.16401E+14	4/1/2014 0:00	12/31/2099 0:00	NULL
Concept poss_eq from	40323929	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	155364009	1/1/1970 0:00	3/11/2016 0:00	U
Concept poss_eq from	40345197	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	266306001	1/1/1970 0:00	3/11/2016 0:00	U
Due to of	4139527	Transient cerebral ischemia due to atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426814001	1/1/1970 0:00	12/31/2099 0:00	NULL
Focus of	42709991	Insertion of pacemaker for control of atrial fibrillation	Procedure	SNOMED	Procedure	S	449863006	1/31/2012 0:00	12/31/2099 0:00	NULL
Has finding site	4242112	Atrial structure	Spec Anatomic Site	SNOMED	Body Structure	S	59652004	1/1/1970 0:00	12/31/2099 0:00	NULL
Is a	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S	40593004	1/1/1970 0:00	12/31/2099 0:00	NULL
Is a	4068155	Atrial arrhythmia	Condition	SNOMED	Clinical Finding	S	17366009	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	40323929	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	155364009	1/1/1970 0:00	3/11/2016 0:00	U
Mapped from	2617597	Patient with heart failure and atrial fibrillation documented to be on warfarin therapy	Observation	HCPCS	HCPCS	NULL	G8183	1/1/1970 0:00	11/11/2014 0:00	D
Mapped from	45576876	Unspecified atrial fibrillation	Condition	ICD10CM	5-char billing code	NULL	I48.91	12/30/2006 0:00	12/31/2099 0:00	NULL
Mapped from	45500085	Atrial fibrillation	Condition	Read	Read	NULL	G573000	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	45611600	Atrial Fibrillation	Condition	MeSH	Main Heading	NULL	D001281	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	40345197	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	266306001	1/1/1970 0:00	3/11/2016 0:00	U
Mapped from	45951191	Atrial Fibrillation	Condition	CIEL	Diagnosis	NULL	148203	11/3/2007 0:00	12/31/2099 0:00	NULL
Mapped from	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code	NULL	427.31	1/1/1970 0:00	12/31/2099 0:00	NULL
Maps to	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - HOI	500002401	OMOP Atrial Fibrillation 1	Condition	Cohort	Cohort	C	500002401	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - HOI	500001801	OMOP Qt Prolongation/Torsade De Pointes 1	Condition	Cohort	Cohort	C	500001801	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21005673	Prevention of Thromboembolism in Chronic Atrial Fibrillation	Drug	Indication	Indication	C	5673	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21003176	Tachyarrhythmia	Drug	Indication	Indication	C	3176	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21001542	Supraventricular Tachycardia	Drug	Indication	Indication	C	1542	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21001594	Disease of Cardiovascular System	Drug	Indication	Indication	C	1594	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - MedDRA eq	35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4117112	Controlled atrial fibrillation	Condition	SNOMED	Clinical Finding	S	300996004	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4119601	Lone atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233910005	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4232697	Persistent atrial fibrillation	Condition	SNOMED	Clinical Finding	S	440059007	1/31/2009 0:00	12/31/2099 0:00	NULL
Subsumes	4141360	Chronic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426749004	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	44782442	Atrial fibrillation with rapid ventricular response	Condition	SNOMED	Clinical Finding	S	1.20041E+14	1/31/2014 0:00	12/31/2099 0:00	NULL
Subsumes	4199501	Rapid atrial fibrillation	Condition	SNOMED	Clinical Finding	S	314208002	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4119602	Non-rheumatic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233911009	1/1/1970 0:00	12/31/2099 0:00	NULL

Descendant concepts

Ancestry Relationships: Higher-Level Relationships
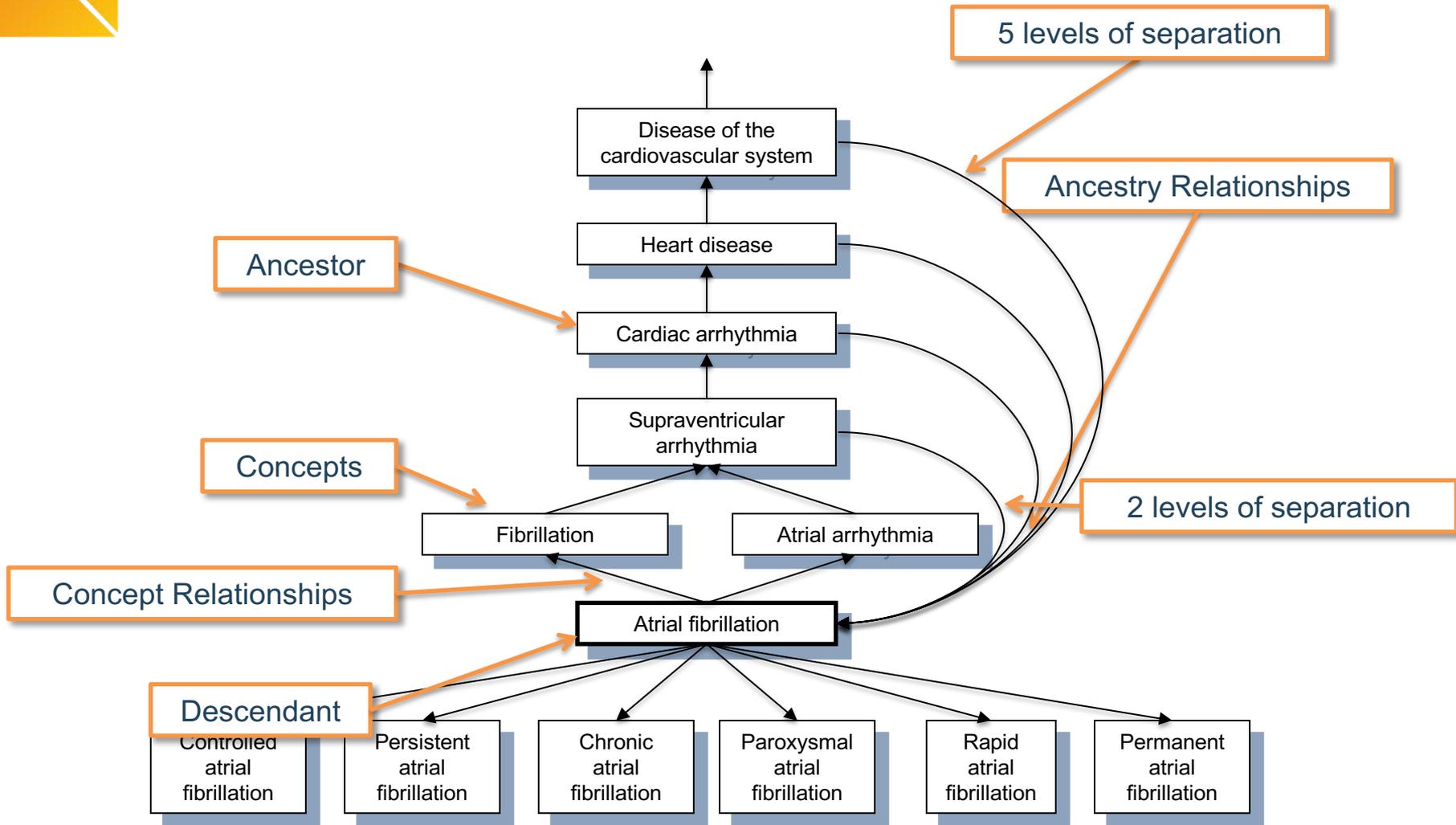
Exploring Ancestors of a Concept

```sql
SELECT max_levels_of_separation, c.*
FROM concept_ancestor ca, concept c
WHERE ca.descendant_concept_id = 313217 /* Atrial fibrillation */
  AND ca.ancestor_concept_id = c.concept_id
ORDER BY max_levels_of_separation
;
```

Hold the descendant

Query Concept

Standard Concepts

Query Concept

max_levels _of_separation	concept _id	concept_na...	domain_i	vocabulary	concept_class	standard _concept
0	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S
0	35204953	Atrial fibrillation	Condition	MedDRA	PT	C
1	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S
1	4068155	Atrial arrhythmia	Condition	SNOMED	Clinical Finding	S
1	35204969	Cardiac fibrillation	Condition	MedDRA	PT	C
2	4248028	Supraventricular arrhythmia	Condition	SNOMED	Clinical Finding	S
2	35204952	Arrhythmia supraventricular	Condition	MedDRA	PT	C
2	35202454	Rate and rhythm disorders NEC	Condition	MedDRA	HLT	C
3	44784217	Cardiac arrhythmia	Condition	SNOMED	Clinical Finding	S
3	35202455	Supraventricular arrhythmias	Condition	MedDRA	HLT	C
4	321588	Heart disease	Condition	SNOMED	Clinical Finding	S
4	35204989	Cardiac disorder	Condition	MedDRA	PT	C
4	35202050	Cardiac arrhythmias	Condition	MedDRA	HLGT	C
5	4103183	Cardiac finding	Condition	SNOMED	Clinical Finding	S
5	440142	Disorder of mediastinum	Condition	SNOMED	Clinical Finding	S
5	134057	Disorder of cardiovascular system	Condition	SNOMED	Clinical Finding	S
5	35204998	Cardiovascular disorder	Condition	MedDRA	PT	C
5	37219970	Mediastinal disorder	Condition	MedDRA	PT	C
5	37622411	Phlebosclerosis	Condition	MedDRA	PT	C
5	35202457	Cardiac disorders NEC	Condition	MedDRA	HLT	C
6	4115390	Mediastinal finding	Condition	SNOMED	Clinical Finding	S
6	4023995	Cardiovascular finding	Condition	SNOMED	Clinical Finding	S

Exploring Descendants of a Concept

Hold the ancestor

```
SELECT max_levels_of_separation, c.*

FROM concept_ancestor ca, concept c

WHERE ca.ancestor_concept_id = 44784217 /* cardiac arrythmia */
  AND ca.descendant_concept_id = c.concept_id

ORDER BY max_levels_of_separation

;
```

MAX_LEVELS_OF_SEPARATION	CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT
0	44784217	Cardiac arrhythmia	Condition	SNOMED	Clinical Finding	S
1	313224	Anomalous atrioventricular excitation	Condition	SNOMED	Clinical Finding	S
1	315643	Tachyarrhythmia	Condition	SNOMED	Clinical Finding	S
1	316429	Premature beats	Condition	SNOMED	Clinical Finding	S
1	316999	Conduction disorder of the heart	Condition	SNOMED	Clinical Finding	S
1	321042	Cardiac arrest	Condition	SNOMED	Clinical Finding	S
1	4030583	Pacemaker twiddler's syndrome	Condition	SNOMED	Clinical Finding	S
1	4057008	Accelerated atrioventricular conduction	Condition	SNOMED	Clinical Finding	S
1	4086313	Withdrawal arrhythmia	Condition	SNOMED	Clinical Finding	S
1	4088507	Ventricular escape complex	Condition	SNOMED	Clinical Finding	S
1	4088986	Atrial escape complex	Condition	SNOMED	Clinical Finding	S
1	4091901	Aberrant premature complexes	Condition	SNOMED	Clinical Finding	S
1	4092011	Aberrantly conducted complex	Condition	SNOMED	Clinical Finding	S
1	4124704	Postoperative sinoatrial disease	Condition	SNOMED	Clinical Finding	S
1	4143042	Ectopic beats	Condition	SNOMED	Clinical Finding	S
1	4164083	Ectopic rhythm	Condition	SNOMED	Clinical Finding	S
1	4172863	Fetal dysrhythmia	Condition	SNOMED	Clinical Finding	S
1	4173170	Neonatal dysrhythmia	Condition	SNOMED	Clinical Finding	S
1	4175473	Atrioventricular dissociation	Condition	SNOMED	Clinical Finding	S
1	4185572	Ventricular arrhythmia	Condition	SNOMED	Clinical Finding	S
1	4217221	Nodal rhythm disorder	Condition	SNOMED	Clinical Finding	S
1	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S
1	4228448	Bradyarrhythmia	Condition	SNOMED	Clinical Finding	S

Let's find Upper Gastrointestinal Bleeding

1. Find some initiation concept

```sql
SELECT * FROM concept WHERE concept_name = 'Upper gastrointestinal bleeding';
```

concept_id	concept_name	concept_level	concept_class	vocabulary_id	concept_code
42891225	Upper gastrointestinal bleeding	1	Lowest Level Term	15	10071910

2. Find standard concepts

```sql
SELECT * FROM concept WHERE lower(concept_name) LIKE '%upper gastrointestinal%'
    AND domain_id = 'Condition' AND standard_concept = 'S';
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
4308202	Acute upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	38938002
4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
4115581	Finding of upper gastrointestinal gas	Condition	SNOMED	Clinical Finding	S	300370006
4103011	Chronic upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	25349007
4012503	Excessive upper gastrointestinal gas	Condition	SNOMED	Clinical Finding	S	162076009
4000609	Disorder of upper gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119291004
4332645	Upper gastrointestinal hemorrhage associated with hypercoagulability state	Condition	SNOMED	Clinical Finding	S	430349003

Going up the hierarchy: Finding the right concept

Hold the descendant

```
SELECT max_levels_of_separation, c.*
FROM concept_ancestor ca, concept c
WHERE ca.descendant_concept_id = 4332645 /* Upper gastrointestinal hemorrhage
  associated...*/
  AND ca.ancestor_concept_id = c.concept_id
  ORDER BY max_levels_of_separation;
```

max_levels _of_separation	concept _id	concept_name	domain _id	vocabulary _id	concept_class _id	standard _concept	concept _code
0	4332645	Upper gastrointestinal hemorrhage associated with hypercoag	Condition	SNOMED	Clinical Finding	S	430349003
1	35708054	Gastritis haemorrhagic	Condition	MedDRA	PT	C	10017866
1	4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
1	35707871	Upper gastrointestinal haemorrhage	Condition	MedDRA	PT	C	10046274
2	35707864	Gastrointestinal haemorrhage	Condition	MedDRA	PT	C	10017955
2	4000609	Disorder of upper gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119291004
2	35707858	Intestinal haemorrhage	Condition	MedDRA	PT	C	10059175
2	35702752	Gastritis (excl infective)	Condition	MedDRA	HLT	C	10017854
2	192671	Gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	74474003
3	37604042	Gastrointestinal haemorrhages	Condition	MedDRA	HLT	C	10052742
3	37622518	Haemorrhage	Condition	MedDRA	PT	C	10055798
3	437312	Bleeding	Condition	SNOMED	Clinical Finding	S	131148009
3	4198525	Disorder of upper digestive tract	Condition	SNOMED	Clinical Finding	S	50410009
3	37622515	Extravasation blood	Condition	MedDRA	PT	C	10015867
3	4000610	Disorder of gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119292006
3	35702116	Gastrointestinal inflammatory conditions	Condition	MedDRA	HLGT	C	10017969
3	35702743	Intestinal haemorrhages	Condition	MedDRA	HLT	C	10022653
3	35702744	Non-site specific gastrointestinal haemorrhages	Condition	MedDRA	HLT	C	10017958
4	35702114	Gastrointestinal haemorrhages NEC	Condition	MedDRA	HLGT	C	10017959
4	4304916	Gastrointestinal tract finding	Condition	SNOMED	Clinical Finding	S	386618008
4	35702767	Nausea and vomiting symptoms	Condition	MedDRA	HLT	C	10028817

Going down: Checking the right content

```sql
SELECT max_levels_of_separation, c.*
FROM concept_ancestor ca, concept c
WHERE ca.ancestor_concept_id = 4291649 /* Upper gastrointestinal hemorrhage */
  AND ca.descendant_concept_id = c.concept_id
ORDER BY max_levels_of_separation;
```

max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
0	4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
1	4318535	Duodenal hemorrhage					
1	23245	Esophageal bleeding					
1	4308202	Acute upper gastrointestinal hemorrhage					
1	4271696	Peptic ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	64121000
1	4103011	Chronic upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	25349007
1	26727	Hematemesis	Condition	SNOMED	Clinical Finding	S	8765009
1	4332645	Upper gastrointestinal hemorrhage associated with hypercoag	Condition	SNOMED	Clinical Finding	S	430349003
1	193250	Gastric hemorrhage	Condition	SNOMED	Clinical Finding	S	61401005
2	4131525	Hemorrhagic gastropathy	Condition	SNOMED	Clinical Finding	S	413218001
2	4204041	Hematemesis - cause unknown	Condition	SNOMED	Clinical Finding	S	308904008
2	4134808	Hemorrhagic duodenopathy	Condition	SNOMED	Clinical Finding	S	413212000
2	4260059	Hemorrhagic gastroenteritis	Condition	SNOMED	Clinical Finding	S	409506009
2	4099014	Duodenal ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	27281001
2	46270145	Gastric hemorrhage due to atrophic gastritis	Condition	SNOMED	Clinical Finding	S	1.5072E+14
2	4096032	Duodenal hematoma	Condition	SNOMED	Clinical Finding	S	262843005
2	4174044	Chronic peptic ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	49232000
2	4095555	Esophageal hematoma	Condition	SNOMED	Clinical Finding	S	262790002
2	46269904	Hemorrhage of duodenum co-occurrent and due to diverticulo	Condition	SNOMED	Clinical Finding	S	1.0866E+15
2	45768629	Gastric hemorrhage due to erosive gastritis	Condition	SNOMED	Clinical Finding	S	7.071E+12

Concept 4291649 and all its descendants comprise Upper GI Bleeding

Exercise: Find Standard Concept ID for Conditions

- Asthma   317009

- Plague   434271

- Ingrown toenail   4065236   4290993

- Your favorite condition here

Does it Work that Way with Drugs?

- Codes
  - NDC, GPI, Multilex, HCPCS, etc.
- Concepts
  - Drug products (Generic and Brand)
  - Drug ingredients
  - Drug Classes
- Relationships
- Ancestry

Drug Hierarchy
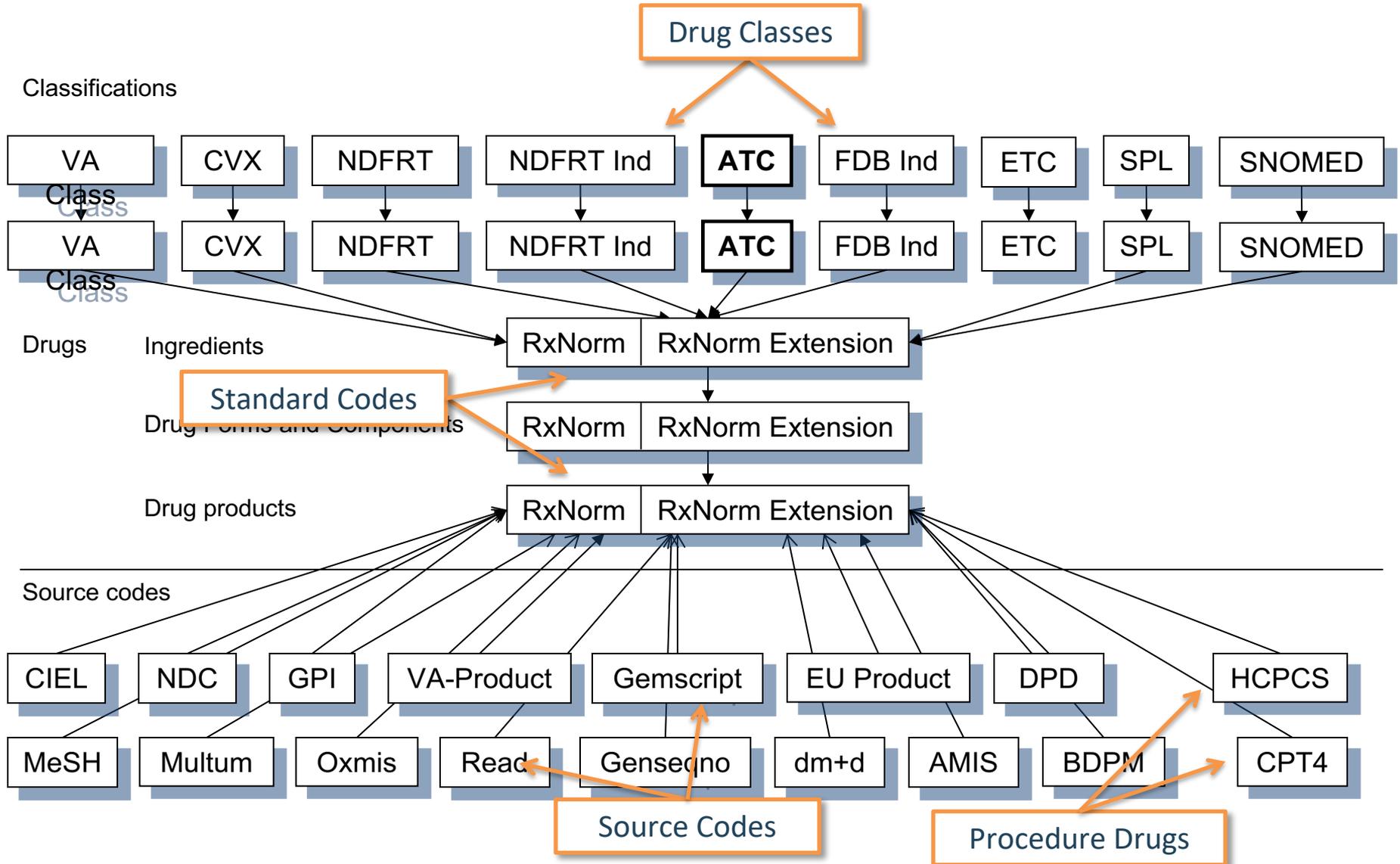
Lunch

Please return in 1 hour

Let's find Warfarin

1. Find active compound Warfarin by keyword

```
SELECT * FROM concept WHERE lower(concept_name) = 'warfarin';
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
21253542	Warfarin	Drug	dm+d	VTM	NULL	48603004
40772658	Warfarin	Measurement	LOINC	LOINC Hierarchy	C	LP16309-4
1847834	WARFARIN	Drug	DA_France	Ingredient	NULL	OMOP13547
4293218	WARFARIN	Drug	NDFRT	Pharma Preparation	NULL	N0000148057
35715898	WARFARIN	Drug	LPD_Australia	Ingredient	NULL	OMOP582219
43081820	warfarin	Drug	Multilex	Ingredient	NULL	2849
4187015	Warfarin	Drug	SNOMED	Substance	NULL	372756006
43343324	Warfarin	Drug	AMT	AU Substance	NULL	2714011000036109
4174989	Warfarin	Drug	SNOMED	Pharma/Biol Product	NULL	48603004
4325514	Warfarin	Drug	NDFRT	Chemical Structure	C	N0000006403
1310149	Warfarin	Drug	RxNorm	Ingredient	S	11289
21600965	warfarin	Drug	ATC	ATC 5th	C	B01AA03
45618204	Warfarin	Drug	MeSH	Main Heading	NULL	D014859

Let's find Clopidogrel

1. Find drug product containing Clopidogrel by NDC code:

Bristol Meyer Squibb's Plavix 75mg capsules: NDC 67544050474

```sql
SELECT * FROM concept WHERE concept_code= '67544050474';
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
45867731	clopidogrel 75 MG Oral Tablet [Plavix]	Drug	NDC	11-digit NDC	NULL	67544050474	2014-07-01	2099-12-31	NULL

```sql
SELECT * FROM concept_relationship WHERE concept_id_1=45867731 and
  relationship_id='Maps to';
```

concept_id_1	concept_id_2	relationship_id	valid_start_date	valid_end_date	invalid_reason
45867731	1322185	Maps to	2015-01-29	2099-12-31	NULL

```sql
SELECT * FROM concept WHERE concept_id=1322185;
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
1322185	clopidogrel 75 MG Oral Tablet [Plavix]	Drug	RxNorm	Branded Drug	S	213169	1970-01-01	2099-12-31	NULL

Let's find Clopidogrel ingredient

2. Find ingredient Clopidogrel as Ancestor of drug product

```sql
SELECT a.max_levels_of_separation, c.*
FROM concept_ancestor ca, concept c
WHERE ca.descendant_concept_id = 1322185 /* clopidogrel 75 MG Oral Tablet [Plavix] */
  AND ca.ancestor_concept_id = c.concept_id;
ORDER BY max_levels_of_separation;
```

max_levels _of_separation	concept _id	concept_name	domain _id	vocabulary _id	concept_class_id	standard _concept	concept _code
0	1322185	clopidogrel 75 MG Oral Tablet [Plavix]	Drug	RxNorm	Branded Drug	S	213169
0	19075601	clopidogrel 75 MG Oral Tablet	Drug	RxNorm	Clinical Drug	S	309362
1	40095879	clopidogrel Oral Tablet [Plavix]	Drug	RxNorm	Branded Drug Form	S	368301
1	19120256	clopidogrel 75 MG [Plavix]	Drug	RxNorm	Branded Drug Comp	S	573094
1	1322187	clopidogrel 75 MG	Drug	RxNorm	Clinical Drug C	S	329449
1	40095878	clopidogrel Oral Tablet	Drug	RxNorm		S	374583
2	36222254	clopidogrel Oral Product	Drug	RxNorm	Clinical Dose Group	C	1163766
2	36229332	Plavix Pill	Drug	RxNorm	Branded Dose Group	C	1181791
2	36229331	Plavix Oral Product	Drug	RxNorm	Branded Dose Group	C	1181790
2	36222255	clopidogrel Pill	Drug	RxNorm	Clinical Dose Group	C	1163767
2	1322184	clopidogrel	Drug	RxNorm	Ingredient	S	32968
3	46319141	CLOPIDOGREL - clopidogrel tablet, film coated	Drug	SPL	Prescription Drug	C	52adfb2c-2062-495c-9954-39eeecae2b41
3	4279519	PLATELET AGGREGATION INHIBITORS	Drug	VA Class	VA Class	C	BL117
3	45796809	clopidogrel 75mg/1 ORAL TABLET, FILM COATED [clopidogrel t	Drug	SPL	Prescription Drug	C	b4e53c96-e280-47c6-baa0-ec676e041d8d
3	45798740	clopidogrel bisulfate 75mg/1 ORAL TABLET, FILM COATED	Drug	SPL	Prescription Drug	C	c7fa330d-d8f1-487e-a730-bafae123e9a8
3	21600985	Platelet aggregation inhibitors excl. heparin	Drug	ATC	ATC 4th	C	B01AC

Clopidogrel

Drug classes

78

Check out Ingredients

3. Check Descendants (other drug products containing Warfarin and Dabigatran)

```sql
SELECT max_levels_of_separation, c.*
FROM concept_ancestor ca, concept c
WHERE ca.ancestor_concept_id = 1310149 /* Warfarin or 1322185 Clopidogrel*/
  AND ca.descendant_concept_id = c.concept_id
ORDER BY max_levels_of_separation;
```

concept_id	concept_name	vocabulary_id	concept_class_id
1310149	Warfarin	RxNorm	Ingredient
36221229	Jantoven Pill	RxNorm	Branded Dose Group
40163559	Warfarin Sodium 6 MG	RxNorm	Clinical Drug Comp
40163544	Warfarin Sodium 3 MG [Jantoven]	RxNorm	Branded Drug Comp
21134746	Warfarin 0.2 MG/ML	RxNorm Extension	Clinical Drug Comp
21105414	Warfarin 5 MG/ML	RxNorm Extension	Clinical Drug Comp
36221228	Jantoven Oral Product	RxNorm	Branded Dose Group
40163565	Warfarin Sodium 7.5 MG	RxNorm	Clinical Drug Comp
21115236	Warfarin 0.3 MG/ML	RxNorm Extension	Clinical Drug Comp
40163509	Warfarin Sodium 1 MG	RxNorm	Clinical Drug Comp
21156284	1 ML Warfarin 0.02 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
21095537	Warfarin 0.3 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
21105427	Warfarin 0.4 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
21046557	Warfarin 1 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
40093133	Warfarin Oral Tablet [Coumadin]	RxNorm	Branded Drug Form
40093134	Warfarin Oral Tablet [Jantoven]	RxNorm	Branded Drug Form
21077698	1 ML Warfarin 1 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
40163534	Warfarin Sodium 2.5 MG Oral Tablet	RxNorm	Clinical Drug
40163530	Warfarin Sodium 2 MG/ML Injectable Solution	RxNorm	Clinical Drug
21066136	Warfarin 5 MG Oral Tablet [Marevan]	RxNorm Extension	Branded Drug
40163542	Warfarin Sodium 3 MG Oral Tablet [Jantoven]	RxNorm	Branded Drug
21116822	1 ML Warfarin 0.6 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21175784	1 ML Warfarin 0.1 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
21175783	1 ML Warfarin 0.832 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug

concept_id	concept_name	vocabulary_id	concept_class_id
1322184	clopidogrel	RxNorm	Ingredient
21043471	clopidogrel Oral Suspension	RxNorm Extension	Clinical Drug Form
36229332	Plavix Pill	RxNorm	Branded Dose Group
21043470	clopidogrel Oral Solution	RxNorm Extension	Clinical Drug Form
21023802	clopidogrel Injectable Solution	RxNorm Extension	Clinical Drug Form
21023806	clopidogrel 5 MG	RxNorm Extension	Clinical Drug Comp
1322187	clopidogrel 75 MG	RxNorm	Clinical Drug Comp
21141600	clopidogrel 1 MG/ML	RxNorm Extension	Clinical Drug Comp
36222254	clopidogrel Oral Product	RxNorm	Clinical Dose Group
21092477	clopidogrel 5 MG/ML	RxNorm Extension	Clinical Drug Comp
21177192	100 ML clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21047899	1 ML clopidogrel 5 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21121870	clopidogrel 5 MG/ML Oral Suspension	RxNorm Extension	Clinical Drug
21063106	clopidogrel 75 MG Oral Tablet [Grepid]	RxNorm Extension	Branded Drug
1322190	clopidogrel 300 MG Oral Tablet [Plavix]	RxNorm	Branded Drug
21121869	clopidogrel 75 MG Injectable Solution	RxNorm Extension	Clinical Drug
21053280	clopidogrel 6 MG Injectable Solution	RxNorm Extension	Clinical Drug
21023810	clopidogrel 4 MG Injectable Solution	RxNorm Extension	Clinical Drug
21106783	1 ML clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
19075601	clopidogrel 75 MG Oral Tablet	RxNorm	Clinical Drug
21102364	clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Clinical Drug
40095879	clopidogrel Oral Tablet [Plavix]	RxNorm	Branded Drug Form
40095878	clopidogrel Oral Tablet	RxNorm	Clinical Drug Form
21088717	100 ML clopidogrel 15 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug

Find members of Drug Classes

4. Check Ingredient Descendants of Drug Class Anticoagulants

```sql
SELECT max_levels_of_separation,  c.*
FROM concept_ancestor ca, concept c
WHERE ca.ancestor_concept_id = 21600961 /* ATC Antithromboic Agent */
  AND ca.descendant_concept_id = c.concept_id
  AND c.concept_class_id = 'Ingredient'
ORDER BY max_levels_of_separation;
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id
46275677	cangrelor	Drug	RxNorm	Ingredient
45892847	edoxaban	Drug	RxNorm	Ingredient
1322184	clopidogrel	Drug	RxNorm	Ingredient
44818499	vorapaxar	Drug	RxNorm	Ingredient
43013024	apixaban	Drug	RxNorm	Ingredient
42898933	defibrotide	Drug	RxNorm	Ingredient
42801108	Protein C	Drug	RxNorm	Ingredient
40241331	rivaroxaban	Drug	RxNorm	Ingredient
1310149	Warfarin	Drug	RxNorm	Ingredient
40241186	Ticagrelor	Drug	RxNorm	Ingredient
40228152	dabigatran etexilate	Drug	RxNorm	Ingredient
40163718	prasugrel	Drug	RxNorm	Ingredient
35604848	selexipag	Drug	RxNorm	Ingredient
19136187	Streptokinase	Drug	RxNorm	Ingredient
19129274	reviparin	Drug	RxNorm	Ingredient

- Metformin   1503297
- Tolazamide   1502809
- Telmisartan   1317640
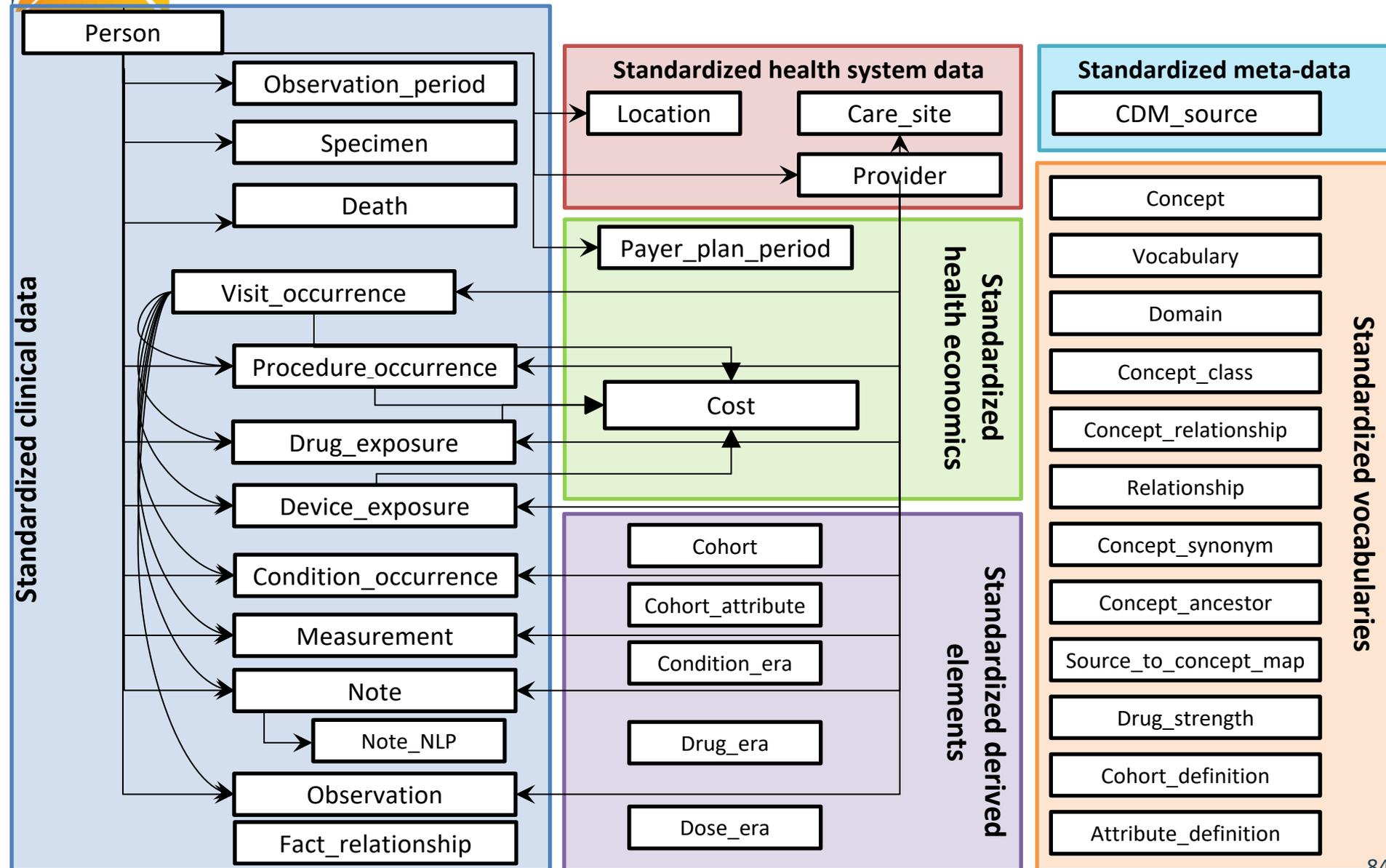- Your favorite ingredient here

- A10AE06
- 686450400
- A10BD14
- Your favorite drug here

35602717

19080217

Common Data Model

History of the model
In depth discussion of model
Era discussion

CDM Version 5 Key Domains

Standardized clinical data

Person

Observation_period

Specimen

Death

Visit_occurrence

Procedure_occurrence

Drug_exposure

Device_exposure

Condition_occurrence

Measurement

Note

Note_NLP

Observation

Fact_relationship

Standardized health system data

Location

Care_site

Provider

Standardized health economics

Payer_plan_period

Cost

Standardized derived elements

Cohort

Cohort_attribute

Condition_era

Drug_era

Dose_era

Standardized meta-data

CDM_source

Standardized vocabularies

Concept

Vocabulary

Domain

Concept_class

Concept_relationship

Relationship

Concept_synonym

Concept_ancestor

Source_to_concept_map

Drug_strength

Cohort_definition

Attribute_definition

OMOP CDM Principles

- OMOP model is an information model
  - Vocabulary(Conceptual) and Data Model are blended
  - Domain-oriented concepts
- Patient centric
- Accommodates data from various sources
- Preserves data provenance
- Extendable
- Evolving

OMOP CDM Standard Domain Features

Feature	Description and purpose	Field name convention	Example
Patient centric	Every domain table has **patient identifier**. Patient data can be retrieved independently from other domains.	**person_id**	person_id 123
Unique domain identifier	Every domain table has a unique primary key to identify domain **entities**	<entity>**_id**	condition_occurrence_id 470985
Standard concept from a respective vocabulary domain	Integration with the vocabulary. Foreign key into the Standard Vocabulary for **Standard Concept**	<entity>**_concept_id**	condition_concept_id 313217 (SNOMED "Atrial Fibrillation")
Source concept from a respective vocabulary domain	Provenance. Foreign key into the Standard Vocabulary for **Source Concept**	<entity>**_source_concept_id**	condition_source_concept_id 44821957 (ICD9CM "Atrial Fibrillation")
Source value	Provenance. Verbatim information from the source data, **not to be used** by any standard analytics	<entity>**_source_value**	condition_source_value 427.31 (ICD9CM "Atrial Fibrillation")
Source type	Provenance. Foreign key into the Vocabulary for the **origin of the**	<entity>**_type_concept_id**	condition_type_concept_id 38000199 ("Inpatient header – primary")

PERSON

Person

- person_id
- gender_concept_id
- year_of_birth
- month_of_birth
- day_of_birth
- birth_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

- Need to create one unique record per person (not multiple rows per move)

- Vocabulary for gender, race, ethnicity: HL7 administrative

- No history of location/demographics: need to select latest available

- Location peculiarity: foreign key to the LOCATION table that contains one record per each unique location

- Year of birth required…day/month optional

LOCATION

- Contains one record per each unique location

- Location is highly variable across sources, of limited use thus far

OBSERVATION_PERIOD

Observation_Period

- observation_period_id
- person_id
- observation_period_start_date
- observation_period_end_date
- period_type_concept_id

- Spans of time where data source has capture of data

- Required to run analytical methods

- One person may have multiple periods if there is interruption in data capture

- Challenge: determine observation periods based on the source data

DEATH

- Can have death without cause

- Can only have 1 death per person

VISIT_OCCURRENCE

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
- admitting_source_concept_id
- admitting_source_value
- discharge_to_concept_id
- discharge_to_source_value
- preceding_visit_occurrence_id

- Visits <> 'Encounters':
  - claims often need to be consolidated to minimize double-counting
  - inpatient transitions are not covered

- Visit Types
  - Inpatient
  - Emergency room
  - Inpatient/Emergency
  - Outpatient
  - Long-term care

- Vocabulary: OMOP

- Other attributes: time of visit start/end, provider, admitting source, discharge disposition

CONDITION_OCCURRENCE

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
stop_reason
provider_id
visit_occurrence_id
condition_source_value
condition_source_concept_id
condition_status_source_value
condition_status_concept_id

- Vocabulary: SNOMED -> classification

- Data sources:
  – Billing diagnosis (inpatient, outpatient)
  – Problem list

- Individual records <> distinct episodes

'Dirty' Conditions

Codes mapped to domains other than the original source domain

Description	AUT MESUR CHIMIO PROPHYL
CIM10	0Z2920
CIM10 Description	OTH PROPHYLACTIC CHEMO
Maps to	**DRUG**

Description	EXSPEC DEPS AUT MAL PREC
CIM10	0Z1380
CIM10 Description	SPEC SCR OTHER SPEC DIS
Maps to	**OBSERVATION**

Description	GRSS CONSTT FORTUITMT
CIM10	0Z3300
CIM10 Description	PREG STATE INCIDENT
Maps to	**CONDITION**

DRUG_EXPOSURE

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
drug_source_value
drug_source_concept_id
route_source_value
dose_unit_source_value

- Vocabulary: RxNorm-> classifications by drug class and indication

- Data sources:
  - Pharmacy dispensing
  - Prescriptions written
  - Medication history

- Source fields may vary, but so inference of drug exposure end may vary

'Dirty' Drugs

Drug Source Description	Form Desc	Admin Route Description	Generic Name	Maps To
OBSERVATION PERIOD	Miscellaneous	Unspecified	Documentation	**OBSERVATION**
LUMBAR DDS BELT	Miscellaneous	Unspecified	Unspecified	**DEVICE**
JOBST KNEE HIGH COMPRESSION STOCKING	Miscellaneous	Unspecified	Antiembolism stockings	**DEVICE**
MASKS	Miscellaneous	Unspecified	Masks	**DEVICE**
PEN NEEDLES	Miscellaneous	Unspecified	Needle	**DEVICE**

PROCEDURE_OCCURRENCE

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
- procedure_source_value
- procedure_source_concept_id
- modifier_source_value

- Vocabularies: CPT-4, HCPCS, ICD-9 Procedures, ICD-10 Procedures, LOINC, SNOMED

- Procedures have the least standardized vocabularies that causes some redundancy

DEVICE_EXPOSURE

- OMOP CDM is the only data model supporting devices
- Accommodates FDA unique device identifiers (UDI) even though most data sources don't have them yet

MEASUREMENT

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
- measurement_source_value
- measurement_source_concept_id
- unit_source_value
- value_source_value

- Entity-Attribute-Value (EAV) design

- Vocabulary: LOINC, SNOMED

- Data sources: structured, quantitative measures, such as laboratory tests

- Measures have associated units
  - Measurement units vocabulary: UCUM

- No free format for measurement results

Measurement Data Issues

- ## The unit of measure is inconsistent in the source data

  - Makes evaluation and studies hard to do

```sql
select distinct unit_source_value
from measurement
where measurement_concept_id IN
 (
    SELECT concept_id
    FROM concept
    WHERE concept_name like '%LDL%'
      AND standard_concept = 'S'
      AND domain_id = 'Measurement'
 )
```

unit_source_value
N < 3.5
g/l
NULL
< 3,2
ng/ml
mmol/l
N < 3.6

Measurement Data Issues

```
select distinct(round (value_as_number/10))*10 as value,
        count (*)
from measurement
where measurement_concept_id in
  (
    SELECT concept_id
    FROM concept
    WHERE concept_name like '%LDL%'
        AND standard_concept = 'S'
        AND domain_id = 'Measurement'
  )
group by value
```

value	↓ Σ ▽ ⊡	count	↓ Σ ▽ ⊡
0		39784	
10		302	
180		1	
20		1	
120		1	
30		2	
NULL		2	
4380		1	
50		2	
60		1	
430		1	

OBSERVATION

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
observation_source_value
observation_source_concept_id
unit_source_value
qualifier_source_value

- Catch-all EAV design to capture all other data:
  - observation: 'question'
  - value: 'answer'
    - Can be numeric, concept, or string (e.g. free text)

- Instrument for CDM extension, playpen

- Not all 'questions' are standardized, source value can accommodate 'custom' observations (particularly pertinent in registries)

SPECIMEN

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

- To capture biomarkers / tissue bank

NOTE

Note

- note_id
- person_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
- note_source_value
- visit_occurrence_id

- To capture unstructured free text

NOTE_NLP

Note_NLP
🔑 note_nlp_id
note_id
section_concept_id
snippet
offset
lexical_variant
note_nlp_concept_id
note_nlp_source_concept_id
nlp_system
nlp_date
nlp_date_time
term_exists
term_temporal
term_modifiers

- The NOTE_NLP table will encode all output of NLP on clinical notes. Each row represents a single extracted term from a note.

Health Economics

Payer_Plan_Period

- 🔑 payer_plan_period_id
- person_id
- payer_plan_period_start_date
- payer_plan_period_end_date
- payer_source_value
- plan_source_value
- family_source_value
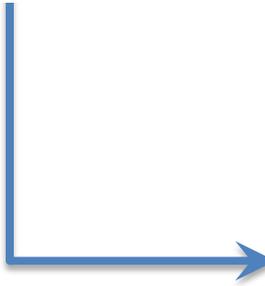
Cost

- 🔑 cost_id
- cost_event_id
- cost_domain_id
- cost_type_concept_id
- currency_concept_id
- total_charge
- total_cost
- total_paid
- paid_by_payer
- paid_by_patient
- paid_patient_copay
- paid_patient_coinsurance
- paid_patient_deductible
- paid_by_primary
- paid_ingredient_cost
- payer_plan_period_id
- amount_allowed
- revenue_code_concept_id
- revenue_code_source_value
- drg_concept_id
- drg_source_value

- All costs consolidated into one table COST table

- Costs tied to respective observation records

- Domain is determined by cost_domain_id (e.g. visit, condition, etc.)

Cohorts
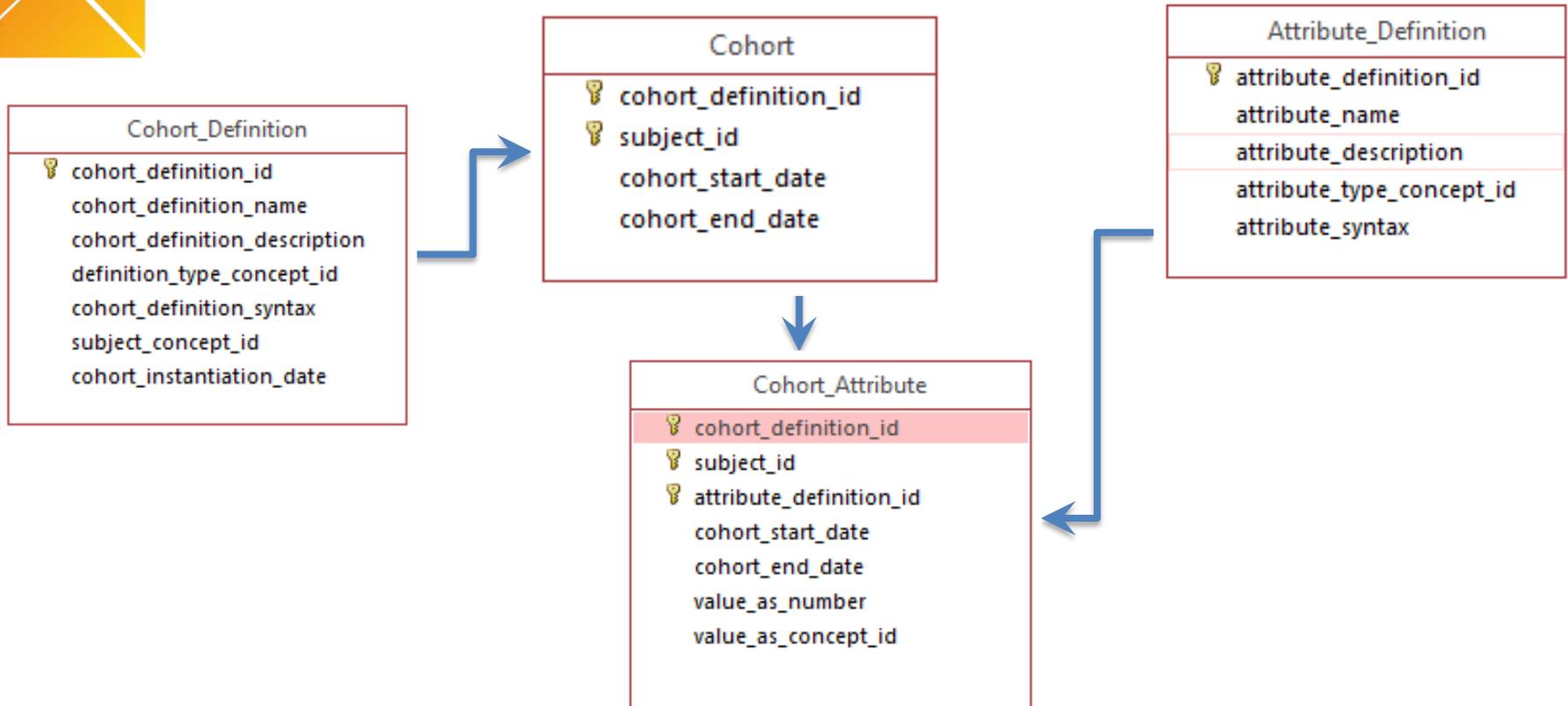
1. **COHORT** table contains records of subjects that satisfy a given set of criteria for a duration of time.
2. The definition of the cohort is contained within the **COHORT_DEFINITIO**N table. It provides a standardized structure for maintaining the rules governing the inclusion of a subject into a cohort, and can store programming code to instantiate the cohort within the OMOP CDM.
3. **COHORT_ATTRIBUTE** table contains attributes associated with each subject within a cohort, as defined by a given set of criteria for a duration of time.
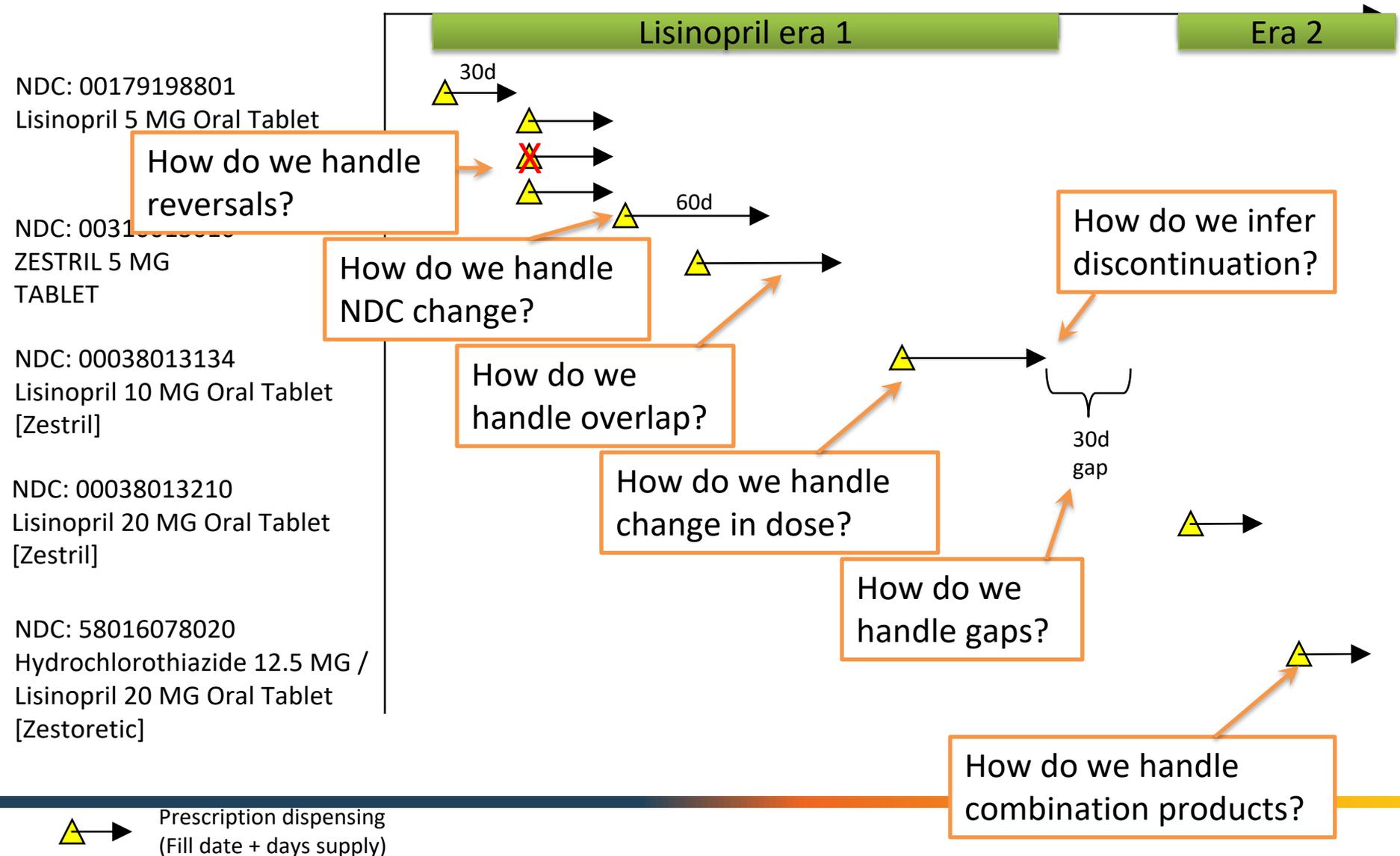4. The definition of the Cohort Attribute is contained in the **ATTRIBUTE_DEFINITION** table.

DRUG_ERA

Drug_Era
🔑 drug_era_id
person_id
drug_concept_id
drug_era_start_date
drug_era_end_date
drug_exposure_count
gap_days

- Standardized inference of length of exposure to product for all active ingredients

- Derived from records in DRUG_EXPOSURE under certain rules to produce continuous Drug Eras

Illustrating inferences needed within longitudinal pharmacy claims data for one patient

Person Timeline

NDC: 00179198801
Lisinopril 5 MG Oral Tablet

How do we handle reversals?

NDC: 00310013020
ZESTRIL 5 MG TABLET

NDC: 00038013134
Lisinopril 10 MG Oral Tablet [Zestril]

NDC: 00038013210
Lisinopril 20 MG Oral Tablet [Zestril]

NDC: 58016078020
Hydrochlorothiazide 12.5 MG / Lisinopril 20 MG Oral Tablet [Zestoretic]

Lisinopril era 1

Era 2

30d

60d

How do we handle NDC change?

How do we handle overlap?

How do we handle change in dose?

How do we infer discontinuation?

30d gap

How do we handle gaps?

How do we handle combination products?

Prescription dispensing
(Fill date + days supply)

ETL: Real world scenario

PharMetrics Plus

CLAIMS

pat_id	claimno	from_dt	to_dt	diagprc_ind	Diag_admit	diag1
05917921689	IPA333393946	1/5/2006	1/5/2006	1	41071	41071

LRx/Dx

MEDICAL_CLAIMS

md_clm_id	ims_pat_nbr	dt_of_service	rxer_id	diag_cd
95963982102	80445908	8/1/2012 0:00	680488	41071

German DA

Problem Events

db_country	international_practice_num	international_doctor_num	international_patient_num	age
GE	GE6326	GE8784	GE46478747	

Diagnosis

db_country	international_diagnosis_num	diagnosis_num	icd10_4_c
GE	GE2397573	2397573	I21.4

Ambulatory EMR

Problem

Patient_id_synth	Diag_dt	Icd10_cd
271138	4/11/2013	I214

4 real observational databases, all containing an inpatient admission for a patient with a diagnosis of 'acute subendocardial infarction'
- Not a single table name the same…
- Not a single variable name the same….
- Different table structures (rows vs. columns)
- Different conventions (with and without decimal points)
- Different coding schemes (ICD9 vs. ICD10)

What does it mean to ETL to OMOP CDM? Standardize **structure** and **content**

PharMetrics Plus
Inpatient Claims

pat_id	claimno	from_dt	to_dt	diagprc_ind	Diag_admit
05917921689	IPA333393946	1/5/2006	1/5/2006	1	41071

Structure optimized for large-scale analysis for clinical characterization, population-level estimation, and patient-level prediction

PharMetrics Plus
CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
05917921689	1/5/2006	41071	Inpatient claims - primary position

Content using international vocabulary standards that can be applied to any data source

PharMetrics Plus
CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
05917921689	1/5/2006	41071	Inpatient claims - primary position	44825429	444406

OMOP CDM = Standardized structure: same tables, same fields, same datatypes, same conventions across disparate sources

PharMetrics Plus
CLAIMS

pat_id	claimno	from_dt	to_dt	diagprc_ind	Diag_admit	diag1
05917921689	IPA333393946	1/5/2006	1/5/2006	1	41071	41071

LRx/Dx
MEDICAL_CLAIMS

md_clm_id	ims_pat_nbr	dt_of_service	rxer_id	diag_cd
95963982102	80445908	8/1/2012 0:00	680488	41071

German DA
Problem Events

db_country	international_practice_num	international_doctor_num	international_patient_num	age_at_event	date_of_event	international_diagnosis_num
GE	GE6326	GE8784	GE46478747	20	11/19/2014 0:00	GE2397573

Diagnosis

db_country	international_diagnosis_num	diagnosis_num	icd10_4_code	icd10_3_text	diagnosis_confidence
GE	GE2397573	2397573	I21.4	Non-ST elevation (NSTEMI) myocardial infarction	Confirmed

Ambulatory EMR
Problem
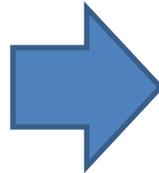
Patient_id_synth	Diag_dt	Icd10_cd
271138	4/11/2013	I214

PharMetrics Plus: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
157033702	1/5/2006	41071	Inpatient claims - primary position

LRX/DX: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
80445908	8/1/2012	41071	Primary Condition

German DA : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
46478747	11/19/2014	I21.4	EHR problem list entry

Ambulatory EMR : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
271138	4/11/2013	I214	Primary Condition

- Consistent structure optimized for large-scale analysis
- Structure preserves all source content and provenance

OMOP CDM = Standardized content: common vocabularies across disparate sources

PharMetrics Plus: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
05917921689	1/5/2006	41071	Inpatient claims - primary position	44825429	444406

LRx/Dx: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
80445908	8/1/2012	41071	Primary Condition	44825429	444406

German DA : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
6478747	11/19/2014	I21.4	EHR problem list entry	4557208	444406

Ambulatory EMR : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
271138	4/11/2013	I214	Primary Condition	4557208	444406

- Standardize across vocabularies to a common referent standard (ICD9/10→SNOMED)
- Source codes mapped into each domain standard so that now you can talk across different languages

- Standardize source codes to be uniquely defined across all vocabularies
- No more worries about formatting or code overlap

112

Standards

- Patients without transaction

- Cleaning dirty data
  - Patient IDs reused
  - Bogus code records (e.g. '000')

- How to handle tobacco information

THEMIS

https://github.com/OHDSI/CommonDataModel/wiki

113

CDM Version Control

- Working group meets once a month to discuss proposed changes to the CDM

- All CDM documentation, versions, and proposals located on Github
  - https://github.com/OHDSI/CommonDataModel
  - Proposals tracked and discussed as Github issues

- Meeting information can be found on the working group wiki page

- Please contact Clair Blacketer (mblacke@its.jnj.com) for more information

Break

Please return in 15 minutes

CDM Examples

Leveraging OHDSI Tools (GitHub /Forums/ Working Group) Exercises

Data Used for Demonstration

- Medicare Claims Synthetic Public Use Files (SynPUFs)
  - synthetic US Medicare insurance claims database
  - Medicare is a government based insurance program for primarily 65 and older but also individuals with disabilities
  - SynPUF not for research but rather demonstration/development purposes
  - Has been converted to the Common Data Model

CMS.gov
Centers for Medicare & Medicaid Services

Data Used for Demonstration

- Five types of data:

	DE-SynPUF	Unit of record	Number of Records 2008	Number of Records 2009	Number of Records 2010
1	*Beneficiary Summary*	Beneficiary	2,326,856	2,291,320	2,255,098
2	*Inpatient Claims*	claim	547,800	504,941	280,081
3	*Outpatient Claims*	claim	5,673,808	6,519,340	3,633,839
4	*Carrier Claims*	claim	34,276,324	37,304,993	23,282,135
5	*Prescription Drug Events (PDE)*	event	39,927,827	43,379,293	27,778,849

SynPUF High Level Diagram

SYNPUF DIAGRAM

- Beneficiary Summary
- Inpatient Claims
- Outpatient Claims
- Carrier Claims
- Prescription Drug Events (PDE)
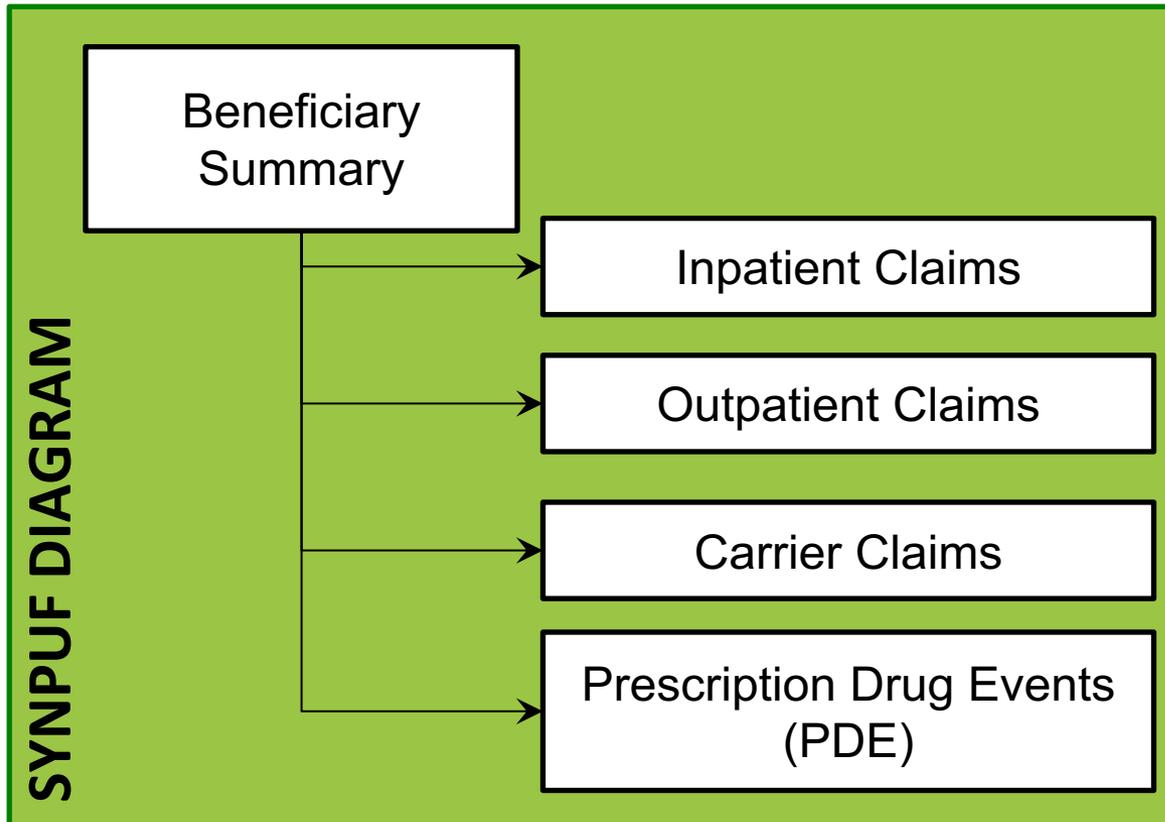
Mapping SynPUF to CDM

SynPUF

CDM

120

OHDSI in a Box

CDM Database:
pgAdmin III New Server

- Click on PGAdmin

CDM Database: Connect

- Password: `ohdsi`

CDM Database:
Open SQL Sheet

CDM Database: Ready

```
set search_path to 'public', 'ohdsi';
```

Open Up SQL File

- 🌐 Open up Google Chrome on Remote Desktop

- Navigate to:

  ## https://goo.gl/qdjLBp

```
/*****************************/
/* Standard Vocabulary Introduction */
/*****************************/

/* select from concept table */
SELECT * FROM concept WHERE concept_id = 313217

/* or... */
SELECT * FROM concept WHERE concept_code = '49436004';

/* select from vocabulary table */
SELECT * FROM vocabulary
order BY vocabulary_id;

SELECT * FROM concept
WHERE concept_name = 'Atrial fibrillation'
AND vocabulary_id = 'SNOMED';

/* Concept ID vs Concpet Code */
SELECT * FROM concept WHERE concept_code = '1001';

/* Concept by name */
SELECT * FROM concept WHERE concept_name = 'Atrial fibrillation';

/* Find relationship for Atrial fibrillation */
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957
ORDER BY relationship_id;
```

Some Example Questions

Ex 0   Finding Warfarin

Ex 1   New Users of Warfarin

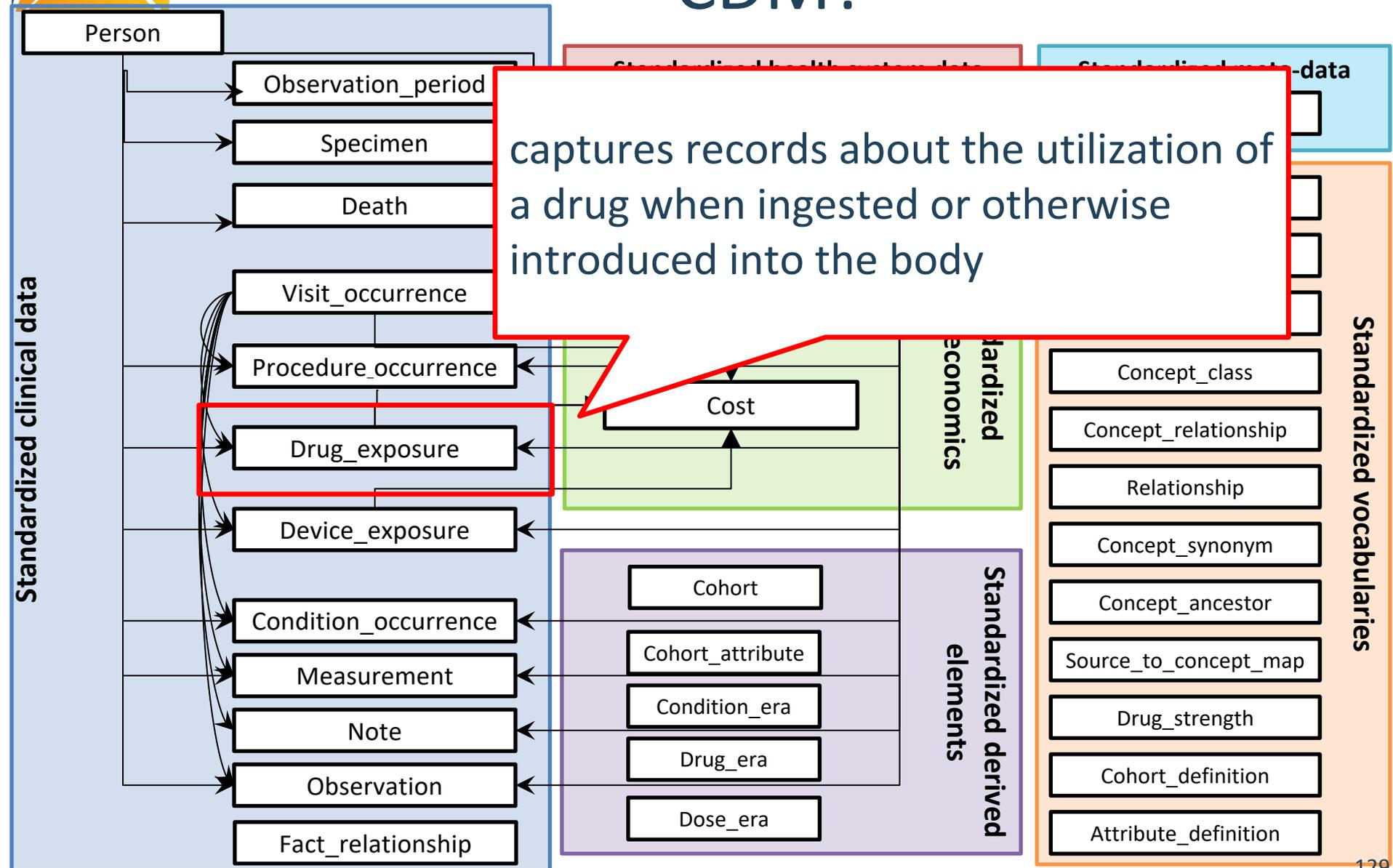**Ex 2**   New Users of Warfarin who are >=65?

Ex 3   New Users of Warfarin with prior Atrial Fibrillation?

Warfarin Exposure

- Warfarin is a blood thinner that is used to treat/prevent blood clots.

  – Where do you find drug data in the CDM?

  – What codes do I use to define drugs?

Where are Drug Exposures in the CDM?

Standardized clinical data

- Person
- Observation_period
- Specimen
- Death
- Visit_occurrence
- Procedure_occurrence
- Drug_exposure
- Device_exposure
- Condition_occurrence
- Measurement
- Note
- Observation
- Fact_relationship

Standardized health system data

Standardized meta-data

Standardized economics

- Cost

Standardized derived elements

- Cohort
- Cohort_attribute
- Condition_era
- Drug_era
- Dose_era

Standardized vocabularies

- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor
- Source_to_concept_map
- Drug_strength
- Cohort_definition
- Attribute_definition

captures records about the utilization of a drug when ingested or otherwise introduced into the body

How do I define Warfarin?

- When raw data is transformed into the CDM raw source codes are transformed into standard OMOP Vocabulary concepts

- In the CDM, we no longer care what source codes existed in the raw data, we just need to use concept identifiers

- We can use the OMOP Vocabulary to identify all concepts that contain the ingredient warfarin

How do I define Warfarin?
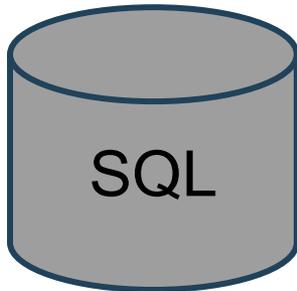
- Writing SQL Statement

- OHDSI Tool ATLAS

Finding Warfarin

```
/**********************************************************************
*           (Exercise 0) Finding Warfarin
**********************************************************************

/*Just looking for the ingredient concept*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE DRUG_CONCEPT_ID = 1310149 /*warfarin*/;
```

0 individuals

```
/*Looking for drugs associated with the ingredient*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
);


/*looking for anticoagulants, a class of drugs warfarin belongs*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 4283987 /*ANTICOAGULANTS (VA Class)*/
);
```

Finding Warfarin

```
/*******************************************************************************
*          (Exercise 0) Finding Warfarin
*******************************************************************************

/*Just looking for the ingredient concept*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE DRUG_CONCEPT_ID = 1310149 /*warfarin*/;

/*Looking for drugs associated with the ingredient*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
);

/*looking for anticoagulants, a class of drugs warfarin belongs*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 4283987 /*ANTICOAGULANTS (VA Class)*/
);
```

0 individuals

25,602 individuals

Finding Warfarin

```
/****************************************************************************
*          (Exercise 0) Finding Warfarin
****************************************************************************

/*Just looking for the ingredient concept*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE DRUG_CONCEPT_ID = 1310149 /*warfarin*/;
```

0 individuals

```
/*Looking for drugs associated with the ingredient*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
);
```

25,602 individuals

```
/*looking for anticoagulants, a class of drugs warfarin belongs*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 4283987 /*ANTICOAGULANTS (VA Class)*/
);
```

38,027 individuals

Some Example Questions

Ex 0   Finding Warfarin
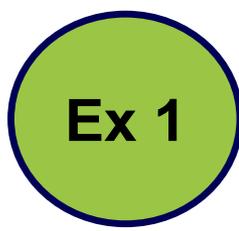
Ex 1   New Users of Warfarin

Ex 2   New Users of Warfarin who are >=65?

Ex 3   New Users of Warfarin with prior Atrial Fibrillation?
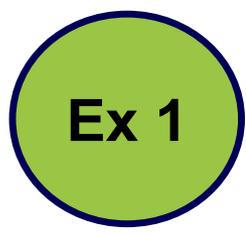
How do I define new users of a drug?

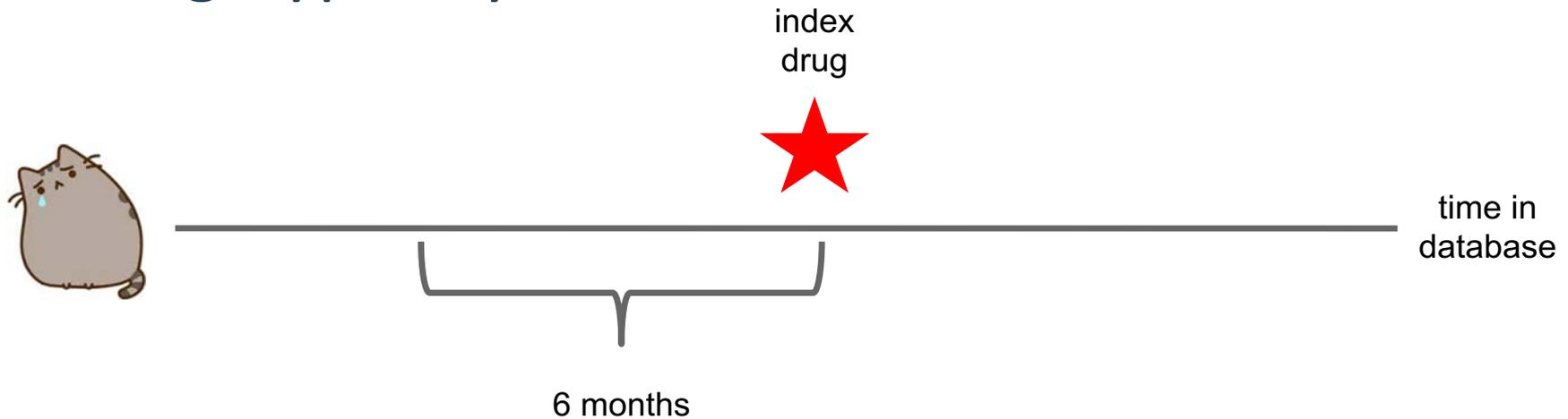Someone who has recently started taking the drug, typically with a 6 or 12 month wash out

| 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |

How do I define new users of a drug?

Someone who has recently started taking the drug, typically with a 6 or 12 month wash out
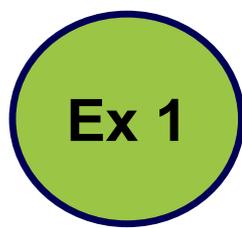

index drug

time in database

6 months

What is Needed in the CDM?

- **OMOP Vocabulary**
  to find the concepts

- **CDM Table DRUG_EXPOSURE**
  to find individuals with exposure

- **CDM Table OBSERVATION_PERIOD**
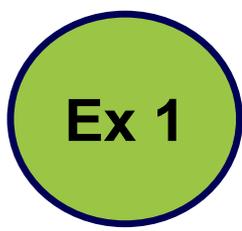 to know people's time within the database

New Users of Warfarin

```sql
/************************************************************************
 *    (Exercise 1) Warfarin New Users
 ************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```

Step 1: Get the codes you need

```
/*****************************************************************************
*    (Exercise 1) Warfarin New Users
*****************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```
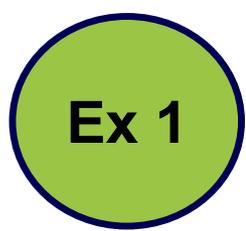
Step 2: Find Drug Exposures

```
/*******************************************************************************
 *    (Exercise 1) Warfarin New Users
 ******************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```

Step 3: Find New Users

```
/*******************************************************************************
*    (Exercise 1) Warfarin New Users
*******************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```
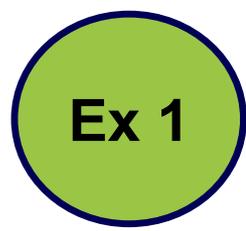
New Users of Warfarin

```
/**********************************************************************
*    (Exercise 1) Warfarin New Users
**********************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```

New Users of Warfarin

Try running this on your own!

How many people do you get?
18,080 individuals

Some Example Questions

Ex 0 Finding Warfarin

Ex 1 New Users of Warfarin

Ex 2 New Users of Warfarin who are >=65?
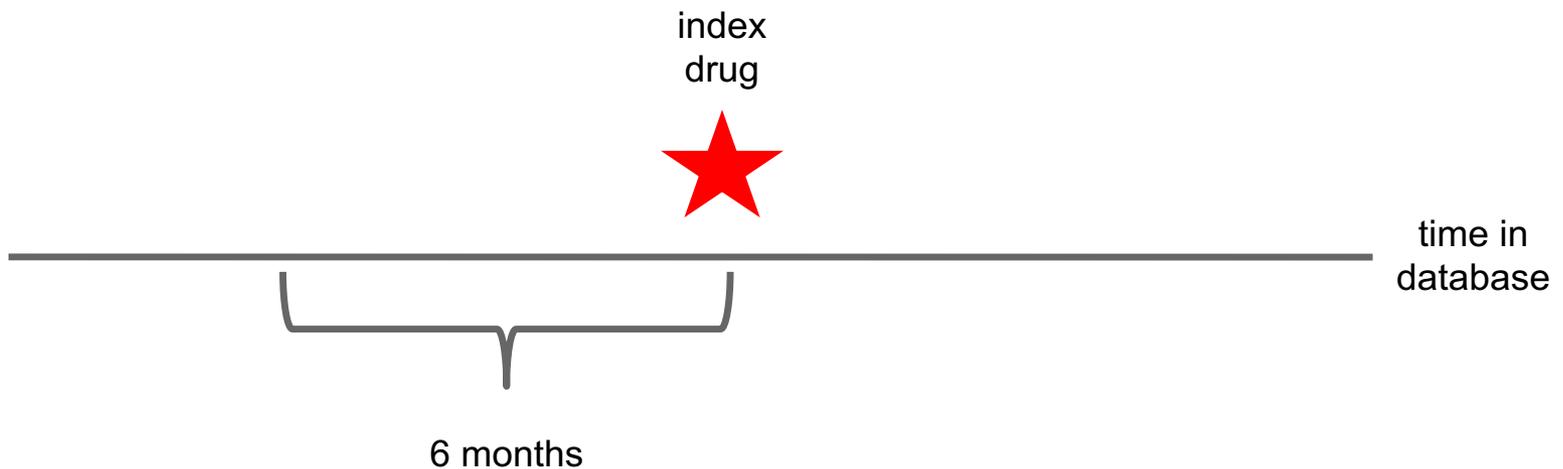
Ex 3 New Users of Warfarin with prior Atrial Fibrillation?

How do I define new users of warfarin who are >=65?

Someone who has recently started taking the drug, typically with a 6 or 12 month wash out

index drug

>=65 years old

6 months

time in database

What is Needed in the CDM?

- **OMOP Vocabulary**
  to find the concepts

- **DRUG_EXPOSURE**
  to find individuals with exposure

- **OBSERVATION_PERIOD**
  to know people's time within the database

- **PERSON**
  to know year of birth

```
/*****************************************************************************
*         (Exercise 2) Warfarin New Users 65 or Older at Index
*****************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
        (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX,
        EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH AS AGE_AT_INDEX
FROM CTE_DRUG_INDEX i
        JOIN OBSERVATION_PERIOD op
                ON op.PERSON_ID = i.PERSON_ID
                AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        JOIN PERSON p
                ON p.PERSON_ID = i.PERSON_ID
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
AND EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH >= 65
ORDER BY i.PERSON_ID
```

Step 2: Add the Person Table to calculate age

Ex 2

```
/**************************************************************************
*        (Exercise 2) Warfarin New Users 65 or Older at Index
**************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
        (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX,
        EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH AS AGE_AT_INDEX
FROM CTE_DRUG_INDEX i
        JOIN OBSERVATION_PERIOD op
                ON op.PERSON_ID = i.PERSON_ID
                AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        JOIN PERSON p
                ON p.PERSON_ID = i.PERSON_ID
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
AND EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH >= 65
ORDER BY i.PERSON_ID
```

New Users of Warfarin >= 65 years of age

Try running this on your own!

```
/********************************************************************
 *       (Exercise 2) Warfarin New Users 65 or Older at Index
 ********************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
        (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX,
        EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH AS AGE_AT_INDEX
FROM CTE_DRUG_INDEX i
        JOIN OBSERVATION_PERIOD op
                ON op.PERSON_ID = i.PERSON_ID
                AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        JOIN PERSON p
                ON p.PERSON_ID = i.PERSON_ID
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
AND EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH >= 65
ORDER BY i.PERSON_ID
```

How many people do you get?

Some Example Questions

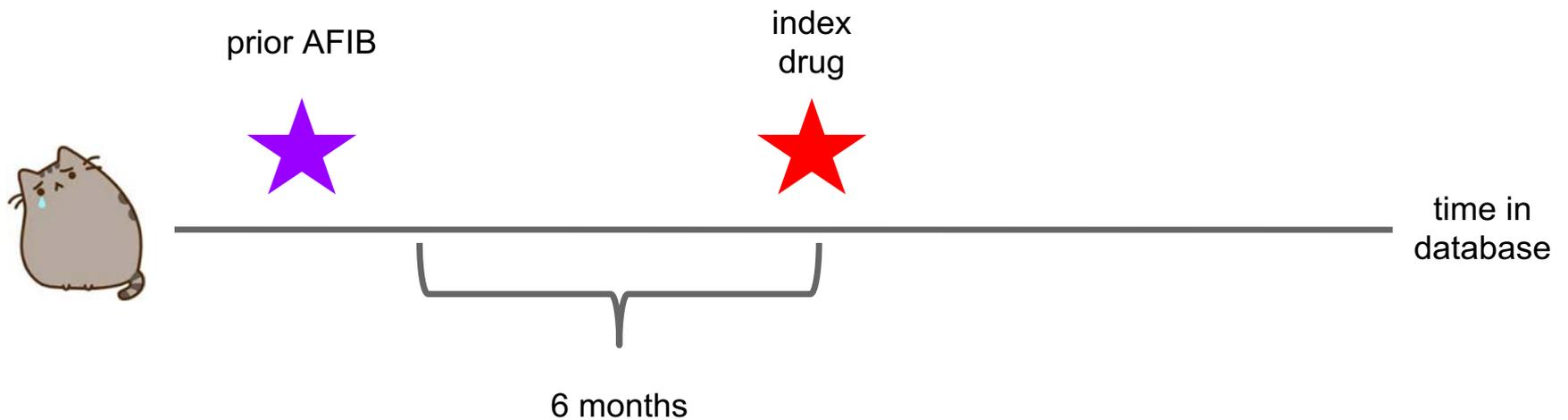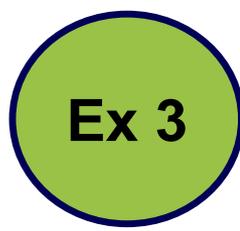**Ex 0** Finding Warfarin

Ex 1 New Users of Warfarin

Ex 2 New Users of Warfarin who are >=65?

Ex 3 New Users of Warfarin with prior Atrial Fibrillation?

How do I define new users of Warfarin with prior Atrial Fibrillation?

prior AFIB

index drug

time in database

6 months

What is Needed in the CDM?

- **OMOP Vocabulary**
  to find the concepts

- **DRUG_EXPOSURE**
  to find individuals with exposure

- **OBSERVATION_PERIOD**
  to know people's time within the database

- **PERSON**
  to know year of birth

- **CONDITION_OCCURRENCE**
  to find presence of a disease

```
/*******************************************************************************
*        (Exercise 3) Warfarin New Users With Prior AFIB
*******************************************************************************/


WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
),
CTE_DRUG_NEW_USERS AS (
        SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
                (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
        FROM CTE_DRUG_INDEX i
                JOIN OBSERVATION_PERIOD op
                        ON op.PERSON_ID = i.PERSON_ID
                        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
)
SELECT nu.*, MIN(nu.INDEX_DATE-co.CONDITION_START_DATE) AS DAYS_OF_CLOSEST_AFIB_PRIOR_TO_INDEX
FROM CTE_DRUG_NEW_USERS nu
        JOIN CONDITION_OCCURRENCE co
                ON co.PERSON_ID = nu.PERSON_ID
                AND co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_DATE
WHERE co.CONDITION_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID =  313217 /*Atrial fibrillation*/
)
AND co.CONDITION_START_DATE < nu.INDEX_DATE
GROUP BY nu.PERSON_ID, nu.INDEX_DATE, nu.OBSERVATION_PERIOD_START_DATE, nu.OBSERVATION_PERIOD_END_DATE, nu.DAYS_BEFORE_INDEX
ORDER BY nu.PERSON_ID
```
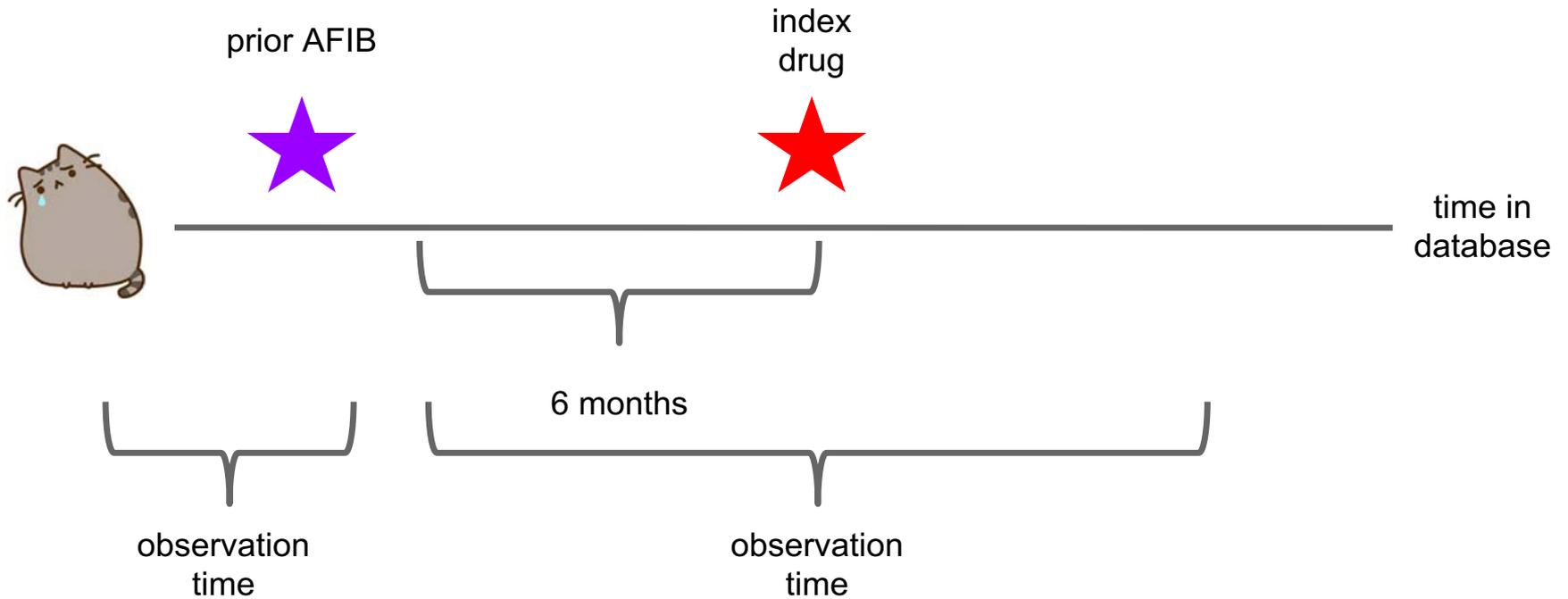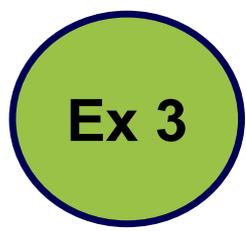
Step 2: Define Atrial Fibrillation

```
/*******************************************************************
 *       (Exercise 3) Warfarin New Users With Prior AFIB
 *******************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
),
CTE_DRUG_NEW_USERS AS (
        SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
                (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
        FROM CTE_DRUG_INDEX i
                JOIN OBSERVATION_PERIOD op
                        ON op.PERSON_ID = i.PERSON_ID
                        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
)
SELECT nu.*, MIN(nu.INDEX_DATE-co.CONDITION_START_DATE) AS DAYS_OF_CLOSEST_AFIB_PRIOR_TO_INDEX
FROM CTE_DRUG_NEW_USERS nu
        JOIN CONDITION_OCCURRENCE co
                ON co.PERSON_ID = nu.PERSON_ID
                AND co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_DATE
WHERE co.CONDITION_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID =  313217 /*Atrial fibrillation*/
)
AND co.CONDITION_START_DATE < nu.INDEX_DATE
GROUP BY nu.PERSON_ID, nu.INDEX_DATE, nu.OBSERVATION_PERIOD_START_DATE, nu.OBSERVATION_PERIOD_END_DATE, nu.DAYS_BEFORE_INDEX
ORDER BY nu.PERSON_ID
```

```
/*********************************************************************
*       (Exercise 3) Warfarin New Users With Prior AFIB
*********************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
),
CTE_DRUG_NEW_USERS AS (
        SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
                (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
        FROM CTE_DRUG_INDEX i
                JOIN OBSERVATION_PERIOD op
                        ON op.PERSON_ID = i.PERSON_ID
                        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND
        WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
)
SELECT nu.*, MIN(nu.INDEX_DATE-co.CONDITION_START_DATE) AS DAYS_OF_CLOSEST_AFIB_PRIOR
FROM CTE_DRUG_NEW_USERS nu
        JOIN CONDITION_OCCURRENCE co
                ON co.PERSON_ID = nu.PERSON_ID
                AND co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_DATE
WHERE co.CONDITION_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 313217 /*Atrial fibrillation*/
)
AND co.CONDITION_START_DATE < nu.INDEX_DATE
GROUP BY nu.PERSON_ID, nu.INDEX_DATE, nu.OBSERVATION_PERIOD_START_DATE, nu.OBSERVATION_PERIOD_END_DATE, nu.DAYS_BEFORE_INDEX
ORDER BY nu.PERSON_ID
```

Keeps condition within the same observable time, exclude if you want all time prior

How do I define new users of Warfarin with prior Atrial Fibrillation?

prior AFIB

index drug

time in database

6 months

observation time

observation time

New Users of Warfarin with prior Atrial Fibrillation

```sql
/*****************************************************************************
*       (Exercise 3) Warfarin New Users With Prior AFIB
*****************************************************************************/


WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
),
CTE_DRUG_NEW_USERS AS (
        SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
                (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
        FROM CTE_DRUG_INDEX i
                JOIN OBSERVATION_PERIOD op
                        ON op.PERSON_ID = i.PERSON_ID
                        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
)
SELECT nu.*, MIN(nu.INDEX_DATE-co.CONDITION_START_DATE) AS DAYS_OF_CLOSEST_AFIB_PRIOR_TO_INDEX
FROM CTE_DRUG_NEW_USERS nu
        JOIN CONDITION_OCCURRENCE co
                ON co.PERSON_ID = nu.PERSON_ID
                AND co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_DATE
WHERE co.CONDITION_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID =  313217 /*Atrial fibrillation*/
)
AND co.CONDITION_START_DATE < nu.INDEX_DATE
GROUP BY nu.PERSON_ID, nu.INDEX_DATE, nu.OBSERVATION_PERIOD_START_DATE, nu.OBSERVATION_PERIOD_END_DATE, nu.DAYS_BEFORE_INDEX
ORDER BY nu.PERSON_ID
```

Try on your own!

- Warfarin New Users 65 or Older at Index with Prior Atrial Fibrillation

8,207 individuals

- Bonus: Clopidogrel New Users 65 or Older at Index with Prior Atrial Fibrillation

3,148 individuals

Queries Can Be Automated

- Open up Google Chrome

- Navigate to: http://127.0.0.1:8080/atlas/

- Example cohort:
  "Warfarin New Users 65 or Older at Index with Prior Atrial Fibrillation"