



THE DRUG DISCOVERY
ENGINE FOR **ALS**

OHDSI Training Day 2

Data Quality & Concept Sets

Danielle Boyce, MPH, DPA
Principal Investigator, RWE



Gratitude

- Parsa Mirhaji
- Rimma Belenkaya
- Chenyre Okpara
- Jessica Sommer
- Ferris Hussein
- Pavel Goriacko
- OHDSI Symposium Trainers for sharing lecture materials which were adapted for this training

...and all of you!

HOUSEKEEPING

- We are recording so you can go back and watch later
- Slides will be on the GitHub
- Please open up:
 - Search/Atlas
 - Athena: <https://athena.ohdsi.org/>
 - Databricks, if applicable
 - Code snippets and keywords here:
<https://boycelab.github.io/OHDSITrainTheTrainer-6WK/exercises/day-02-vocab-dqd/>
- CDM specs: <https://ohdsi.github.io/CommonDataModel/cdm54.html> -
 - **We will drop these links in the chat**
- Please put questions in the chat.
 - Check out Book of OHDSI and OHDSI Forums!
 - If we cannot get to your question today, we will follow up before next week.

“PHONE A FRIEND”:
dboyce@als.net

Schedule

Time	Session Title
9:30 am – 10:15 am	Welcome, Kahoot!, homework review
10:15 am – 10:45 am	Data quality, including Data Quality Dashboard
10:45 am – 11:15 am	Atlas tour
11:15 am – 11:30 am	Break
11:30 am – 12:30 pm	Hands-on: concept set development
12:30 pm – 12:45 pm	Review & discussion
12:45 pm – 1:00 pm	Homework, next steps, preview of next week

Kahoot!

THE DRUG DISCOVERY
ENGINE FOR ALS

ALS
THERAPY DEVELOPMENT
INSTITUTE

Homework Question 1

Exercise 5.1 What is the Standard Concept ID for “Gastrointestinal hemorrhage”?

Homework Question 2

Exercise 5.2 Which ICD-10CM codes map to the Standard Concept for “Gastrointestinal hemorrhage”? Which ICD-9CM codes map to this Standard Concept?

A note about 5.3

Exercise 5.3 What are the MedDRA preferred terms that are equivalent to the Standard Concept for “Gastrointestinal hemorrhage”?

So I went to the forums

Questions about book exercises ch. 4 and 5

General



JDufhcc

Mar 2021

Mar 2021

1 / 3

Mar 2021

I'm very new to OHDSI, and am taking an introductory class. I'm curious to know about two answer choices in Chapter 4 and one in Chapter 5.

In Chapter 4, the listed answer in exercise 4.2 for PERIOD_TYPE_CONCEPT_ID is 44814722 (period while enrolled in insurance; a type concept / observation period type). I had thought 4085479 (private insurance held; an observation/clinical finding). I went with the latter because it was a Standard Concept, rather than the former, which was nonstandard.

Also in Chapter 4, exercise 4.3, for DRUG_TYPE_CONCEPT_ID I had listed 4126021 (prescription; observation/attribute; a Standard Concept), whereas the answer was 38000177 (prescription written; type concept/drug type), which was again a nonstandard concept.

I would appreciate clarification on the rationale in these cases!

In Chapter 5, exercise 5.3, I could not find any reference to the MedDRA terms in either ATLAS or Athena. What is the best way to search these terms?

Apr 2021



So I went to the forums

 **OHDSI** Questions about book exercises ch. 4 and 5

General

sts ●

 Chapter 5

MedDRA is licence restricted, so the concepts can't be seen without having a licence.

I'll create an issue [here](#) 2

A screenshot of a web browser window. The address bar at the top shows the URL: https://github.com/OHDSI/TheBookOfOhdsi/issues/137. Below the address bar is the GitHub header with the organization logo (OHDSI), the repository name ('TheBookOfOhdsi'), and a search bar with the placeholder 'Type / to search'. The main content area displays the issue details for issue #137. At the bottom of the page, there is a navigation bar with links: Code, Issues (highlighted in orange), Pull requests, Actions, Projects, Security, and Insights.

Meddra shouldn't be a part of the excercise as it's linence restriction vocabulary #137

Open



dimshitc opened on Apr 1, 2021

•

<https://forums.ohdsi.org/t/questions-about-book-exercises-ch-4-and-5/13676>

Chapter 5, exercise 5.3.

Meddra shouldn't be a part of the excercise as it's linence restricted vocabulary



Add a comment

Write

Preview

H B I | ⌂ <> 🔗 | ⏪ ⏴ ⏵ ⏹ | @ ↗ ↘ ↙ ↖

Well Done!



THE DRUG DISCOVERY
ENGINE FOR **ALS**

ALS
THERAPY DEVELOPMENT
INSTITUTE

“We release the stress of the week, trust in what we’ve learned so far, and step into this session with confidence, openness, and renewed energy.”

Today: Data Quality & Concept Sets

After today's training, you will be able to:

- Describe key OHDSI data quality principles including the Kahn framework, related tools, and their relevance to CDM curation.
- Construct and evaluate concept sets in Atlas.
- Explain how standardized vocabularies and concept sets enable consistent, reproducible cohort definitions in OHDSI.
- Recognize how concept set design and data quality practices together support trustworthy observational research.

Data Quality

Challenges: Data collection

- **Source data collection**: Health care data are collected to support patient care or to bill payors rather than for research.

Challenges: Data collection

- **Source data collection:** Health care data are collected to support patient care or to bill payors rather than for research.
 - The opportunities for errors of omission and distortion are greater than when data are collected for research.

Challenges: Data collection

- **Source data collection:** Health care data are collected to support patient care or to bill payors rather than for research.
 - The opportunities for errors of omission and distortion are greater than when data are collected for research.
 - The power to standardize and improve data collection methods is less than when it is collected for research.

Challenges & Solutions: Data collection

Source data are heterogeneous: OMOP to the rescue!

- Equivalent codes get mapped to a standard concept.
- Standard representation yields
 - *Semantic interoperability*
 - *Common schema to write code against*
 - *The ability to leverage concept relationships in queries*

Challenges & Solutions: Data normalization

BUT!

Mapping source data to the OMOP CDM is complex!

Challenges & Solutions: Data normalization

BUT!

Mapping source data to the OMOP CDM is complex!

- It is easy to make mistakes when writing ETL code
 - The Rabbit-in-a-Hat tool supports the creation of unit tests – small bits of code that checks whether it functions as intended.

Challenges & Solutions: Data normalization

BUT!

Mapping source data to the OMOP CDM is complex!

- It is easy to make mistakes when writing ETL code
 - The Rabbit-in-a-Hat tool supports the creation of unit tests – small bits of code that checks whether it functions as intended.
 - Various studies have shown that with scrupulous attention, data can be transformed to the CDM with very little information loss.
 - These studies are cited in the Book of OHDSI

Challenges & Solutions: Data normalization

Mapping source data to the OMOP CDM is complex!

BUT!

- Even when coding mistakes are not made, there are many cases where there is more than one defensible way to do the right thing.

Challenges & Solutions: Data

The screenshot shows a web browser window with the URL <https://ohdsi.github.io/Themis/> in the address bar, which is circled in red. The page itself is a GitHub repository for 'THEMIS Repository'. The header includes links for 'Nav', 'THEMIS GitHub', 'CDM Docs', 'DQD Docs', 'Feedback', and a search bar. The main content area features a large heading 'THEMIS' and a 'Summary' section describing the project's overview.

THEMIS Conventions

- General Conventions ▾
- CDM Tables ▾
- Tag Browser ▾

THEMIS

Summary: This page provides an overview of the THEMIS project.

Table of Contents

- [Introduction](#)
 - What falls under Themis and what does not
- [Topic Processing](#)
- [Getting Involved](#)

Introduction

OHDSI Community: Themis makes decisions for the good of the whole community. We must compromise. We can always revisit and modify the convention. Don't let perfect be the enemy of great. And interoperability between different OMOP CDMs is great!

What falls under Themis and what does not

Some questions regarding the CDM would fall under the CDM WG domain, Vocab Team's domain, and/or other areas of the OHDSI community.

Challenges & Solutions: Data collection and normalization

- Healthcare data are prone to omissions and distortions
- Mapping source data to CDM is complex
- There are an enormous number of concepts in each domain and datasets are often very large

Kahn harmonized framework for data quality

Kahn and colleagues did an excellent job of synthesizing the terminology and categories used to conceptualize the data quality errors that affect RWD.

SOURCE: [A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data - PubMed](#)

Kahn framework for data quality

- **Conformance:** Do data values adhere to specified standards and formats?

Kahn framework for data quality

- **Conformance:** Do data values adhere to specified standards and formats?
- **Completeness:** Is a particular variable present OR does it contain all recorded values?

Kahn framework for data quality

- **Conformance:** Do data values adhere to specified standards and formats?
- **Completeness:** Is a particular variable present OR does it contain all recorded values?
- **Plausibility:** Are data values believable?

Kahn framework for data quality

Conformance: Adherence to specified standards and formats

- Value
- Relational
- Computation

Completeness: Variable presence OR capture of all recorded values

Plausibility: Values believability

- Uniqueness
- Atemporal
- Temporal

Kahn framework for data quality

- **Verification:** assesses expected values and distributions using resources within the local environment.
- **Validation:** assesses alignment of data values with respect to relevant external benchmarks such as across multiple data sites

Other challenges: Expectations

- People bring the same expectations to healthcare data quality as they do to assessing data collected explicitly for research.
 - The criteria for assessing clinical data warehouse should not be perfection, it should transparency.
 - The goals should be to identify where there might be problems due to collection or ETL coding errors or divergence from conventions and to facilitate actions that address those problems.
- Understanding data provenance completely is desirable, but it might not be necessary for a fulsome assessment of relevant DQ problems when producing RWE.

Goals: Assess whether data are fit for use

- FDA RWE program
- OHDSI Data Quality Assessment

JCO® Clinical Cancer Informatics
An American Society of Clinical Oncology

FREE ACCESS | EDITORIAL

The Bar Is High for Real-World Data

Authors: Catherine C. Lerro, PhD



AUTHORS INFO & A

JCO Clin Cancer Inform 8, e1

Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

August 2023
Real-World Data/Real-World Evidence (RWD/RWE)

data sources or databases is an important step in assessing a study's feasibility. Such evaluations of data sources serve as a way for the sponsor and FDA to determine if it is feasible to use the data to address the research question without impacting the precision of a potential study without.

ARTICLE



Cancer Clinical Trials and Real-World Data

Authors: Catherine C. Lerro, PhD, and Donna R. Rivera, PharmD, MSc, FISPE

THE DRUG DISCOVERY
ENGINE FOR ALS

ALS
THERAPY DEVELOPMENT
INSTITUTE

Where to begin with Data Quality?

CATEGORIES	CONTEXTS	
	Verification	Validation
	Plausibility	?
	Conformance	?
Completeness		?

Where to begin with Data Quality?

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

Data Quality Check

An aggregated summary statistic that can be computed from the data to which a decision threshold can be applied to determine if the statistic meets expectation.

An example data quality check...

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

The number and percent of records with
a value in the **YEAR_OF_BIRTH** field of the
PERSON table less than **1850**.

...which we can make more generic...

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

The number and percent of records

with a value in the ***CDM field*** of the

CDM table less than ***a low value.***

...and apply to a different example.

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

The number and percent of records with

a value in the **DAY_SUPPLY** field of the

DRUG_EXPOSURE table less than **0**.

What if we add units?

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

For a *measurement with associated unit*,

the number and percent of records with a

value in the *CDM field* of the *CDM table*

less than *a low value*.

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

What if we add units?

For Hemoglobin A1c with unit of percent,

the number and percent of records with a

value in the VALUE_AS_NUMBER field of the

MEASUREMENT table less than 4.

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

An example completeness check...

The number and percent of records which are not mapped into a standard concept in the **CONDITION_CONCEPT_ID** field of the **CONDITION_OCCURRENCE** table.

...which we can make more generic...

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

The number and percent of records which
are not mapped into a standard concept in
the *CDM field* of the *CDM table*.

...and apply to a different example.

	Verification	Validation
Plausibility	?	?
Conformance	?	?
Completeness	?	?

- The number and percent of records which are not mapped into a standard concept in the **UNIT_CONCEPT_ID** field of the **MEASUREMENT** table.

Data Quality Check Types

Check Type	Check Description
Person Completeness	The number and percent of persons in a database that do not have at least one record in the <i>CDM table</i> .
Is Required	The number and percent of records with a NULL value in a <i>CDM field</i> of a <i>CDM table</i> that is considered not nullable.
Is Foreign Key	The number and percent of records that have a value in a foreign key <i>CDM field</i> of a <i>CDM table</i> that does not exist in the <i>foreign key table</i> .
Is Standard Valid Concept	The number and percent of records that do not have a standard, valid concept in the <i>CDM field</i> <i>CDM table</i> .
Plausible Temporal After	The number and percent of records with a value in a <i>CDM field</i> of a <i>CDM table</i> that occurs prior to a <i>plausible date</i> .
...	
Plausible Value Low	For a given <i>CONCEPT_ID</i> and <i>UNIT_CONCEPT_ID</i> pair, the number and percent of records with a value lower than the <i>plausible low value</i> .
Plausible Gender	For a given <i>CONCEPT_ID</i> , the number and percent of records associated with persons with an <i>implausible gender</i> .

Data Quality Check *Types*

Data Quality Check *Types*

	Verification	Validation
Plausibility	6	1
Conformance	7	1
Completeness	4	1

20 Check *Types*

Data Quality Check *Totals*

	Verification	Validation
Plausibility	1878	287
Conformance	681	104
Completeness	386	15

Total 3,351 Checks

Data Quality Check *Totals*

	Verification	Validation
Plausibility	1878	287
Conformance	Total 3,351 Checks	
Completeness	386	15

Data Quality Check

An aggregated summary statistic that can be computed from the data

to which a decision threshold can be applied to determine if the statistic meets expectation.

Data Quality Check Thresholds

Check Category	Check Description	Check Result
Verification - Plausibility	The number and percent of records with a value in the YEAR_OF_BIRTH field of the PERSON table less than 1850.	0%
Verification - Plausibility	The number and percent of records with a value in the DAYS_SUPPLY field of the DRUG_EXPOSURE table less than 0.	0%
Verification - Plausibility	For Hemoglobin A1c percent, the number and percent of records with a value in the VALUE_AS_NUMBER field of the MEASUREMENT table less than 4.	0.01%
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field CONDITION_CONCEPT_ID in the CONDITION_OCCURRENCE table.	0.02%
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field UNIT_CONCEPT_ID in the MEASUREMENT table.	93.66%

Data Quality Check Thresholds

Check Category	Check Description	Check Result
Verification - Plausibility	The number and percent of records with a value in the YEAR_OF_BIRTH field of the PERSON table less than 1850.	0%
Verification - Plausibility	How do we decide if these results are 'good enough'?	
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field CONDITION_CONCEPT_ID in the CONDITION_OCCURRENCE table.	0.01%
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field UNIT_CONCEPT_ID in the MEASUREMENT table.	0.02%
		93.66%

Data Quality Check *Thresholds*

Check Category	Check Description	Check Result	Decision Threshold	Pass / Fail
Verification - Plausibility	The number and percent of records with a value in the YEAR_OF_BIRTH field of the PERSON table less than 1850.	0%	0%	PASS
Verification - Plausibility	The number and percent of records with a value in the DAYS_SUPPLY field of the DRUG_EXPOSURE table less than 0.	0%	1%	PASS
Verification - Plausibility	For Hemoglobin A1c percent, the number and percent of records with a value in the VALUE_AS_NUMBER field of the MEASUREMENT table less than 4.	0.01%	5%	PASS
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field CONDITION_CONCEPT_ID in the CONDITION_OCCURRENCE table.	0.02%	5%	PASS
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field UNIT_CONCEPT_ID in the MEASUREMENT table.	93.66%	5%	FAIL



Data Quality Check *Thresholds*

Check Category	Check Description	Check Result	Decision Threshold	Pass / Fail	
Verification - Plausibility	The number and percent of records with a value in the YEAR_OF_BIRTH field of the PERSON table less than 1850.	0%	0%	PASS	
Verification - Plausibility	The number and percent of records with a value in the DAYS_SUPPLY field of the DRUG_EXPOSURE table less than 0.	0%	1%	PASS	
Verification - Plausibility	For Hemoglobin A1c percent, the number and percent of records with a value in the VALUE_AS_NUMBER field of the MEASUREMENT table less than 4.	0.01%	5%	PASS	
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field CONDITION_CONCEPT_ID in the CONDITION_OCCURRENCE table.	0.02%	5%	PASS	
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field UNIT_CONCEPT_ID in the MEASUREMENT table.	93.66%	95%	PASS	

Data Quality Dashboard

OHDSI / DataQualityDashboard

Unwatch 7 Star 5 Fork 6

Code Issues 16 Pull requests 0 Actions Projects 1 Wiki Security Insights Settings

A tool to help improve data quality standards in observational data science. <https://ohdsi.github.io/DataQualityDa...> Edit

data-quality Manage topics

150 commits 4 branches 0 releases 1 environment 6 contributors Apache-2.0

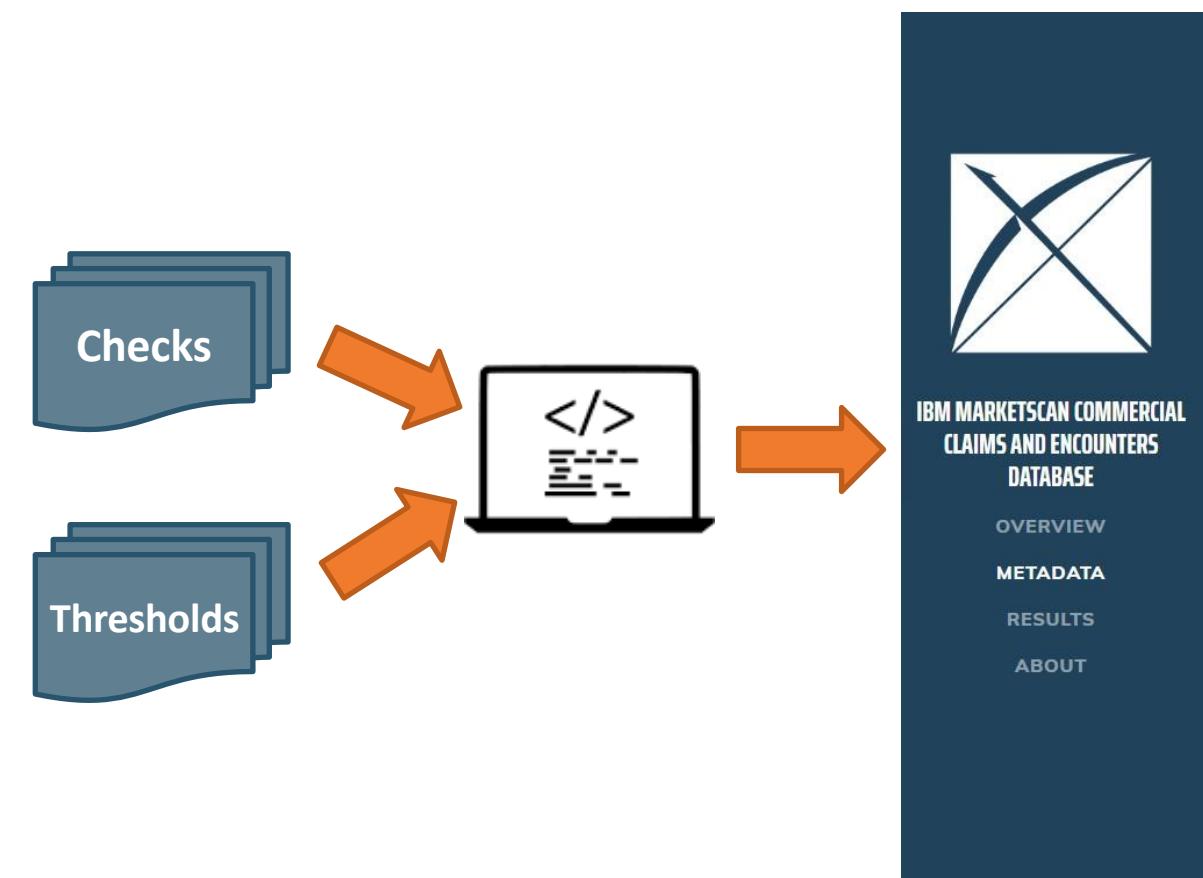
Branch: master New pull request Create new file Upload files Find File Clone or download

File	Commit Message	Time Ago
R	Disable scientific notation in execution	12 hours ago
docs	Added tablesToExclude parameter to allow skipping of tables if known ...	13 days ago
extras	Added screenshot	18 days ago
inst	Remove duplicates from concept level	4 days ago
man	Added tablesToExclude parameter to allow skipping of tables if known ...	13 days ago
tests	Added missing dbType in test	18 days ago
.Rbuildignore	updated R package wrappers	2 months ago
.gitignore	Fixes #30, making autocommit setting specific to connectionDetails db...	18 days ago
.travis.yml	Added devtools to travis. Added travis and codecov statuses to readme...	18 days ago
DESCRIPTION	Added pkgdown documentation. Added one testthat test.	19 days ago
DataQualityDashboard.Rproj	updated R package wrappers	2 months ago
LICENSE	Initial commit	3 months ago
NAMESPACE	Added pkgdown documentation. Added one testthat test.	19 days ago

[https://github.com/OHDSI/
DataQualityDashboard](https://github.com/OHDSI/DataQualityDashboard)



Data Quality Dashboard



RESULTS

IBM MARKETSCAN COMMERCIAL CLAIMS AND ENCOUNTERS DATABASE

Results generated at 2019-09-06 22:20:12 in 7 hours

Show 5 entries

STATUS	CONTEXT	CATEGORY	SUBCATEGORY	LEVEL	DESCRIPTION	% RECORDS
PASS	Verification	Completeness	None	FIELD	The number and percent of records with a NULL value in the range_high of the MEASUREMENT. (Threshold=100%).	82.14%
PASS	Verification	Completeness	None	FIELD	The number and percent of records with a NULL value in the visit_detail_id of the MEASUREMENT. (Threshold=100%).	80.90%
PASS	Verification	Completeness	None	FIELD	The number and percent of records with a NULL value in the value_source_value of the MEASUREMENT. (Threshold=100%).	79.89%
PASS	Validation	Completeness	None	TABLE	The number and percent of persons in the CDM that do not have at least one record in the DEVICE_EXPOSURE table (Threshold=100%).	76.70%
FAIL	Verification	Plausibility	Atemporal	CONCEPT	For the combination of CONCEPT_ID 3016049 (Testosterone Free [Mass/volume] in Serum or Plasma) and UNIT_CONCEPT_ID 8845 (picogram per milliliter), the number and percent of records that have a value less than 5.00e+00. (Threshold=1%).	72.43%

Showing 126 to 130 of 3,351 entries

Previous 1 ... 25 26 27 ... 671 Next

RESULTS



SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-22 14:15:06 in 29 mins

[Column visibility](#)[CSV](#)

Show 5 entries

Search:

STATUS	CONTEXT	CATEGORY	SUBCATEGORY	LEVEL	DESCRIPTION	% RECORDS
FAIL	Validation	Completeness	None	TABLE	The number and percent of persons in the CDM that do not have at least one record in the DEVICE_EXPOSURE table (Threshold=0%).	100.00%
+ FAIL	Validation	Completeness	None	TABLE	The number and percent of persons in the CDM that do not have at least one record in the VISIT_DETAIL table (Threshold=0%).	100.00%
+ FAIL	Validation	Completeness	None	TABLE	The number and percent of persons in the CDM that do not	100.00%

SYNTHEA SYNTHETIC HEALTH DATABASE

[OVERVIEW](#)[METADATA](#)[RESULTS](#)[ABOUT](#)

Study specific checks

%% per month	Max monthly %	Person count	Description
	26.81	92,019,885	Depressive Disorder
	6.64	15,969,198	Depressive disorder 440383
	6.64	15,686,275	311 (ICD9CM) Depressive disorder, not elsewhere classified
	0.46	188,230	F328 (ICD10CM) Other depressive episodes
	0.38	94,693	F3289 (ICD10CM) Other specified depressive episodes
	3.10	12,010,783	Adjustment disorder with mixed emotional features 433454
	3.07	9,839,712	30928 (ICD9CM) Adjustment disorder with mixed anxiety and depressed mood
	3.03	2,049,618	F4323 (ICD10CM) Adjustment disorder with mixed anxiety and depressed mood
	0.04	121,453	3091 (ICD9CM) Prolonged depressive reaction
	3.17	9,237,192	Dysthymia 433440

Case Study: Data Quality Checks – Cure ID

- SQL scripts
- Clinical validity: after generation, number of unique patients in the cohort with certain concepts e.g. acetaminophen
- “Easy” fix: ventilator mappings, procedures vs device, unmapped drugs

Is it data quality or is it your concept sets/cohort definition?

- You will learn this next week!
- Generate the cohort – is it a lot more/less people than expected?

Data Quality Summary

- Most observational healthcare data were not collected for research.
- Data quality checks are an integral part of research. Data quality must be assessed to determine whether the data are of sufficient quality for research purposes.
- We should assess data quality for the purpose of research in general, and critically in the context of a specific study.
- Some aspects of data quality can be assessed automatically through large sets of predefined rules, for example those in the Data Quality Dashboard.
- Other tools exist to evaluate the mapping of codes relevant for a particular study.
- Sometimes we may choose to make our own tools within the context of a particular study

Atlas

THE DRUG DISCOVERY
ENGINE FOR ALS

ALS
THERAPY DEVELOPMENT
INSTITUTE

Concept Sets

Concept Set

- **Exclude:** Exclude this concept (and any of its descendants if selected) from the concept set.
- **Descendants:** Consider not only this concept, but also all of its descendants.
- **Mapped:** Allow to search for non-standard concepts.

What's the deal with the “mapped” field?

Concept Set #1888951
created by anonymous on 2025-11-12 11:29, modified by anonymous on 2025-11-12 11:29

New Concept Set Diabetes 

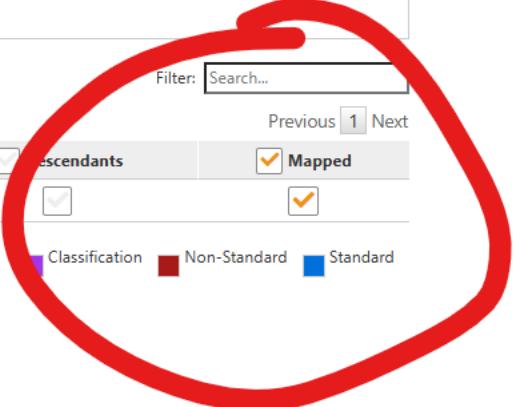
Concept Set Expression 

Enter the concept set description here

Show 50 entries Filter: Search... Previous 1 Next

<input type="checkbox"/> Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	<input type="checkbox"/> Exclude	<input type="checkbox"/> Descendants	<input checked="" type="checkbox"/> Mapped
<input checked="" type="checkbox"/> 201820	73211009	Diabetes mellitus	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Classification Non-Standard Standard

 Remove Selected Add Concepts

Case Study: Semaglutide & Naion



Original Investigation



Semaglutide and Nonarteritic Anterior Ischemic Optic Neuropathy

Cindy X. Cai, MD, MS^{1,2}; Michelle Hribar, PhD^{3,4}; Sally Baxter, MD, MSc^{5,6}; *et al*

[» Author Affiliations](#) | [Article Information](#)

RELATED ARTICLES

FIGURES

SUPPLEMENTAL CONTENT

Key Points

Question Is semaglutide use associated with the nonarteritic anterior ischemic optic neuropathy (NAION)?

README



🔗 Estimation of risk of NAION and other vision disorders from exposure to semaglutide

Study Status Results Available

- Analytics use case(s): Population-Level Estimation
- Study type: Clinical Application
- Tags: Eye Care & Vision Research, Type 2 Diabetes
- Study lead: Cindy X. Cai
- Study lead forums tag: [\[Cindy X. Cai\]](#)
- Study start date: -
- Study end date: -
- Protocol: [HTML Document](#)
- Publications: -
- Results explorer: <https://results.ohdsi.org/>

OHDSI network study for population-level effect estimation of risk of NAION and other vision disorders from exposure to semaglutide.

🔗 How to run the study

The following instructions will guide you through the process of setting up your system to run this network study.

8.2 Study Population

We will include all adults (≥ 18 years) with type 2 diabetes mellitus in each data source who meet the cohort entry criteria for each exposure cohort (detailed below).

8.3 Exposures

The exposure is receipt of semaglutide (GLP-1 RA), dulaglutide (GLP-1RA), exenatide (GLP-1 RA), empagliflozin (a SGLT2 inhibitor), sitagliptin (a DPP4 inhibitor), or glipizide (a sulfonylurea) based on the definitions of LEGEND-T2DM. (<https://ohdsi-studies.github.io/LegendT2dm/Protocol>) For the new-user active comparator cohort design, these will include patients with T2DM with prior metformin monotherapy who initiate treatment with one or more medications. Patients have to have one year of observation prior to the index date, no prior drug exposure to a comparator second-line or anti-diabetic agents, and at most 30 days of insulin exposure before the index date.

As a sensitivity analysis, we will examine patients with each of these anti-diabetic medication exposures but without the second-line therapy requirement, meaning that they can be on another anti-diabetic medication at the time of cohort entry.

8.4 Outcomes

The outcomes are NAION and DR worsening.

NAION will have two definitions, one designed to be more specific and one designed to be more sensitive:

- NAION-specific:
 1. 2 instances of ION diagnosis codes on separate days (within 90 days of each other)
 2. If the 2 diagnoses of ION are preceded by the following diagnoses within the prior 60 days, then use the occurrence of these prior diagnosis to mark the start date of the outcome
 - Visual field defect
 - Optic disc disorder
 - Includes concepts of optic papillitis (concept ID 435269), and “other disorders of optic disc”
 - Optic neuritis
 - Includes “Unspecified” and “Other” optic neuritis
 - Optic disc edema

Protocol Template for Observational Study

Directions for Using this Template:

- This template should be used for studies involving only observational interactions. If your study also involves testing an intervention, use the Protocol Template for Interventional Behavioral Studies or Interventional Clinical Trials as appropriate. Choose the appropriate sections that

5.1 Inclusion Criteria

In order to be eligible to participate in this study, an individual must meet all of the following criteria:

1. <Insert Text>

Create a numbered list of criteria that an individual must meet to be eligible to participate in the study.

Some criteria to consider for inclusion are: provision of appropriate consent and assent, willingness and ability to participate in study procedures, age range, gender, health status, diagnosis or symptoms, background medical treatment, and laboratory ranges. Additional criteria should be included as appropriate for the study design and risk.

5.2 Exclusion Criteria

An individual who meets any of the following criteria will be excluded from participation in this study:

1. <Insert Text>

Create a numbered list of criteria that would exclude an individual from study enrollment. Some criteria to consider for exclusion are: pre-existing conditions or concurrent diagnoses, concomitant use of medication(s) or devices, other factors that would cause harm or increased risk to the participant or close contacts, or preclude the participant's full adherence with or completion of the study. Additional criteria should be included as appropriate for the study design and risk.

B.1.2 Additional Inclusion Criteria

I. Prior metformin use

Entry events with any of the following criteria:

1. having at least 1 drug era of 'Metformin', starting anytime up to 90 days before cohort entry start date; allow events outside observation period; with era length \geq 90 days.
2. having at least 3 drug exposures of 'Metformin', starting anytime on or before cohort entry start date; allow events outside observation period.

II. No prior GLP-1 RA exposure

Entry events having no drug exposures of 'GLP-1 receptor agonists excluding semaglutide', starting anytime on or before cohort entry start date; allow events outside observation period.

III. No prior DPP4 inhibitor exposure

Entry events having no drug exposures of 'DPP4 inhibitors', starting anytime on or before cohort entry start date; allow events outside observation period.

IV. No prior SGLT-2 inhibitor exposure

Entry events having no drug exposures of 'SGLT2 inhibitors', starting anytime on or before cohort entry start date; allow events outside observation period.

V. No prior SU exposure

Entry events having no drug exposures of 'Sulfonylureas', starting anytime on or before cohort entry start date; allow events outside observation period.

VI. No prior other anti-diabetic exposure

Entry events having no drug exposures of 'Other anti-diabetics', starting anytime on or before cohort entry start date; allow events outside observation period.

VII. No prior insulin use or combo initiation: Proxy for < 30 days drug era anytime before index and no combination use on index

Please open Search/Atlas

Let's make a concept set for sulfonylureas!

B.1.8 Concept: Sulfonylureas

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1594973	chlorpropamide	2404	RxNorm	NO	YES	NO
1597756	glimepiride	25789	RxNorm	NO	YES	NO
1560171	glipizide	4821	RxNorm	NO	YES	NO
19097821	gliquidone	25793	RxNorm	NO	YES	NO
1559684	glyburide	4815	RxNorm	NO	YES	NO
1502809	tolazamide	10633	RxNorm	NO	YES	NO
1502855	tolbutamide	10635	RxNorm	NO	YES	NO



SEARCH

DOWNLOAD



Danielle Boyce



SEARCH BY KEYWORD

sulfonylureas

sulfonylureas XClassification X

● DOMAIN

● CONCEPT

 Classification (5) Non-standard (42) Standard (22)

● CLASS

● VOCAB

● VALIDITY

DOWNLOAD RESULTS

Show by 15 ▾ items Total 5 items

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
21600749	A10BB	Sulfonylureas	ATC 4th	Classification	Valid	Drug	ATC
21600767	A10BD02	metformin and sulfonylureas; oral	ATC 5th	Classification	Valid	Drug	ATC
21600766	A10BD01	phenformin and sulfonylureas; oral	ATC 5th	Classification	Valid	Drug	ATC
37071719	LP390555-3	Sulfonylurea Urine Drug toxicology	LOINC Hierarchy	Classification	Valid	Measurement	LOINC
37052577	LP390554-6	Sulfonylurea Serum or Plasma Drug toxicology	LOINC Hierarchy	Classification	Valid	Measurement	LOINC

[← Sulfonylureas](#)

DETAILS

Domain ID Drug

Concept Class ID ATC 4th

Vocabulary ID ATC



Concept ID 21600749

Concept code A10BB

Validity Valid

Classification

Valid start 01-Jan-1970

Valid end 31-Dec-2099

TERM CONNECTIONS (14)

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
Is a	BLOOD GLUCOSE LOWERING DRUGS, EXCL. INSULINS	21600744	ATC
Subsumes	acetohexamide; oral	21600762	ATC
	carbutamide; oral	21600755	ATC
	chlorpropamide; oral	21600751	ATC
	glibenclamide; oral	21600750	ATC
	glibornuride; oral	21600753	ATC
	gliclazide; oral	21600758	ATC
	glimepiride; oral	21600761	ATC
	glipizide; oral	21600756	ATC

B.1.8 Concept: Sulfonylureas

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1594973	chlorpropamide	2404	RxNorm	NO	YES	NO
1597756	glimepiride	25789	RxNorm	NO	YES	NO
4560474	tolazamide	4804	RxNorm	NO	YES	NO



SEARCH BY KEYWORD

chlorpropamide

chlorpropamide XStandard XDrug X● DOMAIN

filter

 Condition (6) Drug (75) Measurement (12) Observation (1) Condition Status (0) Condition/Device (0)● CONCEPT● CLASS

DOWNLOAD RESULTS

Show by 15 items Total 75 items

1 2 3 4 5 >

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
1594973	2404	chlorpropamide	Ingredient	Standard	Valid	Drug	RxNorm
44055054	OMOP10496 85	Chlorpropamide 100 MG [Apo Chlorpropamide]	Branded Drug Comp	Standard	Valid	Drug	RxNorm Extension
44041960	OMOP10365 91	Chlorpropamide 250 MG [Apo Chlorpropamide]	Branded Drug Comp	Standard	Valid	Drug	RxNorm Extension
44044524	OMOP10391 55	Chlorpropamide Oral Tablet [Apo Chlorpropamide]	Branded Drug Form	Standard	Valid	Drug	RxNorm Extension
1594979	315647	chlorpropamide 100 MG	Clinical Drug Comp	Standard	Valid	Drug	RxNorm
19088817	332808	chlorpropamide 125 MG	Clinical Drug Comp	Standard	Valid	Drug	RxNorm
19088819	332810	chlorpropamide 200 MG	Clinical Drug Comp	Standard	Valid	Drug	RxNorm
1594980	315648	chlorpropamide 250 MG	Clinical Drug Comp	Standard	Valid	Drug	RxNorm



chlorpropamide

DETAILS

Domain ID Drug

Concept Class ID Ingredient

Vocabulary ID RxNorm



Concept ID 1594973

Concept code 2404

Validity Valid

Concept Standard

Valid start 01-Jan-1970

Valid end 31-Dec-2099

TERM CONNECTIONS (44)



HIERARCHY

RELATED CONCEPTS

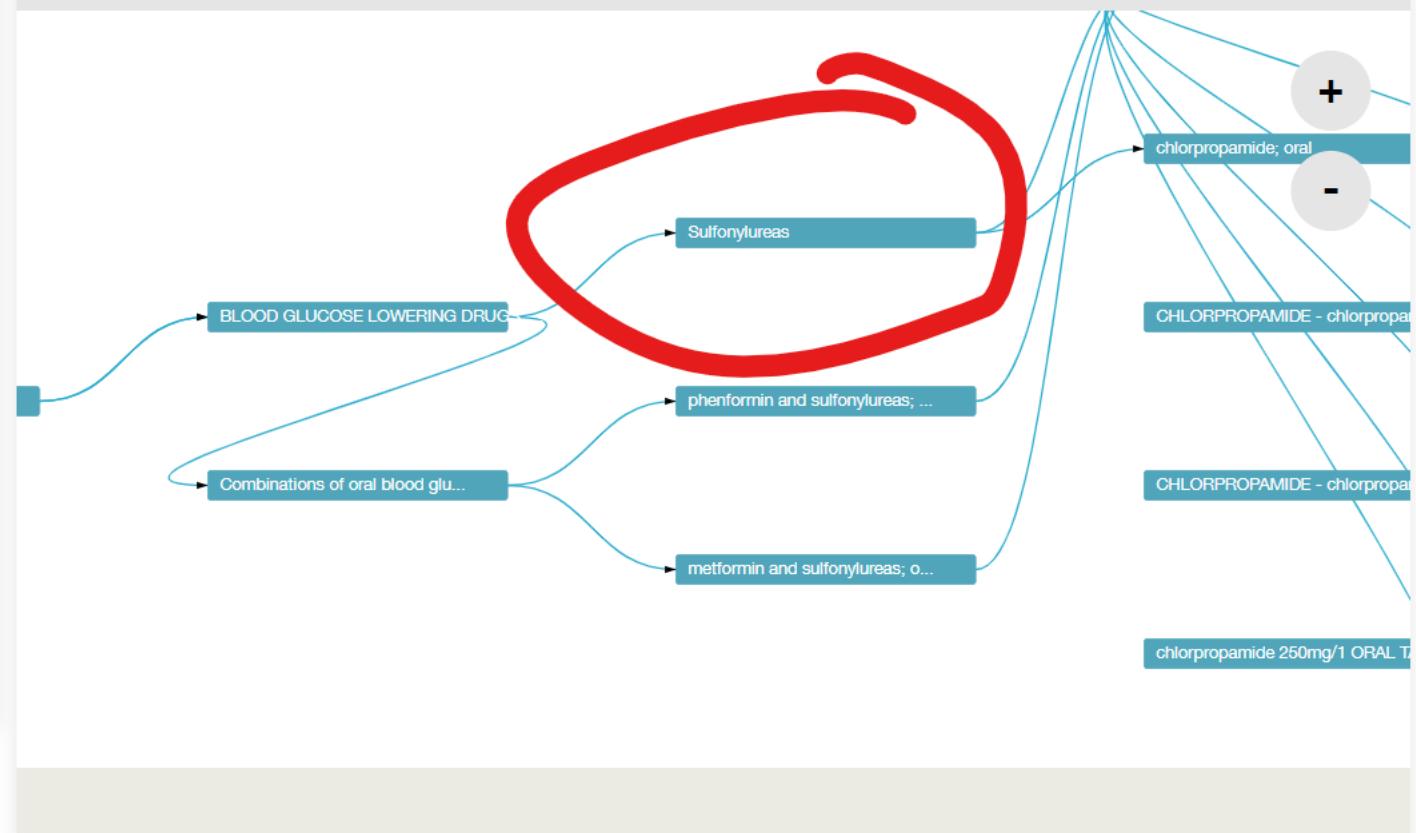
NUMBER OF PARENT LEVELS

10



LEVEL OF DETAILS

MAXIMUM





DETAILS

Domain ID Drug

Concept Class ID Ingredient

Vocabulary ID RxNorm

Concept ID 1594973

Concept code 2404

Validity Valid

Concept Standard

Valid start 01-Jan-1970

Valid end 31-Dec-2099

TERM CONNECTIONS (59)



HIERARCHY

RELATED CONCEPTS

RELATIONSHIP

Has brand name
(OMOP)

RELATES TO

Abemide

CONCEPT ID

35128307

VOCABULARY

RxNorm Extension

Apo Chlorpropamide

44013371

RxNorm Extension

Chloronase

40803418

RxNorm Extension

Diabetoral

40804601

RxNorm Extension

Diabinese

19021491

RxNorm

Insulase

19071136

RxNorm

Novo-Propamide

44020815

RxNorm Extension

Ingredient of
(RxNorm)

Chlorpropamide 50 MG/ML

21023912

RxNorm Extension

Chlorpropamide Oral Suspension

21141703

RxNorm Extension

chlorpropamide / metformin Oral Product

36214817

RxNorm

Now find this in the CDM using Databricks

Run all ✓ Just now (<1s) prod_deid_clinical. omop_deid New SQL editor: ON ⭐

```
1 | SELECT *
2 | FROM concept
3 | WHERE concept_id = 1594973;
```

Add parameter

Table +

	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date
1	1594973	chlorpropamide	Drug	RxNorm	Ingredient	S	2404	1970-01-01

▶ Run selected

Just now (3s)

prod_deid_clinical. omop_deid

New SQL editor: ON



Edit

```
4  
5  -- All descendant concepts of this concept_id  
6  SELECT  
7  ..ca.descendant_concept_id,  
8  ..d.concept_name AS descendant_concept_name,  
9  ..d.domain_id AS descendant_domain_id,
```

Add parameter

Table

Q T E

	A ^B C descendant_concept_name	A ^B C descendant_domain_id	A ^B C descendant_vocabulary_id	1 ² 3 min_levels_of_separation
1	chlorpropamide	Drug	RxNorm	
2	chlorpropamide 100 MG	Drug	RxNorm	
3	chlorpropamide 250 MG	Drug	RxNorm	
4	chlorpropamide 125 MG	Drug	RxNorm	
5	chlorpropamide 200 MG	Drug	RxNorm	
6	chlorpropamide 50 MG	Drug	RxNorm	
7	chlorpropamide 250 MG [Insulase]	Drug	RxNorm	
8	chlorpropamide 100 MG [Diabinese]	Drug	RxNorm	
9	chlorpropamide 250 MG [Diabinese]	Drug	RxNorm	
10				

To Atlas!

THE DRUG DISCOVERY
ENGINE FOR ALS

ALS
THERAPY DEVELOPMENT
INSTITUTE

-  Home
-  Data Sources
-  Search

Cohorts  Concept Sets Criteria Definitions NLP Definitions Cohort Upload Generated CohortsData Extraction  Data Baskets Extraction History New Extraction New Concept Set

sulfonylureas - DB



Concept
Set
Expression

Included
Concepts 

Included
Source
Codes

Recommend

Export

Import

Compare

Versions

Messages

Enter the concept set description here

Show 50  entries

Filter:

Previous Next

<input checked="" type="checkbox"/> Concept Id	 Concept Code	 Concept Name	 Domain	 Standard Concept Caption	<input checked="" type="checkbox"/> Exclude	<input checked="" type="checkbox"/> Descendants	<input checked="" type="checkbox"/> Mapped
--	--	--	--	--	---	---	--

No data available in table

 Classification  Non-Standard  Standard

 Remove Selected

 Add Concepts

[Home](#)[Data Sources](#)[Search](#)[Cohorts](#)[Concept Sets](#)[Criteria Definitions](#)[NLP Definitions](#)[Cohort Upload](#)[Generated Cohorts](#)[Data Extraction](#)[Data Baskets](#)[Extraction History](#)[New Extraction](#)

Search

1594973

[Advanced Options](#)

View record count for: [Training] Databricks Prod Deidentified

[Show columns▼](#)[Copy](#)[CSV](#)

Show 50 entries

Filter: Search...

Showing 1 to 1 of 1 entries

Previous 1 Next

Vocabulary	Id	Code	Name	Class	RC	DRC	PC	DPC	Domain	Vocabulary
RxNorm (1)	1594973	2404	chlorpropamide	Ingredient	0	0	0	0	Drug	RxNorm

Showing 1 to 1 of 1 entries

Previous 1 Next

[T Vocabulary](#)[RxNorm \(1\)](#)[T Class](#)[Ingredient \(1\)](#)[T Domain](#)[Drug \(1\)](#)[T Standard Concept](#)[Standard \(1\)](#)[T Invalid Reason](#)[Valid \(1\)](#)[T Has Records](#)[false \(1\)](#)

 Home
 Data Sources
 Search

Cohorts ^

Concept Sets

Criteria Definitions

NLP Definitions

Cohort Upload

Generated Cohorts

Data Extraction ^

Data Baskets

Extraction History

New Extraction

Standard Analytics ^

Vocabulary > Concept

chlorpropamide

Details

Related Concepts

Hierarchy

Record Counts

Drilldown Report

VIEW:

Full Hierarchy

Parents

Current Concept

Children

↑ Parents

Show columns▼

Copy

CSV

Show 25 ▾ entries

Filter: Search...

Previous 1 Next

Vocabulary	Id	Code	Name	Class	RC	DRC	Distance	Domain	Vocabulary
ATC (3)	<input checked="" type="checkbox"/>	21600749	A10BB	Sulfonylureas	ATC 4th	0	0	1	Drug ATC
Class	<input checked="" type="checkbox"/>	21600767	A10BD02	metformin and sulfonylureas; systemic	ATC 5th	0	0	1	Drug ATC
ATC 5th (2)	<input checked="" type="checkbox"/>	21600766	A10BD01	phenformin and sulfonylureas; oral	ATC 5th	0	0	1	Drug ATC
ATC 4th (1)									
Has Records									
false (3)									
Has Descendant Records									
false (3)									

Showing 1 to 3 of 3 entries

Previous 1 Next

Concept Set #145

created by boyced03 on 2025-11-13 14:46, modified by boyced03 on 2025-11-13 14:46

sulfonylureas - DB

Concept Set Expression Included Concepts (6200) Included Source Codes Recommend Export Import Compare Versions Messages

Enter the concept set description here

Show 50 entries Filter: Search...

Showing 1 to 1 of 1 entries Previous 1 Next

<input checked="" type="checkbox"/>	Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	<input checked="" type="checkbox"/> Exclude	<input checked="" type="checkbox"/> Descendants	<input checked="" type="checkbox"/> Mapped
<input checked="" type="checkbox"/>	21600749	A10BB	Sulfonylureas	Drug	Classification	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Classification Non-Standard Standard

Remove Selected Add Concepts

THE DRUG DISCOVERY
ENGINE FOR ALS

ALD
THERAPY DEVELOPMENT
INSTITUTE

B.1.8 Concept: Sulfonylureas

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1594973 ✓	chlorpropamide	2404	RxNorm	NO	YES	NO
1597756 ✓	glimepiride	25789	RxNorm	NO	YES	NO
1560171 ✓	glipizide	4821	RxNorm	NO	YES	NO
19097821	gliquidone	25793	RxNorm	NO	YES	NO
1559684 ✓	glyburide	4815	RxNorm	NO	YES	NO
1502809	tolazamide	10633	RxNorm	NO	YES	NO
1502855	tolbutamide	10635	RxNorm	NO	YES	NO

<input checked="" type="checkbox"/>	1560171	4821	glipizide	Ingredient	52,215	189,479	1	Drug	RxNorm
<input checked="" type="checkbox"/>	1597756	25789	glimepiride	Ingredient	49,467	187,304	1	Drug	RxNorm
<input checked="" type="checkbox"/>	1559684	4815	glyburide	Ingredient	9,428	24,058	1	Drug	RxNorm
<input checked="" type="checkbox"/>	1594973	2404	chlorpropamide	Ingredient	0	16	1	Drug	RxNorm

[Home](#)[Data Sources](#)[Search](#)[Cohorts](#)[Concept Sets](#)[Criteria Definitions](#)[NLP Definitions](#)[Cohort Upload](#)[Generated Cohorts](#)[Data Extraction](#)[Data Baskets](#)

Search

19097821

[Advanced Options](#)

View record count for: [Training] Databricks Prod Deidentified

[Show columns▼](#)[Copy](#)[CSV](#)

Show 50 entries

Filter: Search...

Showing 1 to 1 of 1 entries

Previous 1 Next

Vocabulary	<input checked="" type="checkbox"/>	Id	Code	Name	Class	RC	DRC	PC	DPC	Domain	Vocabulary
RxNorm (1)	<input checked="" type="checkbox"/>	19097821	25793	gliquidone	Ingredient	0	0	0	0	Drug	RxNorm

Showing 1 to 1 of 1 entries

Previous 1 Next

[Vocabulary](#)

Id

Code

Name

Class

RC

DRC

PC

DPC

Domain

Vocabulary

[RxNorm \(1\)](#)[Class](#)[Ingredient \(1\)](#)[Domain](#)[Drug \(1\)](#)[Standard Concept](#)[Standard \(1\)](#)[Invalid Reason](#)

New Definition

Copy

Edit

Showing 1 to 1 of 1 entries

Vocabulary	<input type="checkbox"/>	Id	Code	Name	Class	RC	DRC	PC	DPC	Domain
RxNorm (1)	<input checked="" type="checkbox"/>	19097821	25793	gliquidone	Ingredient	0	0	0	0	Drug

Vocabulary

RxNorm (1)

Class

Ingredient (1)

Domain

Drug (1)

Standard Concept

Standard (1)

Invalid Reason

Valid (1)

Has Records

false (1)

Has Descendant

Records

false (1)

Has Person Count

false (1)

Has Descendant

Person Count

false (1)

Showing 1 to 1 of 1 entries

Classification
 Non-Sta

**THE DRUG D
ENGINE FOR**

Select Concept Set

sulfonylureas - DB (Repository) ▾



Exclude



Descendants



Mapped

Preview...

Add To Concept Set

ALS
THERAPY DEVELOPMENT
INSTITUTE

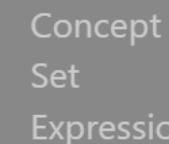
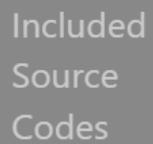
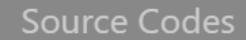
 Home Data Sources SearchCohorts  Concept Sets Criteria Definitions NLP Definitions Cohort Upload Generated CohortsData Extraction  Data Baskets Extraction History New ExtractionStandard Analytics  Concept Set #146

created by boyced03 on 2025-11-13 15:46, modified by boyced03 on 2025-11-13 15:46

sulfonylureas - DB



Optimize

 Concept Set Expression Included Concepts 0 Included Source Codes Recommend Export Import Compare Versions Messages Concept Identifiers Source Codes Concept Set RepositoryEnter Concept Identifiers ([click to clear](#))

1594973

1597756

1560171

19097821

1559684



Exclude



Descendants



Mapped

 Add To Concept Set

Concept Set Expression

Included Concepts

5761

Enter the concept set description here

Show 50 entries

Showing 1 to 7 of 7 entries

<input checked="" type="checkbox"/>	Concept Id	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
-------------------------------------	------------	--------------	------	------------	----------	-------------	--------

1502809	10633	tolazamide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1502855	10635	tolbutamide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1559684	4815	glyburide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1560171	4821	glipizide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1594973	2404	chlorpropamide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1597756	25789	glimepiride	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
19097821	25793	gliquidone	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

ious 1 Next

Mapped

<input checked="" type="checkbox"/>	Concept Id	Concept Name	Code	Vocabulary	Classification	Non-Standard	Standard
<input checked="" type="checkbox"/>	1502809	tolazamide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	1502855	tolbutamide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	1559684	glyburide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	1560171	glipizide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	1594973	chlorpropamide	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	1597756	glimepiride	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	19097821	gliquidone	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Remove Selected

Add Concepts

Classification Non-Standard Standard

LS
EVELOPMENT
TUTE



metformin

DETAILS

Domain ID	Drug
Concept Class ID	Ingredient
Vocabulary ID	RxNorm
Concept ID	1503297
Concept code	6809
Validity	Valid
Concept	Standard
Valid start	01-Jan-1970
Valid end	31-Dec-2099

TER

RE

Ha
(O)

Concept Set Expression Included Concepts 0 Included Source Codes Recommend Export Import Compare Versions Message

Cohorts ▾

Concept Sets

Criteria Definitions

NLP Definitions

Cohort Upload

Generated Cohorts

Data Extraction ▾

Data Baskets

Extraction History

New Extraction

Standard Analytics ▾

Concept Identifiers Source Codes Concept Set Repository

Enter Source Codes ([click to clear](#))

6809

Search concepts

Show columns▼ Copy CSV Show 25 entries

Showing 1 to 2 of 2 entries

	Id	Code	Name	Class
<input checked="" type="checkbox"/>	35956905	6809	MAGEB2 (MAGE family member B2) gene variant measurement	Genetic Variation
<input checked="" type="checkbox"/>	1503297	6809	metformin	Ingredient

Vocabulary

RxNorm (1)

OMOP Genomic (1)

Class

Ingredient (1)

Genetic Variation (1)

Domain

Showing 1 to 2 of 2 entries

	Id	Code	Name	Class
<input checked="" type="checkbox"/>	35956905	6809	MAGEB2 (MAGE family member B2) gene variant measurement	Genetic Variation
<input checked="" type="checkbox"/>	1503297	6809	metformin	Ingredient

BACK
15 IN
MINS



Hands-on Exercises

WARM UP: NEW CONCEPT SET

- Make a new concept set for standard condition concept for “ingrown toenail”
- Call it “ingrown_toenail_yourinitials” and save in Search
- Put in the chat when you are done

Let's do the other concept sets

The screenshot shows a web browser window with the URL https://ohdsi-studies.github.io/SemaglutideNaion/protocol.html#B1_New_user_of_semaglutide_as_2nd-line_treatment_with_prior_T2DM_and_prior_metformin. A red arrow points from the text "B.1.5 Concept: DPP4 inhibitors" to the corresponding table on the right.

B.1.5 Concept: DPP4 inhibitors

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
43013884	alogliptin	1368001	RxNorm	NO	YES	NO
40239216	linagliptin	1100699	RxNorm	NO	YES	NO
40166035	saxagliptin	857974	RxNorm	NO	YES	NO
1580747	sitagliptin	593411	RxNorm	NO	YES	NO
19122137	vildagliptin	596554	RxNorm	NO	YES	NO

B.1.6 Concept: semaglutide

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
793143	semaglutide	1991302	RxNorm	NO	YES	NO

B.1.7 Concept: SGLT2 inhibitors

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
43526465	canagliflozin	1373458	RxNorm	NO	YES	NO
44785829	dapagliflozin	1488564	RxNorm	NO	YES	NO
45774751	empagliflozin	1545653	RxNorm	NO	YES	NO
793293	ertugliflozin	1992672	RxNorm	NO	YES	NO

B.1.8 Concept: Sulfonylureas

Volunteers?

Sneak preview of next week

The screenshot shows a software interface for defining cohorts. On the left, a sidebar lists various menu items: Home, Data Sets, Search, Cohorts (with sub-options: Concept Sets, Criteria Definitions, NLP Definitions, Cohort Upload, Generated Cohorts), Data Extraction (with sub-options: Data Baskets, Extraction History, New Extraction), and Standard Analytics. The main workspace displays a cohort definition titled "B.1 New user of semaglutide as 2nd-line treatment with prior T2DM and prior metformin". The title bar includes a bell icon and a creation timestamp: "created by boycead05 on 2023-11-15 10:11". Below the title is a search bar containing the query "[semaNAION] New user of semaglutide as 2nd-line treatment with prior T2DM and prior metformin". A navigation bar at the top of the main area includes tabs for Definition, Concept Sets, Generation, Samples, Reporting, Export, IRB, Versions, and Messages (with a count of 3). The "Definition" tab is active. A large text input field below the tabs contains the placeholder "Enter a cohort definition description here". A blue header bar labeled "Cohort Entry Events" contains the text "Events having any of the following criteria:". Below this, a dropdown menu shows "a drug exposure of semaglutide" and a plus sign "+". Another plus sign is visible below the first one. At the bottom, there are fields for "with continuous observation of at least 365 days before and 0 days after event index date" and "Limit initial events to: earliest event per person". The interface is branded with NYU Langone Health and Databricks logos.

What we learned today

- How the Kahn Data Quality Framework structures data quality assessment
- How the Data Quality Dashboard (DQD) operationalizes the Kahn framework
- How to build effective concept sets in ATLAS
- Best practices for validating concept sets and data quality

Homework

Readings: Review The Book of OHDSI Chapter 10,
Defining Cohorts

Review the NAION protocol, B.1 New user of semaglutide as 2nd-line treatment with prior T2DM and prior metformin

<https://ohdsi-studies.github.io/SemaglutideNaion/protocol.html>



THE DRUG DISCOVERY
ENGINE FOR **ALS**

Thank you. Questions?

Danielle Boyce

dboyce@als.net

