OHDSI ATLAS 简介

翁春华
哥伦比亚大学医疗信息学系副教授
美国医学信息学院当选成员

Chunhua Weng, PhD, FACMI
Associate Professor of Biomedical Informatics
Columbia University, New York City

August 21, 2017, Hangzhou, China, OHDSI-China Meeting

讲座结束后，希望您能够

1. 熟悉ATLAS 的适用和功能

2. 使用ATLAS 选择有特定临床特征的病人组并对他们进行分析

Why OHDSI?

- 观察性研究(Observational Studies)是任何随机临床测试试验(RCT)的必要先导(Precursor)；前者发现问题，提出假设，积累证据；后者用最严格的实验设计验证假设。

- 利用OHDSI 的大数据和工具，研究人员可以

  - 刻画临床特征 (what happened to the patients?)
    - 糖尿病人吃什么药？那些有并发症？他们对治疗的响应如何？
  - 基于人群的估计 (what are the causal effect?)
    - 哪种糖尿病治疗方案最佳？
  - 基于病人个体的预测 (what will happen to patient X?)
    - 哪个病人更适合哪种治疗方案？

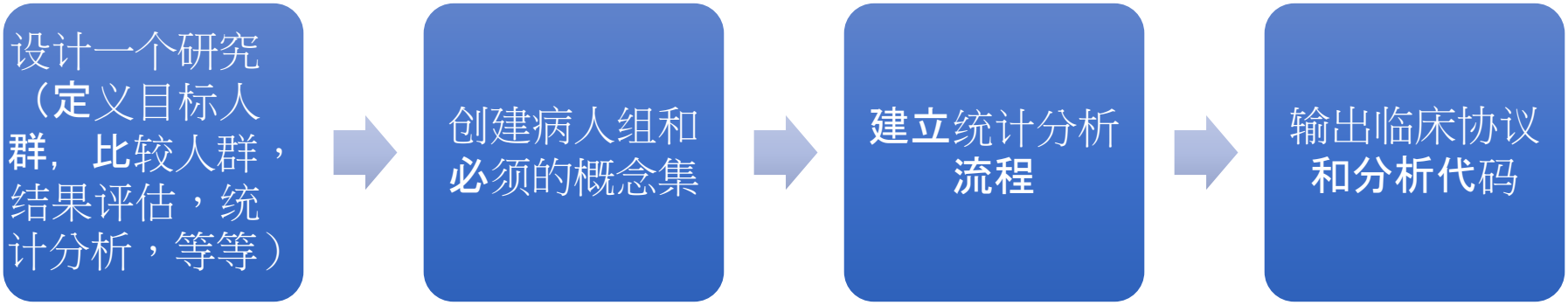
用观察型研究指导临床测试设计

- 设计应用符合研究的病人条件（inclusion and exclusion criteria）
- 进行观察性研究来指导RCT 设计和目标制定
- 展示药物在临床的使用及其安全性

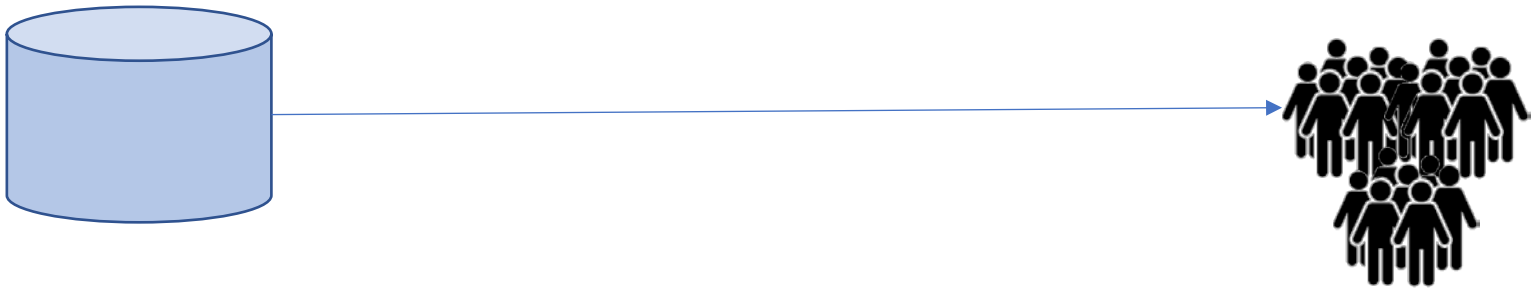
ATLAS 是什么？

- 一个免费网络开源软件工具
- 定义查询病人的条件 （比如，近五年内有二型糖尿病但没有高血压）
- 运用查询条件找到符合条件的病人
- 分析病人的特征
- 对病人特征的描述可分享，重用，可自动在不同系统间转换 JSON, SQL, etc.

Atlas 工作流


设计一个研究（定义目标人群，比较人群，结果评估，统计分析，等等）
创建病人组和必须的概念集
建立统计分析流程
输出临床协议和分析代码

ATLAS 病人定义的抽象模型：Everything is a concept



对应变量和规则：

Insulin, Metformin and its brand names, etc.

有糖尿病的人群 →

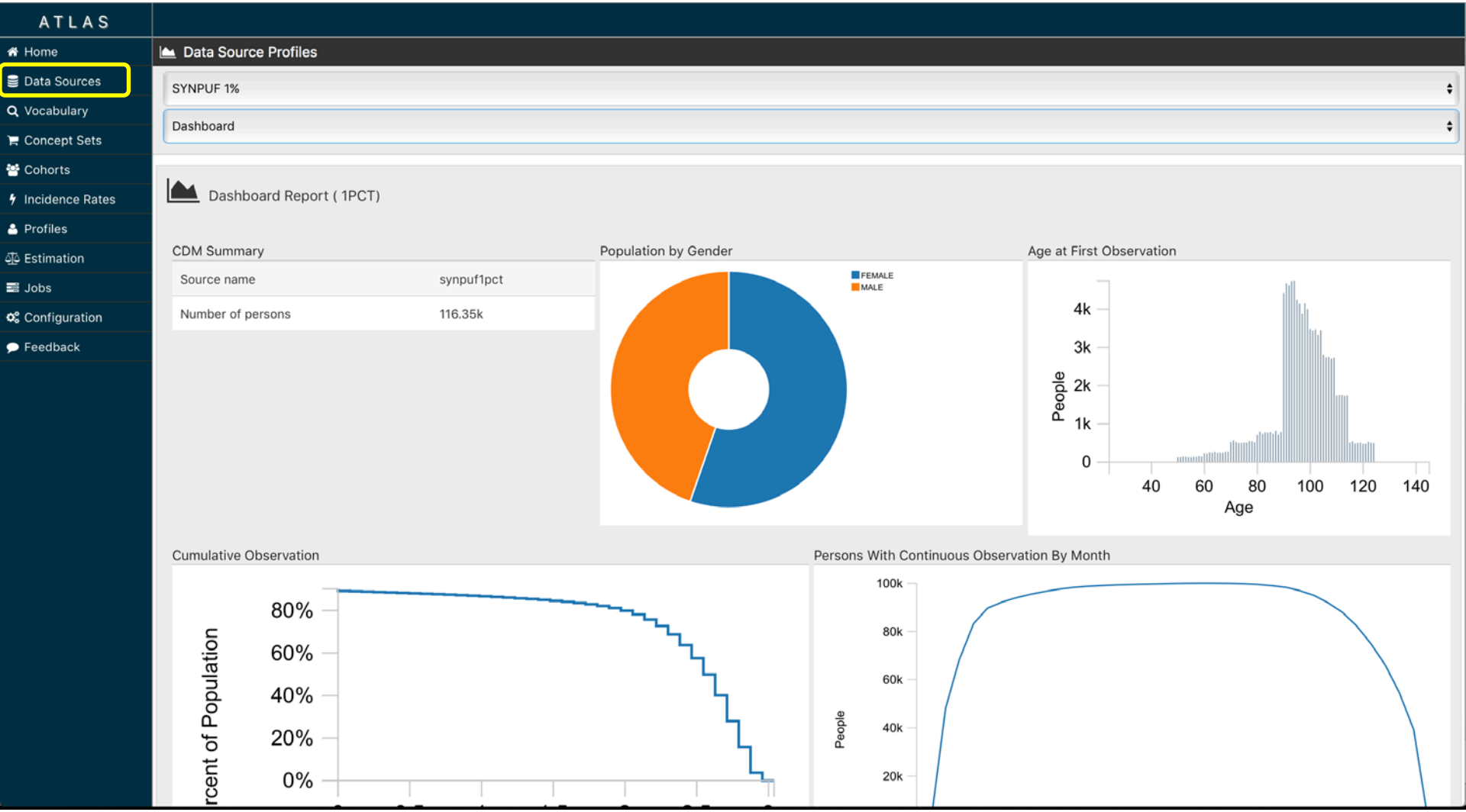
- ✓ **有不正常血糖**的病人
- ✓ **有在服用糖尿病相关**药物的病人
- ✓ 有病历里含有**二型糖尿病**诊断码 (ICD 9, ICD 10, READ, SNOMED) 的病人
- ✓ 有病历里含有**相关文字描述(二型糖尿病， DMII， Type 2 Diabetes)**的病人

等等

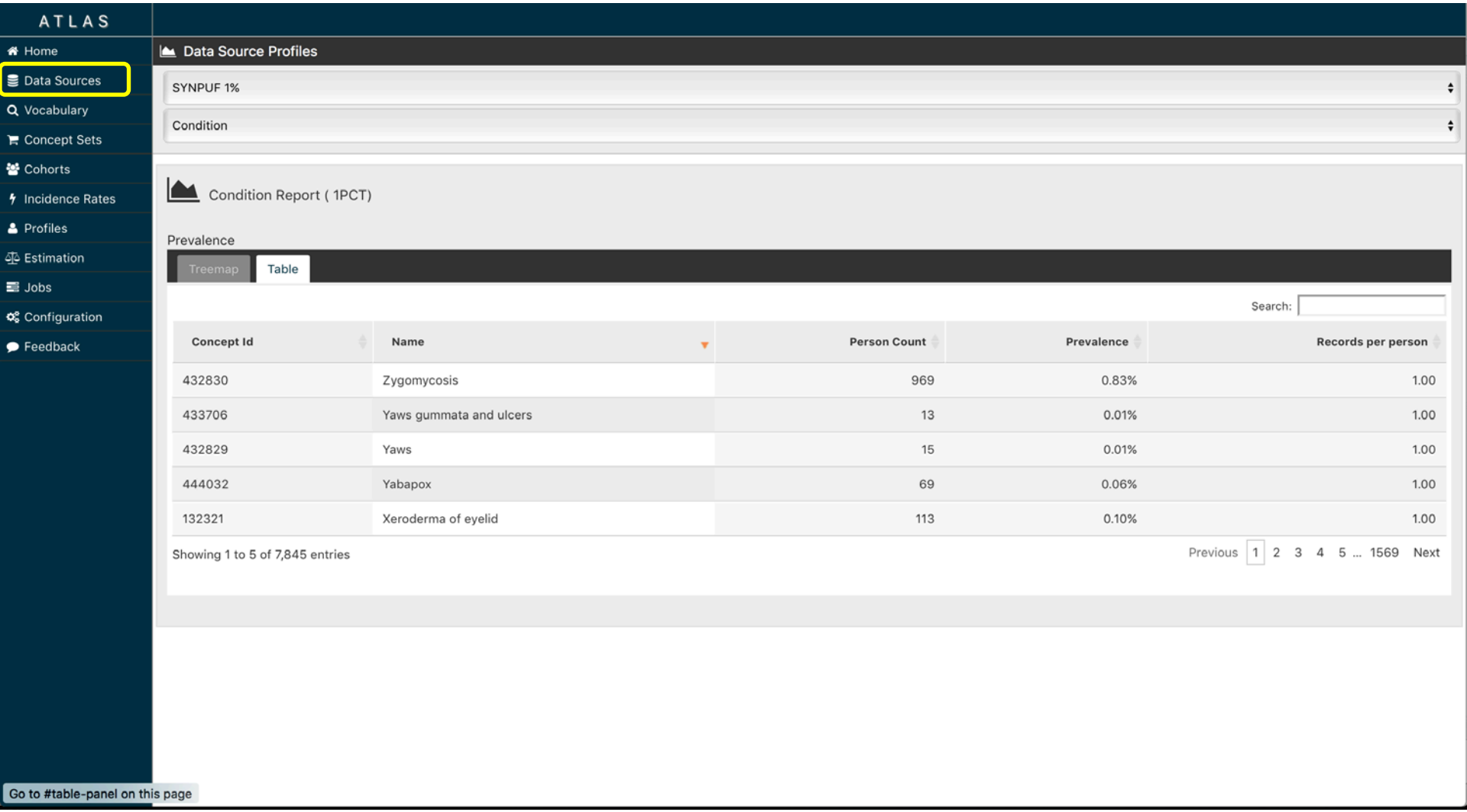
ATLAS 能干什么？

1. 浏览数据源
2. 检索术语 Vocabulary
3. 定义术语集 Concept Set
4. 定义病人组和他们的临床特征
5. 查询数据库找到符合条件的病人组
6. 可视化单独病人的情况
7. 做人群的效果估计

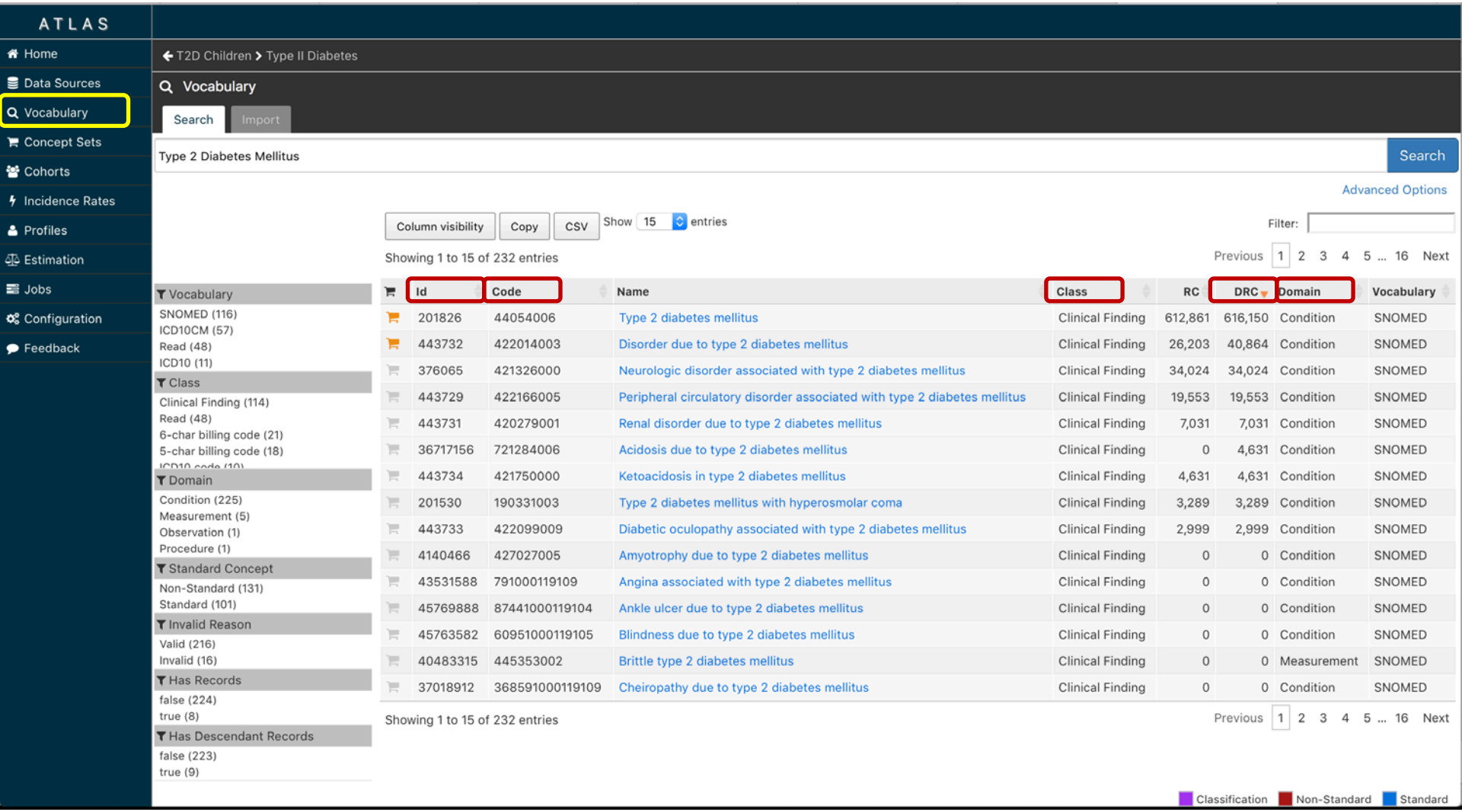
1. 浏览数据源


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Incidence Rates
Profiles
Estimation
Jobs
Configuration
Feedback
Data Source Profiles
SYNPUF 1%
Dashboard
Dashboard Report (1PCT)
CDM Summary
Source name
synpuf1pct
Number of persons
116.35k
Population by Gender
FEMALE
MALE
Age at First Observation
People
4k
3k
2k
1k
0
40
60
80
100
120
140
Age
Cumulative Observation
80%
60%
40%
20%
0%
Persons With Continuous Observation By Month
People
100k
80k
60k
40k
20k

1. 浏览数据源


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Incidence Rates
Profiles
Estimation
Jobs
Configuration
Feedback
Data Source Profiles
SYNPUF 1%
Condition
Condition Report (1PCT)
Prevalence
Treemap
Table
Search:
Concept Id
Name
Person Count
Prevalence
Records per person
432830 Zygomycosis 969 0.83% 1.00
433706 Yaws gummata and ulcers 13 0.01% 1.00
432829 Yaws 15 0.01% 1.00
444032 Yabapox 69 0.06% 1.00
132321 Xeroderma of eyelid 113 0.10% 1.00
Showing 1 to 5 of 7,845 entries
Previous 1 2 3 4 5 ... 1569 Next
Go to #table-panel on this page

2. 检索术语 Vocabulary



2. 检索术语 Vocabulary

- **ID**: a unique concept ID in OHDSI OMOP CDM
- **Concept Code**: concept identifier in the source vocabulary
- **Class**: categories defined by the source vocabulary
- **RC**: The record count. This will show the number of records that are coded with this concept in the
- **DRC**: The descendant record count. The DRC column will show the sum of all descendant concepts that are coded in the CDM.
- **Domain**: categories defined by the OMOP CDM (e.g., Condition, Person, Observation, Specimen, etc)

3. 定义术语集 Concept Set

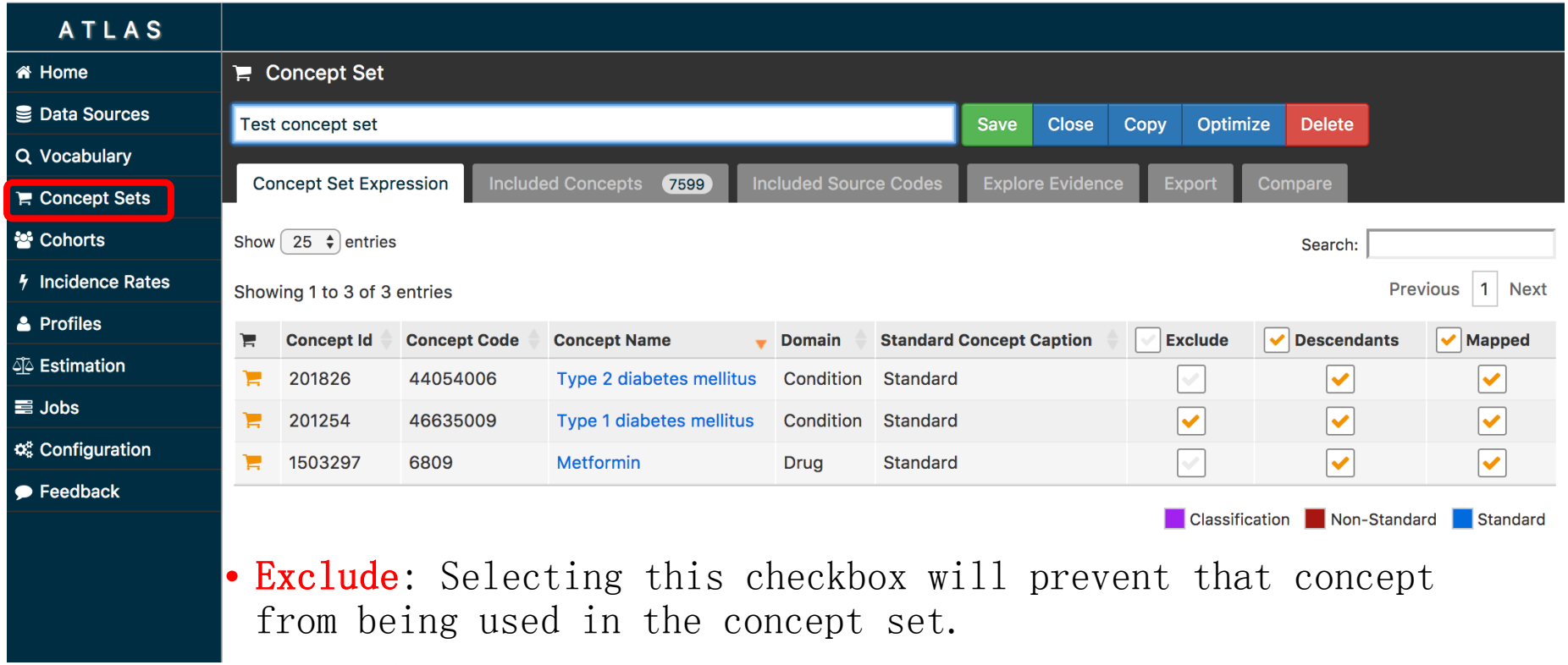
- 什么是 concept set?

They are lists of concepts from the standardized vocabulary that taken together describe a topic of interest for a study.

- 为什么要 concept set?

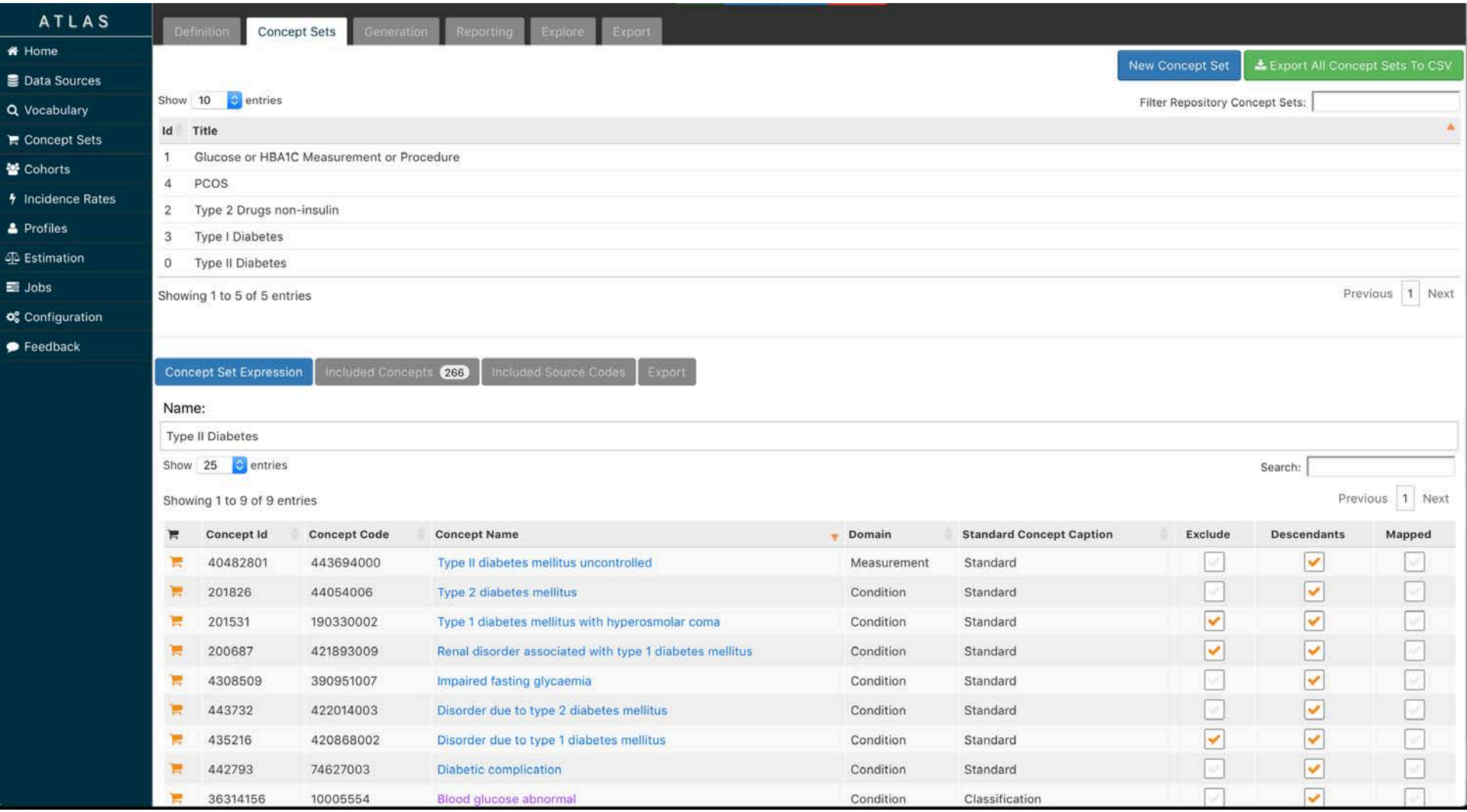
便于随意选择组合概念集以适用于不同场景

3. 定义术语集 Concept Set

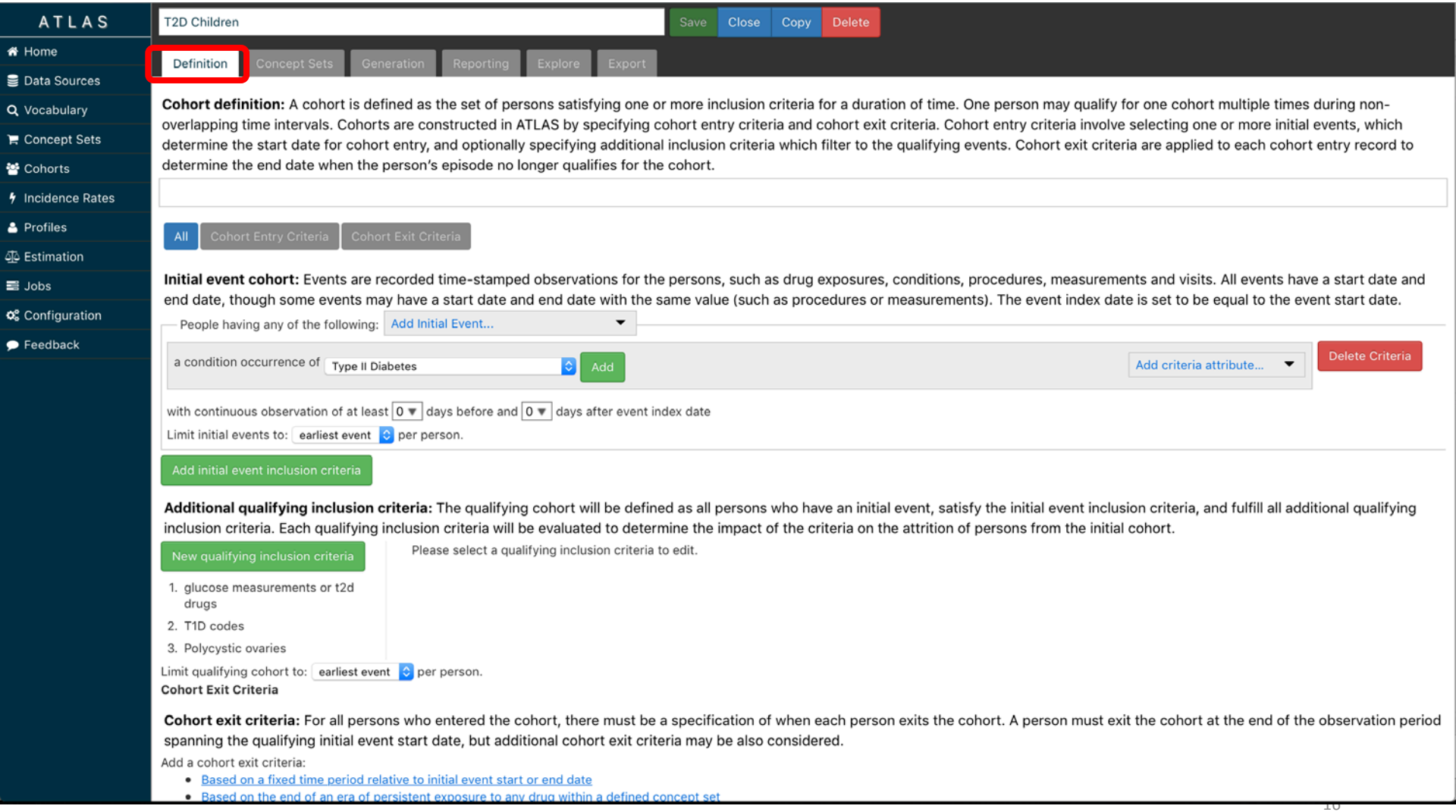


- **Exclude**: Selecting this checkbox will prevent that concept from being used in the concept set.
- **Descendants**: Selecting this check box will use the vocabulary relationships to automatically select all descendants. If this option is used in conjunction with the exclude option, it will exclude the current concept and all descendants.
- **Mapped**: Selecting this check box will use the vocabulary relationships to automatically select all concepts mapped to the selected concept.

3. 定义术语集 Concept Set

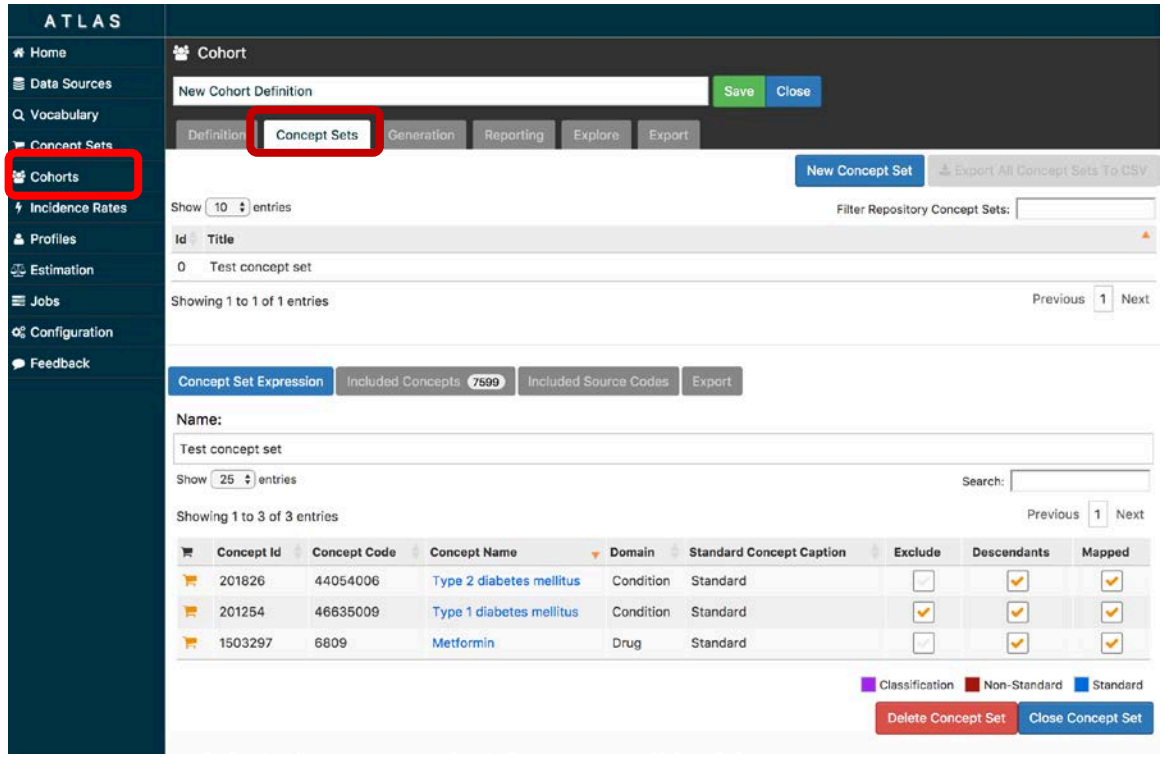


4. 定义病人组和他们的临床特征



Cohorts

- Concept sets which are used in the cohort definition will be listed under the “Concept Sets” tab



定义病人组

- 主要事件 Primary Event (Start Date)
  - ❖Cohort definitions can have lots of rules
  - ❖But the primary event is the bouncer Have to clear this bar for the rest of the rules to come into play
  - ❖Besides being the first rule, the primary event is critical because it sets the *index date*
- 符合的条件 Qualifying Criteria
  - ❖All the other criteria you wish you require of your cohort members
- 退出条件 Exit Criteria (End Date)
  - ❖Defines the end date of the individual in the cohort

Index Date

- The patient's index date (aka cohort start date) is determined by when they satisfy the primary event
- The cohort start date can be limited to just first time a patient meets it or you can count every time they meet it
- Subsequent criteria are very commonly tied relative to the index date

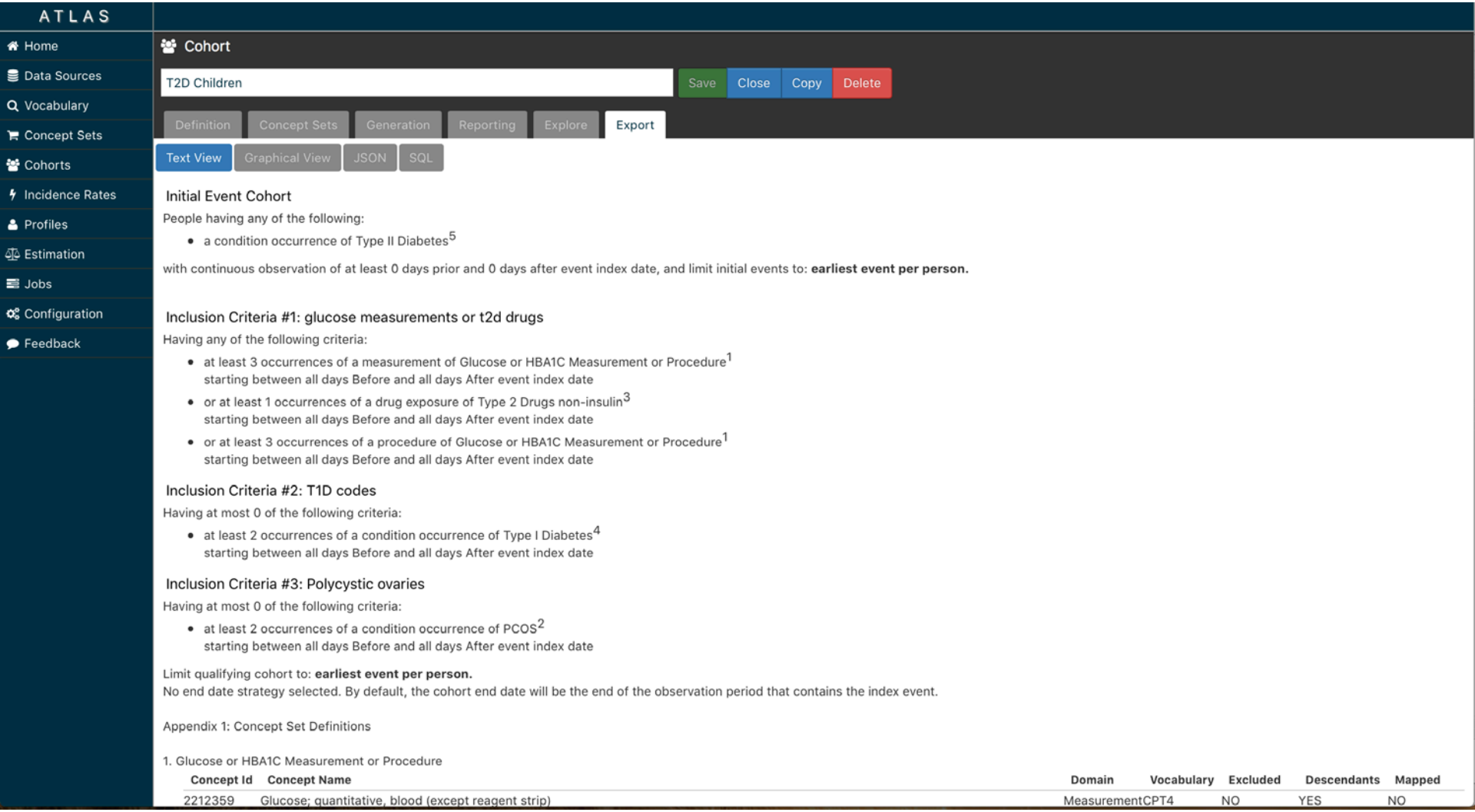
Qualifying Criteria

- All the other criteria you wish you require of your cohort members Noting that it is still the primary event that will mark their point of entry in the cohort
- Can have AND or OR logic
- Can apply the same filters as primary event
- Temporal limitations relative to index

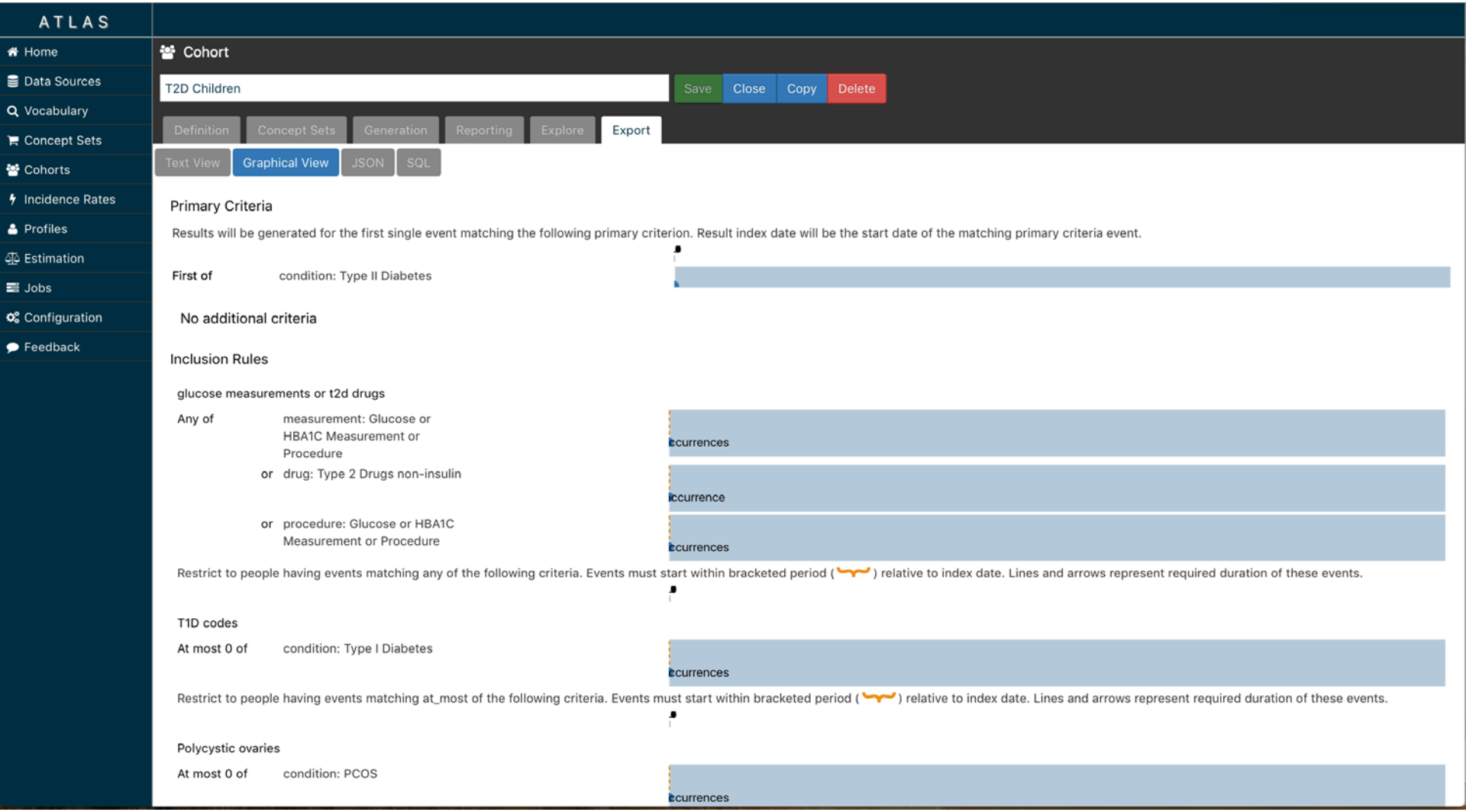
Cohorts 其他工具

- **Reporting** - The reporting tab provides cohort summarization and visualization tools
- **Explore** - 可视化病人信息
- **Export** - 查看源码，XML, JSON, SQL, etc.

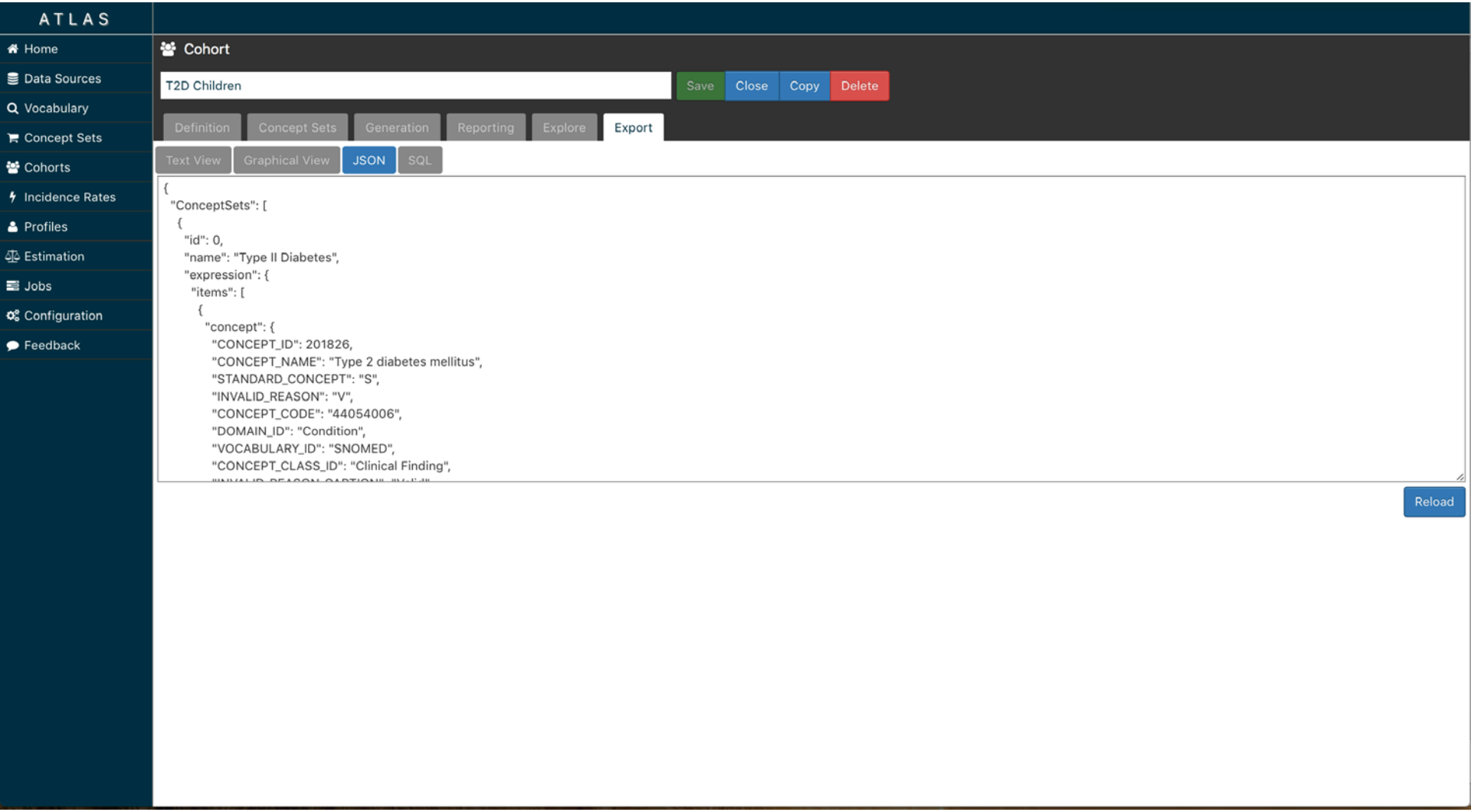
可分享的人群组定义


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Incidence Rates
Profiles
Estimation
Jobs
Configuration
Feedback
Cohort
T2D Children
Save
Close
Copy
Delete
Definition
Concept Sets
Generation
Reporting
Explore
Export
Text View
Graphical View
JSON
SQL
Initial Event Cohort
People having any of the following:
a condition occurrence of Type II Diabetes[5]
with continuous observation of at least 0 days prior and 0 days after event index date, and limit initial events to: earliest event per person.
Inclusion Criteria #1: glucose measurements or t2d drugs
Having any of the following criteria:
at least 3 occurrences of a measurement of Glucose or HBA1C Measurement or Procedure[1]
starting between all days Before and all days After event index date
or at least 1 occurrences of a drug exposure of Type 2 Drugs non-insulin[3]
starting between all days Before and all days After event index date
or at least 3 occurrences of a procedure of Glucose or HBA1C Measurement or Procedure[1]
starting between all days Before and all days After event index date
Inclusion Criteria #2: T1D codes
Having at most 0 of the following criteria:
at least 2 occurrences of a condition occurrence of Type I Diabetes[4]
starting between all days Before and all days After event index date
Inclusion Criteria #3: Polycystic ovaries
Having at most 0 of the following criteria:
at least 2 occurrences of a condition occurrence of PCOS[2]
starting between all days Before and all days After event index date
Limit qualifying cohort to: earliest event per person.
No end date strategy selected. By default, the cohort end date will be the end of the observation period that contains the index event.
Appendix 1: Concept Set Definitions
1. Glucose or HBA1C Measurement or Procedure
Concept Id
Concept Name
Domain
Vocabulary
Excluded
Descendants
Mapped
2212359
Glucose; quantitative, blood (except reagent strip)
Measurement
CPT4
NO
YES
NO

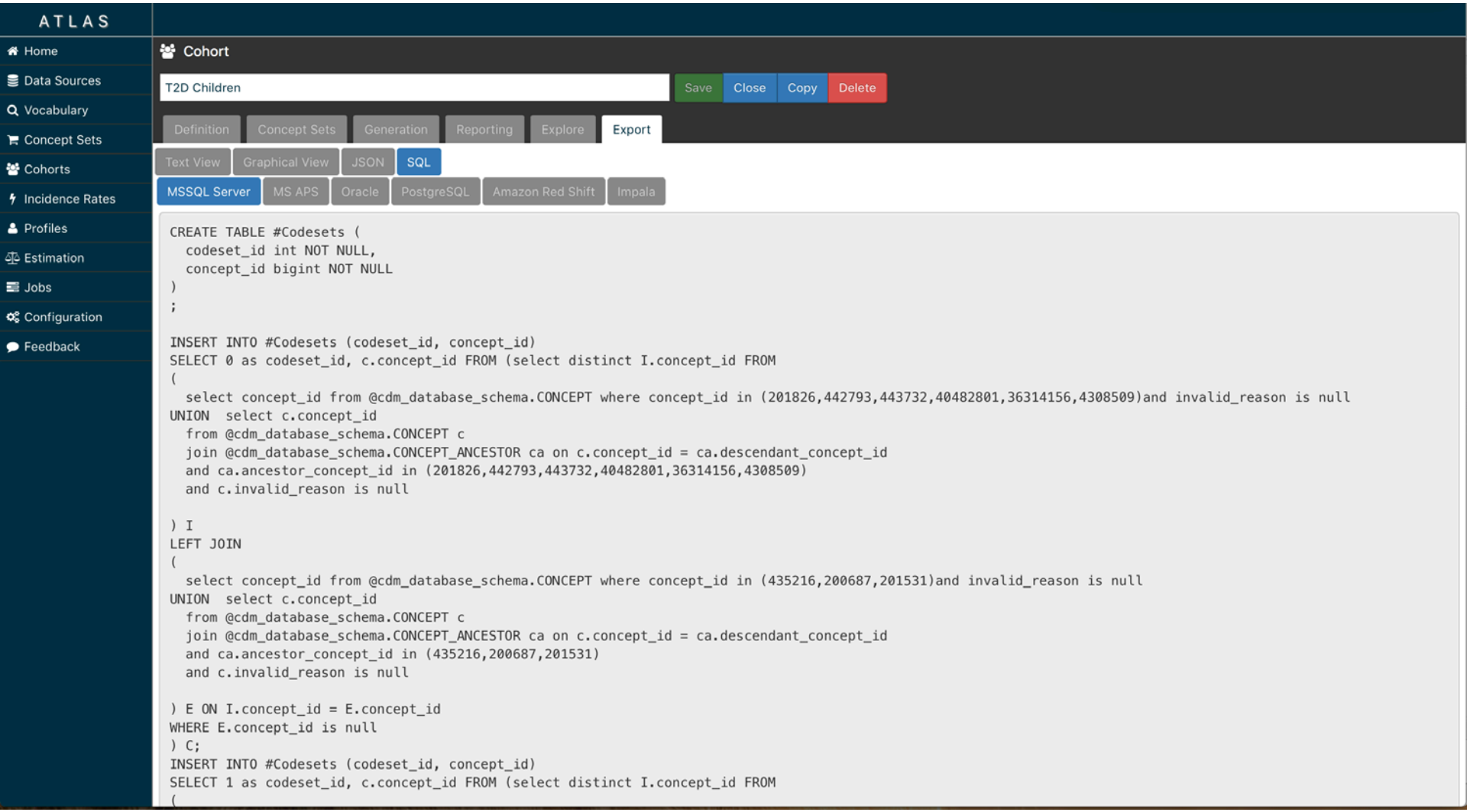
可分享的人群组定义


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Incidence Rates
Profiles
Estimation
Jobs
Configuration
Feedback
Cohort
T2D Children
Save
Close
Copy
Delete
Definition
Concept Sets
Generation
Reporting
Explore
Export
Text View
Graphical View
JSON
SQL
Primary Criteria
Results will be generated for the first single event matching the following primary criterion. Result index date will be the start date of the matching primary criteria event.
First of
condition: Type II Diabetes
No additional criteria
Inclusion Rules
glucose measurements or t2d drugs
Any of
measurement: Glucose or HBA1C Measurement or Procedure
or
drug: Type 2 Drugs non-insulin
or
procedure: Glucose or HBA1C Measurement or Procedure
Restrict to people having events matching any of the following criteria. Events must start within bracketed period ( ) relative to index date. Lines and arrows represent required duration of these events.
T1D codes
At most 0 of
condition: Type I Diabetes
Restrict to people having events matching at_most of the following criteria. Events must start within bracketed period ( ) relative to index date. Lines and arrows represent required duration of these events.
Polycystic ovaries
At most 0 of
condition: PCOS

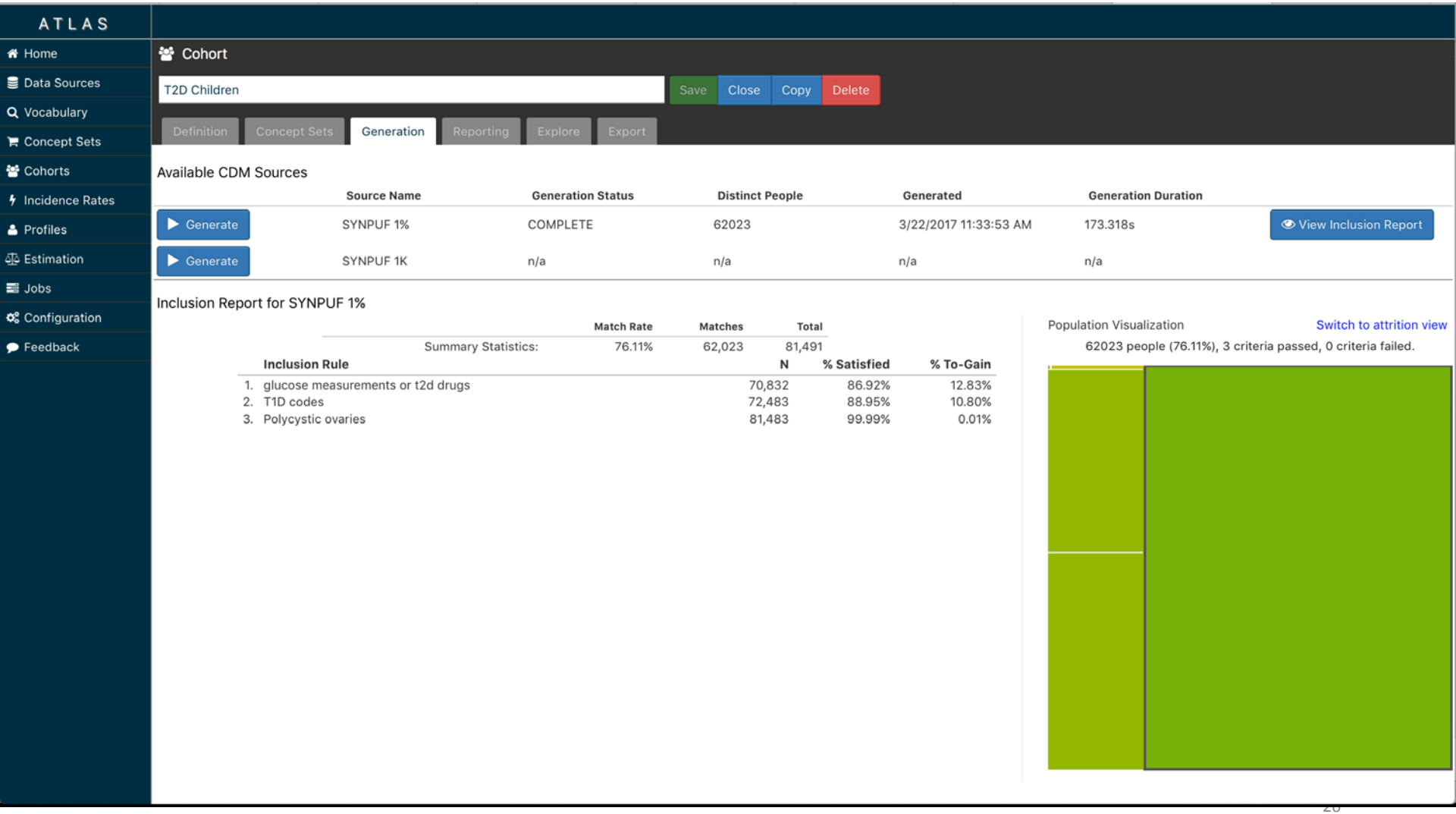
可分享的人群组定义


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Incidence Rates
Profiles
Estimation
Jobs
Configuration
Feedback
Cohort
T2D Children
Save
Close
Copy
Delete
Definition
Concept Sets
Generation
Reporting
Explore
Export
Text View
Graphical View
JSON
SQL
{
"ConceptSets": [
{
"id": 0,
"name": "Type II Diabetes",
"expression": {
"items": [
{
"concept": {
"CONCEPT_ID": 201826,
"CONCEPT_NAME": "Type 2 diabetes mellitus",
"STANDARD_CONCEPT": "S",
"INVALID_REASON": "V",
"CONCEPT_CODE": "44054006",
"DOMAIN_ID": "Condition",
"VOCABULARY_ID": "SNOMED",
"CONCEPT_CLASS_ID": "Clinical Finding",
Reload

可分享的人群组定义


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Incidence Rates
Profiles
Estimation
Jobs
Configuration
Feedback
Cohort
T2D Children
Save
Close
Copy
Delete
Definition
Concept Sets
Generation
Reporting
Explore
Export
Text View
Graphical View
JSON
SQL
MSSQL Server
MS APS
Oracle
PostgreSQL
Amazon Red Shift
Impala
CREATE TABLE #Codesets (
codeset_id int NOT NULL,
concept_id bigint NOT NULL
)
;
INSERT INTO #Codesets (codeset_id, concept_id)
SELECT 0 as codeset_id, c.concept_id FROM (select distinct I.concept_id FROM
(
select concept_id from @cdm_database_schema.CONCEPT where concept_id in (201826,442793,443732,40482801,36314156,4308509)and invalid_reason is null
UNION select c.concept_id
from @cdm_database_schema.CONCEPT c
join @cdm_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
and ca.ancestor_concept_id in (201826,442793,443732,40482801,36314156,4308509)
and c.invalid_reason is null
) I
LEFT JOIN
(
select concept_id from @cdm_database_schema.CONCEPT where concept_id in (435216,200687,201531)and invalid_reason is null
UNION select c.concept_id
from @cdm_database_schema.CONCEPT c
join @cdm_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
and ca.ancestor_concept_id in (435216,200687,201531)
and c.invalid_reason is null
) E ON I.concept_id = E.concept_id
WHERE E.concept_id is null
) C;
INSERT INTO #Codesets (codeset_id, concept_id)
SELECT 1 as codeset_id, c.concept_id FROM (select distinct I.concept_id FROM
(

5. 查询数据库找到符合条件的病人组

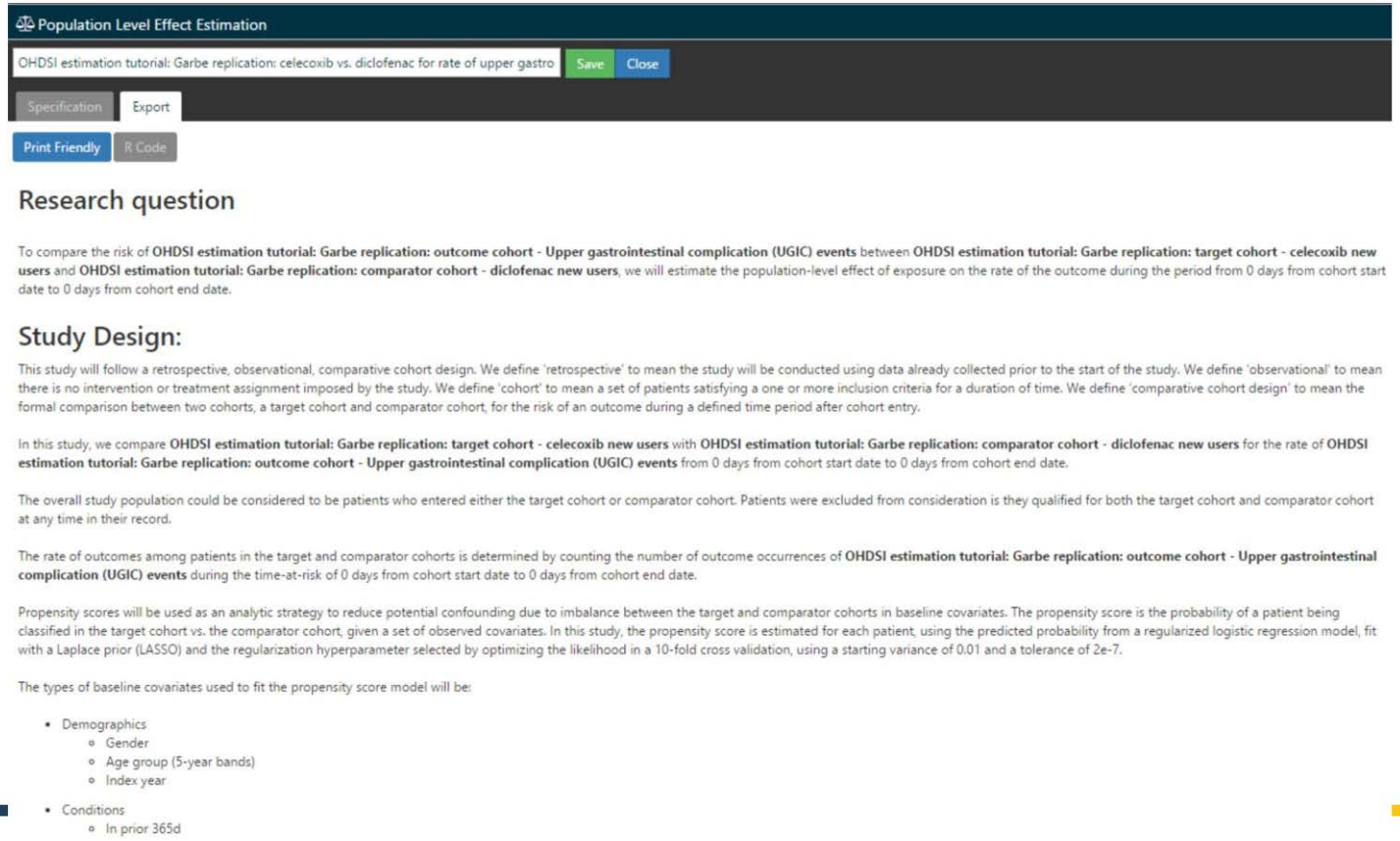


Profiles, Jobs and Configuration

- Profiles - Patient level information visualization
- Jobs - Jobs those are running in the background
- Configuration - Select the “configuration” menu item to review the data sources that have been configured in the source configuration section. This screen will let you review options. At this time, it cannot be used to edit the configuration - that must be done directly in the database.

Export Your Study - Protocol

ATLAS print friendly – the start of your team's protocol



Population Level Effect Estimation

OHDSI estimation tutorial: Garbe replication: celecoxib vs. diclofenac for rate of upper gastro | Save | Close

Specification | Export

Print Friendly | R Code

Research question

To compare the risk of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** between **OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users** and **OHDSI estimation tutorial: Garbe replication: comparator cohort - diclofenac new users**, we will estimate the population-level effect of exposure on the rate of the outcome during the period from 0 days from cohort start date to 0 days from cohort end date.

Study Design:

This study will follow a retrospective, observational, comparative cohort design. We define 'retrospective' to mean the study will be conducted using data already collected prior to the start of the study. We define 'observational' to mean there is no intervention or treatment assignment imposed by the study. We define 'cohort' to mean a set of patients satisfying a one or more inclusion criteria for a duration of time. We define 'comparative cohort design' to mean the formal comparison between two cohorts, a target cohort and comparator cohort, for the risk of an outcome during a defined time period after cohort entry.

In this study, we compare **OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users** with **OHDSI estimation tutorial: Garbe replication: comparator cohort - diclofenac new users** for the rate of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** from 0 days from cohort start date to 0 days from cohort end date.

The overall study population could be considered to be patients who entered either the target cohort or comparator cohort. Patients were excluded from consideration is they qualified for both the target cohort and comparator cohort at any time in their record.

The rate of outcomes among patients in the target and comparator cohorts is determined by counting the number of outcome occurrences of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** during the time-at-risk of 0 days from cohort start date to 0 days from cohort end date.

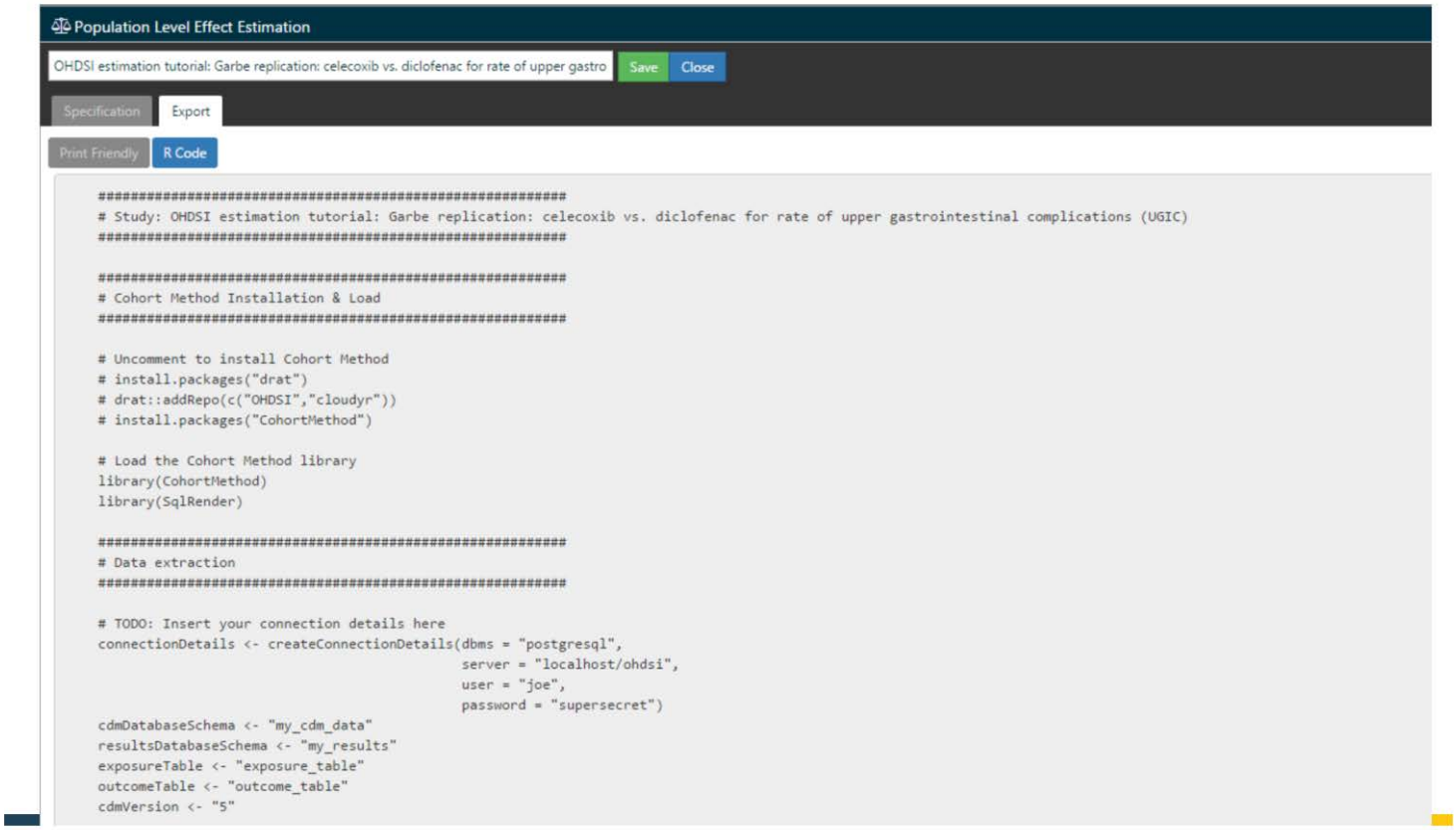
Propensity scores will be used as an analytic strategy to reduce potential confounding due to imbalance between the target and comparator cohorts in baseline covariates. The propensity score is the probability of a patient being classified in the target cohort vs. the comparator cohort, given a set of observed covariates. In this study, the propensity score is estimated for each patient, using the predicted probability from a regularized logistic regression model, fit with a Laplace prior (LASSO) and the regularization hyperparameter selected by optimizing the likelihood in a 10-fold cross validation, using a starting variance of 0.01 and a tolerance of 2e-7.

The types of baseline covariates used to fit the propensity score model will be:

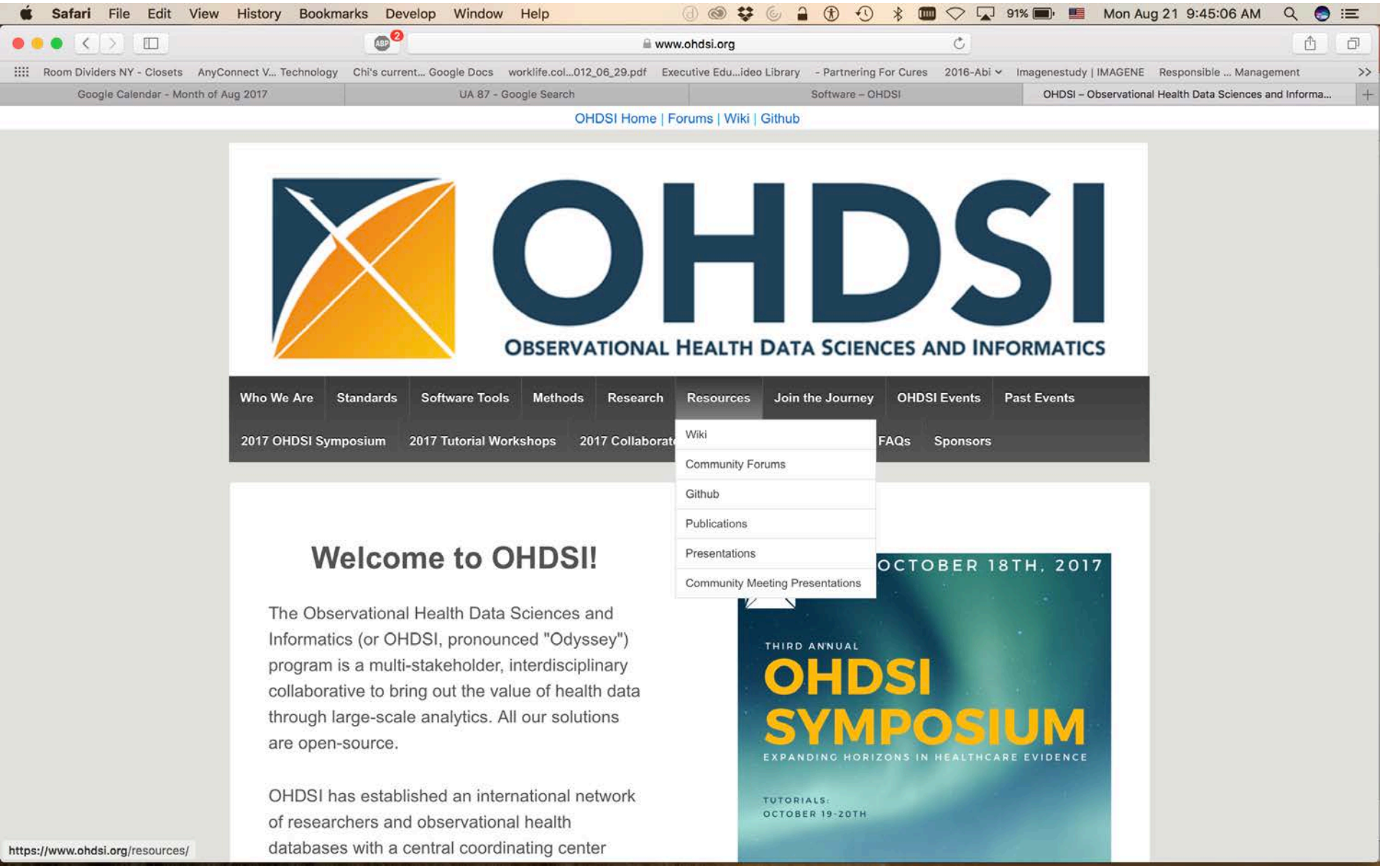
- Demographics
  - Gender
  - Age group (5-year bands)
  - Index year
- Conditions
  - In prior 365d

Export Your Study - R codes

ATLAS R code – the start of your team's implementation



谢谢您的兴趣和**加入！**


Safari File Edit View History Bookmarks Develop Window Help
Mon Aug 21 9:45:06 AM
www.ohdsi.org
Google Calendar - Month of Aug 2017
UA 87 - Google Search
Software – OHDSI
OHDSI – Observational Health Data Sciences and Informa...
OHDSI Home | Forums | Wiki | Github
OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS
Who We Are
Standards
Software Tools
Methods
Research
Resources
Join the Journey
OHDSI Events
Past Events
2017 OHDSI Symposium
2017 Tutorial Workshops
FAQs
Sponsors
Wiki
Community Forums
Github
Publications
Presentations
Community Meeting Presentations
Welcome to OHDSI!
The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.
OHDSI has established an international network of researchers and observational health databases with a central coordinating center
OCTOBER 18TH, 2017
THIRD ANNUAL
OHDSI SYMPOSIUM
EXPANDING HORIZONS IN HEALTHCARE EVIDENCE
TUTORIALS:
OCTOBER 19-20TH
https://www.ohdsi.org/resources/