Observational research results in literature

- Individuals may produce good research studies
- In aggregate, the medical research system is a data-dredging machine

Evidence from literature

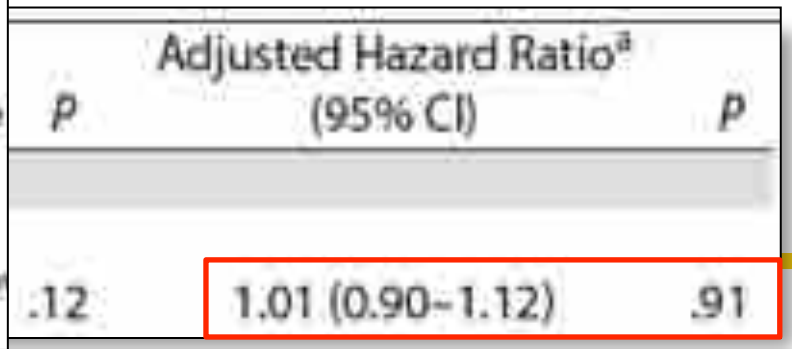
Paper by Lee et al, 2016

- Compare new users of SNRIs (includes duloxetine) vs SSRIs
- Taiwanese insurance claims data
- 12 month washout
- remove people using both drugs
- remove people with a prior history of head injury
- remove people with a prior history of stroke or intracranial hemorrhage
- Propensity score: logistic regression with treatment as dependent variable
- HOI is Stroke: first hospitalization with ICD-9 433,434, or 436
- time-varying Cox regression using 5 PS strata


Focus on Geriatric Psychiatry

Comparison of the Effects of Serotonin-Norepinephrine Reuptake Inhibitors Versus Selective Serotonin Reuptake Inhibitors on Cerebrovascular Events

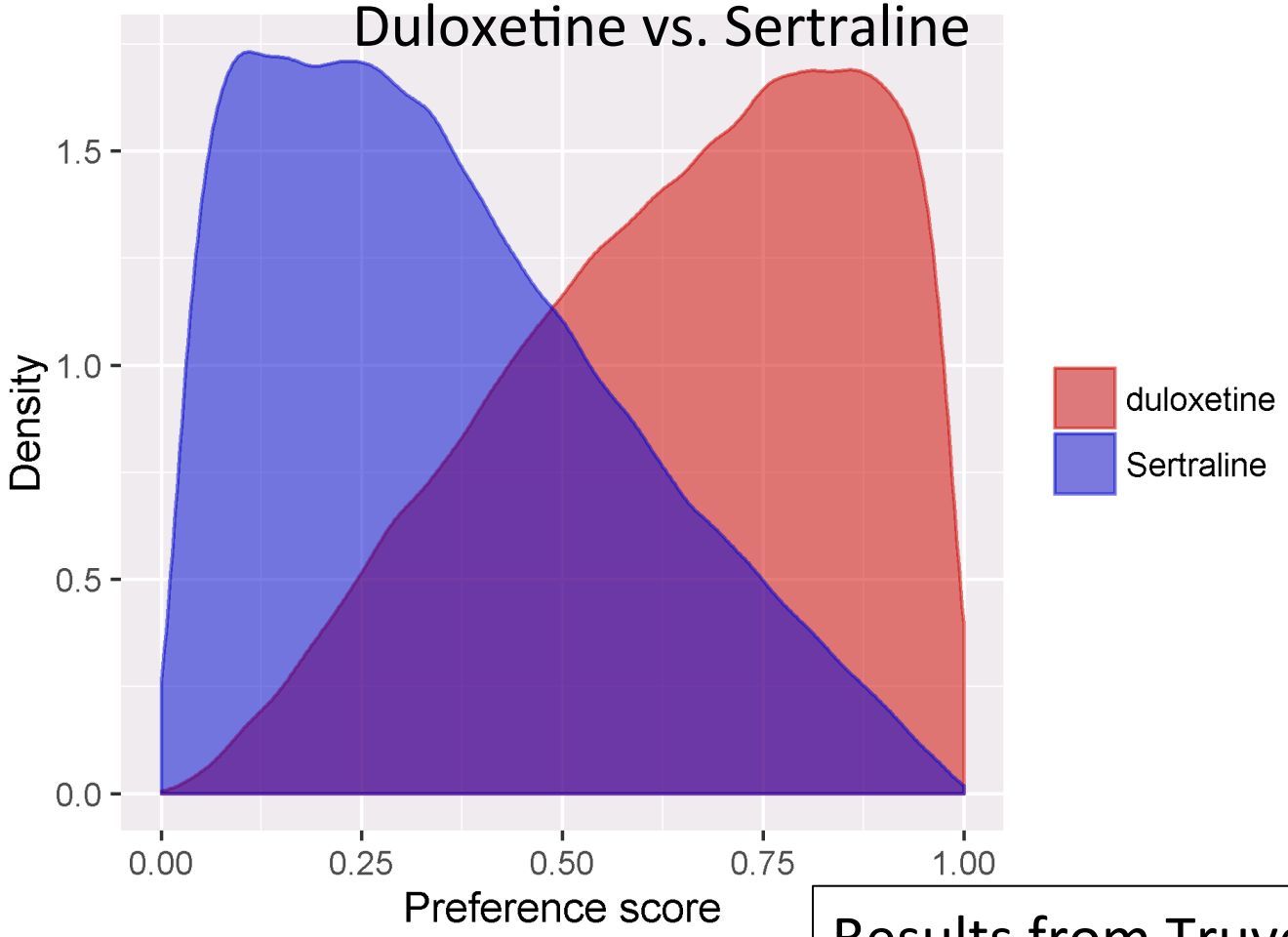
Yen-Chieh Lee, MD[a,+]; Chin-Hsien Lin, MD, PhD[b,+]; Min-Shung Lin, MD[a]; Yun Lu, MSc[c]; Chia-Hsuin Chang, MD, ScD[c,d,+]; and Jou-Wei Lin, MD, PhD[e]


	Adjusted Hazard Ratio[a] (95% CI)	P
.12	1.01 (0.90–1.12)	.91

Repeat the Lee study in OHDSI

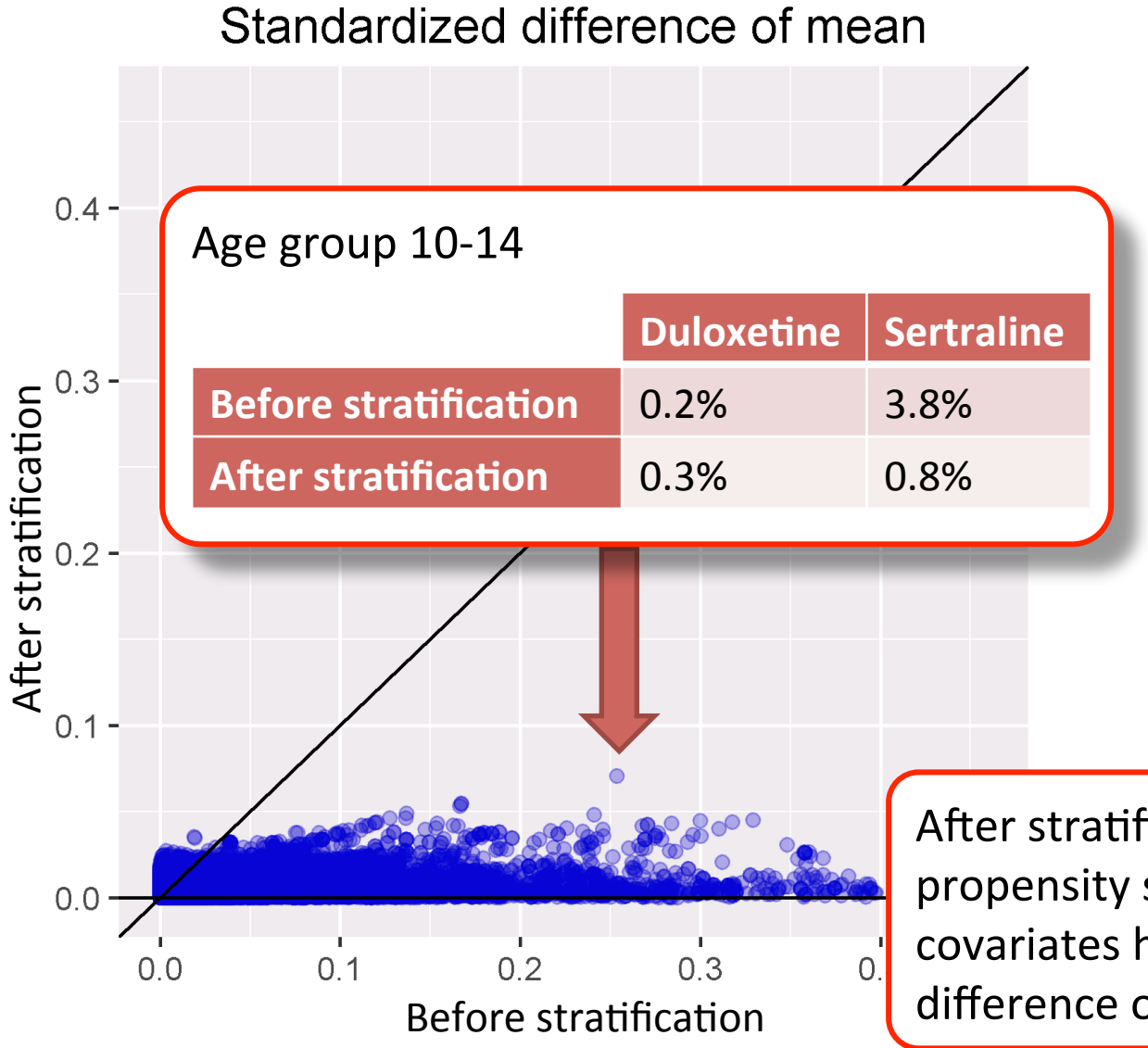
- Still had to infer many design features

Diagnose the propensity score distribution



Results from Truven CCAE
Duloxetine: n = 90,043
Sertraline: n = 175,950

Diagnose covariate balance

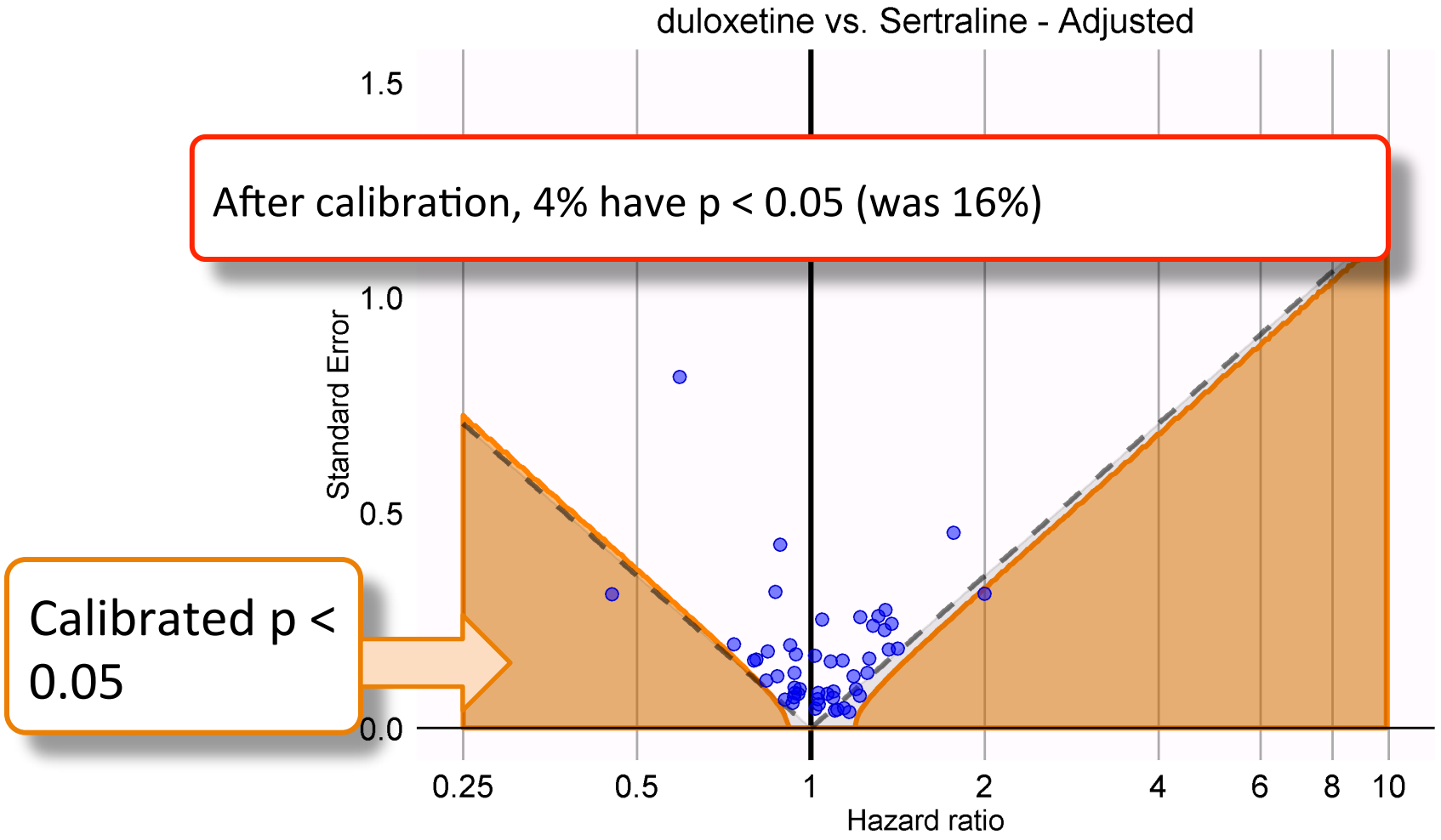


Age group 10-14

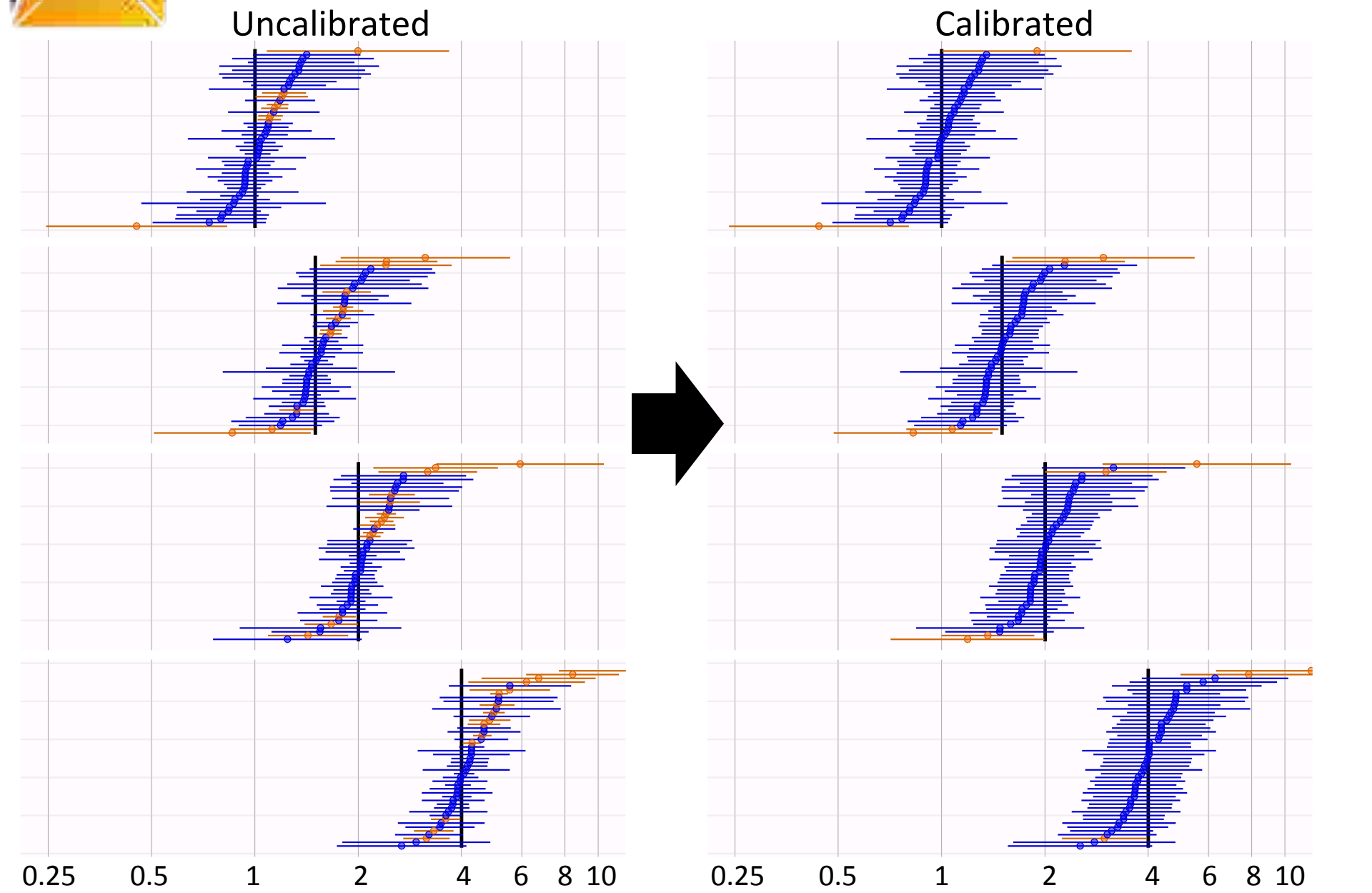
	Duloxetine	Sertraline
Before stratification	0.2%	3.8%
After stratification	0.3%	0.8%

After stratification on the propensity score, all 58,285 covariates have standardized difference of mean < 0.1

P-value calibration

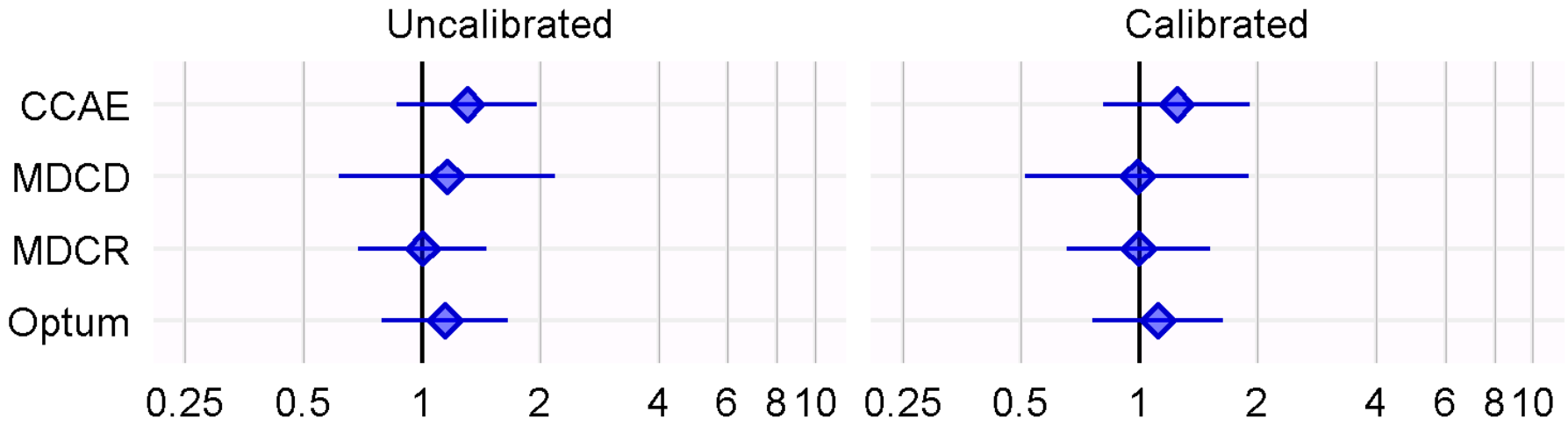


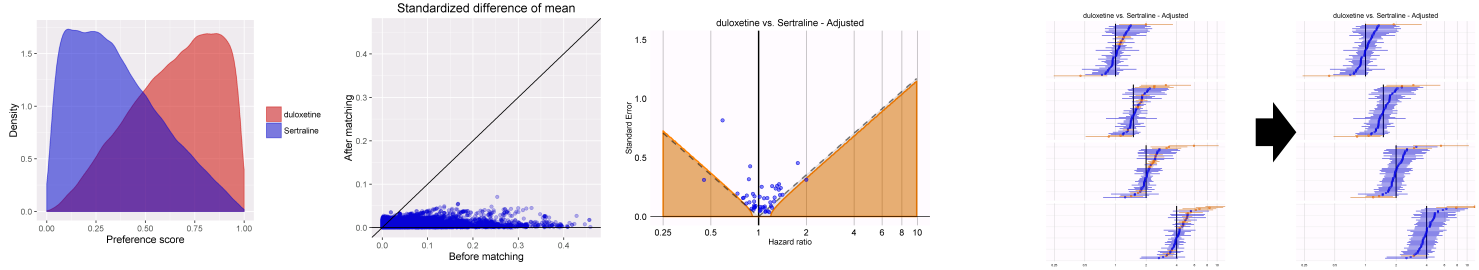
Confidence interval calibration



Proposed evidence for stroke

Duloxetine vs. Sertraline





Results are comparable to Lee et al., but we provide the context to interpret the results

What if we considered all outcomes?

Duloxetine vs. Sertraline for these 22 outcomes:

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

What if we consider all treatments?

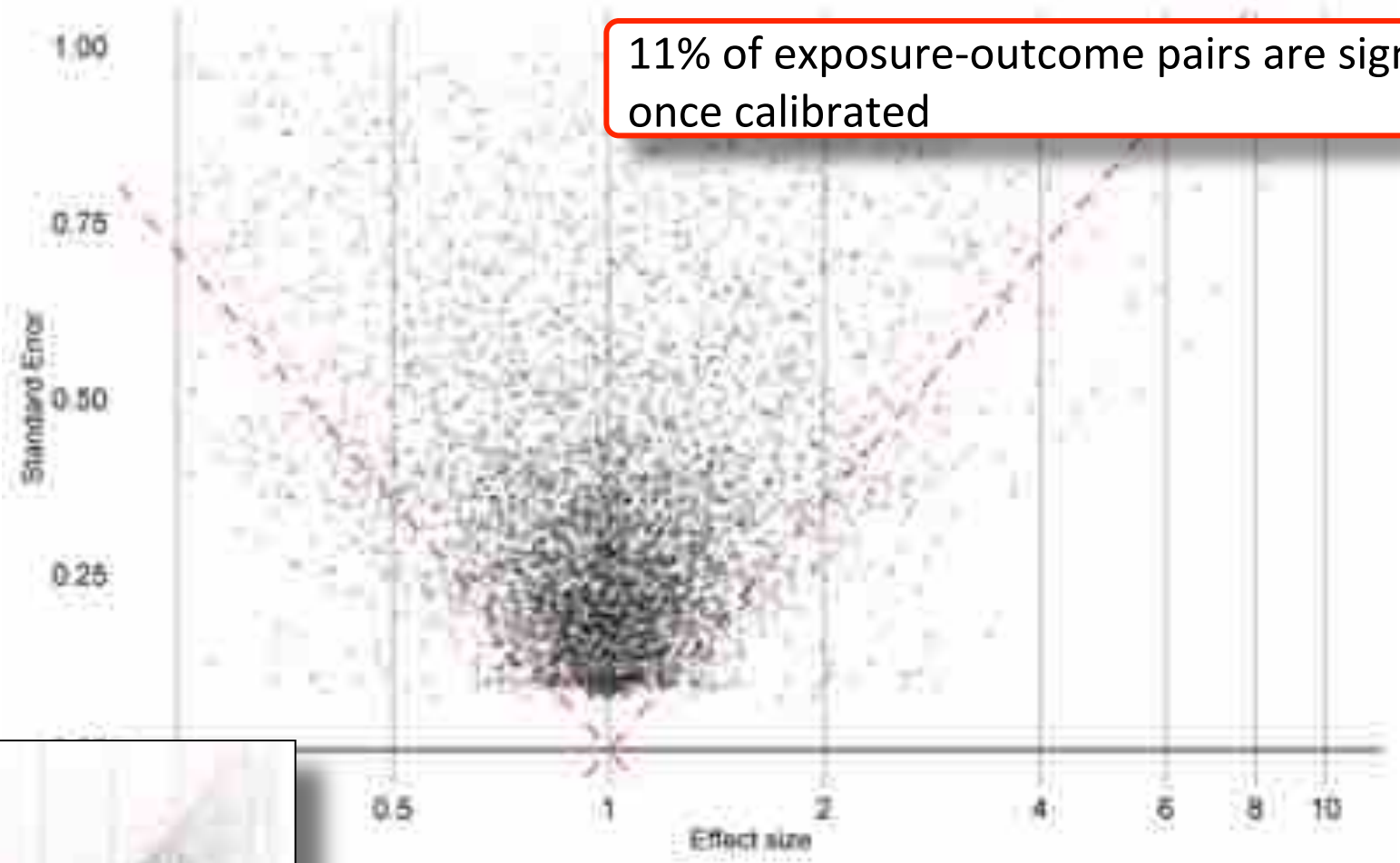
Type	Class	Treatment
Drug	Atypical	Bupropion
Drug	Atypical	Mirtazapine
Procedure	ECT	Electroconvulsive therapy
Procedure	Psychotherapy	Psychotherapy
Drug	SARI	Trazodone
Drug	SNRI	Desvenlafaxine
Drug	SNRI	duloxetine
Drug	SNRI	venlafaxine
Drug	SSRI	Citalopram
Drug	SSRI	Escitalopram
Drug	SSRI	Fluoxetine
Drug	SSRI	Paroxetine
Drug	SSRI	Sertraline
Drug	SSRI	vilazodone
Drug	TCA	Amitriptyline
Drug	TCA	Doxepin
Drug	TCA	Nortriptyline

Large-scale estimation for depression

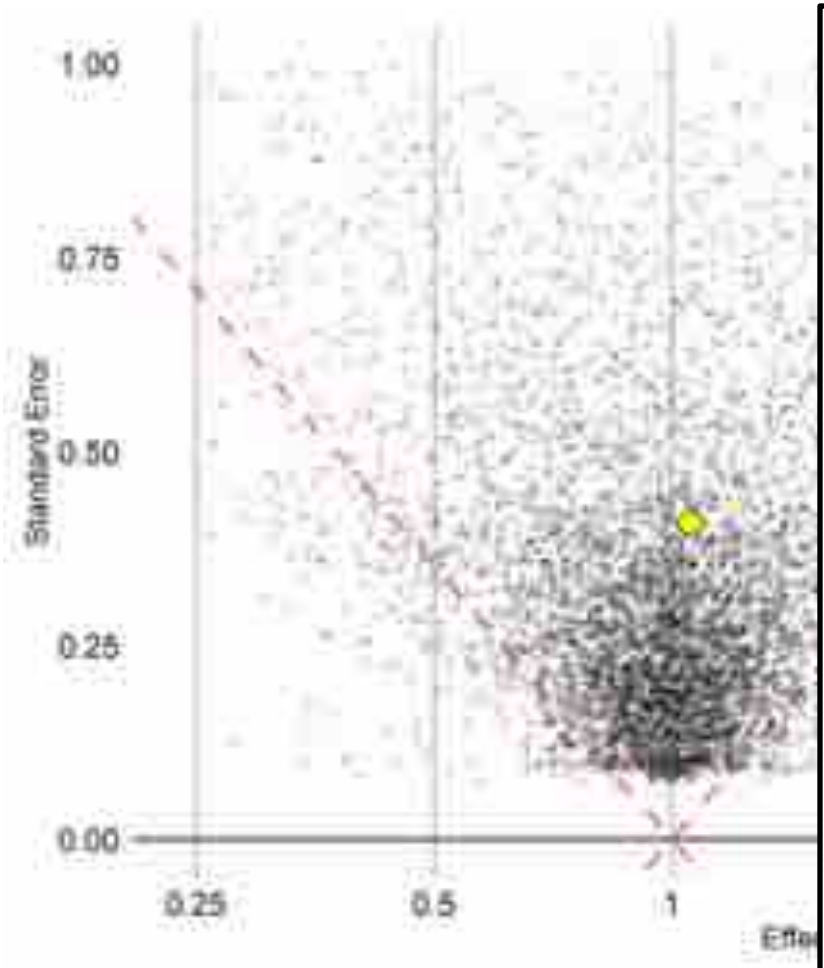
- 17 **treatments**
- 17 * 16 = 272 comparisons
- 22 **outcomes**
- 272 * 22 = 5,984 effect size estimates
- 4 **databases** (Truven CCAE, Truven MDCD, Truven MDCR, Optum)
- 4 * 5,984 = **23,936 estimates**

Estimates are in line with expectations

11% of exposure-outcome pairs are significant once calibrated



Example 1

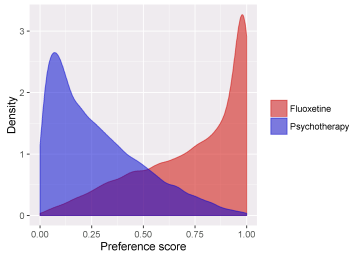


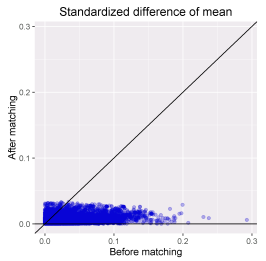
Fluoxetine vs. psychotherapy

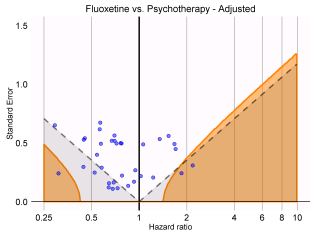
Suicide ideation

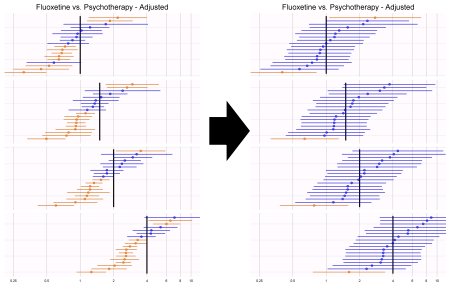
Database: Truven MDCR

Calibrated HR = 1.05 (0.51 – 2.51)

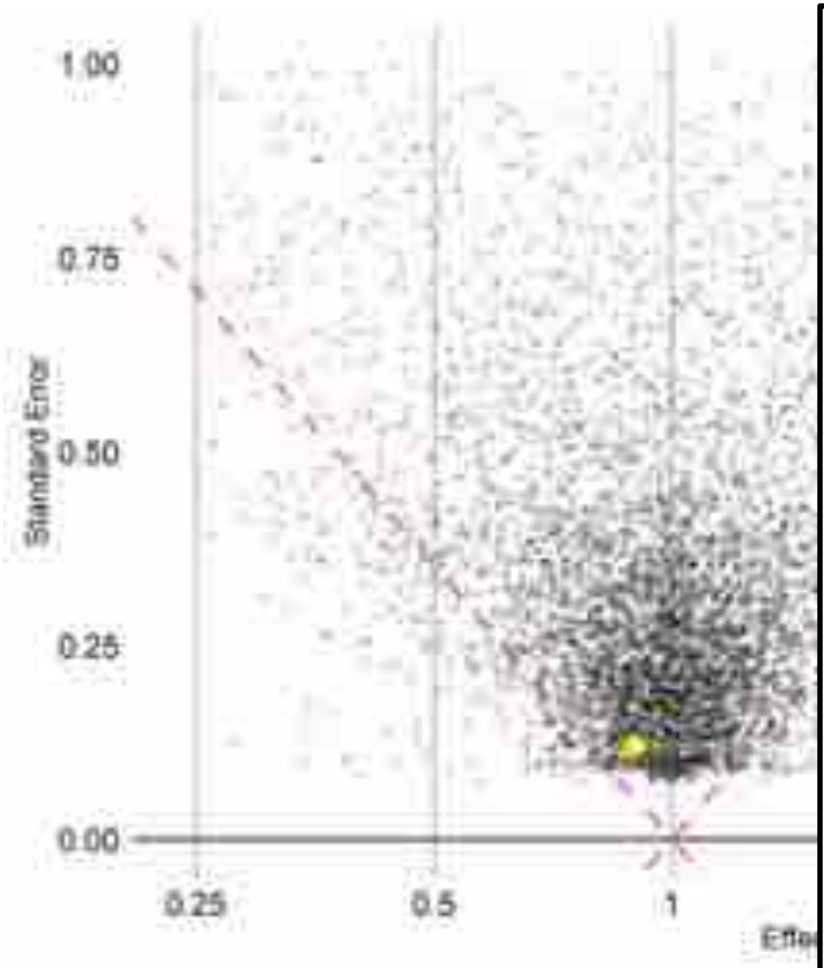








Example 2

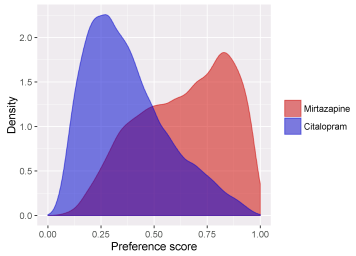


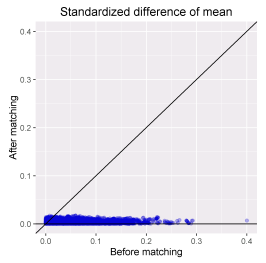
Mirtazapine vs. Citalopram

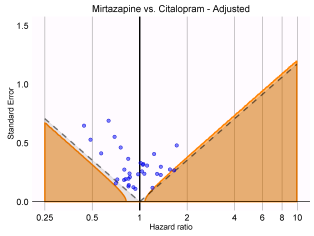
Constipation

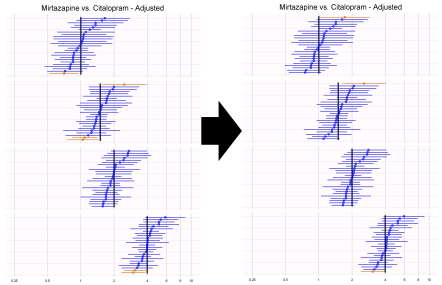
Database: Truven MDCD

Calibrated HR = 0.90 (0.70 – 1.12)

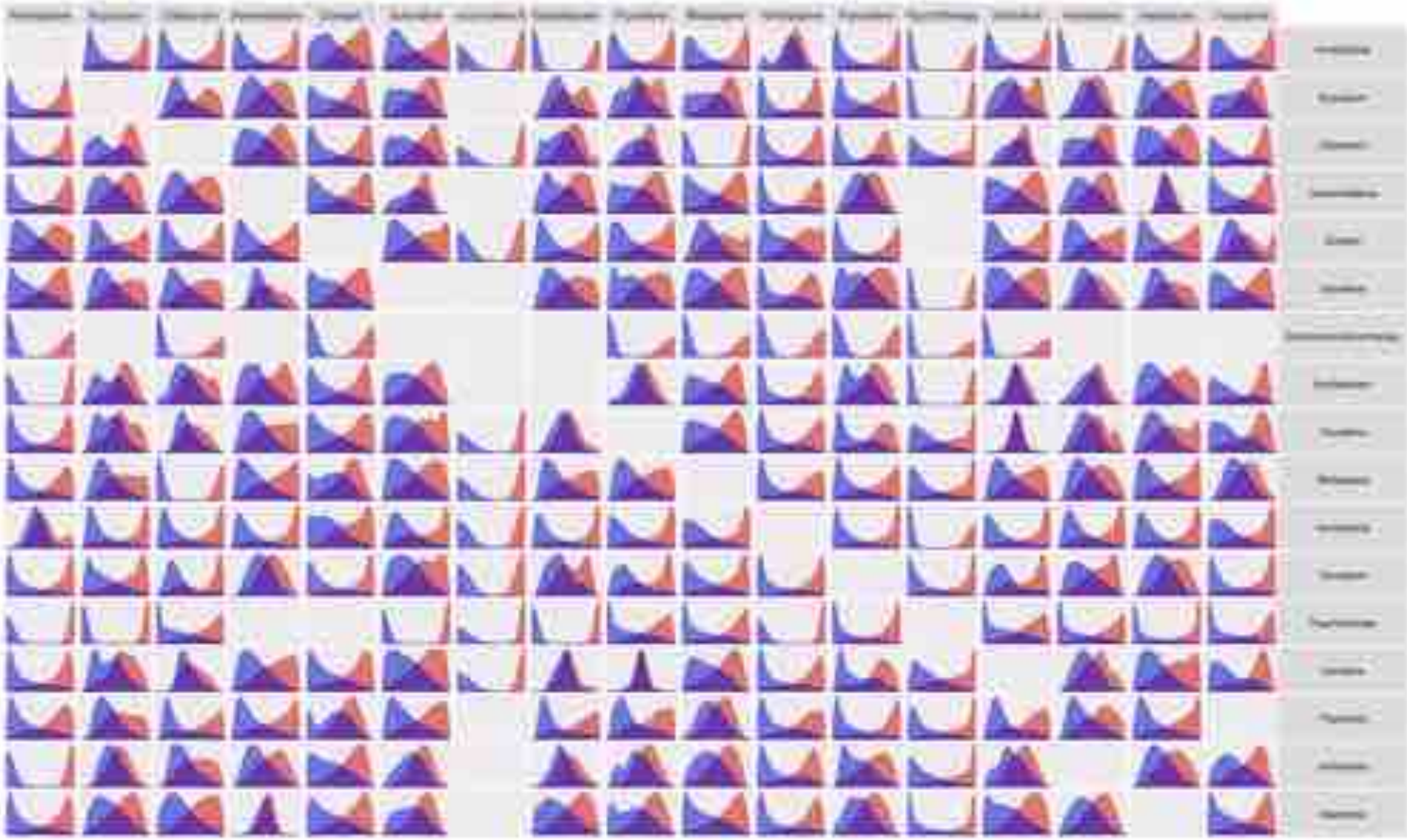








Propensity models for all comparisons (Truven CCAE, one outcome)



Large-scale estimation for depression

- Each estimate produced with same rigor, and could be published as a paper
  - Propensity score adjustment
  - Cox regression
  - Calibrated using negative and positive controls
  - …
- Calibration
  - Even if do not want to calibrate, must look at negative controls

Large-scale estimation for depression

- Not “data-dredging”!
  - Data-dredging is not about what you do but about what you *throw out*
  - This can’t be done for literature
  - Results should be interpreted considering multiple testing
- No reason not to carry out the other studies
  - Do not gain by not seeing them (blinding not relevant)
  - Studies are implicit in the data

Large-scale estimation for depression

- Bespoke studies
  - Wouldn't it be best to optimize each study
  - Never get 10 or 100 parameters right
  - Still good to see the surface
    - Large-scale sensitivity analysis
- At the very least, publish every last parameter so it can be reproduced

OHDSI recommendations for evidence generation

- ✓ Post protocol online
  - Prespecify research objectives and design decisions
- ✓ Make study code open source
  - From CDM to hazard ratios
- ✓ Use validated software
  - OHDSI Methods Library uses unit tests and simulation
- ✓ Replicate across several databases
  - 4 included so far, more will follow

https://github.com/OHDSI/StudyProtocols/LargeScalePopEst

OHDSI recommendations for evidence dissemination

- ✓ Address observation study bias
  - Addressed by adjusting for confounding, and **verifying** bias was addressed. Disseminate your diagnostics and evaluations.
- ✓ Address publication bias
  - Avoided by showing all tests that were performed, not just those that were significant
- ✓ Address CI-hacking
  - Very hard to fine-tune analysis to one specific result

Join the journey

http://ohdsi.org