Biomedical Informatics

discovery and impact

Drawing reproducible conclusions from observational clinical data with OHDSI

George Hripcsak, MD, MS

Biomedical Informatics, Columbia University

Medical Informatics Services, NewYork-Presbyterian

COLUMBIA UNIVERSITY MEDICAL CENTER

NewYork-Presbyterian

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

http://ohdsi.org

OHDSI's global research community

- >200 registered collaborators, 2500 forum participants 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/

Paucity of evidence

A tiny fraction of potential questions have answers

All health outcomes of interest

All drugs

Inability to reproduce evidence

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.

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"

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.

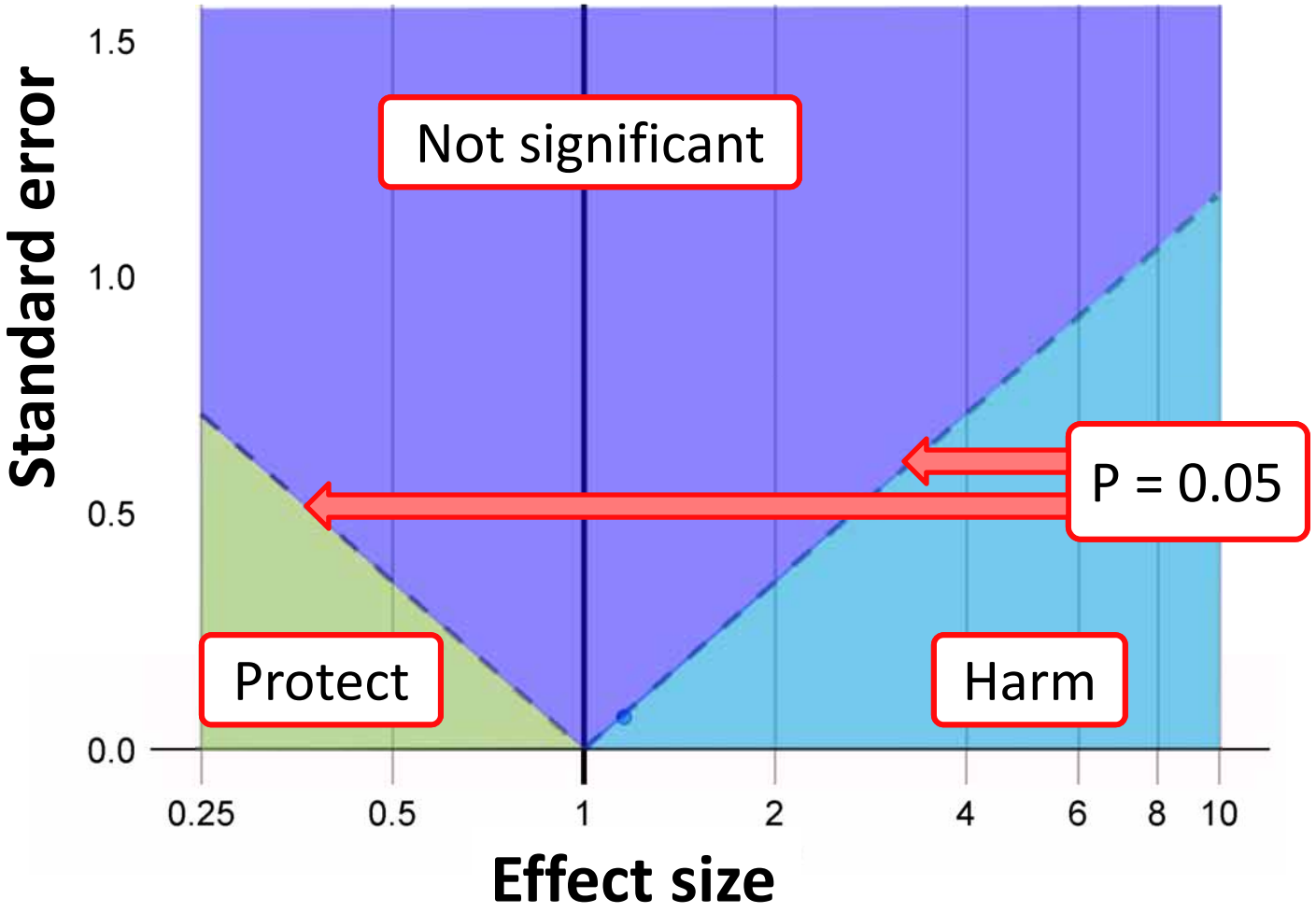
Conclusions The risk of oesophageal cancer increased with 10 or more prescriptions for oral bisphosphonates and with prescriptions over about a five year period. In Europe and North America, the incidence of oesophageal cancer at age 60-79 is typically 1 per 1000 population over five years, and this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

INTRODUCTION

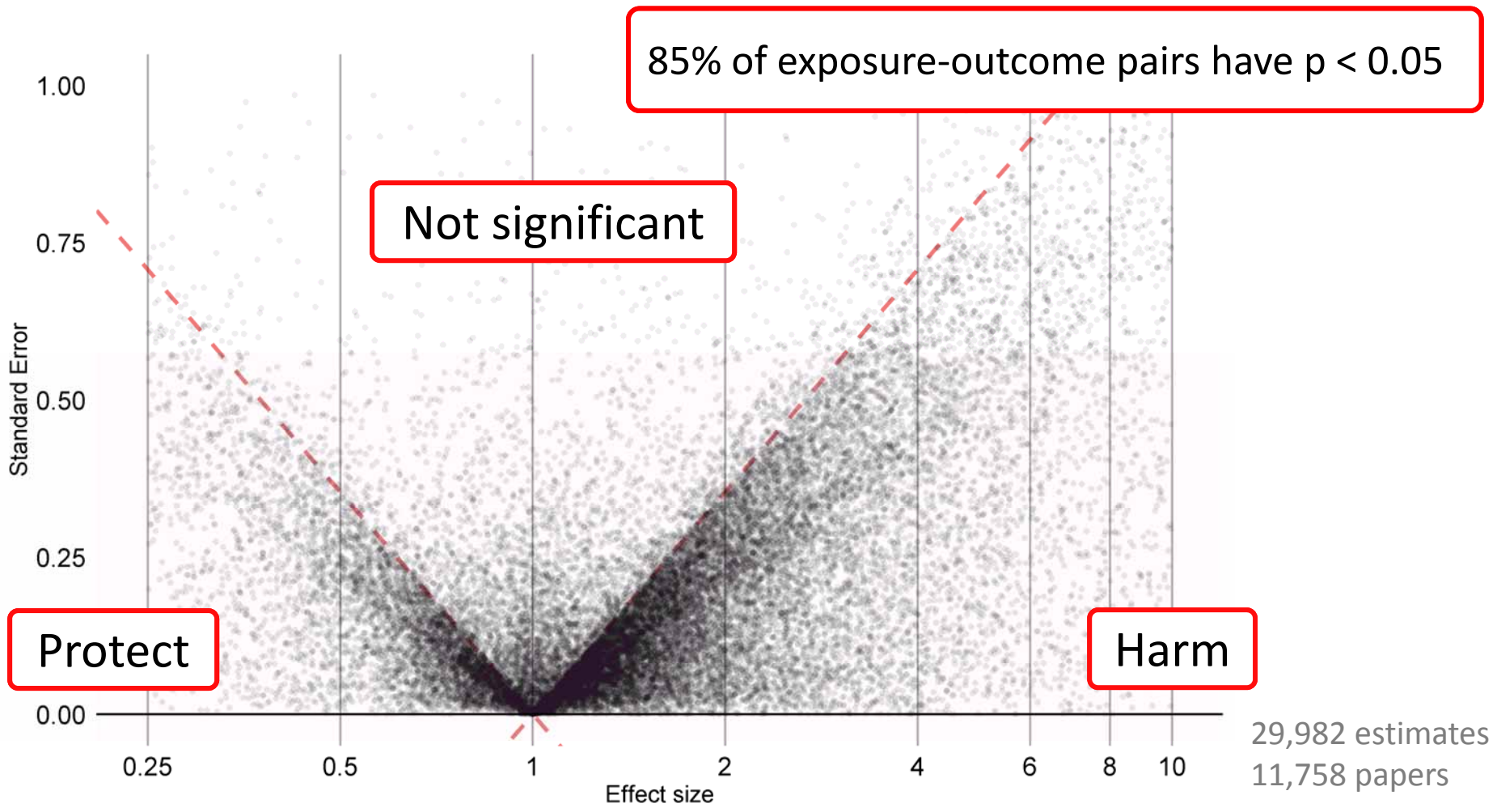
Adverse gastrointestinal effects are common among people who take oral bisphosphonates for the prevention and treatment of osteoporosis; they range from dyspepsia, nausea, and abdominal pain to erosive oesophagitis and oesophageal ulcers.[1] Recent case reports have suggested a possible increase in the 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"

Evaluate observational research

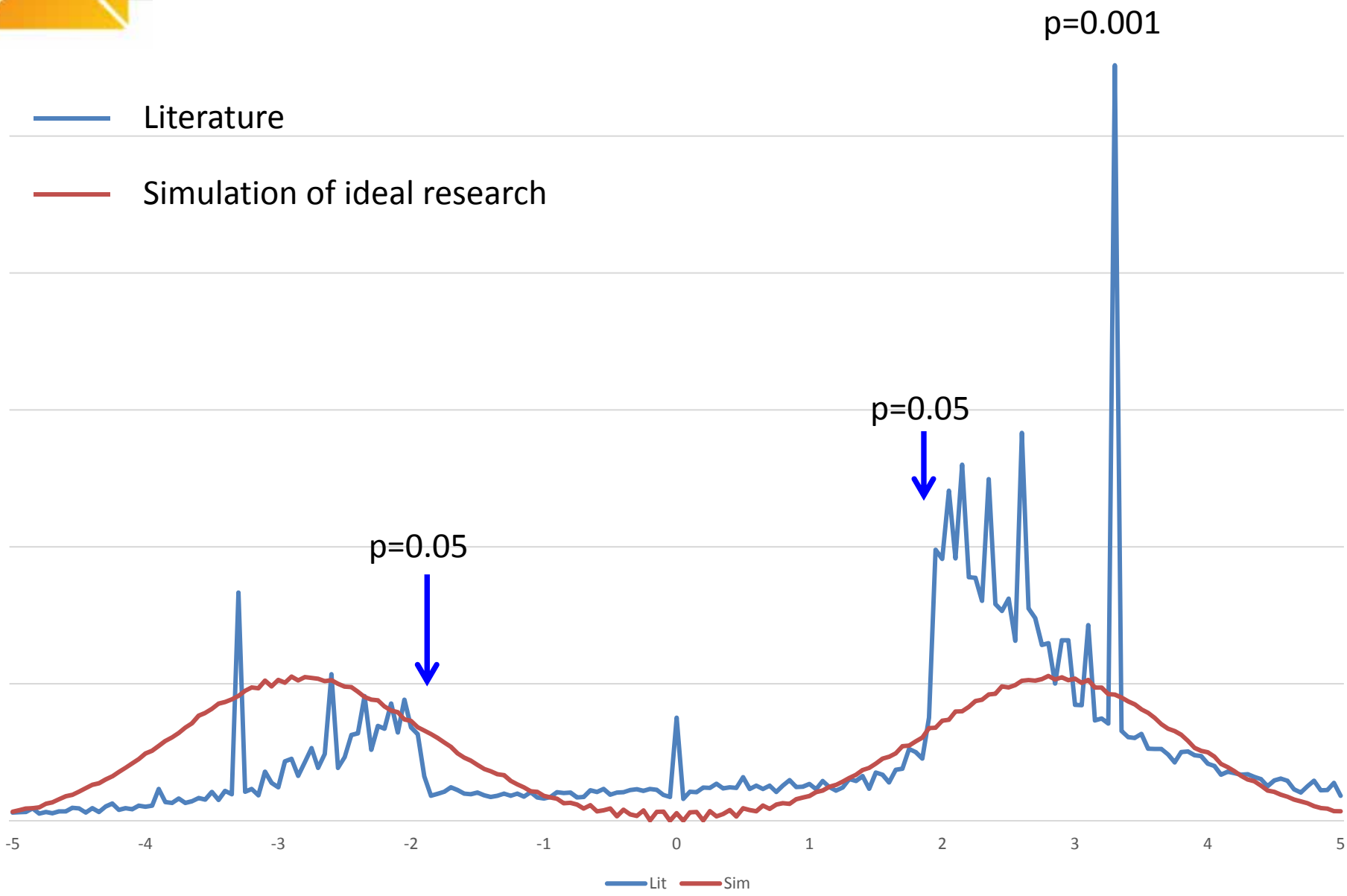


Observational research results in literature



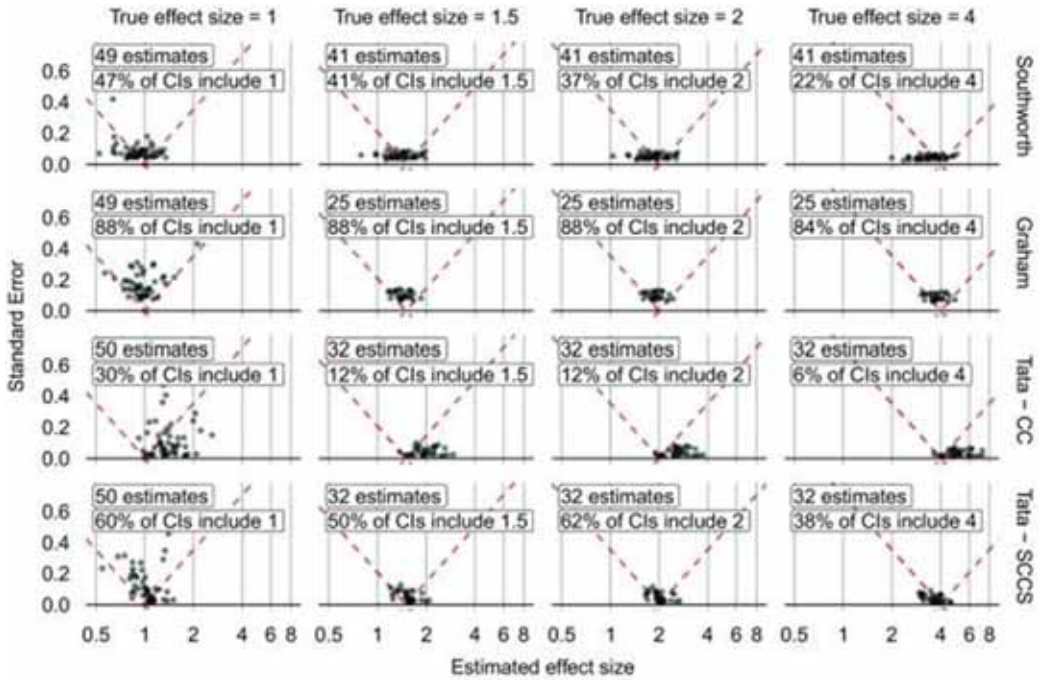
Schuemie, Phil Trans A, 2018

Publication bias



Over-optimism of p-values and CIs

- Look at four published studies
  - Use 50 negative controls to assess CI
- "95%" CIs cover only 30%, 47%, 60%, 88%



Schuemie, PNAS 2018

Consequences

- Overcall positive results
- Throw negative ones away

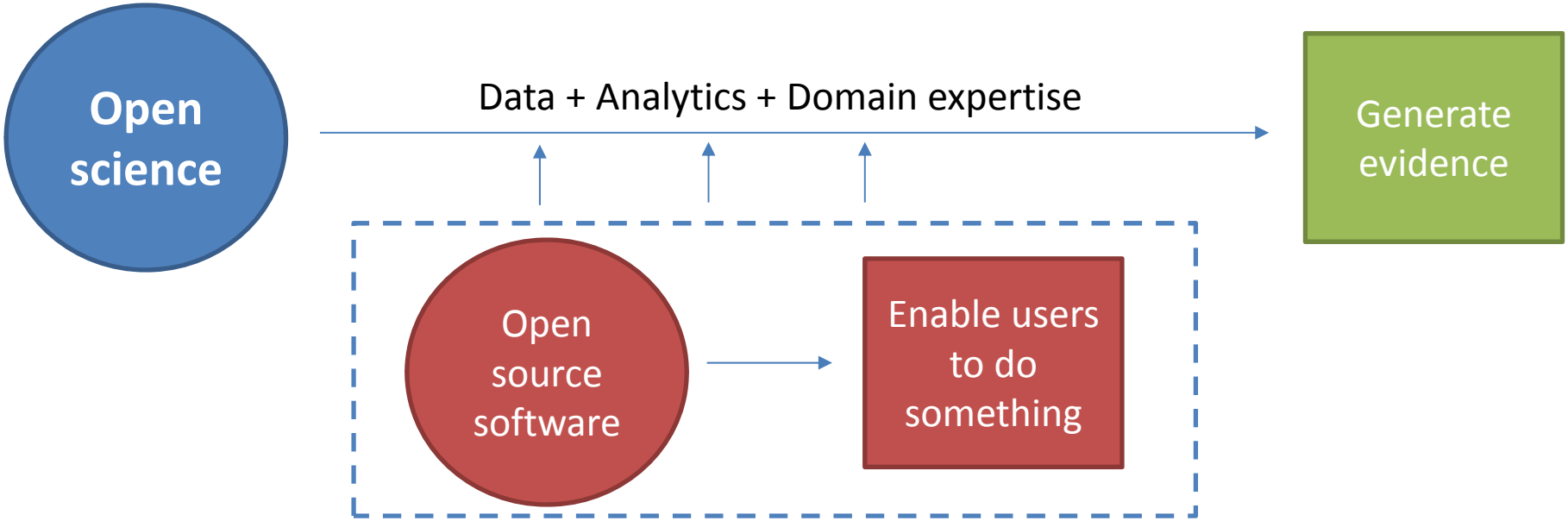
➢ Conflicting, unreliable evidence

➢ Too little evidence and what is there is poor

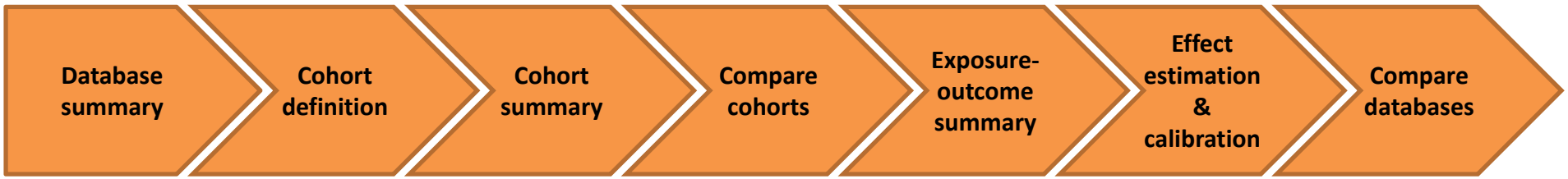
  – Woody Allen:
    "Boy, the food at this place is really terrible."
    "Yeah, I know; and such small portions."

  – It's *just* observational research

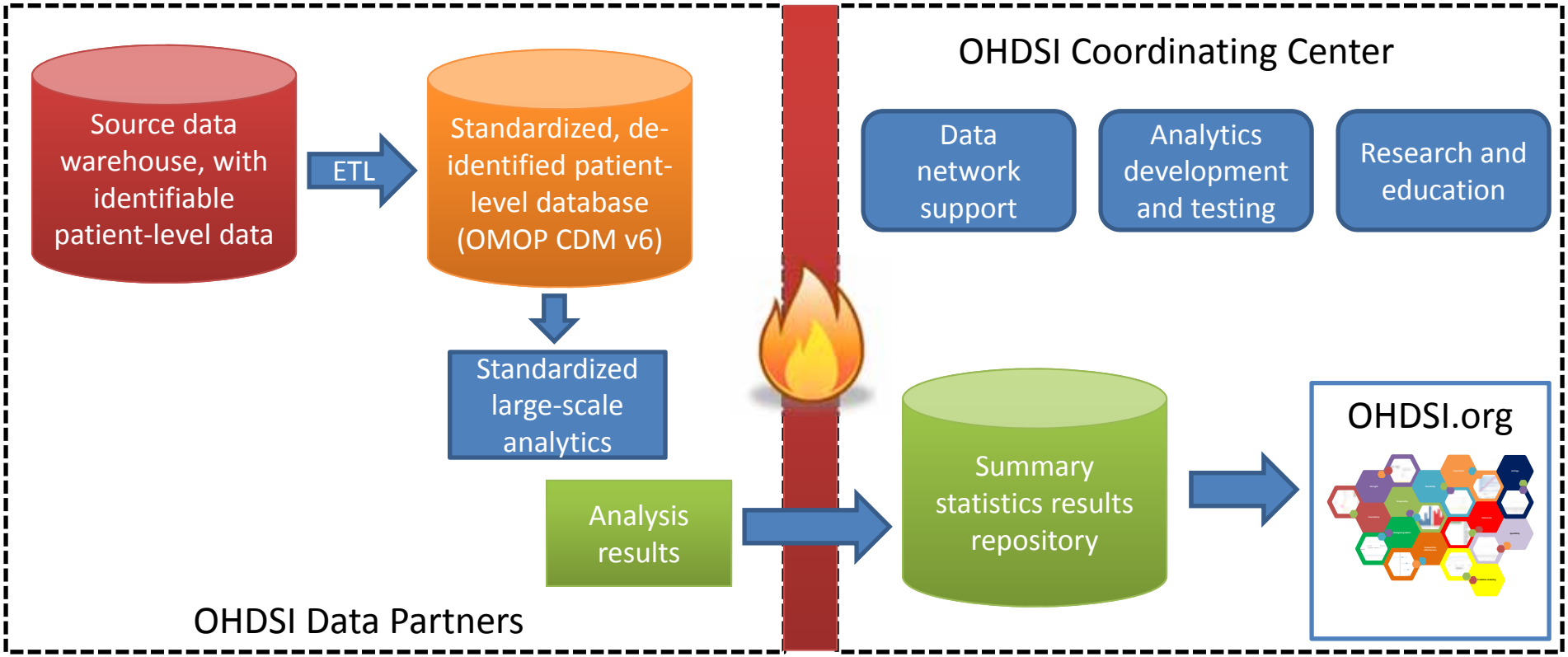
OHDSI is Open Science



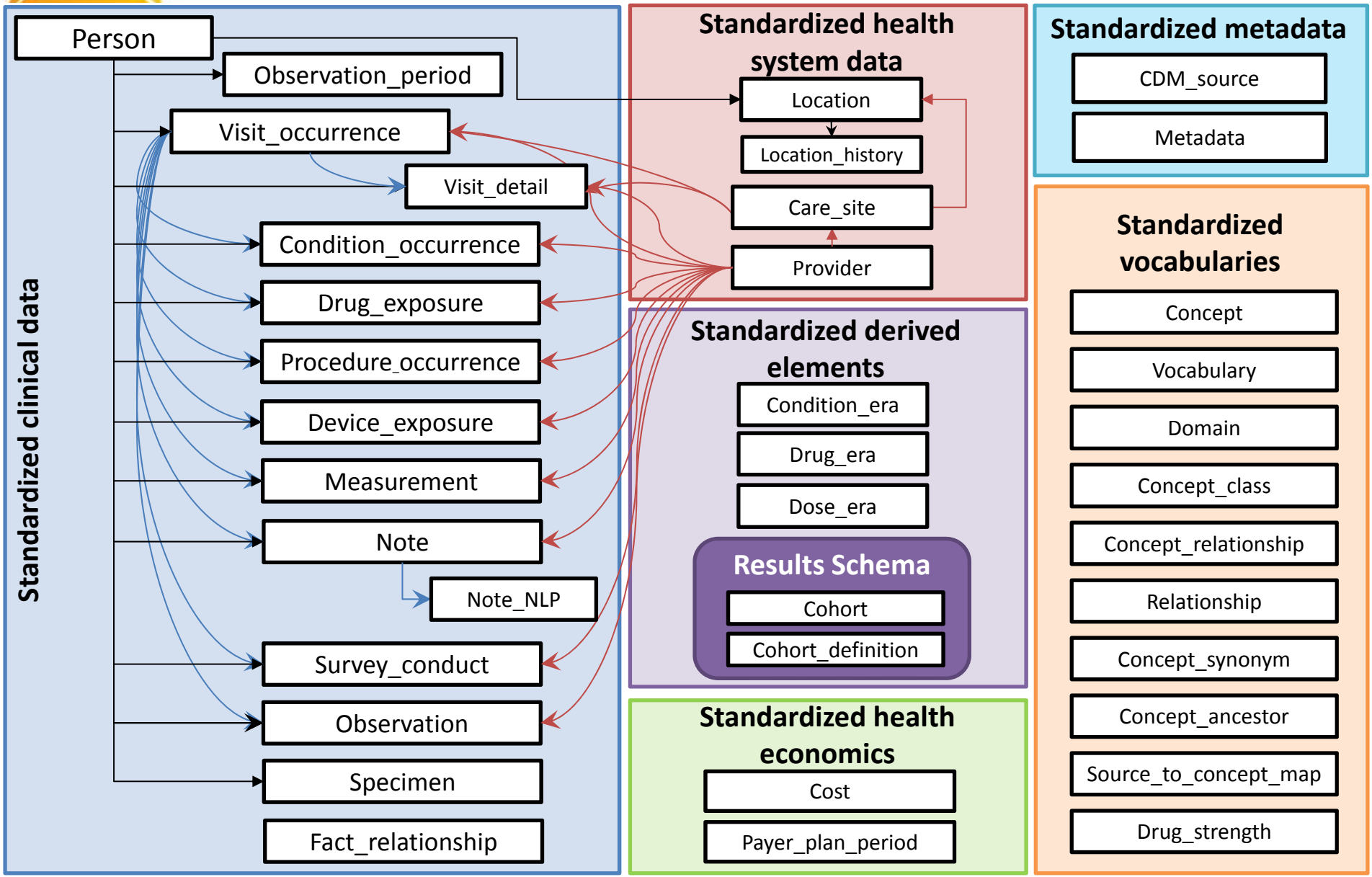
Standardized, transparent workflows



How OHDSI Works



Deep information model
OMOP CDM Version 6

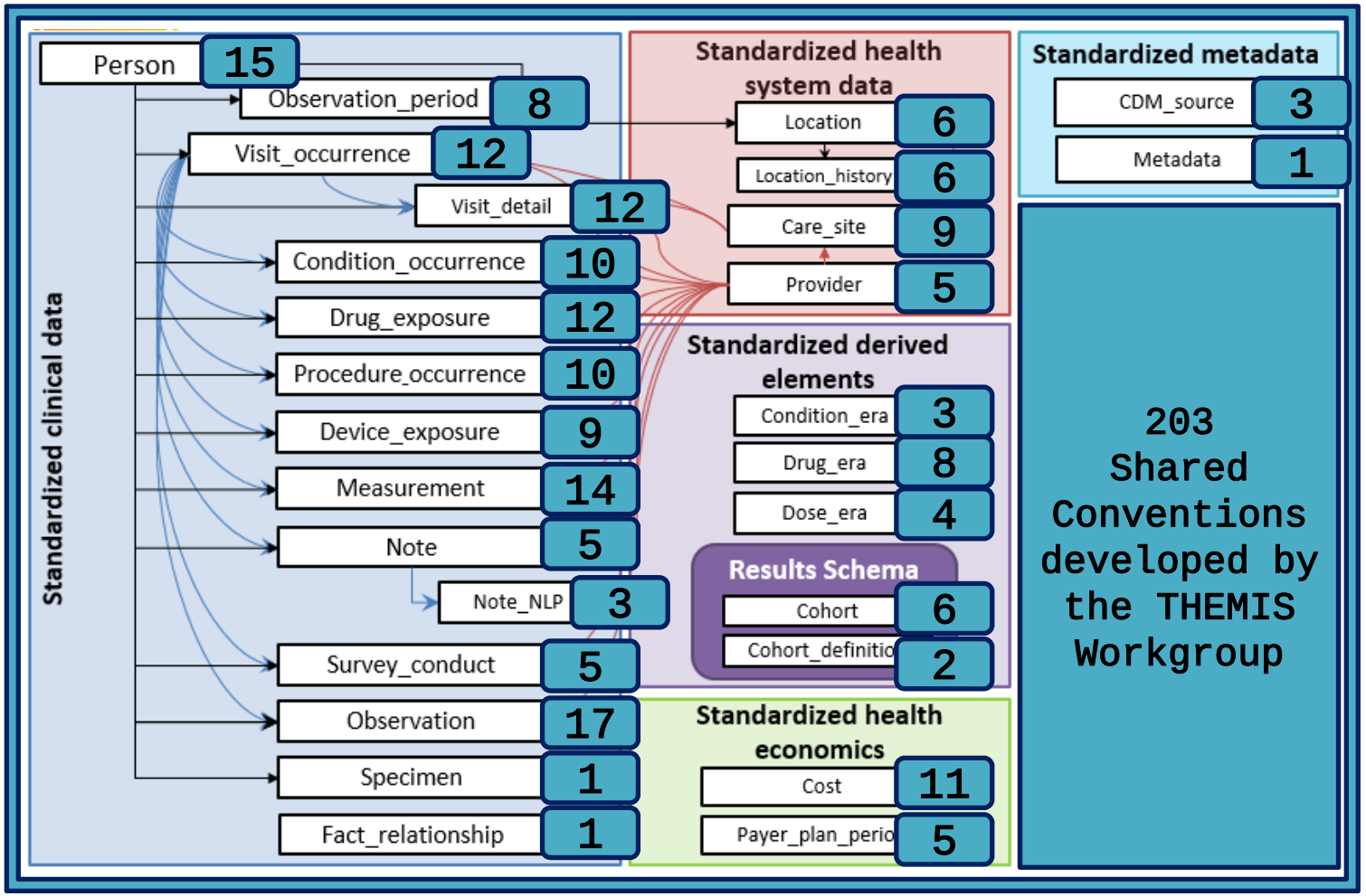


Extensive international vocabularies

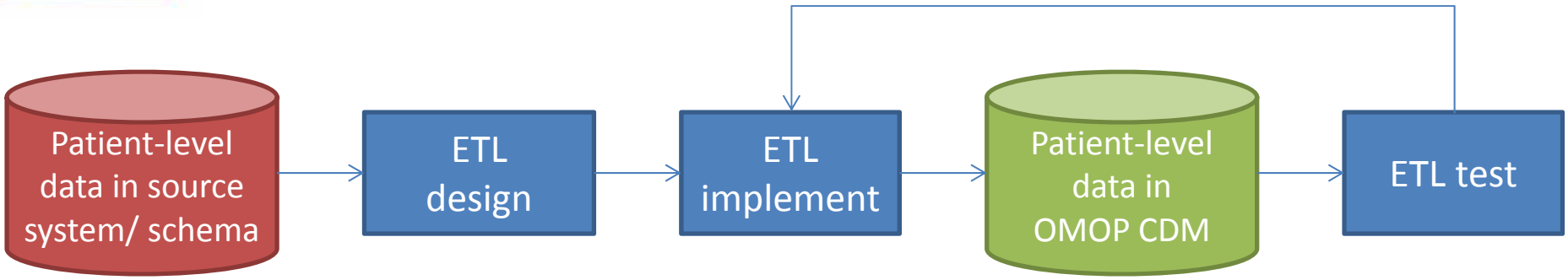
Breakdown of OHDSI concepts by domain, standard class, and vocabulary

Standardized conventions

Need more than vocabulary and schema



Tools to help you map data



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

http://github.com/OHDSI

Data Quality: ACHILLES Heel

Data Quality Messages

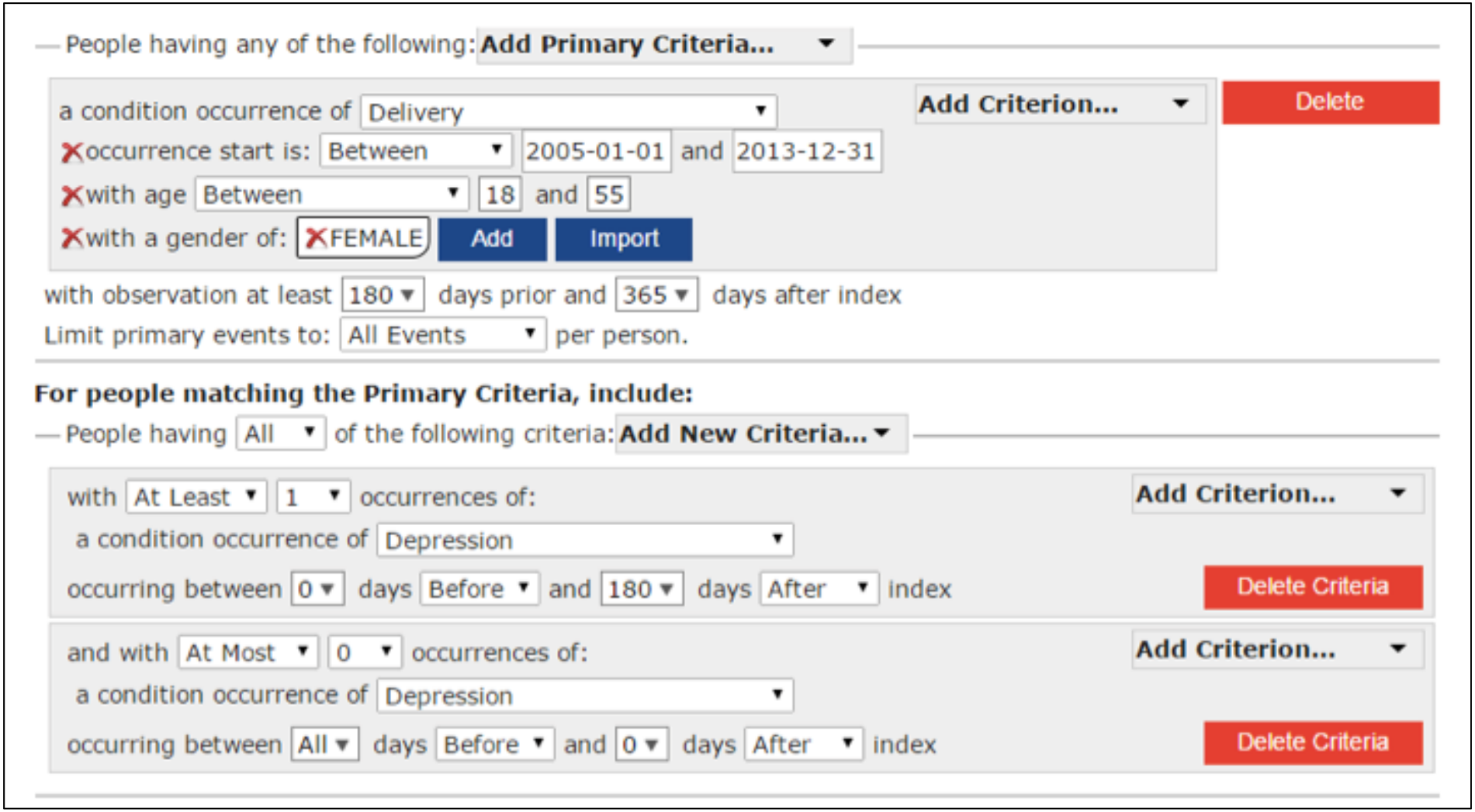
Search: Show / hide columns

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

Tools for analytics: CYCLOPS

- Cyclic Coordinate Descent for Logistic, Poisson and Survival Analysis
  - Open source toolkit
  - 100 million cases x million columns
  - regularized regression
  - R library written in C
  - No one else can do this scale

ATLAS to build, visualize, and analyze cohorts



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?

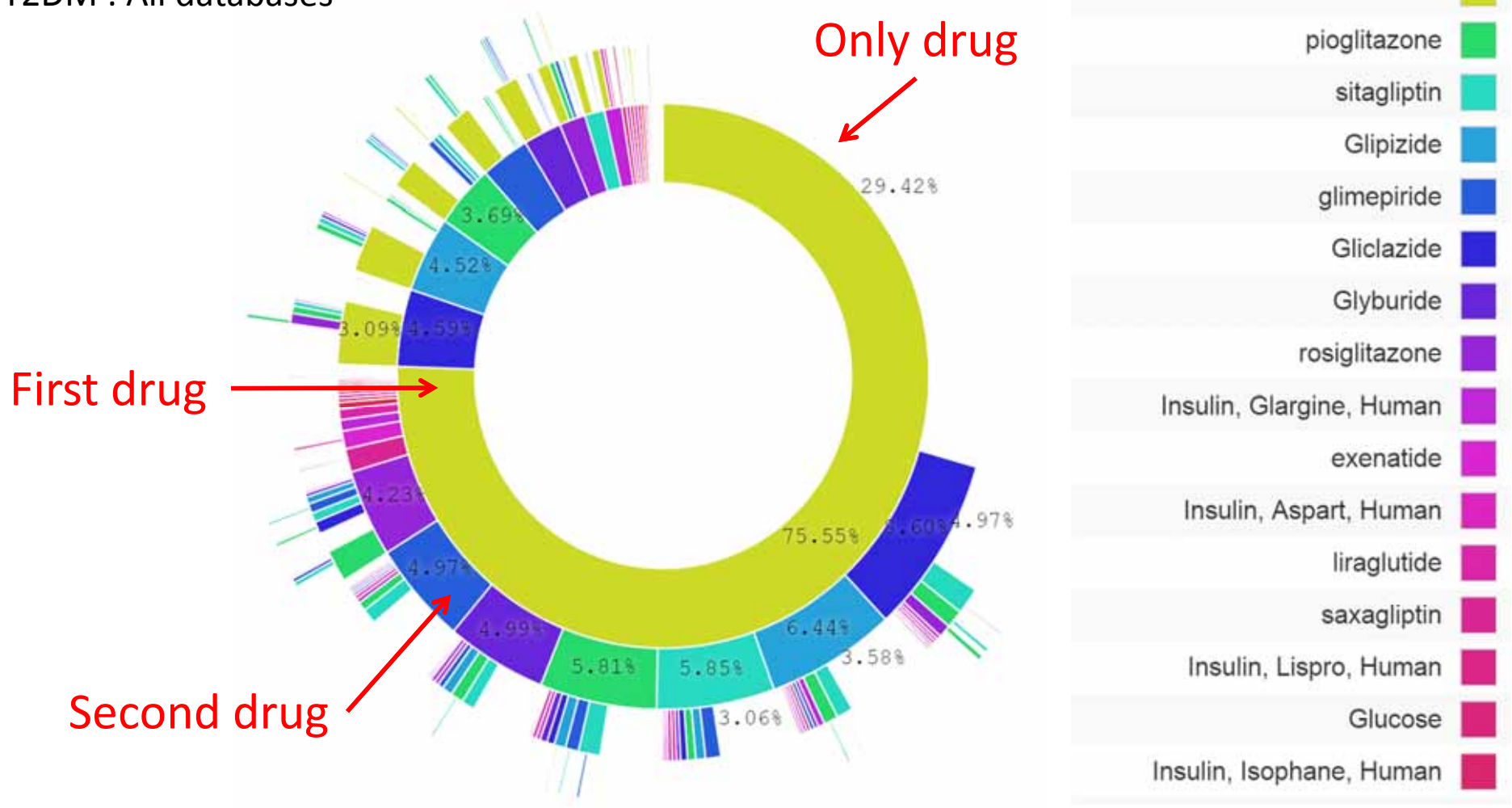
Characterization

- Treatment pathways
  - How are patients currently treated?

Treatment pathways for diabetes

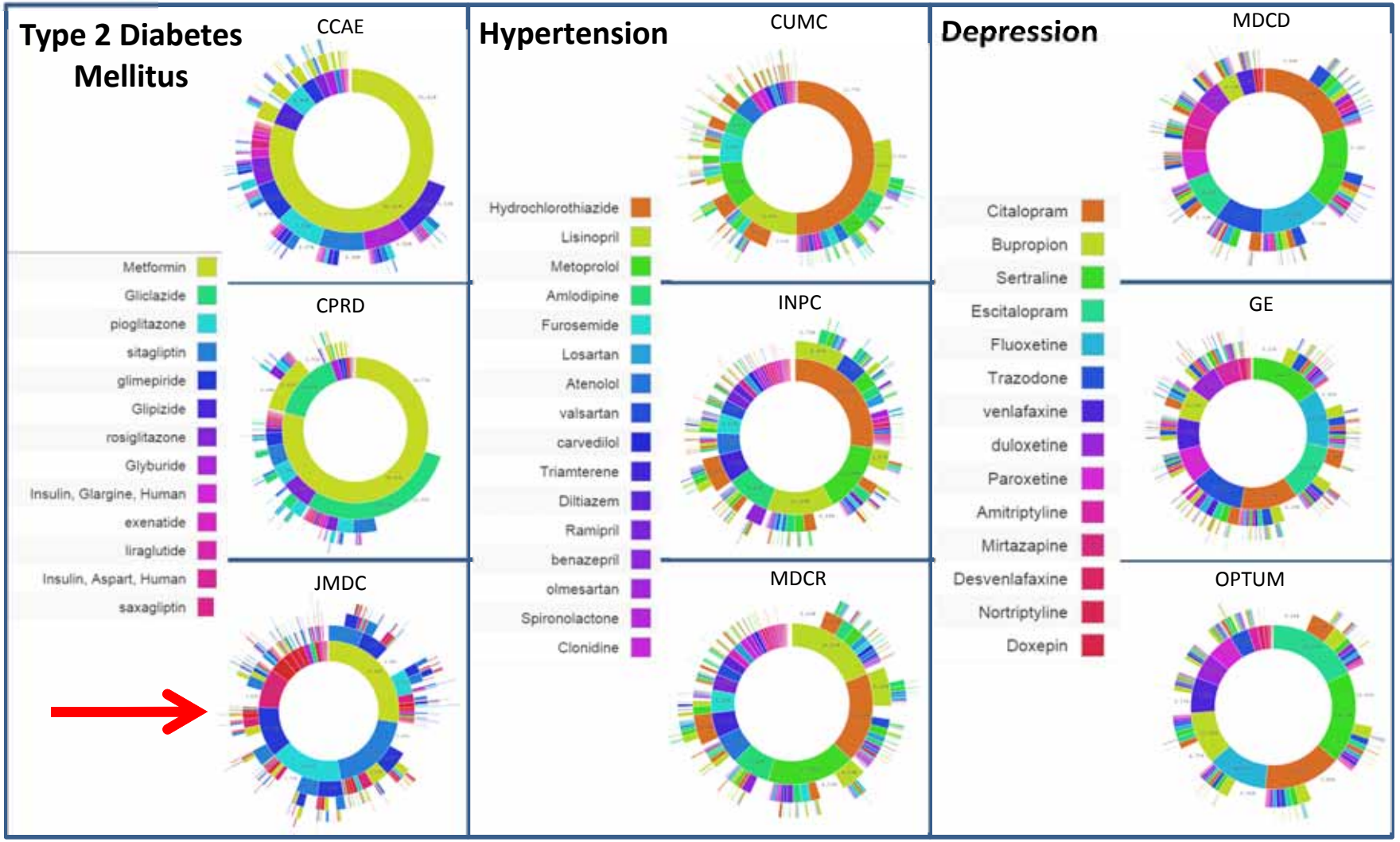
on 240M patients in 5 countries, 12 databases

T2DM : All databases



Treatment pathways

Heterogeneity within and across health systems



25% of HTN patients were on a unique pathway

Hripcsak, PNAS 2016

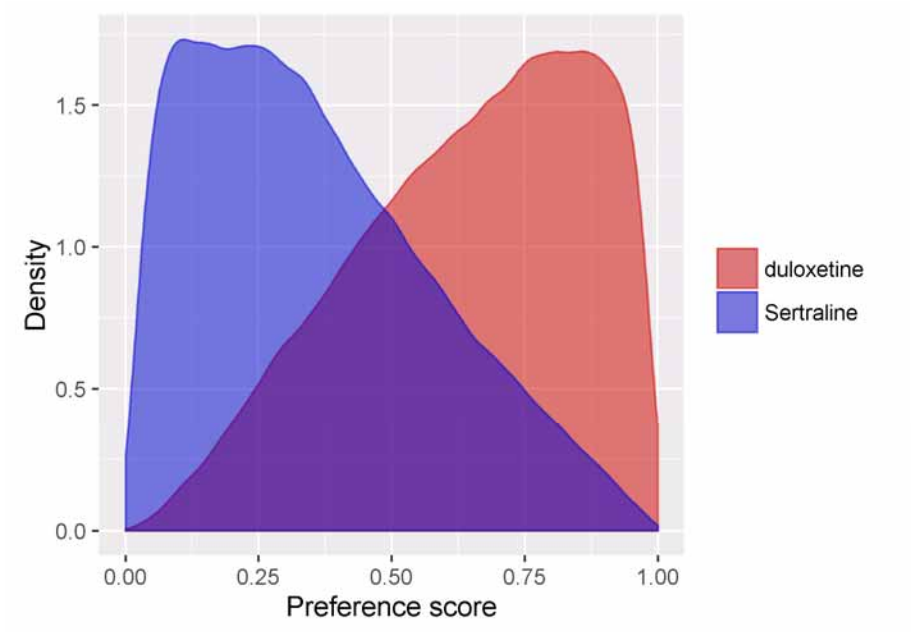
Population-level estimation

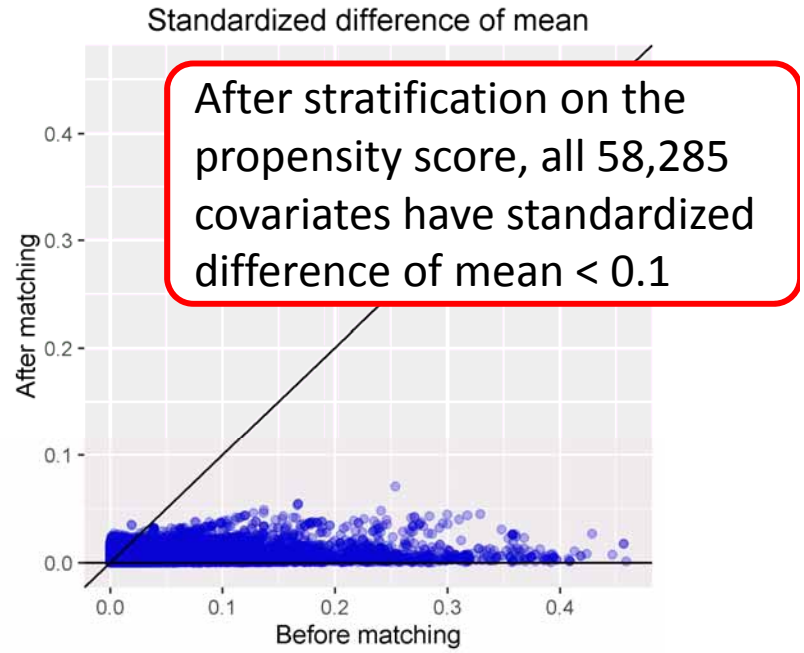
- Causality

OHDSI Reproducible research

1. Address confounding that is measured

- **Propensity stratification**
- ***Systematic* (not manual) variable selection**
  - **Balance 58,285 variables ("Table 1")**



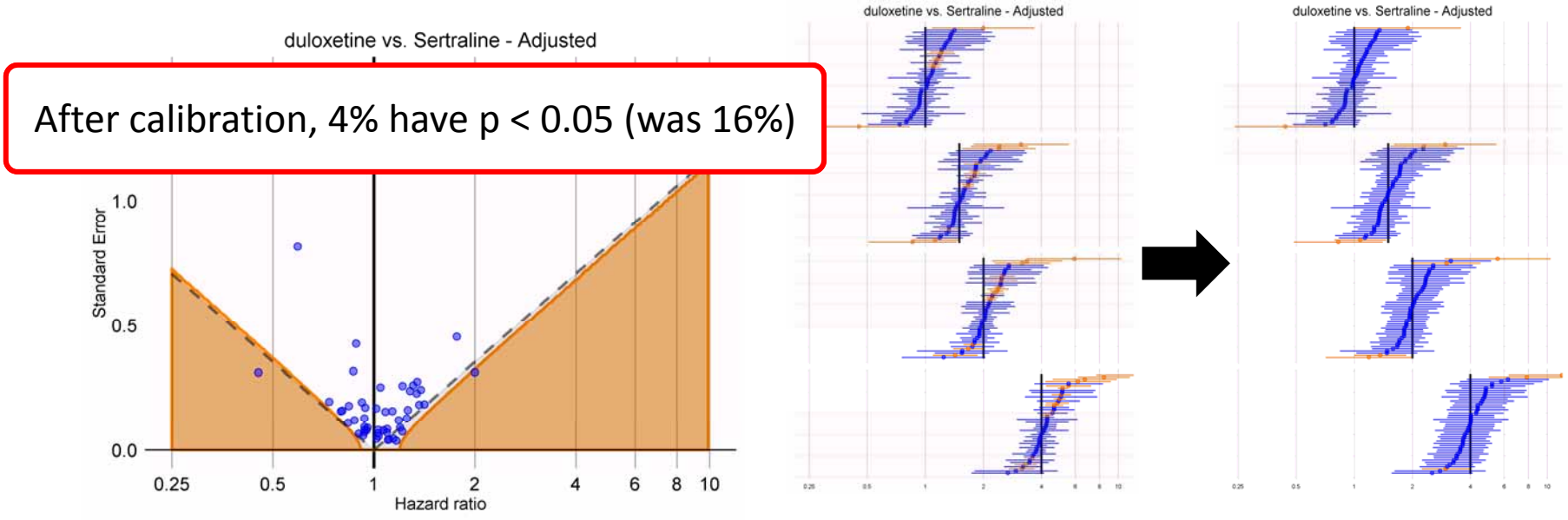


Tian, Int J Epi 2018

OHDSI Reproducible research

2. Unmeasured (residual) **confounding**

- **Confidence interval calibration**
  - **Adjust for all uncertainty, not just sampling**
- **Many (50-100) negative controls**
  - **Unique to OHDSI** (PNAS)

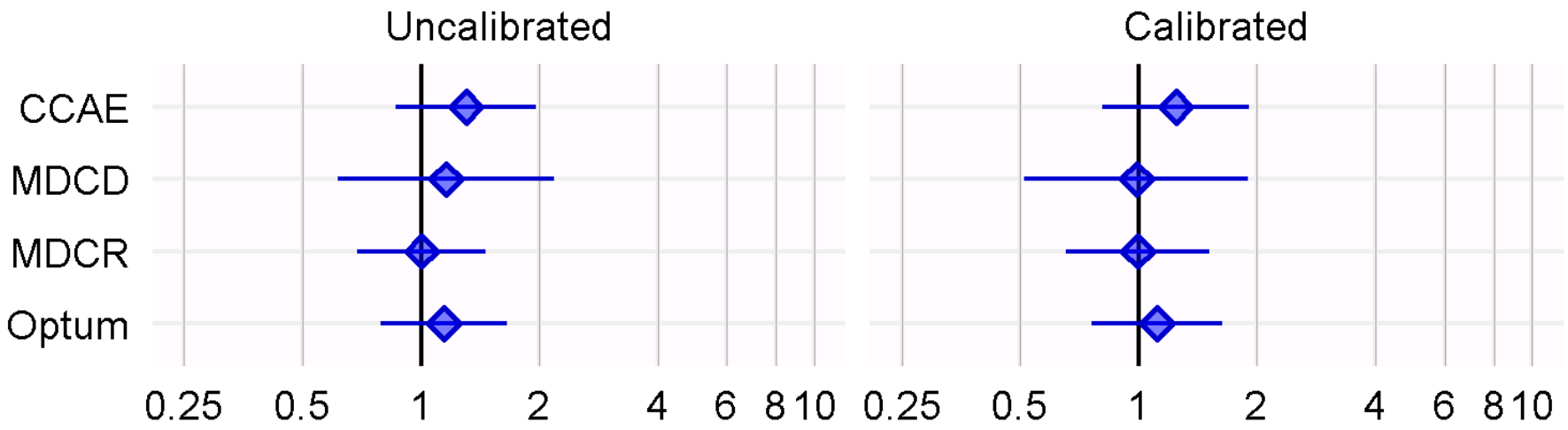


Schuemie, PNAS 2018

OHDSI Reproducible research

3. Multiple databases, locations, practice types

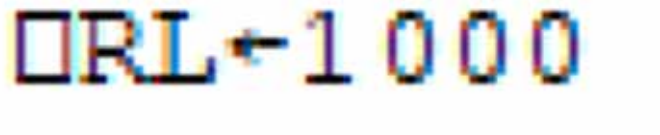
- **Exploit international OHDSI network**



OHDSI Reproducible research

4. Open: publish all

- **Hypotheses**
- **Code**
- **Parameters**
- **Runs**



Generating evidence for US FDA

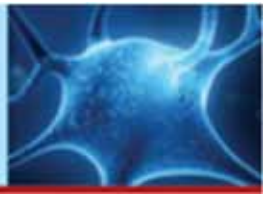
?

Potential Signals of Serious Risks/New Safety Information Identified by the FDA Adverse Event Reporting System (FAERS) between October - December 2015

Keppra (levetiracetam) tablet, oral solution, injection	Angioedema	FDA is evaluating the need for regulatory action.

- Protocol completed, code tested, study announced
  - 7 sites ran the code on 10 databases (5 claims / 5 EHR)
  - 59,367 levetiracetam patients, 74,550 phenytoin patients

Epilepsia®
Official Journal of the International League Against Epilepsy

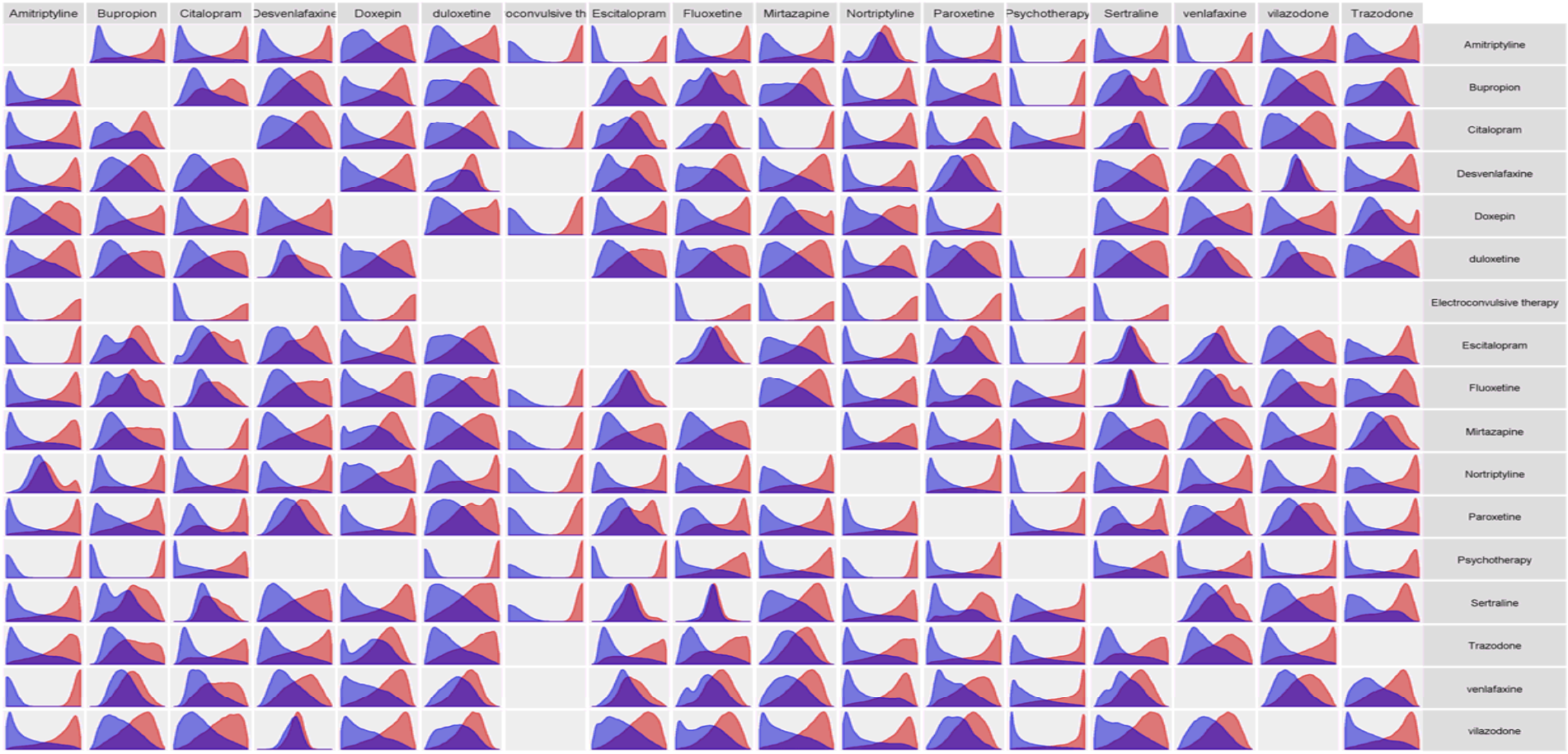


- No difference detected with small confidence interval
- Accepted essentially without revision
  - Add word to title, move diagram from supplement to body

Duke, Epilepsia 2017

OHDSI Reproducible research

5. Carry out on aligned hypotheses at large scale

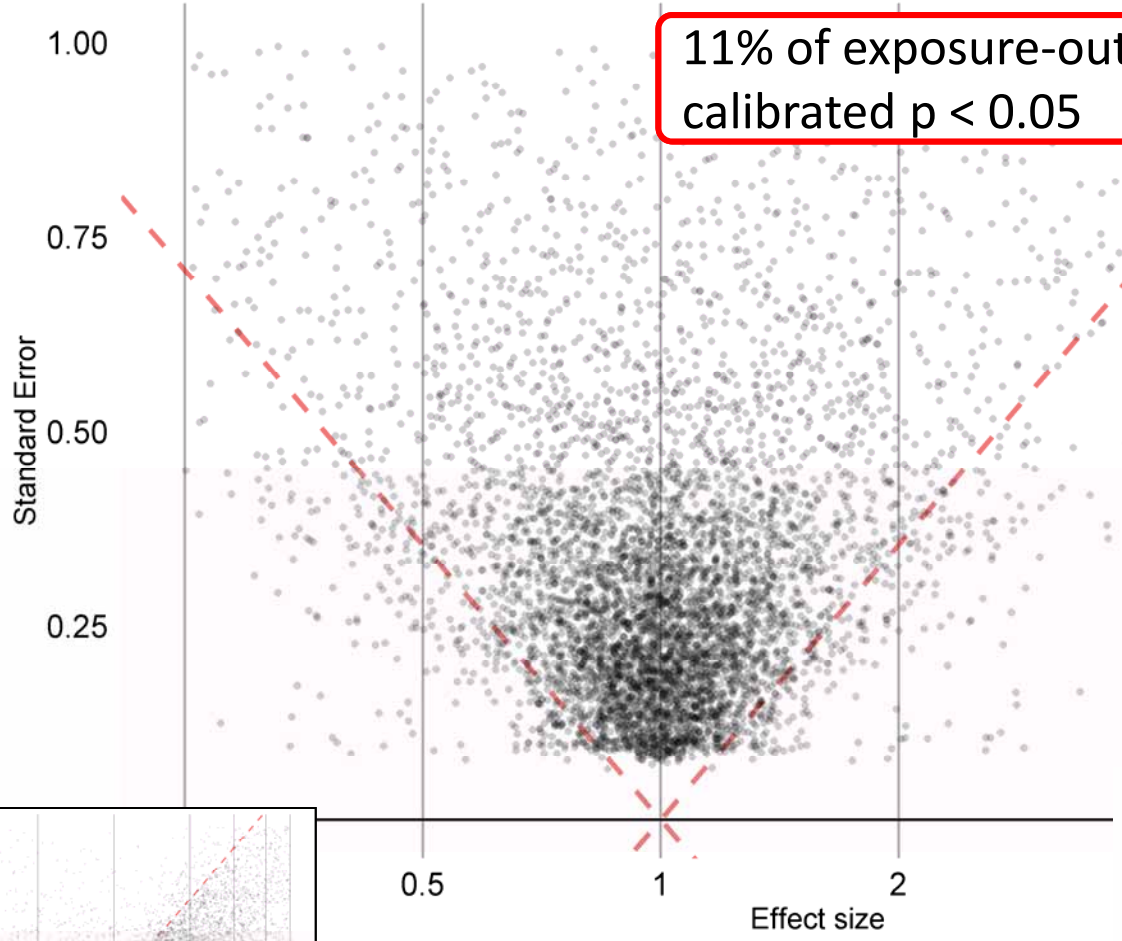


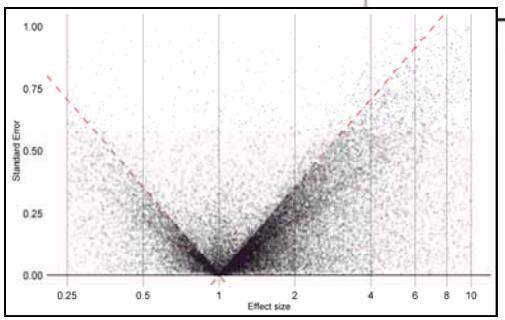
OHDSI Reproducible research

- Full diagnostics at each step
  - Tells us if this a result we can trust

Estimates are in line with expectations

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






PHILOSOPHICAL TRANSACTIONS A

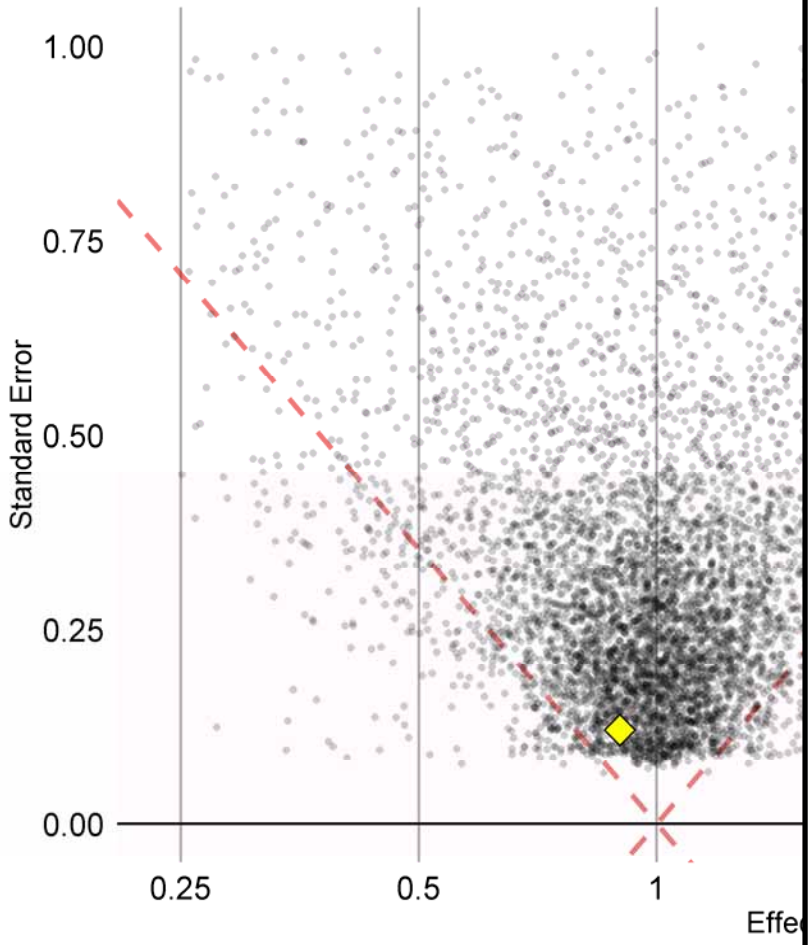
rsta.royalsocietypublishing.org

Research

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

Martijn J. Schuemie, Patrick B. Ryan, George Hripcsak, David Madigan and Marc A. Suchard

Example

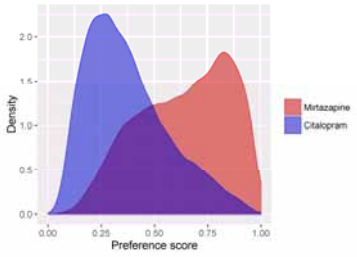


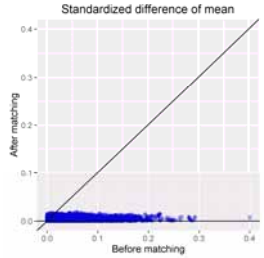
Mirtazapine vs. Citalopram

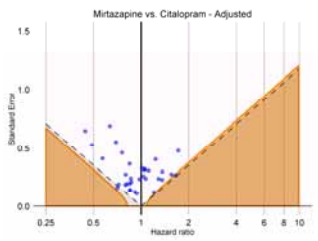
Constipation

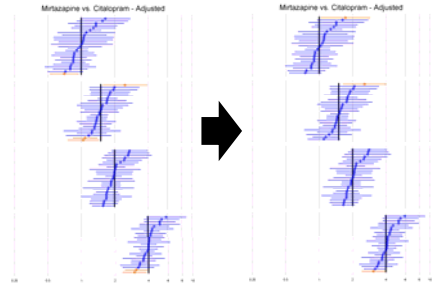
Database: Truven MDCD

Calibrated HR = 0.90 (0.70 – 1.12)









Large-scale estimation

- Not “data-dredging”!
  - Data-dredging is not about what you do but about what you *throw out*
    - This can’t be done for literature
- One-off studies
  - Wouldn’t it be best to optimize each study?
    - Never get 10 or 100 parameters right
  - Cannot study its operating characteristics
    - End up in today’s situation

OHDSI LEGEND Hypertension Study

- What becomes possible?

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 ba prolonged half-life and proven trial r 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 HF*r*EF; 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 HF*r*EF. • 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)	

US hypertension treatment guideline

Whelton et al., Hypertension 2018

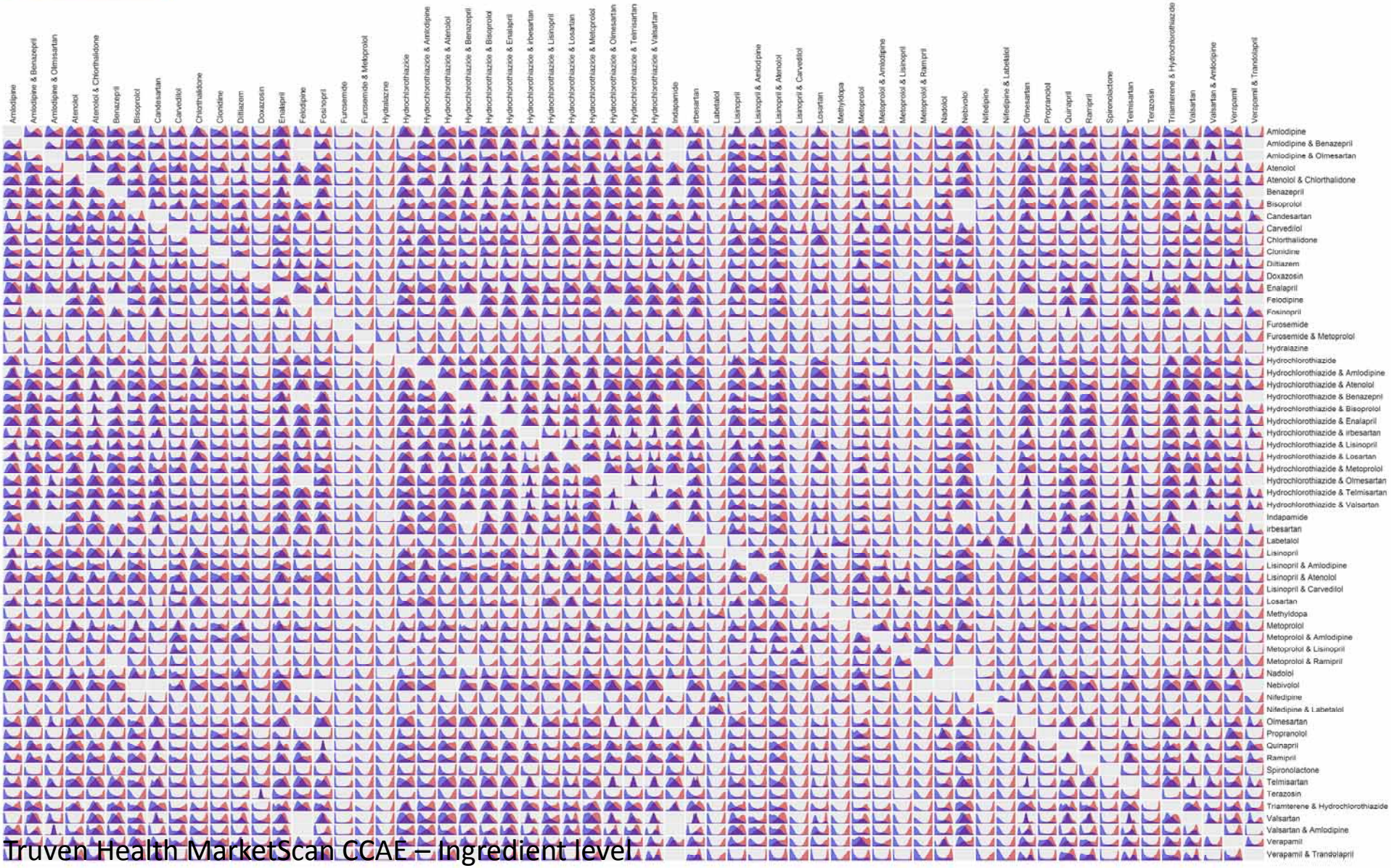
Evidence to support the guideline

- 40 randomized trials
- Only 11% of recommendations based on RCT
- Most decisions are “expert opinion”

Comparisons of hypertension treatments

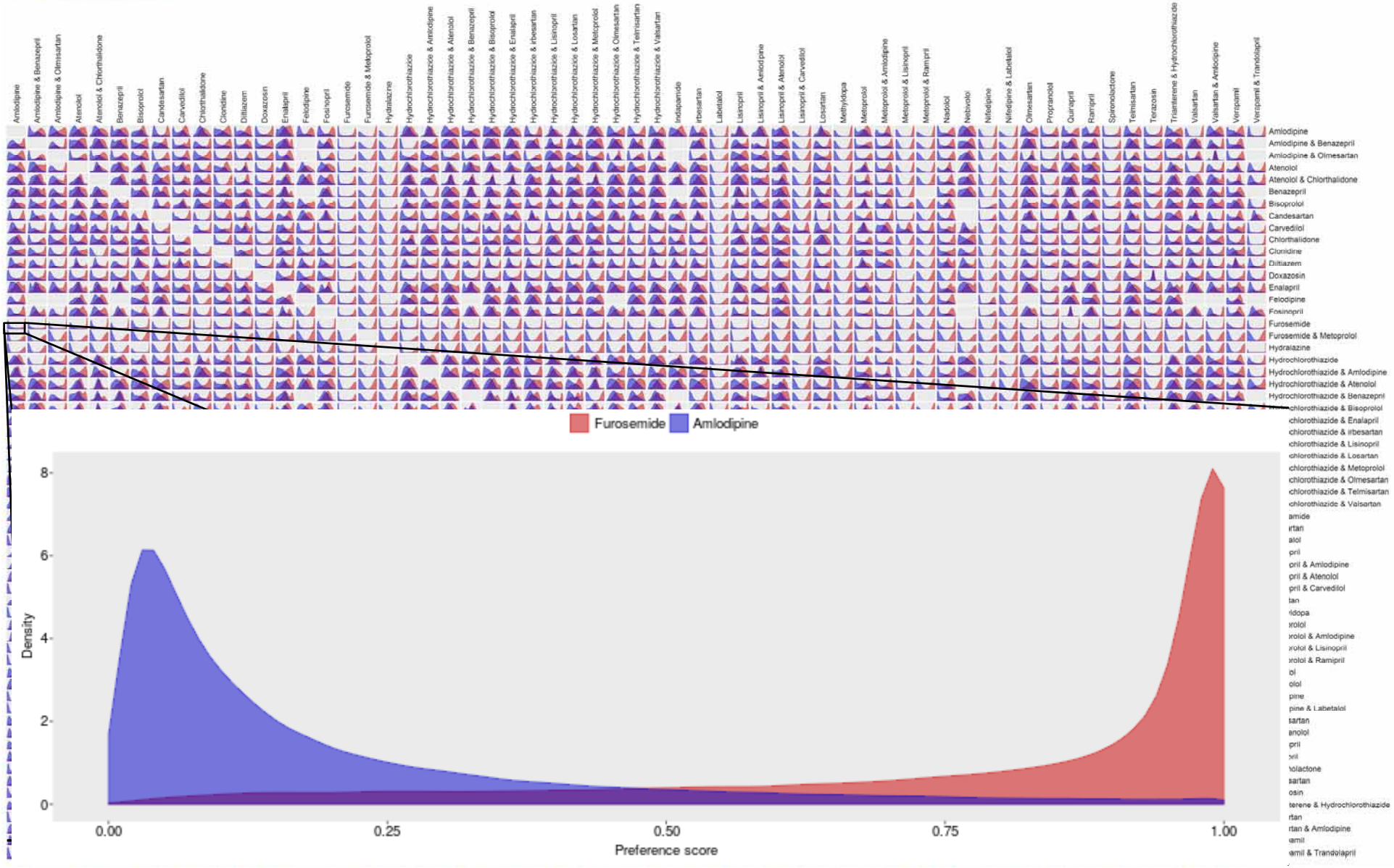
	Theoretical	Observed (n > 2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 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

Each hypothesis a fully executed trial



Truven Health MarketScan CCAE – Ingredient level

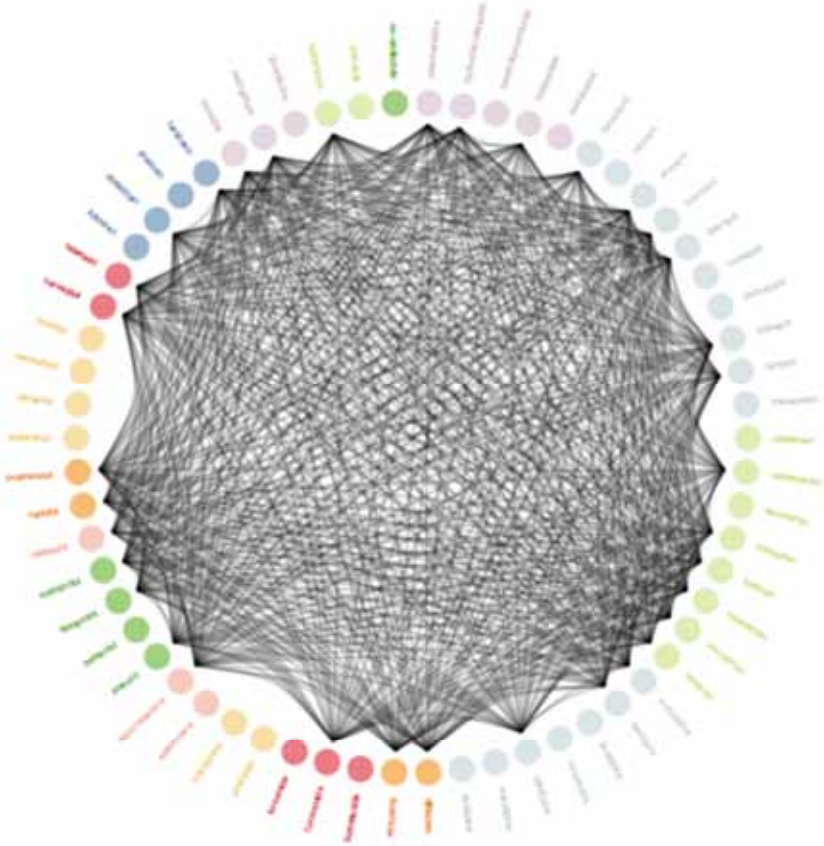
Not all comparisons are valid



LEGEND knowledge base for hypertension

Head-to-head HTN drug comparisons

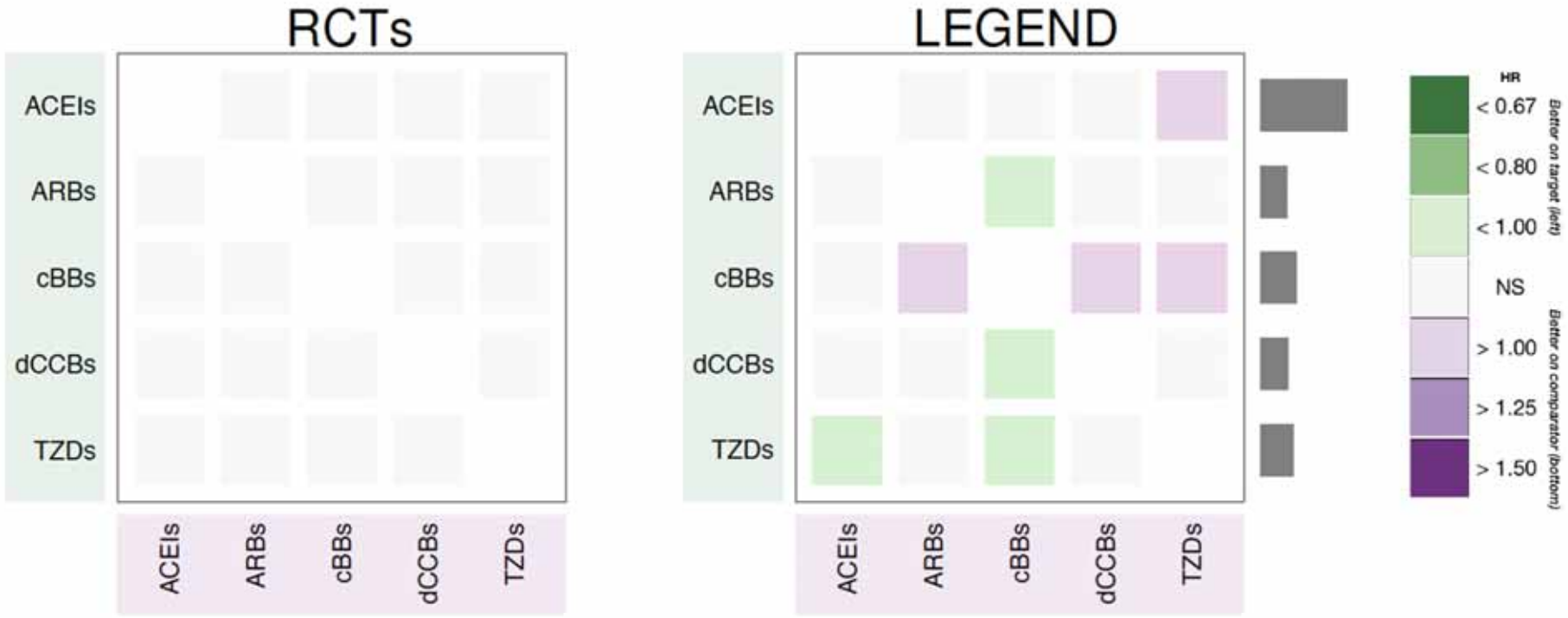
- Trials: 40
- $N = 102 - [1148] - 33K$



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

First-line agents: comparisons from LEGEND

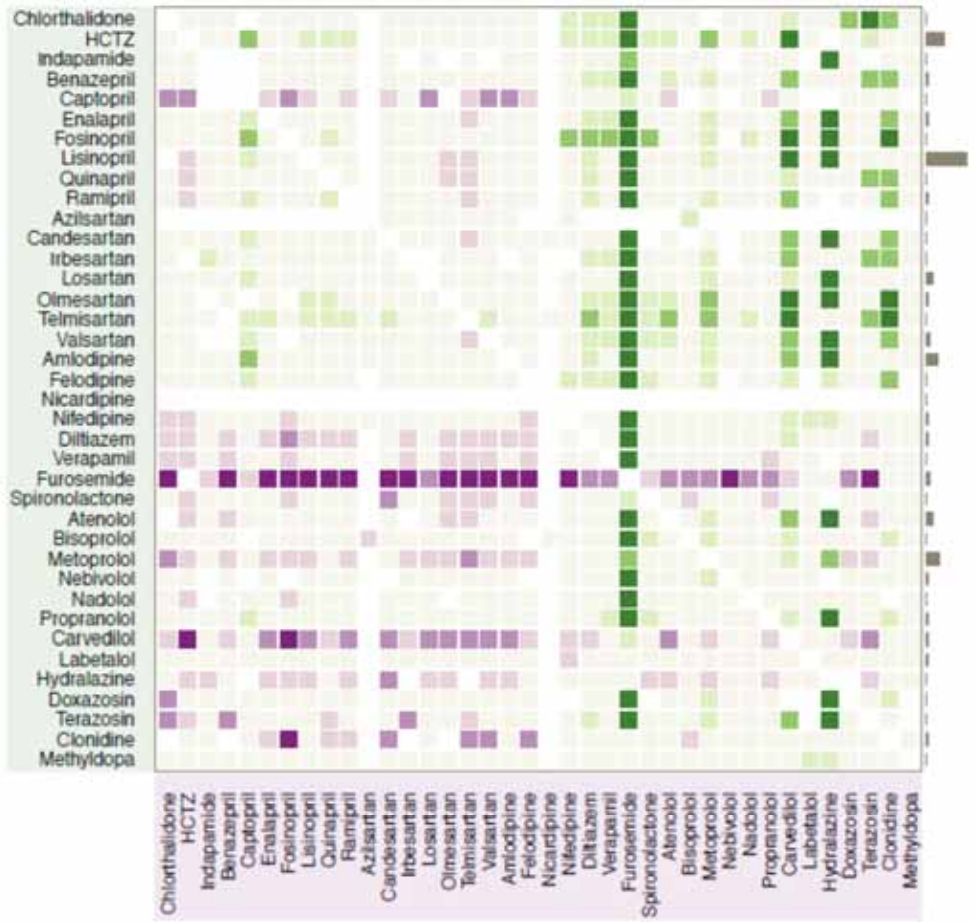
Efficacy outcome: **myocardial infarction**, heart failure, stroke



Data source: meta-analysis, $\sim 1 - 2$M total patients per study

- Beta blockers underperform alternatives
- Unexpected: TZDs $>$ ACEIs. Reliable?

Cardiovascular efficacy by drug

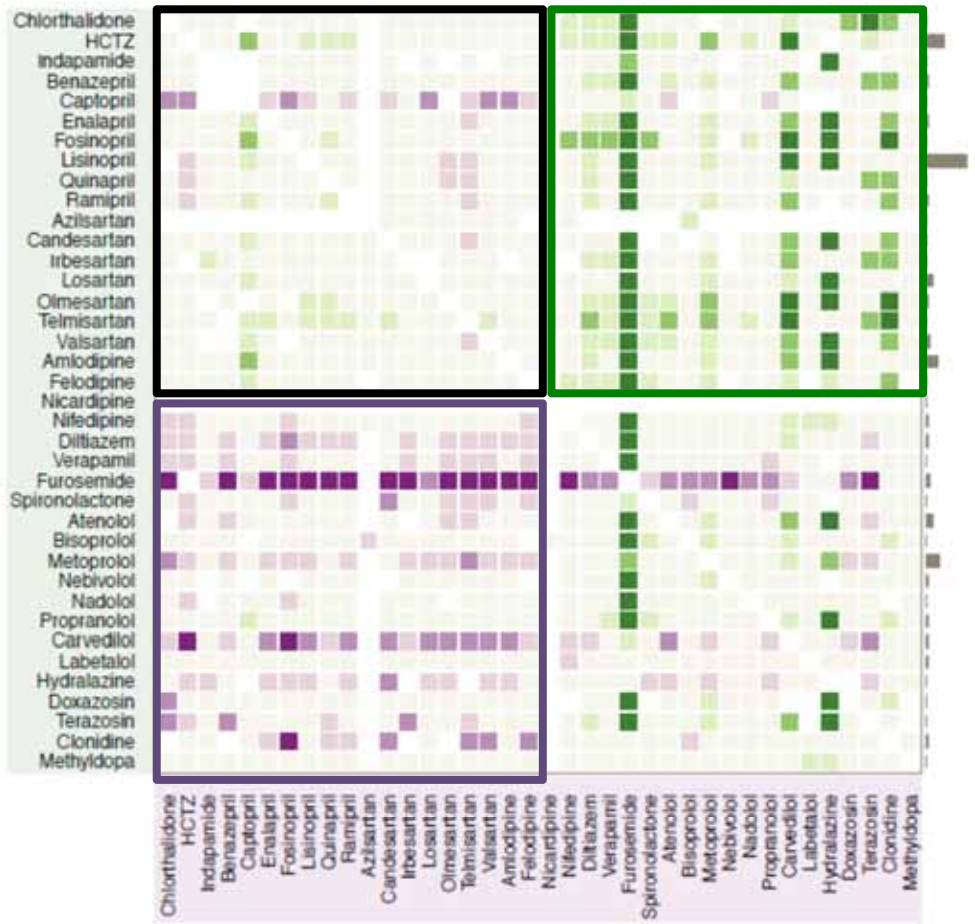


Composite (MI, HF, stroke) outcome in meta-analysis

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

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

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

Chlorthalidone versus hydrochlorthiazide: risk of MI

Risk estimates and meta-analysis across LEGEND databases:

Analysis	Data source	HR	LB	UB	P	Cal.HR	Cal.LB	Cal.UB	Cal.P
PS stratification, on-treatment	CCAE	0.65	0.33	1.14	0.17	0.66	0.37	1.19	0.18
PS stratification, on-treatment	Meta-analysis	0.79	0.54	1.16	0.24	0.81	0.56	1.17	0.30
PS stratification, on-treatment	Optum	0.90	0.52	1.44	0.67	0.93	0.57	1.53	0.82
PS stratification, on-treatment	Panther	0.98	0.05	5.06	0.99	0.91	0.26	3.42	0.96

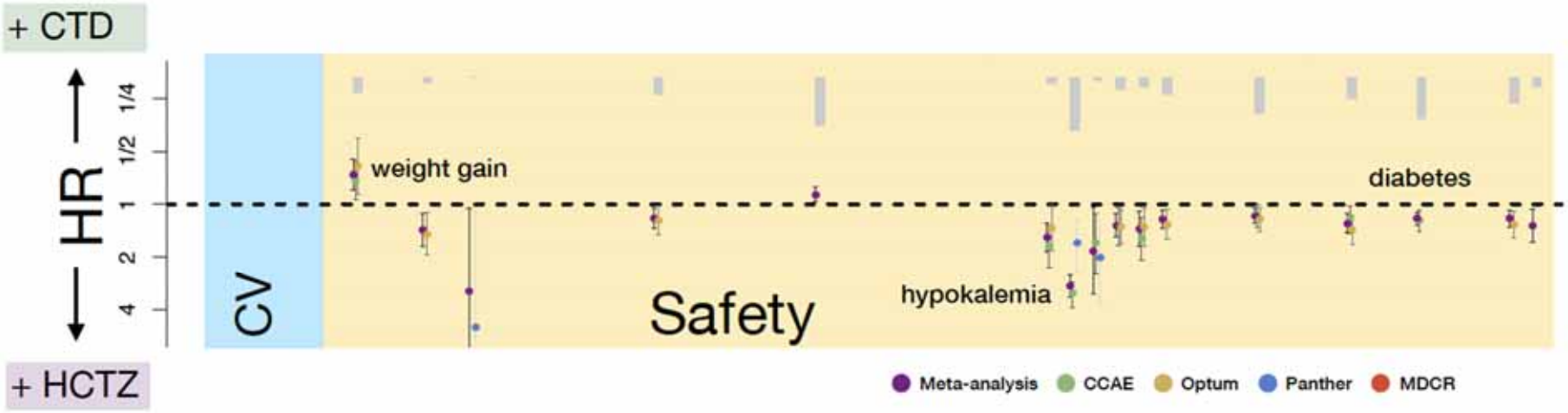
Showing 1 to 4 of 4 entries Previous 1 Next

Power Systematic error Subgroups

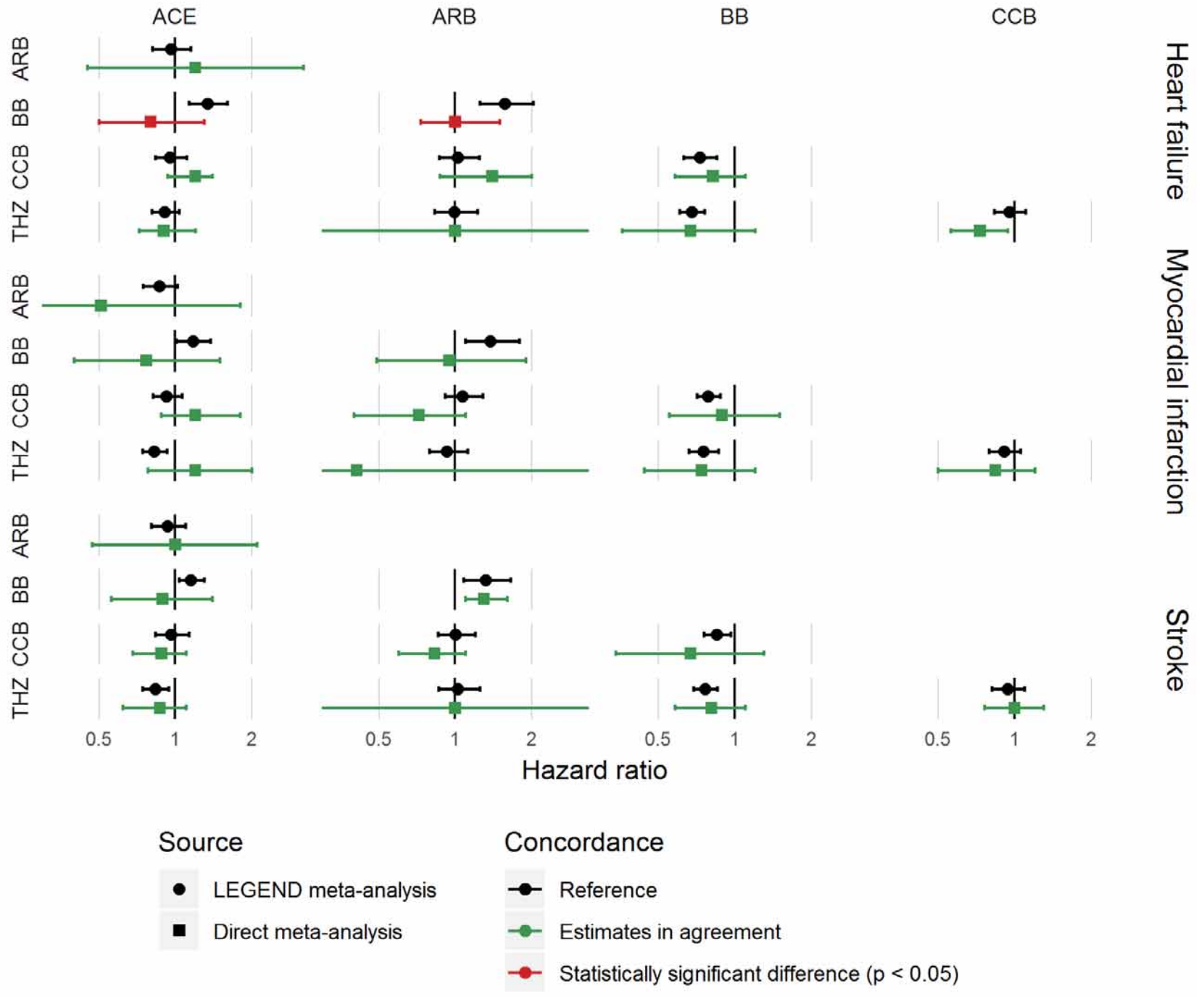
Table 1a. Number of subjects, follow-up time (in years), number of outcome events, and event incidence rate (IR) per 1,000 patient years (PY) in the target (*Chlorthalidone*) and comparator (*Hydrochlorothiazide*) group after stratification, as well as the minimum detectable relative risk (MDRR). Note that the IR does not account for any stratification.

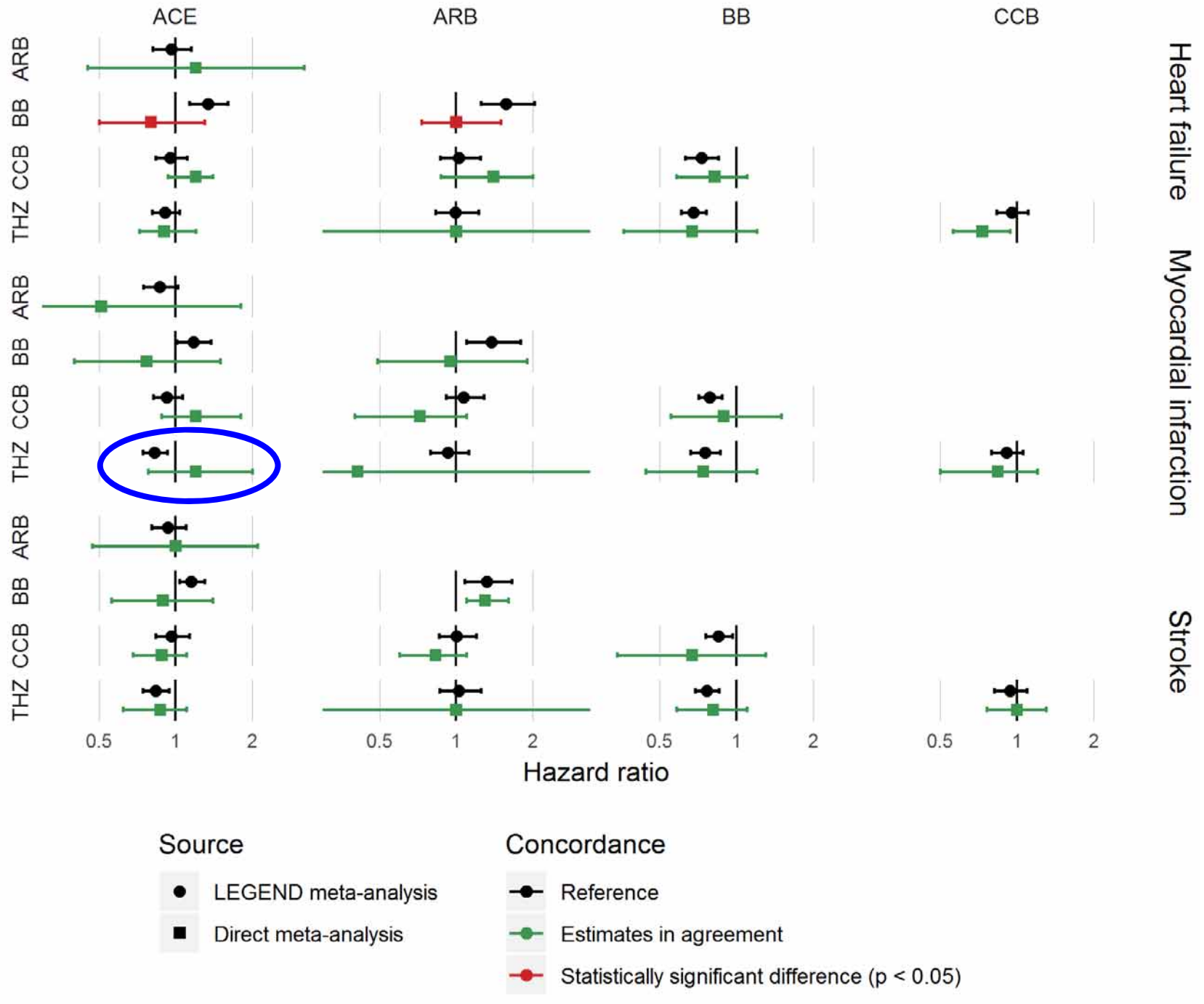
Target subjects	Comparator subjects	Target years	Comparator years	Target events	Comparator events	Target IR (per 1,000 PY)	Comparator IR (per 1,000 PY)	MDRR	i2
25,566	528,202	14,047	339,516	<32	819	<2.28	2.41	>1.58	0.00

CTD vs. HCTZ: safety profile



- Safety favors HCTZ – electrolyte imbalance
- CTD is more potent, longer half-life


ACE
ARB
BB
CCB
Heart failure
Myocardial infarction
Stroke
ARB
BB
CCB
THZ
0.5
1
2
Hazard ratio
Source
LEGEND meta-analysis
Direct meta-analysis
Concordance
Reference
Estimates in agreement
Statistically significant difference (p < 0.05)

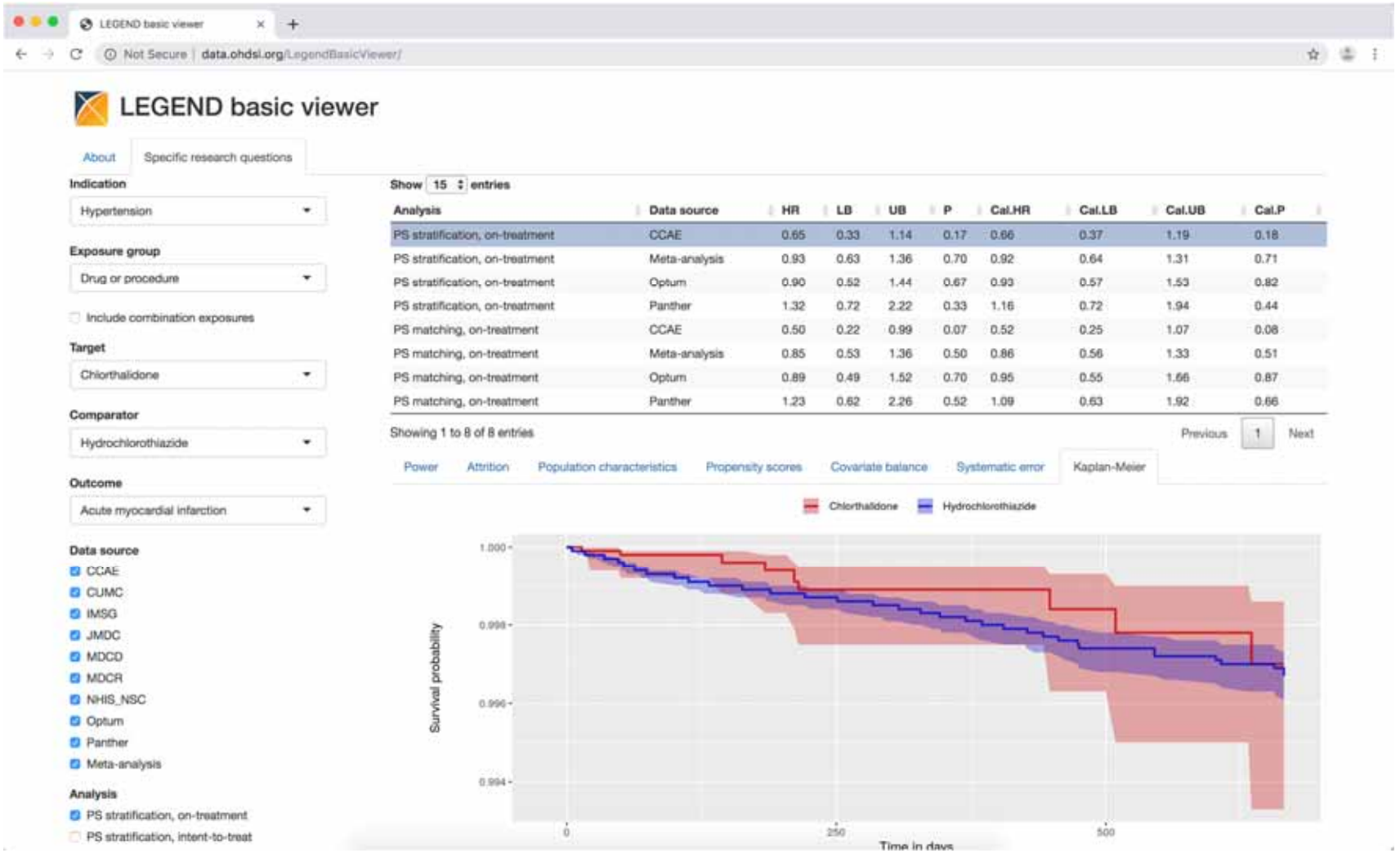

ACE
ARB
BB
CCB
Heart failure
Myocardial infarction
Stroke
ARB
BB
CCB
THZ
0.5
1
2
Hazard ratio
Source
LEGEND meta-analysis
Direct meta-analysis
Concordance
Reference
Estimates in agreement
Statistically significant difference (p < 0.05)

Clinical lessons for hypertension

- OHDSI evidence concordant with RCT meta-analysis
  - Where RCTs exist
  - Larger samples, more diversity, narrower confidence intervals
- Largely supports US guideline
  - First line drugs superior to second
  - Beta blockers not first line (unlike European)
- ACE inhibitors inferior to thiazide diuretics
  - 48% of world starts on ACEi
  - Avoid 1.3 cardiovascular events per 1000 patients (HR = 0.84)
  - Accepted to Lancet

OHDSI open results

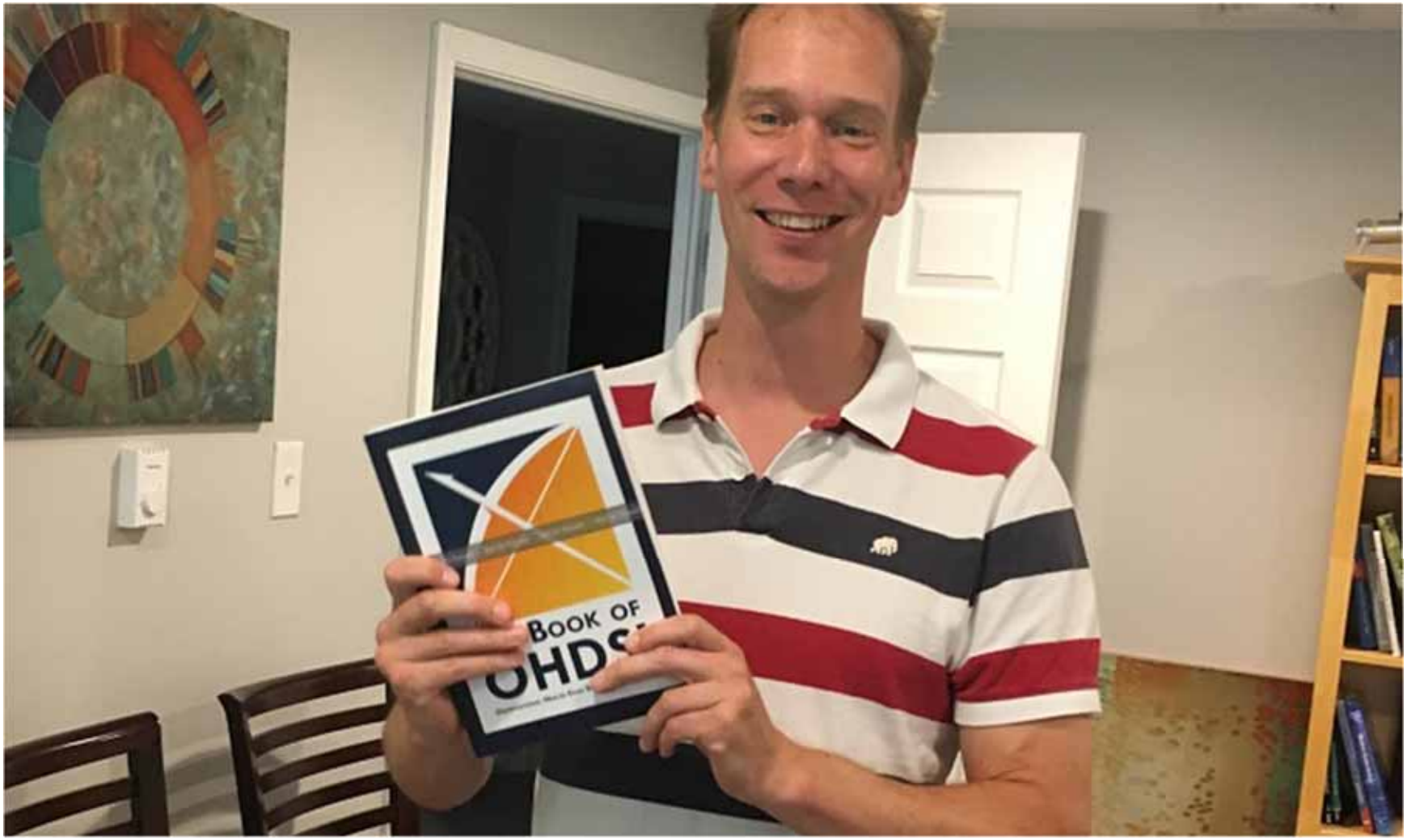
- data.ohdsi.org (500K studies)



Conclusions

- It is feasible to create an enormous international research network
- Sites willing to participate despite privacy …
- Nearing 1/10th world population
- Completely open: Data model, methods, tools
- Concrete approach to address the credibility crisis
- **Must change the way we do observational research**
  - **Large scale to measure and improve reproducibility**

Book of OHDSI



Upcoming meetings

- 2019 OHDSI Symposium
  - 15-17 September 2019
  - Near Washington, DC
    - Bethesda, Maryland, USA
  - No registration fee
- 2020 European OHDSI Symposium
  - 27-29 March 2020
  - University of Oxford, UK

OHDSI.org