Data Availability and Benchmarking in Observational Health Data Sciences and Informatics (OHDSI)

George Hripcsak, MD, MS
Columbia University Irving Medical Center
NewYork-Presbyterian Hospital

International Telecommunication Union &
World Health Organization
Artificial Intelligence for Health

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

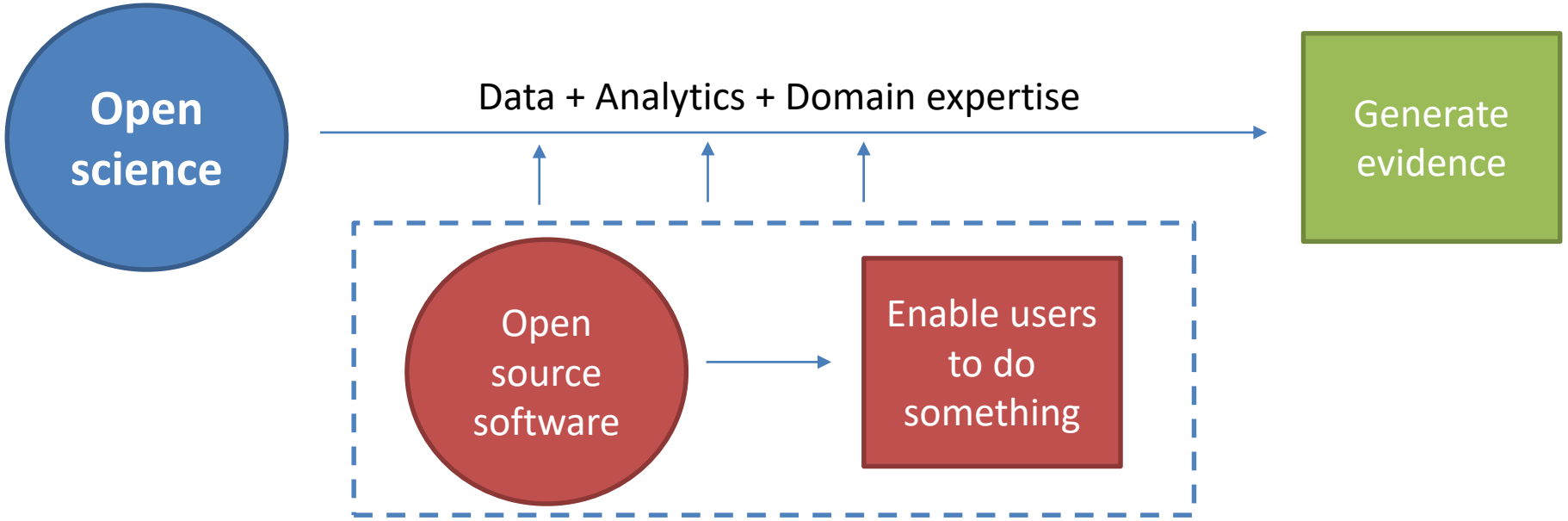
- >200 collaborators from 25 different countries
- Experts in informatics, statistics, epidemiology, clinical sciences
- Active participation from academia, government, industry, providers
- Currently records on about 500 million unique patients in >100 databases

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

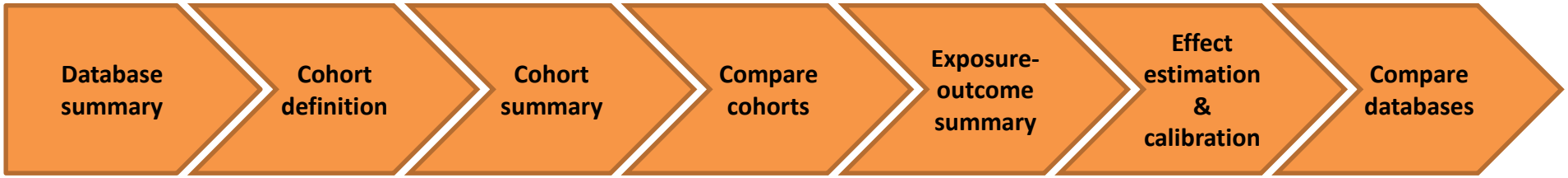
Evidence OHDSI seeks to generate from observational data

- **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?

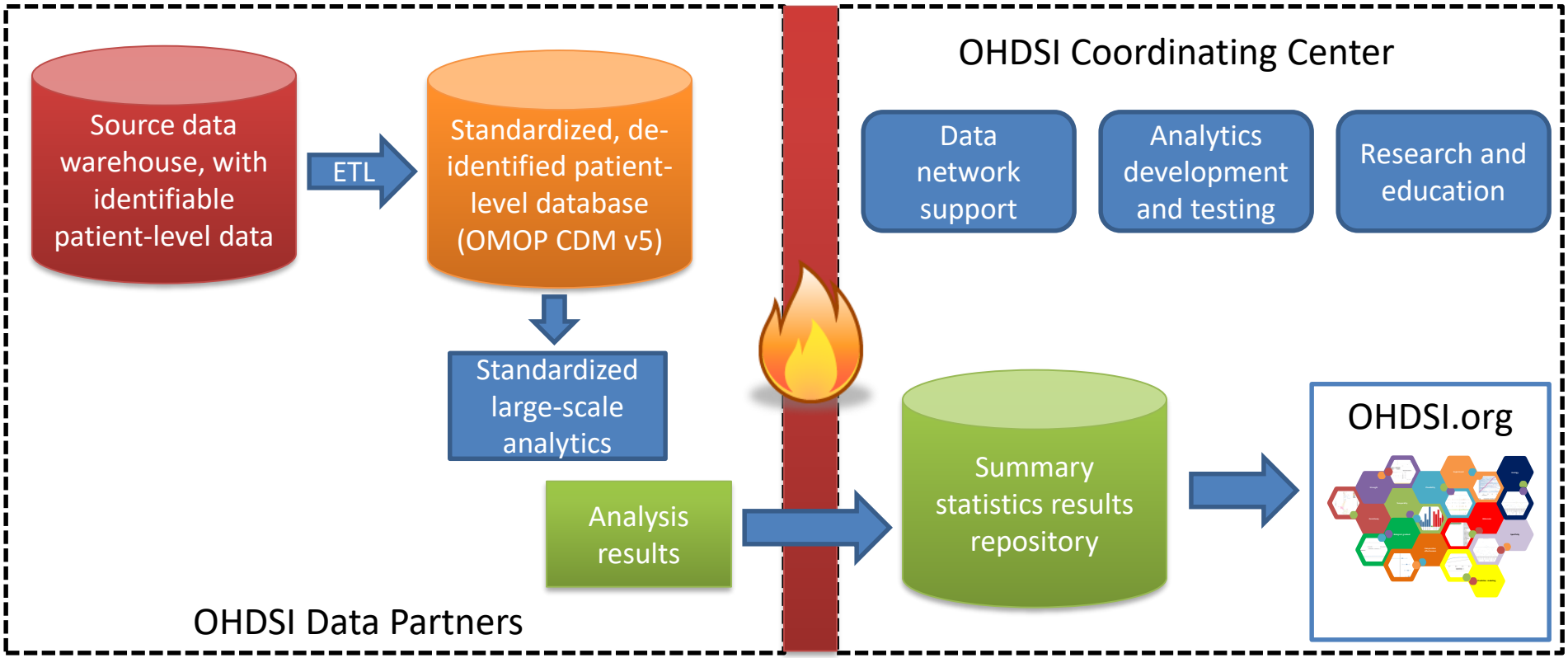
Open Science



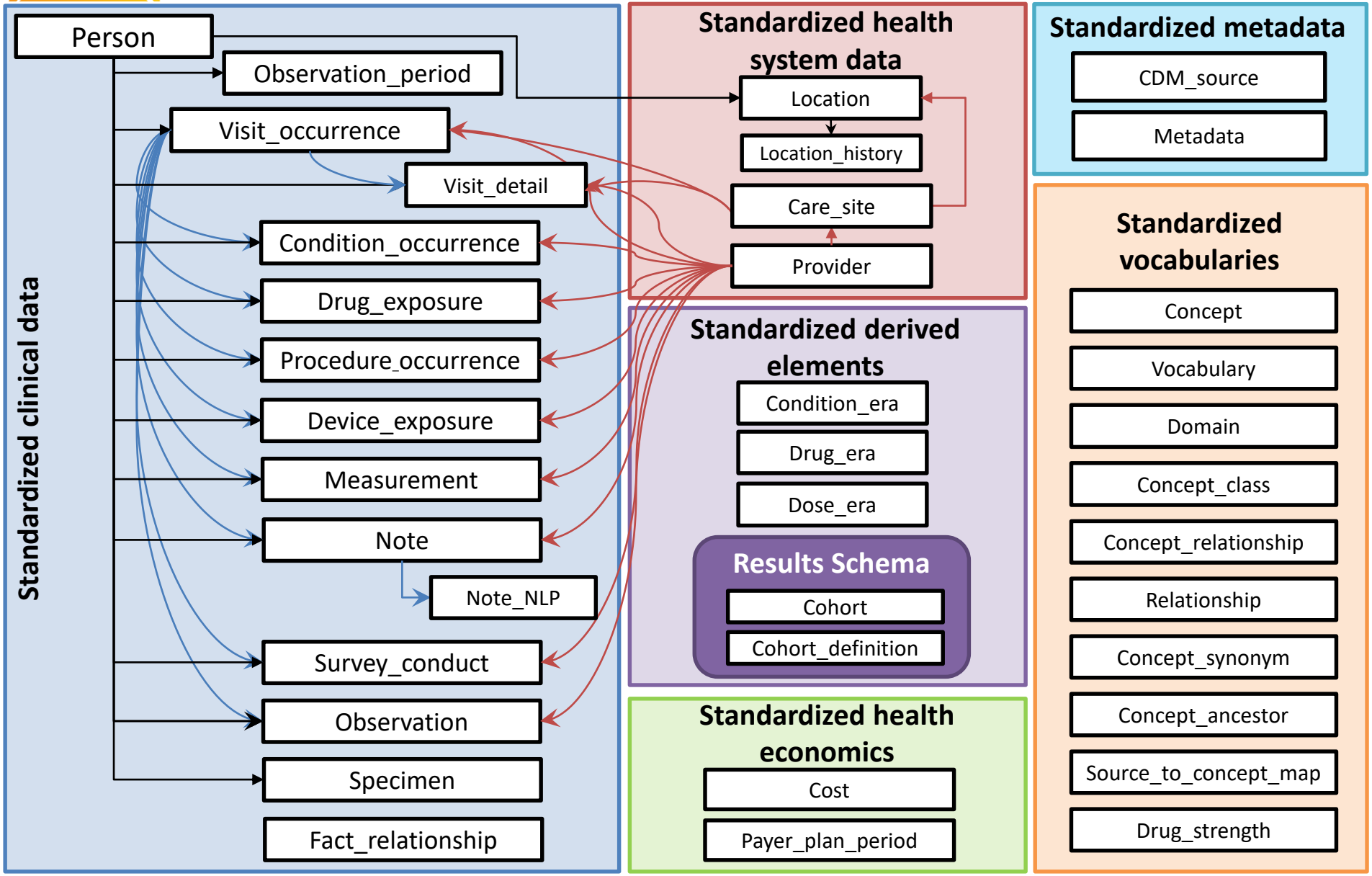
Standardized, transparent workflows



How OHDSI Works

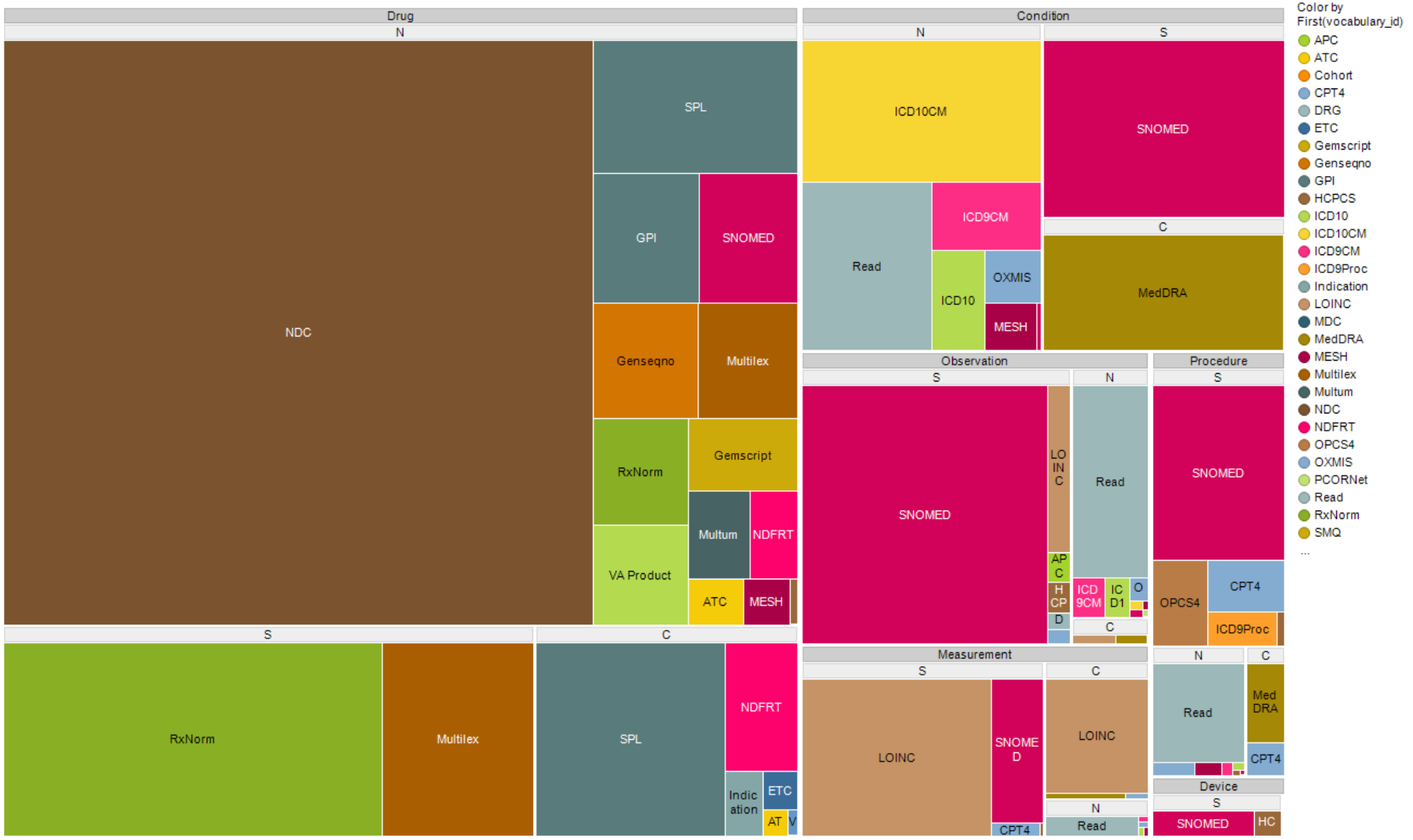


Deep information model
OMOP CDM Version 6

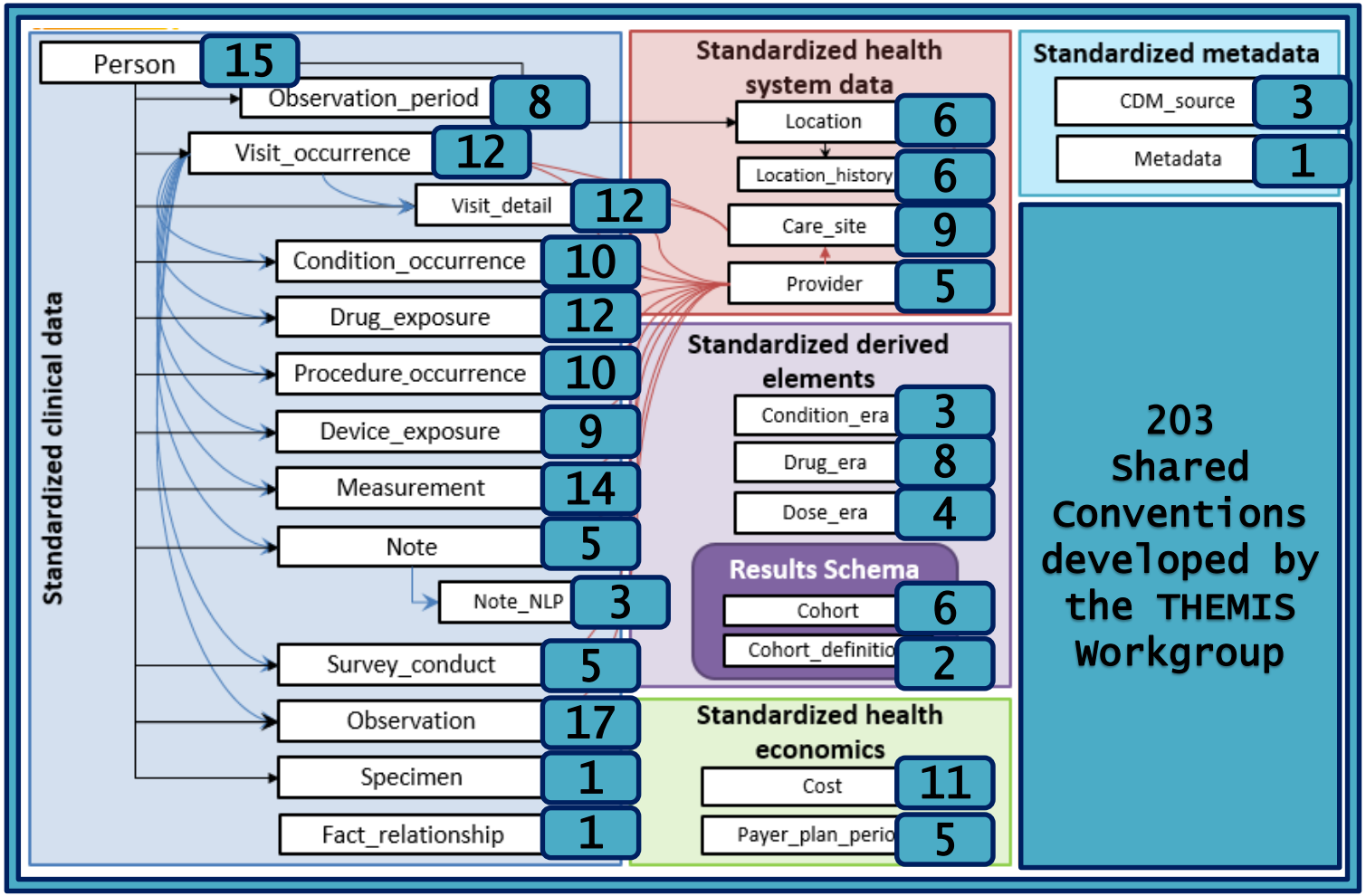


Extensive vocabularies

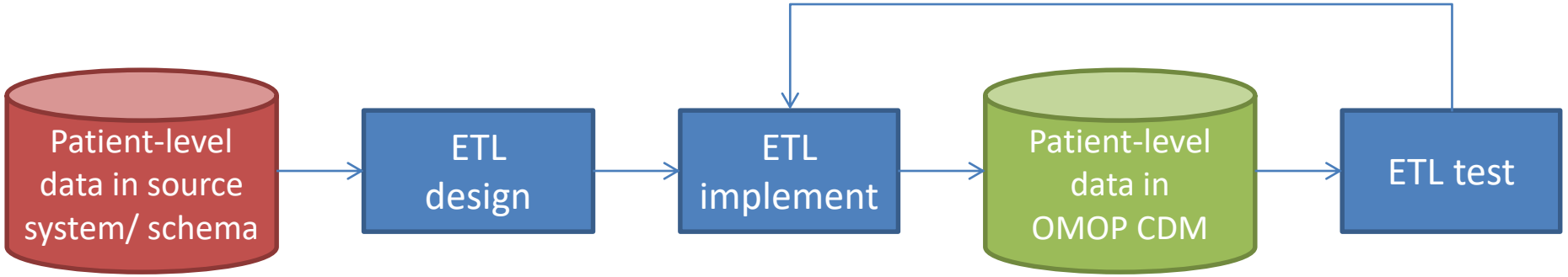
Breakdown of OHDSI concepts by domain, standard class, and vocabulary



Standardized conventions



Preparing your data for analysis



OHDSI tools built to help

WhiteRabbit: profile your source data

RabbitInAHat: map your source structure to CDM tables and fields

ATHENA: standardized vocabularies for all CDM domains

Usagi: map your source codes to CDM vocabulary

CDM: DDL, index, constraints for Oracle, SQL Server, PostgresQL; Vocabulary tables with loading scripts

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

OHDSI Forums: Public discussions for OMOP CDM Implementers/developers

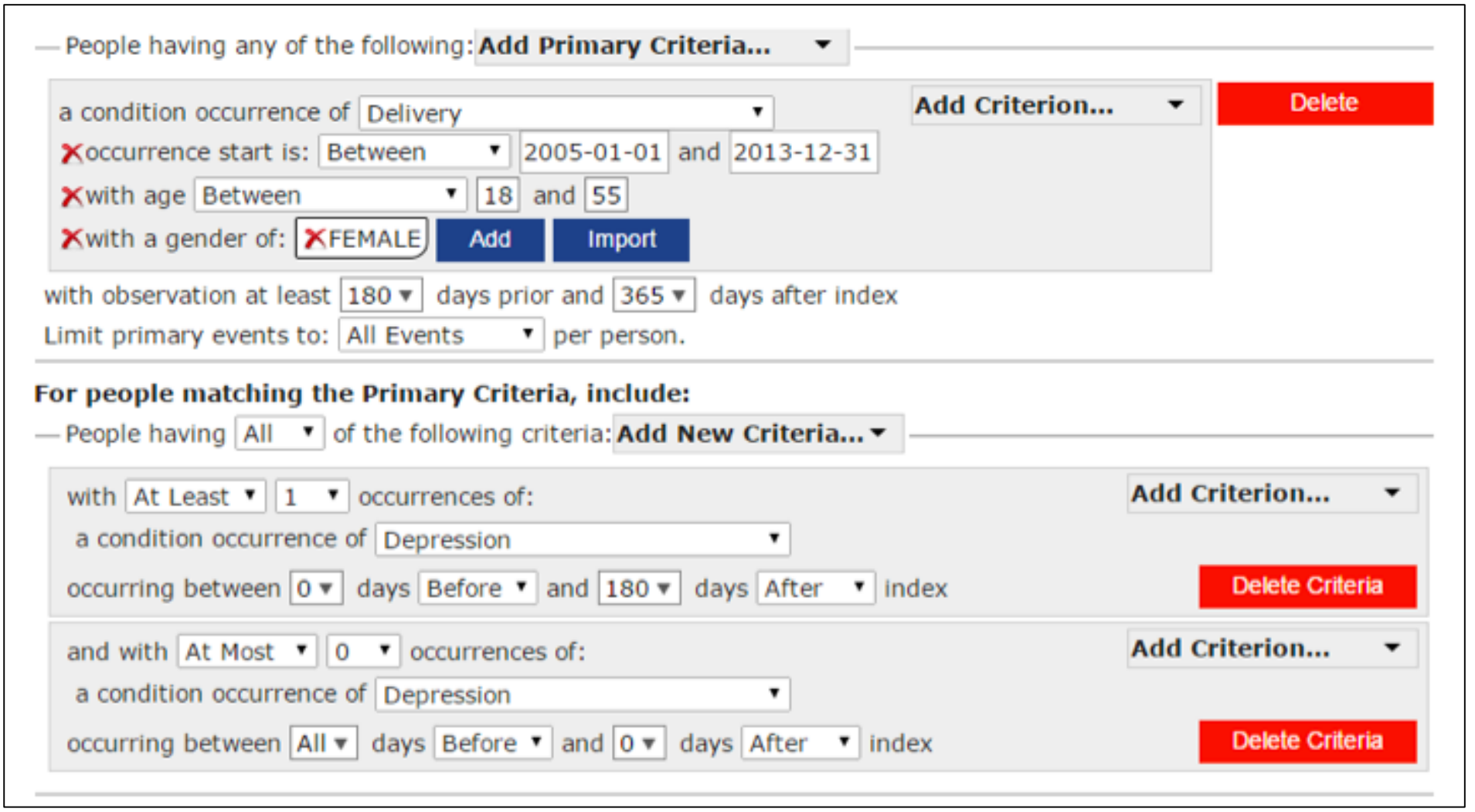
ACHILLES Heel Data Curation

Data Quality Messages

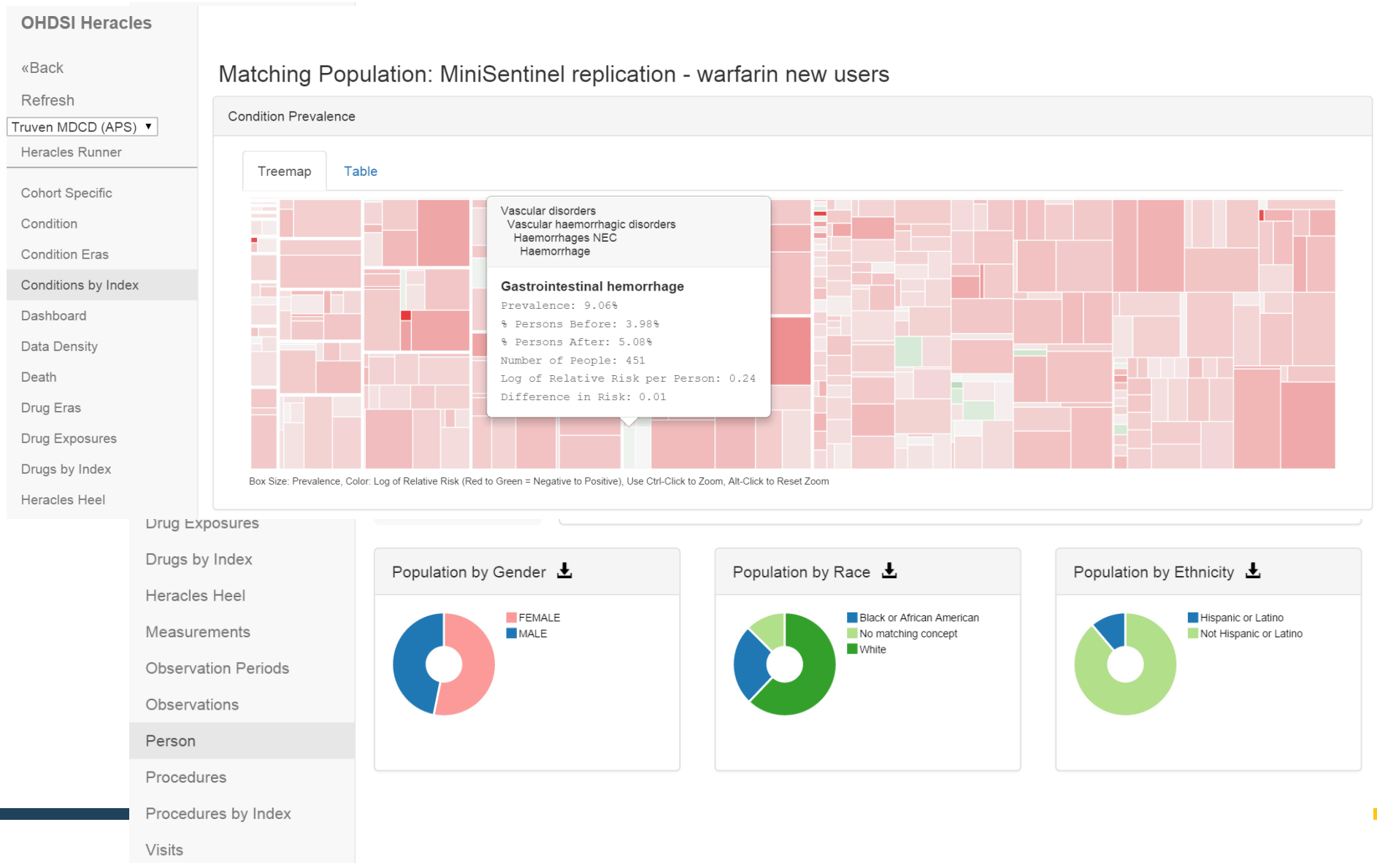
Search: Show / hide columns

Message Type	Message
ERROR	101-Number of persons by age, with age at first observation period; should not have age < 0, (n=848)
ERROR	103 - Distribution of age at first observation period (count = 1); min value should not be negative
ERROR	114-Number of persons with observation period before year-of-birth; count (n=851) should not be > 0
ERROR	206 - Distribution of age by visit_concept_id (count = 7); min value should not be negative
ERROR	301-Number of providers by specialty concept_id; 224 concepts in data are not in correct vocabulary (Specialty)
ERROR	400-Number of persons with at least one condition occurrence, by condition_concept_id; 115 concepts in data are not in correct vocabulary (SNOMED)
ERROR	406 - Distribution of age by condition_concept_id (count = 753); min value should not be negative

ATLAS to build, visualize, and analyze cohorts



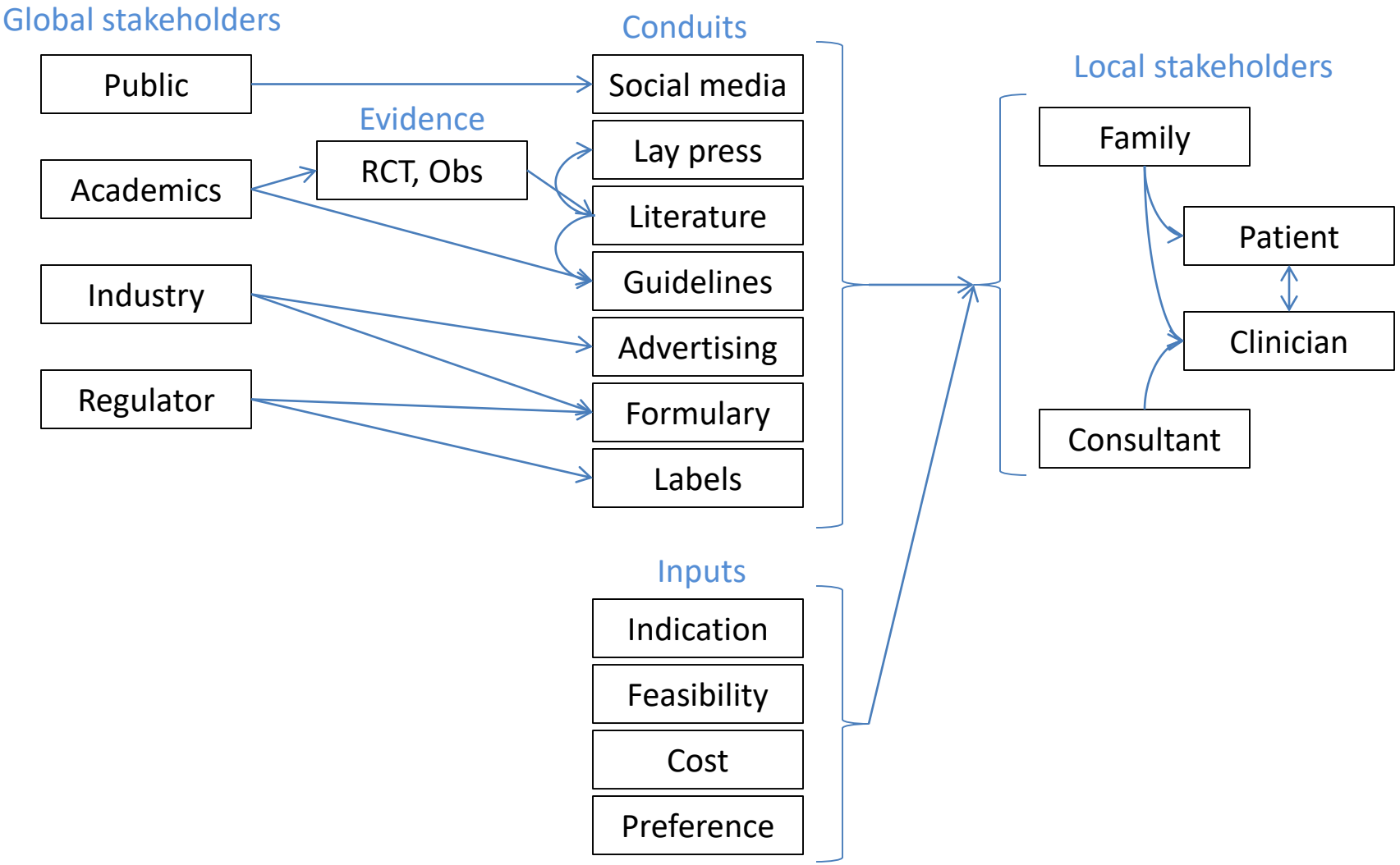
Characterize the cohorts of interest



OHDSI in Action

- Characterization

Treatment Pathways



OHDSI in action:
Chronic disease treatment pathways

- Conceived at AMIA 15Nov2014
- Protocol written, code written and tested at 2 sites 30Nov2014
- Analysis submitted to OHDSI network 2Dec2014
- Results submitted for 7 databases 5Dec2014

OHDSI participating data partners

Abbre-viation	Name	Description	Population, millions
AUSOM	Ajou University School of Medicine	South Korea; inpatient hospital EHR	2
CCAE	MarketScan Commercial Claims and Encounters	US private-payer claims	119
CPRD	UK Clinical Practice Research Datalink	UK; EHR from general practice	11
CUMC	Columbia University Medical Center	US; inpatient EHR	4
GE	GE Centricity	US; outpatient EHR	33
INPC	Regenstrief Institute, Indiana Network for Patient Care	US; integrated health exchange	15
JMDC	Japan Medical Data Center	Japan; private-payer claims	3
MDCD	MarketScan Medicaid Multi-State	US; public-payer claims	17
MDCR	MarketScan Medicare Supplemental and Coordination of Benefits	US; private and public-payer claims	9
OPTUM	Optum ClinFormatics	US; private-payer claims	40
STRIDE	Stanford Translational Research Integrated Database Environment	US; inpatient EHR	2
HKU	Hong Kong University	Hong Kong; EHR	1

Proceedings of the National Academy of Sciences, 2016

PNAS

COLLOQUIUM PAPER

Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak[a,b,c,1], Patrick B. Ryan[c,d], Jon D. Duke[c,e], Nigam H. Shah[c,f], Rae Woong Park[c,g], Vojtech Huser[c,h], Marc A. Suchard[c,i,j,k], Martijn J. Schuemie[c,d], Frank J. DeFalco[c,d], Adler Perotte[a,c], Juan M. Banda[c,f], Christian G. Reich[c,l], Lisa M. Schilling[c,m], Michael E. Matheny[c,n,o], Daniella Meeker[c,p,q], Nicole Pratt[c,r], and David Madigan[c,s]

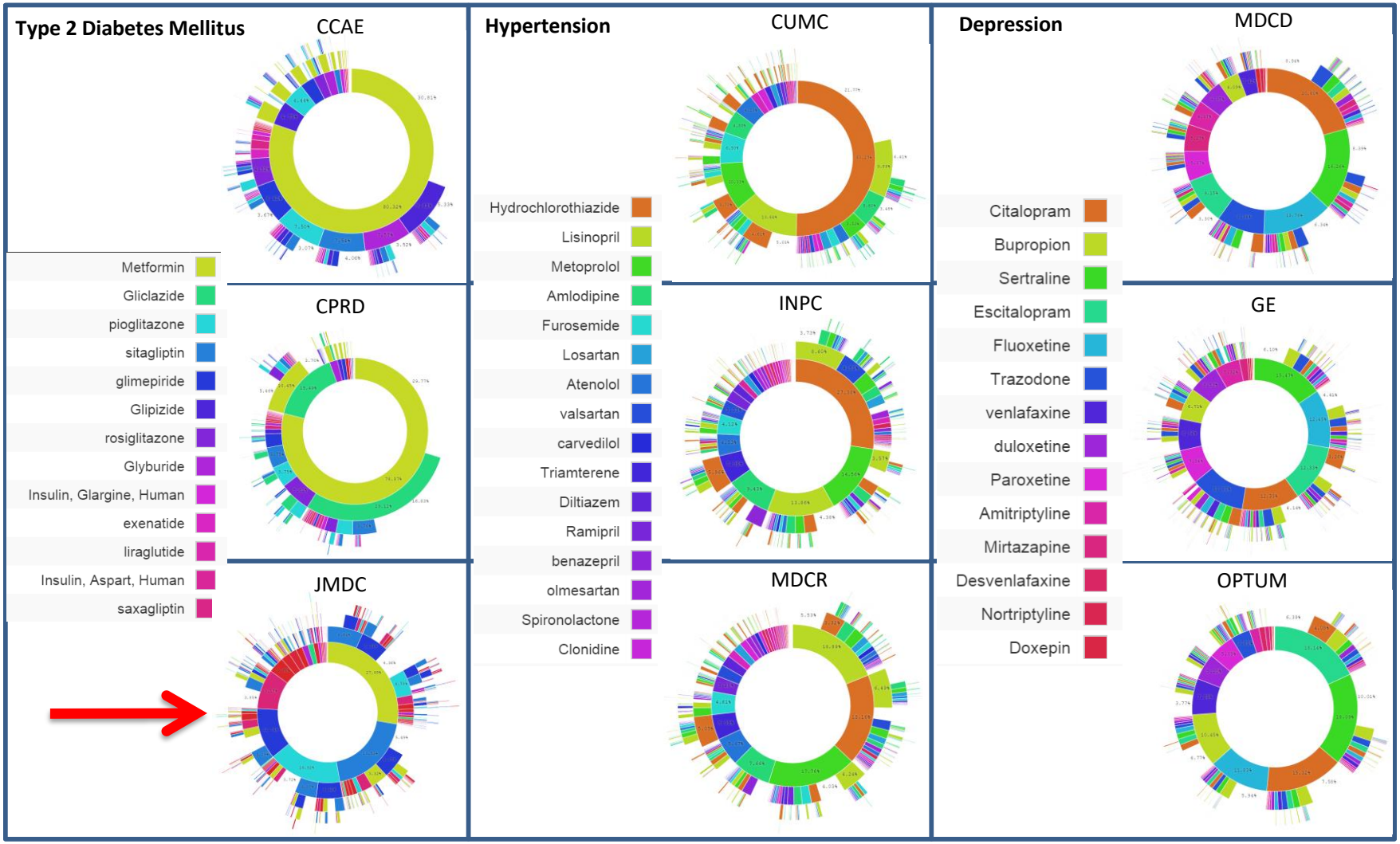
[a]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; [b]Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY 10032; [c]Observational Health Data Sciences and Informatics, New York, NY 10032; [d]Epidemiology Analytics, Janssen Research and Development, Titusville, NJ 08560; [e]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN 46205; [f]Center for Biomedical Informatics Research, Stanford University, CA 94305; [g]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea, 443-380; [h]Lister Hill National Center for Biomedical Communications (National Library of Medicine), National Institutes of Health, Bethesda, MD 20894; [i]Department of Biomathematics, University of California, Los Angeles, CA 90095; [j]Department of Biostatistics, University of California, Los Angeles, CA 90095; [k]Department of Human Genetics, University of California, Los Angeles, CA 90095; [l]Real World Evidence Solutions, IMS Health, Burlington, MA 01809; [m]Department of Medicine, University of Colorado School of Medicine, Aurora, CO 80045; [n]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37212; [o]Geriatric Research, Education and Clinical Center, VA Tennessee Valley Healthcare System, Nashville, TN 37212; [p]Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089; [q]Department of Pediatrics, University of Southern California, Los Angeles, CA 90089; [r]Division of Health Sciences, University of South Australia, Adelaide, SA, Australia 5001; and [s]Department of Statistics, Columbia University, New York, NY 10027

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 5, 2016 (received for review June 14, 2015)

Observational research promises to complement experimental research by providing large, diverse populations that would be infeasible for an experiment. Observational research can test its own clinical hypotheses, and observational studies also can contribute to the design of experiments and inform the generalizability of experimental research. Understanding the diversity of populations

Without sufficiently broad databases available in the first stage, randomized trials are designed without explicit knowledge of actual disease status and treatment practice. Literature reviews are restricted to the population choices of previous investigations, and pilot studies usually are limited in scope. By exploiting the ClinicalTrials.gov national trial registry (9) and electronic health

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



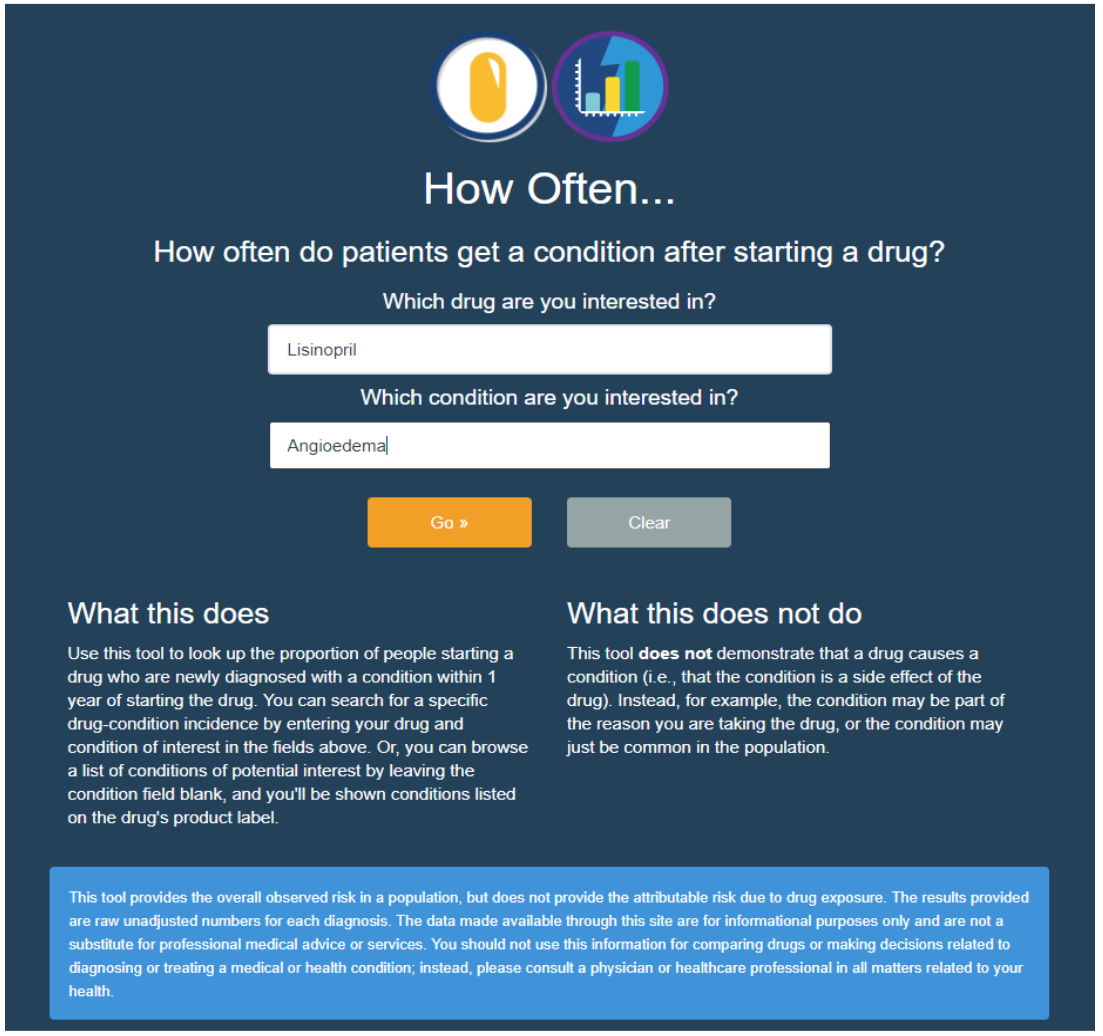
Conclusions: Network research

- It is feasible to encode the world population in a single data model

- Generating evidence is feasible
- Stakeholders willing to share results
- Able to accommodate vast differences in privacy and research regulation

howoften.org

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



OHDSI in Action

- Population-level estimation

What is the quality of the current evidence from observational analyses?

ORIGINAL CONTRIBUTION

JAMA®

Exposure to Oral Bisphosphonates and Risk of Esophageal Cancer

Chris R. Cardwell, PhD
Christian C. Abnet, PhD
Marie M. Cantwell, PhD
Liam J. Murray, MD

Context Use of oral bisphosphonates has increased dra
and elsewhere. Esophagitis is a known adverse effect of
cent reports suggest a link between bisphosphonate us
this has not been robustly investigated.

Objective To investigate the association between bis
ageal cancer.

Design, Setting, and Participants Data were ext

August2010: "Among patients in the UK General Practice Research Database, the use of oral bisphosphonates was not significantly associated with incident esophageal or gastric cancer"

BISPHOSPHONATES INHIBIT OSTEO-clast-mediated bone resorp-

sembles ground alendronate tablets has been found on biopsy in patients with bisphosphonate-related esophagitis, and follow-up endoscopies have shown that abnormalities remain after the esophagitis heals.[6] Reflux esophagitis is an established risk factor for esophageal cancer through the Barrett pathway.[7-9] It is not known whether bisphosphonate-related esophagitis can also increase esophageal cancer risk. However, the US Food and Drug Administration recently reported 23 cases of esophageal cancer (between 1995 and 2008) in patients using the bisphosphonate alendronate and a further 31 cases in pa-

cohort. The incidence of esophageal and gastric cancer
person-years of risk in both the bisphosphonate and c
of esophageal cancer alone in the bisphosphonate a
and 0.44 per 1000 person-years of risk, respectively. T
of esophageal and gastric cancer combined between
phonate use (adjusted hazard ratio, 0.96 [95% confic
risk of esophageal cancer only (adjusted hazard ratio,
val, 0.77-1.49]). There also was no difference in risk of
by duration of bisphosphonate intake.

Conclusion Among patients in the UK General Practic
of oral bisphosphonates was not significantly associate
gastric cancer.

JAMA. 2010;304(6):657-663

Large studies with appropriate comparison groups, adequate follow-up, robust characterization of bisphospho-

termine w
crease esop
dertook su

BMJ

RESEARCH

Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort

Jane Green, clinical epidemiologist,[1] Gabriela Czanner, statistician,[1] Gillian Reeves, statistical epidemiologist,[1] Joanna Watson, epidemiologist,[1] Lesley Wise, manager, Pharmacoepidemiology Research and Intelligence Unit,[2] Valerie Beral, professor of cancer epidemiology[1]

ABSTRACT

Objective To examine the hypothesis that risk of oesophageal, but not of gastric or colorectal, cancer is increased in users of oral bisphosphonates.

Design Nested case-control analysis within a primary care cohort of about 6 million people in the UK, with prospectively recorded information on prescribing of bisphosphonates.

Setting UK General Practice Research Database cohort.

Participants Men and women aged 40 years or over—2954 with oesophageal cancer, 2018 with gastric cancer, and 10 641 with colorectal cancer, diagnosed in 1995-2005; five controls per case matched for age, sex, general practice, and observation time.

Main outcome measures Relative risks for incident invasive cancers of the oesophagus, stomach, and colorectum, adjusted for smoking, alcohol, and body mass index.

Conclusions The risk of oesophageal cancer increased with 10 or more prescriptions for oral bisphosphonates and with prescriptions over about a five year period. In Europe and North America, the incidence of oesophageal cancer at age 60-79 is typically 1 per 1000 population over five years, and this is estimated to increase to about 2 per 1000 with five years' use of oral bisphosphonates.

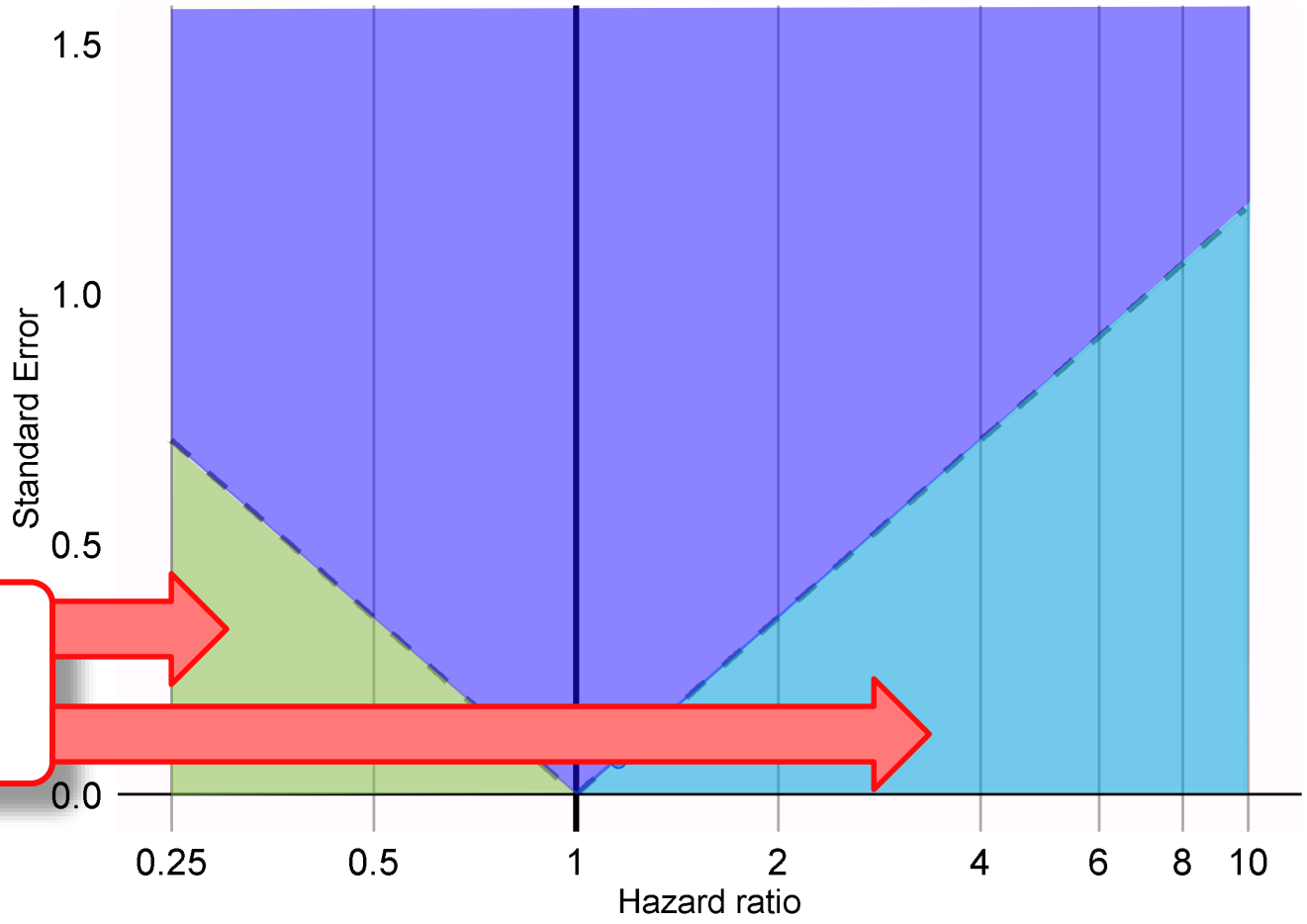
INTRODUCTION

Adverse gastrointestinal effects are common among people who take oral bisphosphonates for the prevention and treatment of osteoporosis; they range from dyspepsia, nausea, and abdominal pain to erosive oesophagitis and oesophageal ulcers.[1] Recent case reports have suggested a possible increase in the risk of oesophageal cancer with use of such bisphosphonate preparations.[2] We report here on the relation between prospectively recorded prescribing information for

Sept2010: "In this large nested case-control study within a UK cohort [General Practice Research Database], we found a significantly increased risk of oesophageal cancer in people with previous prescriptions for oral bisphosphonates"

0.87 (0.64 to 1.19) and 0.87 (0.77 to 1.00). The specificity ... hospital records) are around 95% valid and

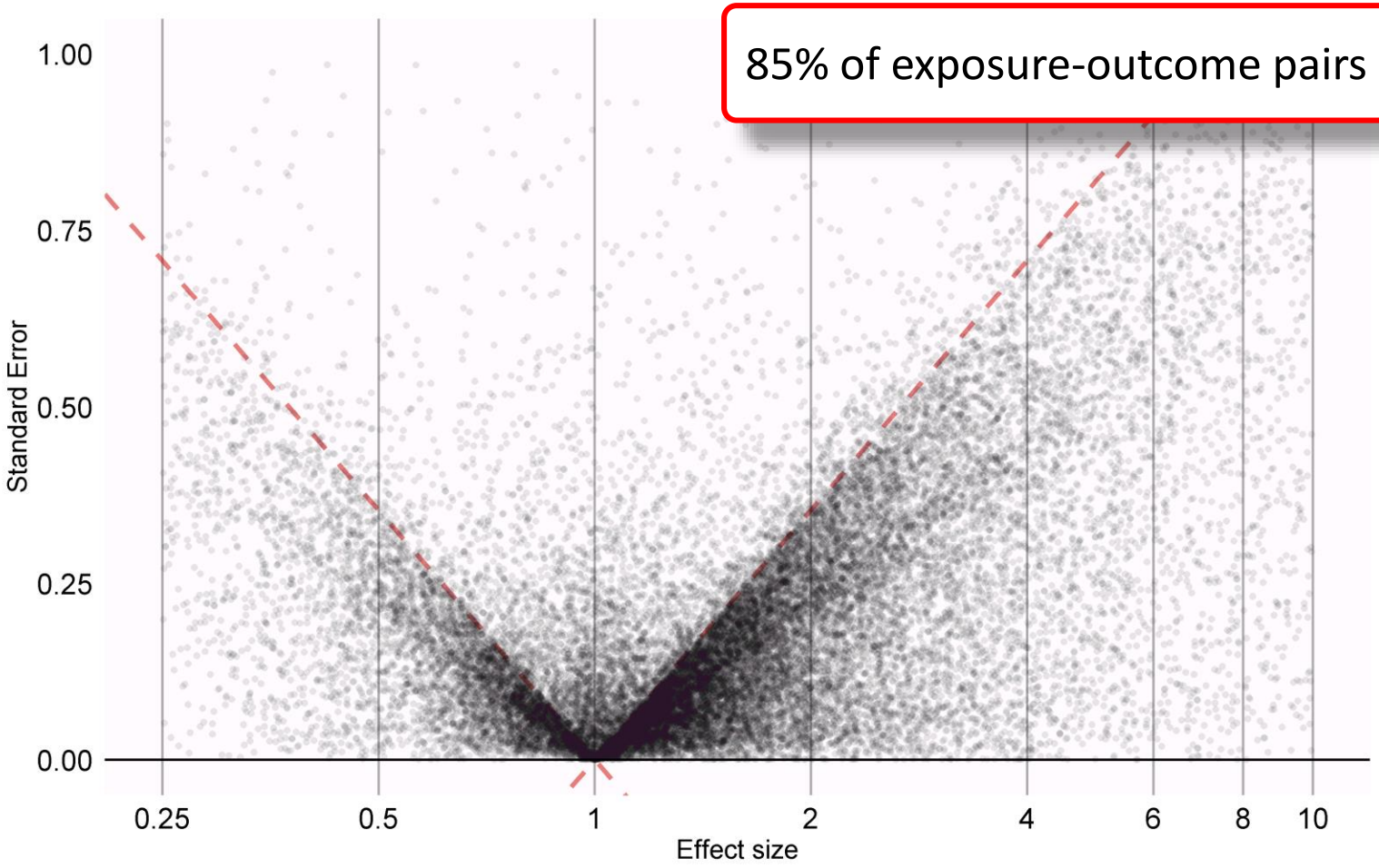
Standard error vs effect size



Statistically significant

Observational research results in literature

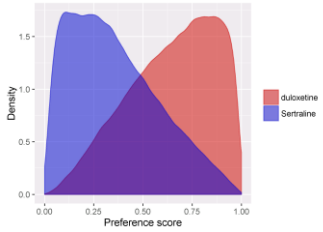
85% of exposure-outcome pairs have $p < 0.05$

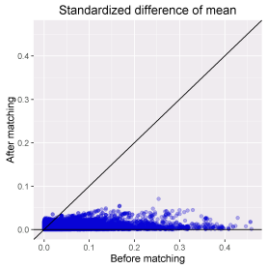


29,982 estimates
11,758 papers

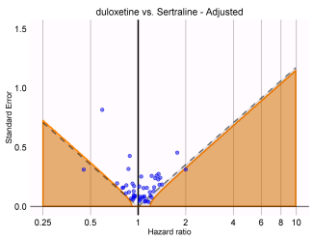
Addressing reproducibility

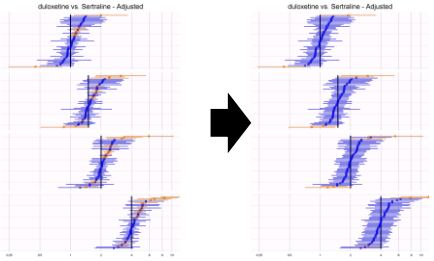
1. **Propensity stratification with *systematic* variable selection: measured confounding**





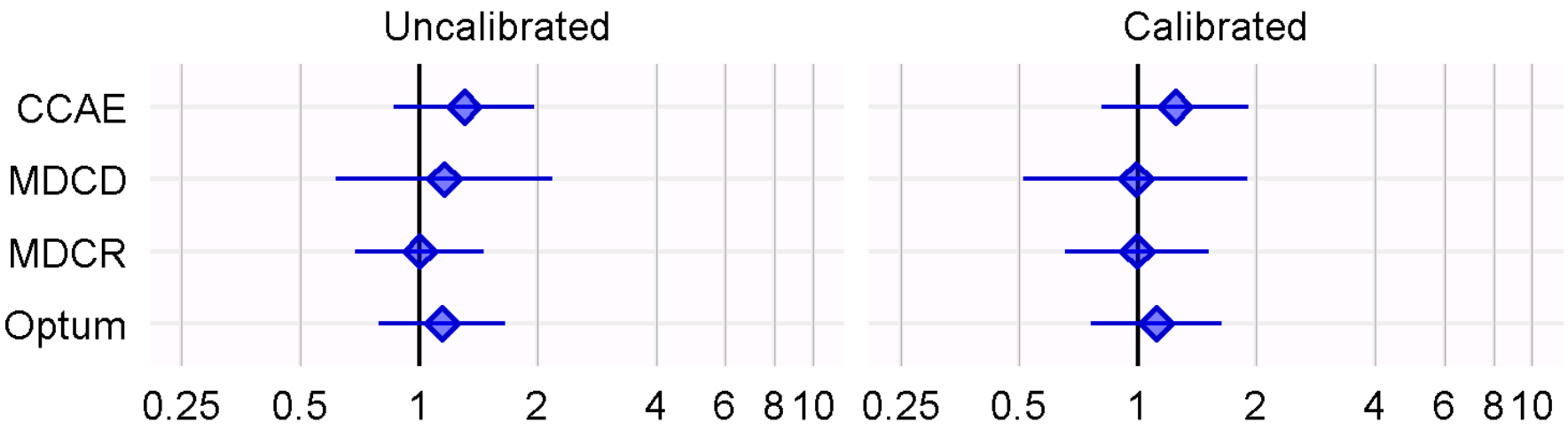
2. Confidence interval calibration using negative controls: unmeasured confounding





Addressing reproducibility

3. Multiple databases, locations, practice types

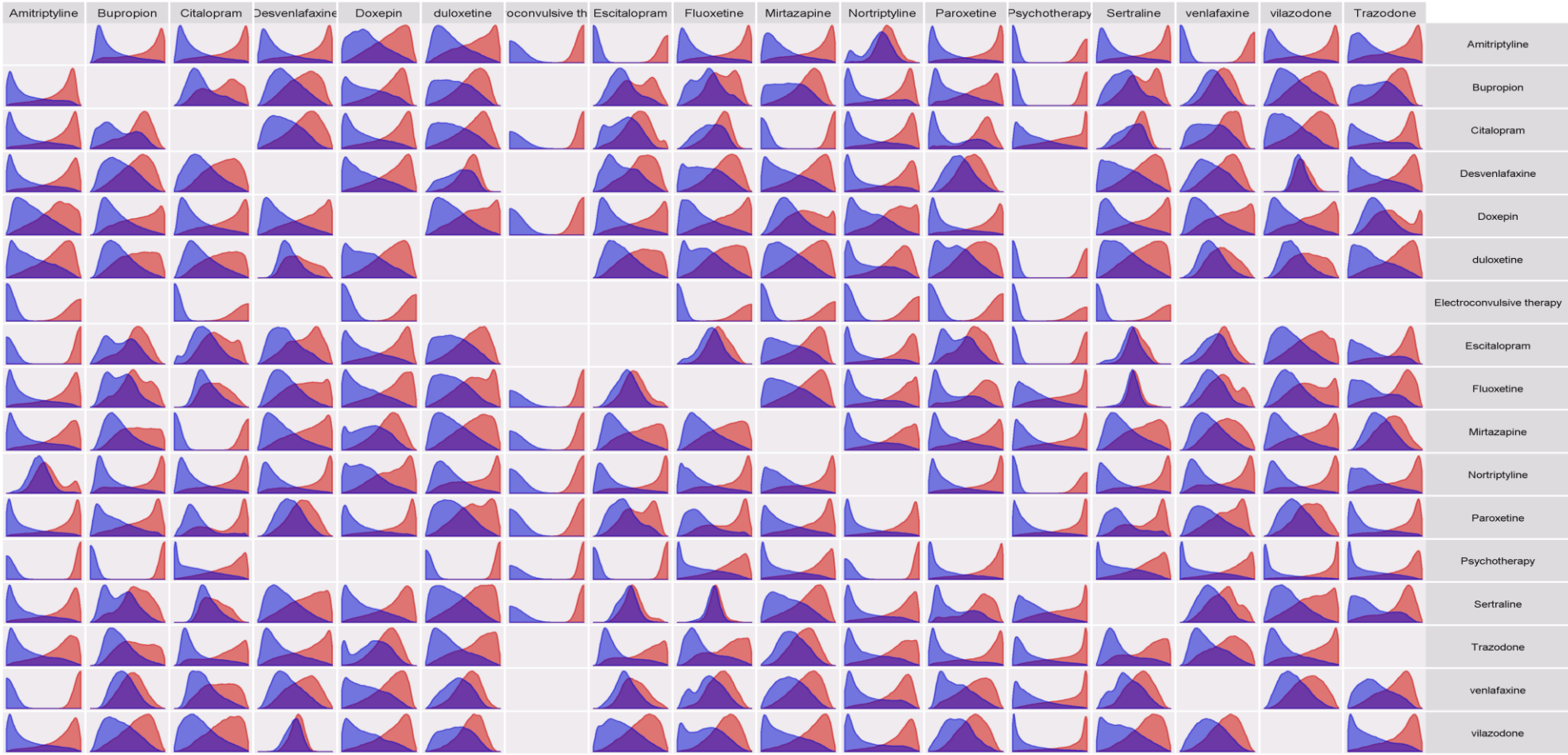


4. Publish all hypotheses, code, parameters, runs

```
□RL←1000
```

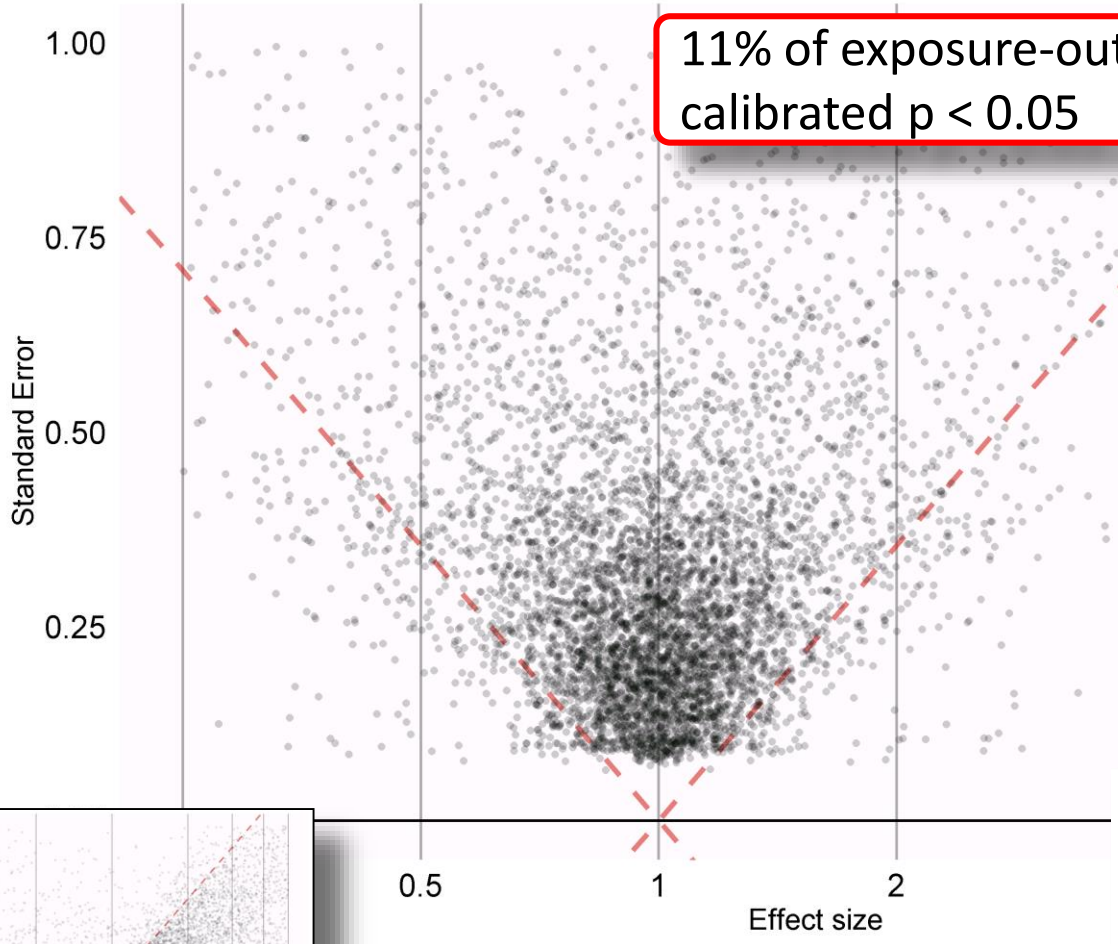
Addressing reproducibility

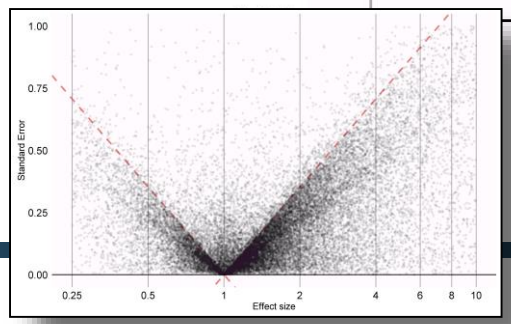
5. Carry out on aligned hypotheses at scale



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

OHDSI LEGEND Hypertension Study

Table 18. Oral Antihypertensive Drugs

Class	Drug	Usual Dose, Range (mg/d)*	Daily Frequency	Comments
Primary agents				
Thiazide or thiazide-type diuretics	Chlorthalidone	12.5–25	1	• Chlorthalidone is preferred on the basis of prolonged half-life and proven trial reduction of CVD. • Monitor for hyponatremia and hypokalemia, uric acid and calcium levels. • Use with caution in patients with history of acute gout unless patient is on uric acid–lowering therapy.
	Hydrochlorothiazide	25–50	1	
	Indapamide	1.25–2.5	1	
	Metolazone	2.5–10	1	
ACE inhibitors	Benazepril	10–40	1 or 2	• Do not use in combination with ARBs or direct renin inhibitor. • There is an increased risk of hyperkalemia, especially in patients with CKD or in those on K^+ supplements or K^+-sparing drugs. • There is a risk of acute renal failure in patients with severe bilateral renal artery stenosis. • Do not use if patient has history of angioedema with ACE inhibitors. • Avoid in pregnancy.
	Captopril	12.5–150	2 or 3	
	Enalapril	5–40	1 or 2	
	Fosinopril	10–40	1	
	Lisinopril	10–40	1	
	Moexipril	7.5–30	1 or 2	
	Perindopril	4–16	1	
	Quinapril	10–80	1 or 2	
	Ramipril	2.5–10	1 or 2	
	Trandolapril	1–4	1	
ARBs	Azilsartan	40–80	1	• Do not use in combination with ACE inhibitors or direct renin inhibitor. • There is an increased risk of hyperkalemia in CKD or in those on K^+ supplements or K^+-sparing drugs. • There is a risk of acute renal failure in patients with severe bilateral renal artery stenosis. • Do not use if patient has history of angioedema with ARBs. Patients with a history of angioedema with an ACE inhibitor can receive an ARB beginning 6 weeks after ACE inhibitor is discontinued. • Avoid in pregnancy.
	Candesartan	8–32	1	
	Eprosartan	600–800	1 or 2	
	Irbesartan	150–300	1	
	Losartan	50–100	1 or 2	
	Olmesartan	20–40	1	
	Telmisartan	20–80	1	
	Valsartan	80–320	1	
CCB—dihydropyridines	Amlodipine	2.5–10	1	• Avoid use in patients with HFrEF; amlodipine or felodipine may be used if required. • They are associated with dose-related pedal edema, which is more common in women than men.
	Felodipine	5–10	1	
	Isradipine	5–10	2	
	Nicardipine SR	5–20	1	
	Nifedipine LA	60–120	1	
	Nisoldipine	30–90	1	
CCB—nondihydropyridines	Diltiazem SR	180–360	2	• Avoid routine use with beta blockers because of increased risk of bradycardia and heart block. • Do not use in patients with HFrEF. • There are drug interactions with diltiazem and verapamil (*CYP3A4* major substrate and moderate inhibitor).
	Diltiazem ER	120–480	1	
	Verapamil IR	40-80	3	
	Verapamil SR	120–480	1 or 2	
	Verapamil-delayed onset ER (various forms)	100–480	1 (in the evening)	

Whelton et al., Hypertension 2018

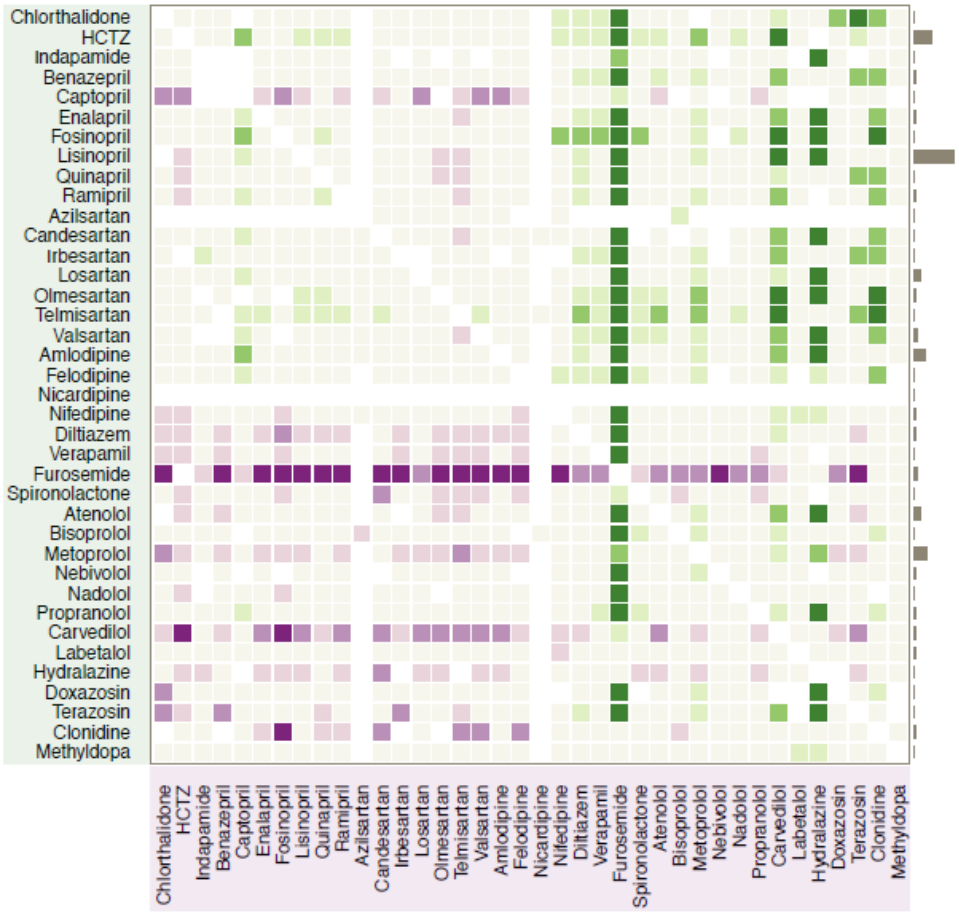
Evidence to support the guideline

- 40 randomized trials
- Most decisions are "expert opinion"

Comparisons of hypertension treatments

	Theoretical	Observed (n > 2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020

Cardiovascular efficacy by drug



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

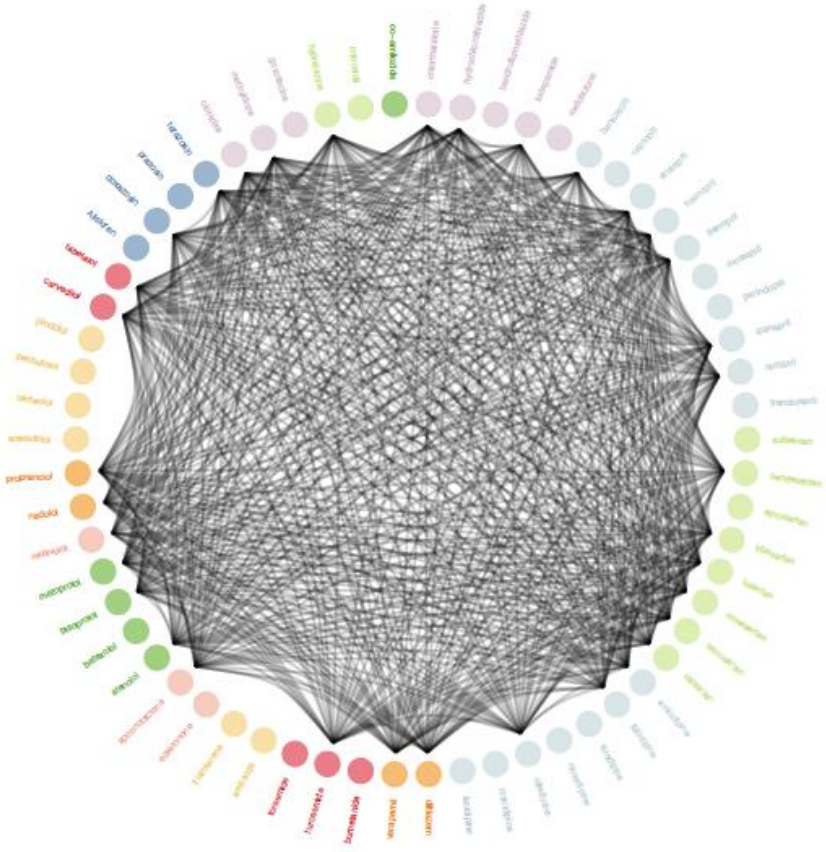
Prescriptions are not written at the class-level; must choose an individual drug for the patient

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

LEGEND knowledge base for hypertension

Head-to-head HTN drug comparisons

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



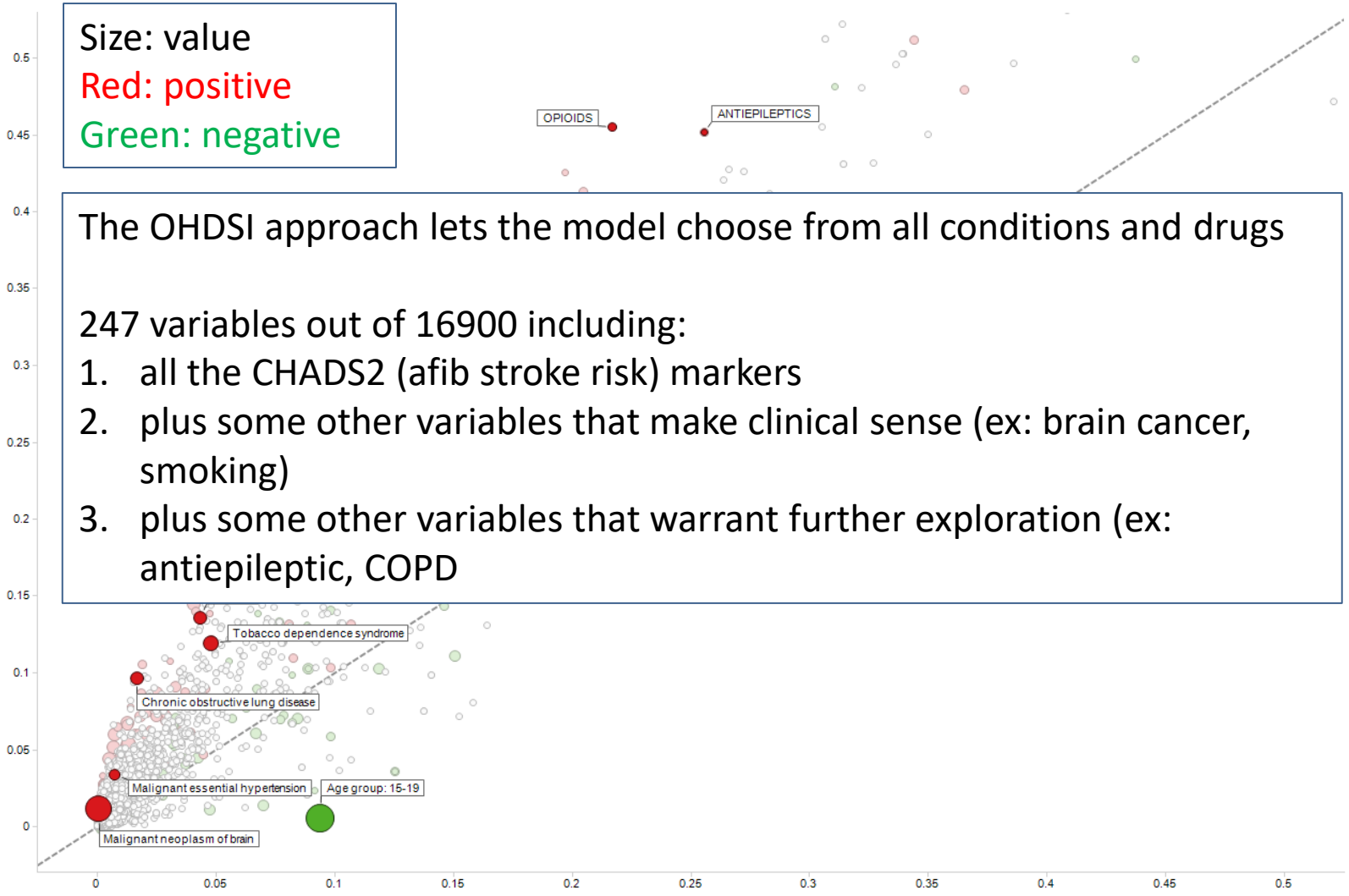
- Comparisons: 10,278
- $N = 3502 - [212K] - 1.9M$

OHDSI in Action

- Patient-level prediction

Patient-level prediction
Stroke risk in atrial fibrillation

Size: value
Red: positive
Green: negative

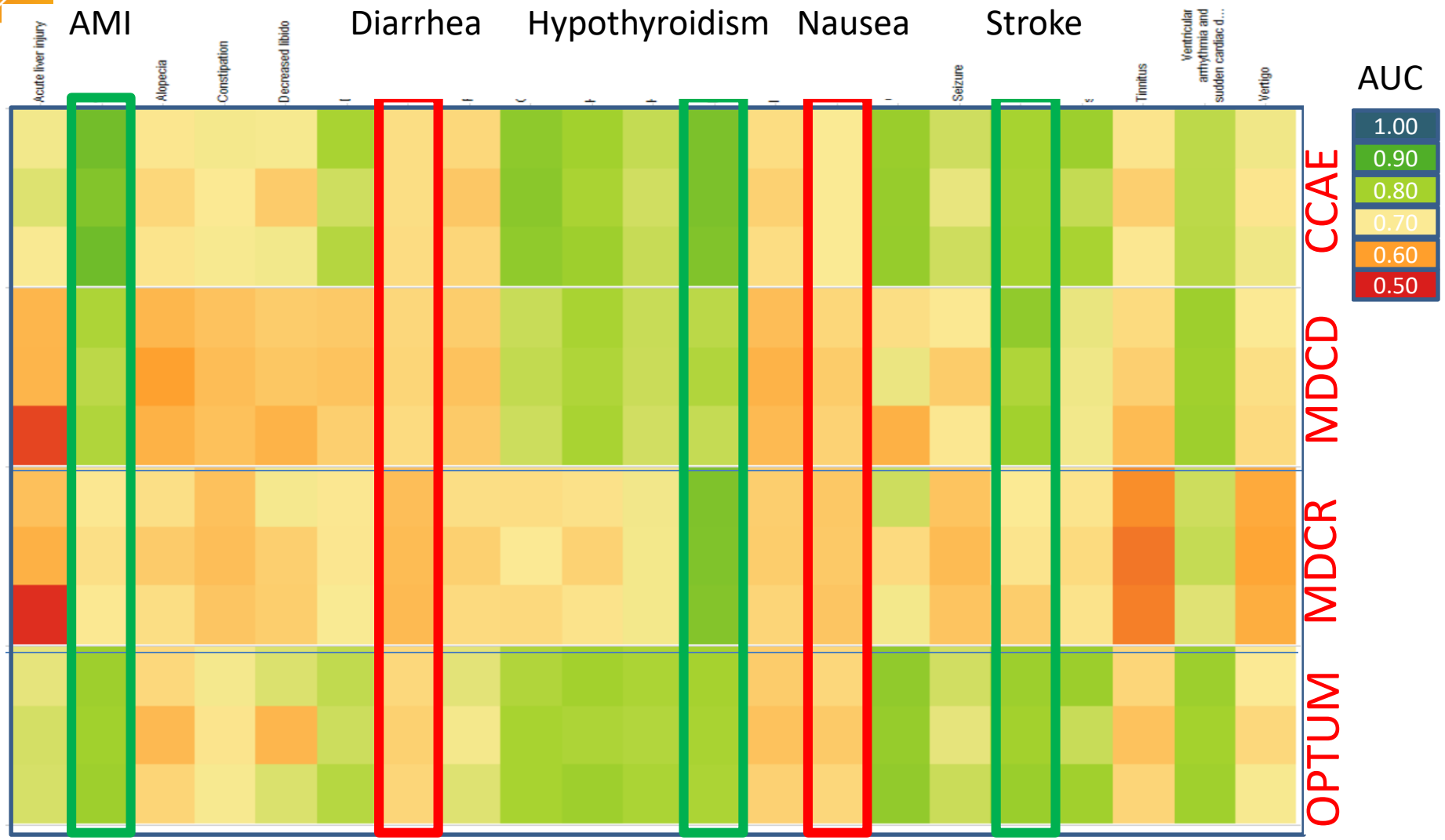


The OHDSI approach lets the model choose from all conditions and drugs

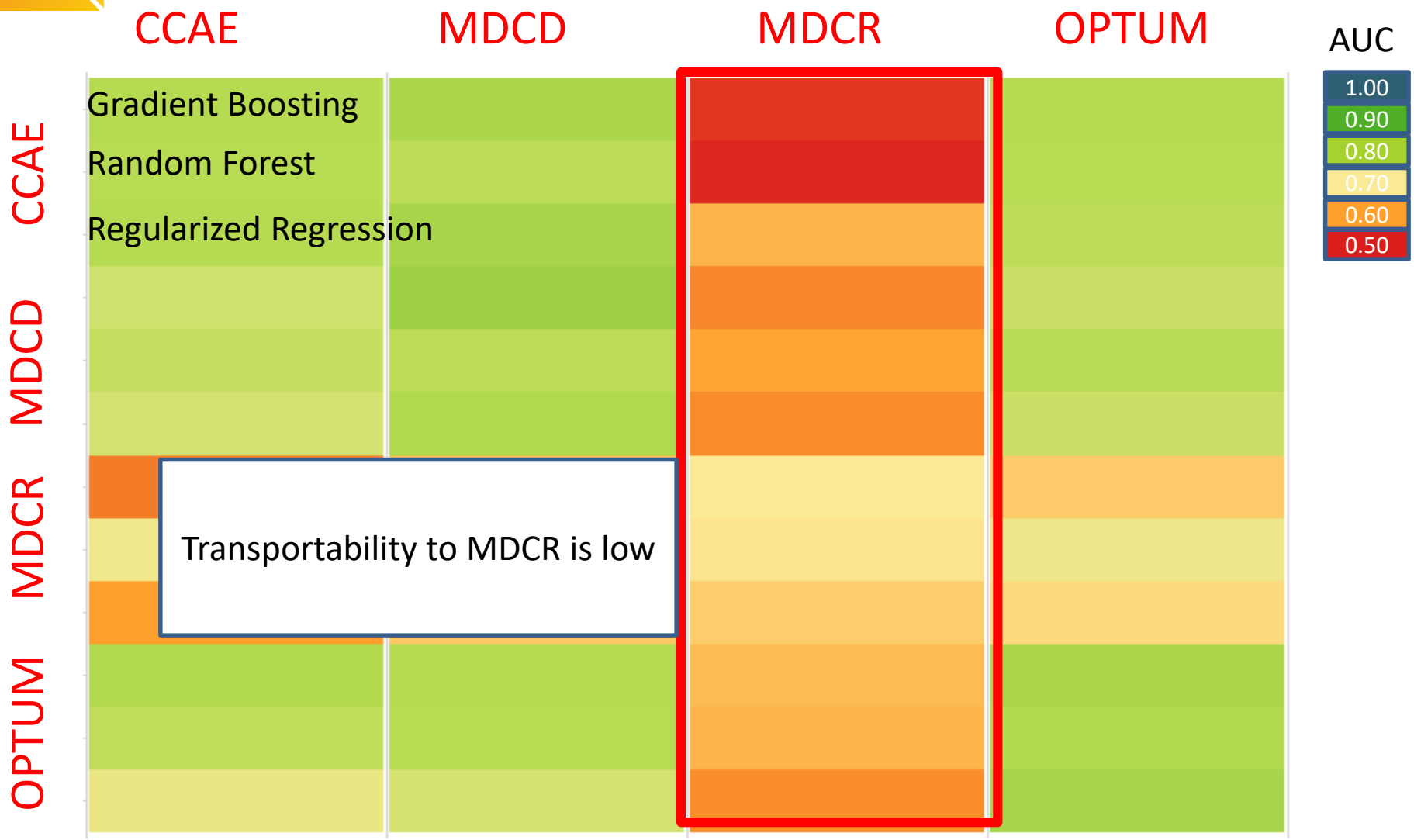
247 variables out of 16900 including:

1. all the CHADS2 (afib stroke risk) 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



Transportability Assessment Stroke



Conclusions

- It is feasible to create an enormous international research network
- Sites will volunteer to run studies
- Completely open
  - Data model, methods, tools
- Concrete approach to address the credibility crisis
- Prediction
  - It's not ROC area
  - New, useful information

Join the journey

http://ohdsi.org