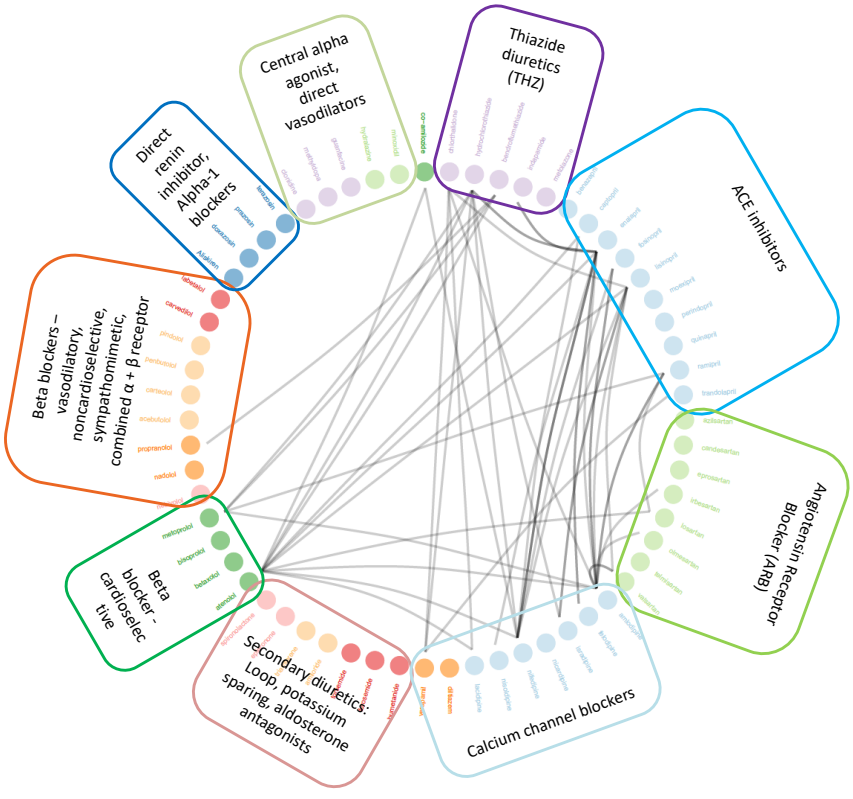
Large-scale evidence generation across a network of databases (LEGEND) for hypertension

Marc A. Suchard, MD, PhD
on behalf of the LEGEND team

Joint Statistical Meetings
3 August 2020

Current knowledge base for hypertension

Head-to-head antihypertensive drug comparisons



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

- Driven primarily by ALLHAT
  - just 3 individual drugs
- Focus: efficacy $\gg$ safety
- New RCTs too expensive

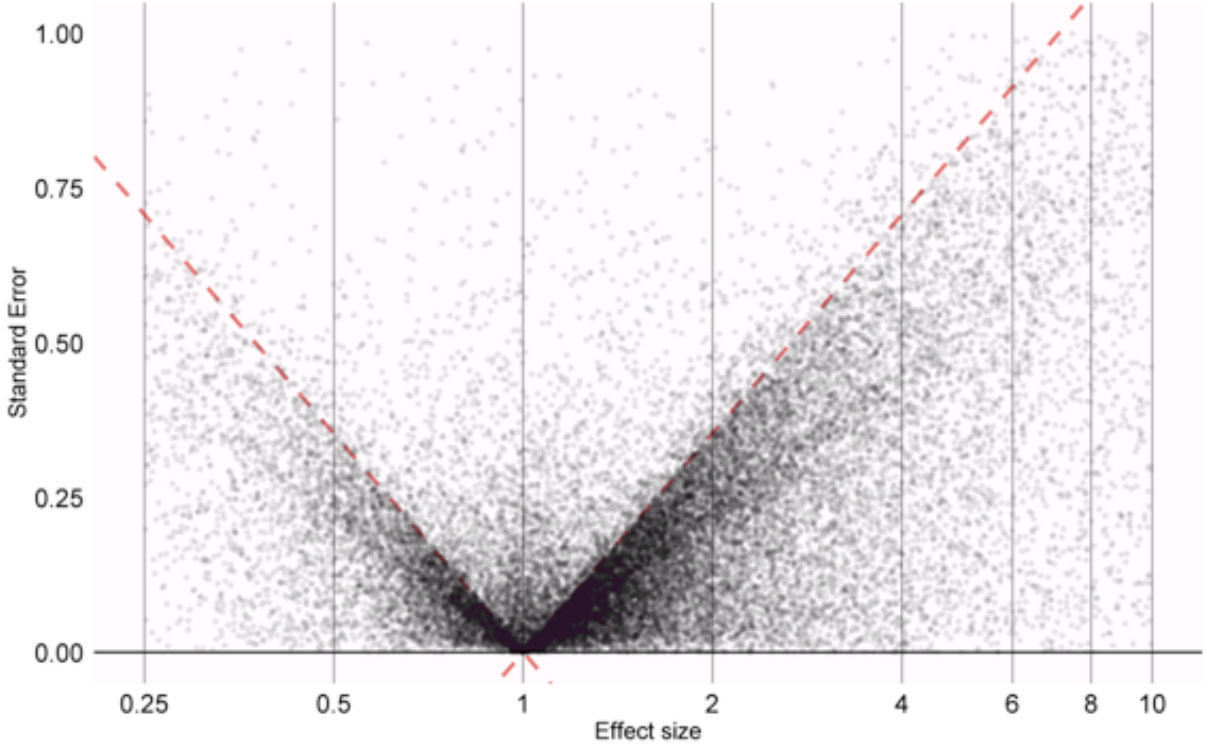
Can we provide

1. reliable / reproducible – concordant extant w/ RCTs
2. rich – across "all" comparators, outcomes
3. relevant – inform practice

evidence?

Observational research estimates in literature

29,982 drug safety estimates
from 11,758 papers



- 85% have reported $p < 0.05$
- Also note unusual peak along boundary

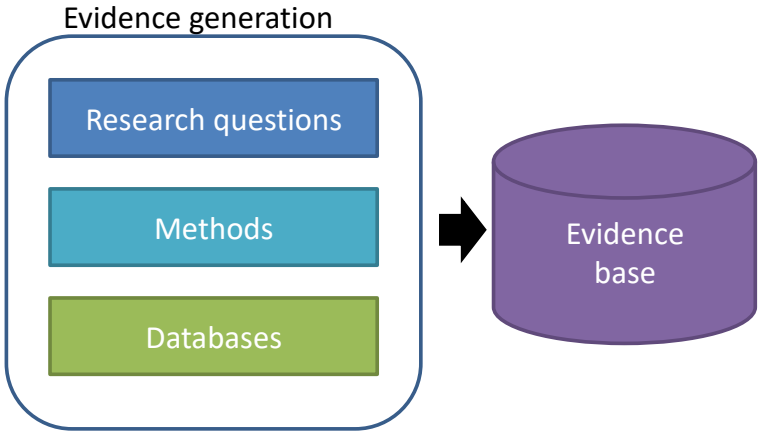
What is going wrong?

- Observational bias (confounding, selection, measurement error)
- Publication bias
- *p*-hacking (one study at a time)
- Reproducibility across populations

A solution?

Large-scale evidence generation across a network of databases (LEGEND)

- Aims to generate reliable evidence on the effects of medical interventions using observational healthcare data
- 10 guiding principles; chief among these:
  - Generate at large-scale (completeness, empirical calibration)
  - Systematically driven by best-practices
  - Disseminate everything (open science)



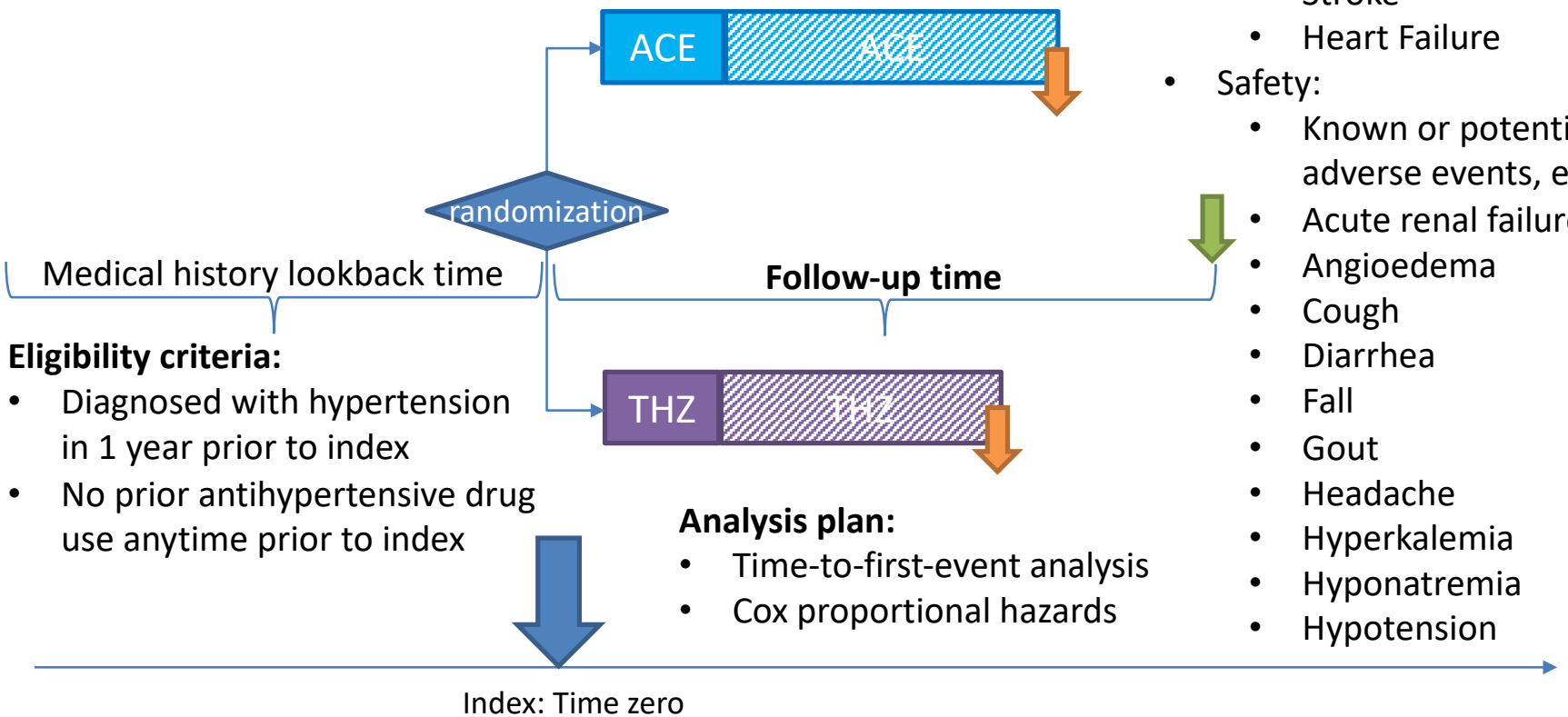
No one person has all the necessary skills

Best-practices: systematic design

Treatment strategies:
- Monotherapy with ACE
- Monotherapy with THZ

Causal contrasts of interest:
- Intent-to-treat effect
- On-treatment effect

Outcomes:
- Efficacy:
  - Myocardial infarction
  - Stroke
  - Heart Failure
- Safety:
  - Known or potential adverse events, e.g.
  - Acute renal failure
  - Angioedema
  - Cough
  - Diarrhea
  - Fall
  - Gout
  - Headache
  - Hyperkalemia
  - Hyponatremia
  - Hypotension



Eligibility criteria:
- Diagnosed with hypertension in 1 year prior to index
- No prior antihypertensive drug use anytime prior to index

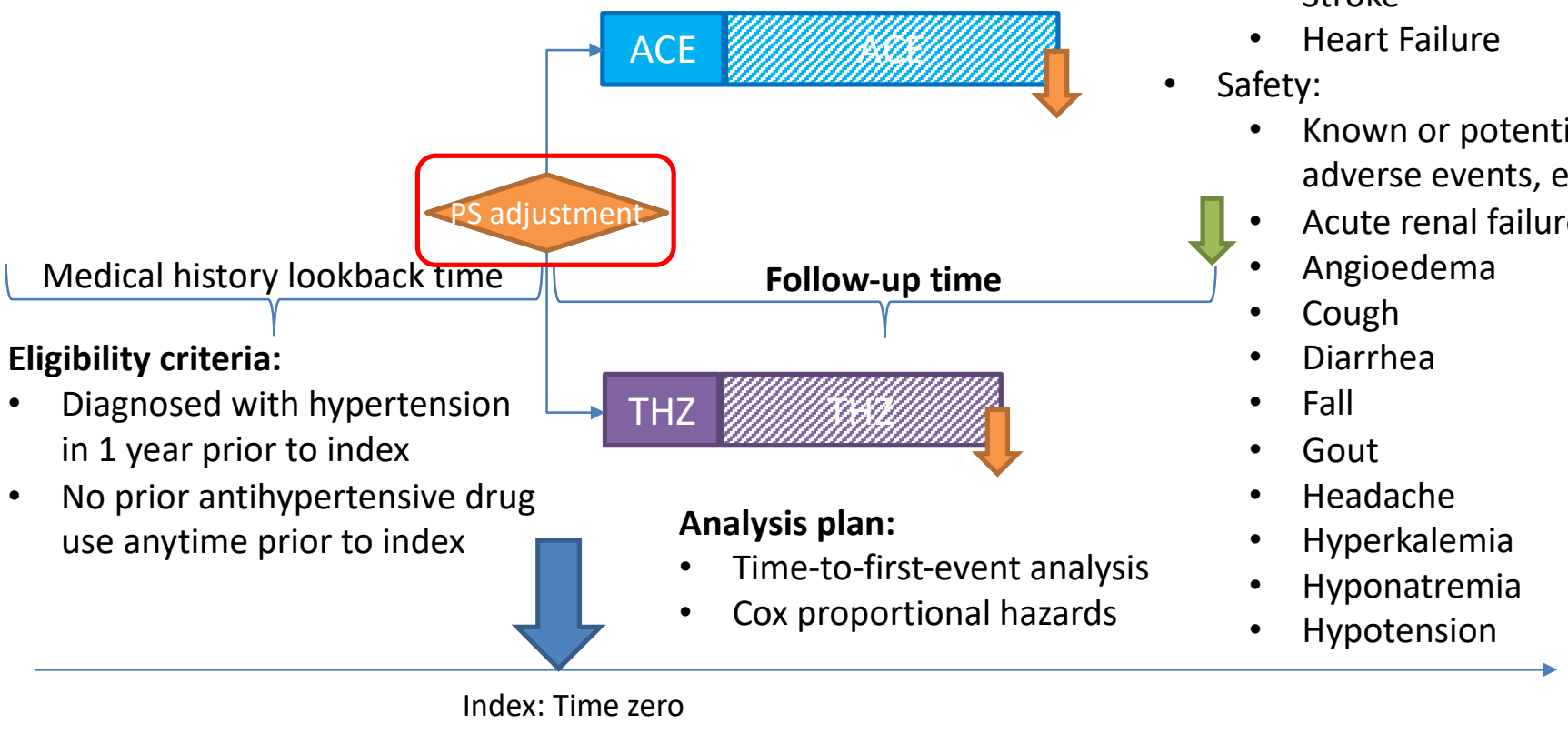
Analysis plan:
- Time-to-first-event analysis
- Cox proportional hazards

Observ. study for comparing two initial therapies

Treatment strategies:
- Monotherapy with ACE
- Monotherapy with THZ

Causal contrasts of interest:
- Intent-to-treat effect
- On-treatment effect

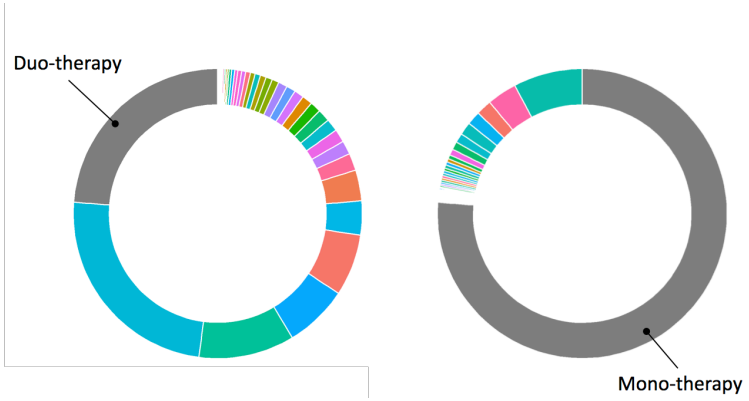
Outcomes:
- Efficacy:
  - Myocardial infarction
  - Stroke
  - Heart Failure
- Safety:
  - Known or potential adverse events, e.g.
  - Acute renal failure
  - Angioedema
  - Cough
  - Diarrhea
  - Fall
  - Gout
  - Headache
  - Hyperkalemia
  - Hyponatremia
  - Hypotension



Eligibility criteria:
- Diagnosed with hypertension in 1 year prior to index
- No prior antihypertensive drug use anytime prior to index

Analysis plan:
- Time-to-first-event analysis
- Cox proportional hazards

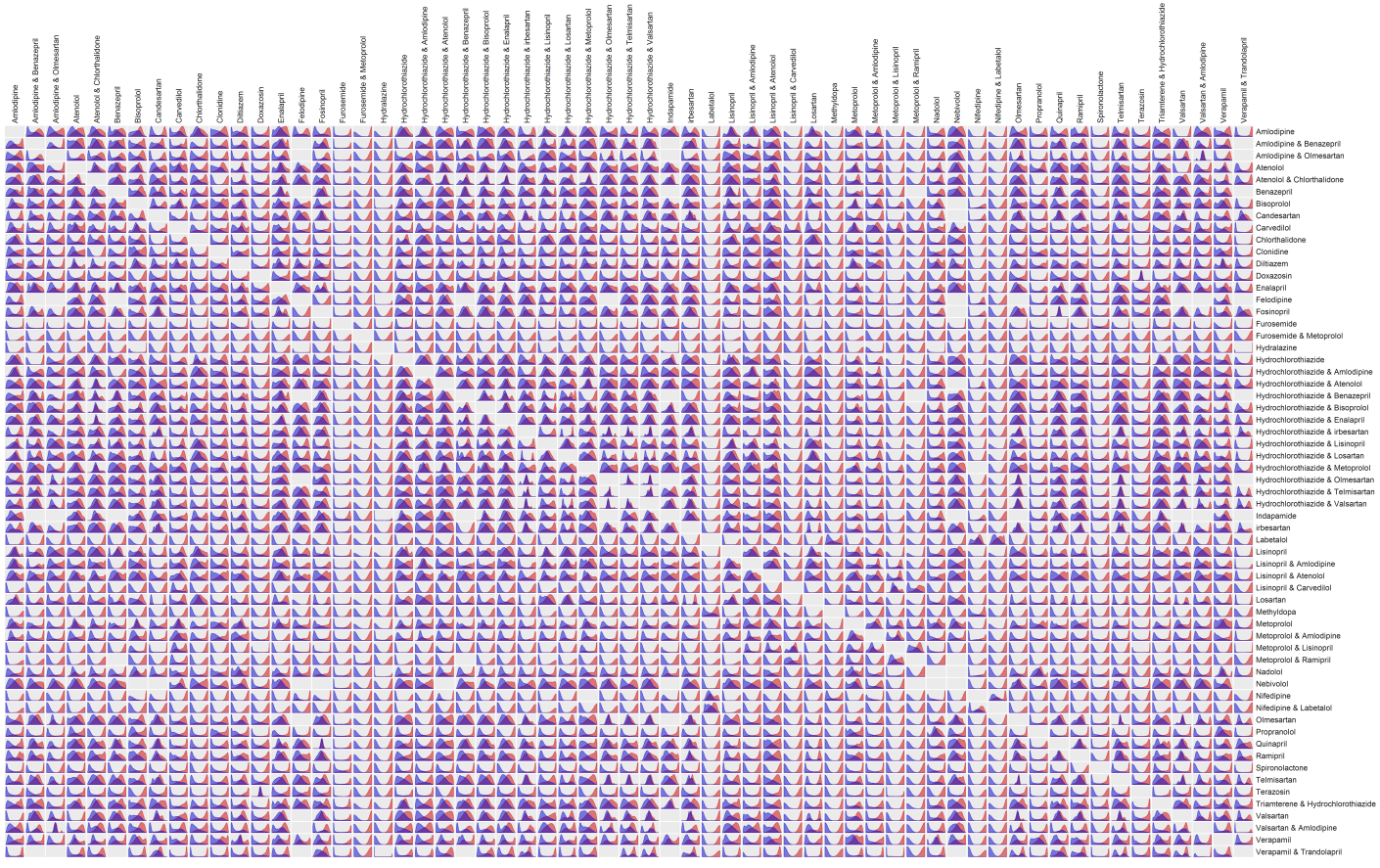
Comparison of hypertension treatments



- 39 mono-drugs, 13 mono-classes
- 58 duo-drugs, 32 duo-classes
- 10,278 comparisons

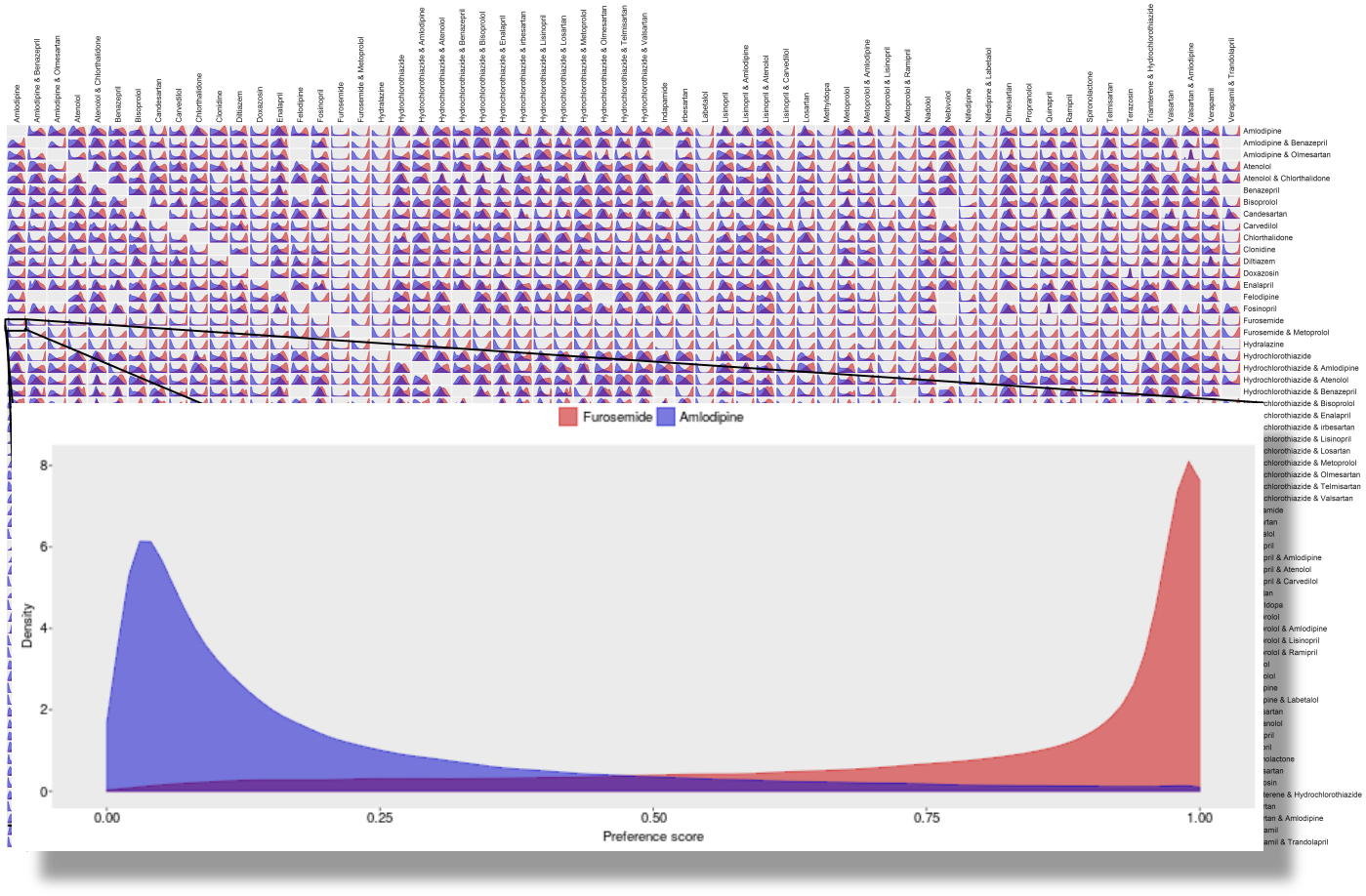
	Theoretical	Observed (n > 2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2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

Best-practices: systematic large-scale PS



- >8,000 (regularized) baseline patient characteristics (all dx, rx, tx)
- Address observed (and some unobserved – BP control) confounding (Tian et al, 2019, IJE)

Of course, not all comparisons are valid



- Evaluation of propensity score (PS) distributions and covariate balance
- Here: poor empirical clinical equipoise

Best-practices: 58 expert-crafted outcomes

- Effectiveness (10): acute MI, heart failure, stroke
- Safety (48): known side-effects

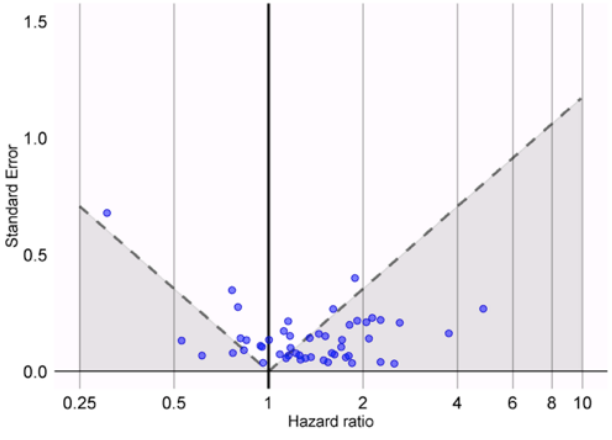
Phenotype	Logical description	Supporting references
Abdominal pain	Abdominal pain condition record of any type; successive records with > 90 day gap are considered independent episodes	4 5 6
Abnormal weight gain	Abnormal weight gain record of any type; successive records with > 90 day gap are considered independent episodes; note, weight measurements not used	7
Abnormal weight loss	Abnormal weight loss record of any type; successive records with > 90 day gap are considered independent episodes; note, weight measurements not used	8
Acute myocardial infarction	Acute myocardial infarction condition record during an inpatient or ER visit; successive records with > 180 day gap are considered independent episodes	9 10 11 12 13 14
Acute pancreatitis	Acute pancreatitis condition record during an inpatient or ER visit; successive records with >30 day gap are considered independent episodes	15 16 17 18
Acute renal failure	A diagnosis of acute renal failure in an inpatient or ER setting; must be at least 30d between inpatient/ER visits to be considered separate episodes	19 20 21 22 23 24 25 26

	Theoretical	Observed (n > 2,500)
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020

Calibrate each study (under null)

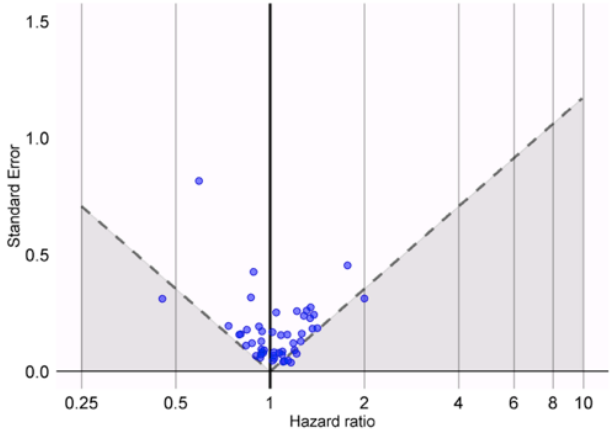
76 negative outcome controls (not caused by either treatment) help expose and control residual bias. Example: ingrown toenail

Crude - Uncalibrated



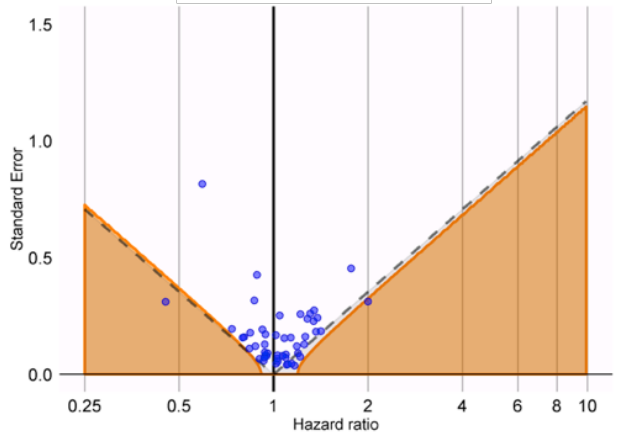
68% have $p < 0.05$

PS stratified - Uncalibrated



16% have $p < 0.05$

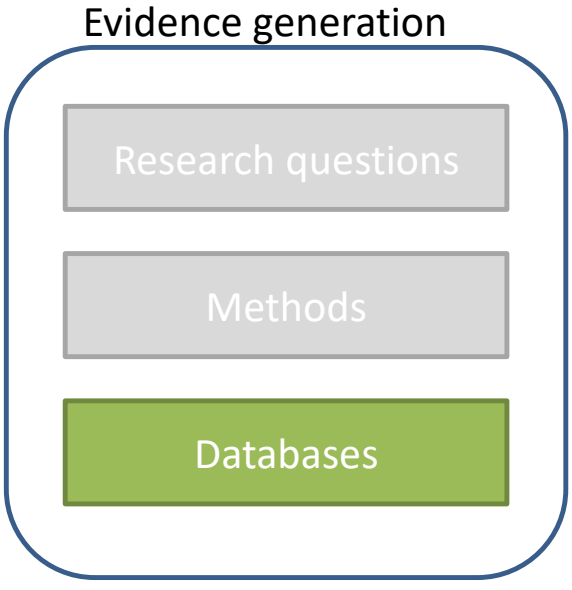
PS stratified - Calibrated



4% have $p < 0.05$

p-value empirical calibration models residual bias as exchangeable and adjusts for a (possibly) non-0 mean. (Schuemie et al, 2018, PNAS)

Network of data sources



Ajou University

Columbia University

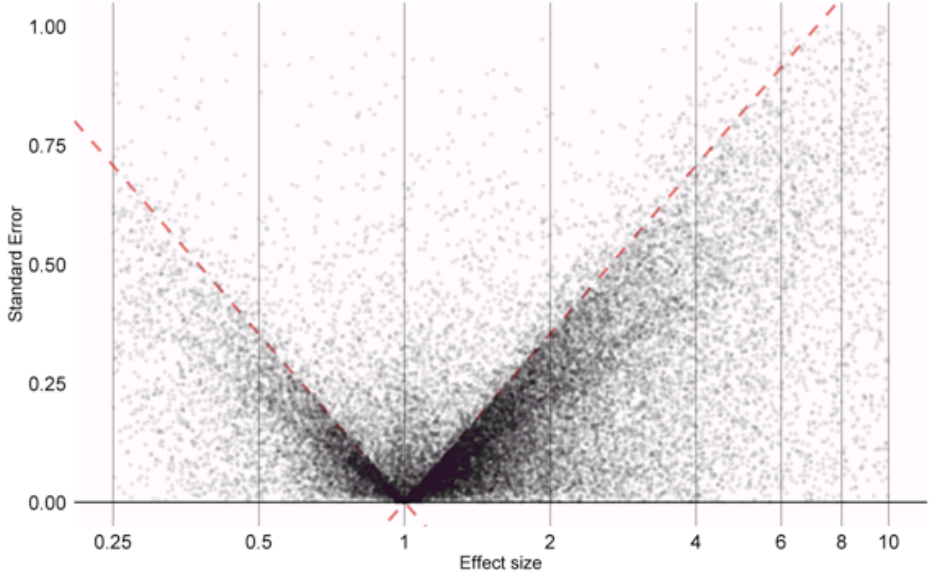
- US insurance databases
  - IBM MarketScan CCAE
  - IBM MarketScan MDCR
  - IBM MarketScan MDCD
  - Optum Clinformatics
- Japanese insurance database: JMDC
- Korean insurance database: NHIS-NSC
- US EHR databases
  - Optum EHR
  - Columbia University Medical Center
- German EHR database: IQVIA DA Germany

Account for population/practice heterogeneity (Madigan et al, 2013, AJE)

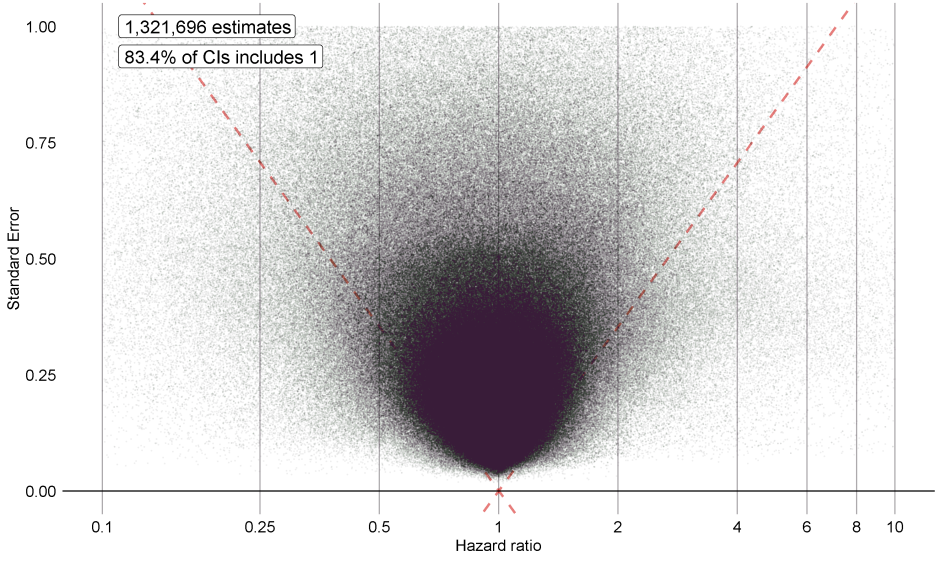
Improve generalizibility

How does LEGEND perform?

Literature

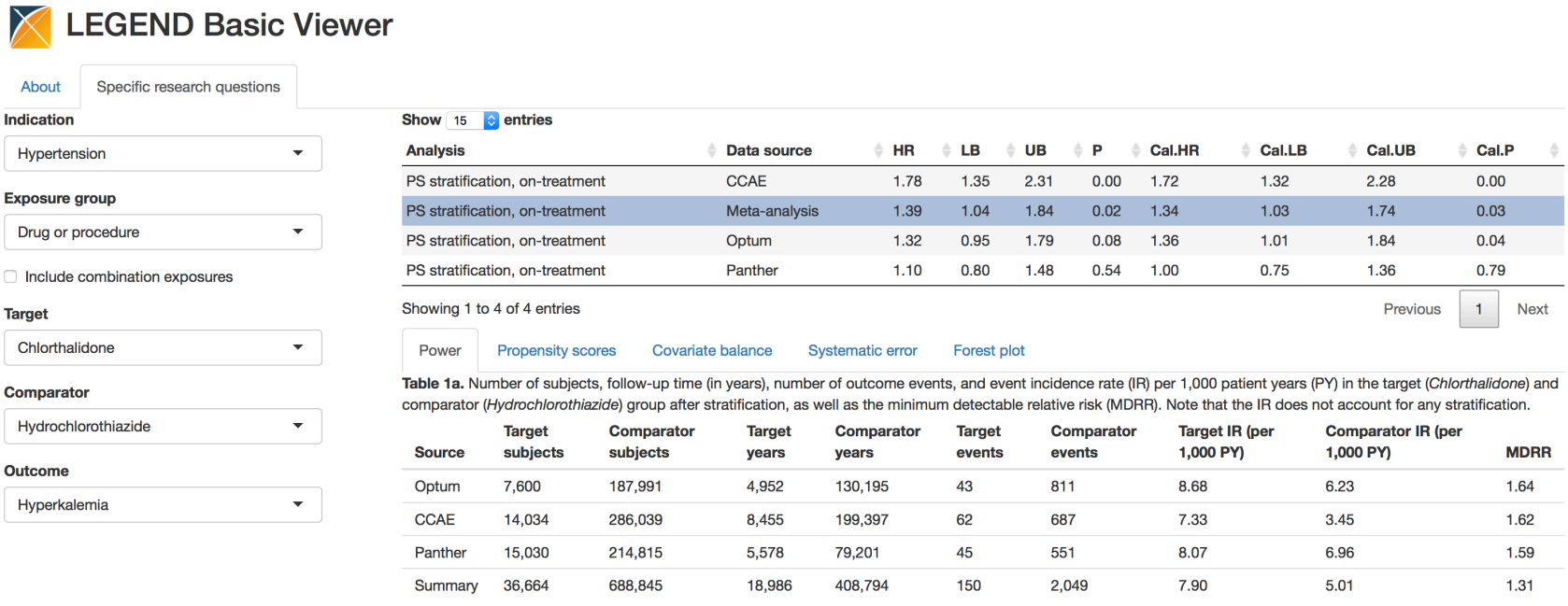


LEGEND



- Best-practices systematic design, evaluation and empirical calibration return near nominal performance
- Provide a more complete and reliable evidence basis

Unbiased LEGEND dissemination



LEGEND Basic Viewer

About | Specific research questions

Indication: Hypertension

Exposure group: Drug or procedure

Include combination exposures

Target: Chlorthalidone

Comparator: Hydrochlorothiazide

Outcome: Hyperkalemia

Show 15 entries

Analysis	Data source	HR	LB	UB	P	Cal.HR	Cal.LB	Cal.UB	Cal.P
PS stratification, on-treatment	CCAE	1.78	1.35	2.31	0.00	1.72	1.32	2.28	0.00
PS stratification, on-treatment	Meta-analysis	1.39	1.04	1.84	0.02	1.34	1.03	1.74	0.03
PS stratification, on-treatment	Optum	1.32	0.95	1.79	0.08	1.36	1.01	1.84	0.04
PS stratification, on-treatment	Panther	1.10	0.80	1.48	0.54	1.00	0.75	1.36	0.79

Showing 1 to 4 of 4 entries — Previous 1 Next

Power | Propensity scores | Covariate balance | Systematic error | Forest plot

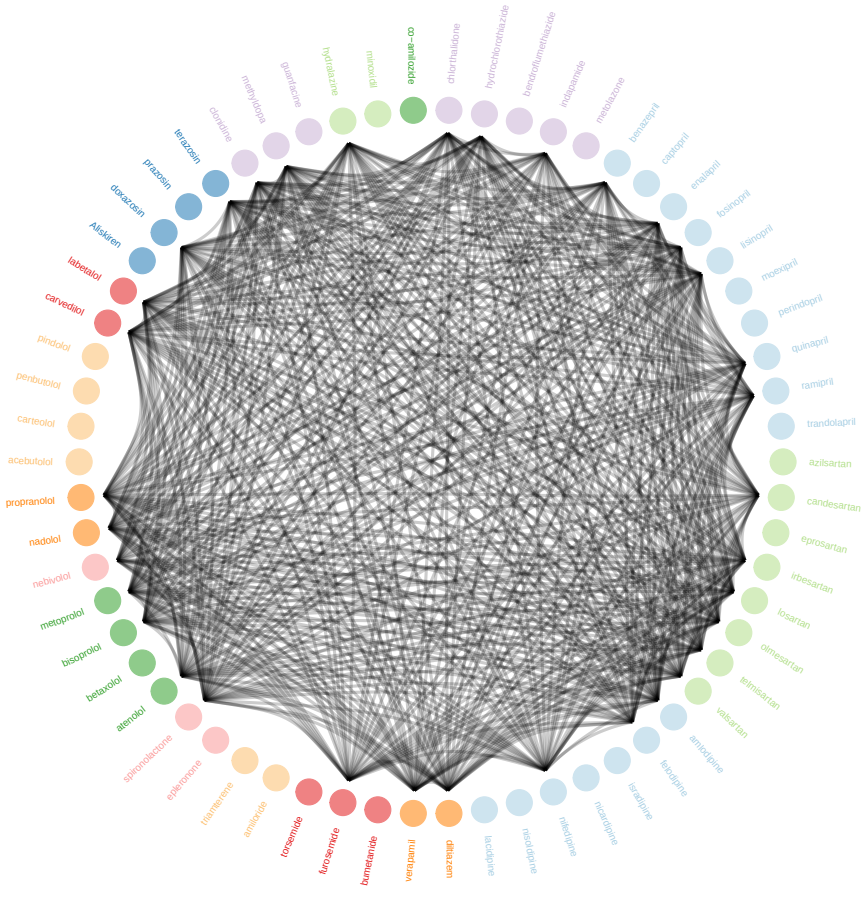
Table 1a. Number of subjects, follow-up time (in years), number of outcome events, and event incidence rate (IR) per 1,000 patient years (PY) in the target (*Chlorthalidone*) and comparator (*Hydrochlorothiazide*) group after stratification, as well as the minimum detectable relative risk (MDRR). Note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

Source	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
Optum	7,600	187,991	4,952	130,195	43	811	8.68	6.23	1.64
CCAE	14,034	286,039	8,455	199,397	62	687	7.33	3.45	1.62
Panther	15,030	214,815	5,578	79,201	45	551	8.07	6.96	1.59
Summary	36,664	688,845	18,986	408,794	150	2,049	7.90	5.01	1.31

- Open source protocol and end-to-end executable code
- `http://data.ohdsi.org/LegendBasicViewer` (all result artifacts for each study)
- `http://data.ohdsi.org/LegendMedCentral` (gimmick)

LEGEND knowledge base for hypertension

Head-to-head HTN drug comparisons

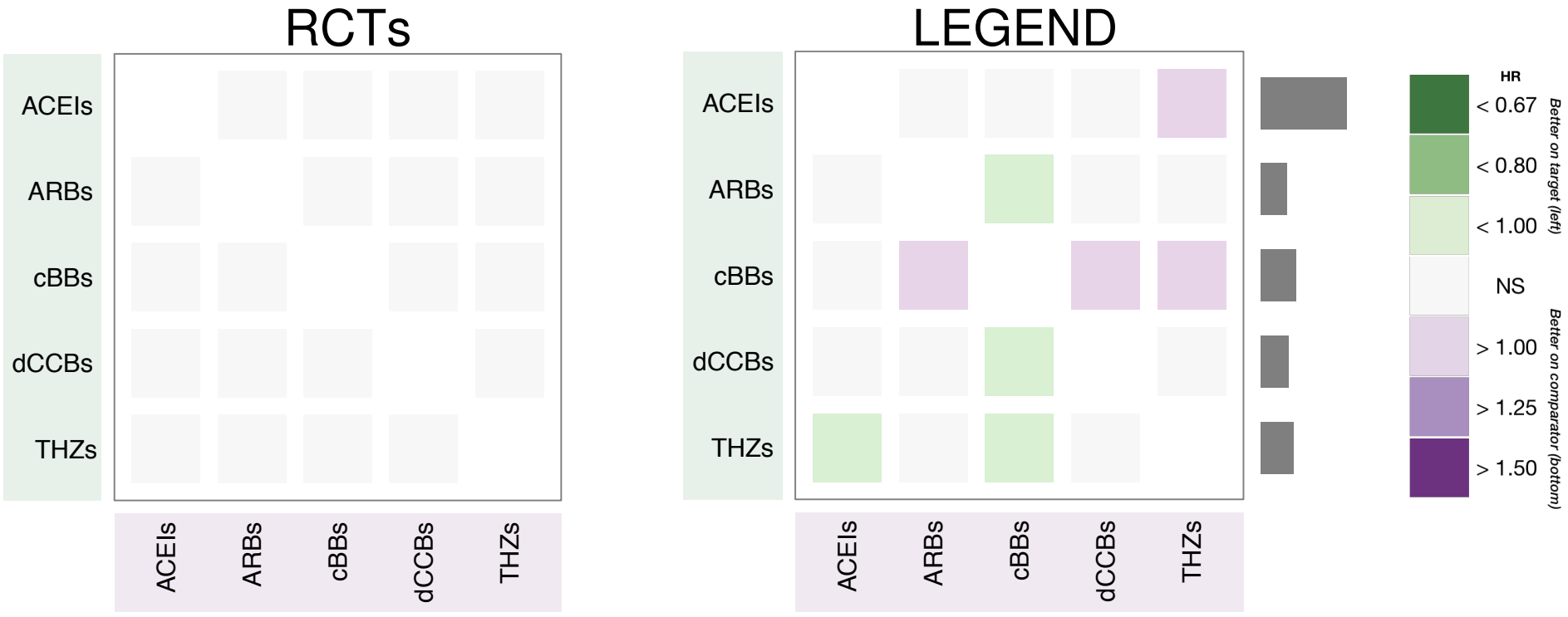


- Trials: 40
- $N = 102 - [1148] - 33K$

- 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: THZs $>$ ACEs. Reliable?

THZs vs. ACEIs: study diagnostics

Large-scale propensity score model controls for observed confounding

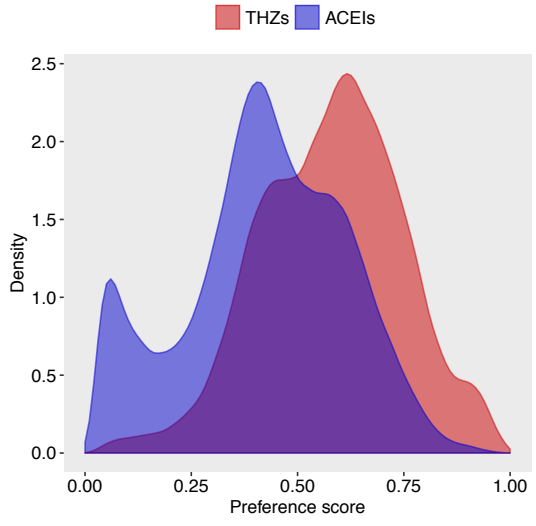


Fig. 2. Preference score distribution for TZDs and ACEIs new-users. The preference score is a transformation of the propensity score that adjusts for prevalence di erences between populations. A higher overlap indicates that subjects in the two populations are more similar in terms of their predicted probability of receiving one treatment over the other.

Cohort stratification / balance:

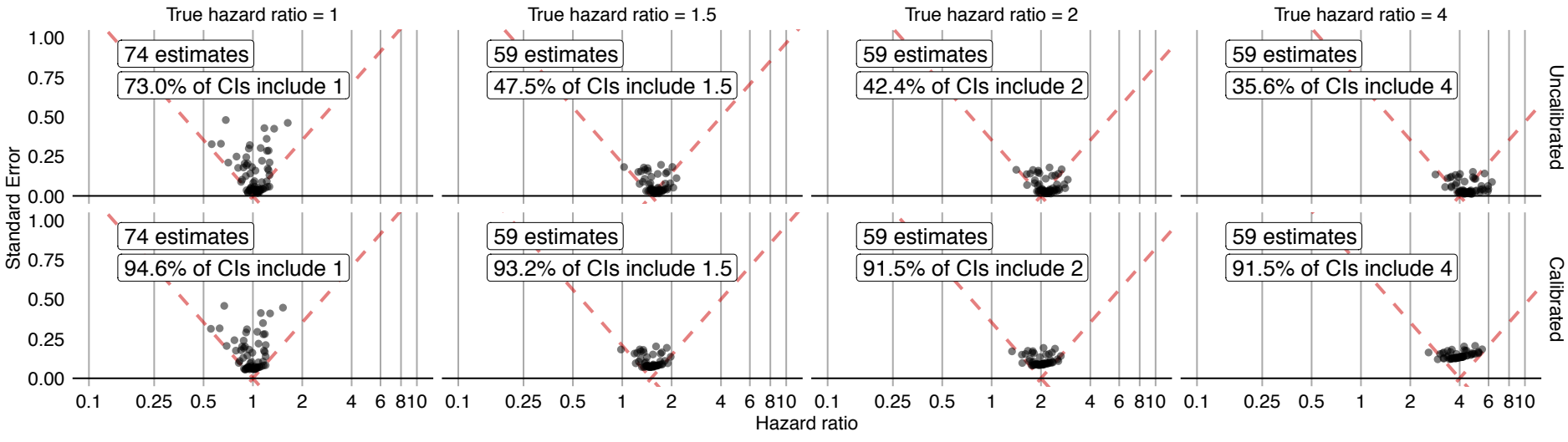
- Achieved across all 10,868 baseline characteristics (CCAE)
- Blood pressure (pop. means in mmHg) (Panther)

	THZs	ACEIs	$\lvert\Delta\rvert$
before	145/89	145/87	0.13
after	145/88	145/87	0.02

No BP measurements used in PS model, but still balanced after stratification

THZs vs. ACEIs: study outcomes

Calibration returns near nominal HR estimate coverage



- Good diagnostics $\rightarrow$ comparable cohorts (observed and unobserved); calibration $\rightarrow$ controls for residual systematic bias
- THZs are more effective than ACEIs in preventing MI

Ideal positioning for COVID-19

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Cold Spring Harbor Laboratory | BMJ | Yale

HOME | ABC

Search

Comment on this paper

Renin-angiotensin system blockers and susceptibility to COVID-19: a multinational open science cohort study

Daniel R Morales, Mitchell M Conover, Seng Chan You, Nicole Pratt, Kristin Kostka, Talita Duarte Salles, Sergio Fernandez Bertolin, Maria Aragon, Scott L. DuVall, Kristine Lynch, Thomas Falconer, Kees van Bochove, Cynthia Sung, Michael E. Matheny, Christophe G. Lambert, Fredrik Nyberg, Thamir M AlShammari, Andrew E. Williams, Rae Woong Park, James Weaver, Anthony G. Sena, Martijn J. Schuemie, Peter R. Rijnbeek, Ross D. Williams, Jennifer C.E Lane, Albert Prats Uribe, Lin Zhang, Carlos Areia, Harlan Krumholz, Daniel Prieto Alhambra, Patrick B Ryan, George Hripcsak, Marc A Suchard

doi: https://doi.org/10.1101/2020.06.11.20125849

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

Abstract | Info/History | Metrics | Preview PDF

Abstract

Introduction: Angiotensin converting enzyme inhibitors (ACEs) and angiotensin receptor blockers (ARBs) could influence infection risk of coronavirus disease (COVID-19). Observational studies to date lack pre-specification, transparency, rigorous ascertainment adjustment and international generalizability, with contradictory results. Methods: Using electronic health records from Spain (SIDIAP) and the United States (Columbia University Irving Medical Center and Department of Veterans Affairs), we conducted a systematic cohort

- Over 1.1 million antihypertensive users
- Active comparator, prevalent-users
- Executed within new EHR data partners
  - SIDIAP (universal primary care in Catalonia)
  - US VA
- End-to-end (almost) transparency

Acknowledgments

LEGEND Scientific Group:

- Martijn J. Schuemie
- Patrick B. Ryan
- Seng Chan You
- Nicole Pratt
- David Madigan
- George Hripcsak
- Marc A. Suchard

Clinical Advisory Team:

- RuiJun Chen
- Jon Duke
- Christian Reich
- Harlan Krumholz

Some of these results in Suchard et al. *Lancet*, 2019