Biomedical Informatics

discovery and impact

OHDSI: Drawing reproducible conclusions from observational clinical data

George Hripcsak, MD, MS

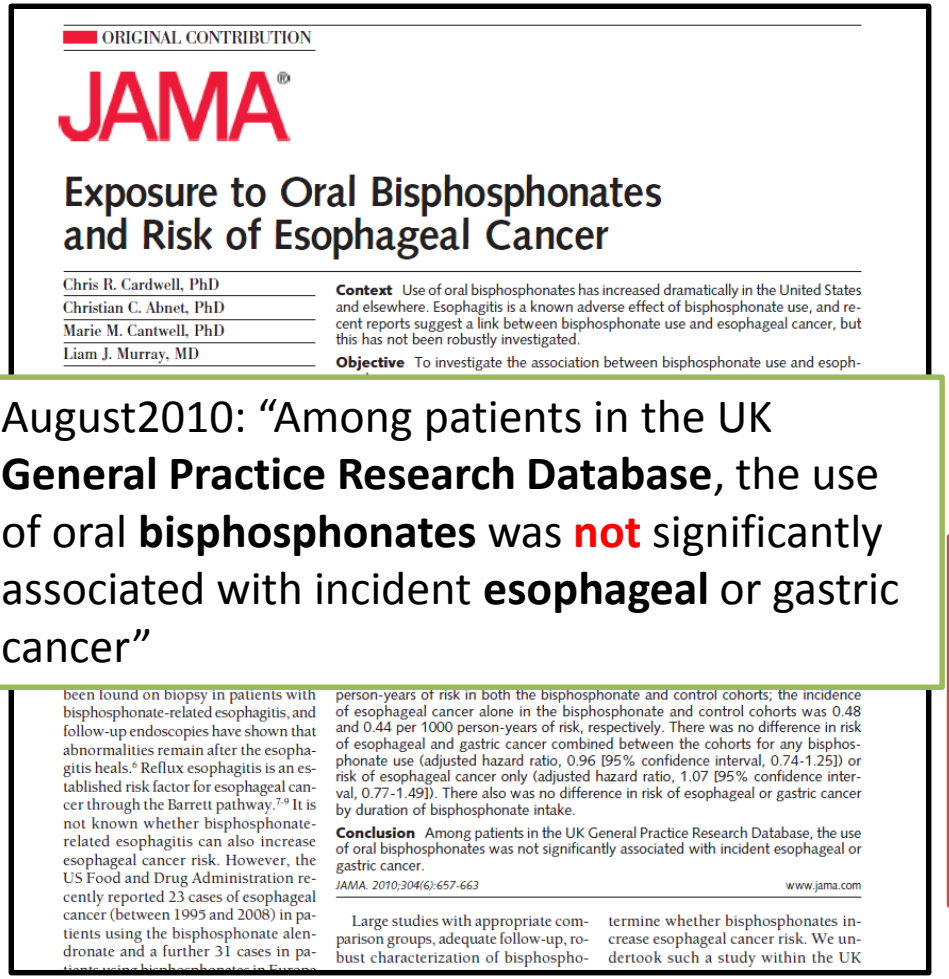
Biomedical Informatics, Columbia University

Medical Informatics Services, NewYork-Presbyterian

Columbia University Medical Center

NewYork-Presbyterian

Drawing reproducible conclusions



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 dramatically in the United States and elsewhere. Esophagitis is a known adverse effect of bisphosphonate use, and recent reports suggest a link between bisphosphonate use and esophageal cancer, but this has not been robustly investigated.

Objective To investigate the association between bisphosphonate use and esoph-

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-

person-years of risk in both the bisphosphonate and control cohorts; the incidence of esophageal cancer alone in the bisphosphonate and control cohorts was 0.48 and 0.44 per 1000 person-years of risk, respectively. There was no difference in risk of esophageal and gastric cancer combined between the cohorts for any bisphosphonate use (adjusted hazard ratio, 0.96 [95% confidence interval, 0.74-1.25]) or risk of esophageal cancer only (adjusted hazard ratio, 1.07 [95% confidence interval, 0.77-1.49]). There also was no difference in risk of esophageal or gastric cancer by duration of bisphosphonate intake.

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

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

Large studies with appropriate comparison groups, adequate follow-up, robust characterization of bisphospho- termine whether bisphosphonates increase esophageal cancer risk. We undertook such a study within the UK

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]

[1]Cancer Epidemiology Unit, University of Oxford, Oxford OX3 7LF
[2]Medicines and Healthcare products Regulatory Agency, Pharmacoepidemiology Research Unit, London SW8 5NQ
Correspondence to: J Green jane.green@ceu.ox.ac.uk

Cite this as: *BMJ* 2010;341:c4444
doi:10.1136/bmj.c4444

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—

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

corticosteroids. Cancers of the stomach and colorectum were not associated with prescription of bisphosphonate: relative risks for one or more versus no prescriptions were 0.87 (0.66 to 1.19) and 0.87 (0.77 to 1.00). The specificity

style data. General Practice Research Database prescription data have been shown to be virtually complete, and the data on incidence of cancer (based on hospital records) are around 95% valid and

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**"

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

Aiming for 1,000,000,000 patient data network

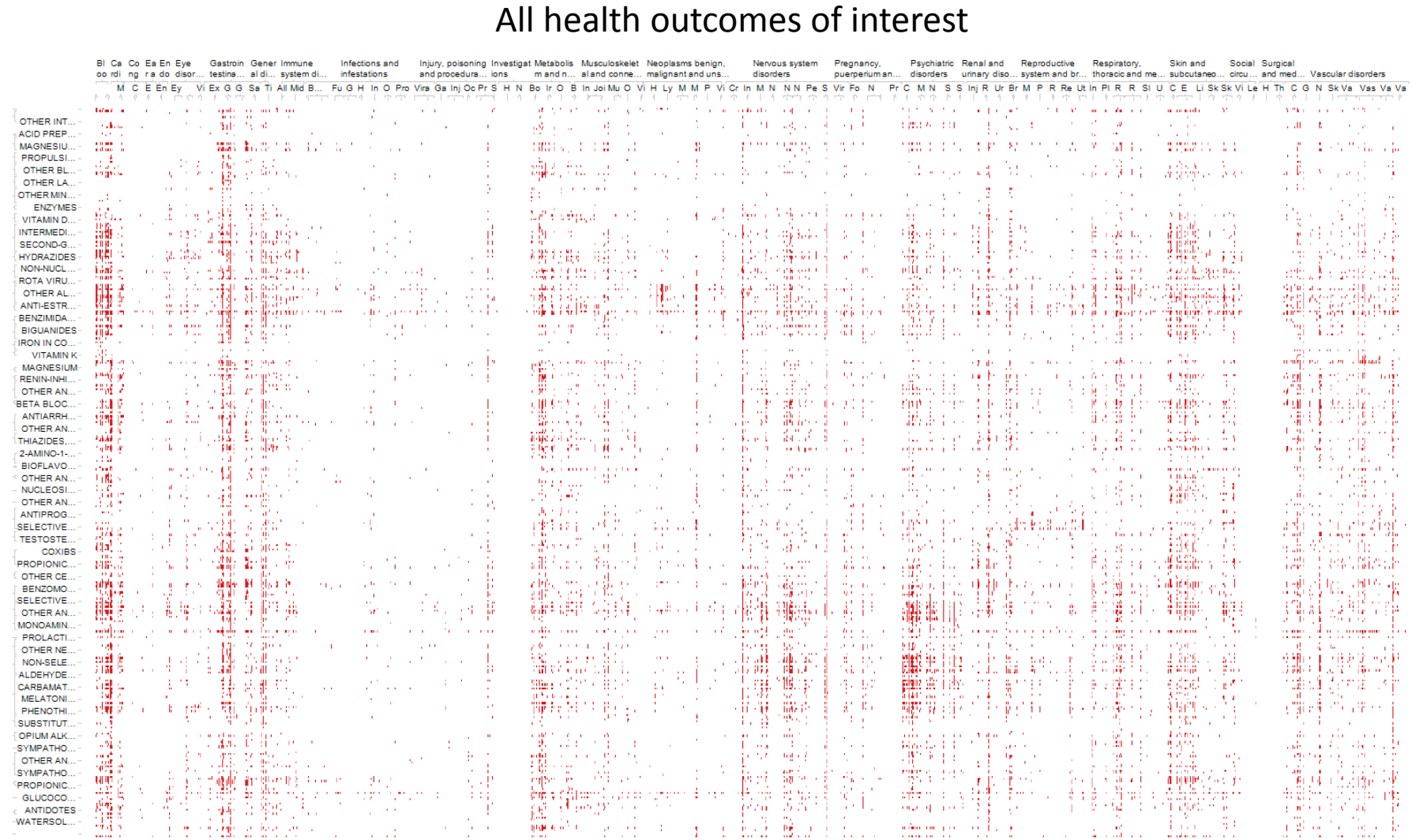
http://ohdsi.org

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
- Over a billion records on >400 million patients in 80 databases

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

Why large-scale analysis is needed in healthcare



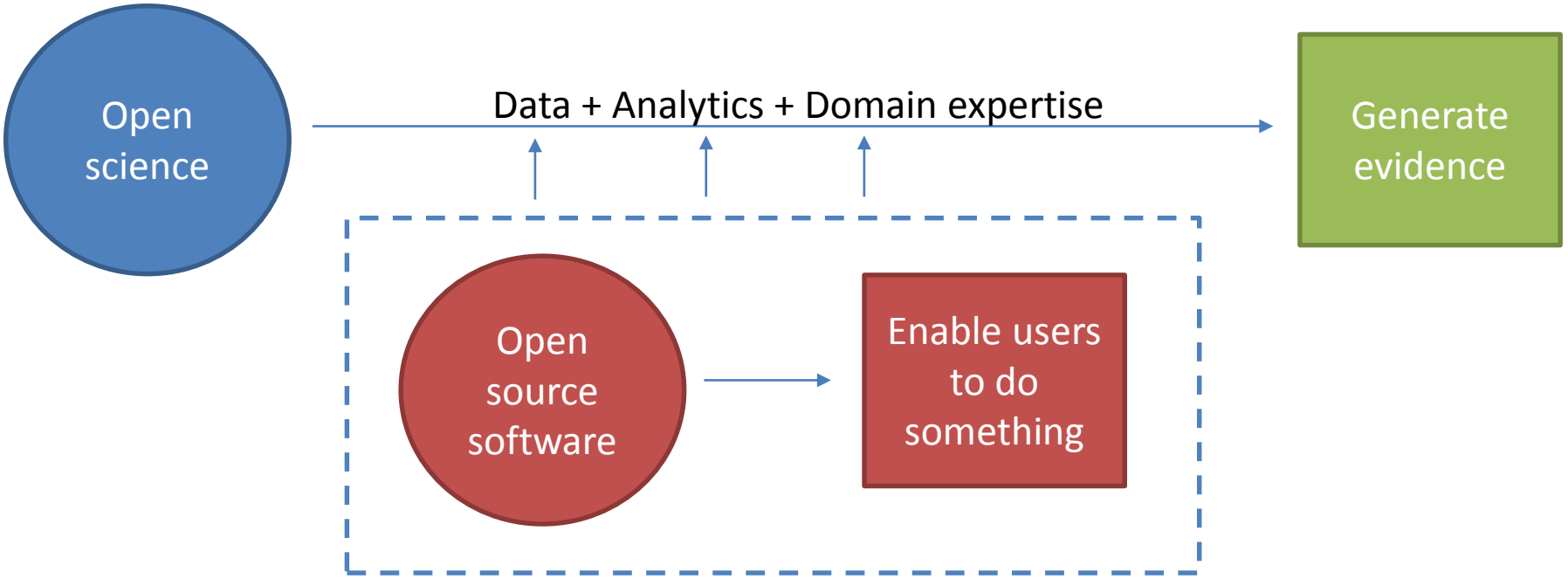
Patient-level predictions for personalized evidence requires big data

2 million patients seem excessive or unnecessary?

- Imagine a provider wants to compare her patient with other patients with the same gender (50%), in the same 10-year age group (10%), and with the same comorbidity of Type 2 diabetes (5%)
- Imagine the patient is concerned about the risk of ketoacidosis (0.5%) associated with two alternative treatments they are considering
- With 2 million patients, you'd only expect to observe 25 similar patients with the event, and would only be powered to observe a relative risk > 2.0

Aggregated data across a health system of 1,000 providers may contain 2,000,000 patients

OHDSI's approach to open science

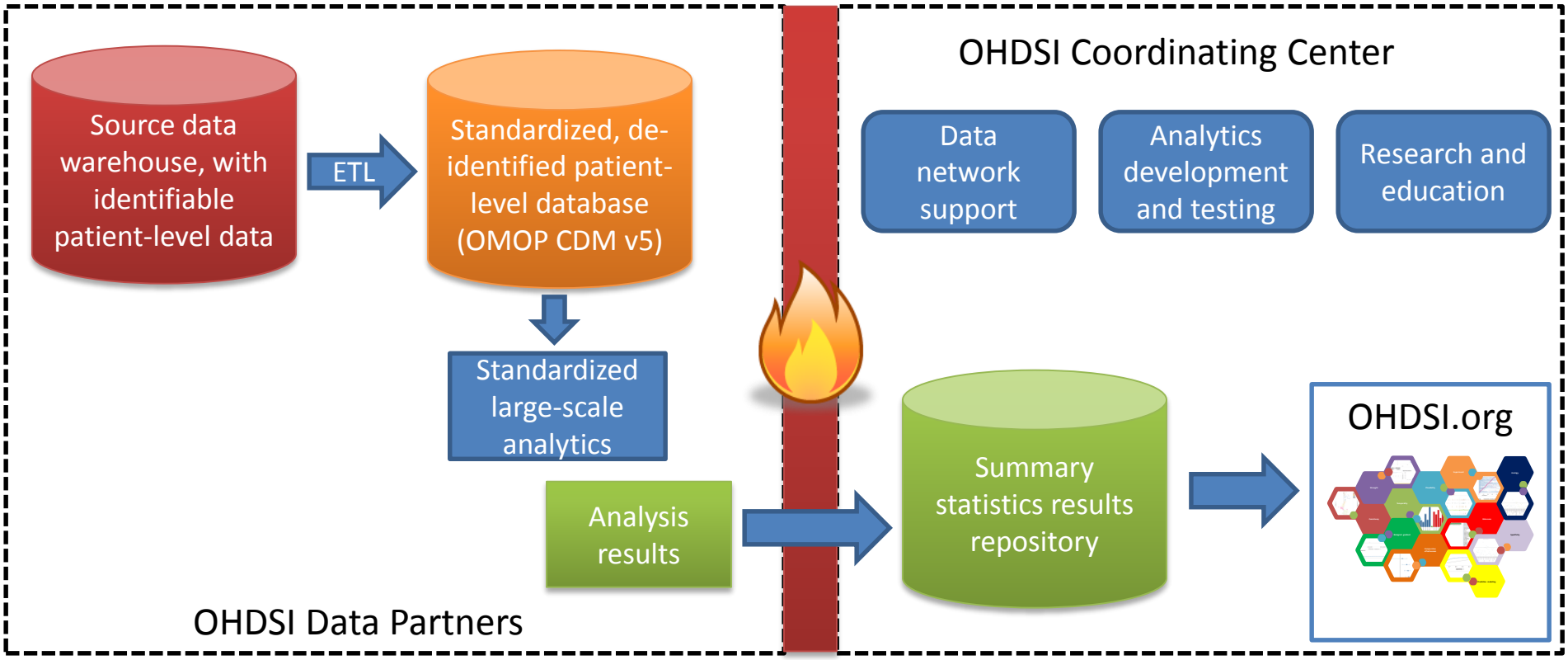


- Open science is about sharing the journey to evidence generation
- Open-source software can be part of the journey, but it's not a final destination
- Open processes can enhance the journey through improved reproducibility of research and expanded adoption of scientific best practices

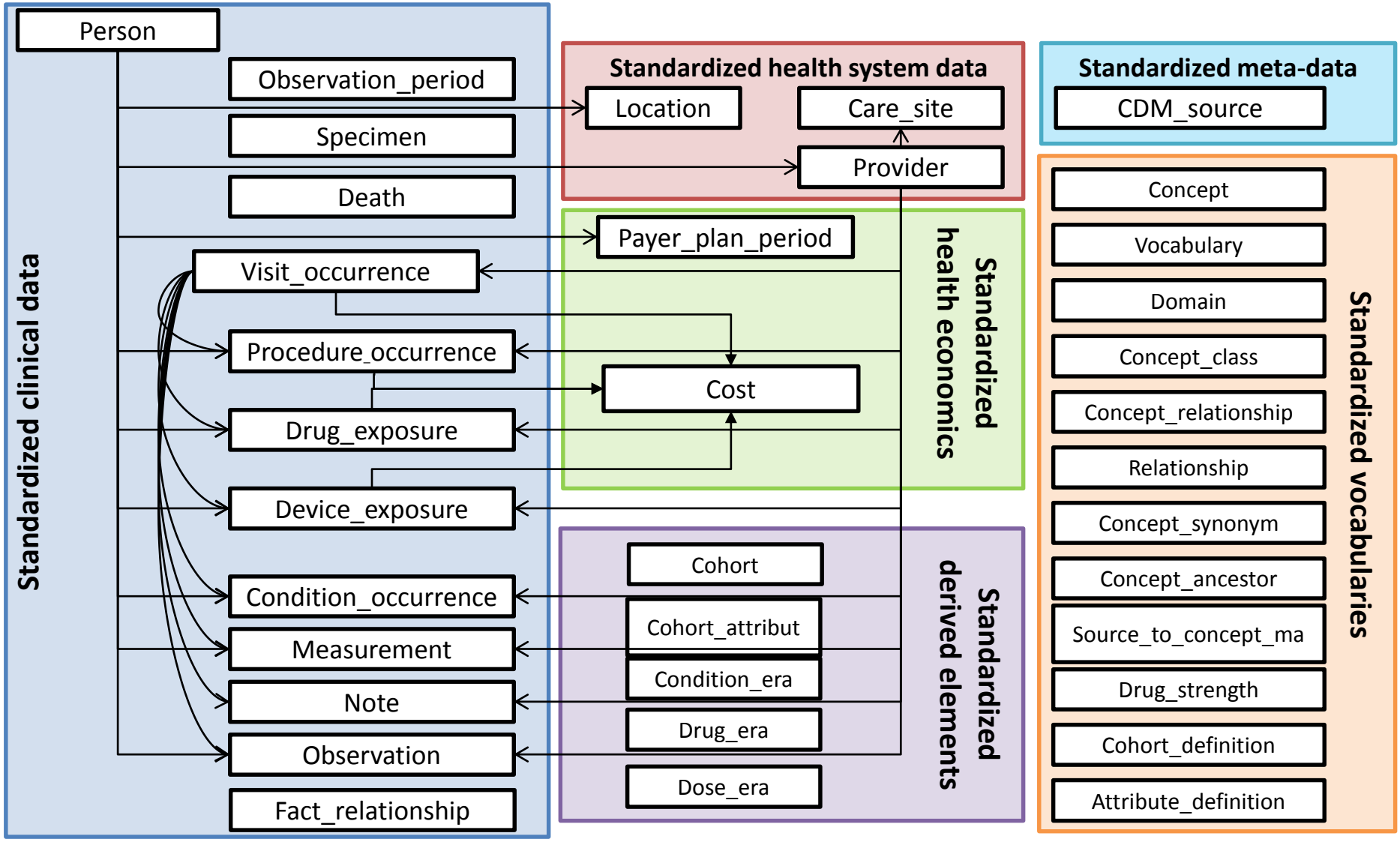
Evidence OHDSI seeks to generate from observational data

- **Clinical characterization**
  - Natural history: Who has diabetes, and who takes metformin?
  - Quality improvement: What proportion of patients with diabetes experience complications?
- **Population-level estimation**
  - Safety surveillance: Does metformin cause lactic acidosis?
  - Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?
- **Patient-level prediction**
  - 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?

How OHDSI Works

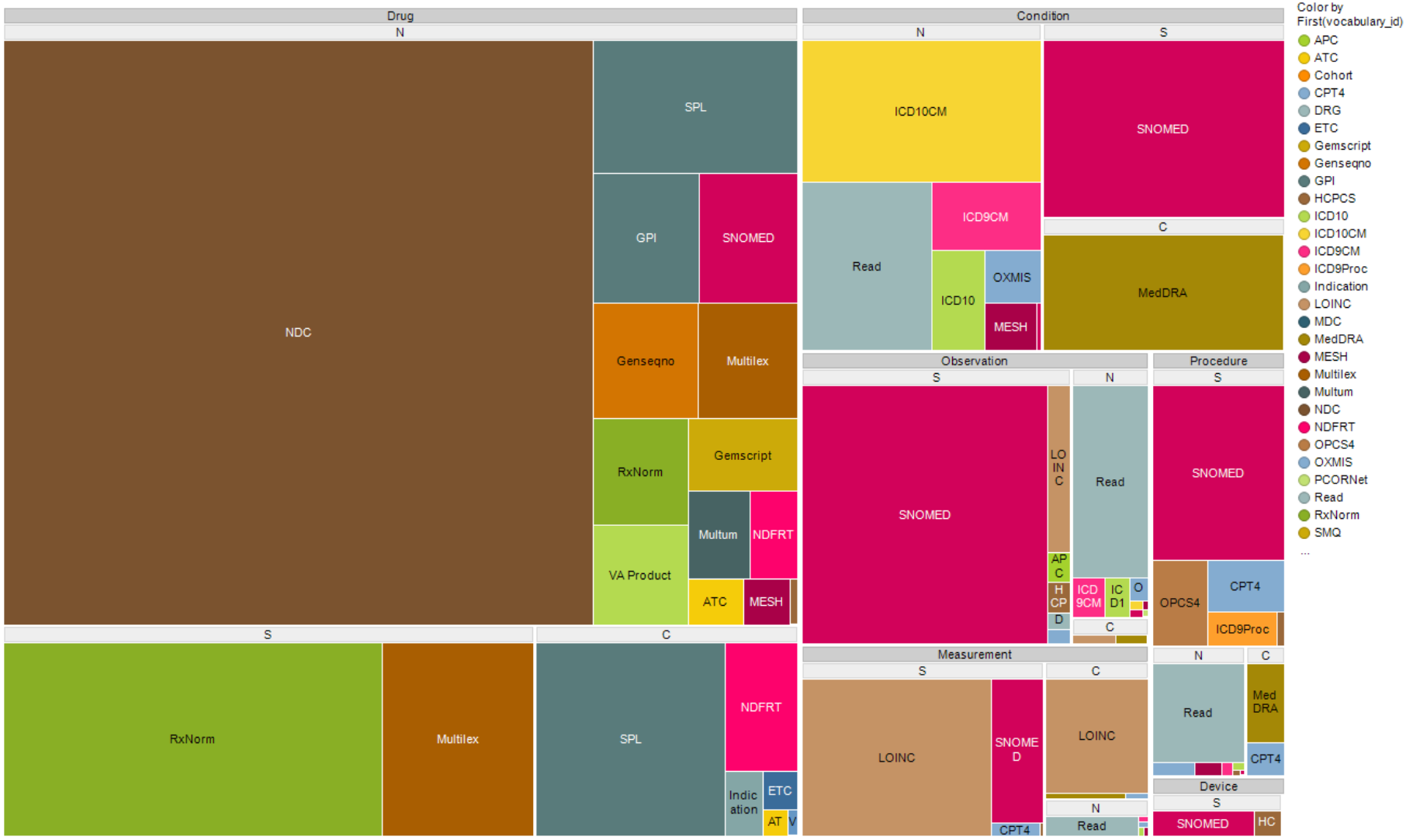


Deep information model
OMOP CDM v5

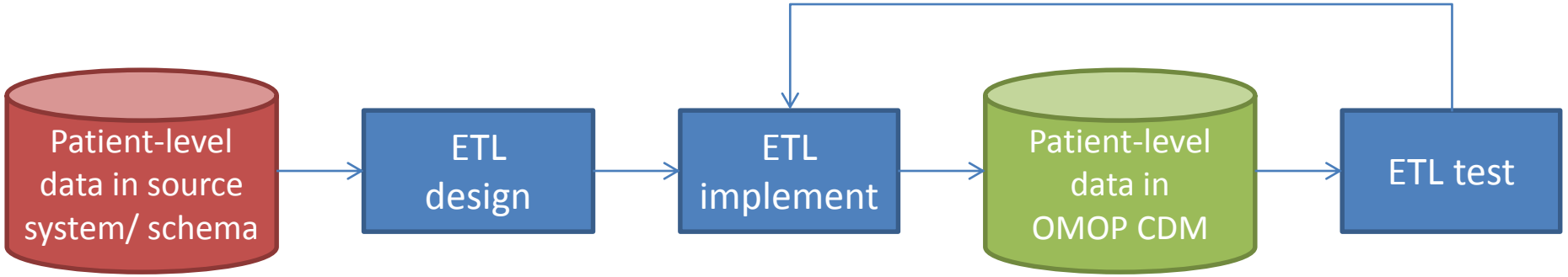


Extensive vocabularies

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



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

http://github.com/OHDSI

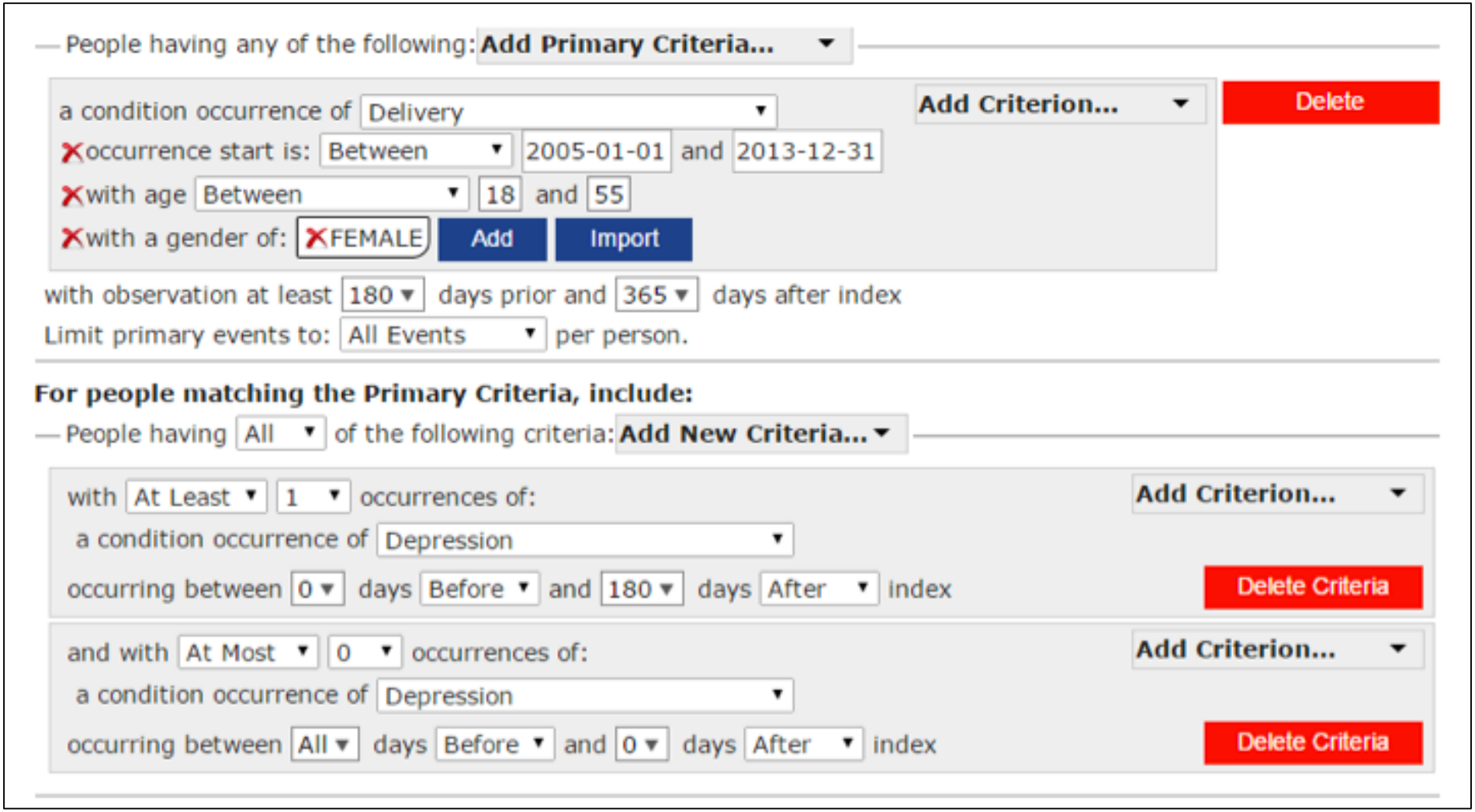
ACHILLES Heel Data Validation

Data Quality Messages

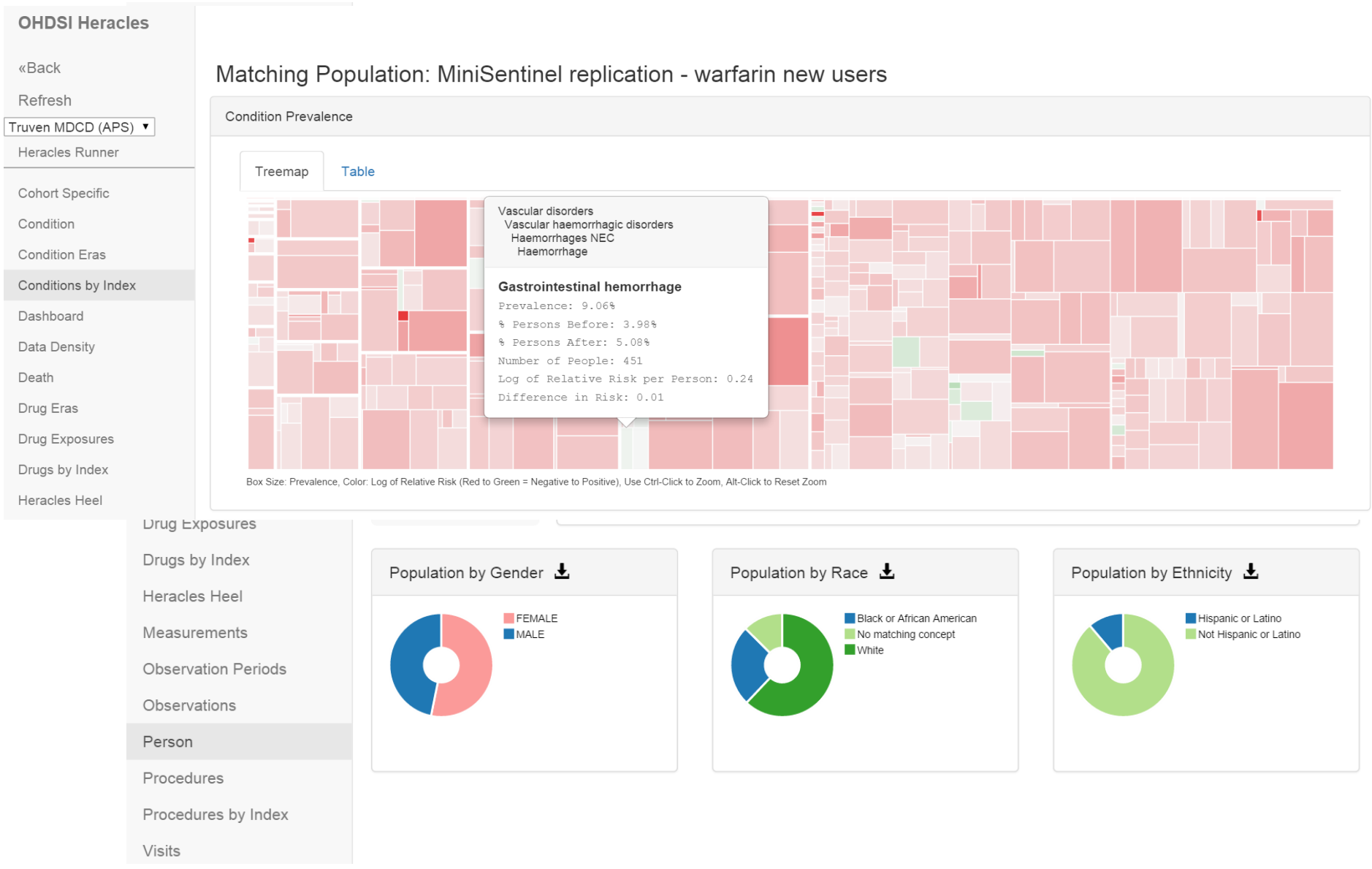
Search: Show / hide columns

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

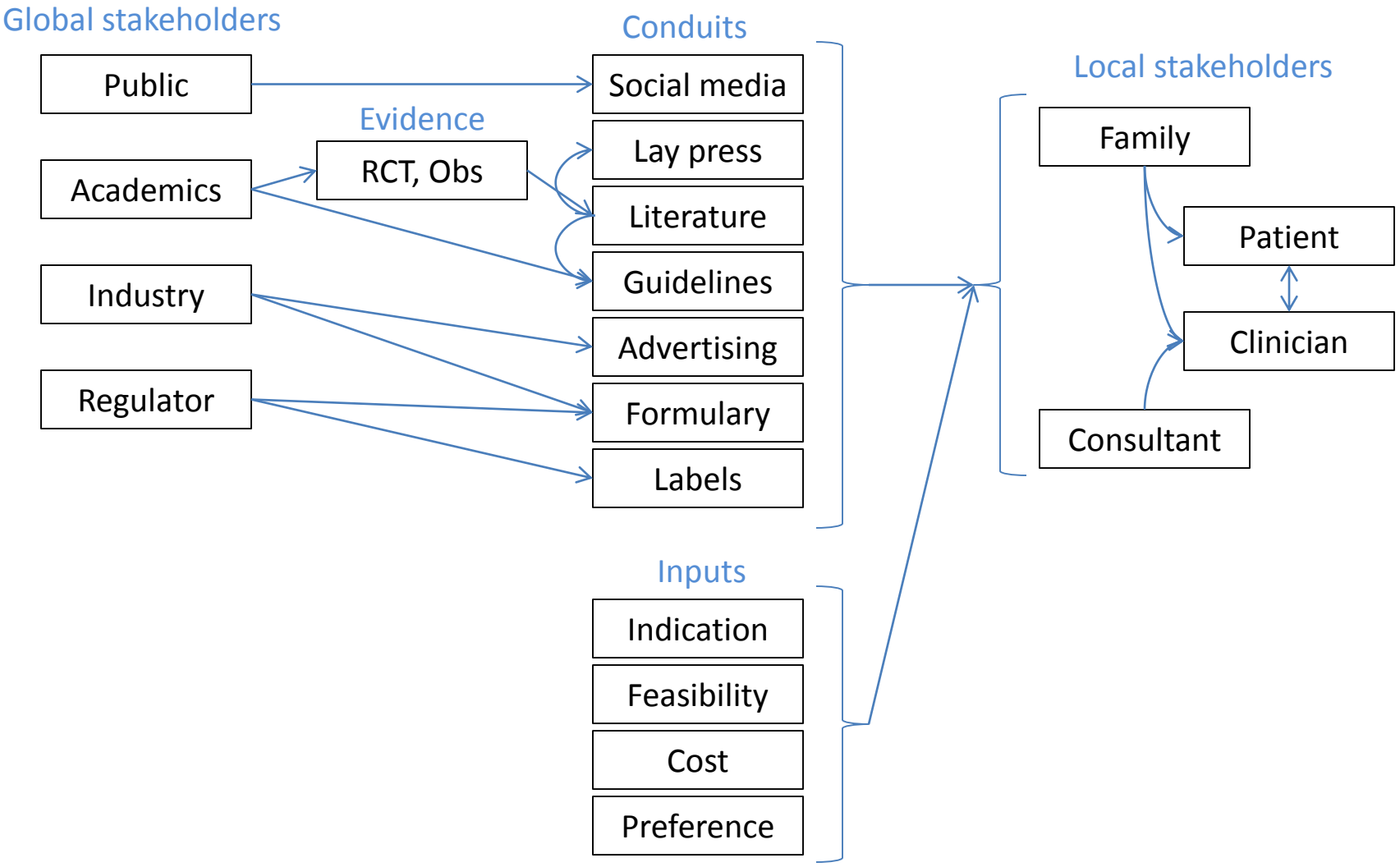


Characterize the cohorts of interest



OHDSI in Action

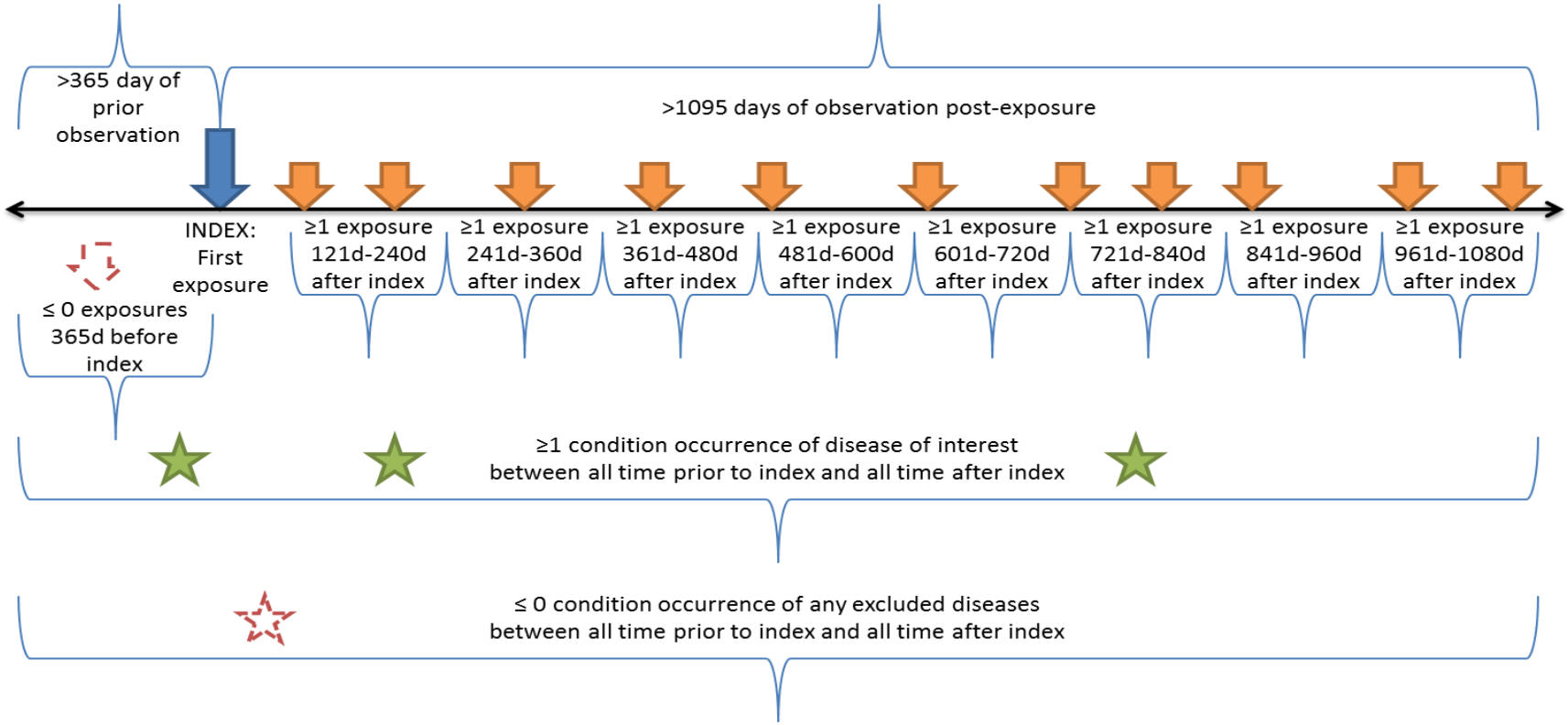
Treatment Pathways



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

Treatment pathway event flow



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

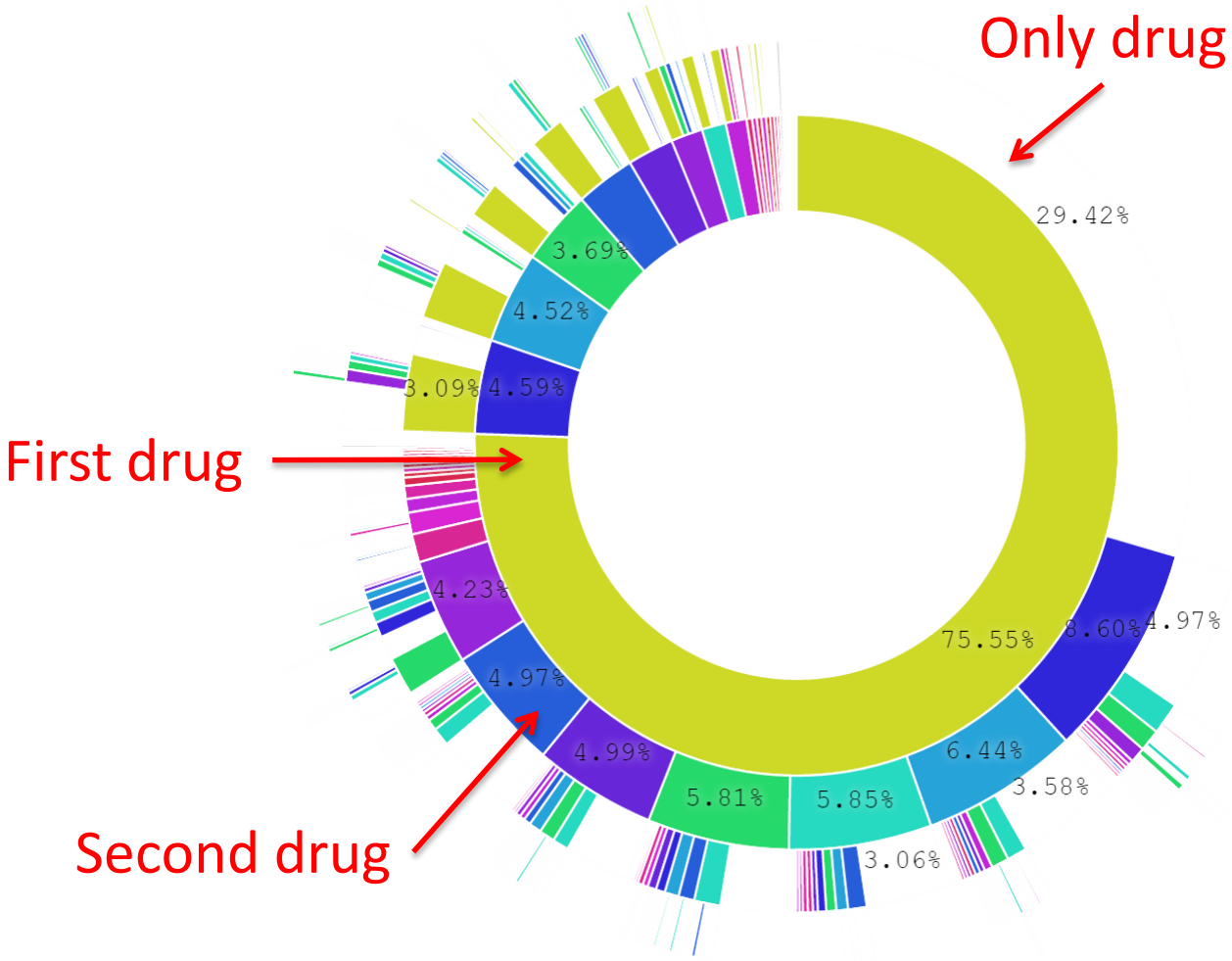
Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 5, 2016 (received for review June 14, 2015)

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

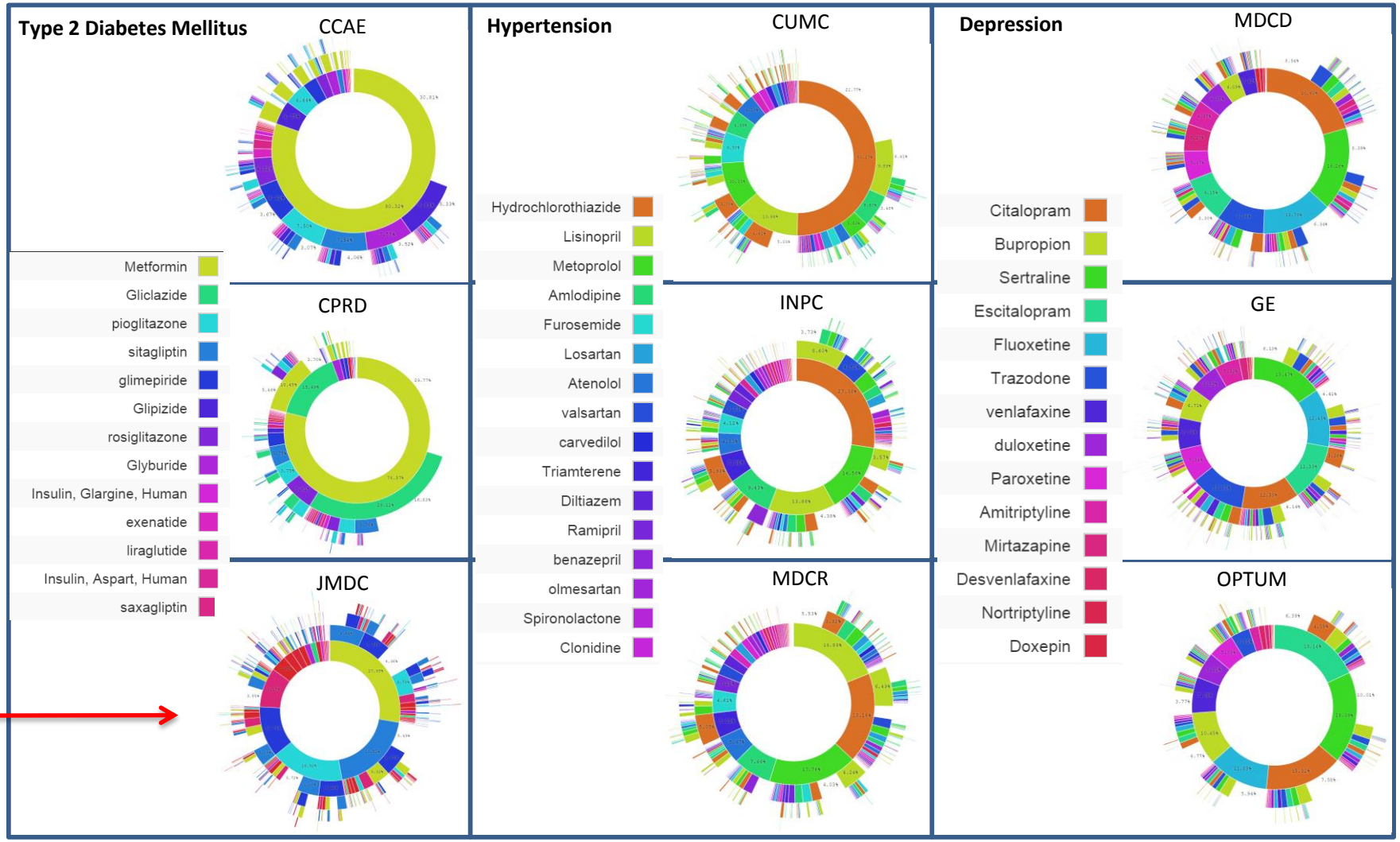
Treatment pathways for diabetes

T2DM : All databases



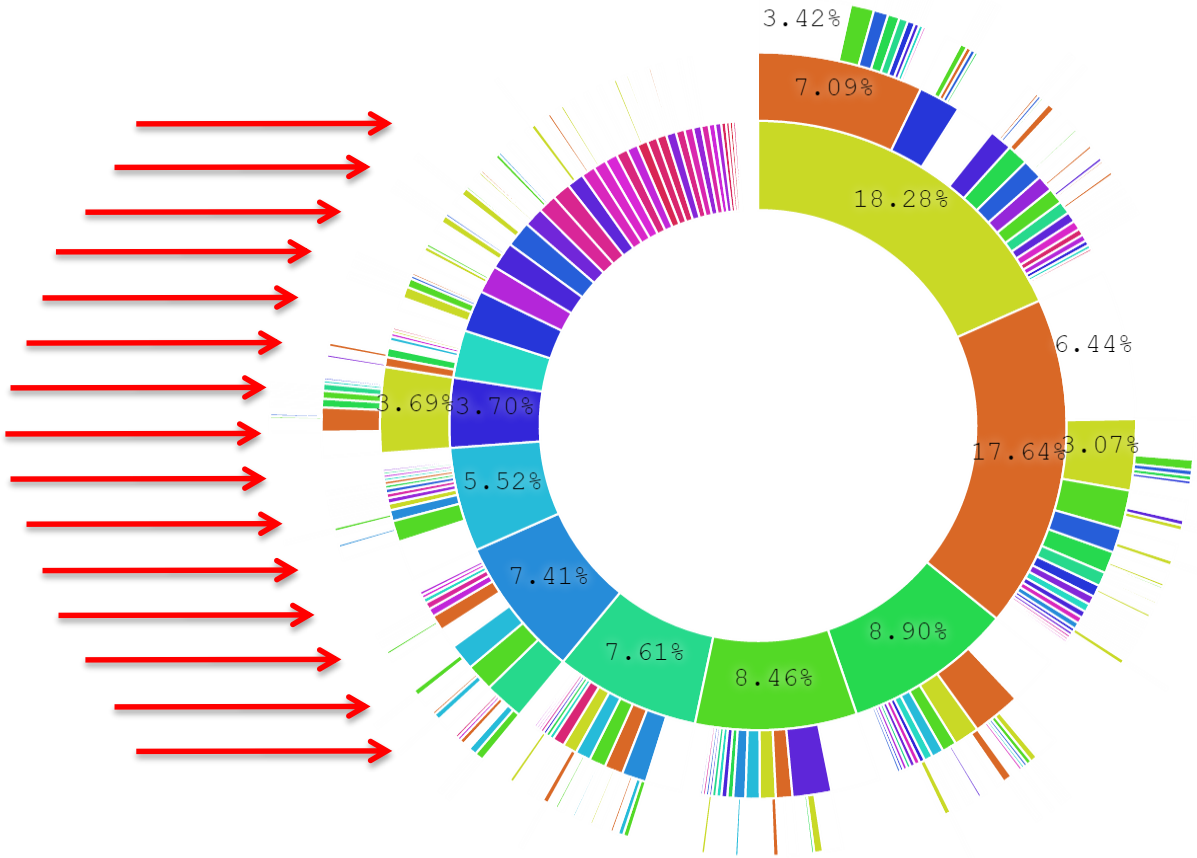
- Metformin
- pioglitazone
- sitagliptin
- Glipizide
- glimepiride
- Gliclazide
- Glyburide
- rosiglitazone
- Insulin, Glargine, Human
- exenatide
- Insulin, Aspart, Human
- liraglutide
- saxagliptin
- Insulin, Lispro, Human
- Glucose
- Insulin, Isophane, Human

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



Patient-level heterogeneity

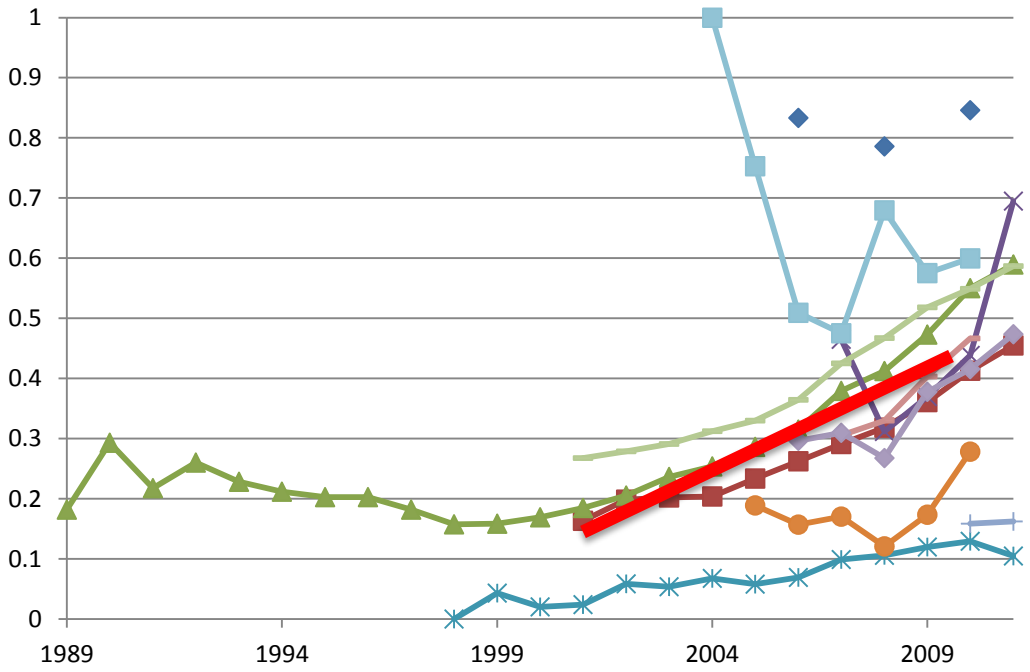
HTN: All databases



25% of HTN patients (10% of others) have a unique path despite 250M pop

Monotherapy – diabetes

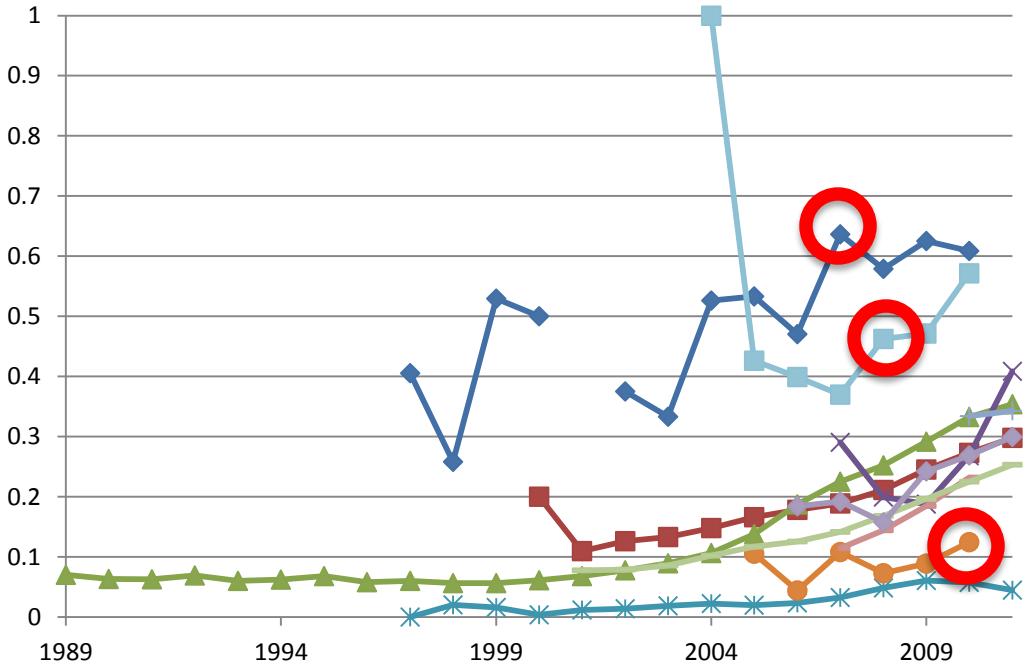
General upward trend in monotherapy



AUSOM (SKorea*)
CCAE (US#)
CPRD (UK*)
CUMC (US*)
GE (US*)
INPC (US*#)
JMDC (Japan#)
MDCD (US#)
MDCR (US#)
OPTUM (US#)
STRIDE (US*)

Monotherapy – HTN

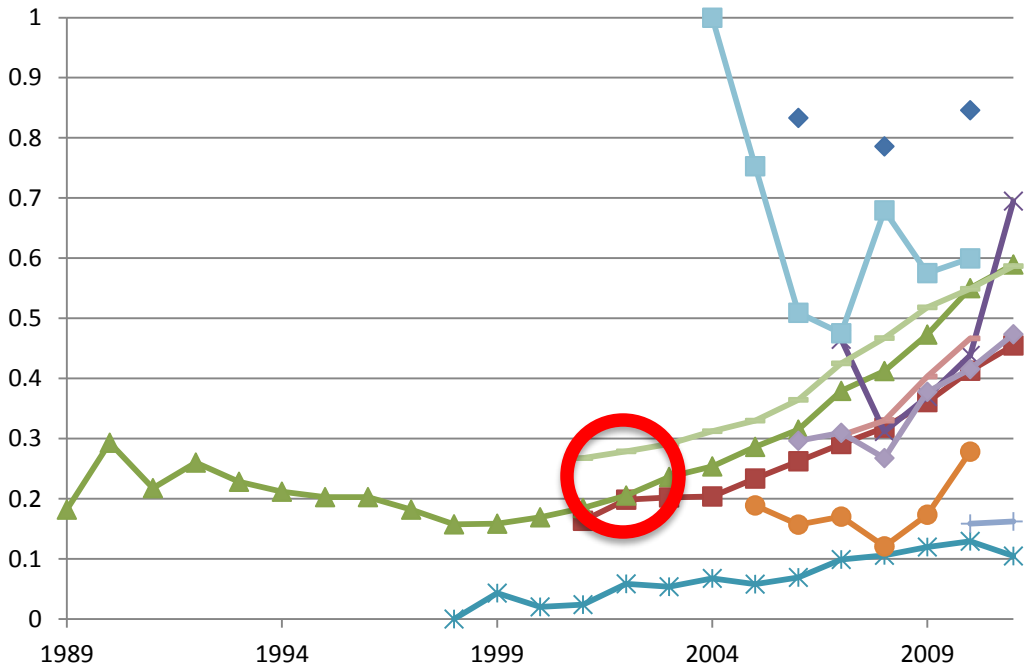
Academic medical centers differ from general practices



AUSOM (SKorea*) CCAE (US#) CPRD (UK*) CUMC (US*)
GE (US*) INPC (US*#) JMDC (Japan#) MDCD (US#)
MDCR (US#) OPTUM (US#) STRIDE (US*)

Monotherapy – diabetes

General practices, whether EHR or claims, have similar profiles



AUSOM (SKorea*) CCAE (US#) CPRD (UK*) CUMC (US*)
GE (US*) INPC (US*#) JMDC (Japan#) MDCD (US#)
MDCR (US#) OPTUM (US#) STRIDE (US*)

Conclusions: Network research

- It is feasible to encode the world population in a single data model
  - Over 1,000,000,000 records by voluntary effort
- Generating evidence is feasible
- Stakeholders willing to share results
- Able to accommodate vast differences in privacy and research regulation

Open science

- Admit that there is a problem
- Study it scientifically
  - Define that surface and differentiate true variation from confounding …
- Total description of every study
- Research into new methods

Take a scientific approach to science

1. Database heterogeneity:
   Holding analysis constant, different data may yield different estimates

Madigan D, Ryan PB, Schuemie MJ et al, American Journal of Epidemiology, 2013
"Evaluating the Impact of Database Heterogeneity on Observational Study Results"

2. Parameter sensitivity:
   Holding data constant, different analytic design choices may yield different estimates

Madigan D, Ryan PB, Schuemie MJ, Therapeutic Advances in Drug Safety, 2013: "Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies"

3. Empirical performance:
   Most observational methods do not have nominal statistical operating characteristics

Ryan PB, Stang PE, Overhage JM et al, Drug Safety, 2013:
"A Comparison of the Empirical Performance of Methods for a Risk Identification System"

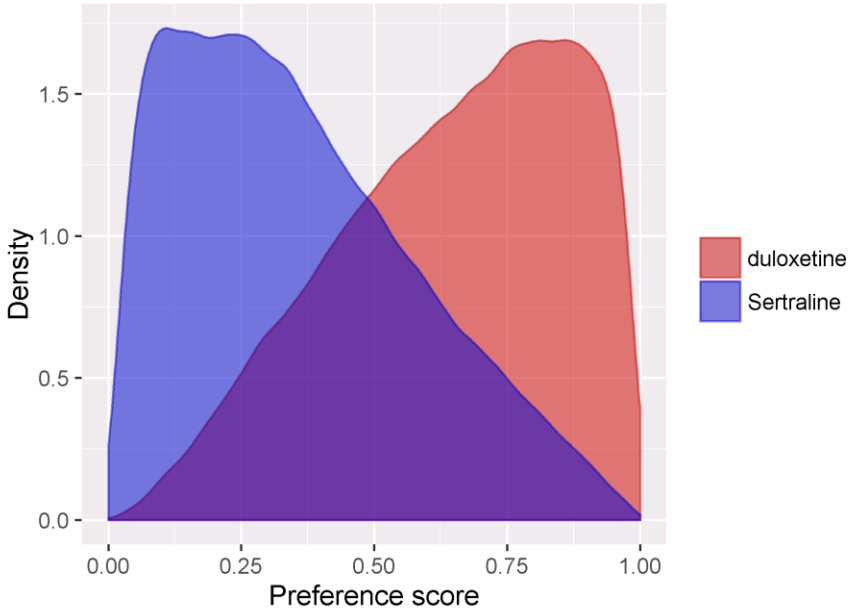
4. Empirical calibration can help restore interpretation of study findings

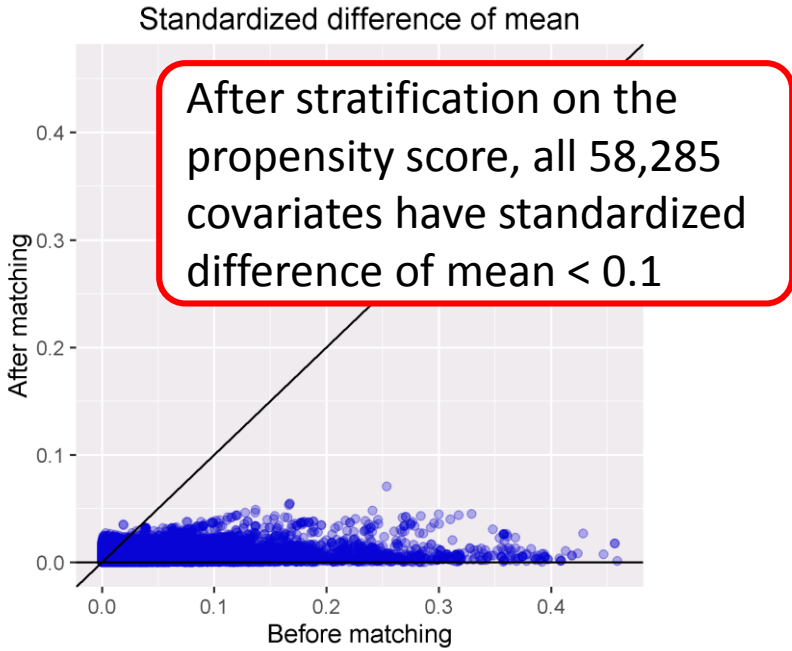
Schuemie MJ, Ryan PB, DuMouchel W, et al, Statistics in Medicine, 2013:
"Interpreting observational studies: why empirical calibration is needed to correct p-values"

Reproducible research

1. Address confounding that is measured

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

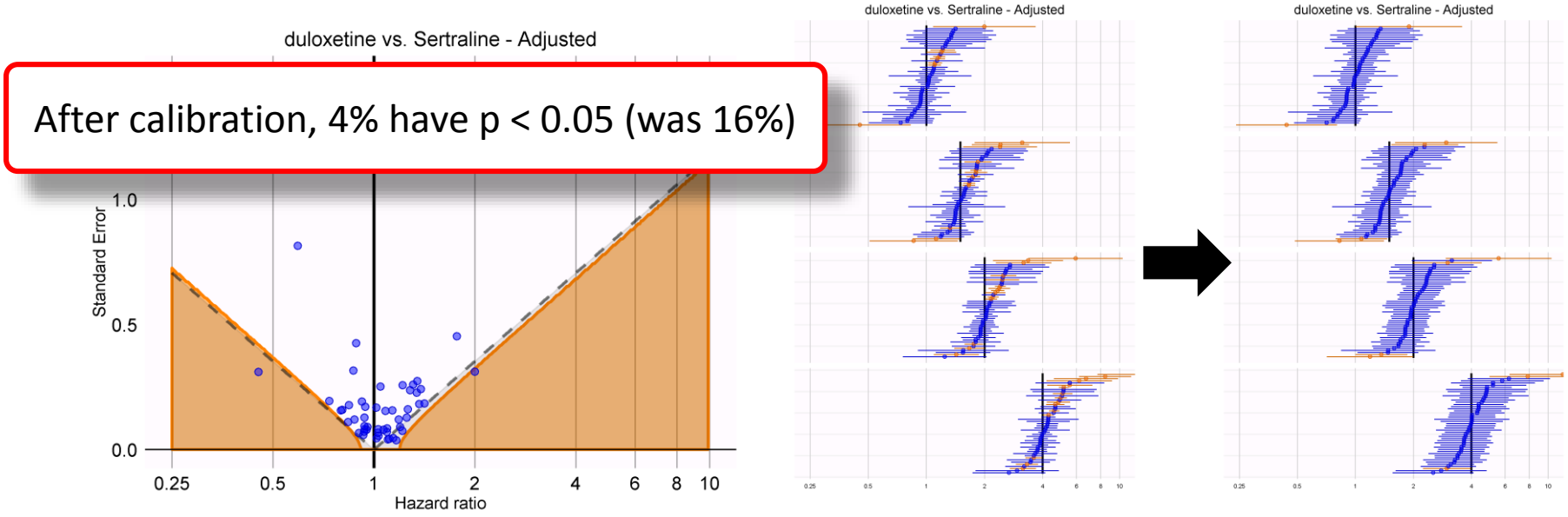




Reproducible research

2. Unmeasured (residual) **confounding**

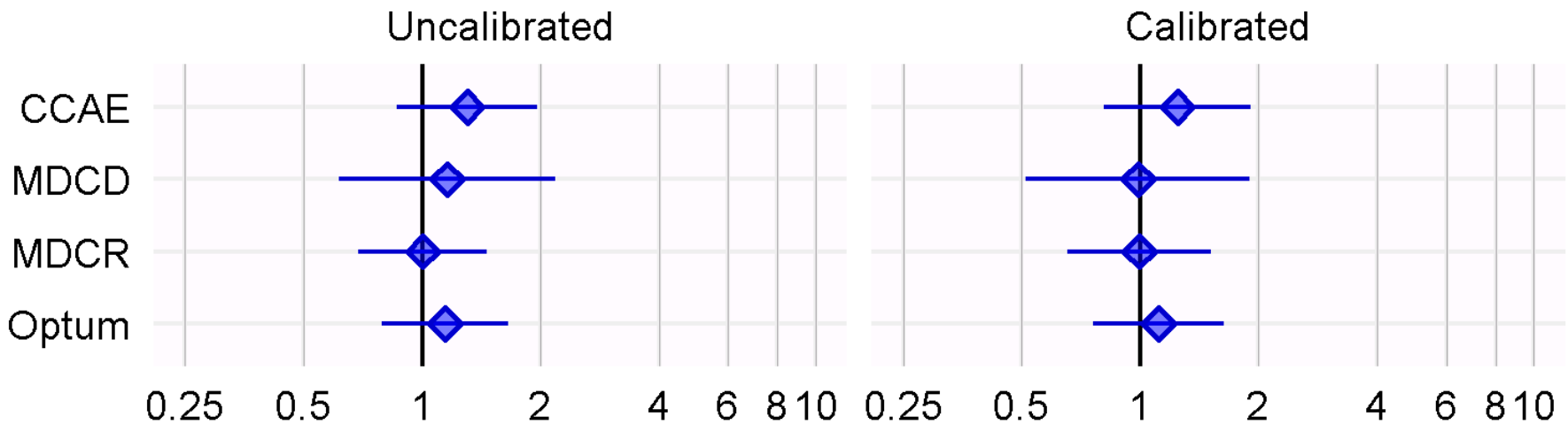
- **Confidence interval calibration**
  - **Adjust for all uncertainty, not just sampling**
- **Many negative controls**
  - **Unique to OHDSI** (PNAS in press)



Reproducible research

3. Multiple databases, locations, practice types

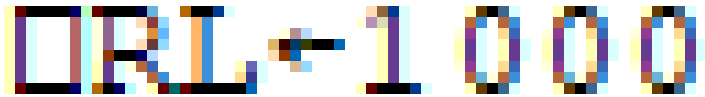
- **Exploit international OHDSI network**



Reproducible research

4. Open: publish all

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



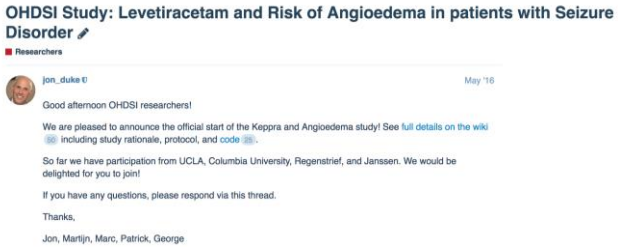
Generating evidence for US FDA

?

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



OHDSI Study: Levetiracetam and Risk of Angioedema in patients with Seizure Disorder

Researchers

jon_duke — May '16

Good afternoon OHDSI researchers!

We are pleased to announce the official start of the Keppra and Angioedema study! See full details on the wiki including study rationale, protocol, and code.

So far we have participation from UCLA, Columbia University, Regenstrief, and Janssen. We would be delighted for you to join!

If you have any questions, please respond via this thread.

Thanks,

Jon, Martijn, Marc, Patrick, George

- 50 viewed protocol, 25 viewed the code, and 7 sites ran the code on 10 databases (5 claims / 5 EHR), 59,367 levetiracetam patients matched with 74,550 phenytoin patients

Generating evidence for US FDA

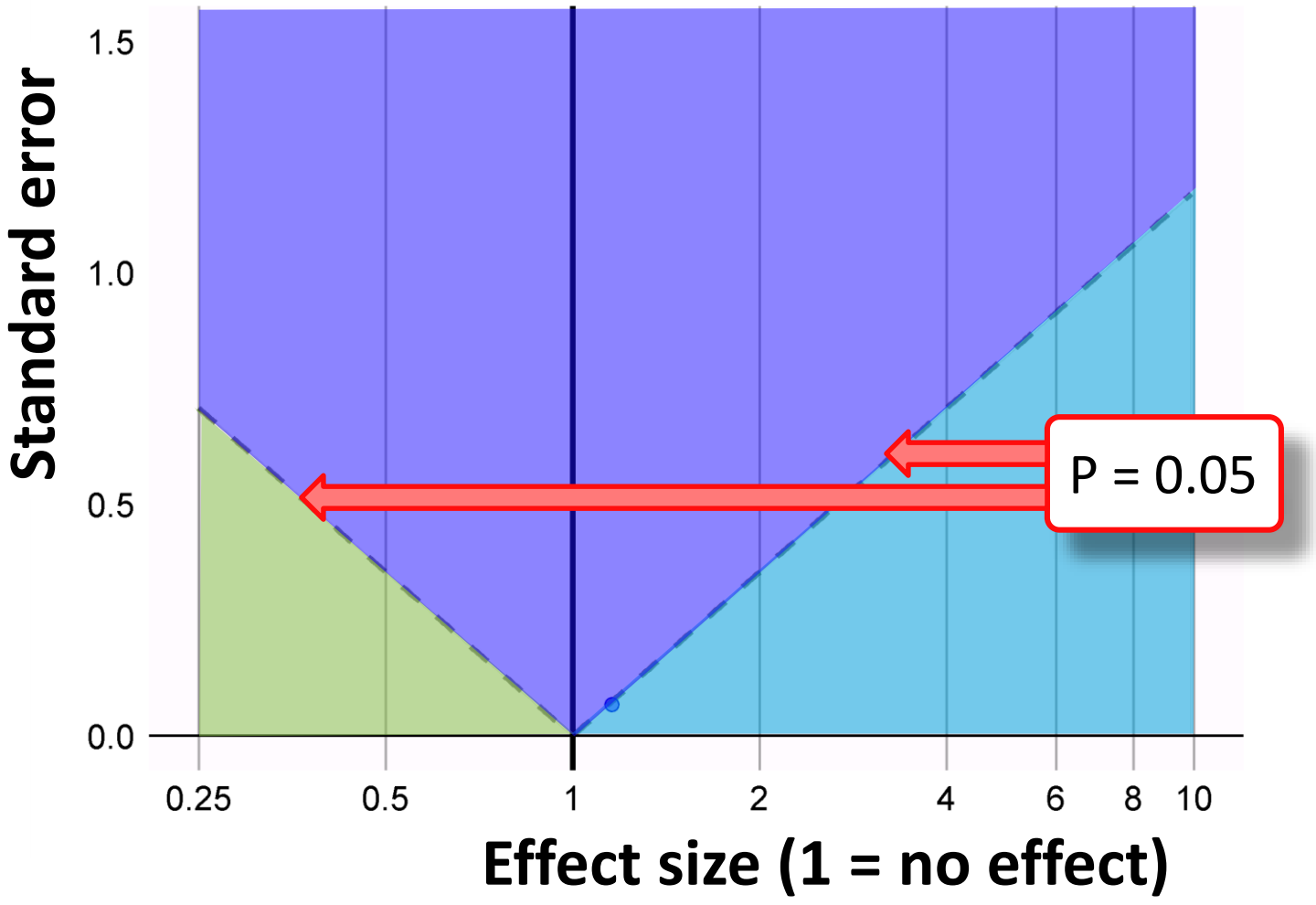
No evidence of increased angioedema risk with levetiracetam use compared with phenytoin use

"The study is focused, appears well designed, and provides new insight that should be of interest to clinicians and regulators... This is an important contribution to improved pharmacovigilance."

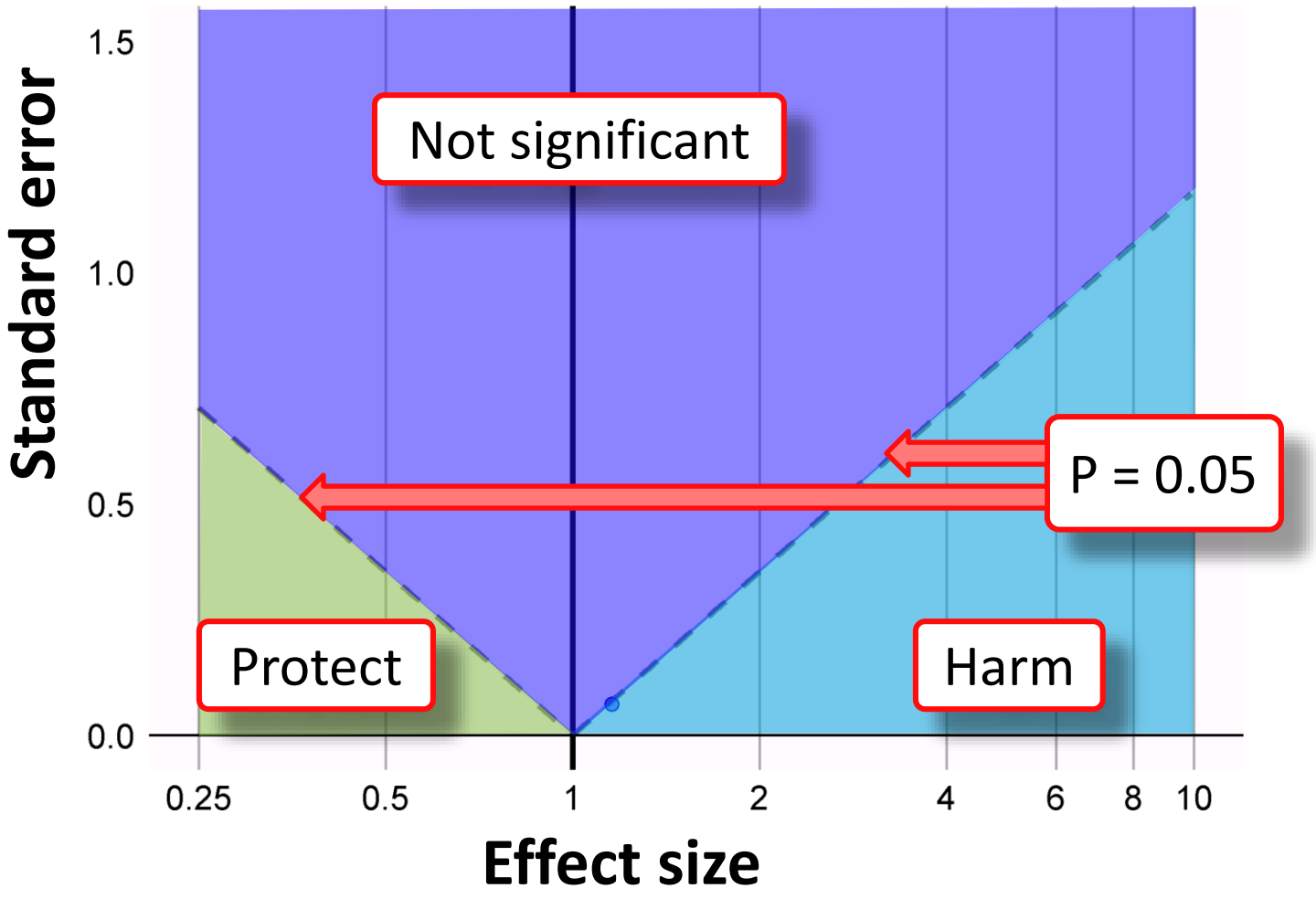
Add word to title, move diagram from supplement to body

How can we improve the literature

Literature

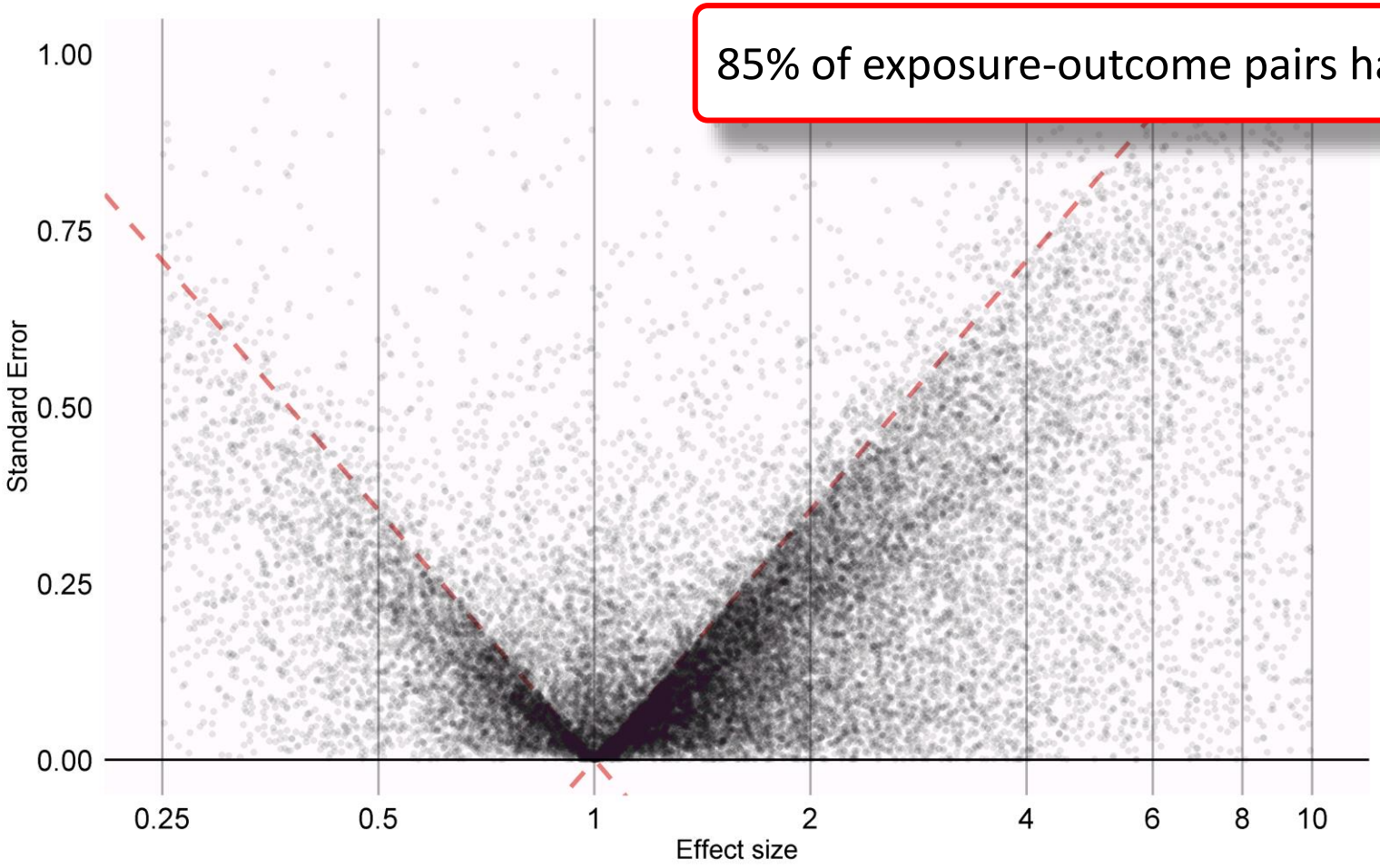

Standard error
1.5
1.0
0.5
0.0
0.25
0.5
1
2
4
6
8
10
Effect size (1 = no effect)
P = 0.05

Literature


Standard error
1.5
1.0
0.5
0.0
Not significant
P = 0.05
Protect
Harm
0.25
0.5
1
2
4
6
8
10
Effect size

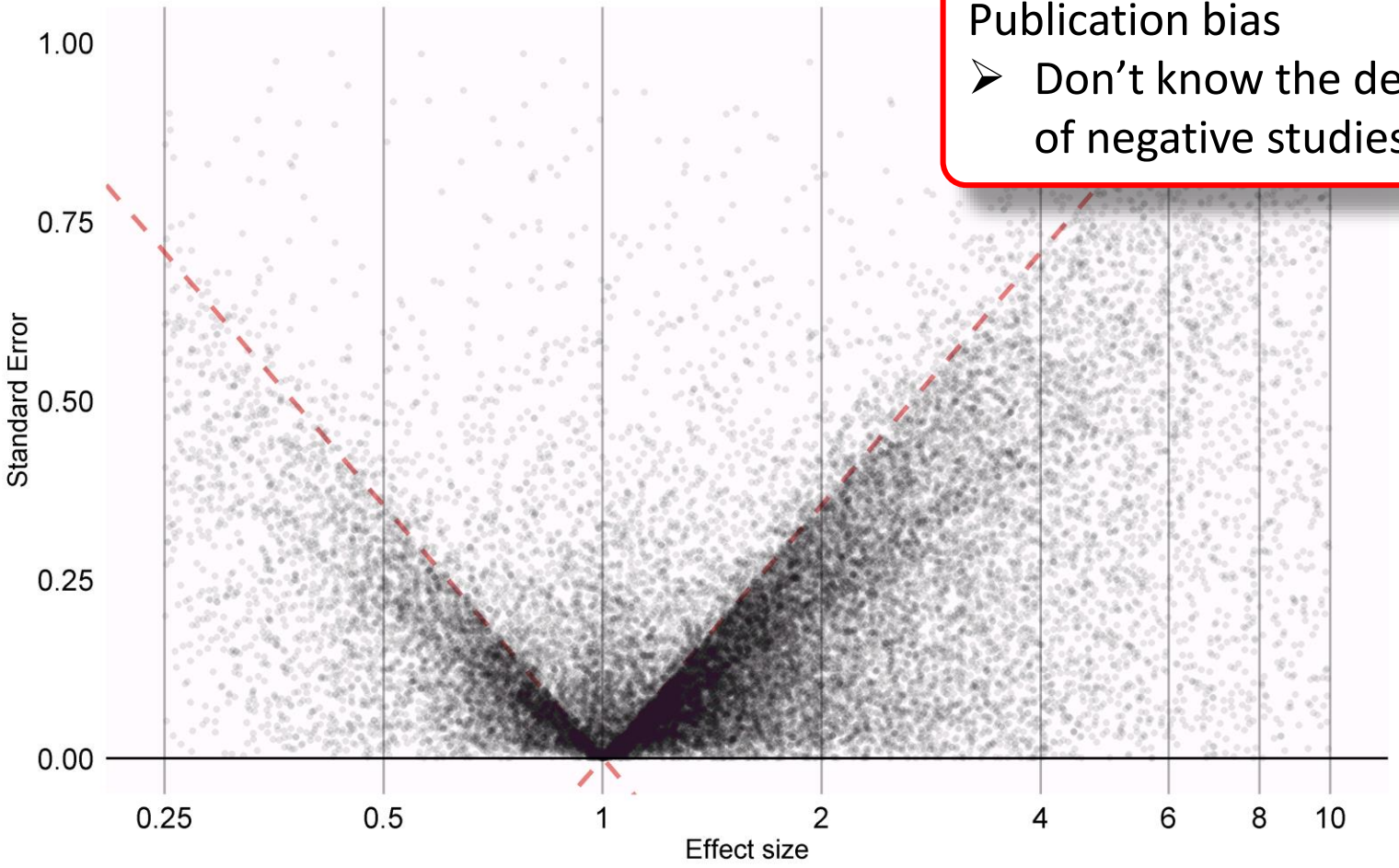
Observational research results in literature

85% of exposure-outcome pairs have p < 0.05



29,982 estimates
11,758 papers

Observational research results in literature

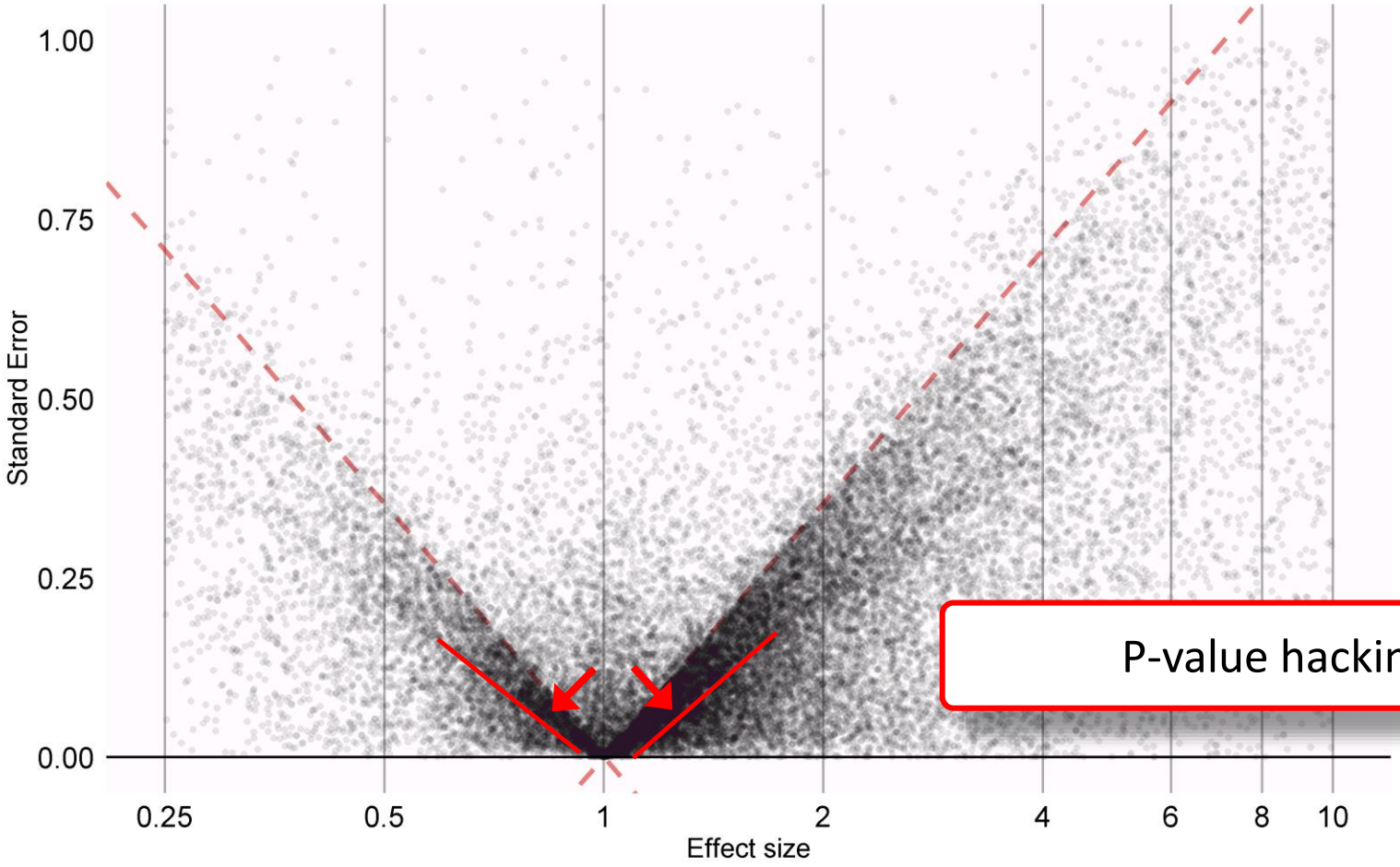


Publication bias

- Don't know the denominator of negative studies.

29,982 estimates
11,758 papers

Observational research results in literature



P-value hacking

29,982 estimates
11,758 papers

Observational research results in literature

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

Look at many outcomes at once

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

Many treatments at once

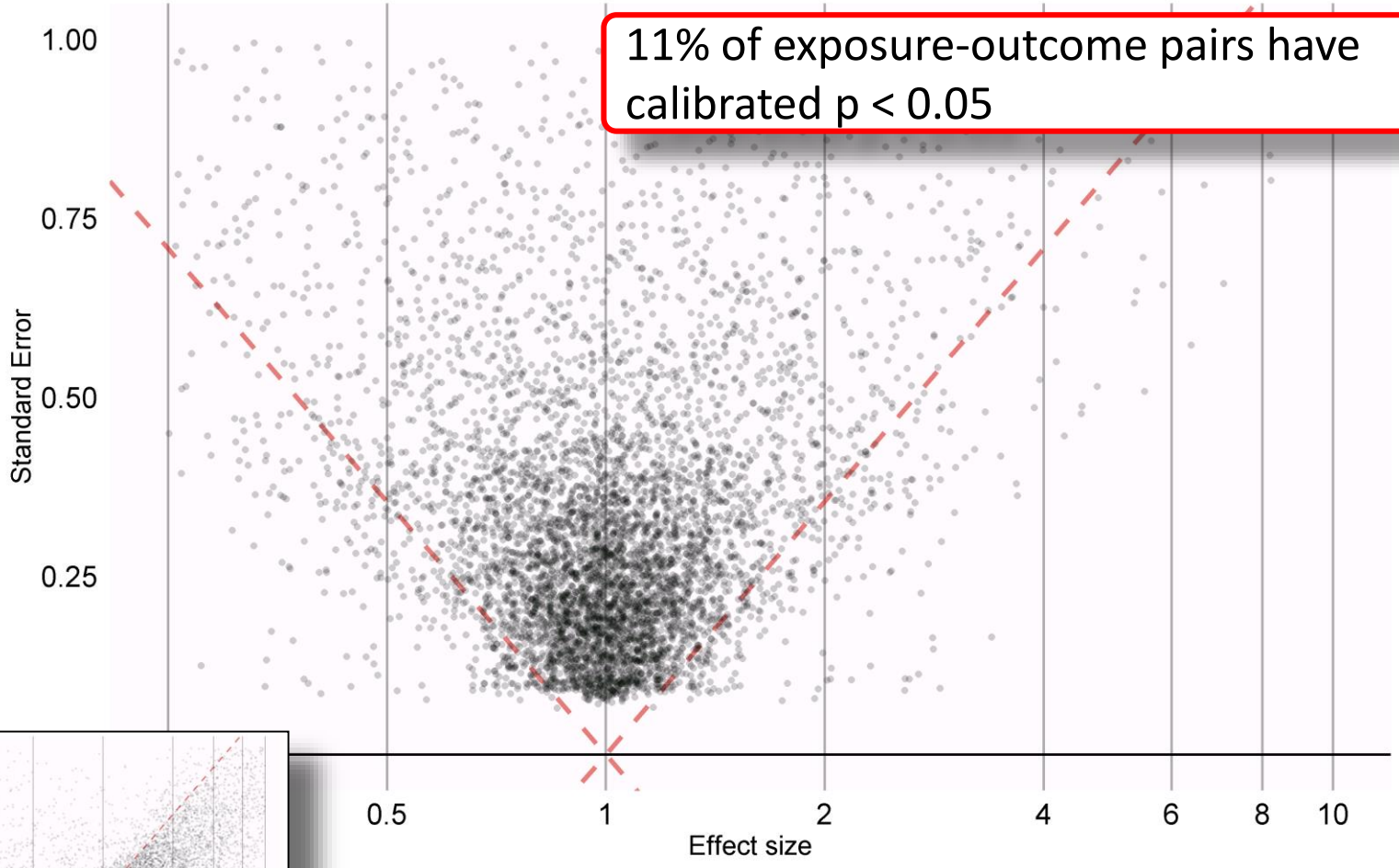
Type	Class	Treatment
Drug	Atypical	Bupropion
Drug	Atypical	Mirtazapine
Procedure	ECT	Electroconvulsive therapy
Procedure	Psychotherapy	Psychotherapy
Drug	SARI	Trazodone
Drug	SNRI	Desvenlafaxine
Drug	SNRI	duloxetine
Drug	SNRI	venlafaxine
Drug	SSRI	Citalopram
Drug	SSRI	Escitalopram
Drug	SSRI	Fluoxetine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

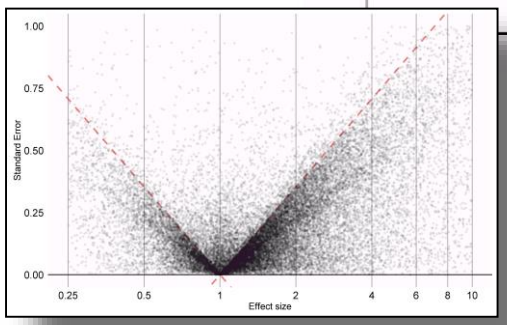
Large-scale estimation for depression

- 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**

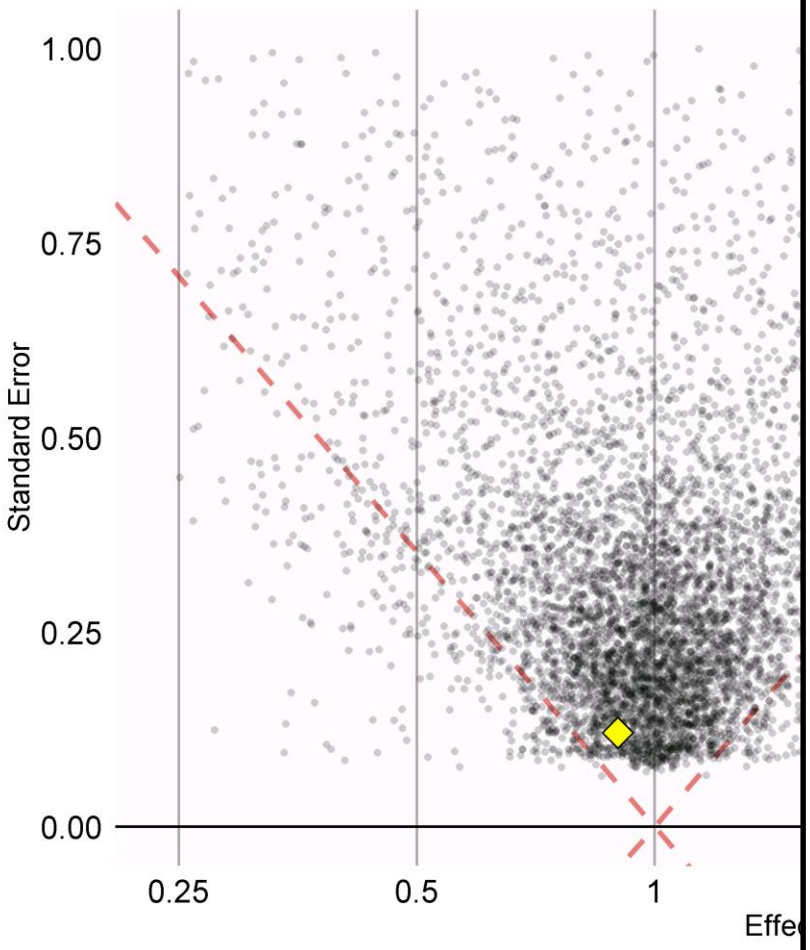
Estimates are in line with expectations

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





Example

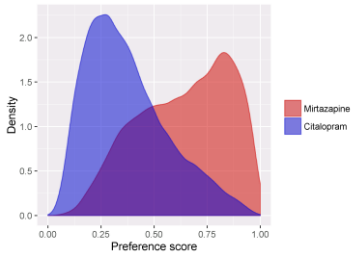


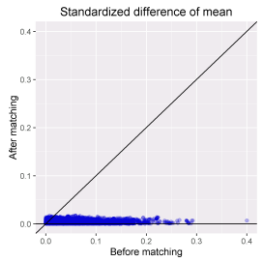
Mirtazapine vs. Citalopram

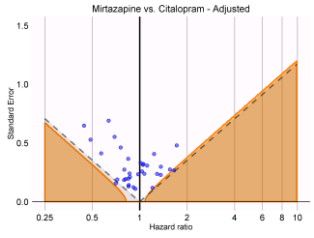
Constipation

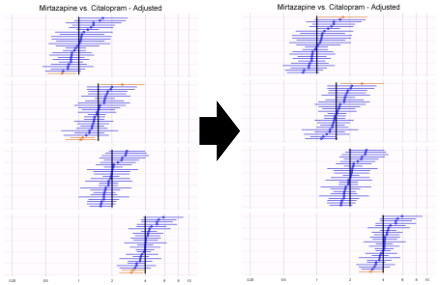
Database: Truven MDCD

Calibrated HR = 0.90 (0.70 – 1.12)

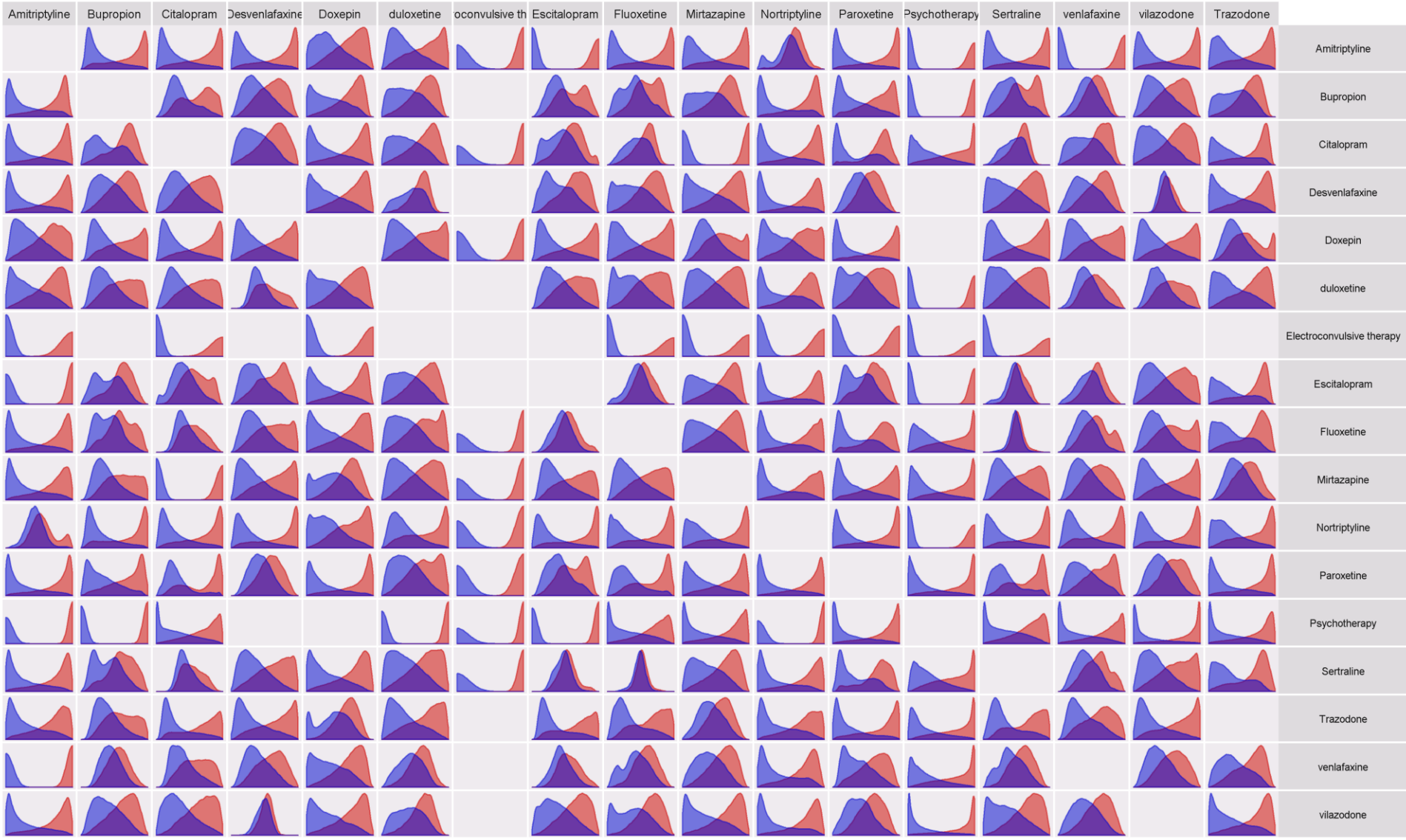








Propensity models for all comparisons (Truven CCAE, one outcome)



Large-scale estimation for depression

- How do we use it? Troll for effects?
- Professor what should I study this year?
  - Simple, go to Pubmed and find the smallest p-values in the literature; surely those must be the most significant things to study
- Which is safer?
  - Seizure in 0.0000000001 to 0.0000000002 (p=0.00001)
  - Seizure in 0 to 0.2 (p=.45)
- Large-scale studies become the literature
  - Come with hypothesis and ask a question

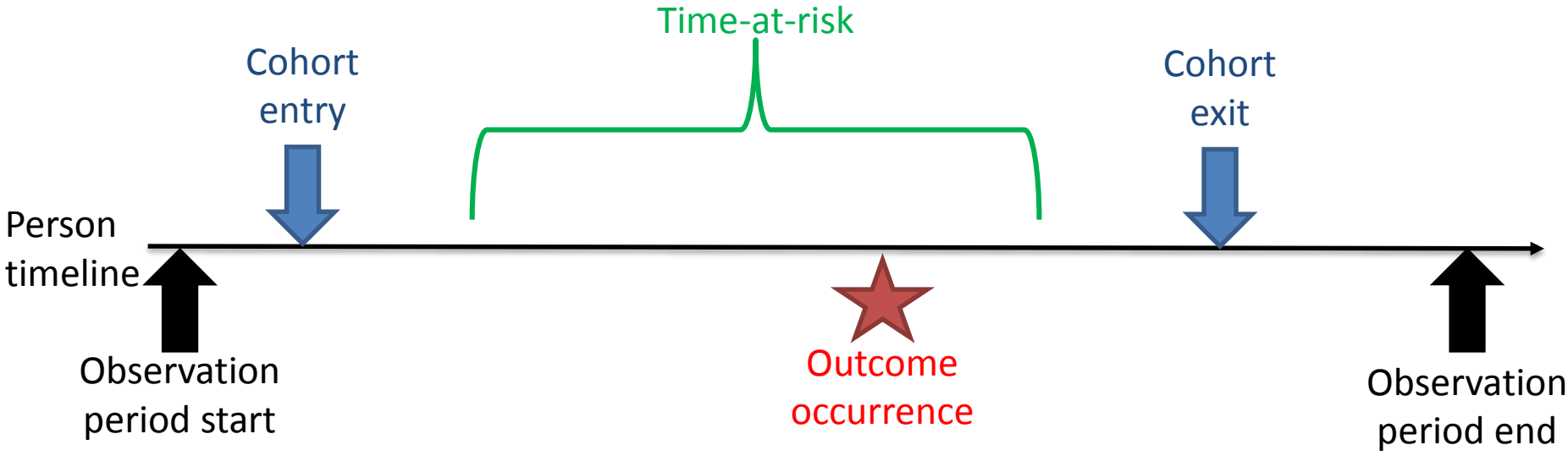
Large-scale estimation for depression

- Not “data-dredging”!
  - Data-dredging is not about what you do but about what you *throw out*
    - This can’t be done for literature
- One-off studies
  - Wouldn’t it be best to optimize each study?
    - Never get 10 or 100 parameters right
  - Still good to see the distribution
- At the very least, publish every last parameter so it can be reproduced

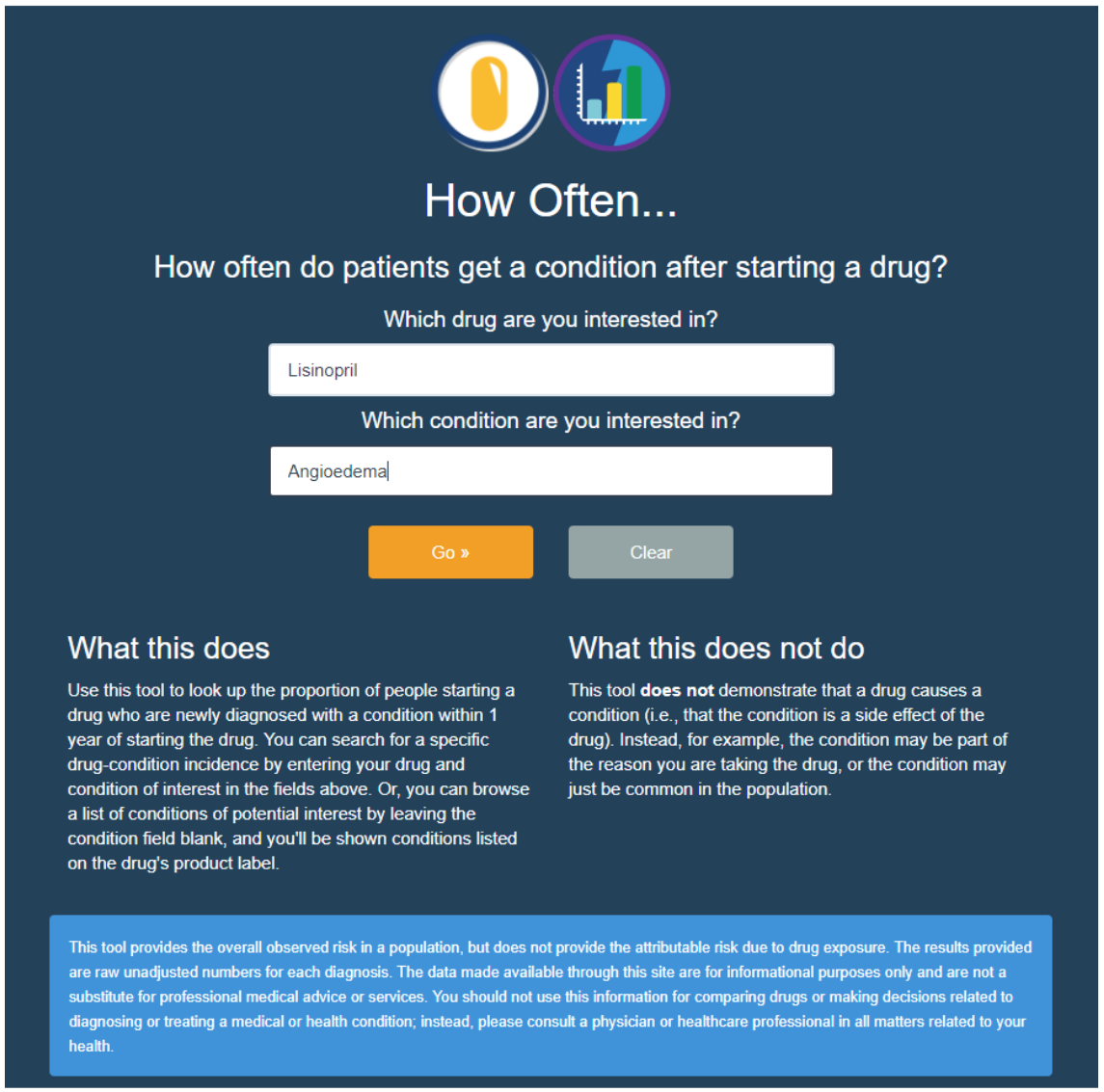
How often

How often do side effects occur?

- New incidence of any condition for any drug on the world market
  - Show range of answers for disparate databases
- Absolute risk (vs. attributable risk)
  - Not know if it is causal or not: MI with statin
- More complicated than it looks
  - Standard framework for reporting incidence



howoften.org



Summary

- Current observational research is suspect
- Large-scale observational research appears to be possible and more reliable than the current approach
- Need to extend to other areas
- Further research on reproducibility