OHDSI Tutorial: Patient-level predictive modelling in observational healthcare data

Faculty:

Peter Rijnbeek (Erasmus MC)

Jenna Reps (Janssen Research and Development)

Joel Swerdel (Janssen Research and Development)

Today’s Agenda

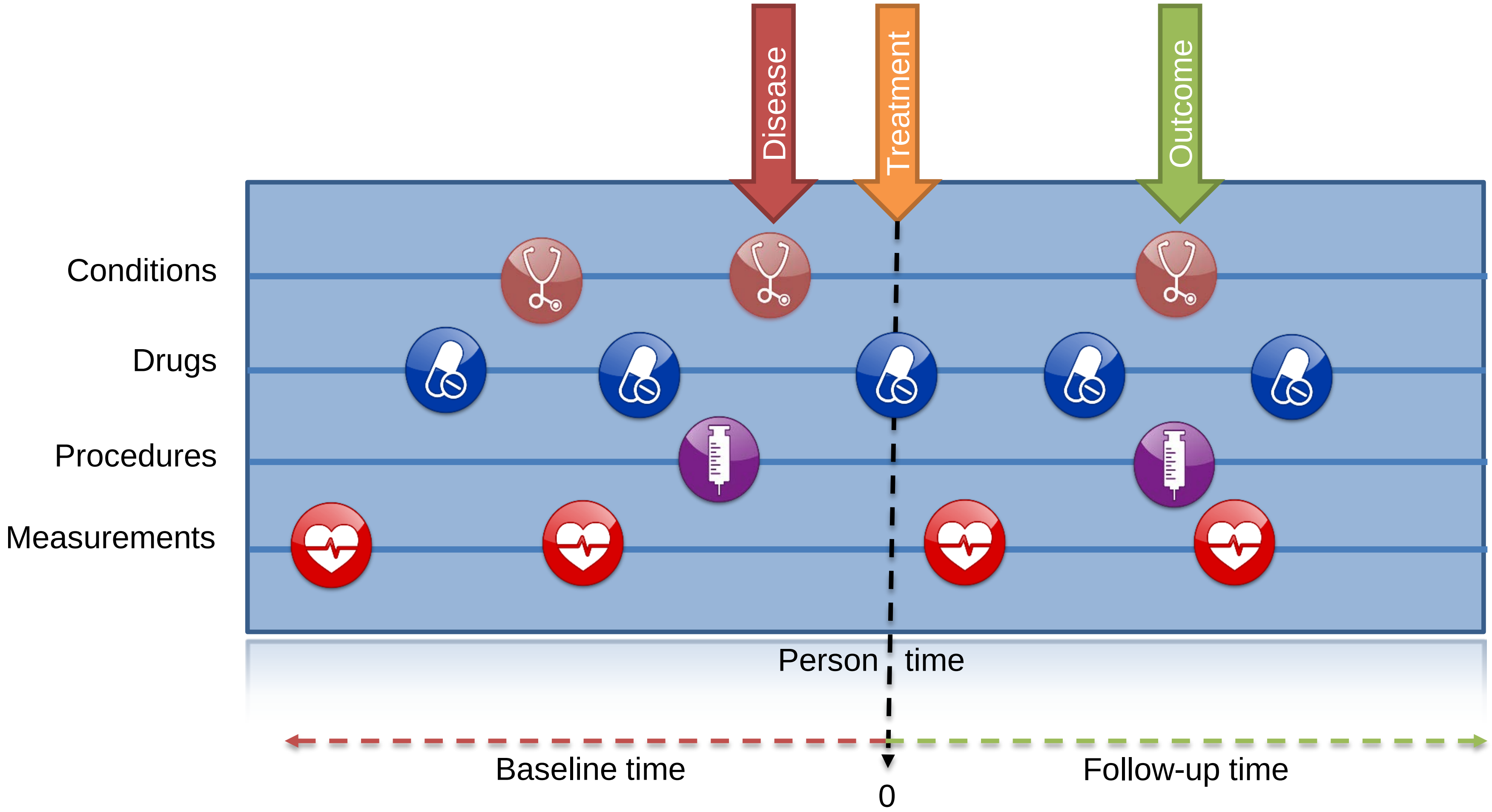
Time	**Topic**
8:45 – 9:00	Welcome, get settled, get laptops ready
9:00 – 10:30	**Presentation:** What is Patient-Level Prediction?
10:30 – 10:45	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 1 Theory
10:45 – 11:45	**Presentation:** Overview of the TRIPOD Statement **Exercise:** Applying TRIPOD to CHADS2
11:45 – 12:30	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 2 Implementation
12:30 – 13:15	Lunch
13:15 – 14:30	**Exercise:** Guided tour through implementing patient-level prediction
14:30 – 14:45	Break
14:45 – 16:30	**Exercise:** Design and implement your own patient-level prediction
16:30 – 17:00	Lessons learned and feedback

OHDSI's Mission

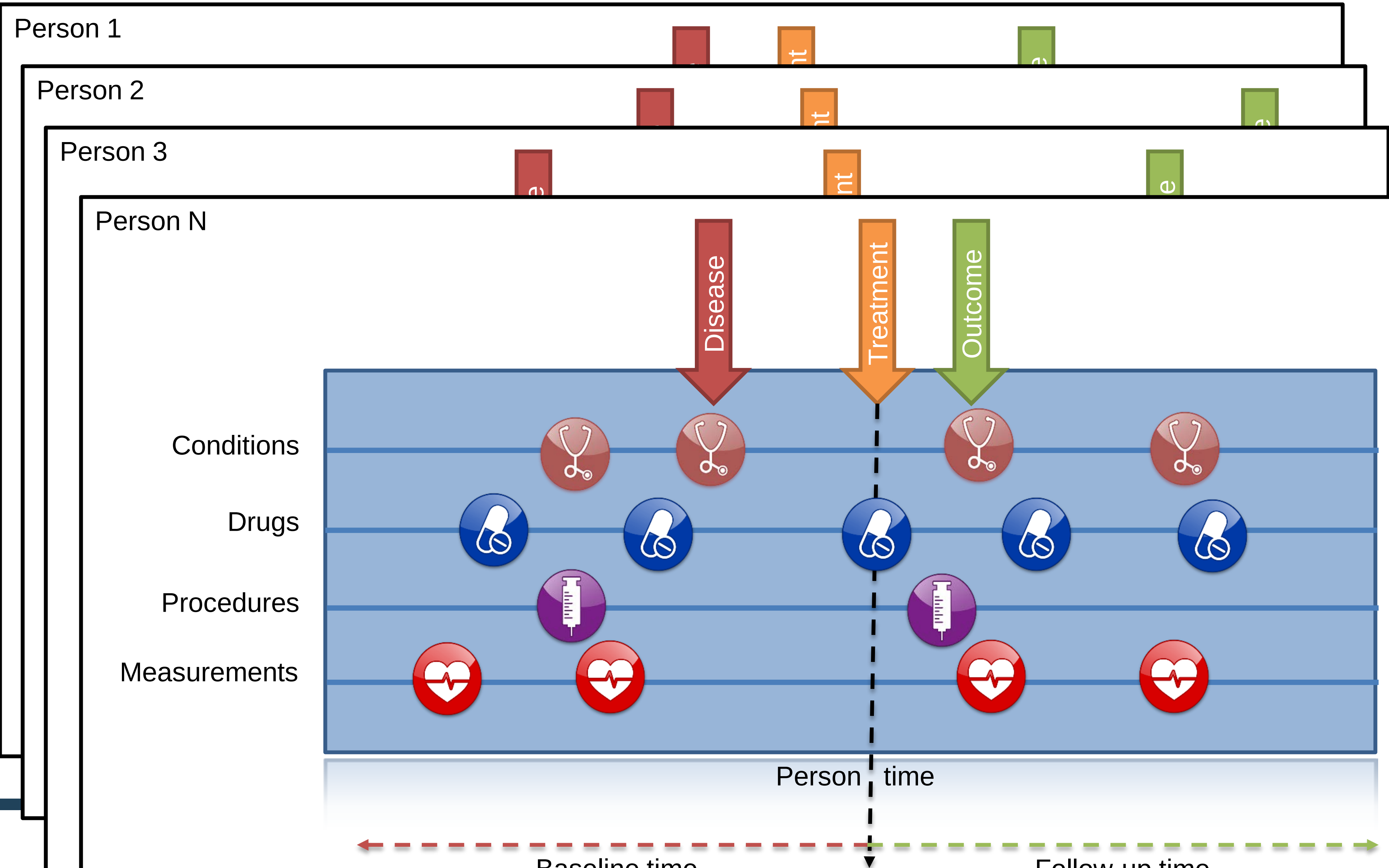
To improve health, by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

Hripcsak G, et al. (2015) Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. Stud Health Technol Inform 216:574–578.

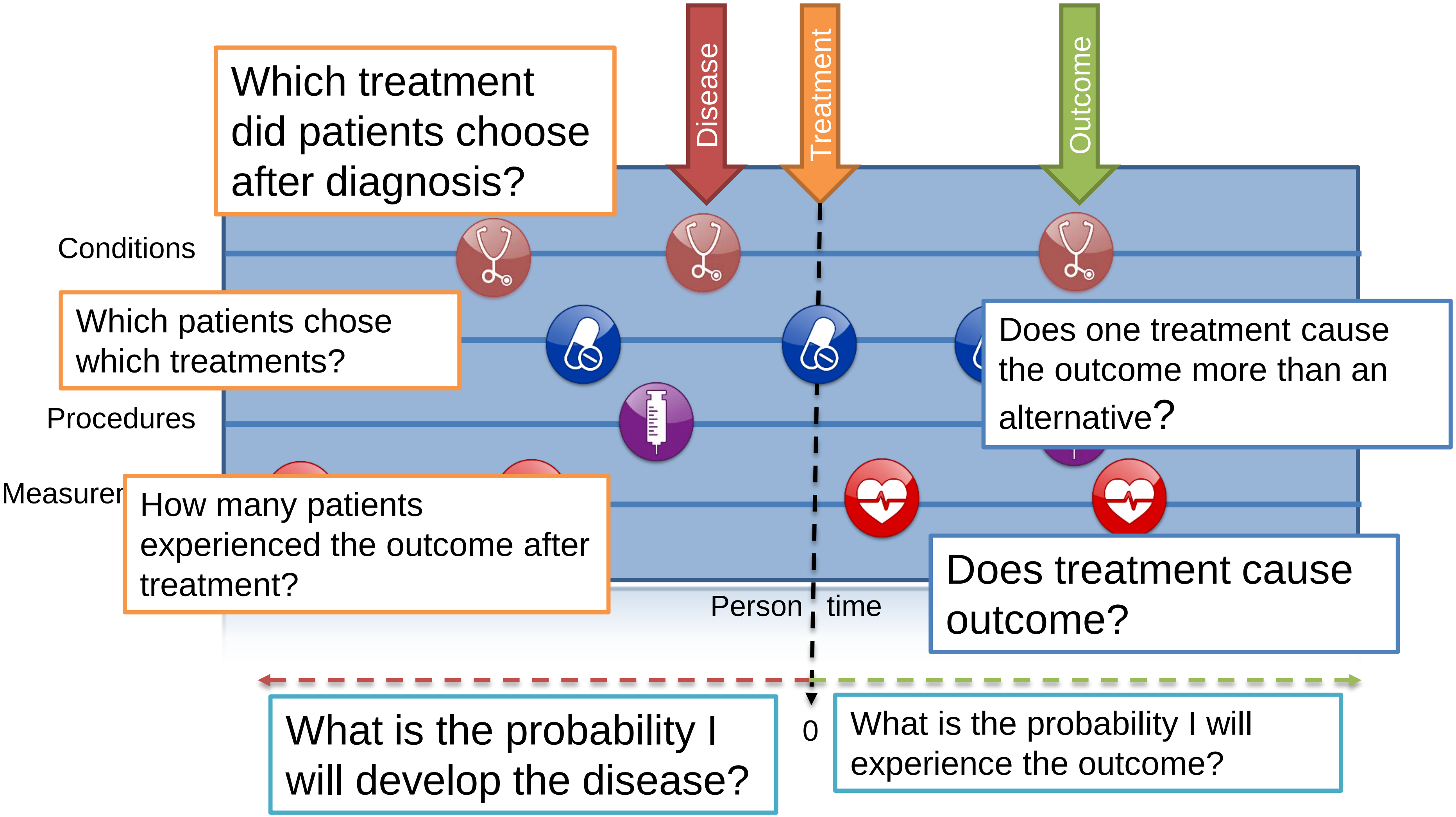
A caricature of the patient journey



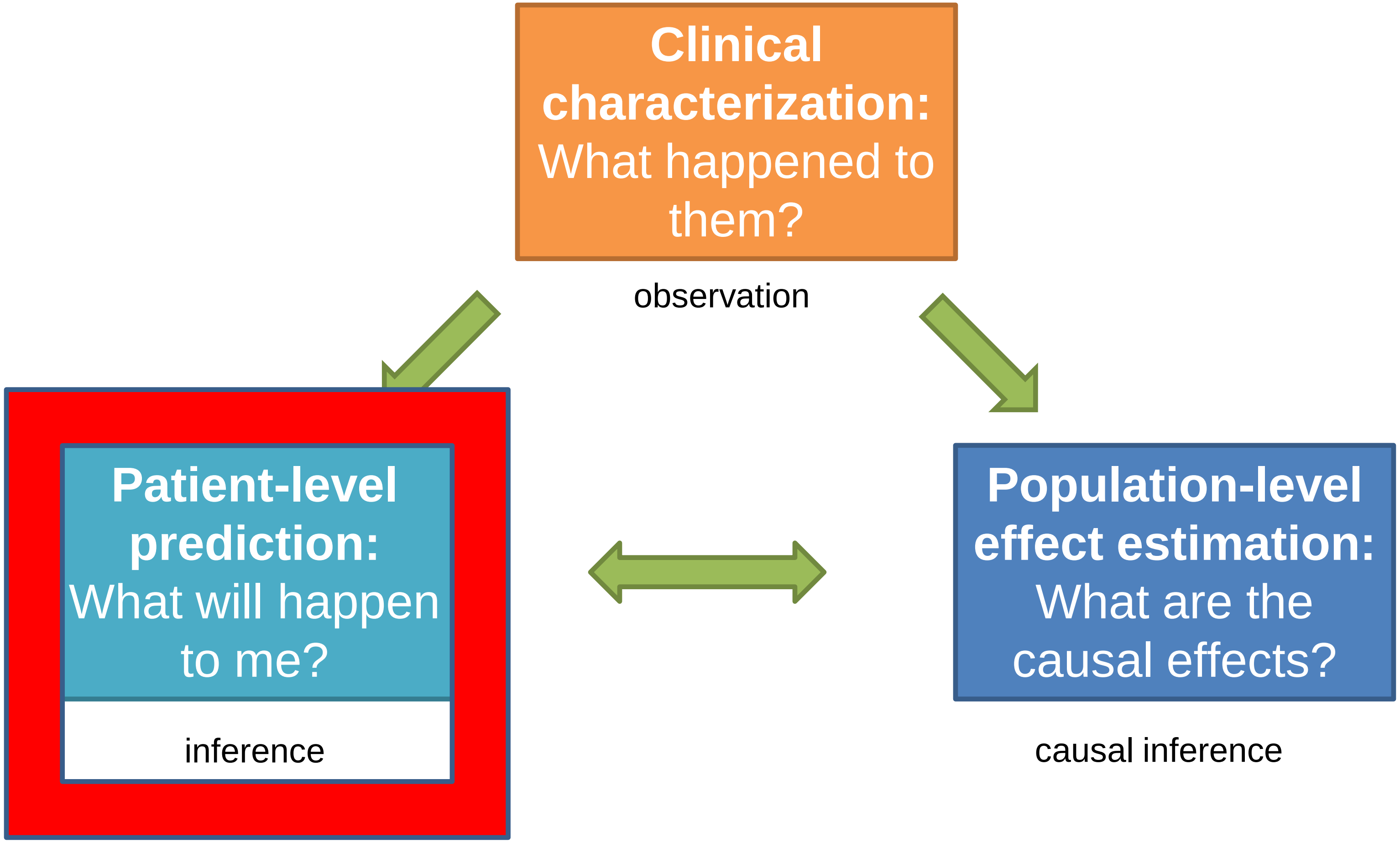
Each observational database is just an (incomplete) compilation of patient journeys



Questions asked across the patient journey



Complementary evidence to inform the patient journey



What is OHDSI's strategy to deliver reliable evidence?

- **Methodological research**
  - Develop new approaches to observational data analysis
  - Evaluate the performance of new and existing methods
  - Establish empirically-based scientific best practices
- **Open-source analytics development**
  - Design tools for data transformation and standardization
  - Implement statistical methods for large-scale analytics
  - Build interactive visualization for evidence exploration
- **Clinical evidence generation**
  - Identify clinically-relevant questions that require real-world evidence
  - Execute research studies by applying scientific best practices through open-source tools across the OHDSI international data network
  - Promote open-science strategies for transparent study design and evidence dissemination

Today’s Agenda

Time	Topic
8:45 – 9:00	Welcome, get settled, get laptops ready
9:00 – 10:30	**Presentation:** What is Patient-Level Prediction?
10:30 – 10:45	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 1 Theory
10:45 – 11:45	**Presentation:** Overview of the TRIPOD Statement **Exercise:** Applying TRIPOD to CHADS2
11:45 – 12:30	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 2 Implementation
12:30 – 13:15	Lunch
13:15 – 14:30	**Exercise:** Guided tour through implementing patient-level prediction
14:30 – 14:45	Break
14:45 – 16:30	**Exercise:** Design and implement your own patient-level prediction
16:30 – 17:00	Lessons learned and feedback

What is Patient-Level Prediction?

Peter Rijnbeek, PhD
Erasmus MC

Learning Objectives

Part 1: Learn what a patient-level prediction model is?

Part 2: Understand the patient-level prediction modelling process

Part 3: Gain insights from a proof-of-concept study in depression patients

Clinicians are confronted with prediction questions on a daily basis. What options do they have?

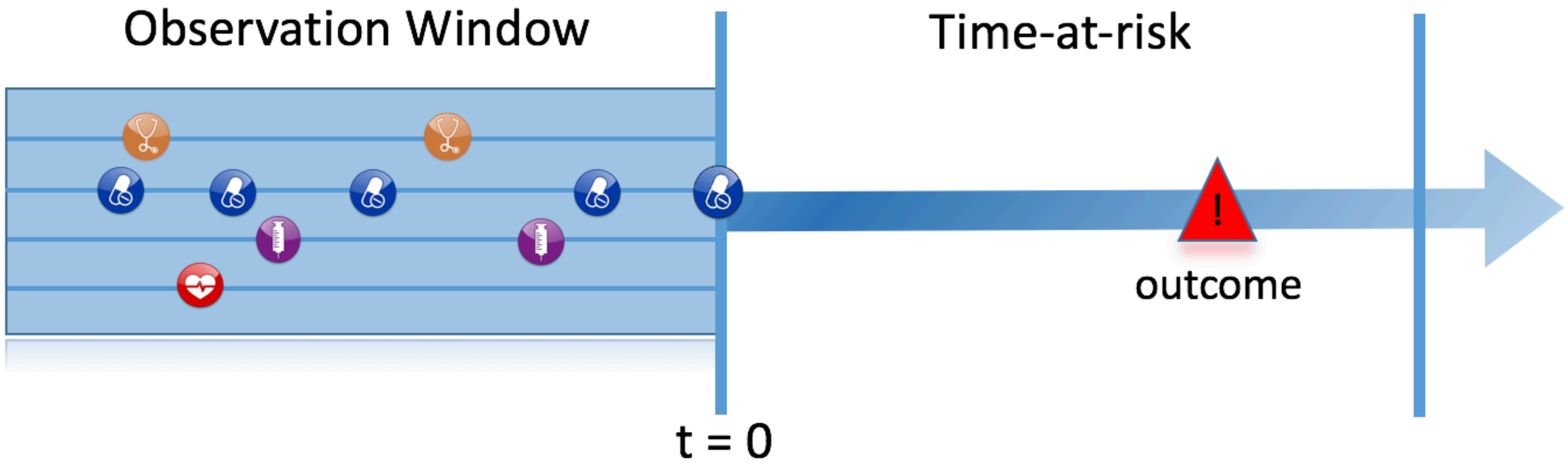
Deny ability to predict at the individual patient level

Quote an overall average to all patients

Utilize knowledge and personal experience

Provide a personalized prediction based on an advanced clinical prediction model

Problem definition



Among a target population (T), we aim to predict which patients at a defined moment in time (t=0) will experience some outcome (O) during a time-at-risk Prediction is done using only information about the patients in an observation window prior to that moment in time.

What are the key inputs to a patient-level prediction study?

Input parameter	Design choice
Target cohort (T)	
Outcome cohort (O)	
Time-at-risk	
Model specification -which model(s)? -which parameters? -which covariates?	

Types of prediction problems in healthcare

Type	Structure	Example
Disease onset and progression	Amongst patients who are newly diagnosed with **<insert your favorite disease>**, which patients will go on to have **<another disease or related complication>** within **<time horizon from diagnosis>**?	Among newly diagnosed AFib patients, which will go onto to have ischemic stroke in next 3 years?
Treatment choice	Amongst patients with **<indicated disease> who are treated with either <treatment 1> or <treatment 2>**, which patients were treated with **<treatment 1>** (**on day 0**)?	Among AFib patients who took either warfarin or rivaroxaban, which patients got warfarin? (as defined for propensity score model)
Treatment response	Amongst patients who are new users of **<insert your favorite chronically-used drug>**, which patients will **<insert desired effect>** in **<time window>**?	Which patients with T2DM who start on metformin stay on metformin after 3 years?
Treatment safety	Amongst patients who are new users of **<insert your favorite drug>,** which patients will experience **<insert your favorite known adverse event from the drug profile>** within **<time horizon following exposure start>**?	Among new users of warfarin, which patients will have GI bleed in 1 year?
Treatment adherence	Amongst patients who are new users of **<insert your favorite chronically-used drug>**, which patients will achieve **<adherence metric threshold>** at **<time horizon>**?	Which patients with T2DM who start on metformin achieve >=80% proportion of days covered at 1 year?

Difference between explanatory models and prediction models

People build a prediction model and make causal claims. This is not correct!

Why

Different interpretations of ”Model”

”Model” is being interpreted differently in Statistics, Epidemiology, and Data Science

- Statistics: models are used to describe data, it is more about data characterization
- Epidemiologist are trained to think about models as tests of hypotheses to perform causal inference
- Data Scientists interpret the word “model” in the context of predicting future events using the available data

It is important we understand what the difference is between explanatory modelling and predictive modelling!

Shmueli, G. 2011. Predictive Analytics in Information Systems Research. MIS Quarterly (35:3), pp. 553-57

Shmueli, G. 2010. To Explain or to Predict?, Statistical Science (25:3), pp. 289-310

Some definitions

Explanatory Model:	Theory-based statistical model for testing causal hypotheses
Explanatory Power:	Strength of the relationship in statistical model
Predictive Model:	Empirical model/algorithm for predicting new observations
Predictive Power:	Ability to accurately predict new observations

You can empirically evaluate the predictive power of explanatory model but you cannot empirically evaluate the explanatory power of a predictive model.

The best explanatory model is not necessary the best predictive model!

You do not have to understand the underlying causes in order to predict well!

Explanatory modelling versus Predictive analytics

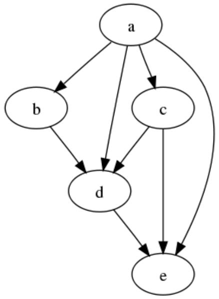


Table 1. Differences Between Explanatory Statistical Modeling and Predictive Analytics

Step	Explanatory	Predictive
Analysis Goal	Explanatory statistical models are used for testing causal hypotheses.	Predictive models are used for predicting new observations and assessing predictability levels.
Variables of Interest	Operationalized variables are used only as instruments to study the underlying conceptual constructs and the relationships between them.	The observed, measurable variables are the focus.
Model Building Optimized Function	In explanatory modeling the focus is on minimizing model bias. Main risks are type I and II errors.	In predictive modeling the focus is on minimizing the combined bias and variance. The main risk is over-fitting.
Model Building Constraints	Empirical model must be interpretable, must support statistical testing of the hypotheses of interest, must adhere to theoretical model (e.g., in terms of form, variables, specification).	Must use variables that are available at time of model deployment.
Model Evaluation	Explanatory power is measured by strength-of-fit measures and tests (e.g., R^2 and statistical significance of coefficients).	Predictive power is measured by accuracy of out-of-sample predictions.

Shmueli, G. 2011. Predictive Analytics in Information Systems Research. MIS Quarterly (35:3), pp. 553-57

Why should we avoid the term "Risk Factor"

"Risk Factor" is an ambiguous term.

A predictive model is not selecting parameters based on their explanatory power but it is **using** association to improve predictive accuracy -> **association does not equal causation**!

If your goal is to search for causal factors you should use population-level effect estimation.

If your goal is to search for association of individual parameters you should use clinical characterization.

We should avoid using the term "risk factors" and use the term predictors to make explicit that we are assessing predictive value.

How to interpret beta values in a logistic regression prediction model?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots$$

Each beta coefficient represents the additional effect of adding that variable to the model, if the effects of all other variables in the model are already accounted for.

→ any change of the model can result in a change of all the beta coefficients

Value	Association	Causation
b = 0	Unknown	Unknown
b <> 0	Yes	Unknown
b > 0	Positively associated under the assumption that all other beta values are fixed. If the variable is correlated to any other variable the direction of the association is unknown	Unknown
b < 0	Negatively associated under the assumption that all other beta values are fixed. If the variable is correlated to any other variable the direction of the association is unknown	Unknown

Why is predictive modelling still valuable?

1. In healthcare the question “What is going to happen to me?” is often more relevant than “Why?”

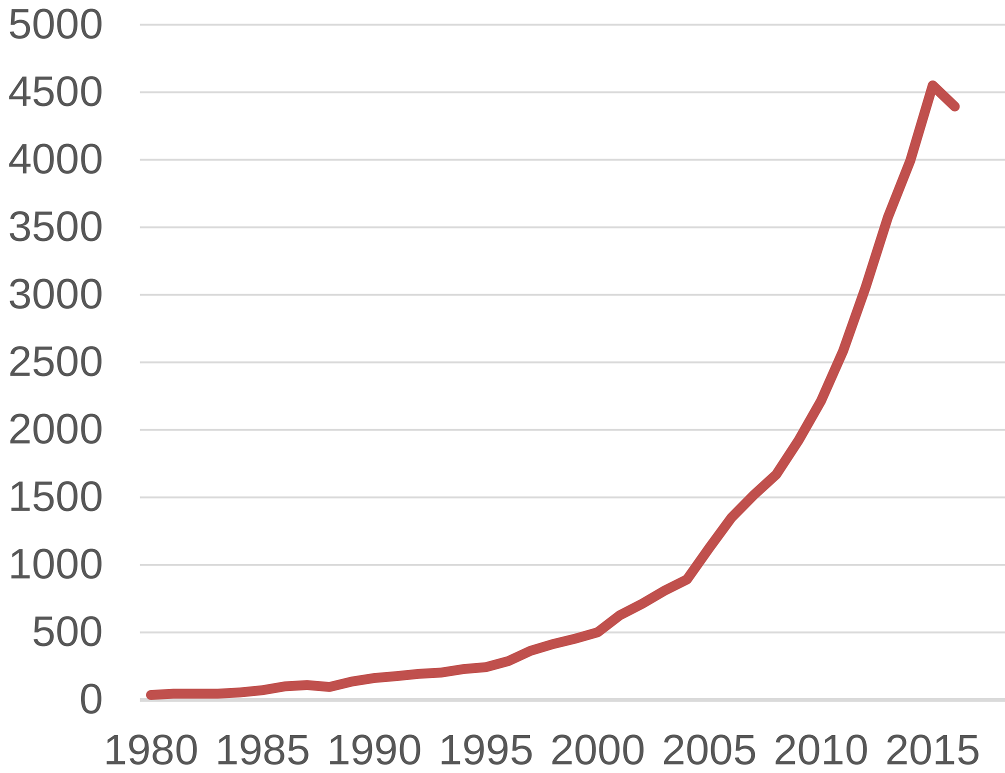
2. Knowing if something is predictable or not based on the available data is valuable on its own.

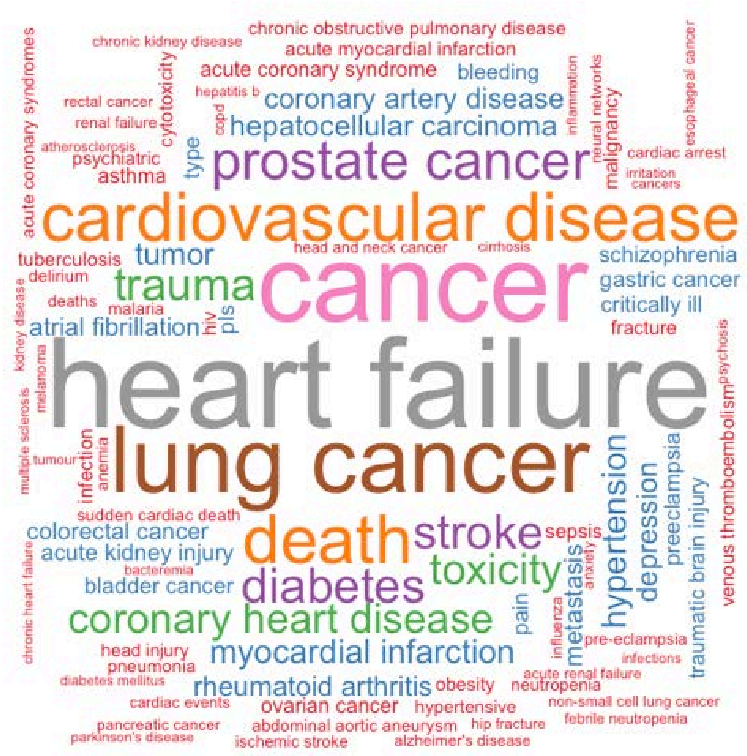
Types of prediction problems in healthcare

Type	Structure	Example
Disease onset and progression	Amongst patients who are newly diagnosed with **<insert your favorite disease>**, which patients will go on to have **<another disease or related complication>** within **<time horizon from diagnosis>**?	Among newly diagnosed AFib patients, which will go onto to have ischemic stroke in next 3 years?
Treatment choice	Amongst patients with **<indicated disease> who are treated with either <treatment 1> or <treatment 2>**, which patients were treated with **<treatment 1>** (**on day 0**)?	Among AFib patients who took either warfarin or rivaroxaban, which patients got warfarin? (as defined for propensity score model)
Treatment response	Amongst patients who are new users of **<insert your favorite chronically-used drug>**, which patients will **<insert desired effect>** in **<time window>**?	Which patients with T2DM who start on metformin stay on metformin after 3 years?
Treatment safety	Amongst patients who are new users of **<insert your favorite drug>,** which patients will experience **<insert your favorite known adverse event from the drug profile>** within **<time horizon following exposure start>**?	Among new users of warfarin, which patients will have GI bleed in 1 year?
Treatment adherence	Amongst patients who are new users of **<insert your favorite chronically-used drug>**, which patients will achieve **<adherence metric threshold>** at **<time horizon>**?	Which patients with T2DM who start on metformin achieve >=80% proportion of days covered at 1 year?

Questions?

Growing interest in prediction modelling



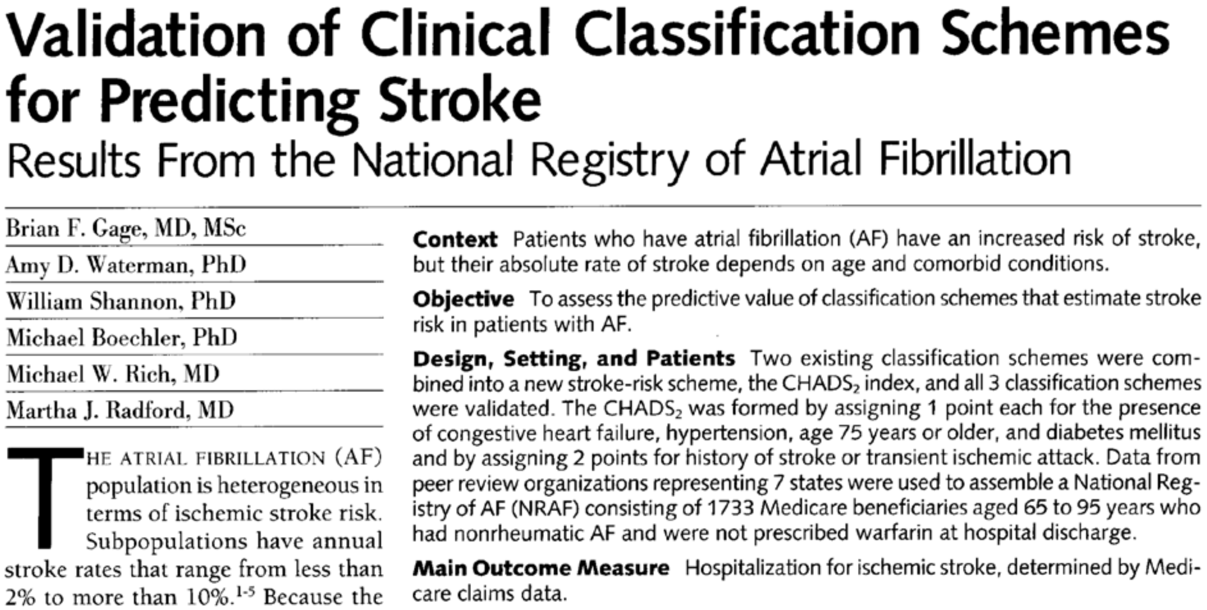


Reviews of published prediction models

- 800 models in individuals with CVD (Sessler 2015)
- 396 models for predicting cardiovascular disease (Damen 2016)
- 111 models for prostate cancer (Shariat 2008)
- 102 models for TBI (Perel 2006)
- 83 models for stroke (Counsell 2001)
- 54 models for breast cancer (Altman 2009)
- 43 models for type 2 diabetes (Collins 2011; van Dieren 2012)
  - 30+ more models have since been published!
- 31 models for osteoporotic fracture (Steurer 2011)
- 29 models in reproductive medicine (Leushuis 2009)
- 26 models for hospital readmission (Kansagara 2011)

Courtesy of Gary Collins

Predicting Stroke in patients with atrial fibrillation



Validation of Clinical Classification Schemes for Predicting Stroke
Results From the National Registry of Atrial Fibrillation

Brian F. Gage, MD, MSc
Amy D. Waterman, PhD
William Shannon, PhD
Michael Boechler, PhD
Michael W. Rich, MD
Martha J. Radford, MD

THE ATRIAL FIBRILLATION (AF) population is heterogeneous in terms of ischemic stroke risk. Subpopulations have annual stroke rates that range from less than 2% to more than 10%.[1-5] Because the

Context Patients who have atrial fibrillation (AF) have an increased risk of stroke, but their absolute rate of stroke depends on age and comorbid conditions.

Objective To assess the predictive value of classification schemes that estimate stroke risk in patients with AF.

Design, Setting, and Patients Two existing classification schemes were combined into a new stroke-risk scheme, the $CHADS_2$ index, and all 3 classification schemes were validated. The $CHADS_2$ was formed by assigning 1 point each for the presence of congestive heart failure, hypertension, age 75 years or older, and diabetes mellitus and by assigning 2 points for history of stroke or transient ischemic attack. Data from peer review organizations representing 7 states were used to assemble a National Registry of AF (NRAF) consisting of 1733 Medicare beneficiaries aged 65 to 95 years who had nonrheumatic AF and were not prescribed warfarin at hospital discharge.

Main Outcome Measure Hospitalization for ischemic stroke, determined by Medicare claims data.

CHADS2	Score
Congestive Heart Failure	1
Hypertension	1
Age $\geq$ 75	1
Diabetes	1
Stroke / TIA	2

How to define the CHADS$_2$ patient-level prediction problem?

Input parameter	Design choice
Target cohort (T)	Patients newly diagnosed with AF
Outcome cohort (O)	Stroke
Time-at-risk	1000 days
Model specification	Logistic Regression using 5 pre-selected covariates

Current status of predictive modelling

Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review

RECEIVED 27 October 2015
REVISED 25 January 2016
ACCEPTED 20 February 2016

INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD UNIVERSITY PRESS

Benjamin A Goldstein[1,2], Ann Marie Navar[2,3], Michael J Pencina[1,2], John PA Ioannidis[4,5]

ABSTRACT

Objective Electronic health records (EHRs) are an increasingly common data source for clinical risk prediction, presenting both unique analytic opportunities and challenges. We sought to evaluate the current state of EHR based risk prediction modeling through a systematic review of clinical prediction studies using EHR data.

Methods We searched PubMed for articles that reported on the use of an EHR to develop a risk prediction model from 2009 to 2014. Articles were extracted by two reviewers, and we abstracted information on study design, use of EHR data, model building, and performance from each publication and supplementary documentation.

Results We identified 107 articles from 15 different countries. Studies were generally very large (median sample size = 26 100) and utilized a diverse array of predictors. Most used validation techniques ($n = 94$ of 107) and reported model coefficients for reproducibility ($n = 83$). However, studies did not fully leverage the breadth of EHR data, as they uncommonly used longitudinal information ($n = 37$) and employed relatively few predictor variables (median = 27 variables). Less than half of the studies were multicenter ($n = 50$) and only 26 performed validation across sites. Many studies did not fully address biases of EHR data such as missing data or loss to follow-up. Average c-statistics for different outcomes were: mortality (0.84), clinical prediction (0.83), hospitalization (0.71), and service utilization (0.71).

Conclusions EHR data present both opportunities and challenges for clinical risk prediction. There is room for improvement in designing such studies.

Goldstein BA, J Am Med Inform Assoc. 2016.

Current status of predictive modelling

- Inadequate internal validation
- Small sets of features
- Incomplete dissemination of model and results
- No transportability assessment
- Impact on clinical decision making unknown

→ Relatively few prediction models are used in clinical practice

OHDSI Mission for Patient-Level Prediction

OHDSI aims to develop a systematic process to learn and evaluate large-scale patient-level prediction models using observational health data in a data network

- Evidence Generation
- Evidence Evaluation
- Evidence Dissemination

Part 2: How to build and validate a prediction model?

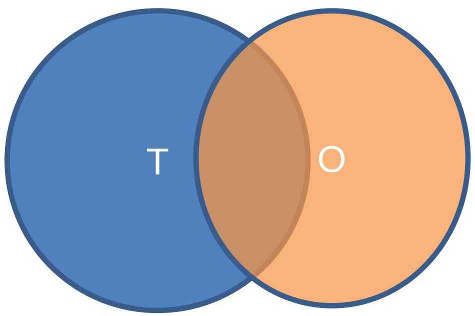
Prediction Model Development

Problem Definition → Data Extraction → Training → Internal Validation → External Validation → Dissemination

Problem pre-specification. A study protocol should unambiguously pre-specify the planned analyses.

Transparency. Others should be able to reproduce a study in every detail using the provided information. All analysis code should be made available as open source on the OHDSI Github.

Prediction Model Development



We extract data for the patients in the Target Cohort (T) and we select all patients that experience the outcome (O)

The Target Cohort (T) and Outcome Cohort (O) can be defined using ATLAS or custom code (see later today).

For model development all outcomes (O) of patients in the Target Cohort (T) are used.

Prediction Model Development

Problem Definition → Data Extraction → Training → Internal Validation → External Validation → Dissemination

Age	Gender	Var 1	...	Var n	Stroke
34	M	1	0	1	1
50	F	0	0	0	1
24	M	0	1	0	0
67	F	1	1	0	0
34	F	0	0	0	0
65	M	1	0	0	1
34	F	0	0	0	0
40	M	0	1	1	0

Data is extracted from the OMOP CDM using the Feature Extraction R-Package.

This allows for specification of the candidate predictors and time windows.

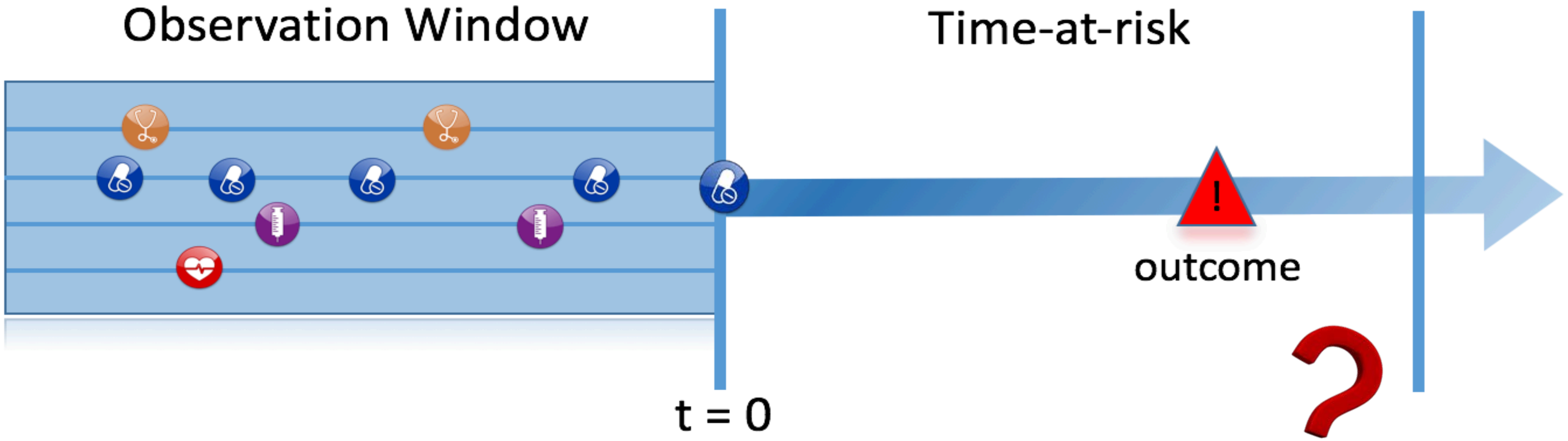
Prediction Model Development

Model training and **Internal validation** is done using a train test split:

1. Person split: examples are assigned randomly to the train or test set, or

2. Time split: a split is made at a moment in time (temporal validation)

Model Training

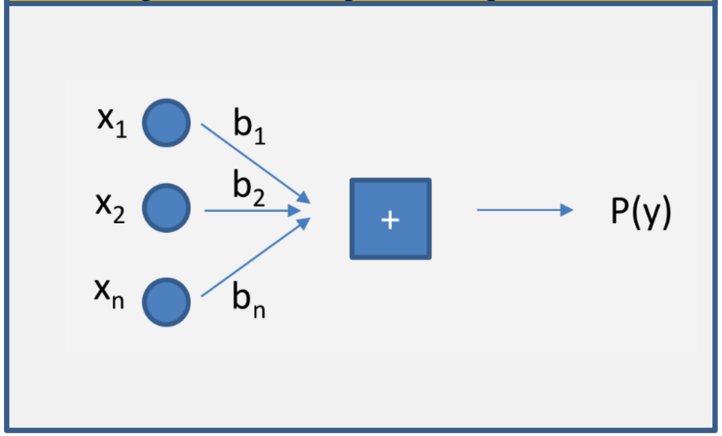


1. Which models?

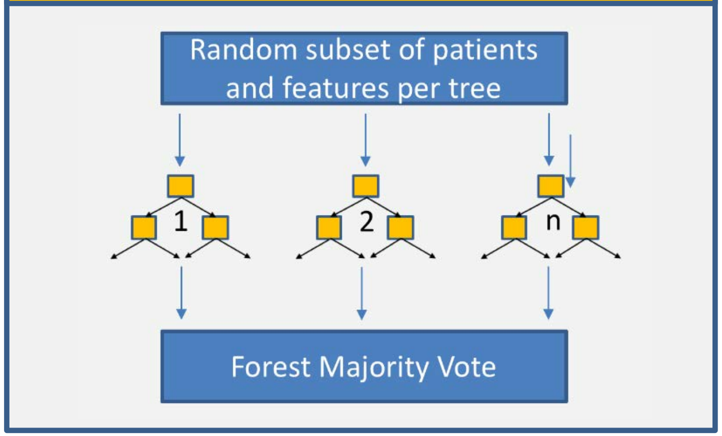
2. How to evaluate the model?

Models and Algorithms

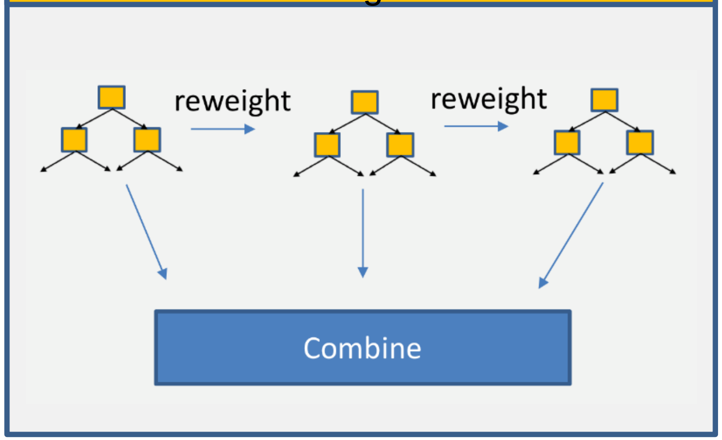
Regularized Logistic Regression



Random Forest



Gradient Boosting Machines



Many other models for example:

K-nearest neighbors
Naïve Bayes
Decision Tree
Adaboost
Neural Network
Etc.

Model selection is an empirical process

The **"No Free Lunch" theorem** states that there is not one model that works best for every problem. The assumptions of a great model for one problem may not hold for another problem.

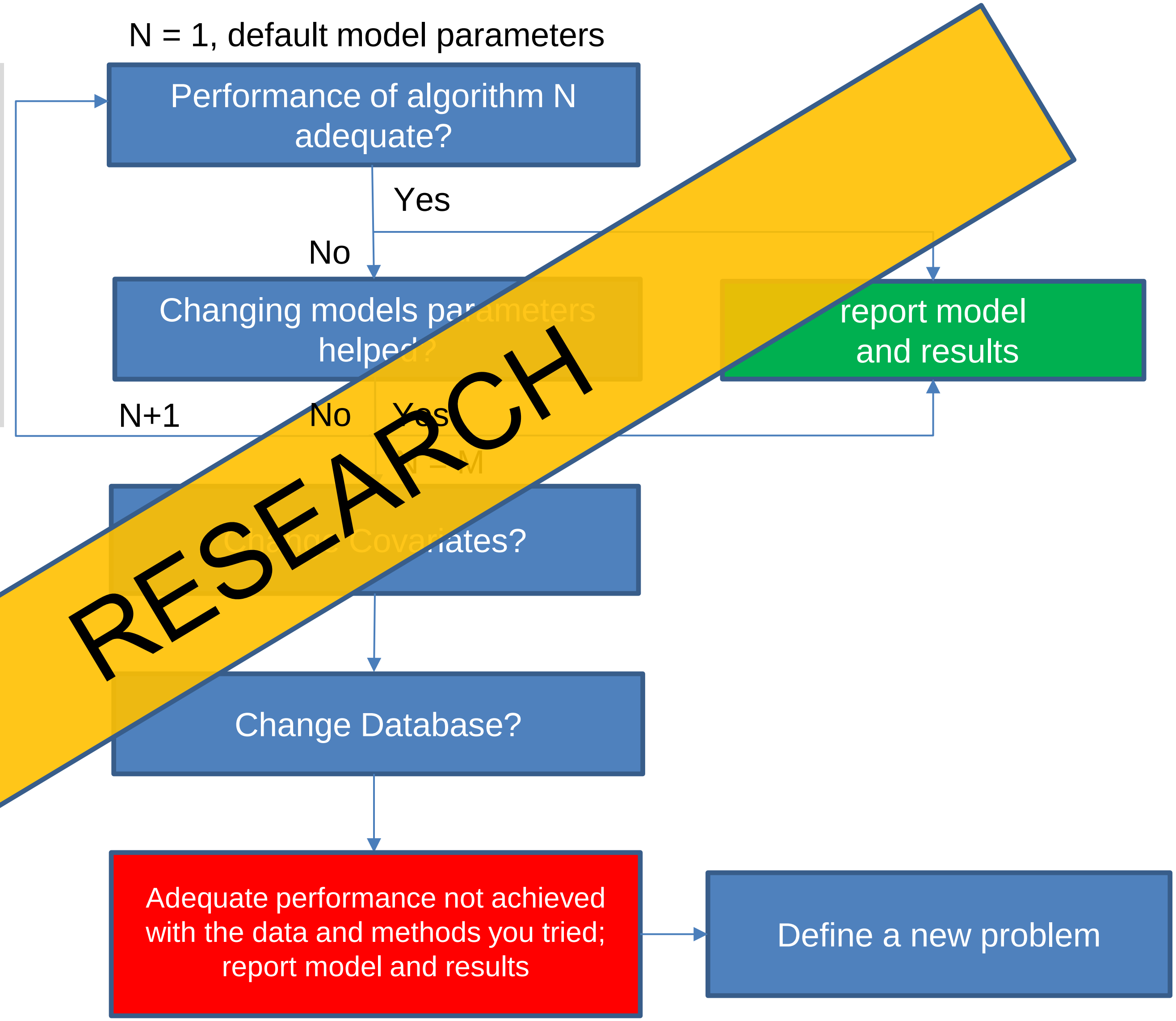
It is common in machine learning to try multiple models and find one that works best for that particular problem.

see no-free-lunch.org

OHDSI Model Selection Strategy

Suggested ordering of available algorithms in PLP package

N	Algorithms
1.	Lasso Logistic Regression
2.	Random Forest
3.	Gradient Boosting Machine
4.	Neural Network
5.	KNN
M.	…



Patient-Level Prediction Roadmap

Evidence Generation

Evidence Evaluation

Evidence Dissemination

Protocol Sharing
CDM Extractions
Code Sharing
Train / Test split

Model Validation

What makes a good model?

Discrimination: differentiates between those with and without the event, i.e. predicts higher probabilities for those with the event compared to those who don't experience the event

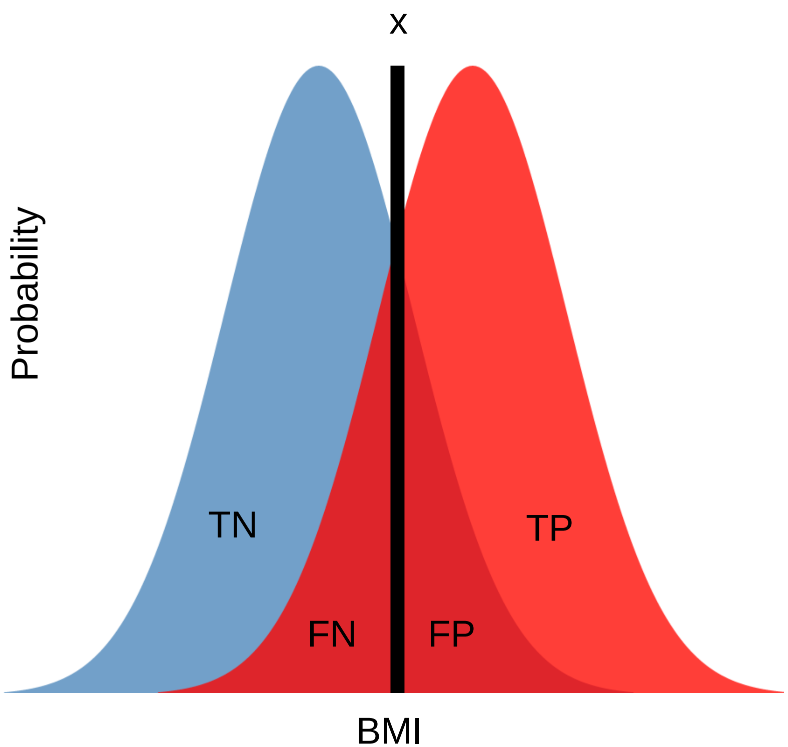
Calibration: estimated probabilities are close to the observed frequency

How to assess discrimination?

Suppose our classifier is simply BMI > x.

Both classes (blue = 0 , red = 1) have their own probability distribution of BMI

The choice of X then determines how sensitive or specific our algorithm is.



		Predicted	
		1	0
Observed	1		
	0		

True Positive Rate (TPR) = TP / (TP + FN)
False Positive Rate (FPR) = FP / (FP + TN)

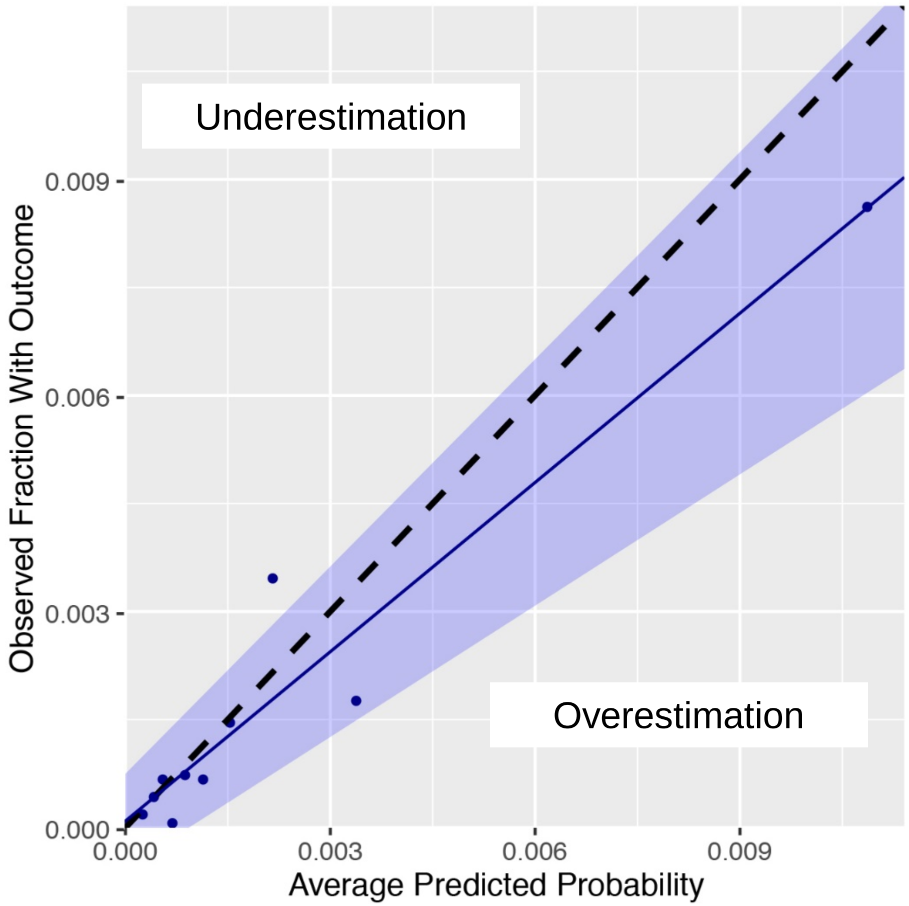
Receiver Operator Characteristic (ROC) curve

Calibration

- Agreement between observed and predicted risk
- We want a model that has good calibration across the range of predictions (not just on average)
- A model is well calibrated if for every 100 individuals given a risk of p% close to p have the event.
- For example, if we predict a 12% risk that an atrial fibrillation patient will have a stroke within 365 days, the observed proportion should be approx. 12 strokes per 100 patients

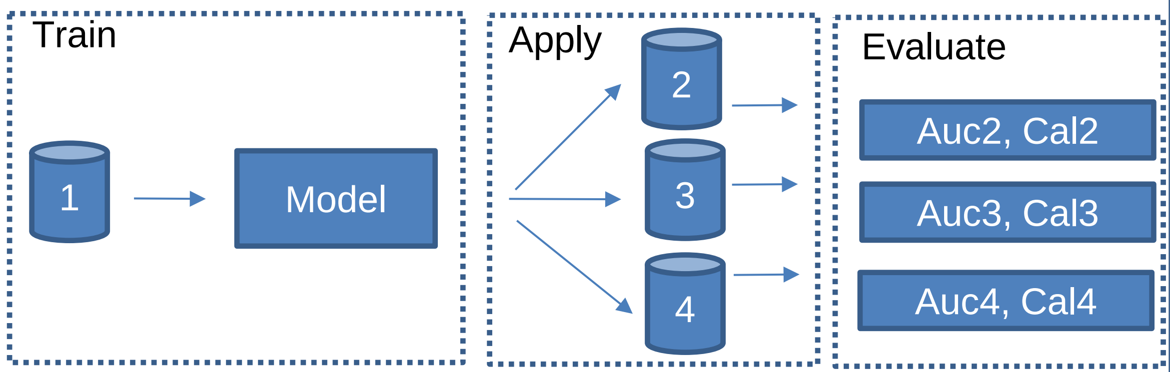
Calibration Assessment

How close is the average predicted probability to the observed fraction with the outcome?



External Validation

External validation is performed using data from multiple populations not used for training.



Patient-Level Prediction Roadmap

Evidence Generation

Evidence Evaluation

Evidence Dissemination

- Protocol Sharing
- CDM Extractions
- Code Sharing
- Train / Test split

- Standardized Process
- Discrimination
- Calibration
- External Validation

Dissemination

Problem Definition → Data Extraction → Training → Internal Validation → External Validation → **Dissemination**

Dissemination of study results should follow the minimum requirements as stated in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [1].

- Internal and external validation
- Sharing of full model details
- Sharing of all analyses code to allow full reproducibility

Website to share protocol, code, models and results for all databases

[1] Moons, KG et al. Ann Intern Med. 2015;162(1):W1-73

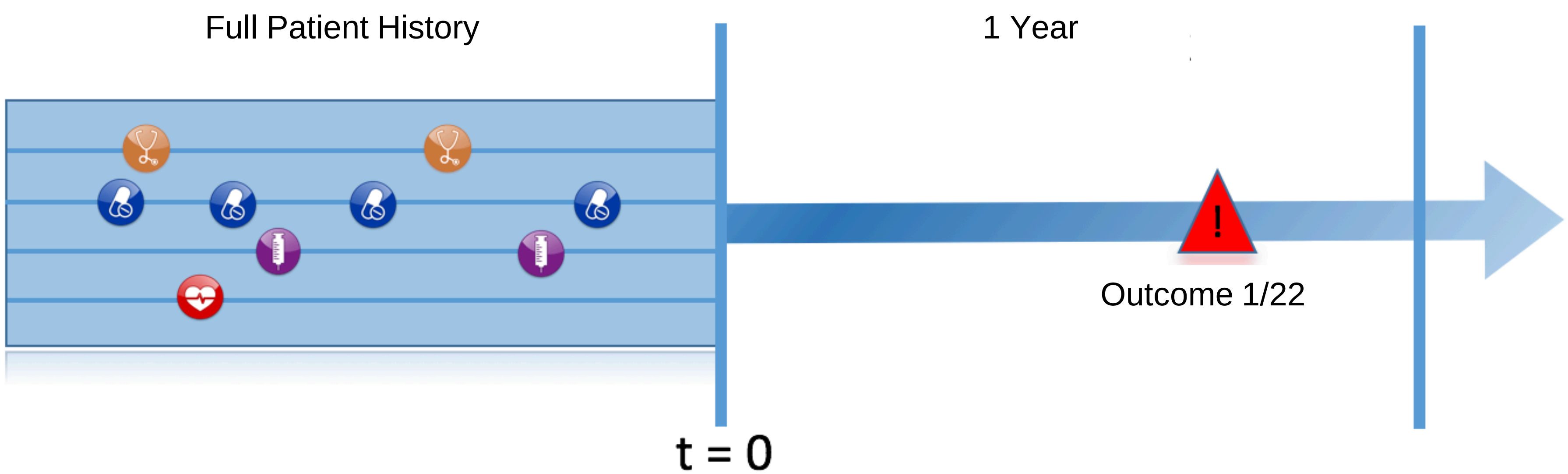
Patient-Level Prediction Roadmap

Evidence Generation	Evidence Evaluation	Evidence Dissemination
Protocol Sharing CDM Extractions Code Sharing Train / Test split	Standardization Discrimination Calibration External Validation	Publications (TRIPOD) Model sharing Full transparency

Questions?

Part 3:
Prediction in Patients with Pharmaceutically Treated Depression

Problem definition



First Pharmaceutically Treated Depression

Among patients **in 4 different databases**, we aim to develop prediction models to predict which patients at a defined moment in time (**First Pharmaceutically Treated Depression Event**) will experience one out of **22 different outcomes** during a time-at-risk (**1 year**). Prediction is done using **all demographics, conditions, and drug use** data prior to that moment in time.

Target (T) Cohort Definition

Patients are included in the cohort of interest at the date of the first occurrence of Pharmaceutically Treated Depression if the following inclusion criteria apply:

1. At least 365 days of history

2. At least 365 days of follow-up or the occurrence of the outcome of interest

3. No occurrence of the event prior to the index date

Setting

Databases

Database	Depression	Stroke
CCAE	659402	1351
MDCD	79818	356
MDCR	57839	874
OPTUM	363051	1183

Data extraction

- All demographics, conditions, drugs
- All 22 outcome cohorts

Training and testing

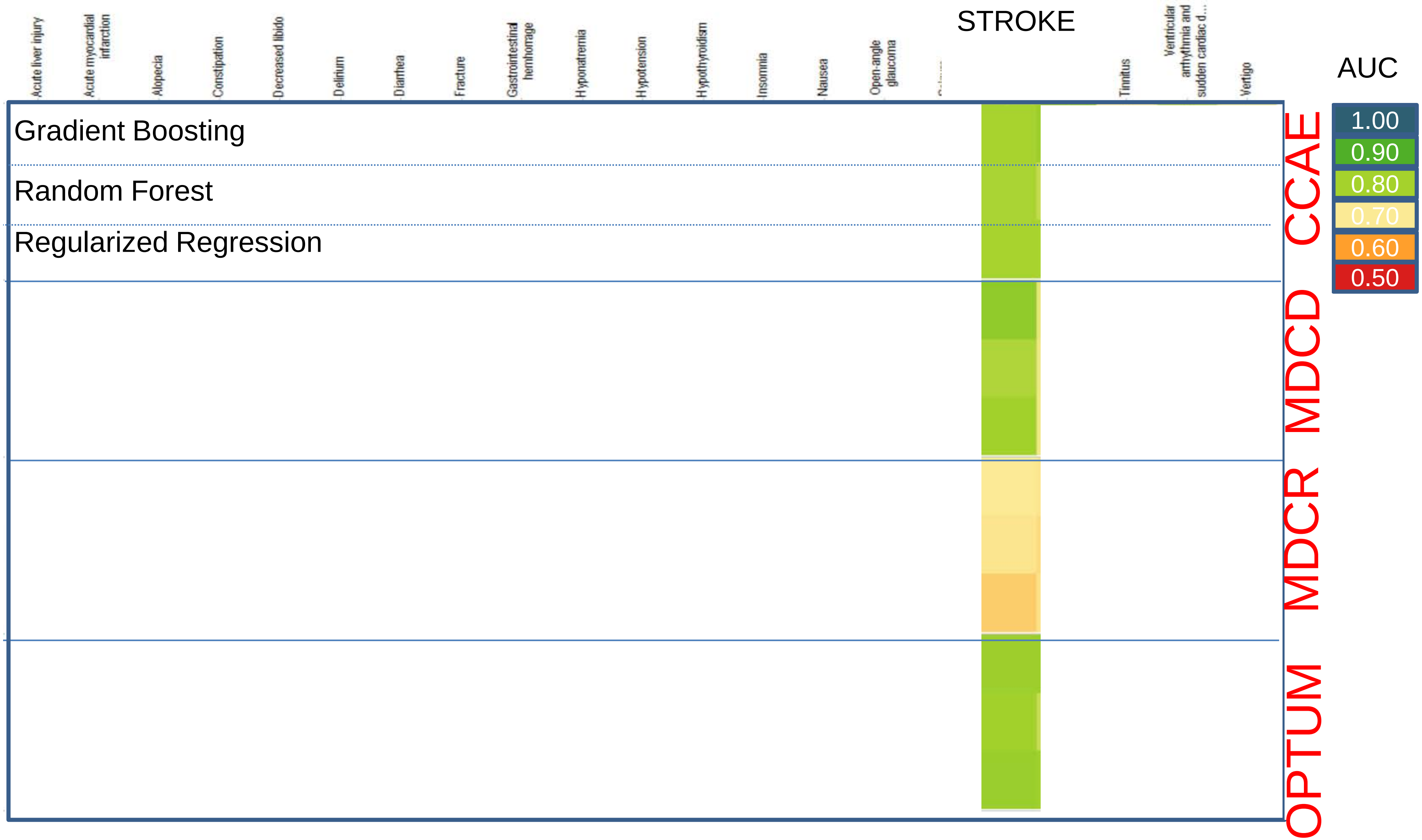
- Time split for training and testing
- Transportability for Stroke

Models

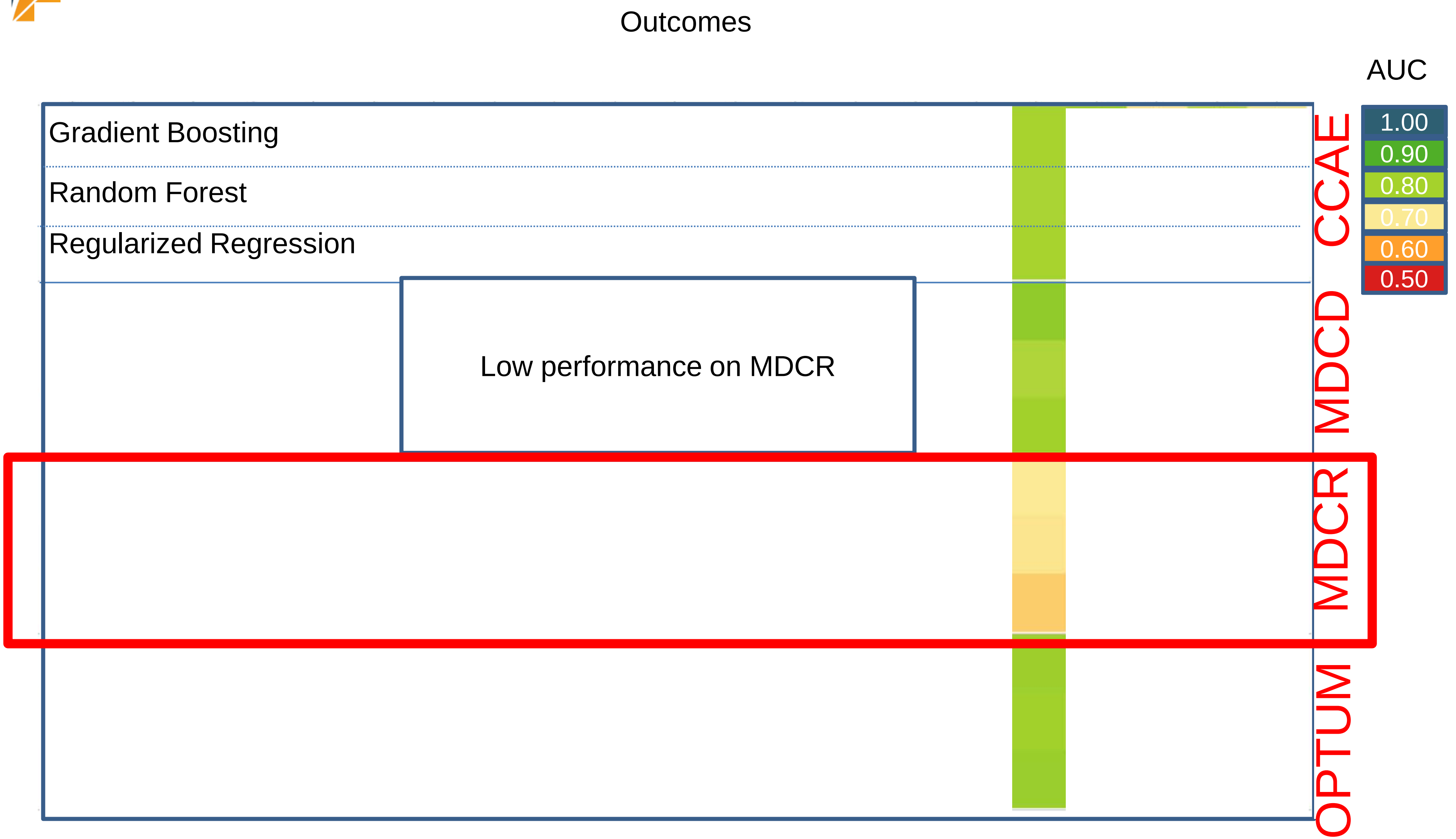
- Gradient Boosting
- Random Forest
- Regularized Regression

Outcomes
Acute liver injury
Acute myocardial infarction
Alopecia
Constipation
Decreased libido
Delirium
Diarrhea
Fracture
Gastrointestinal hemhorrage
Hyperprolactinemia
Hyponatremia
Hypotension
Hypothyroidism
Insomnia
Nausea
Open-angle glaucoma
Seizure
Stroke
Suicide and suicidal ideation
Tinnitus
Ventricular arrhythmia and sudden cardiac death
Vertigo

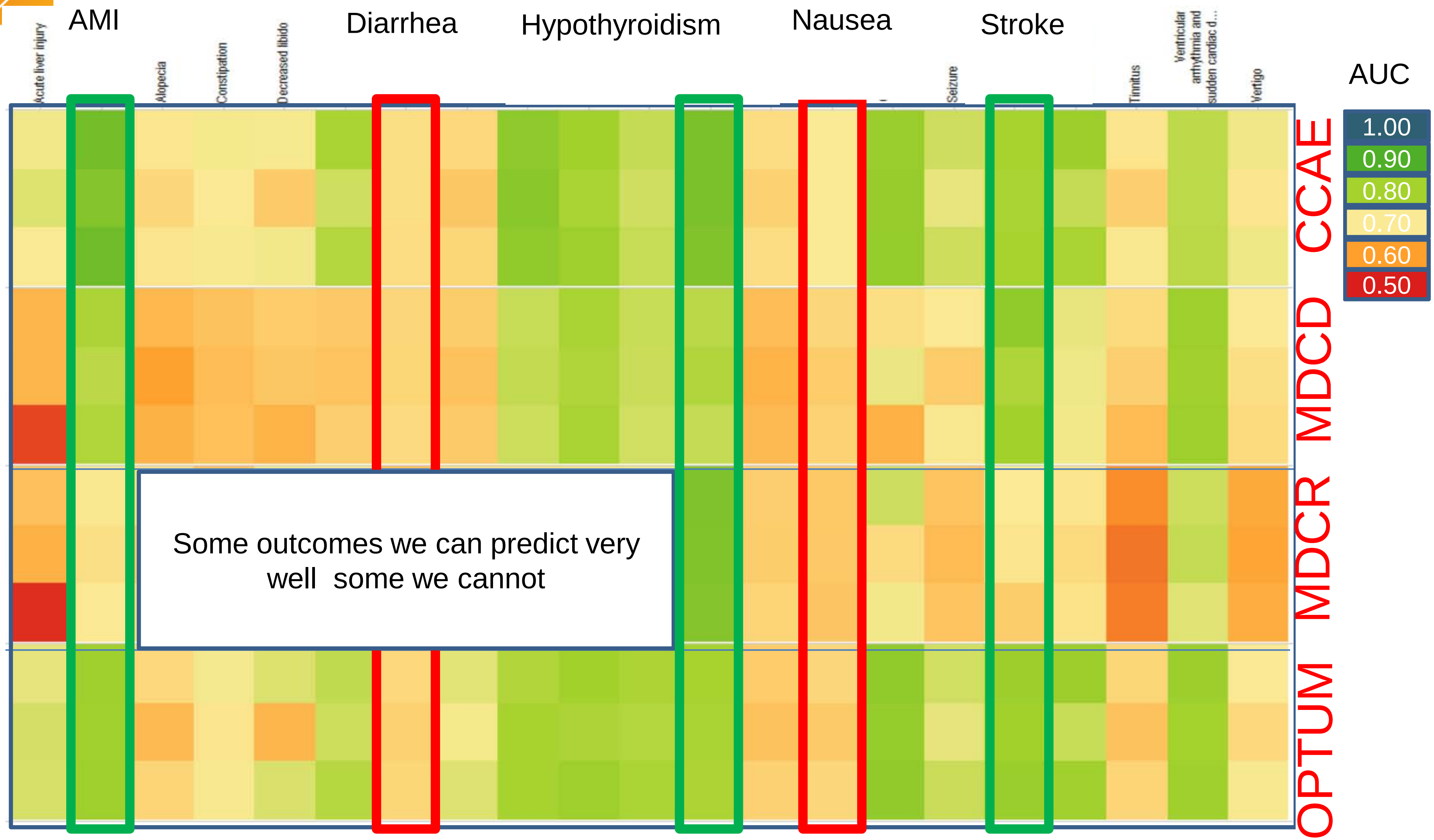
Model Discrimination Stroke



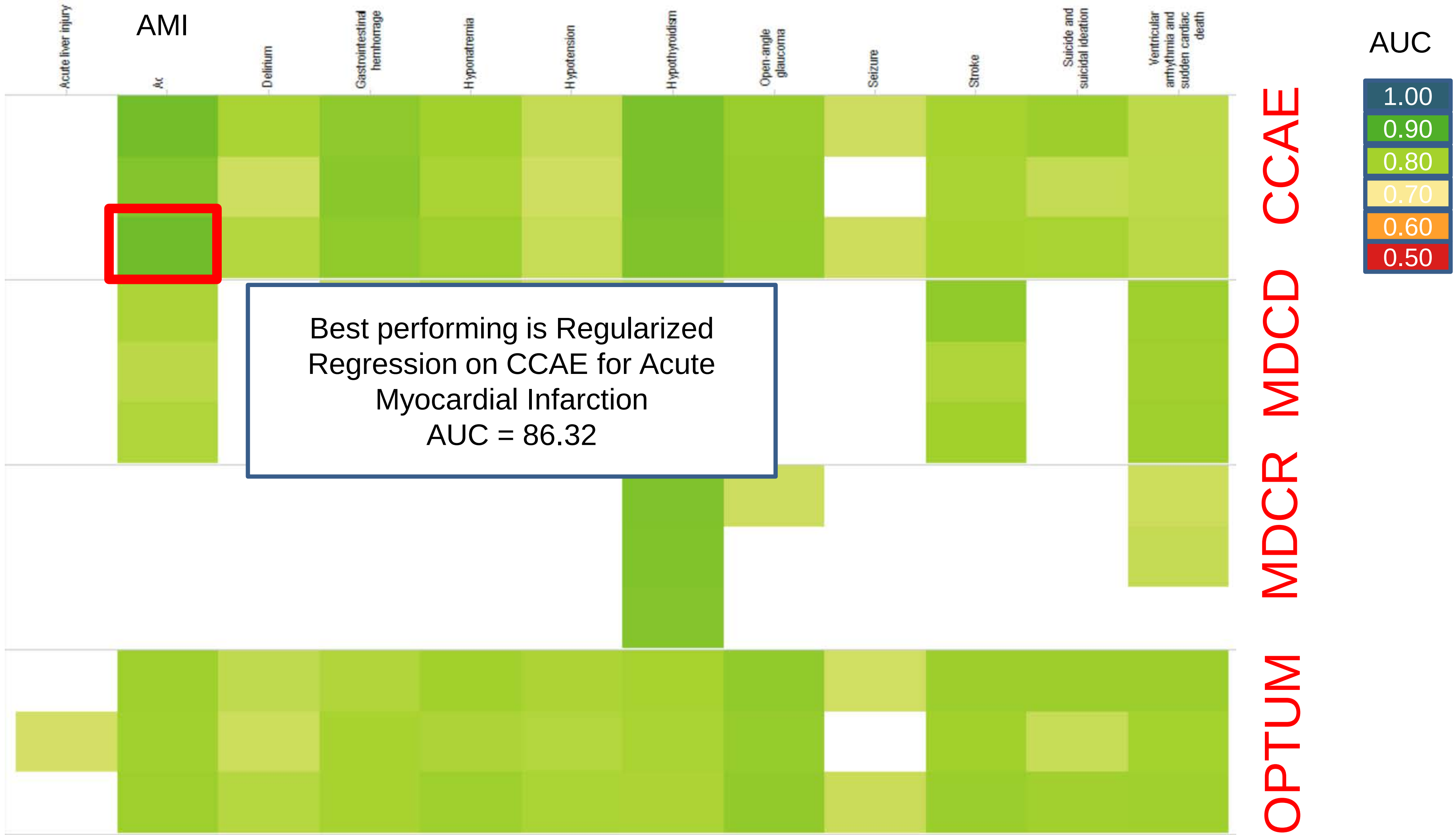
Model Discrimination



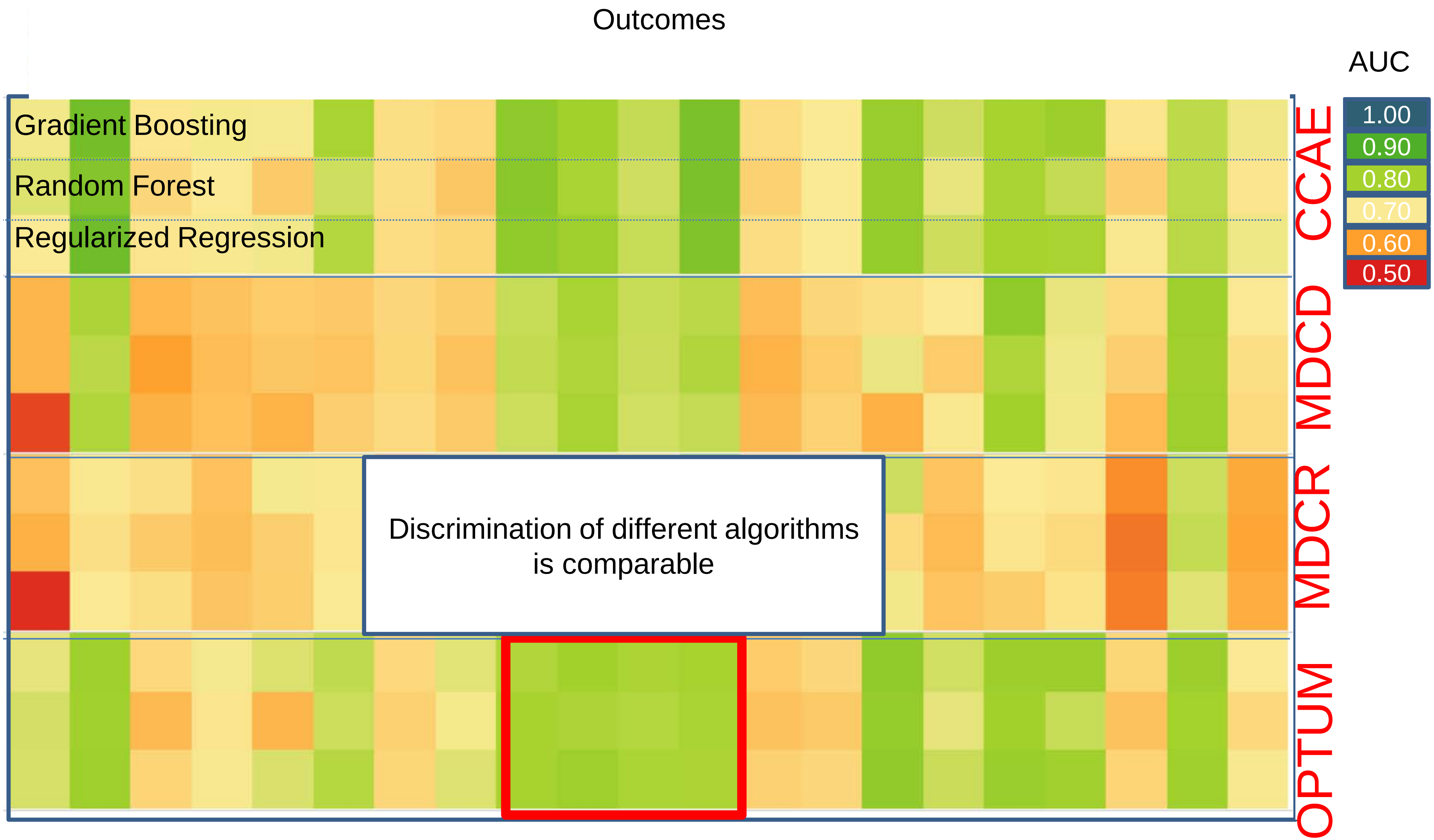
Model Discrimination



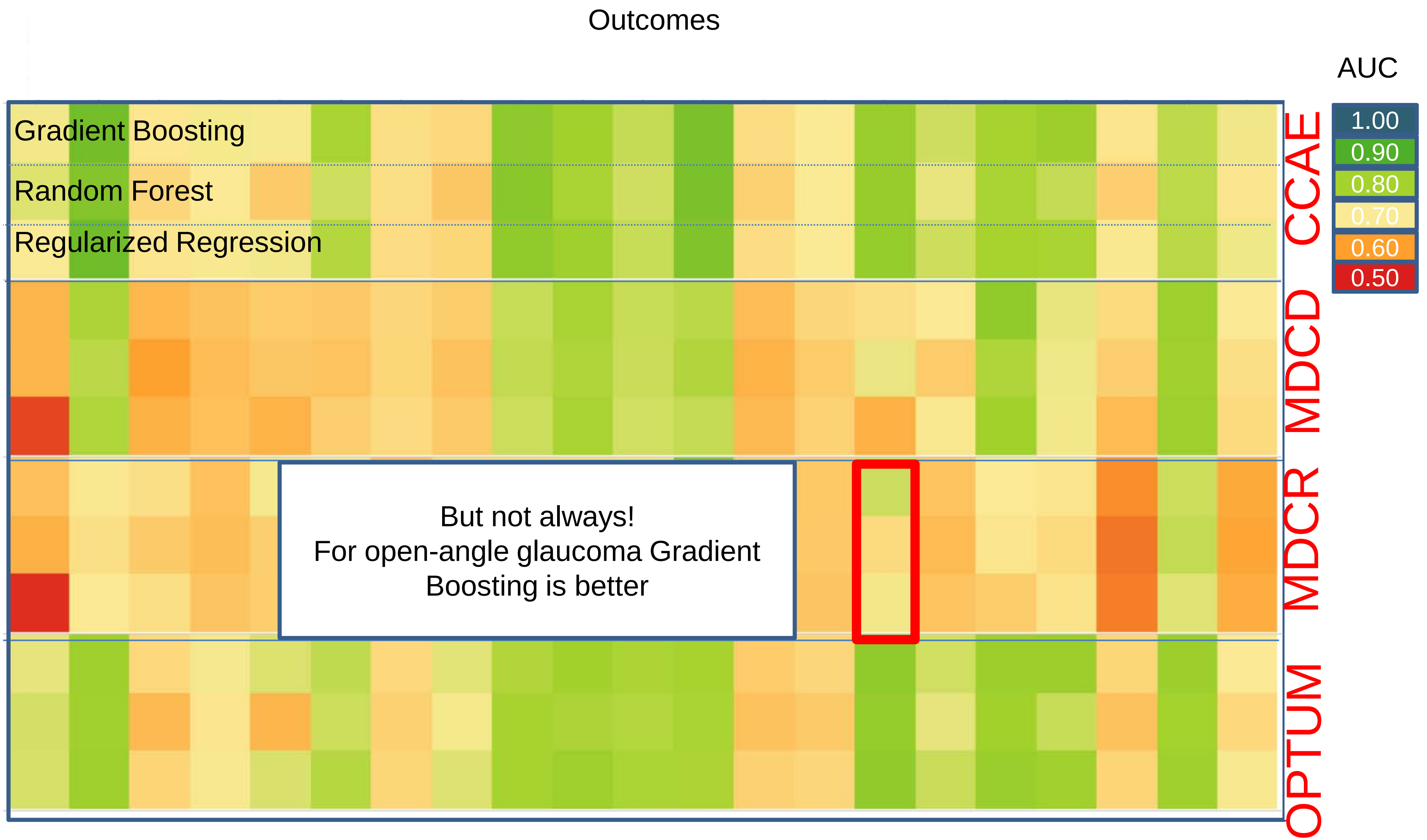
Outcomes with AUC > 0.75



Model Discrimination



Model Discrimination

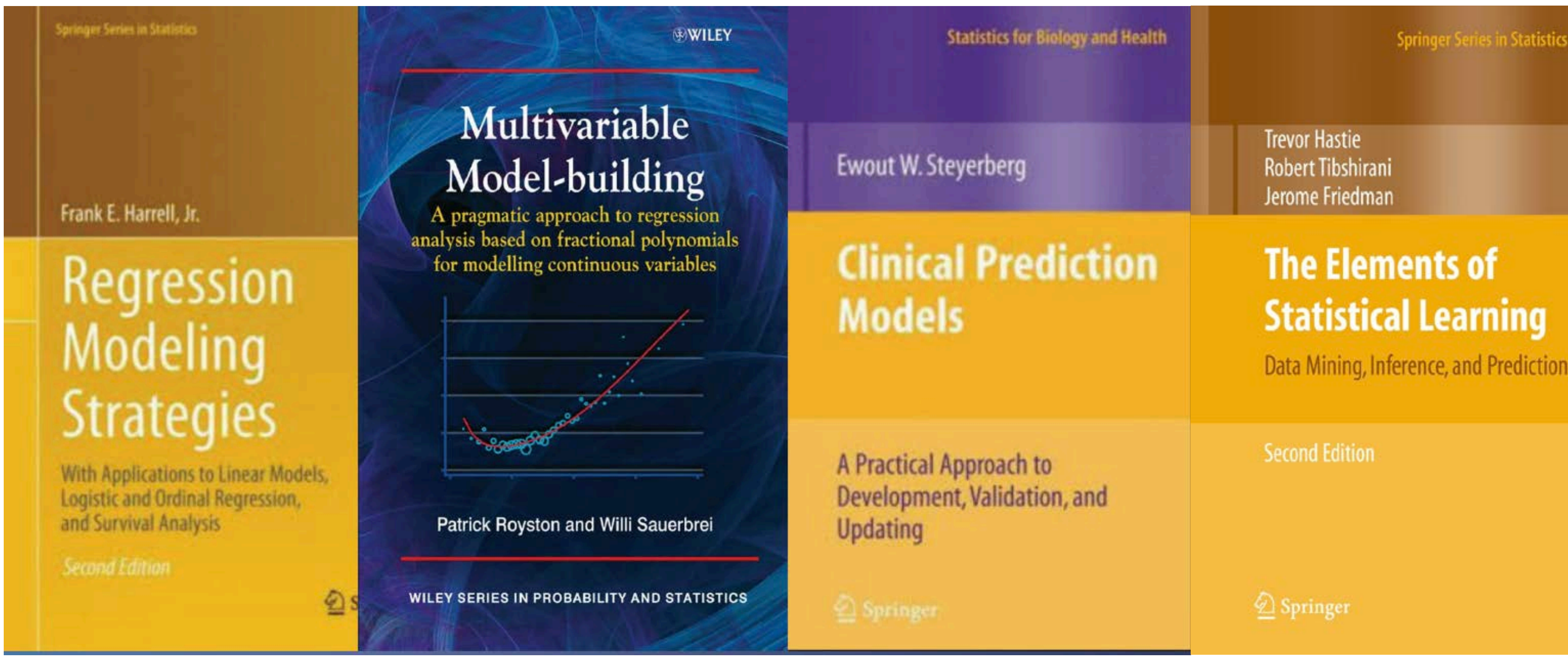


What did we achieve so far?

We showed it is feasible to develop large-scale predictive models for all databases converted to the OMOP CDM. This can now be done for any target cohort (T), outcome (O), and time at risk.

Further Reading if you got very interested!

- Phases of Clinical Prediction Modeling BMJ Series 2009
- Many good textbooks:



Today's Agenda

Time	Topic
8:45 – 9:00	Welcome, get settled, get laptops ready
9:00 – 10:30	**Presentation:** What is Patient-Level Prediction?
10:30 – 10:45	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 1 Theory
10:45 – 11:45	**Presentation:** Overview of the TRIPOD Statement **Exercise:** Applying TRIPOD to CHADS2
11:45 – 12:30	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 2 Implementation
12:30 – 13:15	Lunch
13:15 – 14:30	**Exercise:** Guided tour through implementing patient-level prediction
14:30 – 14:45	Break
14:45 – 16:30	**Exercise:** Design and implement your own patient-level prediction
16:30 – 17:00	Lessons learned and feedback

Learning Goals

1. Our 5-step Framework
2. What goes into PatientLevelPrediction
3. Implement various machine learning techniques
4. What comes out of PatientLevelPrediction
5. Interpretation of model performance metrics and plots

Learning Goals

1. Our 5-step Framework
2. What goes into PatientLevelPrediction
3. Implement various machine learning techniques
4. What comes out of PatientLevelPrediction
5. Interpretation of model performance metrics and plots

Part 1: Our Framework

1. Specify Problem
2. Identify Suitable Data
3. Select Predictor Variables
4. Select Models
5. Validate

1. Specify Problem

Objective:	*Specify the prediction problem in the form: In **[target population]** predict who will develop **[outcome]** during **[time-at-risk]** relative to target index date*
Elements:	▪Define the **target population**, the patients to whom you wish to apply to model. ▪Define the **outcome** for which you wish to predict the risk. ▪Define the **time-at-risk period**; this is the time interval within which you wish to predict the outcome occurring.
Benefit:	**A consistent problem definition increases transparency**

2. Identify Suitable Data

Objective:	*Select the dataset that will be used to develop the model (or try various datasets and pick the best model based on validation)*
Considerations:	▪Check that the target population is of **sufficient size** for model development. ▪Check that there a **sufficient number of outcomes** in the target population during the time at risk. ▪What things are captured in the data (are labs and measurements included?)
Note:	**The prediction ability can depend on the database**

3. Select Predictor Variables

Objective:	*Select from a set of* ***standardized*** *predictor variables or create custom predictors (although we strongly recommend selecting all standardized variables)*
Elements:	▪Can pick different time periods to construct variables prior to target population cohort start date. ▪Can pick from demographics, conditions, drugs, measurements, procedures and observations concepts. ▪Can group concepts based on a hierarchy in the vocabulary.
Benefit:	**The standardisation of variables means the variables are more transparent and reproducible**

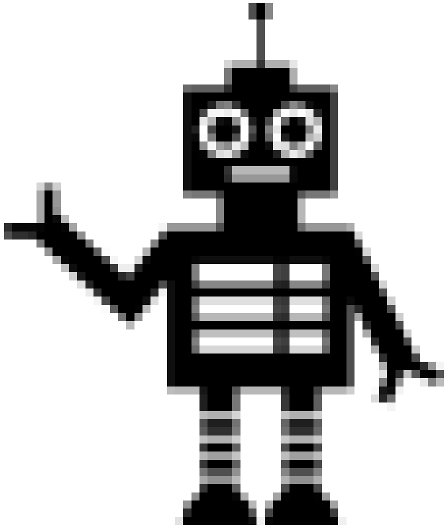
4. Select Models

Objective:	*Select the machine learning models that will be trained, training settings and the model optimisation search strategy.*
Elements:	▪The machine learning model (lasso logistic regression, gradient boosting machine, decision tree, random forest, ada boost, neural network, KNN) ▪The hyper-parameter search grid ▪The test/train split (% and split by person or time)
Benefit:	**It would be useful to explore different models for each prediction problem**

Models in PatientLevelPrediction

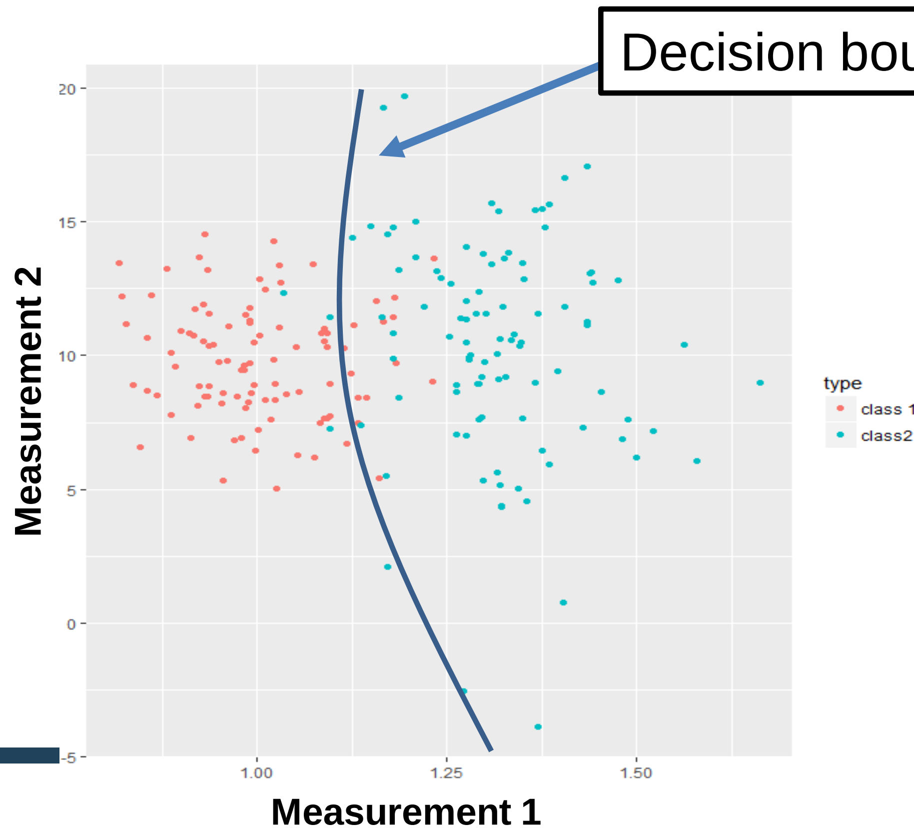
Model
Lasso Logistic Regression
Gradient Boosting Machine
Random Forest
Adaboost
Decision Tree
Neural Network
K-nearest neighbours
Naïve Bayes

Generalised linear models, boosting, bagging, non-parametric, tree based…



Decision Boundary

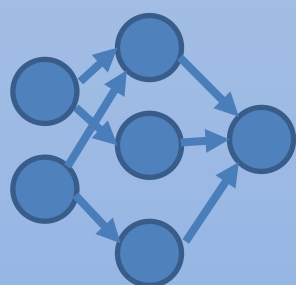
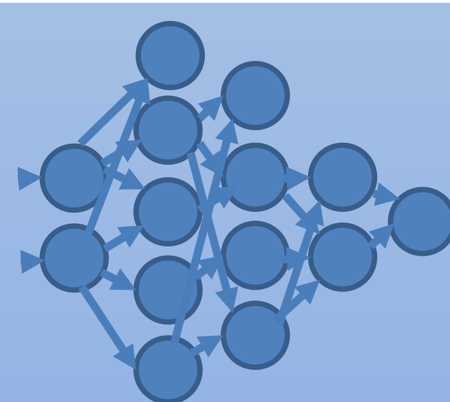
- The aim of a model is to learn a function of the inputs that partitions the classes (the decision boundary)



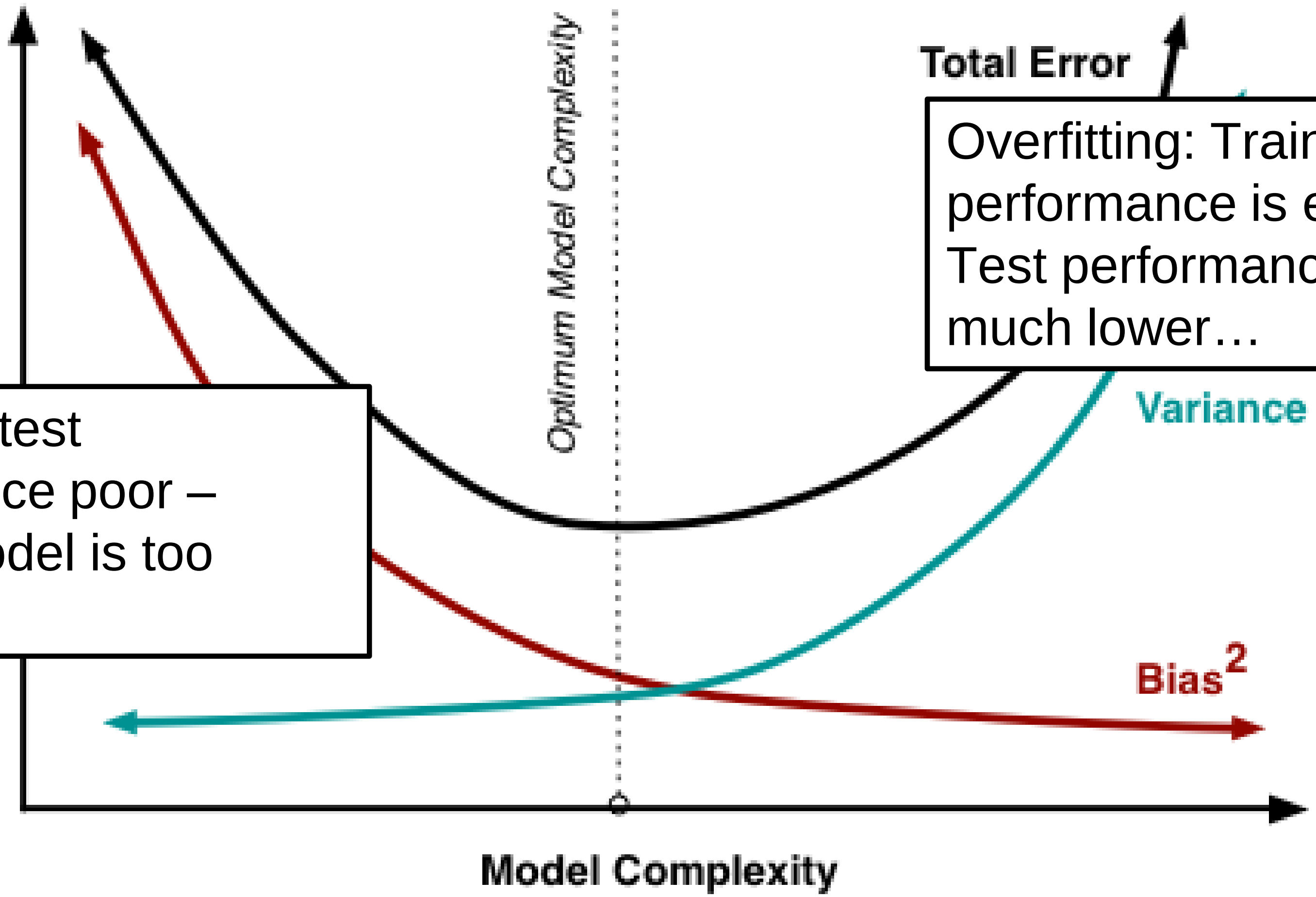
If you had two measurements and wanted to predict the class based on these values then you want to find a function that partitions the blues from the reds

What are hyper-parameters?

- They control the complexity of a model
- E.g., if we wanted to fit a neural network the topology of the network defines the complexity of the model (few layers and a small number of nodes = more simple)

Simple Model	...	Complex Model
	… Complexity →	
High bias/low variance (unlikely to overfit but may not be able to model complexities…)	…	High variance/low bias (prone to overfitting…)

What are hyper-parameters?

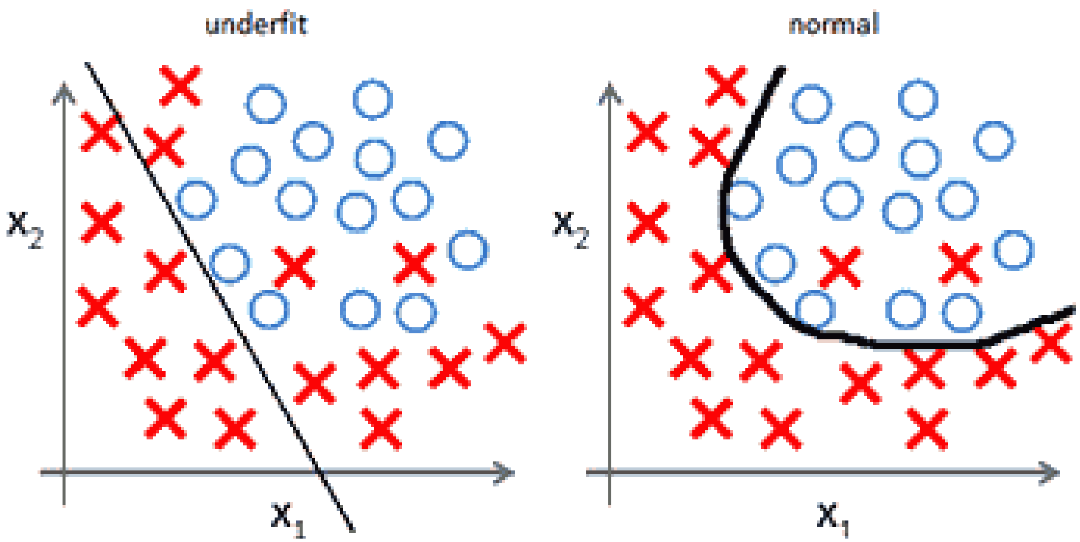


Overfitting: Train performance is excellent, Test performance is much lower…

Train and test performance poor – maybe model is too simple…

Overfitting

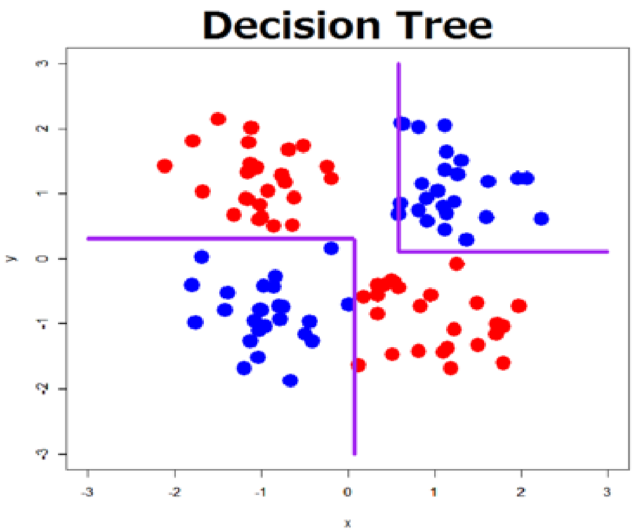
Just right…



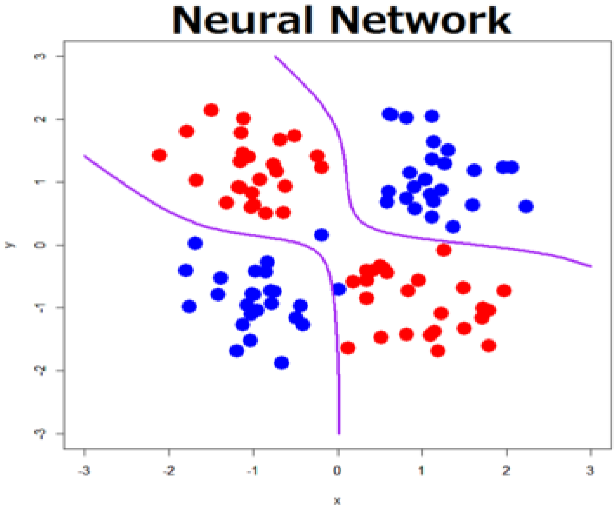
Too simple for this data

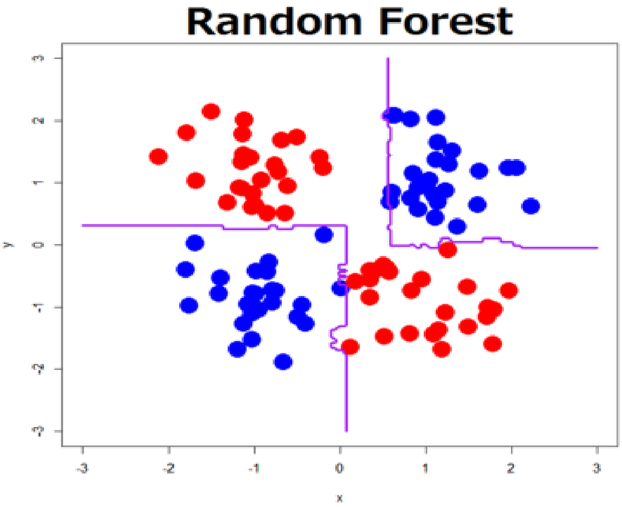
Too complex for this data

Decision boundaries differ



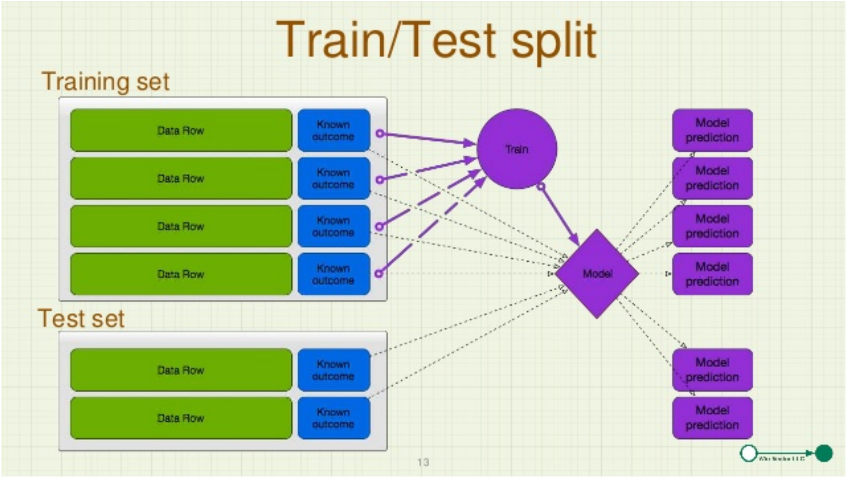
Different machine learning algorithms have different ways to create the decision boundary – no algorithm is always the best, so we recommend as a best practice to implement all the standard algorithms and pick the best one for the specific prediction problem





5. Validate

Generate and validate each model internally and externally



Model training

Internal validation

Externally validate model across OHDSI network

The models can be readily shared across the OHDSI network for validation

Question Break

Any questions about the framework, models or validation?

Today's Agenda

Time	Topic
8:45 – 9:00	Welcome, get settled, get laptops ready
9:00 – 10:30	**Presentation:** What is Patient-Level Prediction?
10:30 – 10:45	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 1 Theory
10:45 – 11:45	**Presentation:** Overview of the TRIPOD Statement **Exercise:** Applying TRIPOD to CHADS2
11:45 – 12:30	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 2 Implementation
12:30 – 13:15	Lunch
13:15 – 14:30	**Exercise:** Guided tour through implementing patient-level prediction
14:30 – 14:45	Break
14:45 – 16:30	**Exercise:** Design and implement your own patient-level prediction
16:30 – 17:00	Lessons learned and feedback

Transparent **R**eporting of a multivariable prediction model for **I**ndividual **P**rognosis **O**r **D**iagnosis (TRIPOD)

Joel Swerdel

Janssen Research and Development

Agenda

- Basics of good reporting for prediction models
- Review of the TRIPOD Statement
- Small group discussion of sample paper
- Large group summary of small group findings

Basics of good reporting for prediction models

- Allows clinicians to decide whether the model is applicable and useful for their patients
- Applicable
  - Was the model developed with patients similar to theirs?
  - Is the data used to inform the model available to them?
- Useful
  - Is the outcome useful to the clinical decisions that need to be made?
  - Can this model be trusted when making a clinical decision?

Basics (cont.)

- Reliable
  - Does the prediction from this model provide high enough sensitivity and specificity?
  - Are the limits of the model well understood?
- Reproducible
  - Are there enough details in the report to reproduce the model?
  - Are there enough details in the report to validate the model?
- **Most models reported in the literature do not provide enough information for model assessment by the reader**

The TRIPOD Statement

- TRIPOD - **T**ransparent **R**eporting of a multivariable prediction model for **I**ndividual **P**rognosis **O**r **D**iagnosis
- Analogous to STROBE (**ST**rengthening the Reporting of **OB**servational studies in **E**pidemiology)
- Developed through the cooperative effort of a 25+ member committee of prediction modeling experts
- Reduced from 76 to 22 items

1 Title

- Concise summary of the model.
- Example: "**Development and validation** of a clinical score to **estimate the probability of coronary artery disease** in **men and women presenting with suspected coronary disease**"

3 Background and Objectives

- What was the goal for developing this model?
- Example: "The aim of this study was to **develop and validate** a **clinical prediction rule** in **women presenting with breast symptoms**, so that a **more evidence based approach to referral**—which would include urgent referral under the 2 week rule—could be implemented as part of clinical practice guidance."

4 Methods - Source of Data

- Gives an indication of both applicability and quality of the data
- Example: “The population based sample used for this report included **2489 men and 2856 women 30 to 74 years old** at the time of their **Framingham Heart Study** examination in 1971 to 1974. Participants attended either the 11th examination of the original Framingham cohort or the initial examination of the Framingham Offspring Study. Similar research protocols were used in each study, and **persons with overt coronary heart disease at the baseline examination were excluded**.”

6 Methods- Outcome

- What was predicted and how was it measured?
- Example: “Breast Cancer Ascertainment: **Incident diagnoses of breast cancer** were **ascertained by self-report on biennial follow up questionnaires** from 1997 to 2005. We learned of deaths from family members, the US Postal Service, and the National Death Index. We identified 1084 incident breast cancers, and **1007 (93%) were confirmed by medical record or by cancer registry data** from 24 states in which 96% of participants resided at baseline.”

7 Methods- Predictors

- What was used to inform the model? When was the data collected?
- Example: “The following data were extracted for each patient: **gender, aspartate aminotransferase in IU/L, alanine aminotransferase in IU/L, aspartate** aminotransferase/alanine aminotransferase ratio, total bilirubin (mg/dl), albumin (g/dl), transferrin saturation (%), mean corpuscular volume (μm3), platelet count (× 103/mm3), and prothrombin time(s). . . . **All laboratory tests were performed within 90 days before liver biopsy**. In the case of repeated test, the results closest to the time of the biopsy were used. No data obtained after the biopsy were used.

10 Methods - Statistics

- What type of model was used and how was performance assessed?
- Example: “We used the **Cox proportional hazards model** in the derivation dataset to estimate the coefficients associated with each potential risk factor [predictor] for the first ever recorded diagnosis of cardiovascular disease for men and women separately.”
- Example: “We assessed the predictive performance of the QRISK2-2011 risk score on the THIN cohort by **examining measures of calibration and discrimination**… **Calibration** of the risk score predictions was assessed by **plotting observed proportions versus predicted probabilities** and by calculating the calibration slope… **Discrimination** … quantified by **calculating the area under the receiver operating characteristic curve statistic**; a value of 0.5 represents chance and 1 represents perfect discrimination.”

15 Results – Model Specification

- What were the predictors and how were they used to inform the final prediction?
- Example:

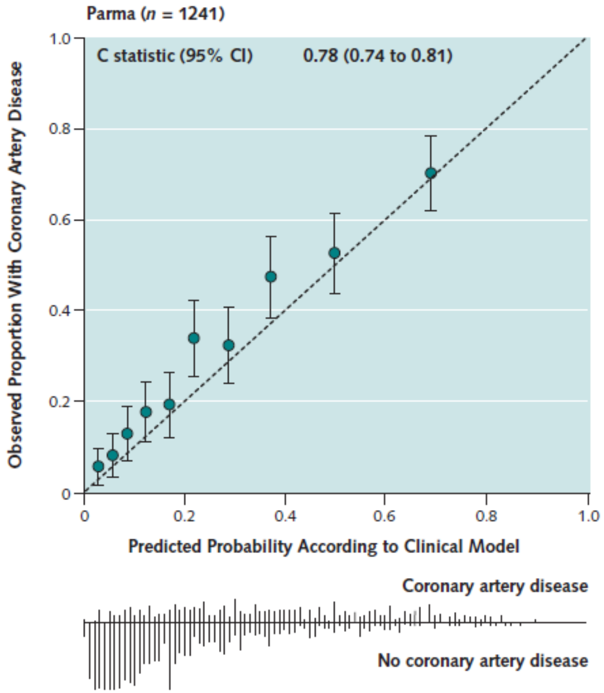
Table 12. Example Table: Presenting the Full Prognostic (Survival) Model, Including the Baseline Survival, for a Specific Time Point*

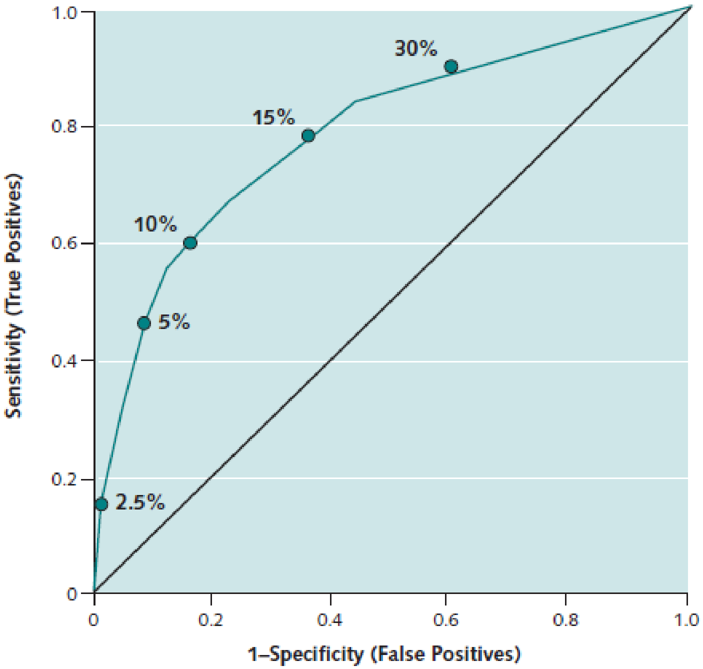
	β Coefficient	SE	*P* Value
Age	0.15052	0.05767	0.009
Age^2	−0.00038	0.00041	0.35
Male sex	1.99406	0.39326	0.0001
Body mass index	0.01930	0.01111	0.08
Systolic blood pressure	0.00615	0.00225	0.006
Treatment for hypertension	0.42410	0.10104	0.0001
PR interval	0.00707	0.00170	0.0001
Significant cardiac murmur	3.79586	1.33532	0.005
Heart failure	9.42833	2.26981	0.0001
Male sex × age^2	−0.00028	0.00008	0.0004
Age × significant murmur	−0.04238	0.01904	0.03
Age × prevalent heart failure	−0.12307	0.03345	0.0002

From reference 402.
* $S_0(10)$ = 0.96337 (10-year baseline survival). β values are expressed per 1-unit increase for continuous variables and for the condition present in dichotomous variables.

16 Results - Performance

- How well did the model perform based on the specified metrics?
- Example:





Small Group Discussion

- Review "Validation of Clinical Classification Schemes for Predicting Stroke Results From the National Registry of Atrial Fibrillation" Gage et al.
- Group assignment for filling in the TRIPOD table
- Grade each item:
  - –A: completely fulfills the requirement
  - –C: partially fulfills the requirement
  - –F: does not fulfill the requirement
- Take about 20 minutes

Today's Agenda

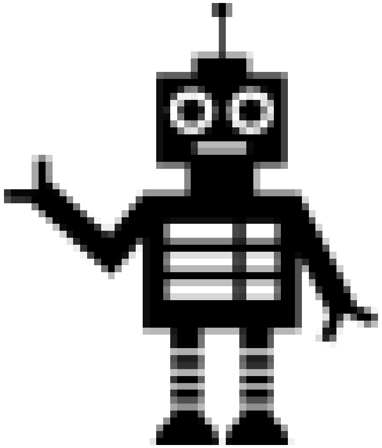
Time	Topic
8:45 – 9:00	Welcome, get settled, get laptops ready
9:00 – 10:30	**Presentation:** What is Patient-Level Prediction?
10:30 – 10:45	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 1 Theory
10:45 – 11:45	**Presentation:** Overview of the TRIPOD Statement **Exercise:** Applying TRIPOD to CHADS2
11:45 – 12:30	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 2 Implementation
12:30 – 13:15	Lunch
13:15 – 14:30	**Exercise:** Guided tour through implementing patient-level prediction
14:30 – 14:45	Break
14:45 – 16:30	**Exercise:** Design and implement your own patient-level prediction
16:30 – 17:00	Lessons learned and feedback

Learning Goals

1. Our 5-step Framework
2. What goes into PatientLevelPrediction
3. Implement various machine learning techniques
4. What comes out of PatientLevelPrediction
5. Interpretation of model performance metrics and plots

Example Task

Objective:	*Reproduce CHADS2 model using the PatientLevelPrediction package…*
Specify Problem:	*Prediction Problem:* In **PLP training: T : patients newly diagnosed with Atrial fibrillation** predict who will develop **PLP training: O - hospitalized ischemic stroke events** during the period from 0 days from cohort start date to 1000 days.
Predictors:	*Predictors:* We will use the 5 variables that are used in the CHAD2 model



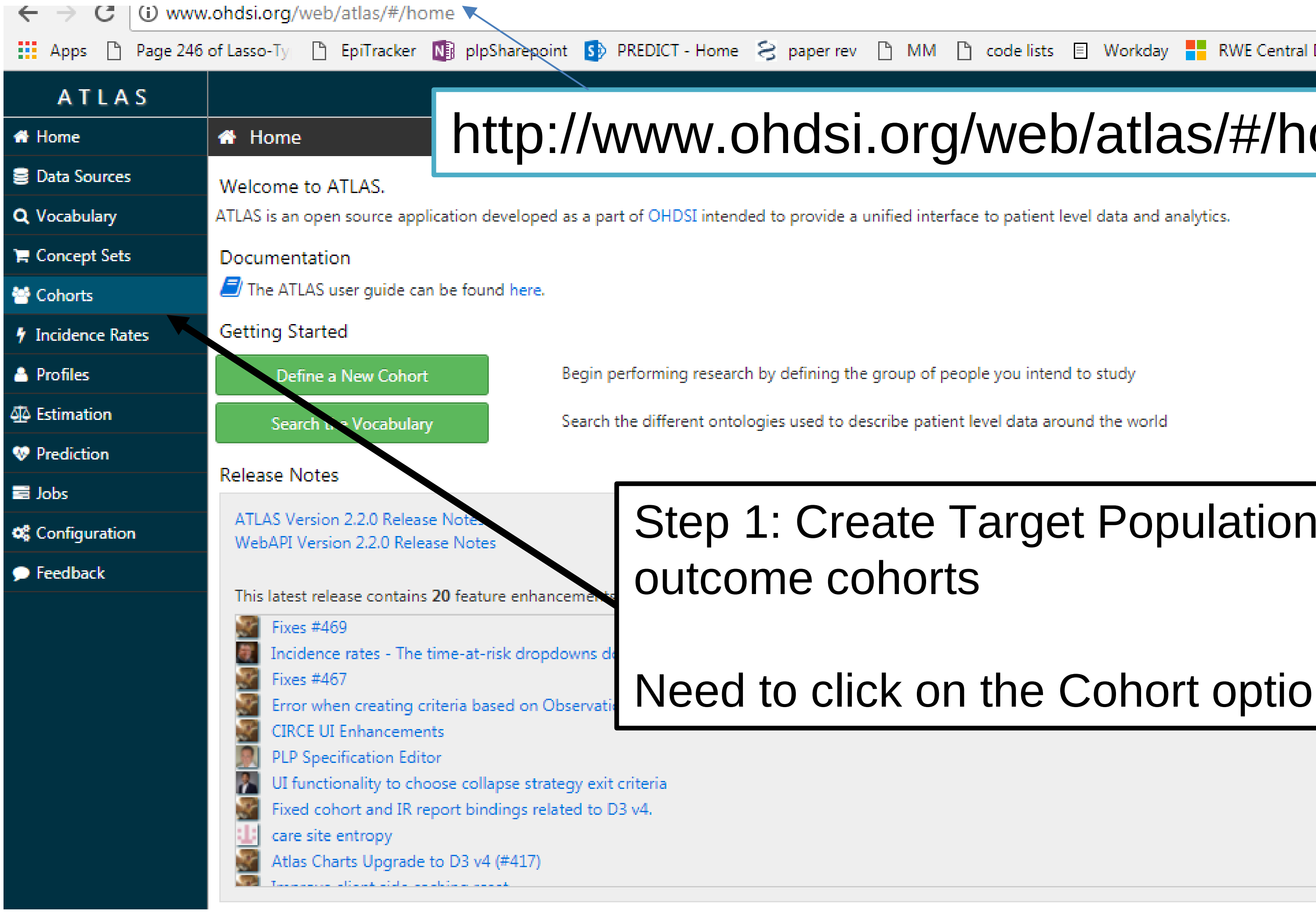
Two Options

Use Atlas form to generate R code

The next few slides will cover the Atlas form options, then we will describe what each R function does

f

Atlas - cohorts

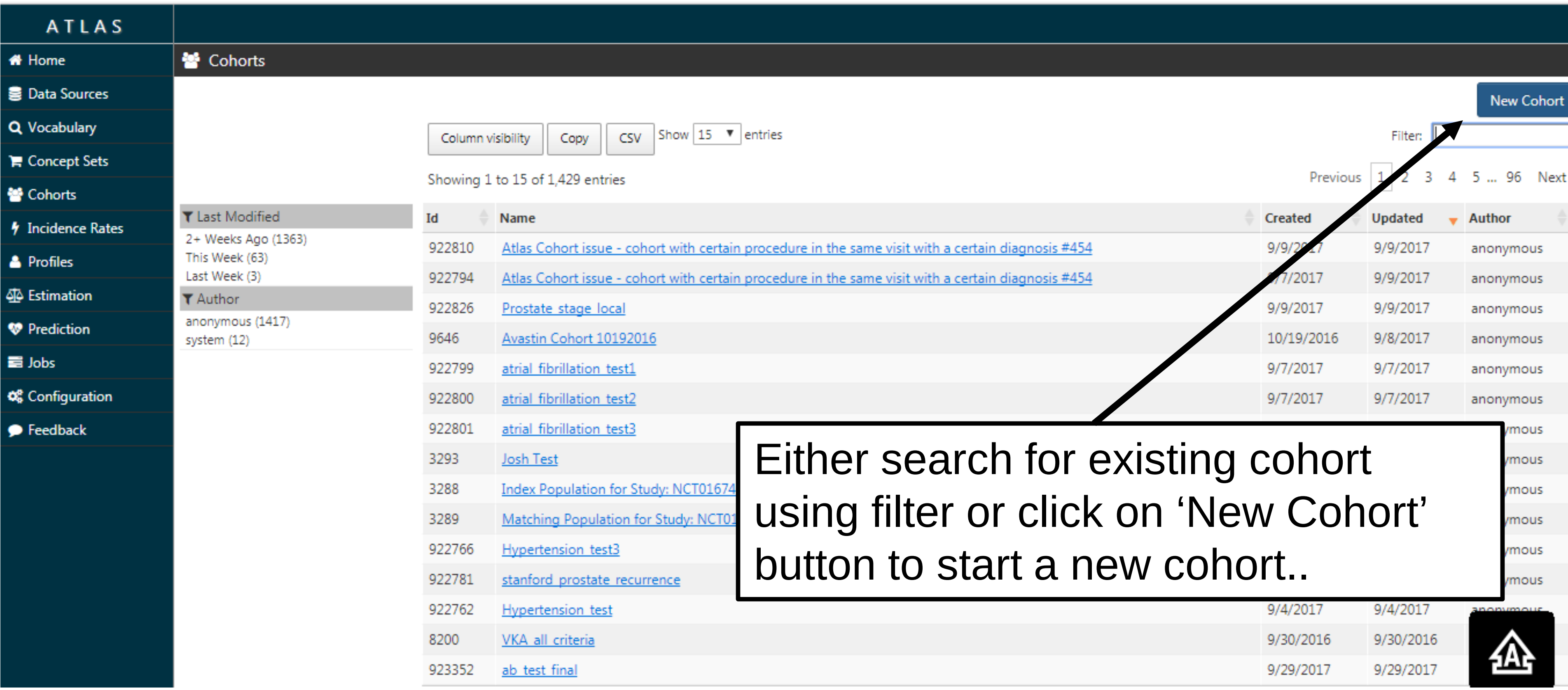


http://www.ohdsi.org/web/atlas/#/home

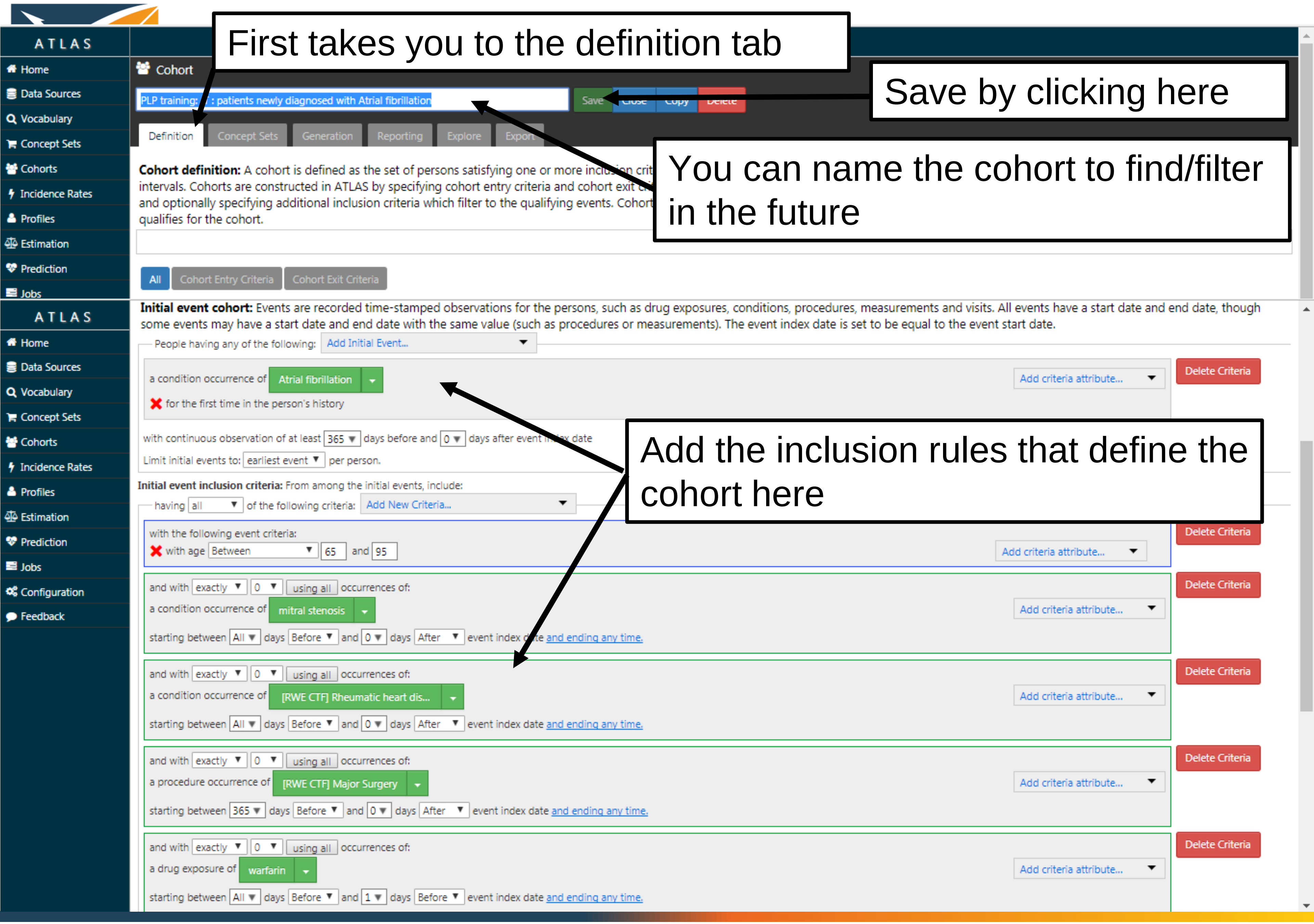
Step 1: Create Target Population and outcome cohorts

Need to click on the Cohort option

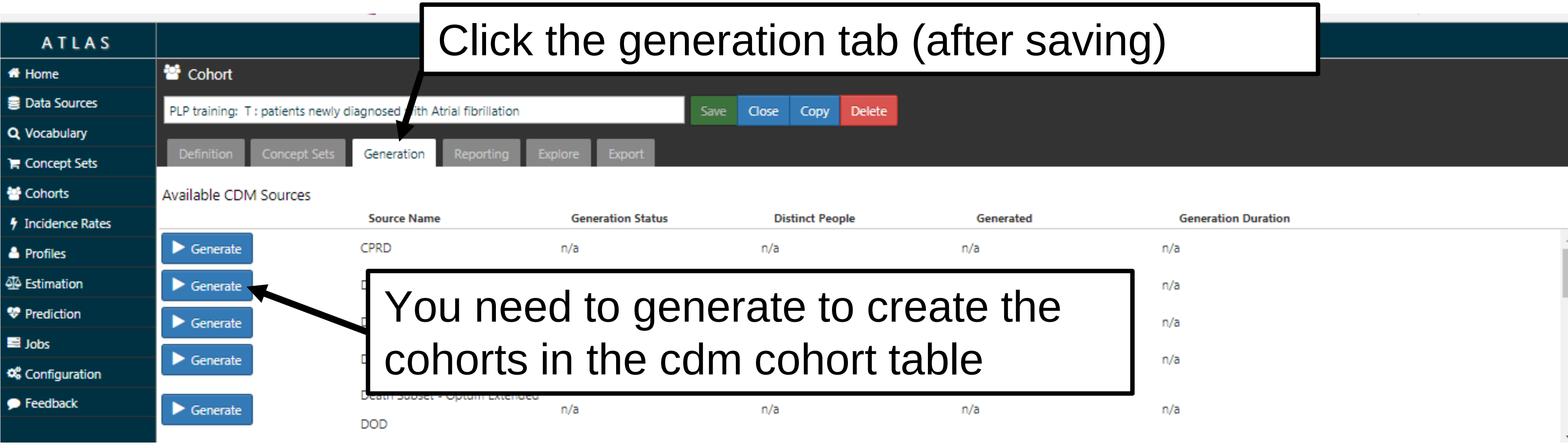
Atlas - cohorts



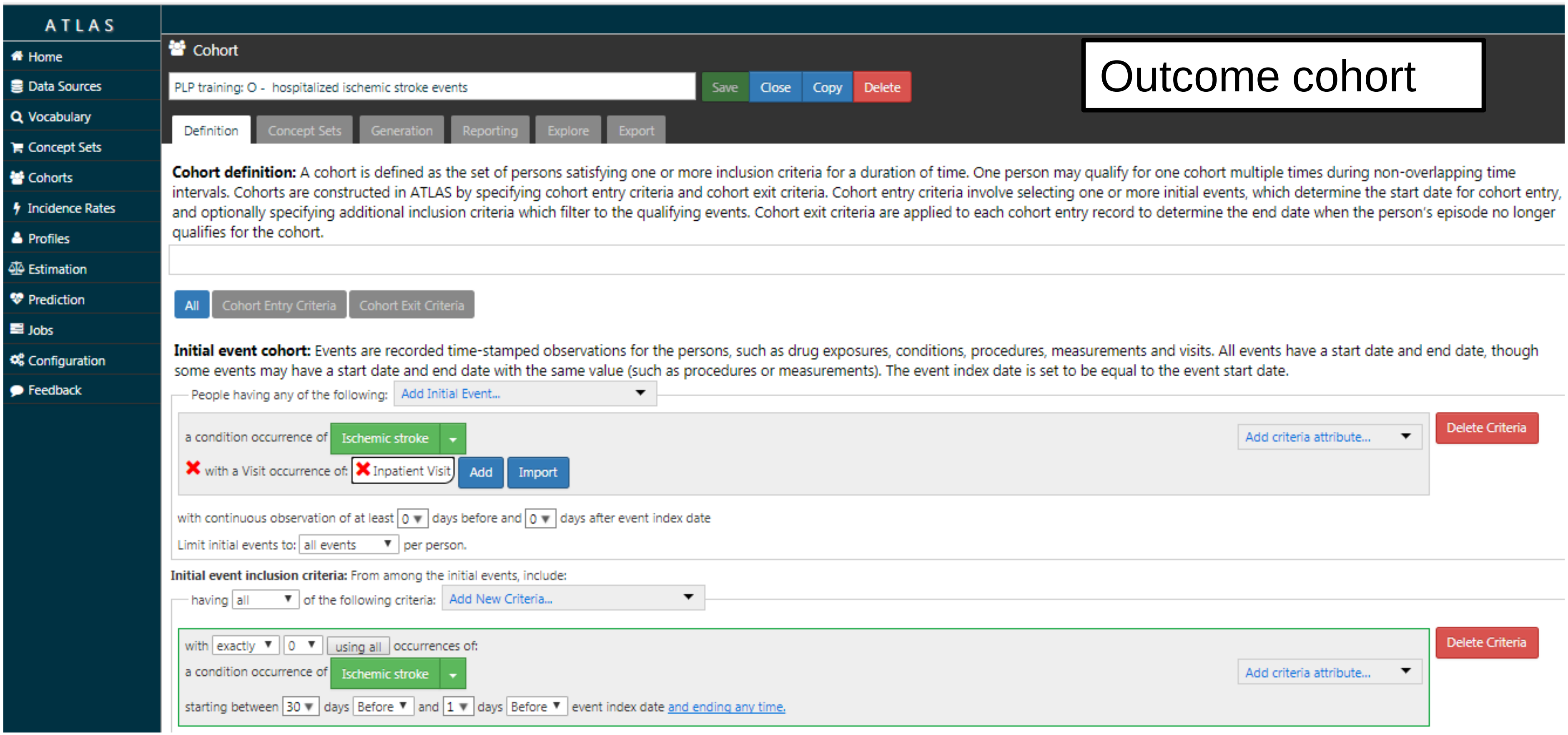
Either search for existing cohort using filter or click on 'New Cohort' button to start a new cohort..


First takes you to the definition tab
Save by clicking here
You can name the cohort to find/filter in the future
Add the inclusion rules that define the cohort here
ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Incidence Rates
Profiles
Estimation
Prediction
Jobs
Configuration
Feedback
Cohort
PLP training: : patients newly diagnosed with Atrial fibrillation
Save
Definition
Concept Sets
Generation
Reporting
Explore
Export
Cohort definition: A cohort is defined as the set of persons satisfying one or more inclusion crit
intervals. Cohorts are constructed in ATLAS by specifying cohort entry criteria and cohort exit cr
and optionally specifying additional inclusion criteria which filter to the qualifying events. Cohort
qualifies for the cohort.
All
Cohort Entry Criteria
Cohort Exit Criteria
Initial event cohort: Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits. All events have a start date and end date, though some events may have a start date and end date with the same value (such as procedures or measurements). The event index date is set to be equal to the event start date.
People having any of the following: Add Initial Event...
a condition occurrence of Atrial fibrillation
for the first time in the person's history
Add criteria attribute...
Delete Criteria
with continuous observation of at least 365 days before and 0 days after event index date
Limit initial events to: earliest event per person.
Initial event inclusion criteria: From among the initial events, include:
having all of the following criteria: Add New Criteria...
with the following event criteria:
with age Between 65 and 95
and with exactly 0 using all occurrences of:
a condition occurrence of mitral stenosis
starting between All days Before and 0 days After event index date and ending any time.
a condition occurrence of [RWE CTF] Rheumatic heart dis...
a procedure occurrence of [RWE CTF] Major Surgery
starting between 365 days Before and 0 days After event index date and ending any time.
a drug exposure of warfarin
starting between All days Before and 1 days Before event index date and ending any time.

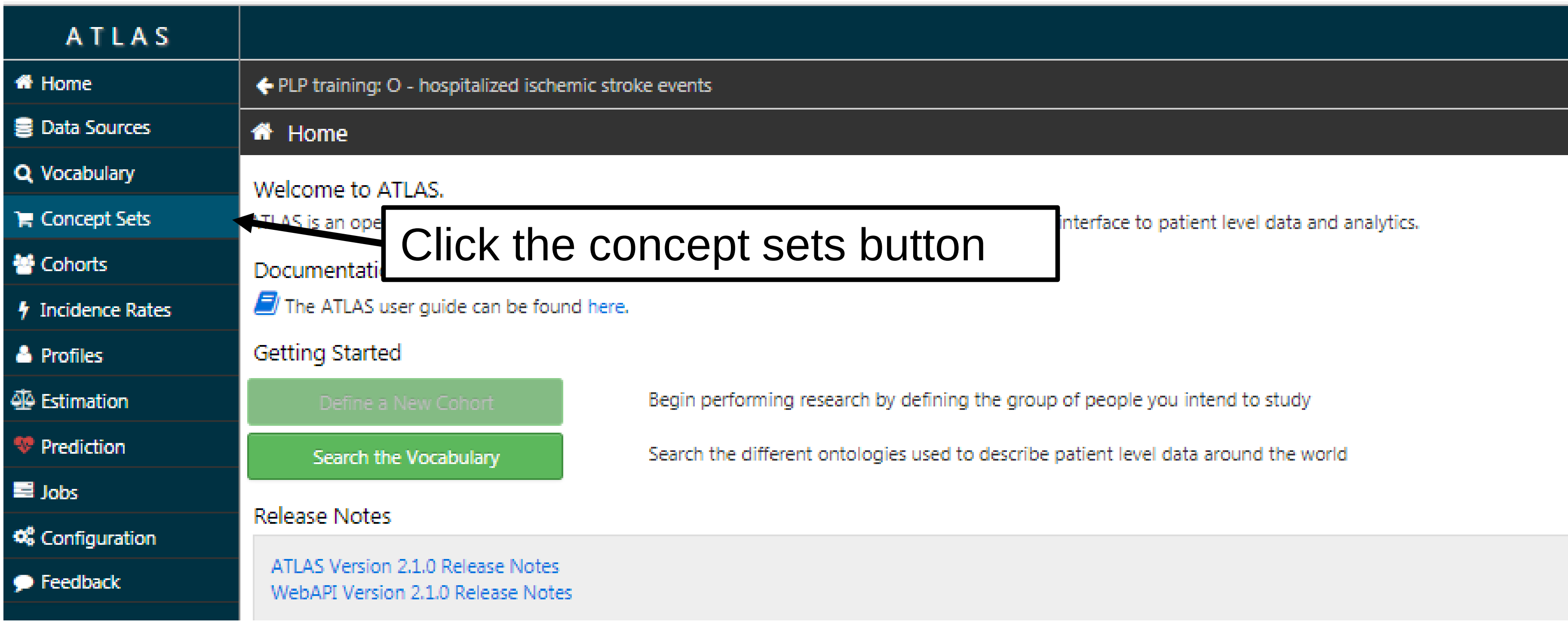
Atlas - cohorts



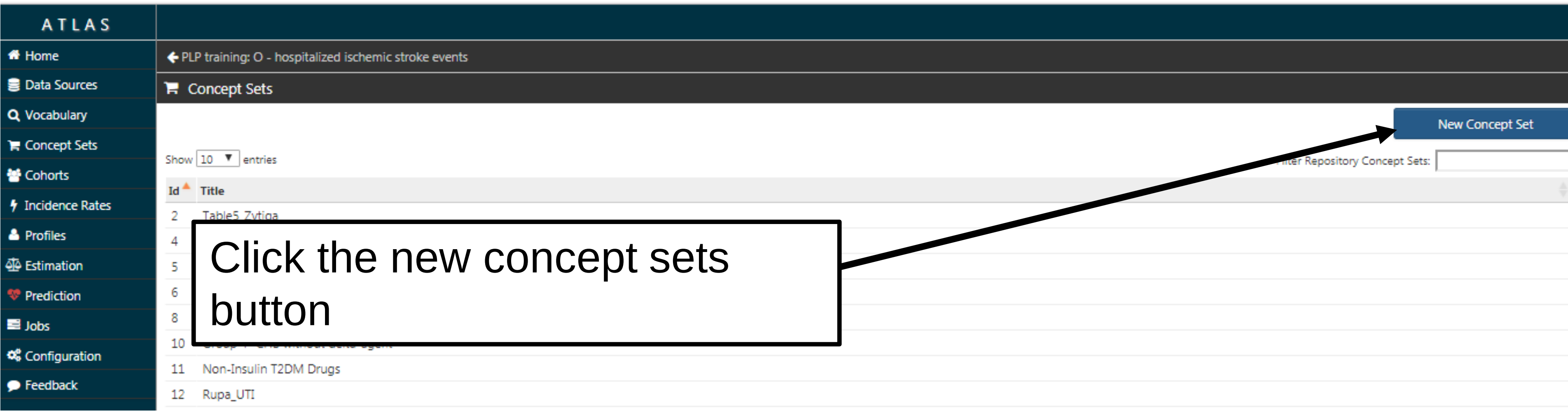
Atlas - cohorts



Atlas – concept sets

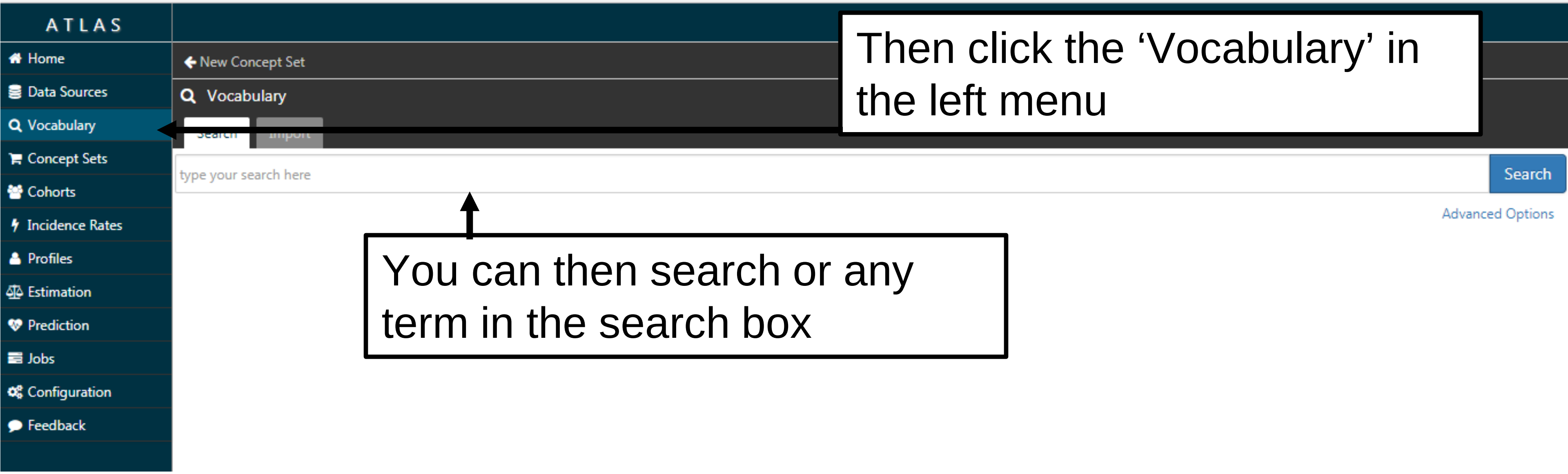


Atlas – concept sets

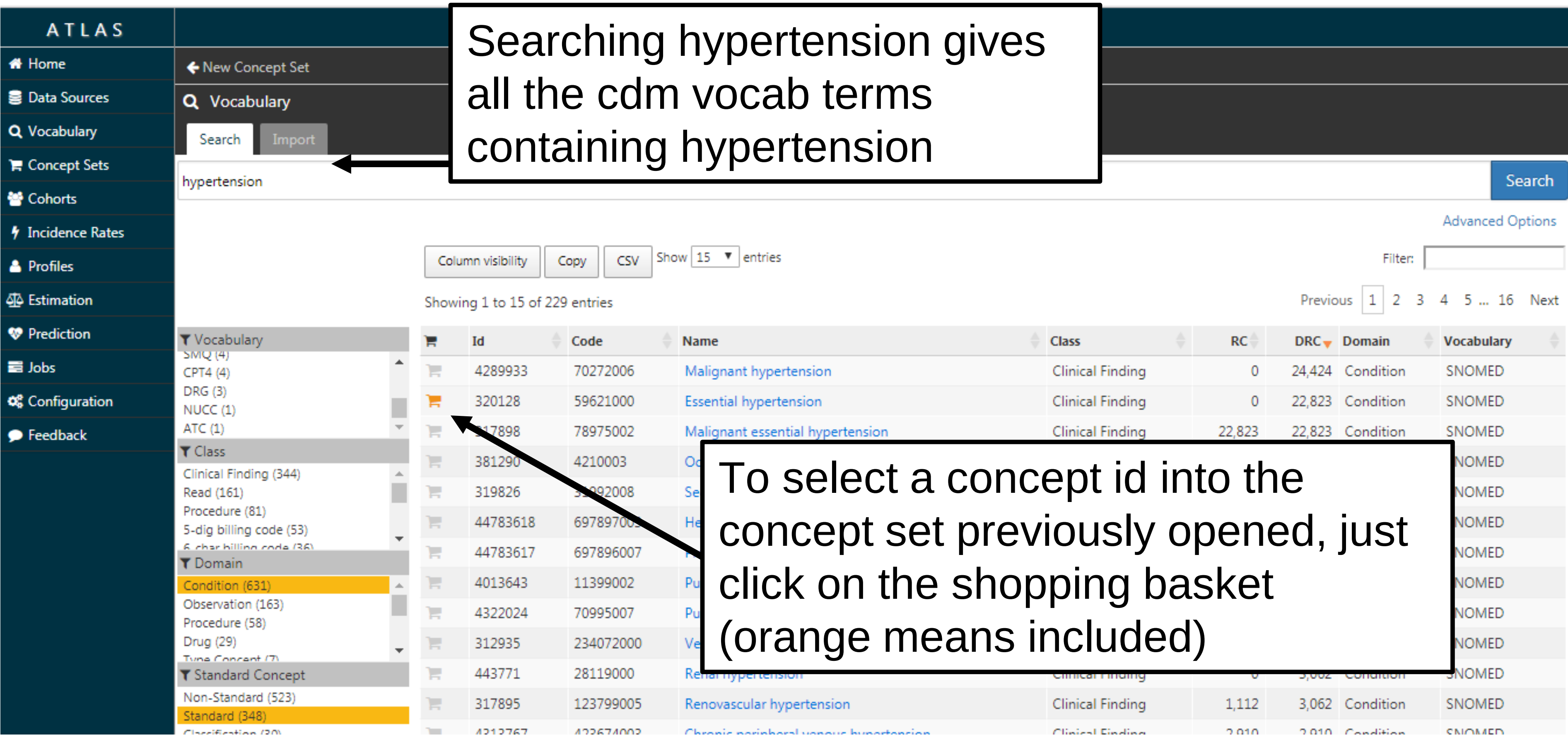


Click the new concept sets button

Atlas – concept sets



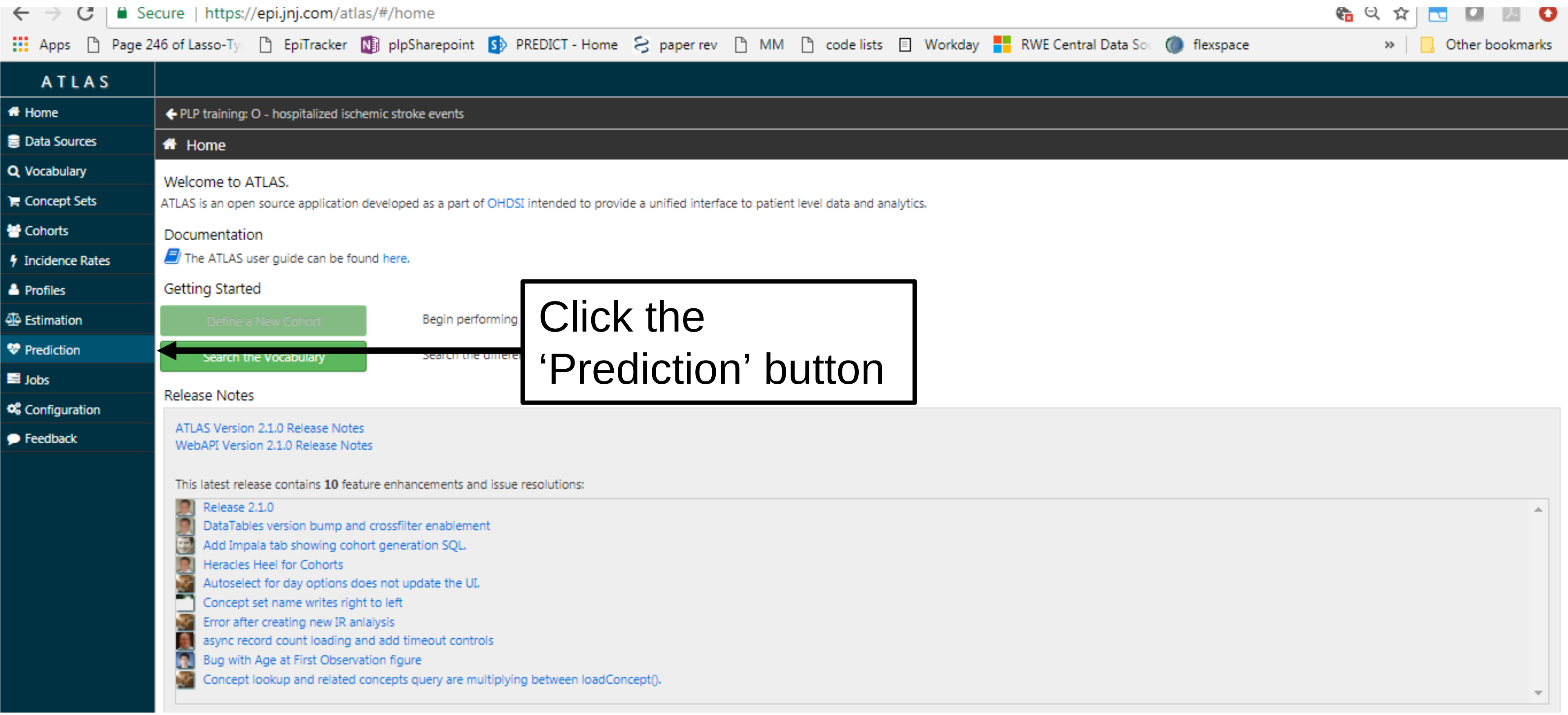
Atlas – concept sets



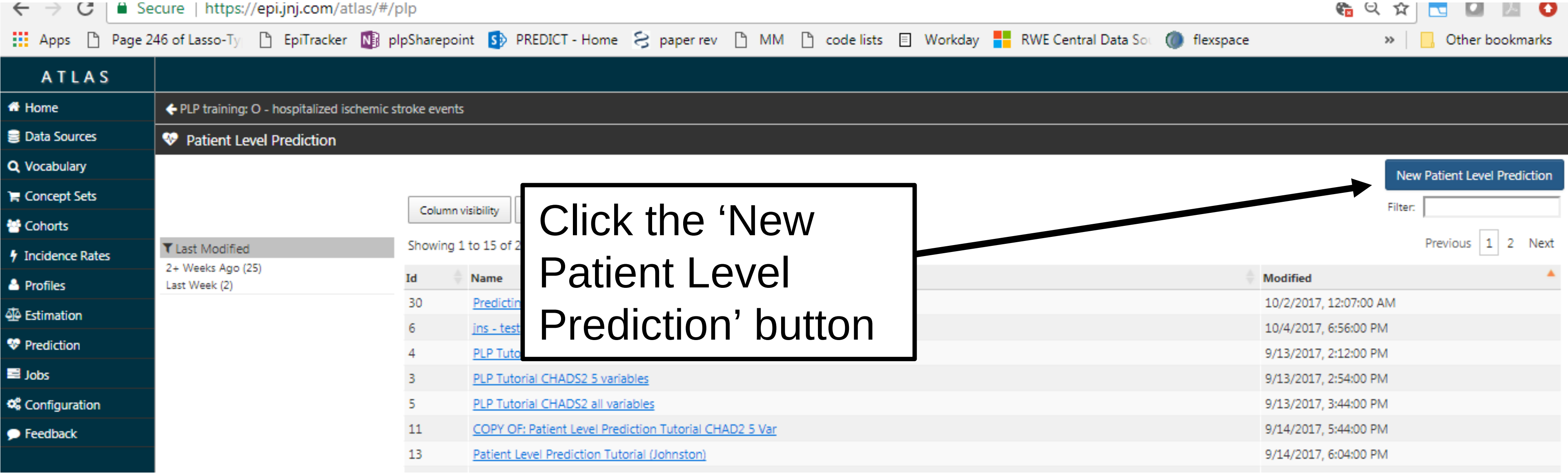
Searching hypertension gives all the cdm vocab terms containing hypertension

To select a concept id into the concept set previously opened, just click on the shopping basket (orange means included)

Atlas - prediction

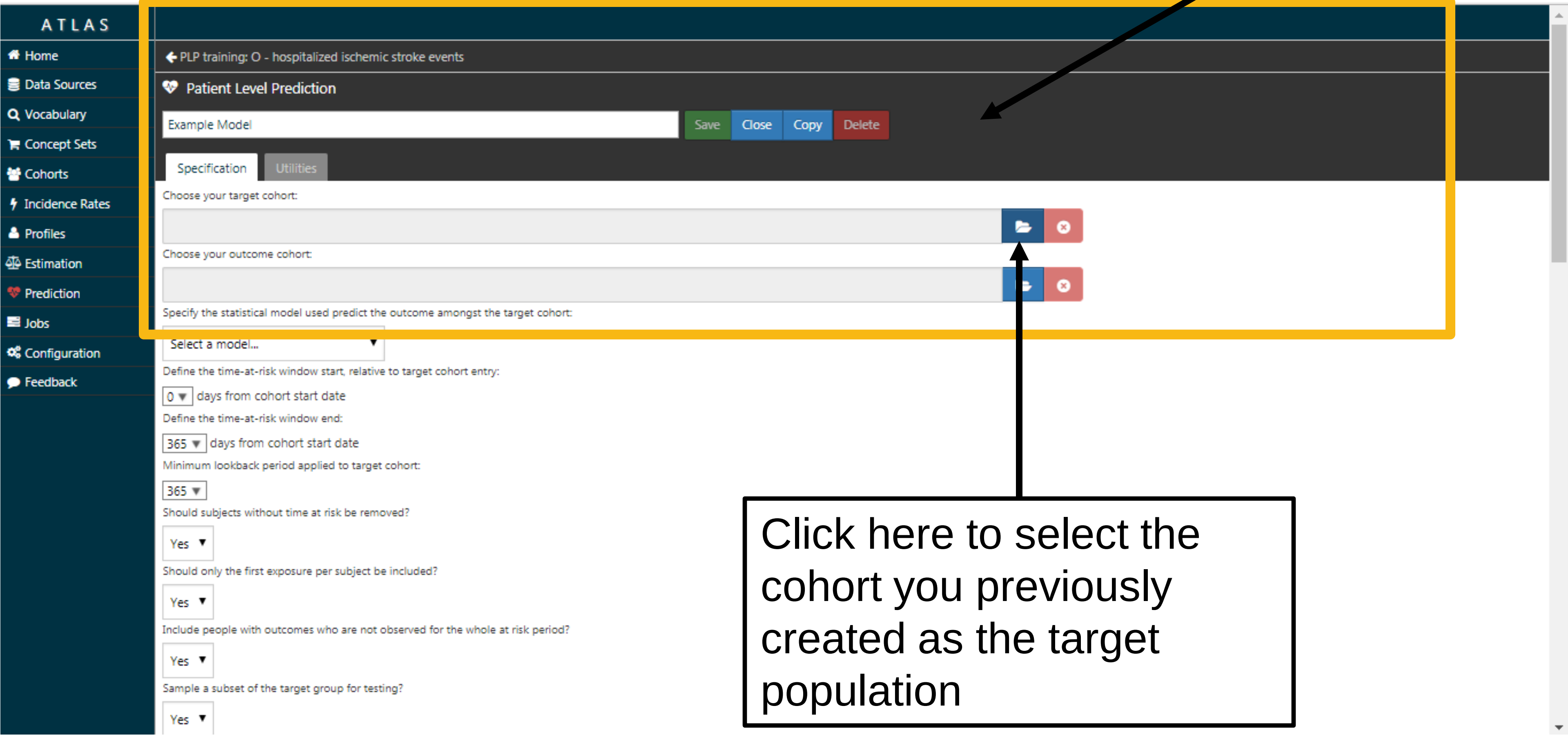


Atlas - prediction



Atlas - predi

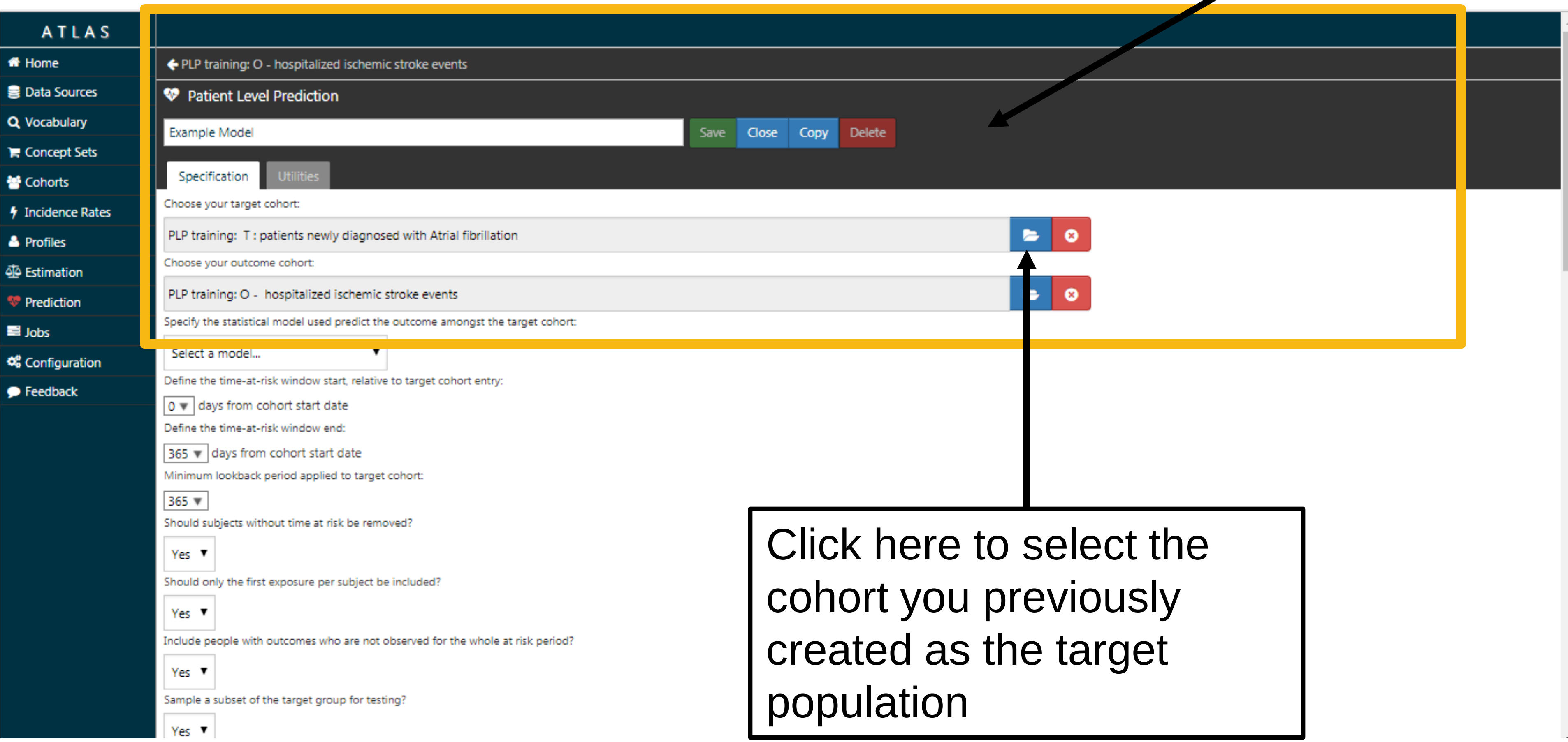
Name the prediction and select the target population/outcome



Click here to select the cohort you previously created as the target population

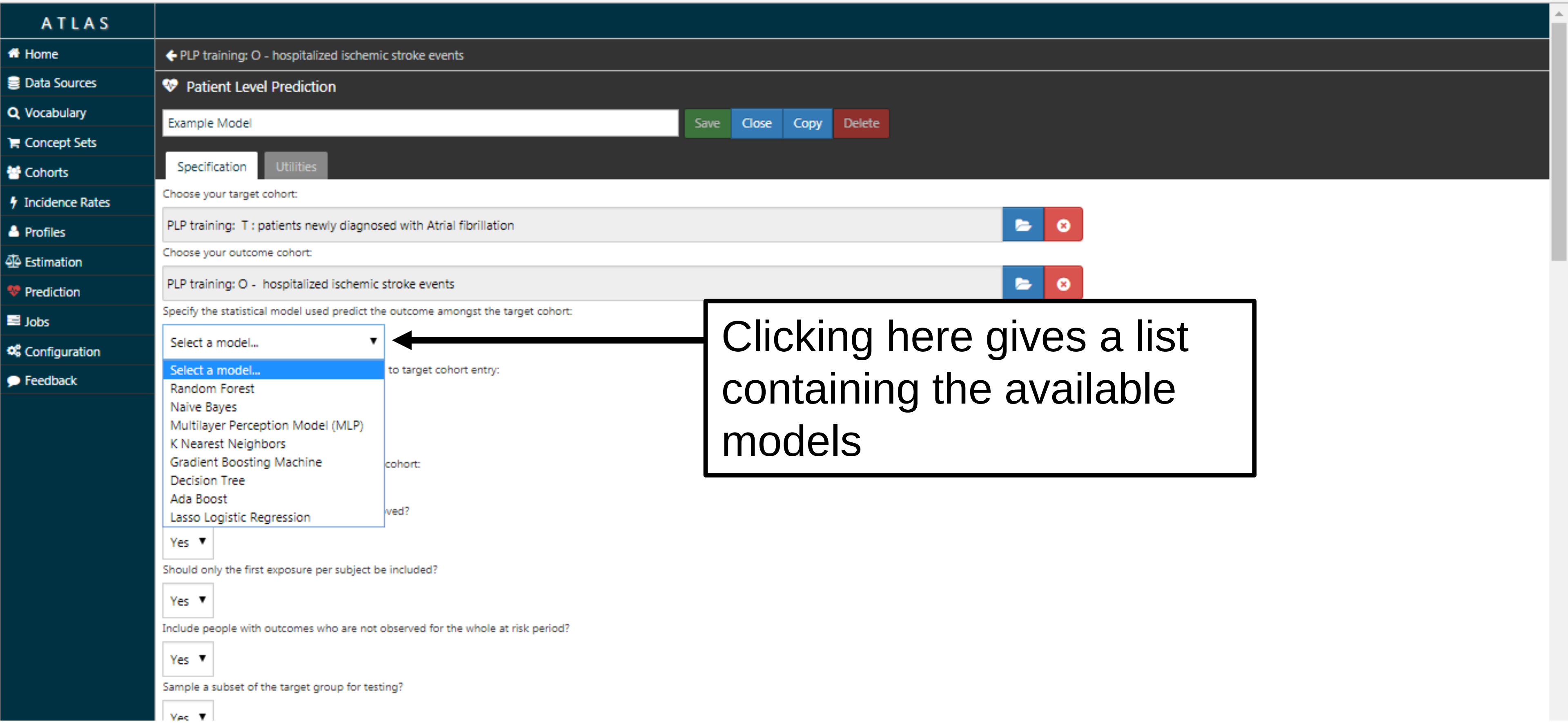
Atlas - predi

Name the prediction and select the target population/outcome

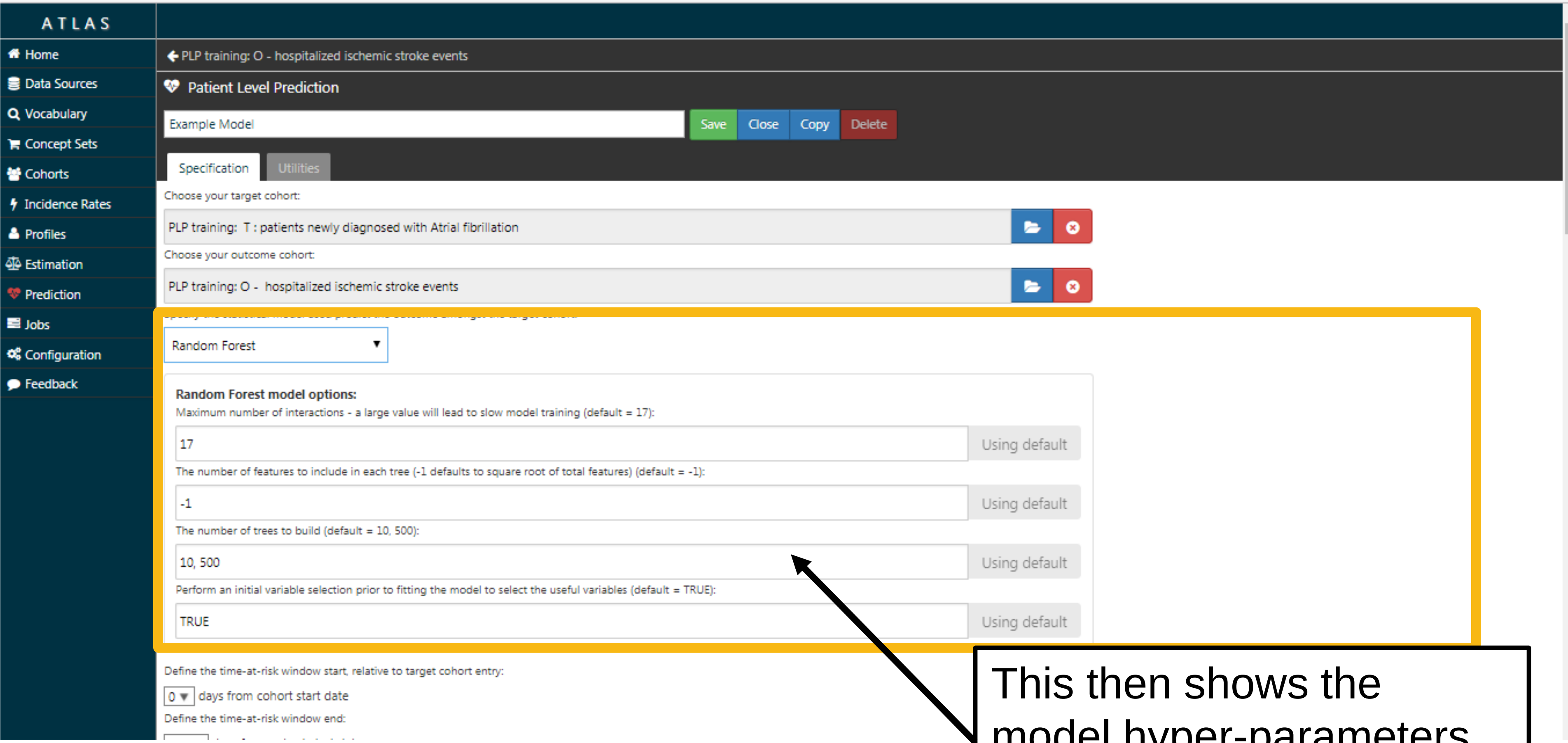


Click here to select the cohort you previously created as the target population

Atlas - prediction

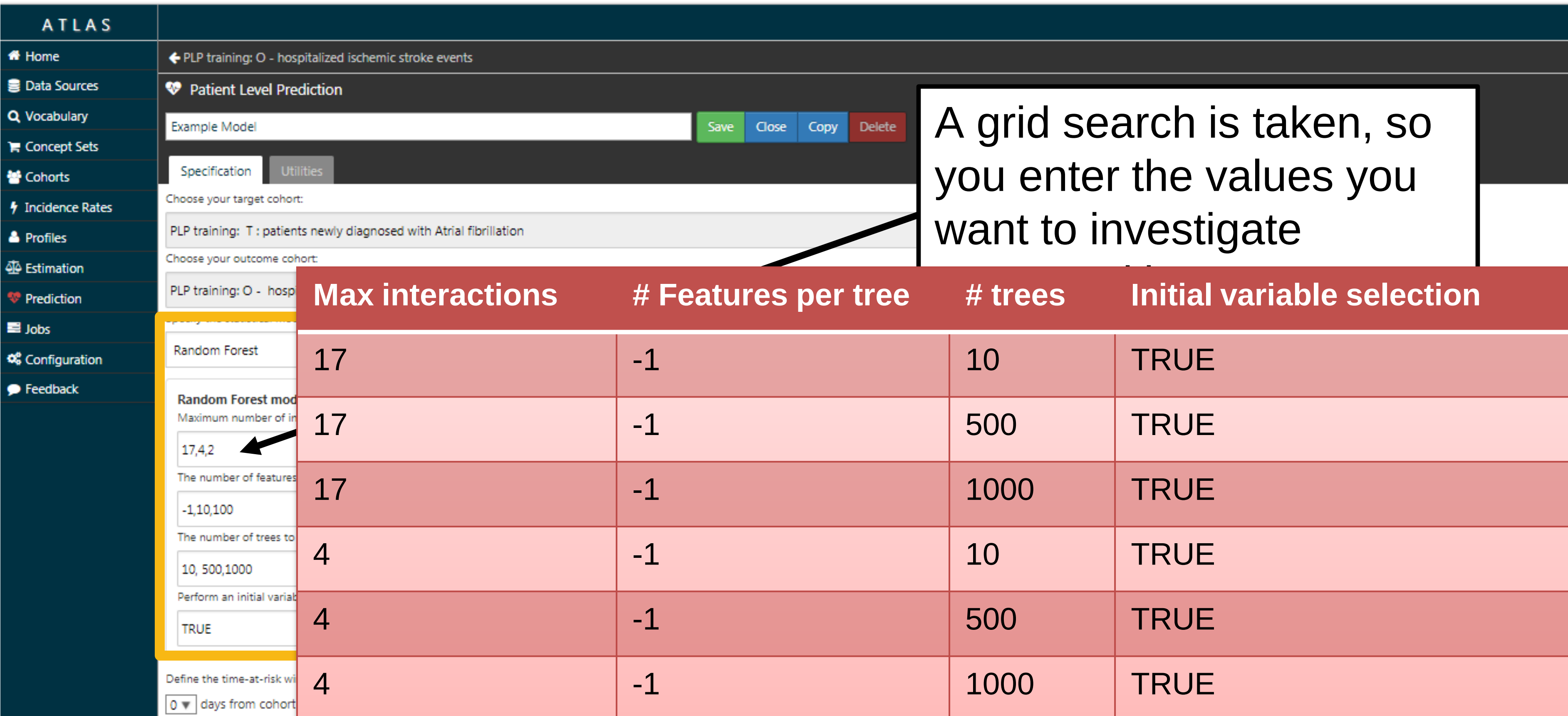


Atlas - prediction



This then shows the model hyper-parameters you can enter

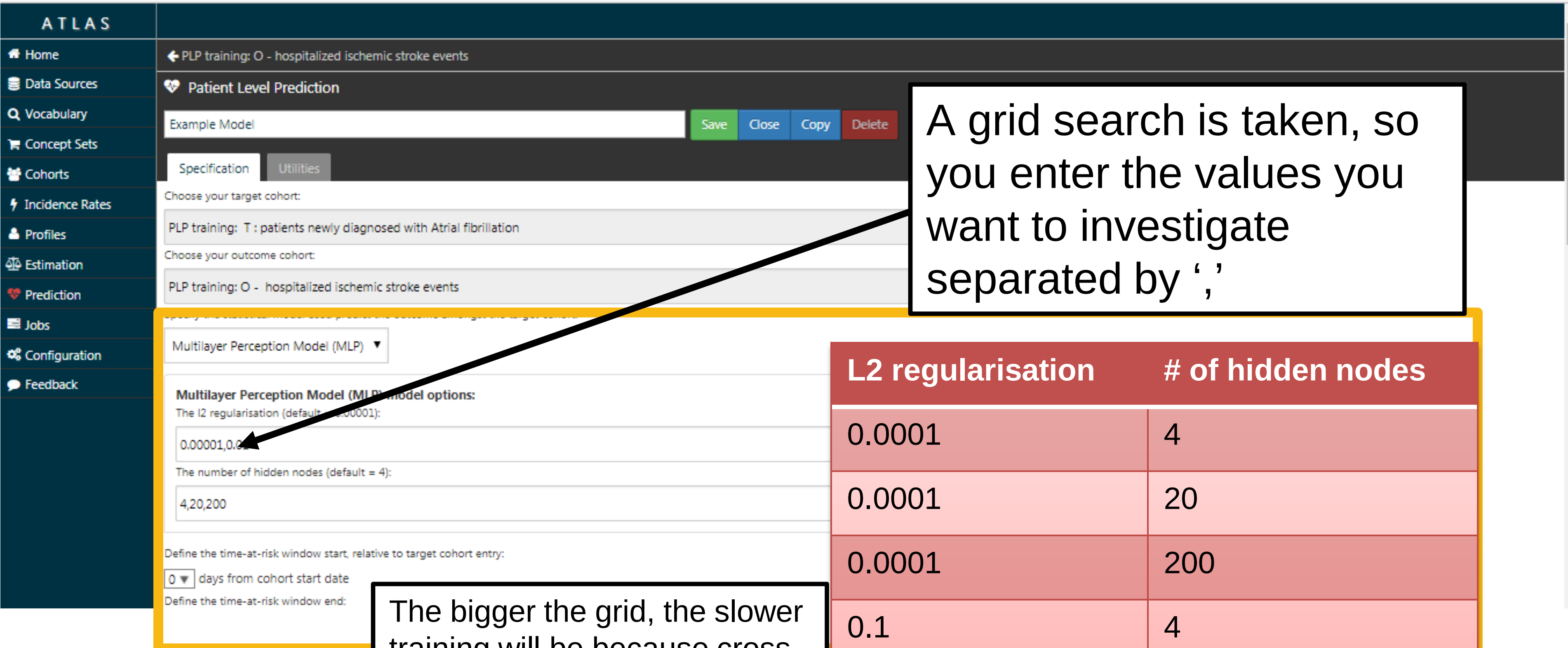
Atlas - prediction



A grid search is taken, so you enter the values you want to investigate

Max interactions	# Features per tree	# trees	Initial variable selection
17	-1	10	TRUE
17	-1	500	TRUE
17	-1	1000	TRUE
4	-1	10	TRUE
4	-1	500	TRUE
4	-1	1000	TRUE
2	-1	10	TRUE
2	-1	500	TRUE
2	-1	1000	TRUE

Atlas - prediction



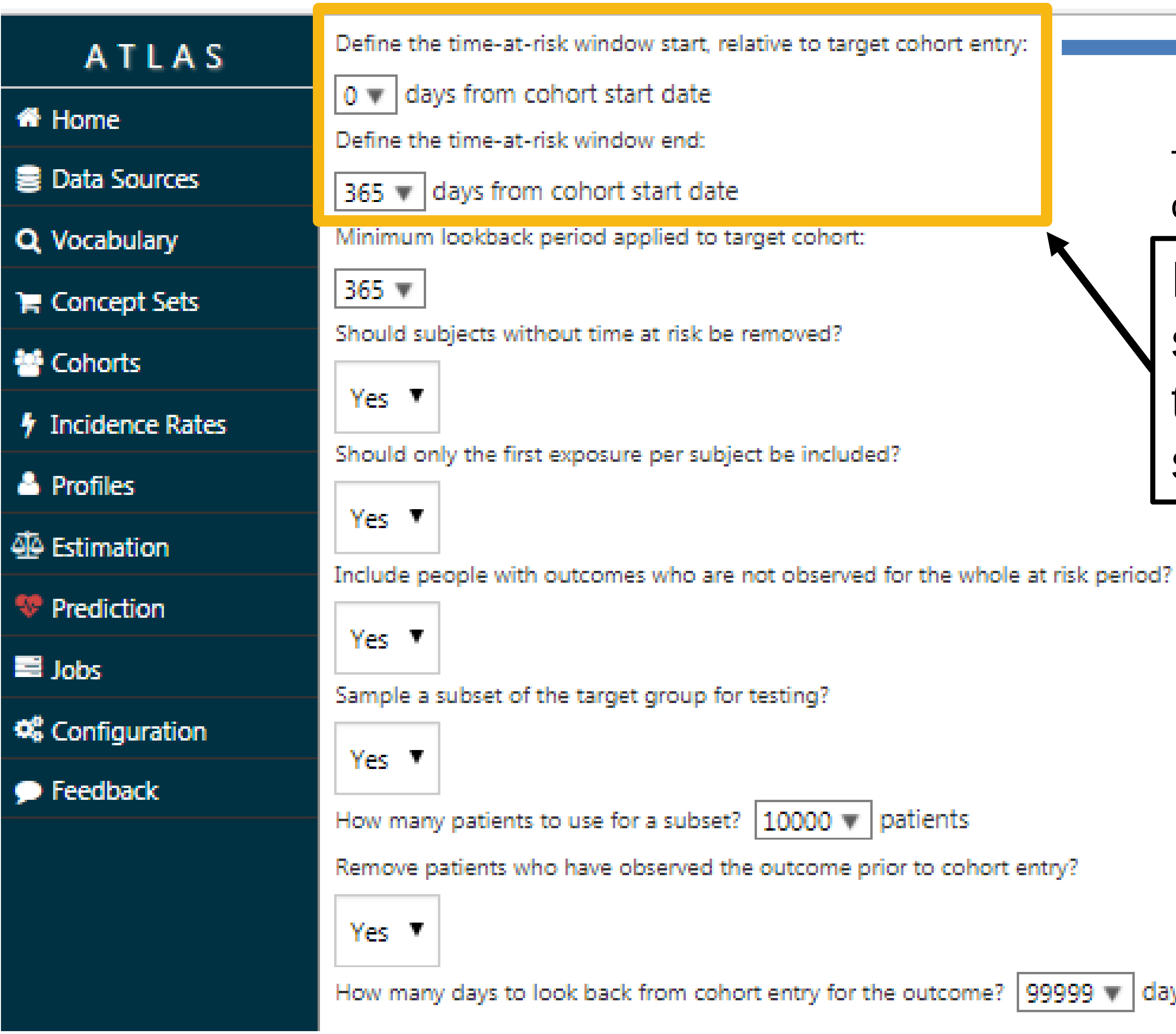
A grid search is taken, so you enter the values you want to investigate separated by ','

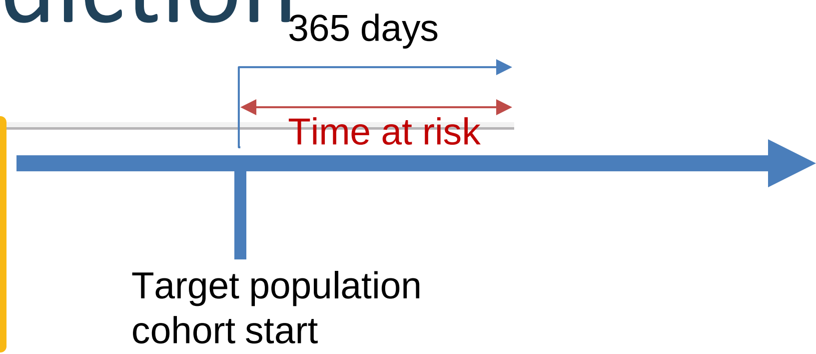
L2 regularisation	# of hidden nodes
0.0001	4
0.0001	20
0.0001	200
0.1	4
0.1	20
0.1	200

The bigger the grid, the slower training will be because cross validation on the test set will be run for each combination

as - prediction

Next define the problem specification elements:

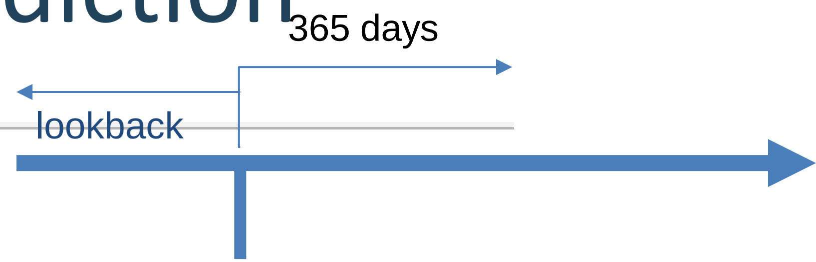


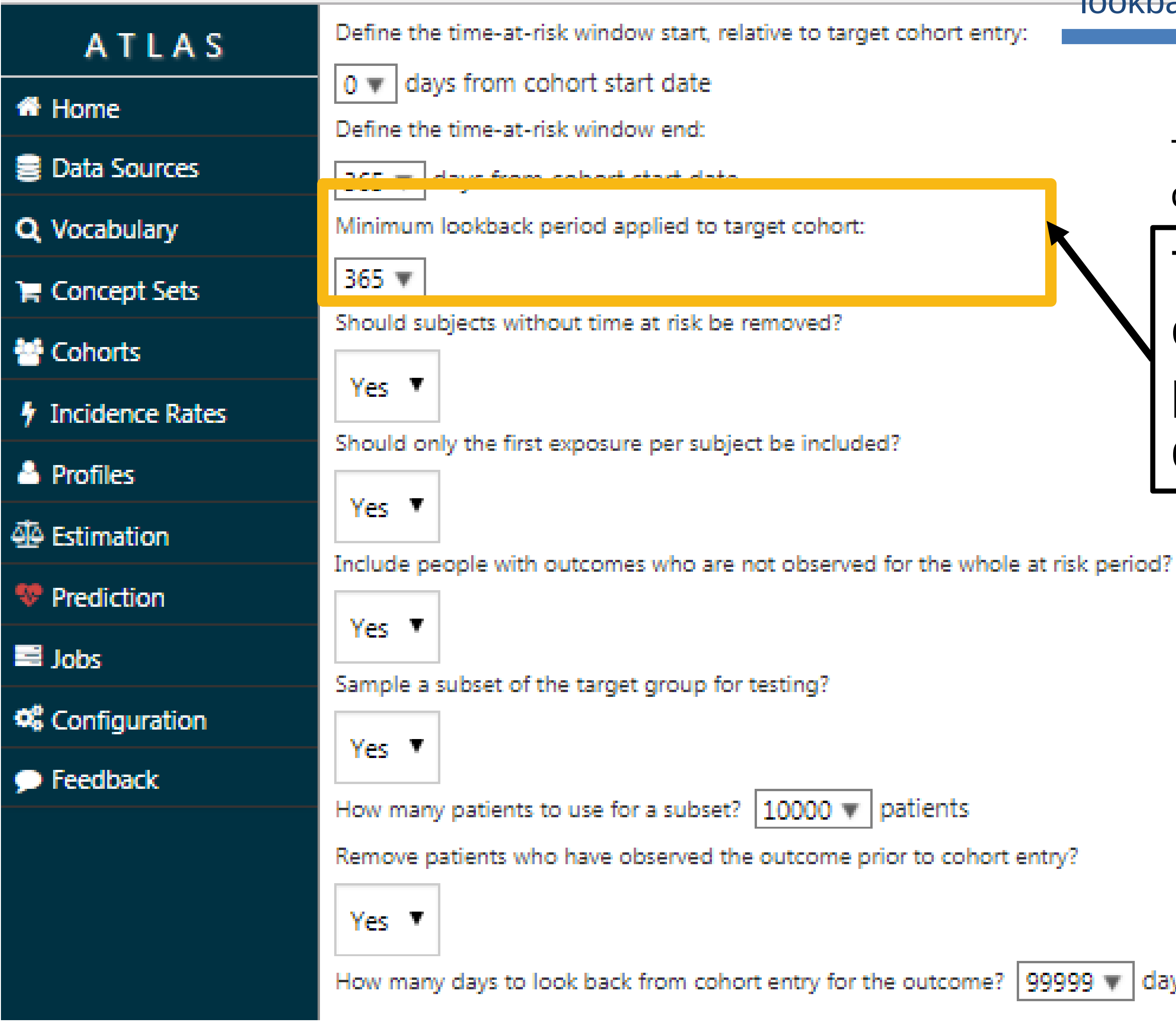


Pick the time-at-risk settings (relative to target population cohort start date)

Next define the problem specification elements:

as - prediction

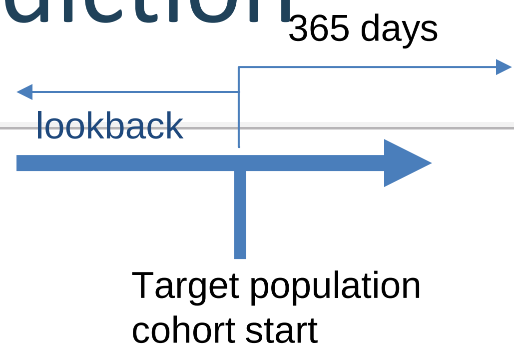


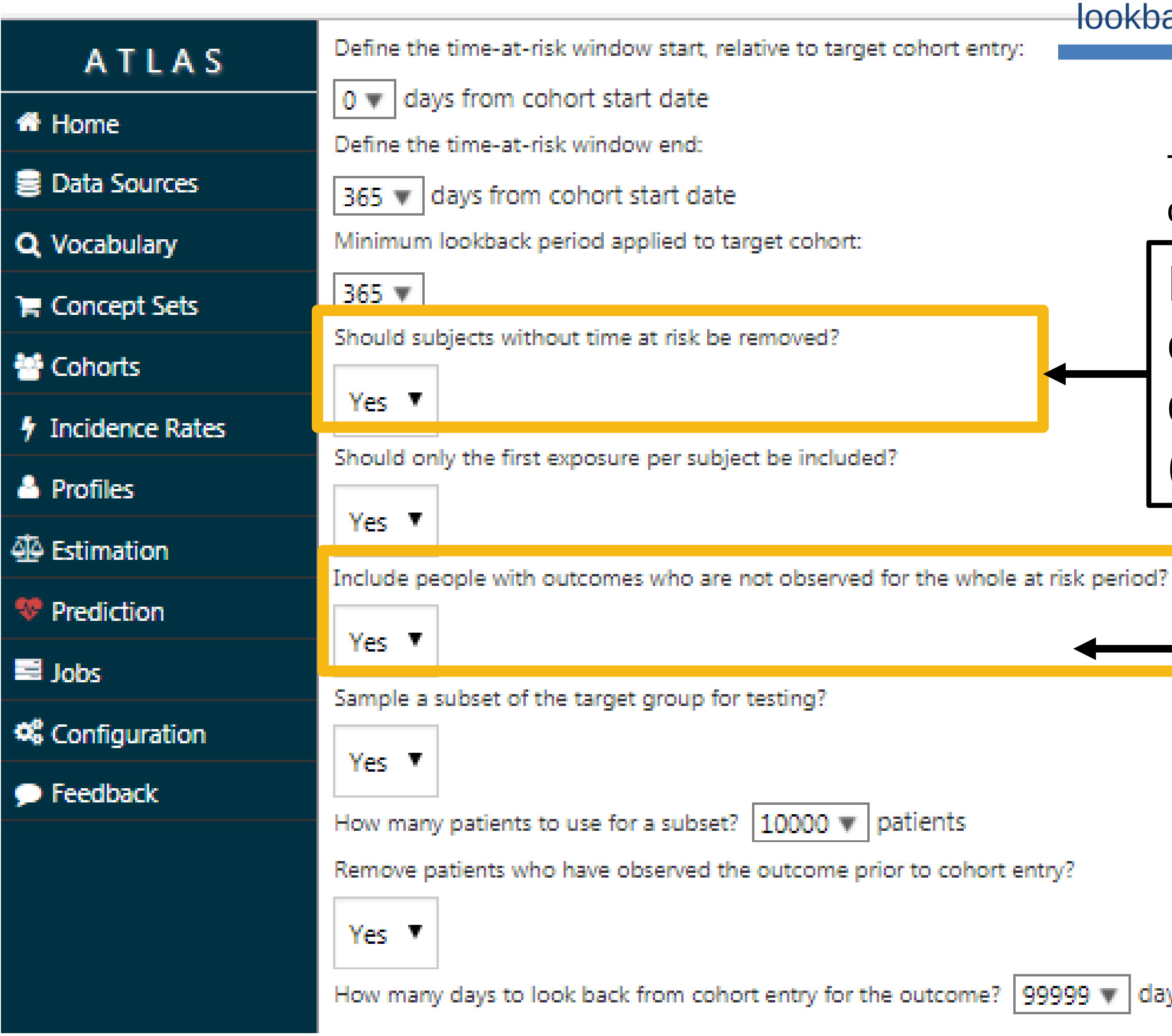


The minimum number of days prior to target population cohort start date to be included

as - prediction

Next define the problem specification elements:



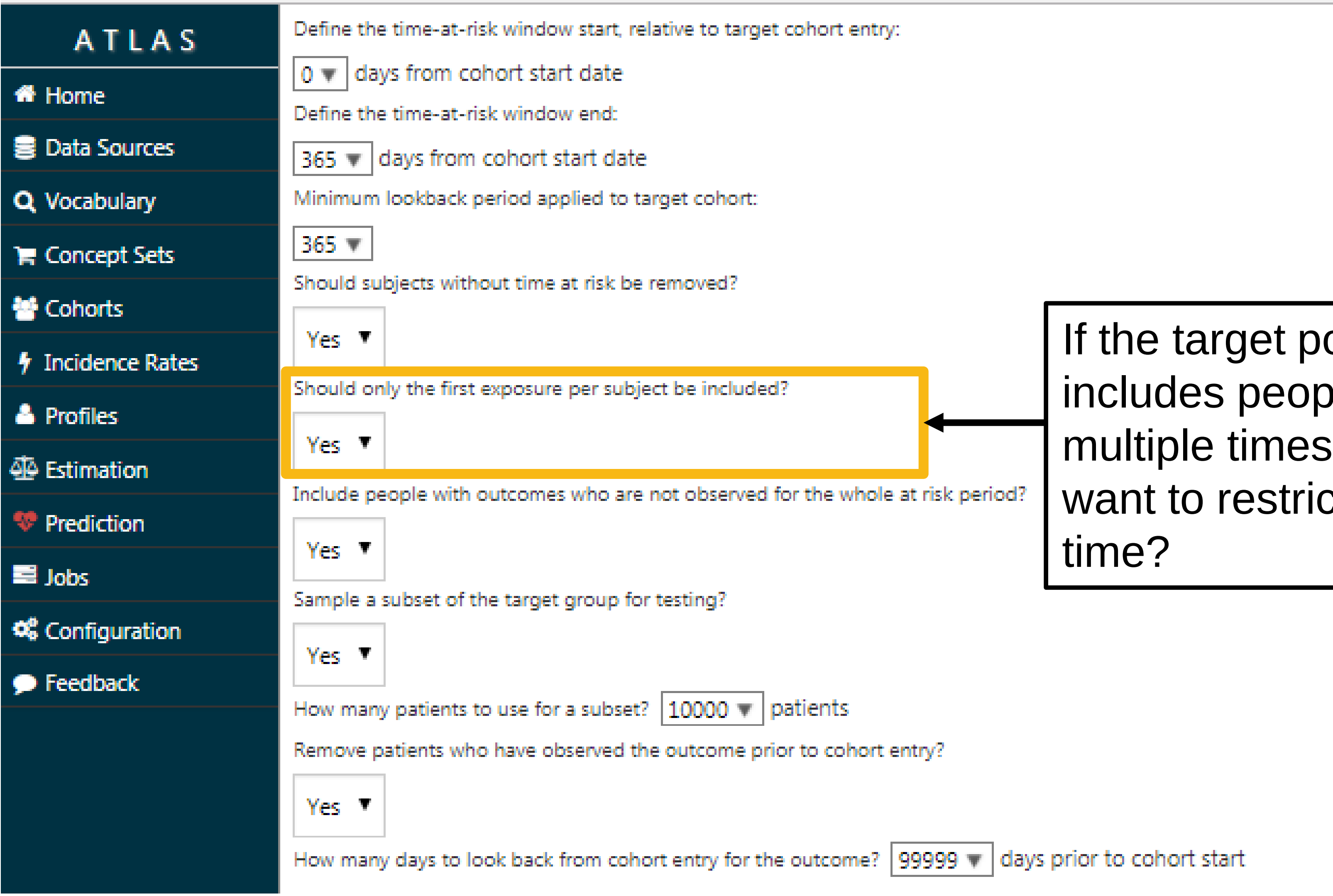


Include people who drop out of the database during the time at risk (noisy labels)?

You can treat outcome and non-outcome people differently (e.g., select this to include outcome people who drop out during time at risk even if you selected yes to removing subjects without time at risk)

Next define the problem specification elements:

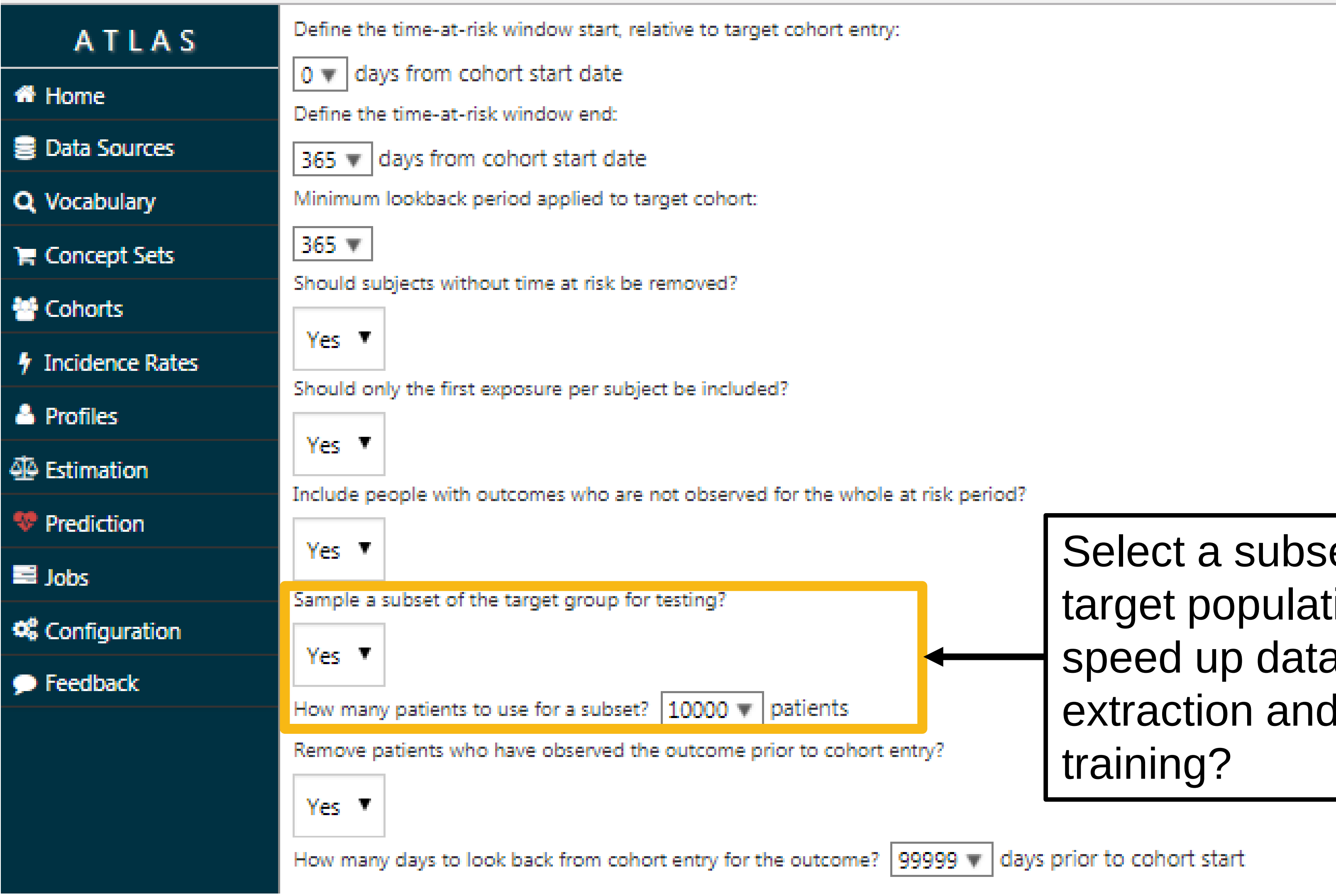
as - prediction



If the target population includes people multiple times, do you want to restrict to first time?

as - prediction

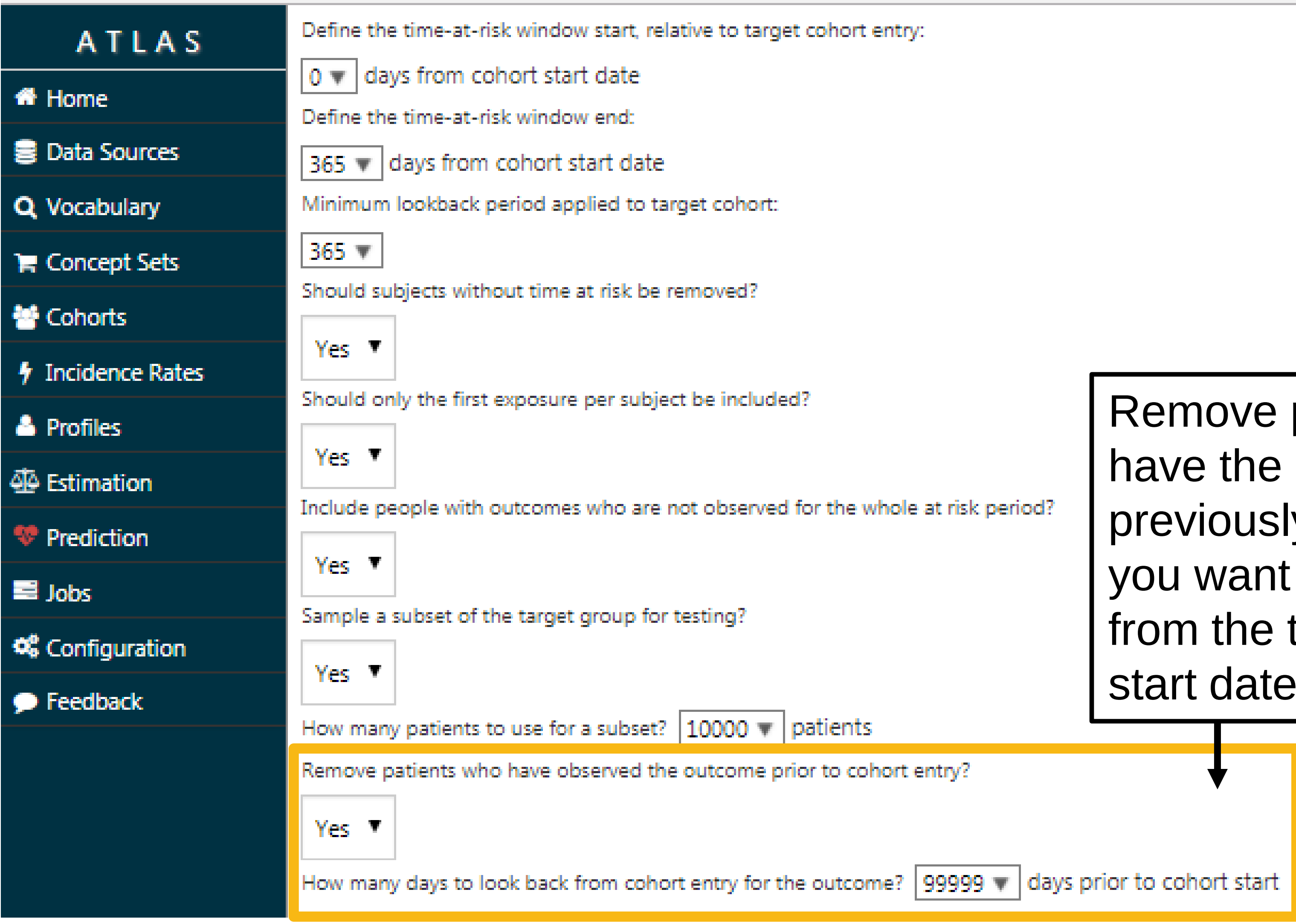
Next define the problem specification elements:



Select a subset of the target population to speed up data extraction and model training?

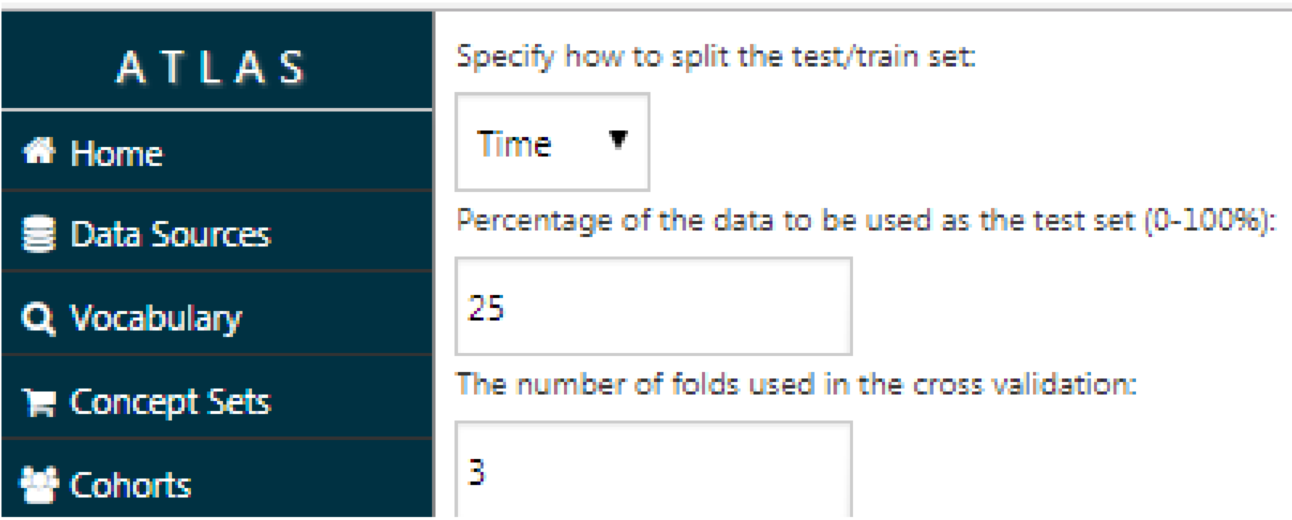
Next define the problem specification elements:

as - prediction



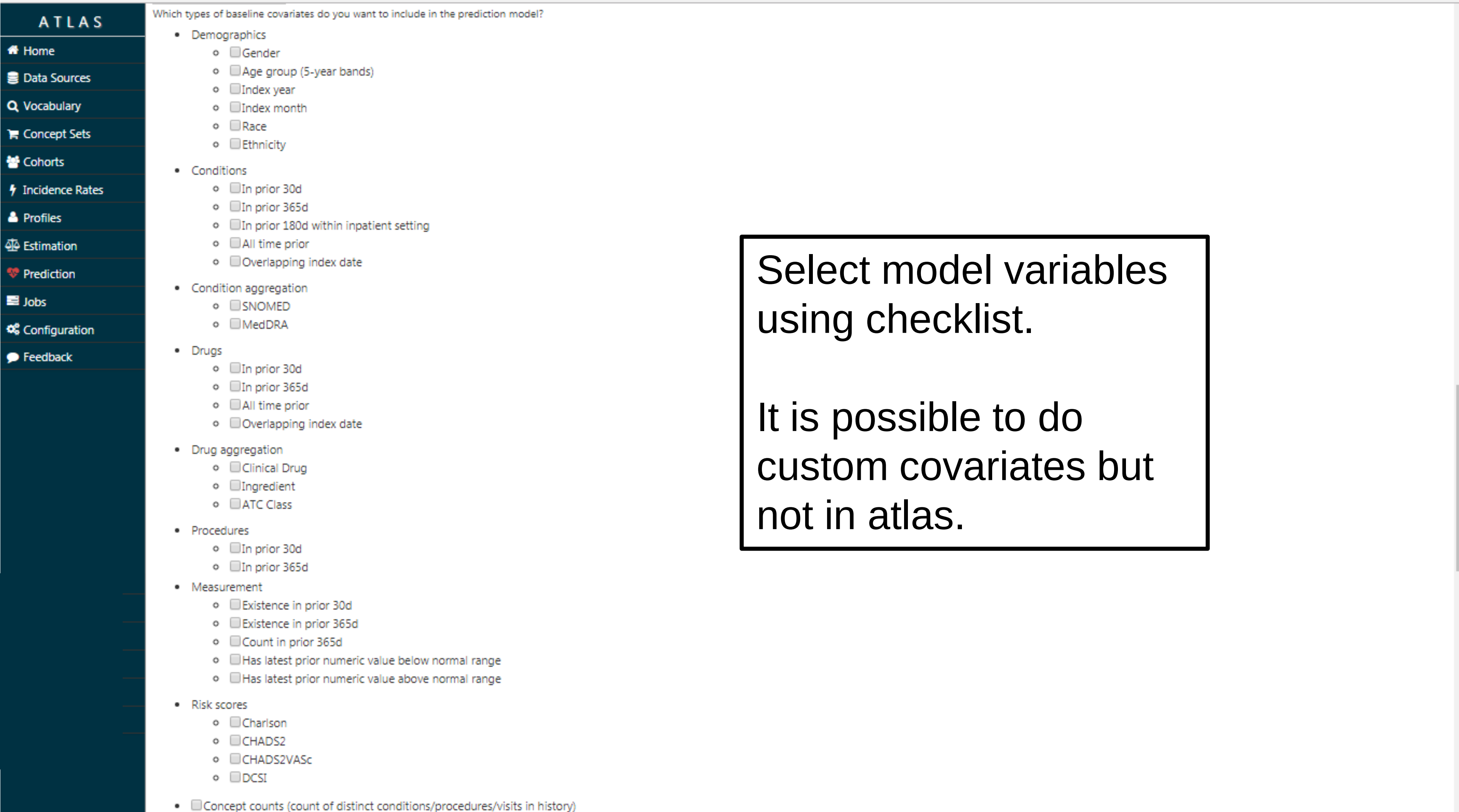
Remove people who have the outcome previously? (how far do you want to look back from the target cohort start date?)

Atlas - prediction



Define the test/train split settings and number of folds used for cross validation during hyper-parameter grid search

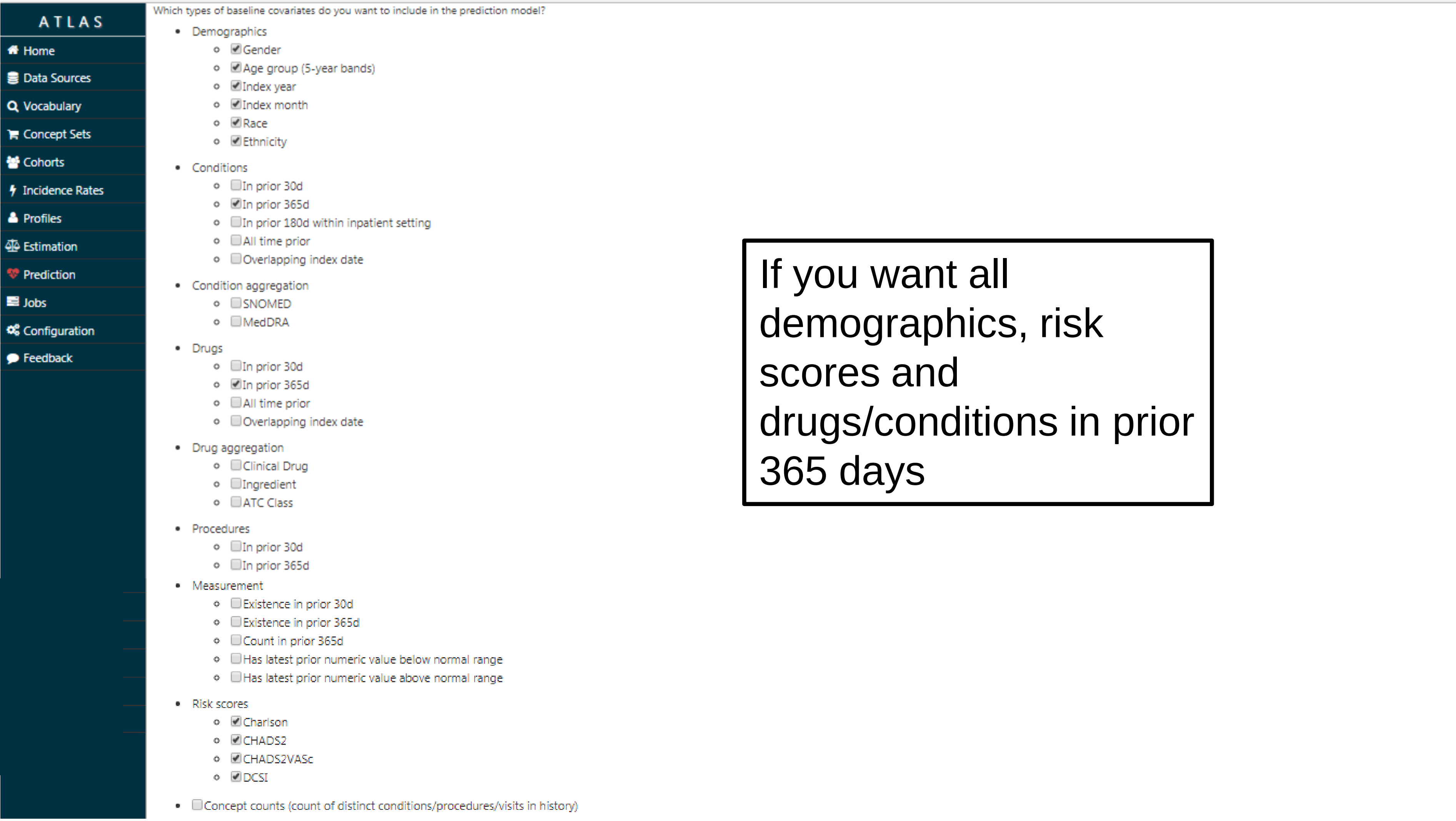
Atlas - prediction



Select model variables using checklist.

It is possible to do custom covariates but not in atlas.

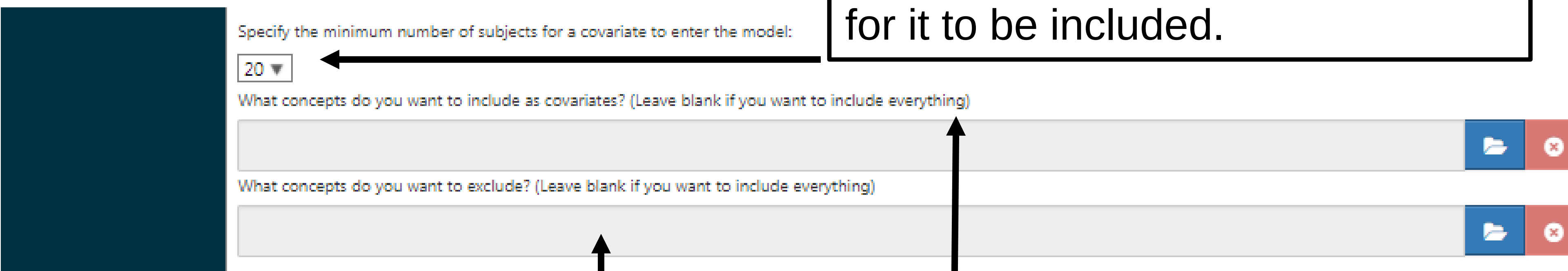
Atlas - prediction



If you want all demographics, risk scores and drugs/conditions in prior 365 days

Atlas - prediction

This sets the minimum number of people in the target population who needs to have the covariate for it to be included.

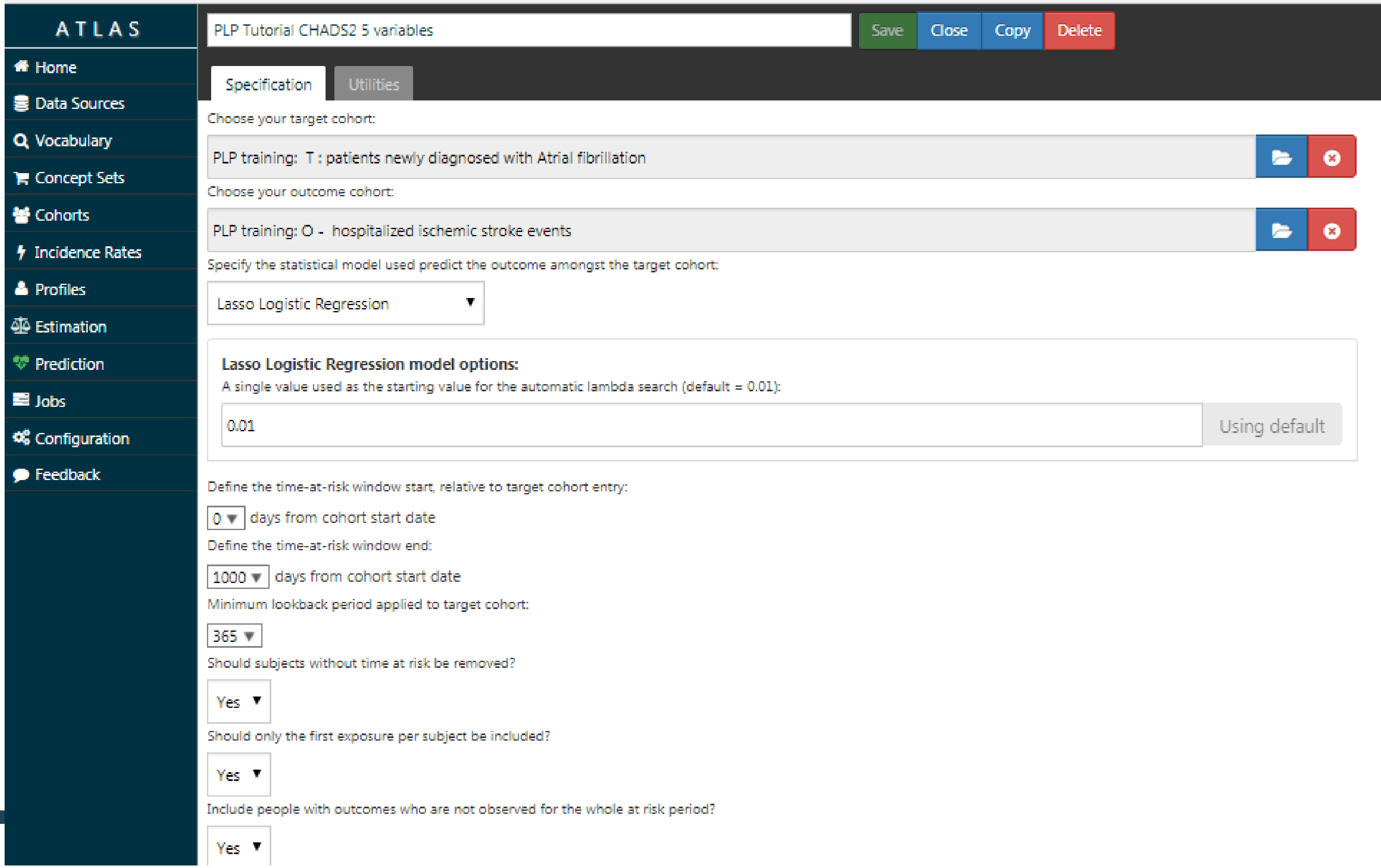


Or make sure some concepts are excluded

You can also restrict to a set of concepts

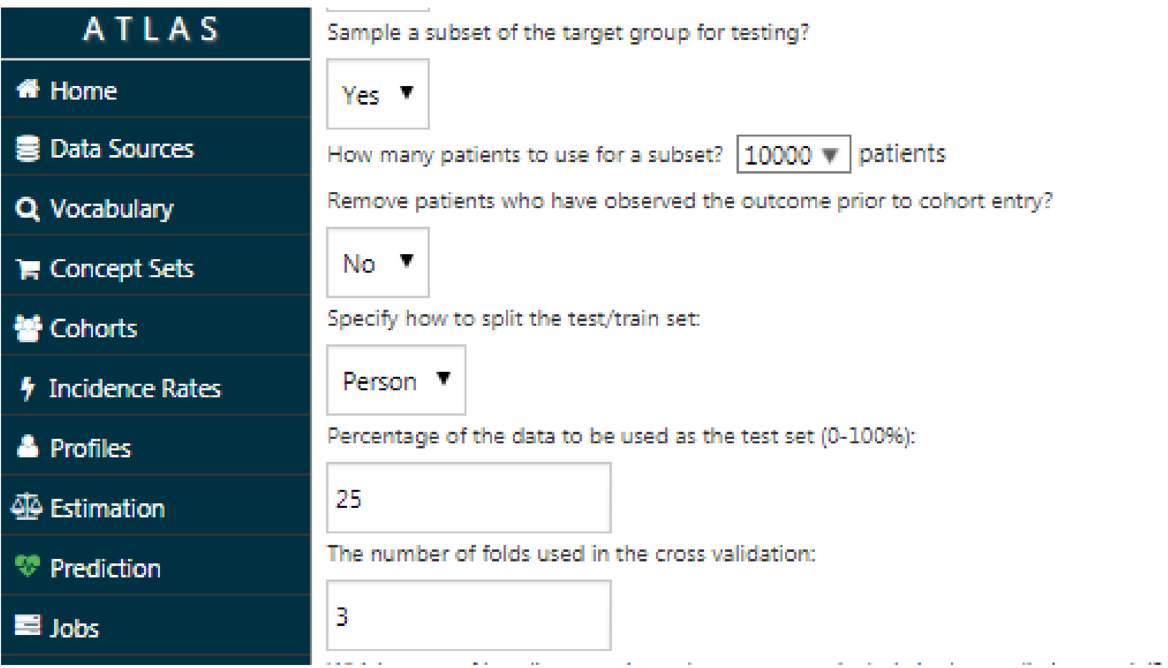
CHADS2 Settings

The settings for the CHADS2 5 variable:



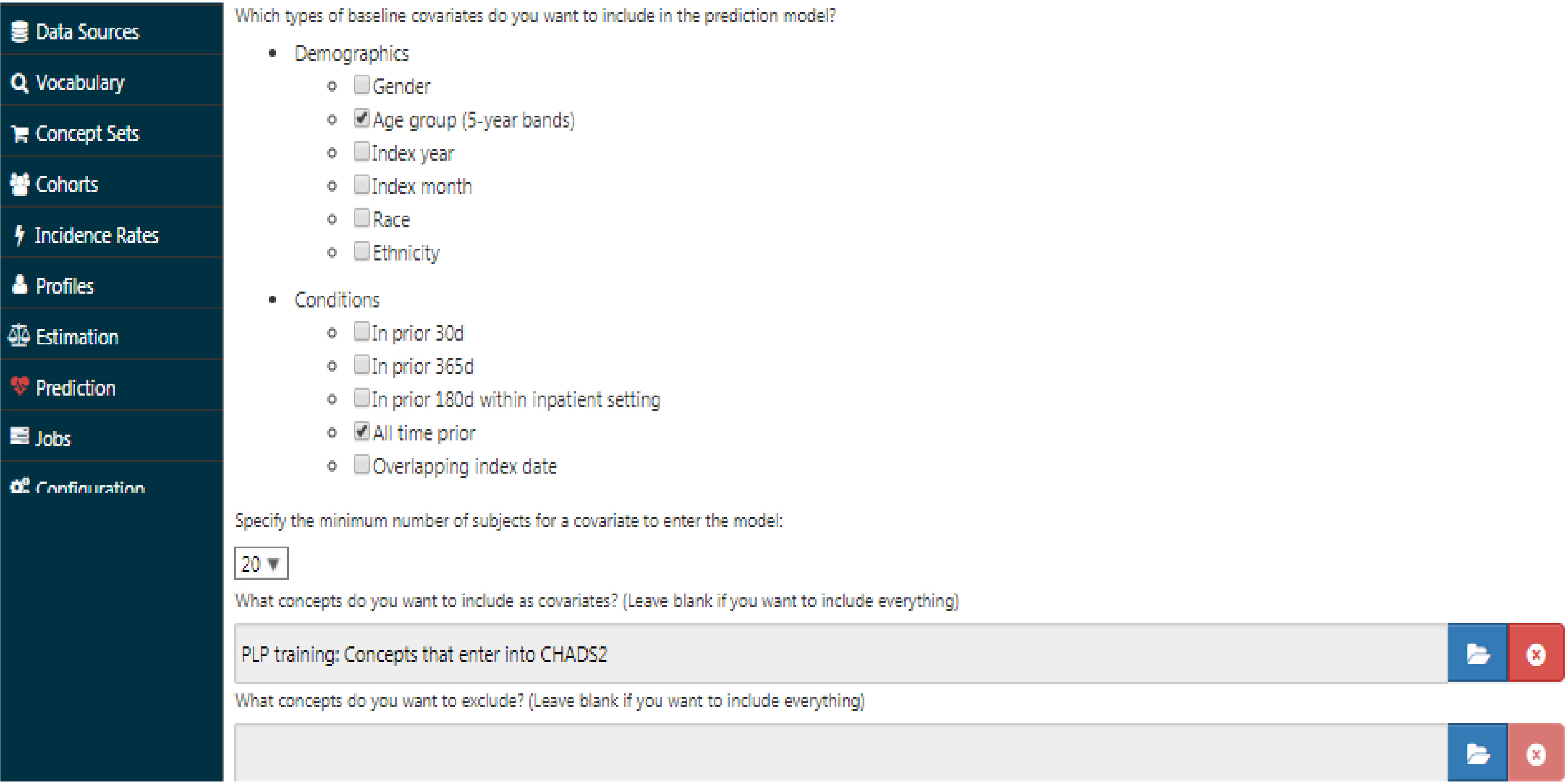
CHADS2 Settings

The settings for the CHADS2 5 variable:

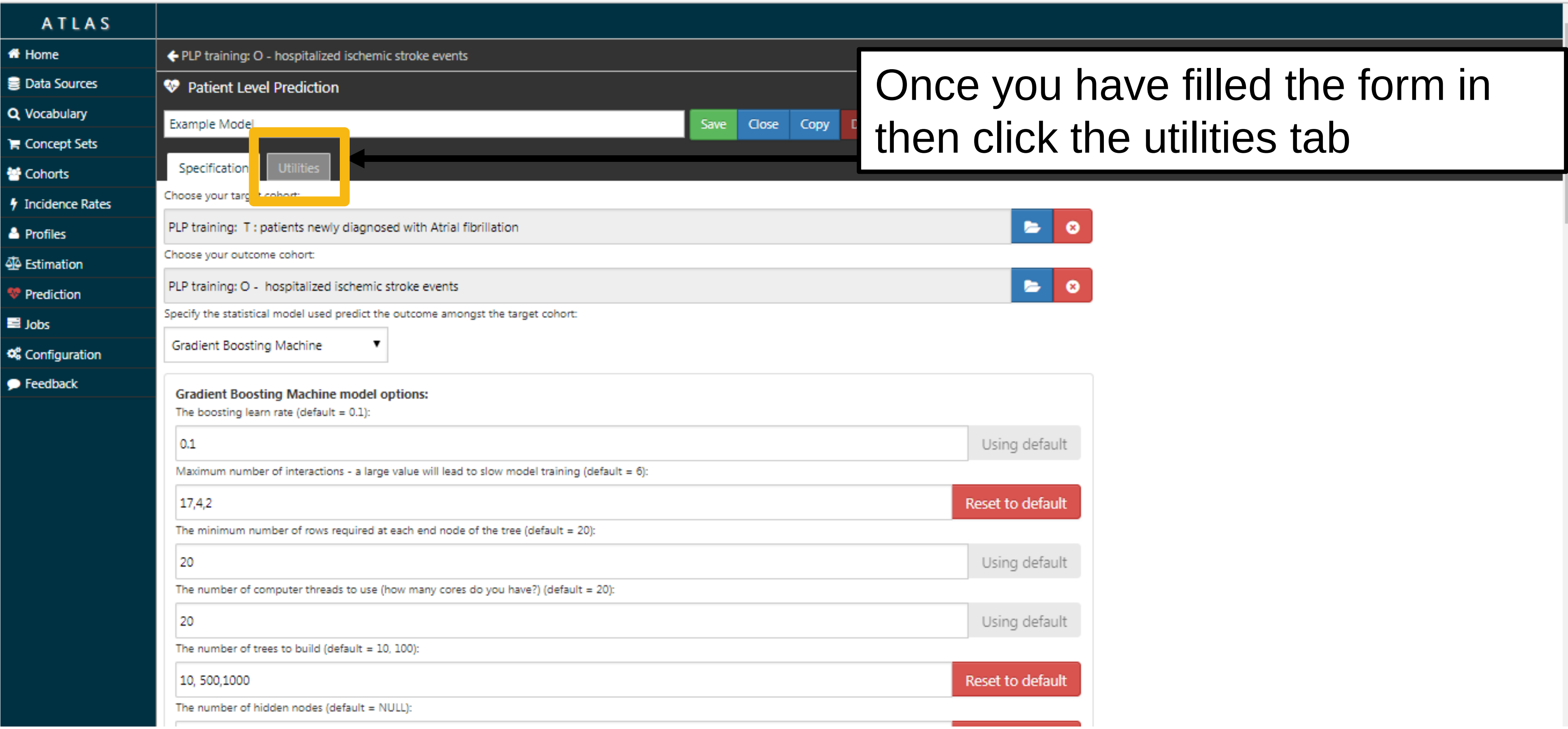


CHADS2 Settings

The settings for the CHADS2 5 variable:

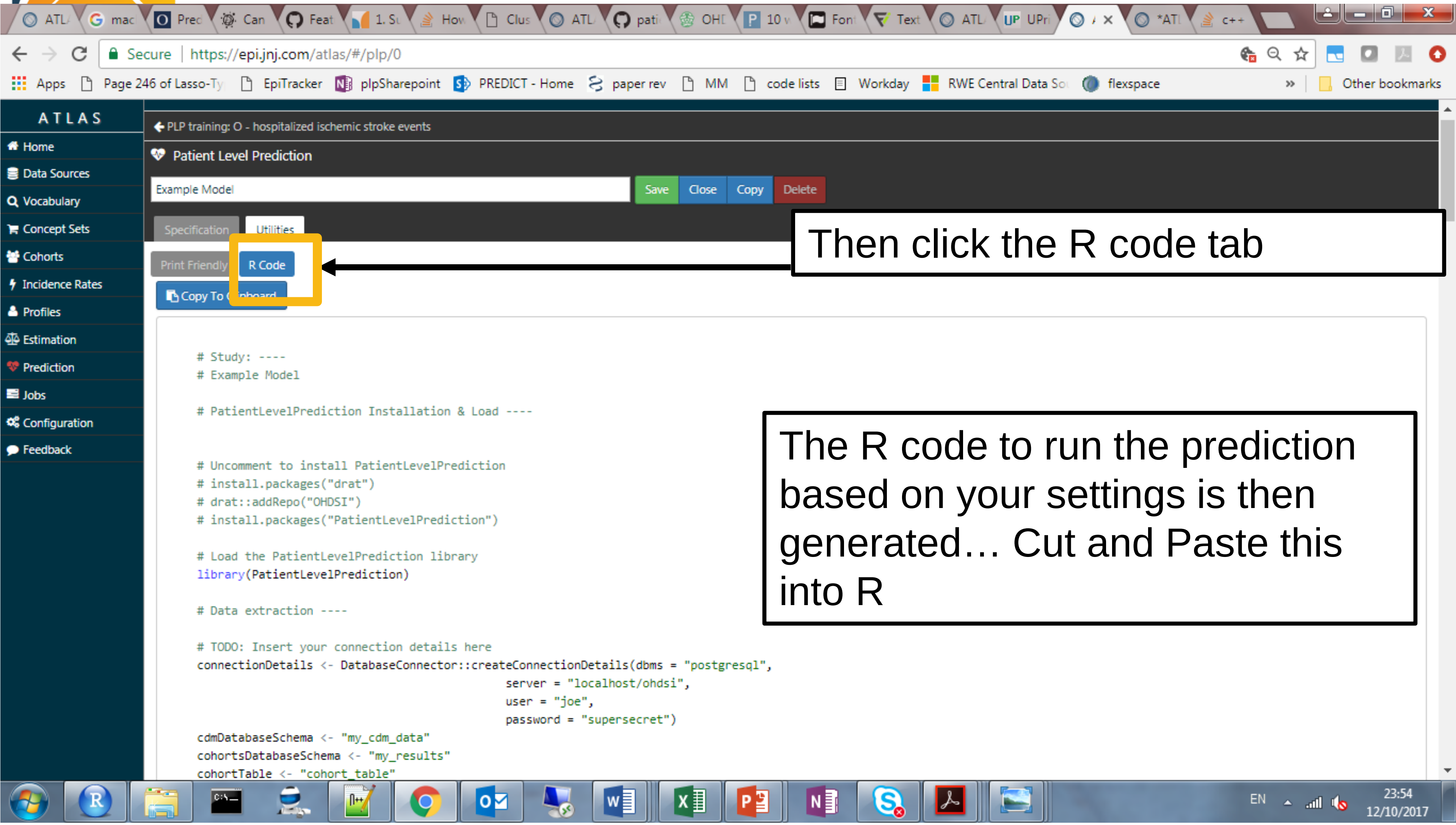


Atlas - prediction



Once you have filled the form in then click the utilities tab

Atlas - prediction

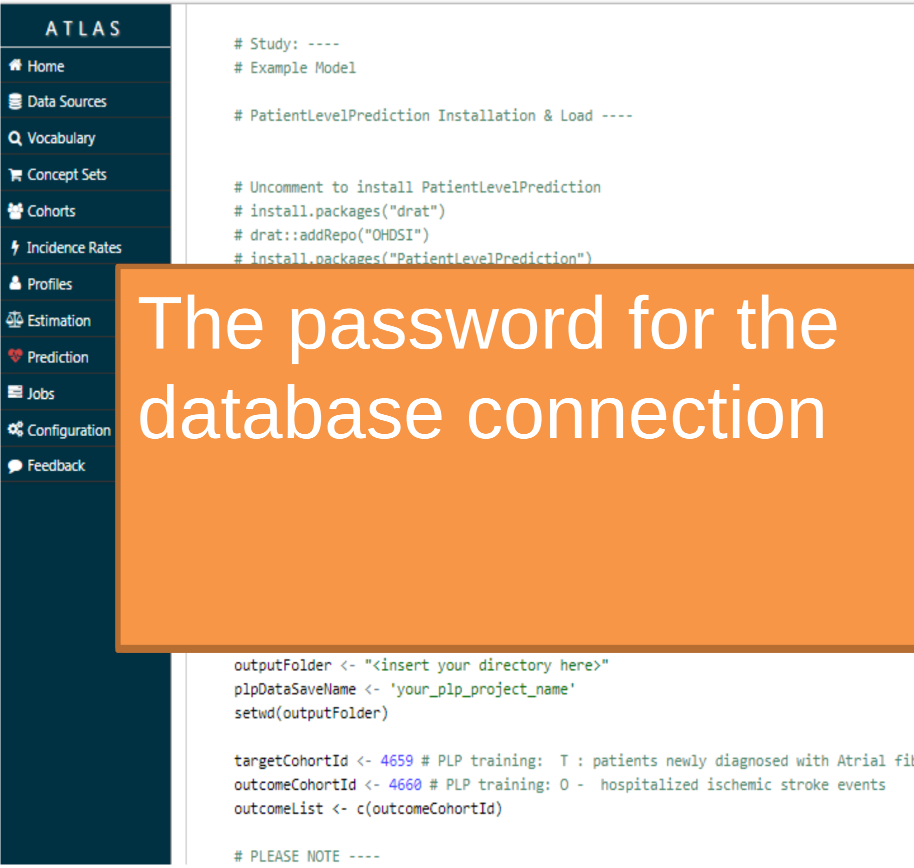


Then click the R code tab

The R code to run the prediction based on your settings is then generated… Cut and Paste this into R

Some manual inputs

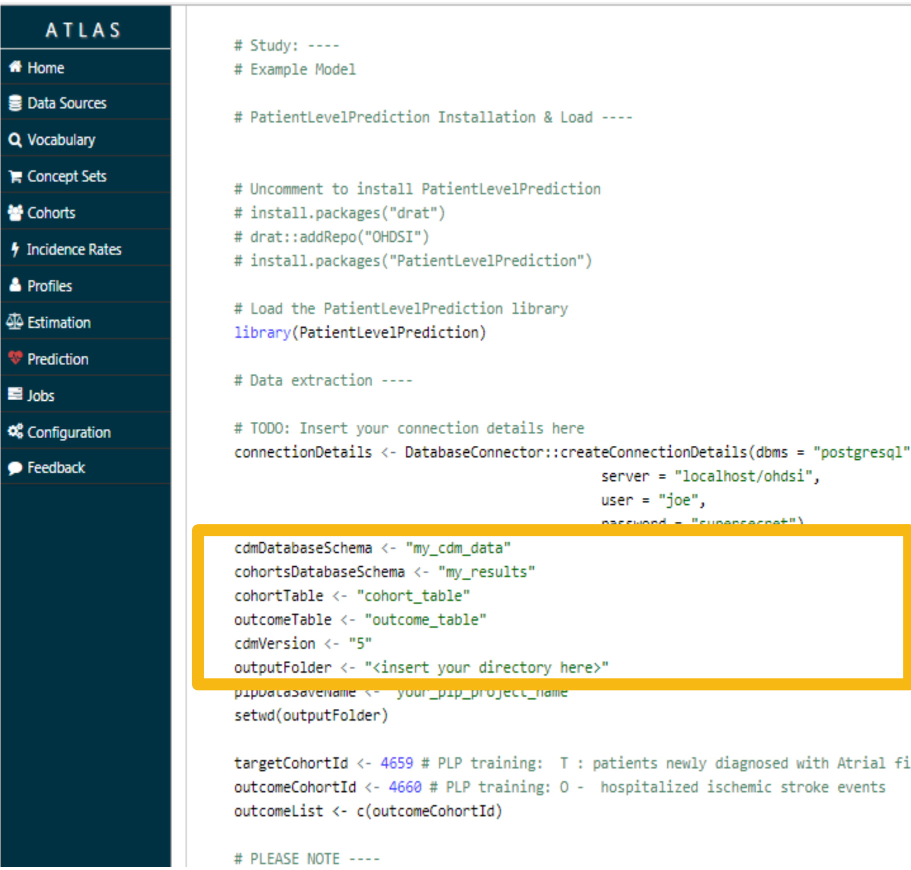
At the top of the Atlas generated R code you will see some required manual inputs



The password for the database connection

```
reateConnectionDetails(dbms = "postgresql",
                       server = "localhost/ohdsi",
                       user = "joe",
                       password = "supersecret")
```

Some manual inputs



At the top of the Atlas generated R code you will see some required manual inputs

```
cdmDatabaseSchema <- "my_cdm_data"
cohortsDatabaseSchema <- "my_results"
cohortTable <- "cohort_table"
outcomeTable <- "outcome_table"
cdmVersion <- "5"
outputFolder <- "<insert your directory here>"
plpDataSaveName <- 'your_plp_project_name'
```

R Functions

1. Create Cohort
2. Database Connections
3. Select Predictor Variables
4. Extract Data
5. Create Study Population
6. Select Model
7. Develop Model + Internal Validation
8. External Validation

R Functions

1	Create Cohort	1	Atlas
2	Database Connections	2	createConnectionDetails()
3	Select Predictor Variables	3	createCovariateSettings()
4	Extract Data	4	getPlpData()
5			()
6			
7			
8			

For examples of using the R code (it has extra flexibilities) see: https://github.com/OHDSI/PatientLevelPrediction/blob/master/inst/doc/BuildingPredictiveModels.pdf

Or type ?PatientLevelPrediction::getPlpData to get the help for getPlpData()

Models and Parameters

Model	setModel Functions	Hyper-parameters
Lasso Logistic Regression	setLassoLogisticRegression()	Variance (regularisation)
Gradient Boosting Machine	setGradientBoostingMachine()	Ntrees (# trees) , max_depth (max interactions), min_rows (regularisation), learn_rate (shinkage - influence decrease per iteration)
Random Forest	setRandomForest()	mtries (predictors per tree) ,ntrees (# trees) ,max_depth (max interactions) ,varImp (feature selection)
Adaboost	setAdaBoost()	n_estimators (interations), learning_rate (shrinkage)
Decision Tree	setDecisionTree()	max_depth (max interactions) , min_samples_split (regularisation) ,min_samples_leaf (regularisation), min_impurity_split ,class_weight
Neural Network	setMLP()	Alpha (regularisation), size (nodes in network)
K-nearest neigh	setKNN()	K (number of neighbours)
Naïve Bayes	setNaiveBayes()	

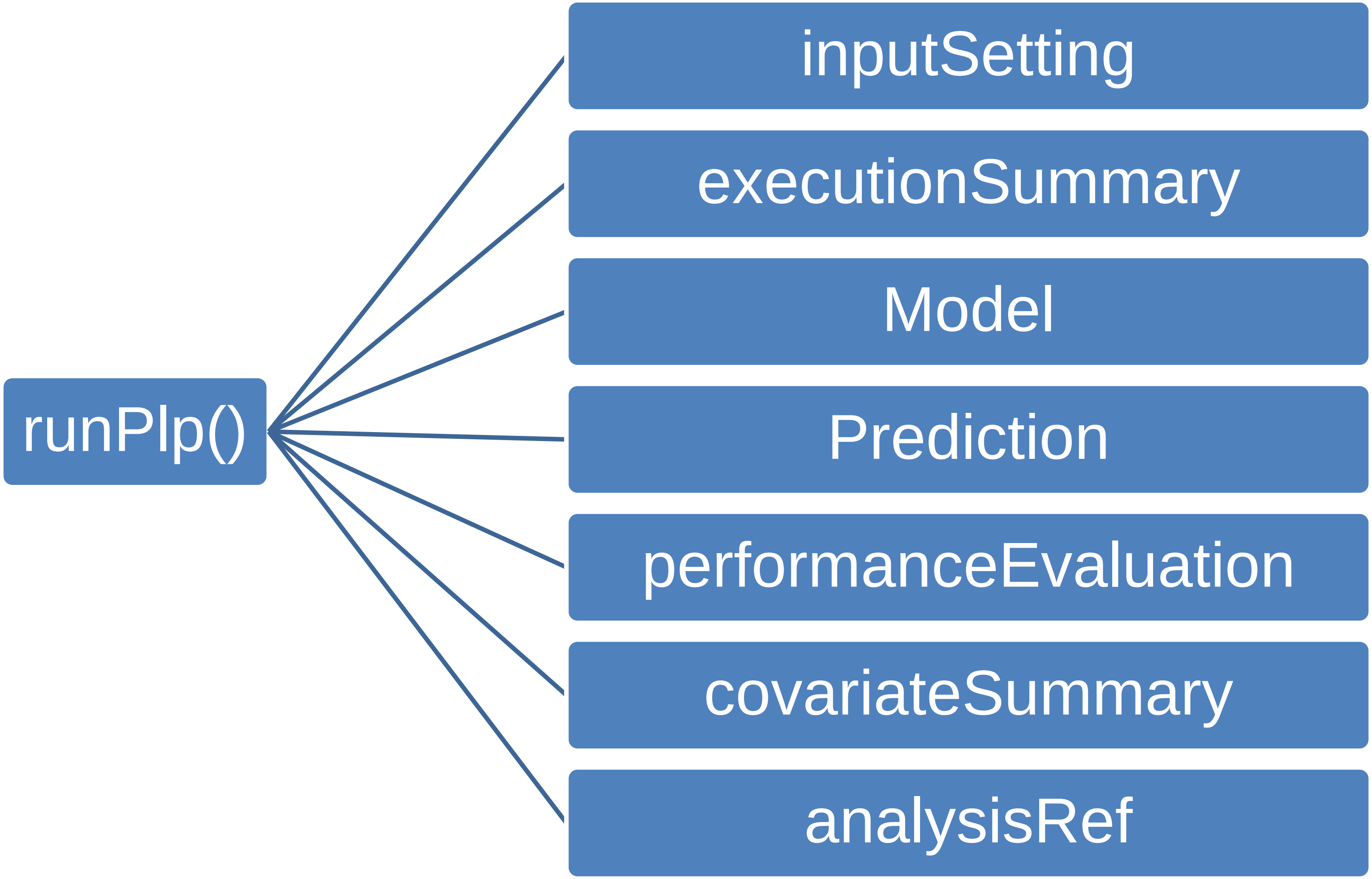
Question Break

Any questions about the implementation, atlas form or R code?

Learning Goals

1. Our 5-step Framework
2. What goes into PatientLevelPrediction
3. Implement various machine learning techniques
4. What comes out of PatientLevelPrediction
5. Interpretation of model performance metrics and plots

Output of runPlp()



Output of runPlp()

runPlp()

inputSetting

this contains information about the settings you selected during model training (e.g., predictor variables, time-at-risk details, minimum prior observation)

Output of runPlp()

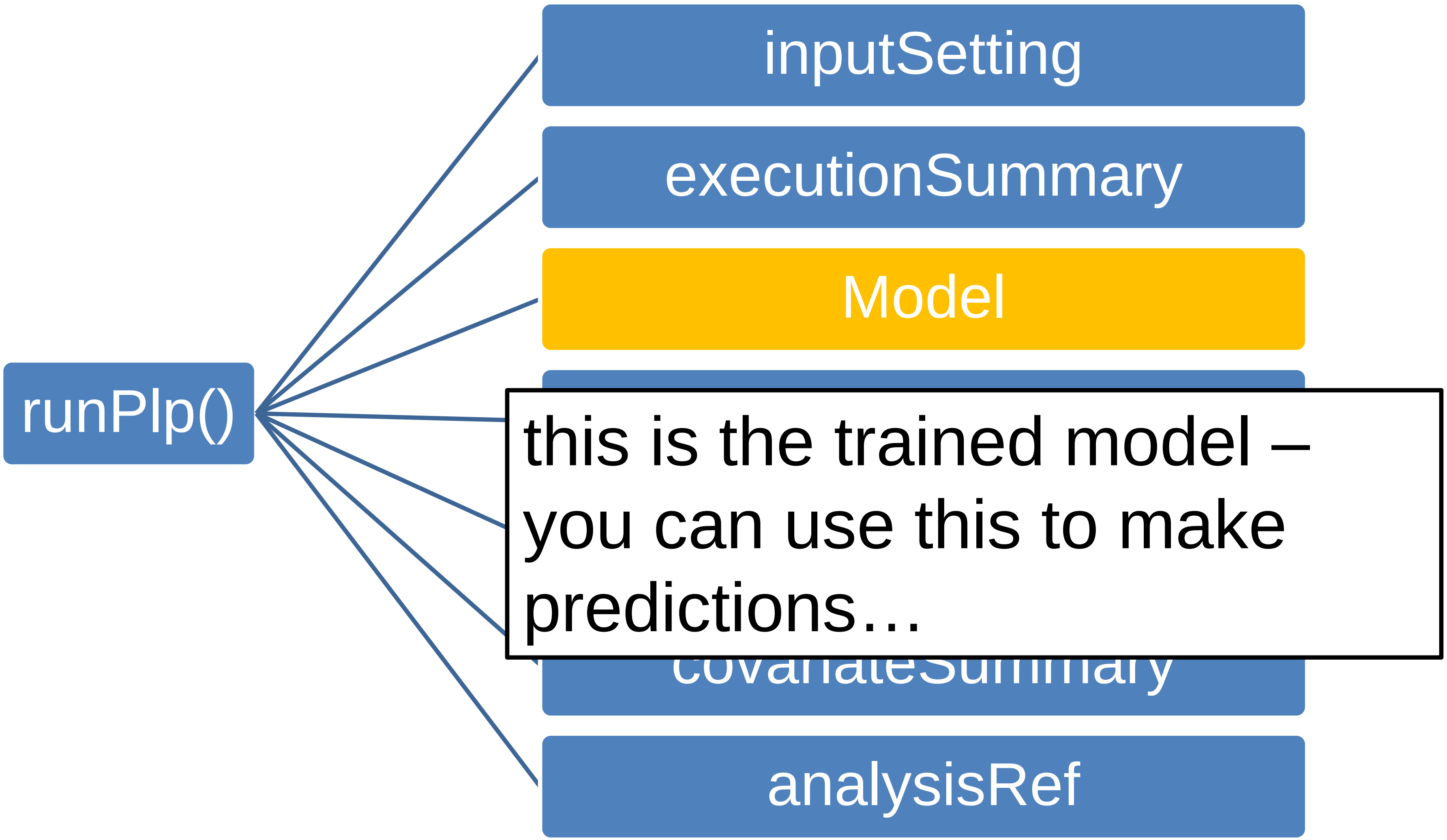
runPlp()

inputSetting

executionSummary

this contains information about the computer used to run the analysis, the R version, the PLP package version, the analysis date and the execution time.

Output of runPlp()



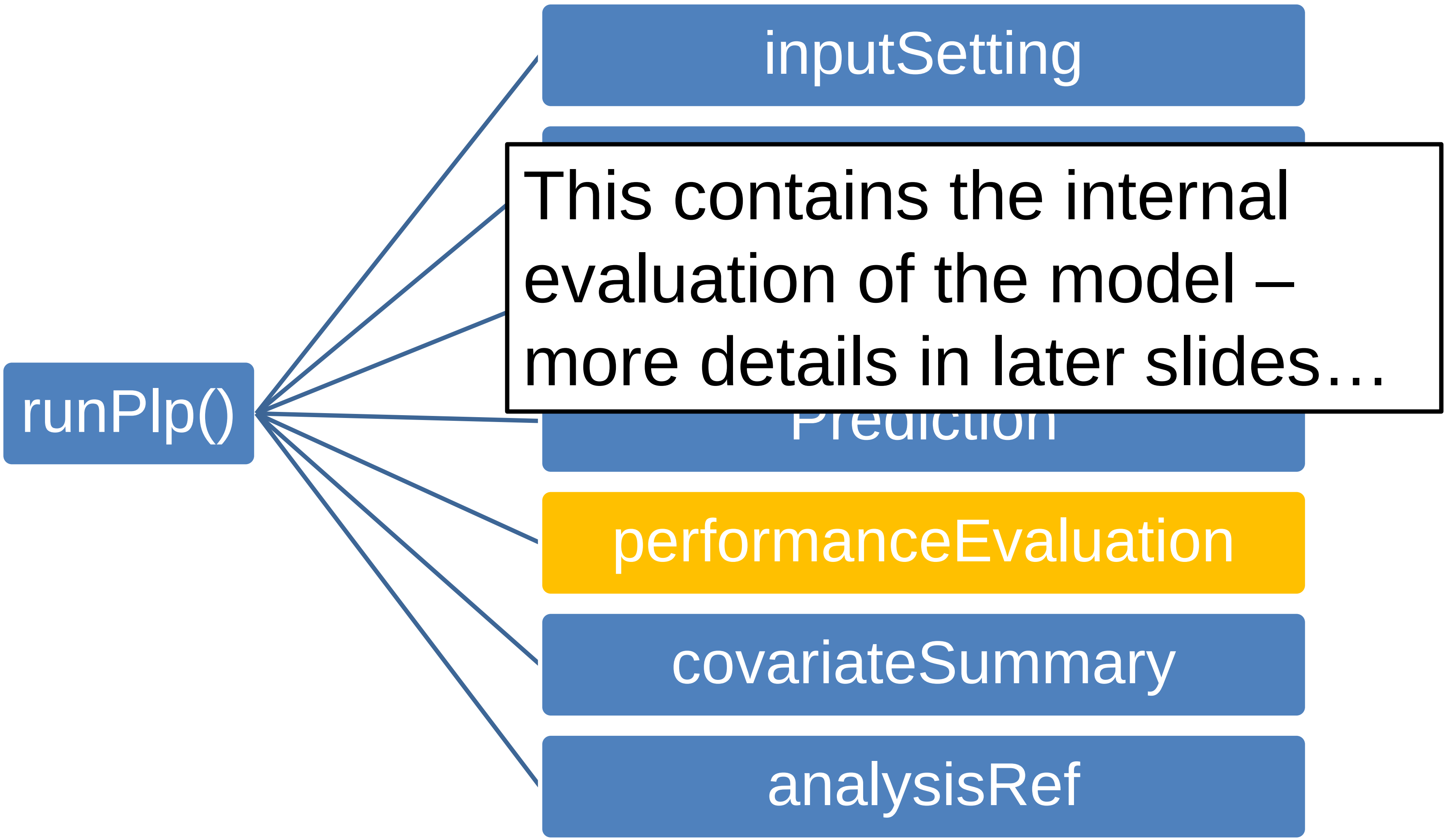
Output of runPlp()

This is a table that contains the predicted risk of the outcome during the time at risk for each person in the target population …

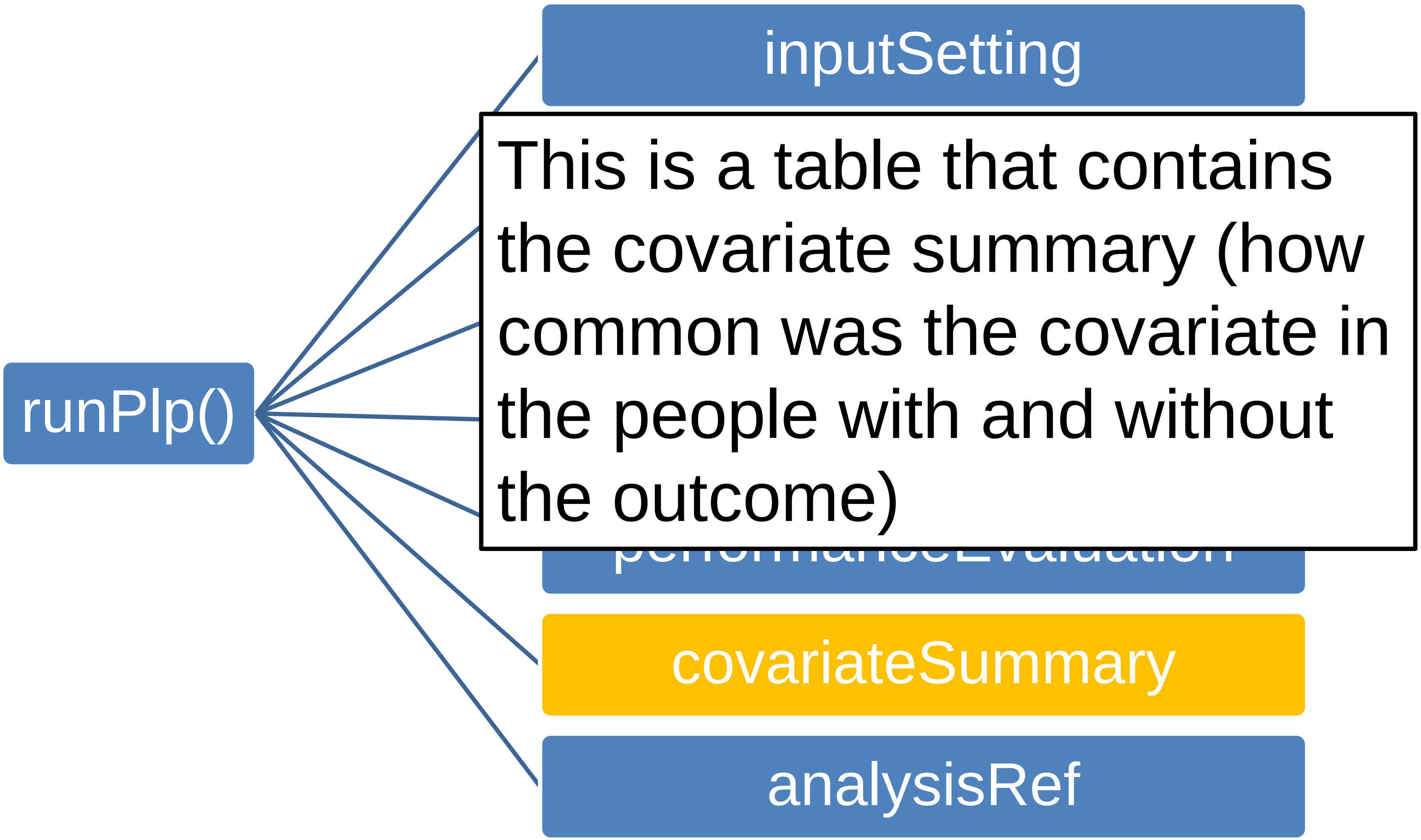
runPlp()

- Prediction
- performanceEvaluation
- covariateSummary
- analysisRef

Explaining performanceEvaluation



Output of runPlp()

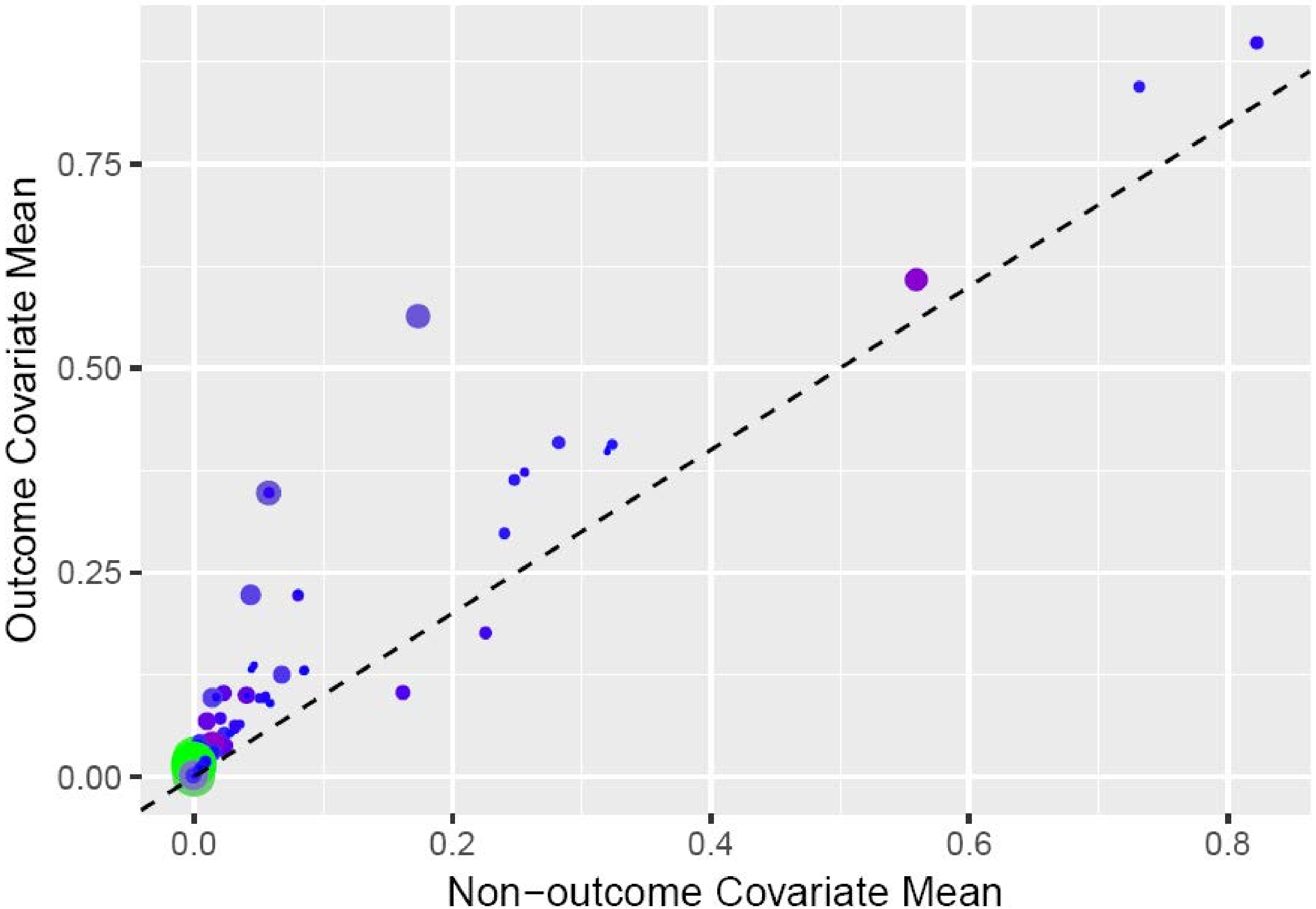


Covariate Summary

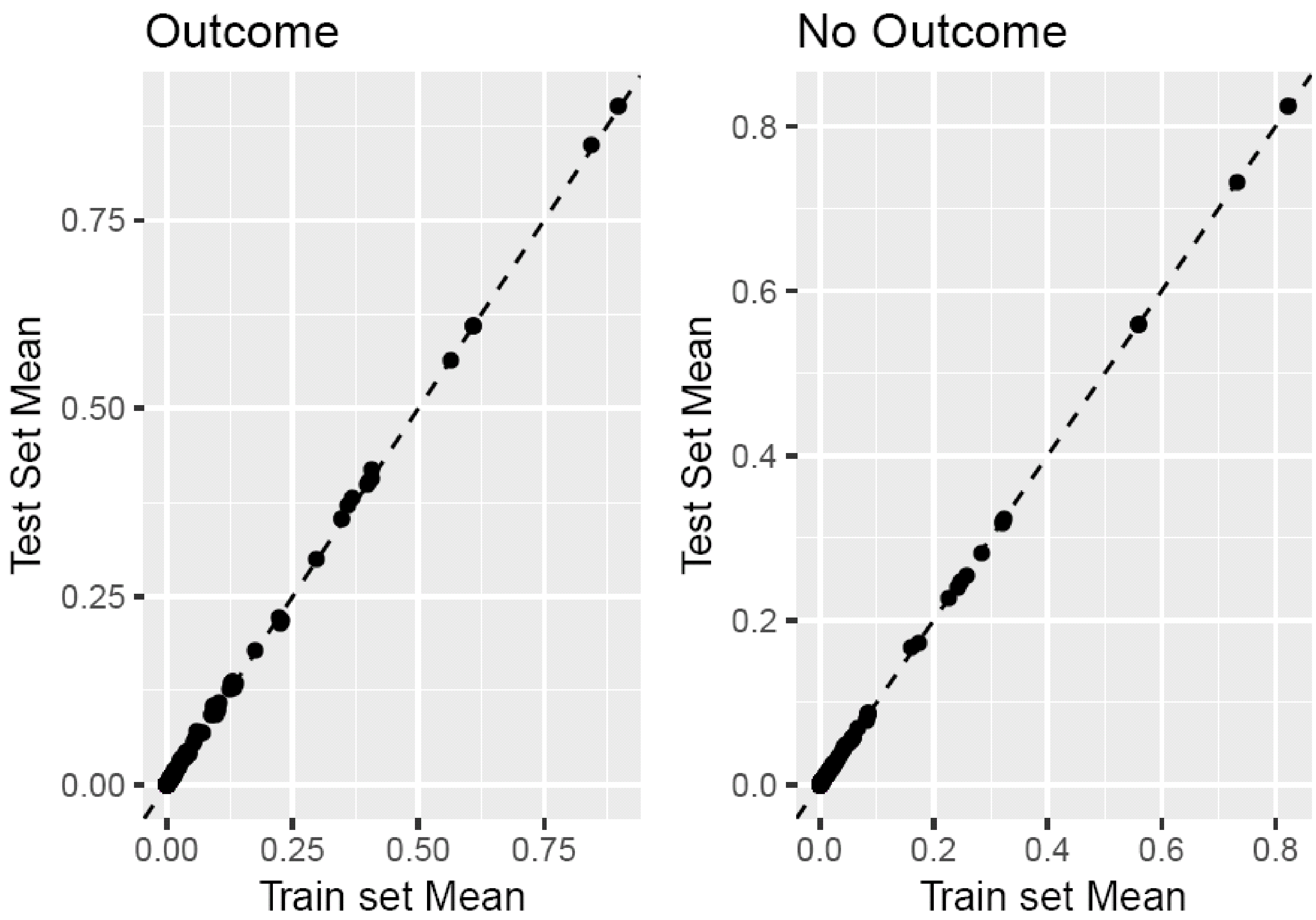
The data characterisation – how often the variable occurred in the people with and without the outcome…

covariateId	covariateName	analysisId	conceptId	covariateValue	CovariateCount	CovariateCountWithOutcome
23	Age group: 65-69	5	0	-3.278466e-01	28223	3509
24	Age group: 70-74	5	0	-1.884020e-01	40549	5939
25	Age group: 75-79	5	0	0.000000e+00	NA	NA
26	Age group: 80-84	5	0	1.604763e-01	46874	10018
27	Age group: 85-89	5	0	5.456322e-01	14581	4216
28	Age group: 90-94	5	0	3.332457e+00	23	23
135601201	Condition era record observed during anytime on or …	201	135601	0.000000e+00	1	0

This can be used to plot variable scatterplot



This can be used to plot variable generalisation plot



Output of runPlp()

runPlp()

inputSetting

executionSummary

Model

This contains the analysis details such as analysis id or name

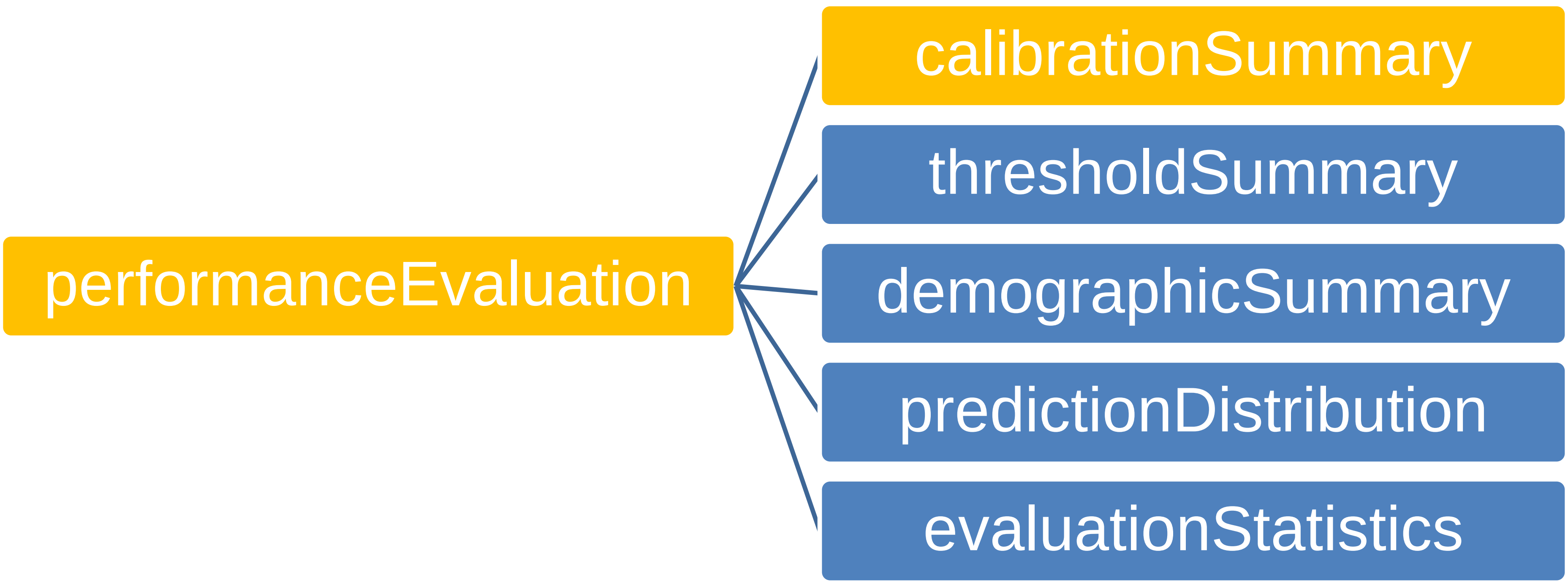
covariateSummary

analysisRef

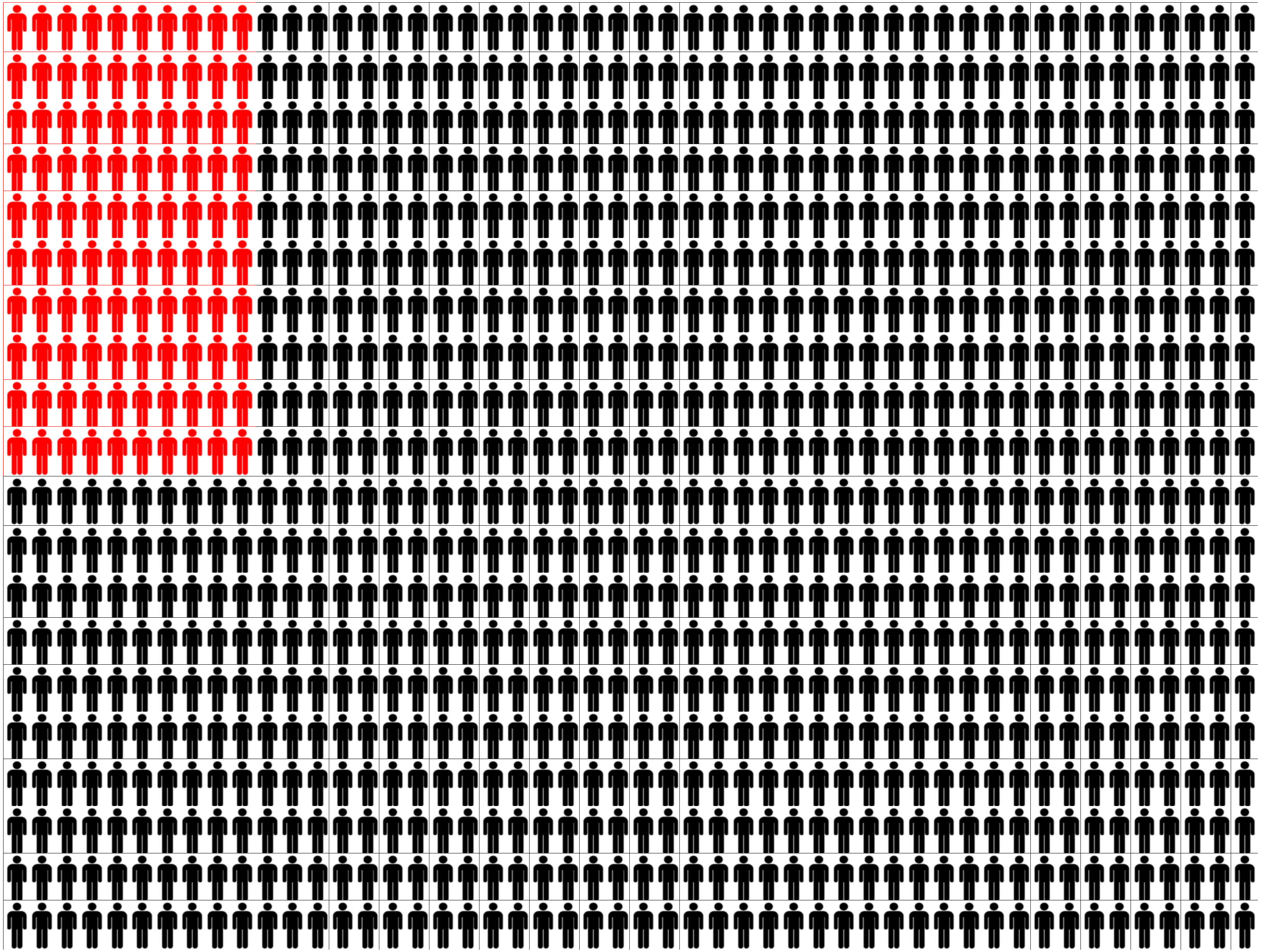
Contents of performanceEvaluation

- performanceEvaluation
  - calibrationSummary
  - thresholdsummary
  - demographicSummary
  - predictionDistribution
  - evaluationStatistics

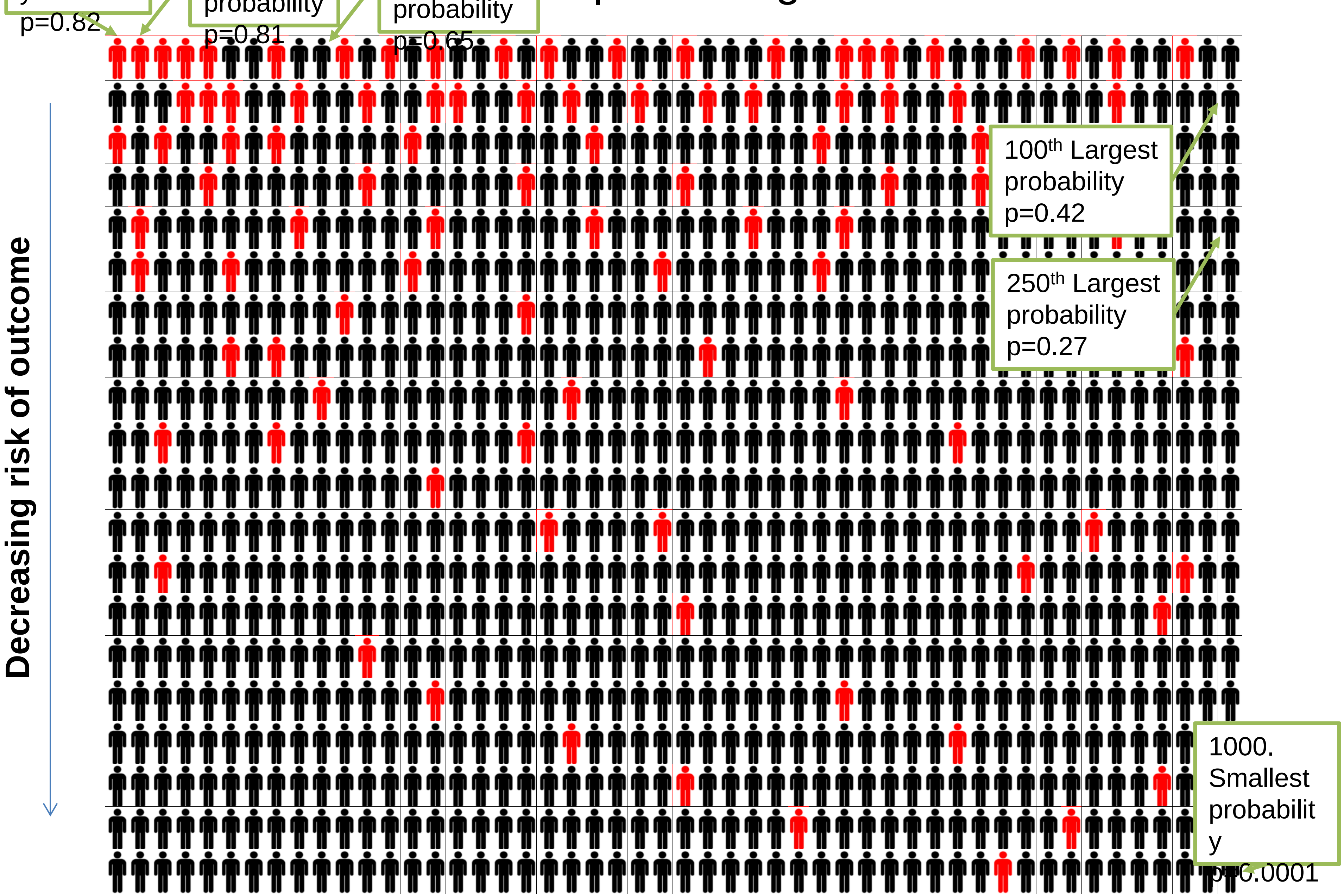
Contents of performanceEvaluation



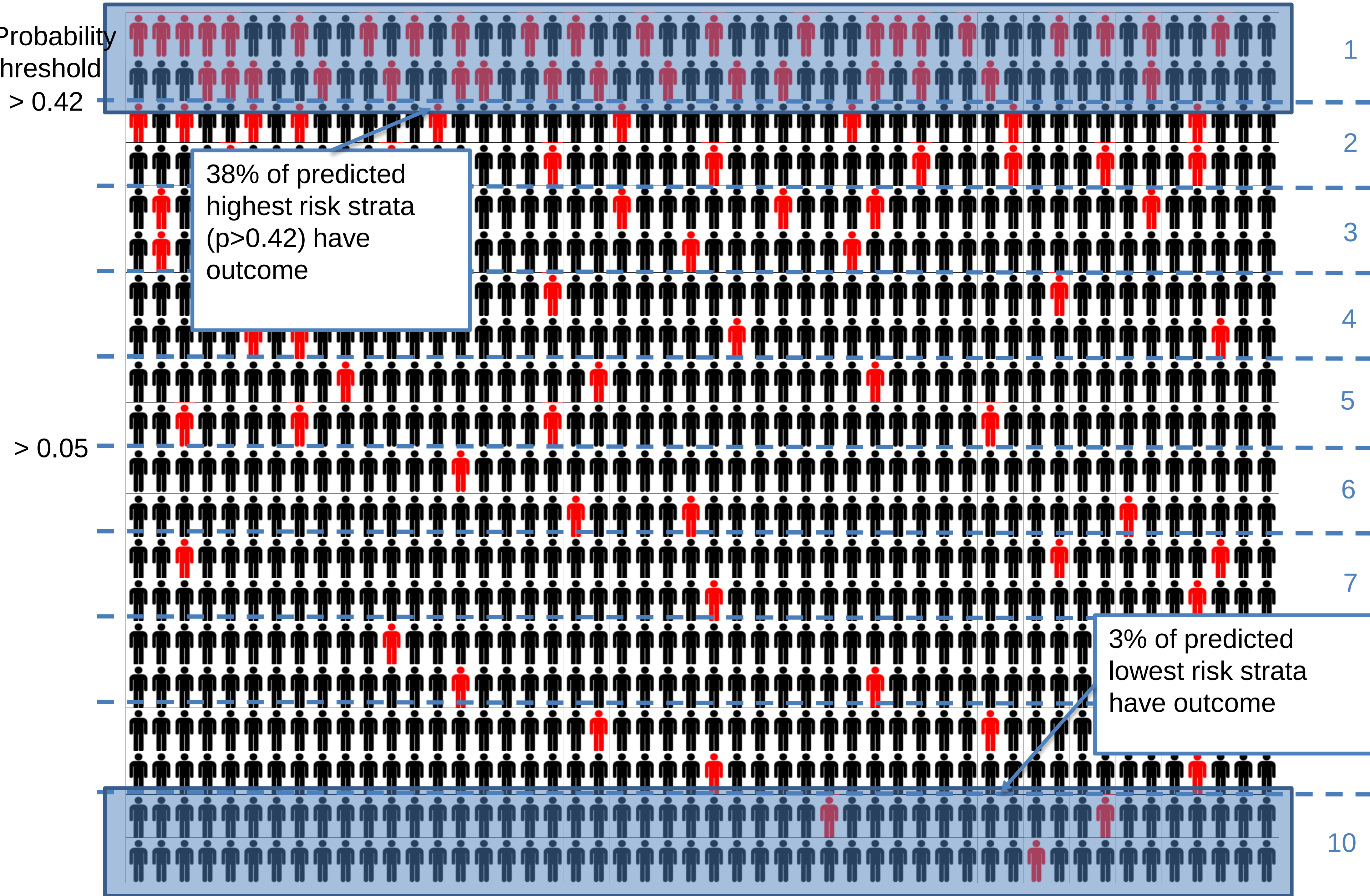
Amongst a target population of 1000 patients, 10% of the patients experience the outcome during the time-at-risk



…di… …odel can rank all 1000 patients by their predicted …ba… experiencing the outcome



Calibration summary: partition target population into 10 strata and compare observed incidence with predicted incidence



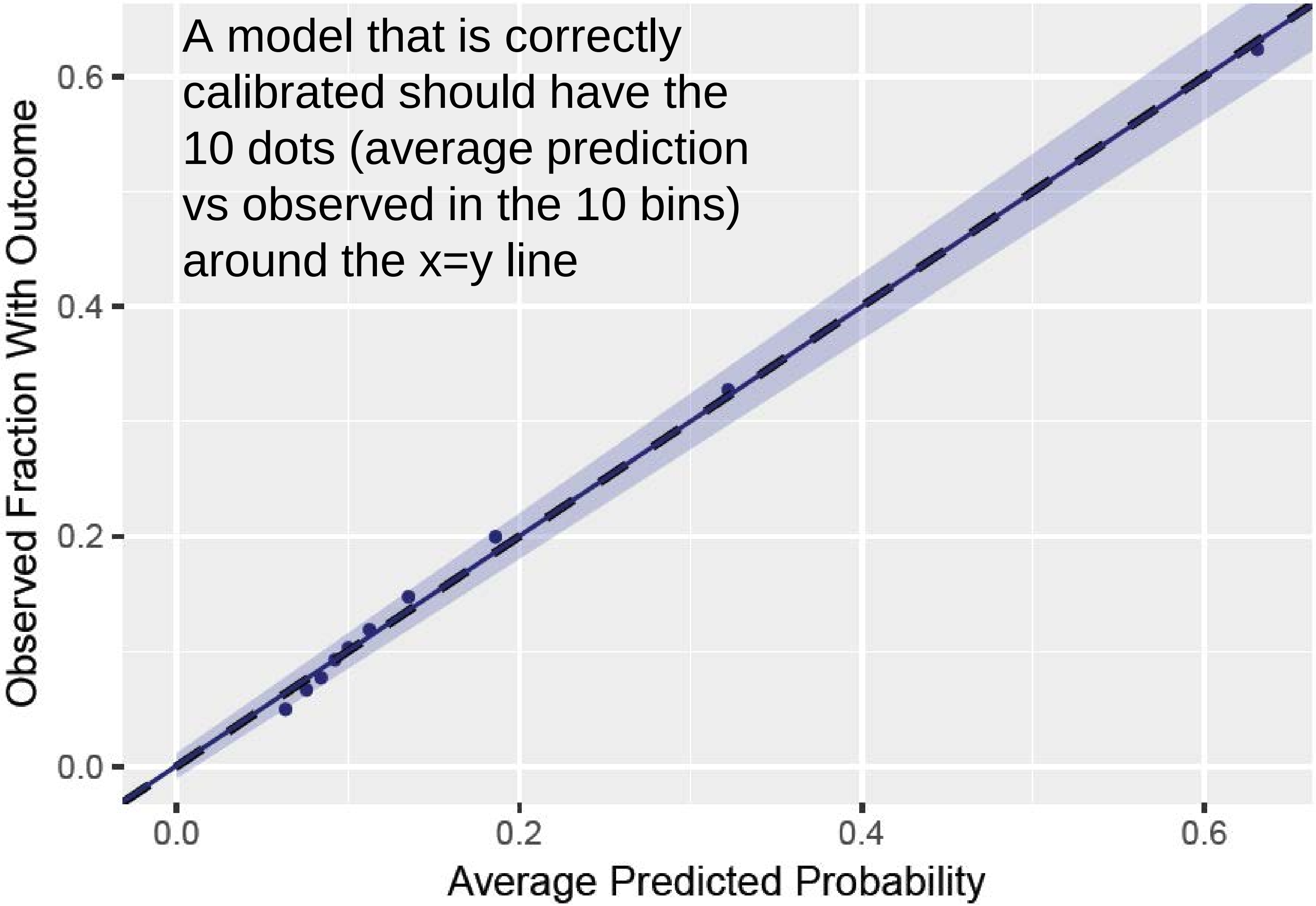
calibrationSummary

For each probability threshold bin we calculate the predicted vs observed outcome occurrence (this is done for the train and the test set)

predictionThreshold	PersonCountAtRisk	PersonCountWithOutcome	averagePredictedProbability	StDevPredictedProbability
0.00000000	5049	255	0.06273506	0.004876098
0.07021915	4746	333	0.07478538	0.001622543
0.07680500	5518	426	0.08406877	0.003091515
0.08746617	3392	299	0.09239939	0.002789445
0.09651138	4854	499	0.10153173	0.002668102

5049 people had a predicted risk between 0 and 0.07 and 255 had the outcome (observed occurrence of 0.05) the average predicted risk in this group was 0.06

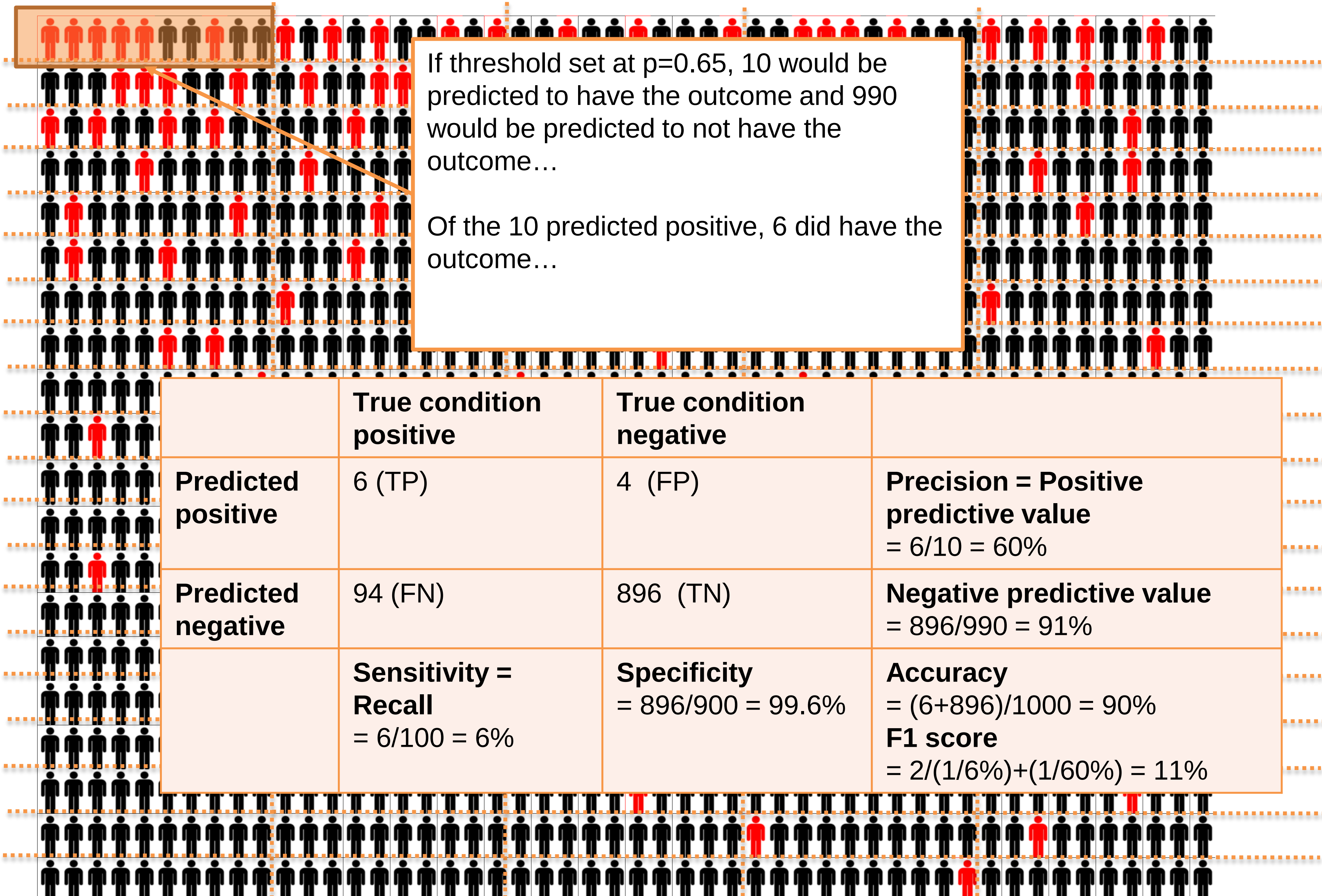
This can be used to plot calibration



Contents of performanceEvaluation

- performanceEvaluation
  - calibrationSummary
  - thresholdSummary
  - demographicSummary
  - predictionDistribution
  - evaluationStatistics

Threshold summary: create 100 cumulative thresholds and evaluate performance of the binary classifier at each threshold

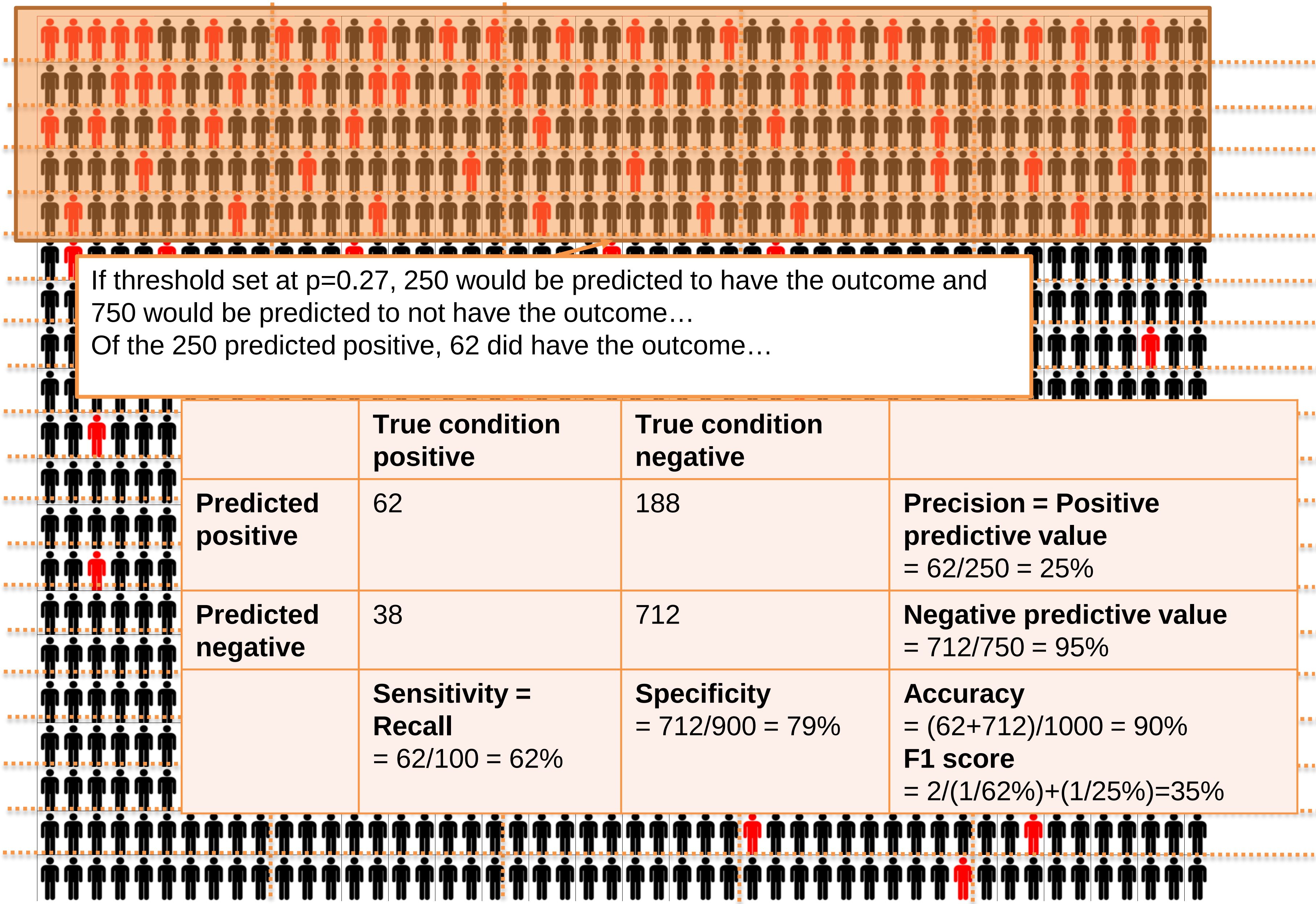


If threshold set at p=0.65, 10 would be predicted to have the outcome and 990 would be predicted to not have the outcome…

Of the 10 predicted positive, 6 did have the outcome…

	True condition positive	**True condition negative**	
Predicted positive	6 (TP)	4 (FP)	**Precision = Positive predictive value** = 6/10 = 60%
Predicted negative	94 (FN)	896 (TN)	**Negative predictive value** = 896/990 = 91%
	Sensitivity = Recall = 6/100 = 6%	**Specificity** = 896/900 = 99.6%	**Accuracy** = (6+896)/1000 = 90% **F1 score** = 2/(1/6%)+(1/60%) = 11%

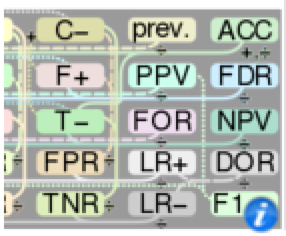
Threshold summary: create 100 cumulative thresholds and evaluate performance of the binary classifier at each threshold



If threshold set at p=0.27, 250 would be predicted to have the outcome and 750 would be predicted to not have the outcome…
Of the 250 predicted positive, 62 did have the outcome…

	True condition positive	**True condition negative**	
Predicted positive	62	188	**Precision = Positive predictive value** = 62/250 = 25%
Predicted negative	38	712	**Negative predictive value** = 712/750 = 95%
	Sensitivity = Recall = 62/100 = 62%	**Specificity** = 712/900 = 79%	**Accuracy** = (62+712)/1000 = 90% **F1 score** = 2/(1/62%)+(1/25%)=35%

Confusion matrix cheatsheet

	True condition				
Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition positive	**True positive**	**False positive** (Type I error)	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
Predicted condition negative	**False negative** (Type II error)	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
	True positive rate (TPR), Recall, Sensitivity, probability of detection = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score = $\frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}}$
	False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	True negative rate (TNR), Specificity (SPC) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		

https://en.wikipedia.org/wiki/Sensitivity_and_specificity

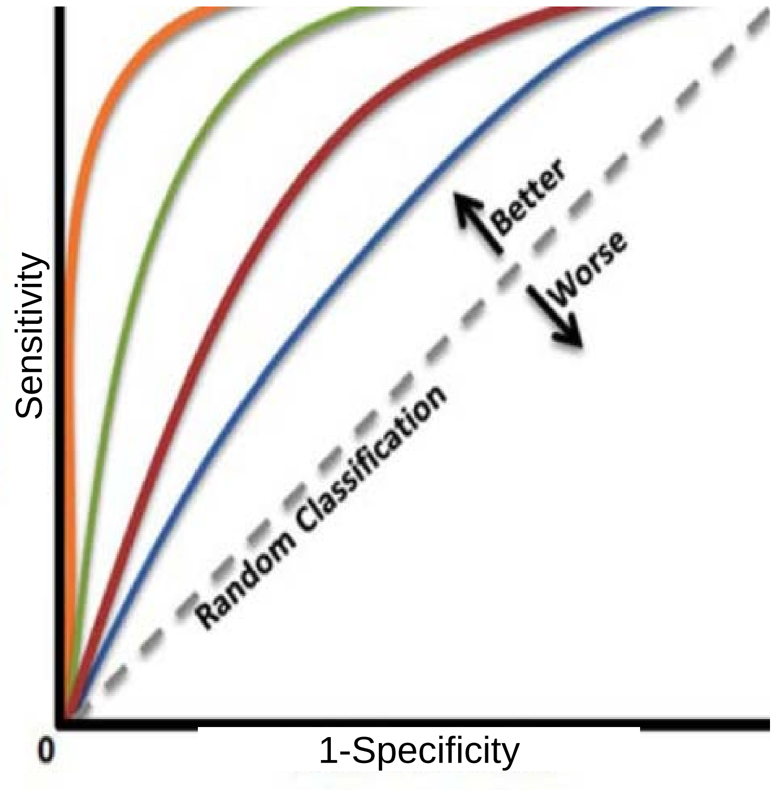
thresholdSummary

Contains general performance metrics such as TP, TN, FP, FN, sensitivity, specificity at the 100 prediction threshold points (more columns than shown…)

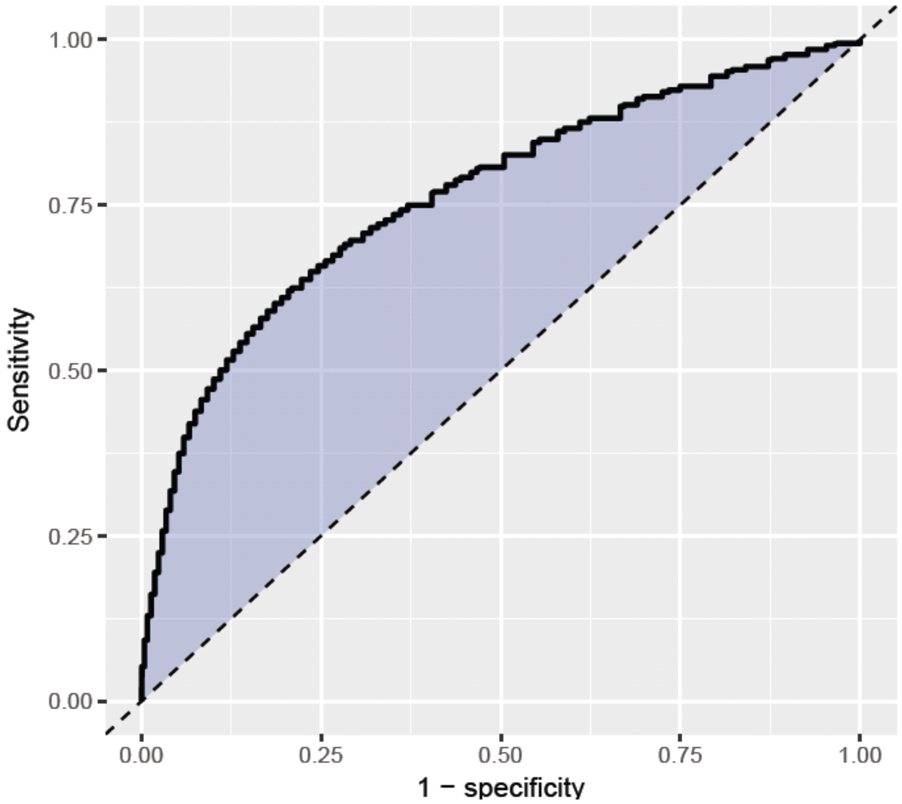
predictionThreshold	preferenceThreshold	positiveCount	negativeCount	trueCount	falseCount	truePositiveCount
0.05655004	0.2147278	45429	1275	8397	38307	8343
0.05655004	0.2147278	45429	1275	8397	38307	8343
0.06047499	0.2269899	45302	1402	8397	38307	8334
0.06151848	0.2302026	44749	1955	8397	38307	8308
0.06446641	0.2391738	43509	3195	8397	38307	8246
0.06446641	0.2391738	43509	3195	8397	38307	8246
0.06459638	0.2395658	42604	4100	8397	38307	8208
0.06459638	0.2395658	42604	4100	8397	38307	8208
0.06686759	0.2463686	42454	4250	8397	38307	8194
0.07021915	0.2562465	41656	5048	8397	38307	8142
0.07312945	0.2646721	41560	5144	8397	38307	8134
0.07330997	0.2651902	40914	5790	8397	38307	8083
0.07355134	0.2658821	39640	7064	8397	38307	8006
0.07355134	0.2658821	39640	7064	8397	38307	8006
0.07355134	0.2658821	39640	7064	8397	38307	8006
0.07368857	0.2662750	38917	7787	8397	38307	7965

This can be used to plot ROC

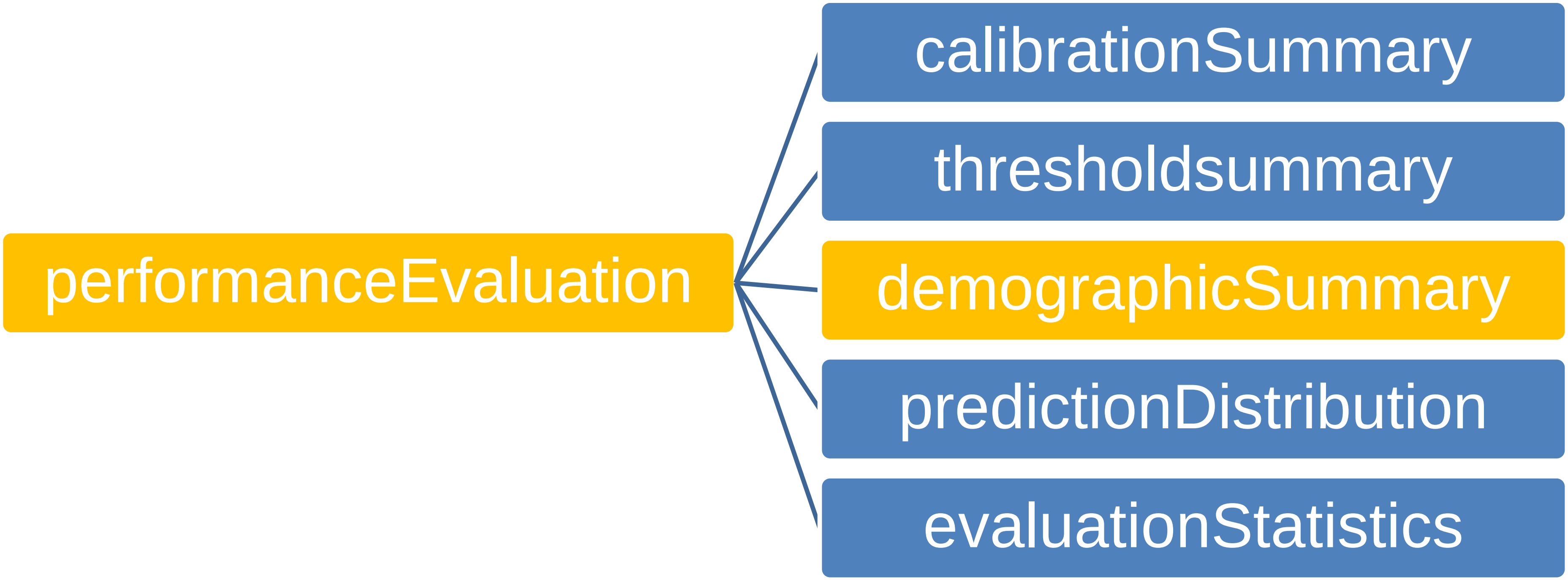
<u>General Performance Chart</u>



<u>Our output for the CHAD2 5 variable model using PLP</u>



Contents of performanceEvaluation



Demographic summary: Stratify the population by age and gender, and compare observed incidence with predicted incidence

demographicSummary

Predicted vs Observed outcome occurrence for each age group and gender

	analysisId	Eval	demographicId	ageId	ageGroup	genId	genGroup	PersonCountAtRisk	PersonCountWithOutcome	averagePredictedProbability
1	20170907211714	train	1	10	Age group: 0-4	8507	Male	NA	NA	NA
2	20170907211714	train	2	11	Age group: 5-9	8507	Male	NA	NA	NA
3	20170907211714	train	3	12	Age group: 10-14	8507	Male	NA	NA	NA
4	20170907211714	train	4	13	Age group: 15-19	8507	Male	5197	792	0.15179358
5	20170907211714	train	5	14	Age group: 20-24	8507	Male	9702	1415	0.14341424
6	20170907211714	train	6	15	Age group: 25-29	8507	Male	5495	742	0.13816323
7	20170907211714	train	7	16	Age group: 30-34	8507	Male	7625	879	0.11606545

In the train set, there was 5179 males ages 15-18 and 792 had the outcome (0.15), the average predicted risk was 0.15 – so the model is well calibrated for this age/gender group!

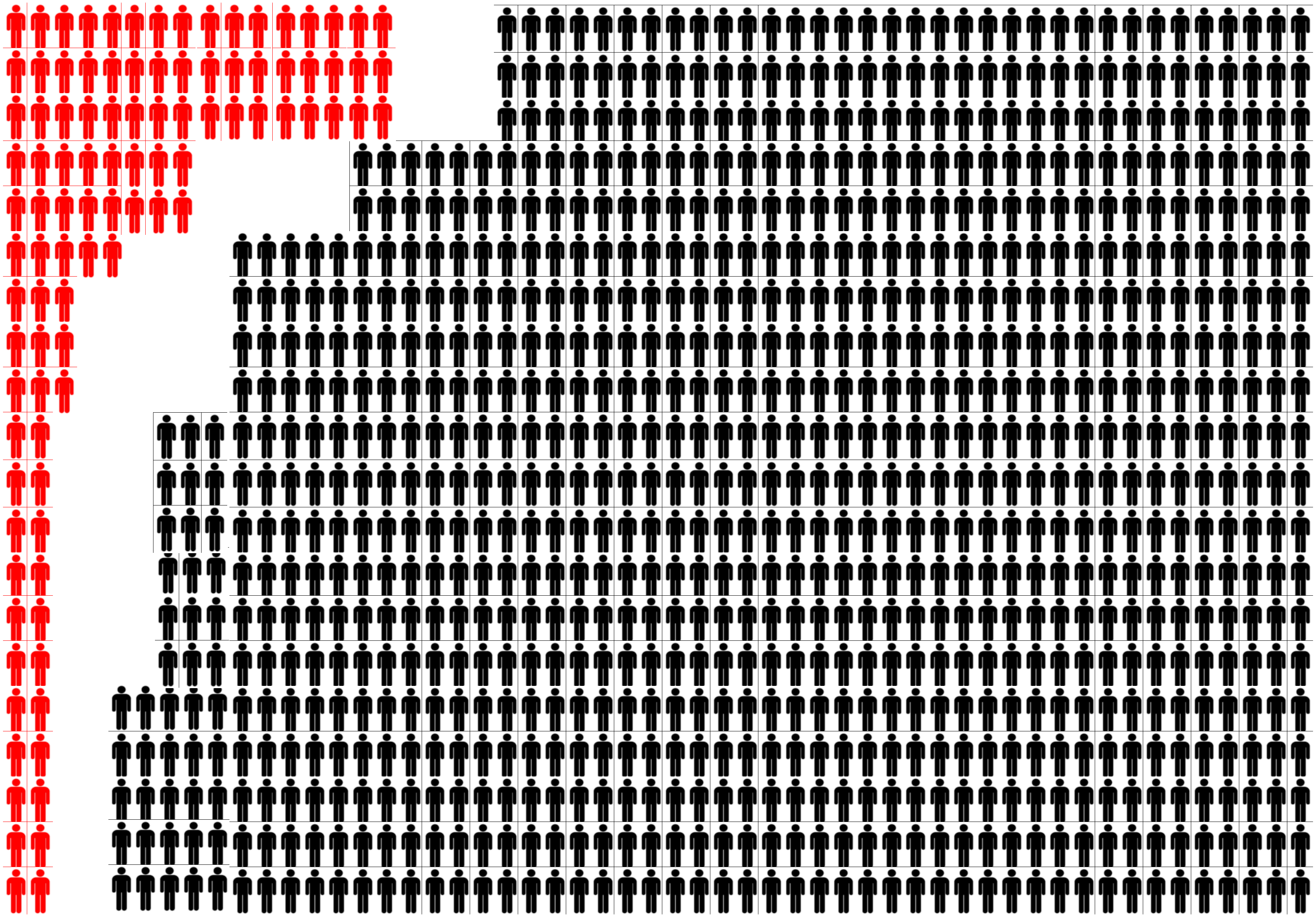
This can be used to plot demographic plot

If the model is calibrate well for each age/gender split then the blue and red lines should be near to each other…

Contents of performanceEvaluation

- performanceEvaluation
  - calibrationSummary
  - thresholdSummary
  - demographicSummary
  - predictionDistribution
  - evaluationStatistics

Prediction distribution: Stratify the population by those with and without outcome, and compare the probability density functions



predictionDistribution

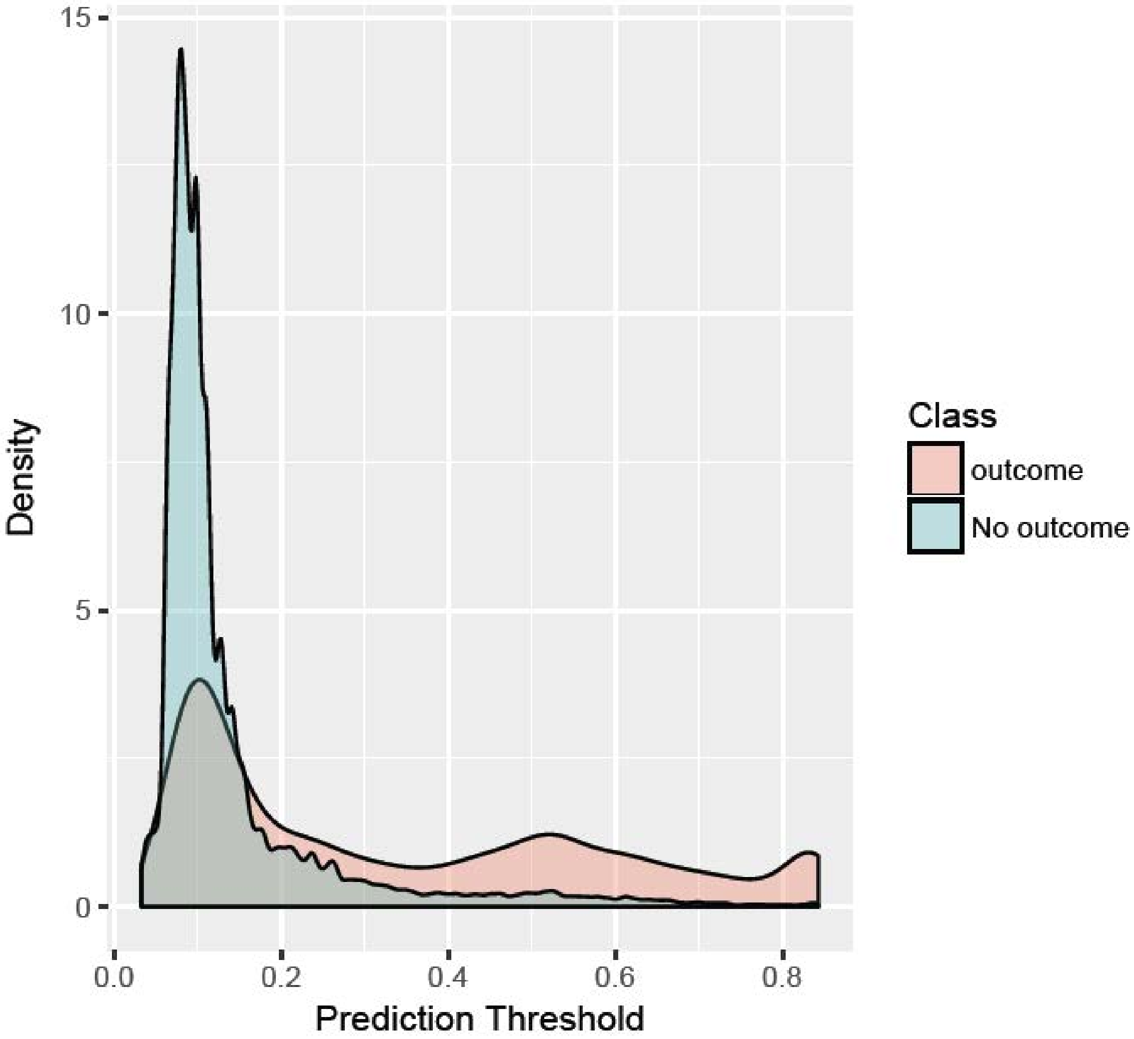
Prediction quantiles for people with and without the outcome (too long to show it all…)

Eval	class	PersonCount	averagePredictedProbability	StDevPredictedProbability	MinPredictedProbability	P05PredictedProbabili
train	0	114920	0.1428148	0.1216245	0.02532292	0.0615184
train	1	25193	0.3485414	0.2607799	0.04756503	0.0736885
test	0	38307	0.1445422	0.1235551	0.03473112	0.0644664
test	1	8397	0.3498708	0.2612192	0.05392288	0.0736885

1 = people with outcome
0 = people without outcome

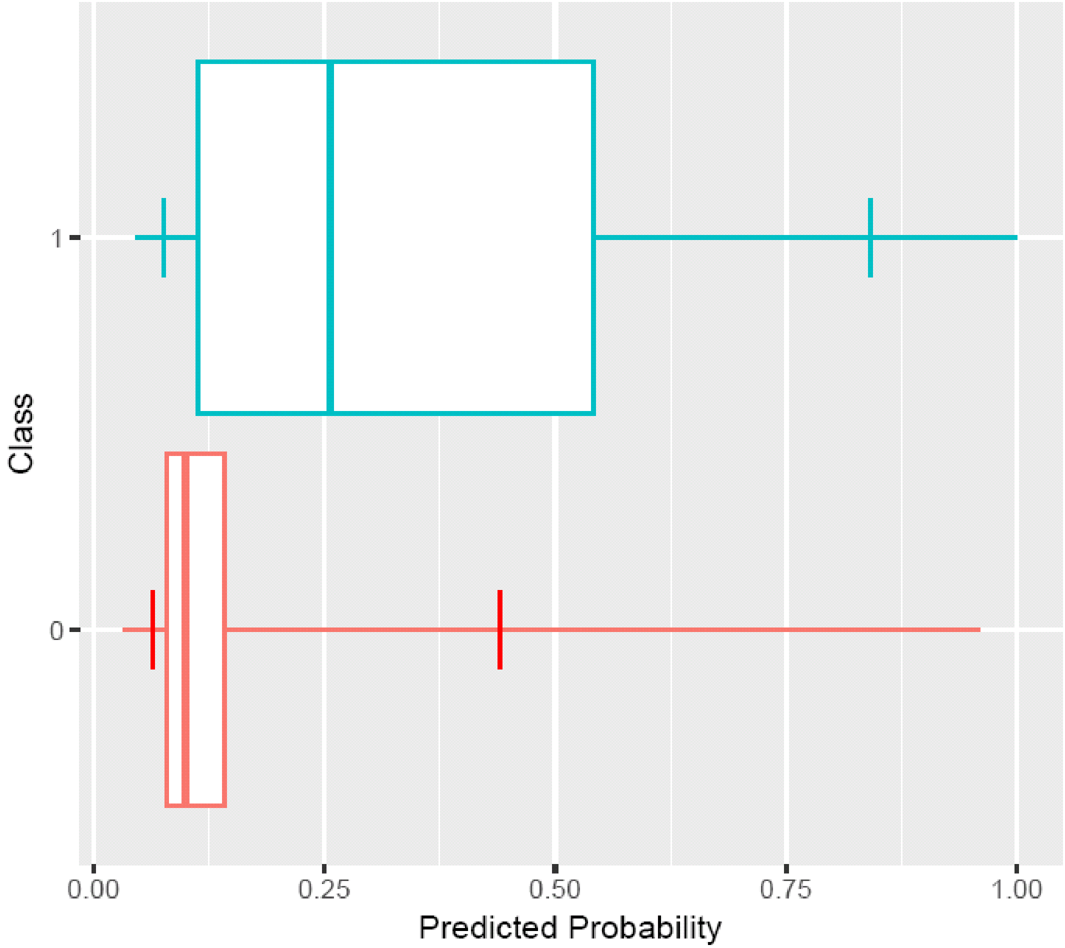
Columns containing: mean, median, 5th percentile, 25th percentile, 75th percentile and 95th percentile

This can be used to plot preference PDF plot



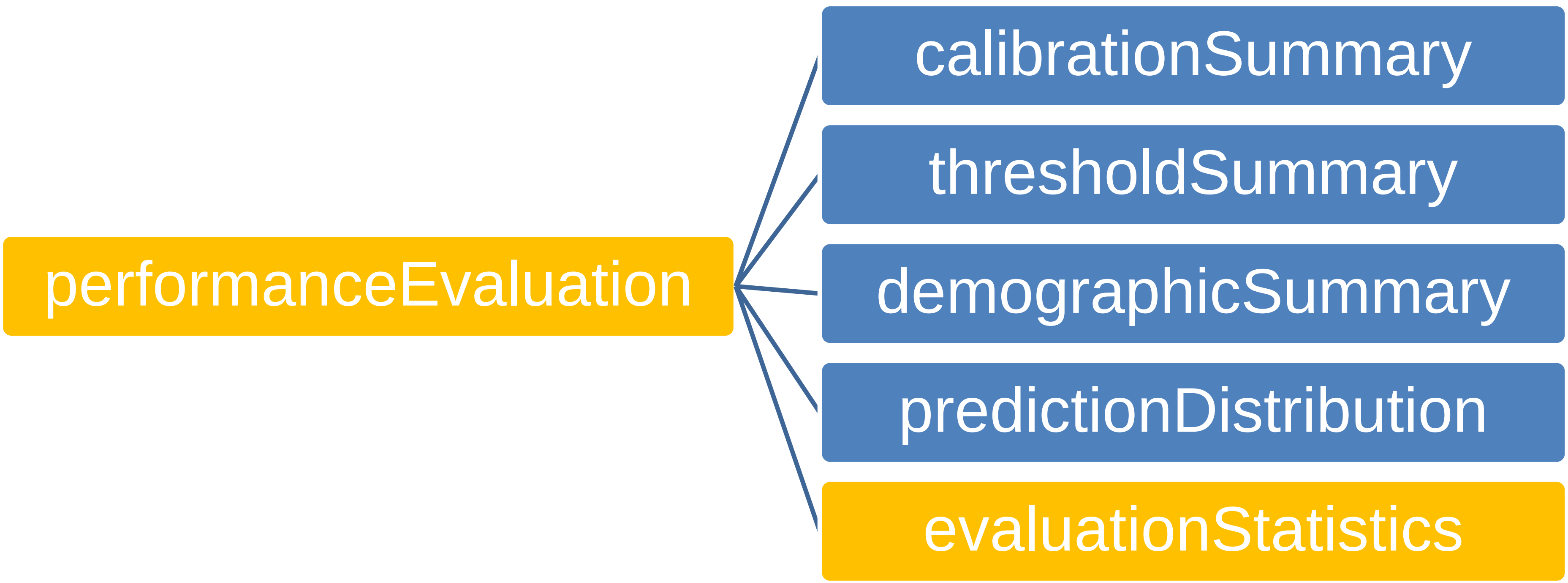
Plot the prediction PDF for the people with the outcome (red) and the people without the outcome (blue)

This can be used to plot prediction distribution plot



This is a summary of the previous plot

Contents of performanceEvaluation



evaluationStatistics

analysisId	Eval	Metric	Value
20170914080427	train	populationSize	140113
20170914080427	train	outcomeCount	25193
20170914080427	train	AUC	0.770716568785114
20170914080427	train	BrierScore	0.11739764180722
27	train	BrierScaled	0.203951252736394
27	train	CalibrationIntercept.Intercept	0.000604616294166257
27	train	CalibrationSlope.Gradient	0.998695289991136
27	test	populationSize	46704
27	test	outcomeCount	8397
27	test	AUC.auc	0.767876247926489
27	test	AUC.auc_lb95ci	0.761865818555656
20170914080427	test	AUC.auc_lb95ci.1	0.773886677297322
20170914080427	test	BrierScore	0.117916015349156
20170914080427	test	BrierScaled	0.206120563897829
20170914080427	test	CalibrationIntercept.Intercept	0.00311558912730724
20170914080427	test	CalibrationSlope.Gradient	0.974553887657516

AUC, brier score and calibration summarised for the train and test sets

Question Break

Any questions about the outputs or visualizations?

Lunch Time

Be back in 45 minutes!

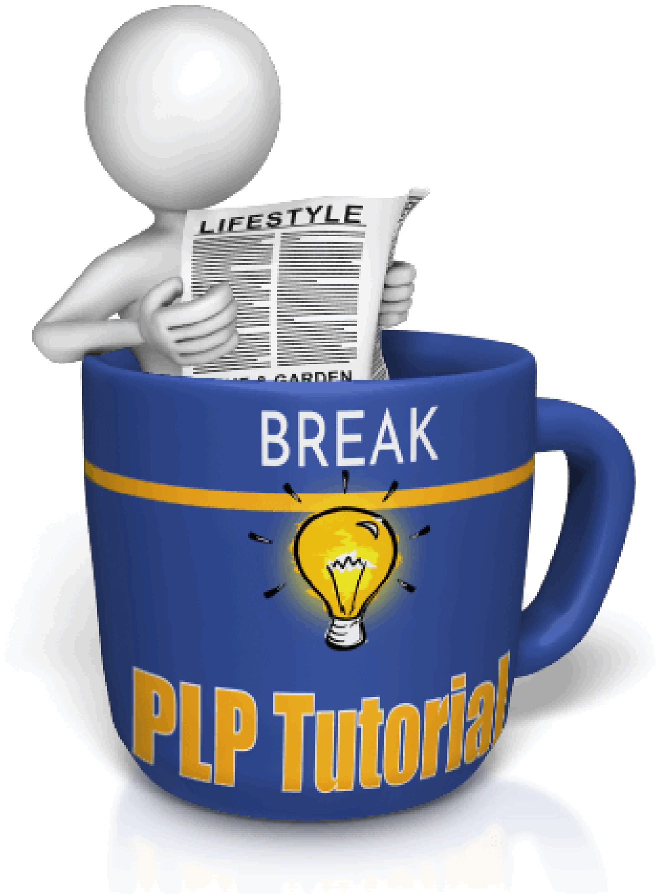
Today's Agenda

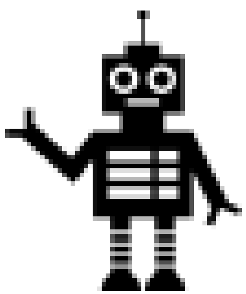
Time	Topic
8:45 – 9:00	Welcome, get settled, get laptops ready
9:00 – 10:30	**Presentation:** What is Patient-Level Prediction?
10:30 – 10:45	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 1 Theory
10:45 – 11:45	**Presentation:** Overview of the TRIPOD Statement **Exercise:** Applying TRIPOD to CHADS2
11:45 – 12:30	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 2 Implementation
12:30 – 13:15	Lunch
13:15 – 14:30	**Exercise:** Guided tour through implementing patient-level prediction
14:30 – 14:45	Break
14:45 – 16:30	**Exercise:** Design and implement your own patient-level prediction
16:30 – 17:00	Lessons learned and feedback

Exercise:

Guided tour through implementing patient-level prediction

Let's take a 15 min break





Task (Modified CHADS2 model)

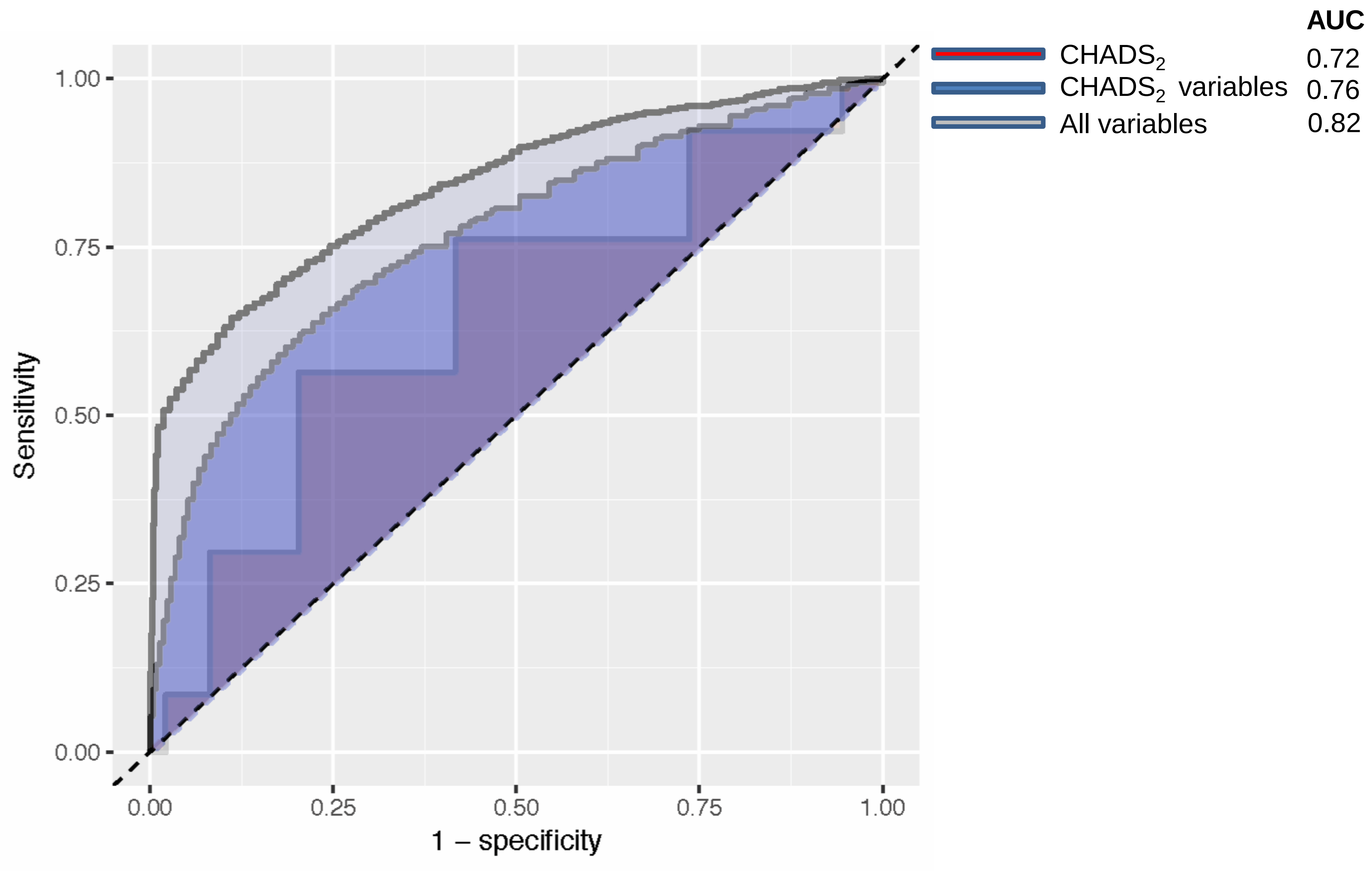
In target population (**PLP training: T : patients newly diagnosed with Atrial fibrillation)** predict who will develop outcome (**PLP training: O - hospitalized ischemic stroke events)** during the period from 0 days from cohort start date to 1000 days.

Example

- We implemented three models in OPTUM for the prediction problem:

1. CHAD2 model
2. PLP model using 5 CHAD2 variables (and descendants)
3. PLP model using all variables

Predicting Stroke in Patients with Atrial Fibrillation: OPTUM results



Today’s Agenda

Time	Topic
8:45 – 9:00	Welcome, get settled, get laptops ready
9:00 – 10:30	**Presentation:** What is Patient-Level Prediction?
10:30 – 10:45	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 1 Theory
10:45 – 11:45	**Presentation:** Overview of the TRIPOD Statement **Exercise:** Applying TRIPOD to CHADS2
11:45 – 12:30	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 2 Implementation
12:30 – 13:15	Lunch
13:15 – 14:30	**Exercise:** Guided tour through implementing patient-level prediction
14:30 – 14:45	Break
14:45 – 16:30	**Exercise:** Design and implement your own patient-level prediction
16:30 – 17:00	Lessons learned and feedback

Exercise:

Design and implement your own patient-level prediction

Exercise

1. Fill in the form to describe your study (15 min)
2. Discuss your study in a group of 4 people (30 min)
3. Select a study to work on with your team
4. Report your progress to the group

Today's Agenda

Time	Topic
8:45 – 9:00	Welcome, get settled, get laptops ready
9:00 – 10:30	**Presentation:** What is Patient-Level Prediction?
10:30 – 10:45	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 1 Theory
10:45 – 11:45	**Presentation:** Overview of the TRIPOD Statement **Exercise:** Applying TRIPOD to CHADS2
11:45 – 12:30	**Presentation:** Learning the OHDSI Patient-Level Prediction Framework - Part 2 Implementation
12:30 – 13:15	Lunch
13:15 – 14:30	**Exercise:** Guided tour through implementing patient-level prediction
14:30 – 14:45	Break
14:45 – 16:30	**Exercise:** Design and implement your own patient-level prediction
16:30 – 17:00	Lessons learned and feedback

Lessons learned and feedback

Peter Rijnbeek[1]
Jenna Reps[2]
Joel Swerdel[2]

1. Department of Medical Informatics, Erasmus MC
2. Janssen Research & Development, LLC

Lessons Learned

Learned the PLP Dance

Educated Fortune Teller

What's next

When you write your JAMA publication;

1. Follow the TRIPOD Statement.
2. Cite our work:

To cite Cyclops in publications use:

Suchard MA, Simpson SE, Zorych I, Ryan P and Madigan D (2013). "Massive parallelization of serial inference algorithms for complex generalized linear models:' *ACM Transactions on Modeling and Computer Simulation, 23,* pp. 10. link

To cite PatientLevelPrediction in publications use:

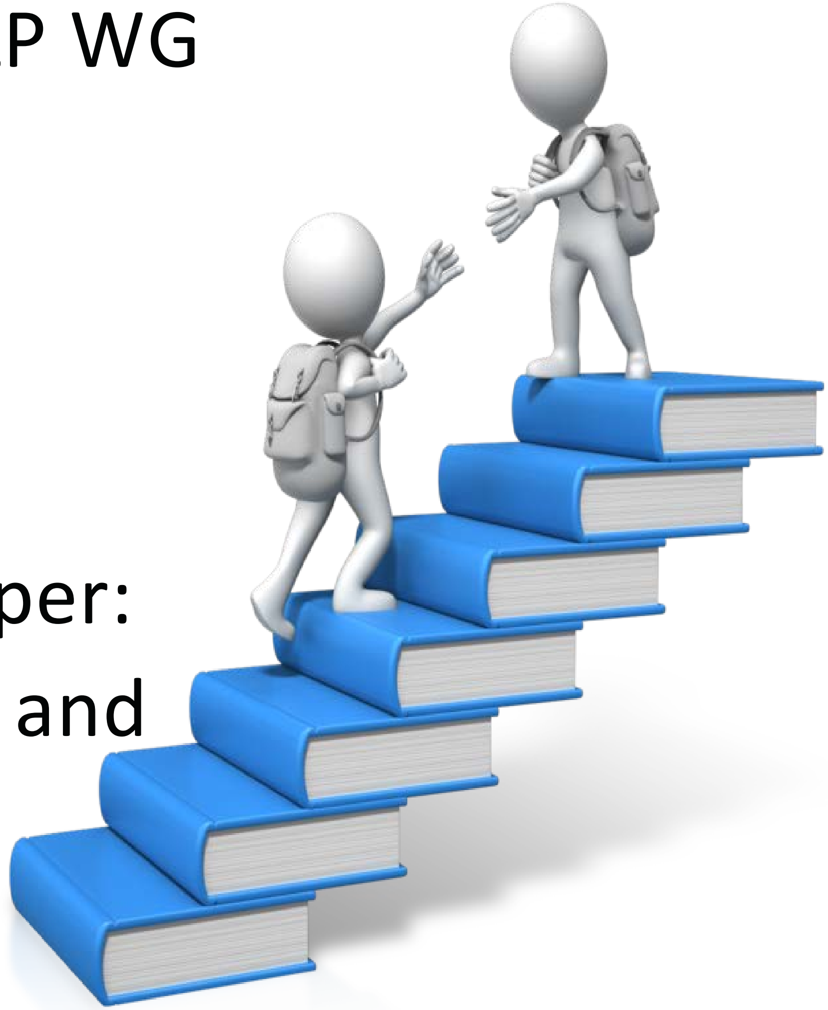
Jenna Reps, Martijn J. Schuemie MJ, Marc A. Suchard, Patrick

B. Ryan P, Peter R. Rijnbeek (2017).

"PatientLevelPrediction: Package for Patient-Level Prediction using data in the OMOP Common Data Model. R package version 1.2.2.

Join the PLP Community

- Bi-weekly meetings of PLP WG
- Researchers Forum (tag patientprediction)
- Become an active developer: add your own algorithms and other features



Continuation of the PLP Journey

Scale up

- Increase the number of database
- Increase the number of cohorts at risk
- Increase the number of outcomes

Method Research

- Performance
- Speed
- Transportability
- Temporal information
- Textual information
- …

Clinical impact for the patient

- How to assess?

Tutorial improvement

We like to hear your feedback on this course:

- What went well?
- What did not?
- What do you like to see added?
- You can give your feedback on the evaluation form.

Questions? Drop us an email

p.rijnbeek@erasmusmc.nl
jreps@its.jnj.com