



OHDSI Tools Ecosystem

Patrick Ryan

Janssen Research and
Development, Columbia
University Medical Center

Gowtham Rao

BlueCrossBlueShield

Kristin Feeney Kostka

Deloitte Consulting LLP

Anthony Sena

Janssen Research and
Development

Greg Klebanov

Odysseus Data Services



Agenda

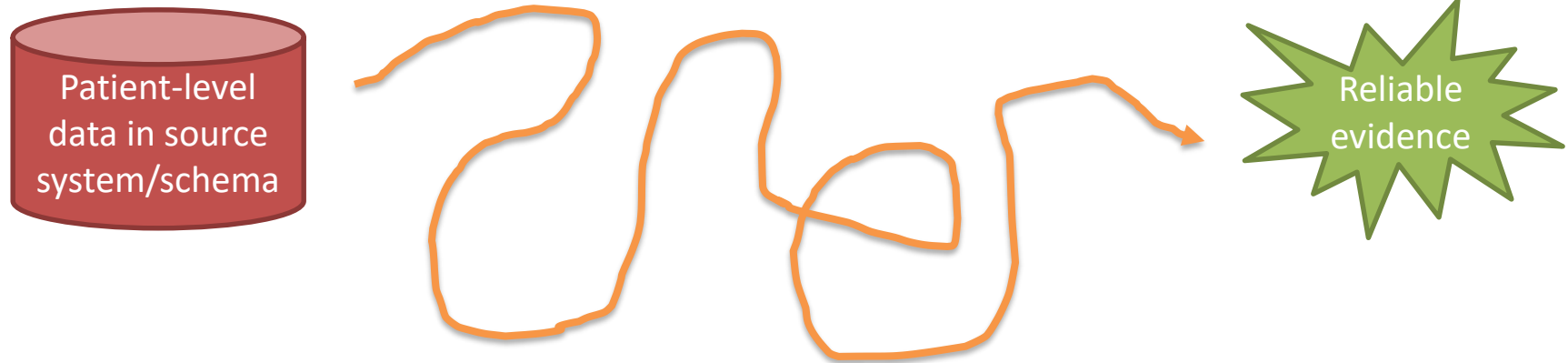
800am-900am	Registration	
900am-1000am	Overview of OHDSI tools ecosystem	Patrick Ryan
1000am-1030am	Vocabulary	Patrick Ryan
1030am-1045am	Break	
1045am-1115am	Data sources (ACHILLES)	Kristin Feeney Kostka
1115am-1230pm	Cohort definition and characterization	Gowtham Rao
1230pm-130pm	Lunch	
130pm-200pm	Incidence rate	Kristin Feeney
200pm-230pm	Population-level effect estimation	Anthony Sena
230pm-300pm	Patient-level prediction	Anthony Sena
300pm-330pm	Break	
330pm-400pm	Network analyses (ARACHNE)	Greg Klebanov
400pm-500pm	Design and implement your own OHDSI study!	Everyone



Overview of the OHDSI tools ecosystem



The journey to real-world evidence

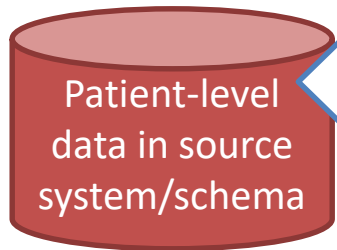




The journey to real-world evidence

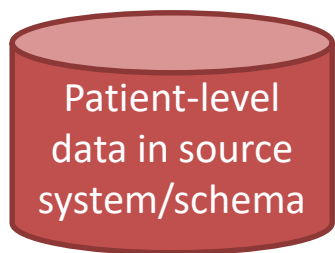
Different types of observational data:

- **Populations**
 - Pediatric vs. elderly
 - Socioeconomic disparities
- **Care setting**
 - Inpatient vs. outpatient
 - Primary vs. secondary care
- **Data capture process**
 - Administrative claims
 - Electronic health records
 - Clinical registries
- **Health system**
 - Insured vs. uninsured
 - Country policies





The journey to real-world evidence



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



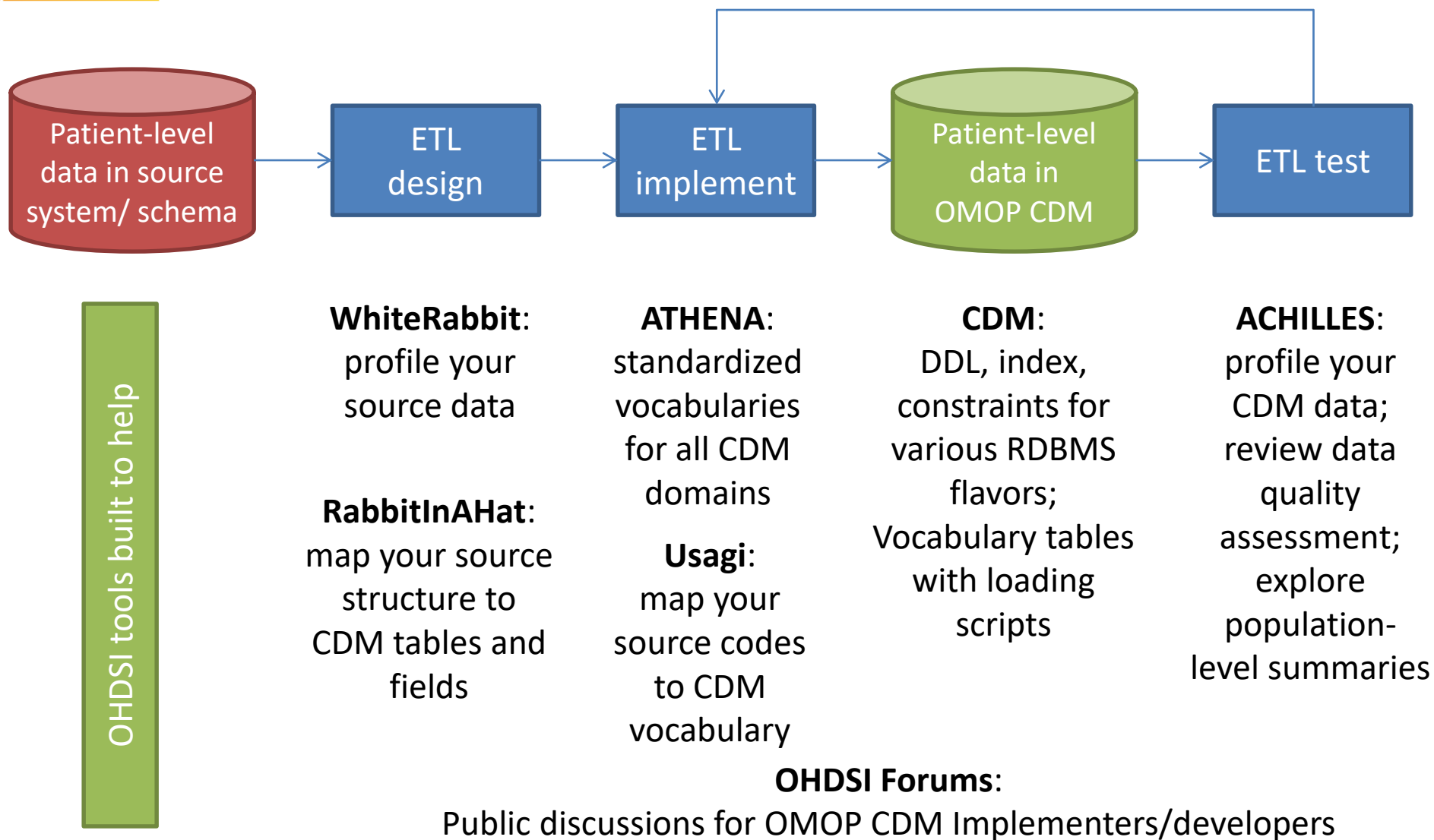


What are your research questions?



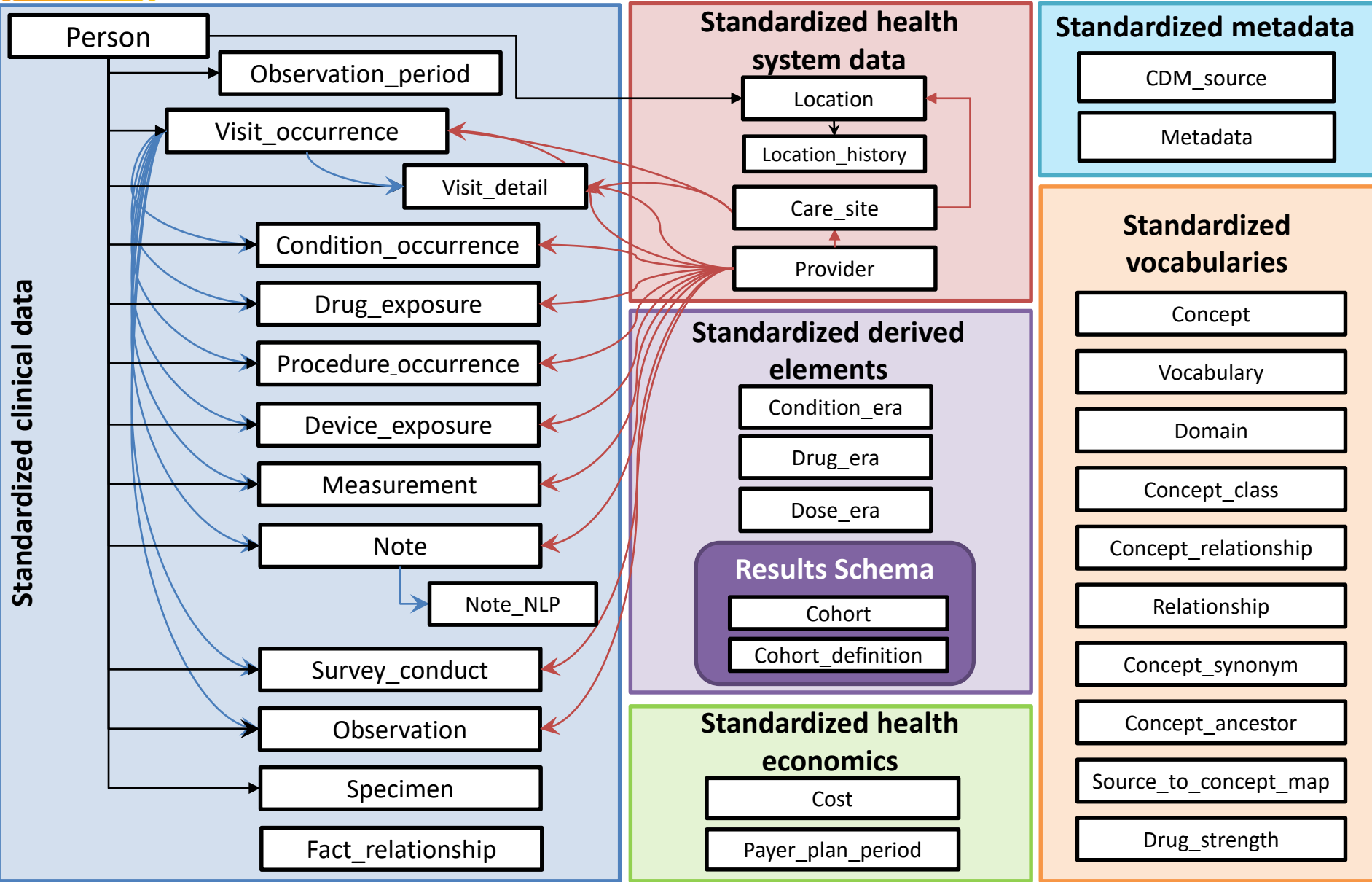


Structuring the journey from source to a common data model



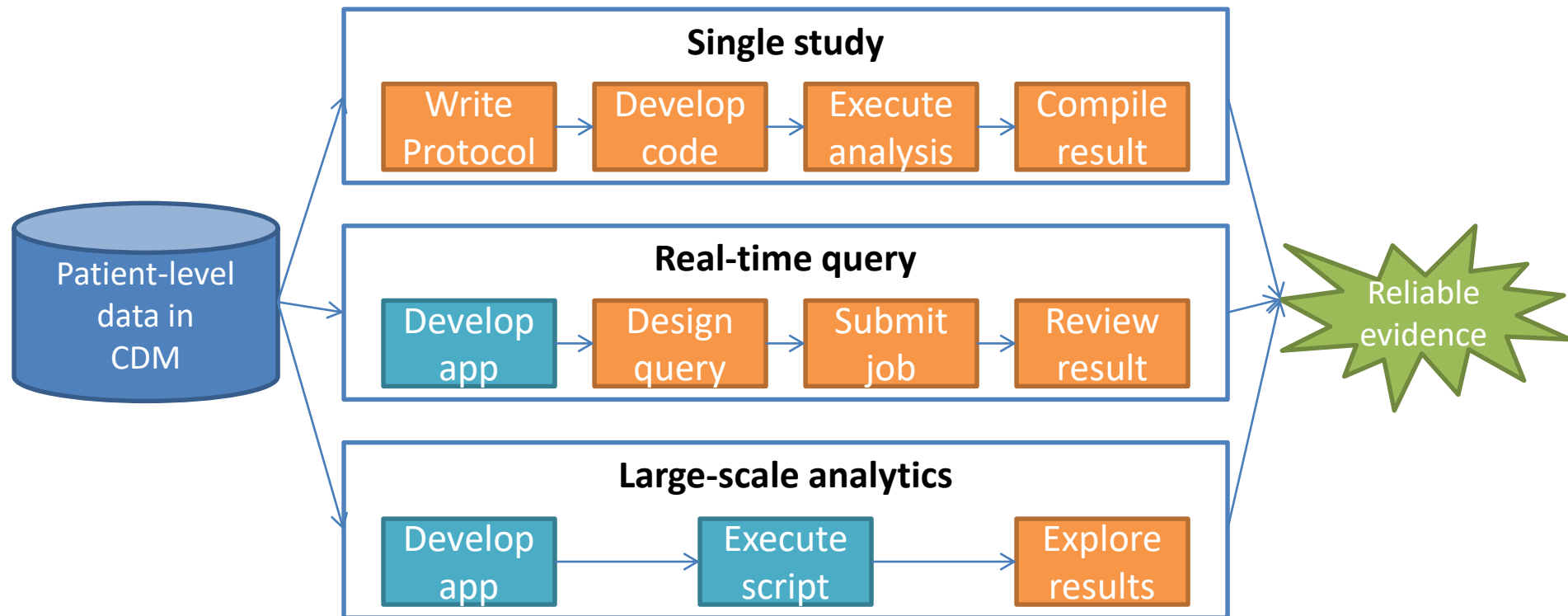


OMOP CDM Version 6





Structuring the journey from a common data model to evidence



One-time

Repeated



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

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates

Profiles

Estimation

Prediction


Jobs

Configuration

Apache 2.0

open source software

provided by




join the journey

Home

Welcome to ATLAS.

ATLAS is an open source application developed as a part of [OHDSI](#) intended to provide a unified interface to patient level data and analytics.

Documentation

 The ATLAS user guide can be found [here](#).

Getting Started

Define a New Cohort

Search the Vocabulary

Begin performing research by defining the group of people you intend to study


Search the different ontologies used to describe patient level data around the world


Release Notes


[ATLAS Version 2.6.0 Release Notes](#)


[WebAPI Version 2.6.0 Release Notes](#)


This latest release contains **23** feature enhancements and issue resolutions:


 [Pathway fixes](#)


 [Moves conceptSets to FeatureExtractionAnalysis level](#)


 [Fixed save btn disabled state in concept set](#)


 [new daimon Temp](#)


 [cohort pathways - confusing save prompt](#)


 [cohort pathway person count issue](#)

 [Data Source order on Cohort Generation tab](#)

 [Excessive generation info polling](#)

 [Not Enough Permissions to Generate Tooltip Error](#)

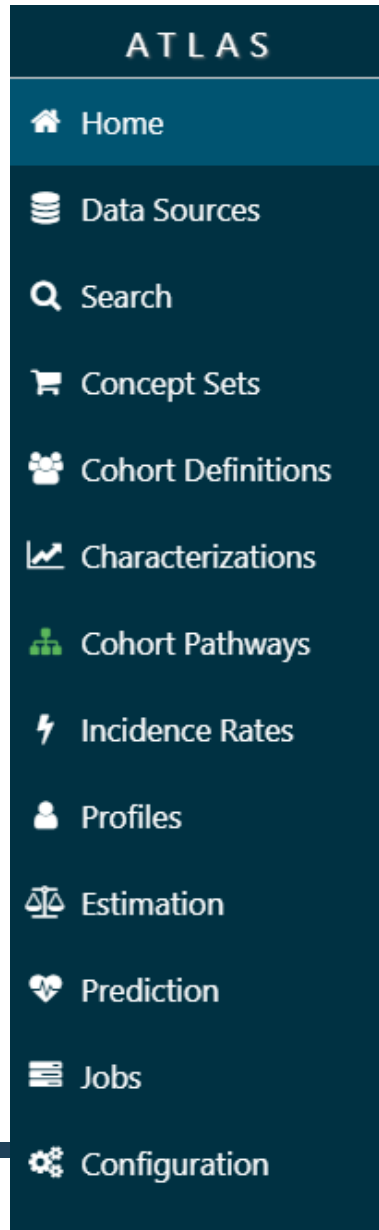
 [Estimation & Prediction Specification Editors](#)

 [Cohort generation time format issue](#)

<http://ohdsi.org/web/ATLAS>




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.



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

The screenshot displays the ARACHNE web application interface. The top navigation bar includes the ARACHNE logo, a search bar, and user information (Gregory Kabanov). The left sidebar contains navigation links: STUDY NOTEBOOK, EXPERT PAGES, DATA CATALOG, and INSIGHTS LIBRARY. The main content area is titled 'STUDIES' and features a table of study records. The table columns are STUDY, LEAD, ROLE, CREATED, TYPE, and STATUS. The studies listed include 'BEST: Pediatric Transfusion and Acute Respiratory Distress Syndrome', 'CLIPOL', 'Demo Study', 'Prediction of tumor response in NSCLC patients', and 'VH Network Study'. Each study entry includes a star icon, a lock icon, and a status indicator (Active or Completed).

STUDY	LEAD	ROLE	CREATED	TYPE	STATUS
BEST: Pediatric Transfusion and Acute Respiratory Distress Syndrome	Gregory Kabanov, Christian Heuch	Data Set Owner, Lead Investigator	15 Oct 2017	Safety and Efficacy	Active
CLIPOL	Eldar Alukhmedov, Christian Heuch, Dmitry Dyrenfyte, Gregory Kabanov, Aaron Galinski	Lead Investigator	10 Mar 2017	Safety and Efficacy	Active
Demo Study	Gregory Kabanov, Christian Heuch	Data Set Owner, Lead Investigator	17 Sep 2017	Sales and Marketing	Completed
Prediction of tumor response in NSCLC patients	Gregory Kabanov, Eldar Alukhmedov, Christian Heuch, Vicky Khanna, Aaron Galinski	Data Set Owner, Lead Investigator	10 Oct 2017	Health Economics and Outcomes	Active
VH Network Study	Viggoch Huse, Gregory Kabanov	Lead Investigator	13 Oct 2017	Other	Active



All of OHDSI tools are open source and freely available



This organization

Search

Pull requests

Issues

Marketplace

Explore



Observational Health Data Sciences and Informatics

<http://ohdsi.org>

Repositories 110

People 94

Teams 17

Projects 1

Settings

Search repositories...

Type: All ▾

Language: All ▾

Customize pinned repositories

New

Atlas

ATLAS is an open source software tool for researchers to conduct scientific analyses on standardized observational data

JavaScript ★ 41 35 Apache-2.0 Updated 8 hours ago



ArachneUI

Network infrastructure for collaborative studies across disparate data nodes and researches

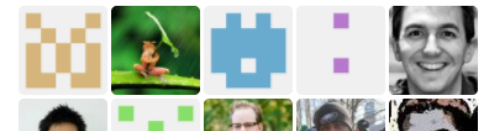


Top languages

R JavaScript Java HTML C++

People

94 >



<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

Table 1. Continued

Characteristic	Dabigatran, % (n=67 207)	Warfarin, % (n=67 207)	Standardized Mean Difference
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			
ACE/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. ACE/ARB indicates angiotensin converting-enzyme inhibitor/angiotensin receptor blocker; NSAIDs, nonsteroidal anti-inflammatory drugs; and SSRI, selective serotonin reuptake inhibitor.

*The CHADS₂ 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

		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:

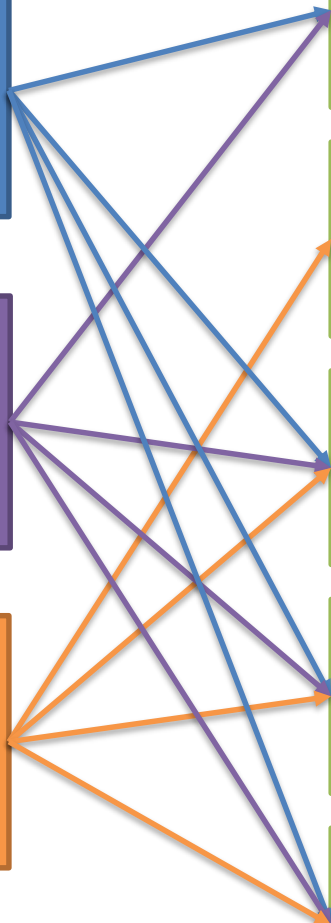
Clinical characterization
Baseline summary of exposures
(treatment utilization)

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

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates


Profiles

Estimation

Apache 2.0

open source software

provided by

 OHDSI

join the journey.

Cohort #2

[OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation

Definition ? Concept Sets Generation Reporting Export Messages 9

enter a cohort definition description here

Cohort Entry Events ?

Events having any of the following criteria:

a drug exposure of warfarin

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.

Restrict initial events

+ Add Initial Event

+ Add attribute...

Delete Criteria



OHDSI in action: Cohort characterization

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates

Profiles

Estimation

Apache 2.0

open source software

provided by

OHDSI

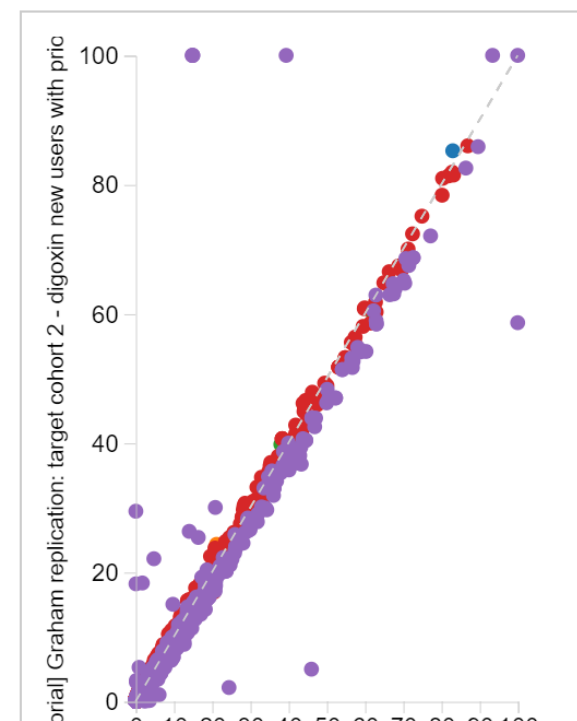
join the journey.

All prevalence covariates

Show 10 entries

Search:

Covariate	[OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation		[OHDSI Ecosystem tutorial] Graham replication: target cohort 2 - digoxin new users with prior atrial fibrillation		Std diff
	Count	Pct	Count	Pct	
age group: 00-04	54	1.37%	21	1.34%	-0.0020
age group: 65-69	611	15.50%	257	16.36%	0.0153
age group: 70-74	836	21.20%	382	24.32%	0.0461
age group: 75-79	792	20.09%	323	20.56%	0.0074
age group: 80-84	689	17.47%	247	15.72%	-0.0304
age group: 85-89	548	13.90%	173	11.01%	-0.0578
age group: 90-94	272	6.90%	114	7.26%	0.0095
age group: 95-99	141	3.58%	54	3.44%	-0.0052
condition era group during day -365 through	4	0.10%	3	0.19%	0.0165





◀ [OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation

⚡ Incidence Rate Analysis

[OHDSI Ecosystem tutorial] Outcomes following warfarin exposure

Definition

Study Cohorts

Target Cohorts

- ✗ #2: [OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation
- ✗ #10: [OHDSI Ecosystem tutorial] Graham replication: target cohort 2 - digoxin new users with prior atrial fibrillation

Add Target Cohort

Outcome Cohorts

- ✖ #3: [OHDSI Ecosystem tutorial] Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting
- ✖ #4: [OHDSI Ecosystem tutorial] Graham replication: outcome cohort #2 - incident intracranial hemorrhage, observed in inpatient setting
- ✖ #5: [OHDSI Ecosystem tutorial] Graham replication: outcome cohort #3 - incident major gastrointestinal (GI) bleeding events, observed in inpatient setting

Add Outcome Cohort

Time At Risk

Time at risk defines the time window relative to the cohort start or end date with an offset to consider the person 'at risk' of the outcome.

- Time at risk starts with plus days.
- Time at risk ends with plus days.

No study window defined. [Add Study Window](#)

Add Study Window

Apache 2.0

open source software

provided by



[join the journey.](#)



OHDSI in action: incidence rate generation

Incidence Rate Analysis

[OHDSI Europe tutorial] Cardiovascular and Bleeding Risks in Elderly Medicare Patients Treated with Aspirin

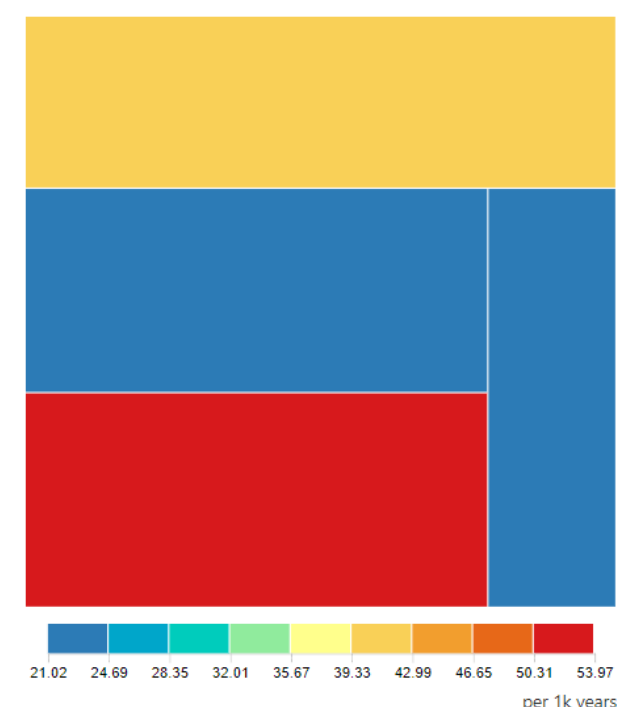
Definition Concept Sets Generation Utilities

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 replication and outcome cohort: [OHDSI Europe tutorial] Graham replication

	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

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates


Profiles

Estimation

Apache 2.0

open source software

provided by

 OHDSI

join the journey.

Population Level Effect Estimation - Comparative Cohort Analysis

[OHDSI Ecosystem tutorial] Warfarin vs. Digoxin for risk of ischemic stroke

Specification Utilities

enter a description here (1000 characters max)



All Comparisons Analysis Settings Evaluation Settings

Comparative Cohort Settings

Comparisons

+ Add Comparison

Show 10 entries Filter:

Remove	Target	Comparator	Outcomes	NC Outcomes	Copy
	[OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation	[OHDSI Ecosystem tutorial] Graham replication: target cohort 2 - digoxin new users with prior atrial fibrillation	[OHDSI Ecosystem tutorial] Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting		

Showing 1 to 1 of 1 entries

Previous 1 Next



OHDSI in action: Population-level effect estimation implementation

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	comparators
Count	17460	17460

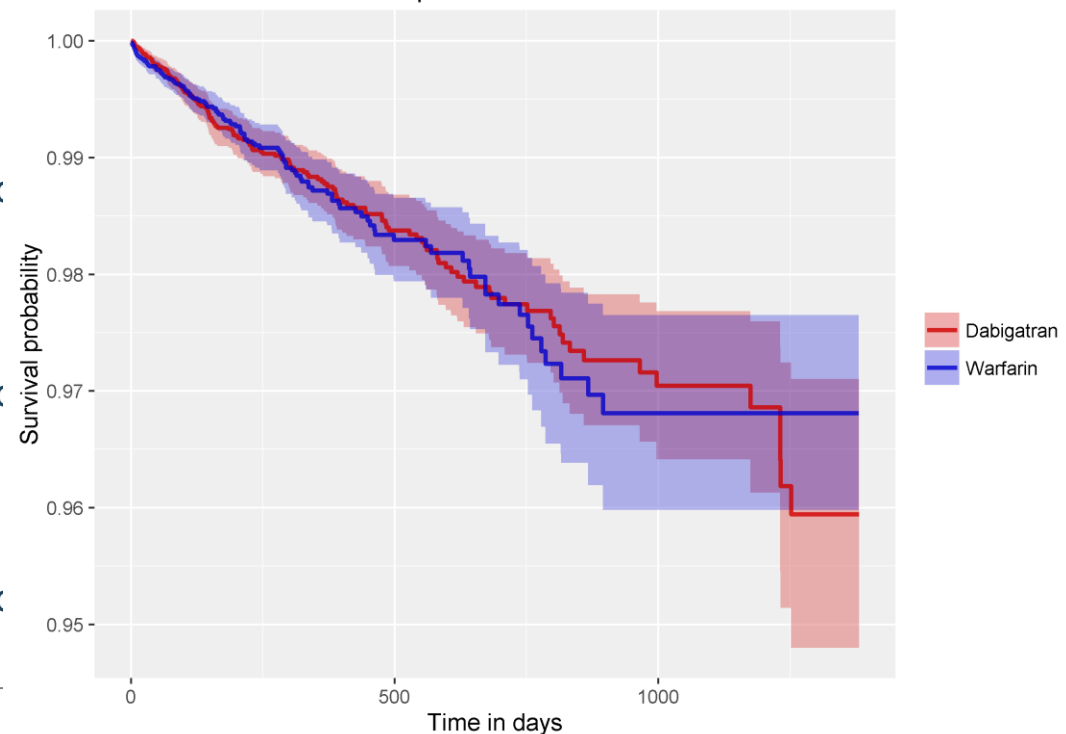
Outcome counts

	treatedPersons	comparators
Count	164	155

Time at risk

	treatedDays	comparators
Days	4912947	3954046

Kaplan-Meier Plot



OHDSI in action:

Patient-level prediction design

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates

Profiles

Estimation


Prediction

Jobs

Apache 2.0

open source software

provided by

 OHDSI

join the journey

♥ Patient Level Prediction

[OHDSI Ecosystem tutorial] Predicting stroke amongst new users of warfarin

Save

Close

Copy

Delete

Specification

Utilities

enter a description here (1000 characters max)

Filter

All

Prediction Problem Settings

Analysis Settings

Execution Settings

Training Settings

♥ Prediction Problem Settings

Target Cohorts

+ Add Target Cohort

Show 10 entries

Filter:

Remove	Name
<div>×</div>	[OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation

Showing 1 to 1 of 1 entries

Previous 1 Next

Outcome Cohorts

+ Add Outcome Cohort

Show 10 entries

Filter:

Remove	Name
<div>×</div>	[OHDSI Ecosystem tutorial] Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting

Previous 1 Next

OHDSI in action:

Patient-level prediction implementation

~/ohdsi_europe_tutorial_stroke - Shiny
http://127.0.0.1:13545 | Open in Browser | Publish

PatientLevelPrediction Explorer Internal Validation External Validation

Evaluation Summary Characterization ROC Calibration Demographics Preference Box Plot Settings

Evaluation Summary


Show entries Search:

	Metric	test	train
AUC.auc	AUC.auc	0.6500	6.74e-01
AUC.auc_lb95ci	AUC.auc_lb95ci	0.6119	6.52e-01
AUC.auc_ub95ci	AUC.auc_ub95ci	0.6881	6.97e-01
BrierScaled	BrierScaled	0.0103	9.37e-03
BrierScore	BrierScore	0.0254	2.54e-02
CalibrationIntercept.Intercept	CalibrationIntercept.Intercept	-0.0140	-2.34e-02
CalibrationSlope.Gradient	CalibrationSlope.Gradient	1.5267	1.89e+00
outcomeCount	outcomeCount	187.0000	5.63e+02
populationSize	populationSize	7134.0000	2.14e+04

Showing 1 to 9 of 9 entries Previous Next



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.

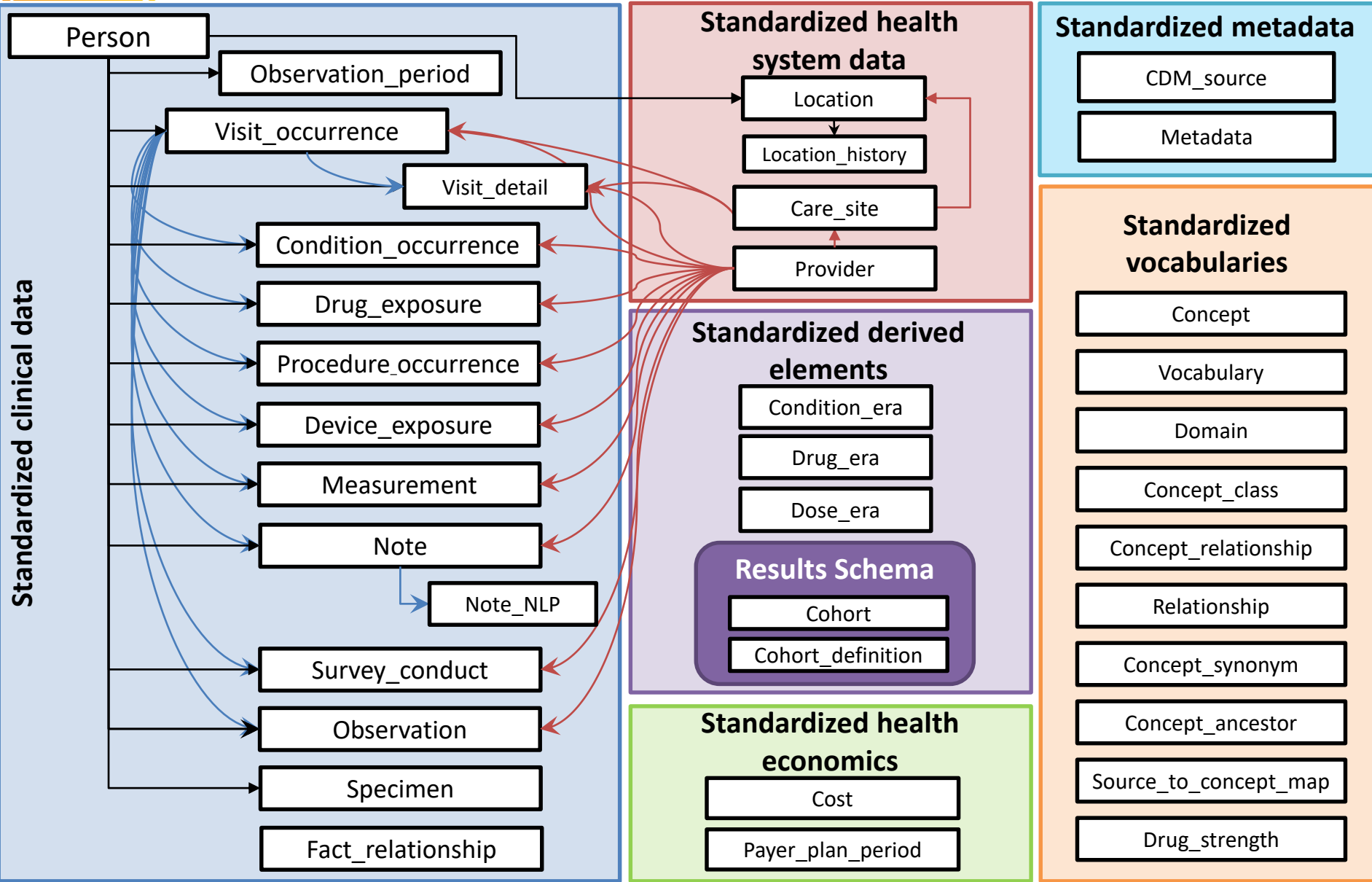
Methods and Results—We formed new-user cohorts of propensity score–matched elderly patients enrolled in Medicare who initiated dabigatran or warfarin for treatment of nonvalvular atrial fibrillation between October 2010 and December 2012. Among 134414 patients with 37587 person-years of follow-up, there 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



OMOP CDM Version 6





OMOP Common Vocabulary Model

What it is

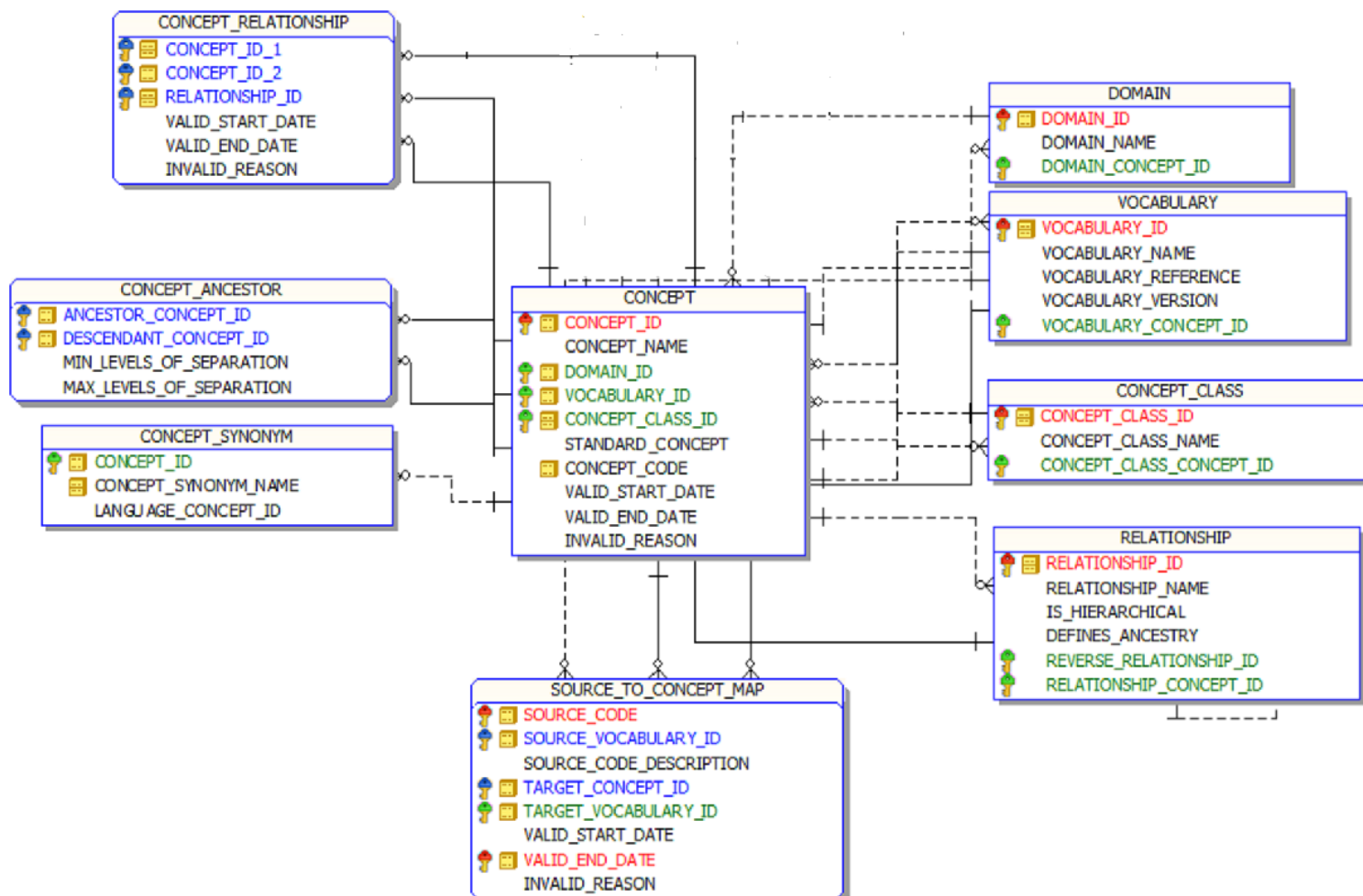
- **Standardized structure** to house existing vocabularies used in the public domain
- **Compiled standards** from disparate public and private sources and some OMOP-grown concepts

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



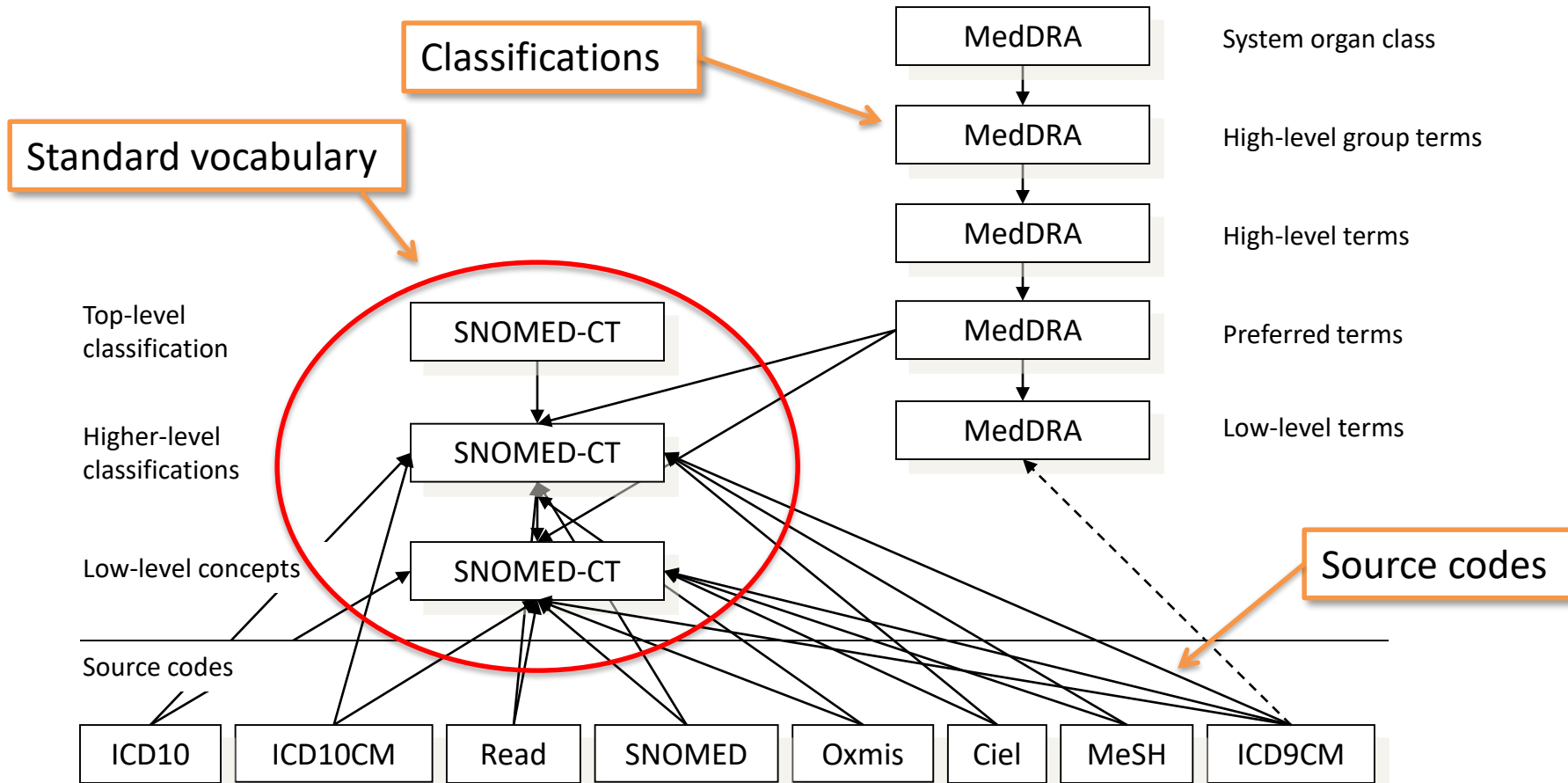
OMOP Vocabulary CDM



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**

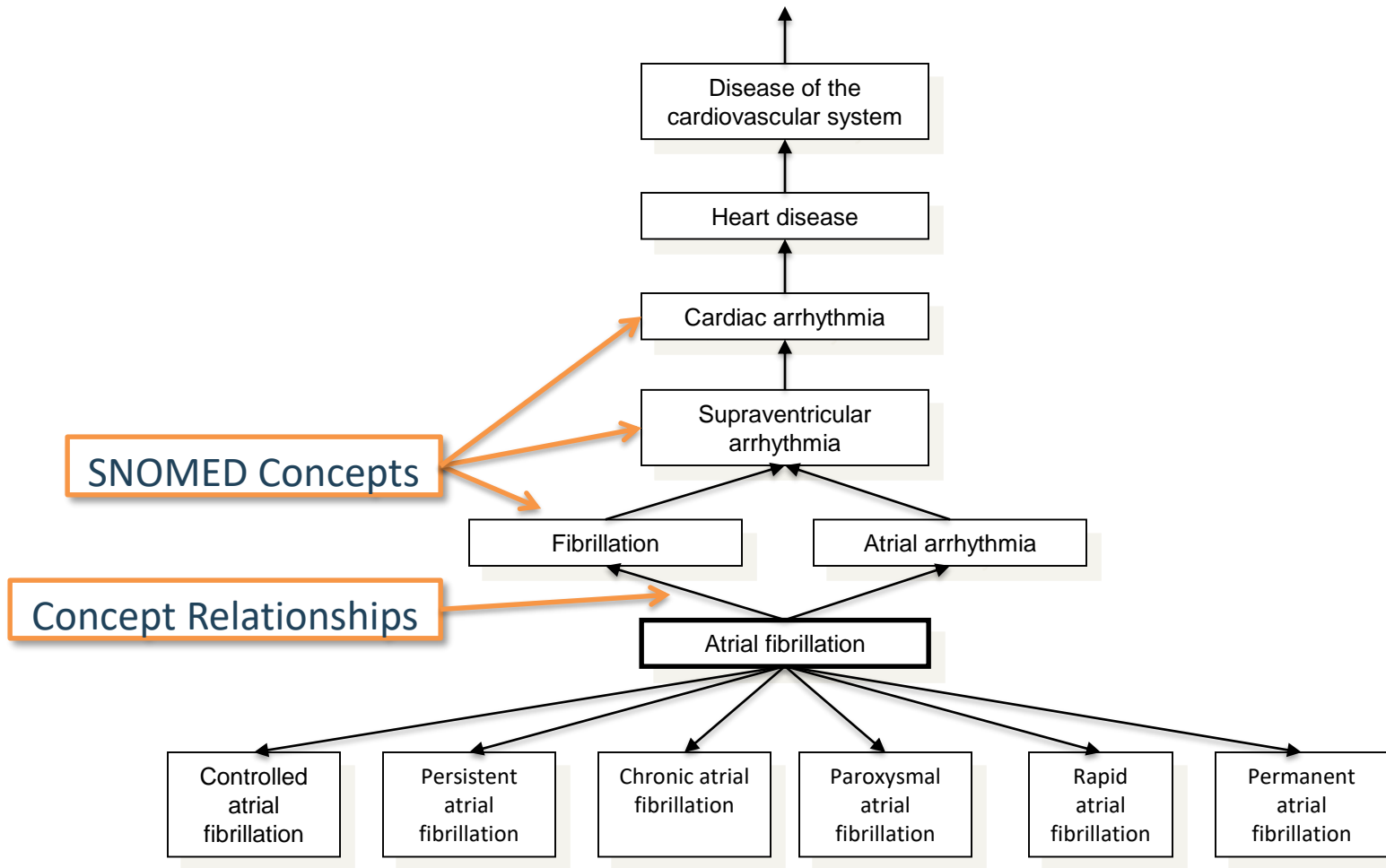


Condition Concepts

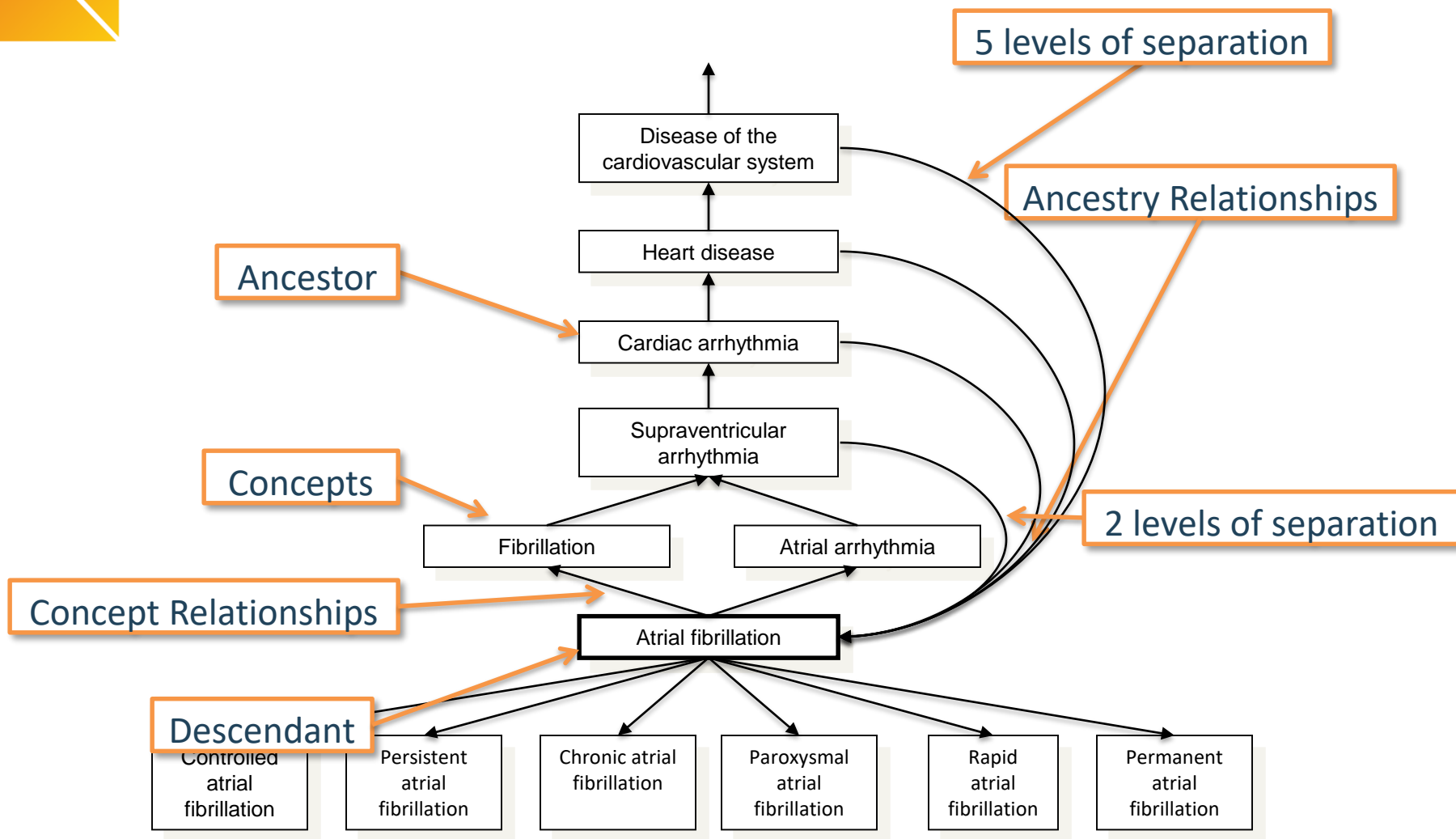




Condition ancestry around “atrial fibrillation”

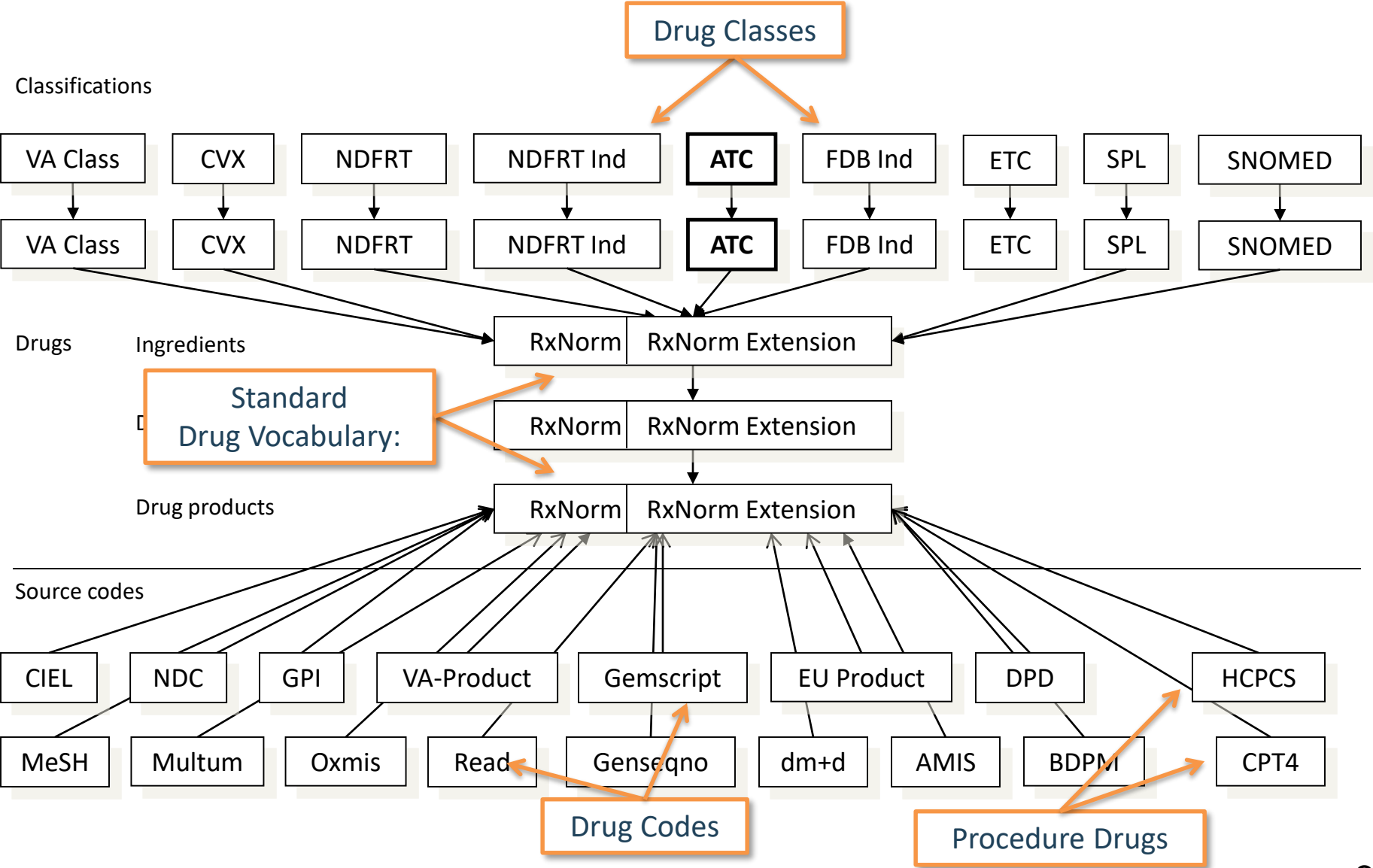


Ancestry Relationships: Higher-Level Relationships





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			

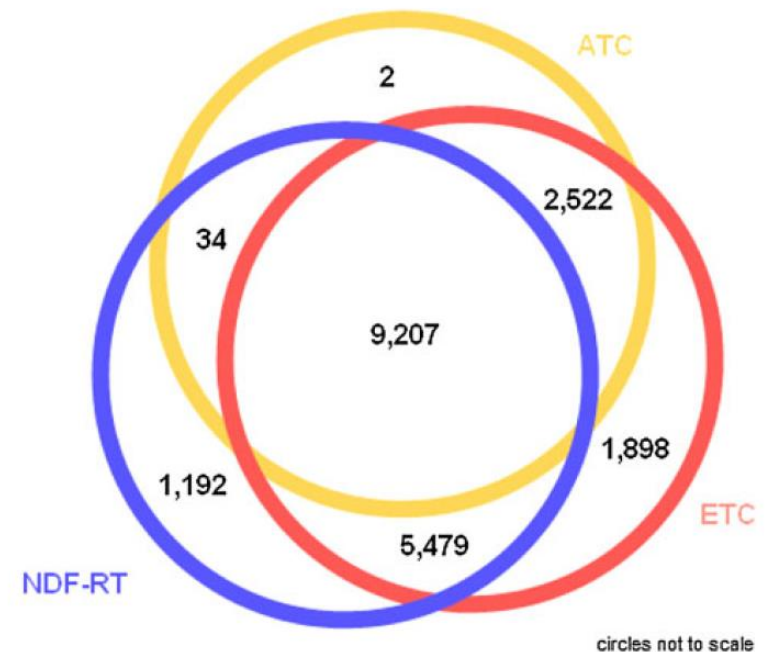


Fig. 1 Overlap in coverage of 'opioid' NDC drug codes by classification system



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

Journal of Biomedical Informatics 45 (2012) 689–696

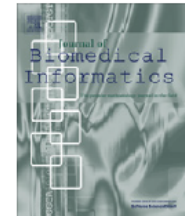


ELSEVIER

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

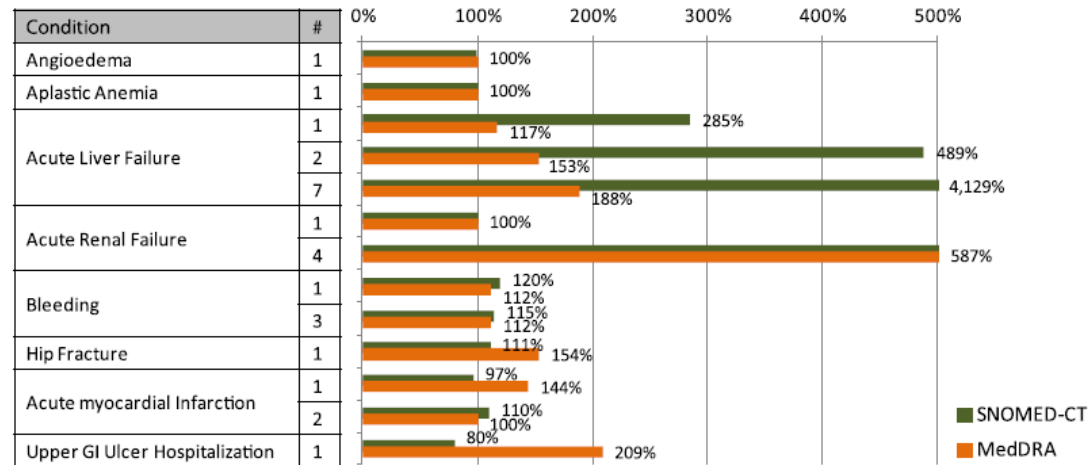
^b Janssen Research & Development, LLC, 1125 Trenton-Harbourton Road, PO Box 200, MS K304, Titusville, NJ 08560, USA

^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

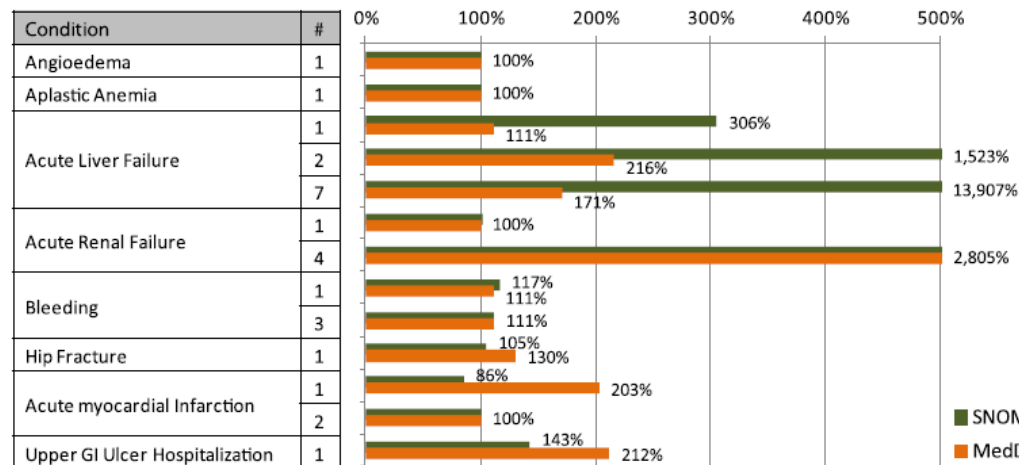


Changing language may change your codelist, that may change your cohort depending on the disease...

Cohort size of HOI in MSLR for different terminologies



Cohort size of HOI in GE for different terminologies





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

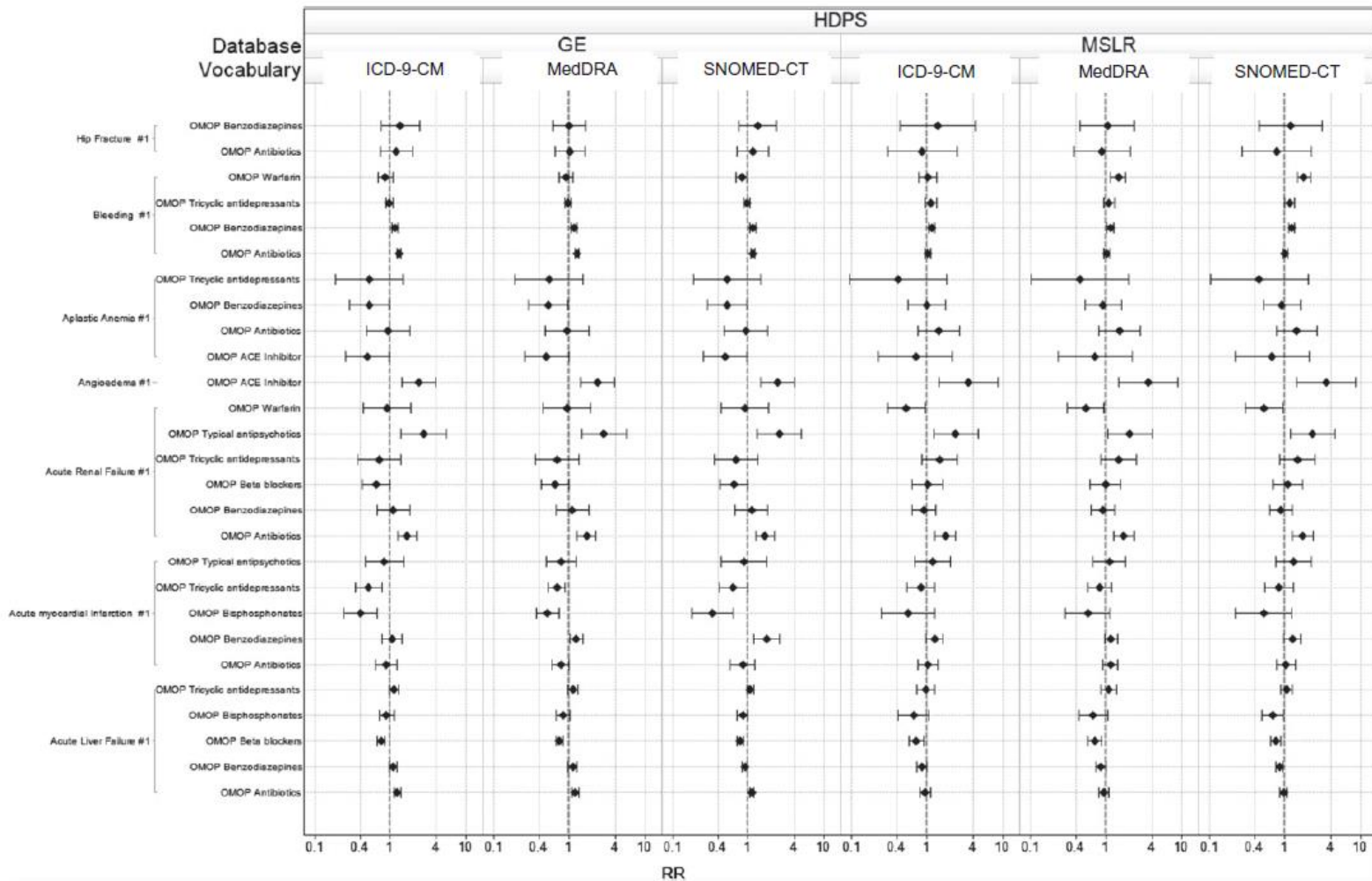


Fig. 3. Effect estimates and 95% confidence intervals for incident user design applied to MSLR and GE using ICD-9-CM, SNOMED-CT, and MedDRA as standard terminologies. Each dot represents the estimate of the effect of an individual HOI-drug combination (on the X-axis).



Demo: Searching the vocabulary in ATLAS

Follow along at:
(during training)


<https://overview.ohdsi.amazingawsdemos.com>

(public demo environment)

<http://ohdsi.org/web/ATLAS>




Demo: vocabulary search

 Vocabulary

Search Import

atrial fibrillation|

 Vocabulary

Search Import

427.31|

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



Demo: vocabulary search results

Vocabulary

Search

Import

atrial fibrillation

Advanced Options

Column visibility

Copy

CSV

Show 15 entries

Filter:

Showing 1 to 15 of 208 entries

Previous 1 2 3 4 5 ... 14 Next

	Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
<div>Vocabulary</div> <div>SNOMED (52)</div> <div>NDC (48)</div> <div>Read (29)</div> <div>MedDRA (14)</div> <div>Indication (12)</div>	21013671	00013671	Ventricular Rate Control in Atrial Fibrillation	Indication	0	136,810,452	Drug	Indication
	4344544	N0000000507	Atrial Fibrillation	Ind / CI	0	133,648,447	Drug	NDFRT
	21013672	00013672	Prevention of Post Cardio-Thoracic Surgery Atrial Fibrillation	Indication	0	85,818,852	Drug	Indication
<div>Class</div> <div>11-digit NDC (48)</div> <div>Read (29)</div> <div>Clinical Finding (27)</div> <div>Procedure (17)</div> <div>UT (12)</div>	313217	49436004	Atrial fibrillation	Clinical Finding	32,938,949	38,400,483	Condition	SNOMED
	35204953	10003658	Atrial fibrillation	PT	0	38,400,483	Condition	MedDRA
	500002401	500002401	OMOP Atrial Fibrillation 1	Cohort	0	38,400,483	Condition	Cohort
<div>Domain</div> <div>Condition (76)</div> <div>Drug (67)</div> <div>Observation (41)</div>	21005673	00005673	Prevention of Thromboembolism in Chronic Atrial Fibrillation	Indication	0	23,812,247	Drug	Indication
	21003018	00003018	Cardioversion of Atrial Fibrillation	Indication	0	5,175,583	Drug	Indication
	21013390	00013390	Prevention of Recurrent Atrial Fibrillation	Indication	0	4,866,791	Drug	Indication



Demo: vocabulary concept selection

Q Vocabulary > Concept

Atrial fibrillation

Details

Related Concepts

Hierarchy

Record Counts

Property	Value
Concept Name	Atrial fibrillation
Domain Id	Condition
Concept Class Id	Clinical Finding
Vocabulary Id	SNOMED
Concept Id	313217
Concept Code	49436004
Invalid Reason	Valid
Standard Concept	Standard



Demo: Concept relationship exploration

Vocabulary > Concept

Atrial fibrillation

Details

Related Concepts

Hierarchy

Record Counts

Column visibility

Copy

CSV

Show 15 entries

Filter:

Showing 1 to 5 of 5 entries

Previous 1 Next

Vocabulary

SNOMED (50)

MedDRA (25)

Indication (18)

CIEL (10)

Read (0)

Standard Concept

Classification (46)

Standard (45)

Non-Standard (37)

Invalid Reason

Valid (122)

Invalid (6)

Class

Clinical Finding (42)

	Id	Code	Name	Class	RC	DRC	Distance	Domain	Vocabulary
	321588	56265001	Heart disease	Clinical Finding	1,212,872	193,512,502	3	Condition	SNOMED
	44784217	698247007	Cardiac arrhythmia	Clinical Finding	655,834	59,608,937	2	Condition	SNOMED
	4248028	72654001	Supraventricular arrhythmia	Clinical Finding	0	44,910,236	2	Condition	SNOMED
	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

Showing 1 to 5 of 5 entries

Previous 1 Next



Demo: Concept hierarchy

Q Vocabulary > Concept

Atrial fibrillation

Details

Related Concepts

Hierarchy

Record Counts

Parents

	<div>Id</div>	<div>Code</div>	<div>Name</div>	<div>Class</div>	<div>RC</div>	<div>DRC</div>	<div>Distance</div>	<div>Domain</div>	<div>Vocabulary</div>
	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

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

Children

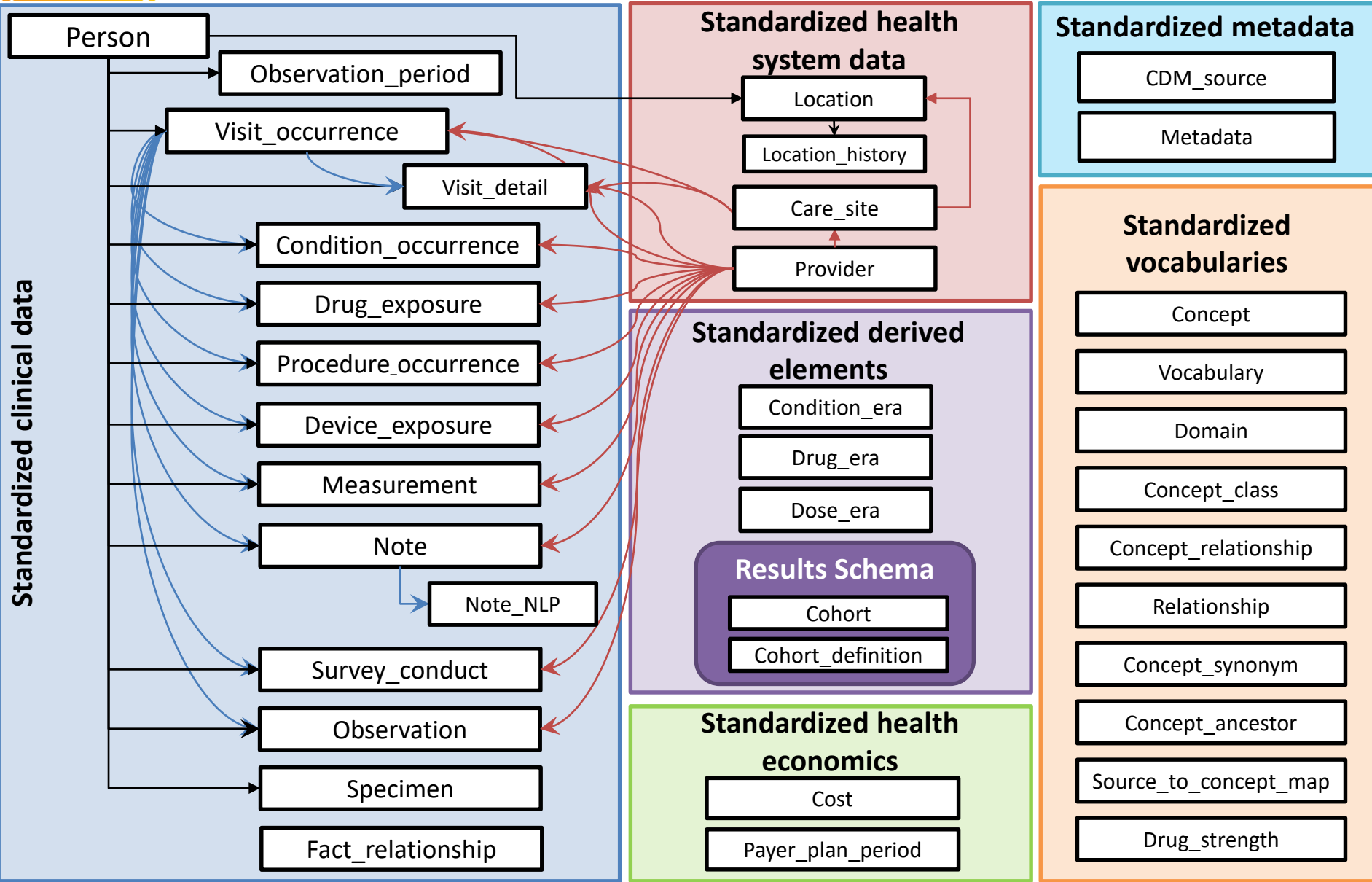
	<div>Id</div>	<div>Code</div>	<div>Name</div>	<div>Class</div>	<div>RC</div>	<div>DRC</div>	<div>Distance</div>	<div>Domain</div>	<div>Vocabulary</div>
	4154290	282825002	Paroxysmal atrial fibrillation	Clinical Finding	2,919,750	2,919,750	1	Condition	SNOMED



Data Sources



OMOP CDM Version 6





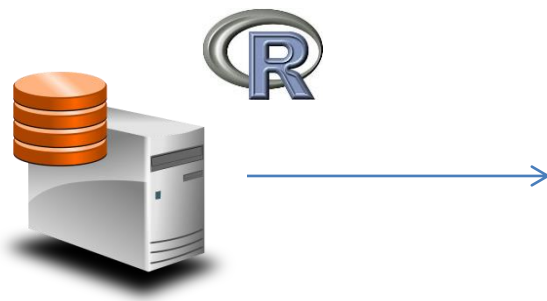
Purpose of Achilles

- 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.

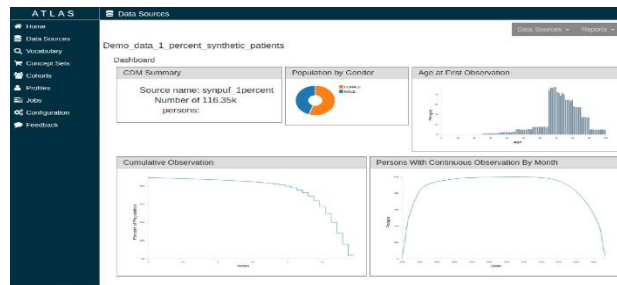


Achilles – Under the Hood

- 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-identified) 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)



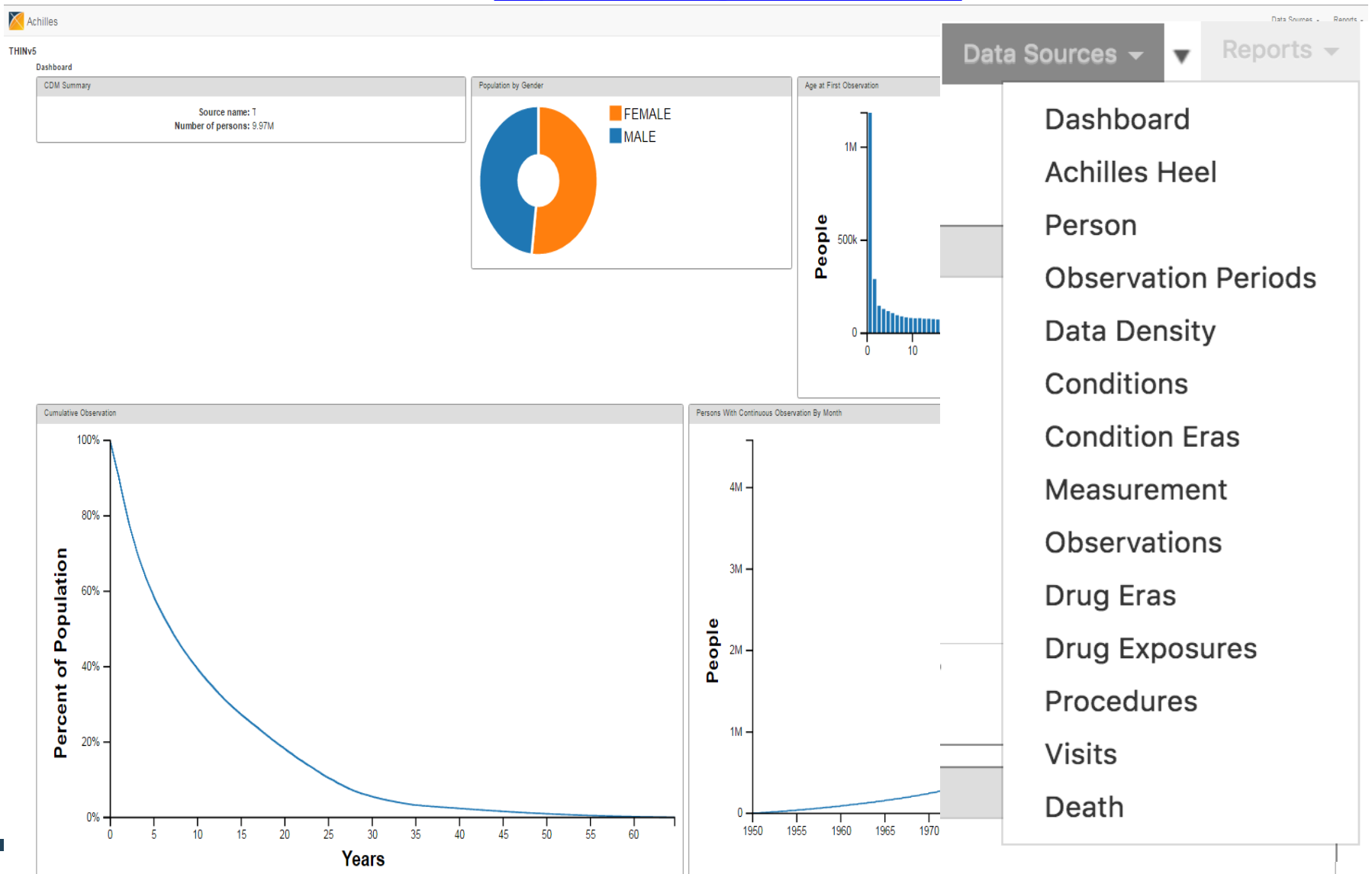
OMOP CDM





Basic Navigation

<http://ohdsi.org/web/ACHILLES>





Examples

ATLAS

Home

Data Sources

Vocabulary

Concept Sets

Cohorts

Profiles

Jobs

Configuration

Feedback

Data Sources

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

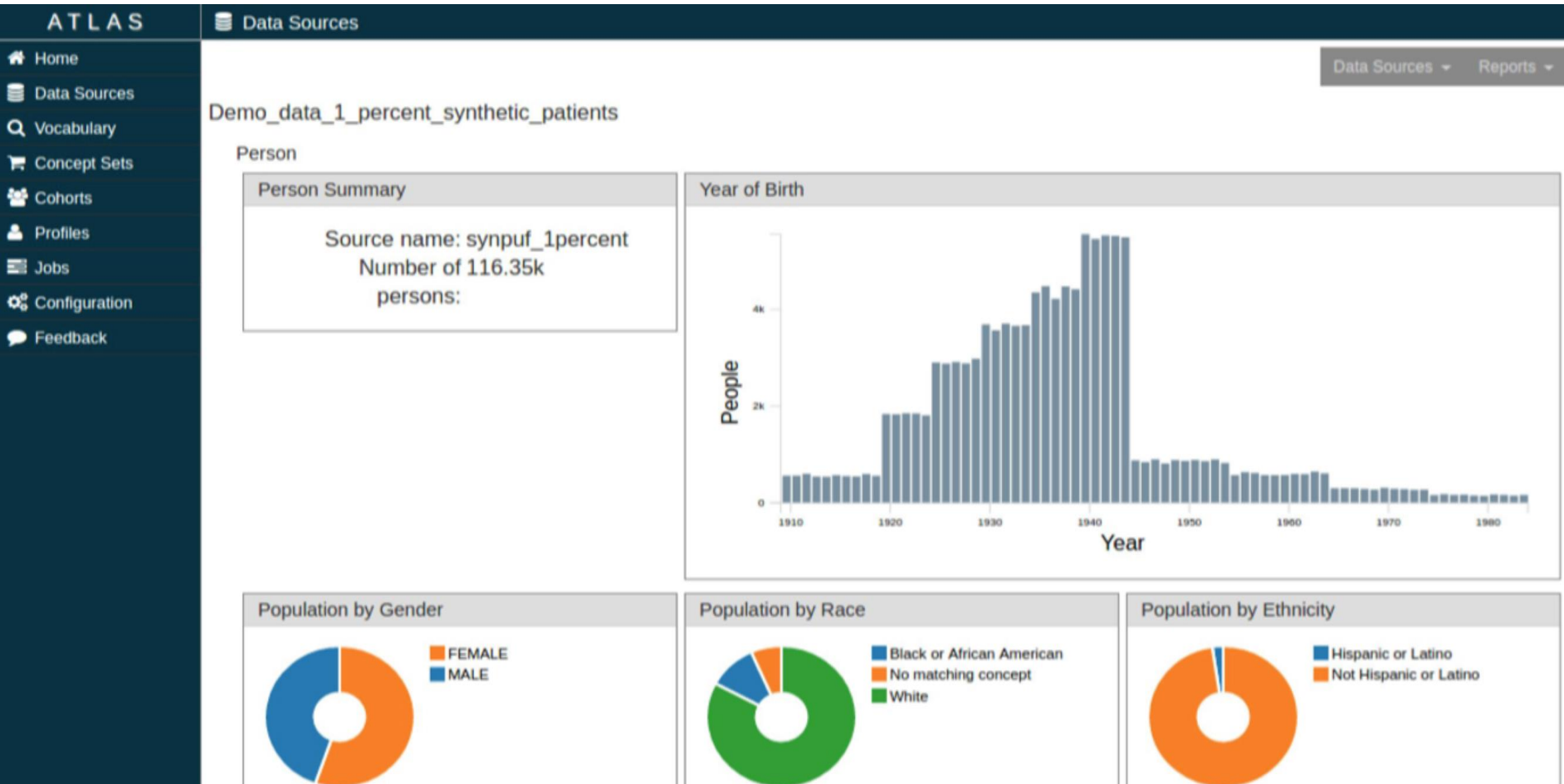
Date

<http://ohdsi.org/web/ATLAS>

54



Examples





Examples

ATLAS

- 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

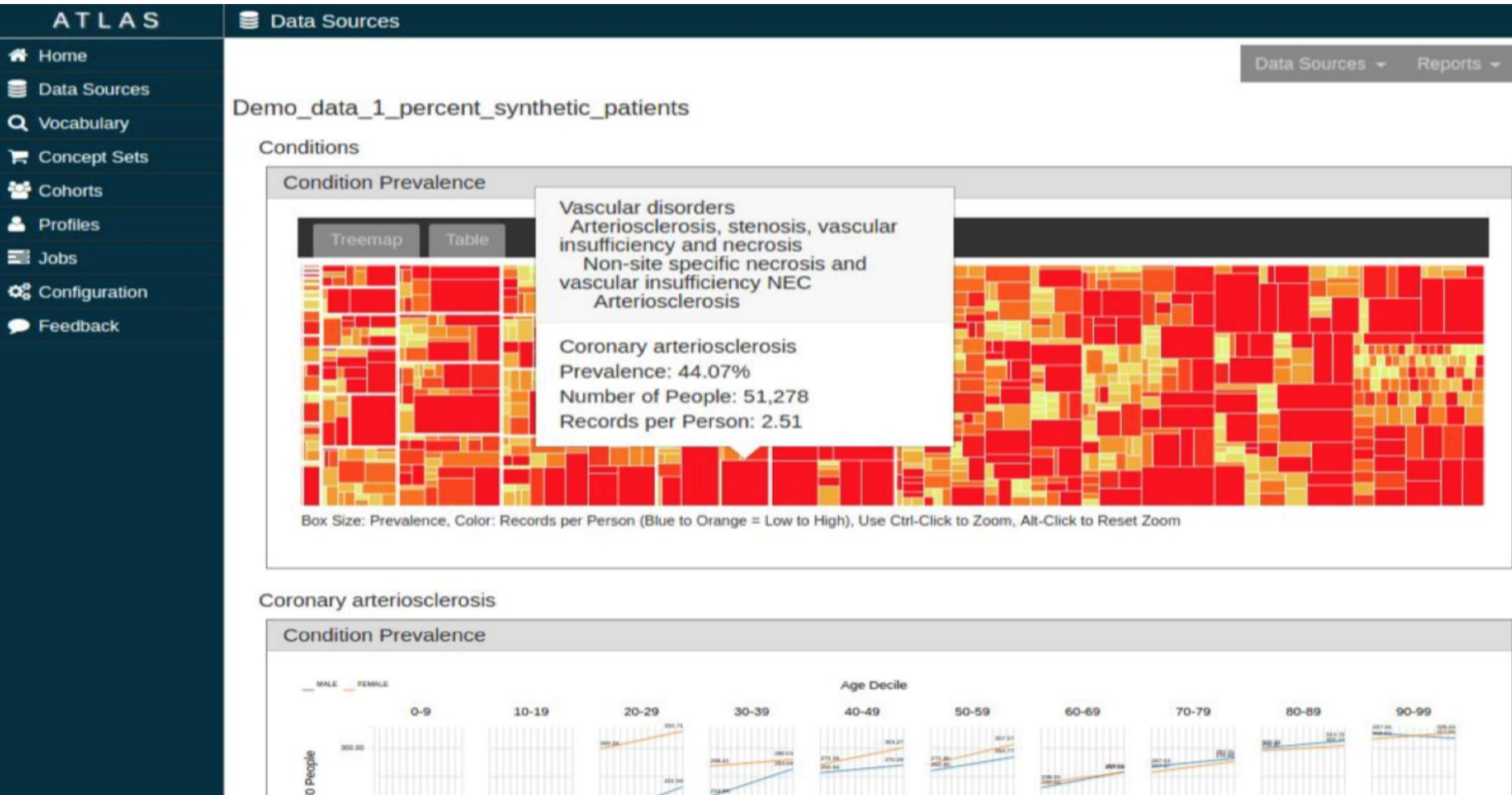
Data Sources

Reports

- Dashboard
- Achilles Heel
- Person
- Observation Periods
- Data Density
- Conditions
- Condition Eras
- Measurement
- Observations
- Drug Eras
- Drug Exposures
- Procedures
- Visits
- Death

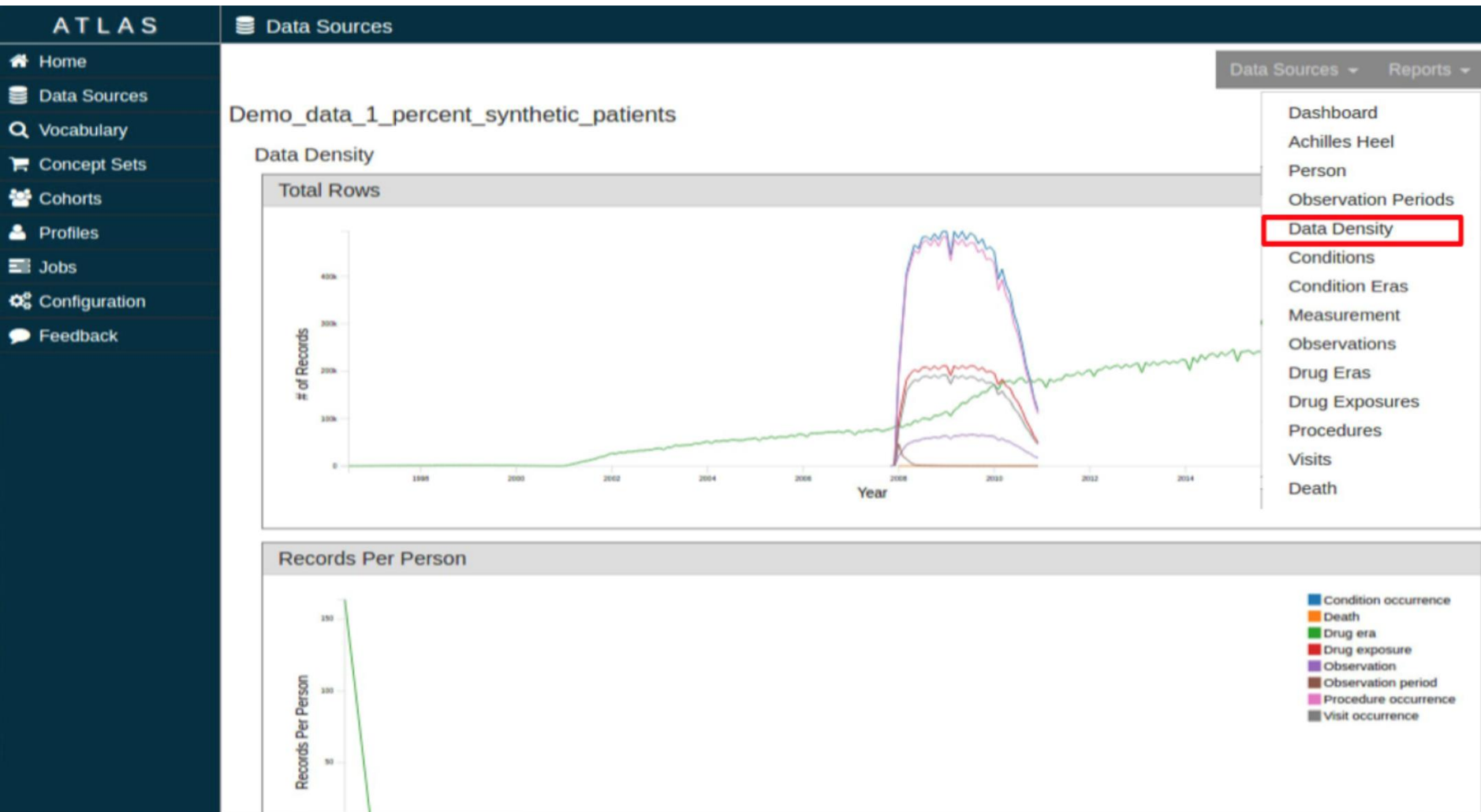


Examples





Examples





Examples

ATLAS

- Home
- Data Sources
- Vocabulary
- Concept Sets
- Cohorts
- Profiles
- Jobs
- Configuration
- Feedback

Data Sources

Demo_data_1_percent_synthetic_patients

Achilles Heel Report

Data Quality Messages

Message Type	Message
ERROR	400-Number of persons with at least one condition occurrence, by condition_id; concepts in data are not in correct vocabulary
ERROR	600-Number of persons with at least one procedure occurrence, by procedure_id; concepts in data are not in correct vocabulary
ERROR	900-Number of persons with at least one drug era, by drug_concept_id; concepts in data are not in correct vocabulary
ERROR	908-Number of drug eras without valid person; count (n=23,452,537) should not be > 0
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

Data Sources

Reports

Dashboard

Achilles Heel

Person

Observation Periods

Data Density

Conditions

Condition Eras

Measurement

Observations

Drug Eras

Drug Exposures

Procedures

Visits

Death



Cohort: Definition and characterization



Construct of a 'Cohort'

Cohort

From Wikipedia, the free encyclopedia

Cohort may refer to:

Phenotype

- ★ **Cohort (statistics)**, a group of subjects with a common defining characteristic for example age group
- **Cohort study**, a form of longitudinal study used in medicine and social science



Defining 'phenotype'

Journal of the American Medical Informatics Association, 0(0), 2017, 1–6

doi: 10.1093/jamia/ocx110

Perspective



OXFORD

Perspective

High-fidelity phenotyping: richness and freedom from bias

George Hripcsak¹ and David J Albers¹

- 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 experts and knowledge engineers, or through diverse forms of machine learning to generate novel representations of data.

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

RECEIVED 8 January 2015

REVISED 14 July 2015

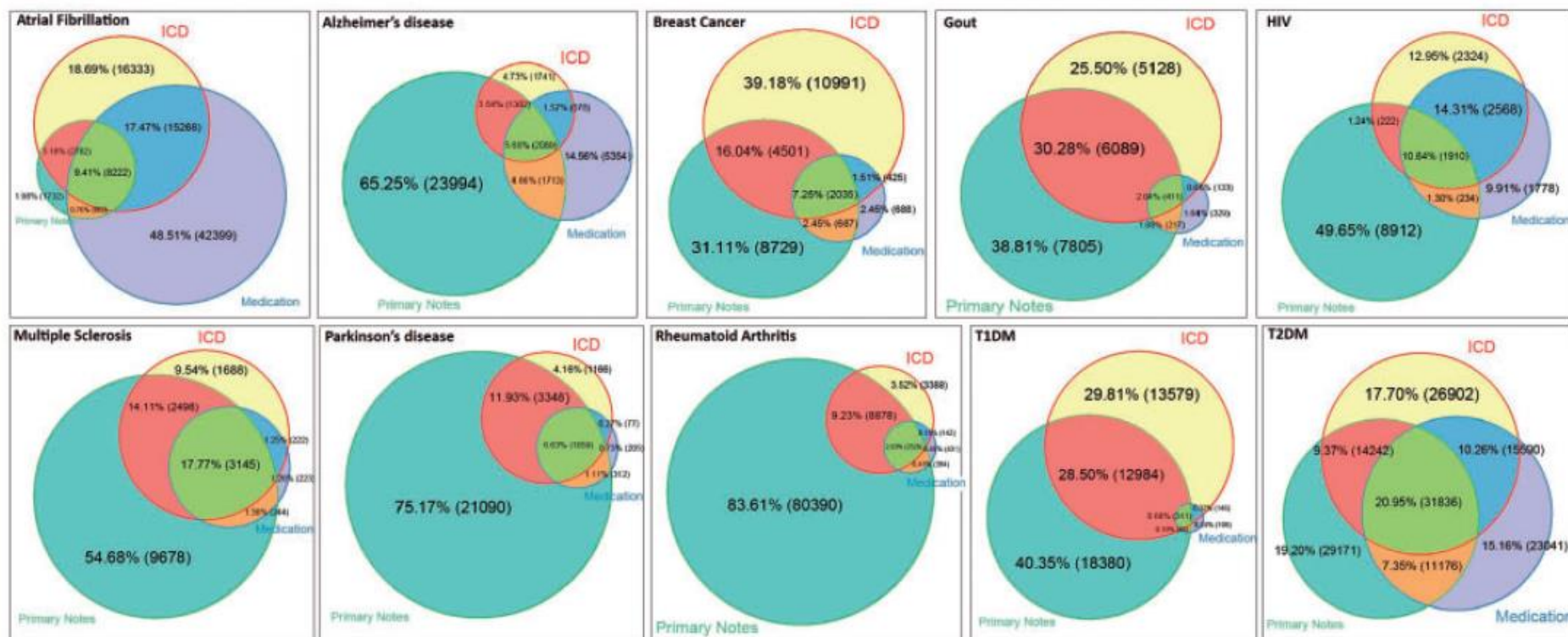
ACCEPTED 15 July 2015

PUBLISHED ONLINE FIRST 2 September 2015



Wei-Qi Wei¹, Pedro L Teixeira¹, Huan Mo¹, 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.





Two Approaches to Phenotyping

Rule-Based
Phenotyping

Probabilistic
Phenotyping

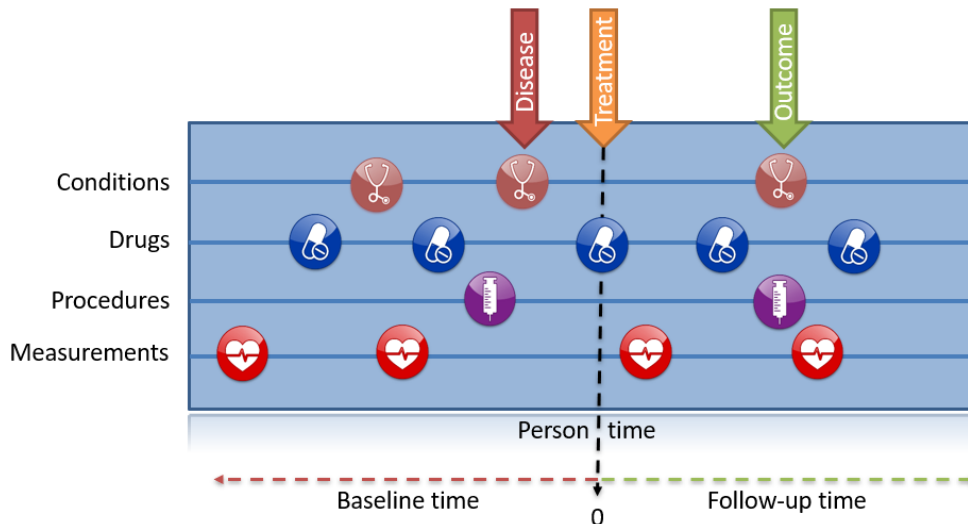


Data are the building blocks for Phenotyping

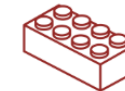
Person level dataset

With time-stamped events

Events organized in domains



Conditions



Drugs



Procedures



Measurements



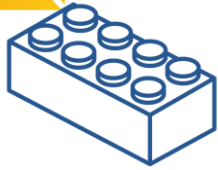
Observations



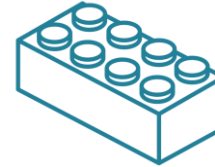
Visits



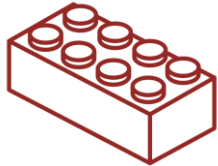
Data are the building blocks for Phenotyping



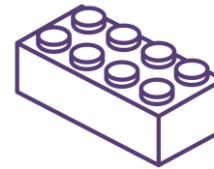
Conditions



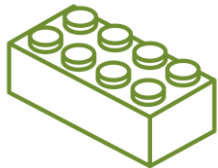
Sponsor



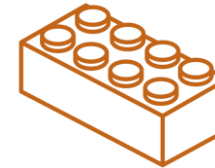
Drugs



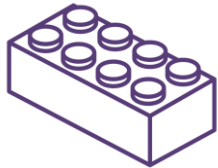
Benefit Plan



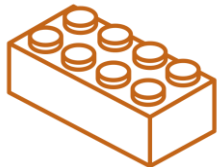
Procedures



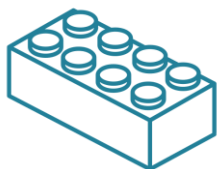
Cost



Measurements



Observations



Visits



OHDSI's definition of 'cohort'

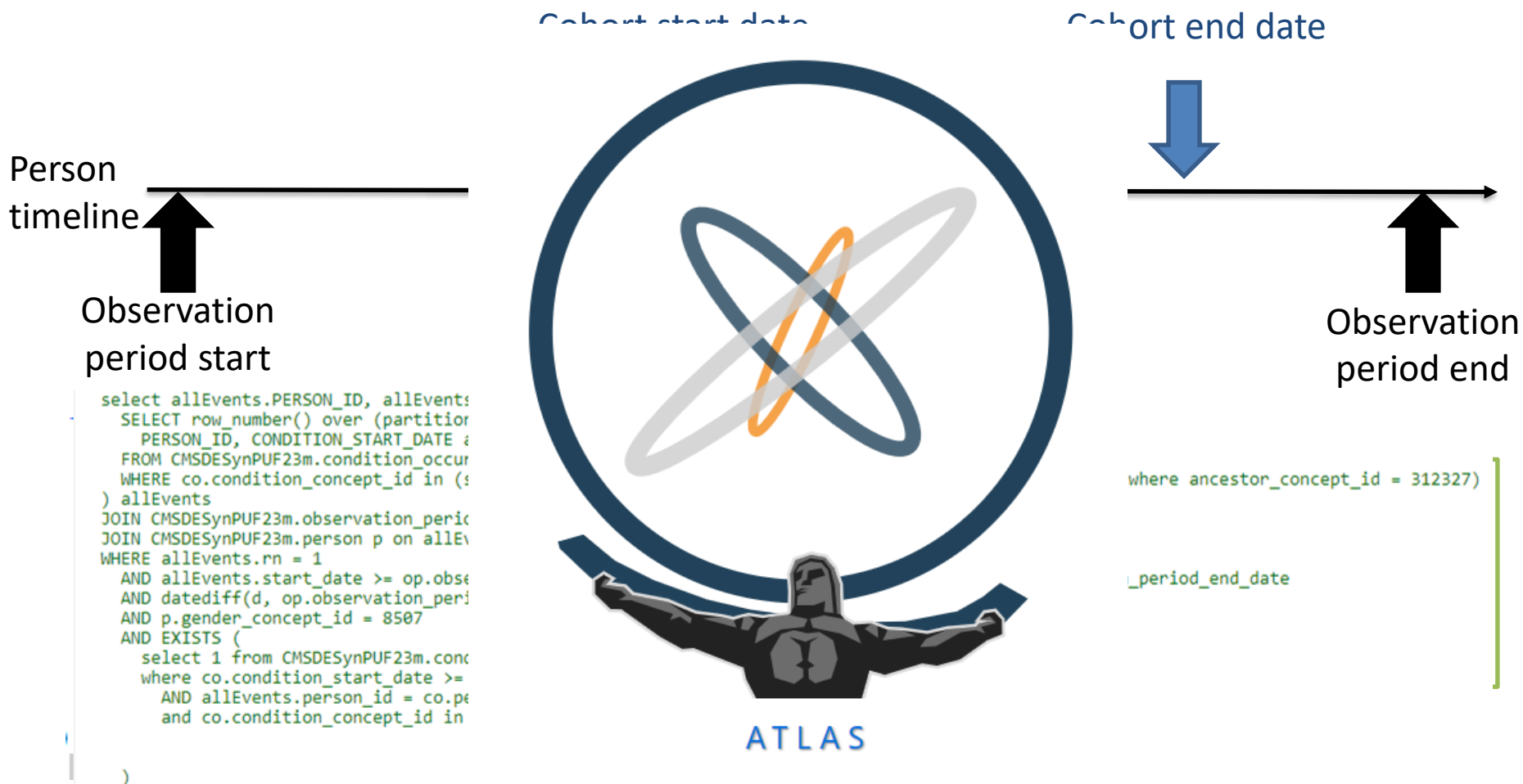
Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

- 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

Cohort = Phenotype for a duration of time

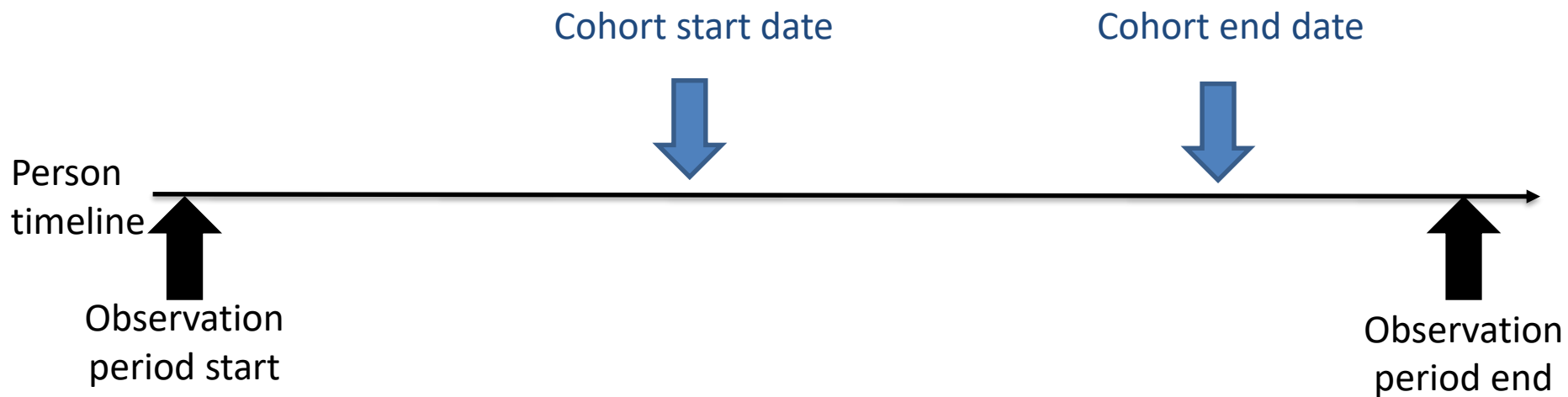


OHDSI's definition of 'cohort'





Anatomy of rule based cohort definition



Criteria: rules, specific to a CDM, that is used to identify records (events) from patient data

Criteria Types:

1. Cohort entry events
2. Inclusion criteria
3. Cohort exit
4. Cohort eras





Rule based cohort definition in Atlas

Cohort Entry Events

Cohort Entry Events: CDM records selected from patient data that qualify the person for presence in a cohort.

+ Add Initial Event ▾

Restrict initial events

Inclusion Criteria

New inclusion criteria

Inclusion Criteria: Use other patient data relative to the initial event to qualify the initial event as valid for inclusion. “Attrition study diagram.”

Limit qualifying events to: earliest event ▾ per person.

Cohort Exit

Event Persistence:

Event will persist until: end of continuous ob

Censoring Events:

Exit Cohort based on the following criteria:

No censoring events selected.

Cohort Exit: Modify how long each event persists the patient’s presence in the cohort, or identify records that censor the event.

- End of observation period
- Fixed time relative to event start date
- Censoring observations, e.g. procedure, or end of therapy

Cohort Eras

- Specify era collapse gap size: 0 ▾ days
- [add trimming options...](#)

Cohort Eras: Chain remaining event dates with an allowable gap with optional left- and/or right-censor the final eras (trimming).



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



Rule based cohort definition in Atlas

Cohort Entry Events

Cohort Entry Events: CDM records selected from patient data that qualify the person for presence in a cohort.

+ Add Initial Event ▾

Restrict initial events

Inclusion Criteria

New inclusion criteria

Inclusion Criteria: Use other patient data relative to the initial event to qualify the initial event as valid for inclusion. “Attrition study diagram.”

Limit qualifying events to: earliest event ▾ per person.

Cohort Exit

Event Persistence:

Event will persist until: end of continuous ob

Censoring Events:

Exit Cohort based on the following criteria:

No censoring events selected.

Cohort Exit: Modify how long each event persists the patient’s presence in the cohort, or identify records that censor the event.

- End of observation period
- Fixed time relative to event start date
- Censoring observations, e.g. procedure, or end of therapy

Cohort Eras

- Specify era collapse gap size: 0 ▾ days
- [add trimming options...](#)

Cohort Eras: Chain remaining event dates with an allowable gap with optional left- and/or right-censor the final eras (trimming).



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.



Demo: Implementing cohorts in ATLAS

Follow along at:

(during training)

<https://overview.ohdsi.amazingawsdemos.com>

(public demo environment)

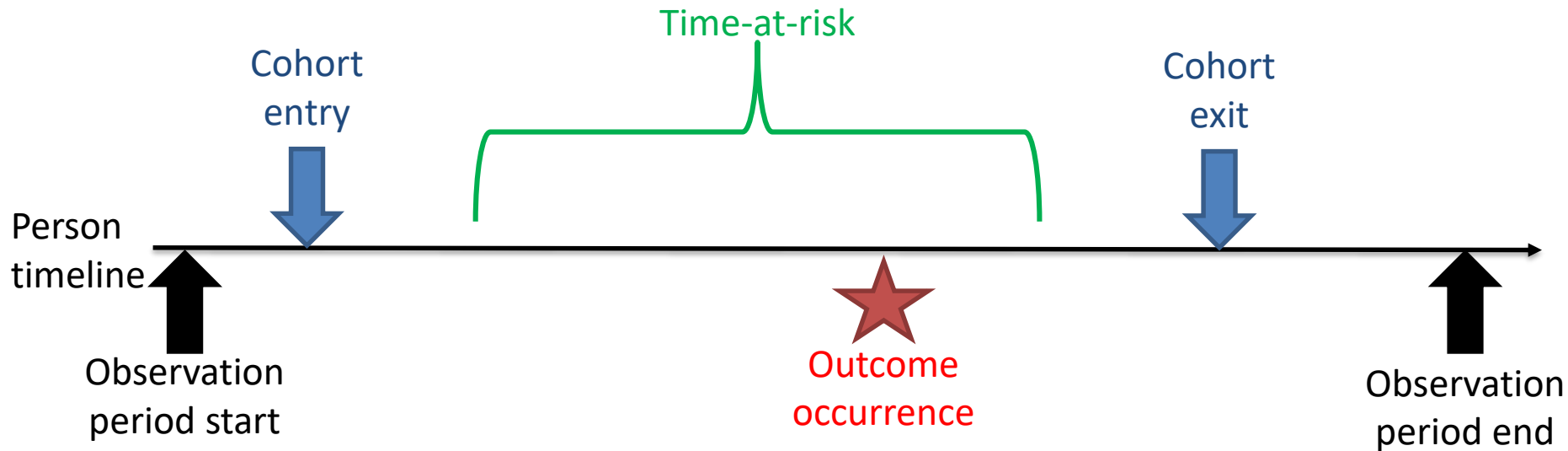
<http://ohdsi.org/web/ATLAS>



Incidence rate



Dissecting the anatomy of incidence



Incidence metrics:

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

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



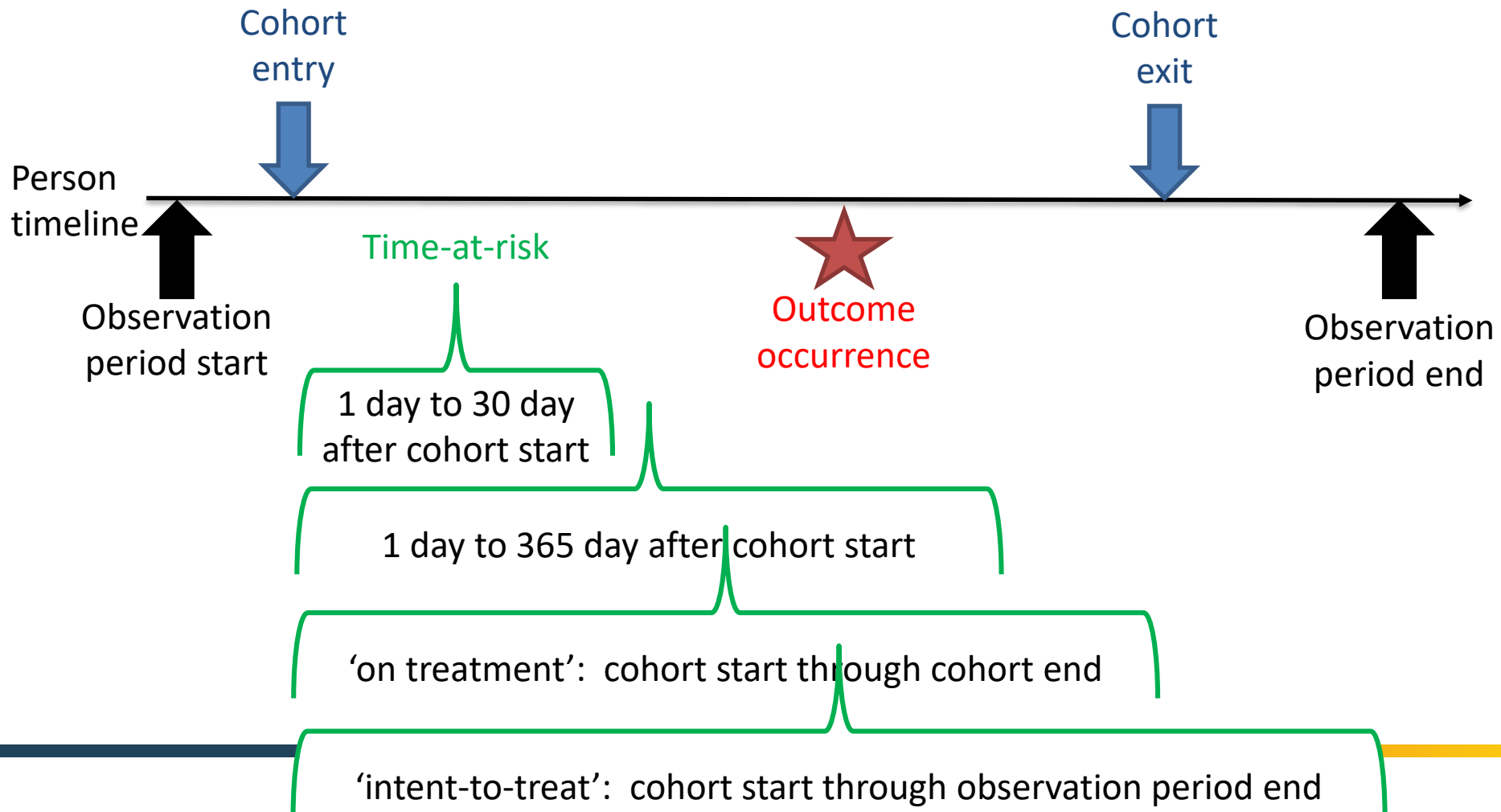
Myriad difficult choices that researchers have to make to produce a 'simple answer'

- 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:
(during training)

<https://overview.ohdsi.amazingawsdemos.com>

(public demo environment)

<http://ohdsi.org/web/ATLAS>



Demo: incidence rate specification

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates

Profiles

Estimation

Prediction

Jobs

Apache 2.0
open source software

provided by
OHDSI
join the journey.

Incidence Rate Analysis

[OHDSI Ecosystem tutorial] Outcomes following warfarin exposure

Generate...

Definition

Concept Sets

Generation

Utilities

Study Cohorts

Target Cohorts

#2: [OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation

#10: [OHDSI Ecosystem tutorial] Graham replication: target cohort 2 - digoxin new users with prior atrial fibrillation

Add Target Cohort

Outcome Cohorts

#3: [OHDSI Ecosystem tutorial] Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting

#4: [OHDSI Ecosystem tutorial] Graham replication: outcome cohort #2 - incident intracranial hemorrhage, observed in inpatient setting

#5: [OHDSI Ecosystem tutorial] Graham replication: outcome cohort #3 - incident major gastrointestinal (GI) bleeding events, observed in inpatient setting

Add Outcome Cohort

Time At Risk

Time at risk defines the time window relative to the cohort start or end date with an offset to consider the person 'at risk' of the outcome.

Time at risk starts with

start date

plus

1

days.

Time at risk ends with

start date

plus

365

days.

No study window defined.

Add Study Window

Stratify Criteria: You can provide optional stratification criteria to the analysis that will divide the population into unique groups based on their satisfied criteria.

New stratify criteria

1. age >= 75

2. Female

3. has heart failure

age >= 75

enter an inclusion rule description

having all of the following criteria:

+ Add criteria to group...



Demo: incidence rate generation

Incidence Rate Analysis

[OHDSI Europe tutorial] Cardiovascular and Bleeding Risks in Elderly Medicare Patients Treated with Oral Anticoagulants



Generate...

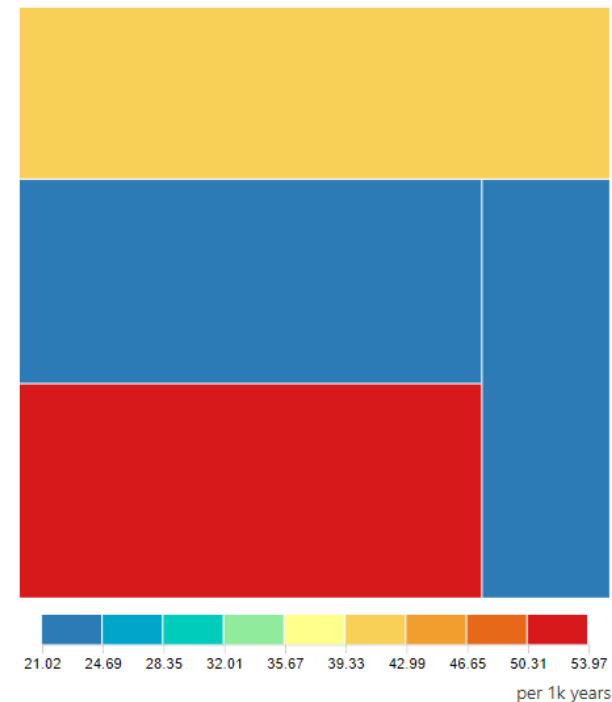
Definition Concept Sets Generation Utilities

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 replication and outcome cohort: [OHDSI Europe tutorial] Graham replication

	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?

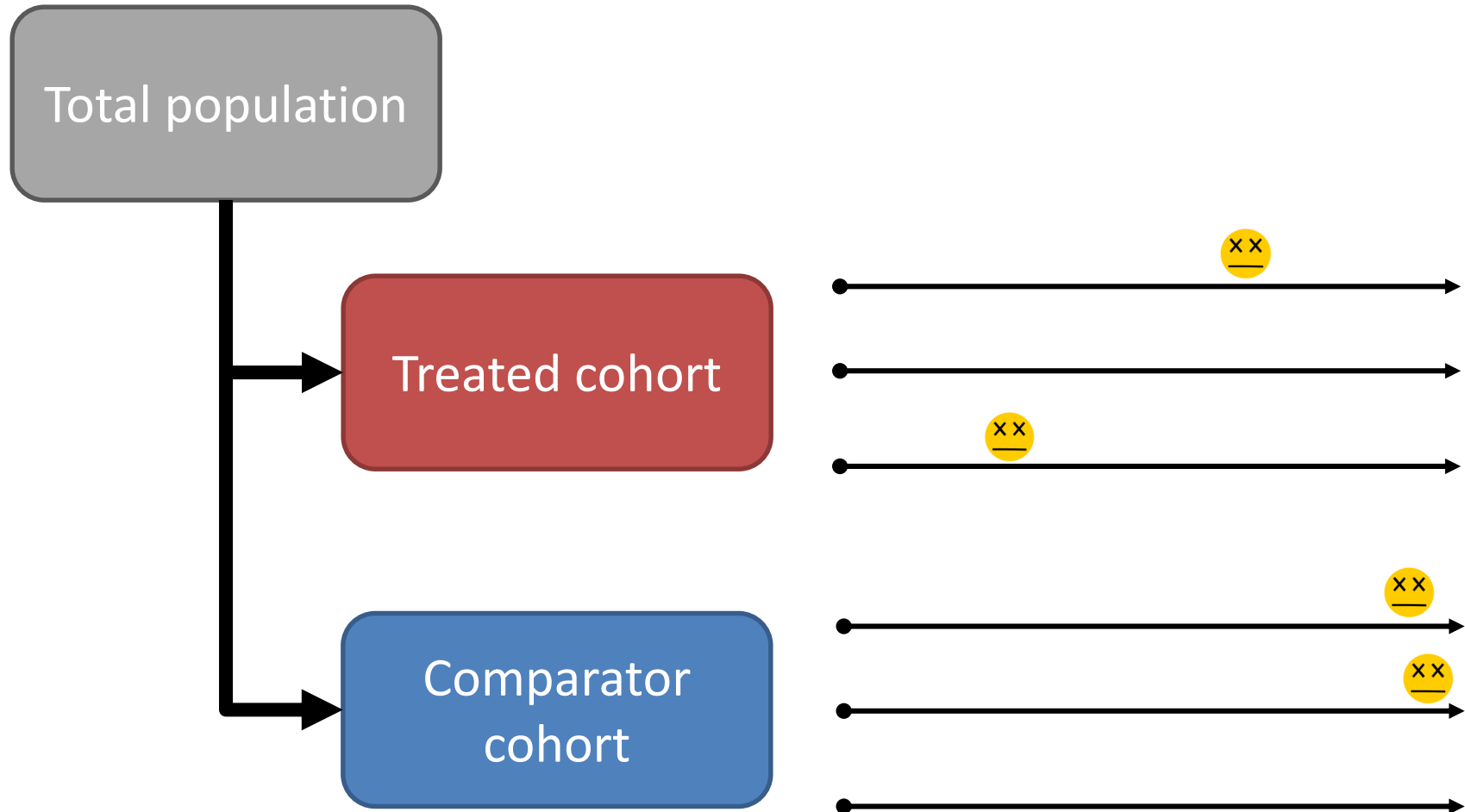
Effect estimation

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

Comparative effect estimation

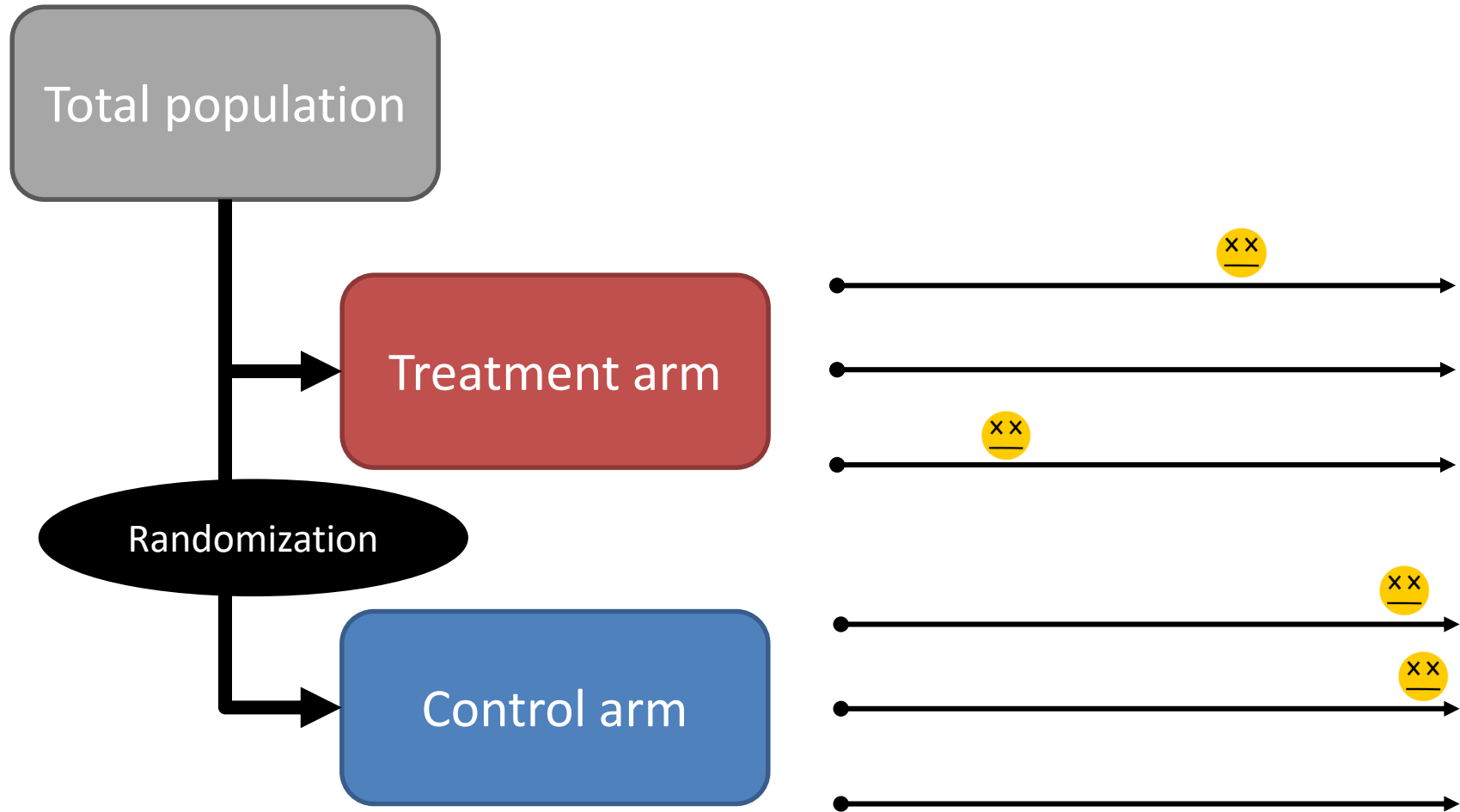


New-user cohort design





Randomized controlled trial





New-user cohort design

Total population

Treatment assignment is not random!

Doctors have reasons why they prescribe a drug to some patients and not to others

comparator cohort

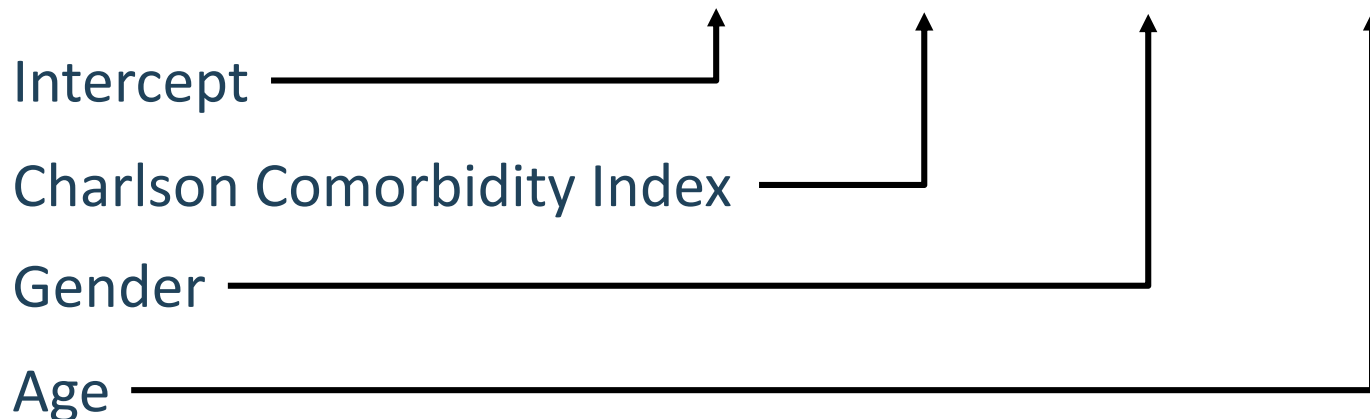




Propensity score (PS)

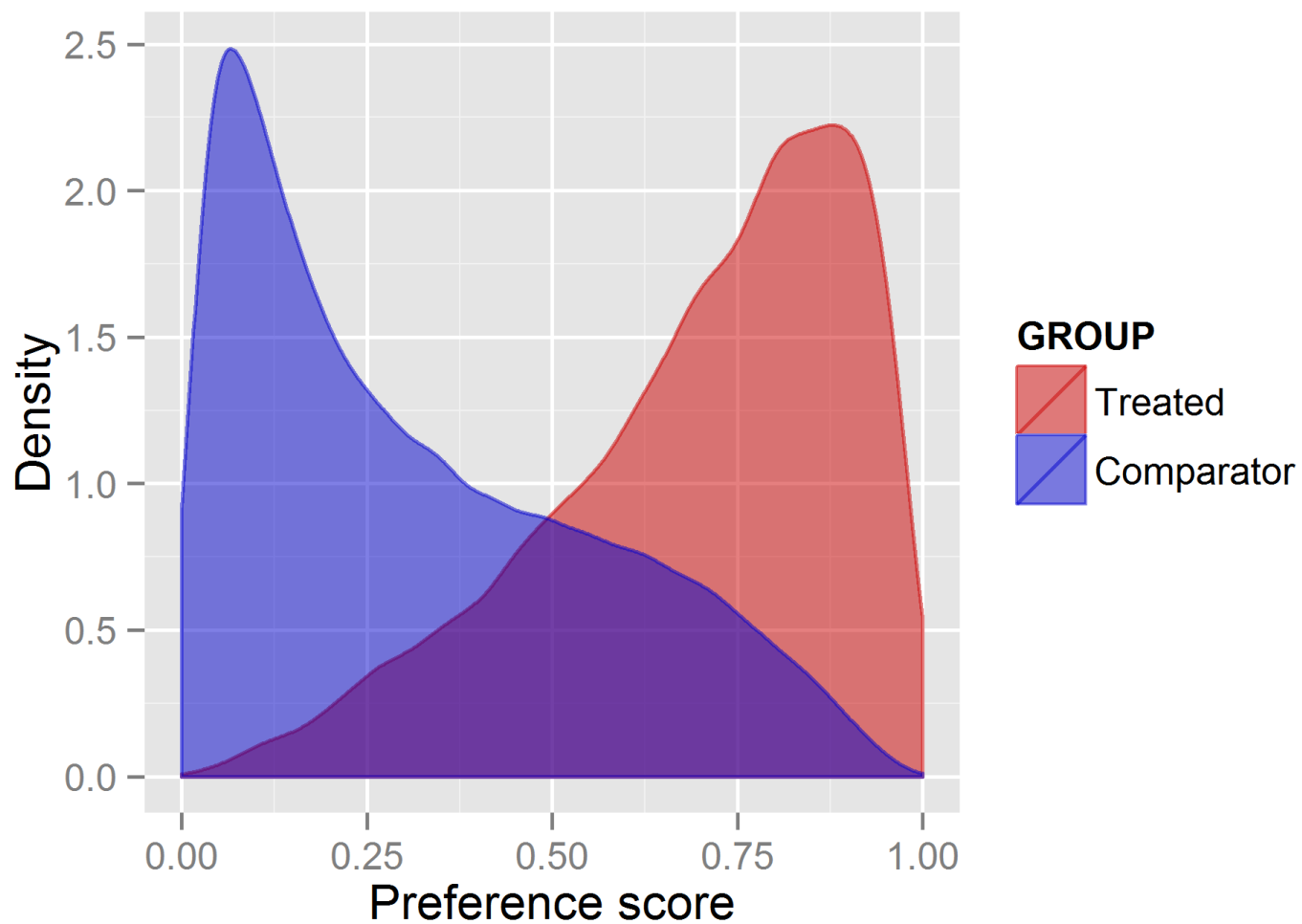
The propensity score is the probability of receiving the treatment, conditional on a set of baseline characteristics

$$P(\text{treatment} \mid X) = f(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots)$$





PS score distribution





Using the PS

- **Trimming**

if $P(\text{treatment})$ is around 50%, treatment assignment 'must be random'

- **Stratification or matching**

only compare subjects to subjects with a similar PS

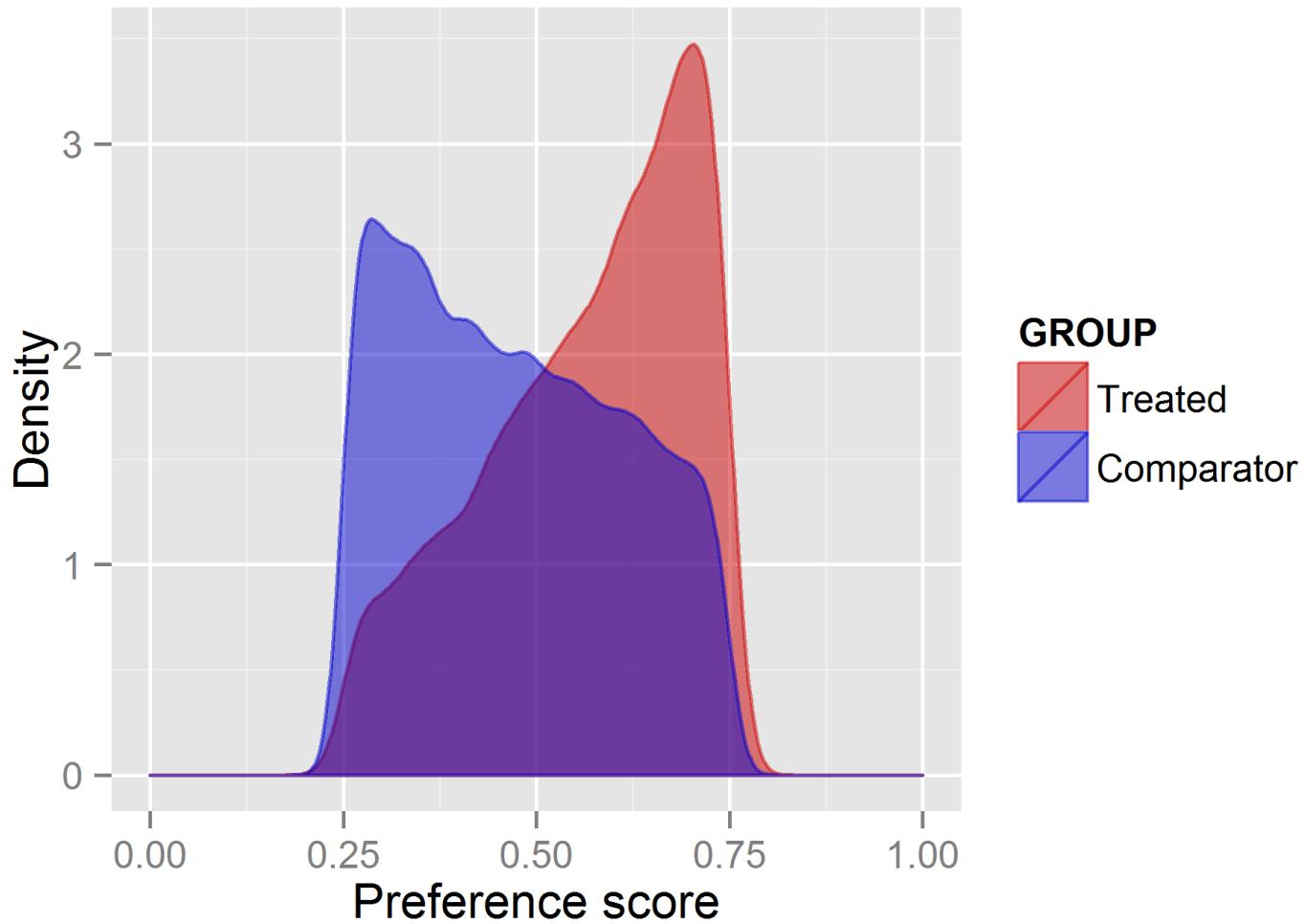
- **Inverse probability weighting**

- **Adding to the outcome model**

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

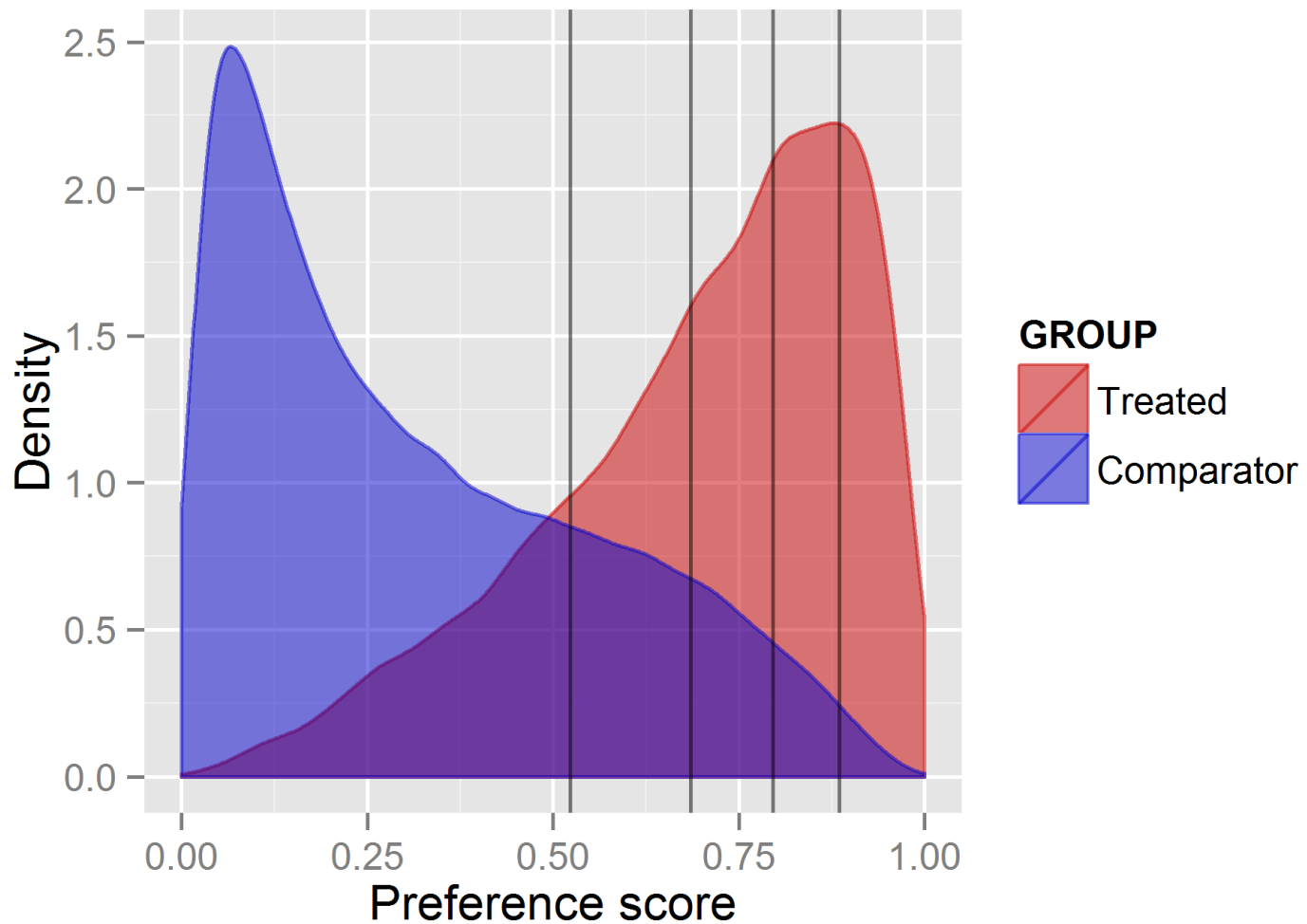


Trimming



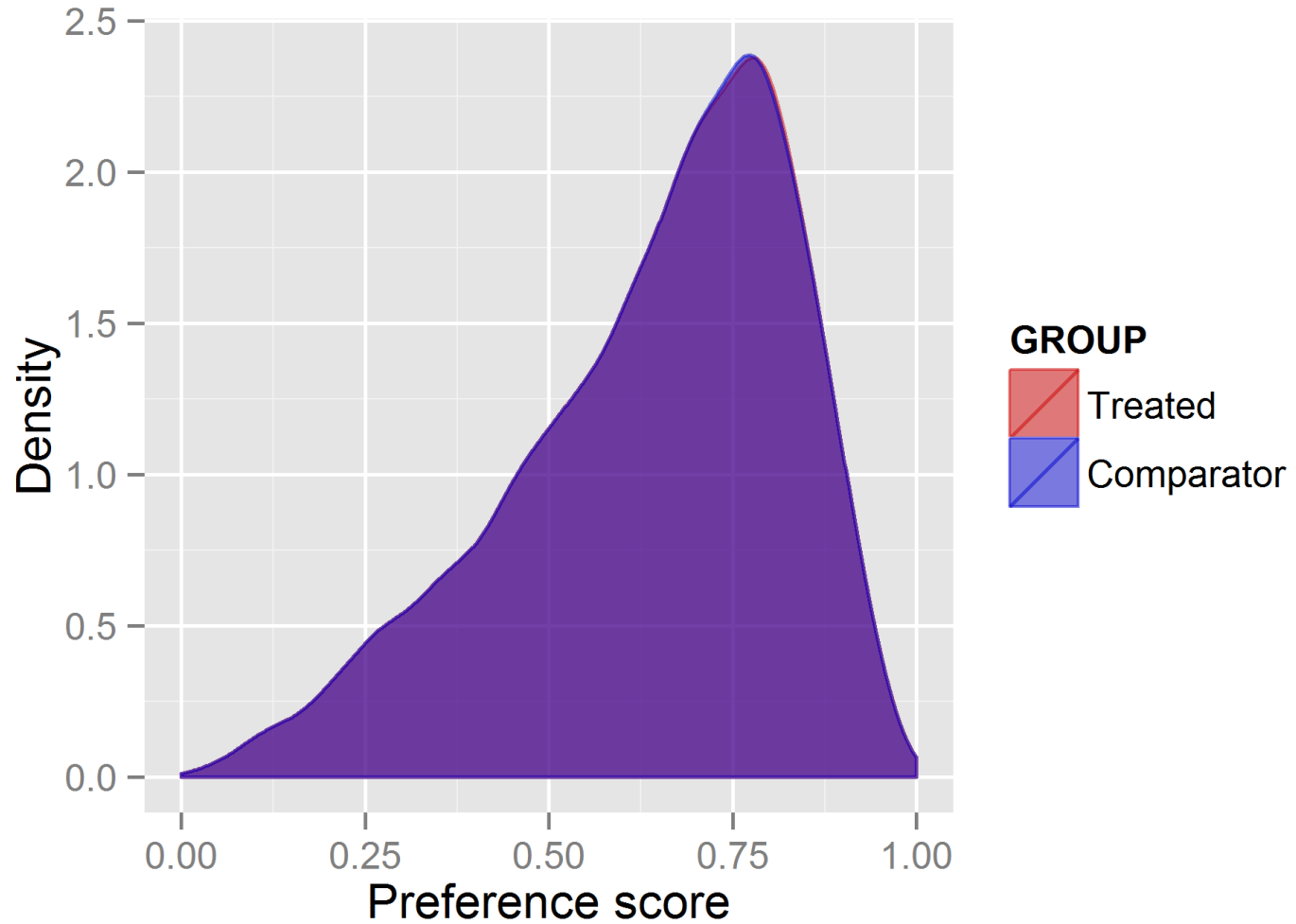


Stratifying





Matching





Which variables go into the PS model?

- Traditional: hard thinking by expert
- High-Dimensional PS: rank many variables (e.g. all drugs, all drug classes, all conditions, all disease classes, all procedures, all observations, all severity indexes) in a regularized regression
 - 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



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 cohort selection strategy

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates

Profiles

Estimation

Prediction

Jobs

Apache 2.0
open source software

Population Level Effect Estimation - Comparative Cohort Analysis

[OHDSI Ecosystem tutorial] Warfarin vs. Dabigatran for risk of ischemic stroke

Comparison

Add or update the target, comparator, outcome(s) cohorts and negative control outcomes

Choose your target cohort:

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

Choose your comparator cohort:

[OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation

Choose your outcome cohorts:

Add Outcome

Show 10 entries

Search:

ID	Name		
3	[OHDSI Ecosystem tutorial] Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting	Edit cohort	Remove

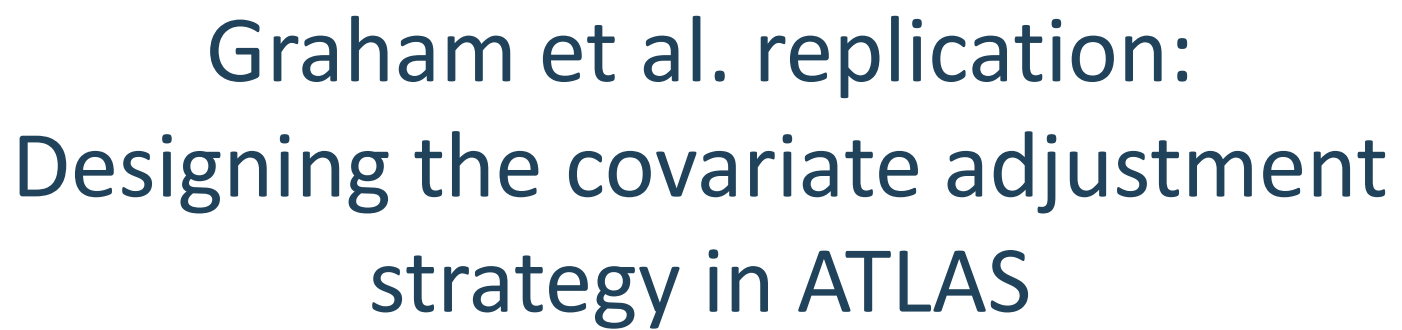
Showing 1 to 1 of 1 entries

Choose your negative control outcomes:



Graham et al. description of the covariate adjustment strategy

¹ 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



Apache 2.0
open source software

Yes ▼ Select Covariates...

[illegible]



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

What concepts do you want to include in baseline covariates in the propensity score model? (Leave blank if you want to include everything)

OHDSI estimation tutorial - Graham replication: covariates to include in PS model



What concepts do you want to exclude from baseline covariates in the propensity score model? (Leave blank if you want to include everything)

OHDSI estimation tutorial - Graham replication: covariates to exclude in PS model



How do you want to restrict your cohorts based on the propensity score distribution?

None ▼

Do you want to perform matching or stratification?

Matching ▼

How many comparator patients do you want to select for each target patient (within a defined caliper)?

1

Do you want to adjust for baseline covariates in the outcome model?

No ▼



Graham et al. replication: Designing a protocol in ATLAS

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates

Profiles

Estimation


Prediction

Jobs

Apache 2.0

open source software

provided by

OHDSI

join the journey

Population Level Effect Estimation - Comparative Cohort Analysis

[OHDSI Ecosystem tutorial] Warfarin vs. Dabigatran for risk of ischemic stroke

Specification

Utilities

Download

Export

Please click the button below to view the full study specifications. Once reviewed, scroll down to download the study package. [Click here to review the full study specifications](#)

Step 1. Review Full Study Specification

Column visibility

Copy

CSV

Show 10 entries

Filter:

Target Cohort Name	Comparator Cohort Name	Outcome Cohort Name	Analysis Name
[OHDSI Ecosystem tutorial] Graham replication: target cohort - dabigatran new users with prior atrial fibrillation	[OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation	[OHDSI Ecosystem tutorial] Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting	Primary analysis

Showing 1 to 1 of 1 entries

Previous 1 Next



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

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates


Profiles

Estimation

Prediction

Jobs


Apache 2.0
open source software

provided by
 **OHDSI**

join the journey.

Step 2. Download the study package

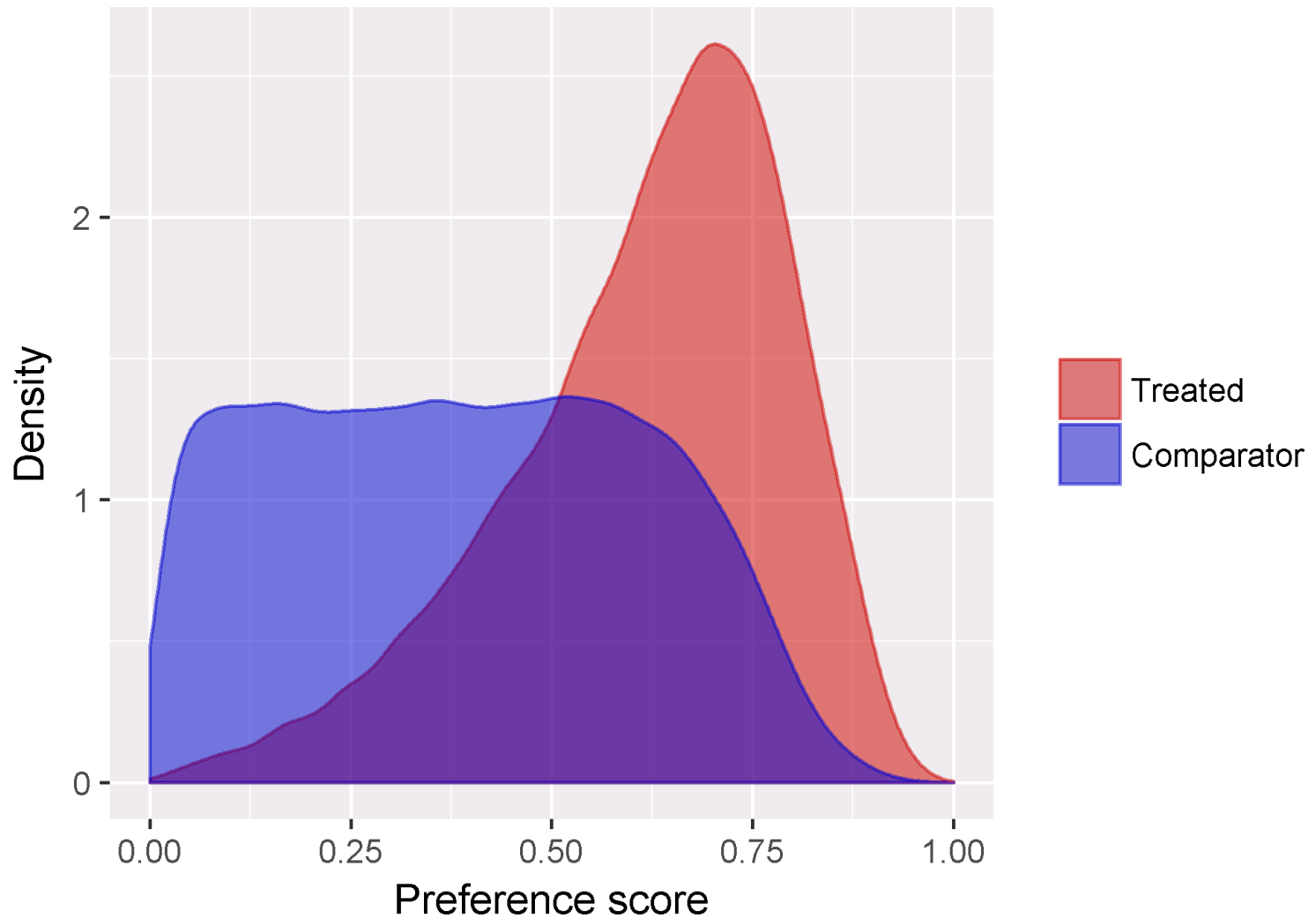
Please provide a name for the study package.

 Download Study Package

Interpreting results

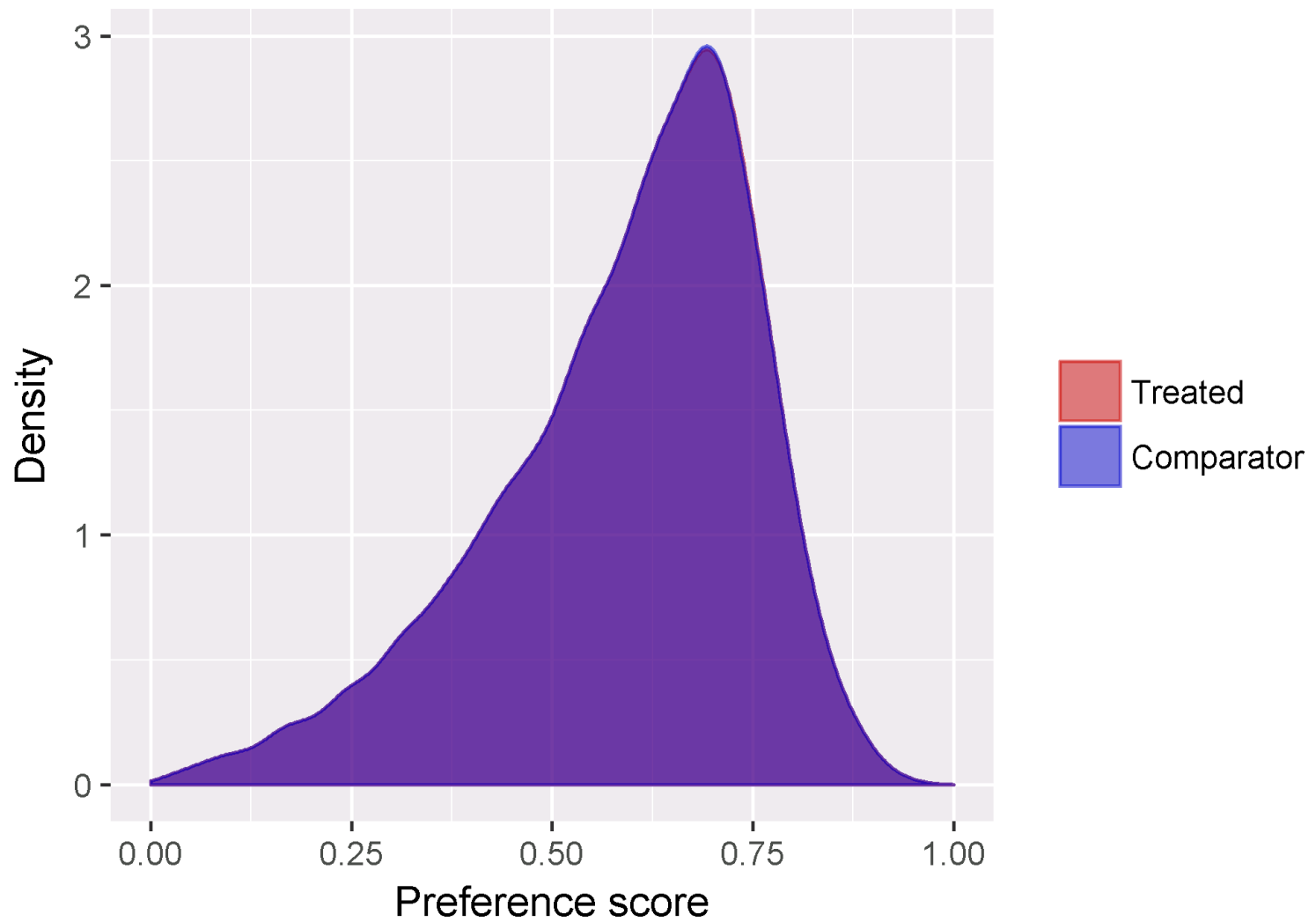


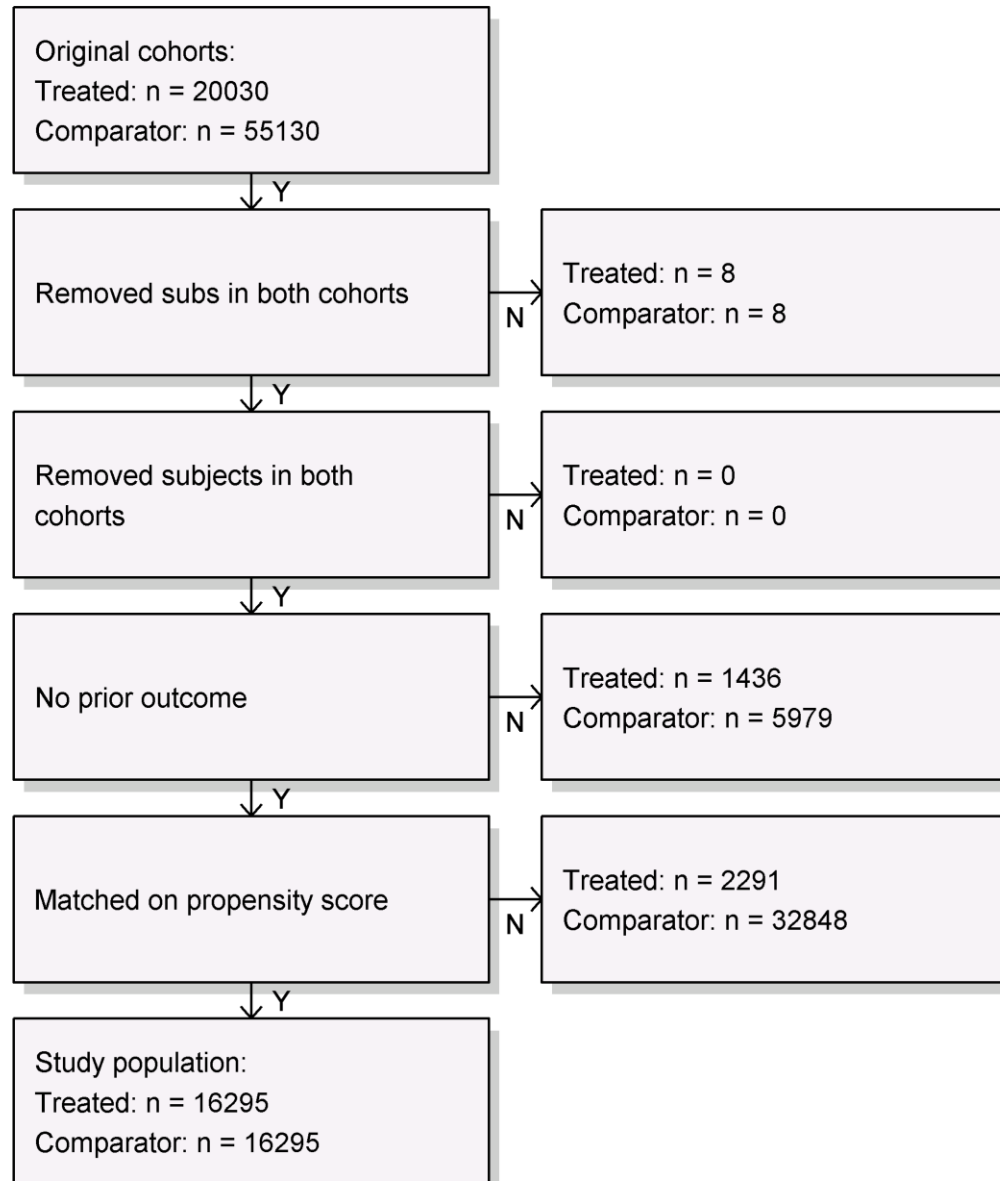
Plot propensity score distribution

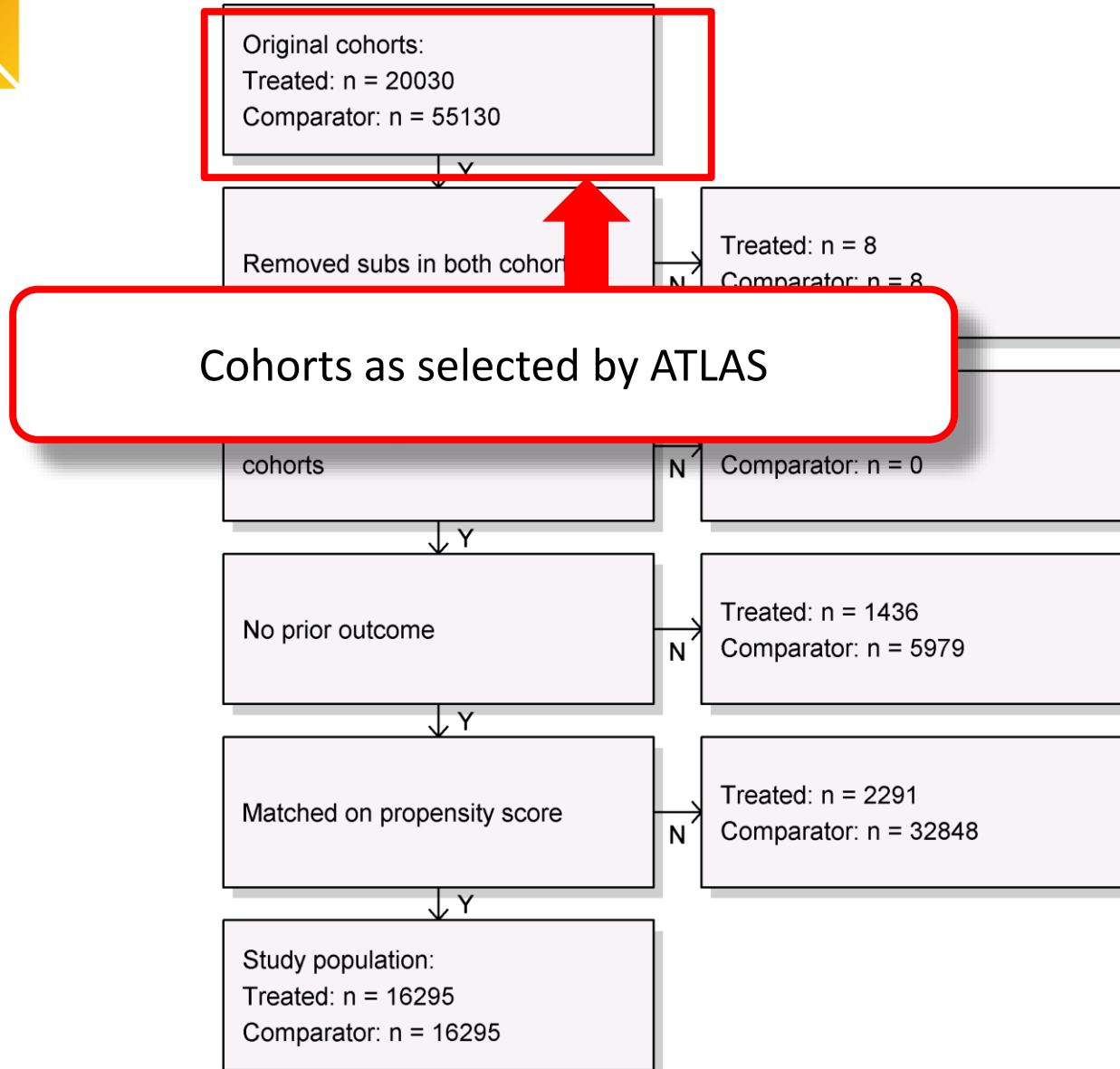


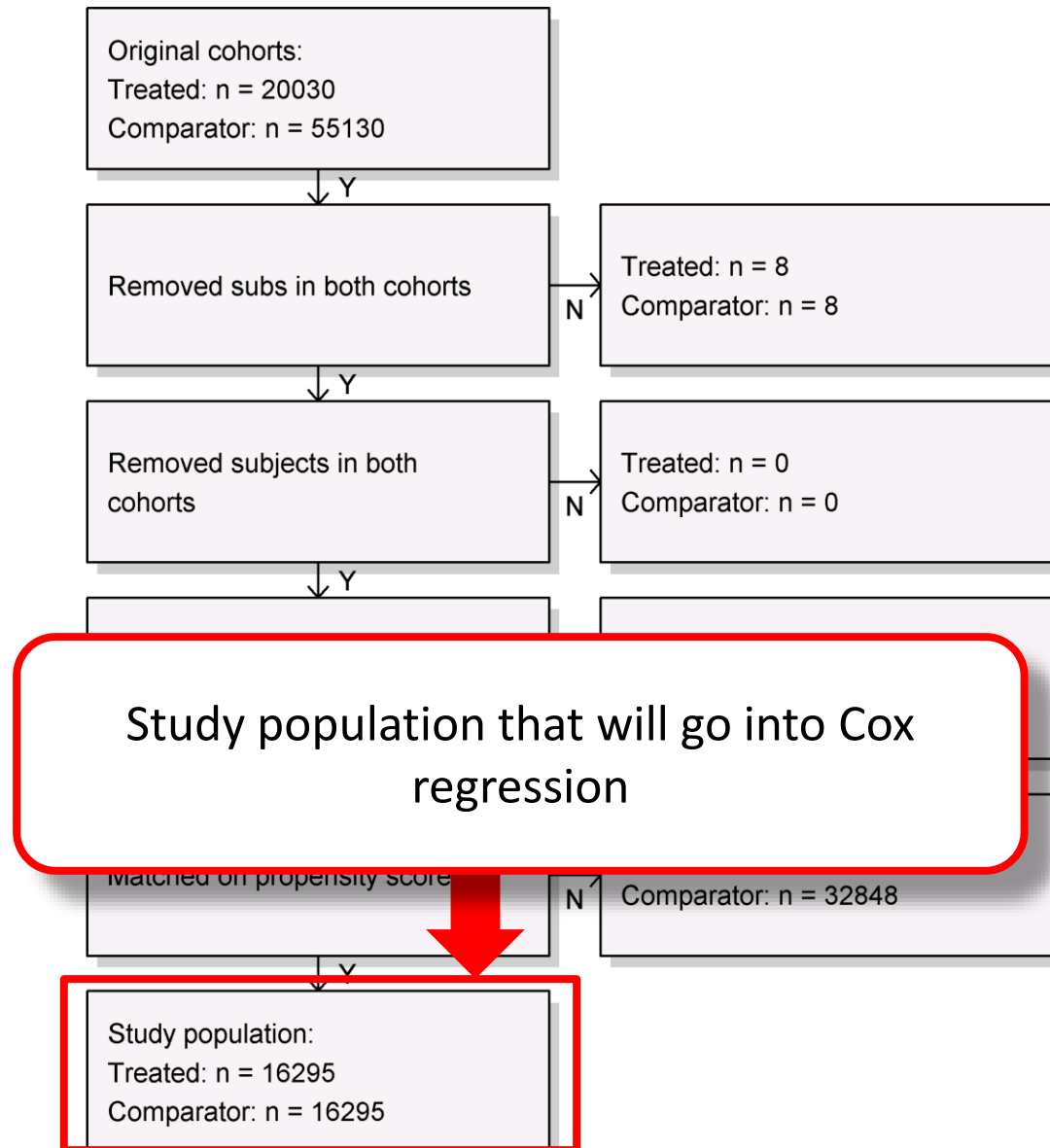


After matching



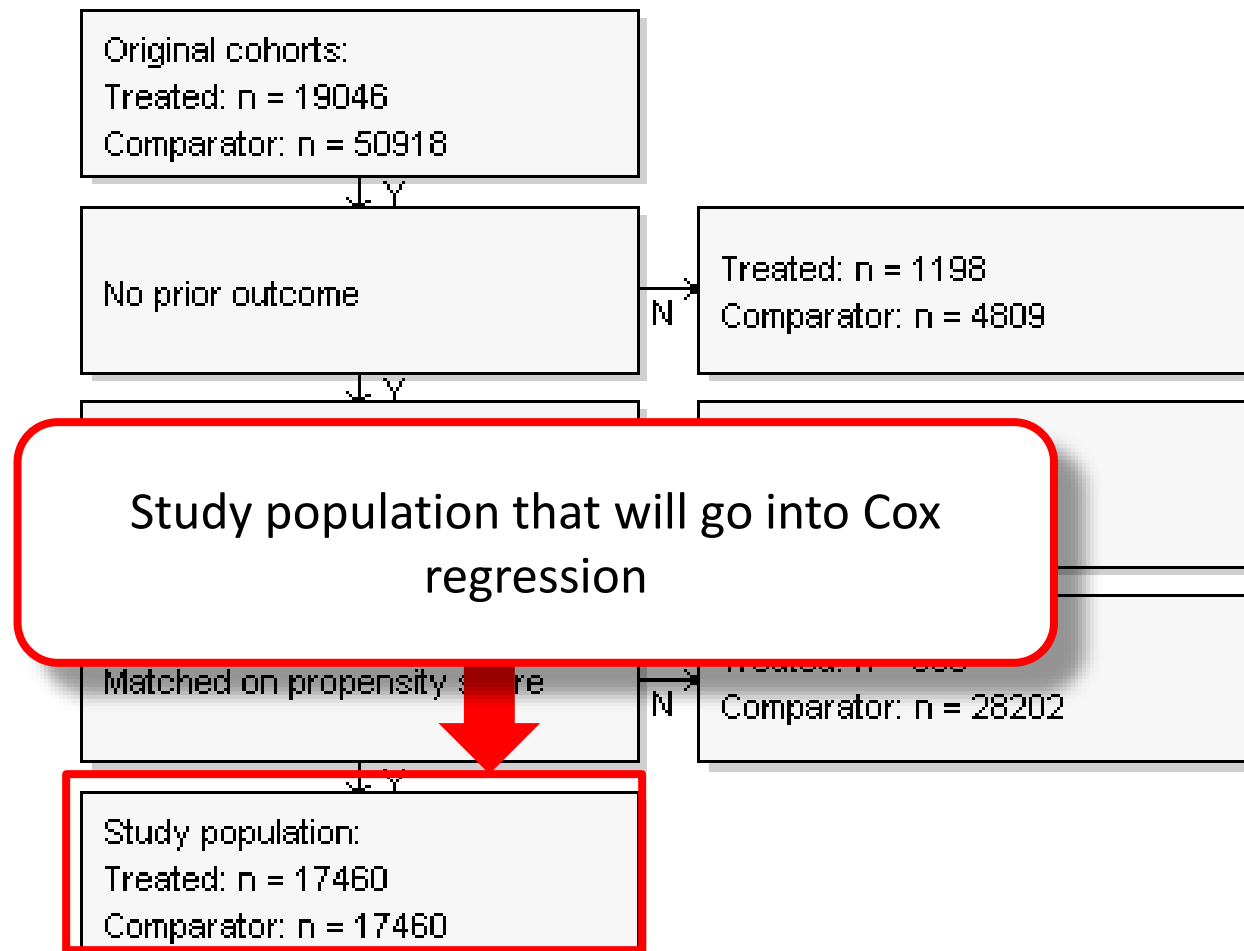








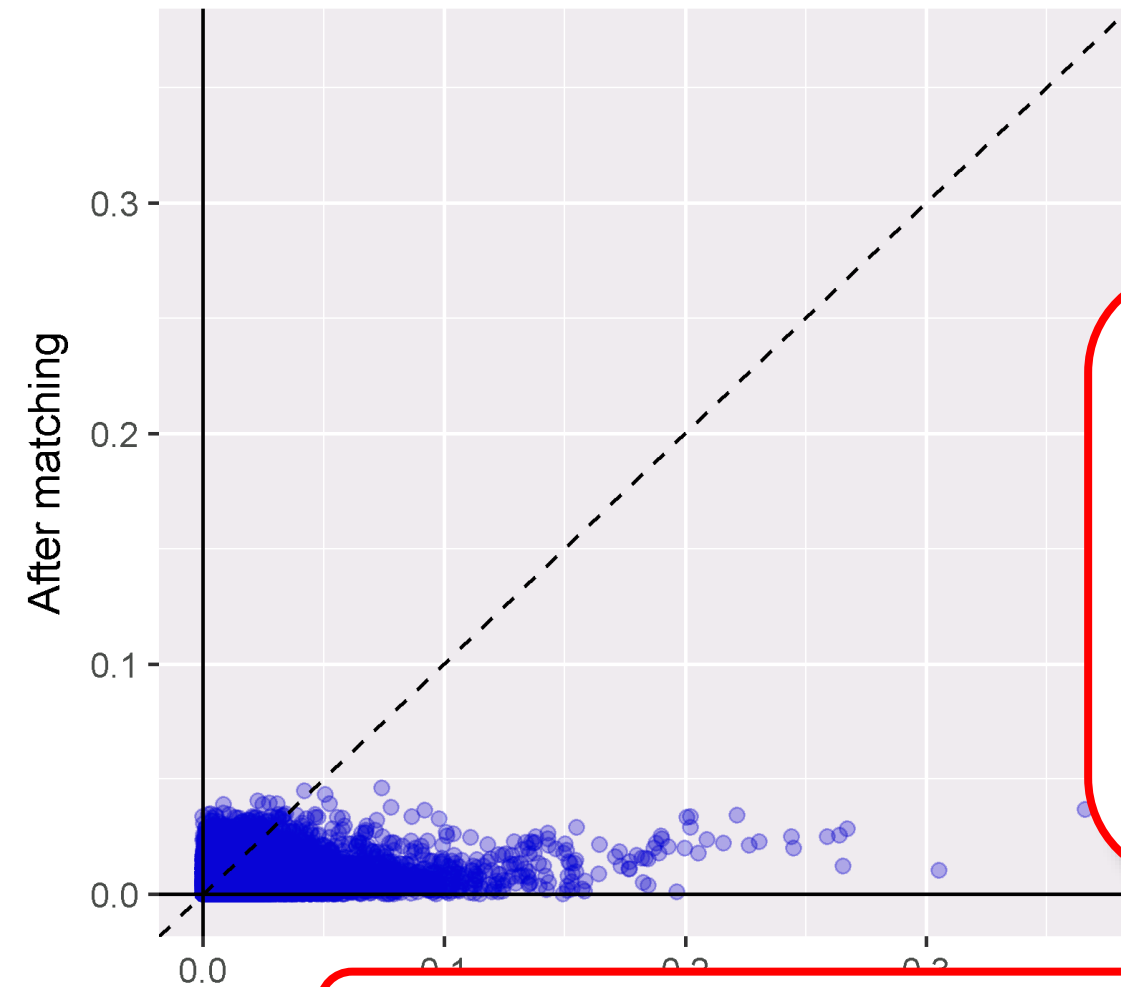
Attrition





Covariate balance

Standardized difference of mean



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

Population counts

	treatedPersons	comparatorPersons	treatedExposures	comparatorExposures
Count	81	73	81	16295

Point estimate and 95% confidence interval

	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	1.18		

Population counts

	treatedPersons
Count	16295

	com
Count	1629

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

Outcome counts

	treatedPersons
Count	81

	comparatorPersons
Count	73

	treat Exposures
Count	81

	comparatorExposures
Count	73

Time at risk

	treatedDays
Days	1682365

	comparatorDays
Days	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
treatment	0.78000	0.51050	1.18316	

Population counts

	treatedPersons	comparatorPersons
Count	16295	16295

Outcome counts

	treatedPersons	comparatorPersons
Count	81	73

Time at risk

	treatedDays	comparatorDays
Days	1682365	1515233

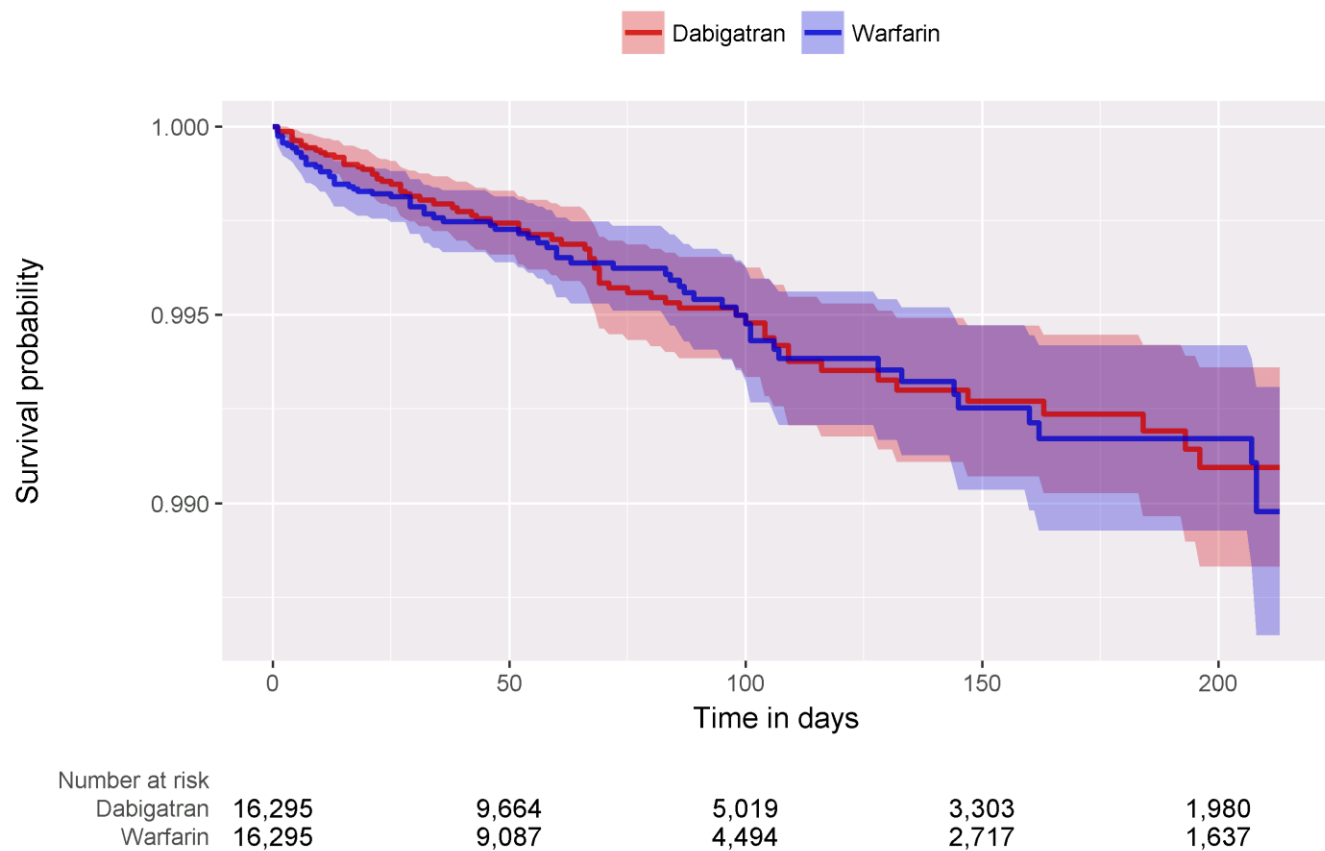
Graham:

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)



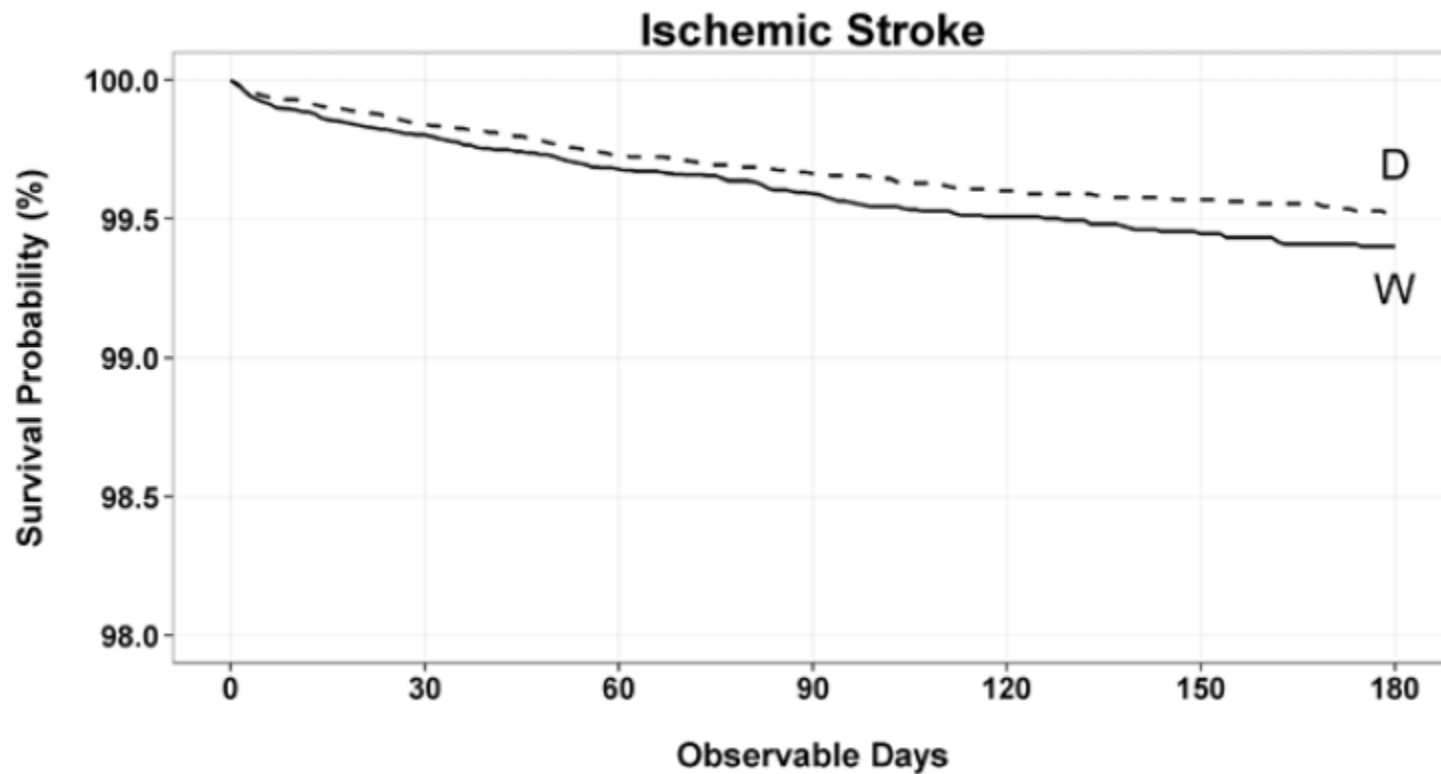
Kaplan Meier plot





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 installation by running:

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

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

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 prediction python code.

Generating R Code With Atlas

ATLAS

- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions
- Characterizations
- Cohort Pathways
- Incidence Rates
- Profiles
- Estimation
- Prediction**
- Jobs

Apache 2.0
open source software
provided by
OHDSI
join the journey.

Patient Level Prediction

[OHDSI Ecosystem tutorial] Predicting stroke amongst new users of warfarin

Specification Utilities

enter a description here (1000 characters max)

All Prediction Problem Settings Analysis Settings Execution Settings Training Settings

Prediction Problem Settings

Target Cohorts

+ Add Target Cohort

Show 10 entries Filter:

Remove	Name
	[OHDSI Ecosystem tutorial] Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation

Showing 1 to 1 of 1 entries

Outcome Cohorts

+ Add Outcome Cohort

Show 10 entries Filter:

Remove	Name
	[OHDSI Ecosystem tutorial] Graham replication: outcome cohort #1 - incident ische

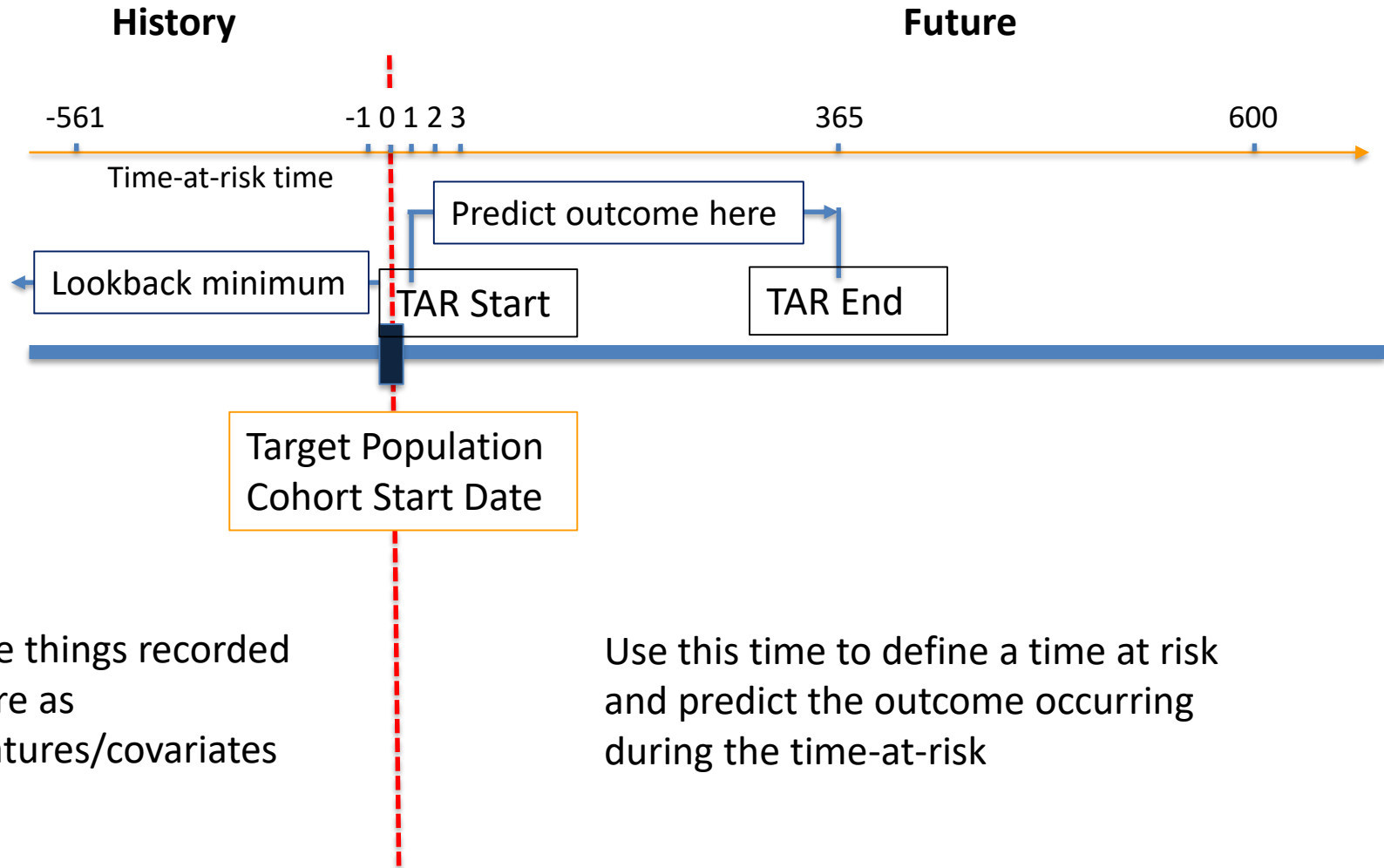
Showing 1 to 1 of 1 entries

Previous 1 Next

Previous 1 Next

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

Prediction Parameters



Prediction Design Choices

The image shows a software interface for configuring prediction models. On the left is a dark sidebar with navigation links. The main area contains a series of configuration options, each with a dropdown menu. Blue arrows point from specific dropdown values to callout boxes on the right that explain their function.

Navigation Menu:

- Data Sources
- Vocabulary
- Concept Sets
- Cohort Definitions
- Incidence Rates
- Profiles
- Estimation
- Prediction
- Jobs
- Configuration
- Feedback

Configuration Options and Annotations:

- Define the time-at-risk window start, relative to target cohort entry:** 1 days from cohort start date → **TAR Start**
- Define the time-at-risk window end:** 365 days from cohort start date → **TAR End**
- Minimum lookback period applied to target cohort:** 365 → **Lookback minimum**
- Should subjects without time at risk be removed?** Yes
- Should only the first exposure per subject be included?** Yes → **If people can be in cohort multiple times – select TRUE to only use first time**
- Include people with outcomes who are not observed for the whole at risk period?** Yes
- Sample a subset of the target group for testing?** No
- Remove patients who have observed the outcome prior to cohort entry?** No → **Select TRUE to remove people who have the outcome before the target population cohort start date**

Prediction Design Choices

Has outcome

Data Sources

Vocabulary

Concept Sets

Cohort Definitions

Incidence Rates

Profiles

Estimation

Prediction

Jobs

Configuration

Feedback

Define the time-at-risk window start, relative to target cohort entry:
1 days from cohort start date

Define the time-at-risk window end:
365 days from cohort start date

Minimum lookback period applied to target cohort:
365

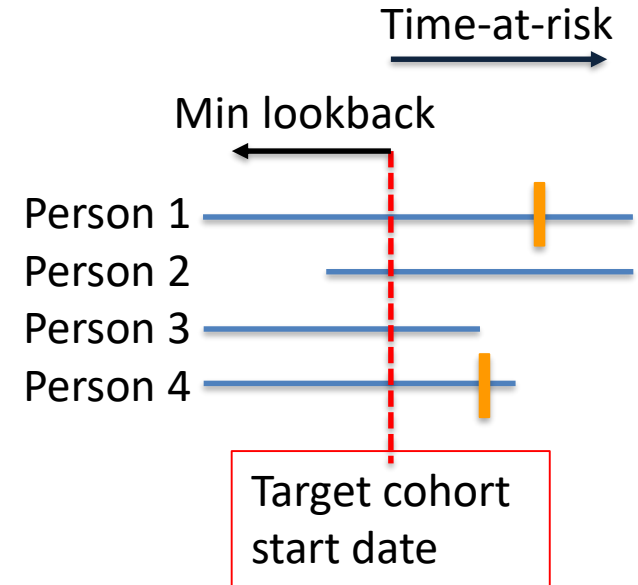
Should subjects without time at risk be removed?
Yes

Should only the first exposure per subject be included?
Yes

Include people with outcomes who are not observed for the whole at risk period?
Yes

Sample a subset of the target group for testing?
No

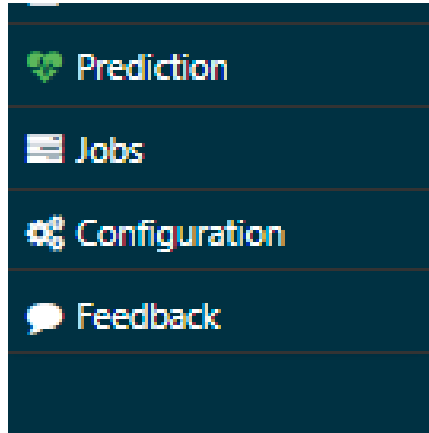
Remove patients who have observed the outcome prior to cohort entry?
No



Person 2 doesn't have min lookback so excluded

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

Training Choices



Specify the statistical model used predict the outcome amongst the target cohort:

Lasso Logistic Regression ▼

Lasso Logistic Regression model option

A single value used as the starting value for the

0.01

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



Specify how to split the test/train set:

Person ▼

Percentage of the data to be used as the test set (0-100%):

25

The number of folds used in the cross validation:

3

Test/train type and %

Number of folds to use when picking optimal hyper-parameters

[Apache 2.0](#)
open source software

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

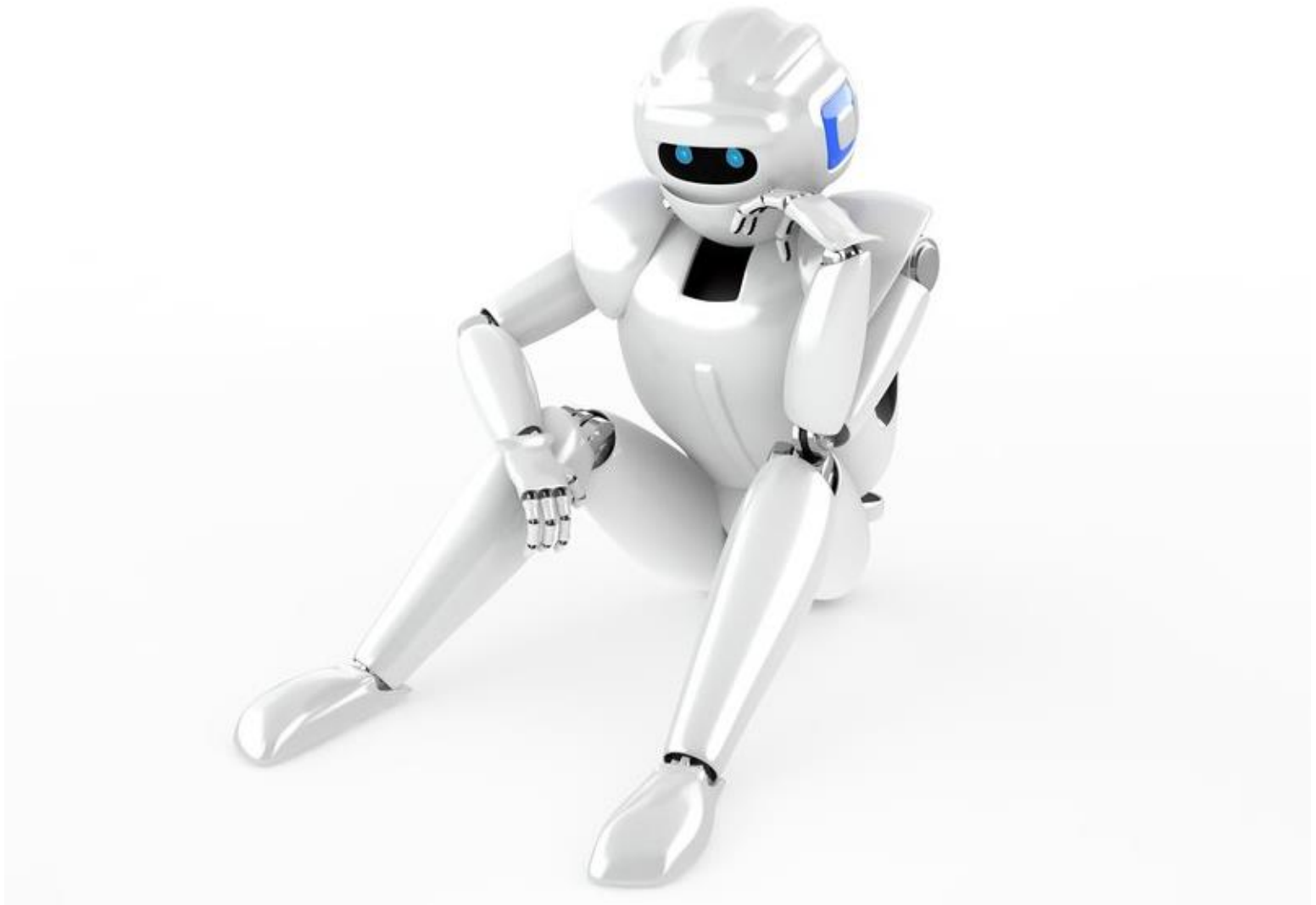
Atlas Demo

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

With

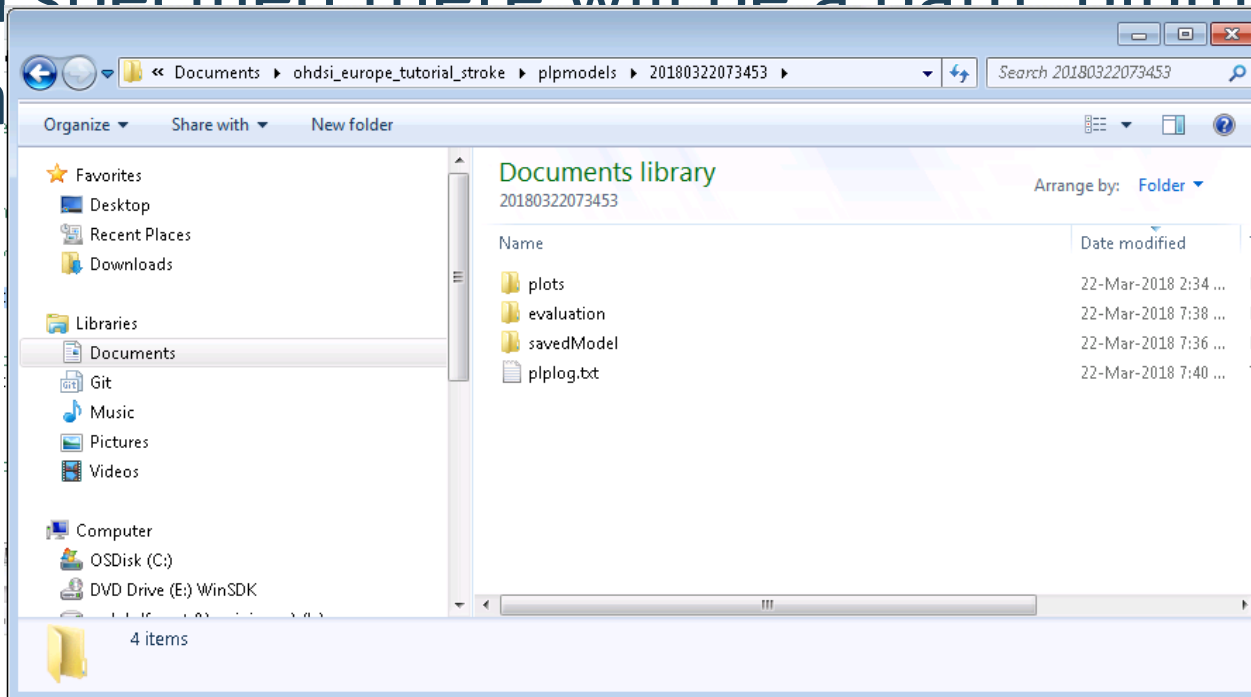
	Option	Choice
Prediction Design Choices	Time at risk start	1
	Time at risk end	365
	Remove prior outcomes	TRUE
	Require time-at-risk	TRUE
	Use all outcomes	TRUE
Training choices	Classifier	Lasso LR
	N-folds	3
	Test %	25
	Split type	Person
Feature Choices	All demo, conditions (+groups), drugs (+groups),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

PatientLevelPrediction Output



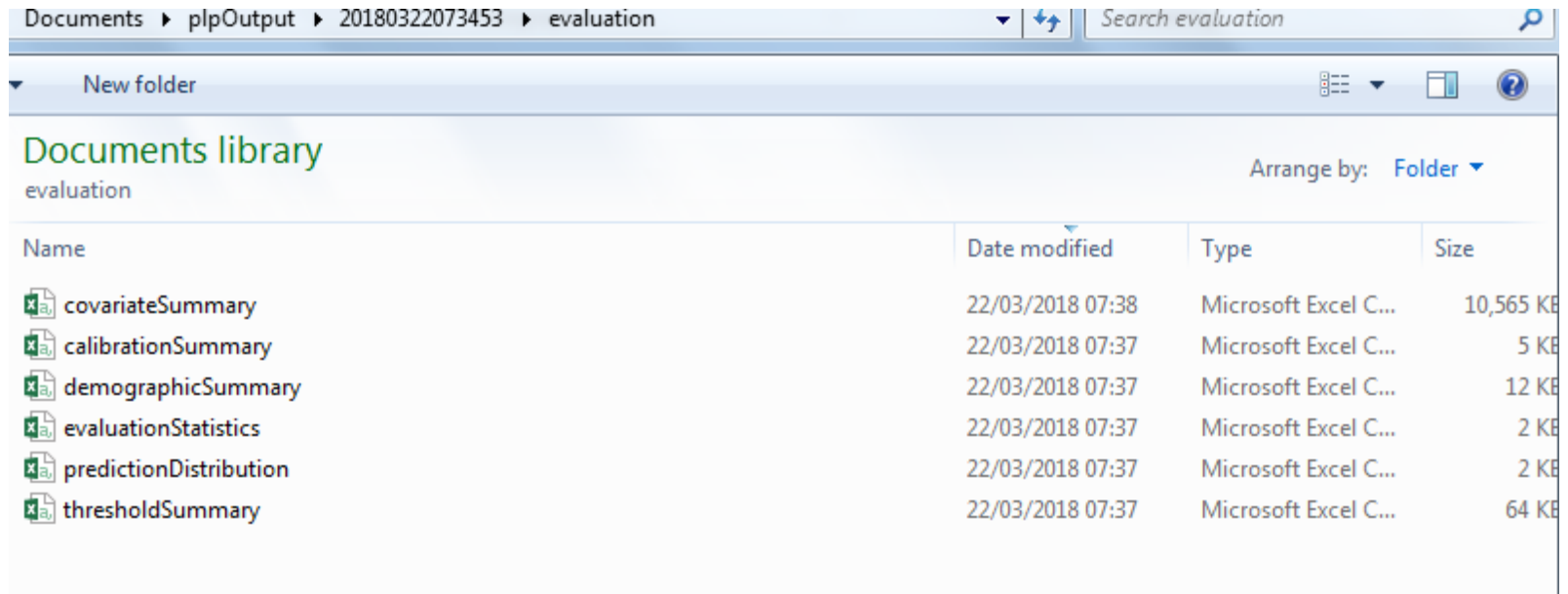
PatientLevelPrediction Output







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



PatientLevelPrediction Output

- Evaluation folder:



Name	Date modified	Type	Size
 covariateSummary	22/03/2018 07:38	Microsoft Excel C...	10,565 KB
 calibrationSummary	22/03/2018 07:37	Microsoft Excel C...	5 KB
 demographicSummary	22/03/2018 07:37	Microsoft Excel C...	12 KB
 evaluationStatistics	22/03/2018 07:37	Microsoft Excel C...	2 KB
 predictionDistribution	22/03/2018 07:37	Microsoft Excel C...	2 KB
 thresholdSummary	22/03/2018 07:37	Microsoft Excel C...	64 KB

PatientLevelPrediction Output

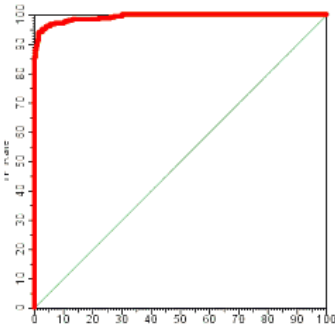
- EvaluationStatistics

populationSize	2.6% of population have outcome	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

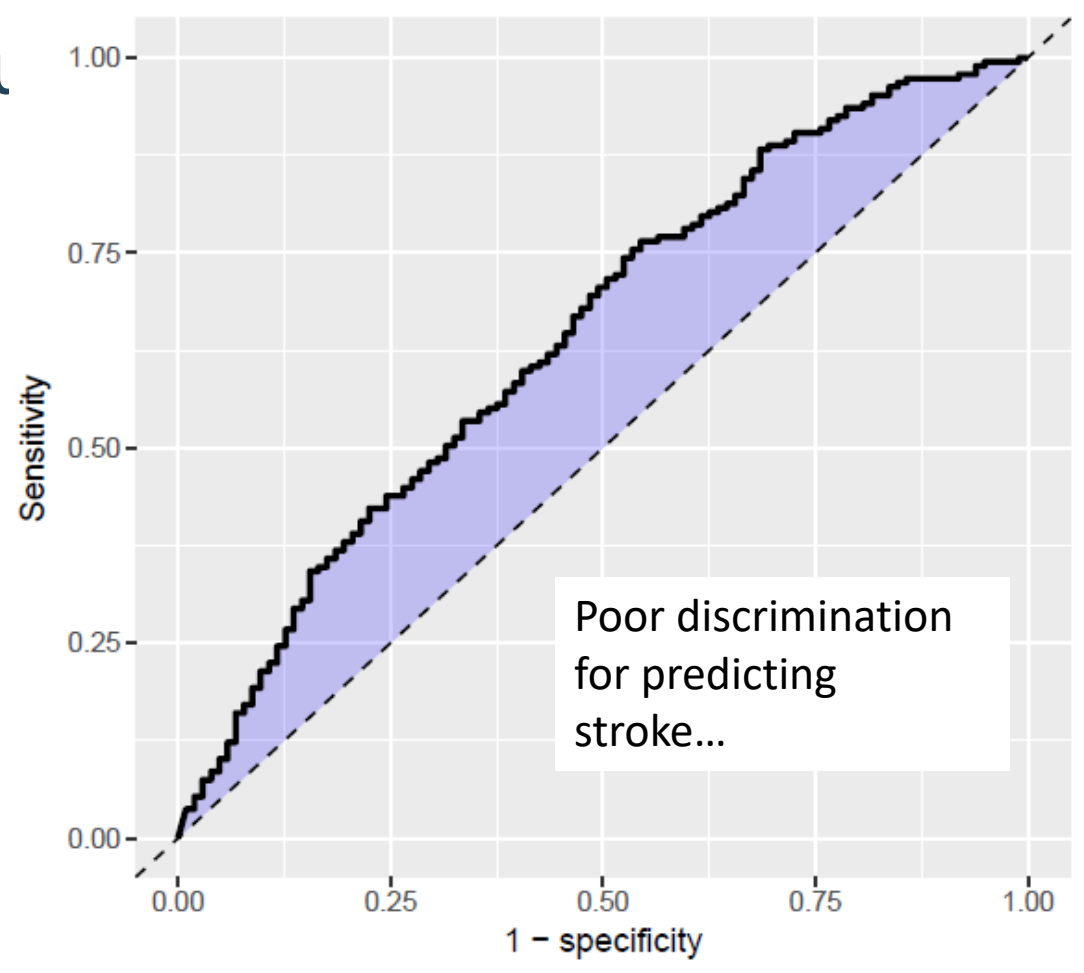
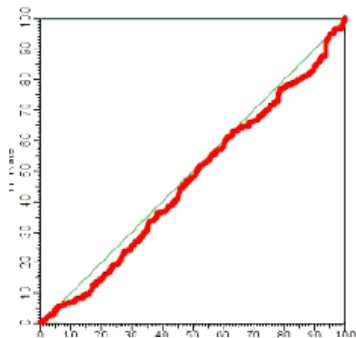
PatientLevelPrediction Output

- ROC Plot: Measure discrimination

Near Perfect discrimination:



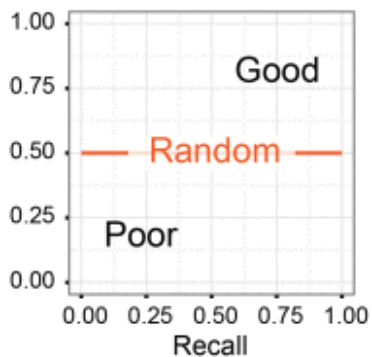
Worse discrimination:



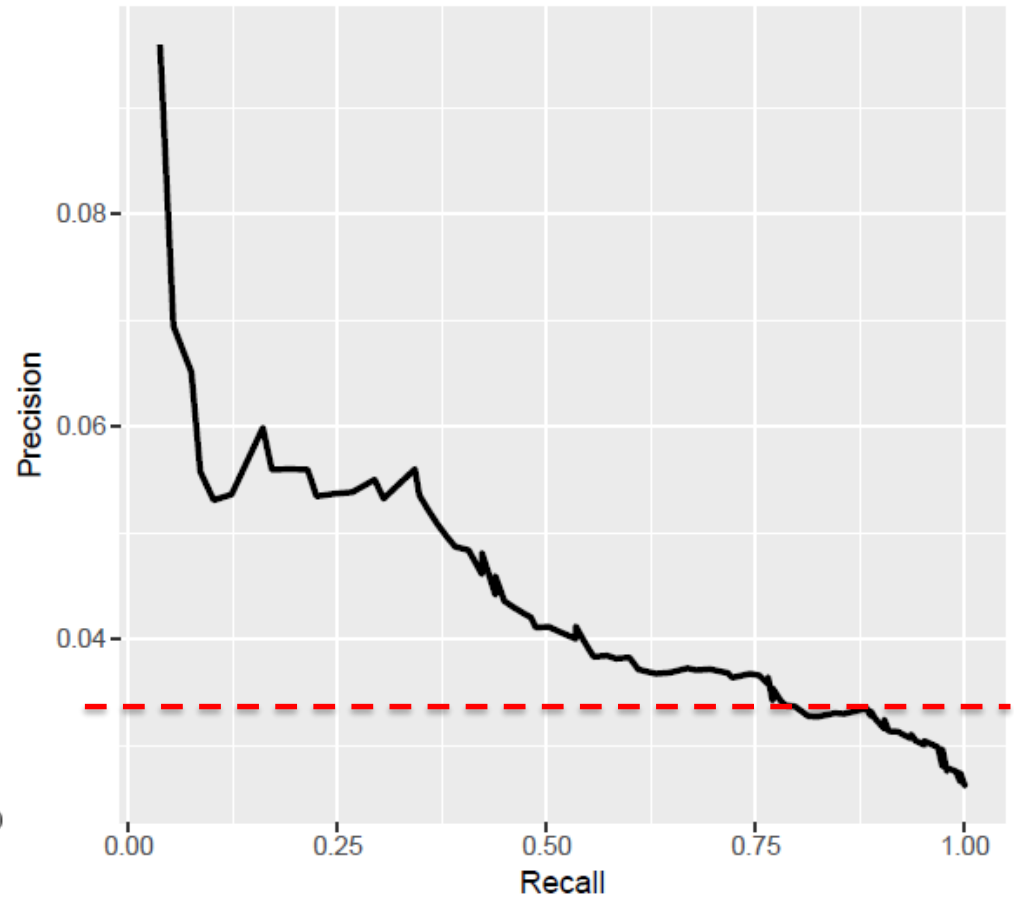
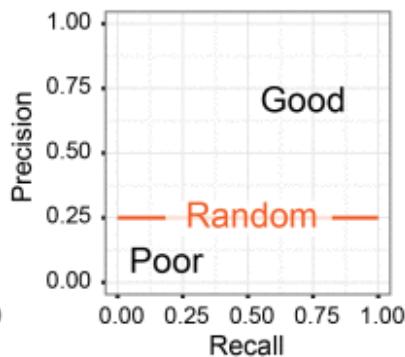
PatientLevelPrediction Output

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

Random classifier (P:N = 1:1)

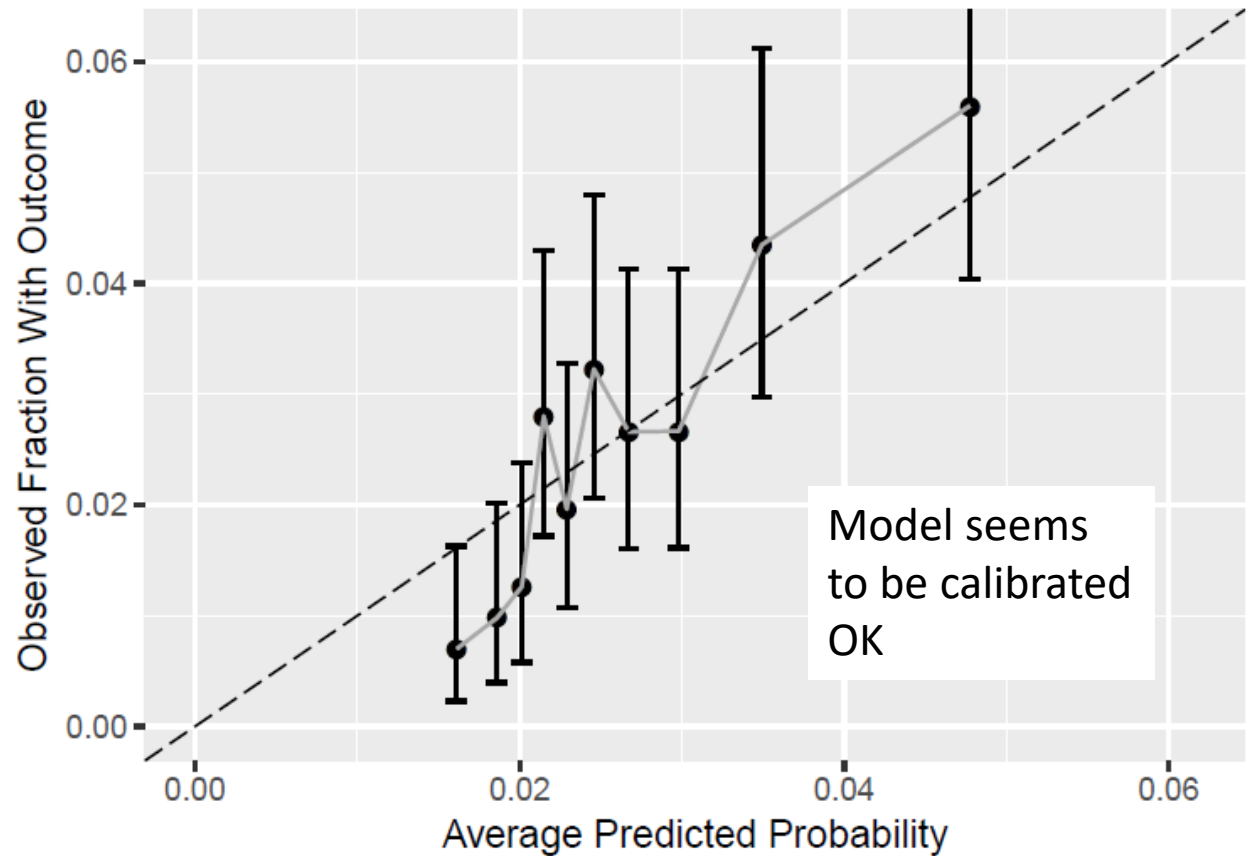


Random classifier (P:N = 1:3)



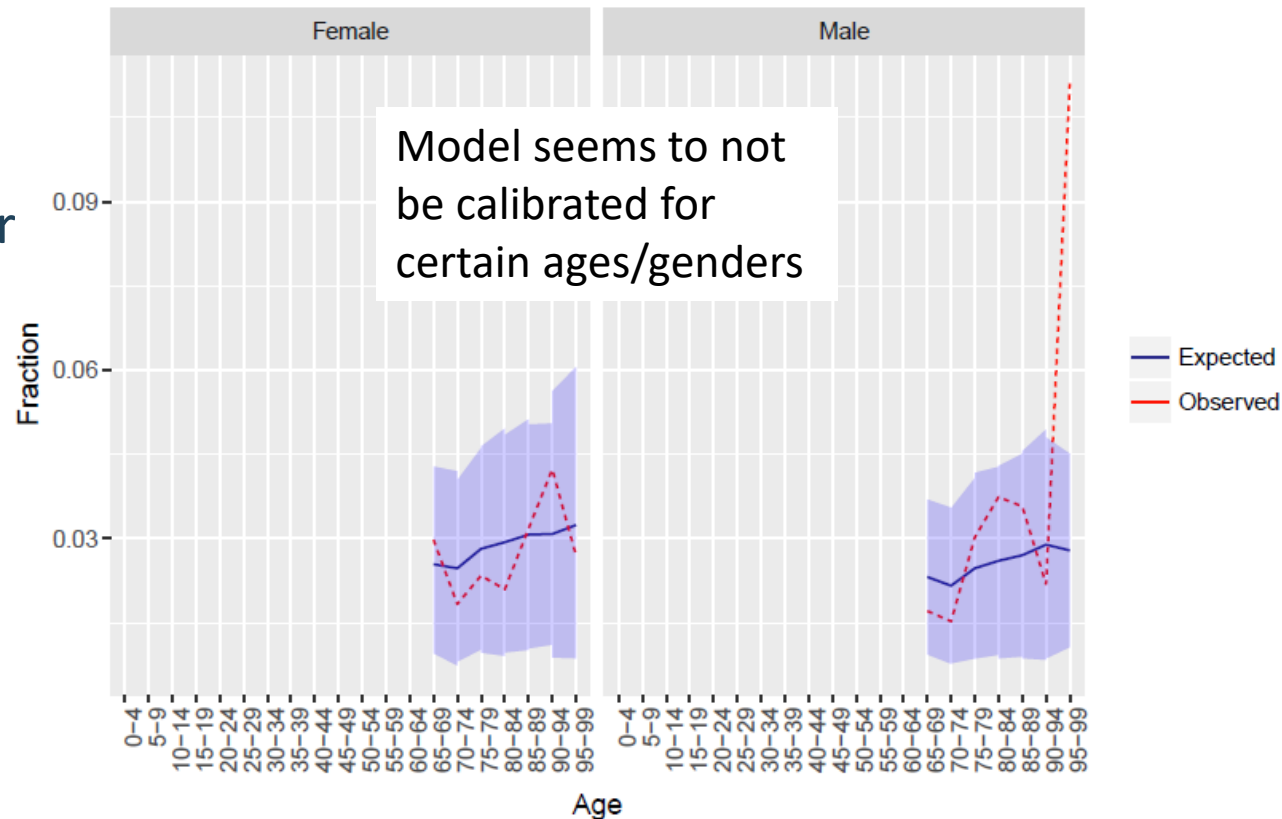
PatientLevelPrediction Output

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



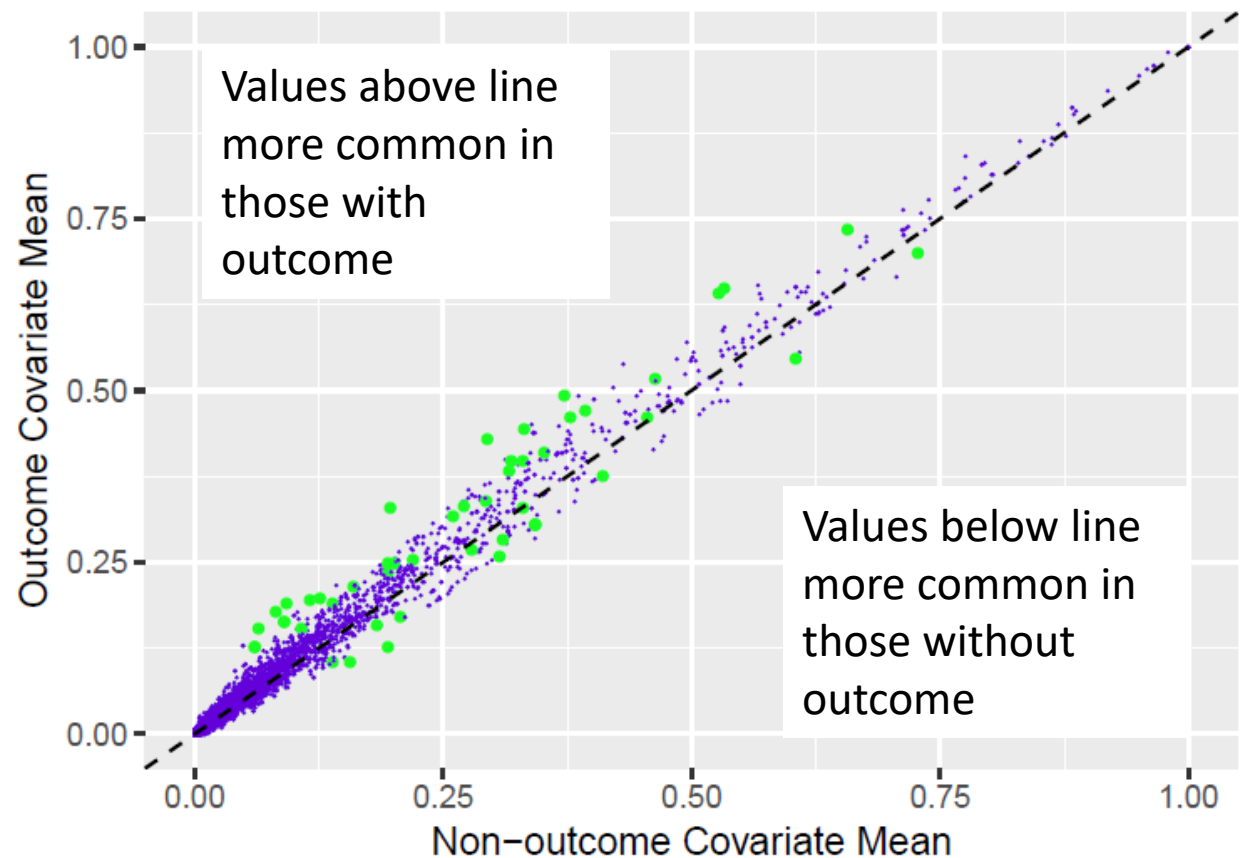
PatientLevelPrediction Output

- 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)`

~/ohdsi_europe_tutorial_stroke - Shiny
http://127.0.0.1:3545 | Open in Browser | Publish

PatientLevelPrediction Explorer Internal Validation External Validation

Evaluation Summary Characterization ROC Calibration Demographics Preference Box Plot Settings

Evaluation Summary

Show entries Search:

	Metric	test	train
AUC.auc	AUC.auc	0.6500	6.74e-01
AUC.auc_lb95ci	AUC.auc_lb95ci	0.6119	6.52e-01
AUC.auc_ub95ci	AUC.auc_ub95ci	0.6881	6.97e-01
BrierScaled	BrierScaled	0.0488	0.0700
BrierScore	BrierScore	0.0000	0.0000
CalibrationIntercept.Intercept	CalibrationIntercept.Intercept	-0.0000	-0.0000
CalibrationSlope.Gradient	CalibrationSlope.Gradient	1.0000	1.0000
outcomeCount	outcomeCount	18	18
populationSize	populationSize	71	71

Showing 1 to 9 of 9 entries

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



Network analyses using ARACHNE



ARACHNE Research Network

An open-source platform to enable federated studies across the OHDSI network

- Collaborative study lifecycle management
- Network Data catalog
- Secure, compliant and trusted data access
- Execute analysis across organizations
- Store analysis aggregate results
- Integration with OHDSI Tools (ATLAS, ACHILLES)
- Support for OHDSI standards and tools





ARACHNE UI

The ARACHNE UI allows access to the following:

- Network Studies
- OHDSI Participants (White pages)
- Network Data Catalog
- Insights Library

ARACHNE

keywords

MY STUDIES

REFINE SEARCH

My studies

My favorites

TYPE

Any

Clinical Trial Design

Clinical Trial Patient Enrollment

Health Economics and Outcomes

Safety and Efficacy

Sales and Marketing

STATUS

Any

Initiate

Active

Completed

PRIVACY

Any

Private

Public

STUDY	LEAD	MY ROLE	CREATED	TYPE	STATUS
Amgen China Demo Nov. 2017	Gregory Kabanov, Shella Smith	Lead Investigator	27 Mar 2017	Clinical Trial Design	Active
Angioedema - Demo	Gregory Kabanov	Lead Investigator	09 Nov 2017	Health Economics and Outcomes	Active
BEST: Platelet Transfusion and Acute Respiratory Distress Syndrome	Gregory Kabanov, Christian Reich, Stelie Smith	Data Set Owner, Lead Investigator	15 Oct 2017	Safety and Efficacy	Active
Demo Study	Yuriy Khoma, Gregory Kabanov, Christian Reich	Data Set Owner, Lead Investigator	12 Sep 2017	Sales and Marketing	Completed
Deployment 1.12 verification	Joe Doe, Pavel Grafkin, Gregory Kabanov	Lead Investigator	27 Apr 2018	Clinical Trial Design	Active
DUBOL	Eldor Allahverdiyev, Christian Reich, Dmitry Dymshyts, Gregory Kabanov, Aaron Galaznik	Lead Investigator	10 Mar 2017	Safety and Efficacy	Active
Early treatment pathways study	Pavel Grafkin	Contributor	15 Feb 2017	Safety and Efficacy	Active
Evidence of Treatment for prostate cancer (PSA)	Gregory Kabanov	Lead Investigator	16 Feb 2017	Safety and Efficacy	Active
FDA Demo Study 1	Joe Doe	Lead Investigator	15 Feb 2017	Health Economics and Outcomes	Active
HEDIS 2017 - annual	Gregory Kabanov	Lead Investigator	18 Mar 2017	Safety and Efficacy	Active
HEDIS 2017 - weekly	Gregory Kabanov	Lead Investigator	18 Mar 2017	Safety and Efficacy	Active
Levetiracetam and Risk of Angioedema in patients with Seizure Disorder	Gregory Kabanov	Lead Investigator	15 Feb 2017	Safety and Efficacy	Initiate

Showing 1 - 12 from 19 results

ARACHNE

ODYSSEUS: SYNPUF 2.3 MIL (SUPERVISED EXAMPLE)

DATA CATALOG

REQUEST ACCESS

GENERAL

DATASOURCE PROFILE

Person

PERSON SUMMARY

SOURCE NAME

Unsupervised 2m

NUMBER OF PERSONS

2.3 mil


YEAR OF BIRTH

POPULATION BY GENDER

POPULATION BY RACE

POPULATION BY ETHNICITY

Study Workbook



ARACHNE
enabling network research

STUDY NOTEBOOK

REFINE SEARCH

My studies ☒

My favorites ☐

TYPE

- ☐ Any
- ☐ Clinical Trial Design
- ☐ Clinical Trial Patient Enrollment
- ☐ Health Economics and Outcomes
- ☐ Safety and Efficacy
- ☐ Sales and Marketing

STATUS

- ☐ Any
- ☐ Initiate
- ☒ Active
- ☐ Completed

PRIVACY

- ☐ Any
- ☐ Private
- ☐ Public

CLEAR

MY STUDIES

STUDY	LEAD	MY ROLE	CREATED	TYPE	STATUS
☆ Amgen China Demo Nov. 2017	Gregory Klebanov, Stelle Smith	Lead Investigator	27 Mar 2017	Clinical Trial Design	Active
★ Angioedema - Demo	Gregory Klebanov	Lead Investigator	09 Nov 2017	Health Economics and Outcomes	Active
★ BEST: Platelet Transfusion and Acute Respiratory Distress Syndrome	Gregory Klebanov, Christian Reich, Stelle Smith	Data Set Owner, Lead Investigator	15 Oct 2017	Safety and Efficacy	Active
★ Demo Study	Yuriy Khoma, Gregory Klebanov, Christian Reich	Data Set Owner, Lead Investigator	12 Sep 2017	Sales and Marketing	Completed
☆ Deployment 1.12 verification	Joe Doe, Pavel Grafkin, Gregory Klebanov	Lead Investigator	27 Apr 2018	Clinical Trial Design	Active
★ DLBCL	Eldar Allakhverdiev, Christian Reich, Dmitry Dymahyts, Gregory Klebanov, Aaron Galaznik	Lead Investigator	10 Mar 2017	Safety and Efficacy	Active
☆ Early treatment pathways study	Pavel Grafkin	Contributor	15 Feb 2017	Safety and Efficacy	Active
☆ Evidence of Treatment for prostate cancer (PSA)	Gregory Klebanov	Lead Investigator	16 Feb 2017	Safety and Efficacy	Active
☆ FDA Demo Study 1	Joe Doe	Lead Investigator	15 Feb 2017	Health Economics and Outcomes	Active
☆ HEDIS 2017 - annual	Gregory Klebanov	Lead Investigator	18 Mar 2017	Safety and Efficacy	Active
☆ HEDIS 2017 - weekly	Gregory Klebanov	Lead Investigator	18 Mar 2017	Safety and Efficacy	Active
☆ Levetiracetam and Risk of Angioedema in patients with Seizure Disorder	Gregory Klebanov	Lead Investigator	15 Feb 2017	Safety and Efficacy	Initiate

Showing 1 - 12 from 19 results

Lists all participant's studies, including those marked as "public" (viewed by anyone) and "private" (viewed by study collaborators only)



Network Study

The ARACHNE Network study is a collaborative workspace for all study documents, design artifacts, analysis code, execution activities as well as results repository

The ARACHNE Network study allows defining and setting a number of attributes, including name, objectives, lifecycle status, visibility (private or public), type of a study

The ARACHNE Network Study allows study members to create and manage a various study-related analyses:

- Simple cohort counts
- Cohort characterization
- Incidence rates
- PLP
- PLE
- Complex custom code

The screenshot displays the ARACHNE Network Study interface. The study title 'OHDSI SYMPOSIUM US 2018 DEMO' is circled in red. The study is set to 'PRIVATE' and 'Completed'. The 'ANALYSES' section is highlighted with a red circle and contains a list of analyses including 'FDA Best Cohort', 'Cohort Characterization', 'PLE 2', 'IR2', 'Cohort', 'Cohort Characterization 2', 'PLE2 -a', 'PLP3', 'PLP v01', 'PLE', and 'Columbia'. The interface also shows a sidebar with navigation options like 'PERSONAL WORKSPACE', 'STUDY NOTEBOOK', 'EXPERT FINDER', 'DATA CATALOG', 'INSIGHTS LIBRARY', and 'ADMIN SETTINGS'.



Study Data Sources and Collaborators

Study Lead Investigator can add data sets by requesting data set access. Data set becomes available for study analysis when approved by the data set owner

Study Lead Investigator can invite other OHDSI collaborators, including granting them the Lead Investigator role

Study Lead Investigators can modify study attributes, manage data access, collaborators

Study collaborators can create new analyses, execute them and review and annotate results

The screenshot shows the ARACHNE web application interface for the OHDSI SYMPOSIUM US 2018 DEMO. The left sidebar contains navigation links: PERSONAL WORKSPACE, STUDY NOTEBOOK, EXPERT FINDER, DATA CATALOG, INSIGHTS LIBRARY, and ADMIN SETTINGS. The main content area is titled 'OHDSI SYMPOSIUM US 2018 DEMO MY STUDIES'. It features a 'START DATE' field (03/17/2018) and an 'END DATE' field (03/21/2018), both of which are circled in red. Below these fields is a 'STUDY OBJECTIVE' section with the text 'this is a demo study used at the BioIT World 2018'. The 'DATA SOURCES' tab is selected, showing a list of data sources with columns for 'DOCUMENTS', 'PARTICIPANTS', and 'DATA SOURCES'. The list includes several entries for 'Odysseus: QA 110K' with status 'APPROVED'. The 'ANALYSES' section on the right lists various analyses, including 'FDA Best Cohort', 'Cohort Characterization', 'PLE 2', 'IR2', 'Cohort', 'Cohort Characterization 2', 'PLE2 -a', 'PLP3', 'PLP v01', 'PLE', and 'Columbia'. The user 'Gregory Klebanov' is logged in, and the interface shows 100% completeness.



Simple Cohort Count Analysis

Allows performing simple counts

Cohorts definitions can be added by:

- Importing from ATLAS instances connected
- Uploading files
- Creating new file

An existing ATLAS defined cohort can be used:

- JSON
- OHDSI.SQL

ARACHNE enabling network research

IBUPROFEN DRUG EXPOSURE V1
MY STUDIES • DEMO STUDY

Created Sep 14, 2017 9:11AM EDT by Gregory Klebanov

STUDY NOTEBOOK
EXPERT FINDER
DATA CATALOG
INSIGHTS LIBRARY

TYPE OF ANALYSIS: Cohort (Simple Counts) [LOCK] [UNLOCK]

CODE FILES LOCK

CODE FILES: ibuprofen_exposure.ohdsi.sql V2 Gregory Klebanov

UPLOAD NEW IMPORT [SCHEDULE] [SUBMIT]

DATA SOURCES STATUS EXECUTE RESULT PUBLISH

No filters applied [FILTER]

5	1 FILE SUBMITTED				Checksum 1b2b840 • Oct 31, 2017 4:27PM EDT
	OHDSI: SynPUF 110K	FINISHED	✓	5 documents	✓
4	1 FILE SUBMITTED				Checksum a883fd3 • Oct 18, 2017 2:31PM EDT
	OHDSI: SynPUF 110K	FINISHED	✓	5 documents	✓
	Odysseus: Unsupervised CDM 2m	FINISHED	✓	5 documents	✓
3	1 FILE SUBMITTED				Checksum a883fd3 • Oct 18, 2017 2:11PM EDT
	Odysseus: Supervised SynPuf CDM 2.3m	AWAITING APPROVAL		No documents	✗
	Odysseus: Unsupervised CDM 2m	FINISHED	✓	5 documents	✓
2	1 FILE SUBMITTED				Checksum a883fd3 • Oct 18, 2017 1:05PM EDT

© 2015-2018 Odysseus Data Services, Inc.

ARACHNE allows simultaneous analysis execution against multiple data sets across the network.

Job submissions can be pre-scheduled



Simple Cohort Count Analysis

- ARACHNE will save a history of jobs executions, including:
 - Submitted Code
 - Execution and approval log
 - Analysis results
- The analysis submissions can be viewed and annotated

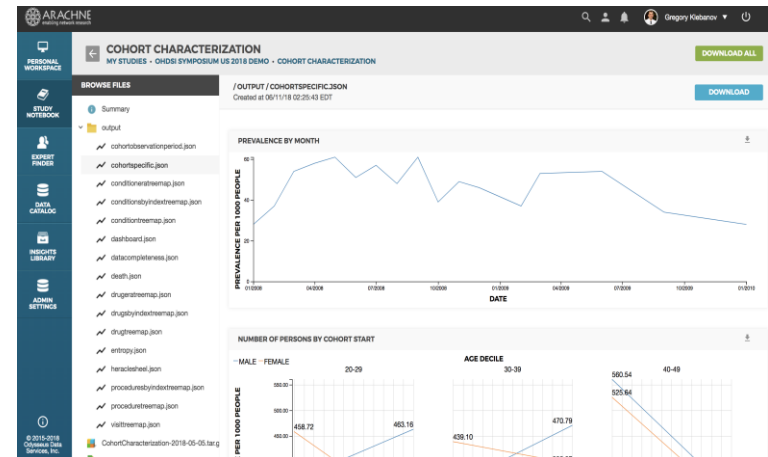
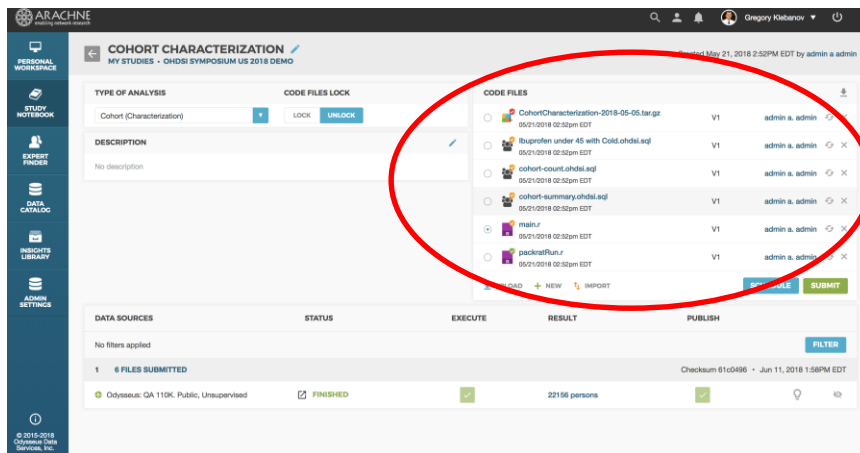


Cohort Characterization Analysis

The Cohort Characterization analysis allows performing Heracles against remote data

ARACHNE will automatically package analyses imported from ATLAS into a self-contained packrat R package

The Heracles execution parameters can be set by modifying the execution R module shell





Incidence Rates

The Incidence Rate analysis allows performing calculating incidence rates against remote data

ARACHNE will automatically package analyses imported from ATLAS into a self-contained packrat R package

ARACHNE web interface showing the 'IR2' analysis setup. The 'FILES' section is highlighted with a red circle, listing various input files for the incidence rate analysis.

FILE	TYPE	OWNER
Angioedema across levetricacetam new users_58_tar...	V1	Gregory Kabanov
Angioedema across levetricacetam new users_58_tar...	V1	Gregory Kabanov
Angioedema across levetricacetam new users_58_out...	V1	Gregory Kabanov
IncidenceRates-v1_0_0_target	V1	Gregory Kabanov
analysisDescription.json	V1	Gregory Kabanov
main.r	V1	Gregory Kabanov
packratRun.r	V1	Gregory Kabanov

ARACHNE web interface showing the 'IR2' analysis results. The 'ANALYSIS' section displays the 'Incidence rates' analysis, and the 'RESULT SUMMARY' table shows the calculated incidence rate.

PERSONS	CASES	PROPORTION [I-I] PER 14 PERSONS	TIME AT RISK (YEARS)	RATE [I-I] PER 14 YEARS
614	1	1.6286645	487	2.03388



PLE and PLE methods

ARACHNE support both Population-Effect Estimation and Patient-Level Prediction methods

ARACHNE will automatically package analyses imported from ATLAS into a self-contained packrat R package

The screenshot shows the ARACHNE PLE interface. The top bar indicates 'PLE' and 'MY STUDIES - OHDSI SYMPOSIUM US 2018 DEMO'. The left sidebar contains navigation options: PERSONAL WORKSPACE, STUDY NOTEBOOK, EXPERT FINDER, DATA CATALOG, INSIGHTS LIBRARY, and ADMIN SETTINGS. The main content area is divided into three sections: TYPE OF ANALYSIS (Population Level Effect Estimation), CODE FILES LOCK (a table with columns for analysis name, date, and user), and CODE FILES (a list of files). A red circle highlights the 'CODE FILES LOCK' section. Below this, there are sections for DATA SOURCES, STATUS, and EXECUTE.





Custom Code

ARACHNE support the execution of the custom SQL or R code

The screenshot shows the ARACHNE web interface with a custom SQL query loaded. The query is titled 'WARFARIN.OHDSI.SQL' and was updated by Gregory Klebanov at 11/8/17 12:52:32 EST. The query text is as follows:

```
1 CREATE TABLE #Codesets (  
2   codeset_id int NOT NULL,  
3   concept_id bigint NOT NULL,  
4 )  
5 ;  
6  
7 INSERT INTO #Codesets (codeset_id, concept_id)  
8 SELECT 0 as codeset_id, c.concept_id FROM (select distinct l.concept_id FROM  
9   l  
10  select concept_id from fcdm_database_schema.CONCEPT where concept_id in (1310149) and invalid_reason is null  
11 UNION select c.concept_id  
12   from fcdm_database_schema.CONCEPT c  
13   join fcdm_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id  
14   and ca.ancestor_concept_id in (1310149)  
15   and c.invalid_reason is null  
16  
17 ) l  
18 ) c;  
19 INSERT INTO #Codesets (codeset_id, concept_id)  
20 SELECT 1 as codeset_id, c.concept_id FROM (select distinct l.concept_id FROM  
21   l  
22   select concept_id from fcdm_database_schema.CONCEPT where concept_id in (313217) and invalid_reason is null  
23 UNION select c.concept_id  
24   from fcdm_database_schema.CONCEPT c  
25   join fcdm_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id  
26   and ca.ancestor_concept_id in (313217)  
27   and c.invalid_reason is null  
28  
29 ) l  
30 ) c;  
31  
32  
33 with primary_events (event_id, person_id, start_date, end_date, op_start_date, op_end_date) as  
34 (  
35   -- Begin Primary Events  
36   select row_number() over (PARTITION BY P.person_id order by P.start_date) as event_id, P.person_id, P.start_date,  
37   from
```

The screenshot shows the ARACHNE web interface for a predictive model analysis titled 'EGFR_PATIENT_LVL_PREDICTION_WINDOW_3(1_YEAR_PREDICTIVE_WINDOW)'. The analysis was created on Oct 10, 2017 12:00PM EDT by Elder Allahverdiev. The interface includes a sidebar with navigation options (STUDY NOTEBOOK, EXPERT FINDER, DATA CATALOG, INSIGHTS LIBRARY) and a main content area with the following sections:

- TYPE OF ANALYSIS:** Patient Level Prediction (LOCK, UNLOCK)
- DESCRIPTION:** Creating models for patient level prediction of disease progression for EGFR positive NSCLC patients. Patients are observed in observation window 2 - 1 year (365 days) after index date. Logistic Regression, Naive Bayes, KNN, Gradient boosting machine, Random forest and multilayer perceptron models are chosen for prediction.
- CODE FILES LOCK:** A table listing code files with their versions and owners.
- DATA SOURCES:** A table showing data sources, their status, and execution results.

CODE FILES	VERSION	OWNER
EGFR_COHORTS_WINDOW_3.ohdsi.sql	V1	Yury Khoma
PatientLevelPrediction_EGFR_WINDOW_3.R	V1	Yury Khoma
PerfMetrics.R	V1	Yury Khoma
README.TXT	V6	Elder Allahverdiev
environmentvars.txt	V1	Yury Khoma

DATA SOURCES	STATUS	EXECUTE	RESULT	PUBLISH
No filters applied				
1 5 FILES SUBMITTED				
Virtual: QuintilesMS P+	FINISHED	✓	5 documents	✓



Demo: Utilizing ARACHNE

Follow along at:

<http://arachne.ohdsi.org/>



Design and implement your
own study!



Questions?

Thanks for joining
the journey!

