JSS RESPONSIVE RELIABLE RESULTS

RESPONSIVE.
RELIABLE.
RESULTS.

ANALYZING REAL-WORLD EVIDENCE FROM LARGE-SCALE POPULATION ELECTRONIC HEALTH RECORDS (EHR)

9400 Henri-Bourassa Blvd. West
St-Laurent (Montreal) Quebec, Canada H4S 1N8
T 514 934-6116 **F** 514 934-9913
jssresearch.com

OUTLINE

- BACKGROUND
- OBJECTIVE
- INTRODUCTION
- OHDSI
- OMOP CDM –DEFINITION & PURPOSE
- SYSTEM & RESOURCES
- ETL PROCESS
- VISUALIZATION OF DATA
- CHALLENGES
- ROAD AHEAD

BACKGROUND

To date, there has not been any systematic attempts to characterize the profile of patients being treated with specific drug. Through this project, data was be collected on all eligible patients diagnosed with rheumatoid arthritis and who will be treated with a specific drug.

Data was collected from Hospital records where physicians treating patients diagnosed with rheumatoid arthritis and who are interested in participating. It was done on a voluntary basis.

It was planned to enroll about 8000 patients in this program. It is expected that about 500 sites will participate and each site should be able to recruit between 5 and 20 patients.

OBJECTIVES

- To characterize the profile of patients being treated
- To follow up on safety parameters in patients treated with Study Drug
- To provide expanded data to health care providers and hospitals to estimate and support appropriate use of drug;

Our primary motive was to standardize the data therefore it can be loaded in CDM and results can be visualized.

Analyzing such a vast data is very cumbersome and time consuming by using legacy programming structure.

INTRODUCTION

Within the last decade, several CDMs have been collaboratively developed and have risen to the level of *de facto* standards for clinical research data. These include

- The Sentinel CDM
- The National Patient-Centered Clinical Research Network (PCORnet)
- The Health Care Systems Research Network (formerly known as the HMO Research Network) Virtual Data Warehouse
- **The Observational Medical Outcomes Partnership (OMOP) CDM**
- The Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM).

OHDSI

OHDSI: Observational Health Data Sciences and Informatics is a research collaborative coordinated through Columbia University.

Multiple geographies: US, Europe, Asia-Pacific, 20 countries. OHDSI collaborators access a network of 600 million patients

-developing analytical methods and tools based on the data standardized to OMOP Common Data Model (CDM) and vocabulary

OMOP CDM –DEFINITION & PURPOSE

The OMOP Common Data Model is an open-source, community standard for observational healthcare data.

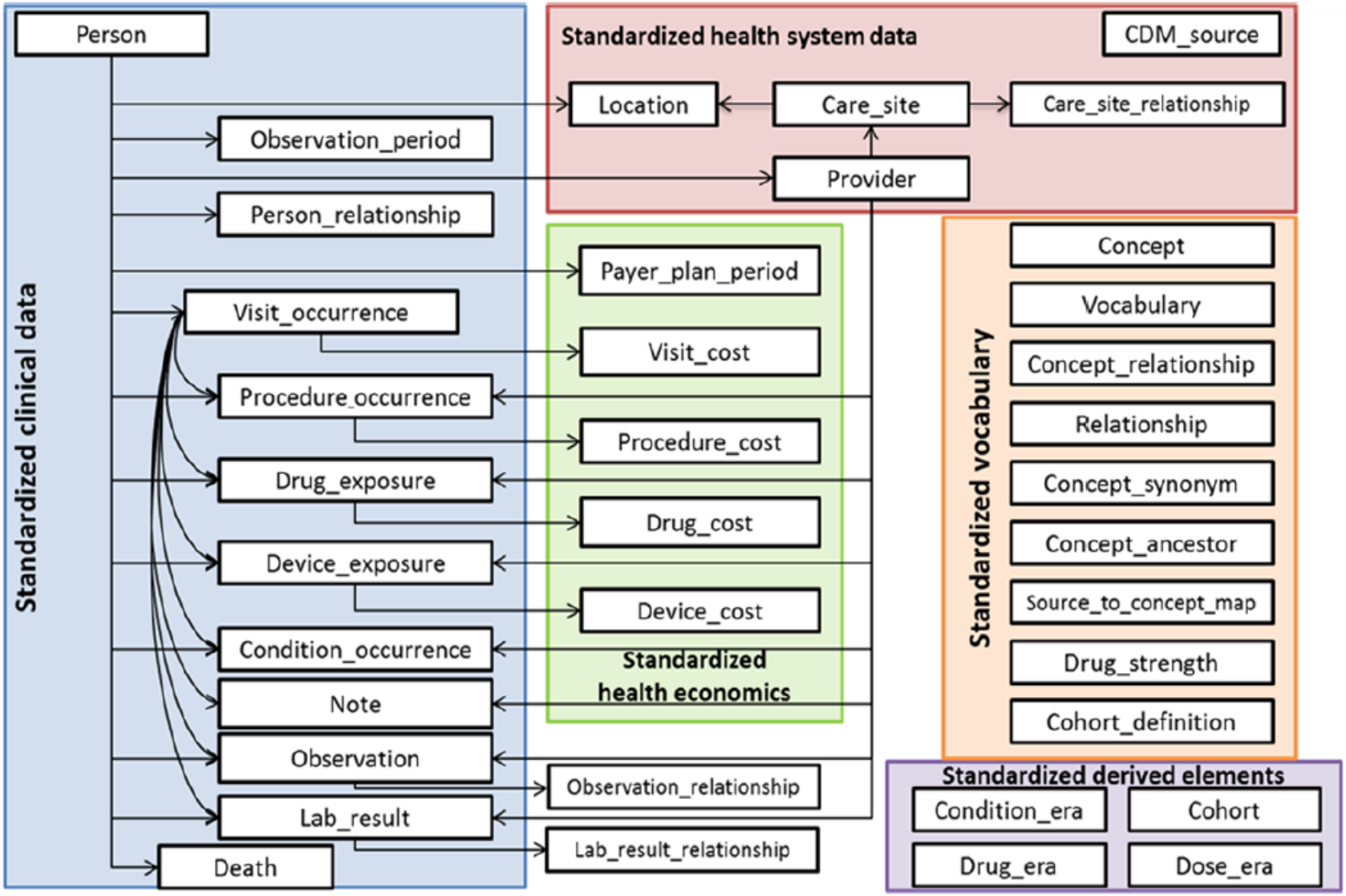
The CDM is designed to support the conduct of research to identify and evaluate associations between interventions (drug exposure and procedures etc.) and outcomes caused by these interventions (condition occurrences, procedures, drug exposure etc.)

- The common data model preserves all source data, including the original source vocabulary codes, but adds the standardized vocabularies.
- Enable standardization of structure and content to support a systematic and reproducible process to efficiently generate evidence
- Support collaborative research

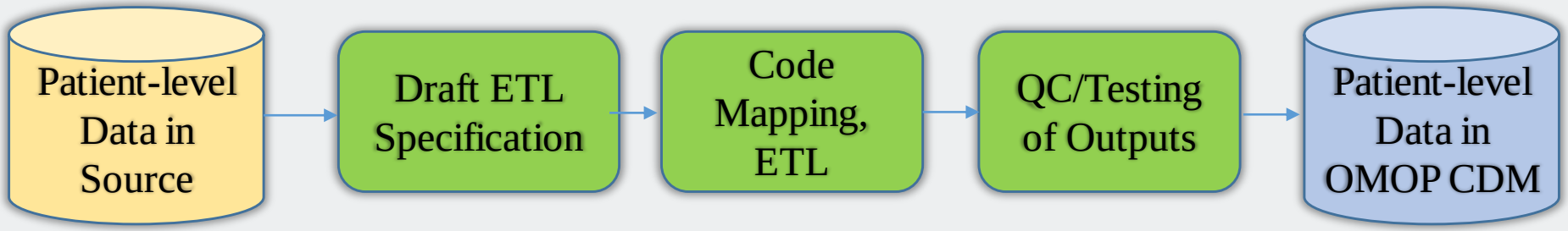
SYSTEM & RESOURCES

- Analyses were conducted on a Microsoft Server 2008 with an Intel i7 processor (Intel Corporation, Santa Clara, California, U.S.), 2.10 GHz, 8-core CPU, and 32 GB of RAM.
- Protocol & CRF
- ETL mapping specs
- OMOP Vocabularies
- SAS Version 9.4
- OMOP Guideline V5.0
- Ohdsi Tools (USAGI, WhiteRabbit and RabbitinAHat)
- PostgreSQL 9.6 Database (Open Source)
- R version 3.4.1
- Achilles for Visualization of Results

OMOP CDM V5.0



CONVERSION PROCESS OVERVIEW



Tools user for Conversion

WhiteRabbit: Profile your source data

RabbitInAHat: Map your source structure to CDM tables and fields

USAGI/ETL : Develop code to map source data Tables with SAS and SQL

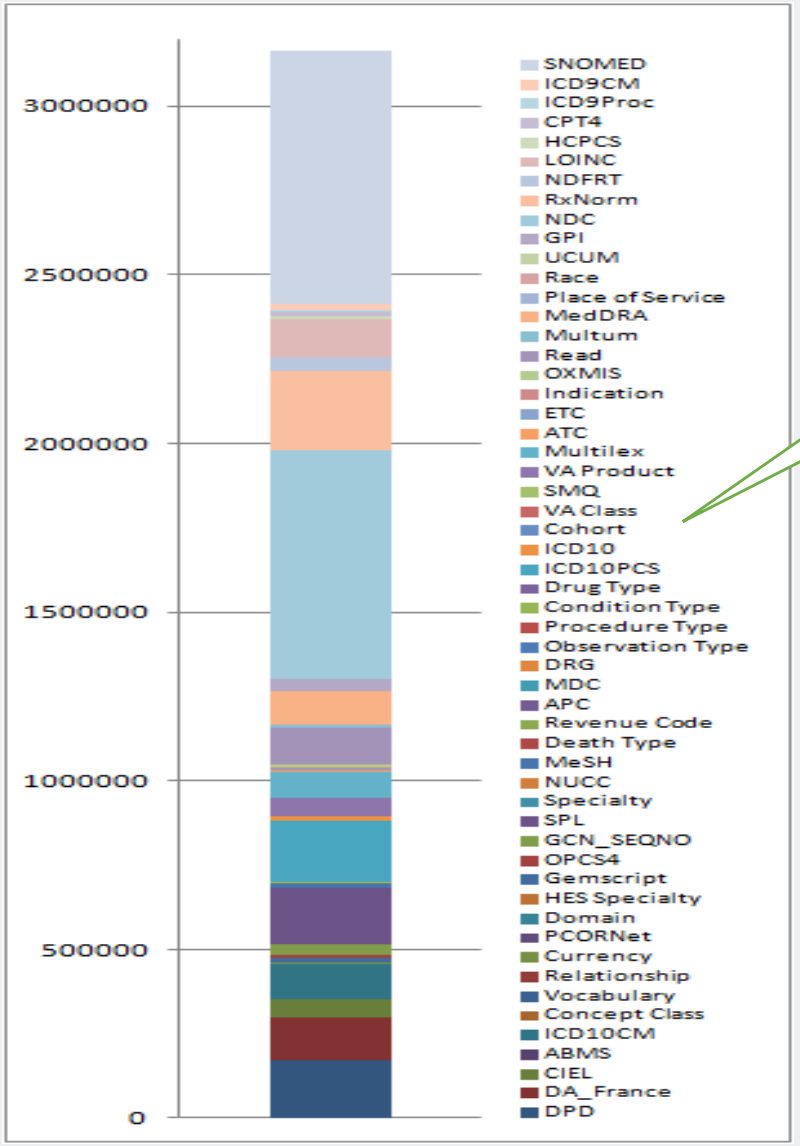
Write Syntax/ Achilles: Review data quality with compliance to OMOP CDM guidelines

Final Outputs: Final Tables ready to be loaded.

ETL CONVERSION- VOCABULARIES

- Downloaded the vocabularies from ATHENA server: OMOP has a server for Vocabularies.
- Some default files: Concepts, Concept relationship and Vocabularies etc.
- The Standardized Vocabularies contains records, or Concepts, that uniquely identify each fundamental unit of meaning used to express clinical information in all domain tables of the CDM.
- Concepts are CDM-specific entities that represent the meaning of a clinical fact.
- Concepts have unique IDs across all domains.

SINGLE CONCEPT REFERENCE TABLE



Vocabularies in a single **CONCEPT** Table

DRAW ETL SPECIFICATION

OHDSI Tools helped us to draw the ETL specification on which we moved ahead for further transformation.

It is equivalent to mapping specification we generally use while mapping data as per CDISC SDTM.

Each CDM-ETL process includes logic to exclude source data that we do not believe is of sufficient quality for research purposes. For example, we applied a consistent set of logic that excluded patients if the source data indicated multiple genders or multiple year of birth records.

WHITE RABBIT

When building the ETL, the data were first categorized with an open source tool called White Rabbit, listing all tables, fields, and distinct values in those fields.

White Rabbit analyzes the structure and content of a database and report is generated which contains all fields, attributes, Max length and frequency in the specific dataset.

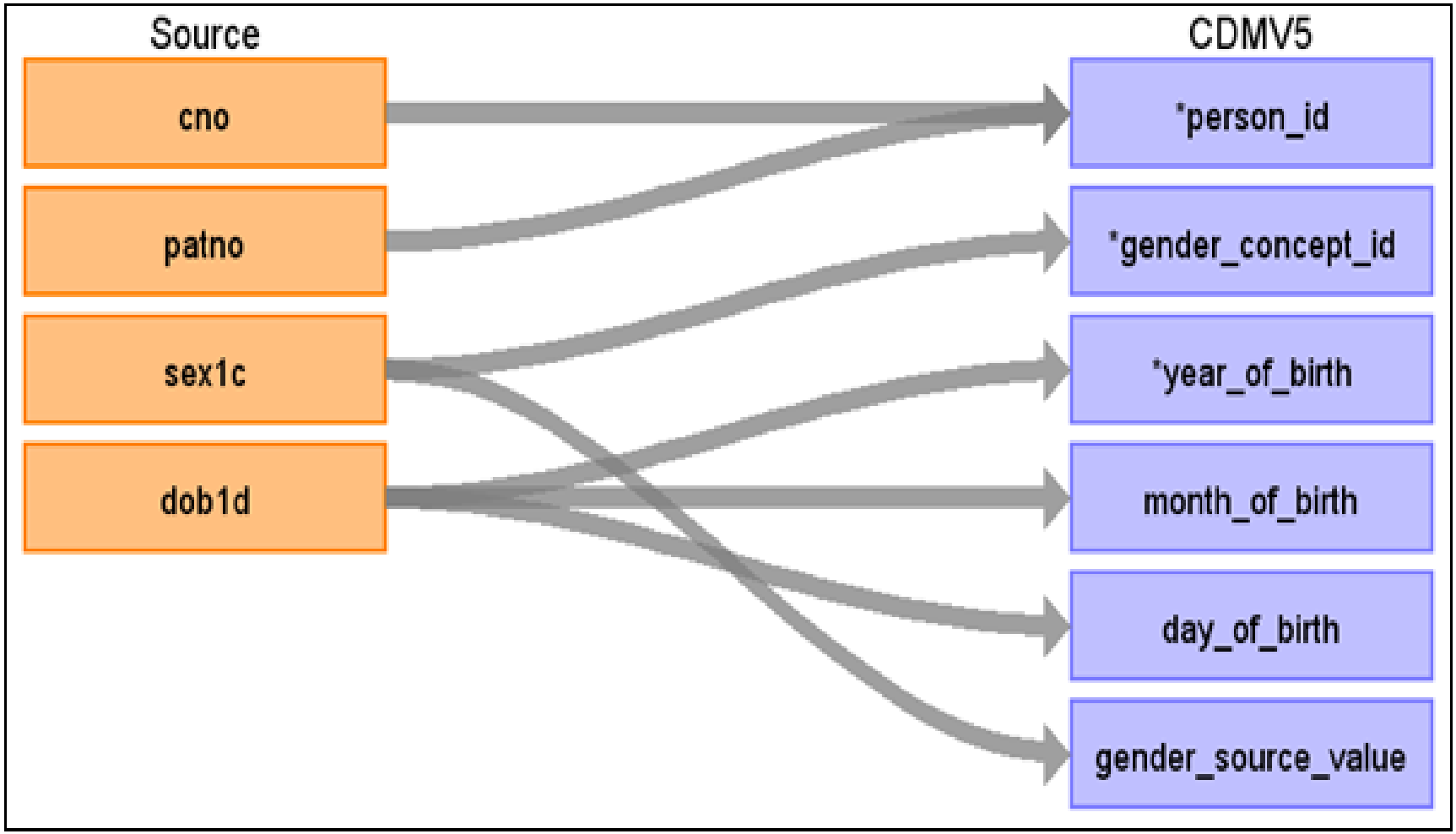
Report need to be uploaded in RabbitInAHat tool.

RABBIT IN A HAT

We documented our ETL with a tool called RabbitInAHat. This interactive application takes the results from WhiteRabbit and allows the user to connect data tables and columns from the raw dataset to where they will map into the OMOP CDM dataset.

- The output RabbitInAHat is a requirements document for building an ETL.
- Helps you define the logic in a consistent way does not generate code to create ETL
- Independent programming to recreate logic with SAS or SQL and iterating until results produced and validated.

Mapping PERSON Table



Order of populating CDM tables

Step 1.

Location

Care_site

Person

Provider

Death

Step 2.

- Visit_occurrence
- Condition_occurrence
- Procedure_occurrence
- Drug_exposure
- Device_exposure
- Observation
- Measurement

Step 3.

- Observation_period

Step 4.

- Condition_era
- Drug_era
- Dose_era

Mapping process

We extensively used Base SAS and SAS Sql to map the Tables according to OMOP CDM guideline.

There were many codes that we could not map for our study, but most of them were not valid codes therefore we assigned their concept id as 0.

Extensively followed the OMOP CDM guideline and generated Tables.

CDM Visualization

- We used PostgreSQL (open source) to upload our tables (changed sas to csv format) by writing scripts first to create tables an then to upload our data.

- To use AchillesWeb to explore the Achilles statistics, we first exported the statistics to a folder JSON files.

- Executed Achilles Heel (component of Achilles), to validate our mapping process

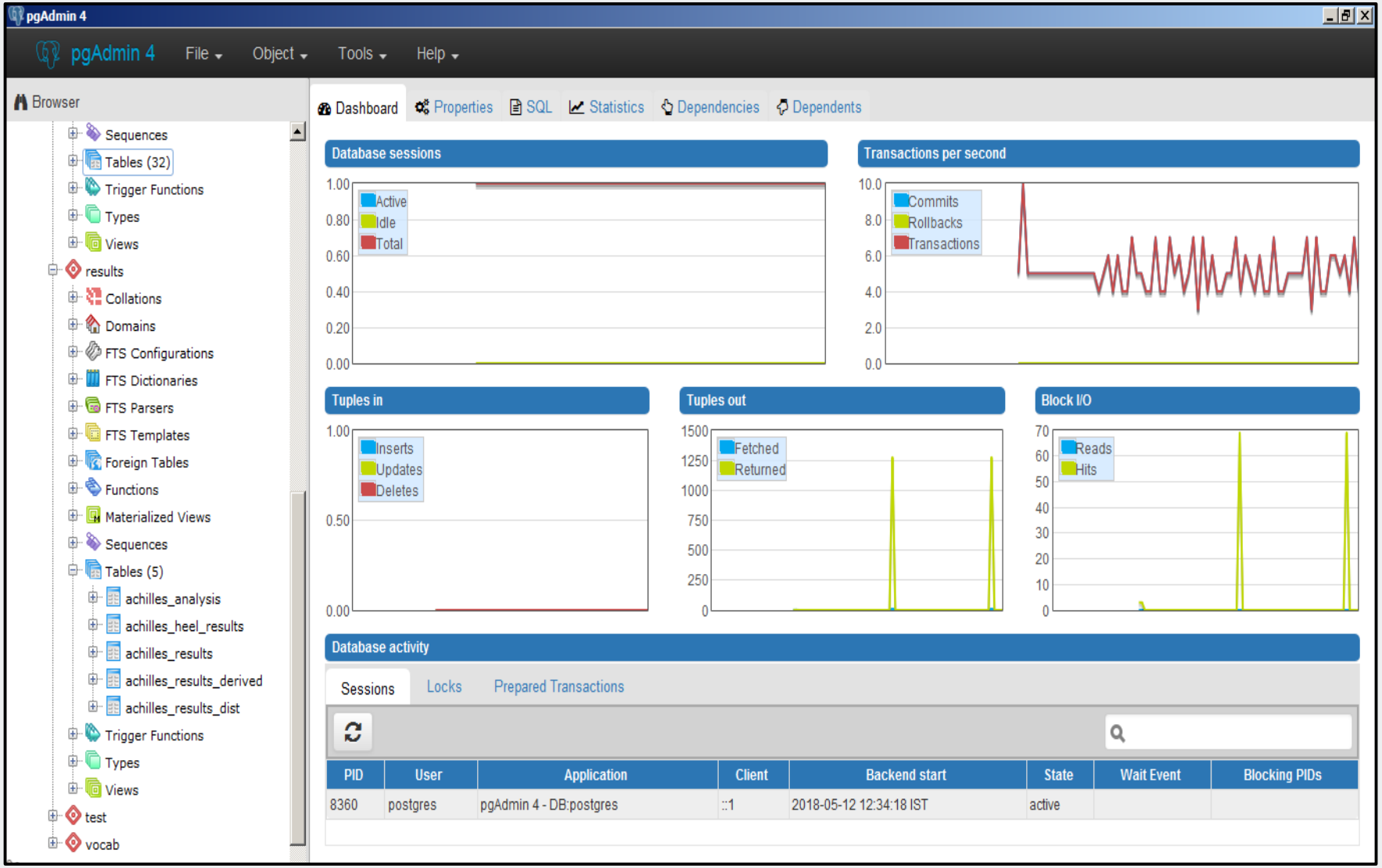
CDM Visualization

- Created a 'data' directory in the root of your AchillesWeb folder and run the Achilles R package exportToJSON method specifying the output path to a subdirectory for the data source.

- Last step was to create a json file at our web server root file and visualize the outputs with the help of local Host.

- Access the AchillesWeb 'index.html' as a web page.

PostgreSQL



ACHILLES

(Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems)

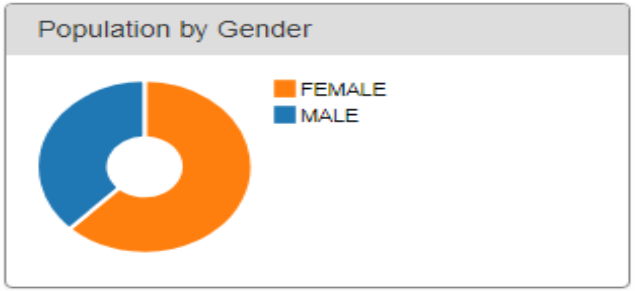
- An open source analytics framework
- Interactively explore population-level summary statistics for the data stored in CDM
  - Profile your CDM data
  - Explore population-level summaries
  - Review data quality assessment

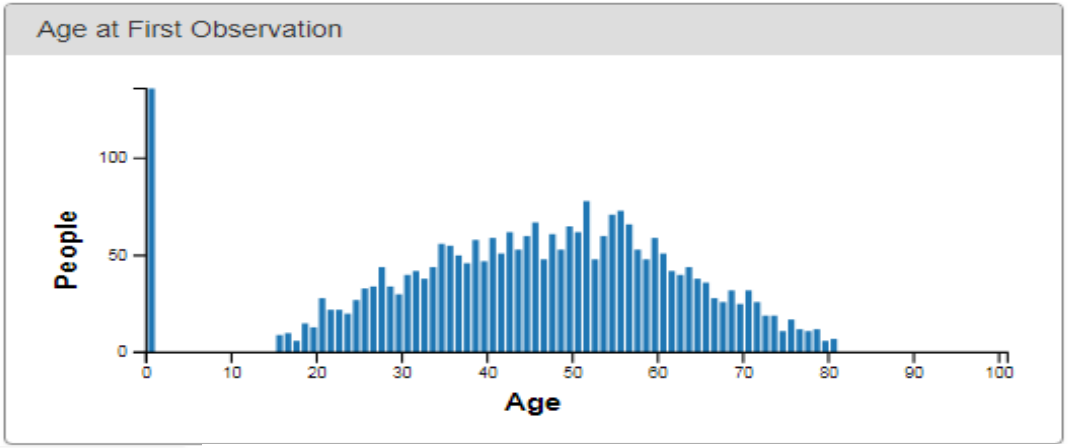
ACHILLES Summary Statistics

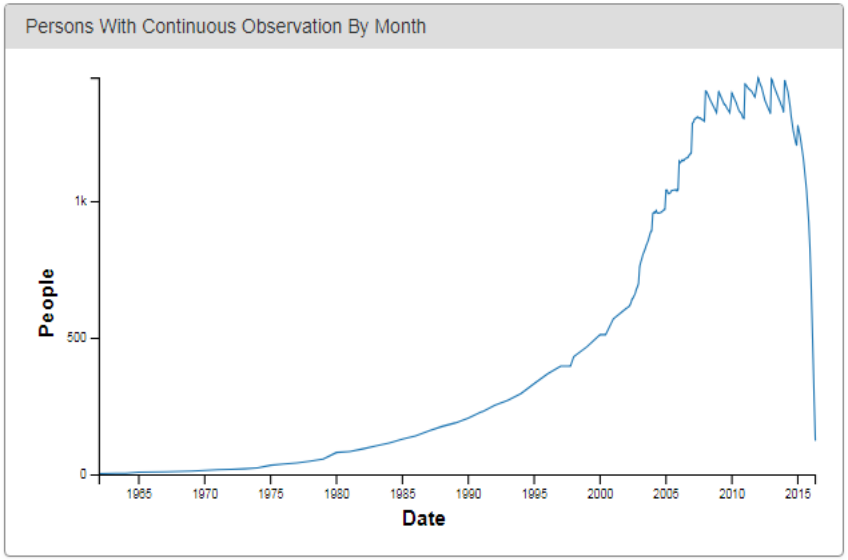
Summary of data set / clinical database

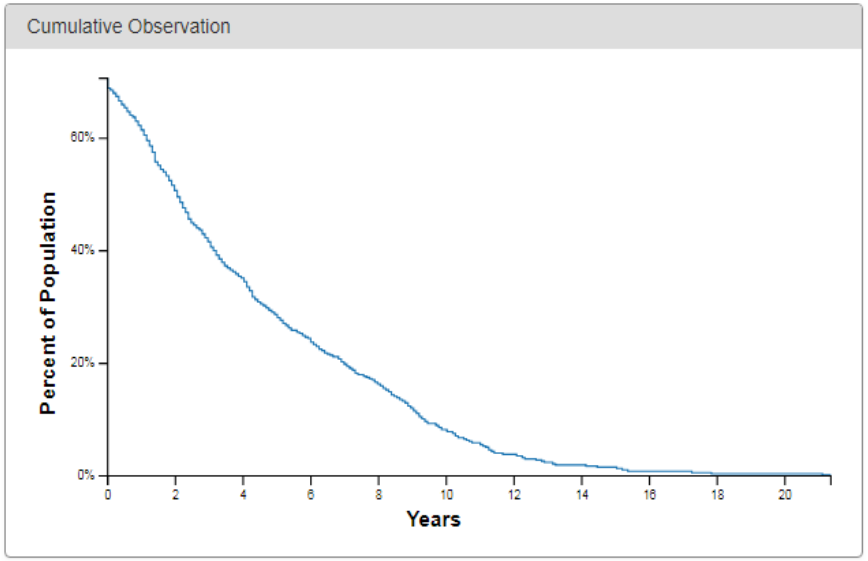
- Size of the database
- Gender Distribution and Age at first observation
- Cumulative /Continuous observation

ACHILLES Summary Statistics 1





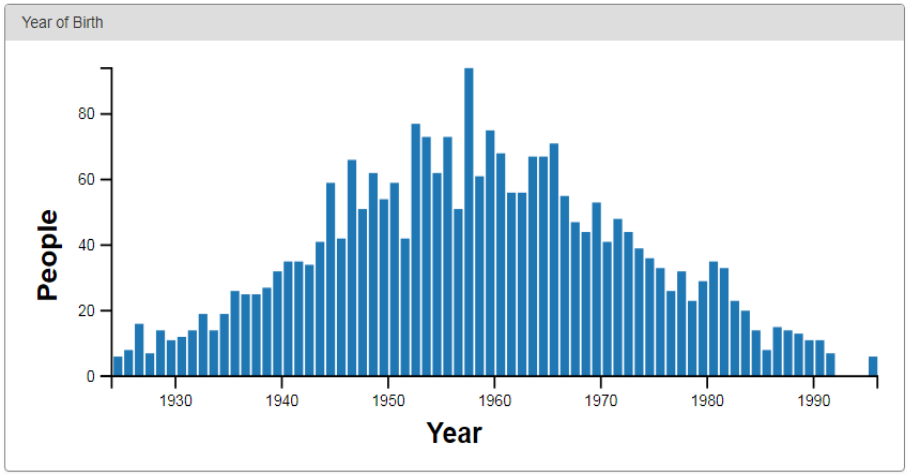


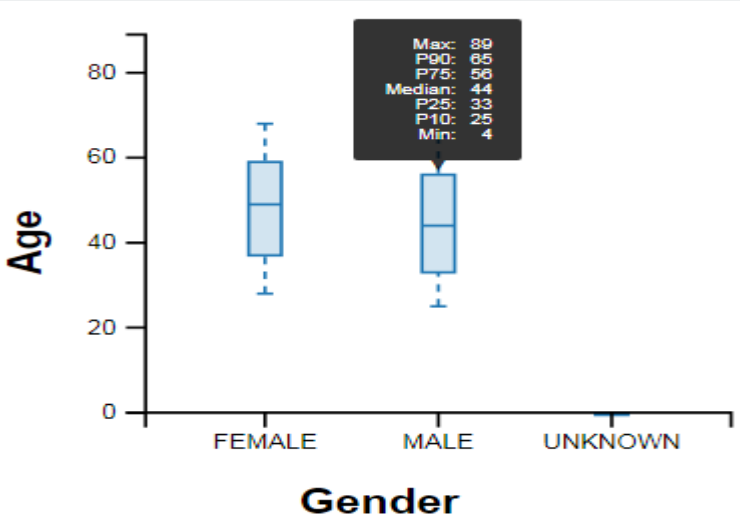


ACHILLES Summary Statistics

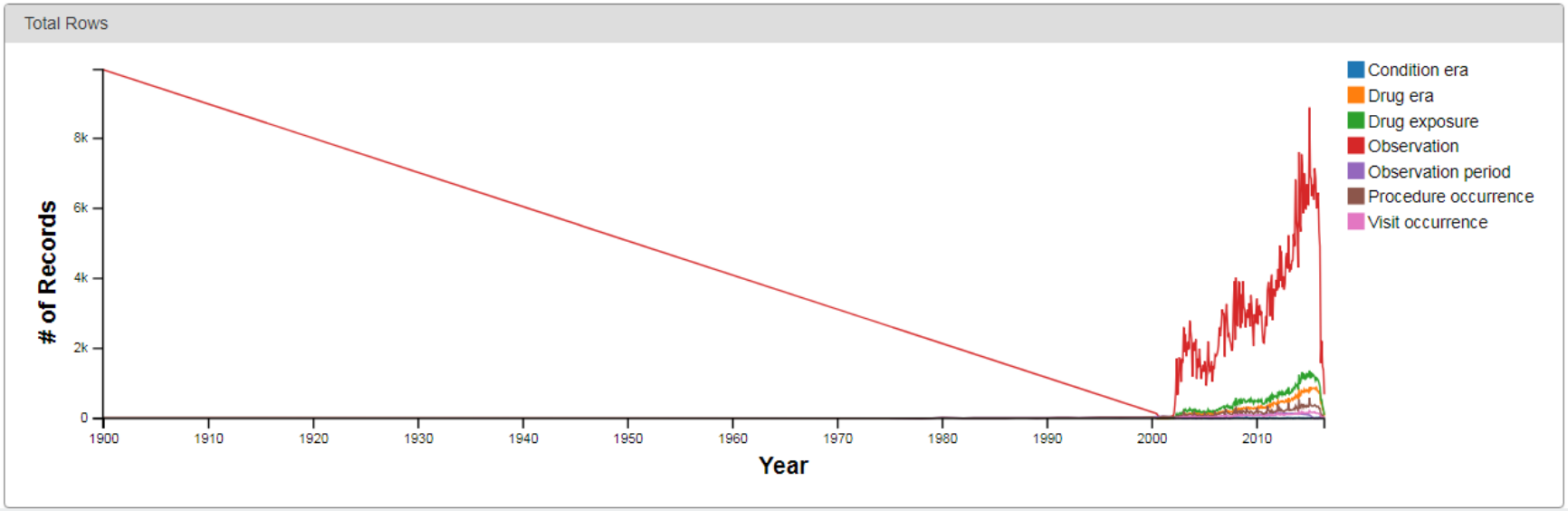
- Person demographic information
- Year of birth, Race and Ethnicity
- Length of Observation patient continued
- Duration of by Gender and Age decile
- Metadata (e.g. observation periods, data density) – Observation periods document time intervals during which health care information captured – Data density describes the unit quantity of records and concepts pertains in each database

Achilles Results

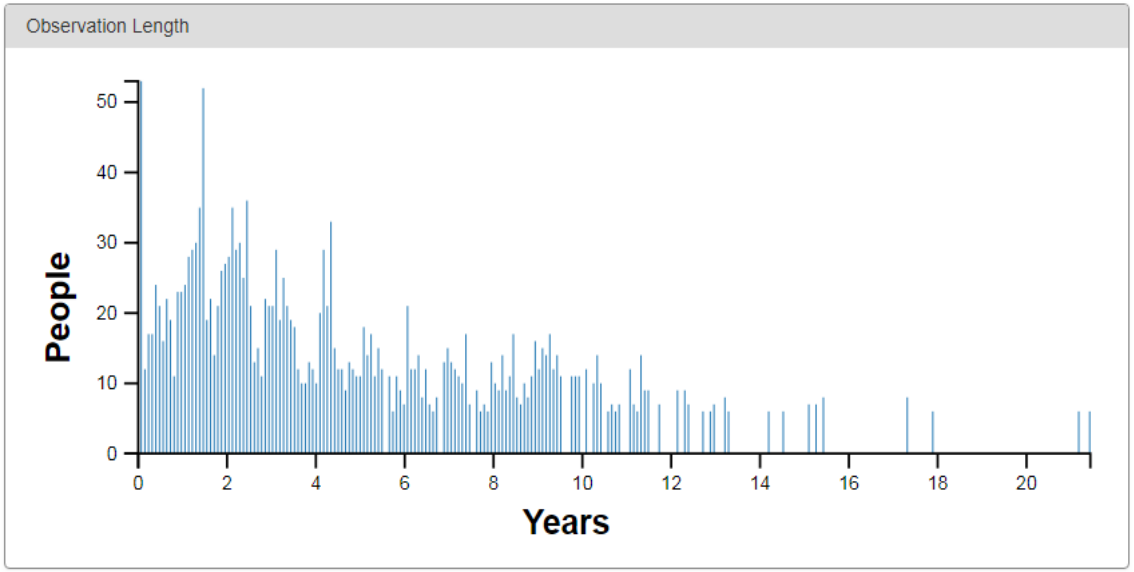


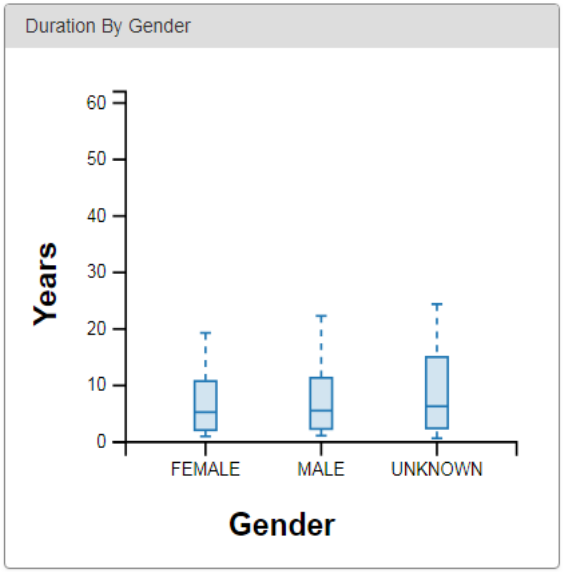


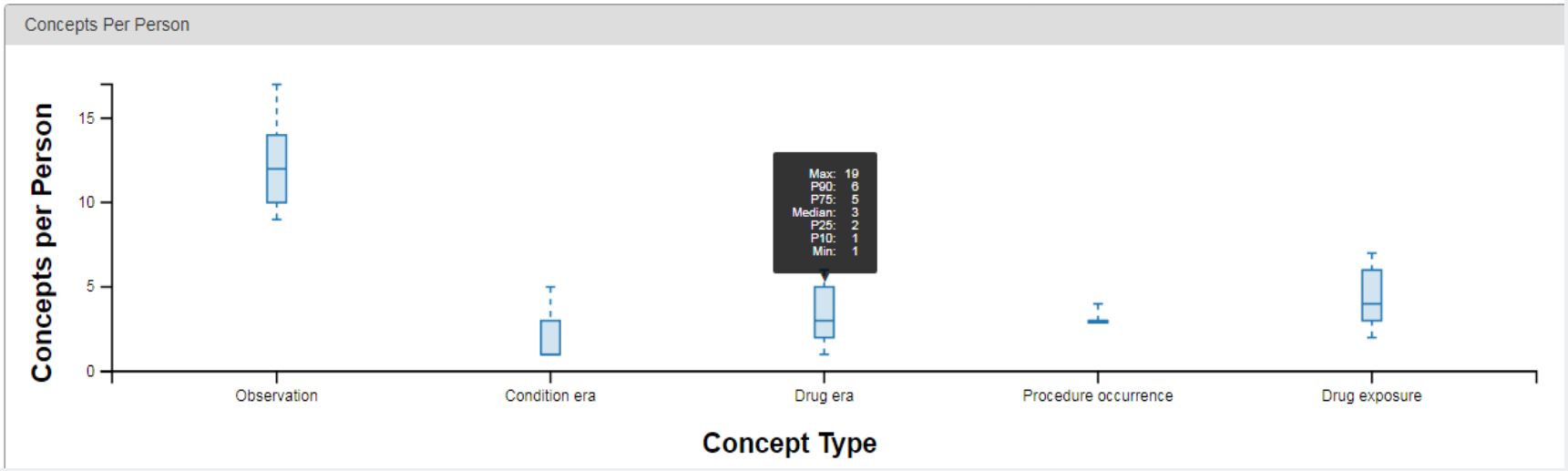
Data Density



Achilles Results



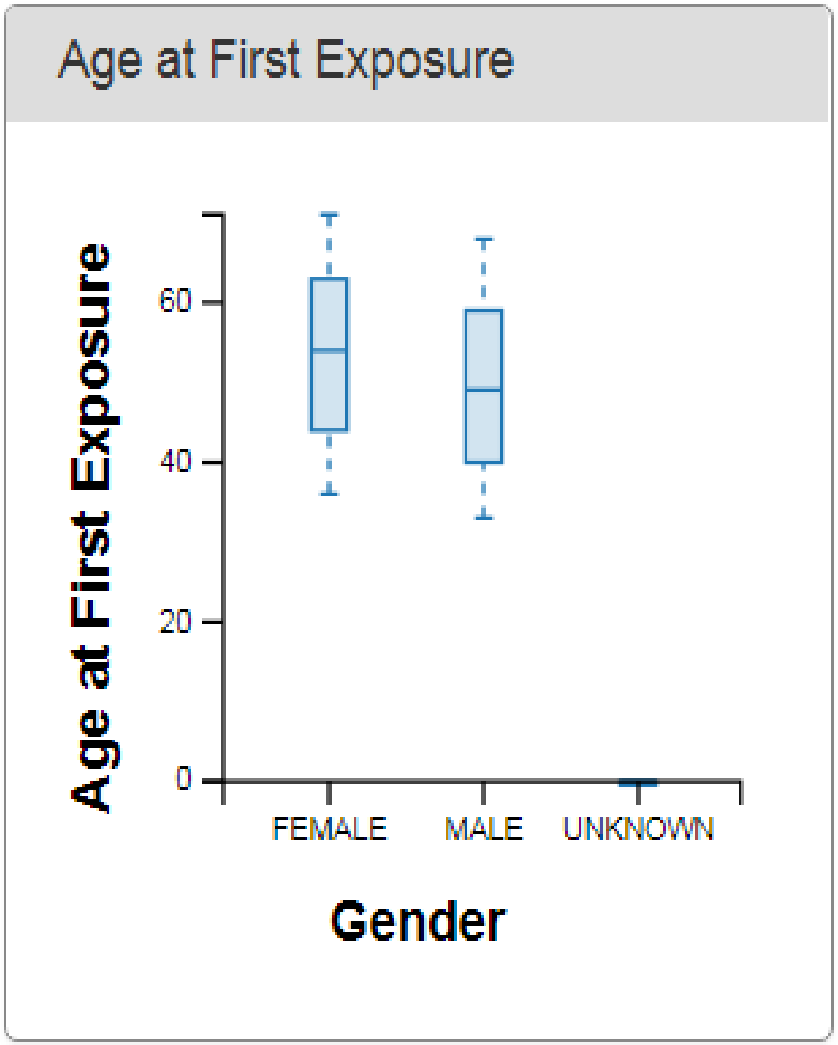


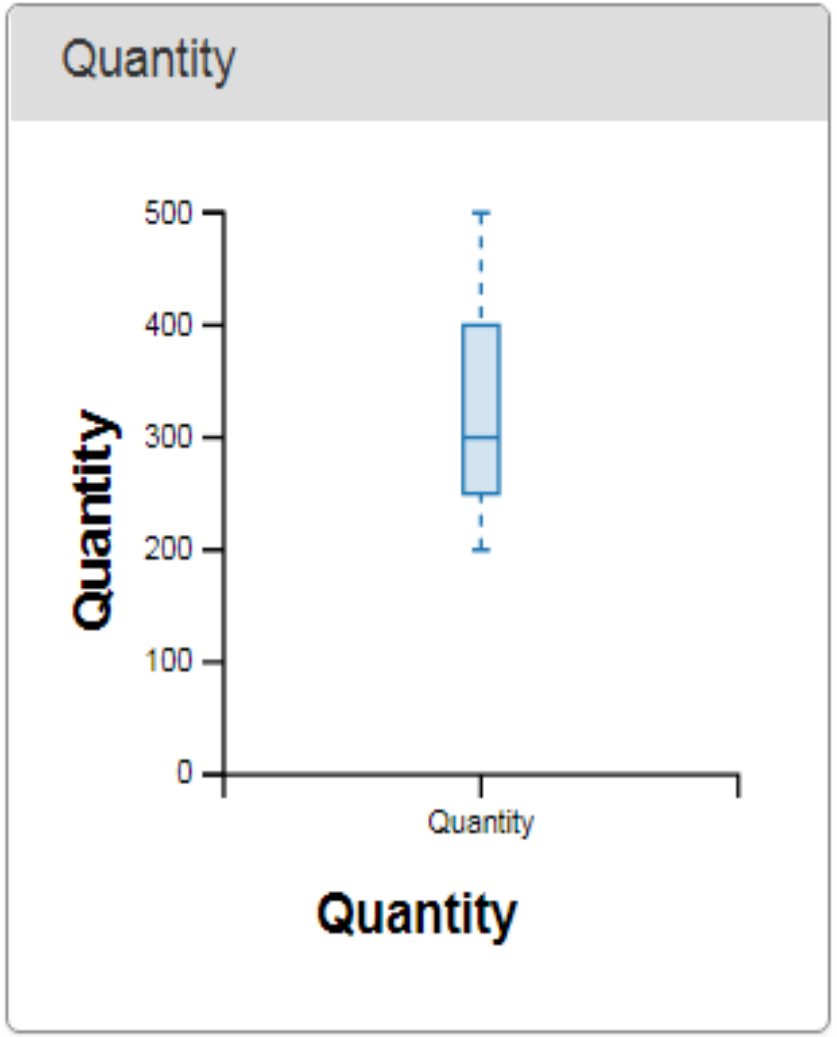


ACHILLES Summary Statistics

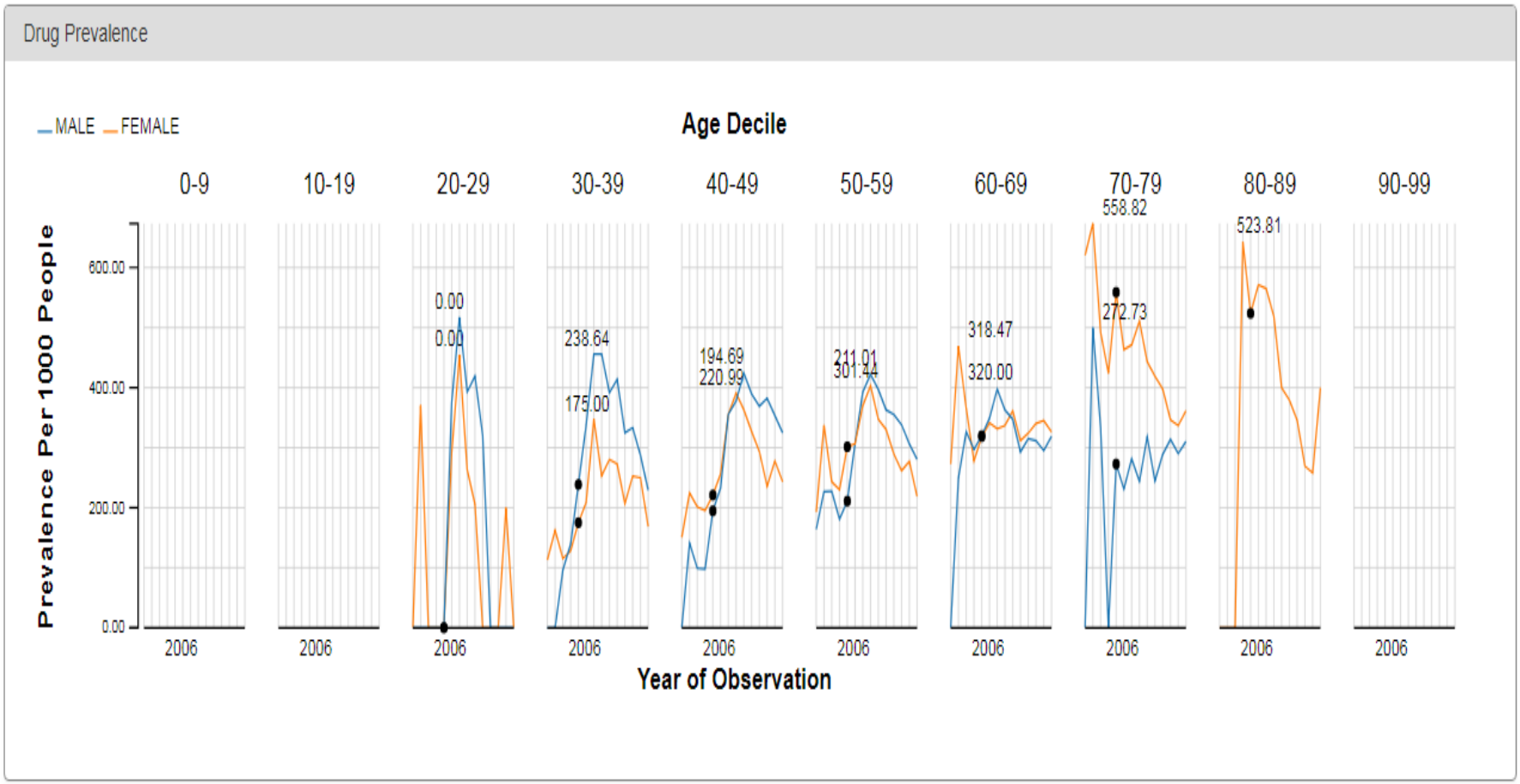
- Drug Exposure
- Age at First Exposure
- Dose descriptive for population
- Dose prevalence (Per 1000 People) by age decile

Achilles Results

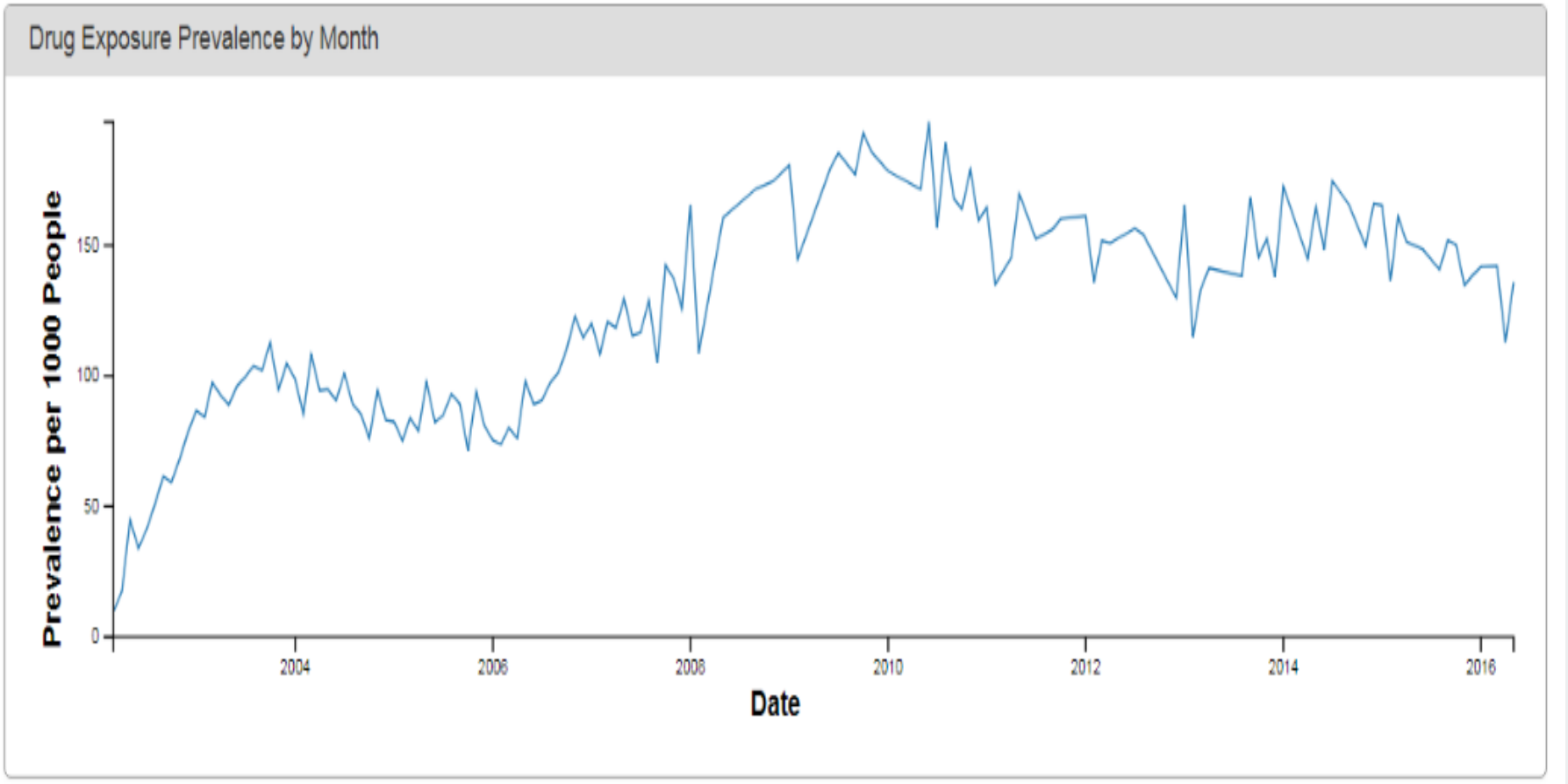




Achilles Results



Achilles Results



ACHILLES Summary Statistics

- Prevalence of condition/condition era/

  observation/drug exposure/drug era/procedure/visit

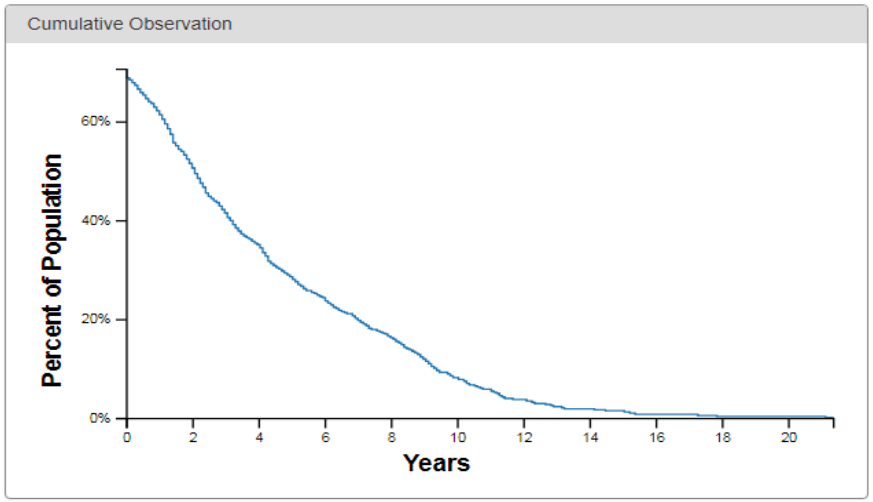
- It can be shown by 3 ways:

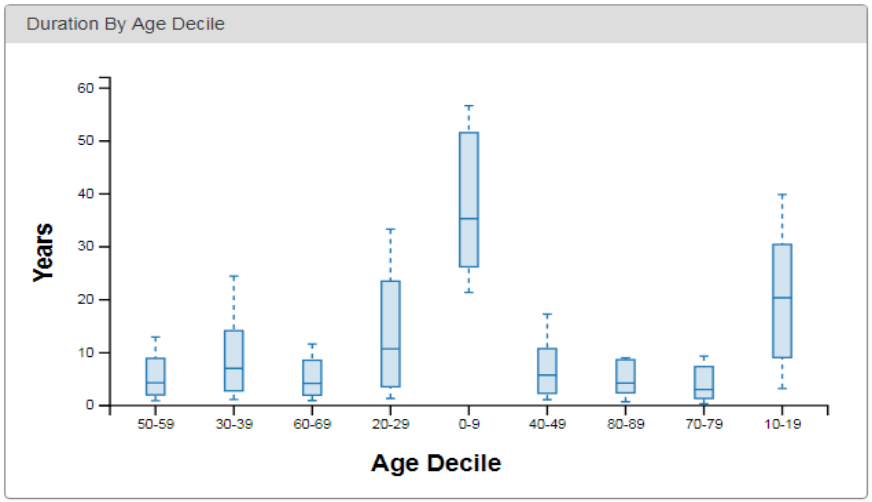
- Tree-map view

- Table view

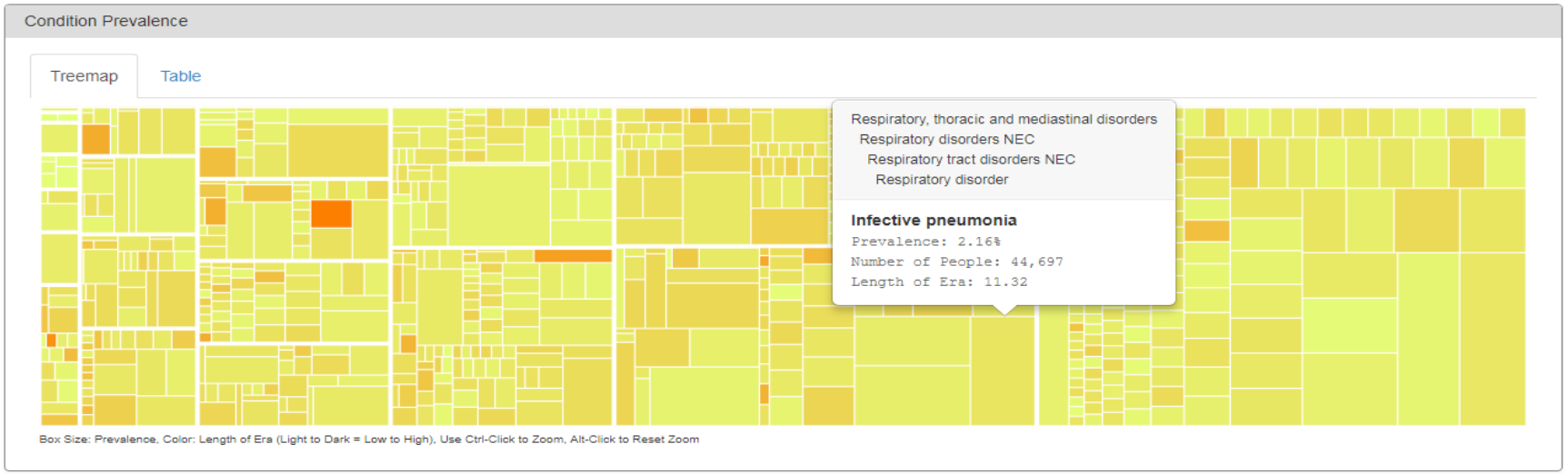
- Drill down view

Achilles Results





Condition Eras



Achilles Results Tree Map

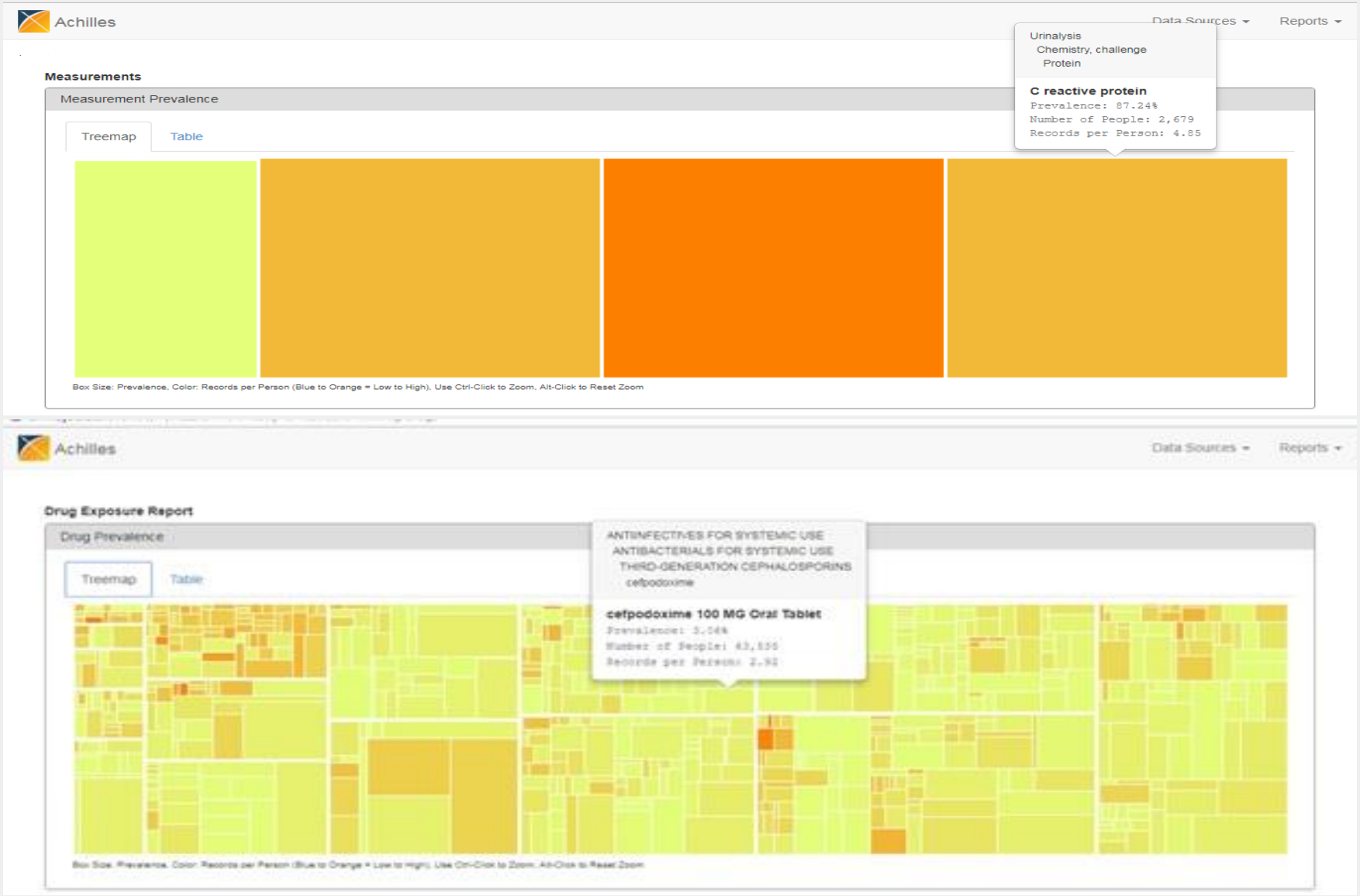


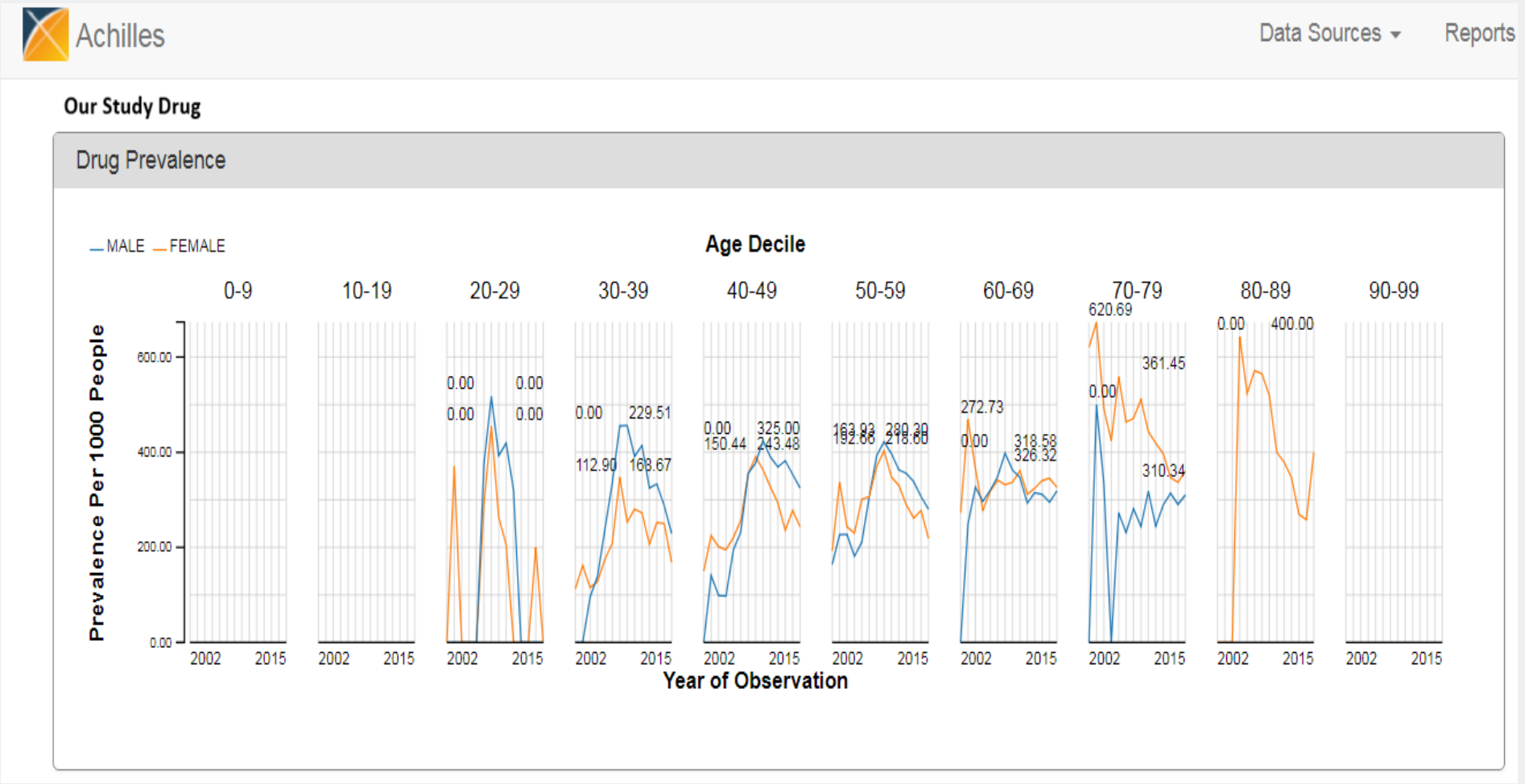
Table view

Search: Show / hide columns

RxNorm	Person Count	Prevalence	Records per Person
Hydroxychloroquine	530	17.26%	2.95
leflunomide	503	16.38%	2.68
Sulfasalazine	286	9.31%	2.79
Naproxen	265	8.63%	2.23
Ibuprofen	201	6.54%	1.87

Previous 1 2 3 4 5 … 19 Next

Drill Down View



Achilles Heel-Data quality tool

Achilles HEEL

- It describes the quality of data
- Report can be generated in various format and further data can be improved by reducing the errors and warnings.
- Equivalent to OpenCDISC Validator report however it is not as detailed.

Achilles Heel-Data quality tool

Achilles Heel Report

Data Quality Messages

Search:

Message Type	Message
ERROR	610-Number of procedure occurrence records outside valid observation period; count (n=28) should not be > 0
ERROR	810-Number of observation records outside valid observation period; count (n=10,006) should not be > 0
ERROR	907- Distribution of drug era length, by drug_concept_id(count = 71); min value should not be negative
ERROR	910-Number of drug eras with end date < start date; count (n=964) should not be > 0
ERROR	1,007- Distribution of condition era length, by condition_concept_id(count = 8); min value should not be negative
ERROR	1,010-Number of condition eras with end date < start date; count (n=8) should not be > 0
NOTIFICATION	Unmapped data over percentage threshold in:DrugExposure
NOTIFICATION	[GeneralPopulationOnly] Not all deciles represented at first observation
NOTIFICATION	[GeneralPopulationOnly] Count of distinct specialties of providers in the PROVIDER table is below threshold
NOTIFICATION	No body weight data in MEASUREMENT table (under concept_id 3,025,315 (LOINC code 29,463-7))
WARNING	4-Number of persons by race; data with unmapped concepts
WARNING	5-Number of persons by ethnicity; data with unmapped concepts
WARNING	500-Number of persons with death, by cause_concept_id; data with unmapped concepts
WARNING	512-Distribution of time from death to last drug(count = 1); max value should not be positive, otherwise its a zombie with data >1mo after death
WARNING	602-Number of persons by procedure occurrence start month, by procedure_concept_id; 2concepts have a 100% change in monthly count of events

Showing 1 to 15 of 24 entries

Print

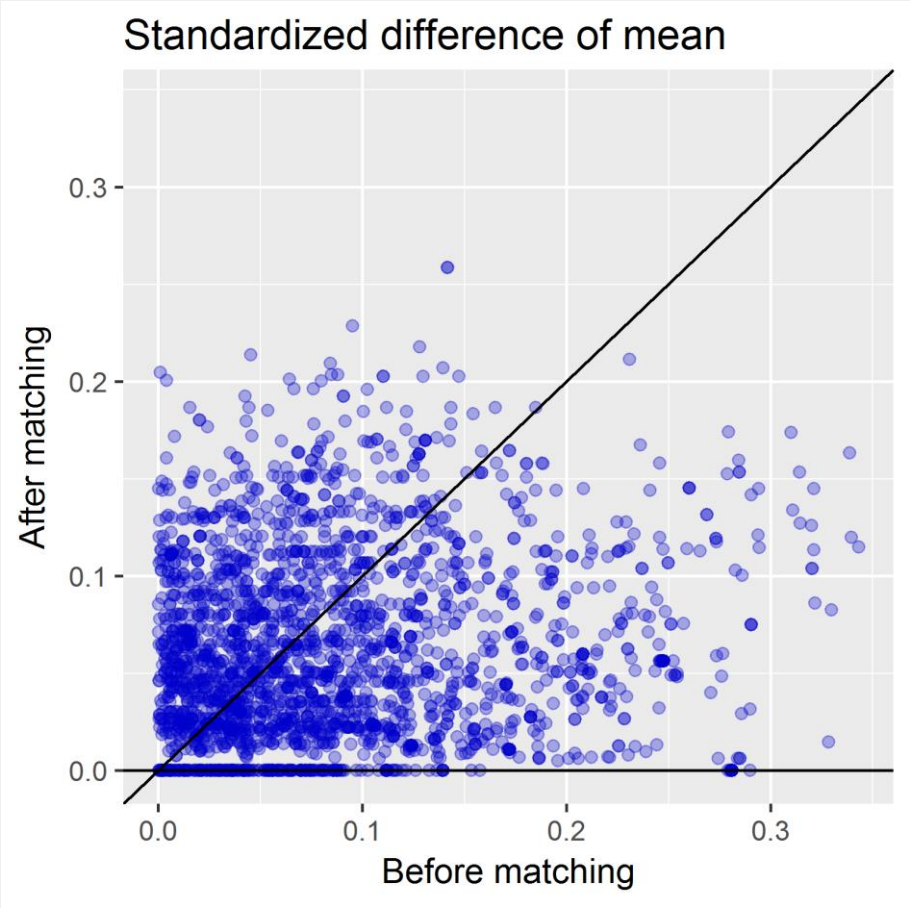
Previous 1 2 Next

ROAD AHEAD

We are planning to go ahead and explore more specific statistical results by using several R packages enabling more in-depth analyses on the data.

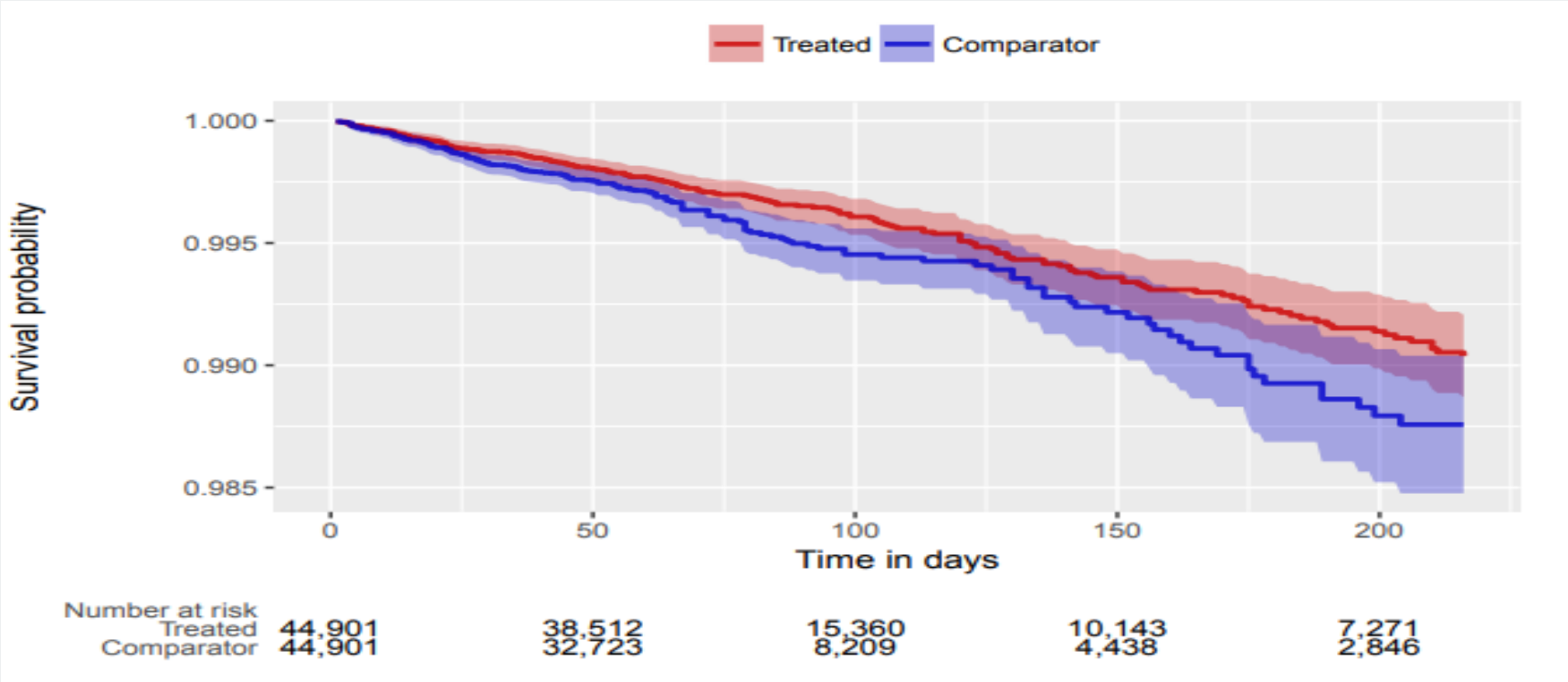
- We can utilize tools like **Cyclops** (Cyclic coordinate descent for logistic, Poisson and survival analysis) for performing large scale regularized regressions and the **CohortMethod** which can use propensity scores to adjust for potential confounders. It uses thousands to millions of covariates that are automatically constructed based on conditions, procedures and drugs in the records of the subjects.

ROAD AHEAD

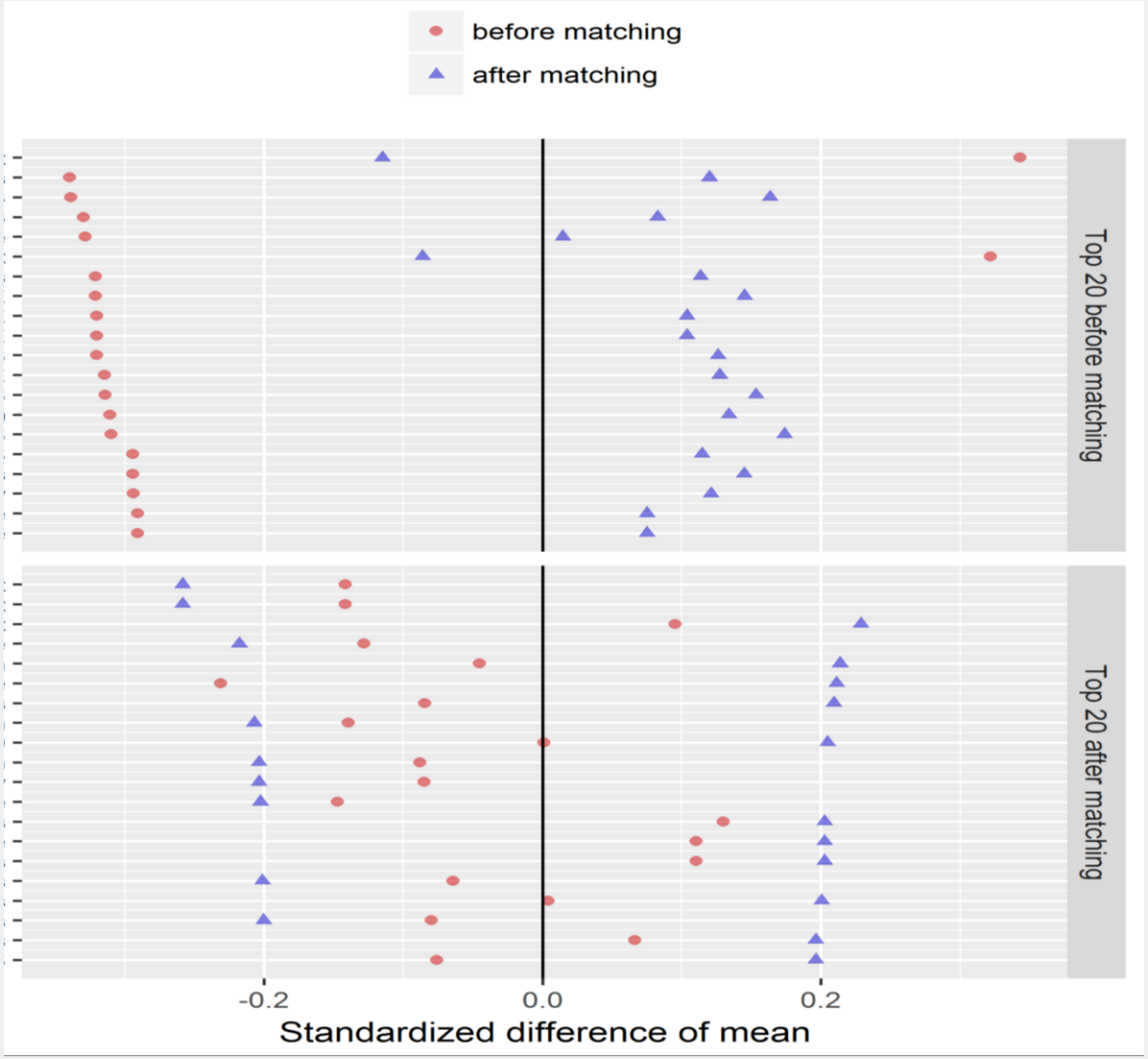


ROAD AHEAD

- **HOMER** (Health Outcomes and Medical Effectiveness Research) enables risk identification and comparative effectiveness studies, with real-time exploration of the effects of medical products.



ROAD AHEAD

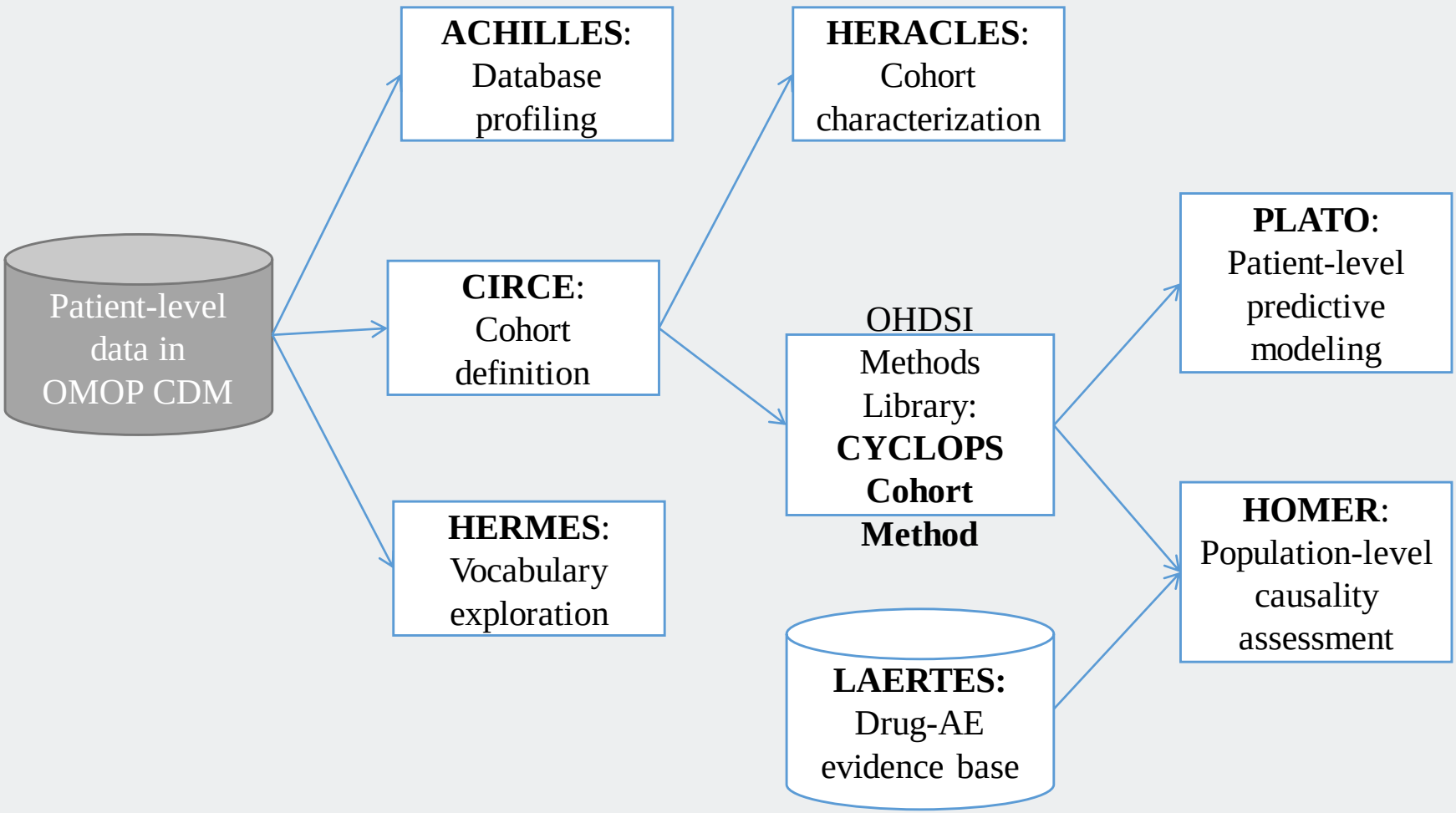


CHALLENGES

Like every new thing we also faced various difficulties during the ETL activities,

- Limited information regarding ETL
- Mapping guideline could be more descriptive
- Forum could be more supportive to commercial entities
- Not Enough material to explained the knowhow of tools
- Require expertise from various fields like programming, medical coding, statistical knowledge and IT skills.

OHDSI tools under development



CONCLUSION

The biggest advantage of using OMOP Common Data Model (OMOP CDM) is that once we have mapped our data, it allows different observational datasets to essentially “speak the same language” when a researcher performs an analysis.

One of the potential costs is loss of information, Not all source codes may map into OMOP Vocabulary concepts. Most loss of information can be attributed to our exclusion rules, which were aimed at improving the quality of the data.

REFERENCES

- www. github.com/OHDSI
- www.ohdsi.org/analytic-tools
- George Hripcsak, Jon D Duke, Nigam H Shah et. al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers; Stud Health Technol Inform. 2015; 216: 574–578.
- Stephanie Reisinger, Gary Schneider, Matthew Reynolds, et al. Will the Growing Reliance on Real-World Data Fuel Fundamental Changes in the Way We Approach Database Analyses. THE EVIDENCE FORUM May 2015
- Rachel E. Sherman, M.D., M.P.H., Steven A. Anderson, Ph.D., M.P.P., et al. Real-World Evidence — What Is It and What Can It Tell Us?. n engl j med 375;23 nejm.org December 8, 2016
- OMOP Common Data Model Specifications Version 5.0 , October 14, 2014

Thank You!

JSS Medical Research Inc.

9400 Henri-Bourassa Blvd. West

St-Laurent, Montreal, Quebec

CANADA H4S 1N8

T. 514 934 6116 • F. 514 934 9913

www.jssresearch.com

JSS Medical Research India Limited

Plot No. 12/2, 6th Floor,

Vatika Mindscapes Tower – B

Sarai Khwaja Metro Station,

Mathura Road, Faridabad-121003

www.jssresearch.com

Follow us on   *for publications and updates*