OHDSI Tools Ecosystem

Patrick Ryan, PhD
Janssen Research and Development, Columbia University Medical Center

Peter Rijnbeek, PhD
Erasmus MC

Martijn Schuemie, PhD
Janssen Research and Development

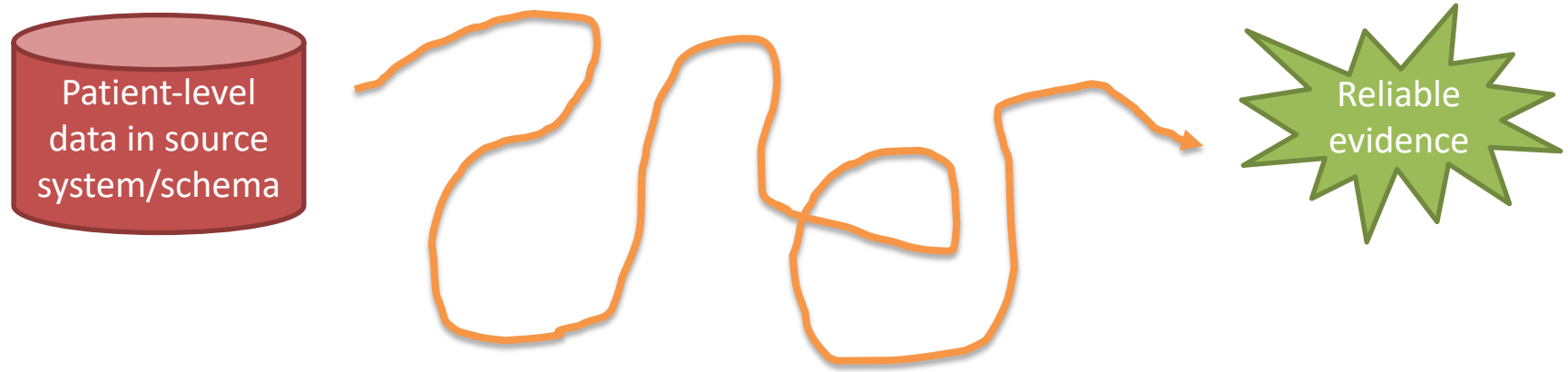
Jenna Reps, PhD
Janssen Research and Development

Agenda

800am-900am	Registration and coffee	
900am-1000am	Overview of OHDSI tools ecosystem	Patrick Ryan
1000am-1030am	Vocabulary	Patrick Ryan
1030am-1045am	Coffee Break	
1045am-1115am	Data sources	Peter Rijnbeek
1115am-1230pm	Cohort definition and characterization	Patrick Ryan
1230pm-130pm	Lunch	
130pm-200pm	Incidence rate	Patrick Ryan
200pm-230pm	Population-level effect estimation	Martijn Schuemie
230pm-300pm	Patient-level prediction	Jenna Reps
300pm-500pm	Design and implement your own OHDSI study!	You!!!
330pm-345pm	Coffee Break	
500pm	Drinks and snacks	

Overview of the OHDSI tools ecosystem

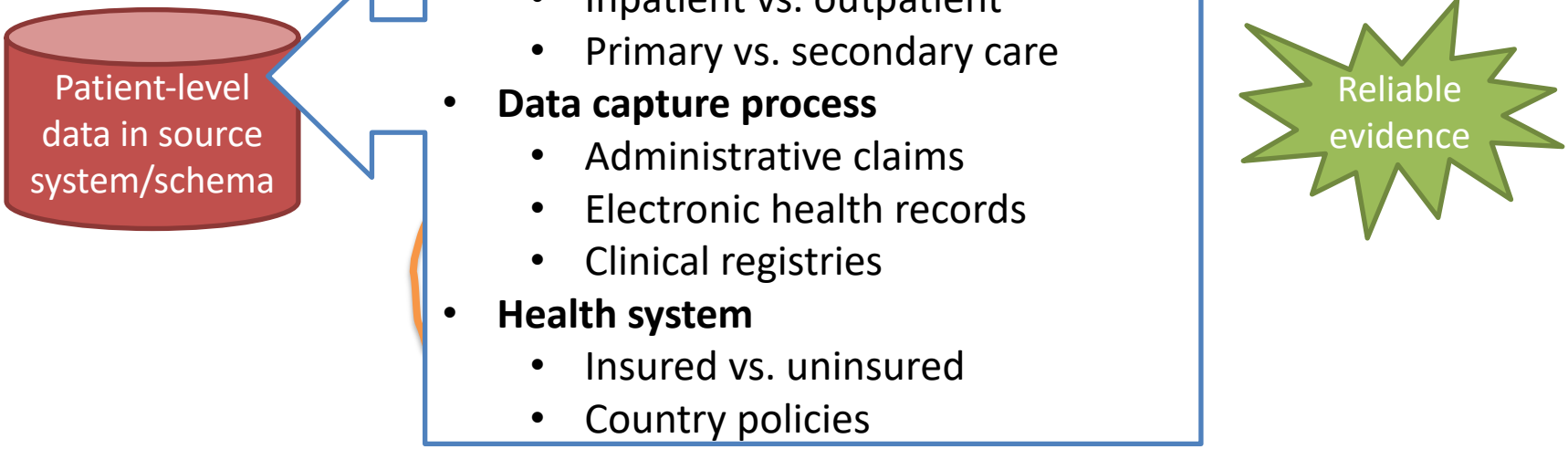
The journey to real-world evidence



One-time

Repeated

The journey to real-world evidence



One-time Repeated

The journey to real-world evidence

Patient-level data in source system/schema

Types of evidence desired:

- **Cohort identification**
  - Clinical trial feasibility and recruitment
- **Clinical characterization**
  - Treatment utilization
  - Disease natural history
  - Quality improvement
- **Population-level effect estimation**
  - Safety surveillance
  - Comparative effectiveness
- **Patient-level prediction**
  - Precision medicine
  - Disease interception

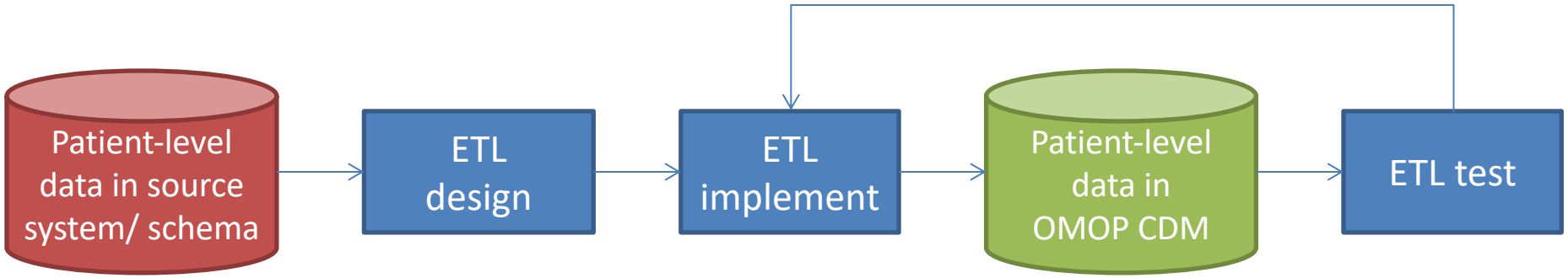
Reliable evidence

One-time

Repeated

What are your research questions?

Structuring the journey from source to a common data model



OHDSI tools built to help

WhiteRabbit: profile your source data

RabbitInAHat: map your source structure to CDM tables and fields

ATHENA: standardized vocabularies for all CDM domains

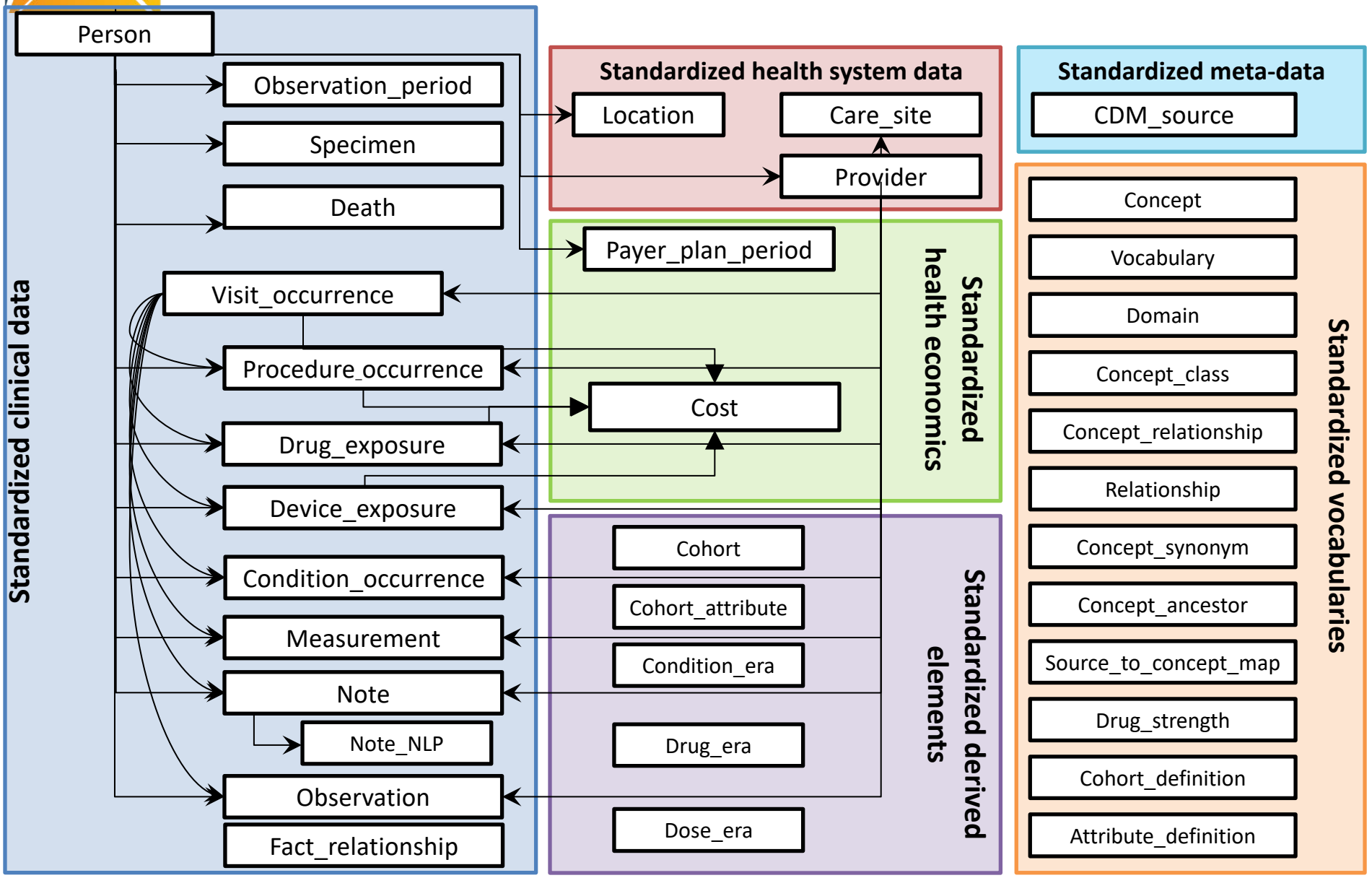
Usagi: map your source codes to CDM vocabulary

CDM: DDL, index, constraints for various RDBMS flavors; Vocabulary tables with loading scripts

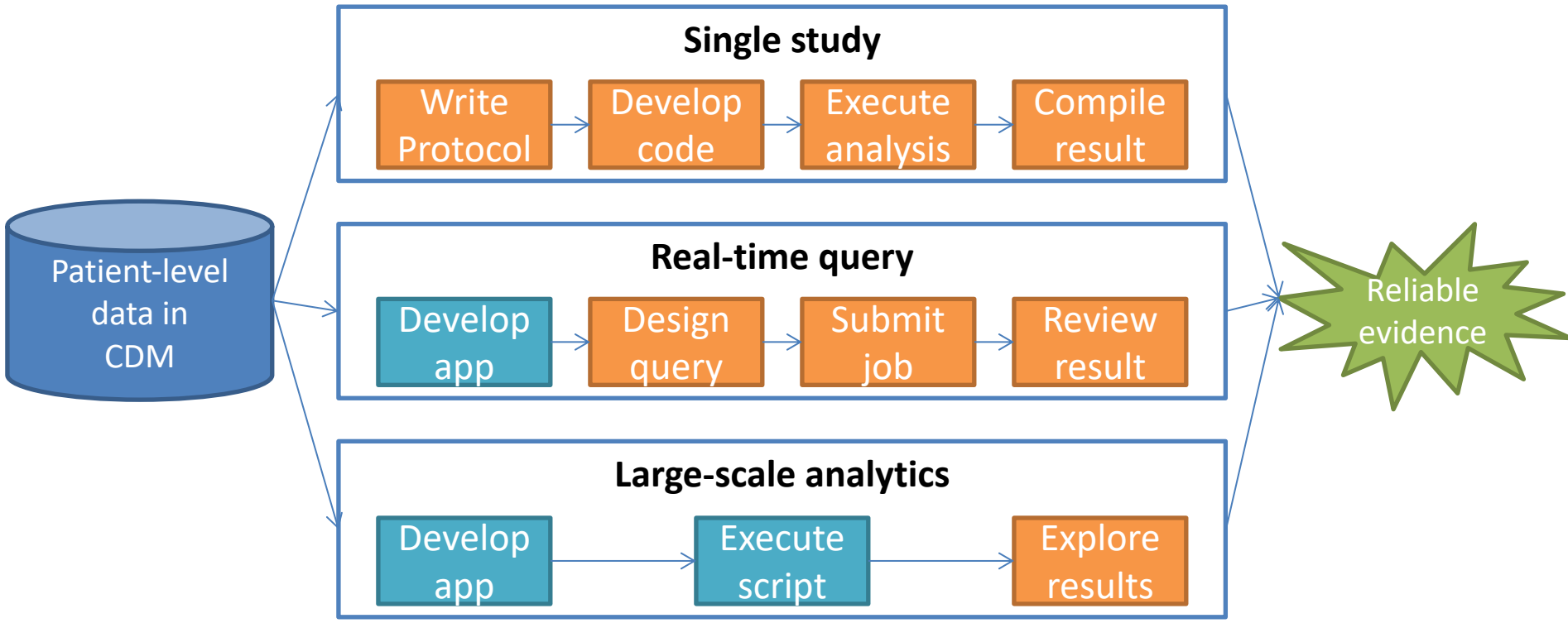
ACHILLES: profile your CDM data; review data quality assessment; explore population-level summaries

OHDSI Forums: Public discussions for OMOP CDM Implementers/developers

OMOP CDM


Standardized clinical data
Person
Observation_period
Specimen
Death
Visit_occurrence
Procedure_occurrence
Drug_exposure
Device_exposure
Condition_occurrence
Measurement
Note
Note_NLP
Observation
Fact_relationship
Standardized health system data
Location
Care_site
Provider
Payer_plan_period
Standardized health economics
Cost
Standardized derived elements
Cohort
Cohort_attribute
Condition_era
Drug_era
Dose_era
Standardized meta-data
CDM_source
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength
Cohort_definition
Attribute_definition

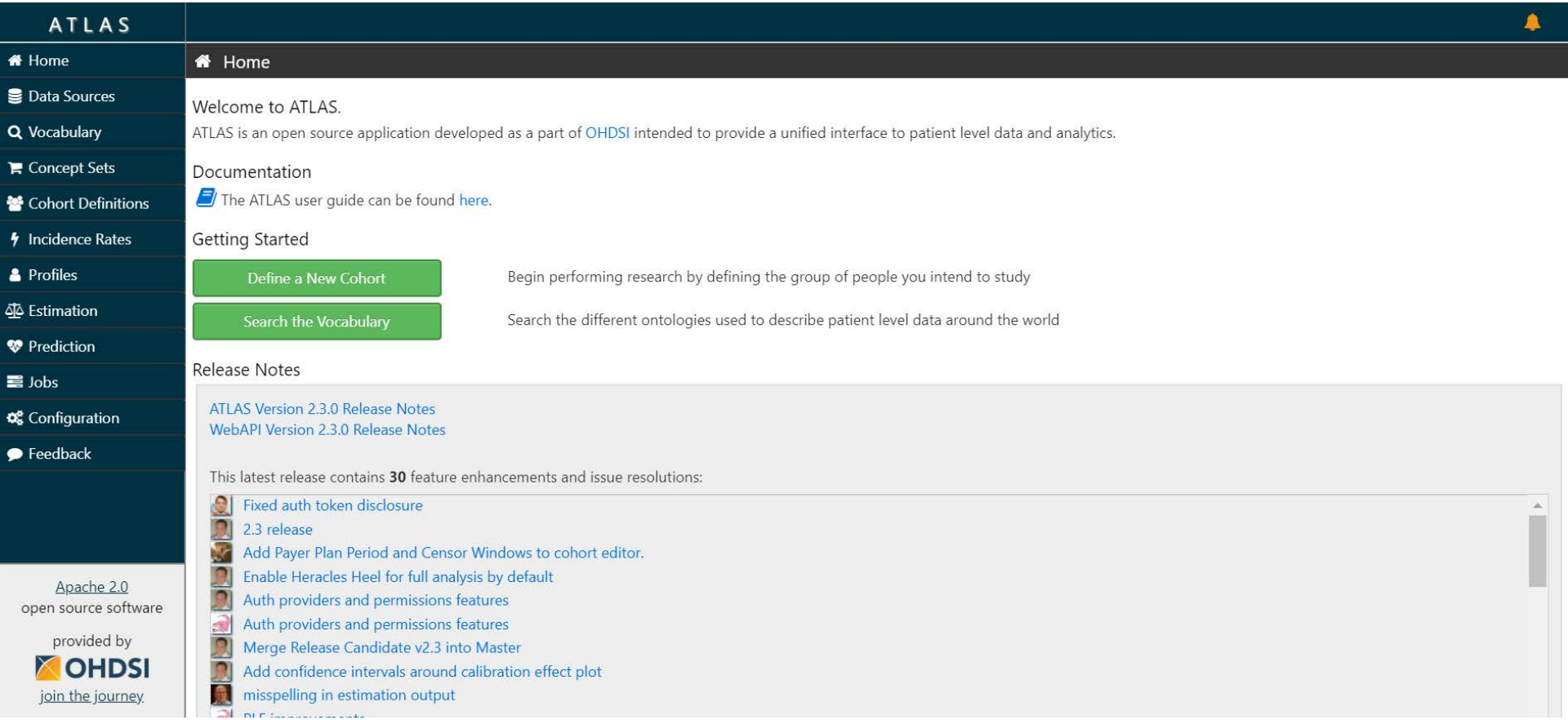
Structuring the journey from a common data model to evidence



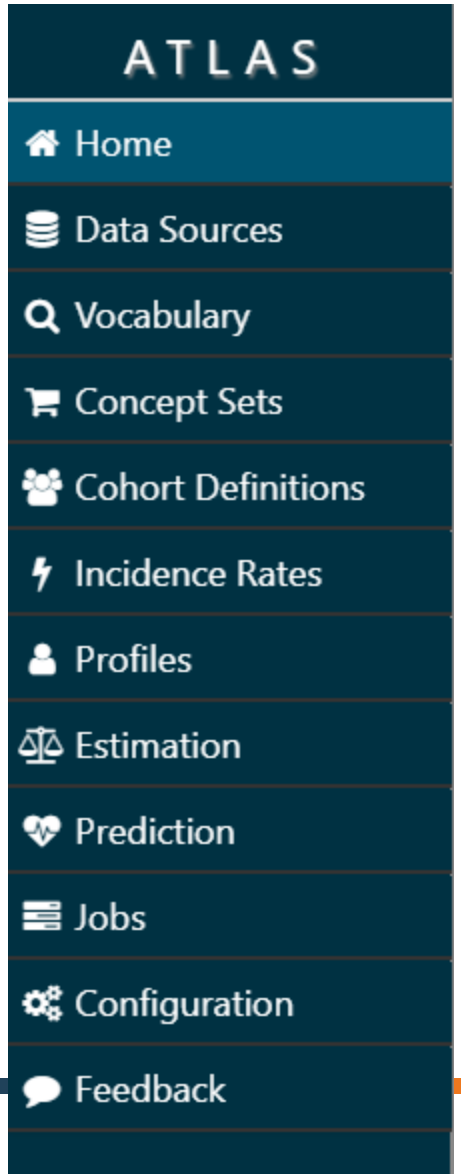
One-time

Repeated

ATLAS – an open-source platform to design and execute observational analyses



Analytic use cases supported in ATLAS



Estimation methods

Cohort Method

New-user cohort studies using large-scale regression for propensity and outcome models

Self-Controlled Case Series

Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.

Self-Controlled Cohort

A self-controlled cohort design, where time preceding exposure is used as control.

IC Temporal Pattern Disc.

A self-controlled design, but using temporal patterns around other exposures and outcomes to correct for time-varying confounding.

Case-control

Case-control studies, matching controls on age, gender, provider, and visit date. Allows nesting of the study in another cohort.

Case-crossover

Case-crossover design including the option to adjust for time-trends in exposures (so-called case-time-control).

Prediction methods

Patient Level Prediction

Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.

Feature Extraction

Automatically extract large sets of features for user-specified cohorts using data in the CDM.

Method characterization

Empirical Calibration

Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.

Method Evaluation

Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods.

Supporting packages

Database Connector

Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.

Sql Render

Generate SQL on the fly for the various SQL dialects.

Cyclops

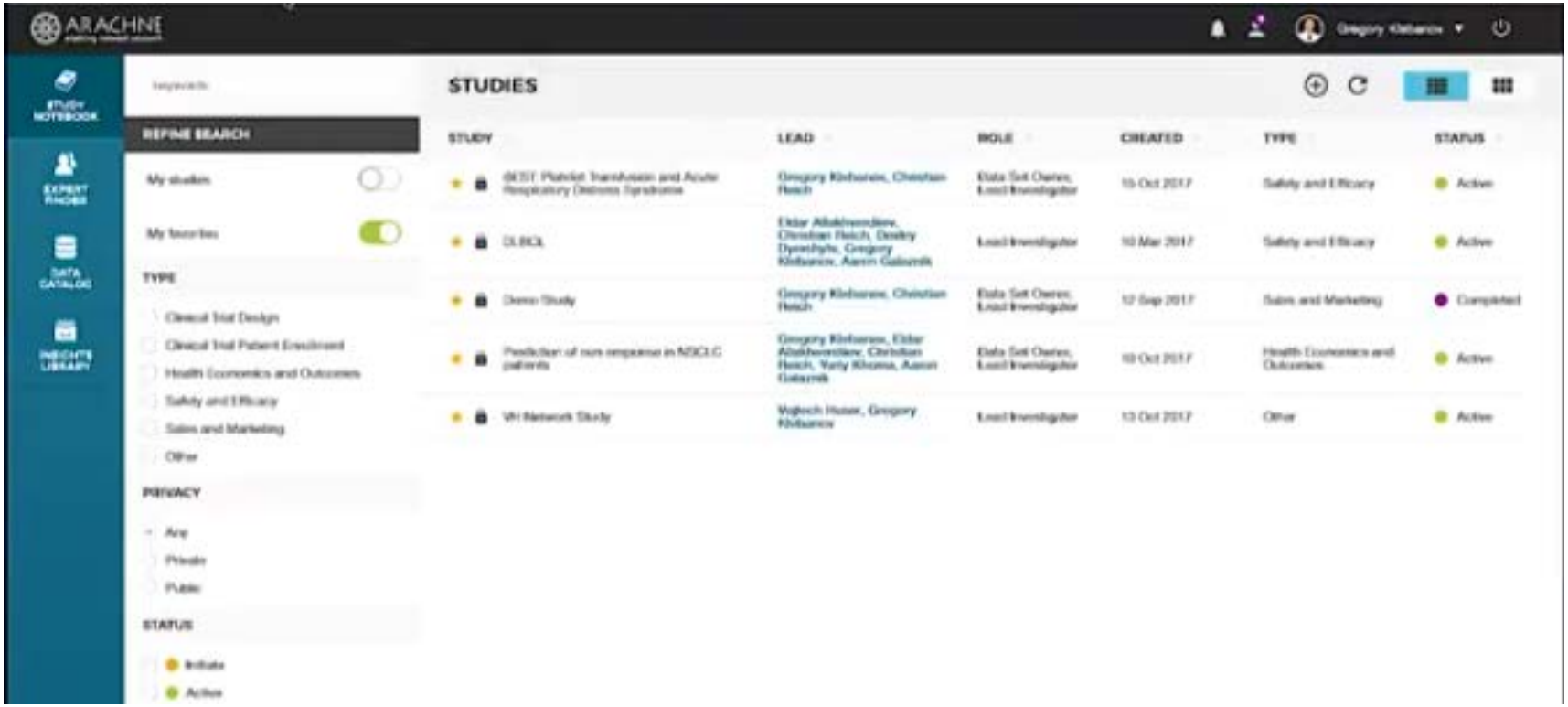
Highly efficient implementation of regularized logistic, Poisson and Cox regression.

Ohdsi R Tools

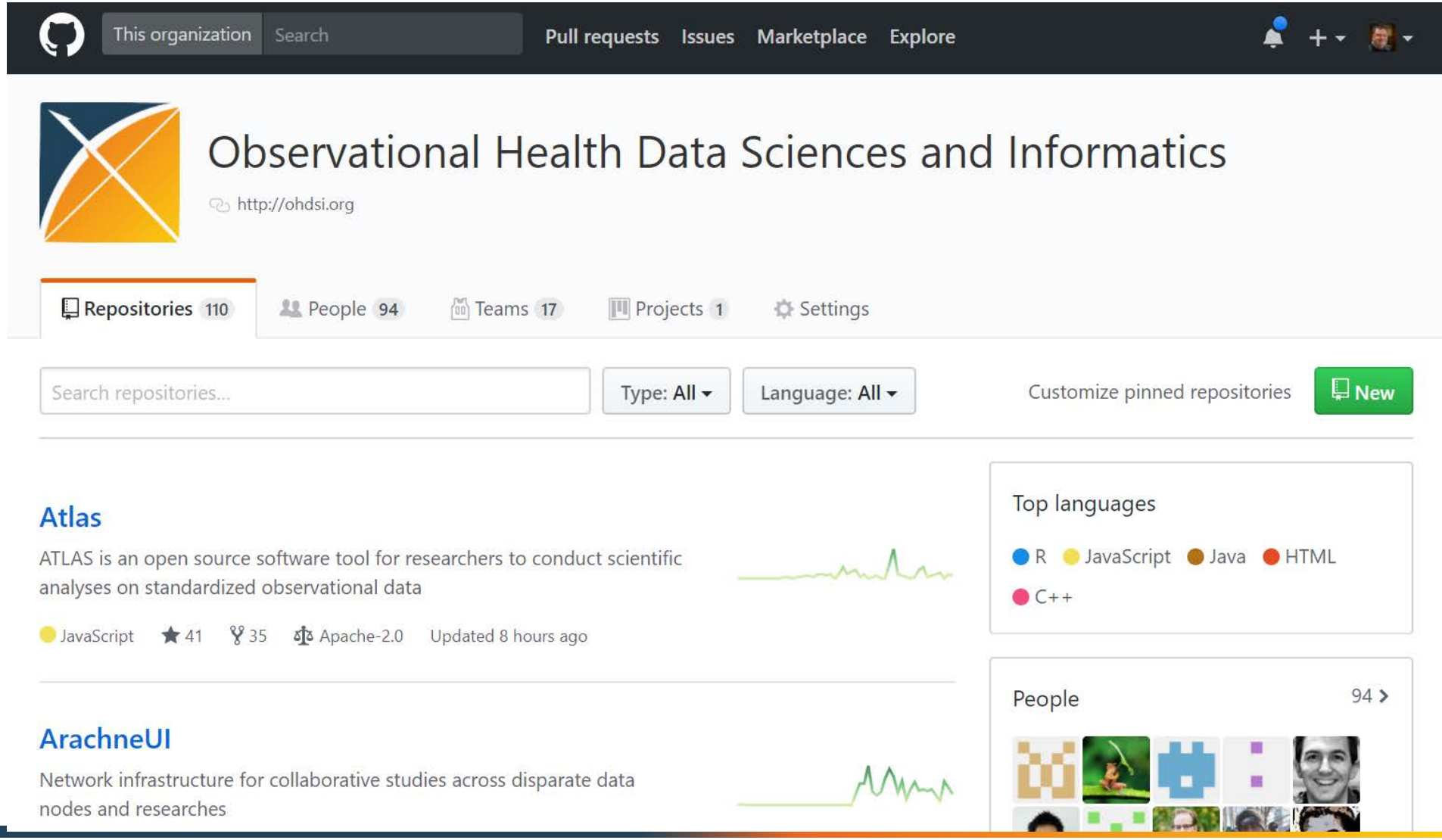
Support tools that didn't fit other categories, including tools for maintaining R libraries.

Under construction

ARACHNE – an open-source platform to enable analyses across the OHDSI network



All of OHDSI tools are open source and freely available



http://github.com/OHDSI

Cardiovascular, Bleeding, and Mortality Risks in Elderly Medicare Patients Treated With Dabigatran or Warfarin for Nonvalvular Atrial Fibrillation

David J. Graham, MD, MPH; Marsha E. Reichman, PhD; Michael Wernecke, BA; Rongmei Zhang, PhD; Mary Ross Southworth, PharmD; Mark Levenson, PhD; Ting-Chang Sheu, MPH; Katrina Mott, MHS; Margie R. Goulding, PhD; Monika Houstoun, PharmD, MPH; Thomas E. MaCurdy, PhD; Chris Worrall, BS; Jeffrey A. Kelman, MD, MMSc

Background—The comparative safety of dabigatran versus warfarin for treatment of nonvalvular atrial fibrillation in general practice settings has not been established.

Methods and Results—We formed new-user cohorts of propensity score–matched elderly patients enrolled in Medicare who initiated dabigatran or warfarin for treatment of nonvalvular atrial fibrillation between October 2010 and December 2012. Among 134414 patients with 37587 person-years of follow-up, there were 2715 primary outcome events. The hazard ratios (95% confidence intervals) comparing dabigatran with warfarin (reference) were as follows: ischemic stroke, 0.80 (0.67–0.96); intracranial hemorrhage, 0.34 (0.26–0.46); major gastrointestinal bleeding, 1.28 (1.14–1.44); acute myocardial infarction, 0.92 (0.78–1.08); and death, 0.86 (0.77–0.96). In the subgroup treated with dabigatran 75 mg twice daily, there was no difference in risk compared with warfarin for any outcome except intracranial hemorrhage, in which case dabigatran risk was reduced. Most patients treated with dabigatran 75 mg twice daily appeared not to have severe renal impairment, the intended population for this dose. In the dabigatran 150-mg twice daily subgroup, the magnitude of effect for each outcome was greater than in the combined-dose analysis.

Conclusions—In general practice settings, dabigatran was associated with reduced risk of ischemic stroke, intracranial hemorrhage, and death and increased risk of major gastrointestinal hemorrhage compared with warfarin in elderly patients with nonvalvular atrial fibrillation. These associations were most pronounced in patients treated with dabigatran 150 mg twice daily, whereas the association of 75 mg twice daily with study outcomes was indistinguishable from warfarin except for a lower risk of intracranial hemorrhage with dabigatran. (***Circulation***. **2015;131:157-164. DOI: 10.1161/CIRCULATIONAHA.114.012061.)**

Key Words: anticoagulant ■ pharmacoepidemiology ■ safety ■ thrombin inhibitor ■ warfarin

- Baseline characterization of target and comparator cohort
- Descriptive summaries of:
  - Demographics
  - Medical history (prior conditions)
  - Medication use (prior drugs)
  - Prior procedures
  - Risk scores

Table 1. Sociodemographic Factors, Medical Conditions, and Medication Use at Baseline in Propensity Score–Matched Medicare Beneficiaries Initiating Dabigatran or Warfarin for Atrial Fibrillation, 2010–2012

Characteristic	Dabigatran, % (n=67 207)	Warfarin, % (n=67 207)	Standardized Mean Difference
Age group, y			
65–74	42	41	0.01
75–84	43	43	0.01
≥85	16	16	0.00
Female sex	51	52	0.01
Race/ethnicity			
White	92	92	0.00
Black	3	3	0.00
Other	5	5	0.00
Medical history			
General			
Diabetes mellitus	33	34	0.00
Hypercholesterolemia	74	74	0.00
Hypertension	87	87	0.00
Kidney failure			
Acute	5	5	0.00
Chronic	13	13	0.00
Obesity	11	11	0.00
Peptic ulcer disease	<1	<1	0.00
Prior bleeding event			
Hospitalized	1	1	0.00
Not hospitalized	3	3	0.01
Smoking	16	16	0.01
Cardiovascular disease			
Acute myocardial infarction			
Past 1–30 d	1	1	0.01
Past 31–183 d	1	1	0.00
Coronary revascularization	16	16	0.01
Heart failure			
Hospitalized	4	4	0.01
Outpatient	14	14	0.00
Other ischemic heart disease	48	49	0.01
Stroke			
Past 1–30 d	2	2	0.00
Past 31–183 d	1	2	0.00
Other cerebrovascular disease	13	13	0.00
Transient ischemic attack	7	7	0.00
Cardioablation	2	2	0.00
Cardioversion	9	9	0.02
Other medical conditions			
Falls	5	5	0.00
Fractures	2	2	0.00
Syncope	10	10	0.00
Walker use	3	3	0.00
$CHADS_2$ score*			
0–1	28	28	0.01
2	40	40	0.00
3	21	21	0.01
≥4	10	11	0.01
HAS-BLED score†			
1	9	9	0.01
2	50	50	0.01
3	32	32	0.01
≥4	9	9	0.00
Medication use			
General			
Estrogen replacement	2	3	0.00
H2 antagonists	5	5	0.00
NSAIDs	15	15	0.00
Proton pump inhibitors	26	27	0.01
SSRI antidepressants	13	13	0.01
Cardiovascular			
ACEI/ARB	59	59	0.00
Antiarrhythmics	25	25	0.01
Anticoagulants (injectable)	7	7	0.01
Antiplatelets	17	17	0.01
β-Blockers	70	71	0.00
Calcium channel blockers	42	42	0.01
Digoxin	17	16	0.00
Diuretics			
Loop	28	28	0.00
Potassium sparing	5	5	0.01
Thiazide	29	29	0.00
Nitrates	10	11	0.01
Statins	57	57	0.00
Fibrates	5	5	0.00
Diabetes related			
Insulin	6	6	0.00
Metformin	13	14	0.00
Sulfonylureas	9	10	0.00
Other	6	6	0.00
Metabolic inhibitors‡			
Amiodarone	10	10	0.00
Dronedarone	5	5	0.02
Verapamil	2	2	0.00
Azole antifungals	<1	<1	0.00

Additional factors included in the propensity score model are shown in the online-only Data Supplement. ACEI/ARB indicates angiotensin converting-enzyme inhibitor/angiotensin receptor blocker; NSAIDs, nonsteroidal anti-inflammatory drugs; and SSRI, selective serotonin reuptake inhibitor.

*The $CHADS_2$ score assigns points for the presence of congestive heart failure, hypertension, age ≥75 y, diabetes mellitus, stroke, or transient ischemic attack.[11]

†The HAS-BLED score assigns points for the presence of hypertension, abnormal renal or liver function, stroke, bleeding history, labile international normalized ratio, age ≥65 y, and antiplatelet drug or alcohol use.[12,13] Labile international normalized ratio could not be determined from claims data and was excluded from our scoring.

‡Days supply of use overlapped with the date of first prescription for warfarin

Table 2. Outcome Event Counts, Incidence Rates, and Adjusted Hazard Ratios With 95% CIs Comparing Propensity Score–Matched New-User Cohorts of Dabigatran and Warfarin Treated for Nonvalvular Atrial Fibrillation, With Warfarin as the Reference Group

	No. of Events		Incidence Rate per 1000 Person-Years	
	Dabigatran	Warfarin	Dabigatran	Warfarin
Primary outcomes				
Ischemic stroke	205	270	11.3	13.9
Major hemorrhage	777	851	42.7	43.9
Gastrointestinal	623	513	34.2	26.5
Intracranial	60	186	3.3	9.6
Intracerebral	44	142	2.4	7.3
Acute myocardial infarction	285	327	15.7	16.9
Secondary outcomes				
All hospitalized bleeds	1079	1139	59.3	58.8
Mortality*	603	744	32.6	37.8

*For 1064 deaths not preceded by a primary study outcome, the adjusted hazard ratio (95% confidence interval [CI]) was 0.89 (0.79–1.00 P=0.051), whereas for 283 deaths occurring within 30 days after a primary outcome, the adjusted hazard ratio (95% CI) was 0.77 (0.61–0.98 P=0.03).

- Incidence rate during target and comparator cohorts based on observing new events during ‘time-at-risk’ for eight selected outcome cohorts

Table 2. Outcome Event Counts, Incidence Rates, and Adjusted Hazard Ratios With 95% CIs Comparing Propensity Score–Matched New-User Cohorts of Dabigatran and Warfarin Treated for Nonvalvular Atrial Fibrillation, With Warfarin as the Reference Group

	[redacted]	Adjusted Hazard Ratio (95% CI)	*P* Value
Primary outcomes			
Ischemic stroke		0.80 (0.67–0.96)	0.02
Major hemorrhage		0.97 (0.88–1.07)	0.50
Gastrointestinal		1.28 (1.14–1.44)	<0.001
Intracranial		0.34 (0.26–0.46)	<0.001
Intracerebral		0.33 (0.24–0.47)	<0.001
Acute myocardial infarction		0.92 (0.78–1.08)	0.29
Secondary outcomes			
All hospitalized bleeds		1.00 (0.92–1.09)	0.97
Mortality*		0.86 (0.77–0.96)	0.006

*For 1064 deaths not preceded by a primary study outcome, the adjusted hazard ratio (95% confidence interval [CI]) was 0.89 (0.79–1.00; *P*=0.051), whereas for 283 deaths occurring within 30 days after a primary outcome, the adjusted hazard ratio (95% CI) was 0.77 (0.61–0.98; *P*=0.03).

- Population-level effect estimation examining temporal association between target and comparator cohorts and eight selected outcome cohorts

The common building block of all observational analysis: cohorts

Required inputs:

Target cohort:
- Person
- cohort start date
- cohort end date

Comparator cohort:
- Person
- cohort start date
- cohort end date

Outcome cohort:
- Person
- cohort start date
- cohort end date

Desired outputs:

Clinical characterization
- Baseline summary of exposures (treatment utilization)

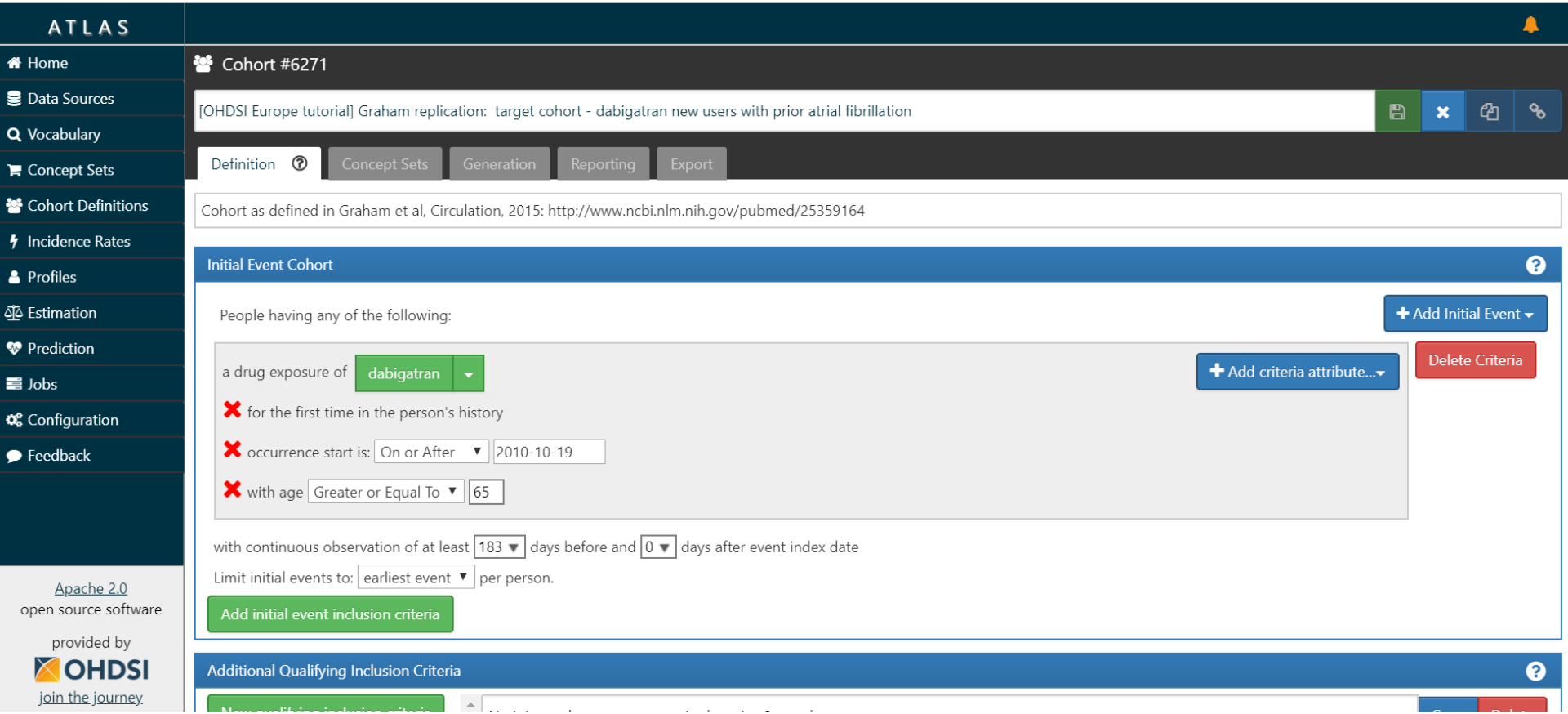
Clinical characterization
- Baseline summary of outcome (disease natural history)

Incidence summary
- Proportion/rate of outcome occurring during time-at-risk for exposure

Population-level effect estimation
- Relative risk (HR, OR, IRR) of outcome occurring during time-at-risk for exposure

Patient-level prediction
- Probability of outcome occurring during time-at-risk for each patient in population

OHDSI in action: Cohort definition



OHDSI in action: Cohort characterization

Inclusion Report | Cohort Features

Feature Report for Truven MDCR

Features are baseline characteristics (e.g collected before /on cohort start)

Demographics | Conditions | Drugs | Procedures | Measurements | Observations | Distributions

Column visibility | Copy | CSV | Show 15 entries

Showing 1 to 15 of 20 entries

Analysis
- Index Month (12)
- Age Group (6)
- Gender (2)

Time Window
- None (20)

Name	Count	% of cohort
age group: 65-69	3,853	19.30
age group: 70-74	4,520	22.60
age group: 75-79	4,551	22.80
age group: 80-84	3,956	19.80
age group: 85-89	2,358	11.80
age group: 90-94	695	3.50
FEMALE	8,781	43.90
MALE	11,249	56.20

OHDSI in action: Cohort characterization in ATLAS

Features are basline characteristics (e.g collected before /on cohort start)

Demographics | Conditions | **Drugs** | Procedures | Measurements | Observations | Distributions

Long Term: 365 day lookback. **Short Term:** 30d lookback. **Overlapping:** Event spans cohort start date.

Column visibility | Copy | CSV | Show 15 ▼ entries | Filter:

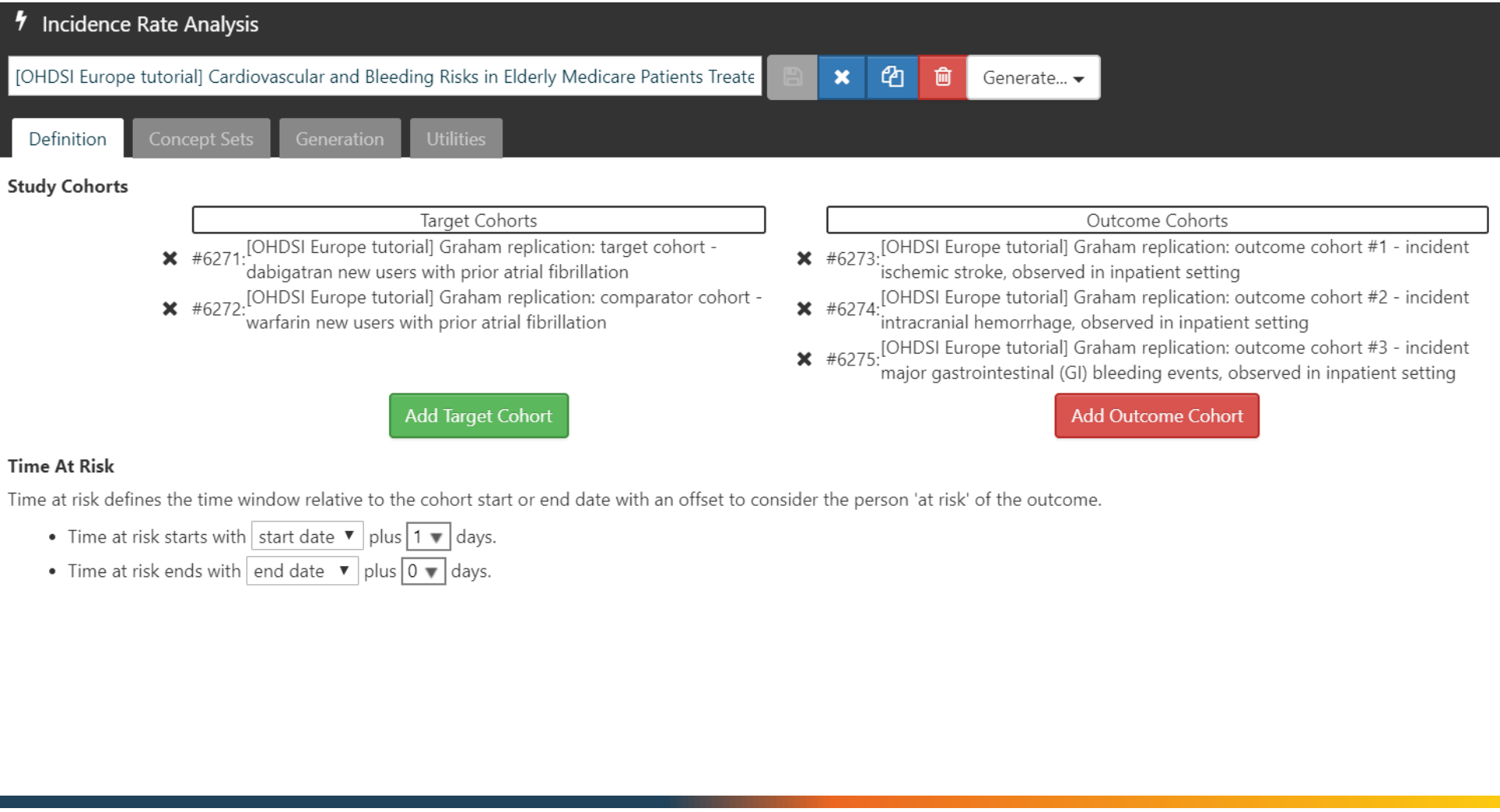
Showing 1 to 15 of 305 entries | Previous 1 2 3 4 5 … 21 Next

Analysis
- Group Era (1025)
- Era (681)

Time Window
- Long Term (708)
- Short Term (537)
- Overlapping (461)

	Concept Name	Time Window	Person Count	% of cohort
Explore	dabigatran etexilate	Long Term	19,975	100.00
Explore	Metoprolol	Long Term	8,820	44.20
Explore	Hydrochlorothiazide	Long Term	5,955	29.90
Explore	Acetaminophen	Long Term	5,739	28.80
Explore	Lisinopril	Long Term	4,935	24.80
Explore	Simvastatin	Long Term	4,851	24.30
Explore	Amlodipine	Long Term	4,808	24.10
Explore	Furosemide	Long Term	4,795	24.10
Explore	Hydrocodone	Long Term	4,590	23.00
Explore	atorvastatin	Long Term	4,422	22.20

OHDSI in action: incidence rate specification



OHDSI in action: incidence rate generation

Incidence Rate Analysis

[OHDSI Europe tutorial] Cardiovascular and Bleeding Risks in Elderly Medicare Patients Treate

Generate...

Definition | Concept Sets | Generation | Utilities

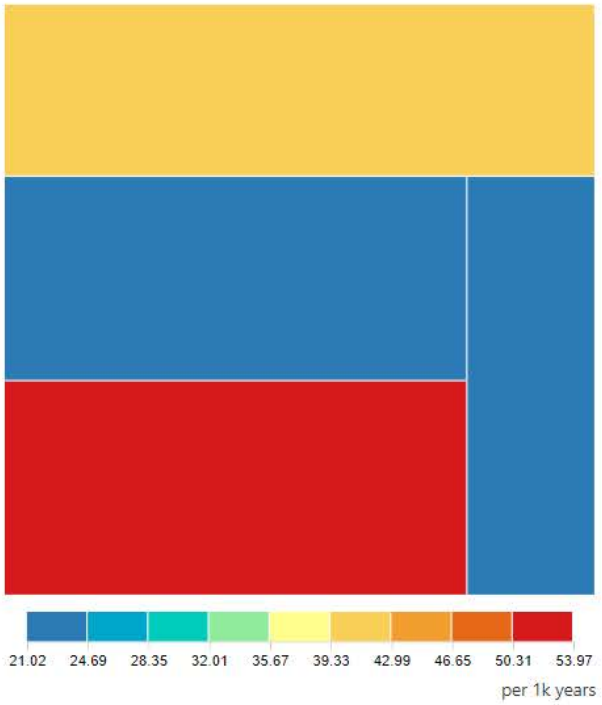
Export Analysis to CSV

Source	Name		Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years	Started	Duration	
TRUVENMDCR_V698	Truven MDCR	Execute	19,288	201	10.42	5,606	35.85	2018-03-21, 15:20	00:01:12	Remove

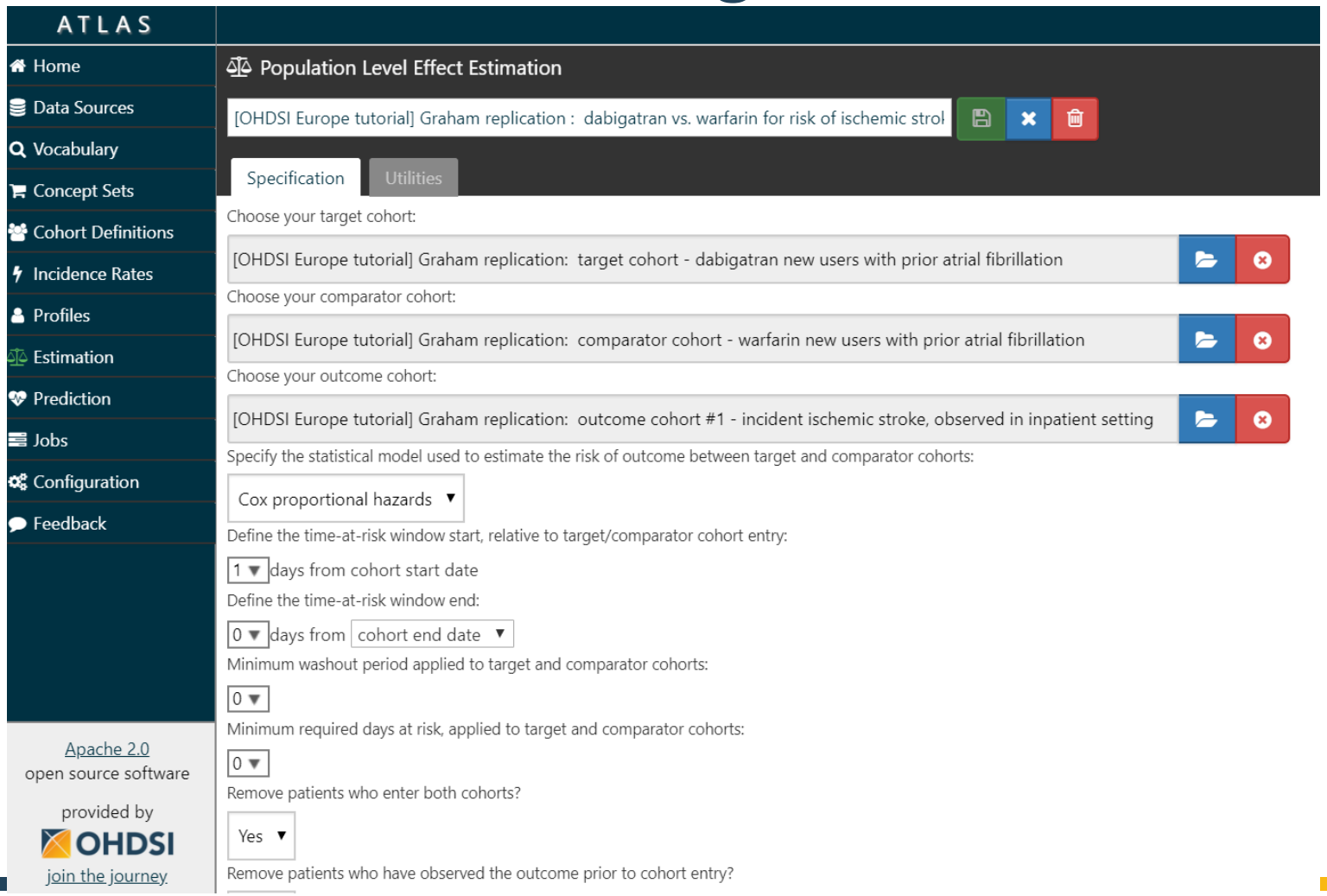
Showing target cohort: [OHDSI Europe tutorial] Graham replicatio and outcome cohort: [OHDSI Europe tutorial] Graham replicatic

	Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
Summary Statistics:	19,288	201	10.42	5,606	35.85

Stratify Rule	N	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
1. Gender = MALE	10,839	99	9.13	3,223	30.72
2. Age >= 75	11,101	150	13.51	3,239	46.31



OHDSI in action: Population-level effect estimation design



OHDSI in action: Population-level effect estimation imeplementation

Model type: cox
Stratified: FALSE
Use covariates: FALSE
Status: OK

	Estimate	lower .95	upper .95
treatment	0.89626	0.71863	1.11829

Population counts

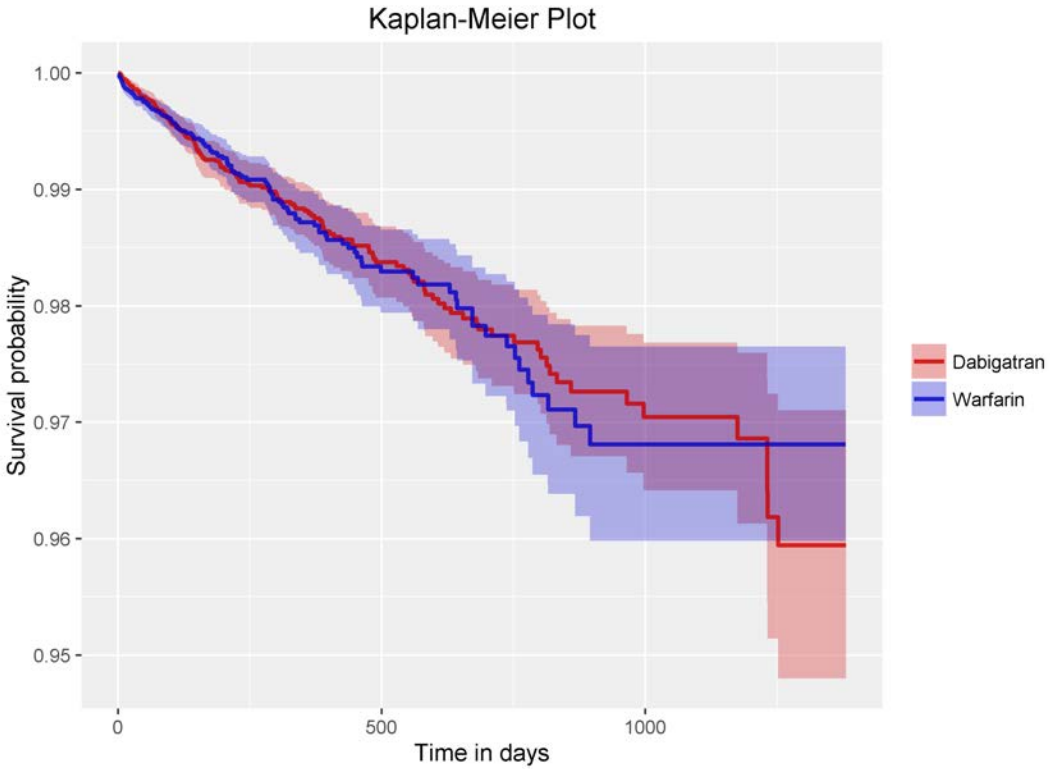
	treatedPersons	comparat
Count	17460	17460

Outcome counts

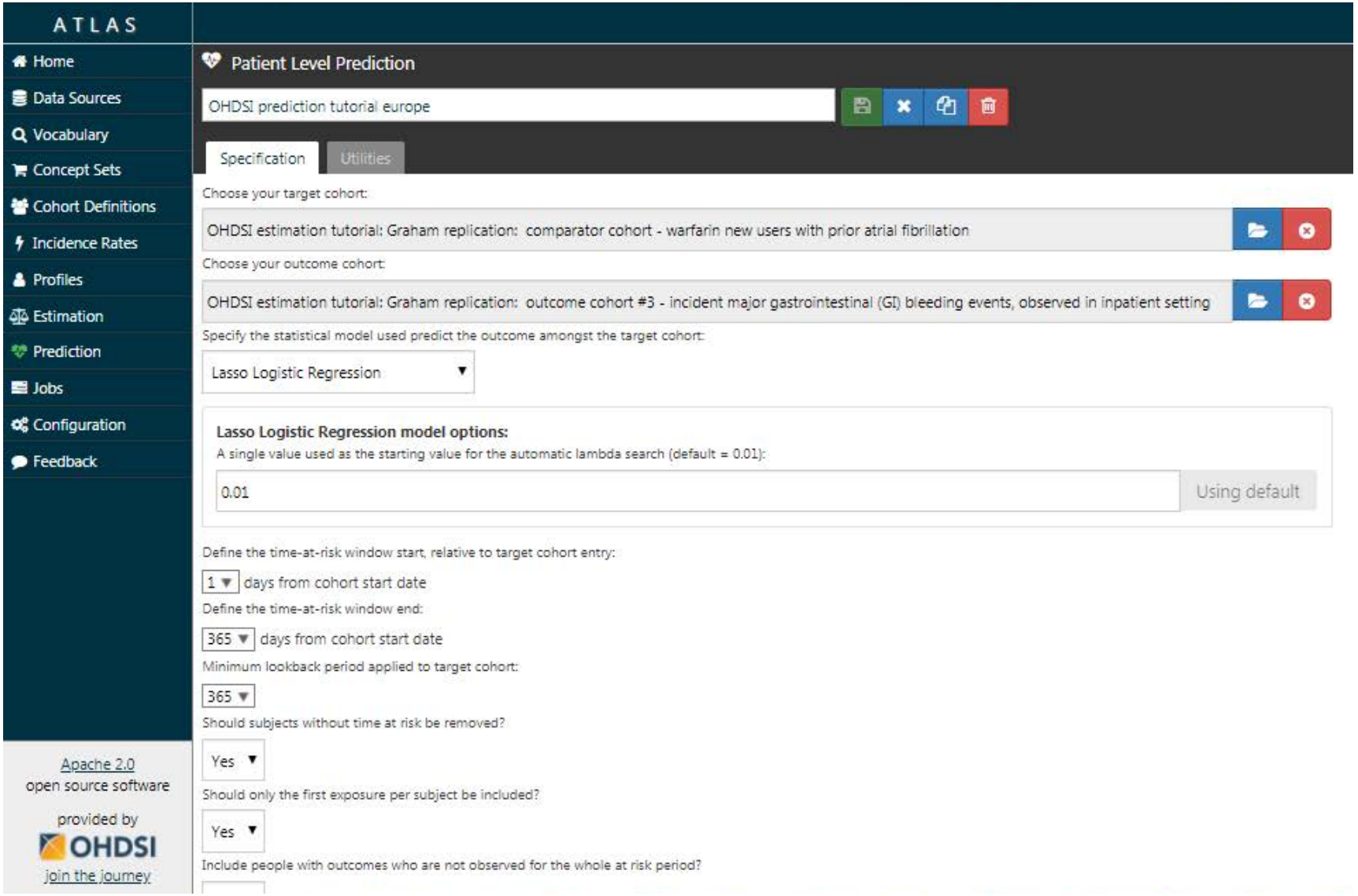
	treatedPersons	comparat
Count	164	155

Time at risk

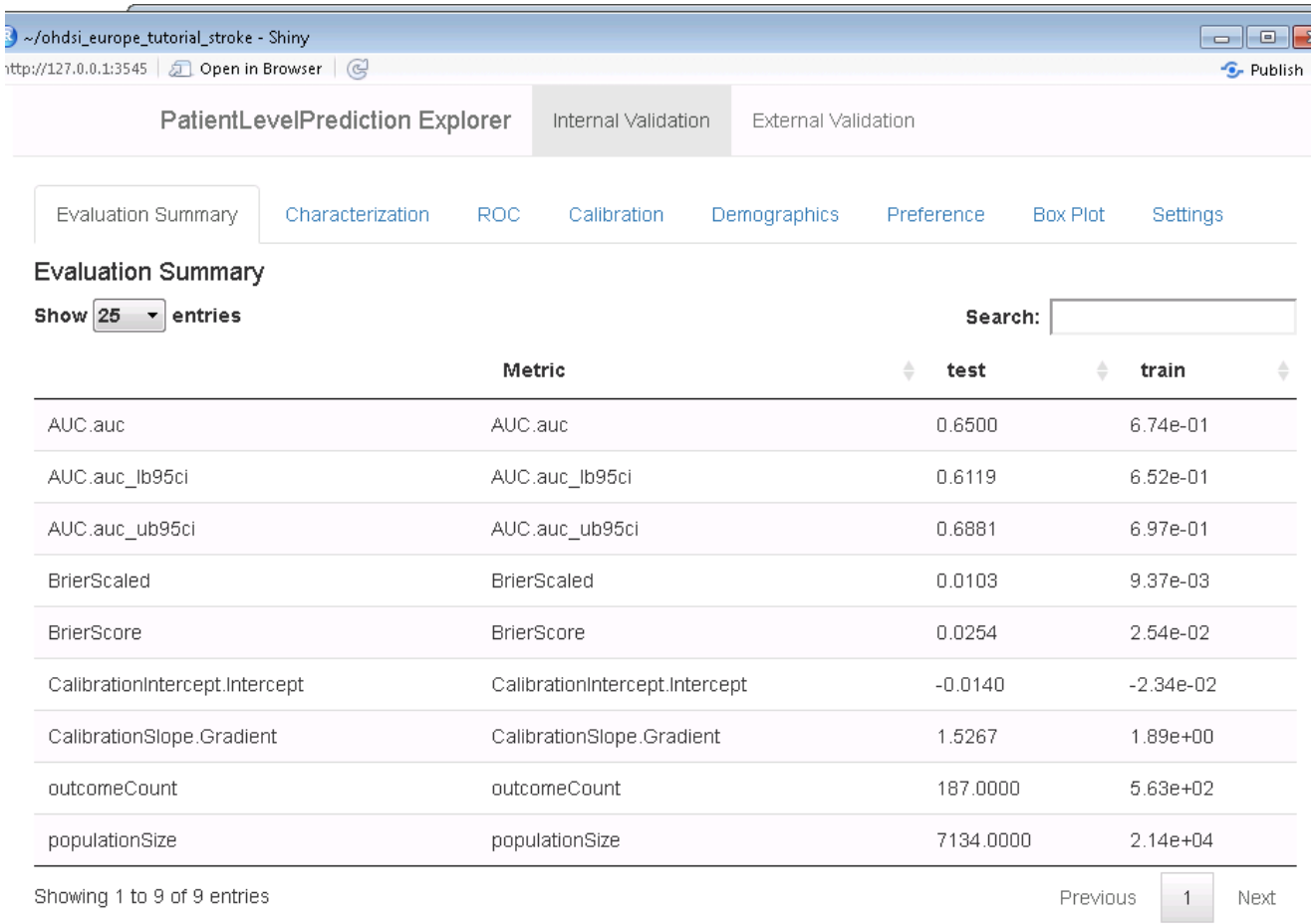
	treatedDays	comparat
Days	4912947	3954046



OHDSI in action:
Patient-level prediction design



OHDSI in action: Patient-level prediction implementation




OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Vocabulary

Everything is a concept….everything needs to be defined in a common language

Cardiovascular, Bleeding, and Mortality Risks in Elderly Medicare Patients Treated With Dabigatran or Warfarin for Nonvalvular Atrial Fibrillation

David J. Graham, MD, MPH; Marsha E. Reichman, PhD; Michael Wernecke, BA; Rongmei Zhang, PhD; Mary Ross Southworth, PharmD; Mark Levenson, PhD; Ting-Chang Sheu, MPH; Katrina Mott, MHS; Margie R. Goulding, PhD; Monika Houstoun, PharmD, MPH; Thomas E. MaCurdy, PhD; Chris Worrall, BS; Jeffrey A. Kelman, MD, MMSc

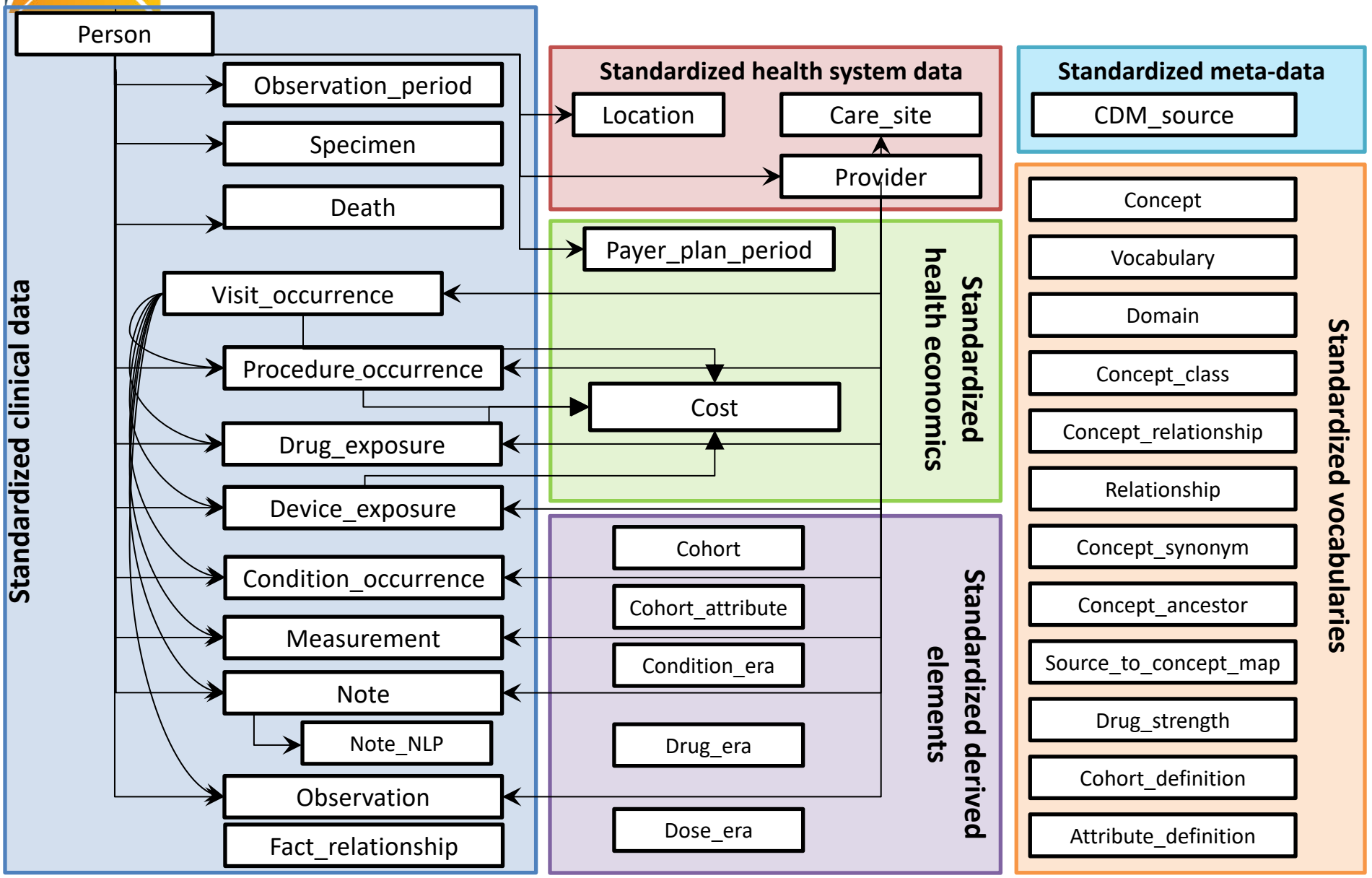
Background—The comparative safety of dabigatran versus warfarin for treatment of nonvalvular atrial fibrillation in general practice settings has not been established.

Methods and Results—We formed new-user cohorts of propensity score–matched elderly patients enrolled in Medicare who initiated dabigatran or warfarin for treatment of nonvalvular atrial fibrillation between October 2010 and December 2012. Among 134 414 patients with 37 587 person-years of follow-up, there were 2715 primary outcome events. The hazard ratios (95% confidence intervals) comparing dabigatran with warfarin (reference) were as follows: ischemic stroke, 0.80 (0.67–0.96); intracranial hemorrhage, 0.34 (0.26–0.46); major gastrointestinal bleeding, 1.28 (1.14–1.44); acute myocardial infarction, 0.92 (0.78–1.08); and death, 0.86 (0.77–0.96). In the subgroup treated with dabigatran 75 mg twice daily, there was no difference in risk compared with warfarin for any outcome except intracranial hemorrhage, in which case dabigatran risk was reduced. Most patients treated with dabigatran 75 mg twice daily appeared not to have severe renal impairment, the intended population for this dose. In the dabigatran 150-mg twice daily subgroup, the magnitude of effect for each outcome was greater than in the combined-dose analysis.

Conclusions—In general practice settings, dabigatran was associated with reduced risk of ischemic stroke, intracranial hemorrhage, and death and increased risk of major gastrointestinal hemorrhage compared with warfarin in elderly patients with nonvalvular atrial fibrillation. These associations were most pronounced in patients treated with dabigatran 150 mg twice daily, whereas the association of 75 mg twice daily with study outcomes was indistinguishable from warfarin except for a lower risk of intracranial hemorrhage with dabigatran. **(*Circulation*. 2015;131:157-164. DOI: 10.1161/CIRCULATIONAHA.114.012061.)**

Key Words: anticoagulant ■ pharmacoepidemiology ■ safety ■ thrombin inhibitor ■ warfarin

OMOP CDM


Standardized clinical data
Person
Observation_period
Specimen
Death
Visit_occurrence
Procedure_occurrence
Drug_exposure
Device_exposure
Condition_occurrence
Measurement
Note
Note_NLP
Observation
Fact_relationship
Standardized health system data
Location
Care_site
Provider
Payer_plan_period
Standardized health economics
Cost
Standardized derived elements
Cohort
Cohort_attribute
Condition_era
Drug_era
Dose_era
Standardized meta-data
CDM_source
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength
Cohort_definition
Attribute_definition

OMOP Common Vocabulary Model

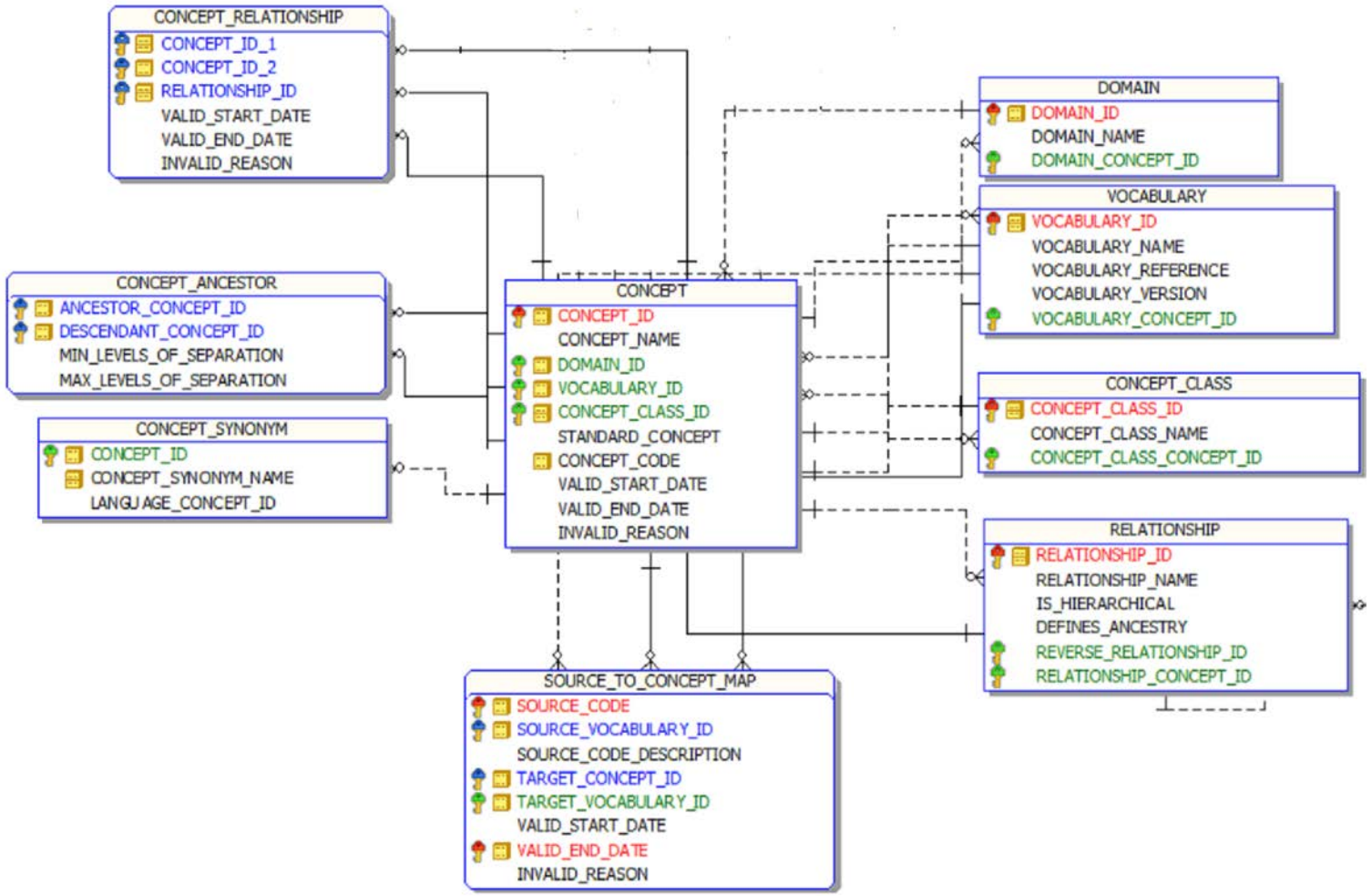
What it is

- **Standardized structure** to house existing vocabularies used in the public domain
- **Compiled standards** from disparate public and private sources and some OMOP-grown concepts

What it's not

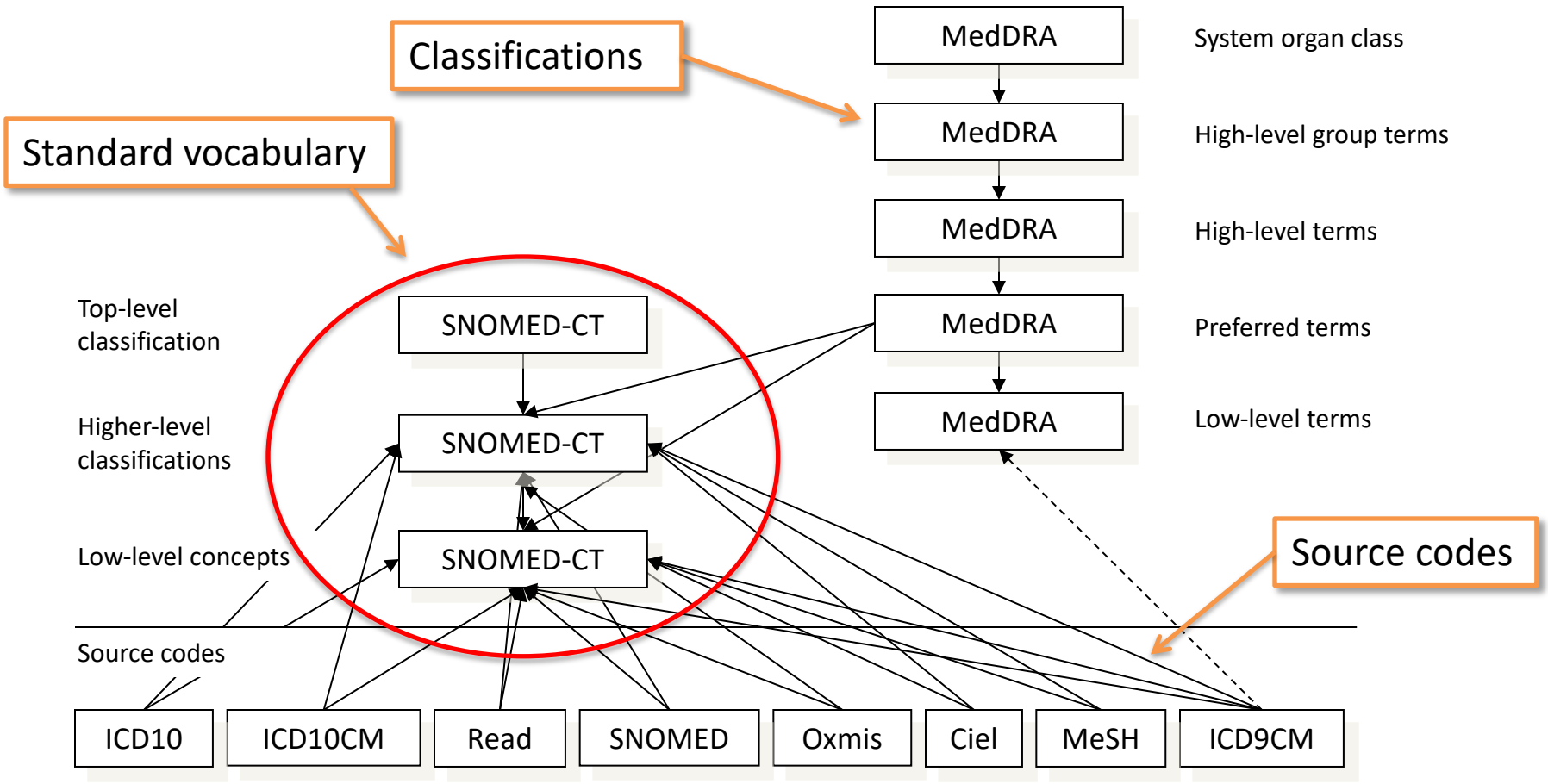
- **Static dataset** – the vocabulary updates regularly to keep up with the continual evolution of the sources
- **Finished product** – vocabulary maintenance and improvement is ongoing activity that requires community participation and support

OMOP Vocabulary CDM

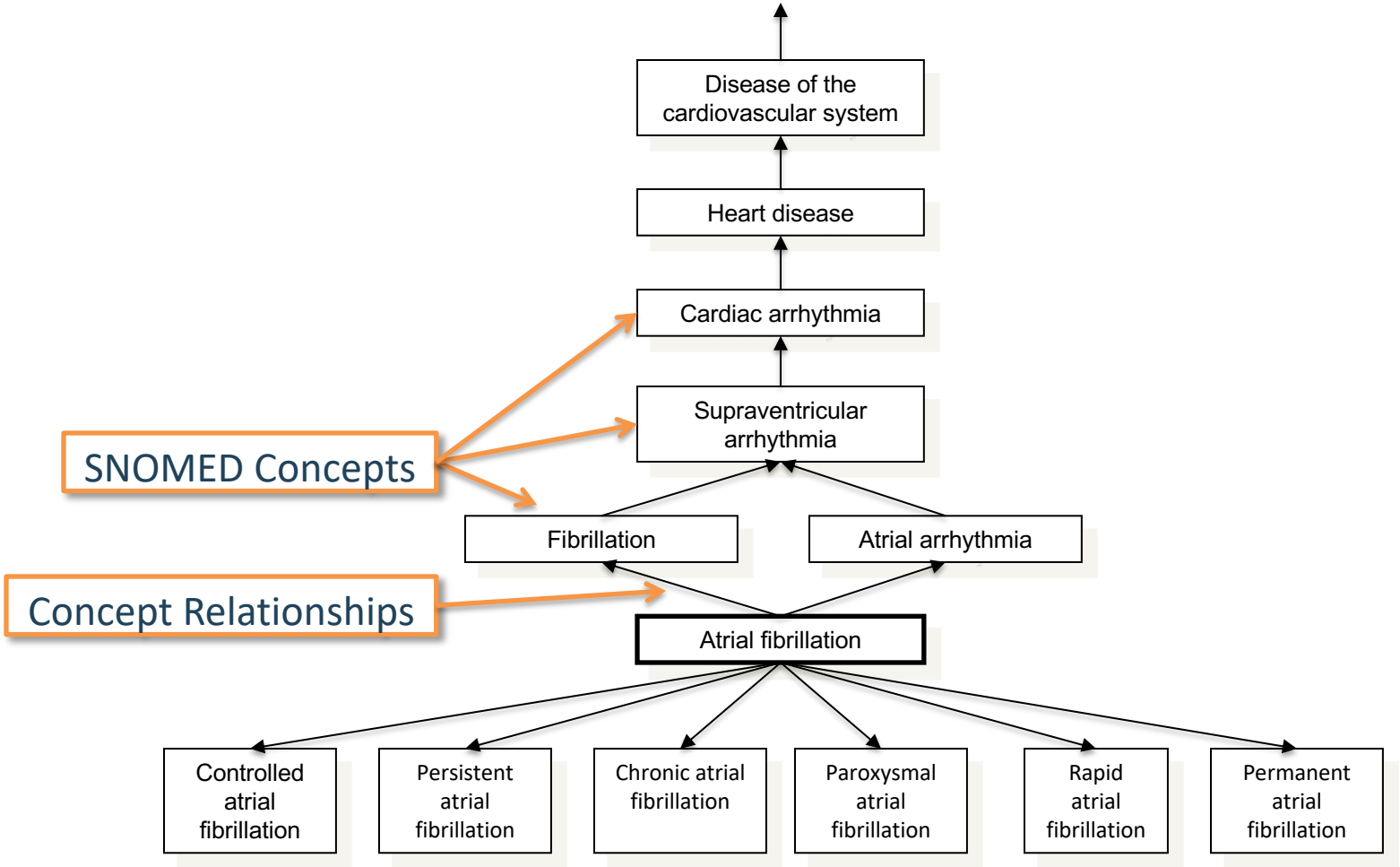


1. All content: concepts in **`concept`** table
2. Direct relationships between concepts listed in **`concept_relationship`**
3. Multi-step hierarchical relationships pre-processed in **`concept_ancestor`**

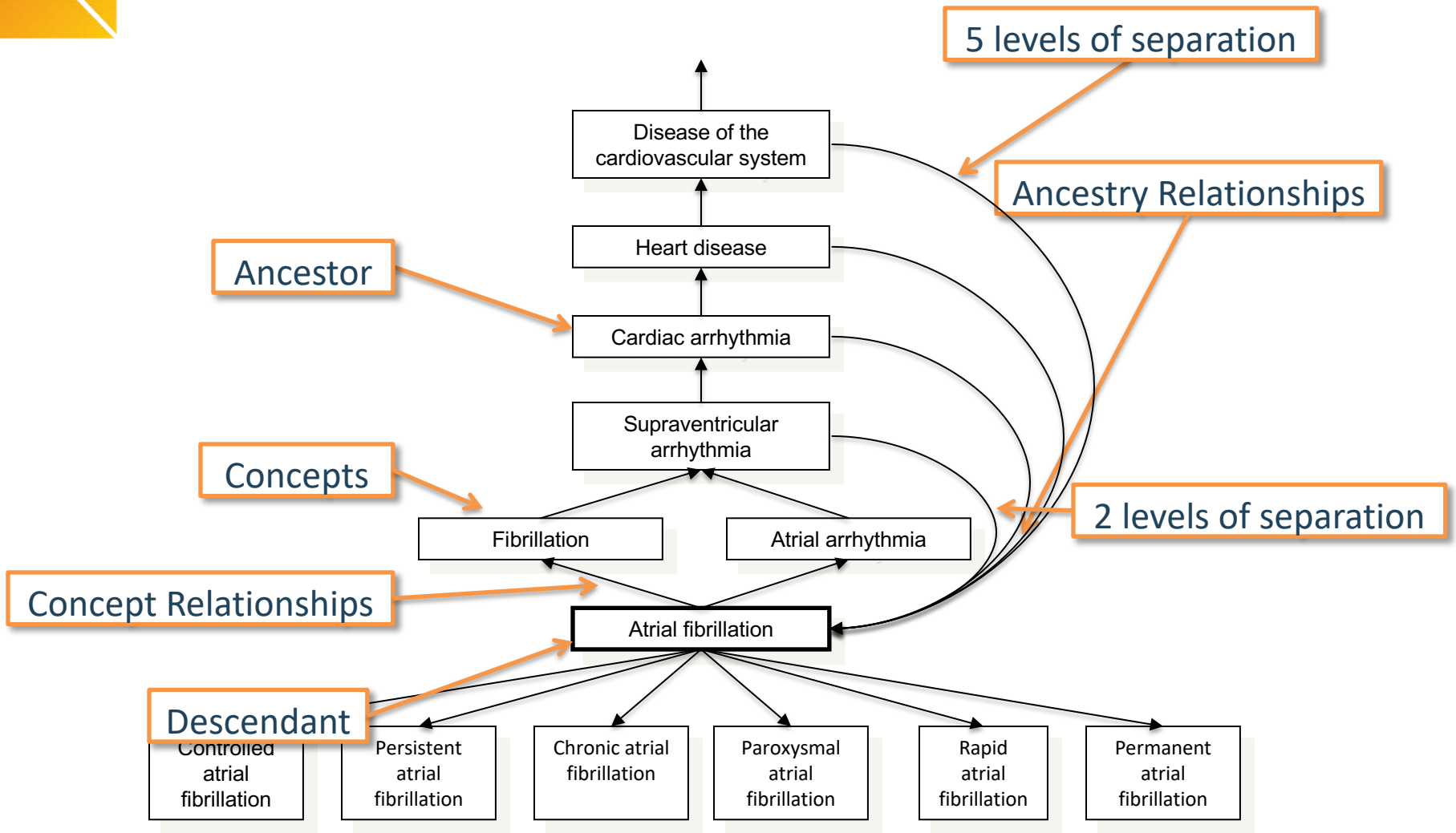
Condition Concepts



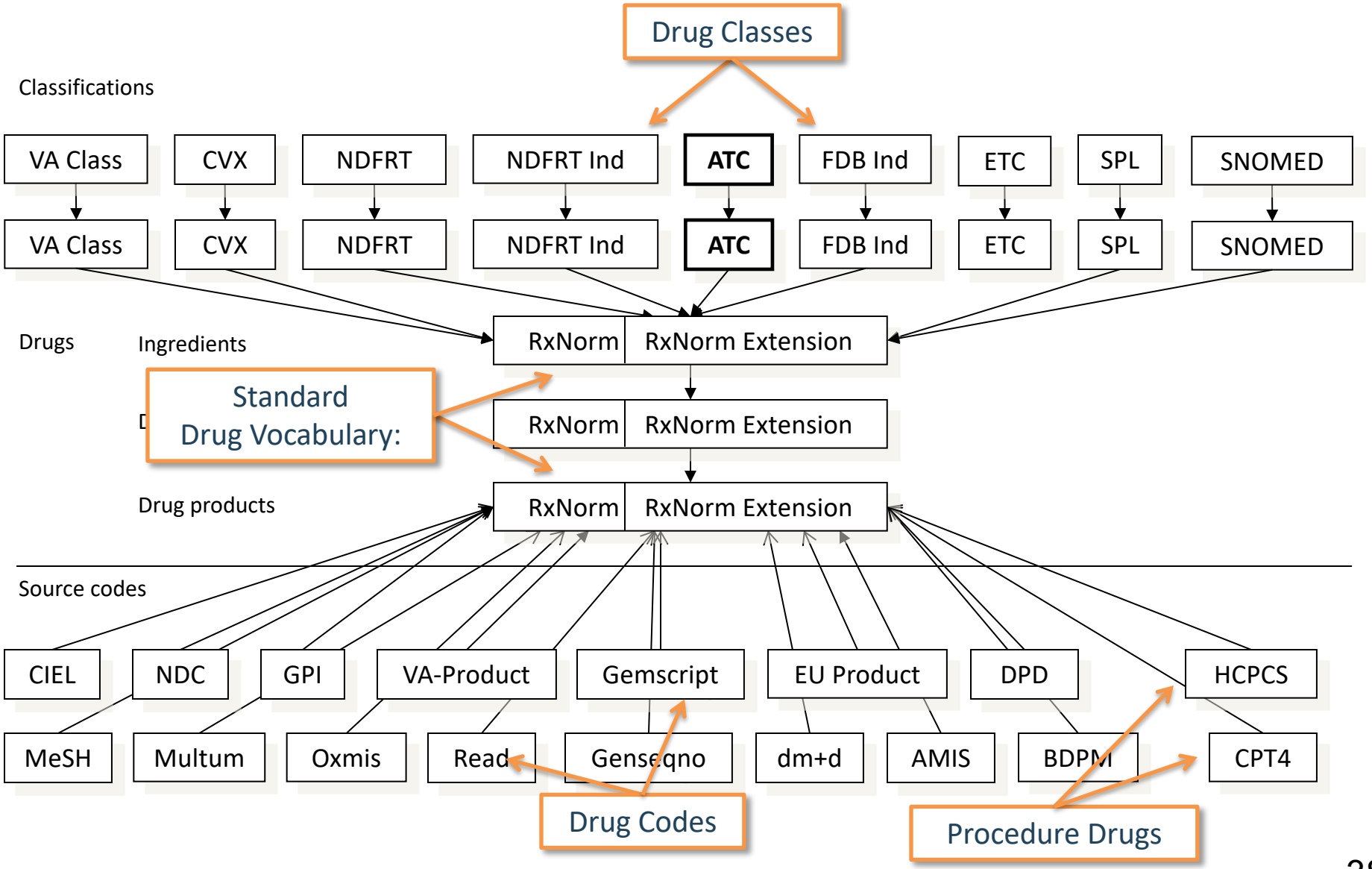
Condition ancestry around "atrial fibrillation"



Ancestry Relationships: Higher-Level Relationships



Drug Hierarchy



Vocabulary classifications improve your efficiency….and your quality!

Health Serv Outcomes Res Method (2013) 13:58–67
DOI 10.1007/s10742-012-0102-1

Applying standardized drug terminologies to observational healthcare databases: a case study on opioid exposure

Frank J. DeFalco · Patrick B. Ryan · M. Soledad Cepeda

- 60% of medication codes and 94% of records are mapped
- 45% of opiate codes covered by one of ATC, ETC, and NDF-RT are covered by all three
  - 15% missed by at least one
- No one classification scheme was better than the others

Vocabulary classifications improve your efficiency….and your quality!

DeFalco HSORM 2013

Table 3 Identification of related 11 digit NDC codes by drug class and vocabulary

Drug class	Vocabulary	System grouping	Ingredients	Clinical drugs	NDC codes	Unique codes
Opioid	ATC	Opioids	23	1,122	11,765	2
Opioid	ETC.	Analgesics–narcotic	20	1,808	19,106	333
Opioid	NDFRT	Opioid agonists	22	1,813	15,912	1,087
Opioid	VA	Opioid analgesics	24	1,750	17,113	450
NSAID	ATC	Antiinflam and antirheumatic products, non-steroids	52			
NSAID	ETC.	NSAID analgesics	23			
NSAID	NDFRT	NSAID analgesics	23			
NSAID	VA	Nonsalicylate NSAIDs, antirheumatic	24			
Antidiabetic	ATC	Drugs used in diabetes	53			
Antidiabetic	ETC.	Oral antidiabetic agents	19			
Antidiabetic	NDFRT	Insulin receptor agonists	42			
Antidiabetic	VA	Oral hypoglycemic agents	18			
Antidepressant	ATC	Antidepressants	47			
Antidepressant	ETC.	Antidepressants	29			
Antidepressant	NDFRT	Serotonin uptake inhibitors, norepinephrine uptake inhibitors, dopamine uptake inhibitors	40			
Antidepressant	VA	Antidepressants	29			

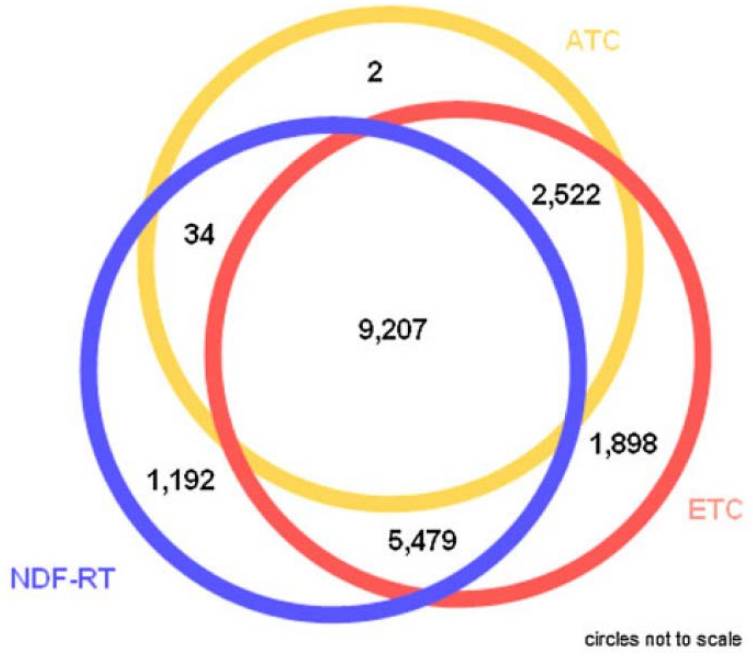


Fig. 1 Overlap in coverage of 'opioid' NDC drug codes by classification system

If we try to speak the same language, will there be loss in translation?

Journal of Biomedical Informatics 45 (2012) 689–696

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases ☆

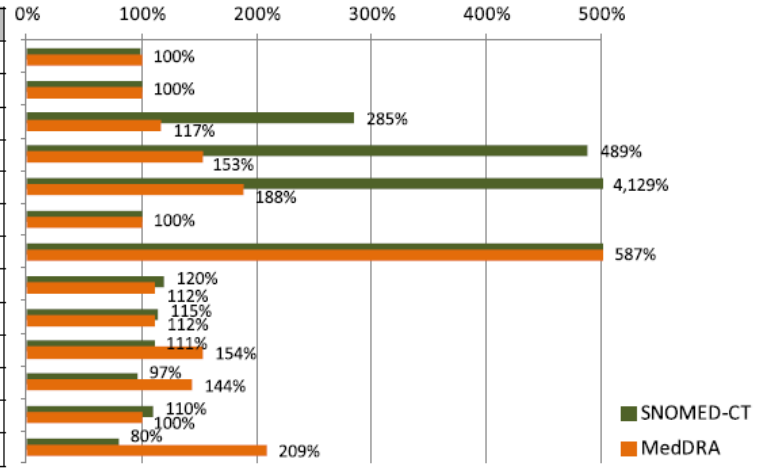
Christian Reich [a,*], Patrick B. Ryan [a,b,1], Paul E. Stang [a,b,1], Mitra Rocca [c,2]

[a] Observational Medical Outcomes Partnership, Foundation for the National Institutes of Health, 9650 Rockville Pike, Bethesda, MD 20814, USA
[b] Janssen Research & Development, LLC, 1125 Trenton-Harbourton Road, PO Box 200, MS K304, Titusville, NJ 08560, USA
[c] Office of Translational Sciences, Center for Drug Evaluation and Research (CDER), US Food and Drug Administration, 10903 New Hampshire Ave., Bldg. 21, Rm. 4608, Silver Spring, MD 20933, USA

Changing language may change your codelist, that may change your cohort depending on the disease…

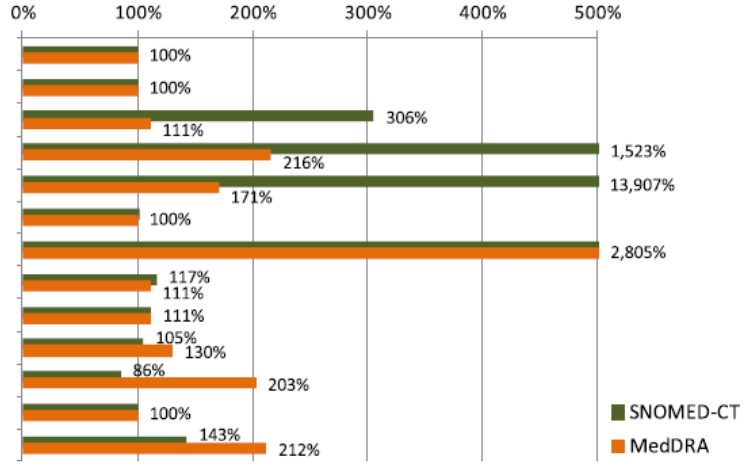
Cohort size of HOI in MSLR for different terminologies

Condition	#
Angioedema	1
Aplastic Anemia	1
Acute Liver Failure	1
	2
	7
Acute Renal Failure	1
	4
Bleeding	1
	3
Hip Fracture	1
Acute myocardial Infarction	1
	2
Upper GI Ulcer Hospitalization	1



Cohort size of HOI in GE for different terminologies

Condition	#
Angioedema	1
Aplastic Anemia	1
Acute Liver Failure	1
	2
	7
Acute Renal Failure	1
	4
Bleeding	1
	3
Hip Fracture	1
Acute myocardial Infarction	1
	2
Upper GI Ulcer Hospitalization	1



…but in practice, running an estimation analysis using source vs. standard vocabulary yields similar results

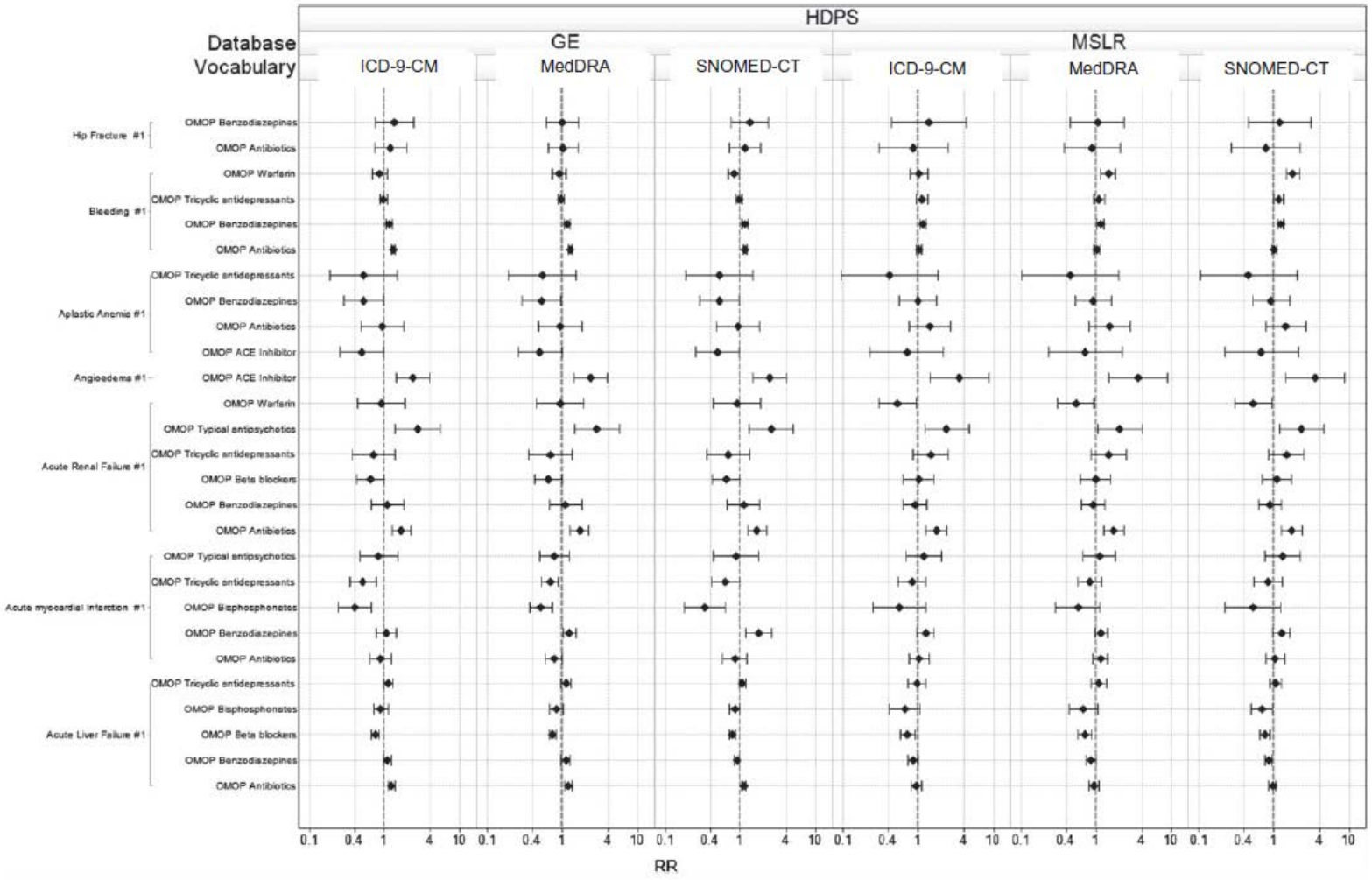
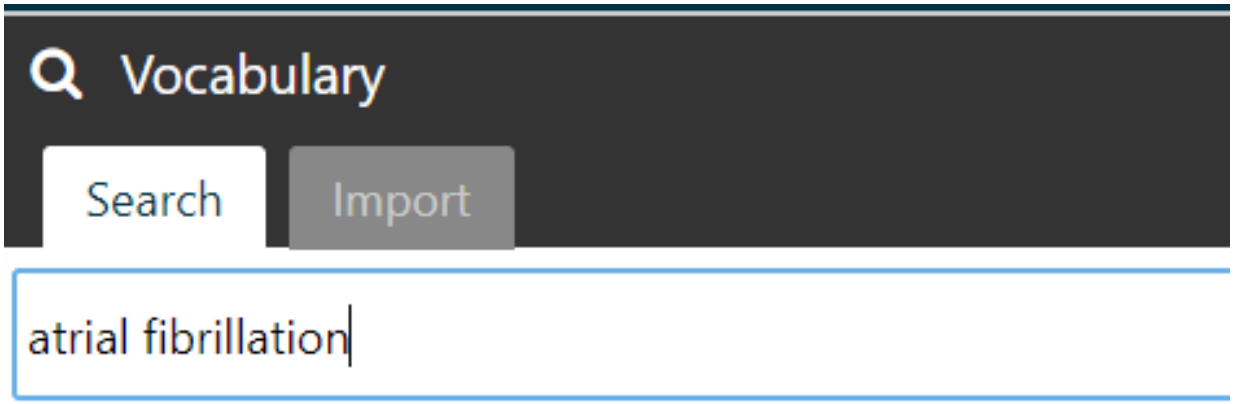


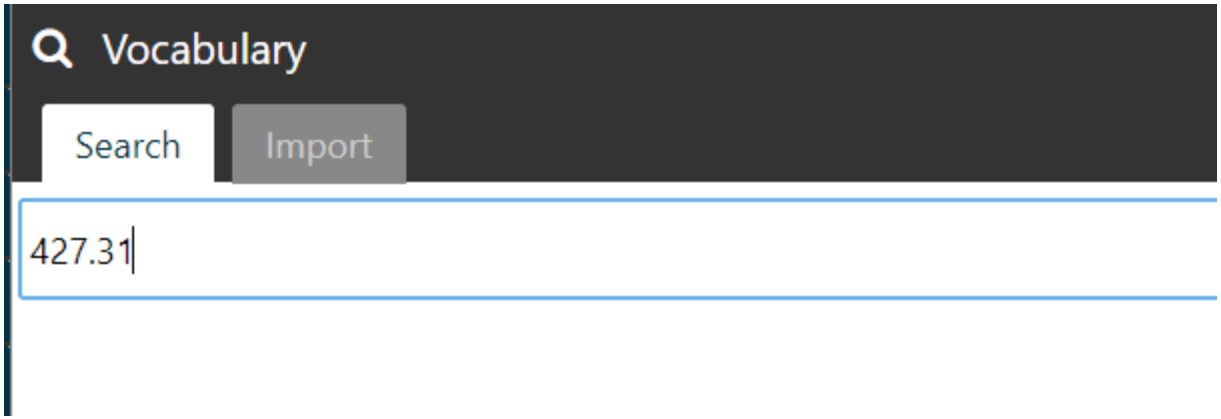
Fig. 3. Effect estimates and 95% confidence intervals for incident user design applied to MSLR and GE using ICD-9-CM, SNOMED-CT, and MedDRA as standard terminologies. Each dot represents the estimate of the effect of an individual HOI-drug combination (on the X-axis).

Demo: Searching the vocabulary in ATLAS

Follow along at:
http://ohdsi.org/web/ATLAS

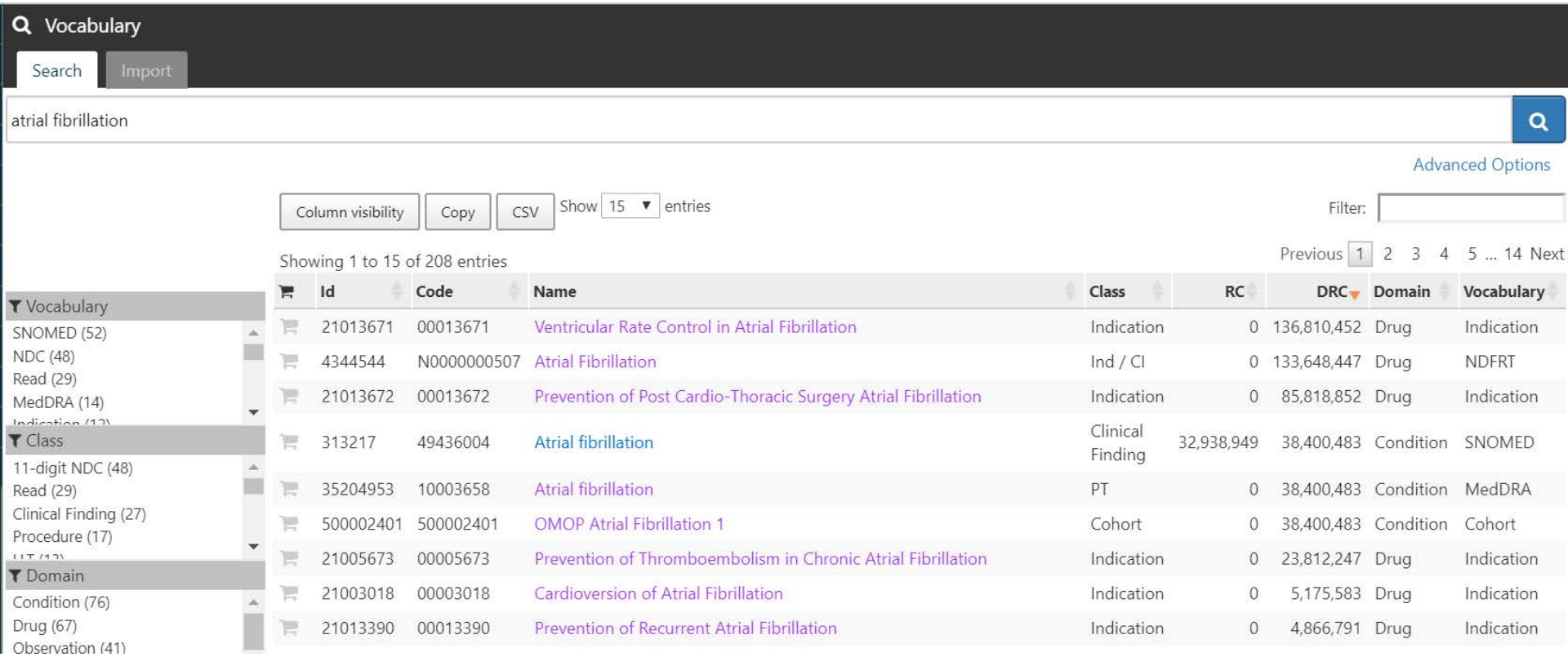
Demo: vocabulary search



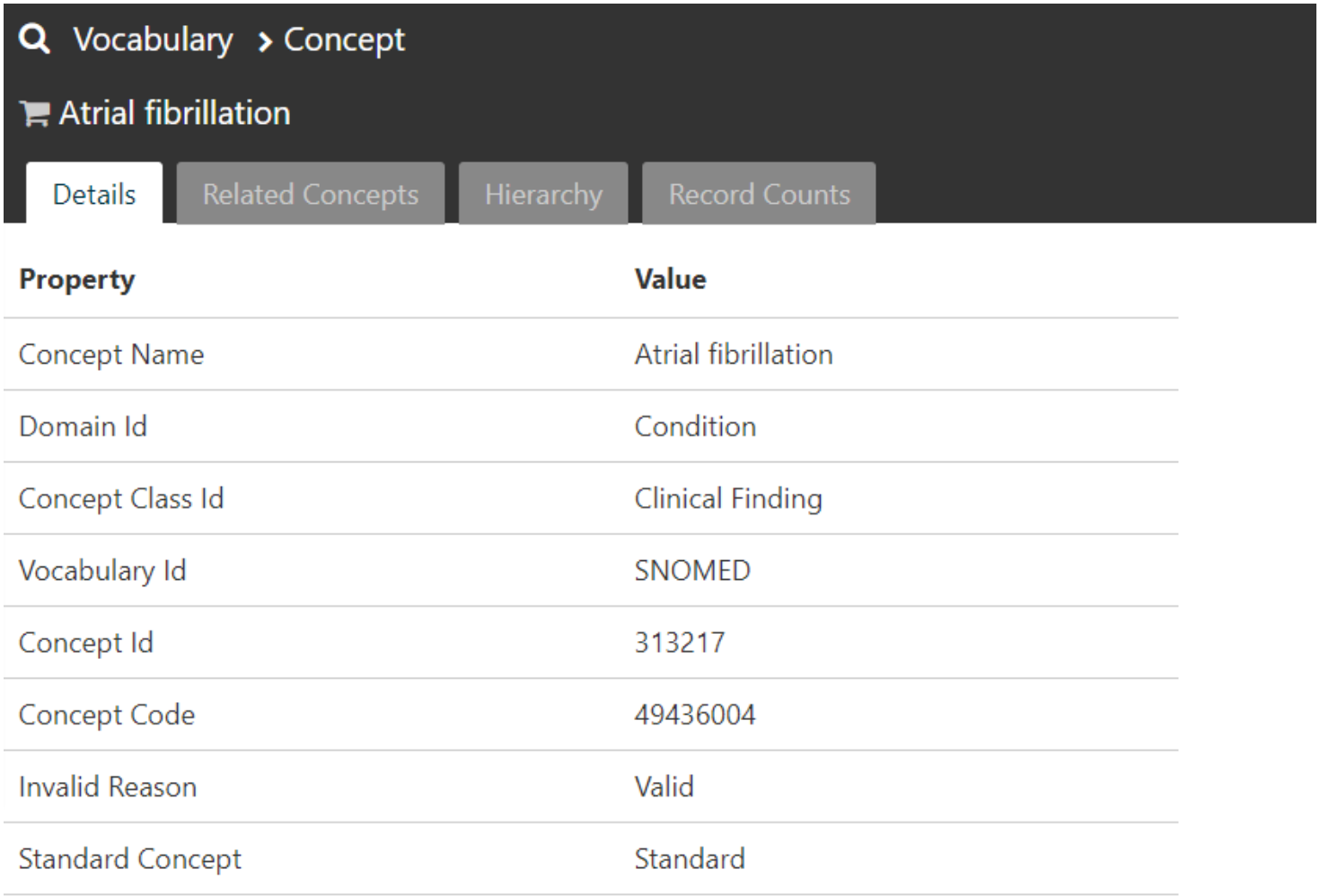


Search for concept names, concept IDs, or source codes across any domain in the OMOP CDM

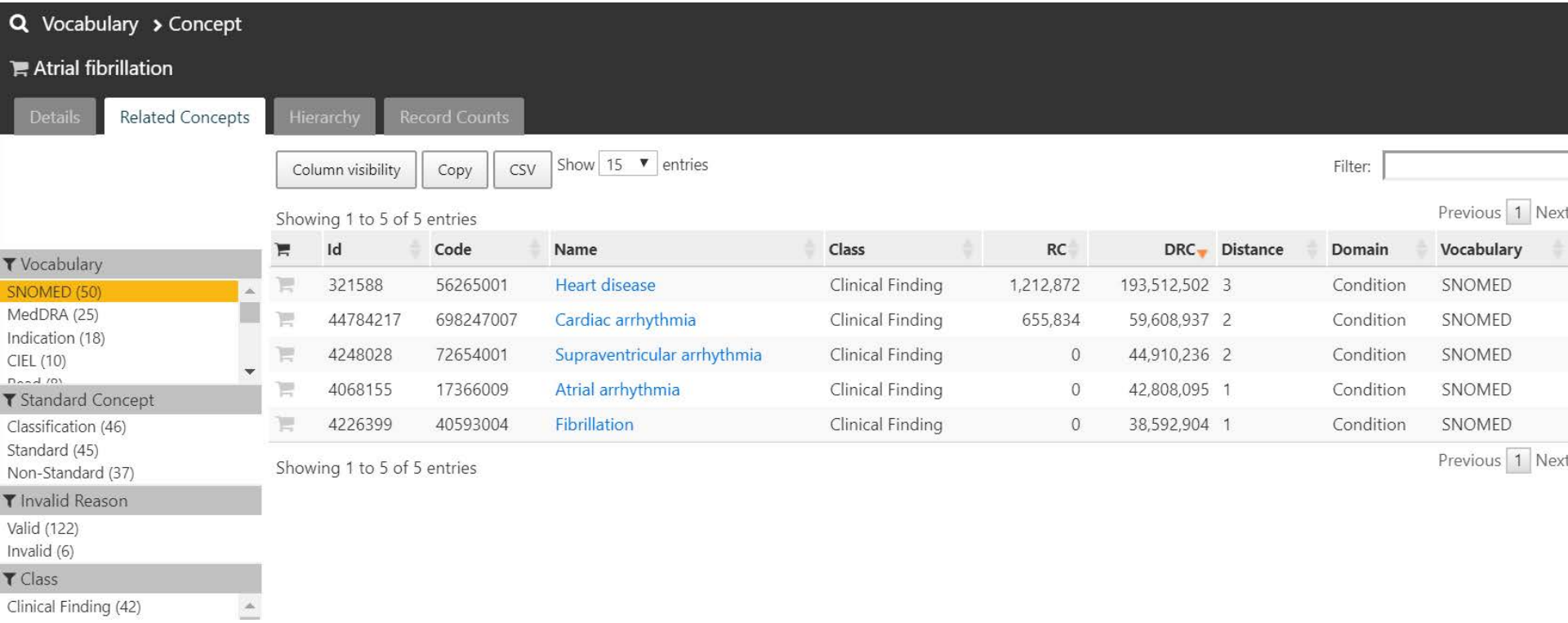
Demo: vocabulary search results



Demo: vocabulary concept selection



Demo: Concept relationship exploration



Demo: Concept hierarchy

Vocabulary › Concept

Atrial fibrillation

Details | Related Concepts | Hierarchy | Record Counts

Parents

Id	Code	Name	Class	RC	DRC	Distance	Domain	Vocabulary
500001801	500001801	OMOP Qt Prolongation/Torsade De Pointes 1	Cohort	0	57,981,637	1	Condition	Cohort
35202455	10042600	Supraventricular arrhythmias	HLT	0	44,910,236	1	Condition	MedDRA
4068155	17366009	Atrial arrhythmia	Clinical Finding	0	42,808,095	1	Condition	SNOMED
4226399	40593004	Fibrillation	Clinical Finding	0	38,592,904	1	Condition	SNOMED
35204969	10061592	Cardiac fibrillation	PT	0	38,592,904	1	Condition	MedDRA
35204953	10003658	Atrial fibrillation	PT	0	38,400,483	1	Condition	MedDRA
500002401	500002401	OMOP Atrial Fibrillation 1	Cohort	0	38,400,483	1	Condition	Cohort

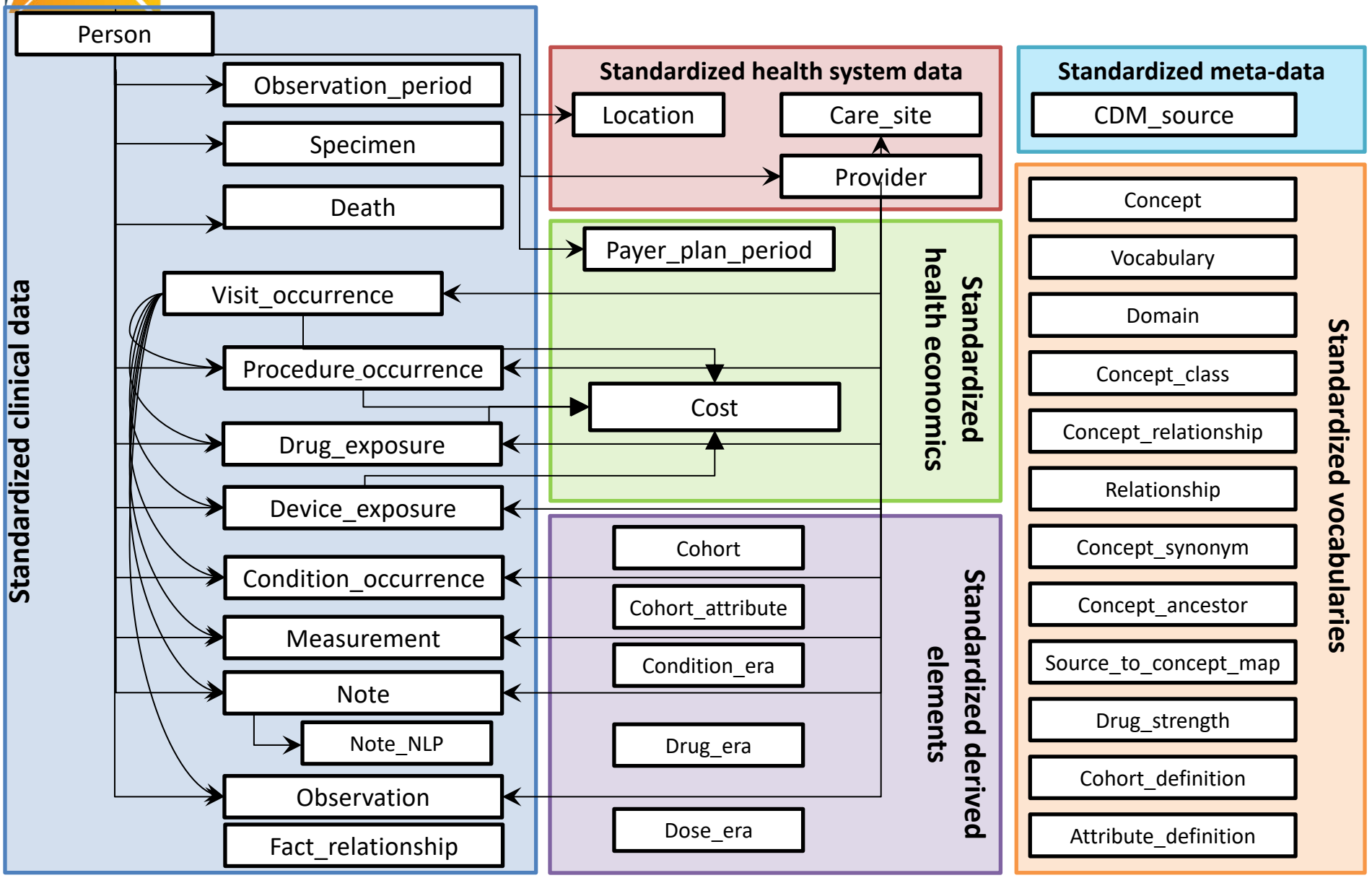
Id	Code	Name	Class	RC	DRC	Distance	Domain	Vocabulary
313217	49436004	Atrial fibrillation	Clinical Finding	32,938,949	38,400,483	1	Condition	SNOMED

Children

Id	Code	Name	Class	RC	DRC	Distance	Domain	Vocabulary
4154290	282825002	Paroxysmal atrial fibrillation	Clinical Finding	2,919,750	2,919,750	1	Condition	SNOMED

Data Sources

CDM Version 5 Key Domains



Purpose of Achilles

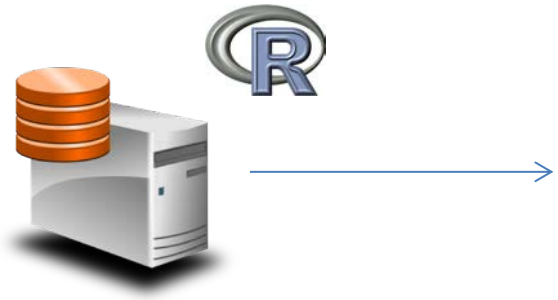
- ACHILLES is a platform which enables the characterization, quality assessment and visualization of observational health databases.

- ACHILLES provides users with an interactive, exploratory framework to assess patient demographics, the prevalence of conditions, drugs and procedures, and to evaluate the distribution of values for clinical observations.

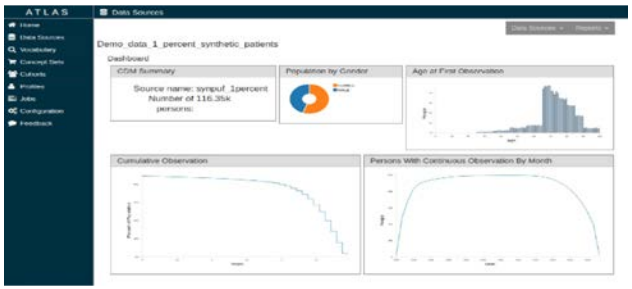
Achilles – Under the Hood

- Two step process
  - Step 1: R Routine running against the local CDM instance. This R routine calculates summary statistics which allow to describe the distribution of patient-level data as well as a generic view on the quality of the data. The output of this step is summarized (and hence de-identfied) set of data stored in results tables in the CDM.

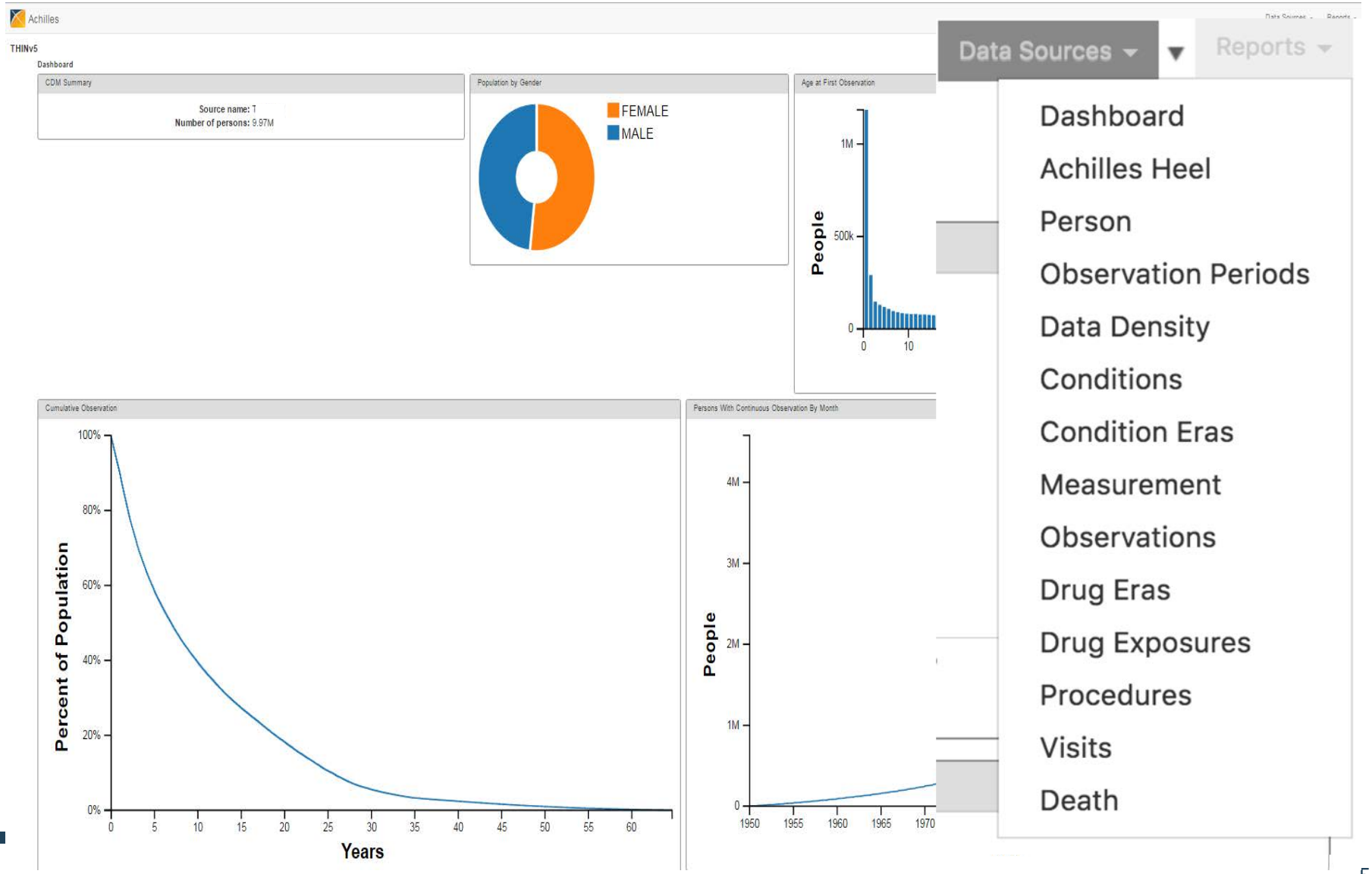
  - Step 2: webapplication which can run standalone from a CDM instance. It requires the ACHILLES R results generated in step 1 as input. The webapplication allows interactive exploration for each of the entities (tables) in the OMOP scheme individually (not possible to query across multiple entities at the same time)



OMOP CDM



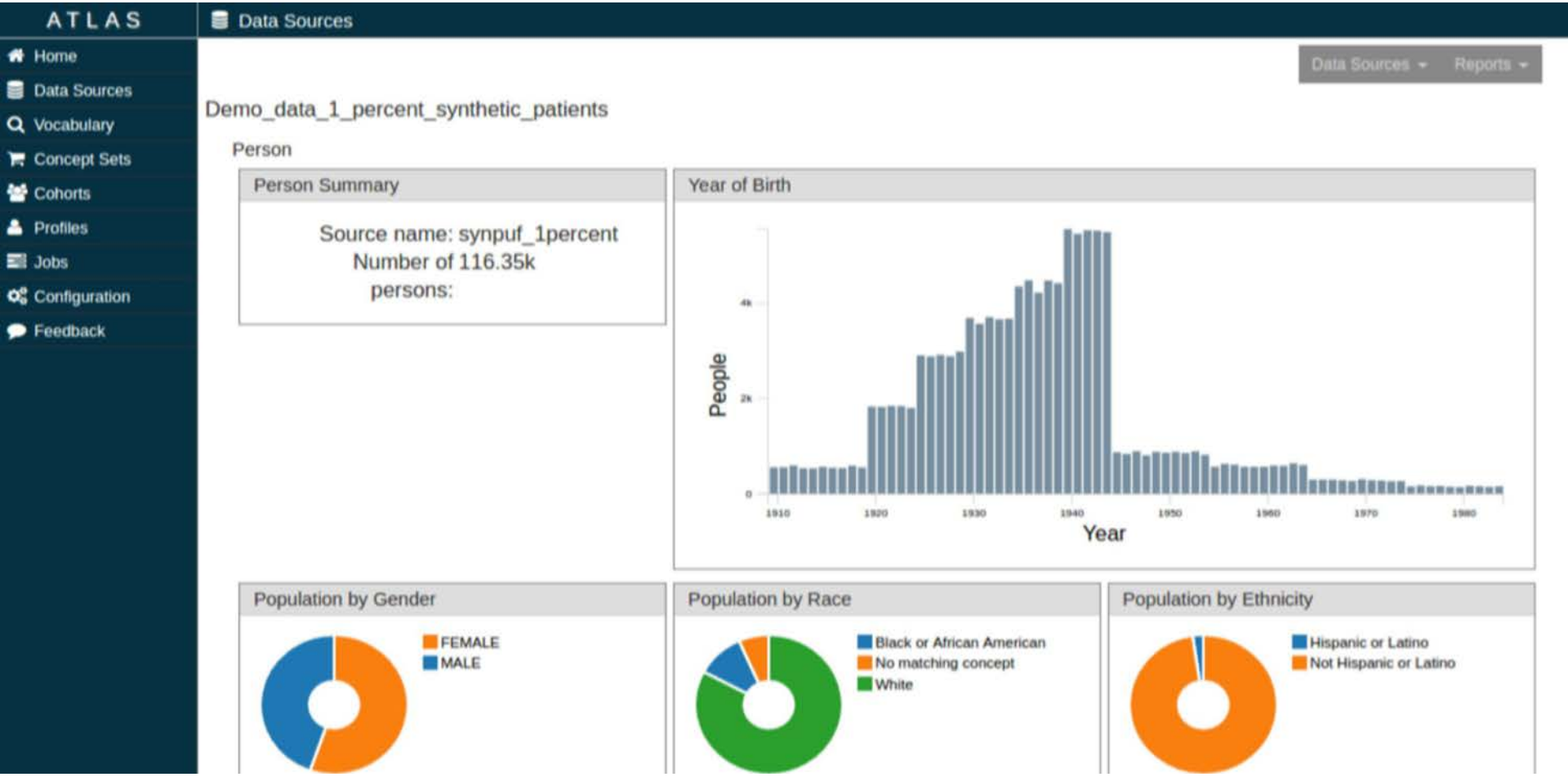
Basic Navigation



Examples


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Dashboard
CDM Summary
Source name: synpuf_1percent
Number of 116.35k
persons:
Population by Gender
FEMALE
MALE
Age at First Observation
People
Age
Cumulative Observation
Percent of Population
Years
Persons With Continuous Observation By Month
People
Date

Examples


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Person
Person Summary
Source name: synpuf_1percent
Number of 116.35k
persons:
Year of Birth
People
Year
Population by Gender
FEMALE
MALE
Population by Race
Black or African American
No matching concept
White
Population by Ethnicity
Hispanic or Latino
Not Hispanic or Latino

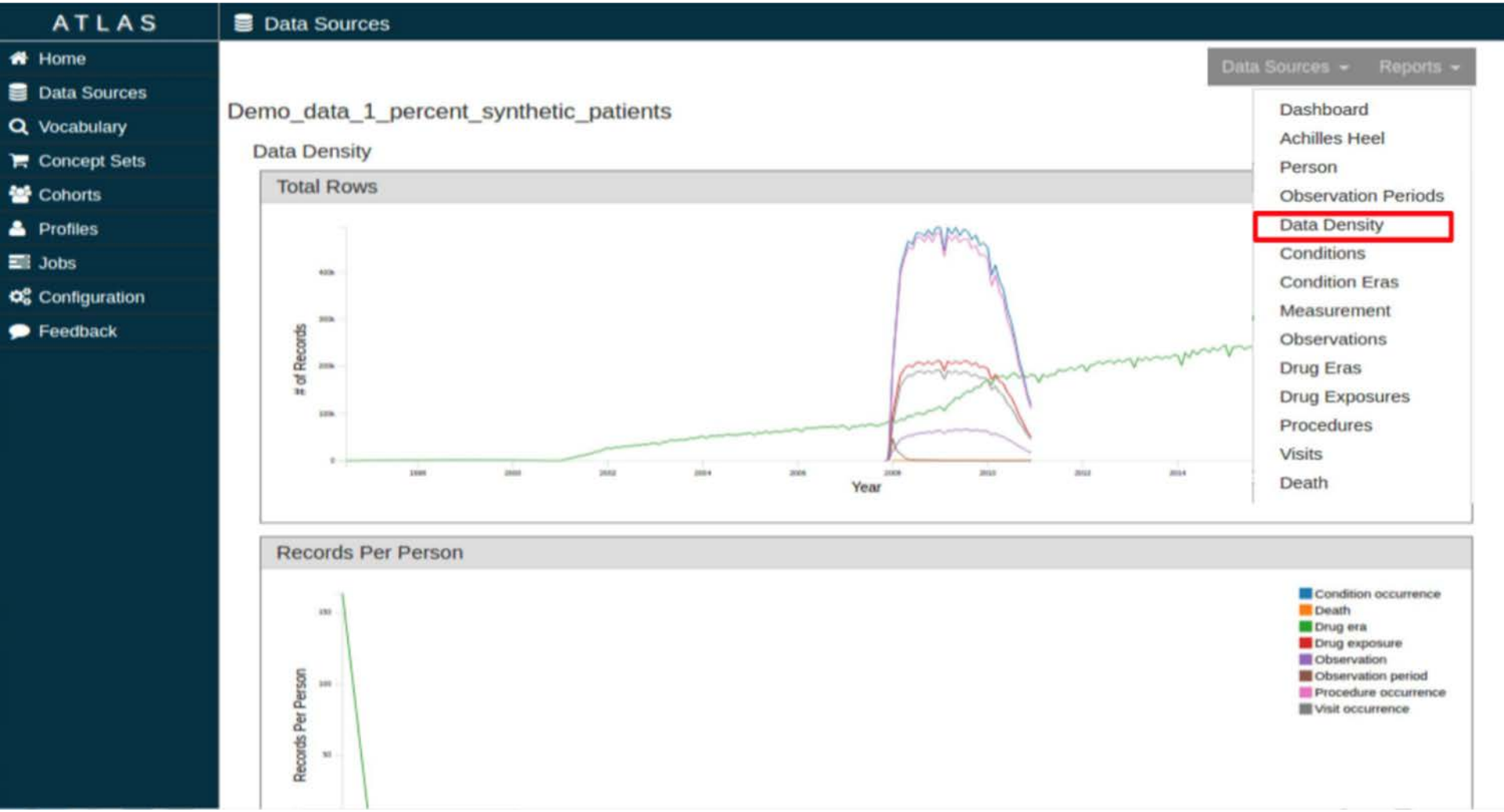
Examples


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Conditions
Condition Prevalence
Treemap
Table
Box Size: Prevalence, Color: Records per Person (Blue to Orange = Low to High), Use Ctrl-Click to Zoom, Alt-Click to Reset Zoom
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death

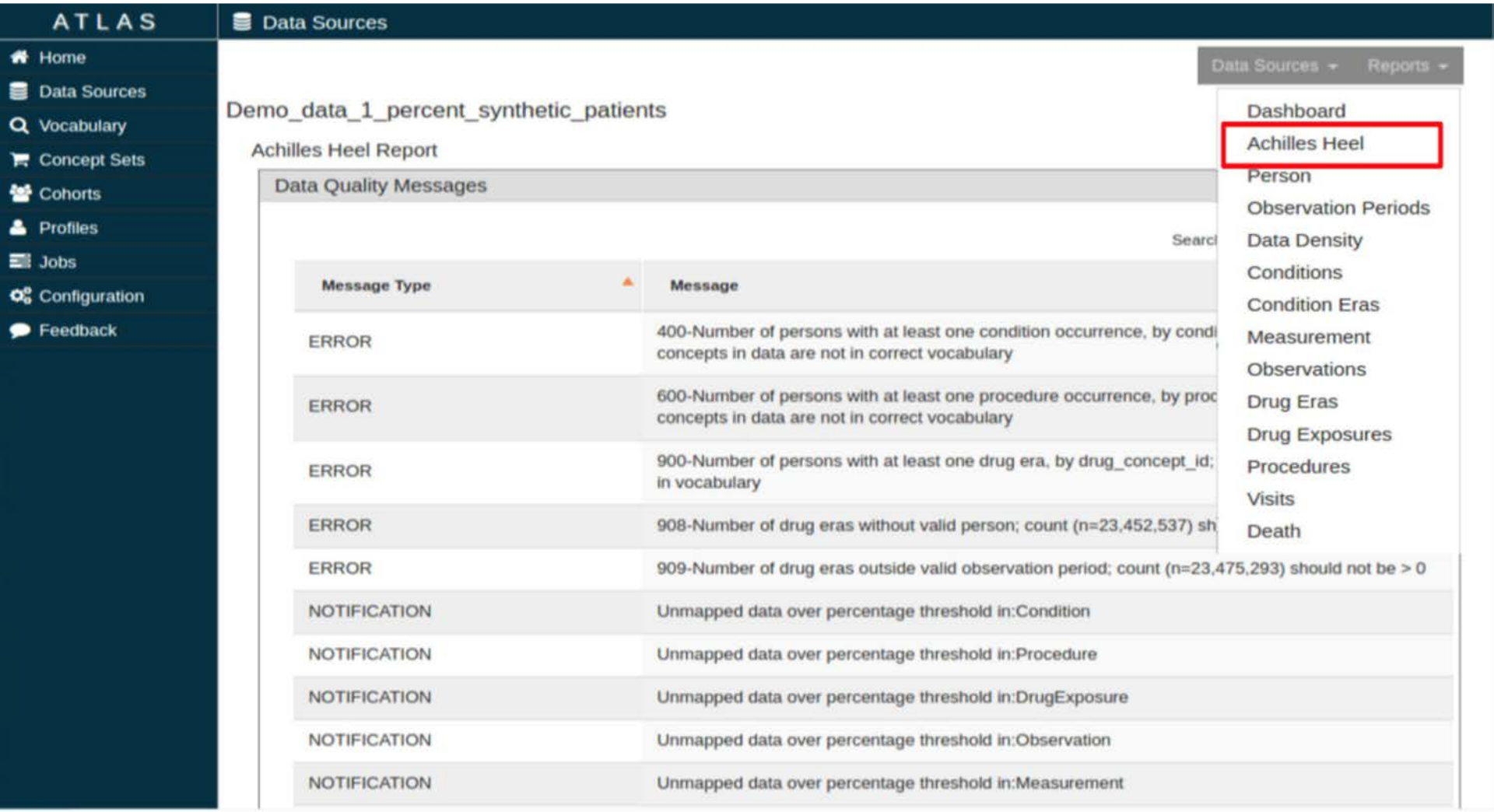
Examples


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Conditions
Condition Prevalence
Treemap
Table
Vascular disorders
Arteriosclerosis, stenosis, vascular insufficiency and necrosis
Non-site specific necrosis and vascular insufficiency NEC
Arteriosclerosis
Coronary arteriosclerosis
Prevalence: 44.07%
Number of People: 51,278
Records per Person: 2.51
Box Size: Prevalence, Color: Records per Person (Blue to Orange = Low to High), Use Ctrl-Click to Zoom, Alt-Click to Reset Zoom
Coronary arteriosclerosis
Condition Prevalence
Age Decile
0-9
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-99

Examples


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Data Density
Total Rows
# of Records
Year
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death
Records Per Person
Records Per Person
Condition occurrence
Death
Drug era
Drug exposure
Observation
Observation period
Procedure occurrence
Visit occurrence

Examples


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Achilles Heel Report
Data Quality Messages
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death
Message Type
Message
ERROR
400-Number of persons with at least one condition occurrence, by condi
concepts in data are not in correct vocabulary
ERROR
600-Number of persons with at least one procedure occurrence, by proc
concepts in data are not in correct vocabulary
ERROR
900-Number of persons with at least one drug era, by drug_concept_id;
in vocabulary
ERROR
908-Number of drug eras without valid person; count (n=23,452,537) sh
ERROR
909-Number of drug eras outside valid observation period; count (n=23,475,293) should not be > 0
NOTIFICATION
Unmapped data over percentage threshold in:Condition
NOTIFICATION
Unmapped data over percentage threshold in:Procedure
NOTIFICATION
Unmapped data over percentage threshold in:DrugExposure
NOTIFICATION
Unmapped data over percentage threshold in:Observation
NOTIFICATION
Unmapped data over percentage threshold in:Measurement

Cohort:
Definition and characterization

Defining ‘phenotype’

Journal of the American Medical Informatics Association, 0(0), 2017, 1–6
doi: 10.1093/jamia/ocx110
Perspective

Perspective

High-fidelity phenotyping: richness and freedom from bias

George Hripcsak[1] and David J Albers[1]

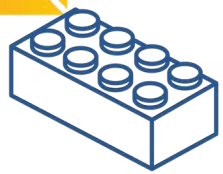
- A phenotype is a specification of an observable, potentially changing state of an organism (as distinguished from the genotype, derived from genetic makeup).
- The term phenotype can be applied to patient characteristics inferred from electronic health record (EHR) data.
- The goal is to draw conclusions about a target concept based on raw EHR data, claims data, or other clinically relevant data.
- Phenotype algorithms – ie, algorithms that identify or characterize phenotypes – may be generated by domain exerts and knowledge engineers, or through diverse forms of machine learning to generate novel representations of data.

Two Approaches to Phenotyping

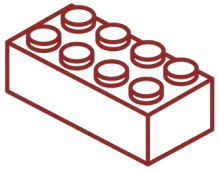
Rule-Based Phenotyping

Probabilistic Phenotyping

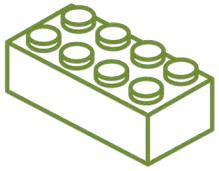
Data are Like Lego Bricks for Phenotypng



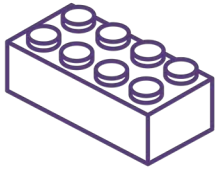
Conditions



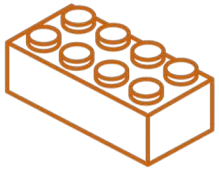
Drugs



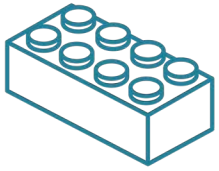
Procedures



Measurements



Observations



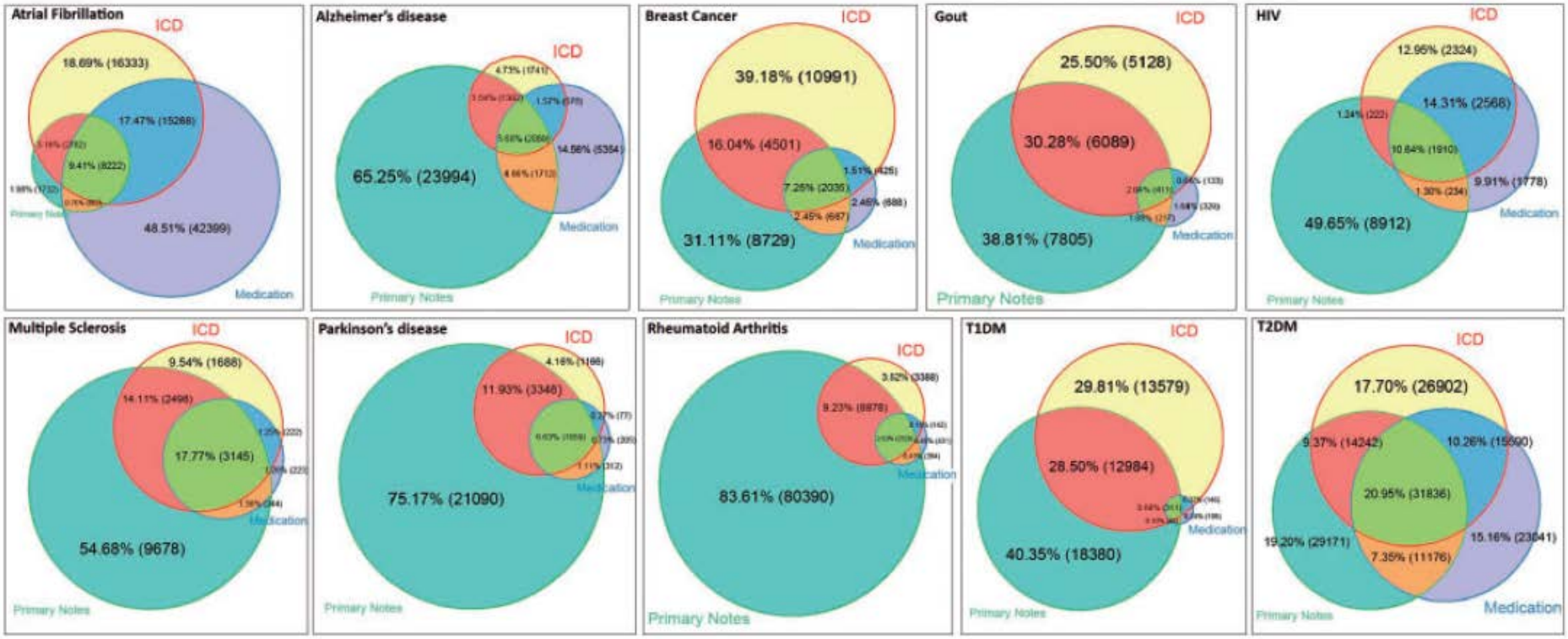
Visits

Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance

RECEIVED 8 January 2015
REVISED 14 July 2015
ACCEPTED 15 July 2015
PUBLISHED ONLINE FIRST 2 September 2015

Wei-Qi Wei[1], Pedro L Teixeira[1], Huan Mo[1], Robert M Cronin[1,2], Jeremy L Warner[1,2], Joshua C Denny[1,2]

Figure 1: Weighted Venn diagrams of the distributions of patients with ICD-9, primary notes, and specific medications. Each color represents a resource. Different area colors represent the number of patients that were found within intersecting resources.



OHDSI's definition of 'cohort'

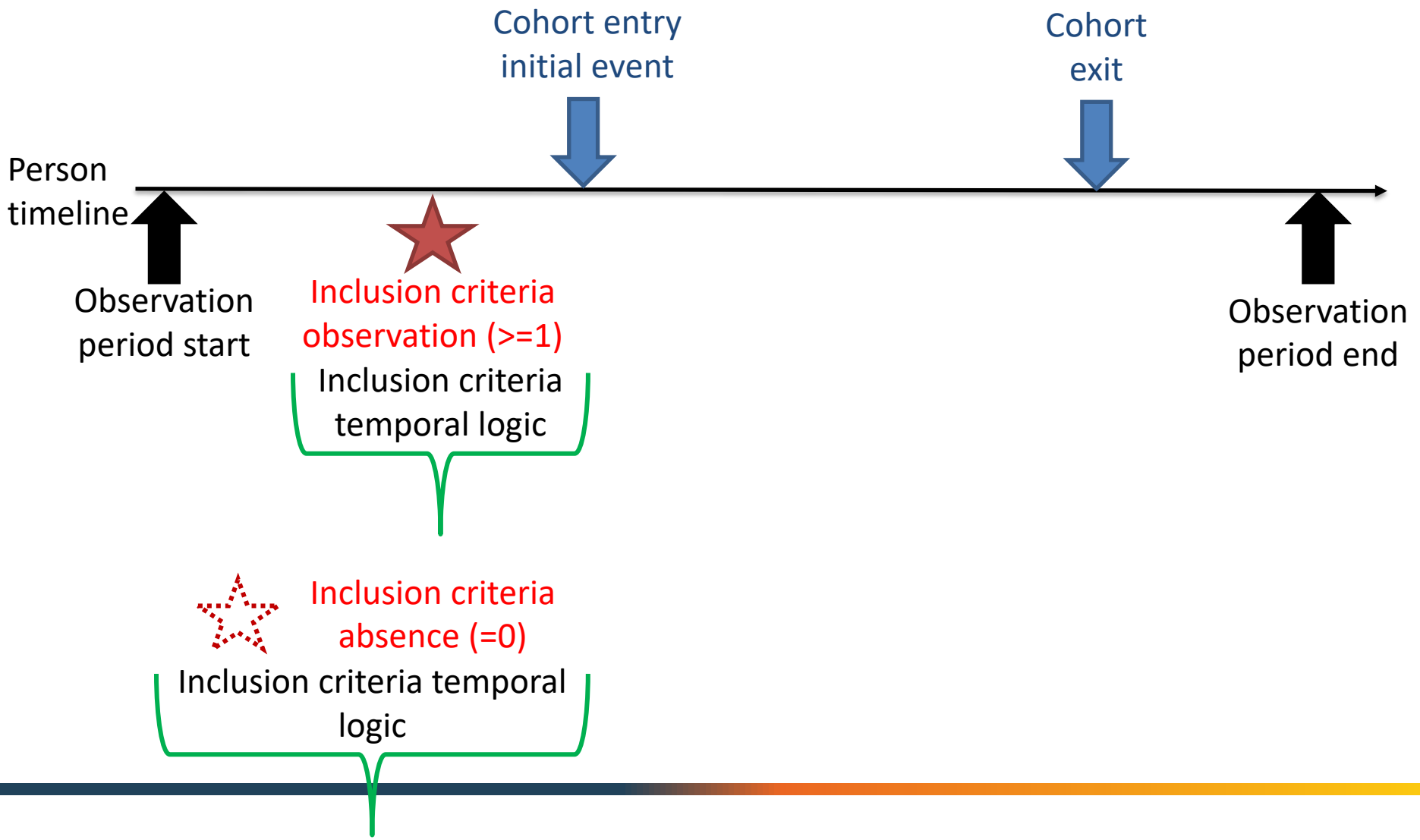
Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

Objective consequences based on this cohort definition:

- One person may belong to multiple cohorts
- One person may belong to the same cohort at multiple different time periods
- One person may not belong to the same cohort multiple times during the same period of time
- One cohort may have zero or more members
- A codeset is NOT a cohort…

  …logic for how to use the codeset in a criteria is required

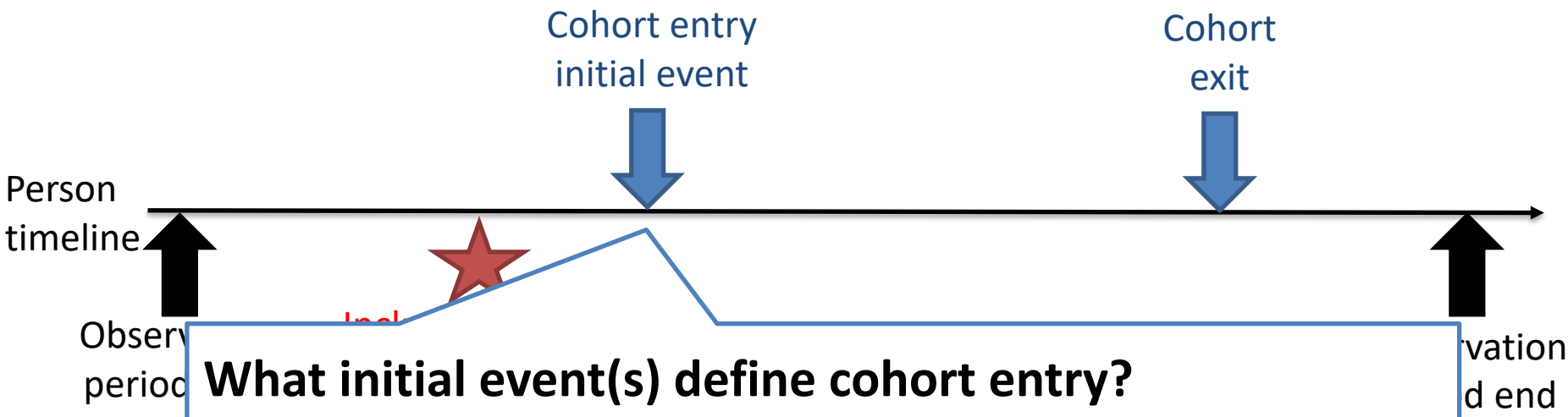
Dissecting the anatomy of a cohort definition



Questions to answer when defining a cohort

- What initial event(s) define cohort entry?
- What inclusion criteria are applied to the initial events?
- What defines a person's cohort exit?

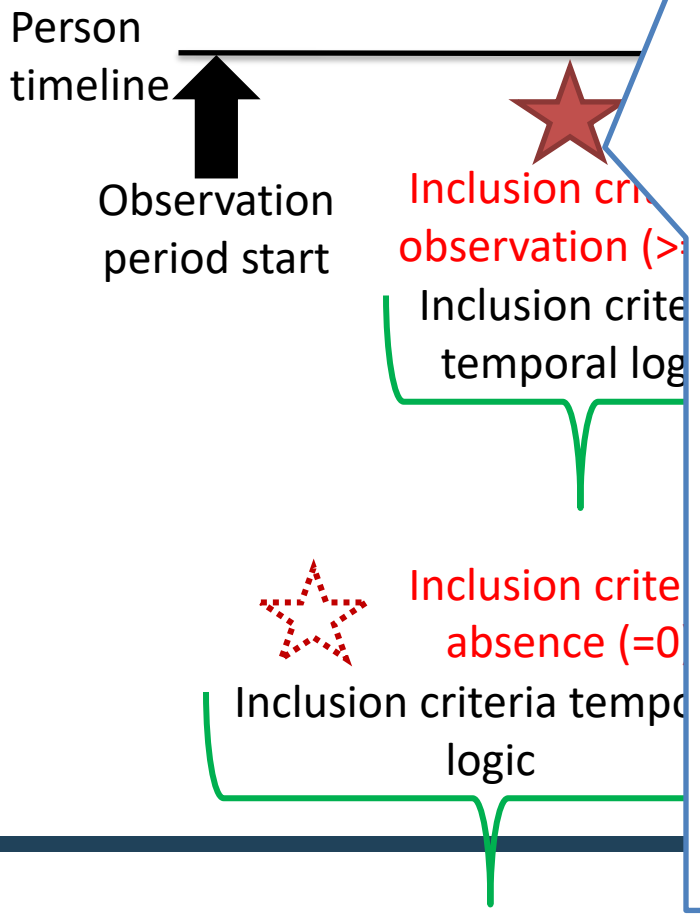
Dissecting the anatomy of a cohort definition



What initial event(s) define cohort entry?

- Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits.
- The event index date is set to be equal to the event start date
- Initial events defined by a domain, conceptset, and any domain-specific attributes required

Dissecting the anatomy of a cohort definition



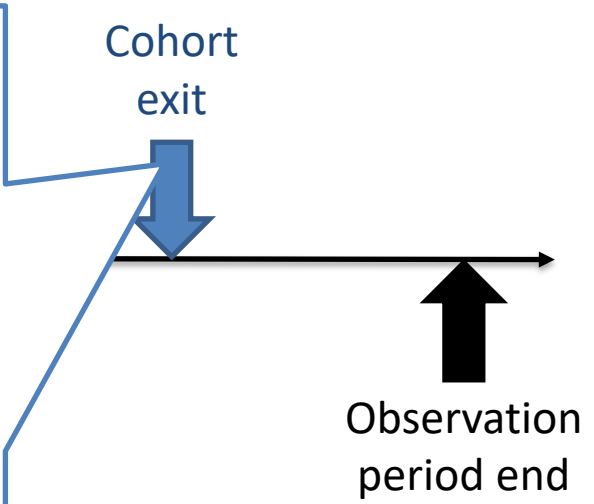
What inclusion criteria are applied to the initial events?

- The qualifying cohort will be defined as all persons who have an initial event and satisfy all qualifying inclusion criteria.
- Each inclusion criteria is defined by domain(s), conceptset(s), domain-specific attributes, and the temporal logic relative to initial events
- Each qualifying inclusion criteria can be evaluated to determine the impact of the criteria on the attrition of persons from the initial cohort (example use case: clinical trial feasibility)

Dissecting the anatomy of a cohort definition

What defines a person's cohort exit?

- Cohort exit signifies when a person no longer qualifies for cohort membership
- Cohort exit can be defined in multiple ways:
  - End of observation period
  - Fixed time interval relative to initial event
  - Last event in a sequence of related observations (ex: persistent drug exposure)
  - Censoring observations
- Cohort exit strategy will impact whether a person can belong to the cohort multiple times during different time intervals



Defining cohort components

- Domain: A Domain defines the set of allowable Concepts for the standardized fields in the CDM tables.
  - Ex: Condition, Drug, Procedure, Measurement
- Conceptset: An expression that defines one or more concepts encompassing a clinical entity of interest
  - Ex: Concepts for T2DM, concepts for antidiabetic drugs
- Domain-specific attribute:
  - Ex: DRUG_EXPOSURE: Days supply; MEASUREMENT: value_as_number, high_range
- Temporal logic: the time intervals within which the relationship between an inclusion criteria and an event is evaluated
  - Ex: Indicated condition must occur during 365d prior to or on exposure start

Cardiovascular, Bleeding, and Mortality Risks in Elderly Medicare Patients Treated With Dabigatran or Warfarin for Nonvalvular Atrial Fibrillation

David J. Graham, MD, MPH; Marsha E. Reichman, PhD; Michael Wernecke, BA; Rongmei Zhang, PhD; Mary Ross Southworth, PharmD; Mark Levenson, PhD; Ting-Chang Sheu, MPH; Katrina Mott, MHS; Margie R. Goulding, PhD; Monika Houstoun, PharmD, MPH; Thomas E. MaCurdy, PhD; Chris Worrall, BS; Jeffrey A. Kelman, MD, MMSc

Background—The comparative safety of dabigatran versus warfarin for treatment of nonvalvular atrial fibrillation in general practice settings has not been established.

Methods and Results—We formed new-user cohorts of propensity score–matched elderly patients enrolled in Medicare who initiated dabigatran or warfarin for treatment of nonvalvular atrial fibrillation between October 2010 and December 2012. Among 134414 patients with 37587 person-years of follow-up, there were 2715 primary outcome events. The hazard ratios (95% confidence intervals) comparing dabigatran with warfarin (reference) were as follows: ischemic stroke, 0.80 (0.67–0.96); intracranial hemorrhage, 0.34 (0.26–0.46); major gastrointestinal bleeding, 1.28 (1.14–1.44); acute myocardial infarction, 0.92 (0.78–1.08); and death, 0.86 (0.77–0.96). In the subgroup treated with dabigatran 75 mg twice daily, there was no difference in risk compared with warfarin for any outcome except intracranial hemorrhage, in which case dabigatran risk was reduced. Most patients treated with dabigatran 75 mg twice daily appeared not to have severe renal impairment, the intended population for this dose. In the dabigatran 150-mg twice daily subgroup, the magnitude of effect for each outcome was greater than in the combined-dose analysis.

Conclusions—In general practice settings, dabigatran was associated with reduced risk of ischemic stroke, intracranial hemorrhage, and death and increased risk of major gastrointestinal hemorrhage compared with warfarin in elderly patients with nonvalvular atrial fibrillation. These associations were most pronounced in patients treated with dabigatran 150 mg twice daily, whereas the association of 75 mg twice daily with study outcomes was indistinguishable from warfarin except for a lower risk of intracranial hemorrhage with dabigatran. (***Circulation***. **2015;131:157-164. DOI: 10.1161/CIRCULATIONAHA.114.012061.)**

Key Words: anticoagulant ■ pharmacoepidemiology ■ safety ■ thrombin inhibitor ■ warfarin

Graham et al. description of the outcomes

Study Outcomes

The primary outcomes were ischemic stroke, major bleeding with specific focus on intracranial and gastrointestinal bleeding, and AMI. Secondary outcomes were all hospitalized bleeding events and mortality. The *International Classification of Diseases, Ninth Revision, Clinical Modification* codes used to define these outcomes are listed in Table II in the online-only Data Supplement. The codes defining ischemic stroke have a positive predictive value (PPV) of 88% to 95%.[18–20] Major bleeding was defined as

Table 2. International Classification of Disease, 9th edition, Clinical Modification (ICD 9-CM) codes used to define study outcomes.

Outcome	ICD-9 Codes	Position	Setting
AMI	410 (all)	1st or 2nd	IP only
Ischemic stroke	433.x1, 434.x (except subcode: x0), 436	1st	IP only

Exercise: Define the outcome cohort for Graham et al.

- What initial event(s) define cohort entry?
- What inclusion criteria are applied to the initial events?
- What defines a person's cohort exit?

Graham et al. description of the cohort(s)

A new-user retrospective cohort design was used to compare patients initiating dabigatran or warfarin for the treatment of nonvalvular AF.[10] We identified all patients with any inpatient or outpatient diagnoses of AF or atrial flutter based on *International Classification of Diseases, Ninth Revision* coding who also filled at least 1 prescription for either drug from October 19, 2010 (US dabigatran approval date) through December 31, 2012, the study end date. Patients were excluded if they had <6 months of enrollment in Medicare before their index dispensing, were aged <65 years, received prior treatment with a study medication or rivaroxaban or apixaban (anticoagulants approved during the study), were in a skilled nursing facility or nursing home, or were receiving hospice care on the date of their cohort-qualifying prescription. Patients were also excluded if they had a hospitalization that extended beyond the index dispensing date. Patients discharged from the hospital on the same day as their index dispensing were included. Patients undergoing dialysis and kidney transplant recipients were also excluded. Additionally, because warfarin is approved for indications other than AF, we excluded patients with diagnoses indicating the presence of mitral valve disease, heart valve repair or replacement, deep vein thrombosis, pulmonary embolism, or joint replacement surgery in the preceding 6 months.

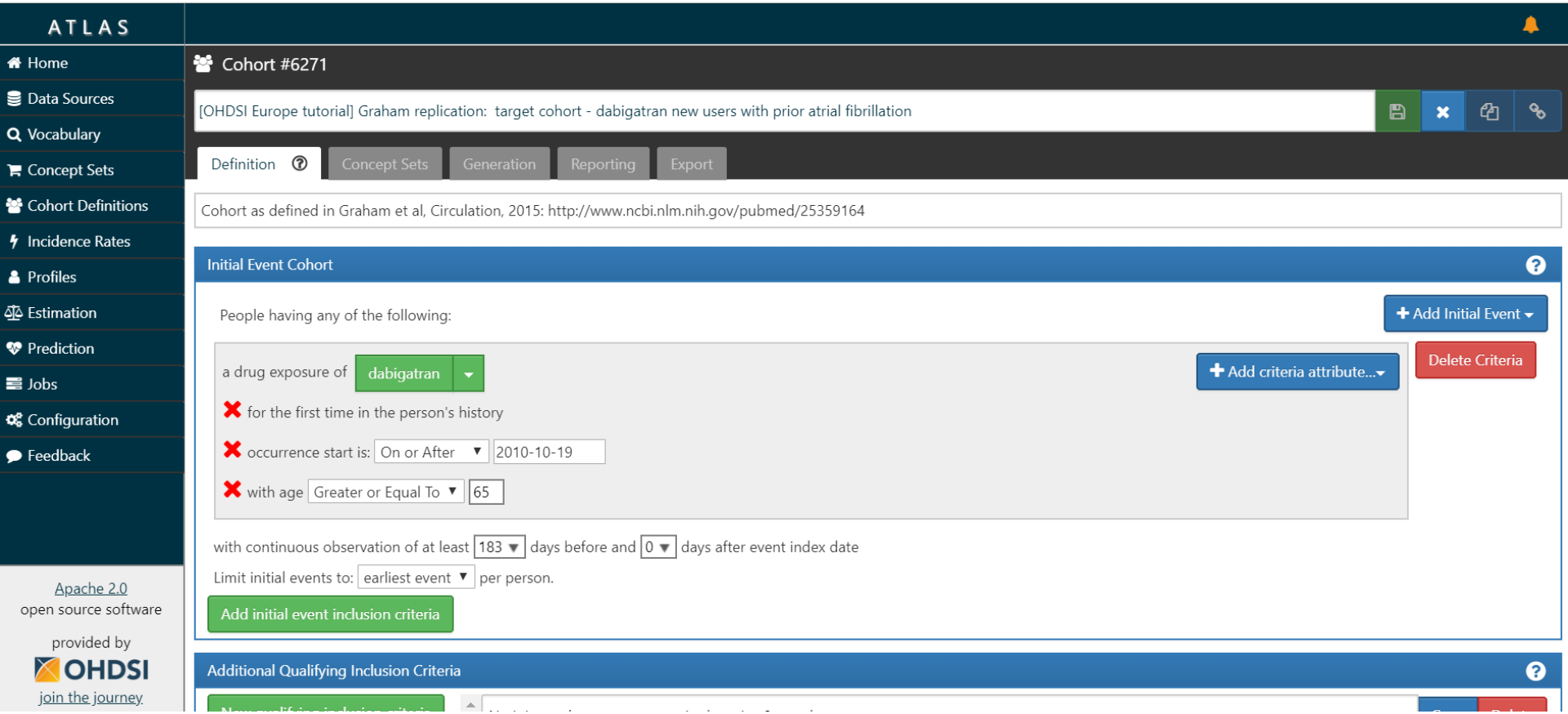
Exercise: Define the target exposure cohort for Graham et al.

- What initial event(s) define cohort entry?
- What inclusion criteria are applied to the initial events?
- What defines a person's cohort exit?

Demo: Implementing cohorts in ATLAS

Follow along at:
http://ohdsi.org/web/ATLAS

Demo: Cohort definition



Demo: Cohort definition – initial event

Definition | Concept Sets | Generation | Reporting | Export

Cohort as defined in Graham et al, Circulation, 2015: http://www.ncbi.nlm.nih.gov/pubmed/25359164

Initial Event Cohort

People having any of the following:

a drug exposure of dabigatran

✖ for the first time in the person's history

✖ occurrence start is: On or After 2010-10-19

✖ with age Greater or Equal To 65

with continuous observation of at least 183 days before and 0 days after event index date

Limit initial events to: earliest event per person.

Add initial event inclusion criteria

Demo: Cohort definition – inclusion criteria

1. Has prior atrial fibrillation or atrial flutter diagnosis
2. Has no prior comparator
3. Has no prior other antico (rivaroxaban
4. Not in a skil nursing hom hospice care
5. Not underg kidney trans
6. No mitral va valve repair, the prior 6 r
7. No deep vei pulmonary e 6 months
8. No joint rep the prior 6 r

Has prior atrial fibrillation or atrial flutter diagnosis

having any of the following criteria:

with at least 1 using all occurrences of:

a condition occurrence of Atrial fibrillation

starting between All days Before and 0 days After event index date and ending any time.

restrict to the same visit occurrence

or with at least 1 using all occurrences of:

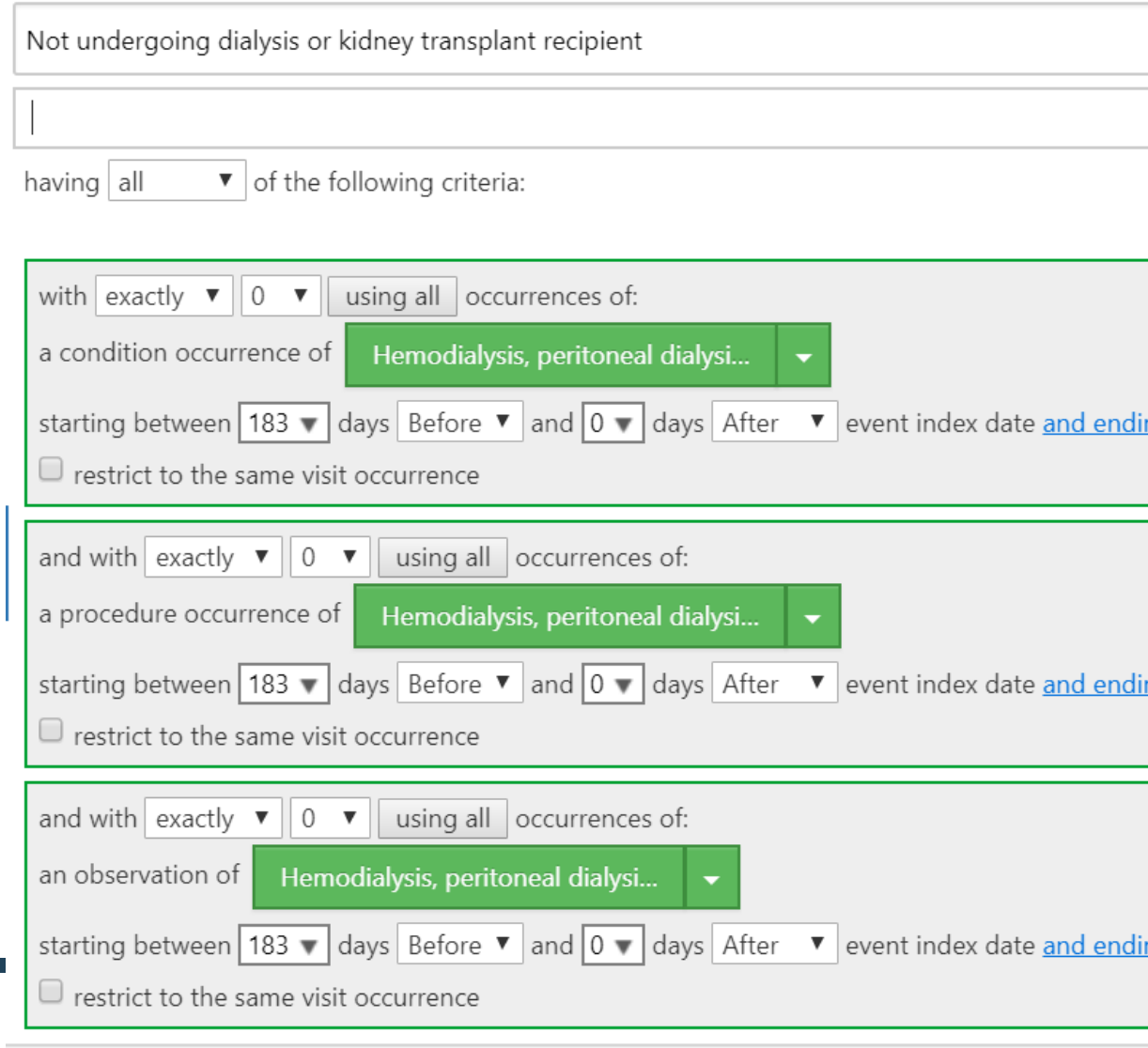
a condition occurrence of Atrial flutter

starting between All days Before and 0 days After event index date and ending any time.

restrict to the same visit occurrence

Demo: Cohort definition – inclusion criteria

3. Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)
4. Not in a skilled nursing facility or nursing home, or receiving hospice care on the index date
5. **Not undergoing dialysis or kidney transplant recipient**
6. No mitral valve disease, heart valve repair, or replacement in the prior 6 months
7. No deep vein thrombosis or pulmonary embolism in the prior 6 months



Demo: Cohort definition – exit criteria

Cohort Exit Criteria

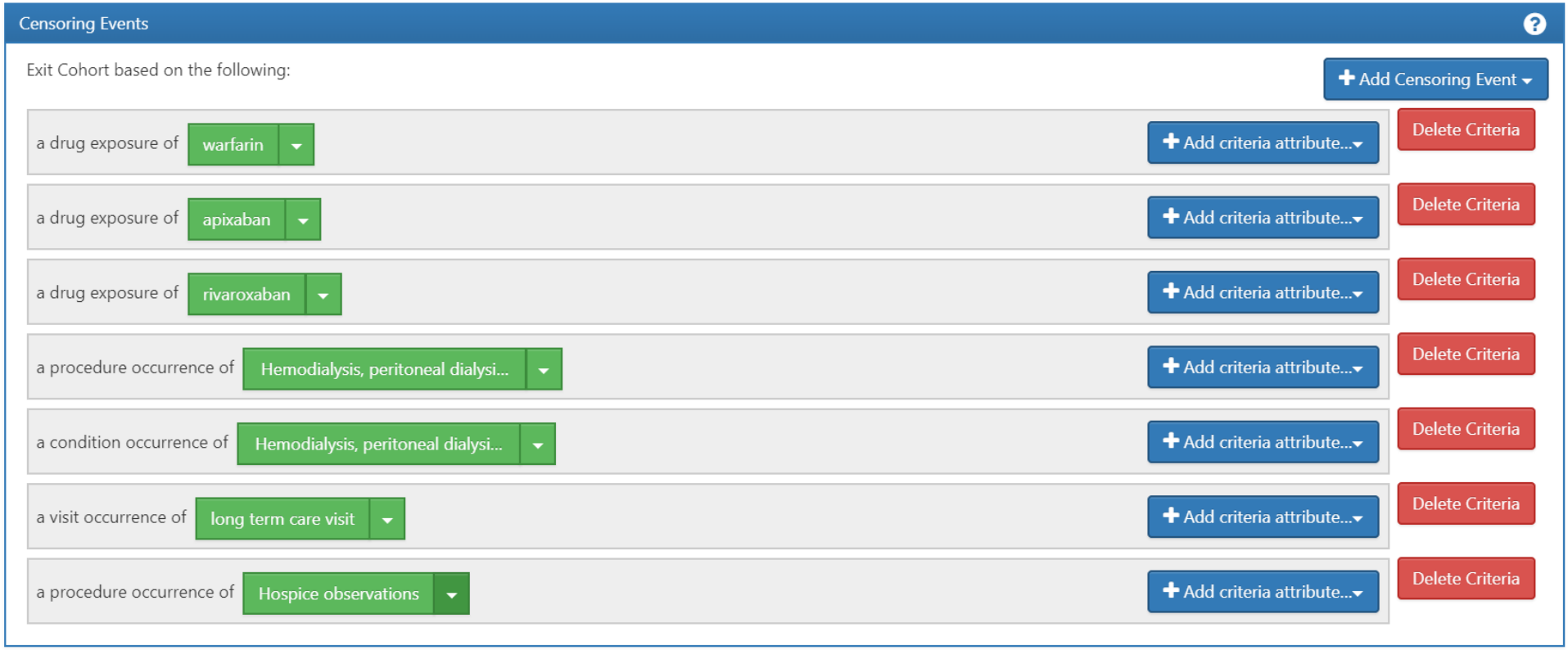
Cohort exit criteria based on the end of an era of persistent exposure to any drug within a defined concept set:
Specify a concept set that contains one or more drugs. A drug era will be derived from all drug exposure events for any of the drugs within the concept set, using the specified persistence window as a maximum allowable gap in days between successive exposure events and adding a specified surveillance window to the final exposure event. If no exposure event end date is provided, then an exposure event end date is inferred to be event start date + days supply in cases when days supply is available or event start date + 1 day otherwise. This cohort exit criteria assures that the cohort end date will be no greater than the drug era end date.

Concept set containing the drug(s) of interest: dabigatran

- Persistence window: allow for a maximum of 3 days between exposure records when inferring the era of persistence exposure
- Surveillance window: add 0 days to the end of the era of persistence exposure as an additional period of surveillance prior to cohort exit.

Remove Cohort Exit Criteria

Demo: Cohort definition – exit criteria censoring events



Demo: Cohort definition – conceptsets

[OHDSI Europe tutorial] Graham replication: target cohort - dabigatran new users with prior atrial fibrillation

Definition | Concept Sets | Generation | Reporting | Export

Show 25 entries

Id	Title
0	dabigatran
1	Atrial fibrillation
2	Atrial flutter
3	rivaroxaban
4	apixaban

Concept Set Expression | Included Concepts 160 | Included Source Codes | Export

Name: dabigatran

Show 25 entries Search:

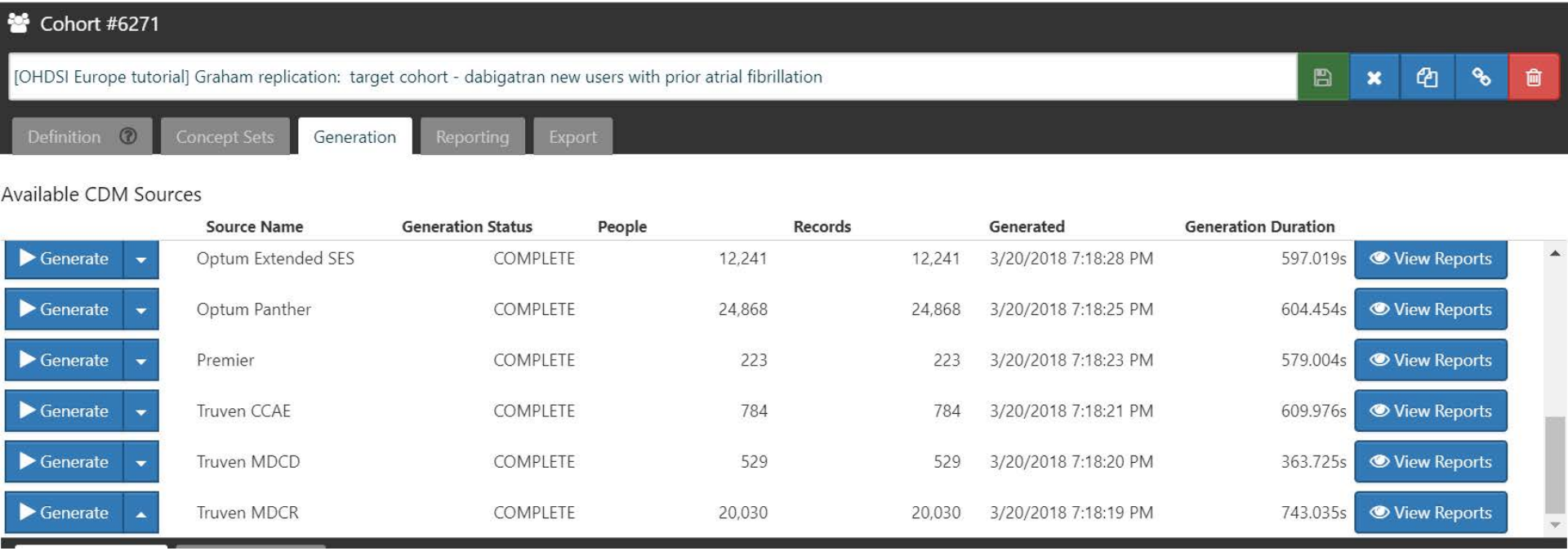
Showing 1 to 1 of 1 entries Previous 1 Next

Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	Exclude	Descendants	Mapped
40228152	1037042	dabigatran etexilate	Drug	Standard	☐	☑	☐

Classification Non-Standard Standard

Delete Concept Set | Close Concept Set

Demo: Cohort generation



Demo: Cohort – inclusion report to evaluate impact of criteria

Inclusion Report | Cohort Features

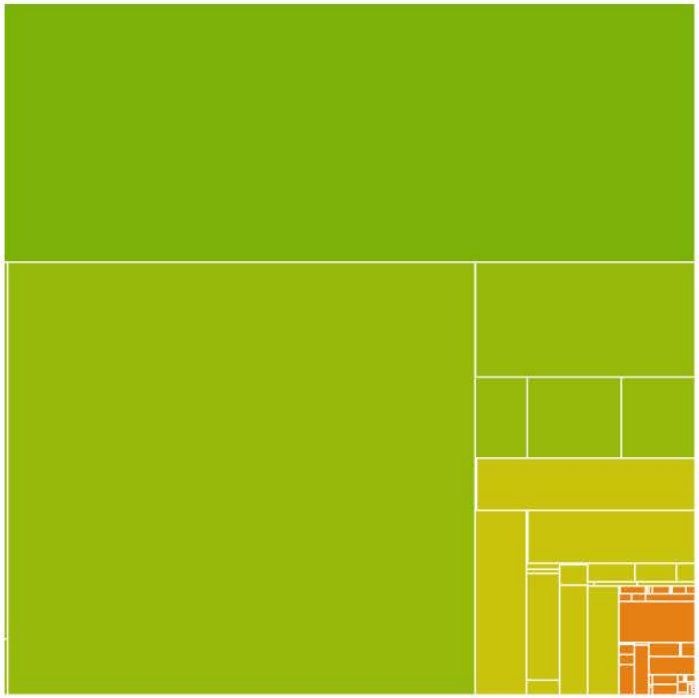
Inclusion Report for Truven MDCR

	Match Rate	Matches	Total
Summary Statistics:	37.42%	20,030	53,523

Inclusion Rule	N	% Satisfied	% To-Gain
1. Has prior atrial fibrillation or atrial flutter diagnosis	48,462	90.54%	5.28%
2. Has no prior treatment with comparator drug (warfarin)	26,010	48.60%	42.32%
3. Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)	51,761	96.71%	1.59%
4. Not in a skilled nursing facility or nursing home, or receiving hospice care on the index date	53,508	99.97%	0.01%
5. Not undergoing dialysis or kidney transplant recipient	53,448	99.86%	0.04%
6. No mitral valve disease, heart valve repair, or replacement in the prior 6 months	51,063	95.40%	1.25%
7. No deep vein thrombosis or pulmonary embolism in the prior 6 months	50,748	94.82%	0.88%
8. No joint replacement surgery in the prior 6 months	52,817	98.68%	0.30%

Population Visualization

Switch to attrition view



Demo: Cohort characterization

Inclusion Report | Cohort Features

Feature Report for Truven MDCR

Features are baseline characteristics (e.g collected before /on cohort start)

Demographics | Conditions | Drugs | Procedures | Measurements | Observations | Distributions

Column visibility | Copy | CSV | Show 15 entries

Showing 1 to 15 of 20 entries

Analysis

Index Month (12)

Age Group (6)

Gender (2)

Time Window

None (20)

Name	Count	% of cohort
age group: 65-69	3,853	19.30
age group: 70-74	4,520	22.60
age group: 75-79	4,551	22.80
age group: 80-84	3,956	19.80
age group: 85-89	2,358	11.80
age group: 90-94	695	3.50
FEMALE	8,781	43.90
MALE	11,249	56.20

Demo: Cohort characterization

Inclusion Report | Cohort Features

Feature Report for Truven MDCR

Features are baseline characteristics (e.g collected before /on cohort start)

Demographics | Conditions | Drugs | Procedures | Measurements | Observations | Distributions

Long Term: 365 day lookback. **Medium Term:** 180d lookback **Short Term:** 30d lookback. **Overlapping:** Event spans cohort start date.

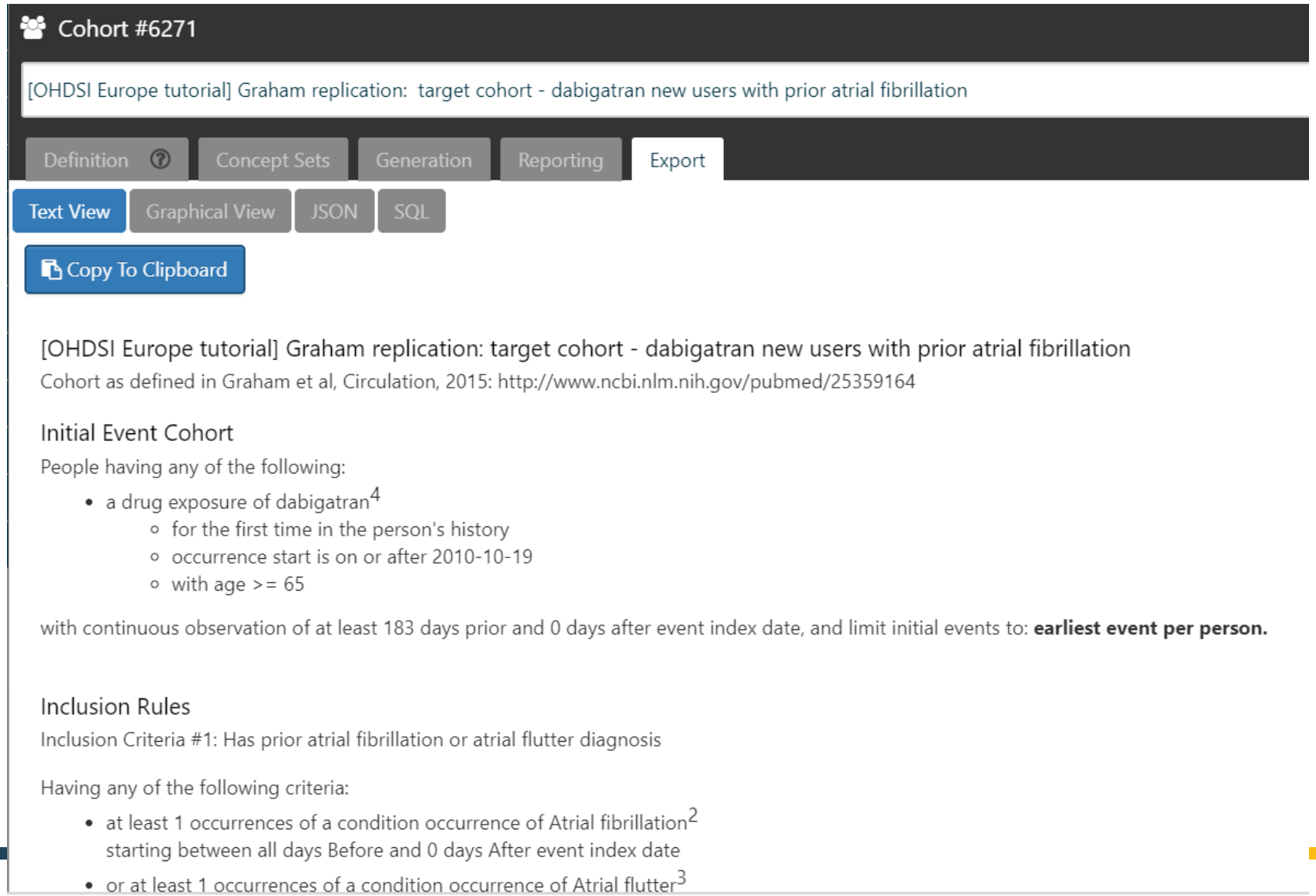
Column visibility | Copy | CSV | Show 15 entries | Filter:

Showing 1 to 15 of 756 entries | Previous 1 2 3 4 5 ... 51 Next

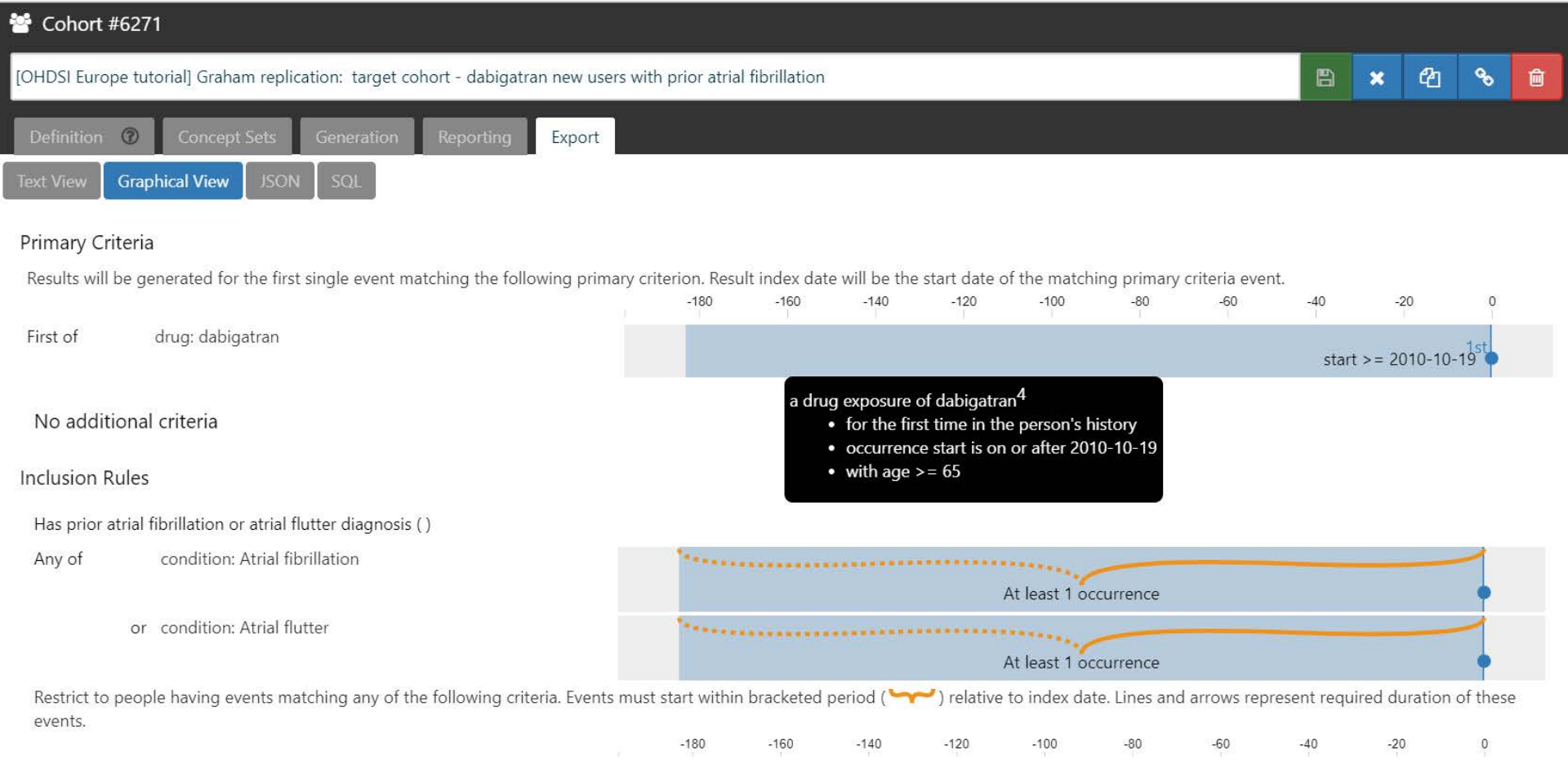
Analysis
Group Era (2500)
Time Window
Long Term (1744)
Short Term (756)

	Concept Name	Time Window	Person Count	% of cohort
Explore	Heart disease	Short Term	18,134	90.60
Explore	Cardiac arrhythmia	Short Term	17,527	87.60
Explore	Supraventricular arrhythmia	Short Term	17,365	86.70
Explore	Atrial arrhythmia	Short Term	17,338	86.60
Explore	Fibrillation	Short Term	16,397	81.90
Explore	Atrial fibrillation	Short Term	16,385	81.90
Explore	Hypertensive disorder	Short Term	8,943	44.70
Explore	Essential hypertension	Short Term	8,931	44.60

Demo: Cohort definition – protocol text



Demo: Cohort definition – graphical view



Demo: Cohort definition – transportable JSON code

[OHDSI Europe tutorial] Graham replication: target cohort - dabigatran new users with prior atrial fibrillation

Definition | Concept Sets | Generation | Reporting | Export

Text View | Graphical View | JSON | SQL


```
{
  "ConceptSets": [
    {
      "id": 0,
      "name": "dabigatran",
      "expression": {
        "items": [
          {
            "concept": {
              "CONCEPT_ID": 40228152,
              "CONCEPT_NAME": "dabigatran etexilate",
              "STANDARD_CONCEPT": "S",
              "INVALID_REASON": "V",
              "CONCEPT_CODE": "1037042",
              "DOMAIN_ID": "Drug",
              "VOCABULARY_ID": "RxNorm",
              "CONCEPT_CLASS_ID": "Ingredient",
              "INVALID_REASON_CAPTION": "Valid",
```


Copy To Clipboard

Demo: Cohort definition – templated SQL

[OHDSI Europe tutorial] Graham replication: target cohort - dabigatran new users with prior atrial fibrillation

Definition | Concept Sets | Generation | Reporting | Export

Text View | Graphical View | JSON | SQL

Template OHDSI.SQL | MSSQL Server | MS APS | Oracle | PostgreSQL | Amazon Red Shift | Impala

Copy To Clipboard

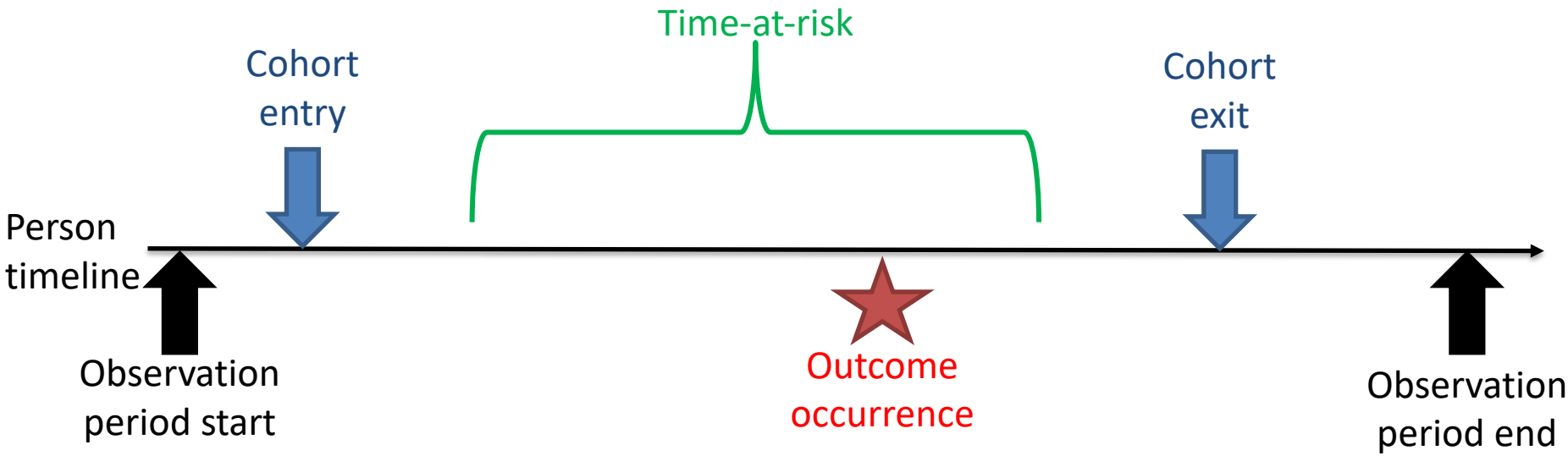
```
CREATE TABLE #Codesets (
  codeset_id int NOT NULL,
  concept_id bigint NOT NULL
)
;

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 @vocabulary_database_schema.CONCEPT where concept_id in (40228152)and invalid_reason is null
UNION  select c.concept_id
  from @vocabulary_database_schema.CONCEPT c
  join @vocabulary_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
  and ca.ancestor_concept_id in (40228152)
  and c.invalid_reason is null

) I
) C;
```

Incidence rate

Dissecting the anatomy of incidence



Incidence metrics:

$$\text{Incidence proportion} = \frac{\text{\# persons in the target cohort who have new outcome occurrence during the time-at-risk}}{\text{\# persons in the target cohort with time-at-risk*}}$$

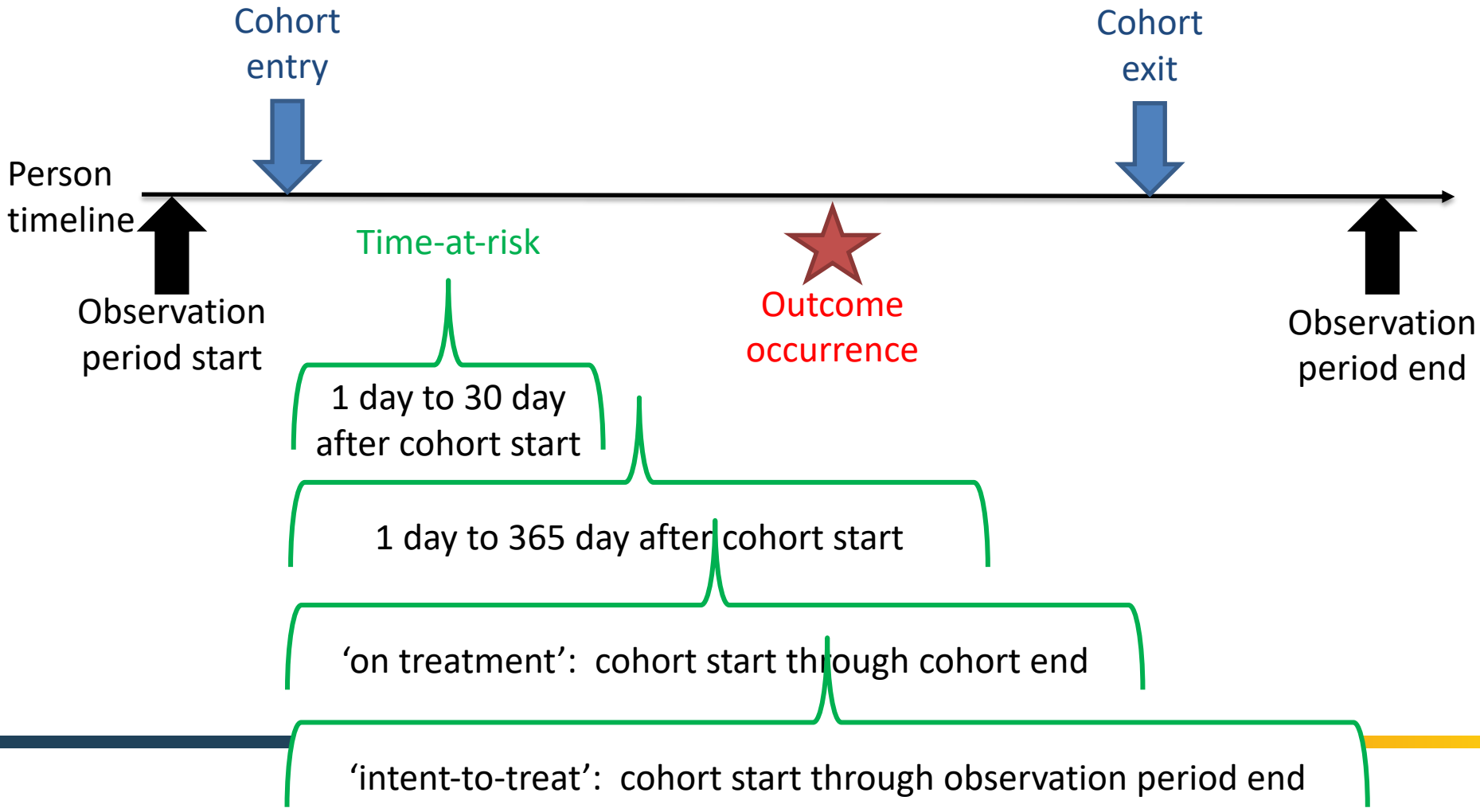
$$\text{Incidence rate} = \frac{\text{\# persons in the target cohort who have new outcome occurrence during the time-at-risk}}{\text{person-time at-risk for persons in the target cohort with time-at-risk*}}$$

Myriad difficult choices that researchers have to make to produce a ‘simple answer’

- How should the target cohort be defined?
- How should the outcome be defined?
- How should the time-at-risk be defined?
- How to account for patients with incomplete time-at-risk?
- Which statistical metrics should be reported?
- Which data should be used?

Myriad difficult choices that researchers have to make to produce a 'simple answer'

- **How should the time-at-risk be defined?**



Decisions for incidence rate estimations in the OHDSI framework

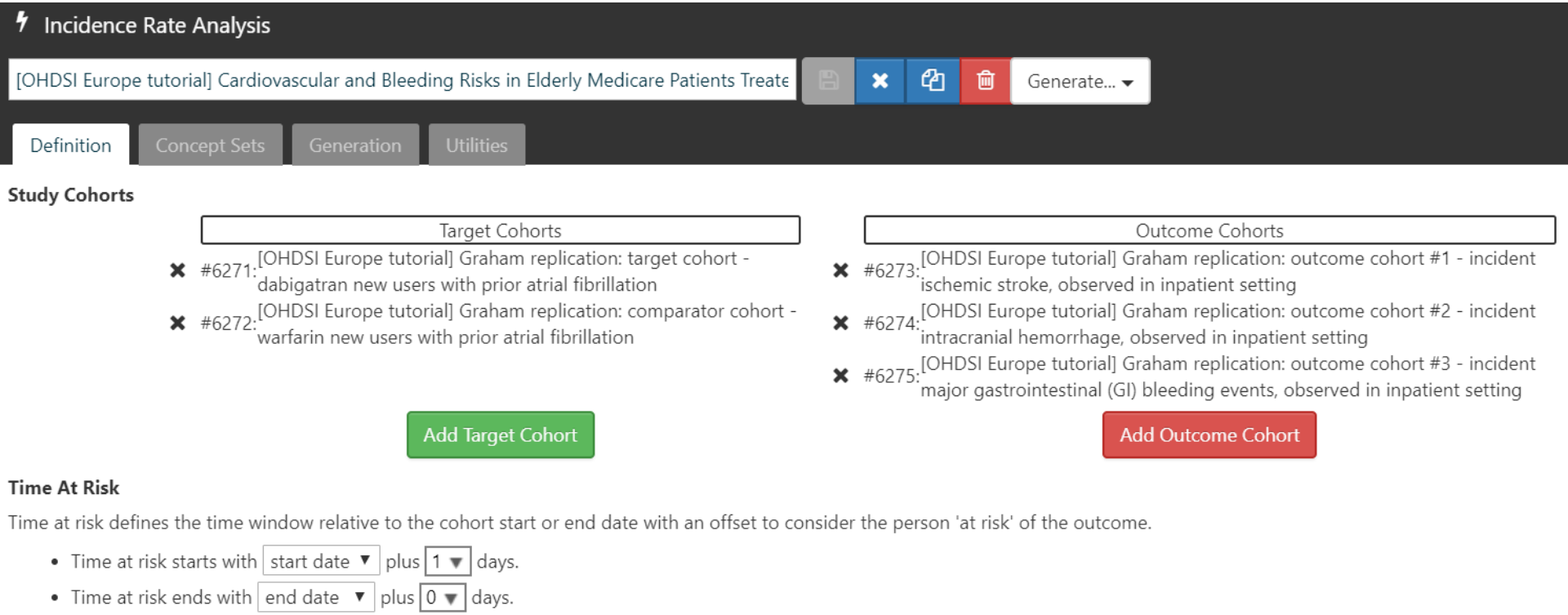
- What's your **T**arget cohort(s)?
- What's your **O**utcome cohort(s)?
- What's your time-at-risk?
- What's your stratification criteria?

Demo: Implementing incidence rates in ATLAS

Follow along at:

http://ohdsi.org/web/ATLAS

Demo: incidence rate specification



Demo: incidence rate generation

Incidence Rate Analysis

[OHDSI Europe tutorial] Cardiovascular and Bleeding Risks in Elderly Medicare Patients Treate | Generate...

Definition | Concept Sets | Generation | Utilities

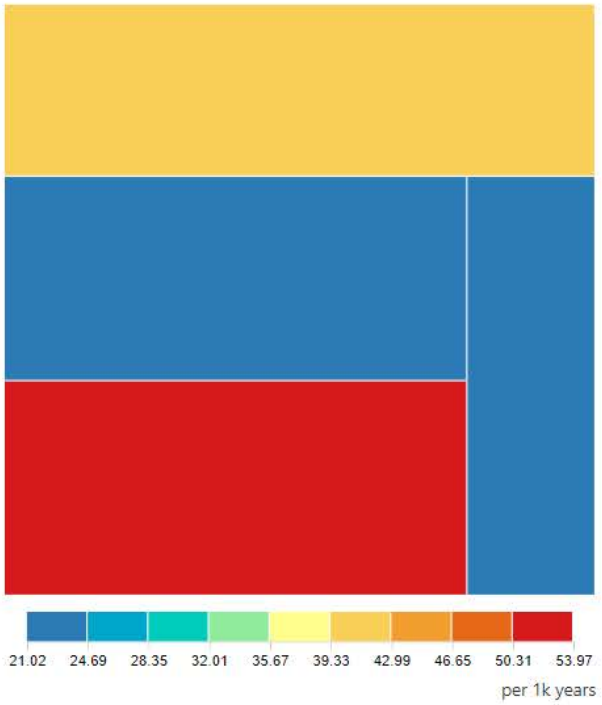
Export Analysis to CSV

Source	Name		Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years	Started	Duration	
TRUVENMDCR_V698	Truven MDCR	Execute	19,288	201	10.42	5,606	35.85	2018-03-21, 15:20	00:01:12	Remove

Showing target cohort: [OHDSI Europe tutorial] Graham replicatio and outcome cohort: [OHDSI Europe tutorial] Graham replicatic

	Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
Summary Statistics:	19,288	201	10.42	5,606	35.85

Stratify Rule	N	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
1. Gender = MALE	10,839	99	9.13	3,223	30.72
2. Age >= 75	11,101	150	13.51	3,239	46.31



Estimation: Population-level effect estimation using the comparative cohort design

Full-day tutorial in 30 minutes

- Will focus on key concepts here
- View video of full day here: https://www.ohdsi.org/past-events/2017-tutorials-population-level-estimation/

Two types of questions

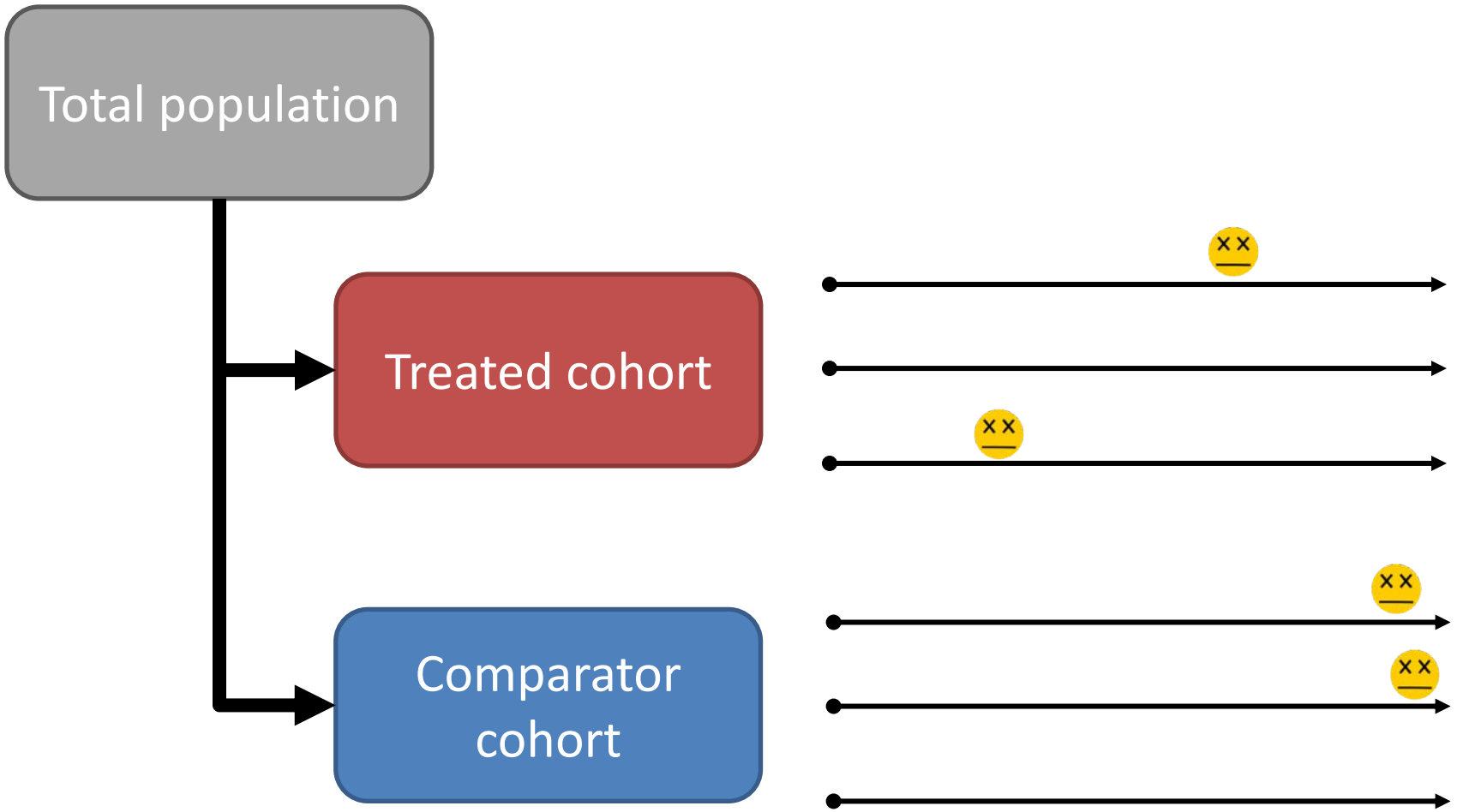
- Does exposure T cause outcome O?

  Effect estimation

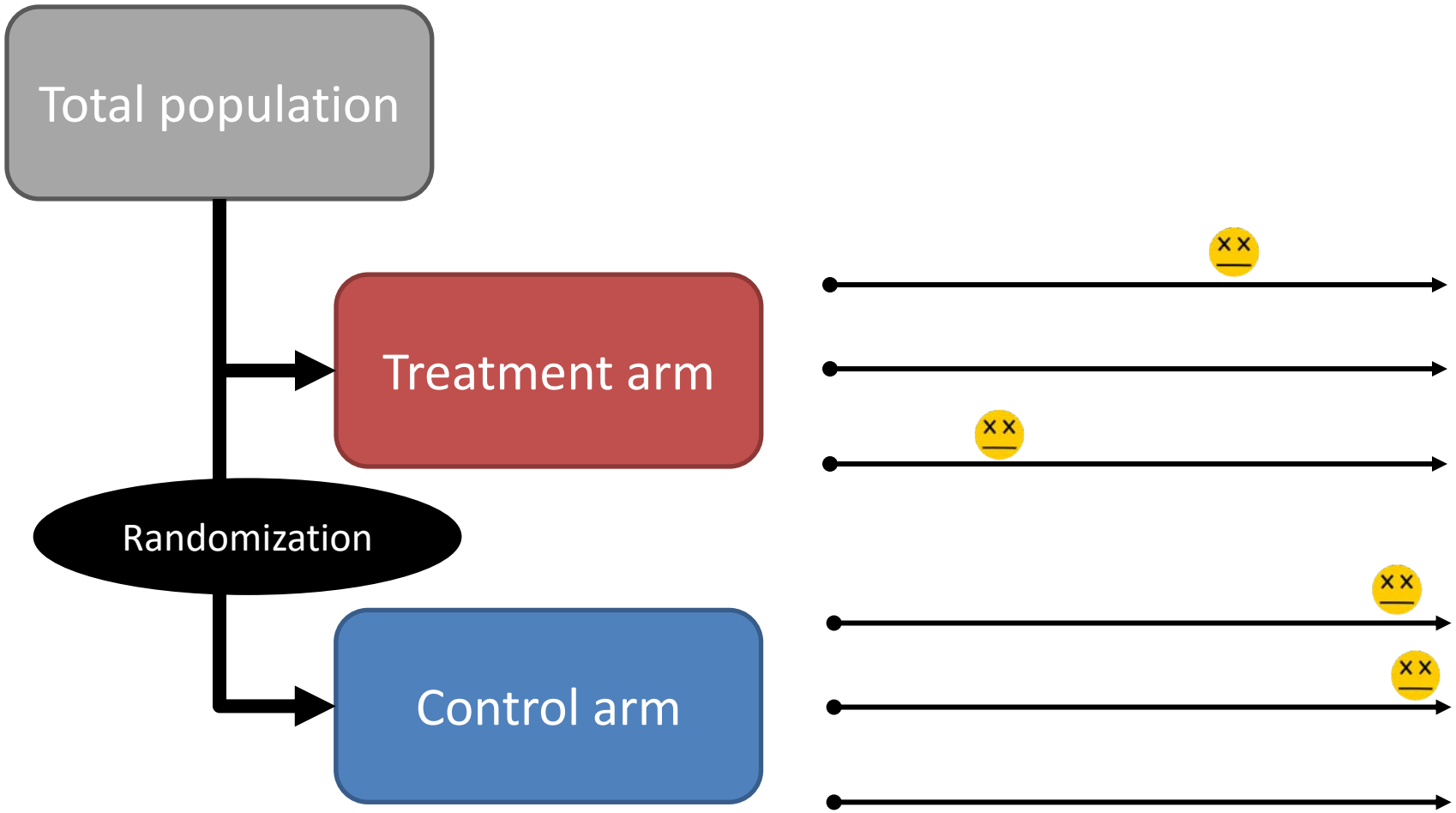
- Does exposure T cause outcome O compared to exposure C?

  Comparative effect estimation

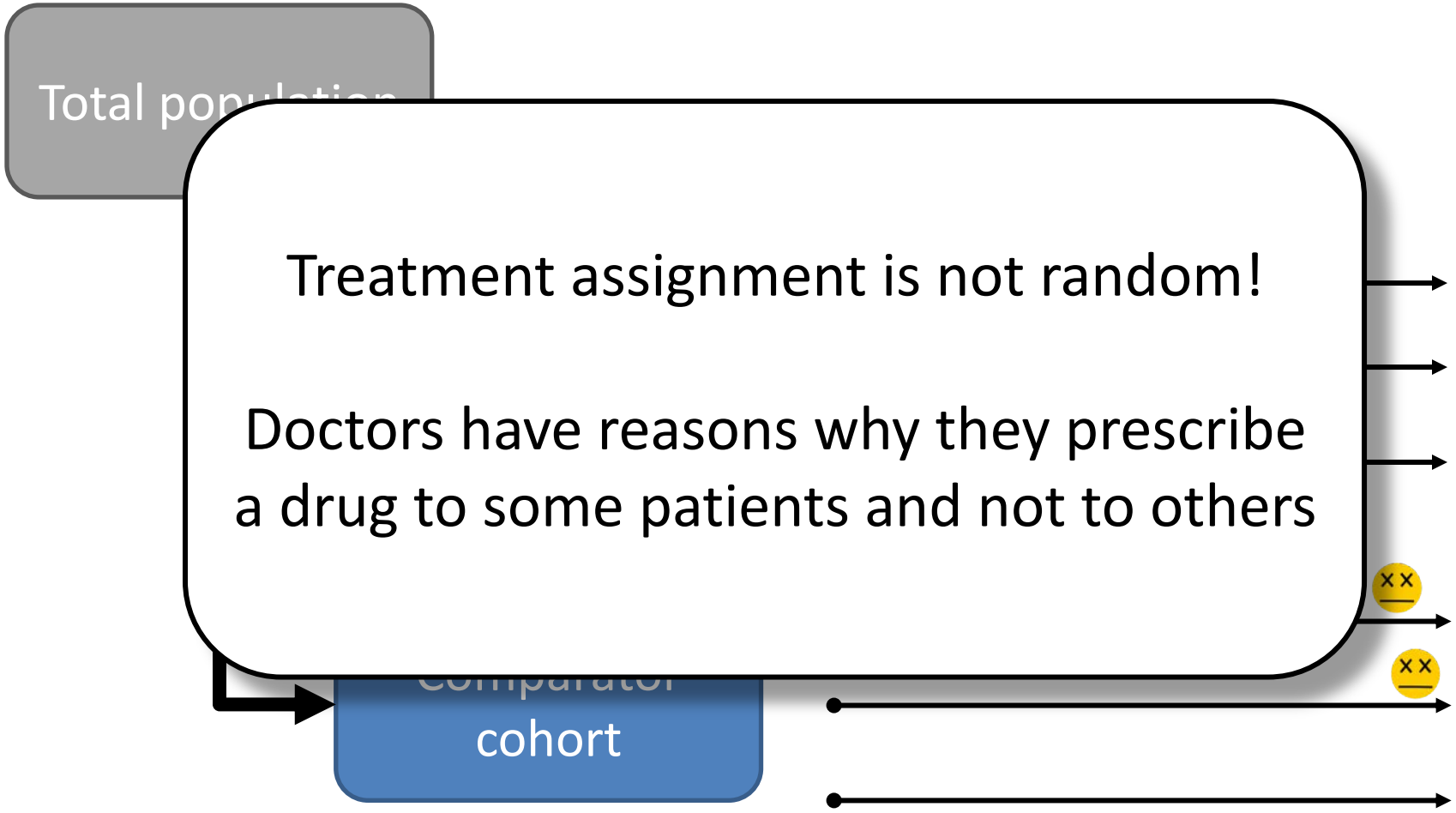
New-user cohort design



Randomized controlled trial



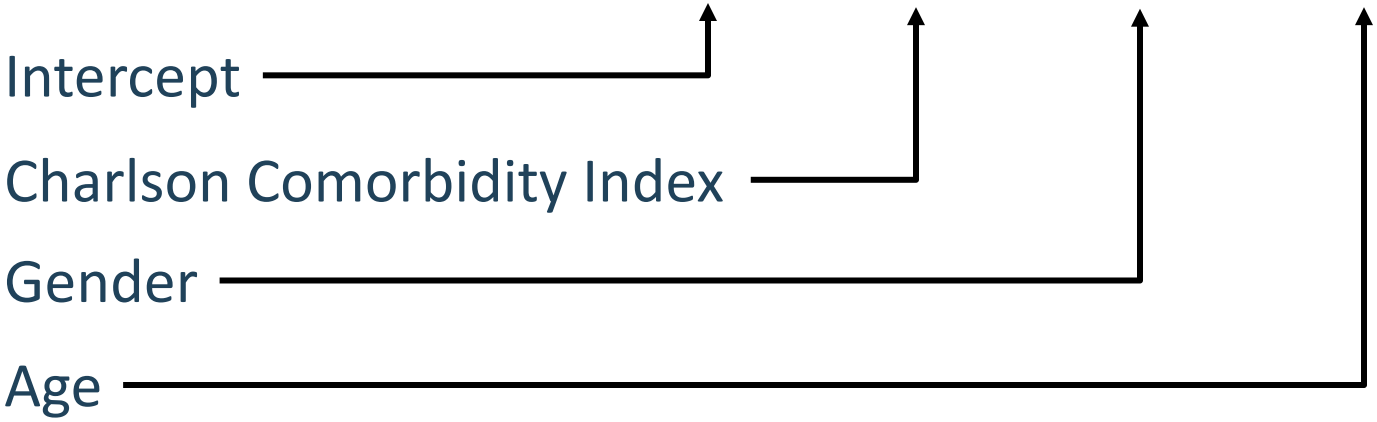
New-user cohort design



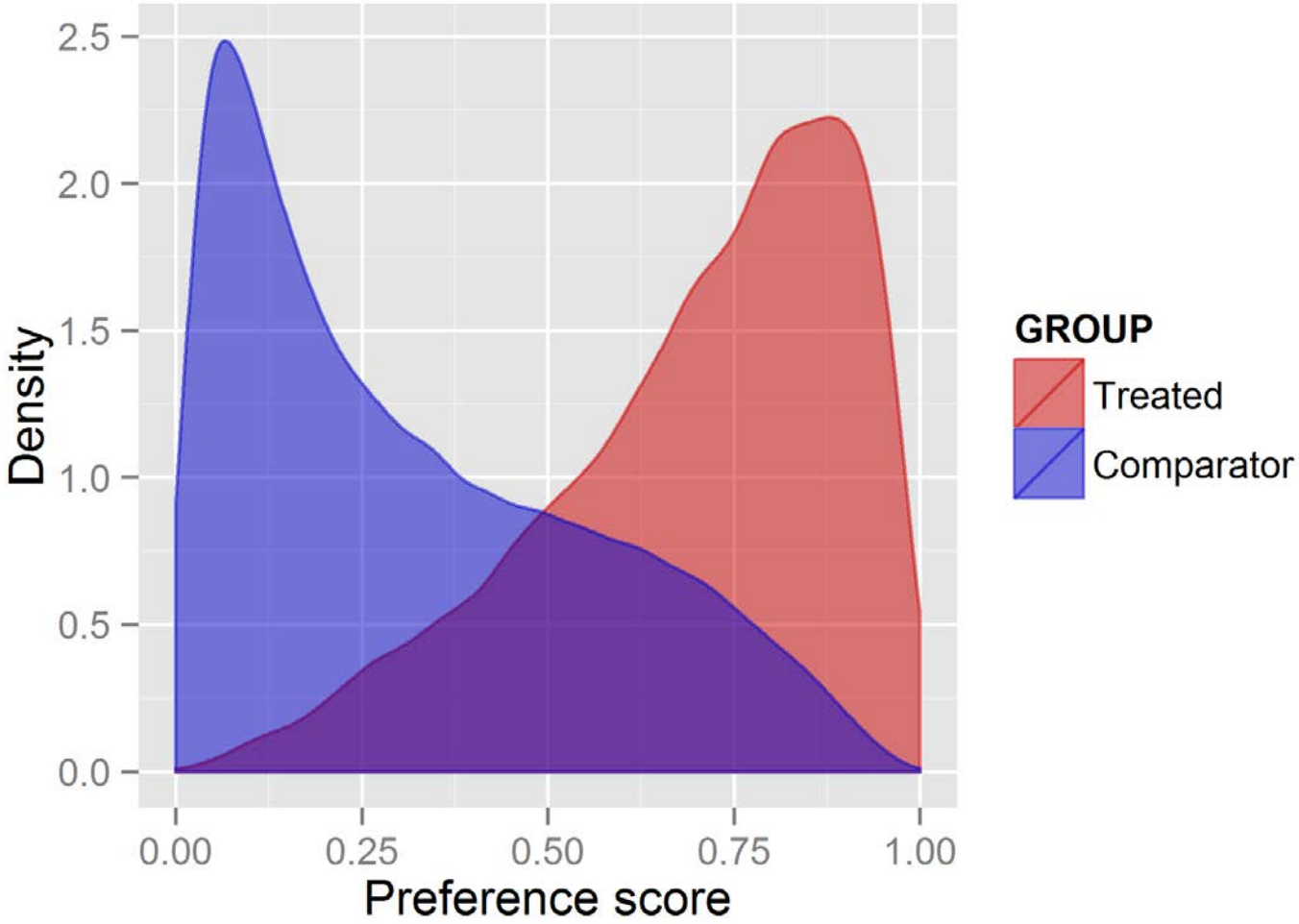
Propensity score (PS)

The propensity score is the probability of receiving the treatment, conditional on a set of baseline characteristics

$$P(\,treatment \mid X) = f(\,\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\,)$$



PS score distribution



Using the PS

- **Trimming**

  if P(treatment) is around 50%, treatment assignment ‘must be random’

- **Stratification or matching**

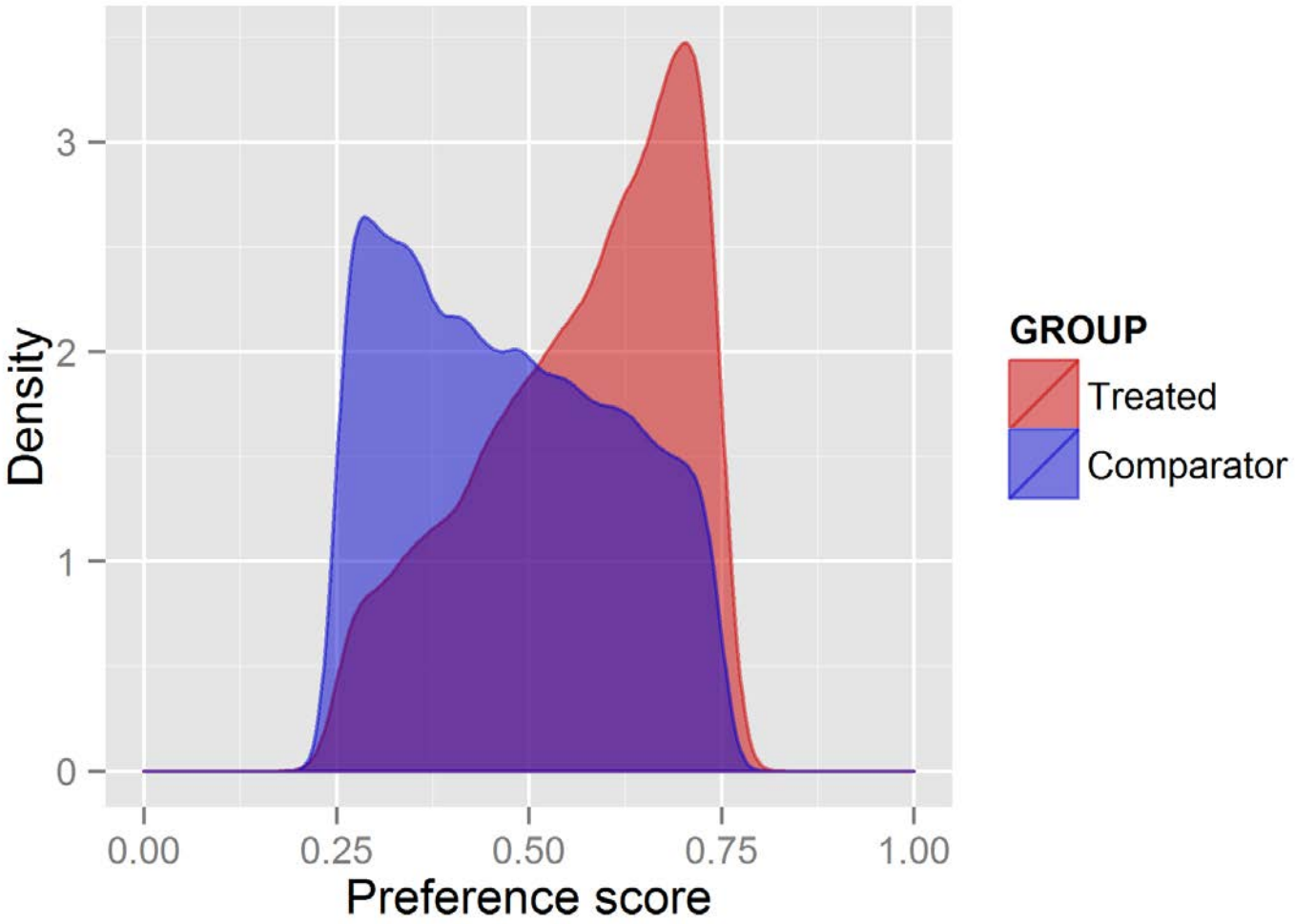
  only compare subjects to subjects with a similar PS

- **Inverse probability weighting**

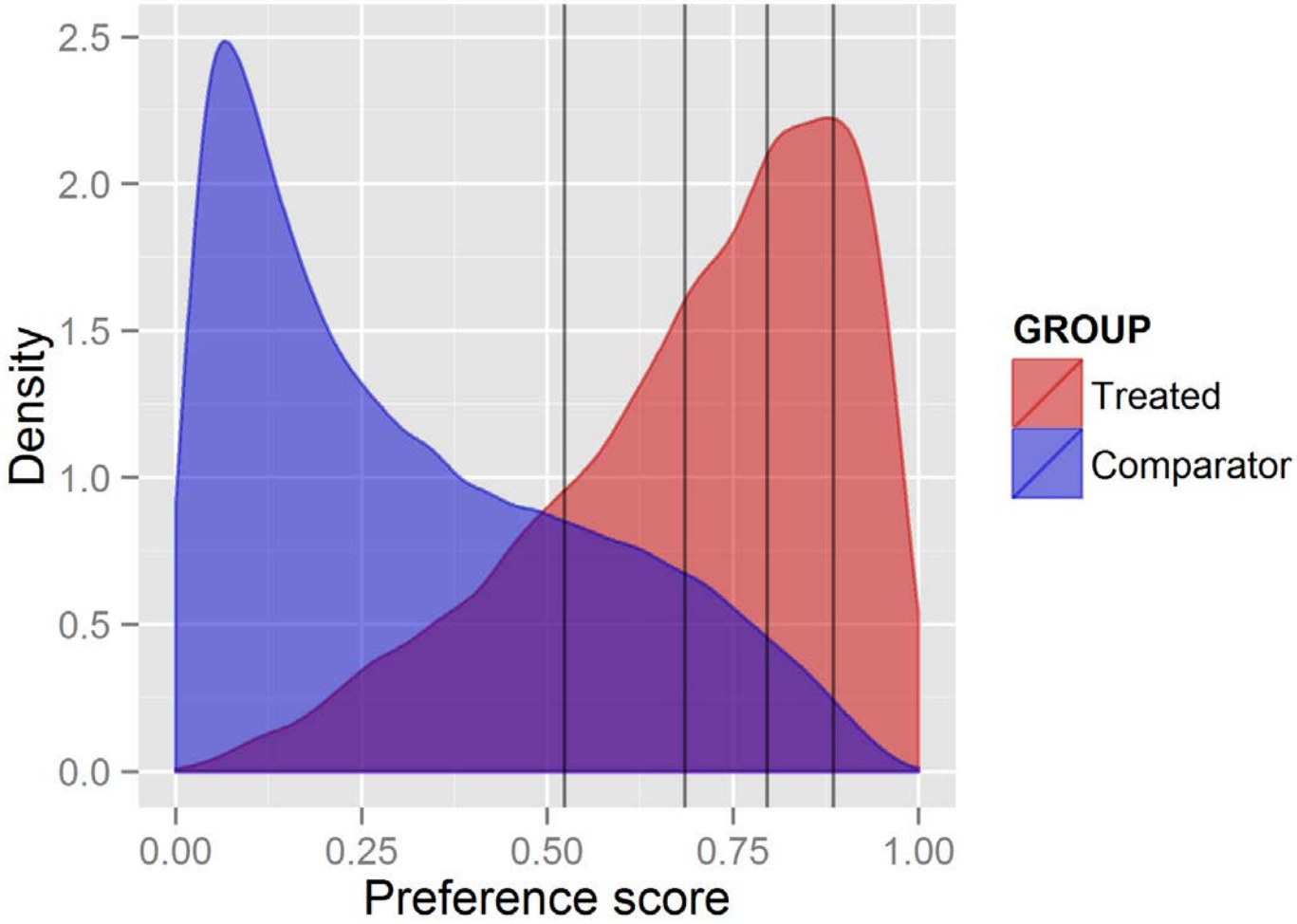
- **Adding to the outcome model**

  correct for the PS in the model used to predict the outcome

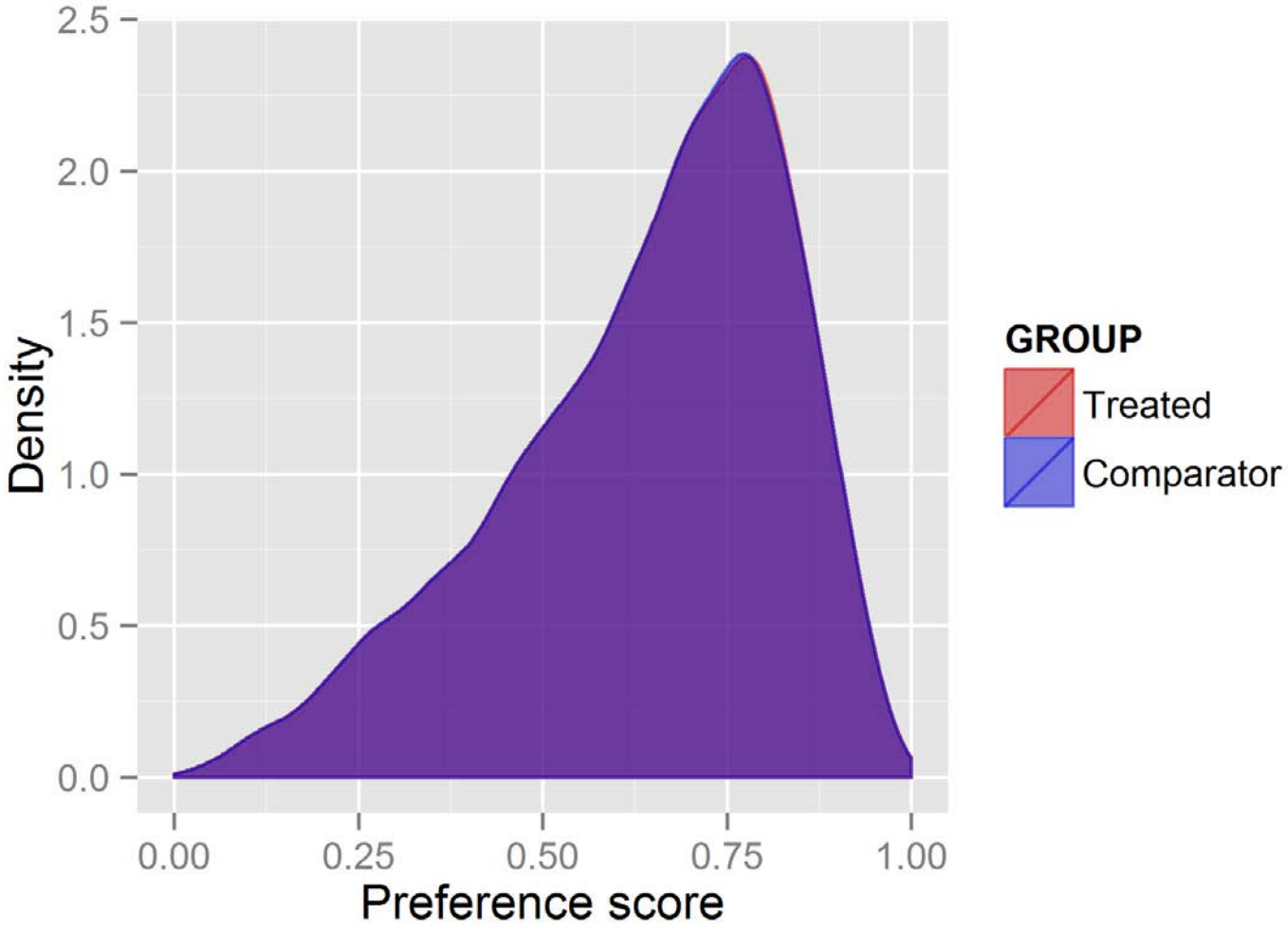
Trimming


Density
3
2
1
0
0.00
0.25
0.50
0.75
1.00
Preference score
GROUP
Treated
Comparator

Stratifying



Matching


2.5
2.0
1.5
1.0
0.5
0.0
Density
0.00
0.25
0.50
0.75
1.00
Preference score
GROUP
Treated
Comparator

Which variables go into the PS model?

- Traditional: hard thinking by expert

- High-Dimensional PS: rank many variables (e.g. all dru re (ar (H

Important: make sure not to put the exposures themselves in the model!

- Our approach: put everything (demographics, all drugs, all drug classes, all conditions, all disease classes, all procedures, all observations, all severity indexes) in a regularized regression

Types of outcome models

- Logistic

  Did the outcome occur yes/no?

- Poisson

  How many times did the outcome occur?

- Cox

  What was the time to the first outcome or end of observation?

- Conditional or non-conditional (Logistic, Poisson, Cox)

  stratify by PS strata or matched sets

Demo: implement Graham study in ATLAS

http://www.ohdsi.org/web/atlas/#/estimation/1735869

Cardiovascular, Bleeding, and Mortality Risks in Elderly Medicare Patients Treated With Dabigatran or Warfarin for Nonvalvular Atrial Fibrillation

David J. Graham, MD, MPH; Marsha E. Reichman, PhD; Michael Wernecke, BA; Rongmei Zhang, PhD; Mary Ross Southworth, PharmD; Mark Levenson, PhD; Ting-Chang Sheu, MPH; Katrina Mott, MHS; Margie R. Goulding, PhD; Monika Houstoun, PharmD, MPH; Thomas E. MaCurdy, PhD; Chris Worrall, BS; Jeffrey A. Kelman, MD, MMSc

Background—The comparative safety of dabigatran versus warfarin for treatment of nonvalvular atrial fibrillation in general practice settings has not been established.

Methods and Results—We formed new-user cohorts of propensity score–matched elderly patients enrolled in Medicare who initiated dabigatran or warfarin for treatment of nonvalvular atrial fibrillation between October 2010 and December 2012. Among 134414 patients with 37587 person-years of follow-up, there were 2715 primary outcome events. The hazard ratios (95% confidence intervals) comparing dabigatran with warfarin (reference) were as follows: ischemic stroke, 0.80 (0.67–0.96); intracranial hemorrhage, 0.34 (0.26–0.46); major gastrointestinal bleeding, 1.28 (1.14–1.44); acute myocardial infarction, 0.92 (0.78–1.08); and death, 0.86 (0.77–0.96). In the subgroup treated with dabigatran 75 mg twice daily, there was no difference in risk compared with warfarin for any outcome except intracranial hemorrhage, in which case dabigatran risk was reduced. Most patients treated with dabigatran 75 mg twice daily appeared not to have severe renal impairment, the intended population for this dose. In the dabigatran 150-mg twice daily subgroup, the magnitude of effect for each outcome was greater than in the combined-dose analysis.

Conclusions—In general practice settings, dabigatran was associated with reduced risk of ischemic stroke, intracranial hemorrhage, and death and increased risk of major gastrointestinal hemorrhage compared with warfarin in elderly patients with nonvalvular atrial fibrillation. These associations were most pronounced in patients treated with dabigatran 150 mg twice daily, whereas the association of 75 mg twice daily with study outcomes was indistinguishable from warfarin except for a lower risk of intracranial hemorrhage with dabigatran. (***Circulation***. **2015;131:157-164. DOI: 10.1161/CIRCULATIONAHA.114.012061.)**

Key Words: anticoagulant ■ pharmacoepidemiology ■ safety ■ thrombin inhibitor ■ warfarin

What is the design used by Graham et al?

Input parameter	Design choice
Target cohort (T)	dabigatran new users with prior atrial fibrillation
Comparator cohort (C)	warfarin new users with prior atrial fibrillation
Outcome cohort (O)	Ischemic stroke
Time-at-risk	1 day after cohort start → cohort end
Model specification	1:1 propensity score-matched univariable conditional Cox proportional hazards

Graham et al. description of the covariate adjustment strategy

To reduce confounding due to imbalance in study covariates, propensity score matching was used.[14–16] Unconditional logistic regression was used to estimate the predicted probability of patients initiating dabigatran therapy given their sociodemographic characteristics, baseline medical comorbidities, medications used during the preceding 6 months, prescriber characteristics, and other potentially relevant variables (Table 1 and Table I in the online-only Data Supplement). Dabigatran users were propensity score matched to warfarin users in a 1:1 ratio with the use of a greedy matching algorithm. The balance of measured covariates between the matched cohorts was assessed with the standardized mean difference, a measure not influenced by sample size and thus useful for comparing cohorts in large observational studies.[17] A standardized mean difference of ≤0.1 indicates a negligible difference in the measured variables between groups.[17]

Graham et al. replication: Designing the covariate adjustment strategy in ATLAS

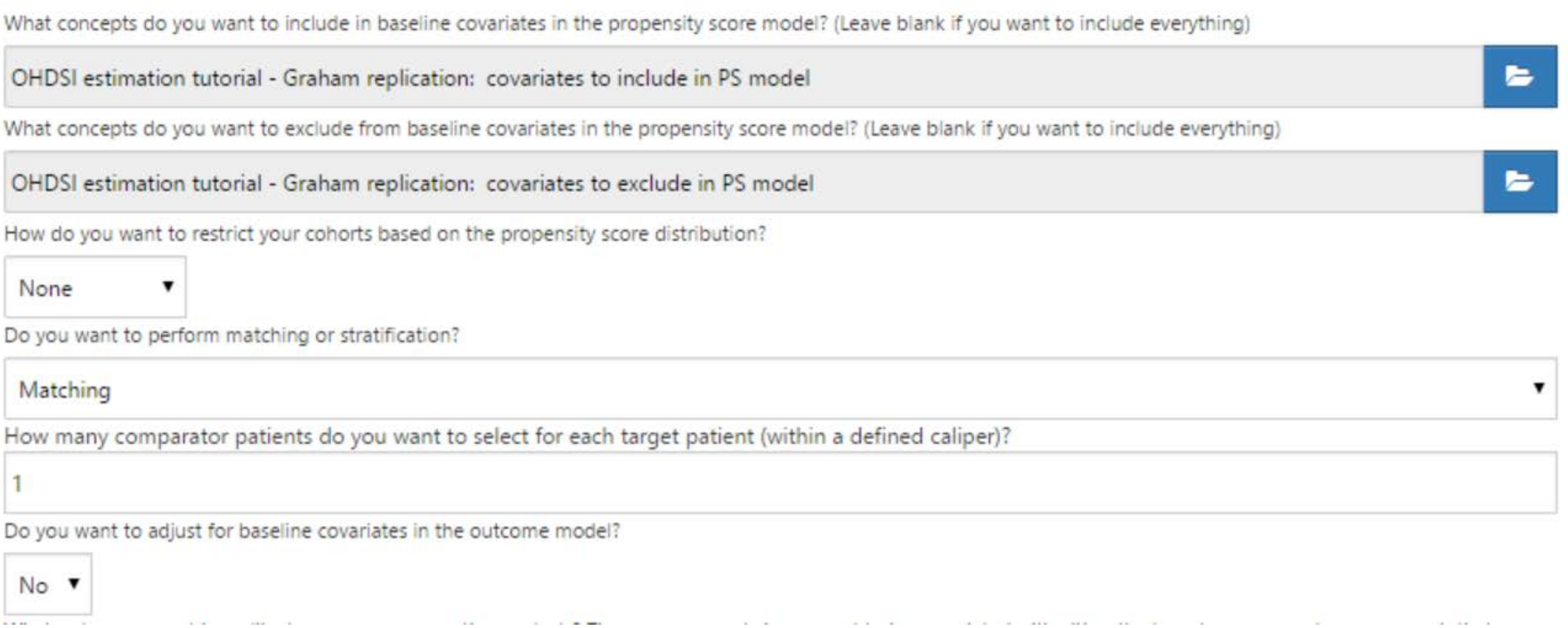
Use propensity score adjustment as a confounding adjustment strategy for baseline covariates?

Yes ▾

Which types of baseline covariates do you want to include in the propensity score model?

- [x] Demographics
  - [x] Gender
  - [x] Age group (5-year bands)
  - [ ] Index year
  - [ ] Index month
  - [x] Race
  - [x] Ethnicity
- [x] Conditions
  - [x] In prior 30d
  - [x] In prior 365d
  - [x] In prior 180d within inpatient setting
  - [ ] All time prior
  - [ ] Overlapping index date
- [x] Condition aggregation
  - [ ] SNOMED
  - [x] MedDRA
- [x] Drugs
  - [ ] In prior 30d
  - [x] In prior 365d
  - [ ] All time prior
  - [ ] Overlapping index date
- [x] Drug aggregation
  - [ ] Clinical Drug
  - [x] Ingredient
  - [x] ATC Class
- [ ] Procedures
  - [ ] In prior 30d
  - [ ] In prior 365d
- [ ] Measurement
  - [ ] Existence in prior 30d
  - [ ] Existence in prior 365d
  - [ ] Count in prior 365d
  - [ ] Has latest prior numeric value below normal range
  - [ ] Has latest prior numeric value above normal range
- [x] Risk scores
  - [ ] Charlson
  - [x] CHADS2
  - [ ] DCSI
- [ ] Concept counts (count of distinct conditions/procedures/visits in history)
- [ ] Interaction terms
  - [ ] By index year
  - [ ] By index month

Graham et al. replication: Designing the covariate adjustment strategy in ATLAS



Graham et al. replication: Designing a protocol in ATLAS

Population Level Effect Estimation

OHDSI estimation tutorial: Graham replication: dabigatran vs warfarin for risk of ischemic str | Save | Close

Specification | Export

Print Friendly | R Code

Research question

To compare the risk of **OHDSI estimation tutorial: Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting** between **OHDSI estimation tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation** and **OHDSI estimation tutorial: Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation**, we will estimate the population-level effect of exposure on the hazards of the outcome during the period from 1 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: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation** with **OHDSI estimation tutorial: Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation** for the hazards of **OHDSI estimation tutorial: Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting** from 1 days from cohort start date to 0 days from cohort end date.

For both cohorts, we impose a requirement that patients must have at least 1 days of continuous observation after the time-at-risk start, 1 days from cohort start date.

The overall study population could be considered to be patients who entered either the target cohort or comparator cohort.

The time-to-event of outcome among patients in the target and comparator cohorts is determined by calculating the number of days from the start of the time-at-risk window, 1 days from cohort start date until the earliest event among 1) the first occurrence of the outcome, **OHDSI estimation tutorial: Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting** before 0 days from cohort end date, 2) the end of the time-at-risk window, 0 days from cohort end date, and 3) the end of the observation period that spans the time-at-risk start.

Patients with **OHDSI estimation tutorial: Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting** prior to target or comparator cohort entry were excluded from consideration.

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)
  - Race
  - Ethnicity

Graham et al. replication: Designing a protocol in ATLAS (2)

- Conditions
  - In prior 30d
  - In prior 365d
  - In prior 180d within inpatient setting
- Condition aggregation
  - MedDRA
- Drugs
  - In prior 365d
- Drug aggregation
  - Ingredient
  - ATC Class
- Risk scores
  - CHADS2

Specific covariates to be included in the propensity score model are labelled **OHDSI estimation tutorial - Graham replication: covariates to include in PS model** as detailed in Appendix 2.

Specific covariates to be excluded from the propensity score model are labelled **OHDSI estimation tutorial - Graham replication: covariates to exclude in PS model** as detailed in Appendix 2.

All covariates that occur in fewer than 10 persons between the target and comparator cohorts combined will be excluded prior to model fitting.

Patients in the target cohort will be matched to patients in the comparator cohort, using a maximum ratio of 1:1 matching, using a greedy matching algorithm, applying a caliper of 0.25 of the standard deviation of the propensity score distribution.

Output and evaluation

Covariate balance will be summarized in tabular form by showing the mean value for all baseline covariates in the target and comparator cohort, with the associated standardized mean difference computed for each covariate.

Once the propensity score model is fit, we will plot the propensity score distribution of the target and comparator cohorts to evaluate the comparability of the two cohorts. The plot will be scaled to the preference score, normalizing for any imbalance in cohort size. The area under the Receiver Operating Characteric (ROC) curve (AUC) will be reported. The covariates selected within the propensity score model, with associated coefficients will also be reported.

A plot showing the propensity score distributions for both cohorts after matching will be provided. Covariate balance will be evaluated by plotting the standardized mean difference of each covariate before propensity score matching against the standardized mean difference for each covariate after propensity score matching.

An attrition diagram will be provided to detail the loss of patients from the original target cohort, **OHDSI estimation tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation**, and comparator cohort **OHDSI estimation tutorial: Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation** to the subpopulations that remain after all design considerations have been applied.

The final outcome model, a conditional Cox proportional hazards model, will be summarized by providing the hazards ratio and associated 95% confidence interval. The number of persons, amount of time-at-risk, and number of outcomes in each cohort will also be reported.

A Kaplan-Meier plot will be generated to characterize the contour of risk over time for the outcome of interest.

Appendix 1: Cohort Definitions

Graham et al. replication: Designing a protocol in ATLAS (3)

Appendix 1: Cohort Definitions

The target cohort **OHDSI estimation tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation** is defined as:

Initial Event Cohort

People having any of the following:

- a drug era of dabigatran[4]
  - for the first time in the person's history
  - era start is on or after 2010-10-19
  - with age at era start >= 65

with continuous observation of at least 183 days prior and 0 days after event index date, and limit initial events to: **earliest event per person.**

Limit cohort of initial events to: **earliest event per person.**

Inclusion Criteria #1: Has prior atrial fibrillation of atrial flutter diagnosis

People having any of the following criteria:

- at least 1 occurrences of a condition occurrence of Atrial fibrillation[2]
  occurring between all days Before and 0 days After event index date
- or at least 1 occurrences of a condition occurrence of Atrial flutter[3]
  occurring between all days Before and 0 days After event index date

Inclusion Criteria #2: Has no prior treatment with comparator drug (warfarin)

People having all of the following criteria:

- exactly 0 occurrences of a drug exposure of warfarin[13]
  occurring between all days Before and 0 days Before event index date

Inclusion Criteria #3: Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)

People having all of the following criteria:

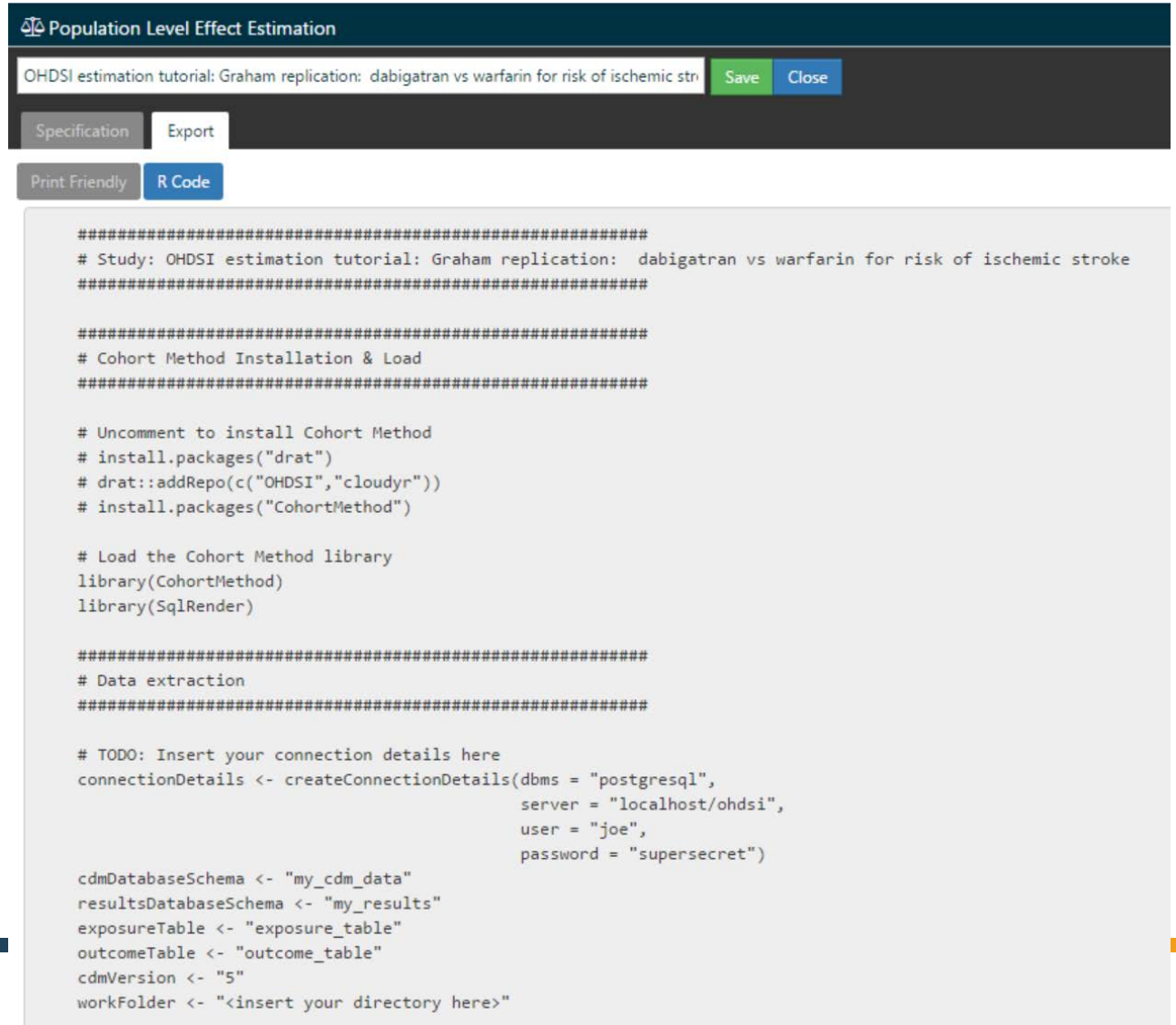
- exactly 0 occurrences of a drug exposure of rivaroxaban[12]
  occurring between all days Before and 0 days After event index date
- and exactly 0 occurrences of a drug exposure of apixaban[1]
  occurring between all days Before and 0 days After event index date

Inclusion Criteria #4: Not in a skilled nursing facility or nursing home, or receiving hospice care on the index date

People having all of the following criteria:

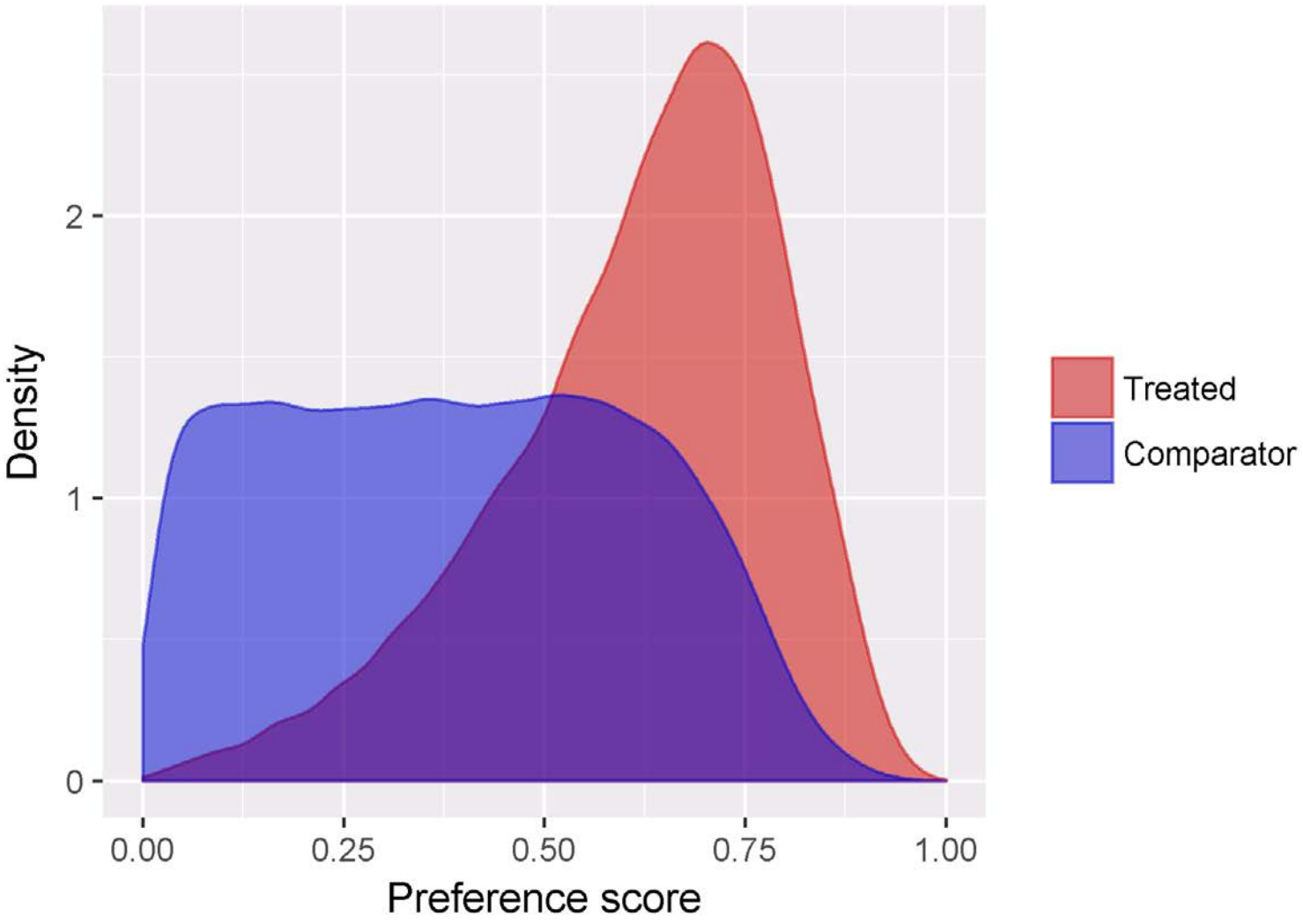
- exactly 0 occurrences of a visit occurrence of long term care visit[10]
  occurring between 0 days Before and 0 days After event index date
- and exactly 0 occurrences of a procedure of Hospice observations[9]
  occurring between all days Before and 0 days After event index date

Graham et al. replication: Designing the source code in ATLAS

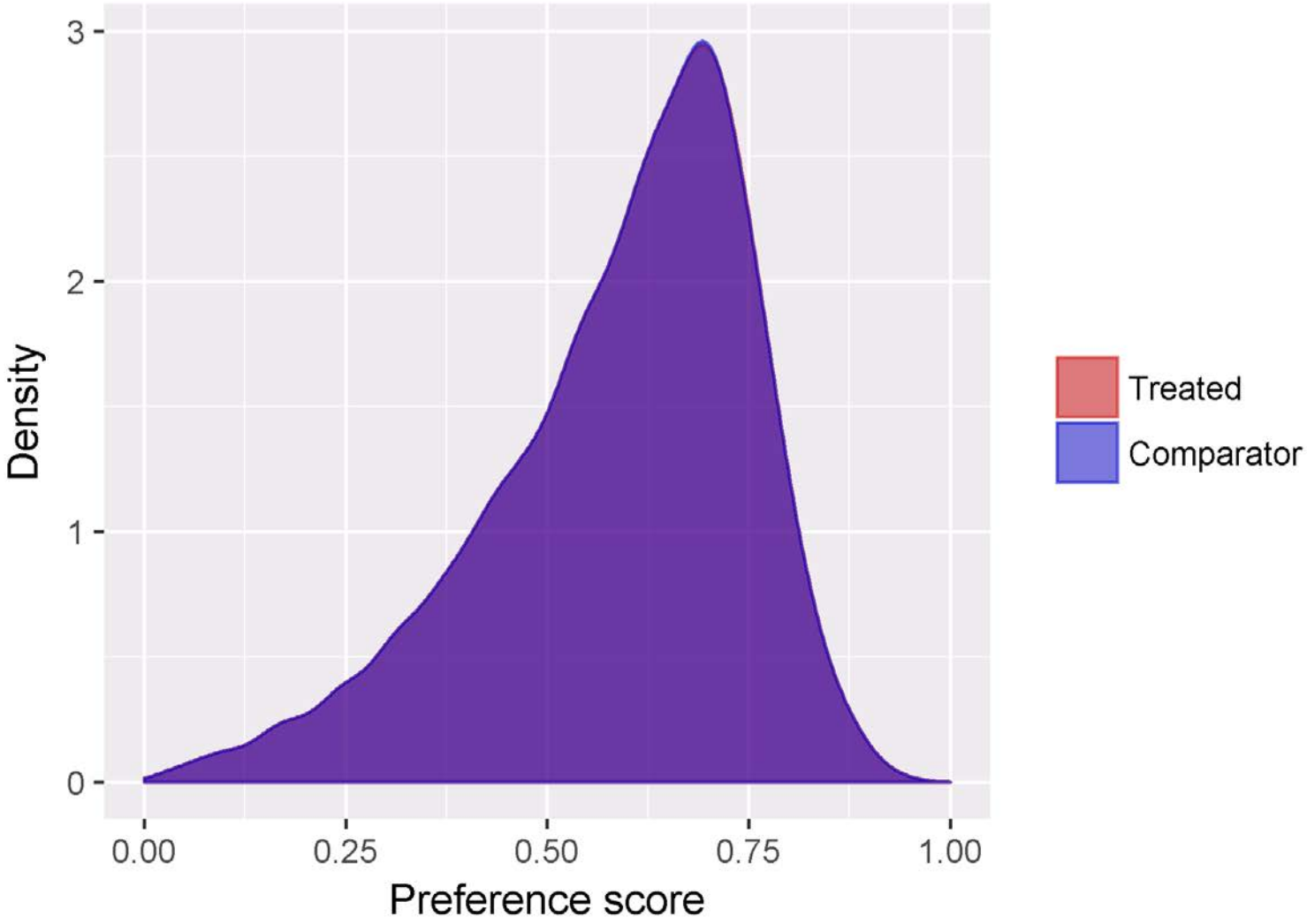


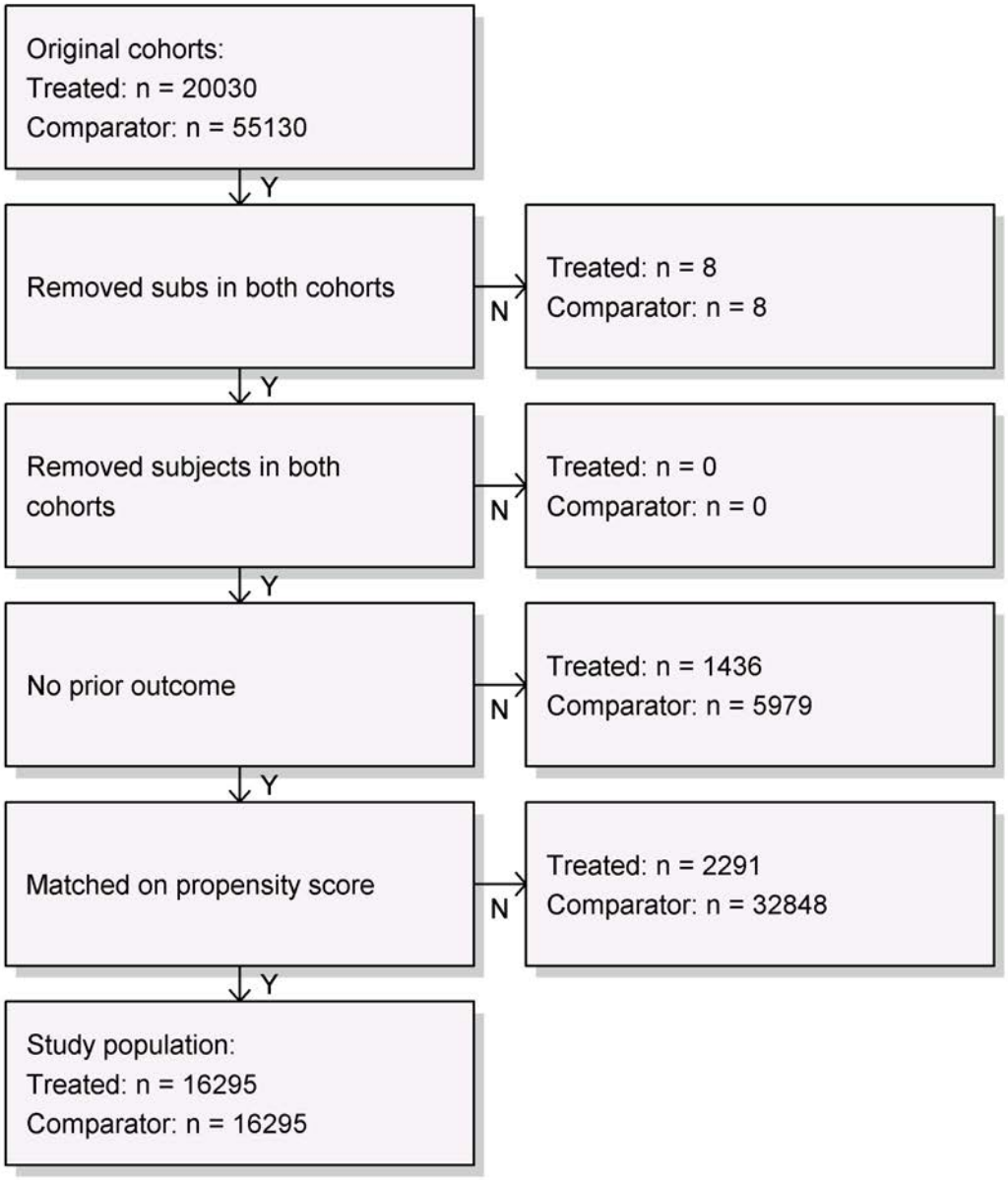
Interpreting results

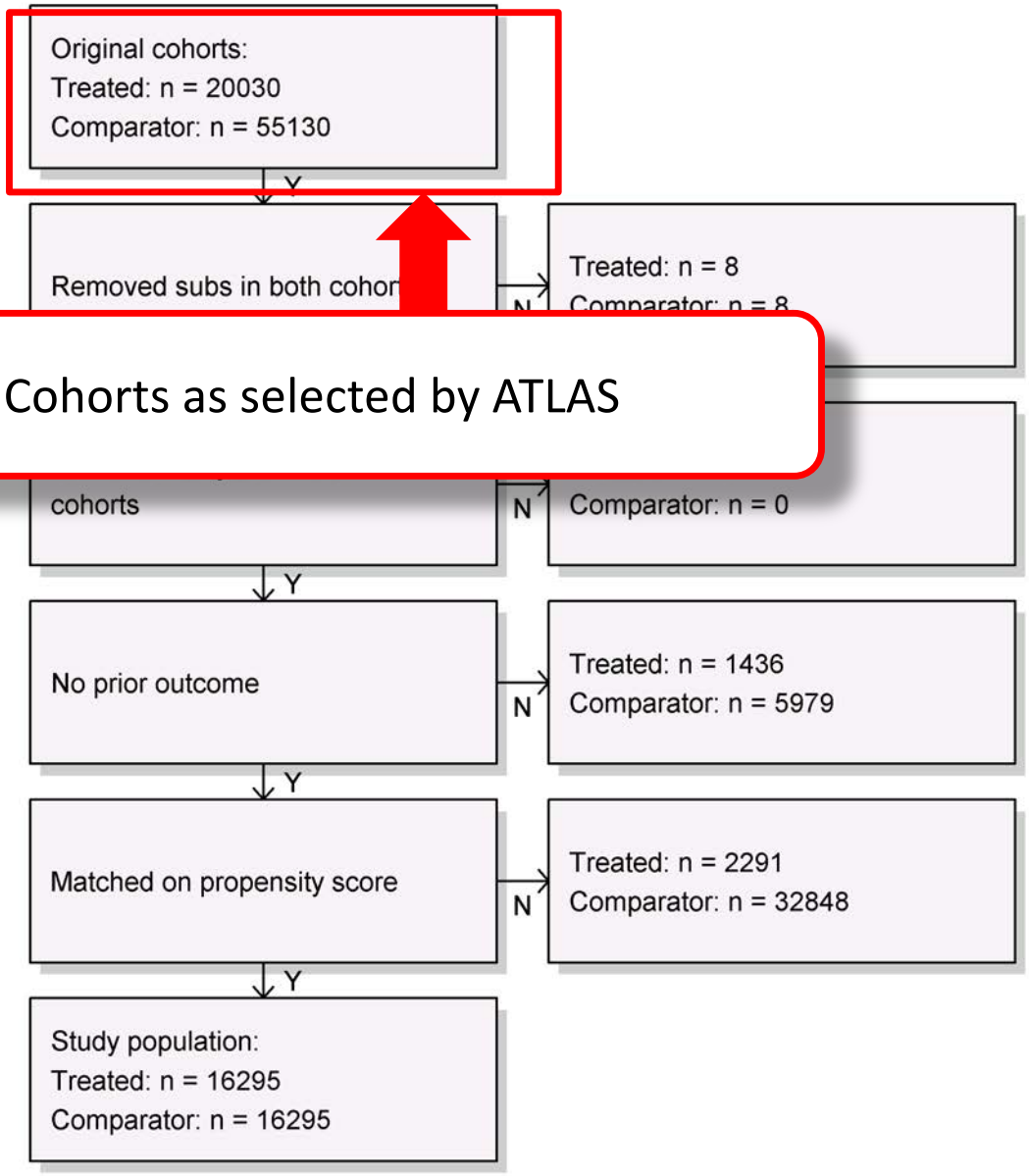
Plot propensity score distribution

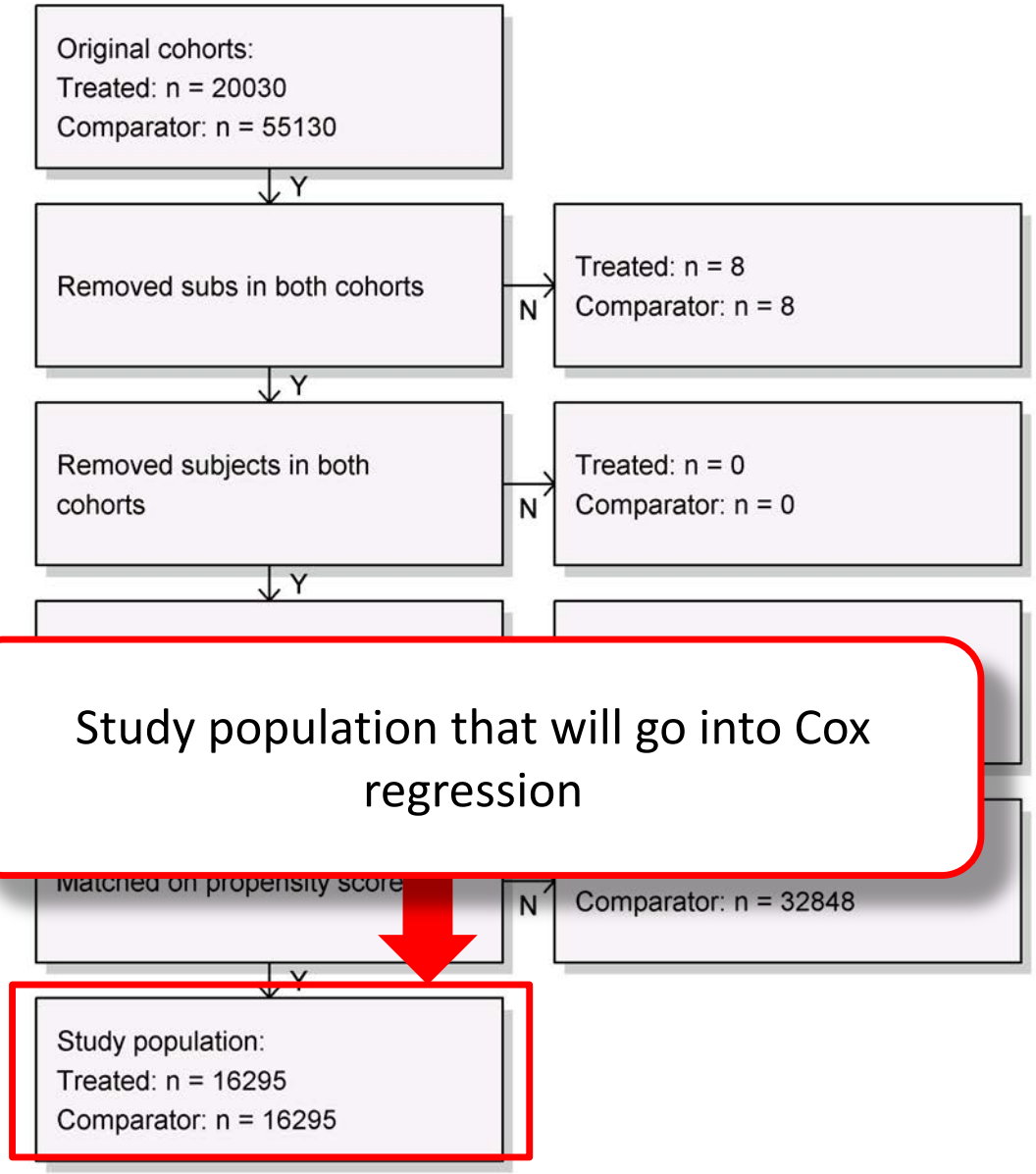


After matching

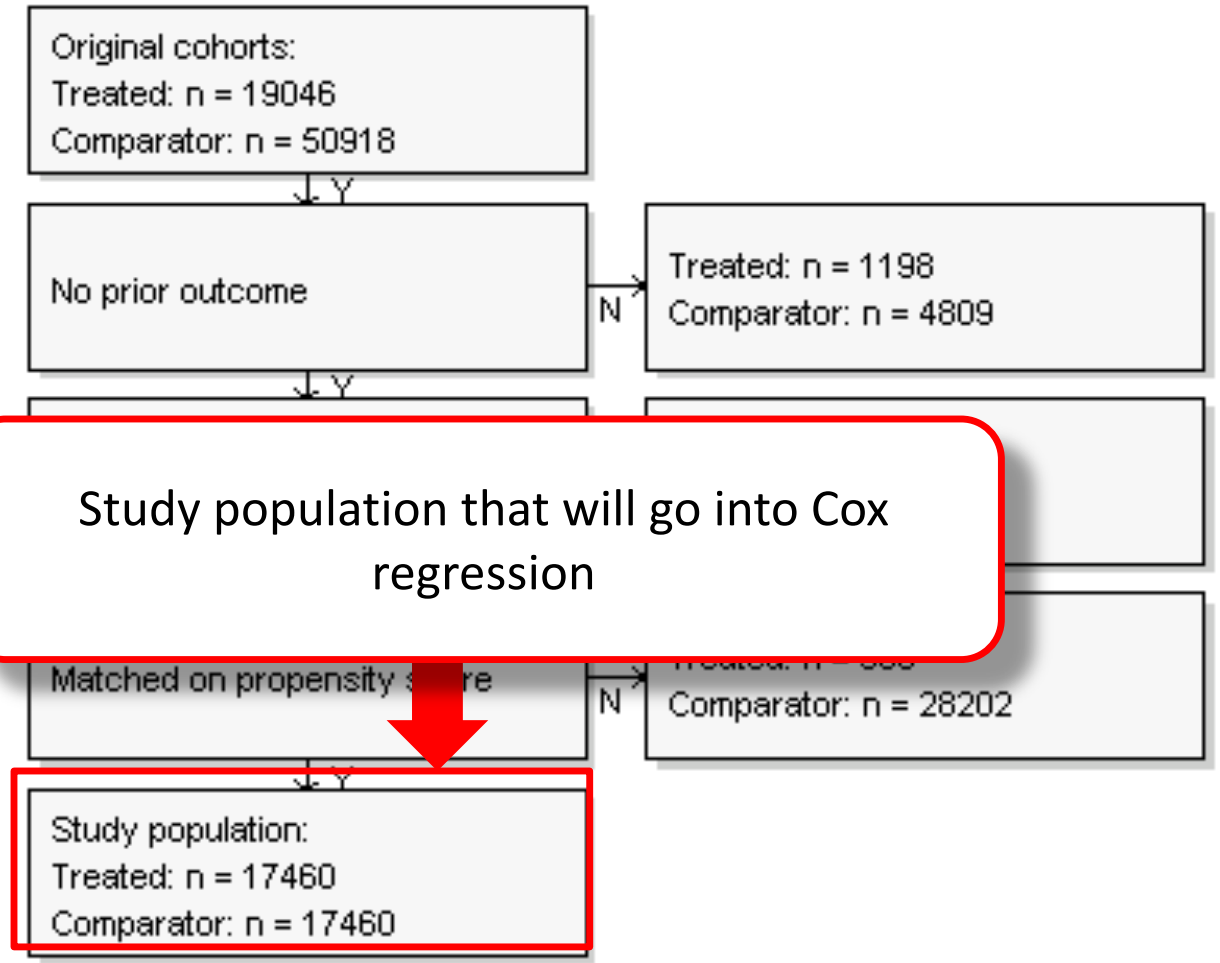



Original cohorts:
Treated: n = 20030
Comparator: n = 55130
Y
Removed subs in both cohorts
N
Treated: n = 8
Comparator: n = 8
Y
Removed subjects in both cohorts
N
Treated: n = 0
Comparator: n = 0
Y
No prior outcome
N
Treated: n = 1436
Comparator: n = 5979
Y
Matched on propensity score
N
Treated: n = 2291
Comparator: n = 32848
Y
Study population:
Treated: n = 16295
Comparator: n = 16295

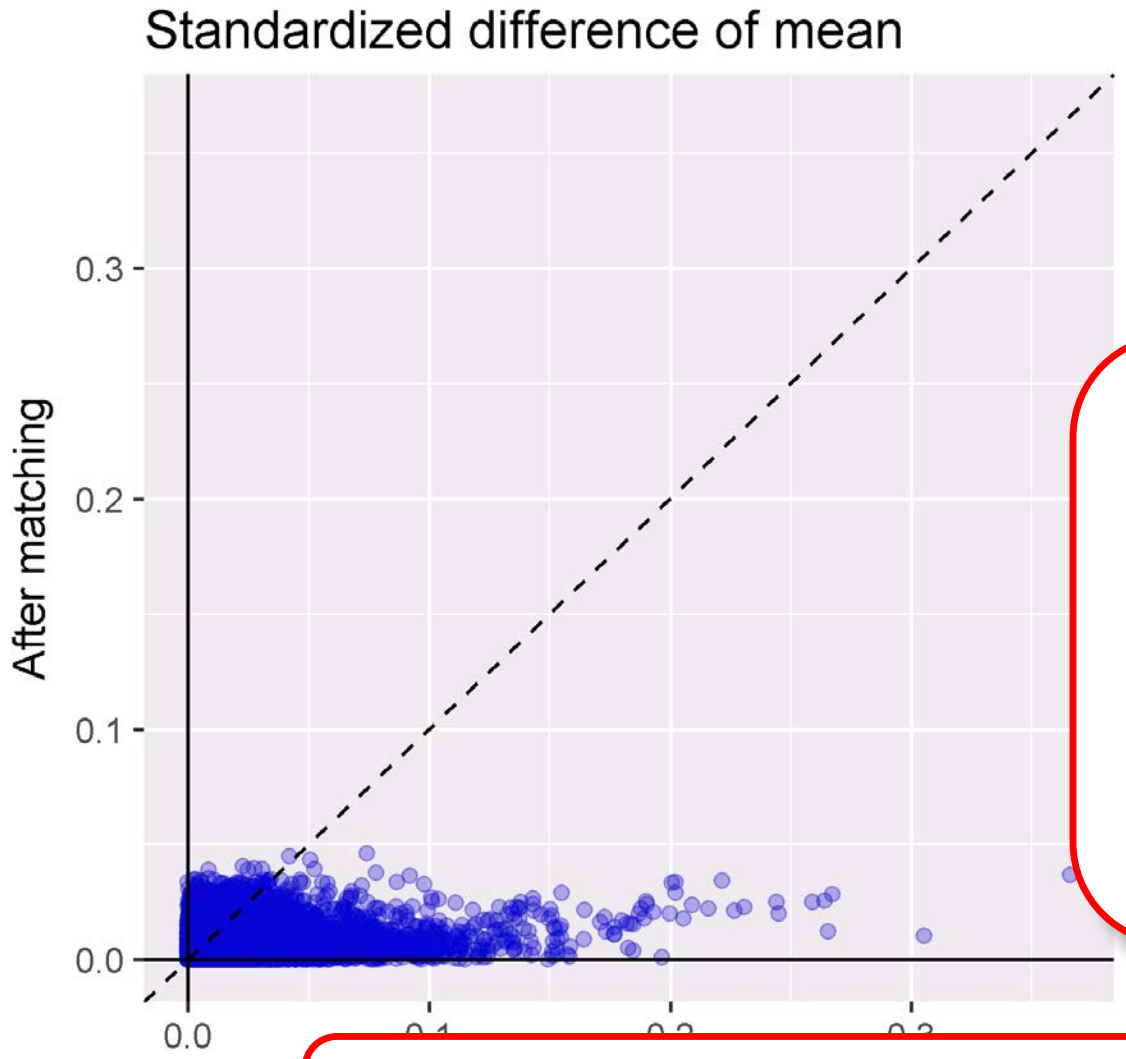

Original cohorts:
Treated: n = 20030
Comparator: n = 55130
Y
Removed subs in both cohor
N
Treated: n = 8
Comparator: n = 8
Cohorts as selected by ATLAS
cohorts
N
Comparator: n = 0
Y
No prior outcome
N
Treated: n = 1436
Comparator: n = 5979
Y
Matched on propensity score
N
Treated: n = 2291
Comparator: n = 32848
Y
Study population:
Treated: n = 16295
Comparator: n = 16295


Original cohorts:
Treated: n = 20030
Comparator: n = 55130
Y
Removed subs in both cohorts
N
Treated: n = 8
Comparator: n = 8
Y
Removed subjects in both cohorts
N
Treated: n = 0
Comparator: n = 0
Y
Study population that will go into Cox regression
N
Comparator: n = 32848
Y
Study population:
Treated: n = 16295
Comparator: n = 16295

Attrition


Original cohorts:
Treated: n = 19046
Comparator: n = 50918
Y
No prior outcome
N
Treated: n = 1198
Comparator: n = 4809
Y
Study population that will go into Cox regression
Matched on propensity
N
Comparator: n = 28202
Y
Study population:
Treated: n = 17460
Comparator: n = 17460

Covariate balance



Most covariates are binary:

$$\frac{\text{abs}(P_{\text{target group}} - P_{\text{comparator group}})}{\text{standard deviation}}$$

Graham: "A standardized mean difference of ≤0.1 indicates a negligible difference."

Inspect the outcome model

Model type: cox
Stratified: TRUE
Use covariates: FALSE
Use inverse probability of treatment weighting: FALSE
Status: OK

	Estimate	lower .95	upper .95	logRr	seLogRr
treatment	0.78000	0.51050	1.18316	-0.24846	0.2144

Population counts

	treatedPersons	comparatorPersons	treatedExposures	comparatorExposures
Count	16295	16295	16295	16295

Outcome counts

	treatedPersons	comparatorPersons	treatedExposures	comparatorExposures
Count	81	73	81	73

Time at risk

	treatedDays	comparatorDays
Days	1682365	1515233

Inspect the outcome model

Model type: cox
Stratified: TRUE
Use covariates: FALSE
Use inverse probability of treatment weighting: FALSE
Status: OK

	Estimate	lower .95	upper .95	logRr	seLogRr
treatment	0.78000	0.51050	1.18316	-0.24846	0.2144

Point estimate and 95% confidence interval

Population counts

	treatedPersons	comparatorPersons	treatedExposures	comparatorExposures
C				16295

			edExposures	comparatorExposures
Count	81	73	81	73

Time at risk

	treatedDays	comparatorDays
Days	1682365	1515233

Inspect the outcome model

Model type: cox
Stratified: TRUE
Use covariates: FALSE
Use inverse probability of treatment weighting: FALSE
Status: OK

	Estimate	lower .95	upper .95	logRr	seLogRr
treatment	0.78000	0.51050	1.18		

Population counts

	treatedPersons	com
Count	16295	1629

Outcome counts

	treatedPersons	comparatorPersons	treat Exposures	comparatorExposures
Count	81	73	81	73

Time at risk

	treatedDays	comparatorDays
Days	1682365	1515233

Target group (dabigatran) has more outcomes, but also more time at risk

Inspect the outcome model

Model type: cox
Stratified: TRUE
Use covariates: FALSE
Use inverse probability of treatment weighting: FA
Status: OK

	Estimate	lower .95	upper .95	logRr
treatment	0.78000	0.51050	1.18316	

Population counts

	treatedPersons	comparatorPers
Count	16295	16295

Outcome counts

	treatedPersons	comparatorPersons	treat Exposures	comparatorExposures
Count	81	73	81	73

Time at risk

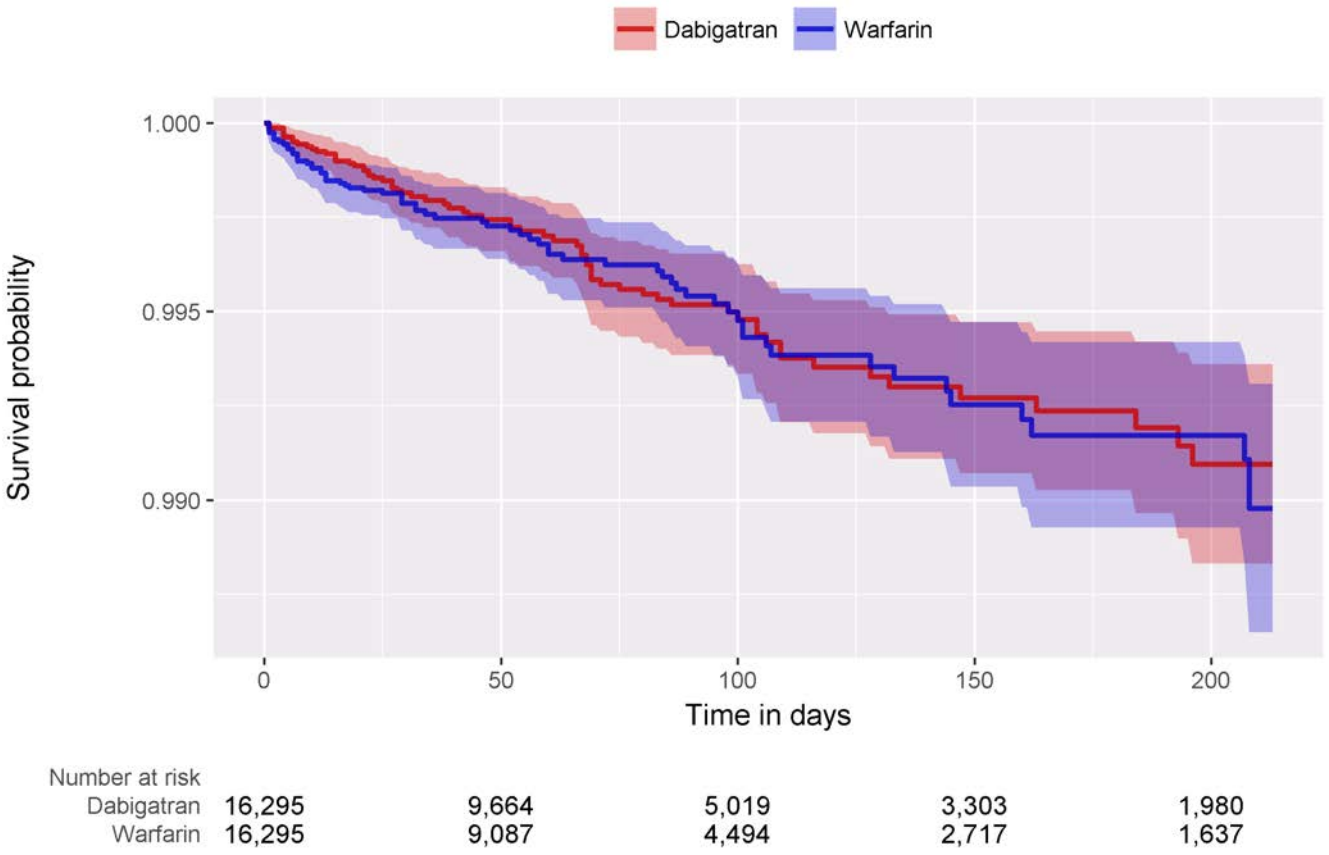
	treatedDays	comparatorDays
Days	1682365	1515233

Graham:

$IR_{dabigatran}$ = 11.3
$IR_{warfarin}$ = 13.9
$HR_{adjusted}$ = 0.80 (0.67–0.96)

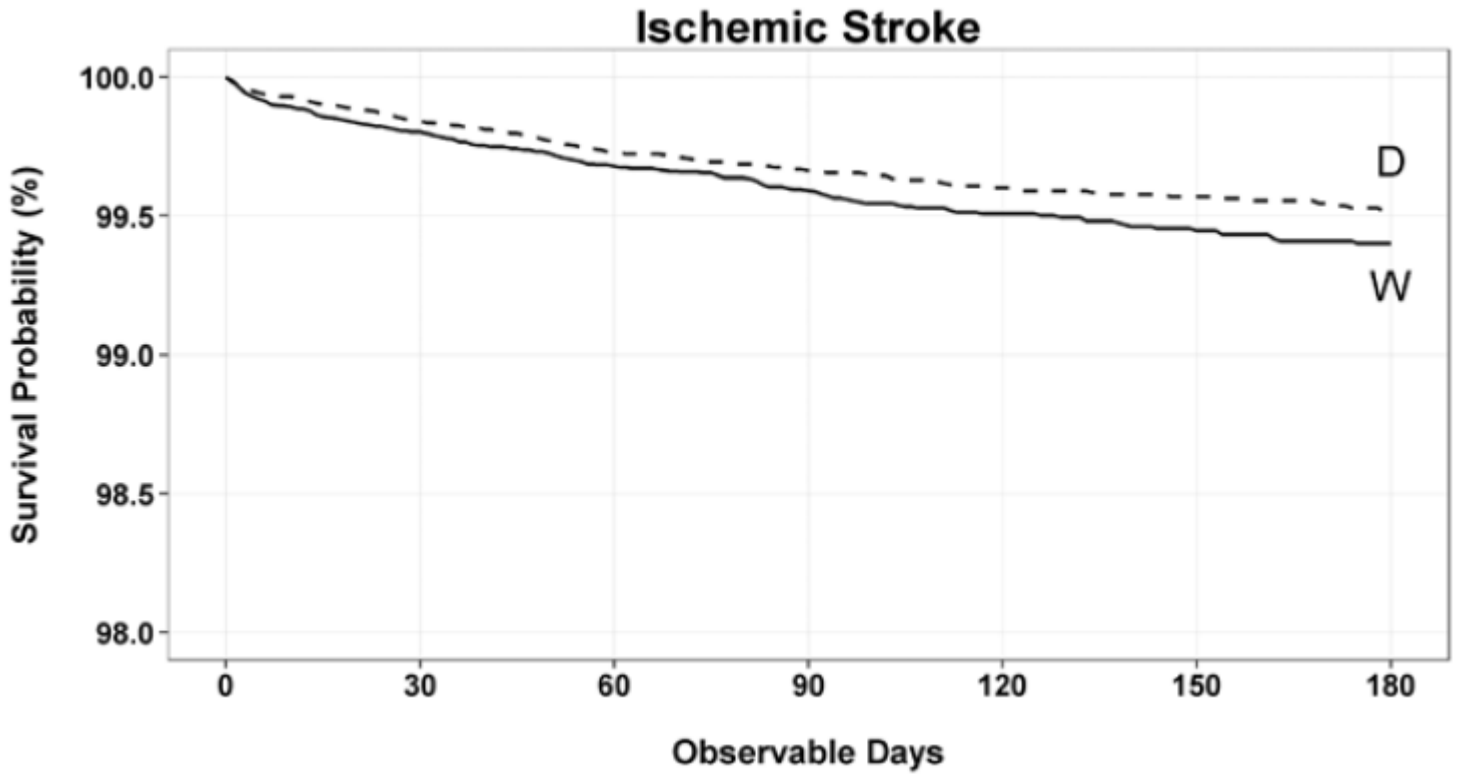
$IR_{dabigatran}$ = 17.6
$IR_{warfarin}$ = 17.6
$HR_{adjusted}$ = 0.78 (0.51 – 1.18)

Kaplan Meier plot



Kaplan Meier plot

Graham:



In conclusion

- ATLAS can
  - Write protocol
  - Generate R code to do a study
- Not shown:
  - Include negative controls & calibrate P-value
  - Synthesize positive controls & calibrate CI
  - Multiple T, C, O
  - Multiple analyses
- Other study designs available in R
  - Self-controlled case series
  - Case-crossover & case-time-control
  - Case-control
  - Self-controlled cohort

Prediction: Patient-level predictive modeling and evaluation

Installing the R Package

Instructions found on the github:
https://github.com/OHDSI/PatientLevelPrediction

1. On Windows, make sure RTools is installed.
2. The DatabaseConnector and SqlRender packages require Java. Java can be downloaded from http://www.java.com.
3. Random forest, Naive Bayes and MLP require python 3.6. Python 3.6 can be downloaded from: https://www.continuum.io/downloads.
4. In R, use the following commands to download and install PatientLevelPrediction:

```
install.packages("drat")
drat::addRepo("OHDSI")
install.packages("PatientLevelPrediction")
```

5. We recommend testing your instalation by running:

```
PatientLevelPrediction::checkPlpInstallation()
```

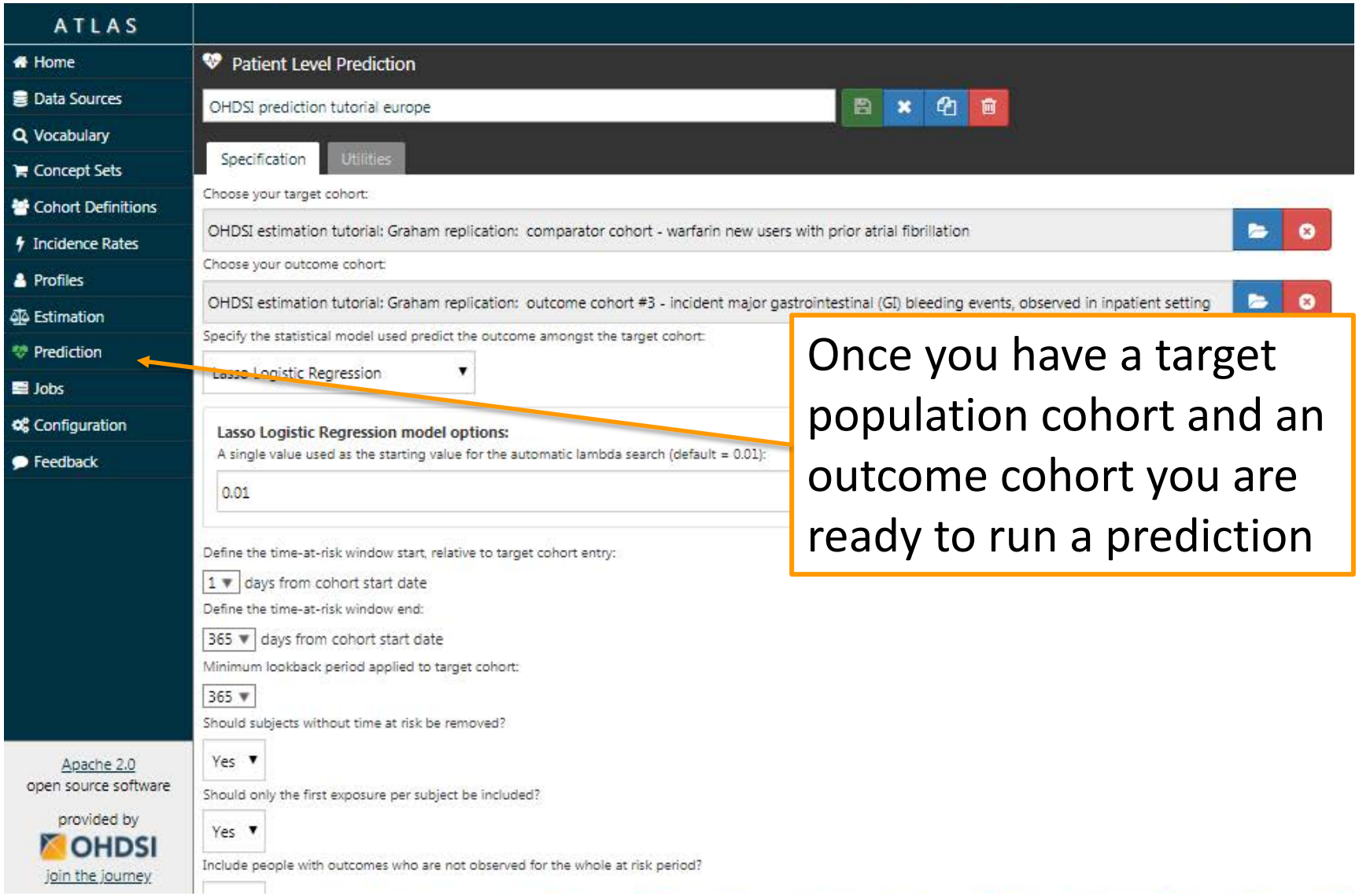
If you have a response other than 1 (indicating everything works), enter the response number in:

```
PatientLevelPrediction::interpretInstallCode()
```

Non-windows users: Please note that the package uses python to implement some of the classifiers. The package pythonInR is used as the interface, and in Linux or Mac OS it uses the same python specified in path (the python that loads when you type the command python). Please make sure the anaconda python is specified in your path rather than any default python (unless it is set up with the following packages), as the packages: numpy, scikit-learn and tensorFlow are required to run the patient level prediciton python code.

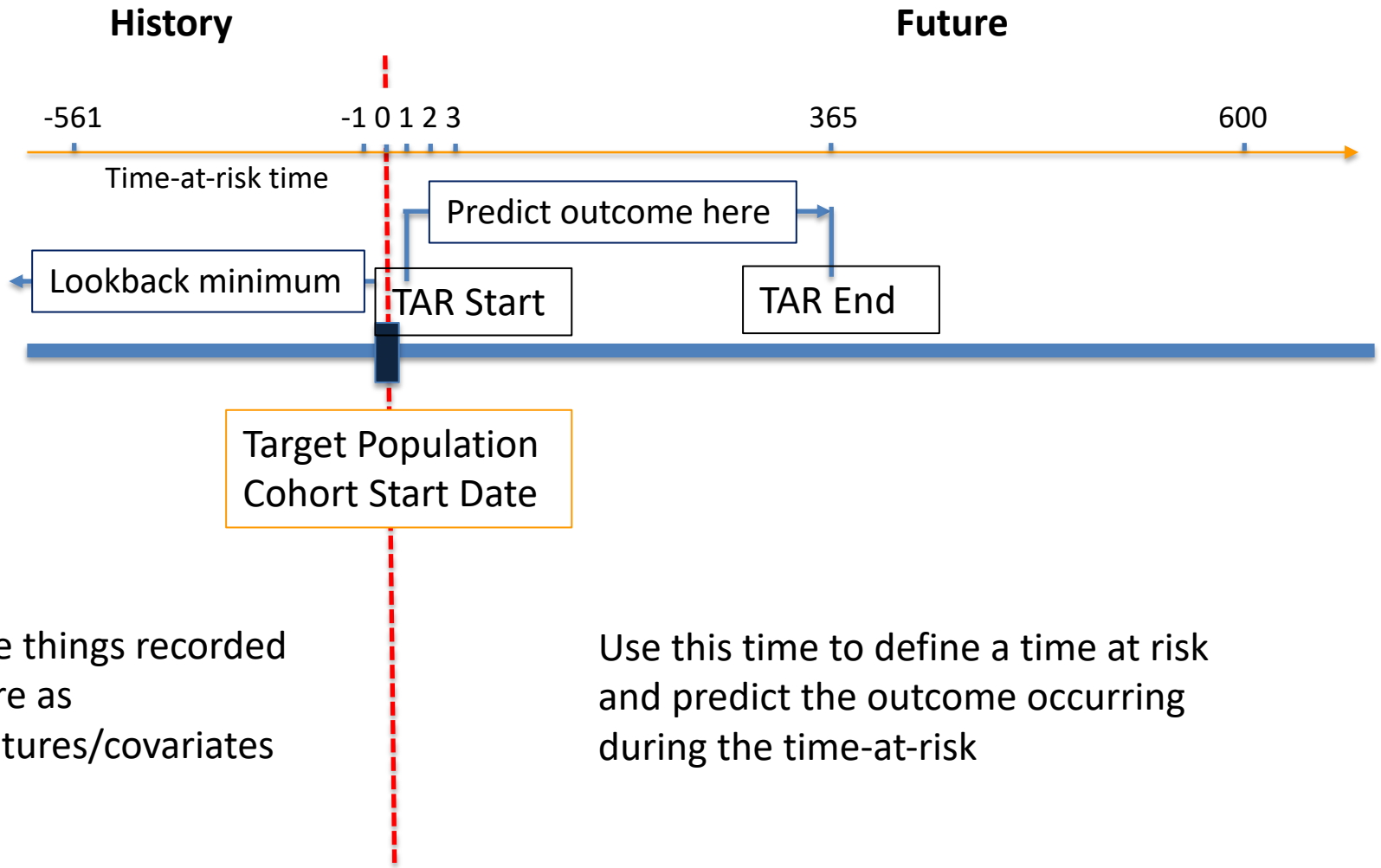
There is a function: checkPlpInstallation() that makes sure you have everything correctly set up

Generating R Code With Atlas



Once you have a target population cohort and an outcome cohort you are ready to run a prediction

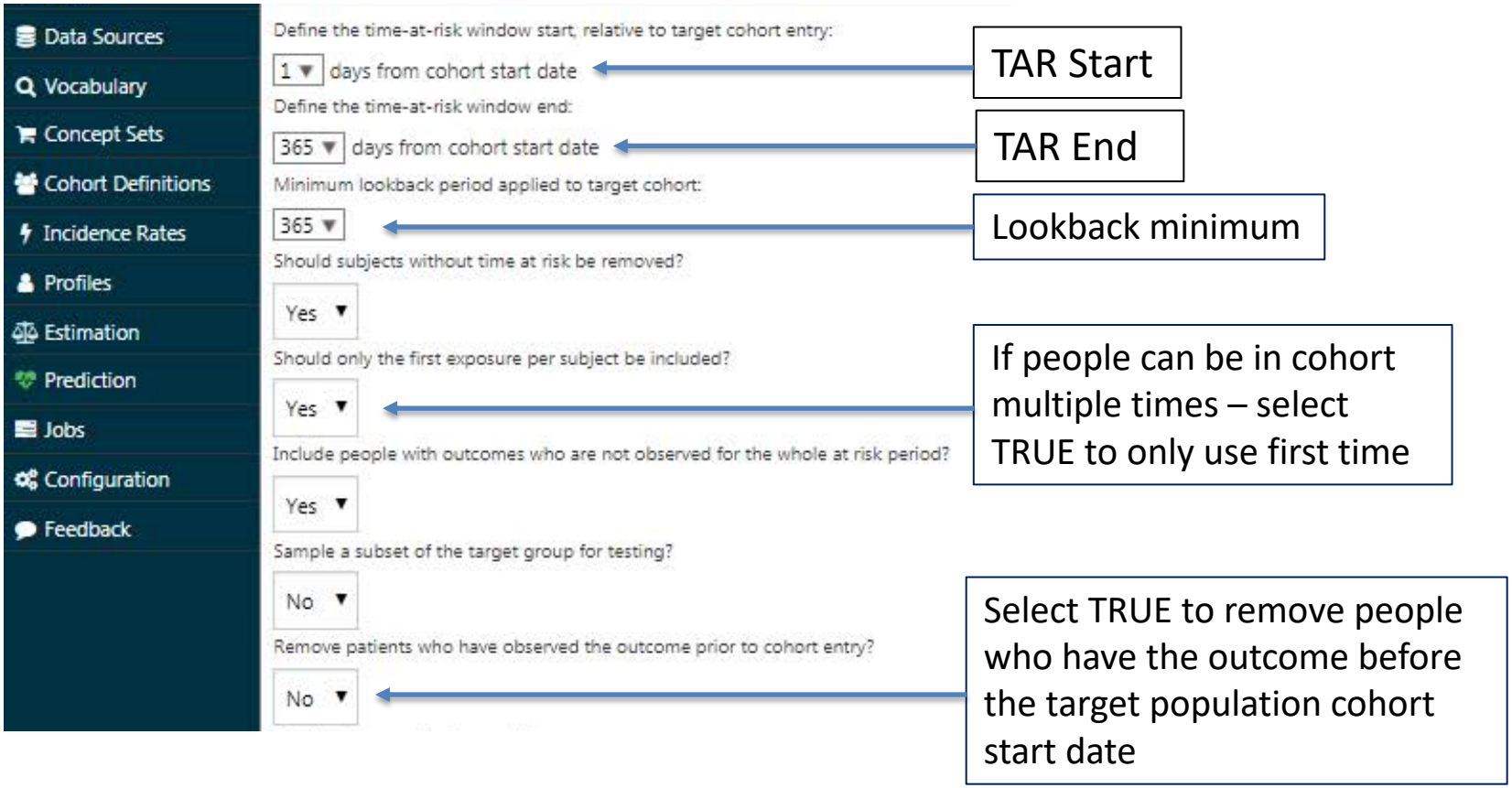
Prediction Parameters



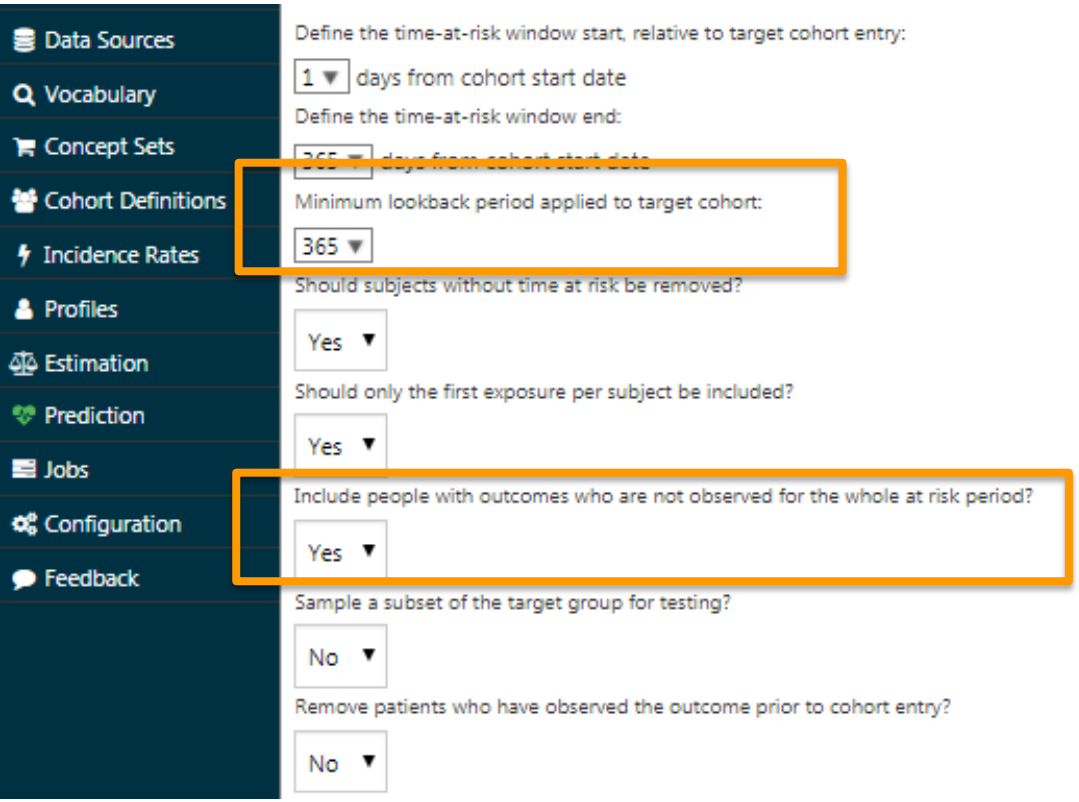
Use things recorded here as features/covariates

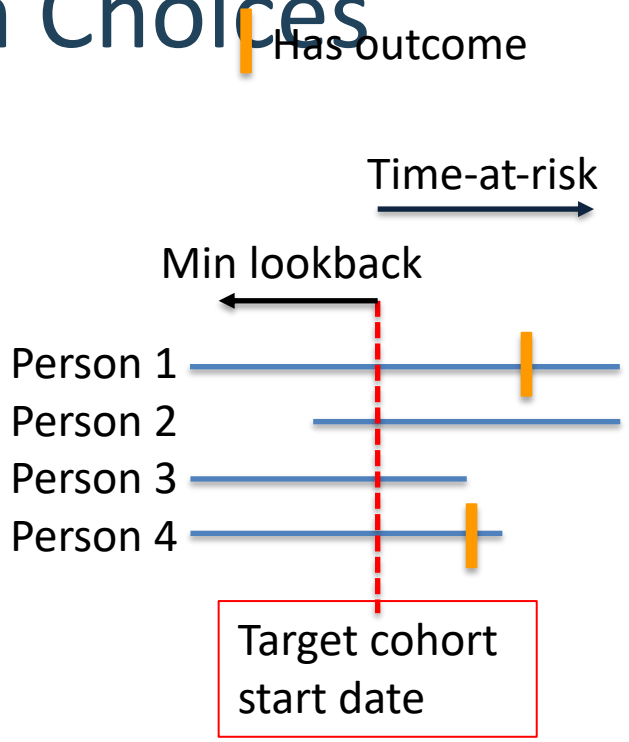
Use this time to define a time at risk and predict the outcome occurring during the time-at-risk

Prediction Deign Choices



Prediction Deign Choices

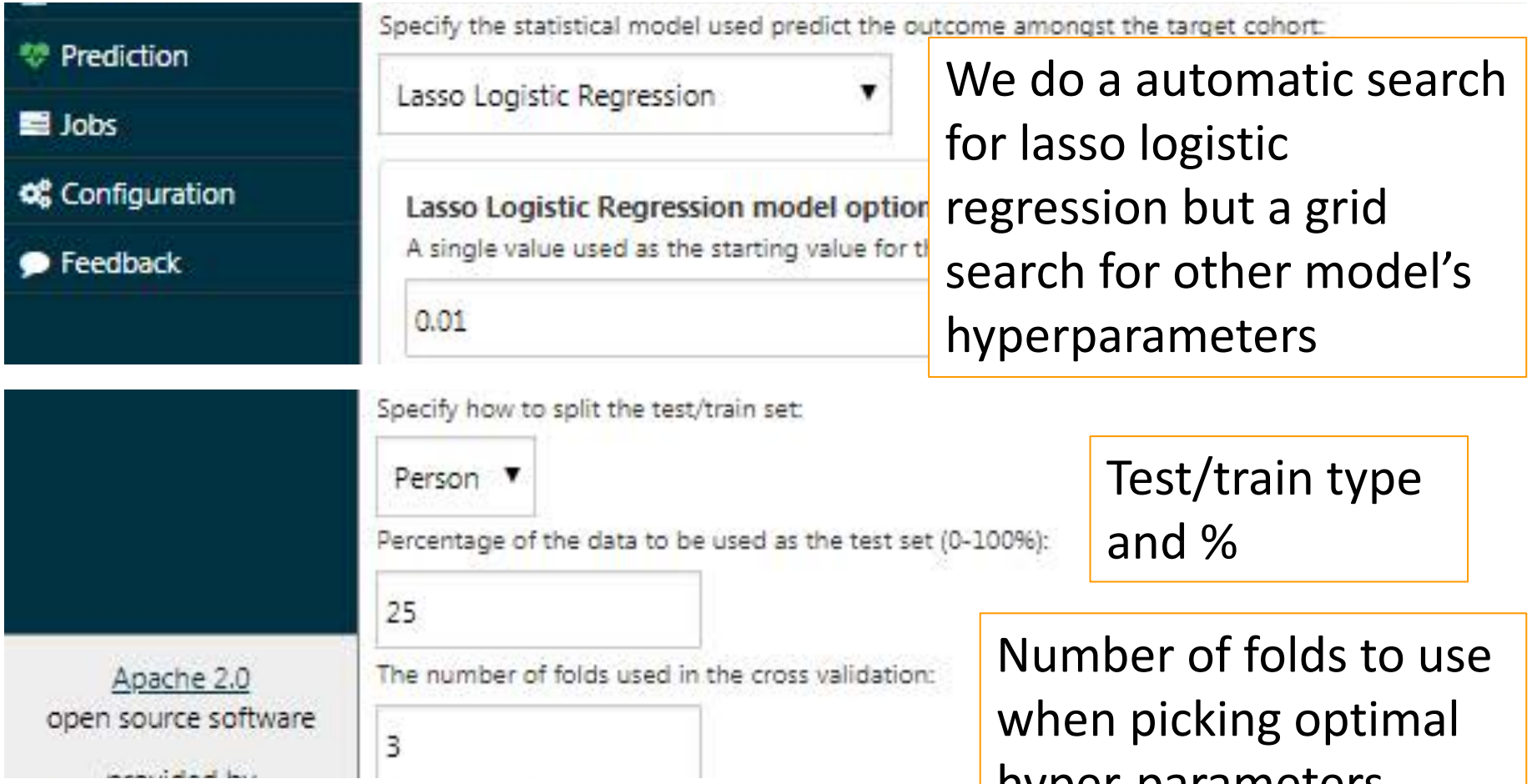




Person 2 doesn't have min lookback so excluded

Persons 3&4 don't have full time-at-risk

Training Choices



We do a automatic search for lasso logistic regression but a grid search for other model's hyperparameters

Test/train type and %

Number of folds to use when picking optimal hyper-parameters

Feature/Covariate Choices

Standardised Features:

- Demographics (e.g., age, gender, ethnicity)
- Conditions (+ condition groups using SNOMED/MEDRA vocabs)
- Drug (+ drug groups)
- Procedures
- Measurements
- Observations
- Counts
- Some existing risk models

  (Flexible times before cohort start date (e.g., -365 to -1 days relative to cohort start date)

Custom Features:

- Can also make any feature you want using R and SQL

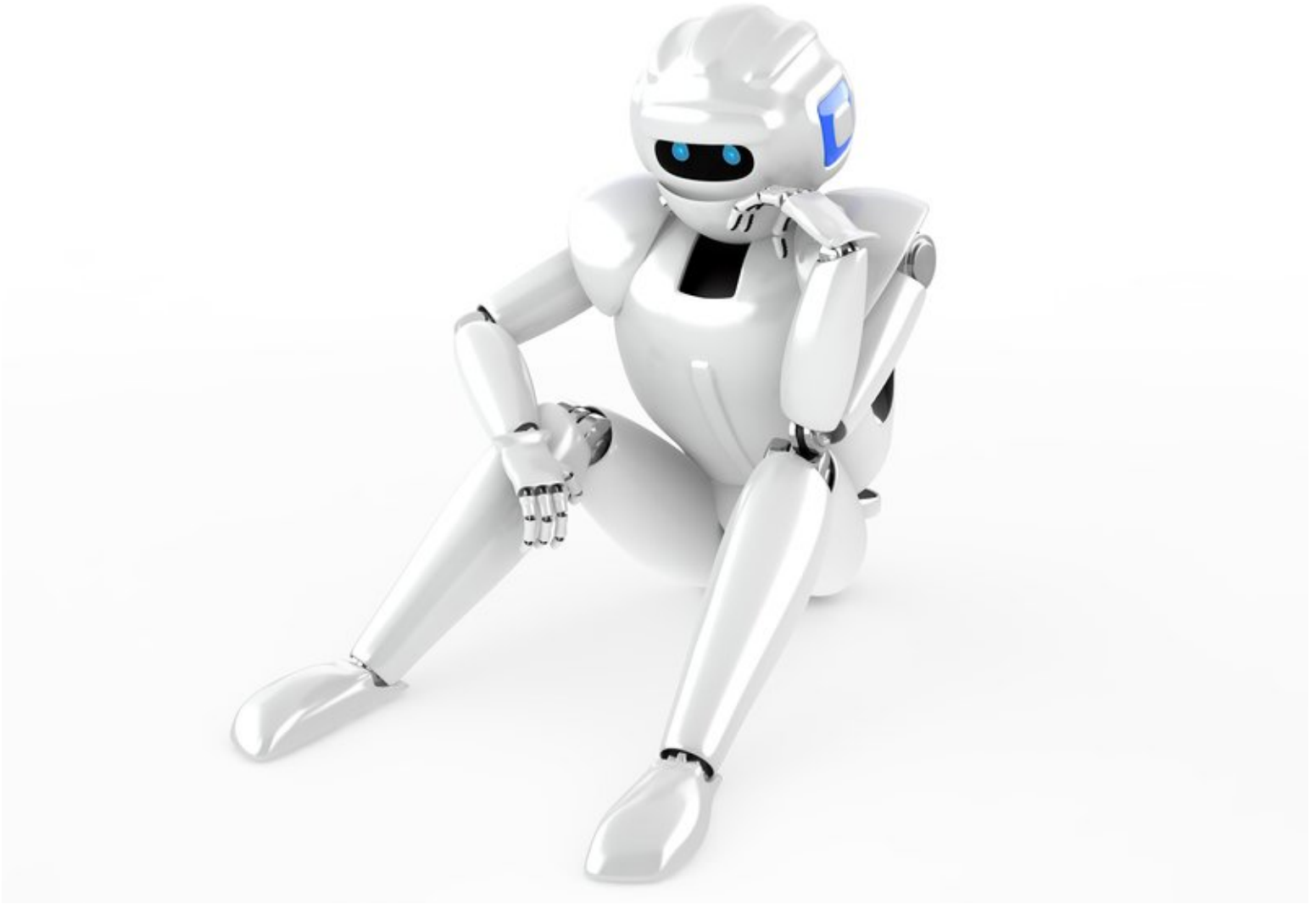
Atlas Demo

I will show how to create the R code to predict bleeding within 1 to 365 days after first time wi

	Option	Choice
Prediction Design Choices	Time at risk start	1
	Time at risk end	365
	Remove prior outcomes	TRUE
	Require time-at-risk	TRUE
	Use all outcomes	TRUE
Training choices	Classifier	Lasso LR
	N-folds	3
	Test %	25
	Split type	Person
Feature Choices	All demo, conditions (+groups), drugs (+groups), measurements, observations, procedures 365 days prior	

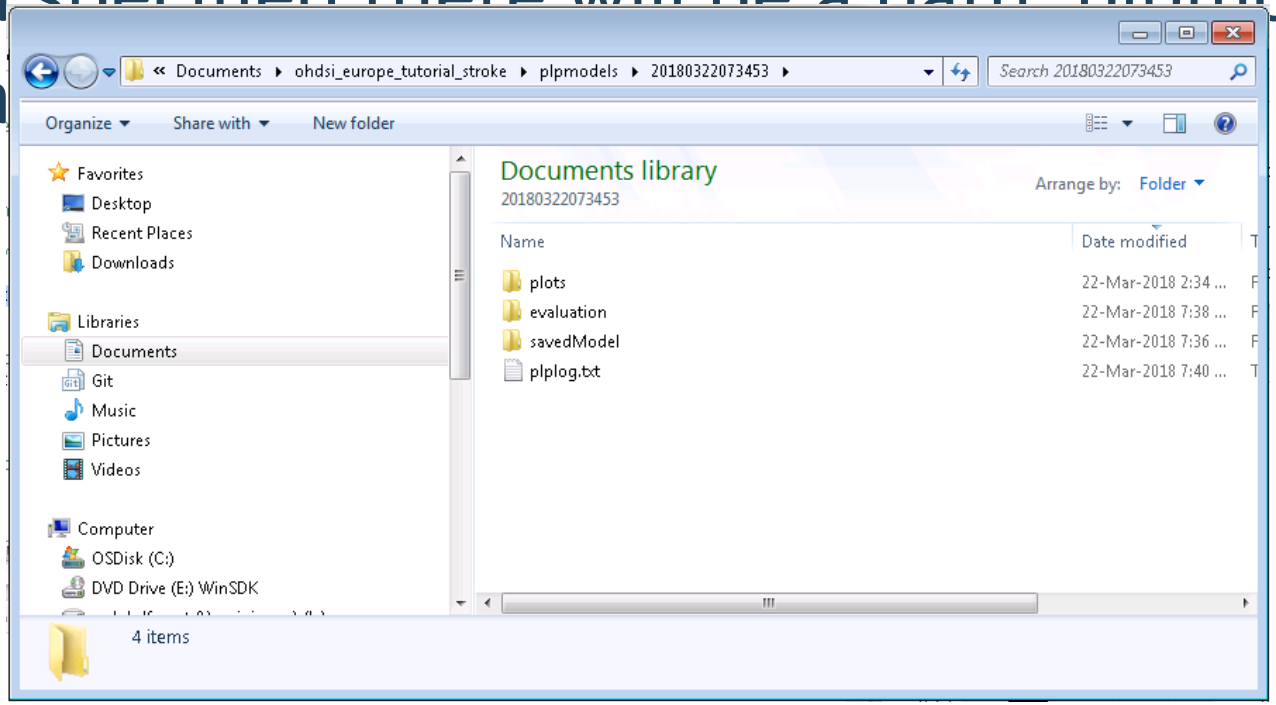
- Go to: http://www.ohdsi.org/web/atlas/#/plp/1735887

PatientLevelPrediction Output



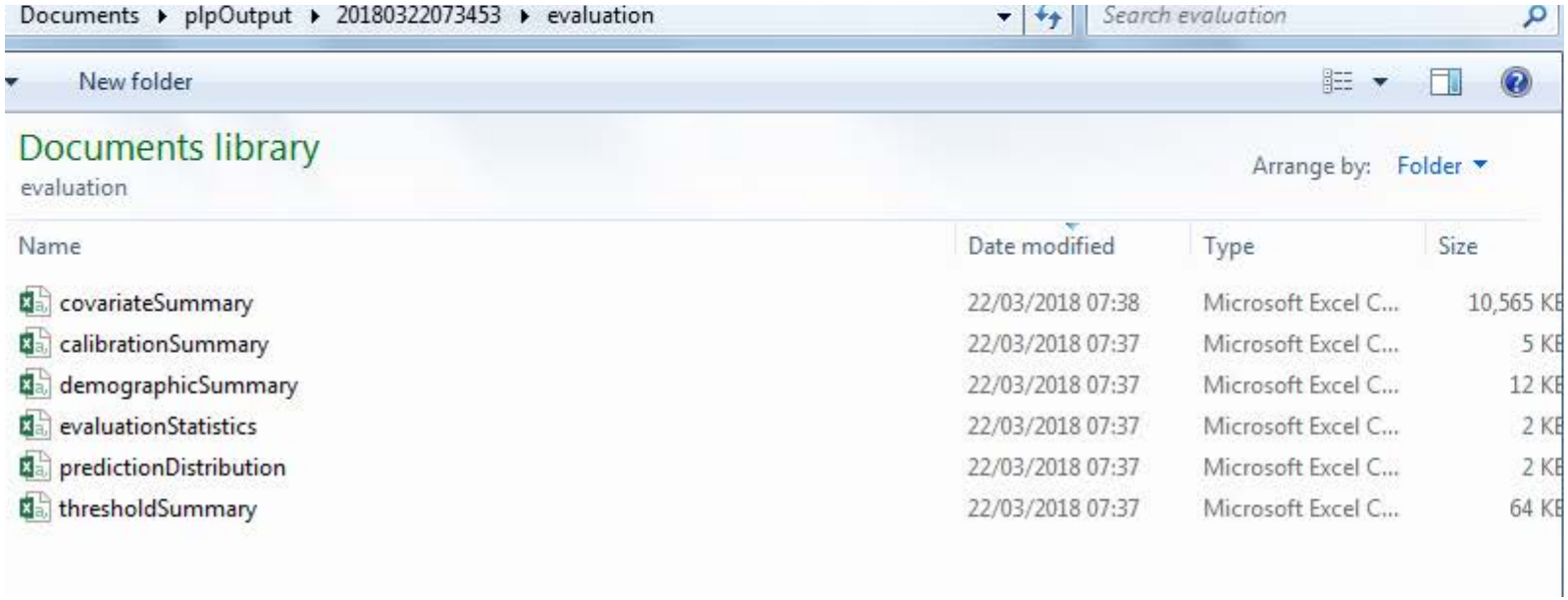
PatientLevelPrediction Output

- When you run the atlas code, in the directory you specified there will be a path: plpmodels-> ana



PatientLevelPrediction Output

- Evaluation folder:



PatientLevelPrediction Output

- EvaluationStatistics

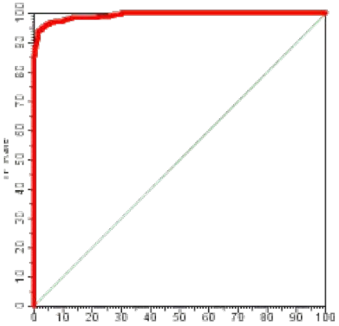
populationSize	7134
outcomeCount	187
AUC.auc	0.649988
AUC.auc_lb95ci	0.611923
AUC.auc_ub95ci	0.688053
BrierScore	0.025352
BrierScaled	0.010316
CalibrationIntercept.Intercept	-0.01395
CalibrationSlope.Gradient	1.526665

2.6% of population have outcome

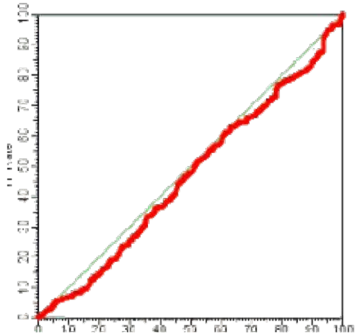
PatientLevelPrediction Output

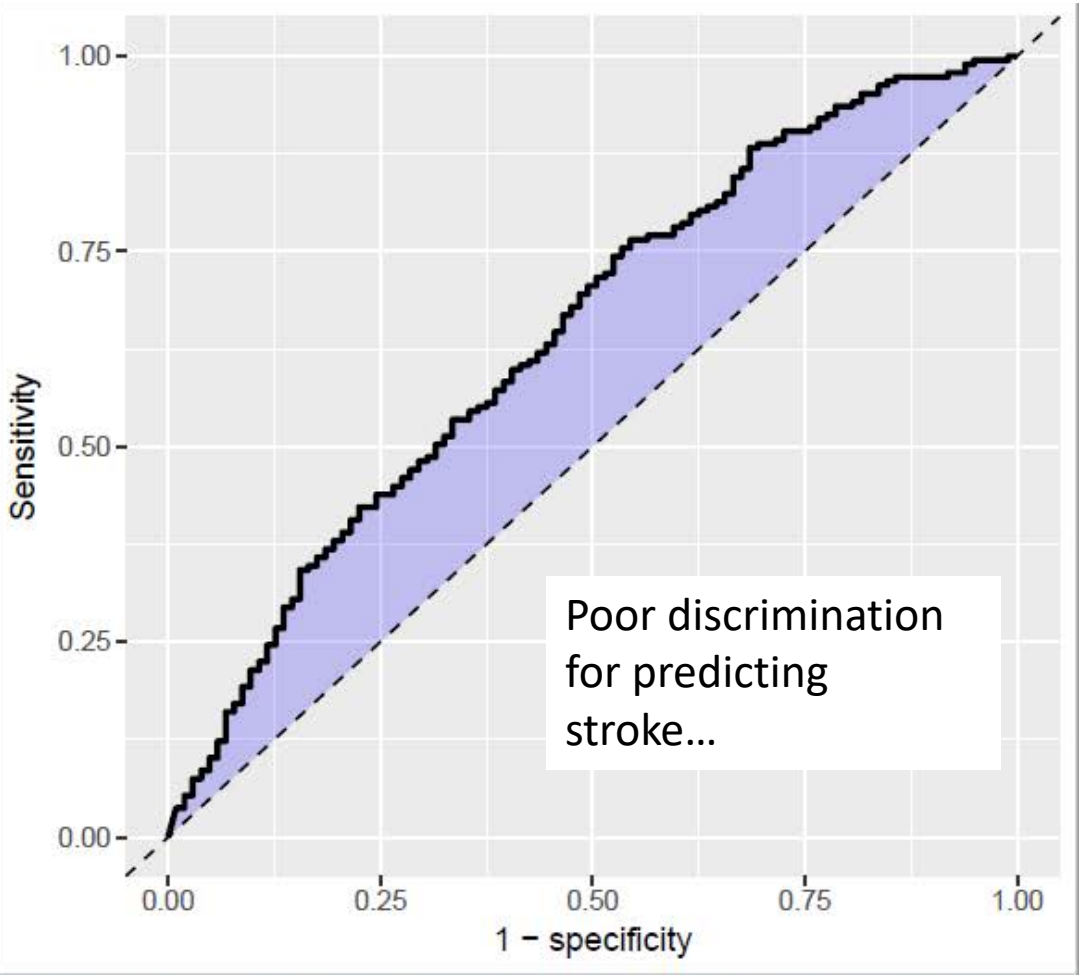
- ROC Plot: Measu discrimination

Near Perfect discrimination:



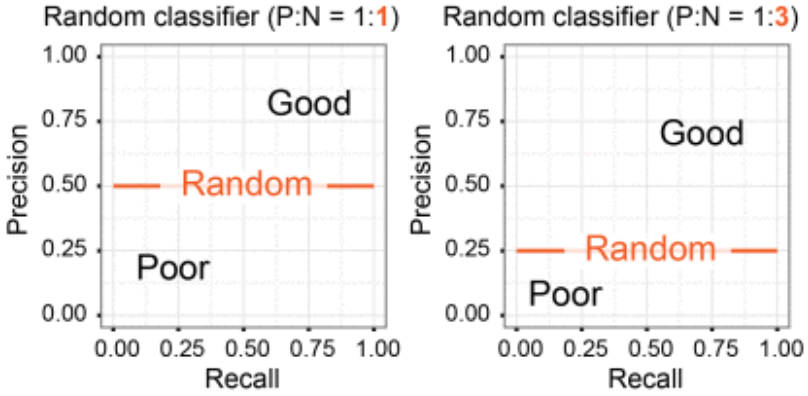
Worse discrimination:

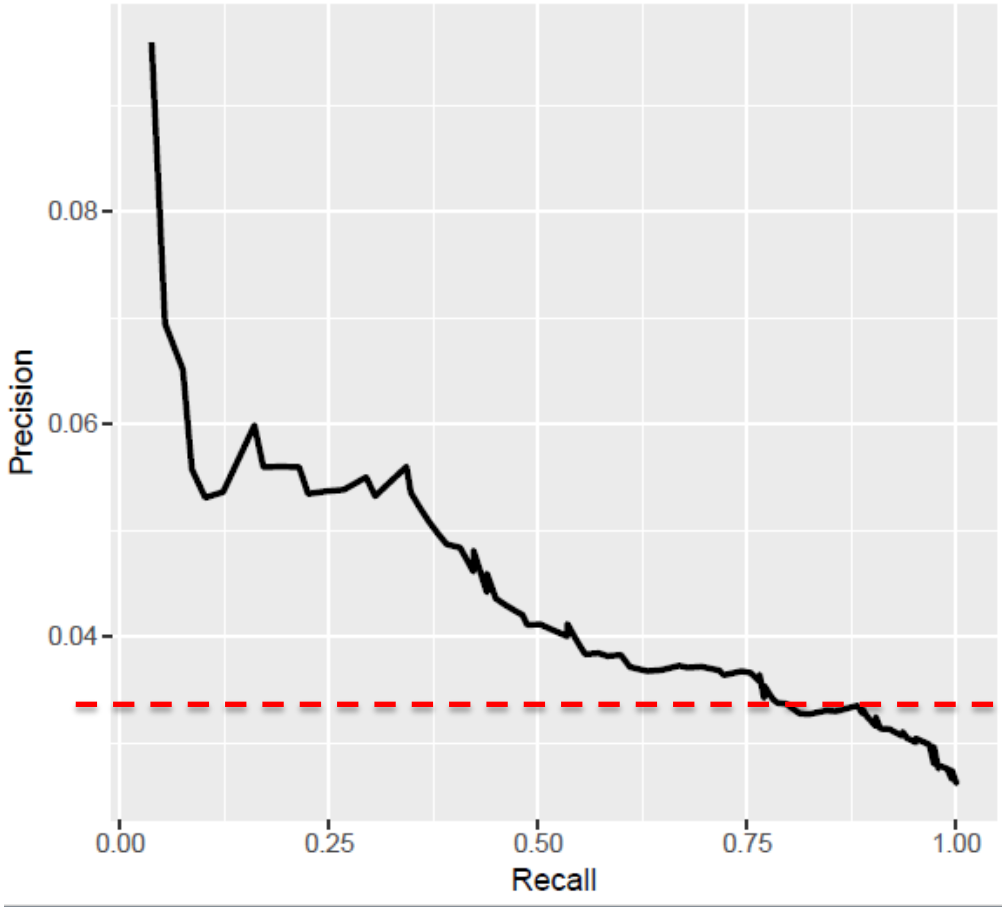




PatientLevelPrediction Output

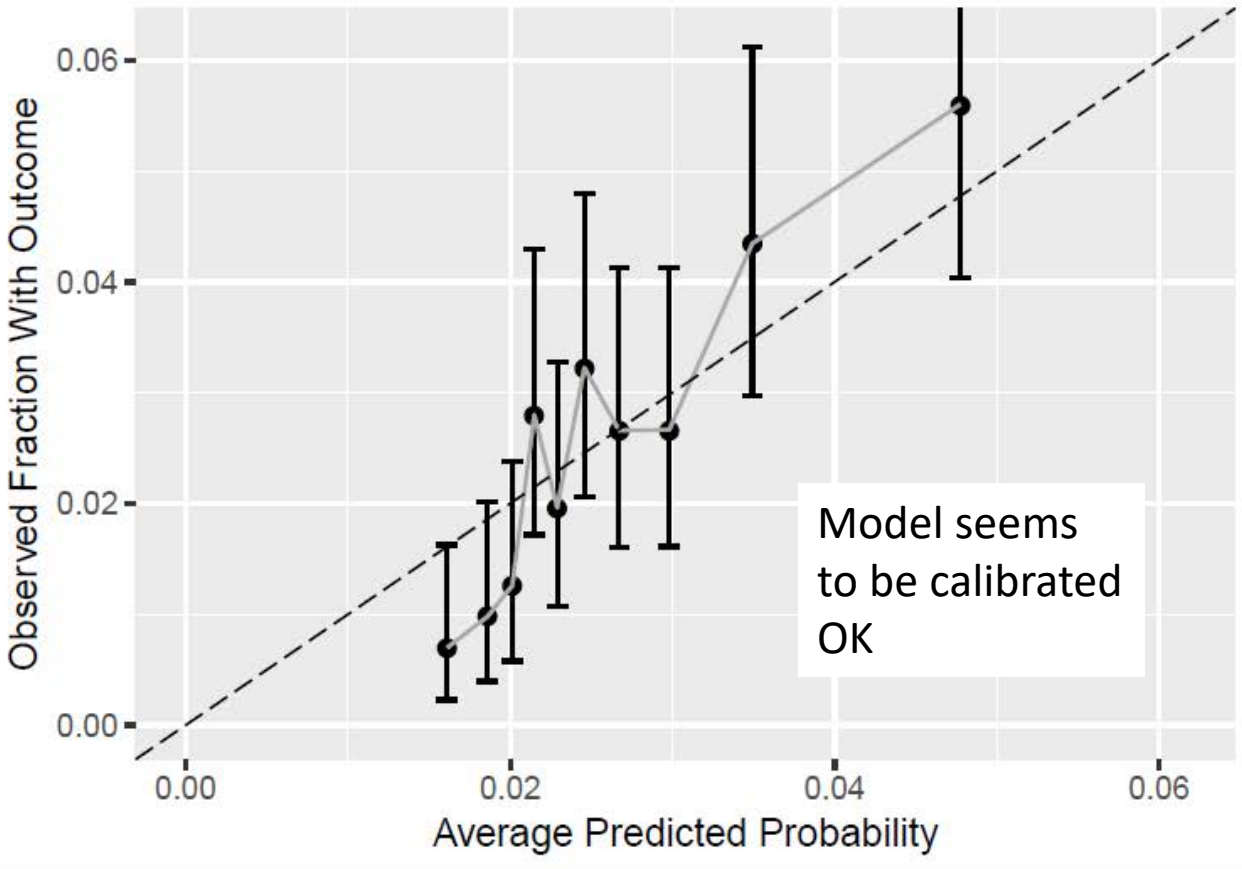
- Precision recall plot: Good to use when the outcome is rare to





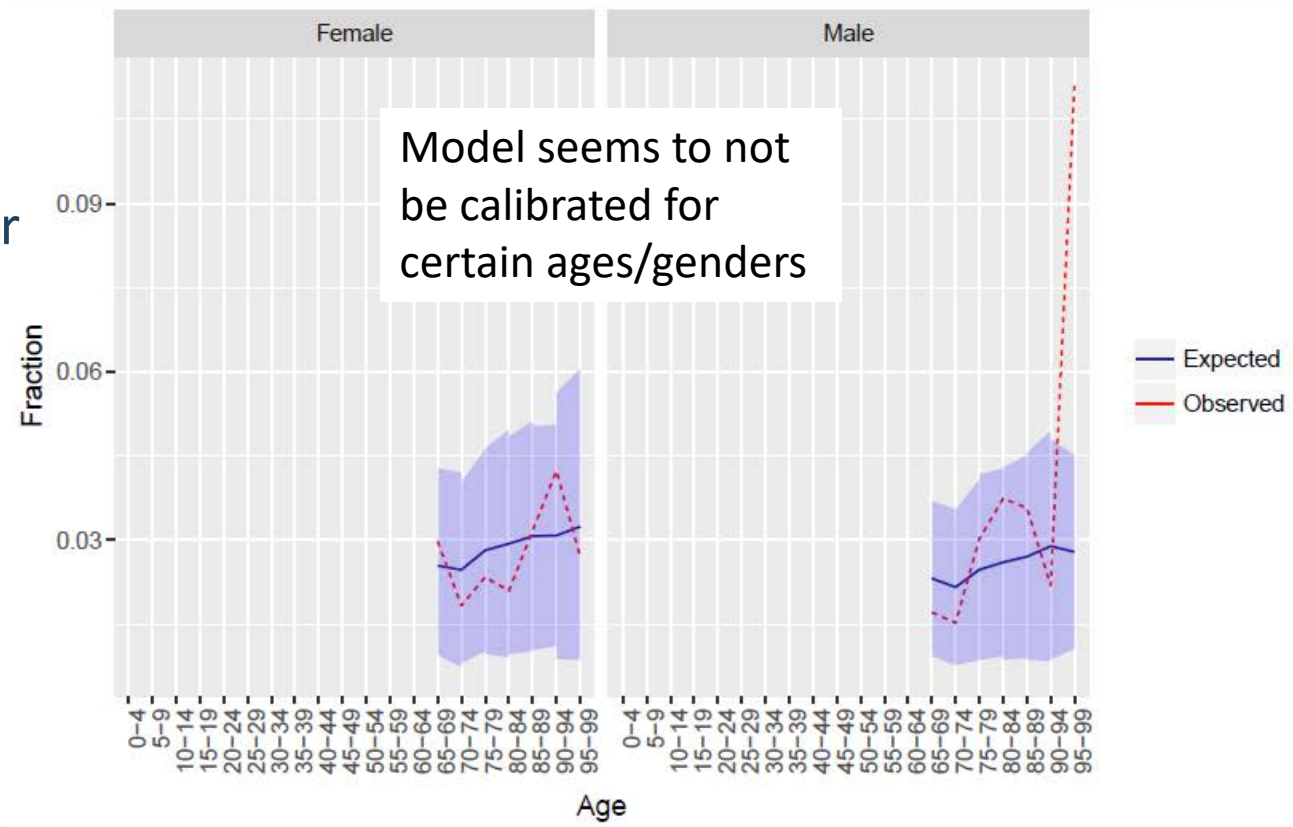
PatientLevelPrediction Output

- Calibration Plot
- Good calibration means dots around x=y line



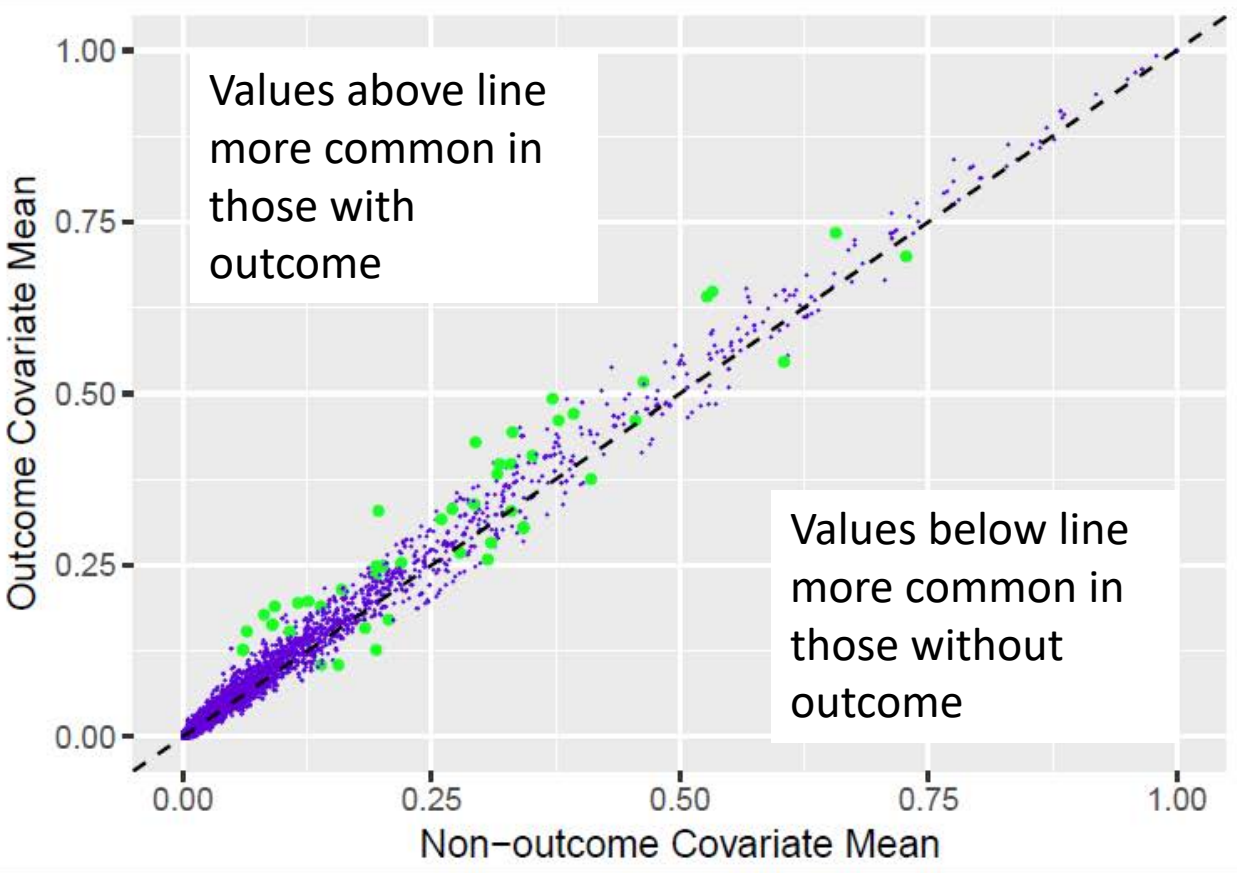
PatientLevelPrediction Output

- Demographic Calibration:
- What expected and observed to be similar across age/gender
- If observed/expected differ than maybe need to treat that strata differently



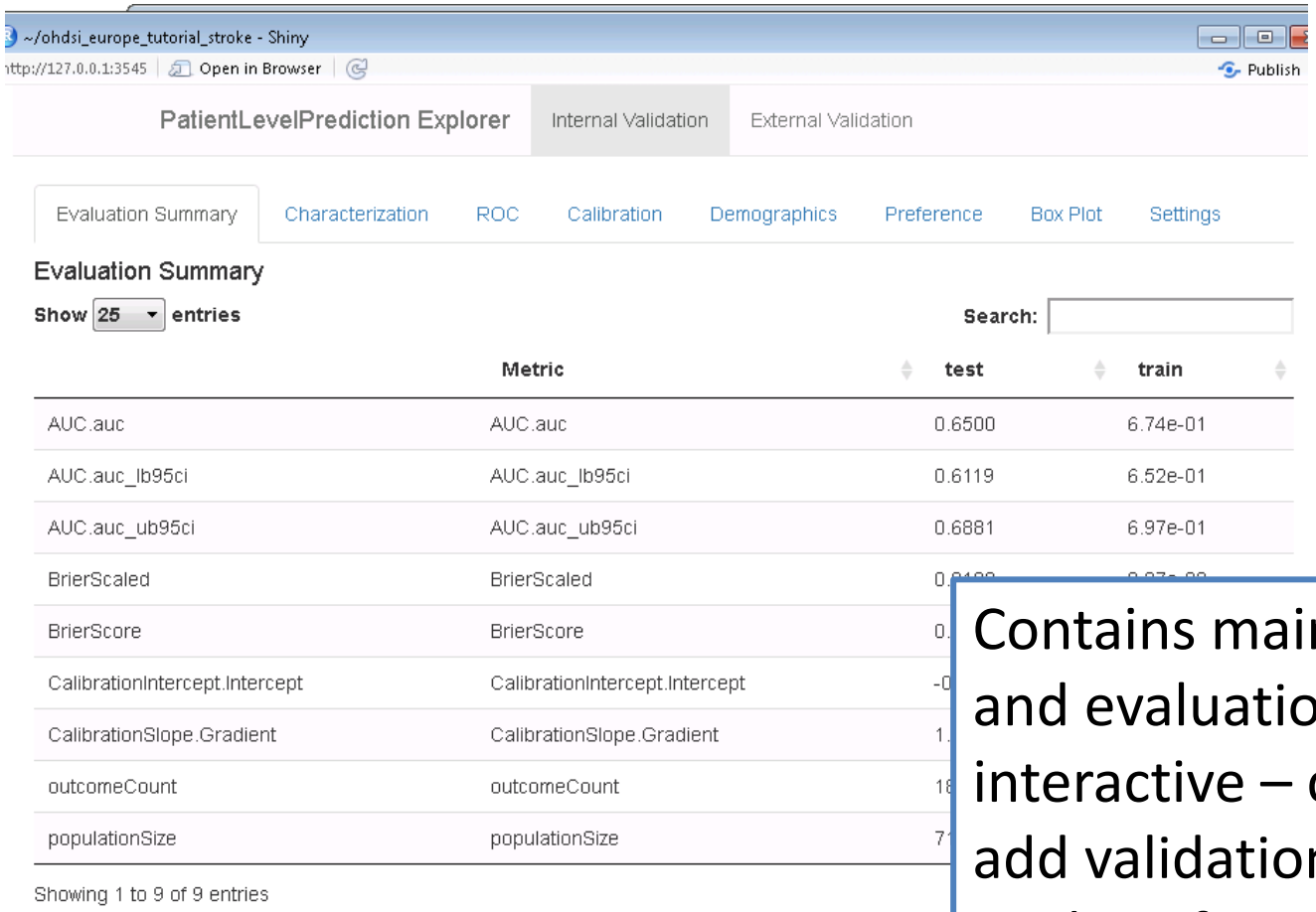
PatientLevelPrediction Output

- Variable scatterplot: shows differences between people with outcome and without outcome



PatientLevelPrediction Shiny View

- PatientLevelPrediction::viewPlp(runPlp = results, validatePlp = externalVal)



Contains main plots and evaluation but is interactive – can also add validation analysis for comparison

Design and implement your own study!

Questions?

Thanks for joining
the journey!