Introduction to AIR·MS

Health Data Literacy

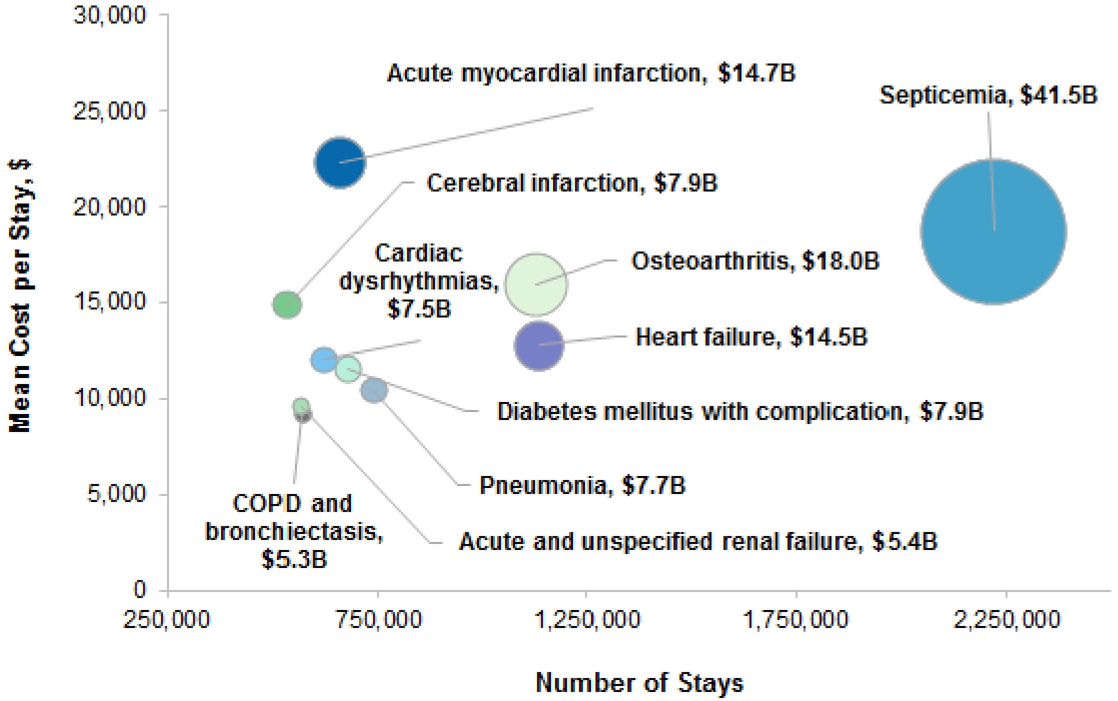
Eugenia Alleva
Andrew Deonarine

Sept 30th, 2025

Hasso Plattner Institute for Digital Health at Mount Sinai

Diagnosis Information

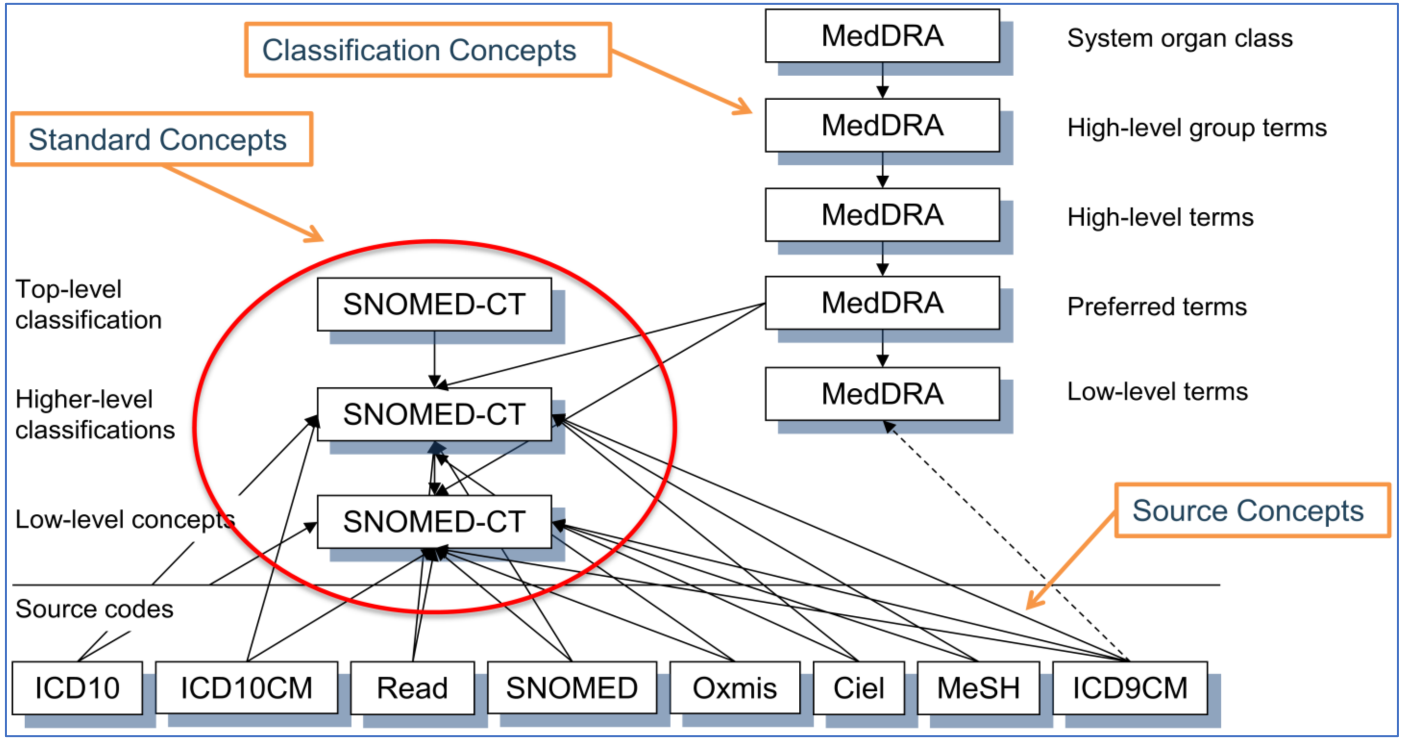
- During a visit (encounter) to the hospital, a patient may be diagnosed with a condition, such as diabetes, pneumonia, etc.
- Typically, diagnoses are stored using standard coding systems such as SNOMED and ICD-10
- These coding systems have been built over years by international consortia (for ICD-10) and are constantly being revised
- **When you look for diagnoses and pull up lists of conditions, you can consult with clinicians (physicians, specialists, etc) who will have a good idea of how frequent conditions are, how they rank, etc. – this is the process of clinically validating your query results**



https://hcup-us.ahrq.gov/reports/statbriefs/sb277-Top-Reasons-Hospital-Stays-2018.jsp

ICD-10 and SNOMED

- ICD-10: International Classification of Diseases, version 10 (developed internationally by the WHO)
- SNOMED: Systematized Nomenclature of Medicine (developed in the USA)
- There are 357,000 unique concepts in SNOMED, ICD-10-CM has about 70,000
- ICD codes are mapped to the low level and higher-level SNOMED-CT codes
- **AIRMS has a custom mapping produced by a 3rd party vendor between Epic IDs and SNOMED / ICD codes**



https://ohdsi.github.io/TheBookOfOhdsi/StandardizedVocabularies.html

Examples – ICD-10

E11: Diabetes mellitus without complications

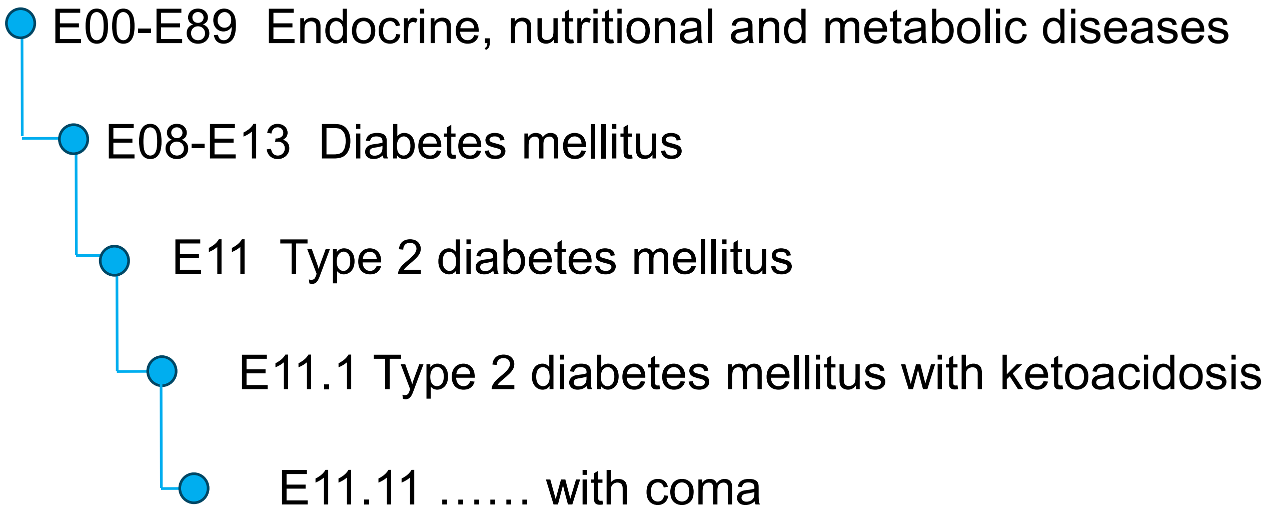
E11.0: Diabetes mellitus with hyperosmolarity

E11.1: Diabetes mellitus with ketoacidosis...

W61.62: Struck by duck

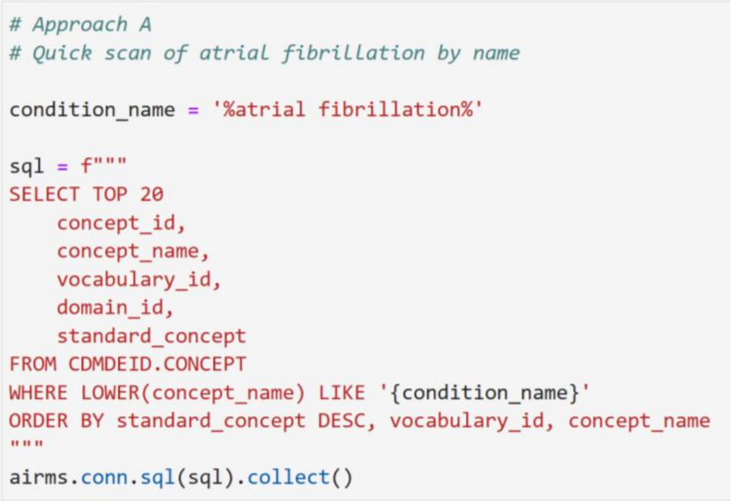
W61.62XA: Struck by duck, initial encounter

W61.62XD: Struck by duck, subsequent encounter...

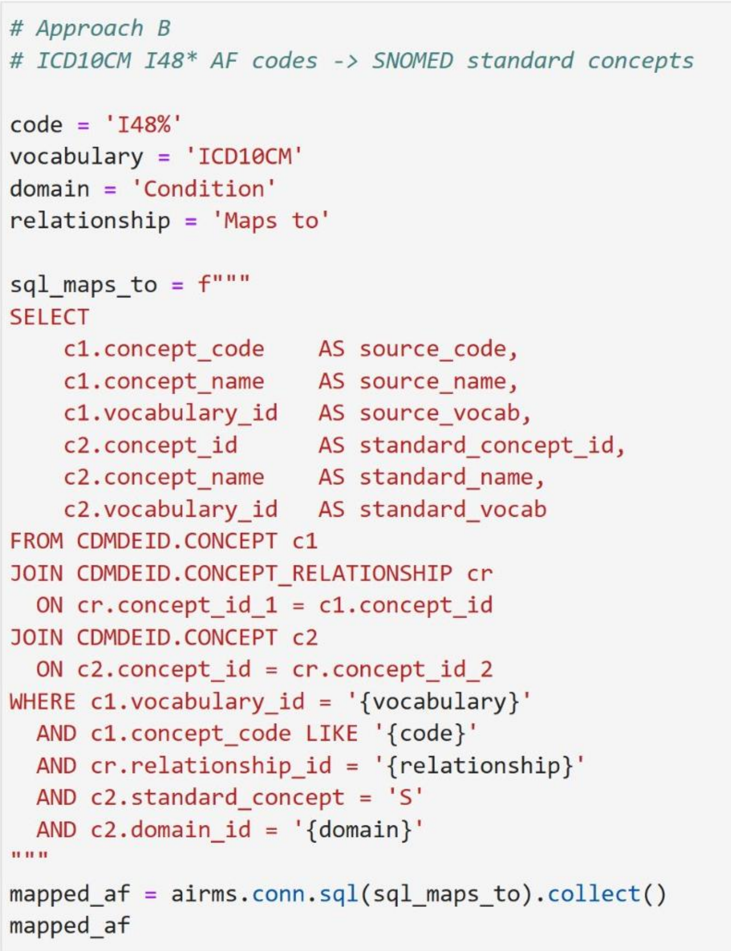


- E00-E89 Endocrine, nutritional and metabolic diseases
  - E08-E13 Diabetes mellitus
    - E11 Type 2 diabetes mellitus
      - E11.1 Type 2 diabetes mellitus with ketoacidosis
        - E11.11 …… with coma

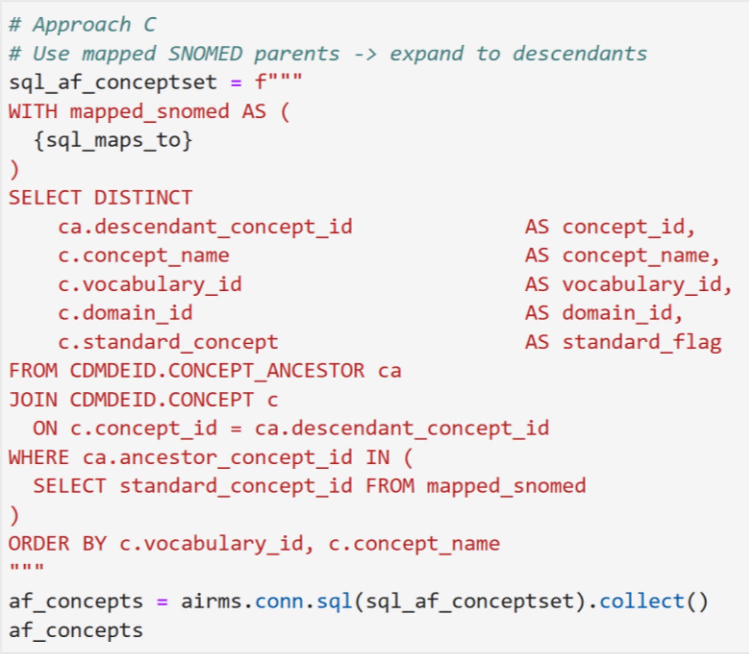
Using the Data: Querying and Retrieving Data



```
# Approach A
# Quick scan of atrial fibrillation by name

condition_name = '%atrial fibrillation%'

sql = f"""
SELECT TOP 20
    concept_id,
    concept_name,
    vocabulary_id,
    domain_id,
    standard_concept
FROM CDMDEID.CONCEPT
WHERE LOWER(concept_name) LIKE '{condition_name}'
ORDER BY standard_concept DESC, vocabulary_id, concept_name
"""
airms.conn.sql(sql).collect()
```

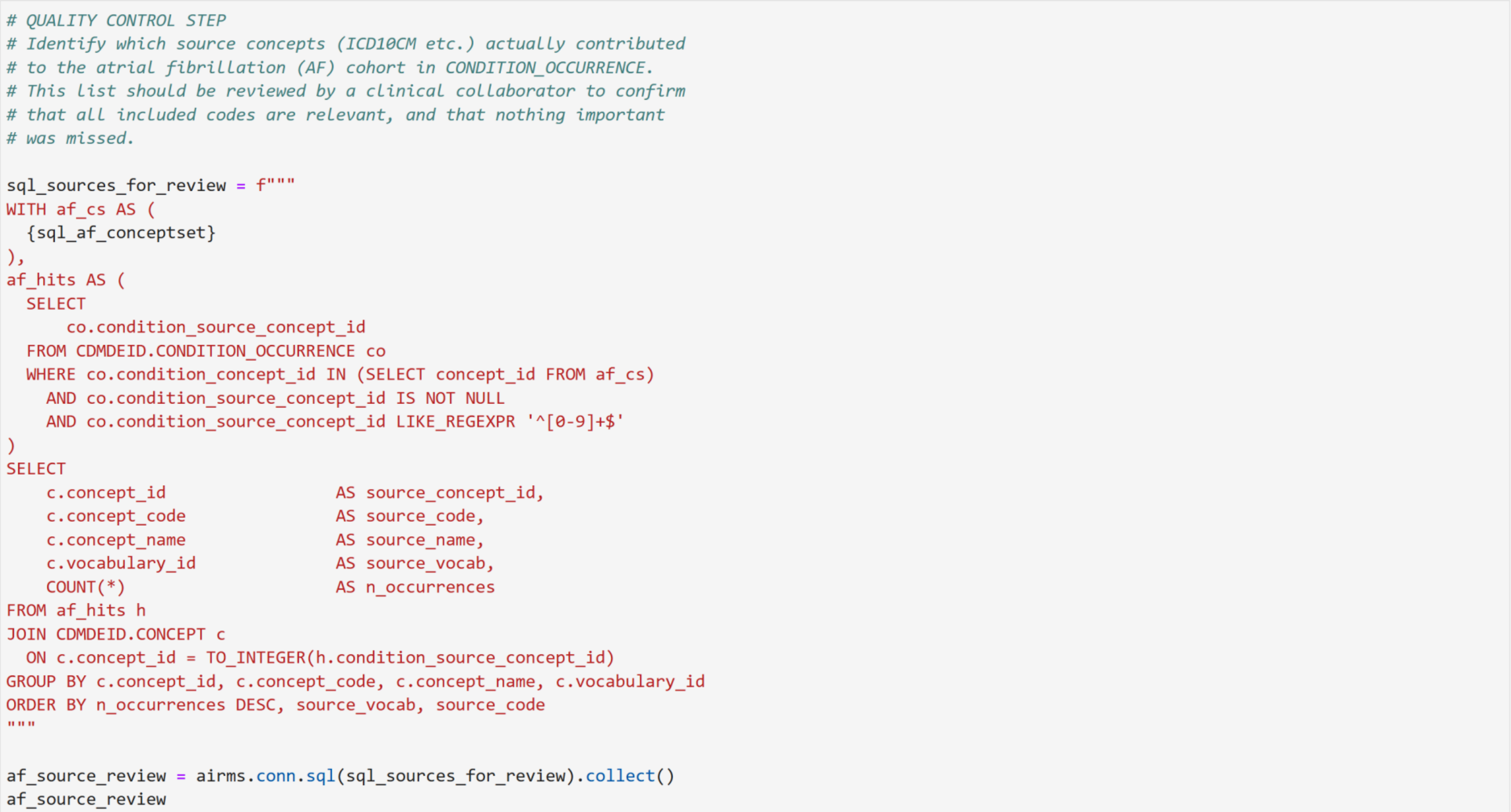


```
# Approach B
# ICD10CM I48* AF codes -> SNOMED standard concepts

code = 'I48%'
vocabulary = 'ICD10CM'
domain = 'Condition'
relationship = 'Maps to'

sql_maps_to = f"""
SELECT
    c1.concept_code    AS source_code,
    c1.concept_name    AS source_name,
    c1.vocabulary_id   AS source_vocab,
    c2.concept_id      AS standard_concept_id,
    c2.concept_name    AS standard_name,
    c2.vocabulary_id   AS standard_vocab
FROM CDMDEID.CONCEPT c1
JOIN CDMDEID.CONCEPT_RELATIONSHIP cr
  ON cr.concept_id_1 = c1.concept_id
JOIN CDMDEID.CONCEPT c2
  ON c2.concept_id = cr.concept_id_2
WHERE c1.vocabulary_id = '{vocabulary}'
  AND c1.concept_code LIKE '{code}'
  AND cr.relationship_id = '{relationship}'
  AND c2.standard_concept = 'S'
  AND c2.domain_id = '{domain}'
"""
mapped_af = airms.conn.sql(sql_maps_to).collect()
mapped_af
```



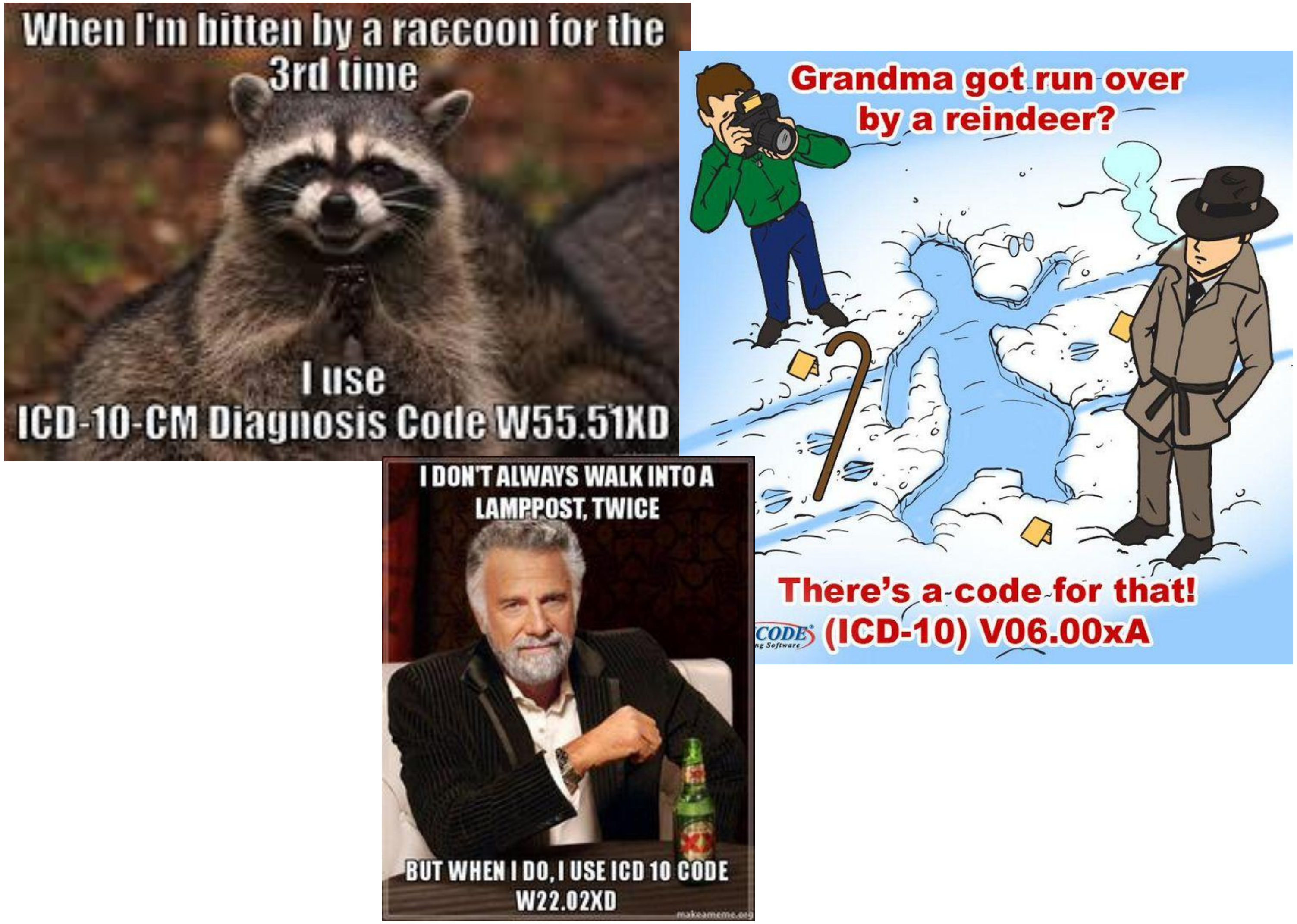
```
# Approach C
# Use mapped SNOMED parents -> expand to descendants
sql_af_conceptset = f"""
WITH mapped_snomed AS (
  {sql_maps_to}
)
SELECT DISTINCT
    ca.descendant_concept_id             AS concept_id,
    c.concept_name                       AS concept_name,
    c.vocabulary_id                      AS vocabulary_id,
    c.domain_id                          AS domain_id,
    c.standard_concept                   AS standard_flag
FROM CDMDEID.CONCEPT_ANCESTOR ca
JOIN CDMDEID.CONCEPT c
  ON c.concept_id = ca.descendant_concept_id
WHERE ca.ancestor_concept_id IN (
  SELECT standard_concept_id FROM mapped_snomed
)
ORDER BY c.vocabulary_id, c.concept_name
"""
af_concepts = airms.conn.sql(sql_af_conceptset).collect()
af_concepts
```

Using the Data: Querying and Retrieving Data



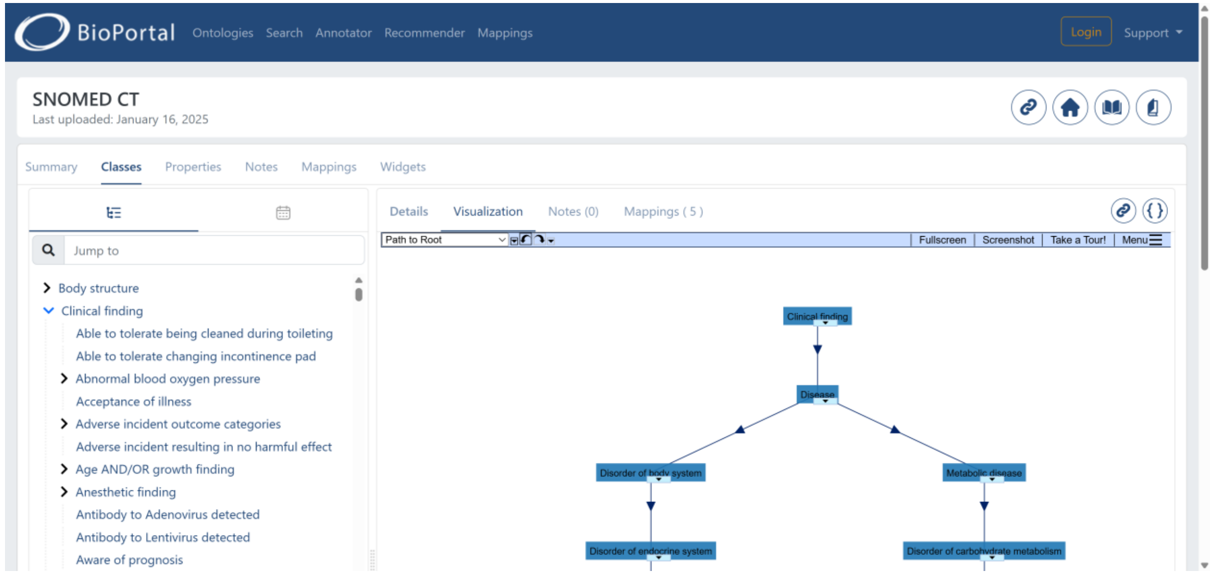
```
# QUALITY CONTROL STEP
# Identify which source concepts (ICD10CM etc.) actually contributed
# to the atrial fibrillation (AF) cohort in CONDITION_OCCURRENCE.
# This list should be reviewed by a clinical collaborator to confirm
# that all included codes are relevant, and that nothing important
# was missed.

sql_sources_for_review = f"""
WITH af_cs AS (
  {sql_af_conceptset}
),
af_hits AS (
  SELECT
      co.condition_source_concept_id
  FROM CDMDEID.CONDITION_OCCURRENCE co
  WHERE co.condition_concept_id IN (SELECT concept_id FROM af_cs)
    AND co.condition_source_concept_id IS NOT NULL
    AND co.condition_source_concept_id LIKE_REGEXPR '^[0-9]+$'
)
SELECT
    c.concept_id                 AS source_concept_id,
    c.concept_code               AS source_code,
    c.concept_name               AS source_name,
    c.vocabulary_id              AS source_vocab,
    COUNT(*)                     AS n_occurrences
FROM af_hits h
JOIN CDMDEID.CONCEPT c
  ON c.concept_id = TO_INTEGER(h.condition_source_concept_id)
GROUP BY c.concept_id, c.concept_code, c.concept_name, c.vocabulary_id
ORDER BY n_occurrences DESC, source_vocab, source_code
"""

af_source_review = airms.conn.sql(sql_sources_for_review).collect()
af_source_review
```

Other ICD-10 Examples


When I'm bitten by a raccoon for the 3rd time
I use ICD-10-CM Diagnosis Code W55.51XD
Grandma got run over by a reindeer?
There's a code for that!
CODE
(ICD-10) V06.00xA
I DON'T ALWAYS WALK INTO A LAMPPOST, TWICE
BUT WHEN I DO, I USE ICD 10 CODE W22.02XD
makeameme.org

SNOMED CT

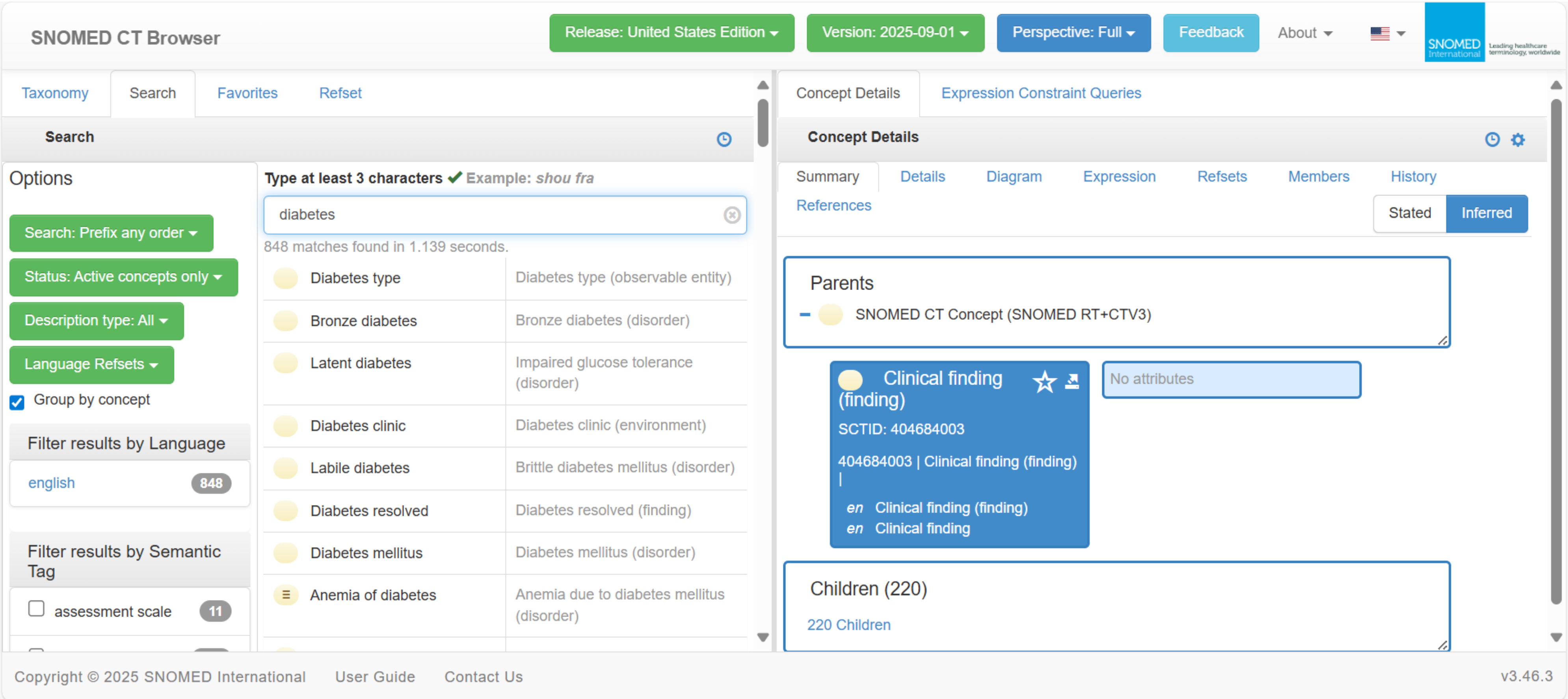
- SNOMED CT stands for Systematized Nomenclature of Medicine - Clinical Terms
- This is a comprehensive terminology for standardizing medical terms, including diagnoses, procedures, and other information
- SNOMED CT is organized into a hierarchical structure
- The main top-level categories include "Clinical Findings", "Procedures", "Organisms", etc.
- Relationships are also defined in SNOMED CT, the most common relationship being "is_a"
- Use Bioportal to browse these relationships: http://purl.bioontology.org/ontology/SNOMEDCT/313436004



SNOMED CT
Last uploaded: January 16, 2025

https://bioportal.bioontology.org/

Examples – SNOMED



SNOMED CT Browser

Search

Options

- Search: Prefix any order
- Status: Active concepts only
- Description type: All
- Language Refsets
- ☑ Group by concept

Filter results by Language

english 848

Filter results by Semantic Tag

□ assessment scale 11

Type at least 3 characters ✓ **Example:** ***shou fra***

diabetes

848 matches found in 1.139 seconds.

Diabetes type	Diabetes type (observable entity)
Bronze diabetes	Bronze diabetes (disorder)
Latent diabetes	Impaired glucose tolerance (disorder)
Diabetes clinic	Diabetes clinic (environment)
Labile diabetes	Brittle diabetes mellitus (disorder)
Diabetes resolved	Diabetes resolved (finding)
Diabetes mellitus	Diabetes mellitus (disorder)
Anemia of diabetes	Anemia due to diabetes mellitus (disorder)

Concept Details

Parents

SNOMED CT Concept (SNOMED RT+CTV3)

Clinical finding (finding)

SCTID: 404684003

404684003 | Clinical finding (finding) |

en Clinical finding (finding)

en Clinical finding

No attributes

Children (220)

220 Children

Copyright © 2025 SNOMED International User Guide Contact Us v3.46.3

https://browser.ihtsdotools.org/

What about Epic IDs?

- Epic has its own internal identification system (Epic ID's) which map to diseases, procedures, medications, etc. **These are not standard codes – for research, we need standardized codes**

- It also has mapping between Epic IDs and SNOMED CT, ICD-10, etc.

- **However, these mapping are commercial, and cannot be used for research purposes, so external consultants and companies have to produce them.**

- This is why you typically won't find a mapping on the Internet of Epic ID to SNOMED, ICD-10, etc.

- AIRMS OMOP contains its own mapping between Epic IDs and SNOMED/ICD, which was produced with the help of a consulting company

Diagnoses

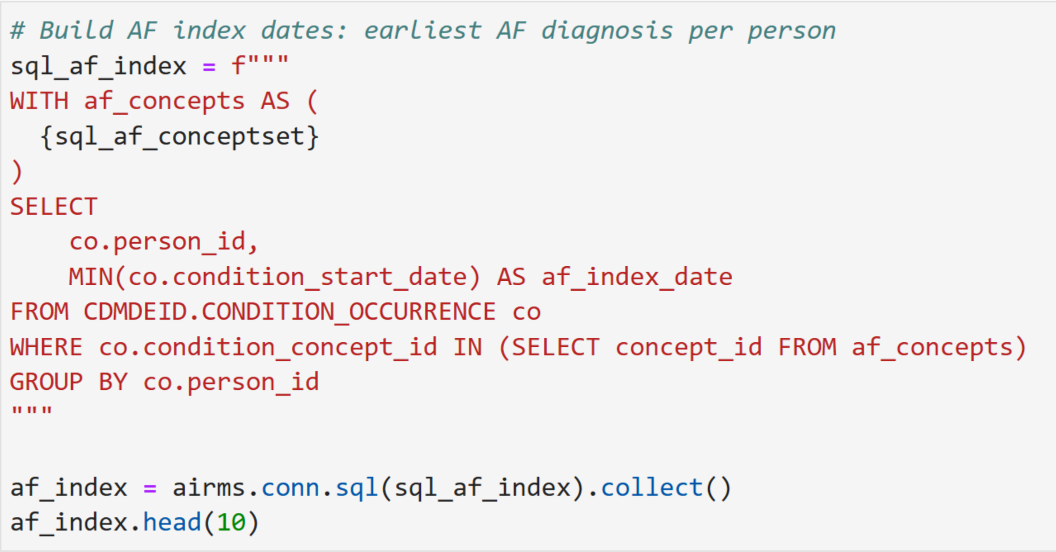
visit_occurrence		condition_occurrence			observation	
Patient A	Outpatient Visit 1				Patient A	Past Medical History F6
						Past Medical History K11
Patient B	Outpatient Visit 2	Patient B	Outpatient Visit 2	Encounter Diagnosis D4	Patient B	Past Medical History D4
				Encounter Diagnosis H8		Past Medical History L12
Patient C	Inpt Hospitalization 3	Patient C	Inpt Hospitalization 3	Encounter Diagnosis A1		
				Encounter Diagnosis B2		
				Hospital Problem C3		
				Hospital Problem D4		
				Problem List E5		
Patient D	Inpt Hospitalization 4	Patient D	Inpt Hospitalization 4	Encounter Diagnosis F6		
				Billing Diagnosis G7		
		Patient E		Problem List I9		
				Problem List J10		

Diagnosis Record Types

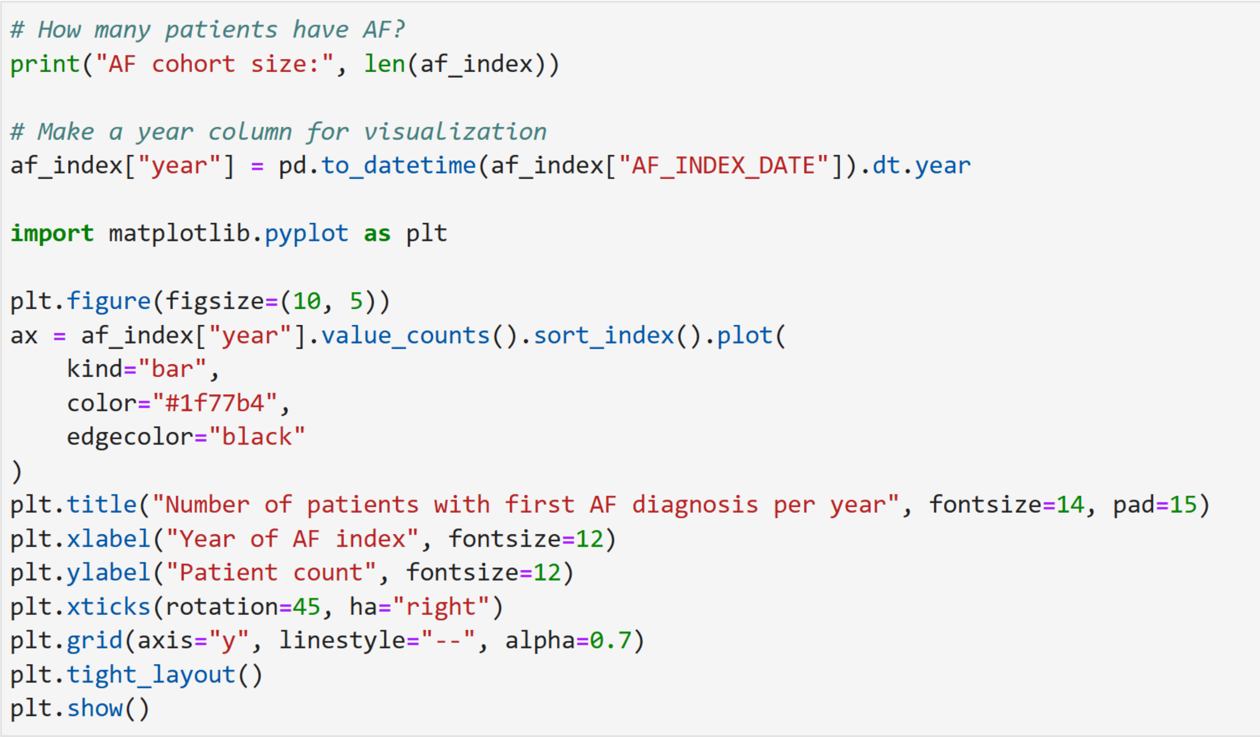
xtn_condition_type_source_concept_id	xtn_condition_type_source_concept_name	row_count
2000000108	Billing Diagnosis	73,040,692
2000000129	Encounter Diagnosis	120,323,291
2000000122	Hospital Problem	3,837,039
2000000120	Problem List	13,032,738

Record counts as of April 21, 2025

Using the Data: Querying and Retrieving Data



```
# Build AF index dates: earliest AF diagnosis per person
sql_af_index = f"""
WITH af_concepts AS (
  {sql_af_conceptset}
)
SELECT
    co.person_id,
    MIN(co.condition_start_date) AS af_index_date
FROM CDMDEID.CONDITION_OCCURRENCE co
WHERE co.condition_concept_id IN (SELECT concept_id FROM af_concepts)
GROUP BY co.person_id
"""

af_index = airms.conn.sql(sql_af_index).collect()
af_index.head(10)
```



```
# How many patients have AF?
print("AF cohort size:", len(af_index))

# Make a year column for visualization
af_index["year"] = pd.to_datetime(af_index["AF_INDEX_DATE"]).dt.year

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 5))
ax = af_index["year"].value_counts().sort_index().plot(
    kind="bar",
    color="#1f77b4",
    edgecolor="black"
)
plt.title("Number of patients with first AF diagnosis per year", fontsize=14, pad=15)
plt.xlabel("Year of AF index", fontsize=12)
plt.ylabel("Patient count", fontsize=12)
plt.xticks(rotation=45, ha="right")
plt.grid(axis="y", linestyle="--", alpha=0.7)
plt.tight_layout()
plt.show()
```

How Do You Pick the Right Codes?

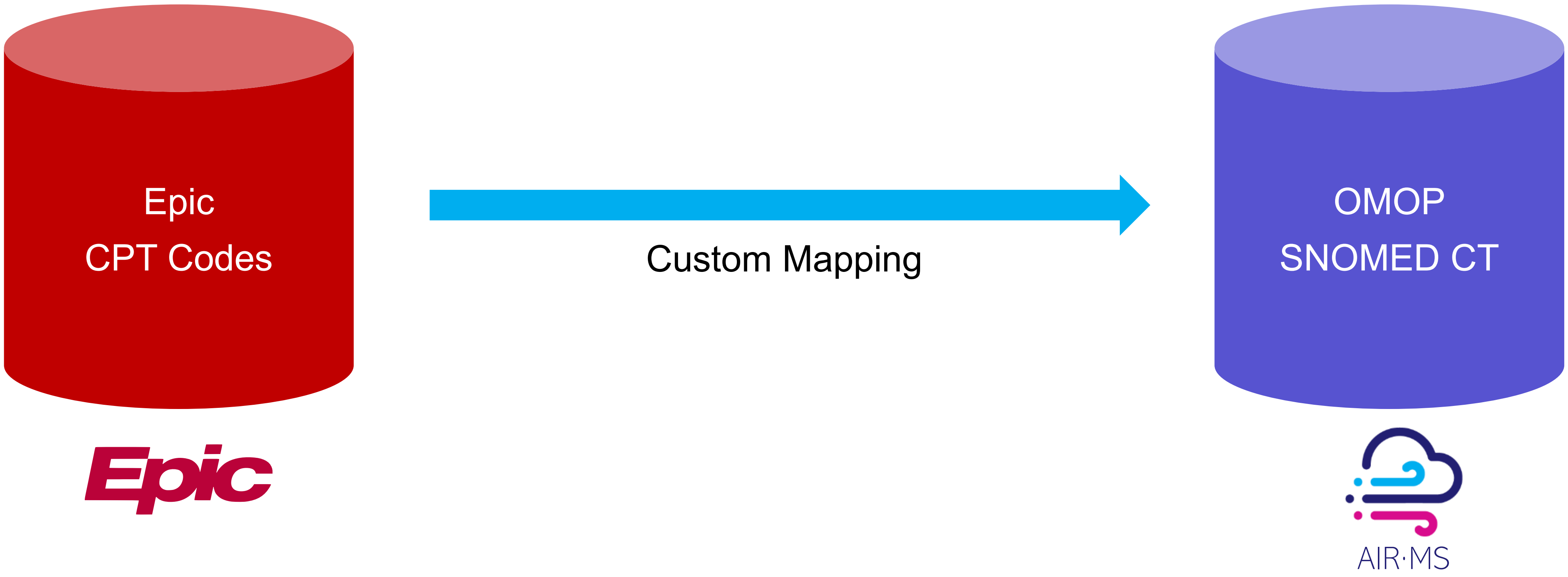
- It's tempting to do a keyword search for a condition, but this is not correct

- For example, if you query for "stroke" you'll miss "cerebral ischemia" (2025 ICD-10-CM Diagnosis Code I67.82)

- **As a result: look to existing published studies, or consult a clinician when building a list of conditions to look for**

- **Clinicians can also help you validate your query results: if you query for a few diseases among a specific demographic (ex. >65 year old males), certain diseases should come up as frequent**

Epic to OMOP Procedure Mapping



Procedure Data: CPT

- CPT stands for Current Procedural Terminology
- A commercial code set of codes maintained by the American Medical Association (AMA), requires a license to use
- These codes have played an important role in medical billing, documentation, and reporting
- Currently there are > 11,000 CPT codes
- Commercial – will need subscription to tools like Codify
- Mount Sinai has a license, which the AIRMS data uses – limitation on distributing mapping

Code Range	Category
00100-01999	Anesthesia
95700-95811	Sleep Medicine Testing Procedures
10004-69990	Surgery
70010-79999	Radiology Procedures
80047-89398	Pathology and Laboratory Procedures
90281-99607	Medicine Services and Procedures
98000-99499	Evaluation and Management
0001F-9007F	Category II Codes
0002M-0020M	Multianalyte Assay
0042T-0987T	Category III Codes

Category II Codes – supplemental performance tracking codes (4 digits with "F" suffix)

Category III Codes – temporary codes for emerging technologies, etc. (4 digits with "T" suffix)

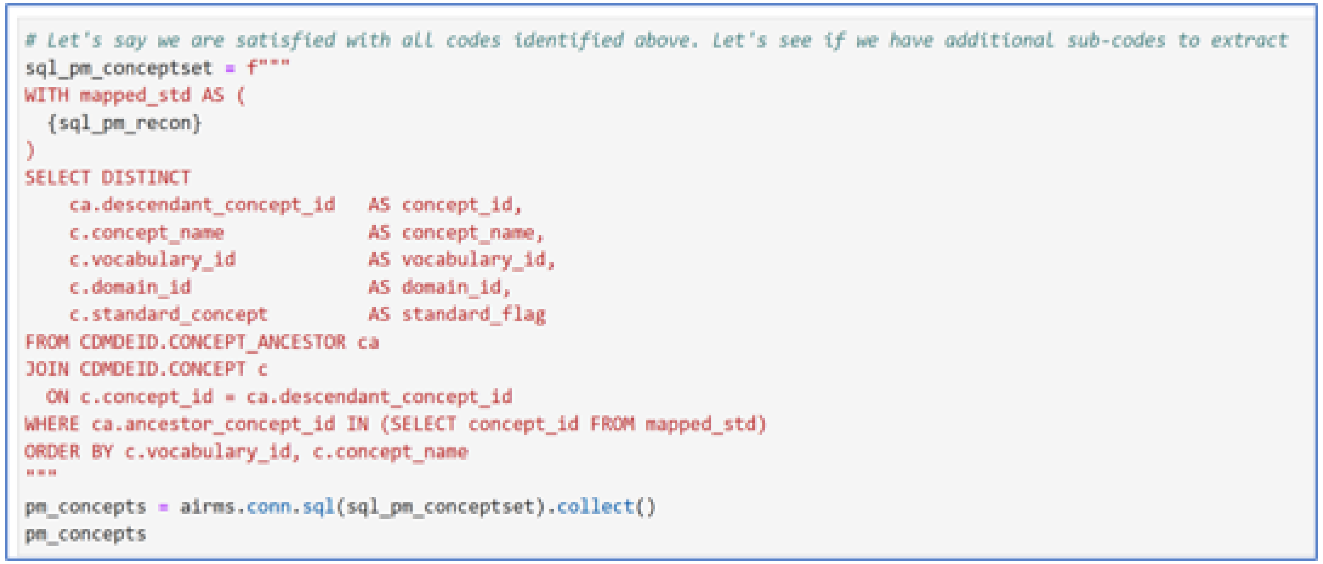
Procedures

Procedure	Surgical Procedure	procedure_occurrence		Flowsheet	
Procedure 1		Patient A	Procedure 1		
Procedure 2		Patient B	Procedure 2		
Procedure 3			Procedure 3		
	Surgical Procedure 4	Patient C	Surgical Procedure 4		
	Surgical Procedure 5	Patient D	Surgical Procedure 5		
		Patient E	Inferred Procedure 6	Patient E	Flowsheet Metric 6-1
					Flowsheet Metric 6-2
					Flowsheet Metric 6-3

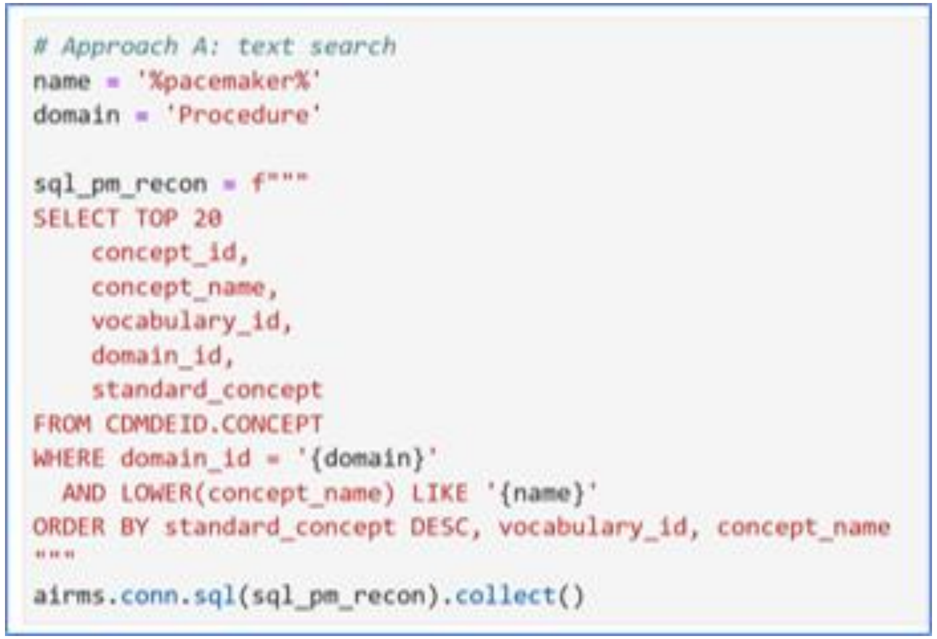
Procedure Record Types

xtn_procedure_type_source_concept_id	xtn_procedure_type_source_concept_name	row_count
2000000111	General Procedure	326,372,912
2002067235	Procedure Inferred from Flowsheet	281,426
2000000097	Surgical Procedure	962,017

Record counts as of April 21, 2025

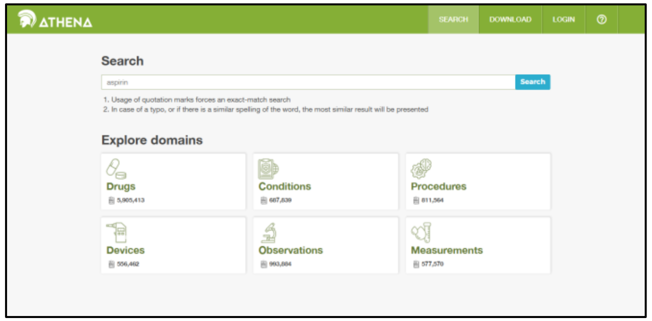


```
# Let's say we are satisfied with all codes identified above. Let's see if we have additional sub-codes to extract
sql_pm_conceptset = f"""
WITH mapped_std AS (
  {sql_pm_recon}
)
SELECT DISTINCT
    ca.descendant_concept_id   AS concept_id,
    c.concept_name             AS concept_name,
    c.vocabulary_id            AS vocabulary_id,
    c.domain_id                AS domain_id,
    c.standard_concept         AS standard_flag
FROM CDMDEID.CONCEPT_ANCESTOR ca
JOIN CDMDEID.CONCEPT c
  ON c.concept_id = ca.descendant_concept_id
WHERE ca.ancestor_concept_id IN (SELECT concept_id FROM mapped_std)
ORDER BY c.vocabulary_id, c.concept_name
"""
pm_concepts = airms.conn.sql(sql_pm_conceptset).collect()
pm_concepts
```



```
# Approach A: text search
name = '%pacemaker%'
domain = 'Procedure'

sql_pm_recon = f"""
SELECT TOP 20
    concept_id,
    concept_name,
    vocabulary_id,
    domain_id,
    standard_concept
FROM CDMDEID.CONCEPT
WHERE domain_id = '{domain}'
  AND LOWER(concept_name) LIKE '{name}'
ORDER BY standard_concept DESC, vocabulary_id, concept_name
"""
airms.conn.sql(sql_pm_recon).collect()
```

Epic to OMOP Medication Mapping



Search

aspirin

1. Usage of quotation marks forces an exact-match search
2. In case of a typo, or if there is a similar spelling of the word, the most similar result will be presented

Explore domains

Drugs 5,905,413

Conditions 687,839

Procedures 811,564

Devices 556,462

Observations 993,884

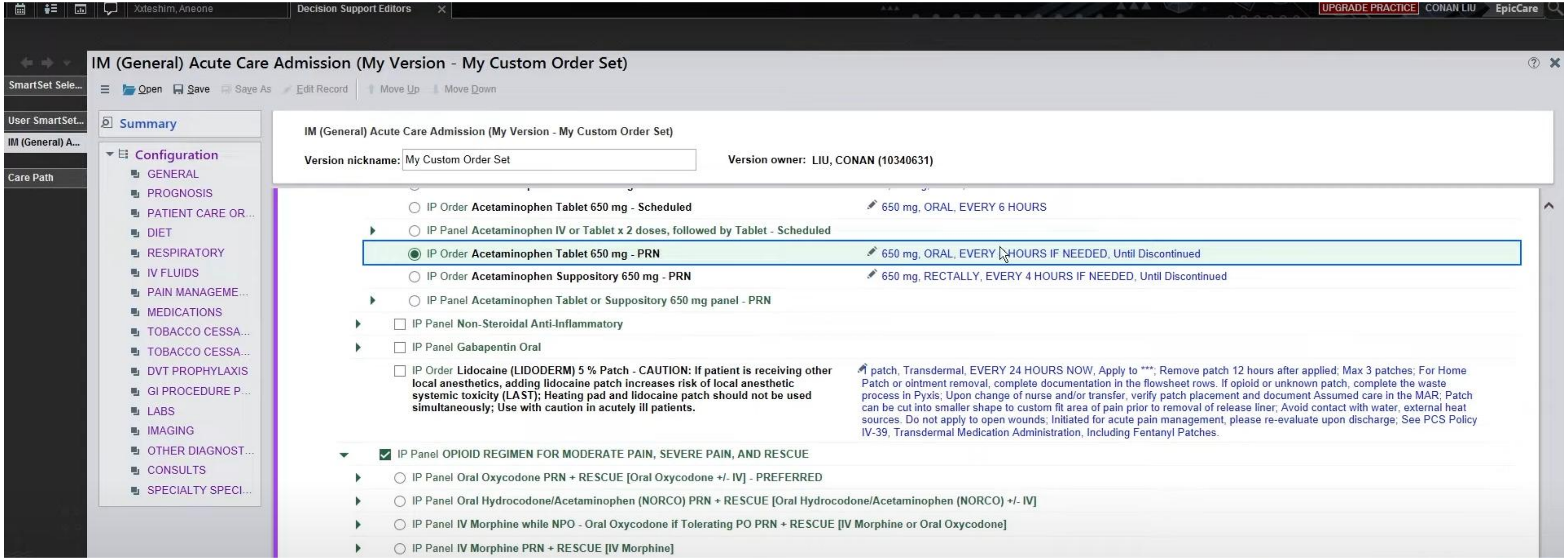
Measurements 577,570

Epic
ATC Codes

https://athena.ohdsi.org/

OMOP
RxNorm

Medication Data



IM (General) Acute Care Admission (My Version - My Custom Order Set)

IM (General) Acute Care Admission (My Version - My Custom Order Set)

Version nickname: My Custom Order Set

Version owner: LIU, CONAN (10340631)

○ IP Order **Acetaminophen Tablet 650 mg - Scheduled** — 650 mg, ORAL, EVERY 6 HOURS

○ IP Panel **Acetaminophen IV or Tablet x 2 doses, followed by Tablet - Scheduled**

● IP Order **Acetaminophen Tablet 650 mg - PRN** — 650 mg, ORAL, EVERY [illegible] HOURS IF NEEDED, Until Discontinued

○ IP Order **Acetaminophen Suppository 650 mg - PRN** — 650 mg, RECTALLY, EVERY 4 HOURS IF NEEDED, Until Discontinued

○ IP Panel **Acetaminophen Tablet or Suppository 650 mg panel - PRN**

□ IP Panel **Non-Steroidal Anti-Inflammatory**

□ IP Panel **Gabapentin Oral**

□ IP Order **Lidocaine (LIDODERM) 5 % Patch - CAUTION: If patient is receiving other local anesthetics, adding lidocaine patch increases risk of local anesthetic systemic toxicity (LAST); Heating pad and lidocaine patch should not be used simultaneously; Use with caution in acutely ill patients.** — patch, Transdermal, EVERY 24 HOURS NOW, Apply to ***; Remove patch 12 hours after applied; Max 3 patches; For Home Patch or ointment removal, complete documentation in the flowsheet rows. If opioid or unknown patch, complete the waste process in Pyxis; Upon change of nurse and/or transfer, verify patch placement and document Assumed care in the MAR; Patch can be cut into smaller shape to custom fit area of pain prior to removal of release liner; Avoid contact with water, external heat sources. Do not apply to open wounds; Initiated for acute pain management, please re-evaluate upon discharge; See PCS Policy IV-39, Transdermal Medication Administration, Including Fentanyl Patches.

☑ IP Panel **OPIOID REGIMEN FOR MODERATE PAIN, SEVERE PAIN, AND RESCUE**

○ IP Panel **Oral Oxycodone PRN + RESCUE [Oral Oxycodone +/- IV] - PREFERRED**

○ IP Panel **Oral Hydrocodone/Acetaminophen (NORCO) PRN + RESCUE [Oral Hydrocodone/Acetaminophen (NORCO) +/- IV]**

○ IP Panel **IV Morphine while NPO - Oral Oxycodone if Tolerating PO PRN + RESCUE [IV Morphine or Oral Oxycodone]**

○ IP Panel **IV Morphine PRN + RESCUE [IV Morphine]**

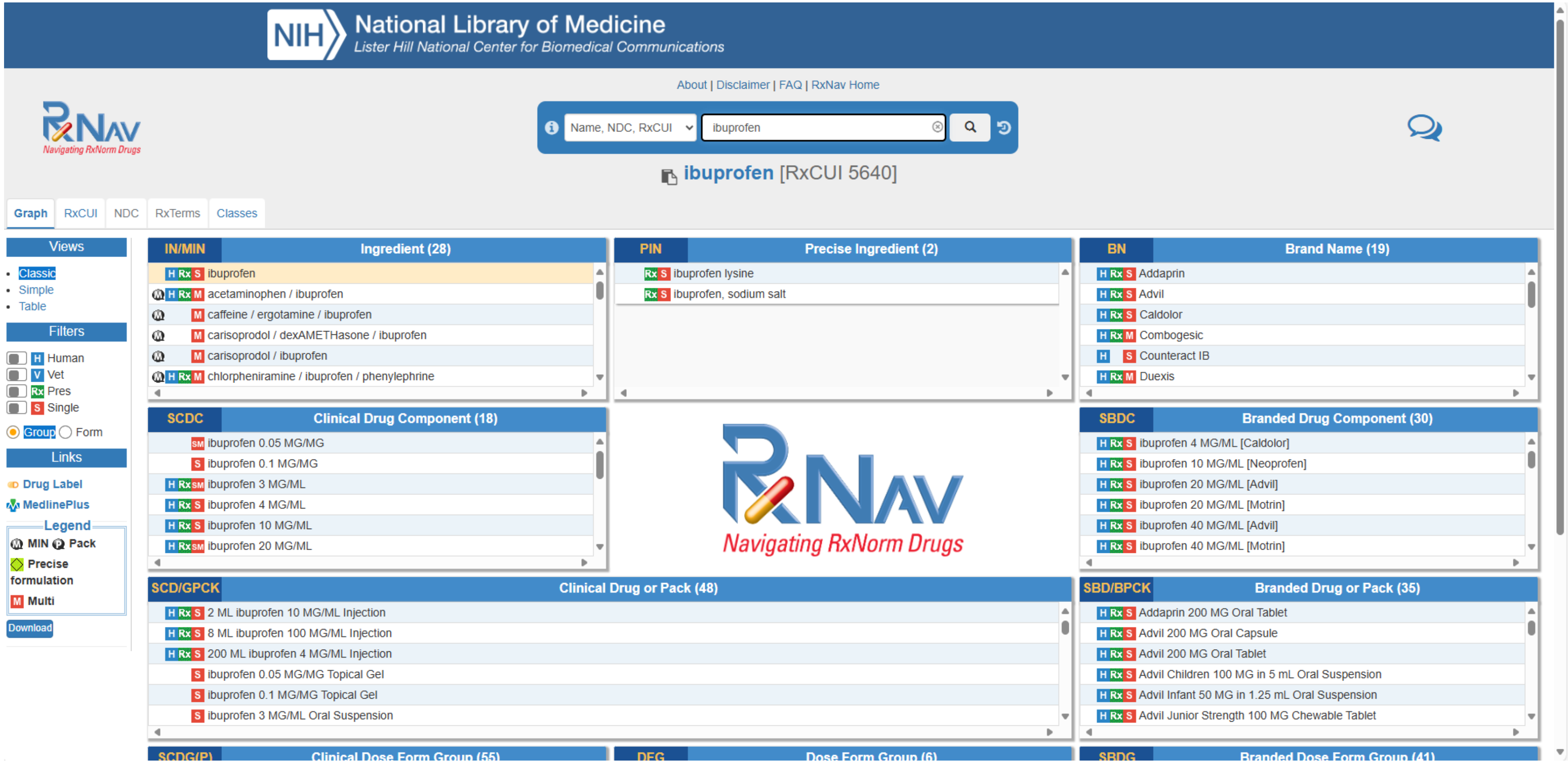
- Medications are stored in AIRMS, and contain a wealth of data
- This information is mapped from ATC codes (in Epic) to RxNorm (in OMOP)

Medication Data: RxNorm

- RxNorm – this coding standard is used in the OMOP CDM
  - It is developed by the National Library of Medicine (NLM)
  - A comprehensive system that presents a standardized way of representing medications
  - They use unique identifiers, namely a Concept Identifier (RXCUI), that is linked to each medication concept.
  - Identifiers are consistent across different sources and versions

```
Ingredient (IN)
  └── Precise Ingredient (PIN)
   └── Clinical Drug Form (CDF)
      └── Clinical Drug (SCD)
         └── Branded Drug Component (SBDC)
            └── Branded Drug (SBD)
               └── Branded Pack (BPCK)

Multiple Ingredients (MIN)
  └── Clinical Drug (SCD)
   └── Branded Drug (SBD)
      └── Pack (BPCK)
```

Medication Data: RxNorm



NIH National Library of Medicine
Lister Hill National Center for Biomedical Communications

About | Disclaimer | FAQ | RxNav Home

RxNav
Navigating RxNorm Drugs

Name, NDC, RxCUI | ibuprofen

ibuprofen [RxCUI 5640]

Graph | RxCUI | NDC | RxTerms | Classes

Views

- Classic
- Simple
- Table

Filters

□ H Human
□ V Vet
□ Rx Pres
□ S Single

● Group ○ Form

Links

Drug Label
MedlinePlus

Legend

MIN Pack
Precise formulation
M Multi

Download

IN/MIN	Ingredient (28)
H Rx S	ibuprofen
H Rx M	acetaminophen / ibuprofen
M	caffeine / ergotamine / ibuprofen
M	carisoprodol / dexAMETHasone / ibuprofen
M	carisoprodol / ibuprofen
H Rx M	chlorpheniramine / ibuprofen / phenylephrine

PIN	Precise Ingredient (2)
Rx S	ibuprofen lysine
Rx S	ibuprofen, sodium salt

BN	Brand Name (19)
H Rx S	Addaprin
H Rx S	Advil
H Rx S	Caldolor
H Rx M	Combogesic
H S	Counteract IB
H Rx M	Duexis

SCDC	Clinical Drug Component (18)
SM	ibuprofen 0.05 MG/MG
S	ibuprofen 0.1 MG/MG
H Rx SM	ibuprofen 3 MG/ML
H Rx S	ibuprofen 4 MG/ML
H Rx S	ibuprofen 10 MG/ML
H Rx SM	ibuprofen 20 MG/ML

SBDC	Branded Drug Component (30)
H Rx S	ibuprofen 4 MG/ML [Caldolor]
H Rx S	ibuprofen 10 MG/ML [Neoprofen]
H Rx S	ibuprofen 20 MG/ML [Advil]
H Rx S	ibuprofen 20 MG/ML [Motrin]
H Rx S	ibuprofen 40 MG/ML [Advil]
H Rx S	ibuprofen 40 MG/ML [Motrin]

SCD/GPCK	Clinical Drug or Pack (48)
H Rx S	2 ML ibuprofen 10 MG/ML Injection
H Rx S	8 ML ibuprofen 100 MG/ML Injection
H Rx S	200 ML ibuprofen 4 MG/ML Injection
S	ibuprofen 0.05 MG/MG Topical Gel
S	ibuprofen 0.1 MG/MG Topical Gel
S	ibuprofen 3 MG/ML Oral Suspension

SBD/BPCK	Branded Drug or Pack (35)
H Rx S	Addaprin 200 MG Oral Tablet
H Rx S	Advil 200 MG Oral Capsule
H Rx S	Advil 200 MG Oral Tablet
H Rx S	Advil Children 100 MG in 5 mL Oral Suspension
H Rx S	Advil Infant 50 MG in 1.25 mL Oral Suspension
H Rx S	Advil Junior Strength 100 MG Chewable Tablet

SCDG(P) Clinical Dose Form Group (55) | DFG Dose Form Group (6) | SBDG Branded Dose Form Group (41)

https://mor.nlm.nih.gov/RxNav/search

RxNorm Examples

1. Ingredient (IN)

The basic chemical component of a drug.

Example: Ibuprofen

2. Precise Ingredient (PIN)

A more specific form of an ingredient, like a salt or ester.

Example: Ibuprofen sodium

3. Multiple Ingredients (MIN)

Represents a combination of ingredients.

Example: Ibuprofen / Hydrocodone

4. Clinical Drug Form (CDF) [Optional level in hierarchy]

Combines an ingredient with a dosage form (without strength).

Example: Ibuprofen Oral Tablet

5. Clinical Drug (SCD = Semantic Clinical Drug)

Includes the ingredient, strength, and dose form, but no brand.

Example: Ibuprofen 200 MG Oral Tablet

6. Branded Drug Component (SBDC = Semantic Branded Drug Component)

A branded version of the SCD with one ingredient.

Example: Advil 200 MG Oral Tablet [Ibuprofen]

7. Branded Drug (SBD = Semantic Branded Drug)

A branded medication with all details (strength, dose form, etc.).

Example: Advil 200 MG Oral Tablet

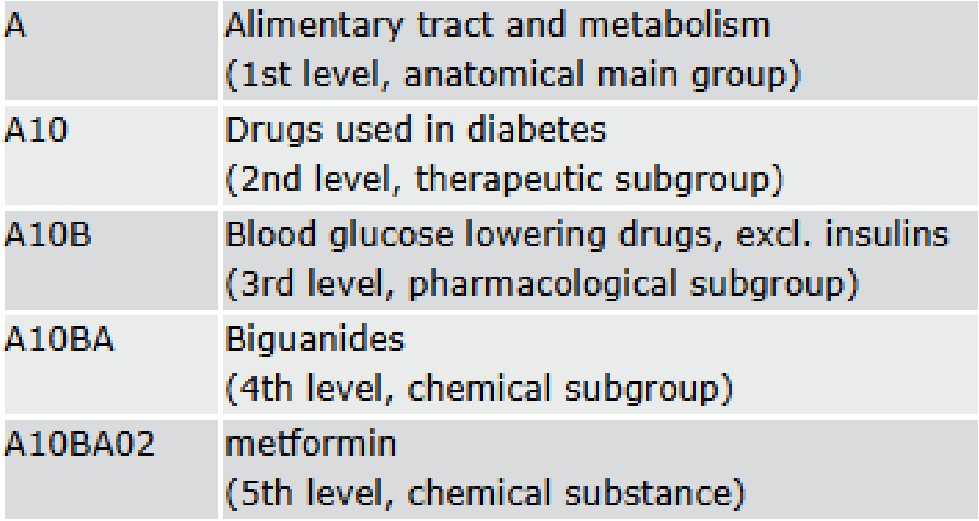
8. Branded Pack (BPCK) / Clinical Pack (GPCK)

Represents packaging of one or more SBDs or SCDs.

Example: Advil 200 MG Oral Tablet Pack

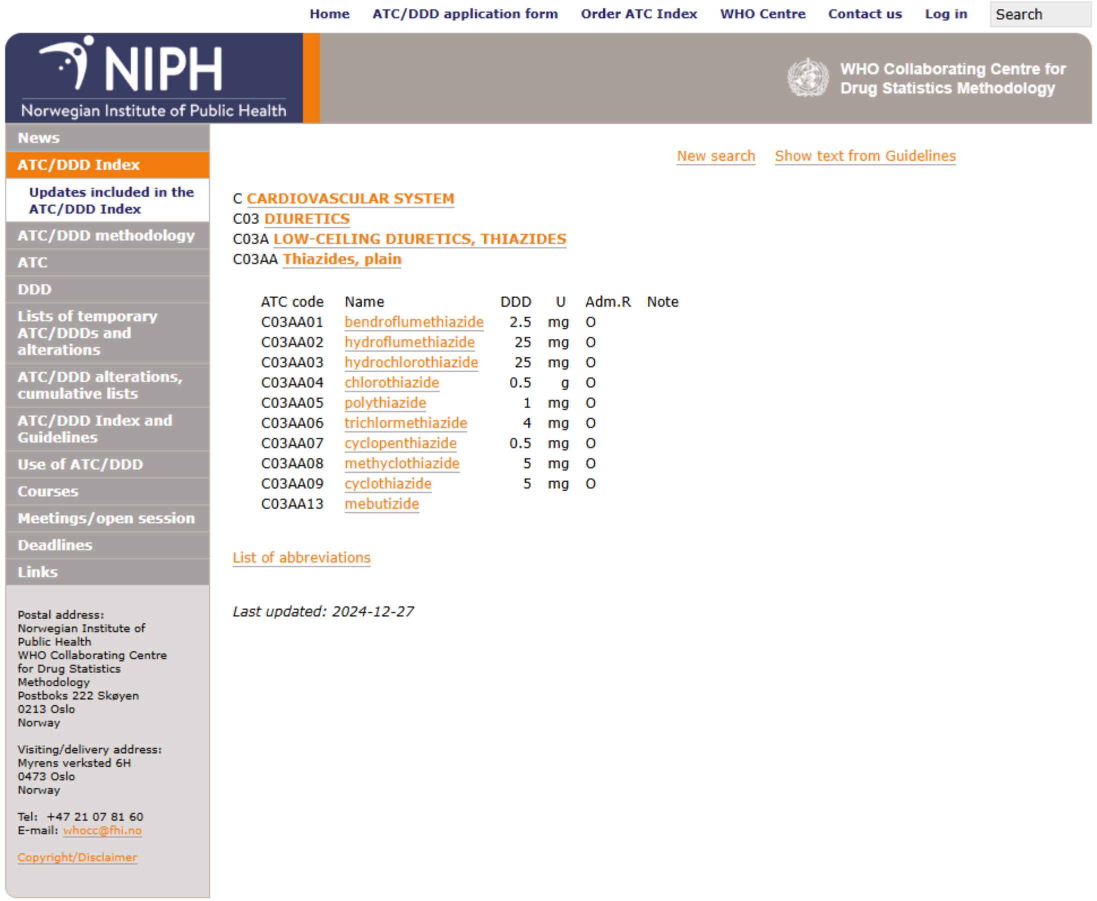
Medication Data: ATC Codes

- Anatomical Therapeutic Chemical (ATC) Classification System - classifies drugs based on the organ system they act on, and their properties (chemical, pharmacological, or therapeutic)
- Produced by the World Health Organization Collaborating Center for Drug Statistics Methodology (WHOCC)
- First published in 1976
- The lowest level of the ATC classification contains 5067 codes
- Simple drug browser available from https://atcddd.fhi.no (Norwegian Institute of Public Health)



A	Alimentary tract and metabolism (1st level, anatomical main group)
A10	Drugs used in diabetes (2nd level, therapeutic subgroup)
A10B	Blood glucose lowering drugs, excl. insulins (3rd level, pharmacological subgroup)
A10BA	Biguanides (4th level, chemical subgroup)
A10BA02	metformin (5th level, chemical substance)

Medication Data: ATC Codes


Home
ATC/DDD application form
Order ATC Index
WHO Centre
Contact us
Log in
Search
NIPH
Norwegian Institute of Public Health
WHO Collaborating Centre for Drug Statistics Methodology
News
ATC/DDD Index
Updates included in the ATC/DDD Index
ATC/DDD methodology
ATC
DDD
Lists of temporary ATC/DDDs and alterations
ATC/DDD alterations, cumulative lists
ATC/DDD Index and Guidelines
Use of ATC/DDD
Courses
Meetings/open session
Deadlines
Links
New search
Show text from Guidelines

C CARDIOVASCULAR SYSTEM
C03 DIURETICS
C03A LOW-CEILING DIURETICS, THIAZIDES
C03AA Thiazides, plain

ATC code	Name	DDD	U	Adm.R	Note
C03AA01	bendroflumethiazide	2.5	mg	O	
C03AA02	hydroflumethiazide	25	mg	O	
C03AA03	hydrochlorothiazide	25	mg	O	
C03AA04	chlorothiazide	0.5	g	O	
C03AA05	polythiazide	1	mg	O	
C03AA06	trichlormethiazide	4	mg	O	
C03AA07	cyclopenthiazide	0.5	mg	O	
C03AA08	methyclothiazide	5	mg	O	
C03AA09	cyclothiazide	5	mg	O	
C03AA13	mebutizide				

List of abbreviations

Last updated: 2024-12-27

Postal address:
Norwegian Institute of
Public Health
WHO Collaborating Centre
for Drug Statistics
Methodology
Postboks 222 Skøyen
0213 Oslo
Norway

Visiting/delivery address:
Myrens verksted 6H
0473 Oslo
Norway

Tel: +47 21 07 81 60
E-mail: whocc@fhi.no

Copyright/Disclaimer

https://atcddd.fhi.no/atc_ddd_index/?code=B02BC&showdescription=no

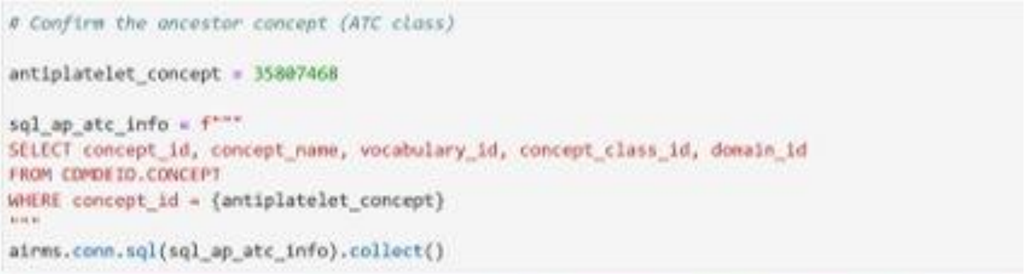
Medications

Medication Order		Medication Dispense		Medication Administration		Immunization Event		drug_exposure		
						Patient A	Vaccine Z	Patient A	Immuniz Z1	Vaccine Z
Patient B	Inpt Med Order 2					Inpt Med Order 2	Vaccine Y	Patient B	Inpt Med Order 2	Vaccine Y
Patient C	Inpt Med Order 3			Inpt Med Order 3	Med Admin 3-1			Patient C	Med Admin 3-1	Drug X
					Med Admin 3-2				Med Admin 3-2	Drug X
Patient D	Inpt Med Order 4	Inpt Med Order 4	Dispense 4-1	Inpt Med Order 4	Med Admin 4-1			Patient D	Med Admin 4-1	Drug W
			Dispense 4-2		Med Admin 4-2				Med Admin 4-2	Drug W
Patient E	Inpt Med Order 5	Inpt Med Order 5	Dispense 5-1					Patient E	Dispense 5-1	Drug V
Patient F	Outpt Med Order 6	Outpt Med Order 6	Dispense 6-1					Patient F	Dispense 6-1	Drug U
Patient G	Outpt Med Order 7	Outpt Med Order 7	Dispense 7-1					Patient G	Dispense 7-1	Drug T
			Dispense 7-2						Dispense 7-1	Drug T
Patient H	Outpt Med Order 8							Patient H	Outpt Med Order 8	Drug S

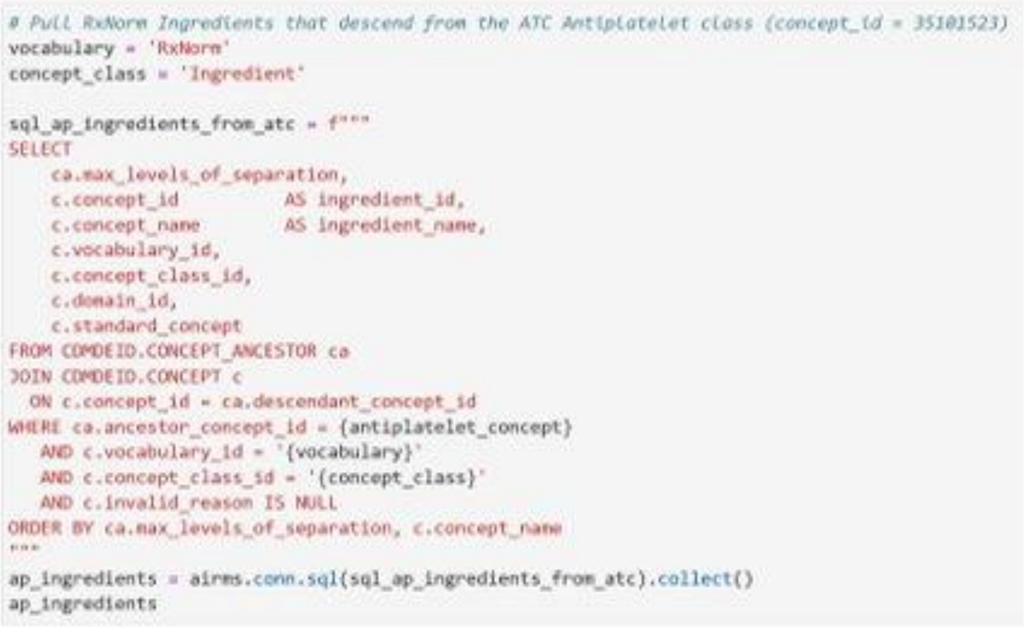
Medication Record Types

xtn_drug_type_source_concept_id	xtn_drug_type_source_concept_name	row_count
2002056202	Immunization Administration	7,835,276
2002056203	Immunization from Medication Order	289,032
2002056204	Immunization from Procedure Order	3,488,313
2000000121	Medication Order	13,670,853
2000000110	Medication Order with Administration	138,120,557
2000000098	Medication Order with Dispense	16,351,836
2000000096	Outpatient Medication Order	48,588,821
2000000109	Outpatient Medication Order with Dispense	151,086

Record counts as of April 21, 2025

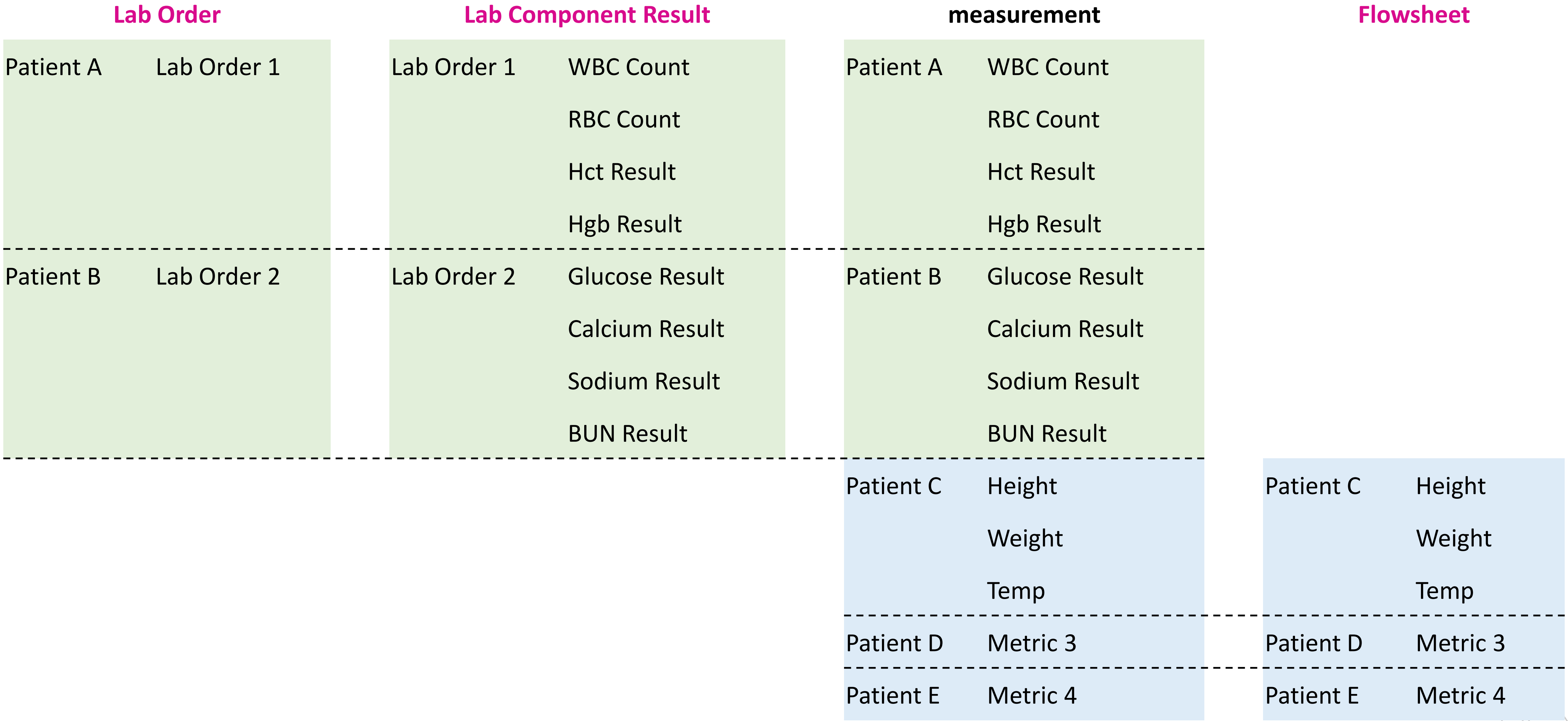


```
# Confirm the ancestor concept (ATC class)

antiplatelet_concept = 35807468

sql_ap_atc_info = f"""
SELECT concept_id, concept_name, vocabulary_id, concept_class_id, domain_id
FROM CDMDEID.CONCEPT
WHERE concept_id = {antiplatelet_concept}
"""
airms.conn.sql(sql_ap_atc_info).collect()
```



```
# Pull RxNorm Ingredients that descend from the ATC Antiplatelet class (concept_id = 35101523)
vocabulary = 'RxNorm'
concept_class = 'Ingredient'

sql_ap_ingredients_from_atc = f"""
SELECT
    ca.max_levels_of_separation,
    c.concept_id          AS ingredient_id,
    c.concept_name        AS ingredient_name,
    c.vocabulary_id,
    c.concept_class_id,
    c.domain_id,
    c.standard_concept
FROM CDMDEID.CONCEPT_ANCESTOR ca
JOIN CDMDEID.CONCEPT c
  ON c.concept_id = ca.descendant_concept_id
WHERE ca.ancestor_concept_id = {antiplatelet_concept}
   AND c.vocabulary_id = '{vocabulary}'
   AND c.concept_class_id = '{concept_class}'
   AND c.invalid_reason IS NULL
ORDER BY ca.max_levels_of_separation, c.concept_name
"""
ap_ingredients = airms.conn.sql(sql_ap_ingredients_from_atc).collect()
ap_ingredients
```

Lab Results & Vital Signs (measurements)

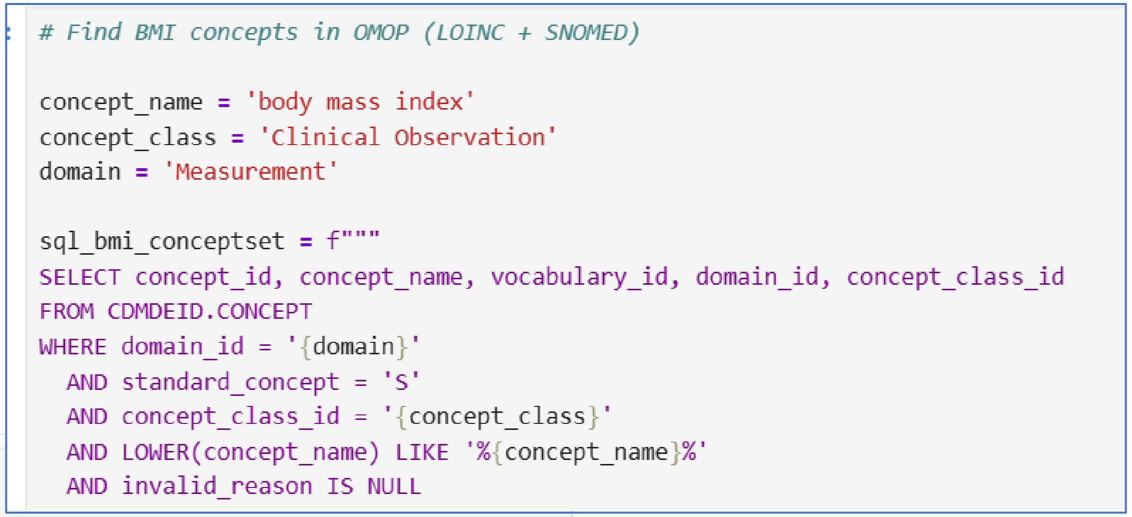


Lab Order		Lab Component Result		measurement		Flowsheet	
Patient A	Lab Order 1	Lab Order 1	WBC Count	Patient A	WBC Count		
			RBC Count		RBC Count		
			Hct Result		Hct Result		
			Hgb Result		Hgb Result		
Patient B	Lab Order 2	Lab Order 2	Glucose Result	Patient B	Glucose Result		
			Calcium Result		Calcium Result		
			Sodium Result		Sodium Result		
			BUN Result		BUN Result		
				Patient C	Height	Patient C	Height
					Weight		Weight
					Temp		Temp
				Patient D	Metric 3	Patient D	Metric 3
				Patient E	Metric 4	Patient E	Metric 4

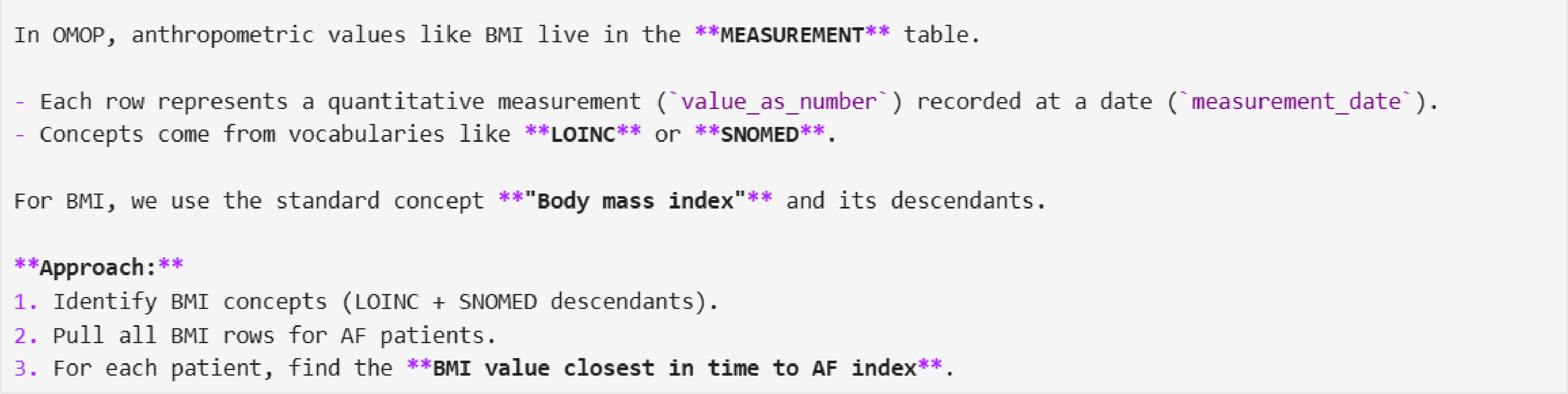
Measurement Record Types

xtn_measurement_type_source_concept_id	xtn_measurement_type_source_concept_name	row_count
2002067233	Flowsheet Measurement	254,696,901
2000000100	Lab Component Result	1,107,913,246
2000000123	Vital Signs	692,312,136

Record counts as of April 21, 2025



```
# Find BMI concepts in OMOP (LOINC + SNOMED)

concept_name = 'body mass index'
concept_class = 'Clinical Observation'
domain = 'Measurement'

sql_bmi_conceptset = f"""
SELECT concept_id, concept_name, vocabulary_id, domain_id, concept_class_id
FROM CDMDEID.CONCEPT
WHERE domain_id = '{domain}'
  AND standard_concept = 'S'
  AND concept_class_id = '{concept_class}'
  AND LOWER(concept_name) LIKE '%{concept_name}%'
  AND invalid_reason IS NULL
```



```
## BMI Measurement near AF Index

In OMOP, anthropometric values like BMI live in the **MEASUREMENT** table.

- Each row represents a quantitative measurement (`value_as_number`) recorded at a date (`measurement_date`).
- Concepts come from vocabularies like **LOINC** or **SNOMED**.

For BMI, we use the standard concept **"Body mass index"** and its descendants.

**Approach:**
1. Identify BMI concepts (LOINC + SNOMED descendants).
2. Pull all BMI rows for AF patients.
3. For each patient, find the **BMI value closest in time to AF index**.
```

Content Standardization
via
Concept Mapping

OMOP Data Example (*synthetic*)

condition_occurrence	Value	Definition
condition_type_concept_id	32827	OMOP's standard record type "EHR encounter record"
xtn_condition_type_source_concept_id	2000000129	MSDW's source record type "Encounter Diagnosis"
visit_occurrence_id	999888777	Unique identifier for the encounter
person_id	987654321	Unique identifier for the patient
provider_id	123456789	Unique identifier for the provider recording the diagnosis
condition_concept_id	4193704	OMOP's identifier for **SNOMED code 313436004**
condition_source_concept_id	2000602205	MSDW's identifier for **Epic diagnosis code 521601**
condition_start_date	1/1/2022	The date of condition onset or documentation per provider
condition_end_date	NULL	The date on which the condition resolved (if any)

concept_relationship	Value	Definition
concept_1	2000602205	MSDW's identifier for **Epic diagnosis code 521601**
relationship_id	Maps to non-standard	Text string denoting the type of **mapping relationship** between concept_id_1 & concept_id_2
concept_2	35206882	OMOP's identifier for **ICD-10-CM code E11.9**

OMOP Data Example (*synthetic*)

		concept			
condition_occurrence	Value	concept_id	vocabulary_id	concept_code	concept_name
condition_type_concept_id	32827	32827	Type Concept	OMOP4976900	EHR encounter record
xtn_condition_type_source_concept_id	2000000129	2000000129	MSDW Src Rec Type	Encounter Diagnosis	Encounter Diagnosis
visit_occurrence_id	999888777				
person_id	987654321				
provider_id	123456789				
condition_concept_id	4193704	4193704	SNOMED	313436004	Type 2 diabetes mellitus without complication
condition_source_concept_id	2000602205	2000602205	EPIC EDG .1	521601	Type 2 diabetes mellitus without complications
condition_start_date	1/1/2022				
condition_end_date	NULL				

concept_relationship	Value				
concept_1	2000602205	2000602205	EPIC EDG .1	521601	Type 2 diabetes mellitus without complications
relationship_id	Maps to non-standard				
concept_2	35206882	35206882	ICD10CM	E11.9	Type 2 diabetes mellitus without complications

Visualizing and Analyzing Data in Jupyter

Using the Data: Querying and Retrieving Data

Age and Sex from PERSON

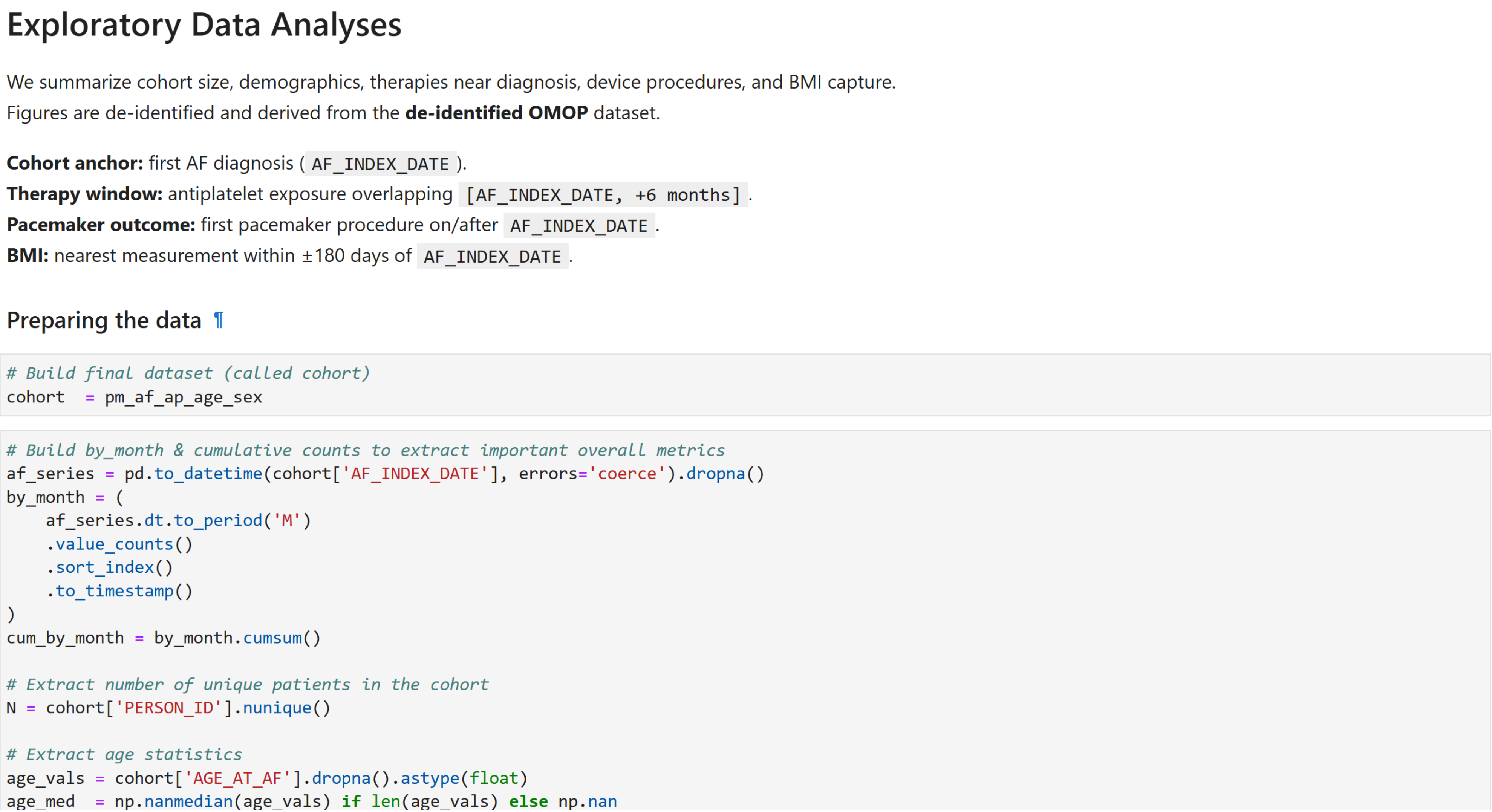
We now extend the cohort with **demographics**:

- **Sex** is stored in `PERSON.gender_concept_id`.
  We join to the CONCEPT table for a readable label.
- **Age at AF index** is calculated from `PERSON.birth_date` relative to `AF_INDEX_DATE`.

This information is critical for describing the cohort and adjusting analyses.

```
# Querying age and sex
sql_pm_af_ap_age_sex = f"""
WITH cohort_base AS (
  {sql_pm_af_ap_bmi}
)
SELECT
    cb.*,
    p.birth_datetime,
    g.concept_name AS sex,
    FLOOR(MONTHS_BETWEEN(p.birth_datetime, cb.af_index_date) / 12) AS age_at_af
FROM cohort_base cb
JOIN CDMDEID.PERSON p
  ON p.person_id = cb.person_id
JOIN CDMDEID.CONCEPT g
  ON g.concept_id = p.gender_concept_id
ORDER BY cb.af_index_date
"""

pm_af_ap_age_sex = airms.conn.sql(sql_pm_af_ap_age_sex).collect()
pm_af_ap_age_sex.head()
```

Using the Data: Querying and Retrieving Data



Exploratory Data Analyses

We summarize cohort size, demographics, therapies near diagnosis, device procedures, and BMI capture.
Figures are de-identified and derived from the **de-identified OMOP** dataset.

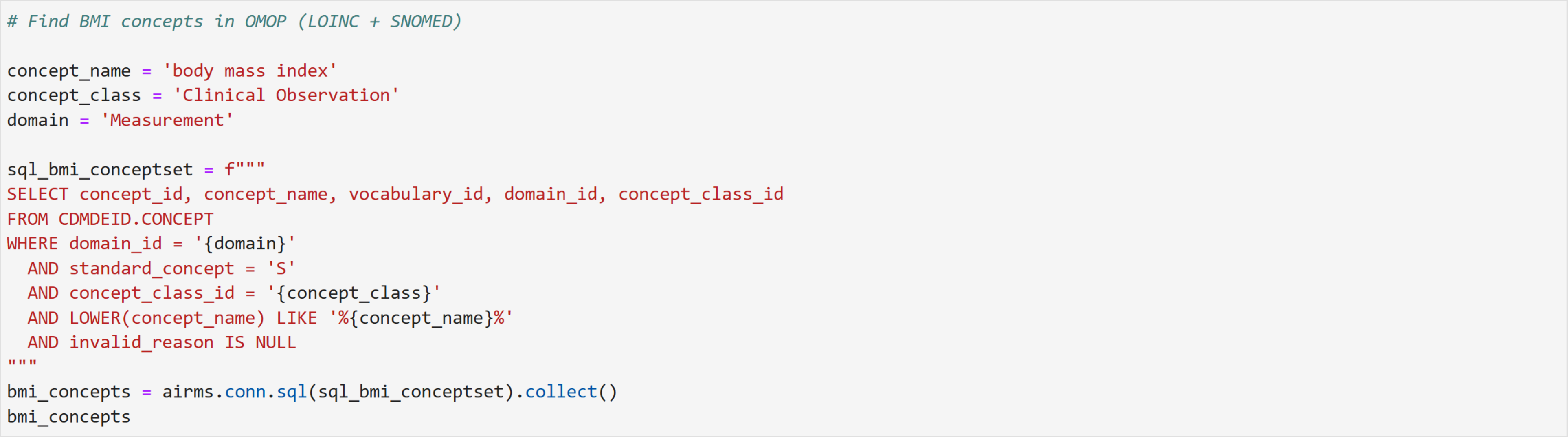
Cohort anchor: first AF diagnosis (`AF_INDEX_DATE`).
Therapy window: antiplatelet exposure overlapping `[AF_INDEX_DATE, +6 months]` .
Pacemaker outcome: first pacemaker procedure on/after `AF_INDEX_DATE` .
BMI: nearest measurement within ±180 days of `AF_INDEX_DATE` .

Preparing the data ¶

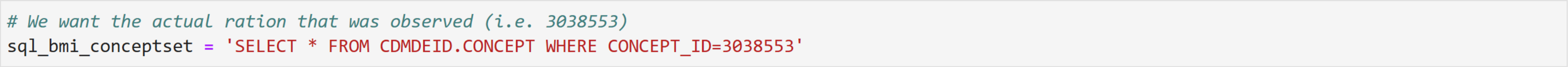
```
# Build final dataset (called cohort)
cohort  = pm_af_ap_age_sex
```

```
# Build by_month & cumulative counts to extract important overall metrics
af_series = pd.to_datetime(cohort['AF_INDEX_DATE'], errors='coerce').dropna()
by_month = (
    af_series.dt.to_period('M')
    .value_counts()
    .sort_index()
    .to_timestamp()
)
cum_by_month = by_month.cumsum()

# Extract number of unique patients in the cohort
N = cohort['PERSON_ID'].nunique()

# Extract age statistics
age_vals = cohort['AGE_AT_AF'].dropna().astype(float)
age_med  = np.nanmedian(age_vals) if len(age_vals) else np.nan
```

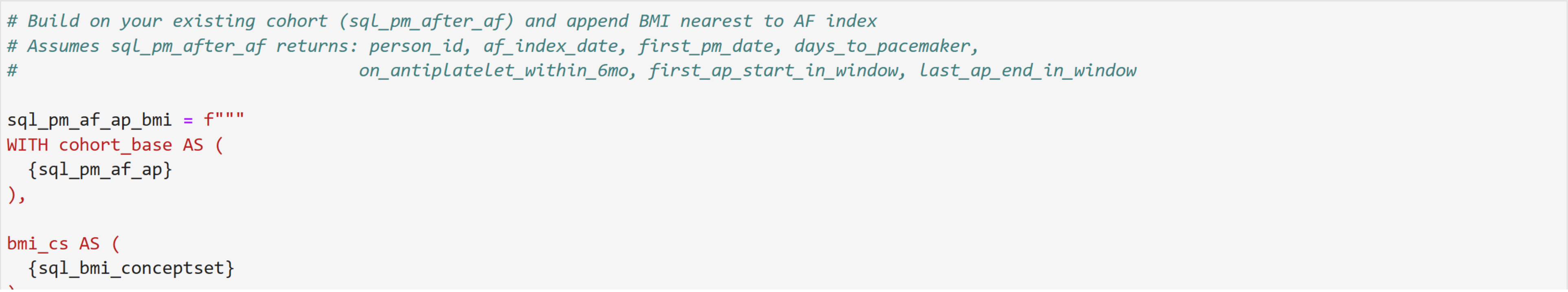
Using the Data: Querying and Retrieving Data



```
# Find BMI concepts in OMOP (LOINC + SNOMED)

concept_name = 'body mass index'
concept_class = 'Clinical Observation'
domain = 'Measurement'

sql_bmi_conceptset = f"""
SELECT concept_id, concept_name, vocabulary_id, domain_id, concept_class_id
FROM CDMDEID.CONCEPT
WHERE domain_id = '{domain}'
  AND standard_concept = 'S'
  AND concept_class_id = '{concept_class}'
  AND LOWER(concept_name) LIKE '%{concept_name}%'
  AND invalid_reason IS NULL
"""
bmi_concepts = airms.conn.sql(sql_bmi_conceptset).collect()
bmi_concepts
```



```
# We want the actual ration that was observed (i.e. 3038553)
sql_bmi_conceptset = 'SELECT * FROM CDMDEID.CONCEPT WHERE CONCEPT_ID=3038553'
```



```
# Build on your existing cohort (sql_pm_after_af) and append BMI nearest to AF index
# Assumes sql_pm_after_af returns: person_id, af_index_date, first_pm_date, days_to_pacemaker,
#                                  on_antiplatelet_within_6mo, first_ap_start_in_window, last_ap_end_in_window

sql_pm_af_ap_bmi = f"""
WITH cohort_base AS (
  {sql_pm_af_ap}
),

bmi_cs AS (
  {sql_bmi_conceptset}
```

Observations

Patient Demographics		Surgical History		Social History		Family History		Patient Allergy		observation	
Patient A	Race									Patient A	Race
	Ethnicity										Ethnicity
	Language Preference										Language Preference
	Sexual Orientation										Sexual Orientation
Patient B	Race									Patient B	Race
	Marital Status										Marital Status
	Gender Identity										Gender Identity
Patient C	Ethnicity									Patient C	Ethnicity
	Religious Affiliation										Religious Affiliation
		Patient A	Procedure Z							Patient A	Procedure Z
			Procedure Y								Procedure Y
		Patient B	Procedure X							Patient B	Procedure X
			Procedure W								Procedure W
				Patient C	Soc Hx Item 1-1					Patient C	Soc Hx Item 1-1
					Soc Hx Item 1-2						Soc Hx Item 1-2
					Soc Hx Item 1-3						Soc Hx Item 1-3
						Patient D	Fam Hx 2-1			Patient D	Fam Hx 2-1
							Fam Hx 2-2				Fam Hx 2-2
								Patient E	Allergy 3-1	Patient E	Allergy 3-1

Observation Record Types

xtn_observation_type_source_concept_id	xtn_observation_type_source_concept_name	row_count
2002056205	Allergy	4,277,961
2000000128	Family History	65,724,728
2002067234	Flowsheet Observation	114,620,841
2000000124	Past Medical History	8,610,801
2000000116	Patient Demographics	64,233,860
2000000118	Social History	145,273,185
2000000126	Surgical History Procedure	3,681,599

Record counts as of April 21, 2025

```
SELECT
  t.xtn_observation_type_source_concept_id
, t.xtn_observation_type_source_concept_name
, FORMAT(COUNT_BIG(*), 'N0') AS row_count
FROM omop.cdm_phi.observation t
WHERE t.observation_id <> 0
GROUP BY
  t.xtn_observation_type_source_concept_id
, t.xtn_observation_type_source_concept_name
ORDER BY
  t.xtn_observation_type_source_concept_name
```

Putting it Together: Queries & Phenotypes

Querying the Data & Phenotypes

- When you're researching a disease (ex. Lyme disease) – just looking for a particular International Classification of Diseases (ICD) code or disease keyword ("lyme") is not enough
  - You need to consult clinical practice guidelines
  - Sometimes there's an initial diagnosis that's wrong – you may need to look for lab results (ex. 2 separate diagnoses of diabetes within 6 months of each other, and an HbA1c level)
  - So, you might write a query that looks for patients with three Type 2 diabetes codes within a 6-month period, and two HbA1c levels above a certain threshold, rather than simply looking for patients with one Type 2 diabetes code
  - When possible, consult physicians and clinicians when devising how to identify patients with a particular disease or phenotype, peer-reviewed publications are also a good source of information
  - For example – NIH N3C specific codes for COVID-19: https://tinyurl.com/4dwruasx

Checklist

Item	Comments
Check if Patient IDs are deleted, merged, orphan, or substituted, where possible check against MPI	MPI = master patient index
When selecting codes, consult a clinician or use peer-reviewed literature	
Do not use keyword searches on vocabularies	Ex. "stroke" vs. "cerebral ischemia"
Use mapping/ vocabulary browser tools when necessary: Athena, Bioportal, RxNav, etc.	Remember: Epic IDs are licensed, so you won't find public mappings for Epic IDs vs. other coding systems
When constructing your digital phenotype, use clinical practice guidelines to inform the development of your phenotype	Get a clinician to help you understand a clinical practice guideline

Thank You