Predicting risk from multiple Observational Health Data Sciences and Informatics (OHDSI) databases

George Hripcsak, MD, MS
Columbia University Irving Medical Center
NewYork-Presbyterian Hospital
Peter Rijnbeek, PhD, Erasmus
Jenna Reps, PhD, Janssen

Observational Health Data Sciences and Informatics (OHDSI, as “Odyssey”)

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

A multi-stakeholder, interdisciplinary, international collaborative with a coordinating center at Columbia University

OHDSI's global research community

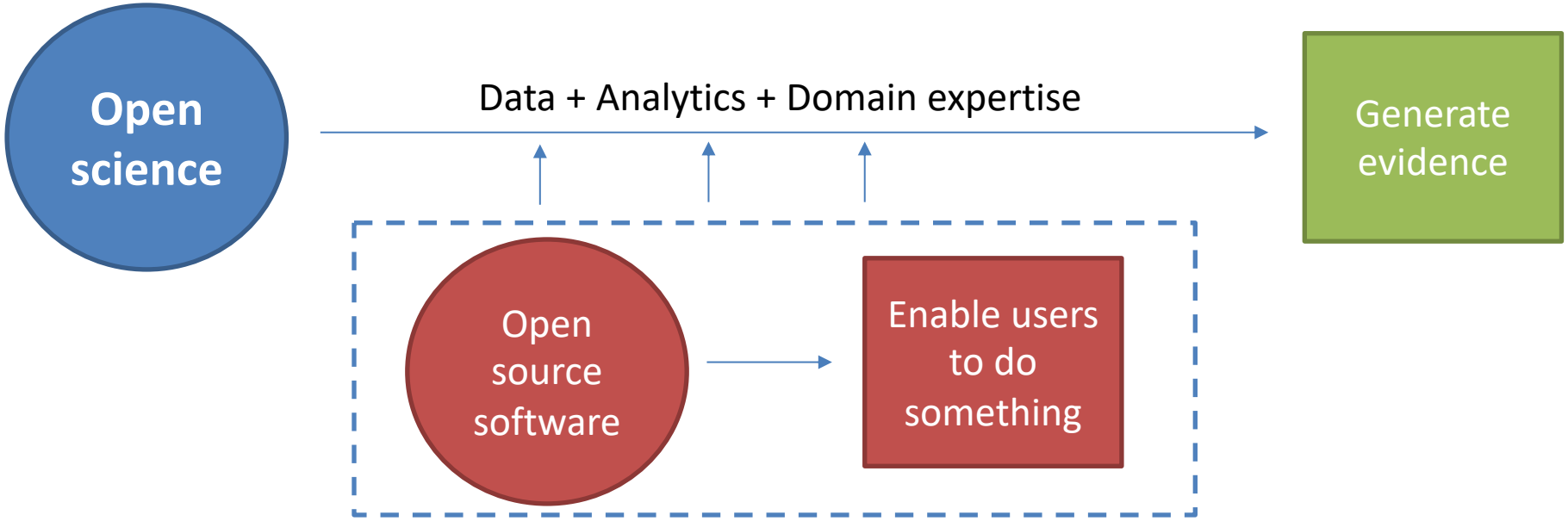
- >200 collaborators from 25 different countries
- Experts in informatics, statistics, epidemiology, clinical sciences
- Active participation from academia, government, industry, providers
- Currently records on about 500 million unique patients in >100 databases

http://ohdsi.org/who-we-are/collaborators/

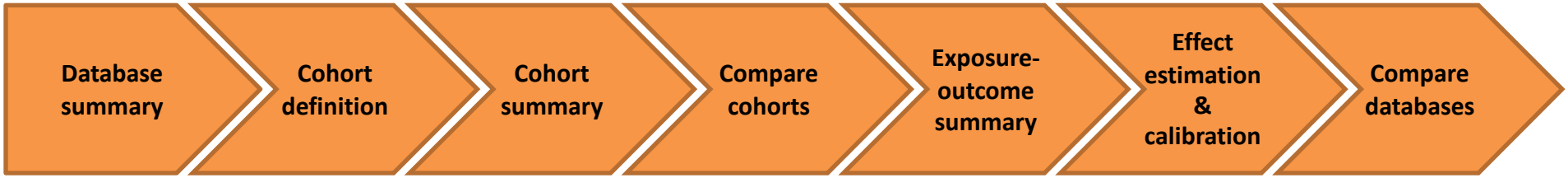
Evidence OHDSI seeks to generate from observational data

- **Clinical characterization - tally**
  - Natural history: Who has diabetes, and who takes metformin?
  - Quality improvement: What proportion of patients with diabetes experience complications?
- **Population-level estimation - cause**
  - Safety surveillance: Does metformin cause lactic acidosis?
  - Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?
- **Patient-level prediction - predict**
  - Precision medicine: Given everything you know about me, if I take metformin, what is the chance I will get lactic acidosis?
  - Disease interception: Given everything you know about me, what is the chance I will develop diabetes?

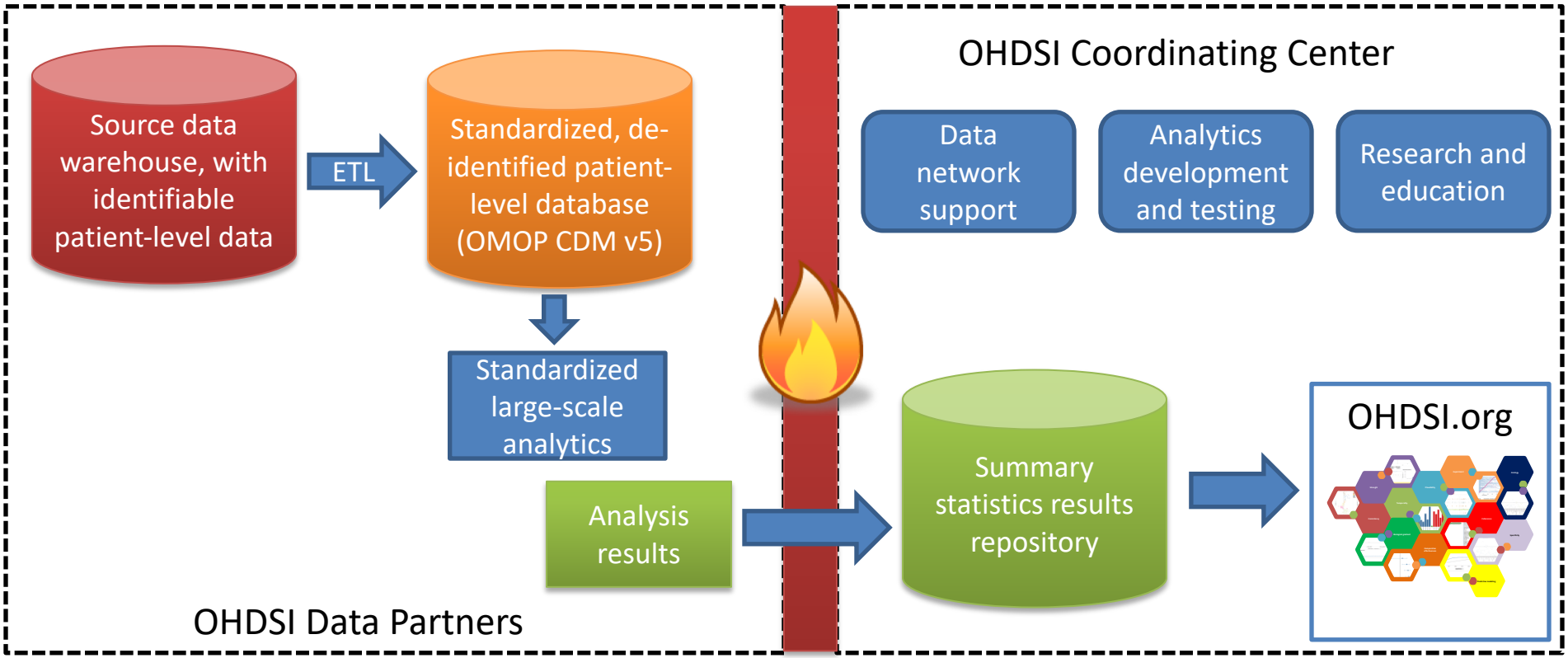
Open Science



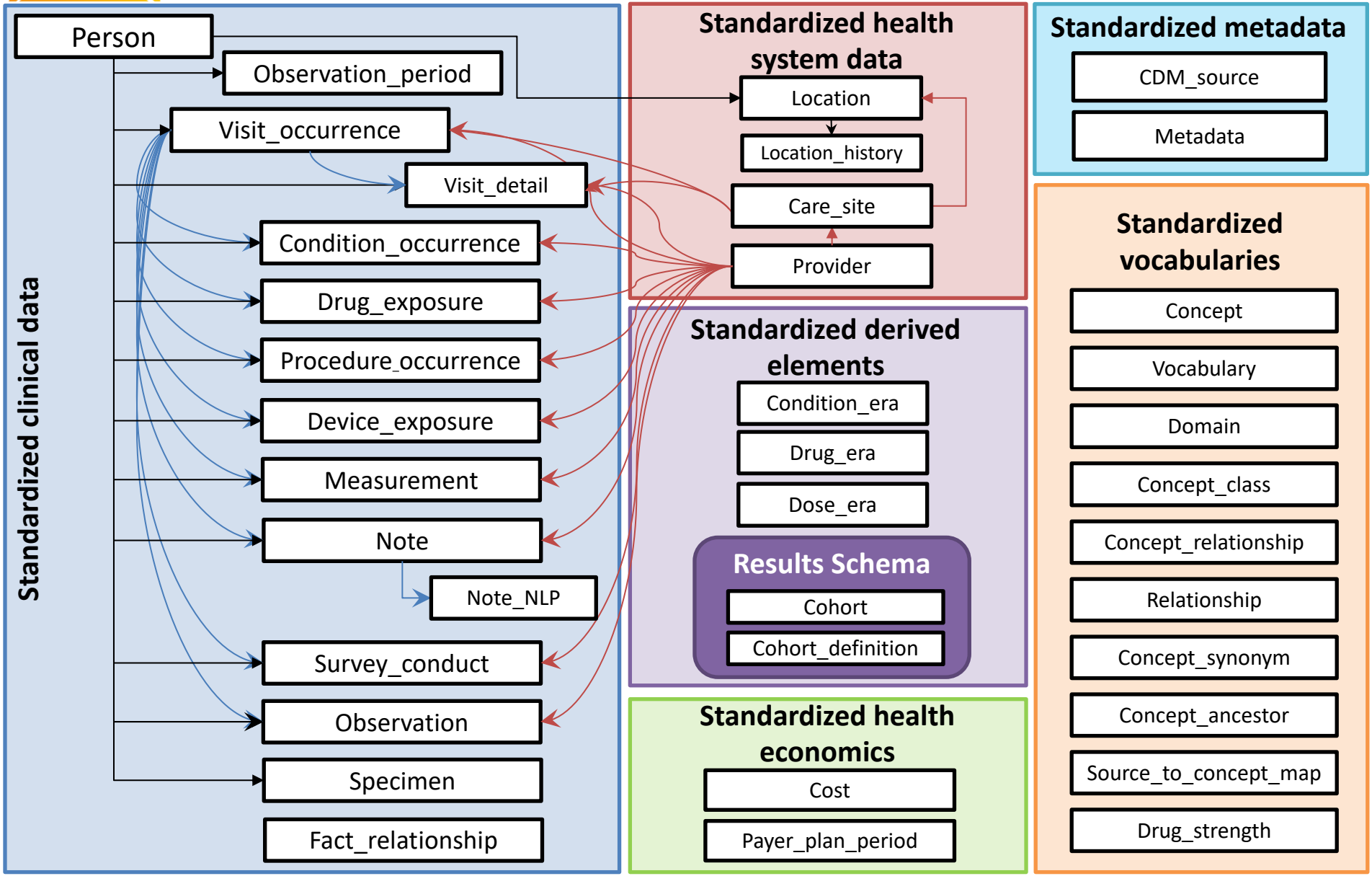
Standardized, transparent workflows



How OHDSI Works

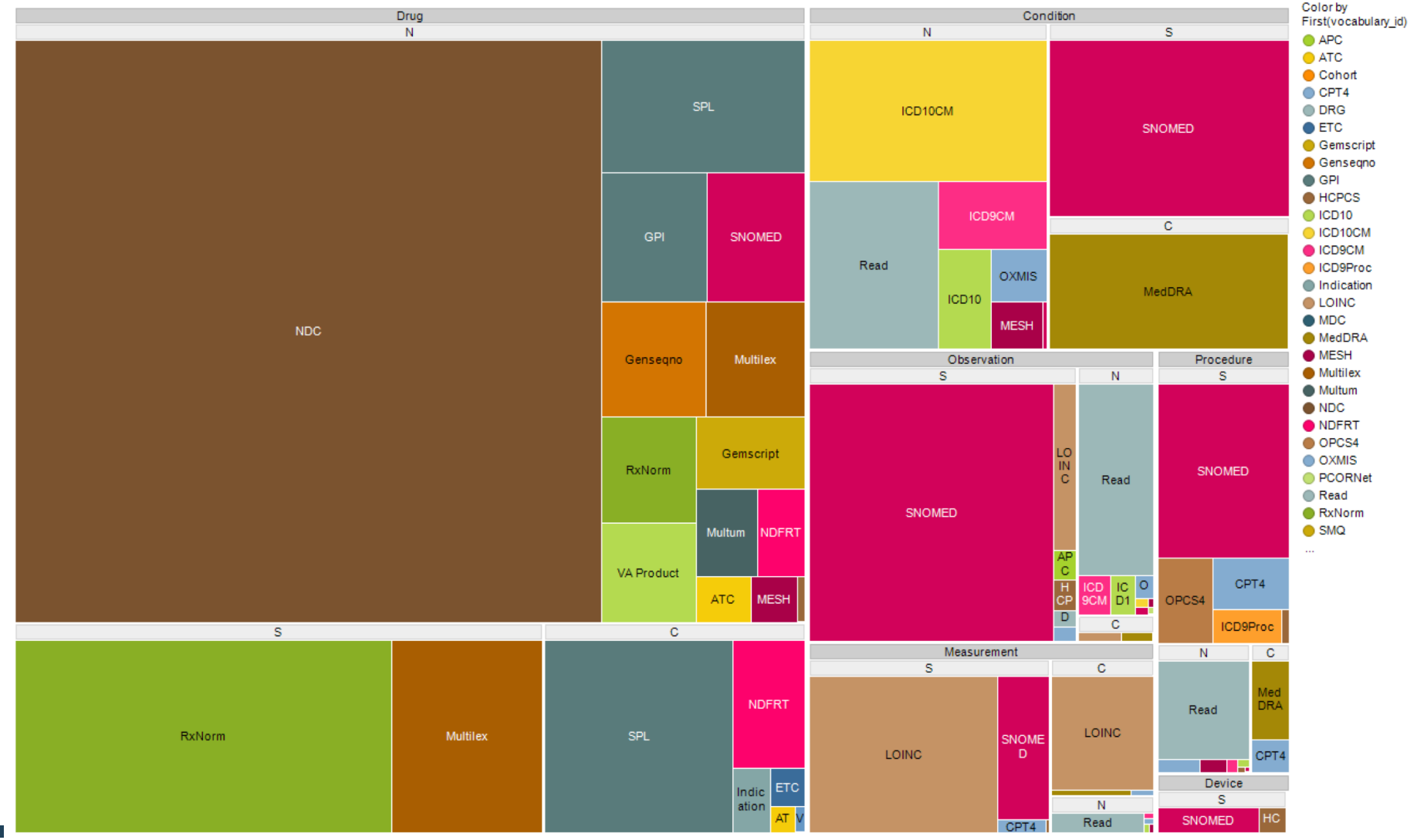


Deep information model
OMOP CDM Version 6

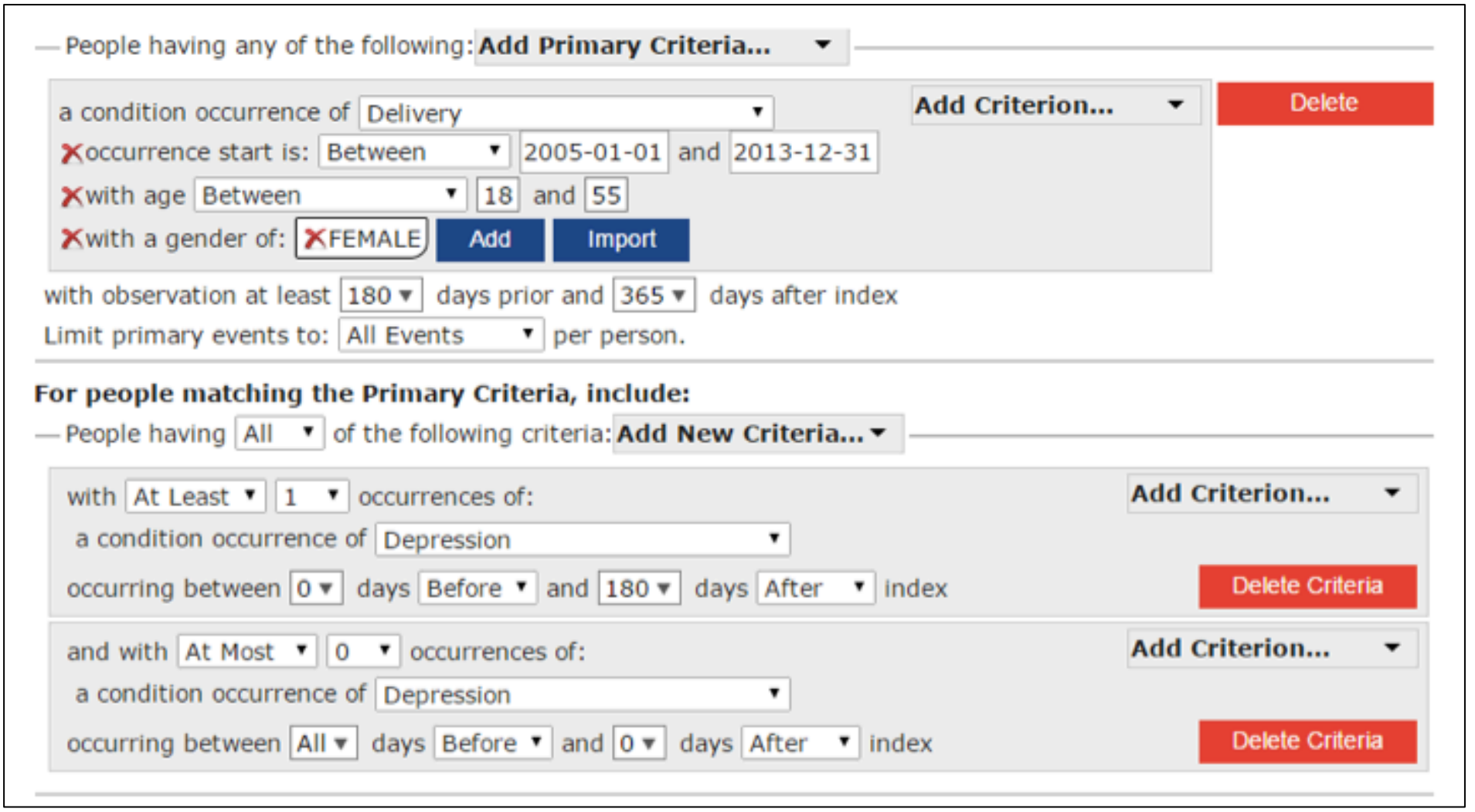


Extensive vocabularies

Breakdown of OHDSI concepts by domain, standard class, and vocabulary

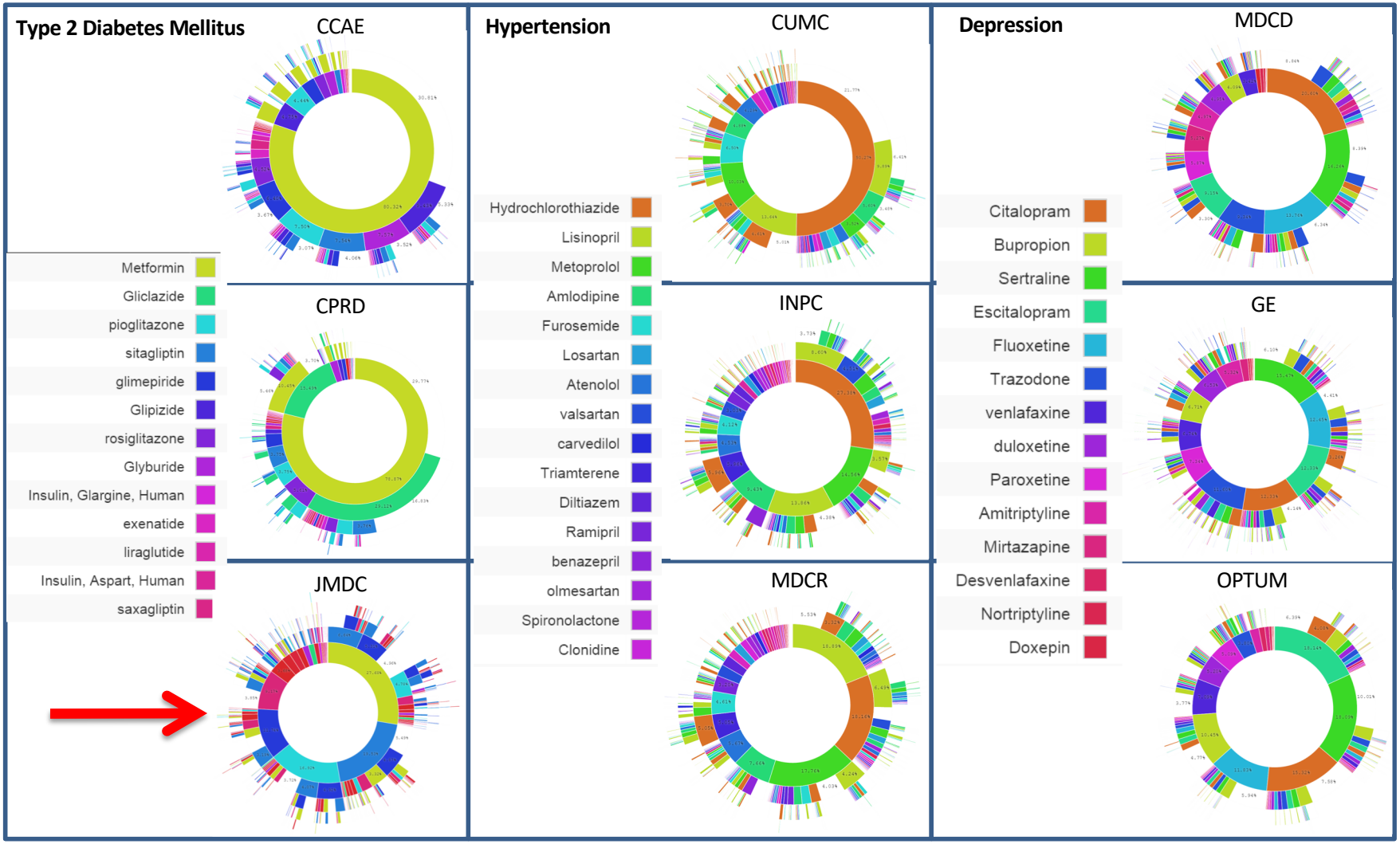


ATLAS to build, visualize, and analyze cohorts



OHDSI in Action

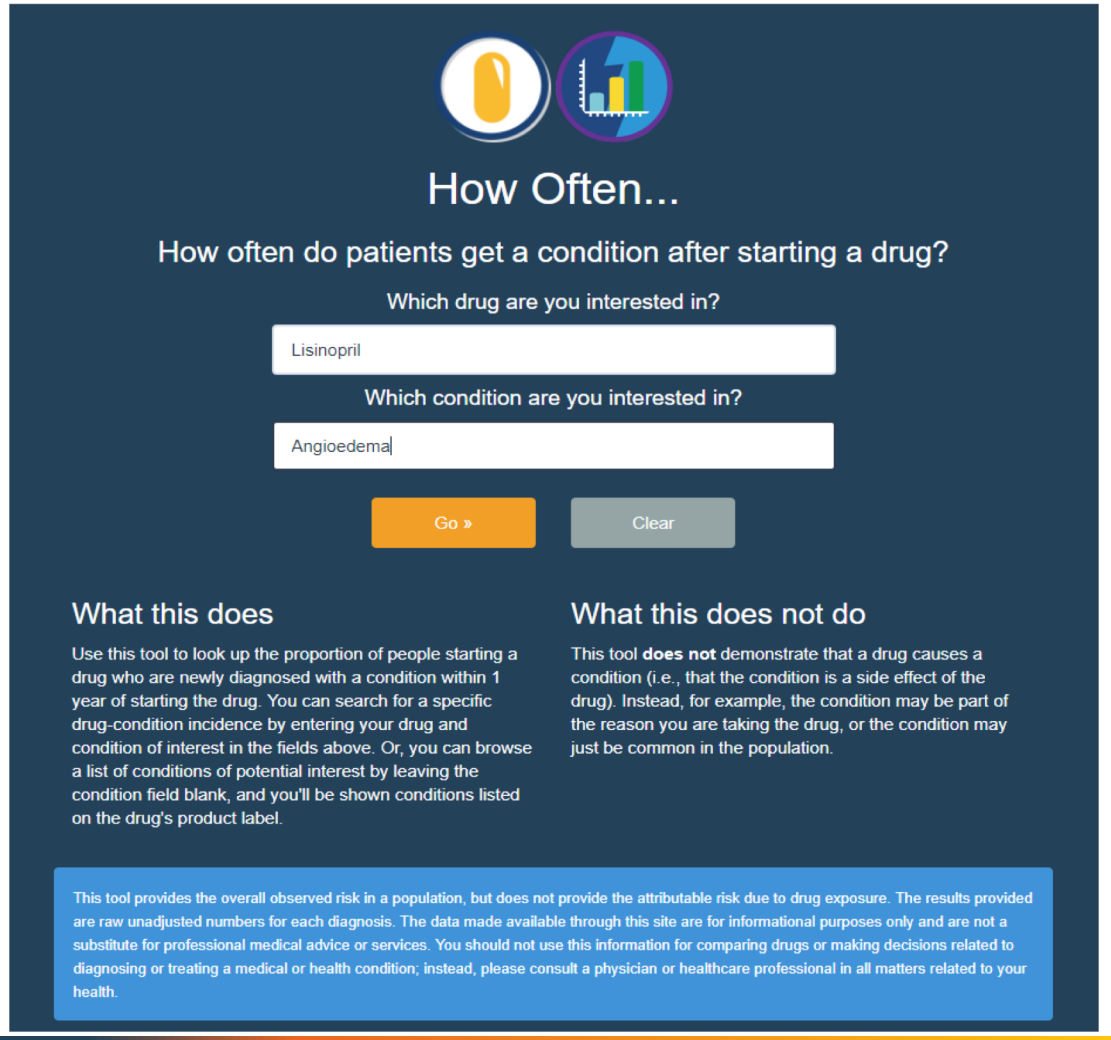
Population-level heterogeneity across systems, and patient-level heterogeneity within systems



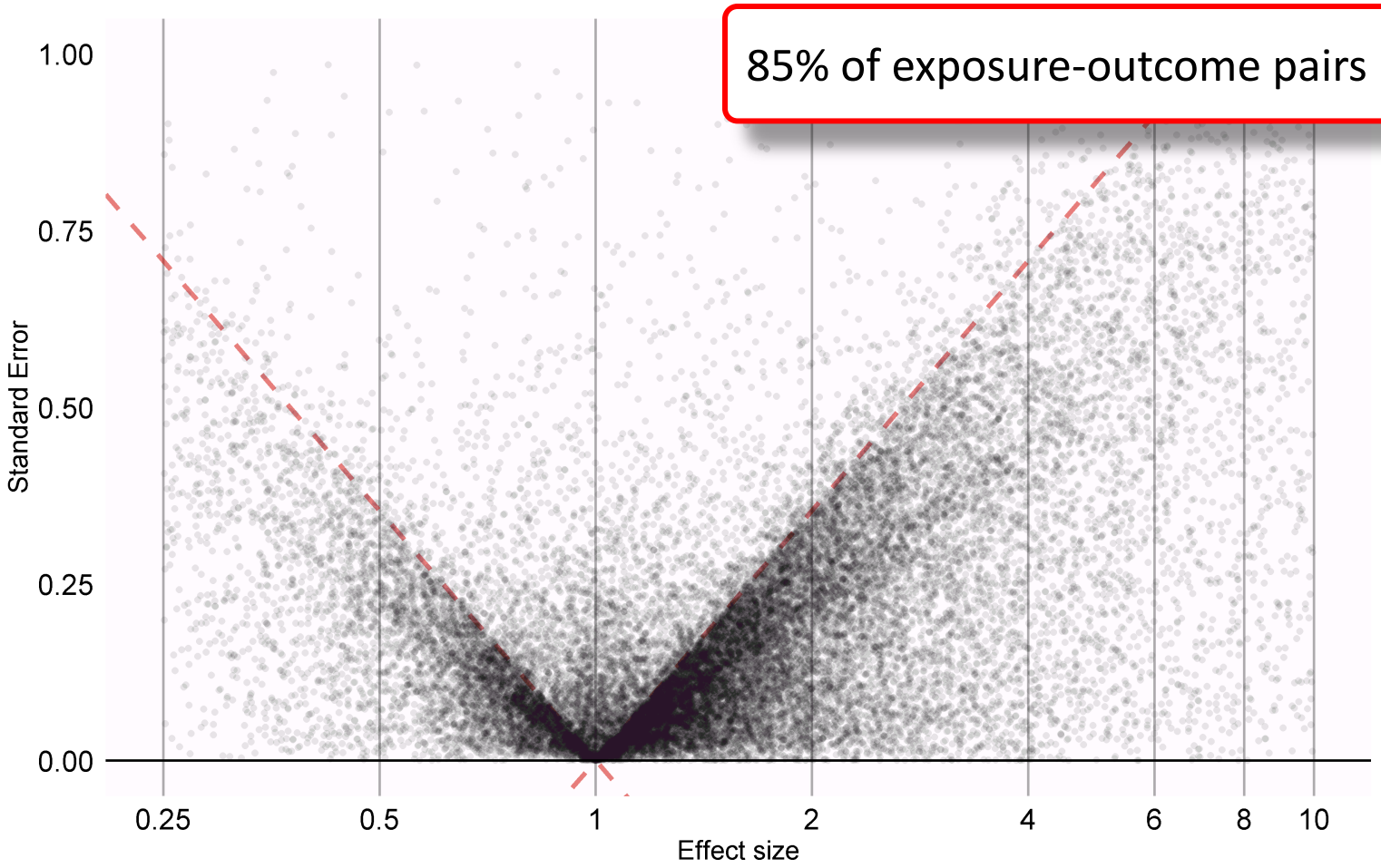
PNAS 2016

howoften.org

- Incidence of side effects
- Any drug on the world market
- Any condition
- Absolute risk
  - Not causal (Characterization)
- On the Internet



Observational research results in literature

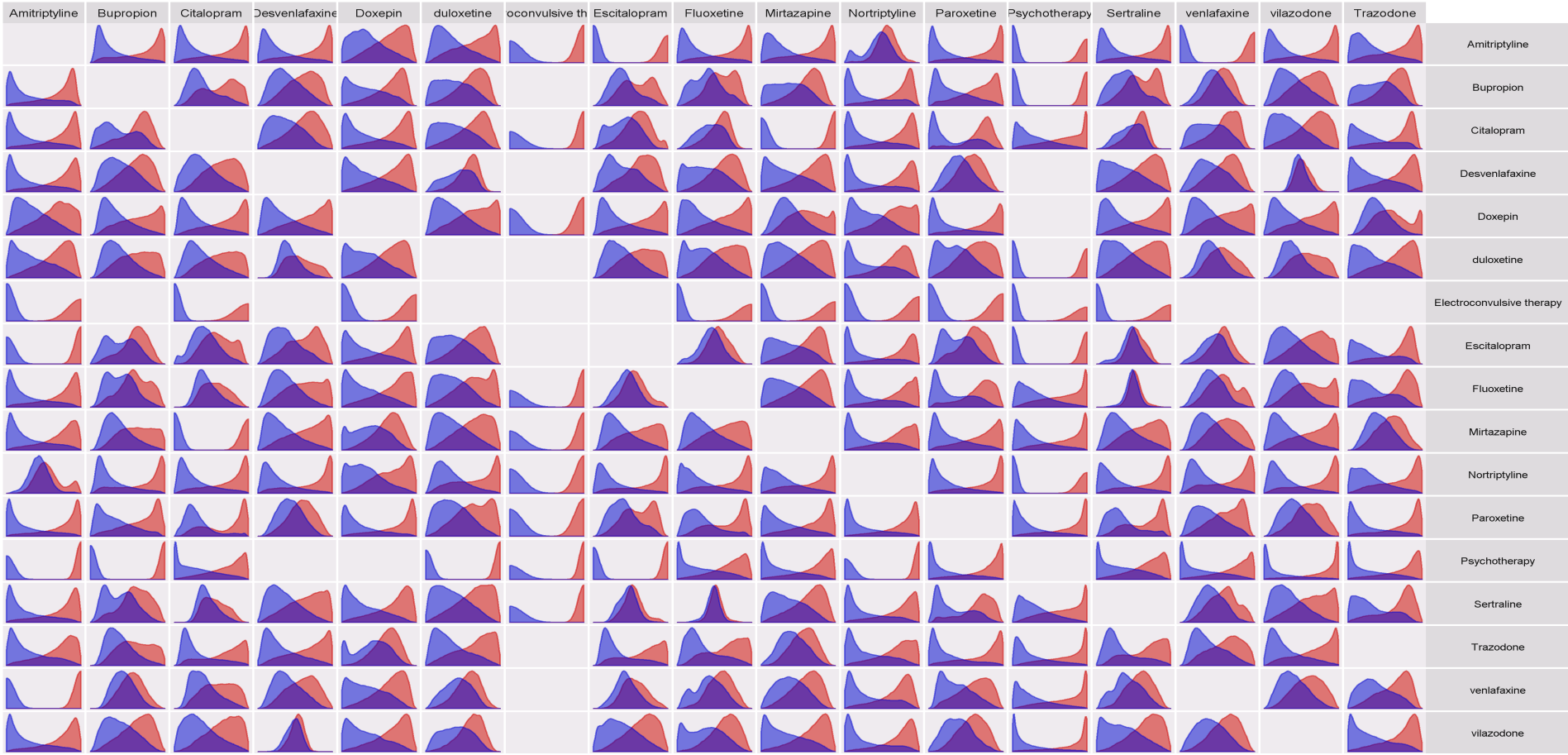


85% of exposure-outcome pairs have p < 0.05

29,982 estimates
11,758 papers

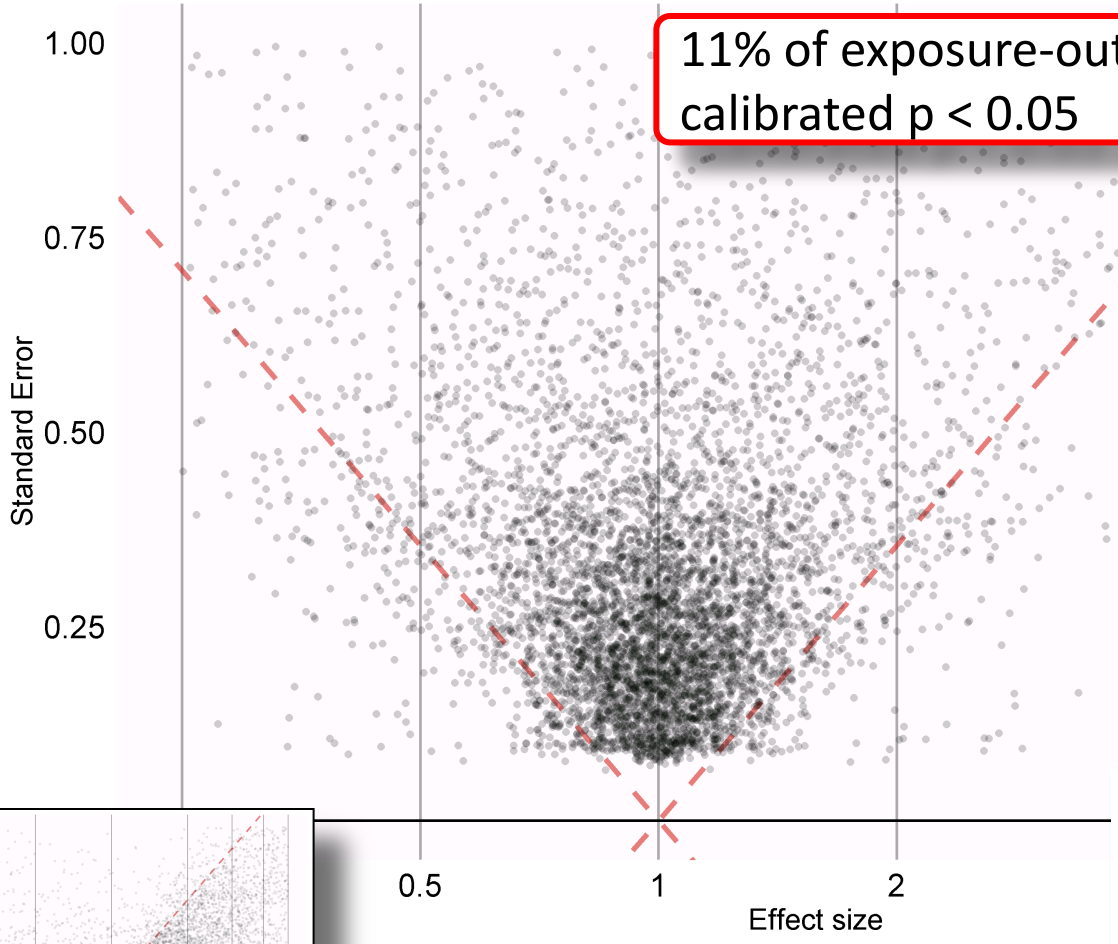
Addressing reproducibility

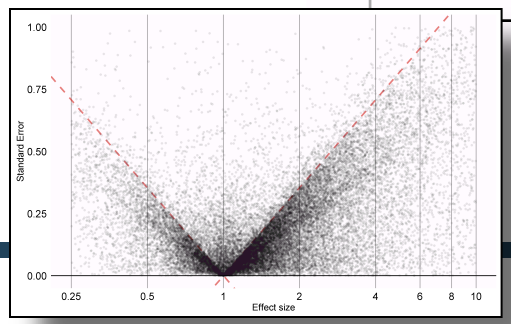
Carry out on aligned hypotheses at scale



Estimates are in line with expectations

11% of exposure-outcome pairs have calibrated p < 0.05





PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Cite this article: Schuemie MJ, Ryan PB, Hripcsak G, Madigan D, Suchard MA. 2018 Improving reproducibility by using high-throughput observational studies with empirical calibration. *Phil. Trans. R. Soc. A* **376**: 20170356.
http://dx.doi.org/10.1098/rsta.2017.0356

Accepted: 8 May 2018

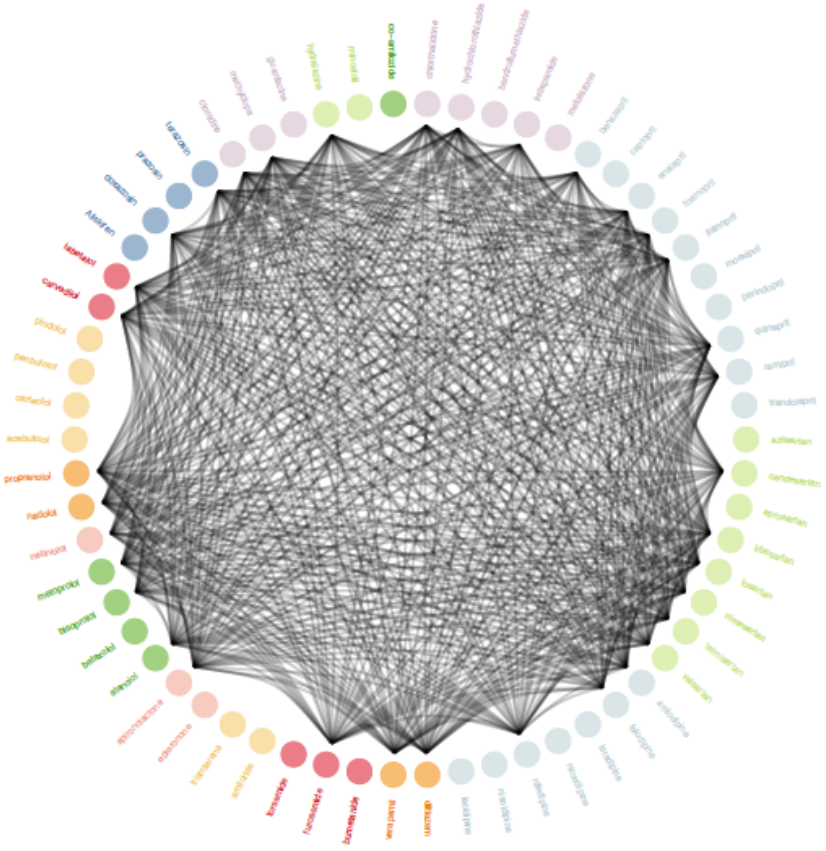
Improving reproducibility by using high-throughput observational studies with empirical calibration

Martijn J. Schuemie[1,2], Patrick B. Ryan[1,2,3], George Hripcsak[1,3,4], David Madigan[1,5] and Marc A. Suchard[1,6,7,8]

[1]Observational Health Data Sciences and Informatics (OHDSI), New York, NY 10032, USA
[2]Epidemiology Analytics, Janssen Research and Development, Titusville, NJ 08560, USA
[3]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032, USA

LEGEND knowledge base for hypertension

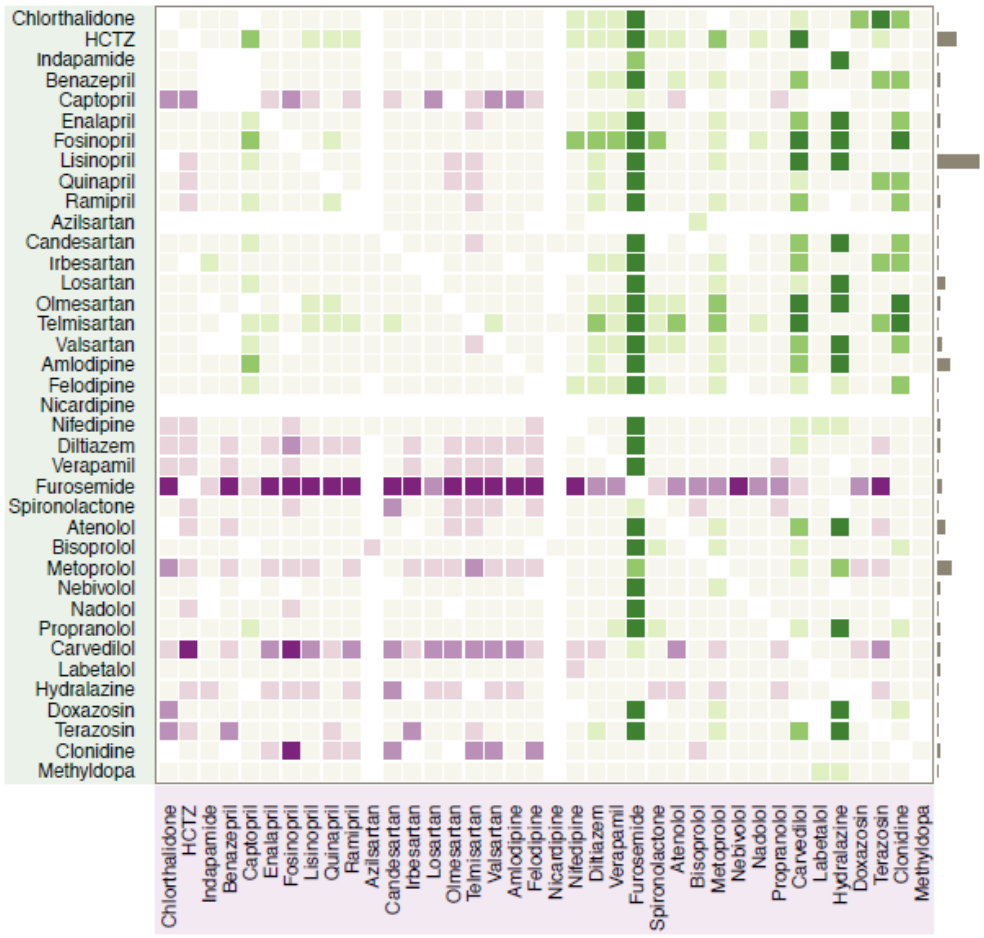
Head-to-head HTN drug comparisons



- Trials: 40
- $N = 102 - [1148] - 33K$

- Comparisons: 10,278
- $N = 3502 - [212K] - 1.9M$

Cardiovascular efficacy by drug



Composite (MI, HF, stroke) outcome in meta-analysis

Prescriptions are not written at the class-level; must choose an individual drug for the patient

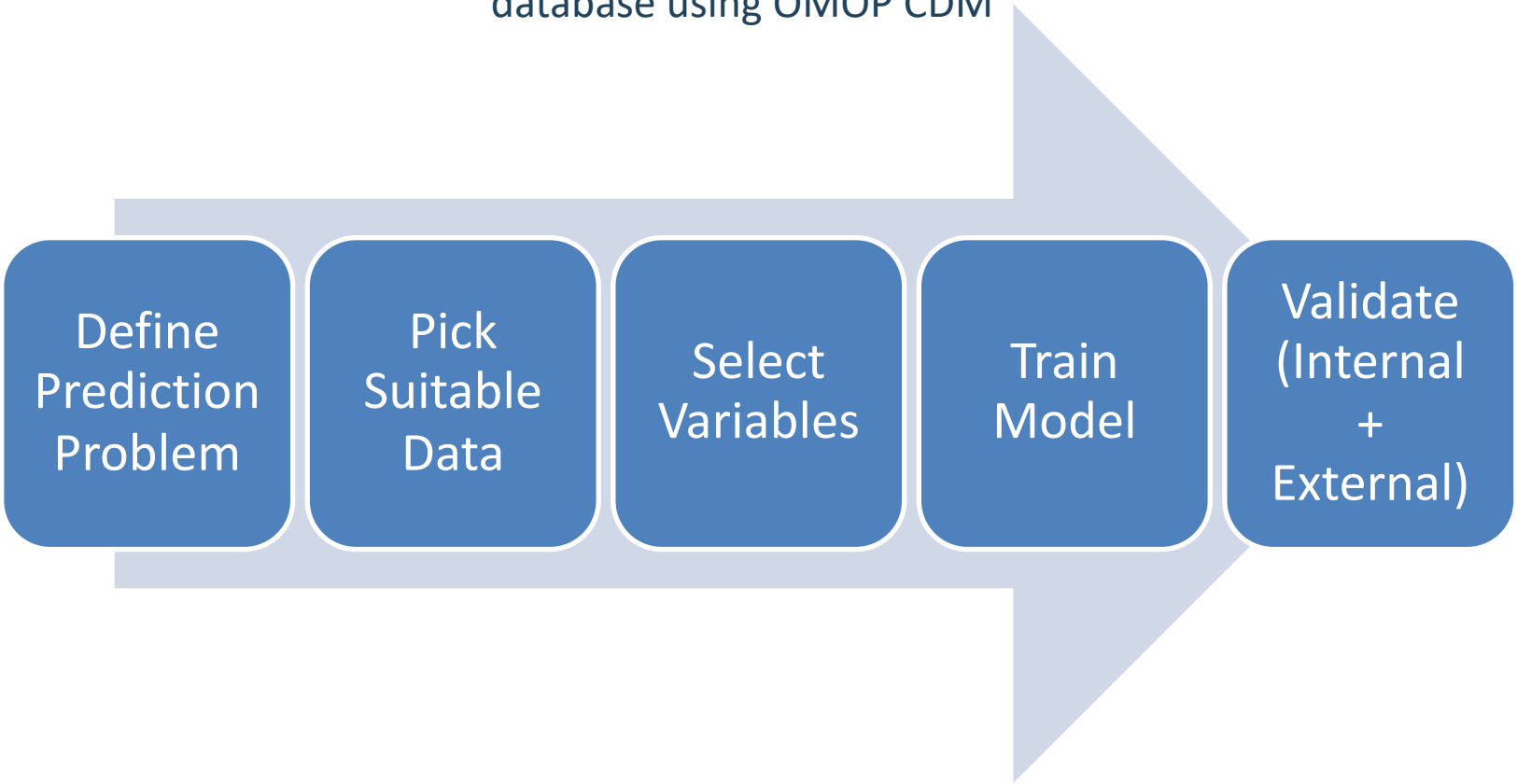
- 1st-line > 2nd-line
- Some within-class differences failed diagnostics, e.g. captopril

OHDSI in Action

- Patient-level prediction

An OHDSI to Patient-Level Prediction

OHDSI established a 5-step standardized framework for developing and evaluating patient-level prediction models, and has released an open-source R package (PatientLevelPrediction) to implement the framework against any observational database using OMOP CDM



Types of prediction problems in healthcare

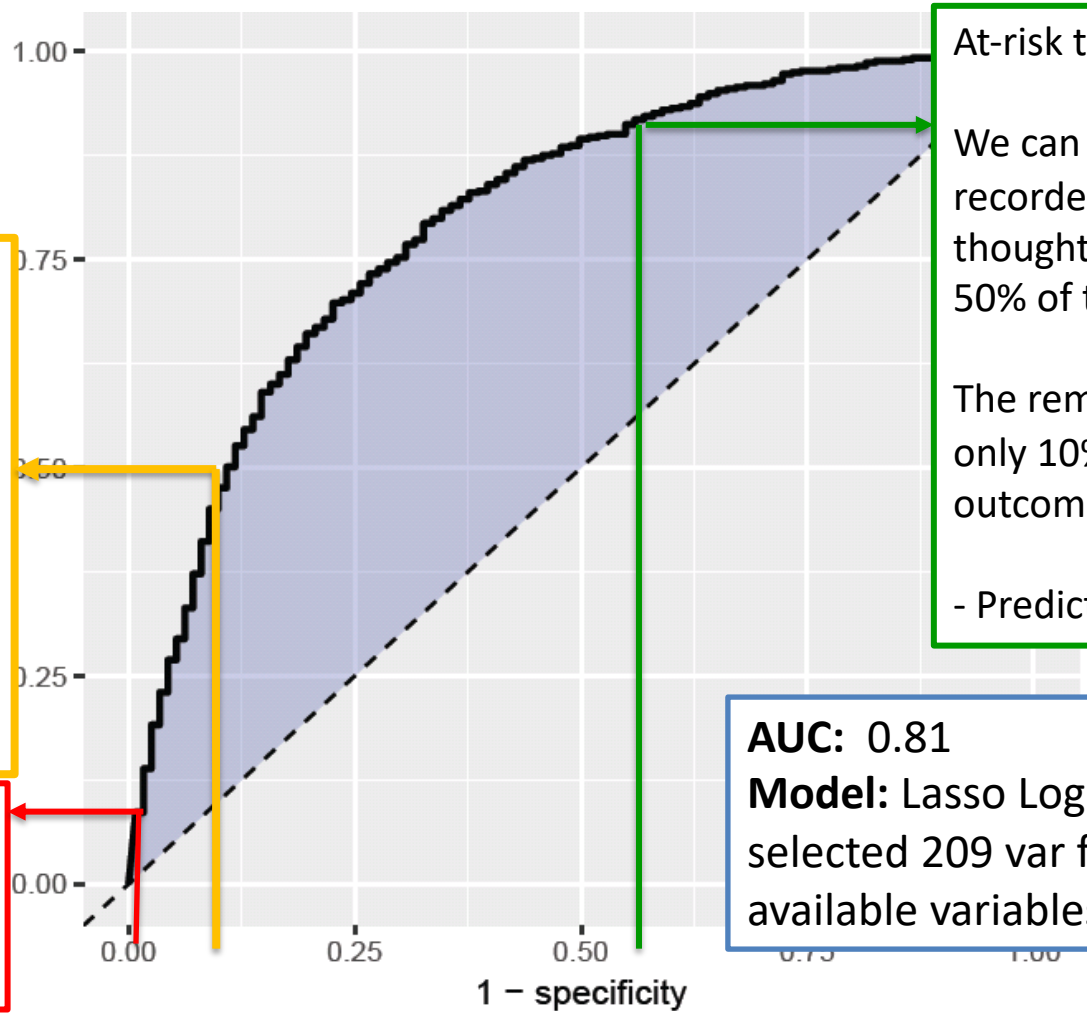
Amongst **<insert your target population>,** which patients will experience **<insert your outcome>** within **<time at risk>**?

Type	Structure	Example
Disease onset and progression	Amongst patients who are newly diagnosed with **<insert disease>**, which patients will go on to have **<another disease or related complication>** within **<time horizon from diagnosis>**?	Among newly diagnosed **depression** patients, which will go onto to have **suicide** in **next 1 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 **MDD patients who took either sertraline or bupropion**, which patients got **sertraline**? (as defined for propensity score model)
Treatment response	Amongst patients who are new users of **<insert chronically-used drug>**, which patients will **<insert desired effect>** in **<time window>**?	Which patients with **depression who start on sertraline do not require a different antidepressant after 1 years**?
Treatment safety	Amongst patients who are new users of **<insert drug>,** which patients will experience **<insert potential adverse event of the drug>** within **<time horizon following**	Among new users of **sertraline**, which patients will have **sexual dysfunction** in **1 year**?

Note: If you want to determine if a variable **causes** the outcome (e.g., a causal risk factor), then you require population-level effect estimation…

NOT Patient-Level Prediction

Internal validation on test set: Model shows good discrimination



At-risk threshold = 6%:
Sensitivity = 50%
Specificity = 89%
+ Predictive Value = 12%

We can predict half of all recorded suicidal thoughts and behaviors using only 12% of the population

At-risk threshold = 20%:
Sensitivity = 9%
Specificity = 99%
+ Predictive Value = 26%

At-risk threshold = 1.5%:

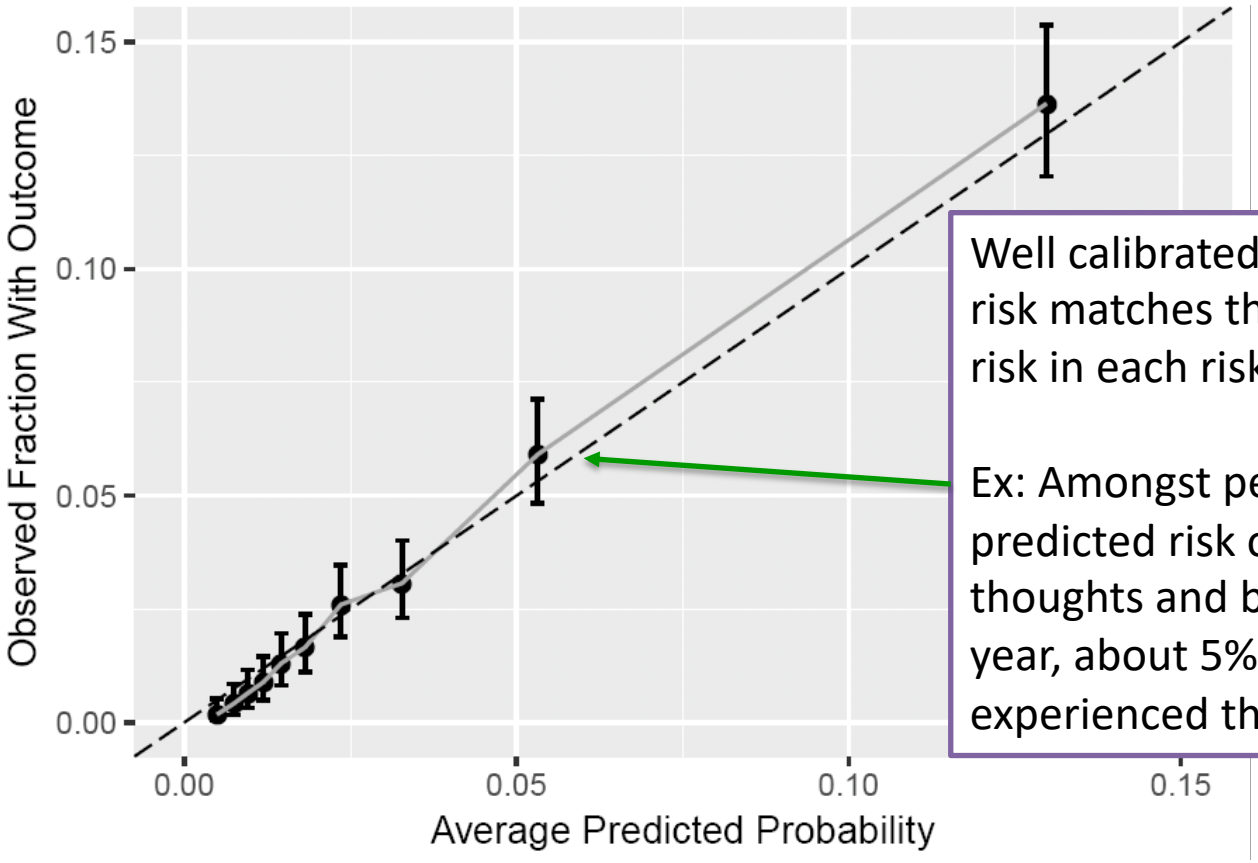
We can predict 90% of all recorded suicidal thoughts and behaviors in 50% of the population.

The remaining 50% have only 10% of the outcomes.

- Predictive Value = 99.3%

AUC: 0.81
Model: Lasso Logistic regression selected 209 var from 53k available variables

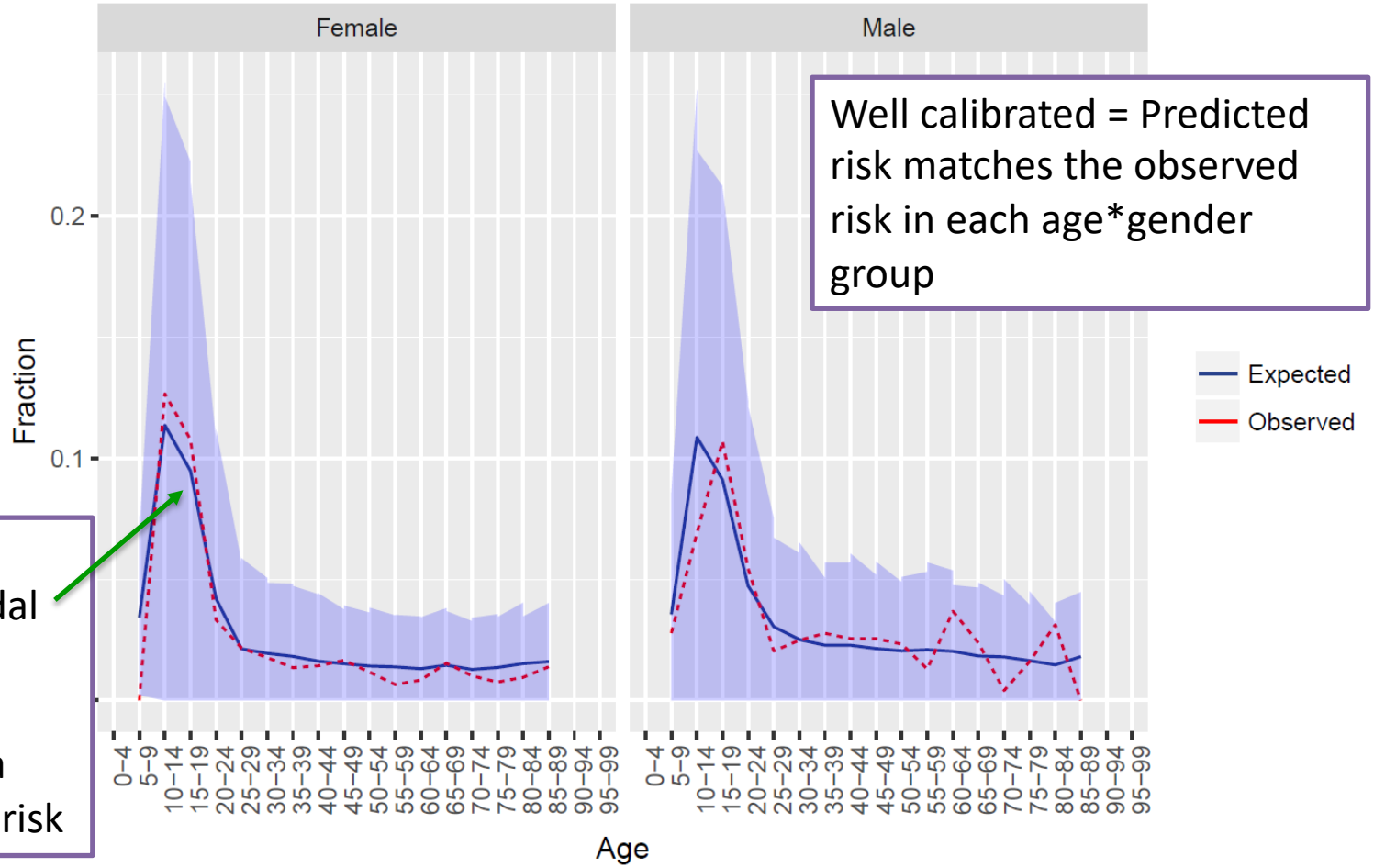
Internal validation on test set: Model shows good calibration across risk profiles



Well calibrated = Predicted risk matches the observed risk in each risk strata

Ex: Amongst people with 5% predicted risk of Suicidal thoughts and behaviors in a year, about 5% actually experienced the event

Internal validation: Model shows good calibration across demographic subgroups



Well calibrated = Predicted risk matches the observed risk in each age*gender group

Young people experience Suicidal thoughts and behaviors more often, and have a higher predicted risk

External validation: Model shows consistent discrimination when applied to other populations

Data type	AUC
Optum (reference)	0.81
US private-payer claims (Truven MarketScan)	0.78
US Medicaid claims	0.70
US Medicare supplemental beneficiary claims	0.70
US electronic health records	0.78
UK electronic health records	0.69

How can these models be useful to a patient?

Outcome of interest	Population average	Patient story	Personalized risk
Suicidal thoughts and behaviors	3.0%	18 year-old female with history of skin cancer and recurrent bouts of anxiety requiring psychotherapy	↑ 14.6%
Hypothyroidism	2.0%		↓ 0.76%
Hyponatremia	1.9%		↓ 0.93%
Sexual dysfunction	1.0%		↓ 0.05%
Seizure	0.60%		0.28%
Gastrointestinal hemhorrage	0.33%		↓ 0.07%
Angle-closure glaucoma	0.06%		0.03%

Rare: 0.01% <= p <0.1%

Uncommon: 0.1% <= p <1%

Common: 1% <= p <10%

Very common: p >= 10%

How can these models be useful to a patient?

Outcome of interest	Population average	Patient story	Personalized risk
Suicidal thoughts and behaviors	3.0%	76 year-old male with liver disease, gout, diverticulitis, who was recently diagnosed with pancreatic cancer	5.18%
Hypothyroidism	2.0%		2.28%
Hyponatremia	1.9%		↑ 23.97%
Sexual dysfunction	1.0%		6.75%
Seizure	0.60%		↑ 10.06%
Gastrointestinal hemhorrage	0.33%		↑ 2.42%
Angle-closure glaucoma	0.06%		↑ 0.15%

Rare: 0.01% <= p <0.1%

Uncommon: 0.1% <= p <1%

Common: 1% <= p <10%

Very common: p >= 10%

How can these models be useful to a patient?

Outcome of interest	Population average	Patient story	Personalized risk
Suicidal thoughts and behaviors	3.0%	79 year-old female with comorbid obesity, Type 2 diabetes mellitus, atrial fibrillation, congestive heart failure, and prior usage of NSAIDs	↓ 0.77%
Hypothyroidism	2.0%		↑ 31.67%
Hyponatremia	1.9%		6.65%
Sexual dysfunction	1.0%		↓ 0.15%
Seizure	0.60%		0.40%
Gastrointestinal hemhorrage	0.33%		0.64%
Angle-closure glaucoma	0.06%		0.07%

Rare: 0.01% <= p <0.1%

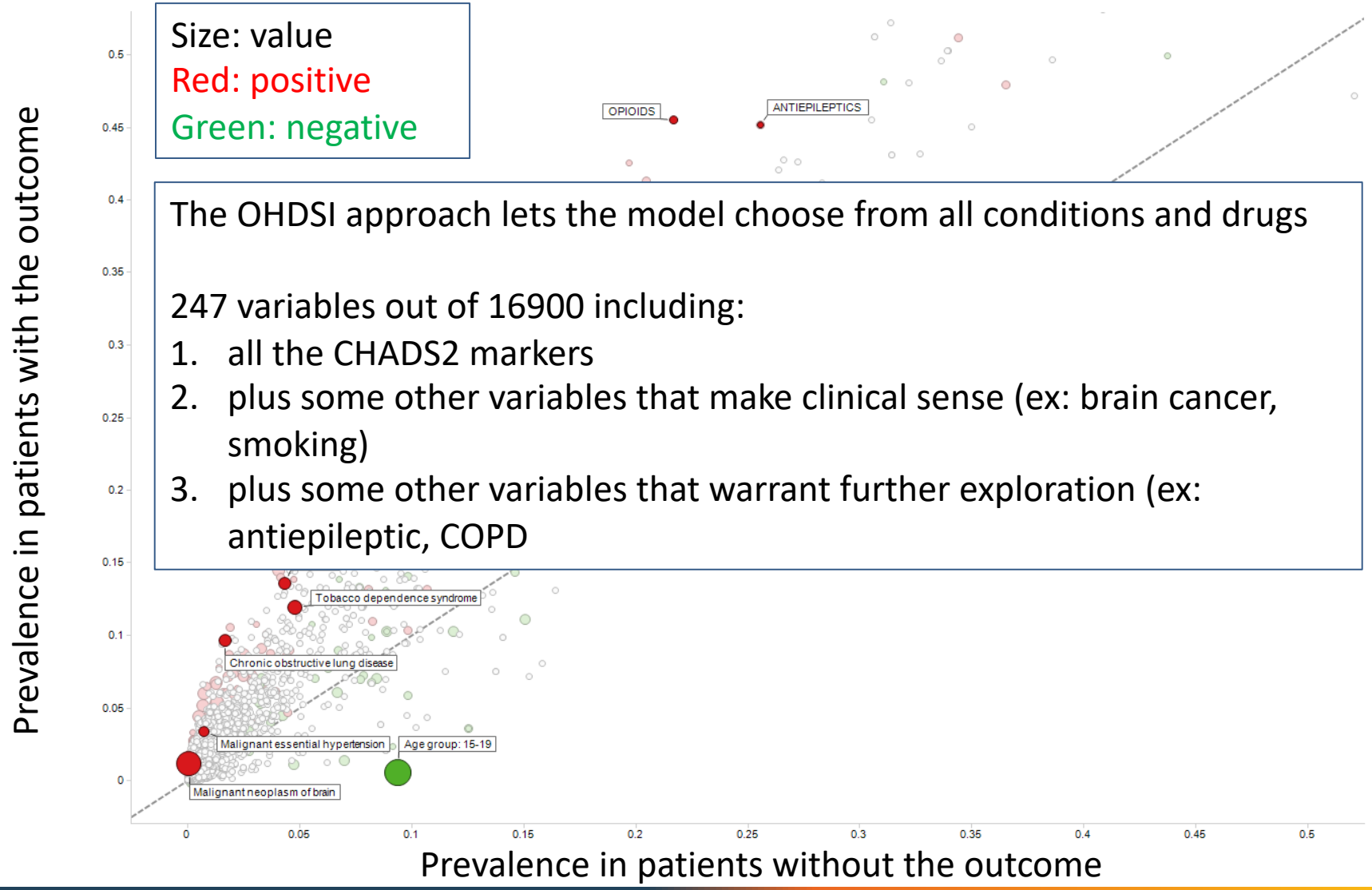
Uncommon: 0.1% <= p <1%

Common: 1% <= p <10%

Very common: p >= 10%

Stroke risk in atrial fibrillation (compare to CHA_2DS_2-VASc Score)

Size: value
Red: positive
Green: negative

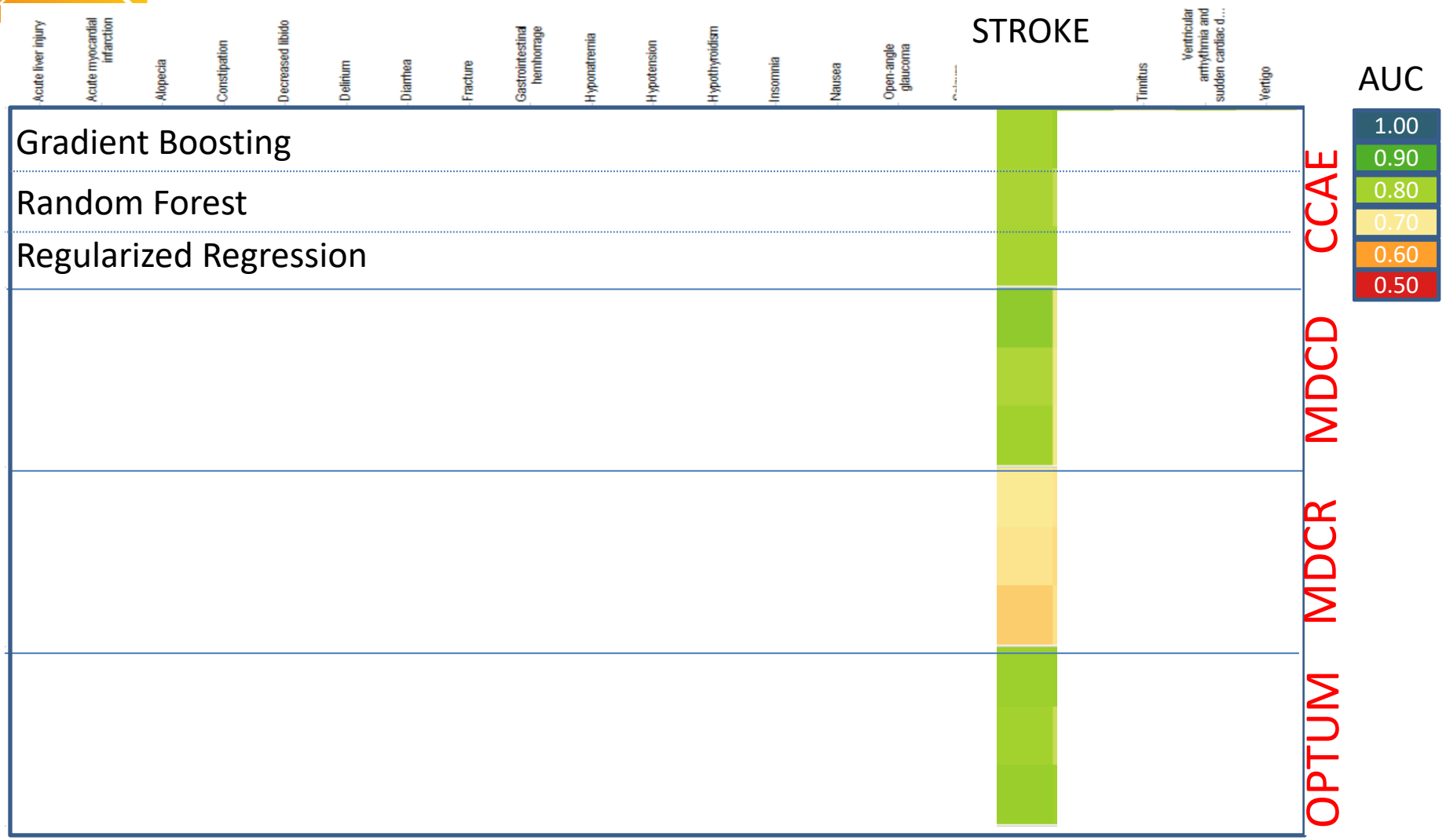


The OHDSI approach lets the model choose from all conditions and drugs

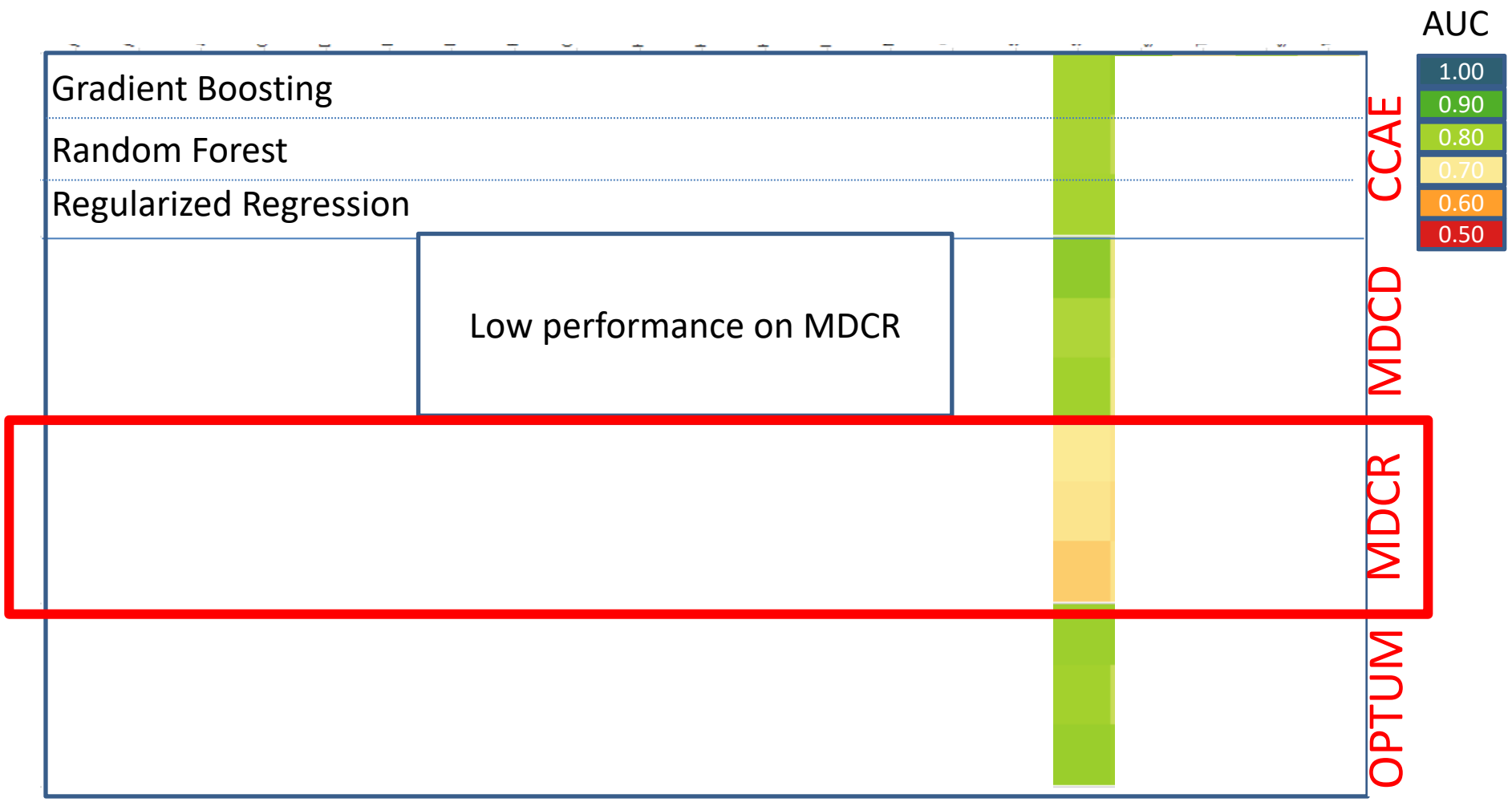
247 variables out of 16900 including:

1. all the CHADS2 markers
2. plus some other variables that make clinical sense (ex: brain cancer, smoking)
3. plus some other variables that warrant further exploration (ex: antiepileptic, COPD

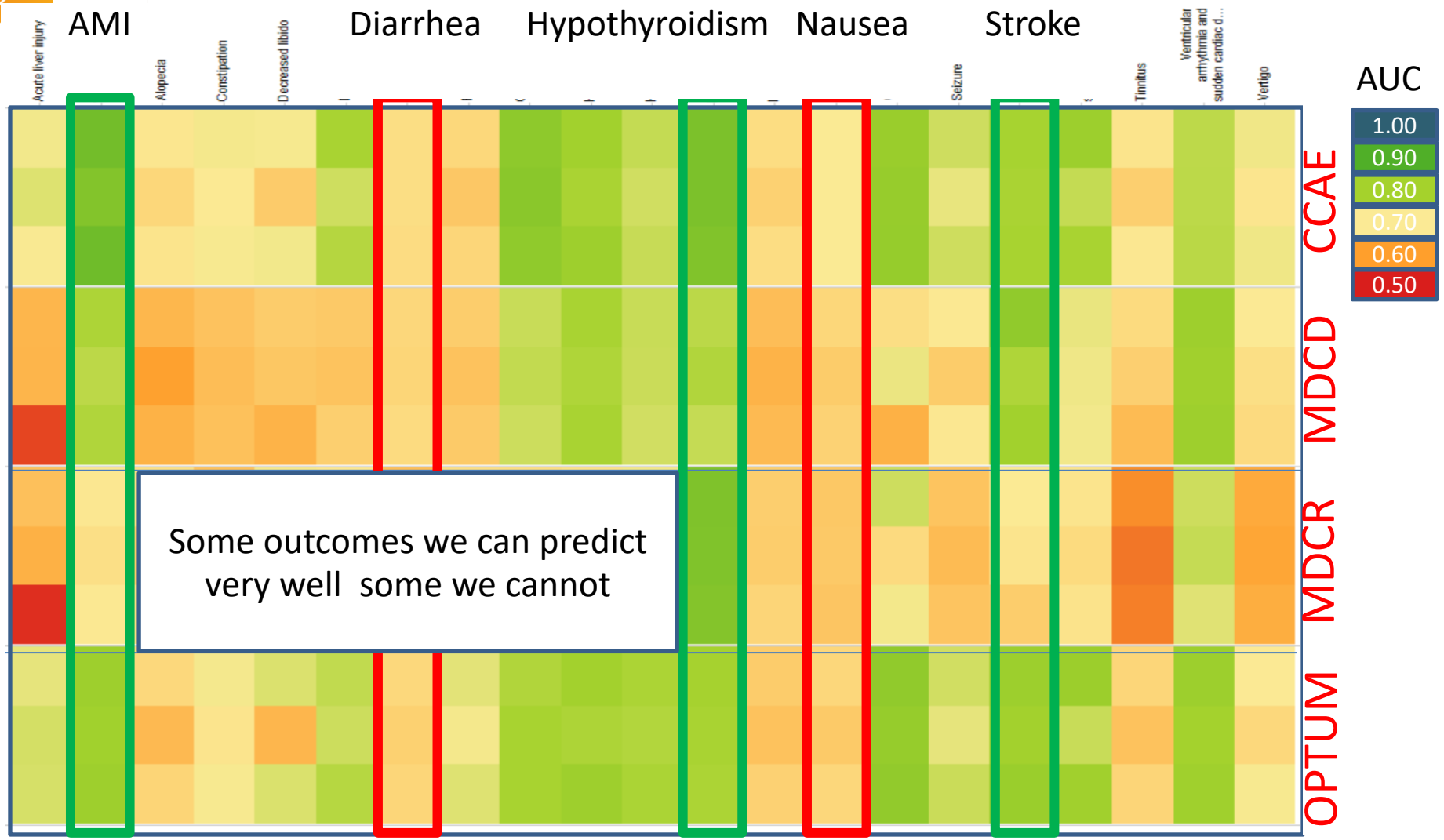
Model Discrimination Stroke



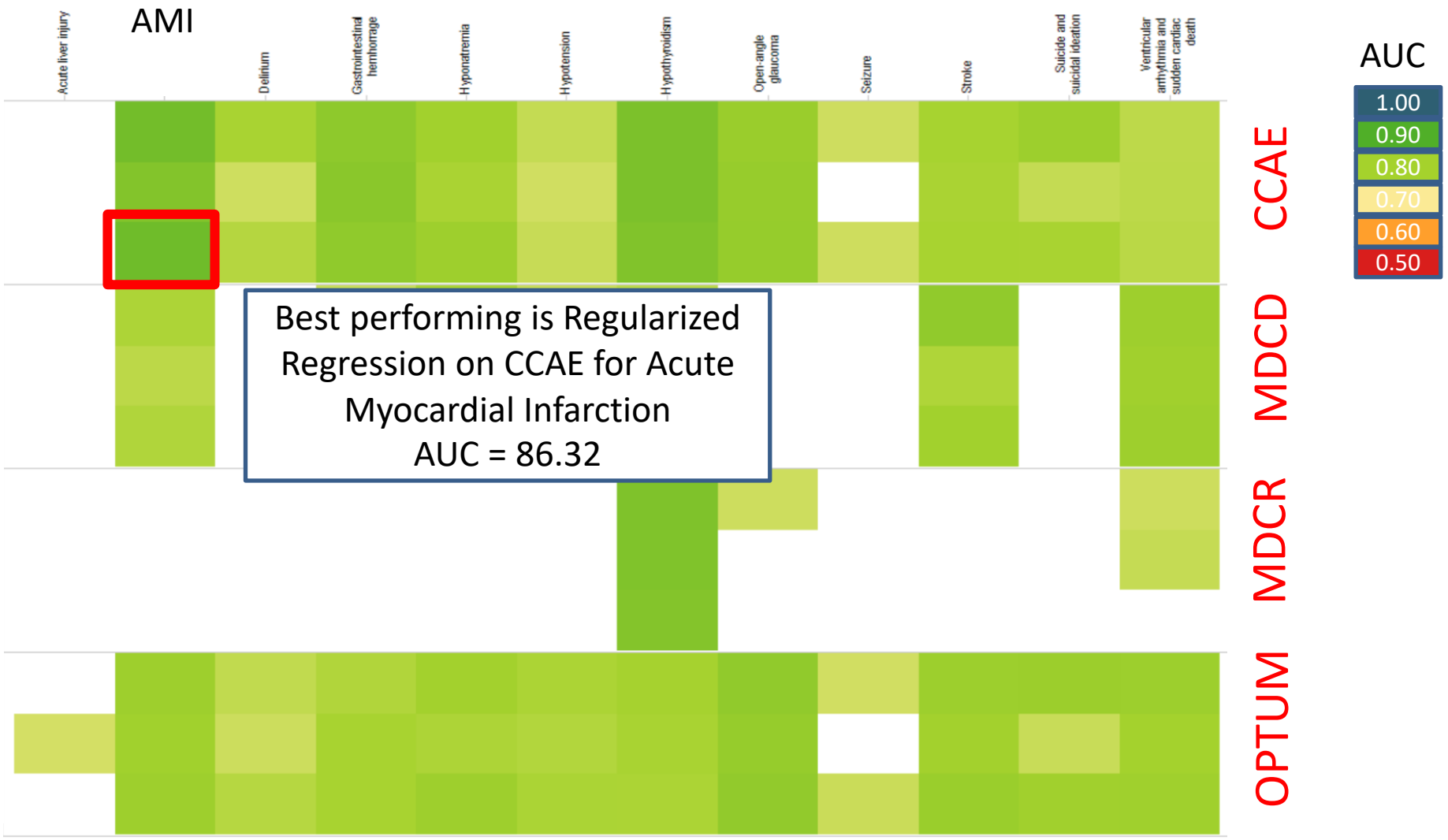
Model Discrimination



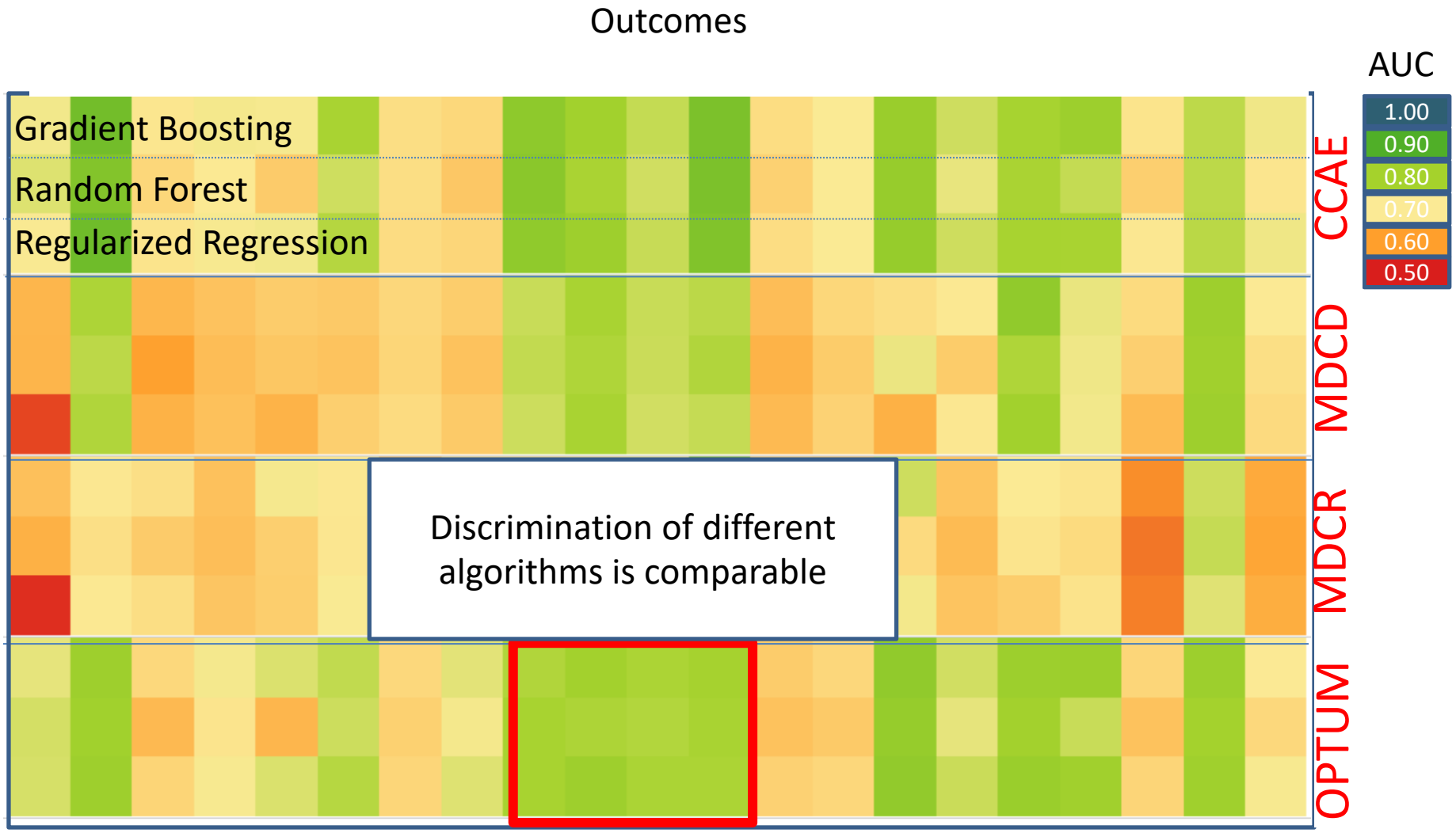
Model Discrimination



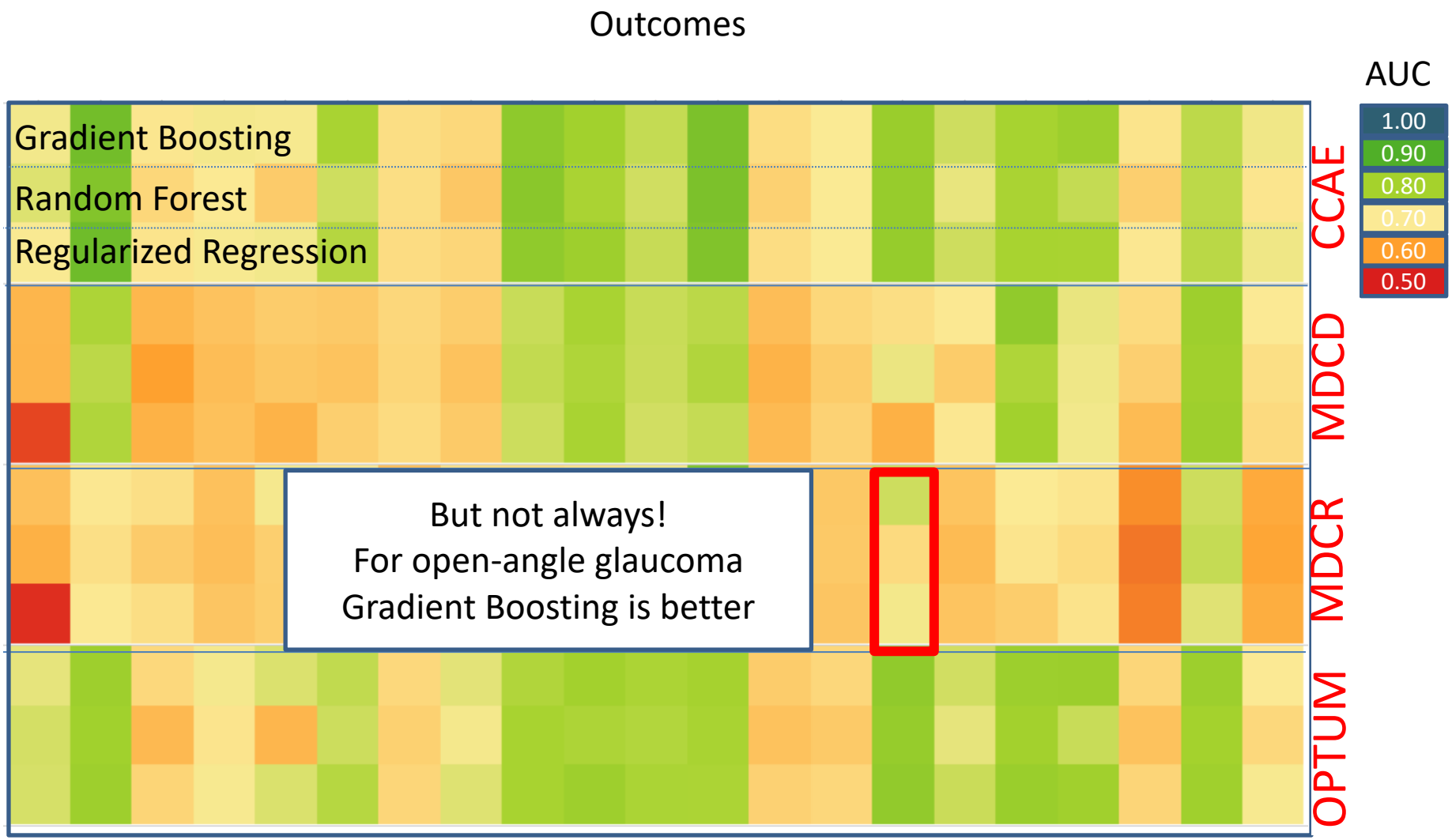
Outcomes with AUC > 0.75



Model Discrimination

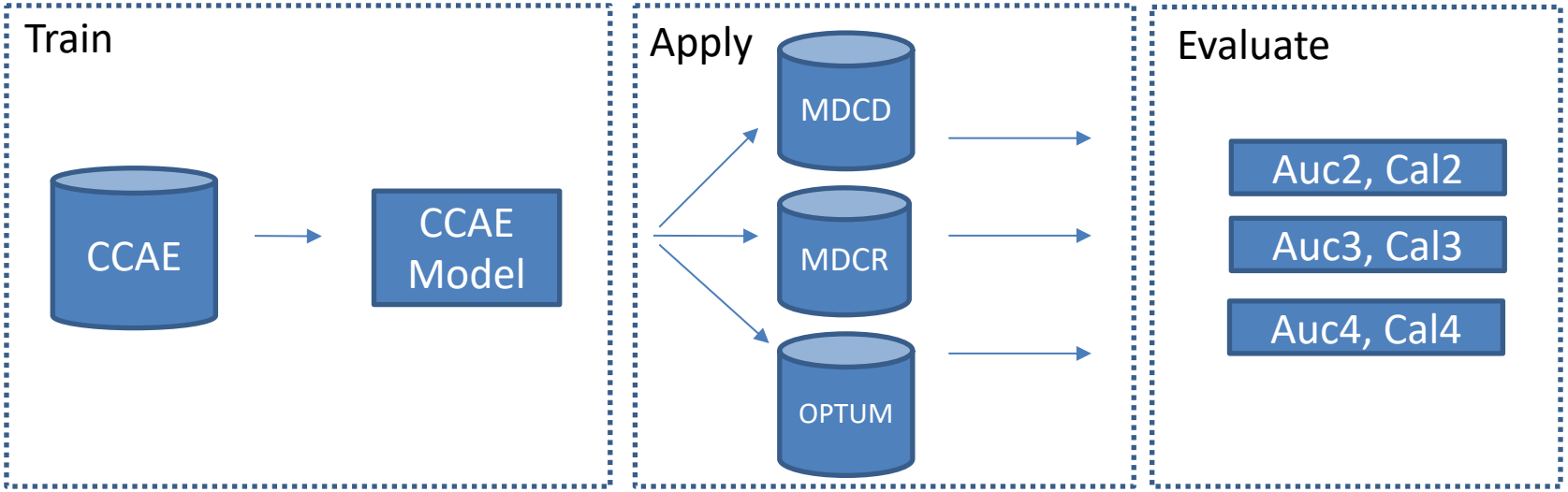


Model Discrimination

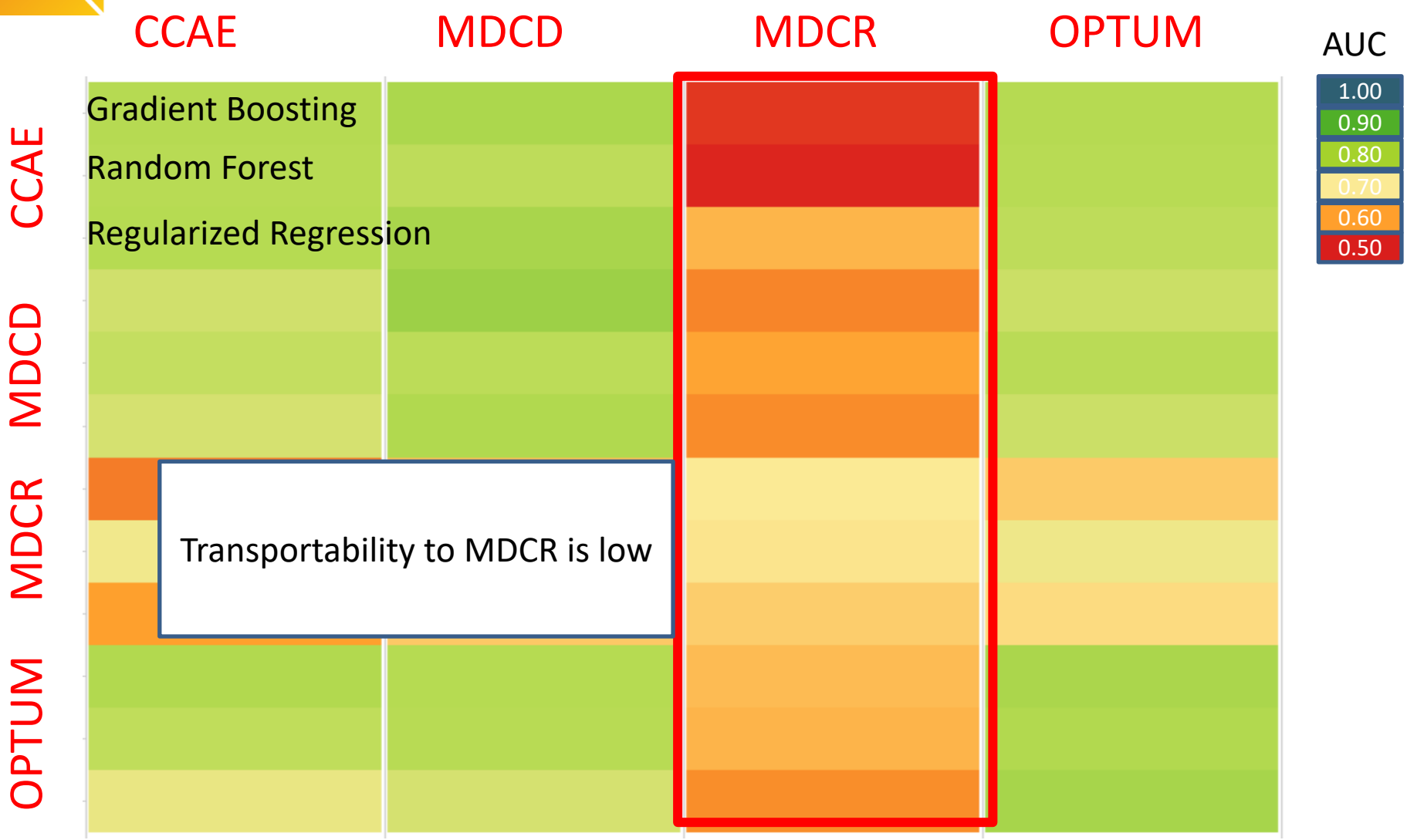


Transportability Assessment

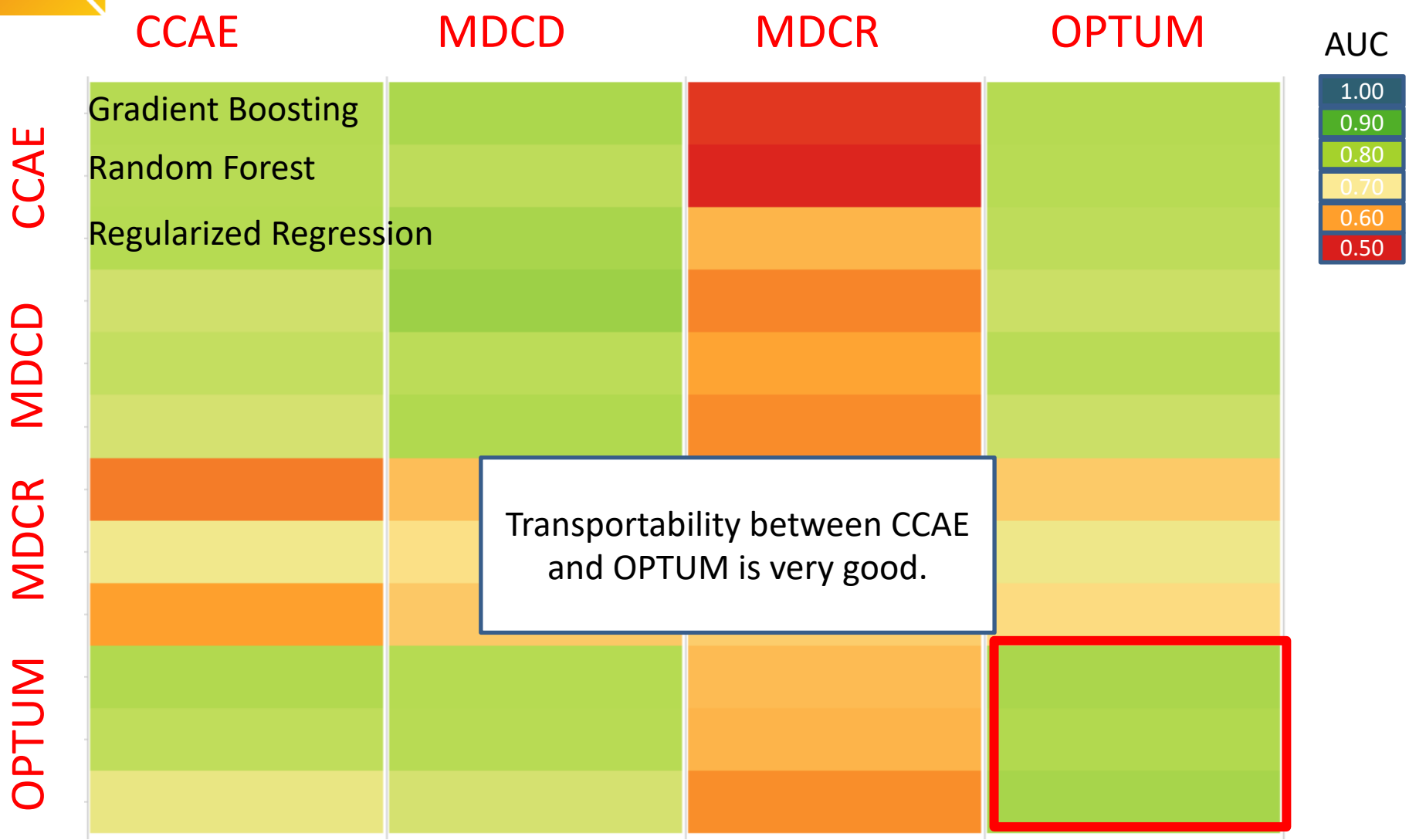
How well do the models perform on other databases?



Transportability Assessment Stroke



Transportability Assessment Stroke



Conclusions

- It is feasible to create an enormous international open research network
  - Sites will volunteer to run studies
- Patient-level prediction can advance the notion of ‘precision medicine’ by identifying the subpopulations at high and low risk and managing treatment decisions accordingly
- This does not have to be a ‘post hoc’ research endeavor but could be integrated into the healthcare delivery system itself
  - At scale

Comments

- Stratified medicine (Iain Buchan)
  - Genomics adds strata
  - Versus N-of-1 and physiology
- If perfect calibration, .95 AUC can get .1 and .9
  - Deep learning can accentuate this
- Cannot predict effect of altering behavior
  - Stop carrying a lighter
  - Predicting risk is not recommending treatment
  - Need population who switched or causality
- Scale to many diseases and populations
  - Must create a repeatable process
  - Needed to study operating characteristics

Join the journey

http://ohdsi.org