OMOP Common Data Model and Standardized Vocabularies

Rimma Belenkaya, M.A., M.S.
Clair Blacketer, MPH, PMP
George Hripcsak, MD, MS
Karthik Natarajan, PhD
Don O’Hara, MS

Christian Reich, MD, PhD
Gowtham Rao, MD, PhD
Don Torok, MS
Mui Van Zandt
Mark Velez, MA
Erica A. Voss, MPH, PMP

Please copy the contents of the USB drive to your hard disk now if VM not set up previously.
You will need ~45GB free disk space available.
After the Tutorials, you will know...

1. What's OMOP, OHDSI?
2. How does the Standardized Vocabulary work?
3. How do I find codes and Concepts?
4. How do I navigate the hierarchy?
5. What is the OMOP CDM?
6. How to use the OMOP CDM
<table>
<thead>
<tr>
<th>Section</th>
<th>Speaker</th>
<th>Time</th>
<th>Room</th>
<th>Item(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration</td>
<td>-</td>
<td>8:00AM - 9:00AM</td>
<td>Glen Foyer</td>
<td>Glen Foyer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1 hour)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>George / Mark / Karthik</td>
<td>9:00AM - 10:00AM</td>
<td>Glen Echo</td>
<td>Introductions and Ground Rules Foundational</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1 hour)</td>
<td></td>
<td>• History of OMOP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Why and How</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Birth of OHDSI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Introduction to OMOP Common Data Model</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OHDSI Community</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Example of Remote Study</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VM Overview</td>
</tr>
<tr>
<td>Break</td>
<td>-</td>
<td>10:00AM - 10:15AM</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15 min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary – Part 1</td>
<td>Christian</td>
<td>10:15AM - 12:30PM</td>
<td>Glen Echo</td>
<td>Basic Relationship, Ancestors, &amp; Descendants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2 hours &amp; 15 min)</td>
<td></td>
<td>How does it work for Drugs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SQL Examples</td>
</tr>
<tr>
<td>Lunch</td>
<td>-</td>
<td>12:30PM - 1:30PM</td>
<td>Glen Foyer</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1 hour)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Agenda (cont.)

<table>
<thead>
<tr>
<th>Section</th>
<th>Speaker</th>
<th>Time</th>
<th>Room</th>
<th>Item(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary – Part 2</td>
<td>Christian</td>
<td>1:30PM - 2:00PM</td>
<td>Glen Echo</td>
<td>Continued</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(30 min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Data Model</td>
<td>Mui / Rimma</td>
<td>2:00PM - 3:15PM</td>
<td>Glen Echo</td>
<td>History of the model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1 hour &amp; 15 min)</td>
<td></td>
<td>In depth discussion of model</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Era discussion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Real World Scenario</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ETL Pitfalls</td>
</tr>
<tr>
<td>Break</td>
<td>-</td>
<td>3:15PM - 3:30PM</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15 min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDM Examples</td>
<td>Erica / Clair</td>
<td>3:30PM - 5:00 PM</td>
<td>Glen Echo</td>
<td>Leveraging OHDSI Tools (GitHub/Forums/Working Group)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1 hour &amp; 30 min)</td>
<td></td>
<td>Exercises</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OHDSI Community</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Conclusion Game</td>
</tr>
</tbody>
</table>
Instructors

<table>
<thead>
<tr>
<th>Rimma Belenkaya, M.A., M.S.</th>
<th>Clair Blacketer, MPH, PMP</th>
<th>George Hripcsak, MD, MS</th>
<th>Karthik Natarajan, PhD</th>
<th>Don O’Hara, MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Image](6x444 to 98x543)</td>
<td>![Image](83x282 to 170x372)</td>
<td>![Image](375x284 to 460x370)</td>
<td>![Image](472x122 to 558x214)</td>
<td>![Image](570x122 to 652x214)</td>
</tr>
<tr>
<td>Christian Reich, MD, PhD</td>
<td>Gowtham Rao, MD, PhD</td>
<td>Don Torok, MS</td>
<td>Mui Van Zandt</td>
<td>Mark Velez, MA</td>
</tr>
<tr>
<td>![Image](179x282 to 269x372)</td>
<td>![Image](83x118 to 168x218)</td>
<td>![Image](179x124 to 265x210)</td>
<td>![Image](280x273 to 360x373)</td>
<td>![Image](375x126 to 459x212)</td>
</tr>
<tr>
<td>Don Torok, MS</td>
<td>Mui Van Zandt</td>
<td>Mark Velez, MA</td>
<td>Erica A. Voss, MPH, PMP</td>
<td></td>
</tr>
<tr>
<td>![Image](275x120 to 366x217)</td>
<td>![Image](470x289 to 559x370)</td>
<td>![Image](179x282 to 269x372)</td>
<td>![Image](83x118 to 168x218)</td>
<td></td>
</tr>
</tbody>
</table>

Acknowledgements:
Anthony Reckard, Mike Warfe, Dmytry Dymshyts, & Michael Goodman
Ground Rules

• We are recording today’s session, so when asking questions wait for a microphone.

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

• Please return the Virtual Machine (VM) distributed today, unless you want to use it for some good purpose.

• If we cannot get the VM working on your machine let’s try to buddy you up. Do not worry the presentation will still walk you through the content.

• If you do not already have the VM set up, please begin copying over the flash drive to your local computer (local drive).
Foundational

What is OMOP/OHDSI?
OMOP Common Data Model (CDM) – Why and How
FDA Regulatory Action over Time

Number of FDA-caused Withdrawals
FDAAA calls for establishing Risk Identification and Analysis System

SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.

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

“(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—

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

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

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

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

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

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

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

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

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)

- Open-source
- Standards-based

Common Data Model

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

Drug

<table>
<thead>
<tr>
<th>Outcome</th>
<th>ACE Inhibitors</th>
<th>Amphotericin B</th>
<th>Antibiotics: sulfa, erythromycin, tetracyclines</th>
<th>Antiepileptics: carbamazepine, phenytoin</th>
<th>Benzodiazepines</th>
<th>Beta blockers</th>
<th>Bisphosphonates: alendronate</th>
<th>Tricyclic antidepressants</th>
<th>Typical antipsychotics</th>
<th>Warfarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angioedema</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aplastic Anemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute Liver Injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bleeding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip Fracture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial Infarction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality after MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal Failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI Ulcer Hospitalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- True positive' benefit
- True positive' risk
- Negative control'

OMOP Methods Library
- Inception cohort
- Case control
- Logistic regression

• 14 methods
• Epidemiology designs
• Statistical approaches adapted for longitudinal data
OMOP Experiment 2 (2011-2012)

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

**Observational data**
- 4 claims databases
- 1 ambulatory EMR

**Drug-outcome pairs**

<table>
<thead>
<tr>
<th></th>
<th>Positives</th>
<th>Negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>165</td>
<td>234</td>
</tr>
<tr>
<td>Myocardial Infarction</td>
<td>36</td>
<td>66</td>
</tr>
<tr>
<td>Upper GI Bleed</td>
<td>24</td>
<td>67</td>
</tr>
<tr>
<td>Acute Liver Injury</td>
<td>81</td>
<td>37</td>
</tr>
<tr>
<td>Acute Renal Failure</td>
<td>24</td>
<td>64</td>
</tr>
</tbody>
</table>
European OMOP Experiment

### Observational data
- Aarhus
- Pedianet
- ARS
- IPCI
- HS
- PHARMO

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

### Drug-outcome pairs

<table>
<thead>
<tr>
<th></th>
<th>Positives</th>
<th>Negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>165</td>
<td>234</td>
</tr>
<tr>
<td>Myocardial Infarction</td>
<td>36</td>
<td>66</td>
</tr>
<tr>
<td>Upper GI Bleed</td>
<td>24</td>
<td>67</td>
</tr>
<tr>
<td>Acute Liver Injury</td>
<td>81</td>
<td>37</td>
</tr>
<tr>
<td>Acute Renal Failure</td>
<td>24</td>
<td>64</td>
</tr>
</tbody>
</table>
Ground Truth for OMOP Experiment

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

<table>
<thead>
<tr>
<th>Event</th>
<th>Positive controls</th>
<th>Negative controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Liver Injury</td>
<td>81</td>
<td>37</td>
<td>118</td>
</tr>
<tr>
<td>Acute Myocardial Infarction</td>
<td>36</td>
<td>66</td>
<td>102</td>
</tr>
<tr>
<td>Acute Renal Failure</td>
<td>24</td>
<td>64</td>
<td>88</td>
</tr>
<tr>
<td>Upper Gastrointestinal Bleeding</td>
<td>24</td>
<td>67</td>
<td>91</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>165</strong></td>
<td><strong>234</strong></td>
<td><strong>399</strong></td>
</tr>
</tbody>
</table>

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

Drugs:
- Isoniazid
- Fluticasone
- Indomethacin
- Clindamycin
- Ibuprofen
- Loratadine
- Sertraline
- Pioglitazone
<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aarhus</td>
<td>Danish national health registry, covering the Aarhus region. Includes inhabitant registry, drug dispensations, hospital claims, lab values, and death registry.</td>
<td>2 M</td>
</tr>
<tr>
<td>ARS</td>
<td>Italian record linkage system covering the Tuscany region, including inhabitant registry, drug dispensations, hospital claims, and death registry</td>
<td>4 M</td>
</tr>
<tr>
<td>Health-Search</td>
<td>Italian general practice database (no children)</td>
<td>1 M</td>
</tr>
<tr>
<td>IPCI</td>
<td>Dutch general practice database</td>
<td>0.75 M</td>
</tr>
<tr>
<td>Pedianet</td>
<td>Italian general practice pediatric database</td>
<td>0.14 M</td>
</tr>
<tr>
<td>PHARMO</td>
<td>Dutch record linkage system. Includes inhabitant registry, drug dispensations, hospital claims, and lab values.</td>
<td>1.28 M</td>
</tr>
</tbody>
</table>
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
Observational Health Data Sciences and Informatics (OHDSI) Plans and Ambitions
Letter Soup

• **OMOP**: ended in 2013 with Symposium

• **IMEDS**: Program at Reagan-Udall Foundation of the FDA
  – Methodological research to inform Industry and Agency
  – Research Lab

• **OHDSI**: Open Research Collaborative started by OMOP PIs and coordinated through Columbia University
  – Multiple stakeholders: academia, government, industry
  – Multiple geographies: US, Europe, Asia-Pacific
  – Multiple disciplines: Statistics, epidemiology, informatics, clinical sciences
  – Maintains OMOP CDM and Vocabularies
• 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

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

Current Approach: “One Study – One Script”

Analytical method: Adherence to Drug

Application to data

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
- Safety Signals
- Source of Business

OHDSI Tools

- Standardized data
- North America
- Southeast Asia
- China
- Europe
- UK
- Japan
- India
- So Africa
- Switzerland
- Italy
- Israel
Analytics can be remote

North America  Southeast Asia  China
Europe  UK  Japan  India
So Africa  Switzerland  Italy  Israel
Analytics can be behind firewall
Network Studies
Networks of networks

Another Network

Network

Coordinating Center

EMR

ISDN

University Medical Center

Inpatient Hospital

Outpatient Hospital

Claims Asset

EMR Asset

Claims Asset

EMR Asset

Claims Asset

EMR Asset

Claims Asset

EMR Asset

Claims Asset

EMR Asset
Evolution of the CDM

OMOP CDM now Version 5, following multiple iterations of implementation, testing, modifications, and expansion based on the experiences of the OMOP community who bring on a growing landscape of research use cases.

http://omop.org/CDM
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
- Standardized health economics
  - Location
  - Care_site
  - Provider
  - Payer_plan_period
  - Cost
- Standardized elements
  - Cohort
  - Cohort_attribute
  - Condition_era
  - Drug_era
  - Dose_era

Standardized derived elements

Standardized meta-data
- CDM_source
- Concept
- Vocabulary
- Domain
- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor
- Source_to_concept_map
- Drug_strength
- Cohort_definition
- Attribute_definition

Standardized health economics
- Standardized vocabularies
## Standard Variable Name Conventions

<table>
<thead>
<tr>
<th>Field name</th>
<th>Purpose</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;entity&gt;_concept_id</code></td>
<td>Foreign key into the Standard Vocabulary for <strong>Standard Concept</strong></td>
<td>condition_concept_id 313217 (SNOMED &quot;Atrial Fibrillation&quot;)</td>
</tr>
<tr>
<td><code>&lt;entity&gt;_source_concept_id</code></td>
<td>Foreign key into the Standard Vocabulary for <strong>Source Concept</strong></td>
<td>condition_source_concept_id 44821957 (ICD9CM &quot;Atrial Fibrillation&quot;)</td>
</tr>
<tr>
<td><code>&lt;entity&gt;_source_value</code></td>
<td>Verbatim information from the source data, <strong>not to be used</strong> by any standard analytics</td>
<td>condition_source_value 427.31 (ICD9CM &quot;Atrial Fibrillation&quot;)</td>
</tr>
<tr>
<td><code>&lt;entity&gt;_type_concept_id</code></td>
<td>Foreign key into the Vocabulary for the <strong>origin of the information</strong></td>
<td>condition_type_concept_id 38000199 (&quot;Inpatient header – primary&quot;)</td>
</tr>
<tr>
<td><code>&lt;entity&gt;_id</code></td>
<td>Unique identifiers for <strong>entities</strong> (row numbers, or IDs imported from source)</td>
<td>person_id 1234567 visit_occurrence_id 7654321 could be a person identifier or an autogenerated number by the CDM builder</td>
</tr>
</tbody>
</table>
All vocabularies stacked up in one table

Vocabulary ID
### What's in a Concept

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONCEPT_ID</td>
<td>313217</td>
</tr>
<tr>
<td>CONCEPT_NAME</td>
<td>Atrial fibrillation</td>
</tr>
<tr>
<td>DOMAIN_ID</td>
<td>Condition</td>
</tr>
<tr>
<td>VOCABULARY_ID</td>
<td>SNOMED</td>
</tr>
<tr>
<td>CONCEPT_CLASS_ID</td>
<td>Clinical Finding</td>
</tr>
<tr>
<td>STANDARD_CONCEPT</td>
<td>S</td>
</tr>
<tr>
<td>CONCEPT_CODE</td>
<td>49436004</td>
</tr>
<tr>
<td>VALID_START_DATE</td>
<td>01-Jan-1970</td>
</tr>
<tr>
<td>VALID_END_DATE</td>
<td>31-Dec-2099</td>
</tr>
<tr>
<td>INVALID_REASON</td>
<td></td>
</tr>
</tbody>
</table>

- **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

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

PostgreSQL

VirtualBox

cdm

webapi

pgAdmin

Broadsea

WebAPI

WebTools

Atlas

Penelope

Calypso

Tomcat

Methods Library

OHDSI R packages

Studio

synpuf_100k

WhiteRabbit

RabbitInAHat
OHDSI in a Box – Setup

1. Open VM VirtualBox Manager

2. Click on New

Name and operating system

Please choose a descriptive name for the new virtual machine and select the type of operating system you intend to install on it. The name you choose will be used throughout VirtualBox to identify this machine.

- Name: OHDSI-1percent
- Type: Linux
- Version: Ubuntu (64-bit)

Memory size

Select the amount of memory (RAM) in megabytes to be allocated to the virtual machine.

The recommended memory size is **1024 MB.**

4 MB **8 MB** 2048 **4096 MB** 8192 MB

- Do not add a virtual hard disk
- Create a virtual hard disk now
- Use an existing virtual hard disk file

ohdsi-1k-Final.vdi (Normal, 30.00 GB)
OHDSI in a Box – Start Up
OHDSI in a Box – Adjust Resolution
OHDSI in a Box – Clipboard
OHDSI in a Box – Timeout
OHDSI in a Box – Ready
CDM Database – pgAdmin III New Server

New Server Registration

Properties
Name: ohdsi
Host: localhost
Port: 5432
Service: 
Maintenance DB: postgres
Username: ohdsi
Password: ohdsi
Store password: 
Colour: 
Group: Servers

SSL

SSH Tunnel

Advanced

OK Cancel
CDM Database – Connect
CDM Database – Open SQL Sheet
CDM Database – Ready

```
select * from concept limit 10;
```
Break

Please return in 15 minutes
Vocabulary

Basic Relationship, Ancestors, & Descendants

How does it work for Drugs

SQL Examples
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
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?

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Events</th>
<th>100,000 days at risk</th>
<th>Patients</th>
<th>Events</th>
<th>100,000 days at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gastrointestinal hemorrhage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis with required diagnosis of atrial fibrillation</td>
<td>10,599</td>
<td>16</td>
<td>1.6</td>
<td>43,541</td>
<td>160</td>
<td>3.5</td>
</tr>
<tr>
<td>Sensitivity analysis without required diagnosis of atrial fibrillation</td>
<td>12,195</td>
<td>19</td>
<td>1.6</td>
<td>119,940</td>
<td>338</td>
<td>3.1</td>
</tr>
<tr>
<td><strong>Intracranial hemorrhage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis with required diagnosis of atrial fibrillation</td>
<td>10,587</td>
<td>8</td>
<td>0.8</td>
<td>43,594</td>
<td>109</td>
<td>2.4</td>
</tr>
<tr>
<td>Sensitivity analysis without required diagnosis of atrial fibrillation</td>
<td>12,182</td>
<td>10</td>
<td>0.9</td>
<td>120,020</td>
<td>204</td>
<td>1.9</td>
</tr>
</tbody>
</table>
All Content in CDM is Coded as Concepts

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

```
SELECT *  
FROM concept  
WHERE concept_id = 313217
```
Dozens of schemes, formats, rules

<table>
<thead>
<tr>
<th>PATH_TO_ROOT</th>
<th>SEQUENCE</th>
<th>IMMEDIATE_PARENT</th>
<th>CODE</th>
<th>CODE_TEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP317</td>
<td>1</td>
<td>LP31755-9</td>
<td>LP31755-9</td>
<td>Microbiology</td>
</tr>
<tr>
<td>LP317</td>
<td>1</td>
<td>LP14559-6</td>
<td>LP14559-6</td>
<td>Microorganism</td>
</tr>
<tr>
<td>LP317</td>
<td>1</td>
<td>LP98185-9</td>
<td>LP98185-9</td>
<td>Bacteria</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LOINC_NL_COMPONENT</th>
<th>PROPERTY</th>
<th>TIME_ASPECT</th>
<th>SYSTEM</th>
<th>SCALE_TYP</th>
<th>METHOD_TYP</th>
<th>CLASS</th>
<th>SOURCE</th>
<th>DATE_LAST_CHANGING</th>
<th>TYCOMMEN</th>
<th>STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>xylose^2H post 25 g xylose PO</td>
<td>MCnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td>CHAL</td>
<td>SH</td>
<td>19961220</td>
<td>ADD</td>
<td>ACTIVE</td>
<td></td>
</tr>
<tr>
<td>xylose^30M post 25 g xylose PO</td>
<td>MCnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td>CHAL</td>
<td>SH</td>
<td>19961220</td>
<td>ADD</td>
<td>ACTIVE</td>
<td></td>
</tr>
<tr>
<td>xylose^post 6H CFst</td>
<td>MCnc</td>
<td>Pt</td>
<td>Tiss</td>
<td>Ord</td>
<td>Immune stain</td>
<td>PATH</td>
<td>SH</td>
<td>SDL-M</td>
<td>20060706</td>
<td>MIN</td>
</tr>
</tbody>
</table>

| CMS32_DESC_LONG_SHORT_DX.xlsx |

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

Standard vocabulary

Classifications

Top-level classification

Higher-level classifications

Low-level concepts

Source codes

SNOMED-CT

MedDRA

System organ class

High-level group terms

High-level terms

Preferred terms

Low-level terms

Source codes

ICD10
ICD10CM
Read
SNOMED
Oxmis
Ciel
MeSH
ICD9CM

Standard vocabulary

Classifications

Top-level classification

Higher-level classifications

Low-level concepts

Source codes
Finding the Right Concept: #1

1. ..if I know the ID

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

<table>
<thead>
<tr>
<th>CONCEPT_ID</th>
<th>CONCEPT_NAME</th>
<th>DOMAIN_ID</th>
<th>VOCABULARY_ID</th>
<th>CONCEPT_CLASS_ID</th>
<th>STANDARD_CONCEPT</th>
<th>CONCEPT_CODE</th>
<th>VALID_START_DATE</th>
<th>VALID_END_DATE</th>
<th>INVALID_REASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>313217</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>49436004</td>
<td>01-Jan-1970</td>
<td>31-Dec-2099</td>
<td></td>
</tr>
</tbody>
</table>

2. ..if I know the code

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

<table>
<thead>
<tr>
<th>CONCEPT_ID</th>
<th>CONCEPT_NAME</th>
<th>DOMAIN_ID</th>
<th>VOCABULARY_ID</th>
<th>CONCEPT_CLASS_ID</th>
<th>STANDARD_CONCEPT</th>
<th>CONCEPT_CODE</th>
<th>VALID_START_DATE</th>
<th>VALID_END_DATE</th>
<th>INVALID_REASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>313217</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>49436004</td>
<td>01-Jan-1970</td>
<td>31-Dec-2099</td>
<td></td>
</tr>
</tbody>
</table>
Concept code 49436004 in SNOMED Browser
SELECT *
FROM concept
WHERE concept_code = '1001';

<table>
<thead>
<tr>
<th>Concept_Name</th>
<th>Concept Class</th>
<th>Vocabulary_ID</th>
<th>Concept_Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antipyrine</td>
<td>Ingredient</td>
<td>RxNorm</td>
<td>1001</td>
</tr>
<tr>
<td>Aceprometazine maleate</td>
<td>Ingredient</td>
<td>BDPM</td>
<td>1001</td>
</tr>
<tr>
<td>Serum</td>
<td>Specimen</td>
<td>CIEL</td>
<td>1001</td>
</tr>
<tr>
<td>methixene hydrochloride</td>
<td>Ingredient</td>
<td>Multilex</td>
<td>1001</td>
</tr>
<tr>
<td>Brompheniramine Maleate, 10 mg/mL injectable solution</td>
<td>Ingredient</td>
<td>Multum</td>
<td>1001</td>
</tr>
<tr>
<td>ABBOTT COLD SORE BALM 4%/0.06% W/ Drug Product</td>
<td>Drug Product</td>
<td>LPD_Australia</td>
<td>1001</td>
</tr>
<tr>
<td>Residential Treatment - Psychiatric</td>
<td>Revenue Code</td>
<td>Revenue Code</td>
<td>1001</td>
</tr>
</tbody>
</table>
Finding the Right Concept: #2

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

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

<table>
<thead>
<tr>
<th>CONCEPT_ID</th>
<th>CONCEPT_NAME</th>
<th>DOMAIN_ID</th>
<th>VOCABULARY_ID</th>
<th>CONCEPT_CLASS_ID</th>
<th>STANDARD_CONCEPT</th>
<th>CONCEPT_CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>313217</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>49436004</td>
</tr>
<tr>
<td>44821957</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>ICD9CM</td>
<td>5-digit billing code</td>
<td>427.31</td>
<td></td>
</tr>
<tr>
<td>35204953</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
<td>10003658</td>
</tr>
<tr>
<td>45500085</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>Read</td>
<td>Read</td>
<td></td>
<td>G573000</td>
</tr>
<tr>
<td>45883018</td>
<td>Atrial fibrillation</td>
<td>Meas Value</td>
<td>LOINC</td>
<td>Answer</td>
<td>S</td>
<td>LA17084-7</td>
</tr>
</tbody>
</table>
Finding the Right Concept: #3

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

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

<table>
<thead>
<tr>
<th>CONCEPT_ID</th>
<th>CONCEPT_NAME</th>
<th>DOMAIN_ID</th>
<th>VOCABULARY_ID</th>
<th>CONCEPT_CLASS_ID</th>
<th>STANDARD_CONCEPT</th>
<th>CONCEPT_CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>44821957</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>ICD9CM</td>
<td>5-dig billing code</td>
<td></td>
<td>427.31</td>
</tr>
</tbody>
</table>

```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957;
```

<table>
<thead>
<tr>
<th>ID_1</th>
<th>ID_2</th>
<th>RELATIONSHIP_ID</th>
<th>VALID_START_DATE</th>
<th>VALID_END_DATE</th>
<th>INVALID_REASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>44821957</td>
<td>21001551</td>
<td>ICD9CM - FDB Ind</td>
<td>01-Oct-13</td>
<td>31-Dec-2099</td>
<td></td>
</tr>
<tr>
<td>44821957</td>
<td>35204953</td>
<td>ICD9CM - MedDRA</td>
<td>01-Jan-70</td>
<td>31-Dec-2099</td>
<td></td>
</tr>
<tr>
<td>44821957</td>
<td>44824248</td>
<td>Is a</td>
<td>01-Oct-14</td>
<td>31-Dec-2099</td>
<td></td>
</tr>
<tr>
<td>44821957</td>
<td>44834731</td>
<td>Is a</td>
<td>01-Oct-14</td>
<td>31-Dec-2099</td>
<td></td>
</tr>
<tr>
<td>44821957</td>
<td>313217</td>
<td>Maps to</td>
<td>01-Jan-70</td>
<td>31-Dec-2099</td>
<td></td>
</tr>
</tbody>
</table>

ICD-9 is not a Standard Concept

Mapping to different vocabularies

Kind of relationship
### Official languages of the EU

<table>
<thead>
<tr>
<th>What is it?</th>
<th>The European Union has 24 official and working languages. They are:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bulgarian  French  Maltese</td>
</tr>
<tr>
<td></td>
<td>Croatian  German  Polish</td>
</tr>
<tr>
<td></td>
<td>Czech  Greek  Portuguese</td>
</tr>
<tr>
<td></td>
<td>Danish  Hungarian  Romanian</td>
</tr>
<tr>
<td></td>
<td>Dutch  Irish  Slovak</td>
</tr>
<tr>
<td></td>
<td>English  Italian  Slovenian</td>
</tr>
<tr>
<td></td>
<td>Estonian  Latvian  Spanish</td>
</tr>
<tr>
<td></td>
<td>Finnish  Lithuanian  Swedish</td>
</tr>
</tbody>
</table>

### 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?

Cheers
Mapping = Translating

Step 1. Find the Source Concept

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

<table>
<thead>
<tr>
<th>CONCEPT_ID</th>
<th>CONCEPT_NAME</th>
<th>DOMAIN_ID</th>
<th>VOCABULARY_ID</th>
<th>CONCEPT_CLASS_ID</th>
<th>STANDARD_CONCEPT</th>
<th>CONCEPT_CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>44821957</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>ICD9CM</td>
<td>5-digit billing code</td>
<td></td>
<td>427.31</td>
</tr>
</tbody>
</table>

Step 2. Translate

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

<table>
<thead>
<tr>
<th>CONCEPT_ID_1</th>
<th>CONCEPT_ID_2</th>
<th>RELATIONSHIP_ID</th>
<th>VALID_START_DATE</th>
<th>VALID_END_DATE</th>
<th>INVALID_REASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>44821957</td>
<td>313217</td>
<td>Maps to</td>
<td>01-Jan-1970</td>
<td>31-Dec-2099</td>
<td></td>
</tr>
</tbody>
</table>

Step 3. Check out the translated Concept

```
SELECT * FROM concept WHERE concept_id = 313217;
```
Exercise: Find Standard Concept ID from Source Concept

ICD-9: '427.31' : 313217
Read: 'G573000' : 313217
ICD-10: 'I48.0' : 4154290 'Paroxysmal Atrial Fibrillation'
Codes Used in the World

• Conditions
  – READ, OXMIS, ICD-9-CM, ICD-10-CM, ICPC, MedDRA, free text in different languages

• Drugs
  – Multilex, dm+d, BDPM, AMIS, AMT, ATC, NPI, NDC, free text in different languages
Reason #2: Disease Hierarchy

- Disease of the cardiovascular system
- Heart disease
- Cardiac arrhythmia
- Supraventricular arrhythmia
- Fibrillation
- Atrial arrhythmia
- Atrial fibrillation
  - Controlled atrial fibrillation
  - Persistent atrial fibrillation
  - Chronic atrial fibrillation
  - Paroxysmal atrial fibrillation
  - Rapid atrial fibrillation
  - Permanent atrial fibrillation
SELECT
*  
FROM
  concept_relationship
WHERE
  concept_id_1 = 313217;

<table>
<thead>
<tr>
<th>CONCEPT_ID_1</th>
<th>CONCEPT_ID_2</th>
<th>RELATIONSHIP_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>313217</td>
<td>4232697</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>4181800</td>
<td>Focus of</td>
</tr>
<tr>
<td>313217</td>
<td>35204953</td>
<td>SNOMED - MedDRA eq</td>
</tr>
<tr>
<td>313217</td>
<td>4203375</td>
<td>Asso finding of</td>
</tr>
<tr>
<td>313217</td>
<td>4141360</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>4119601</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>4117112</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>4232691</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>4139517</td>
<td>Due to of</td>
</tr>
<tr>
<td>313217</td>
<td>4194288</td>
<td>Asso finding of</td>
</tr>
<tr>
<td>313217</td>
<td>44782442</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>44783731</td>
<td>Focus of</td>
</tr>
<tr>
<td>313217</td>
<td>21003018</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>40248987</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>21001540</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>45576876</td>
<td>Mapped from</td>
</tr>
<tr>
<td>313217</td>
<td>44807374</td>
<td>Asso finding of</td>
</tr>
<tr>
<td>313217</td>
<td>21013834</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>21001572</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>21001606</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>21003176</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>4226399</td>
<td>Is a</td>
</tr>
<tr>
<td>313217</td>
<td>500001801</td>
<td>SNOMED - HOI</td>
</tr>
<tr>
<td>313217</td>
<td>500002401</td>
<td>SNOMED - HOI</td>
</tr>
<tr>
<td>313217</td>
<td>4119602</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>40631039</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>4108832</td>
<td>Subsumes</td>
</tr>
<tr>
<td>313217</td>
<td>21013671</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>21013390</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>313217</td>
<td>Maps to</td>
</tr>
<tr>
<td>313217</td>
<td>44821957</td>
<td>Mapped from</td>
</tr>
<tr>
<td>313217</td>
<td>2617597</td>
<td>Mapped from</td>
</tr>
<tr>
<td>313217</td>
<td>45500085</td>
<td>Mapped from</td>
</tr>
<tr>
<td>313217</td>
<td>313217</td>
<td>Mapped from</td>
</tr>
<tr>
<td>313217</td>
<td>45951191</td>
<td>Mapped from</td>
</tr>
<tr>
<td>313217</td>
<td>21013856</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>21001575</td>
<td>SNOMED - ind/CI</td>
</tr>
<tr>
<td>313217</td>
<td>21001594</td>
<td>SNOMED - ind/CI</td>
</tr>
</tbody>
</table>
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;
```

<table>
<thead>
<tr>
<th>relationship_id</th>
<th>concept_id</th>
<th>concept_name</th>
<th>domain_id</th>
<th>vocabulary_id</th>
<th>concept_class_id</th>
<th>standard_concept</th>
<th>concept_code</th>
<th>valid_start_date</th>
<th>valid_end_date</th>
<th>invalid_reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asso finding of</td>
<td>4203375</td>
<td>Family history</td>
<td>Observation</td>
<td>SNOMED</td>
<td>Context-dependent</td>
<td>S</td>
<td>433276002</td>
<td>31-Jan-2009</td>
<td>31-Dec-2009</td>
<td></td>
</tr>
<tr>
<td>Asso finding of</td>
<td>44807374</td>
<td>Atrial fibrillation</td>
<td>Observation</td>
<td>SNOMED</td>
<td>Context-dependent</td>
<td>S</td>
<td>816401000000</td>
<td>1-Apr-2014</td>
<td>31-Dec-2009</td>
<td></td>
</tr>
<tr>
<td>Due to of</td>
<td>4135917</td>
<td>Transient cerebral ischemia due to atrial fibrillation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>426814001</td>
<td>1-Jan-1970</td>
<td>31-Dec-1970</td>
<td></td>
</tr>
<tr>
<td>Focus of</td>
<td>42709991</td>
<td>Insertion of pacemaker for control of atrial fibrillation</td>
<td>Procedure</td>
<td>SNOMED</td>
<td>Procedure</td>
<td>S</td>
<td>449863006</td>
<td>31-Jan-2012</td>
<td>31-Dec-2008</td>
<td></td>
</tr>
<tr>
<td>Focus of</td>
<td>4181800</td>
<td>Maze procedure for atrial fibrillation</td>
<td>Procedure</td>
<td>SNOMED</td>
<td>Procedure</td>
<td>S</td>
<td>429211003</td>
<td>31-Jan-2008</td>
<td>31-Dec-2008</td>
<td></td>
</tr>
<tr>
<td>Focus of</td>
<td>44783781</td>
<td>Provision of written information about atrial fibrillation</td>
<td>Procedure</td>
<td>SNOMED</td>
<td>Procedure</td>
<td>S</td>
<td>699833006</td>
<td>31-Jan-2014</td>
<td>31-Dec-2009</td>
<td></td>
</tr>
<tr>
<td>Has finding site</td>
<td>4242142</td>
<td>Atrial structure</td>
<td>Spec Anatomic Site</td>
<td>SNOMED</td>
<td>Body Structure</td>
<td>S</td>
<td>59652004</td>
<td>1-Jan-1970</td>
<td>31-Dec-1970</td>
<td></td>
</tr>
<tr>
<td>Has finding site</td>
<td>4093357</td>
<td>Cardiac conducting system structure</td>
<td>Spec Anatomic Site</td>
<td>SNOMED</td>
<td>Body Structure</td>
<td>S</td>
<td>24994005</td>
<td>1-Jan-1970</td>
<td>31-Dec-1970</td>
<td></td>
</tr>
<tr>
<td>Is a</td>
<td>4226399</td>
<td>Fibrillation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>40593004</td>
<td>1-Jan-1970</td>
<td>31-Dec-1970</td>
<td></td>
</tr>
<tr>
<td>Is a</td>
<td>4068155</td>
<td>Atrial arrhythmia</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>17366009</td>
<td>1-Jan-1970</td>
<td>31-Dec-1970</td>
<td></td>
</tr>
</tbody>
</table>

**Ancestor concepts**
- Asso finding of 4203375: Family history
- Asso finding of 44807374: Atrial fibrillation
- Asso finding of 4194258: H/O: atrial fibrillation
- Due to of 4135917: Transient cerebral ischemia due to atrial fibrillation
- Focus of 42709991: Insertion of pacemaker for control of atrial fibrillation
- Focus of 4181800: Maze procedure for atrial fibrillation
- Focus of 44783781: Provision of written information about atrial fibrillation

**Descendant concepts**
- Asso finding of 4203375: Family history
- Asso finding of 44807374: Atrial fibrillation
- Asso finding of 4194258: H/O: atrial fibrillation
- Due to of 4135917: Transient cerebral ischemia due to atrial fibrillation
- Focus of 42709991: Insertion of pacemaker for control of atrial fibrillation
- Focus of 4181800: Maze procedure for atrial fibrillation
- Focus of 44783781: Provision of written information about atrial fibrillation

**Find out related concept**
Ancestry Relationships: Higher-Level Relationships

- Ancestry
- Concepts
- Concept Relationships
- Descendant

- Ancestor
- Heart disease
- Cardiac arrhythmia
- Supraventricular arrhythmia
- Fibrillation
- Atrial arrhythmia
- Atrial fibrillation
- Disease of the cardiovascular system

- Controlled atrial fibrillation
- Persistent atrial fibrillation
- Chronic atrial fibrillation
- Paroxysmal atrial fibrillation
- Rapid atrial fibrillation
- Permanent atrial fibrillation

5 levels of separation

2 levels of separation
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;
```

<table>
<thead>
<tr>
<th>max_levels_of_separation</th>
<th>concept_id</th>
<th>concept_name</th>
<th>domain_i</th>
<th>vocabulary</th>
<th>concept_class</th>
<th>standard_concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>313217</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>0</td>
<td>35204953</td>
<td>Atrial fibrillation</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
</tr>
<tr>
<td>1</td>
<td>4226399</td>
<td>Fibrillation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4068155</td>
<td>Atrial arrhythmia</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>35204969</td>
<td>Cardiac fibrillation</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>448028</td>
<td>Supraventricular arrhythmia</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>2</td>
<td>35204952</td>
<td>Arrhythmia supraventricular</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>35202454</td>
<td>Rate and rhythm disorders NEC</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>44784217</td>
<td>Cardiac arrhythmia</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>3</td>
<td>35202455</td>
<td>Supraventricular arrhythmias</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>321588</td>
<td>Heart disease</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>35204989</td>
<td>Cardiac disorder</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>35202050</td>
<td>Cardiac arrhythmias</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>4103183</td>
<td>Cardiac finding</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>440142</td>
<td>Disorder of mediastinum</td>
<td>Condition</td>
<td>SNOMED</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>134057</td>
<td>Disorder of cardiovascular system</td>
<td>Condition</td>
<td>SNOMED</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>35204998</td>
<td>Cardiovascular disorder</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>37219970</td>
<td>Mediastinal disorder</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>37622411</td>
<td>Phleboclesis</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>35202457</td>
<td>Cardiac disorders NEC</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>4115390</td>
<td>Mediastinal finding</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>6</td>
<td>4023995</td>
<td>Cardiovascular finding</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
</tbody>
</table>

Hold the descendant

Query Concept

Standard Concepts

Query Concept
Exploring Descendants of a Concept

```
SELECT max_levels_of_separation, c.*
FROM concept_ancestor ca, concept c
WHERE ca.ancestor_concept_id = 44784217 /* cardiac arrhythmia */
  AND ca.descendant_concept_id = c.concept_id
ORDER BY max_levels_of_separation
```

<table>
<thead>
<tr>
<th>MAX_LEVELS_OF_SEPARATION</th>
<th>CONCEPT_ID</th>
<th>CONCEPT_NAME</th>
<th>DOMAIN_ID</th>
<th>VOCABULARY_ID</th>
<th>CONCEPT_CLASS_ID</th>
<th>STANDARD_CONCEPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>44784217</td>
<td>Cardiac arrhythmia</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>313224</td>
<td>Anomalous atrioventricular excitation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>315643</td>
<td>Tachyarrhythm</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>316429</td>
<td>Premature beats</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>316999</td>
<td>Conduction disorder of the heart</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>321042</td>
<td>Cardiac arrest</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4030583</td>
<td>Pacemaker twiddler's syndrome</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4057008</td>
<td>Accelerated atrioventricular conduction</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4086313</td>
<td>Withdrawal arrhythmia</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4088507</td>
<td>Ventricular escape complex</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4088986</td>
<td>Atrial escape complex</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4091901</td>
<td>Aberrant premature complexes</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4092011</td>
<td>Aberrantly conducted complex</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4124704</td>
<td>Postoperative sinoatrial disease</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4143042</td>
<td>Ectopic beats</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4164083</td>
<td>Ectopic rhythm</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4172863</td>
<td>Fetal dysrhythm</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4173170</td>
<td>Neonatal dysrhythm</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4175473</td>
<td>Atrioventricular dissociation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4185572</td>
<td>Ventricular arrhythmia</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4217221</td>
<td>Nodal rhythm disorder</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4226399</td>
<td>Fibrillation</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>4228448</td>
<td>Bradyarrhythm</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
</tr>
</tbody>
</table>
Let's find Upper Gastrointestinal Bleeding

1. Find some initiation concept

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

<table>
<thead>
<tr>
<th>concept_id</th>
<th>concept_name</th>
<th>domain_id</th>
<th>vocabulary_id</th>
<th>concept_class_id</th>
<th>standard_concept</th>
<th>concept_code</th>
</tr>
</thead>
<tbody>
<tr>
<td>42891225</td>
<td>Upper gastrointestinal bleeding</td>
<td>Condition</td>
<td>MedDRA</td>
<td>LLT</td>
<td>C</td>
<td>10071910</td>
</tr>
</tbody>
</table>

2. Find standard concepts

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

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

<table>
<thead>
<tr>
<th>max_levels</th>
<th>concept_id</th>
<th>concept_name</th>
<th>domain_id</th>
<th>vocabulary_id</th>
<th>concept_class_id</th>
<th>standard_concept</th>
<th>concept_code</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4332645</td>
<td>Upper gastrointestinal hemorrhage associated with hypercoag</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>430349003</td>
</tr>
<tr>
<td>1</td>
<td>35708054</td>
<td>Gastritis haemorrhagic</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
<td>10017866</td>
</tr>
<tr>
<td>1</td>
<td>4291649</td>
<td><strong>Upper gastrointestinal hemorrhage</strong></td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
<td>10046274</td>
</tr>
<tr>
<td>2</td>
<td>35707864</td>
<td>Gastrointestinal haemorrhage</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
<td>10017955</td>
</tr>
<tr>
<td>2</td>
<td>4000609</td>
<td>Disorder of upper gastrointestinal tract</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>119291004</td>
</tr>
<tr>
<td>2</td>
<td>35707858</td>
<td>Intestinal haemorrhage</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
<td>10059175</td>
</tr>
<tr>
<td>2</td>
<td>35702752</td>
<td>Gastritis (excl infective)</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
<td>10017854</td>
</tr>
<tr>
<td>2</td>
<td>192671</td>
<td>Gastrointestinal hemorrhage</td>
<td>Condition</td>
<td>MedDRA</td>
<td>Clinical Finding</td>
<td>S</td>
<td>74474003</td>
</tr>
<tr>
<td>3</td>
<td>37604042</td>
<td>Gastrointestinal haemorrhages</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
<td>10052742</td>
</tr>
<tr>
<td>3</td>
<td>37622518</td>
<td>Haemorrhage</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
<td>10055798</td>
</tr>
<tr>
<td>3</td>
<td>437312</td>
<td>Bleeding</td>
<td>Condition</td>
<td>MedDRA</td>
<td>Clinical Finding</td>
<td>S</td>
<td>131148009</td>
</tr>
<tr>
<td>3</td>
<td>4198525</td>
<td>Disorder of upper digestive tract</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>50410009</td>
</tr>
<tr>
<td>3</td>
<td>37622515</td>
<td>Extravasation blood</td>
<td>Condition</td>
<td>MedDRA</td>
<td>PT</td>
<td>C</td>
<td>10015867</td>
</tr>
<tr>
<td>3</td>
<td>4000610</td>
<td>Disorder of gastrointestinal tract</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>119292006</td>
</tr>
<tr>
<td>3</td>
<td>35702116</td>
<td>Gastrointestinal inflammatory conditions</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
<td>10017969</td>
</tr>
<tr>
<td>3</td>
<td>35702743</td>
<td>Intestinal haemorrhages</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
<td>10022653</td>
</tr>
<tr>
<td>3</td>
<td>35702744</td>
<td>Non-site specific gastrointestinal haemorrhages</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
<td>10017958</td>
</tr>
<tr>
<td>4</td>
<td>35702114</td>
<td>Gastrointestinal haemorrhages NEC</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLGT</td>
<td>C</td>
<td>10017959</td>
</tr>
<tr>
<td>4</td>
<td>4304916</td>
<td>Gastrointestinal tract finding</td>
<td>Condition</td>
<td>SNOMED</td>
<td>Clinical Finding</td>
<td>S</td>
<td>386618008</td>
</tr>
<tr>
<td>4</td>
<td>35702767</td>
<td>Nausea and vomiting symptoms</td>
<td>Condition</td>
<td>MedDRA</td>
<td>HLT</td>
<td>C</td>
<td>10028817</td>
</tr>
</tbody>
</table>

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.ancestor_concept_id = 4291649 /* Upper gastrointestinal hemorrhage */
    AND ca.descendant_concept_id = c.concept_id
ORDER BY max_levels_of_separation;

Going down: Checking the right content
Concept 4291649 and all its descendants comprise Upper GI Bleeding
Lunch

In Glen Foyer, please return in 1 hour
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

Classifications
- VA Class
- CVX
- NDFRT
- NDFRT Ind
- ATC
- FDB Ind
- ETC
- SPL
- SNOMED

Drugs
- Source codes
- CIEL
- NDC
- GPI
- VA-Product
- Gemscript
- EU Product
- DPD
- HCPCS
- MeSH
- Multum
- Oxmis
- Read
- Genseqno
- dm+d
- AMIS
- BDPM
- CPT4

Ingredients
- Standard Drug Vocabulary:
  - RxNorm
  - RxNorm Extension

Drug products
- Drug Codes
- Procedure Drugs
Let's find Warfarin

1. Find active compound Warfarin by keyword

```sql
SELECT * FROM concept WHERE lower(concept_name) = 'warfarin';
```
Let's find Clopidogrel

1. Find drug product containing Clopidogrel by NDC code:
   Bristol Meyer Squibb's Plavix 75mg capsules: NDC 67544050474

   SELECT * FROM concept WHERE concept_code = '67544050474';

<table>
<thead>
<tr>
<th>concept_id</th>
<th>concept_name</th>
<th>domain_id</th>
<th>vocabulary_id</th>
<th>concept_class_id</th>
<th>standard_concept</th>
<th>concept_code</th>
<th>valid_start_date</th>
<th>valid_end_date</th>
<th>invalid_reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>45867731</td>
<td>clopidogrel 75 MG Oral Tablet [Plavix]</td>
<td>Drug</td>
<td>NDC</td>
<td>11-digit NDC</td>
<td>NULL</td>
<td>67544050474</td>
<td>2014-07-01</td>
<td>2099-12-31</td>
<td>NULL</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>concept_id_1</th>
<th>concept_id_2</th>
<th>relationship_id</th>
<th>valid_start_date</th>
<th>valid_end_date</th>
<th>invalid_reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>45867731</td>
<td>1322185</td>
<td>Maps to</td>
<td>2015-01-29</td>
<td>2099-12-31</td>
<td>NULL</td>
</tr>
</tbody>
</table>

   SELECT * FROM concept WHERE concept_id = 1322185;

<table>
<thead>
<tr>
<th>concept_id</th>
<th>concept_name</th>
<th>domain_id</th>
<th>vocabulary_id</th>
<th>concept_class_id</th>
<th>standard_concept</th>
<th>concept_code</th>
<th>valid_start_date</th>
<th>valid_end_date</th>
<th>invalid_reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1322185</td>
<td>clopidogrel 75 MG Oral Tablet [Plavix]</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Branded Drug</td>
<td>S</td>
<td>213169</td>
<td>1970-01-01</td>
<td>2099-12-31</td>
<td>NULL</td>
</tr>
</tbody>
</table>
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;
```
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;
```

<table>
<thead>
<tr>
<th>concept_id</th>
<th>concept_name</th>
<th>vocabulary_id</th>
<th>concept_class_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1310149</td>
<td>Warfarin</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>36221229</td>
<td>Jantoven Pill</td>
<td>RxNorm</td>
<td>Branded Dose Group</td>
</tr>
<tr>
<td>40163559</td>
<td>Warfarin Sodium 6 MG</td>
<td>RxNorm</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>40163544</td>
<td>Warfarin Sodium 3 MG</td>
<td>RxNorm</td>
<td>Branded Dose Group</td>
</tr>
<tr>
<td>21134746</td>
<td>Warfarin 0.2 MG/ML</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>21105414</td>
<td>Warfarin 5 MG/ML</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>36221228</td>
<td>Jantoven Oral Product</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>40163555</td>
<td>Warfarin Sodium 7.5 MG</td>
<td>RxNorm</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>21115236</td>
<td>Warfarin 0.3 MG/ML</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>40163509</td>
<td>Warfarin Sodium 1 MG</td>
<td>RxNorm</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>21156284</td>
<td>1 ML Warfarin 0.02 MG/ML Oral Solution</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
<tr>
<td>21095537</td>
<td>Warfarin 0.3 MG/ML Oral Solution</td>
<td>RxNorm Extension</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>21105427</td>
<td>Warfarin 0.4 MG/ML Oral Solution</td>
<td>RxNorm Extension</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>21064557</td>
<td>Warfarin 1 MG/ML Oral Solution</td>
<td>RxNorm Extension</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>40093133</td>
<td>Warfarin Oral Tablet [Coumadin]</td>
<td>RxNorm</td>
<td>Branded Drug Form</td>
</tr>
<tr>
<td>40093134</td>
<td>Warfarin Oral Tablet [Jantoven]</td>
<td>RxNorm</td>
<td>Branded Drug Form</td>
</tr>
<tr>
<td>21077698</td>
<td>1 ML Warfarin 1 MG Oral Solution</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
<tr>
<td>40163594</td>
<td>Warfarin Sodium 2.5 MG Oral Tablet</td>
<td>RxNorm</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>40163530</td>
<td>Warfarin Sodium 2 MG/ML Injectable Solution</td>
<td>RxNorm Extension</td>
<td>Branded Drug</td>
</tr>
<tr>
<td>21066136</td>
<td>Warfarin 5 MG Oral Tablet [Marevan]</td>
<td>RxNorm Extension</td>
<td>Branded Drug</td>
</tr>
<tr>
<td>40163542</td>
<td>Warfarin Sodium 3 MG Oral Tablet [Jantoven]</td>
<td>RxNorm</td>
<td>Branded Drug</td>
</tr>
<tr>
<td>21116822</td>
<td>1 ML Warfarin 0.6 MG/ML Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
<tr>
<td>21175784</td>
<td>1 ML Warfarin 0.1 MG/ML Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
<tr>
<td>21175783</td>
<td>1 ML Warfarin 0.832 MG/ML Oral Solution</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
<tr>
<td>1322184</td>
<td>clopidogrel</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>21043471</td>
<td>clopidogrel Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Form</td>
</tr>
<tr>
<td>36229332</td>
<td>Plavix Pill</td>
<td>RxNorm</td>
<td>Branded Dose Group</td>
</tr>
<tr>
<td>21043470</td>
<td>clopidogrel Oral Solution</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Form</td>
</tr>
<tr>
<td>21023802</td>
<td>clopidogrel Injectable Solution</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Form</td>
</tr>
<tr>
<td>21023806</td>
<td>clopidogrel 5 MG</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>1322187</td>
<td>clopidogrel 75 MG</td>
<td>RxNorm</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>21141600</td>
<td>clopidogrel 1 MG/ML</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>36222254</td>
<td>clopidogrel Oral Product</td>
<td>RxNorm</td>
<td>Clinical Dose Group</td>
</tr>
<tr>
<td>21092477</td>
<td>clopidogrel 5 MG/ML</td>
<td>RxNorm Extension</td>
<td>Clinical Drug Comp</td>
</tr>
<tr>
<td>21177192</td>
<td>100 ML clopidogrel 1 MG/ML Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
<tr>
<td>21047899</td>
<td>1 ML clopidogrel 5 MG/ML Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
<tr>
<td>21121870</td>
<td>clopidogrel 5 MG/ML Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>21063106</td>
<td>clopidogrel 75 MG Oral Tablet [Grepid]</td>
<td>RxNorm Extension</td>
<td>Branded Drug</td>
</tr>
<tr>
<td>1322190</td>
<td>clopidogrel 300 MG Oral Tablet [Plavix]</td>
<td>RxNorm</td>
<td>Branded Drug</td>
</tr>
<tr>
<td>21121869</td>
<td>clopidogrel 75 MG Injectable Solution</td>
<td>RxNorm Extension</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>21053280</td>
<td>clopidogrel 6 MG Injectable Solution</td>
<td>RxNorm Extension</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>21023810</td>
<td>clopidogrel 4 MG Injectable Solution</td>
<td>RxNorm Extension</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>21106783</td>
<td>1 ML clopidogrel 1 MG/ML Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
<tr>
<td>19075601</td>
<td>clopidogrel 75 MG Oral Tablet</td>
<td>RxNorm</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>21102364</td>
<td>clopidogrel 1 MG/ML Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>40095879</td>
<td>clopidogrel Oral Tablet [Plavix]</td>
<td>RxNorm</td>
<td>Branded Drug Form</td>
</tr>
<tr>
<td>40095878</td>
<td>clopidogrel Oral Tablet</td>
<td>RxNorm</td>
<td>Clinical Drug</td>
</tr>
<tr>
<td>21088717</td>
<td>100 ML clopidogrel 15 MG/ML Oral Suspension</td>
<td>RxNorm Extension</td>
<td>Quant Clinical Drug</td>
</tr>
</tbody>
</table>
Find members of Drug Classes

4. Check Ingredient Descendants of Drug Class Anticoagulants

```
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;
```

<table>
<thead>
<tr>
<th>concept_id</th>
<th>concept_name</th>
<th>domain_id</th>
<th>vocabulary_id</th>
<th>concept_class_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1344992</td>
<td>Iloprost</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>19084670</td>
<td>Bivalirudin</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>19069137</td>
<td>Bemiparin</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1315865</td>
<td>Fondaparinux</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1350310</td>
<td>Clopidogrel</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>19026343</td>
<td>Danaparoid</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>40163718</td>
<td>Prasugrel</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>19098548</td>
<td>Tenecteplase</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1322180</td>
<td>Clopidogrel</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1367571</td>
<td>Heparin</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1310149</td>
<td>Warfarin</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1308473</td>
<td>Tinzaparin</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1322199</td>
<td>Eptifibatide</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>40241331</td>
<td>Rivaroxaban</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>19024191</td>
<td>Reteplose</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1301065</td>
<td>Dalteparin</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>1731597</td>
<td>Drotrecogin alfa</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>35594848</td>
<td>Ecolaxin cog</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
<tr>
<td>25690488</td>
<td>Ecolaxin og</td>
<td>Drug</td>
<td>RxNorm</td>
<td>Ingredient</td>
</tr>
</tbody>
</table>
Exercise:
Find Standard Concept ID

• Metformin  1503297
• Tolazamide  1502809
• Telmisartan  1317640
• Your favorite ingredient here
Exercise:
Find Standard Concept ID

• 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
- Payer_plan_period
- Cost

Standardized health economics:
- Cohort
- Cohort_attribute
- Condition_era
- Drug_era
- Dose_era

Standardized meta-data:
- CDM_source
- Concept
- Vocabulary
- Domain
- Concept_class
- Concept_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
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description and purpose</th>
<th>Field name convention</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient centric</td>
<td>Every domain table has <strong>patient identifier</strong>. Patient data can be retrieved independently from other domains.</td>
<td>person_id</td>
<td>person_id 123</td>
</tr>
<tr>
<td>Unique domain identifier</td>
<td>Every domain table has a unique primary key to identify domain <strong>entities</strong></td>
<td>&lt;entity&gt;_id</td>
<td>condition_occurrence_id 470985</td>
</tr>
<tr>
<td>Standard concept from a respective vocabulary domain</td>
<td>Integration with the vocabulary. Foreign key into the Standard Vocabulary for <strong>Standard Concept</strong></td>
<td>&lt;entity&gt;_concept_id</td>
<td>condition_concept_id 313217 (SNOMED &quot;Atrial Fibrillation&quot;)</td>
</tr>
<tr>
<td>Source concept from a respective vocabulary domain</td>
<td>Provenance. Foreign key into the Standard Vocabulary for <strong>Source Concept</strong></td>
<td>&lt;entity&gt;_source_concept_id</td>
<td>condition_source_concept_id 44821957 (ICD9CM &quot;Atrial Fibrillation&quot;)</td>
</tr>
<tr>
<td>Source value</td>
<td>Provenance. Verbatim information from the source data, <strong>not to be used</strong> by any standard analytics</td>
<td>&lt;entity&gt;_source_value</td>
<td>condition_source_value 427.31 (ICD9CM &quot;Atrial Fibrillation&quot;)</td>
</tr>
<tr>
<td>Source type</td>
<td>Provenance. Foreign key into the Vocabulary for the <strong>origin of the</strong></td>
<td>&lt;entity&gt;_type_concept_id</td>
<td>condition_type_concept_id 38000199 (&quot;Inpatient header – primary&quot;)</td>
</tr>
</tbody>
</table>
PERSON

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

• Visit Types
  – Inpatient
  – Emergency room
  – Inpatient/Emergency - **new**
  – Outpatient
  – Long-term care

• Vocabulary: OMOP

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

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

- Procedures have the least standardized vocabularies that causes some redundancy
CONDITION_OCCURRENCE

- 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

<table>
<thead>
<tr>
<th>Description</th>
<th>AUT MESUR CHIMIO PROPHYL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIM10</td>
<td>0Z2920</td>
</tr>
<tr>
<td>CIM10 Description</td>
<td>OTH PROPHYLACTIC CHEMO</td>
</tr>
<tr>
<td>Maps to</td>
<td>DRUG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>EXSPEC DEPS AUT MAL PREC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIM10</td>
<td>0Z1380</td>
</tr>
<tr>
<td>CIM10 Description</td>
<td>SPEC SCR OTHER SPEC DIS</td>
</tr>
<tr>
<td>Maps to</td>
<td>OBSERVATION</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>GRSS CONSTT FORTUITMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIM10</td>
<td>0Z3300</td>
</tr>
<tr>
<td>CIM10 Description</td>
<td>PREG STATE INCIDENT</td>
</tr>
<tr>
<td>Maps to</td>
<td>CONDITION</td>
</tr>
</tbody>
</table>
DRUG_EXPOSURE

• 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

<table>
<thead>
<tr>
<th>Drug Source Description</th>
<th>Form Desc</th>
<th>Admin Route Description</th>
<th>Generic Name</th>
<th>Maps To</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBSERVATION PERIOD</td>
<td>Miscellaneous</td>
<td>Unspecified</td>
<td>Documentation</td>
<td>OBSERVATION</td>
</tr>
<tr>
<td>LUMBAR DDS BELT</td>
<td>Miscellaneous</td>
<td>Unspecified</td>
<td>Unspecified</td>
<td>DEVICE</td>
</tr>
<tr>
<td>JOBST KNEE HIGH COMPRESSION STOCKING</td>
<td>Miscellaneous</td>
<td>Unspecified</td>
<td>Antiembolism stockings</td>
<td>DEVICE</td>
</tr>
<tr>
<td>MASKS</td>
<td>Miscellaneous</td>
<td>Unspecified</td>
<td>Masks</td>
<td>DEVICE</td>
</tr>
<tr>
<td>PEN NEEDLES</td>
<td>Miscellaneous</td>
<td>Unspecified</td>
<td>Needle</td>
<td>DEVICE</td>
</tr>
</tbody>
</table>
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

• 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')
```
Measurement Data Issues

```sql
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
```
OBSERVATION

• 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

- To capture biomarkers / tissue bank
NOTE

• To capture unstructured free text
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

- 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.)
OMOP CDM Service Tables

• **CDM_SOURCE**
  – Provenance, integration, metadata
  – Future extension to individual domains

• **FACT_RELATIONSHIP**
  – Linkage between related observations
  – Example: systolic and diastolic blood pressure
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_DEFINITION** 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

- 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

NDC: 00310013010
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]

How do we handle reversals?

How do we handle NDC change?

How do we handle overlap?

How do we handle change in dose?

How do we handle gaps?

How do we handle combination products?

How do we handle discontinuation?

How do we handle gap?
ETL: Real world scenario

PharMetrics Plus
CLAIMS

<table>
<thead>
<tr>
<th>pat_id</th>
<th>claimno</th>
<th>from_dt</th>
<th>to_dt</th>
<th>diagprc_ind</th>
<th>Diag_admit</th>
<th>diag1</th>
</tr>
</thead>
<tbody>
<tr>
<td>05917921689</td>
<td>IPA333393946</td>
<td>1/5/2006</td>
<td>1/5/2006</td>
<td>1</td>
<td>41071</td>
<td>41071</td>
</tr>
</tbody>
</table>

LRx/Dx
MEDICAL_CLAIMS

<table>
<thead>
<tr>
<th>md_clm_id</th>
<th>ims_pat_nbr</th>
<th>dt_of_service</th>
<th>rxer_id</th>
<th>diag_cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>95963982102</td>
<td>80445908</td>
<td>8/1/2012 0:00</td>
<td>680488</td>
<td>41071</td>
</tr>
</tbody>
</table>

German DA
Problem Events

<table>
<thead>
<tr>
<th>db_country</th>
<th>international_practice_num</th>
<th>international_doctor_num</th>
<th>international_patient_num</th>
<th>age_at_event</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE</td>
<td>GE6326</td>
<td>GE8784</td>
<td>GE46478747</td>
<td>20</td>
</tr>
</tbody>
</table>

Diagnosis

<table>
<thead>
<tr>
<th>db_country</th>
<th>international_diagnosis_num</th>
<th>diagnosis_num</th>
<th>icd10_4_code</th>
<th>icd10_3_text</th>
<th>diagnosis_confi</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE</td>
<td>GE2397573</td>
<td>2397573</td>
<td>I21.4</td>
<td>Non-ST elevation myocardial infarction</td>
<td>Confirmed</td>
</tr>
</tbody>
</table>

Ambulatory EMR
Problem

<table>
<thead>
<tr>
<th>Patient_id_synth</th>
<th>Diag_dt</th>
<th>icd10_cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>271138</td>
<td>4/11/2013</td>
<td>I214</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>pat_id</th>
<th>claimno</th>
<th>from_dt</th>
<th>to_dt</th>
<th>diagprc_ind</th>
<th>Diag_admit</th>
</tr>
</thead>
<tbody>
<tr>
<td>05917921689</td>
<td>IPA333393946</td>
<td>1/5/2006</td>
<td>1/5/2006</td>
<td>1</td>
<td>41071</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>05917921689</td>
<td>1/5/2006</td>
<td>41071</td>
<td>Inpatient claims - primary position</td>
</tr>
<tr>
<td>05917921689</td>
<td>1/5/2006</td>
<td>41071</td>
<td>Inpatient claims - 1st position</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
<th>CONDITION_SOURCE_CONCEPT_ID</th>
<th>CONDITION_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>05917921689</td>
<td>1/5/2006</td>
<td>41071</td>
<td>Inpatient claims - primary position</td>
<td>44825429</td>
<td>444406</td>
</tr>
</tbody>
</table>
OMOP CDM = Standardized structure:
same tables, same fields, same datatypes,
same conventions across disparate sources

### PharMetrics Plus: CONDITION_OCCURRENCE

<table>
<thead>
<tr>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>157033702</td>
<td>1/5/2006</td>
<td>41071</td>
<td>Inpatient claims - primary position</td>
</tr>
<tr>
<td>157033702</td>
<td>1/5/2006</td>
<td>41071</td>
<td>Inpatient claims - 1st position</td>
</tr>
</tbody>
</table>

### LRX/DX: CONDITION_OCCURRENCE

<table>
<thead>
<tr>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>80445908</td>
<td>8/1/2012</td>
<td>41071</td>
<td>Primary Condition</td>
</tr>
</tbody>
</table>

### German DA : CONDITION_OCCURRENCE

<table>
<thead>
<tr>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>46478747</td>
<td>11/19/2014</td>
<td>I21.4</td>
<td>EHR problem list entry</td>
</tr>
</tbody>
</table>

### Ambulatory EMR : CONDITION_OCCURRENCE

<table>
<thead>
<tr>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>271138</td>
<td>4/11/2013</td>
<td>I214</td>
<td>Primary Condition</td>
</tr>
</tbody>
</table>

- Consistent structure optimized for large-scale analysis
- Structure preserves all source content and provenance
OMOP CDM = Standardized content: common vocabularies across disparate sources

• 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

<table>
<thead>
<tr>
<th>PharMetrics Plus: CONDITION_OCCURRENCE</th>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
<th>CONDITION_SOURCE_CONCEPT_ID</th>
<th>CONDITION_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>05917921689</td>
<td>1/5/2006</td>
<td>41071</td>
<td>Inpatient claims - primary position</td>
<td>4482542</td>
<td>444406</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LRx/Dx: CONDITION_OCCURRENCE</th>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
<th>CONDITION_SOURCE_CONCEPT_ID</th>
<th>CONDITION_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>80445908</td>
<td>8/1/2012</td>
<td>41071</td>
<td>Primary Condition</td>
<td>4482542</td>
<td>444406</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>German DA : CONDITION_OCCURRENCE</th>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
<th>CONDITION_SOURCE_CONCEPT_ID</th>
<th>CONDITION_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>6478747</td>
<td>11/19/2014</td>
<td>121.4</td>
<td>EHR problem list entry</td>
<td>4557208</td>
<td>444406</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ambulatory EMR : CONDITION_OCCURRENCE</th>
<th>PERSON_ID</th>
<th>CONDITION_START_DATE</th>
<th>CONDITION_SOURCE_VALUE</th>
<th>CONDITION_TYPE_CONCEPT_ID</th>
<th>CONDITION_SOURCE_CONCEPT_ID</th>
<th>CONDITION_CONCEPT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>271138</td>
<td>4/11/2013</td>
<td>1214</td>
<td>Primary Condition</td>
<td>4557208</td>
<td>444406</td>
<td></td>
</tr>
</tbody>
</table>
Lesson Learned

• Date Shifting
  – Added logic to shift date of actual patient transactions

• Encrypt/De-identify Provider or Plan information within a link dataset
  – Encrypted provider ID information when linked claims with EMR dataset

• Privacy ICD9/10 Codes
  – Removal of ICD9/10 codes that are considered privacy issues, such as death or sexual abuse
  – Using “fake” date in Death table to indicate a death

• Data unable to leave a specific country

• Pilot Patients
  – Removal of patients that were are “dummy patients”
Lesson Learned (cont.)

• Patients without transaction
  – Adding an observation period
• Local country vocabulary mapping
• Local knowledge of each countries health system
• Knowledge of local data and business rules
• Ability to extract patient level data
• Cleaning dirty data
• Standardize measurement and unit of measure
• Source field not transferring to OMOP CDM
What makes OMOP CDM unique

- Specialized CDM - reflective of clinical domain, granular, well structured
- Vocabulary - uniformly structured and well curated
- Information Model - formalized connection between data model and conceptual model (Vocabulary)
- Specialized yet Extendable – new attributes and concepts can be added
- Supportive Community of developers and researchers
- Development driven by analytic use cases
Break

Please return in 15 minutes
CDM Examples

Leveraging OHDSI Tools
(GitHub /Forums/
Working Group)
Exercises
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](https://github.com/OHDSI/CommonDataModel)
  – Proposals tracked and discussed as Github issues

• Meeting information can be found on the working group [wiki page](https://github.com/OHDSI/CommonDataModel)

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

1. Download
   
   http://athena.ohdsi.org

2. Rebuild (not for the faint of heart)
   
   https://github.com/OHDSI/Vocabulary-v5.0

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

Data Used for Demonstration

- Five types of data:

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

SynPUF High Level Diagram

Beneficiary Summary

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

**SynPUF**

- Beneficiary Summary
  - Inpatient Claims
  - Outpatient Claims
  - Carrier Claims
  - Prescription Drug Events (PDE)

**CDM**

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

- Standardized health system data
  - Location
  - Care_site
  - Provider
  - Payer_plan_period
  - Location
  - Care_site
  - Provider

- Standardized health economics
  - Cohort
  - Cohort_attribute
  - Condition_era
  - Drug_era
  - Dose_era
Some Example Questions

Finding Warfarin

New Users of Warfarin

New Users of Warfarin who are >=65?

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?

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 concepts 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
OHDSI in a Box
CDM Database – Connect
CDM Database – Open SQL Sheet
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) */);
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*/;

/*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)*/
);
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.
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 period.
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

Finding Warfarin

New Users of Warfarin

New Users of Warfarin who are >=65?

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.
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
Step 1: Start with the previous query

```sql
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

```sql
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,
       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!

```sql
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
```
Some Example Questions

Finding Warfarin

New Users of Warfarin

New Users of Warfarin
who are >=65?

New Users of Warfarin
with prior Atrial Fibrillation?
How do I define new users of Warfarin with prior Atrial Fibrillation?
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
Step 1: Start with the Ex 1 query

/********************
 * (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
    WHERE i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
    )

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
    WHERE co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_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

```sql
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 3: Prior Atrial Fibrillation

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?

Ex 3
New Users of Warfarin with prior 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
/

Try running this on your own!

How many people do you get?

10,005 individuals
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”
**Cohort definition:** A cohort is defined as the set of persons satisfying one or more inclusion criteria for a duration of time. Cohort entry criteria involve selecting one or more initial events, which determine the start date for the cohort. Cohort exit criteria determine the end date when the person's episode no longer qualifies for the cohort.

### Available CDM Sources

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Generation Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHDSI CDM V5 Database</td>
<td>COMPLETE</td>
</tr>
</tbody>
</table>

**Distinct People:** 8207

- **with age Greater or Equal To 65**
- **a condition occurrence of Atrial Fibrillation**
- **starting between All days Before and 1 days Before** event index date and ending any time.
Conclusion

CDM standardizes the structure of data.

OMOP is now OHDSI.

Source data is preserved in the CDM.

CONCEPT_ANCESTOR table helps you find what you are looking for.

OMOP Vocabulary is used to standardize the terminology.

OHDSI is an Open Source Collaborative Community.

CDM can be used for many types of data (e.g., claims, EHR, survey, labs, etc.).
Extra Slides
OHDSI in a Box – International Keyboards