OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Lessons from the Observational Medical Outcomes Partnership: Opportunities for Exploring Healthcare Databases to Study the Effects of Medical Products

Patrick Ryan
on behalf of the OMOP research team
29 August 2012

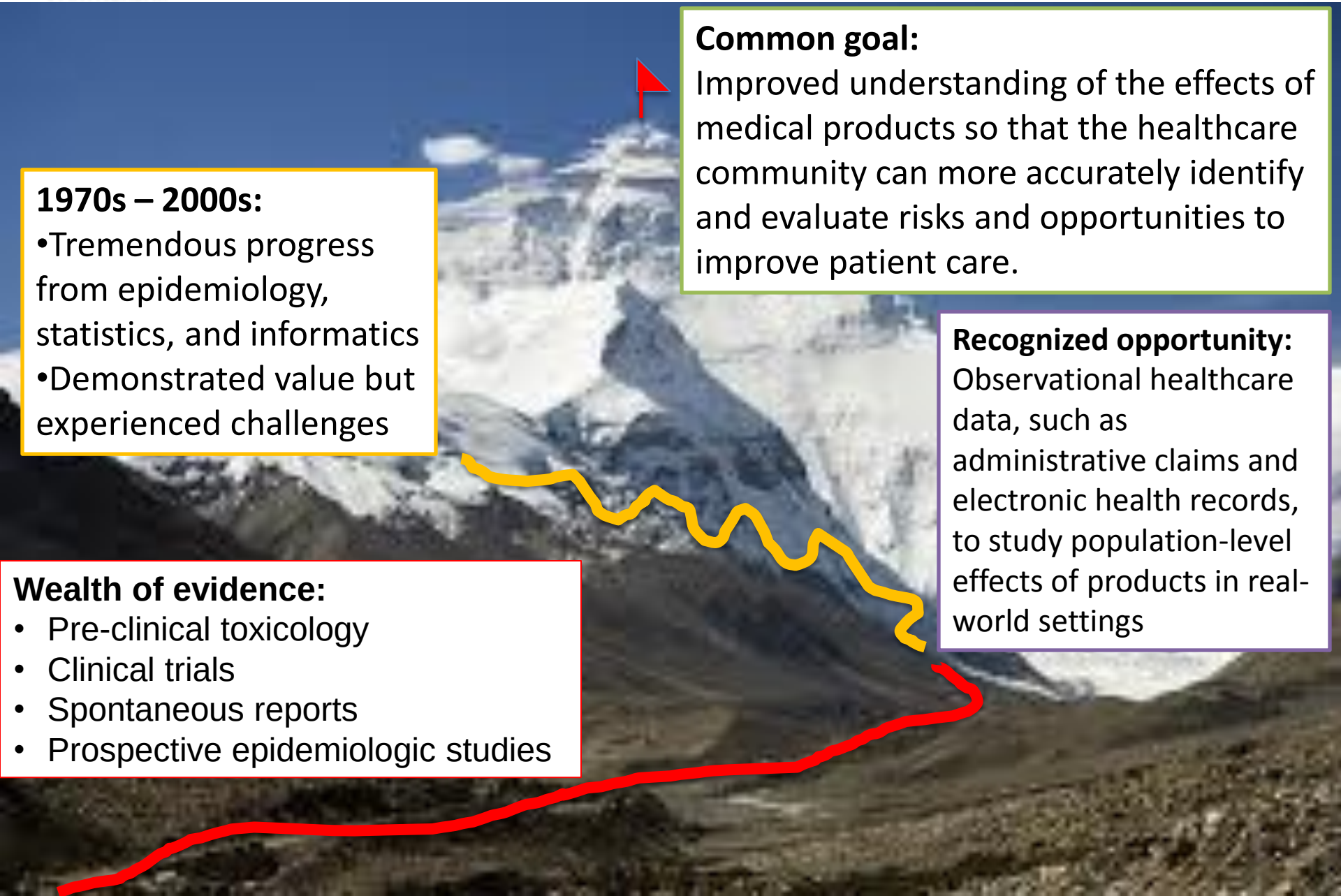
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Observational Medical Outcomes Partnership

- ***Public-Private Research Partnership established to inform the appropriate use of observational healthcare databases for studying the effects of medical products:***
  - Conducting methodological research to empirically evaluate the performance of various analytical methods on their ability to identify true associations and avoid false findings
  - Developing tools and capabilities for transforming, characterizing, and analyzing disparate data sources across the health care delivery spectrum
  - Establishing a shared resource so that the broader research community can collaboratively advance the science

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

A shared journey to learning about medical products



Common goal:
Improved understanding of the effects of medical products so that the healthcare community can more accurately identify and evaluate risks and opportunities to improve patient care.

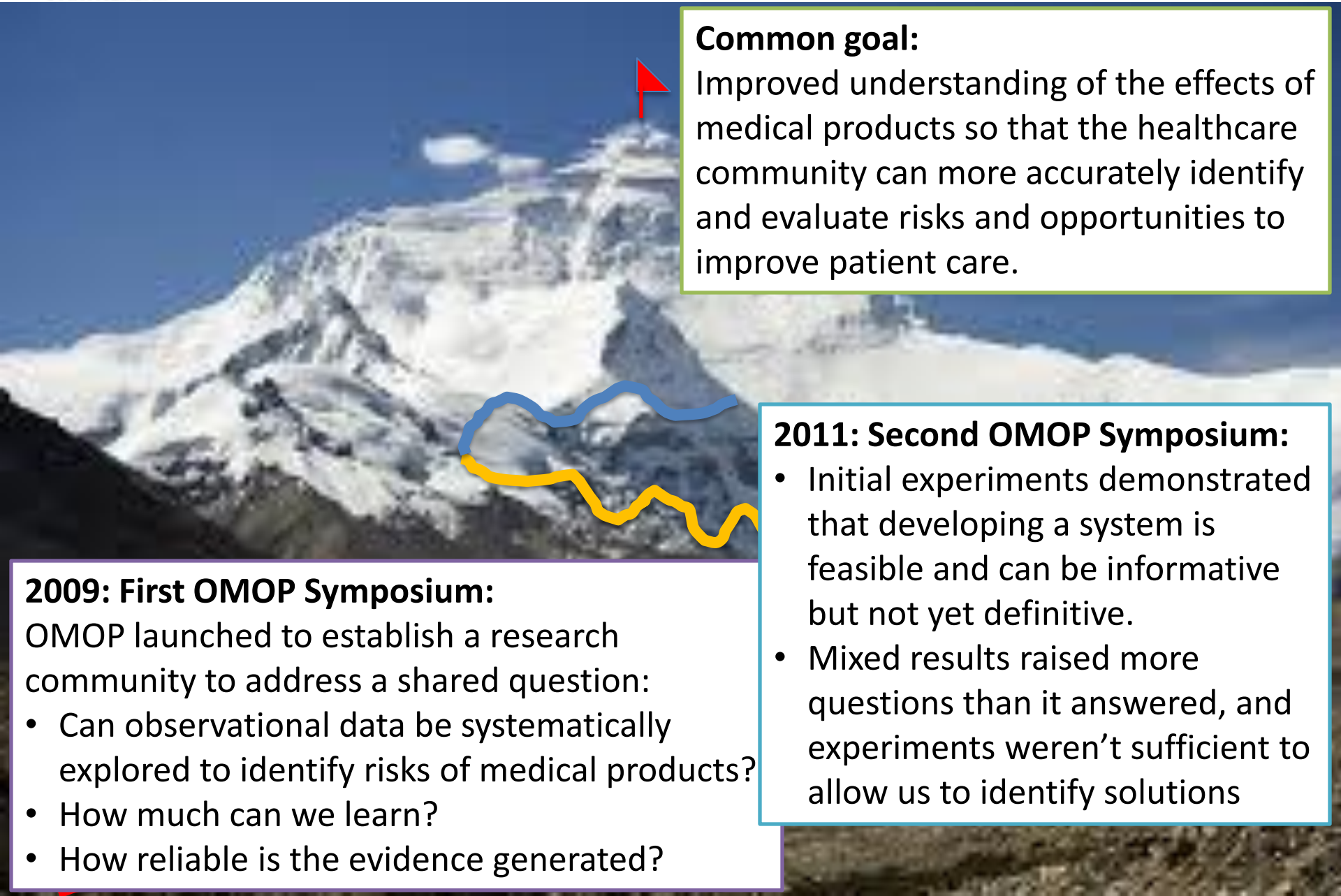
1970s – 2000s:
- Tremendous progress from epidemiology, statistics, and informatics
- Demonstrated value but experienced challenges

Recognized opportunity:
Observational healthcare data, such as administrative claims and electronic health records, to study population-level effects of products in real-world settings

Wealth of evidence:
- Pre-clinical toxicology
- Clinical trials
- Spontaneous reports
- Prospective epidemiologic studies

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

A shared journey to learning about medical products



Common goal:
Improved understanding of the effects of medical products so that the healthcare community can more accurately identify and evaluate risks and opportunities to improve patient care.

2009: First OMOP Symposium:
OMOP launched to establish a research community to address a shared question:

- Can observational data be systematically explored to identify risks of medical products?
- How much can we learn?
- How reliable is the evidence generated?

2011: Second OMOP Symposium:

- Initial experiments demonstrated that developing a system is feasible and can be informative but not yet definitive.
- Mixed results raised more questions than it answered, and experiments weren't sufficient to allow us to identify solutions

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

A shared journey to learning about medical products

We still have a long way to go, and the future directions are not certain, but it is clear we can only continue to make progress if we work together as a research community toward our common goals

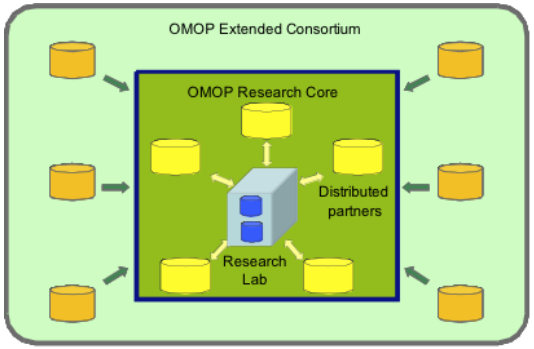
Common goal:
Improved understanding of the effects of medical products so that the healthcare community can more accurately identify and evaluate risks and opportunities to improve patient care.

2012: Third OMOP Symposium:

- Expanded experiments have yielded more promising results
- Started to develop practical insights for how to build a risk identification system and how to interpret individual study results

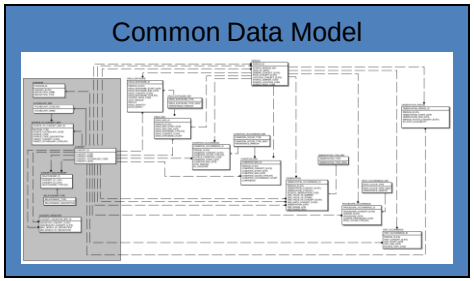
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

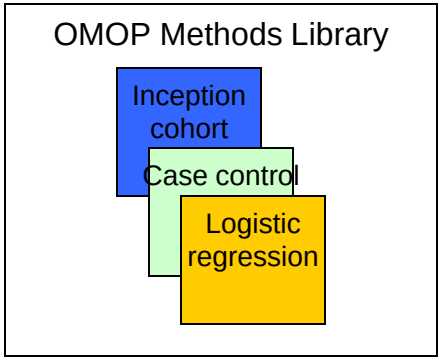
OMOP 2010/2011 Research Experiment



- 10 data sources
- Claims and EHRs
- 200M+ lives

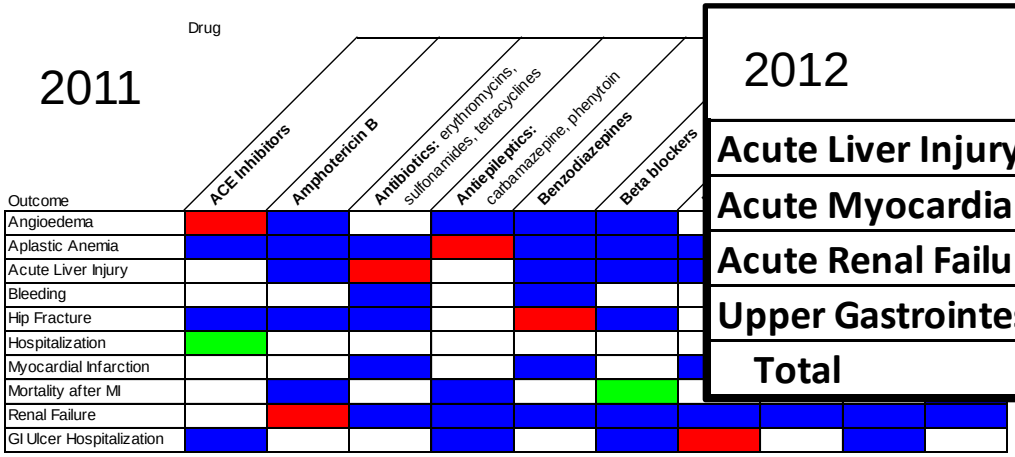
- Open-source
- Standards-based





- 14 methods
- Epidemiology designs
- Statistical approaches adapted for longitudinal data

2011



2012	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

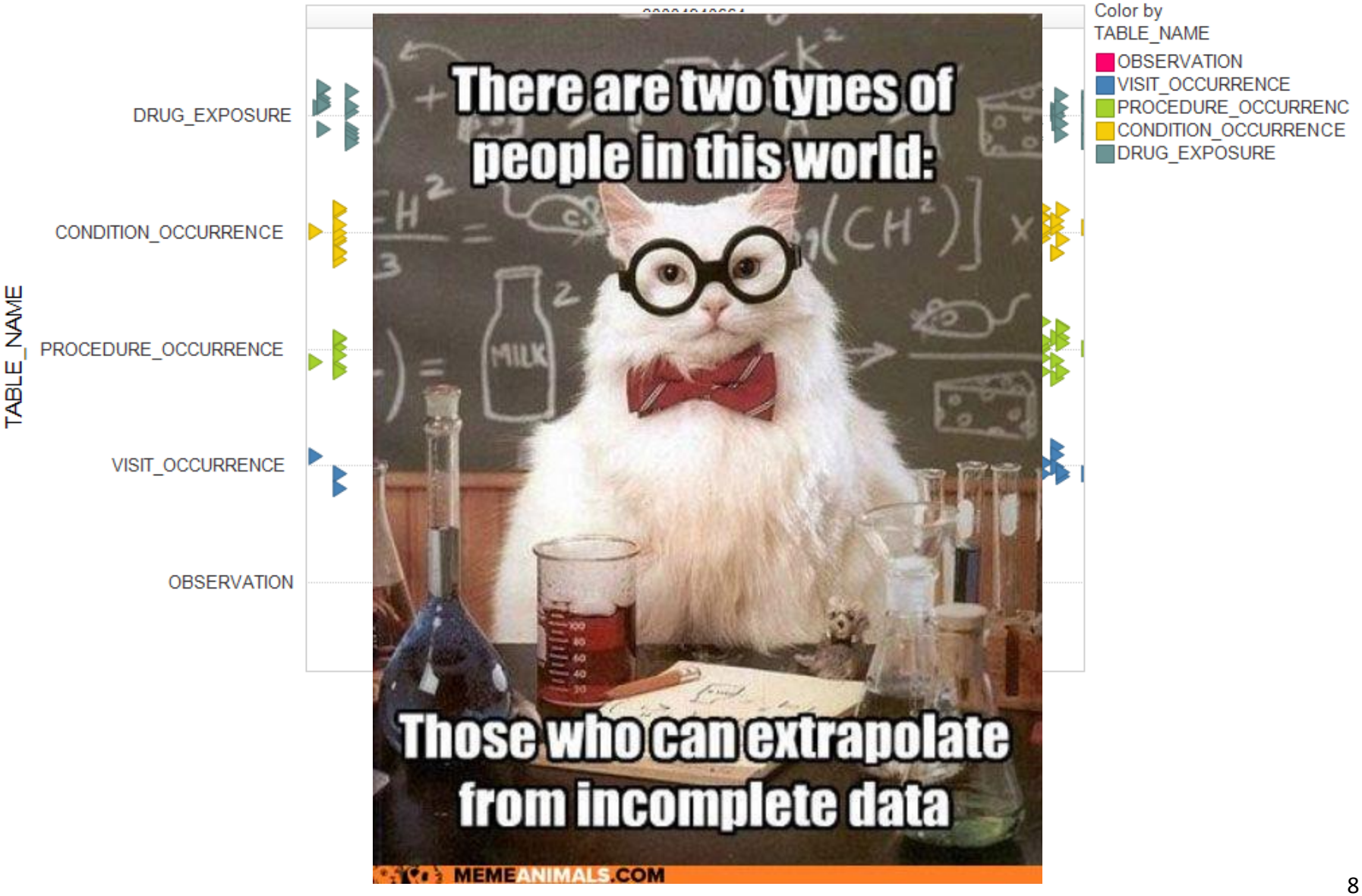
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Within OMOP, enough research areas for all backgrounds and interests

- Methods development
  - Estimate average treatment effects
  - Predict patient outcomes
- Methods implementation
  - Transform conceptual ideas into scalable computationally efficient applications
  - Contribute to open-source solutions within community of users to characterize, visualize, and analyze longitudinal observational data
- Methods evaluation
  - Measure and compare performance of different algorithms across an array of different databases, outcomes, exposures
  - Design and implement simulations to model real-world data and inject patterns of interest
  - Develop and apply metrics for empirical assessment of methods operating characteristics
- System optimization

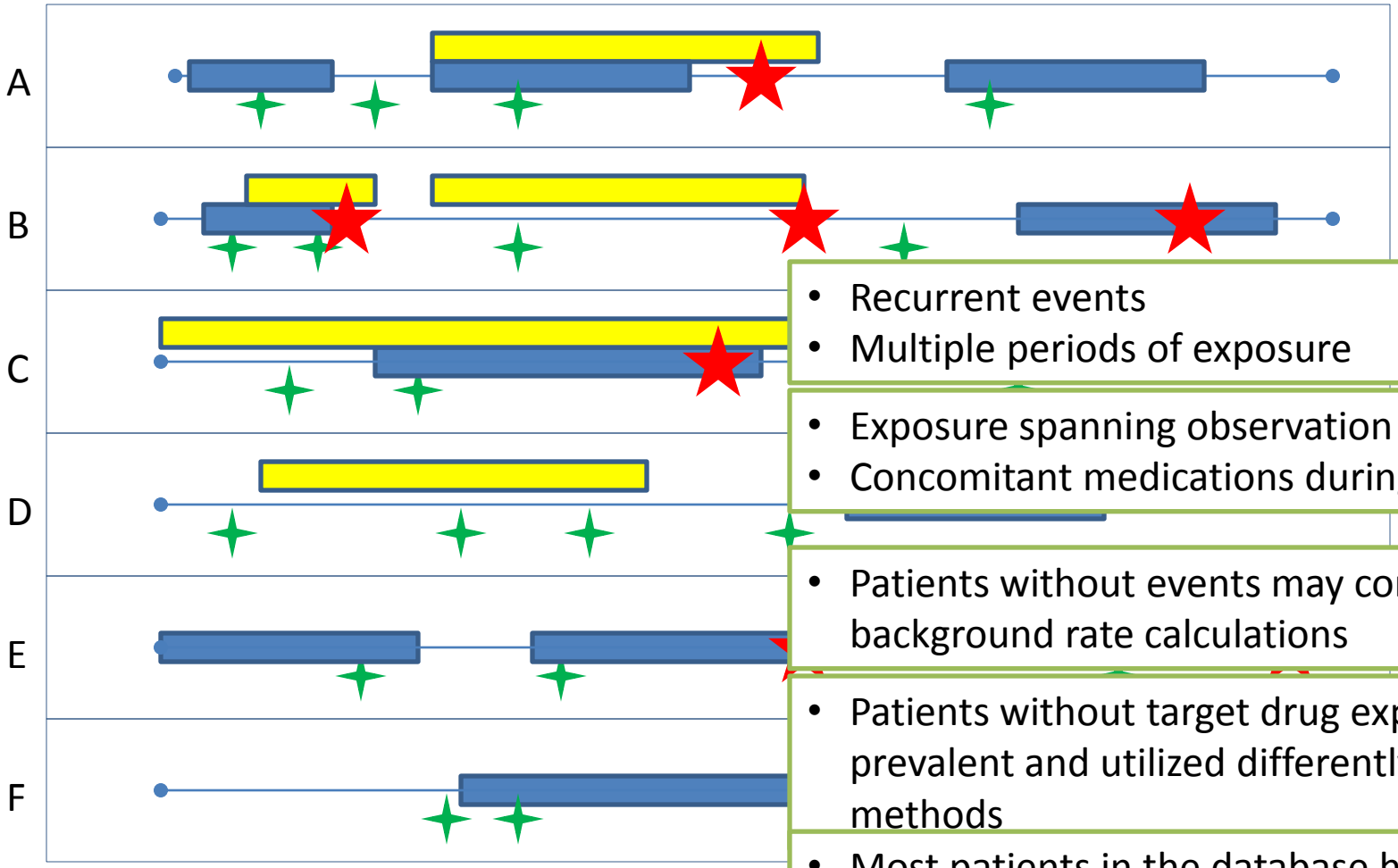
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

A couple years in the life of a patient in an observational healthcare database



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Patient profiles in observational data



- Recurrent events
- Multiple periods of exposure

- Exposure spanning observation period
- Concomitant medications during events

- Patients without events may contribute to background rate calculations

- Patients without target drug exposure are prevalent and utilized differently across all methods

- Most patients in the database have neither the target drug nor the target outcome

Target condition

Other conditions

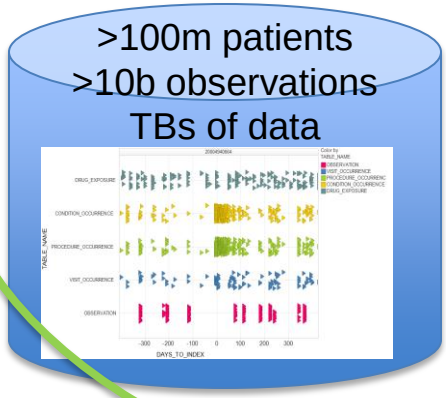
Other drugs

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Opportunities through the data analytics process

Required solution for public health

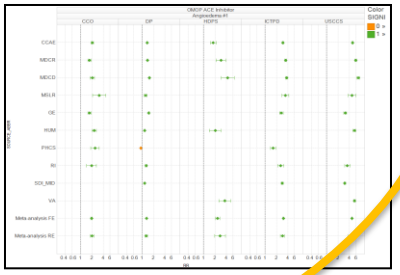
Raw observational data



Pre-processed analytical dataset

Outcome: Stroke	CHF	Hypertension	Age>=75	Diabetes	Prior stroke
1	1	0	0	0	1
0	1	1	0	0	0
0	1	1	1	0	1
1	1	1	0	1	0
0	0	1	0	0	0
1	1	1	1	0	0
0	0	0	1	1	0

Statistical summary



GIANT opportunity for the data scientist

Typical comfort zone for the statistician

Outstanding questions:
How can we efficiently learn from observational healthcare data?
How reliable is the evidence that we generate?

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Summary of methods tested in OMOP 2011/2012 experiment

Method	Abbreviation	Parameter combinations tested	Collaborator
Cohort	CM	126	OMOP Team
Case-control	CC	384	OMOP Team
Self-controlled case series	SCCS	560	OMOP Team
Observational screening	OS	54	UBC/ProSanos, GlaxoSmithKline
Temporal pattern discovery	ICTPD	42	Uppsala Monitoring Centre
Disproportionality analysis	DP	48	OMOP Team
Longitudinal Gamma Poisson Shrinker	LGPS	32	Erasmus MC

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Data used for new user cohort design to estimate average treatment effect

New user design

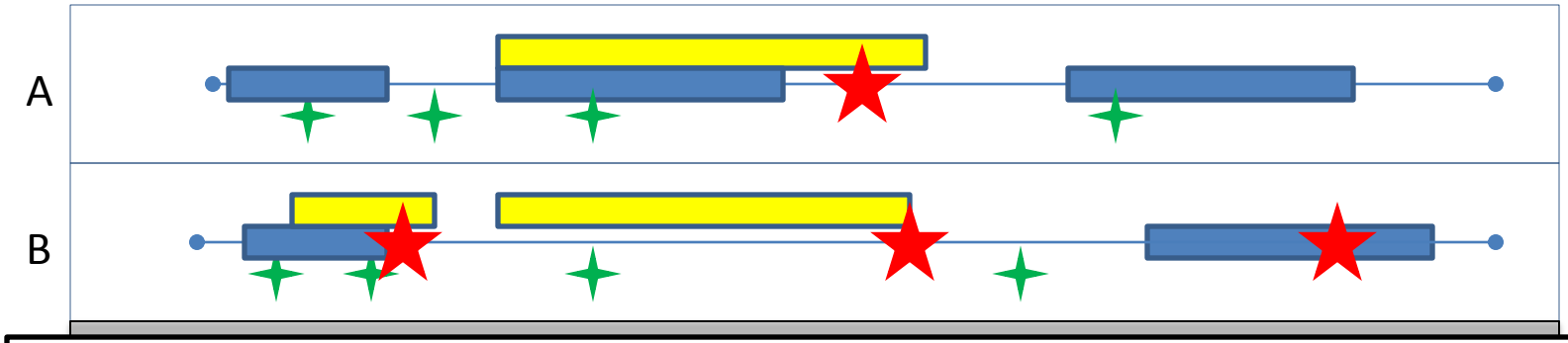
- Focus on comparing rates of events among patients exposed to target drug, relative to rates of events among patients in some referent comparator group
- Relative risk can be adjusted for baseline covariates through various strategies, including propensity score
- Define cohorts based on index exposure (first use after washout period)

- Define cohorts based on index exposure (first use after washout period)
- Observations prior to index may be used as covariates
- Observations on or after index, except for incident outcome, are not considered in analysis (e.g. no time-varying covariates)

Target drug
Target condition
Other conditions
Other drugs

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Data used for self-controlled case series to estimate average treatment effects



Univariate self-controlled case series :

- focus on time exposed/unexposed to target drug and occurrences of target condition (do not consider comorbidities or concomitant medications)

Patient A: 1 event in 3mo exposed; 0 events in 6mo unexposed
Patient B: 2 events in 4mo exposed; 1 event in 5mo unexposed

Odds ratio estimated by maximizing the likelihood

Select SCCS parameters:

- **Events to use**: first occurrence or all occurrence?
- **Surveillance window**: first 30d after drug start, length of exposure + 30d, all time post-exposure start? Is date of dispensing consider exposed or unexposed time?
- **Multivariate model** (MSCCS): condition on all other drugs?

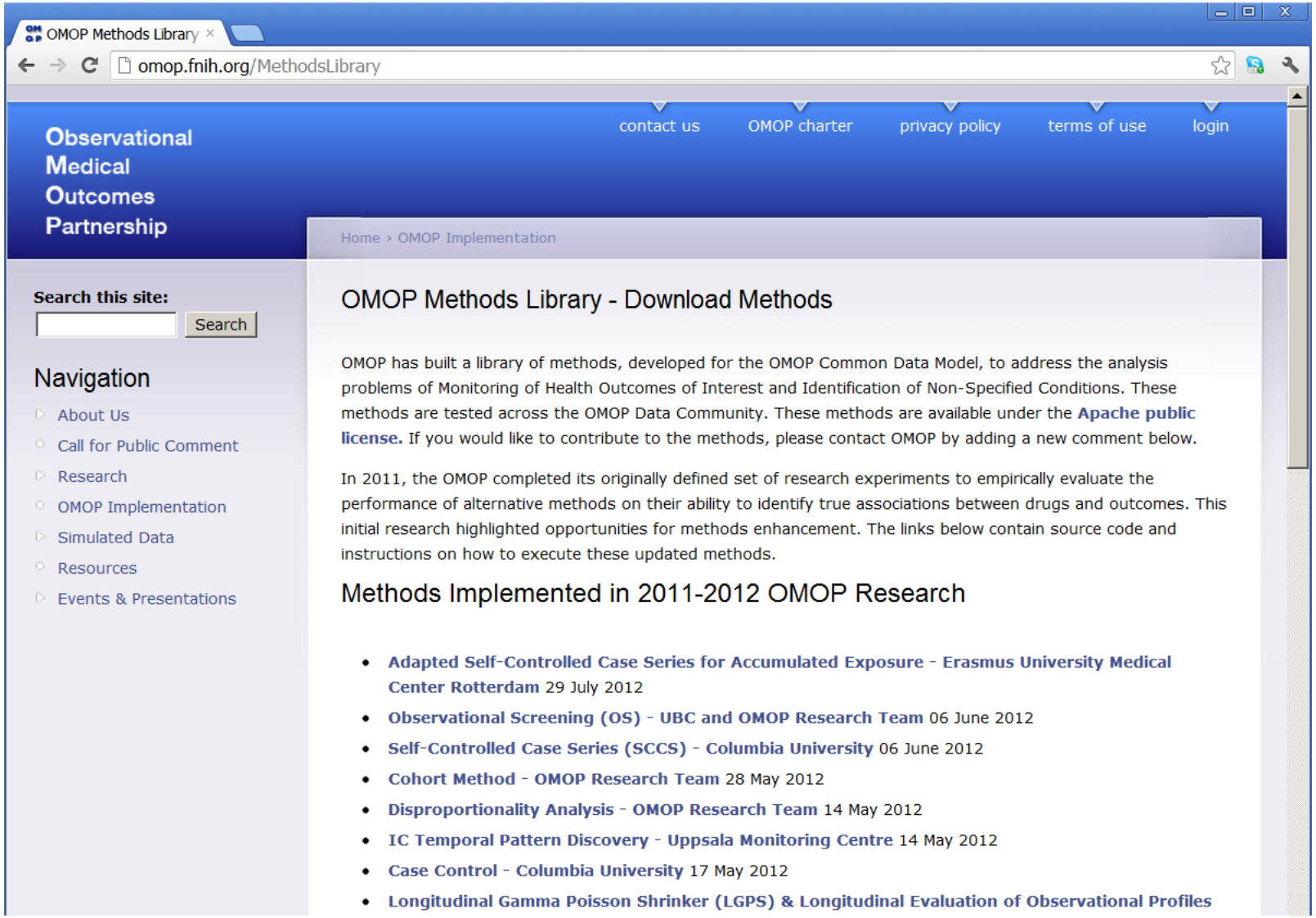
ition

itions

Other drugs

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

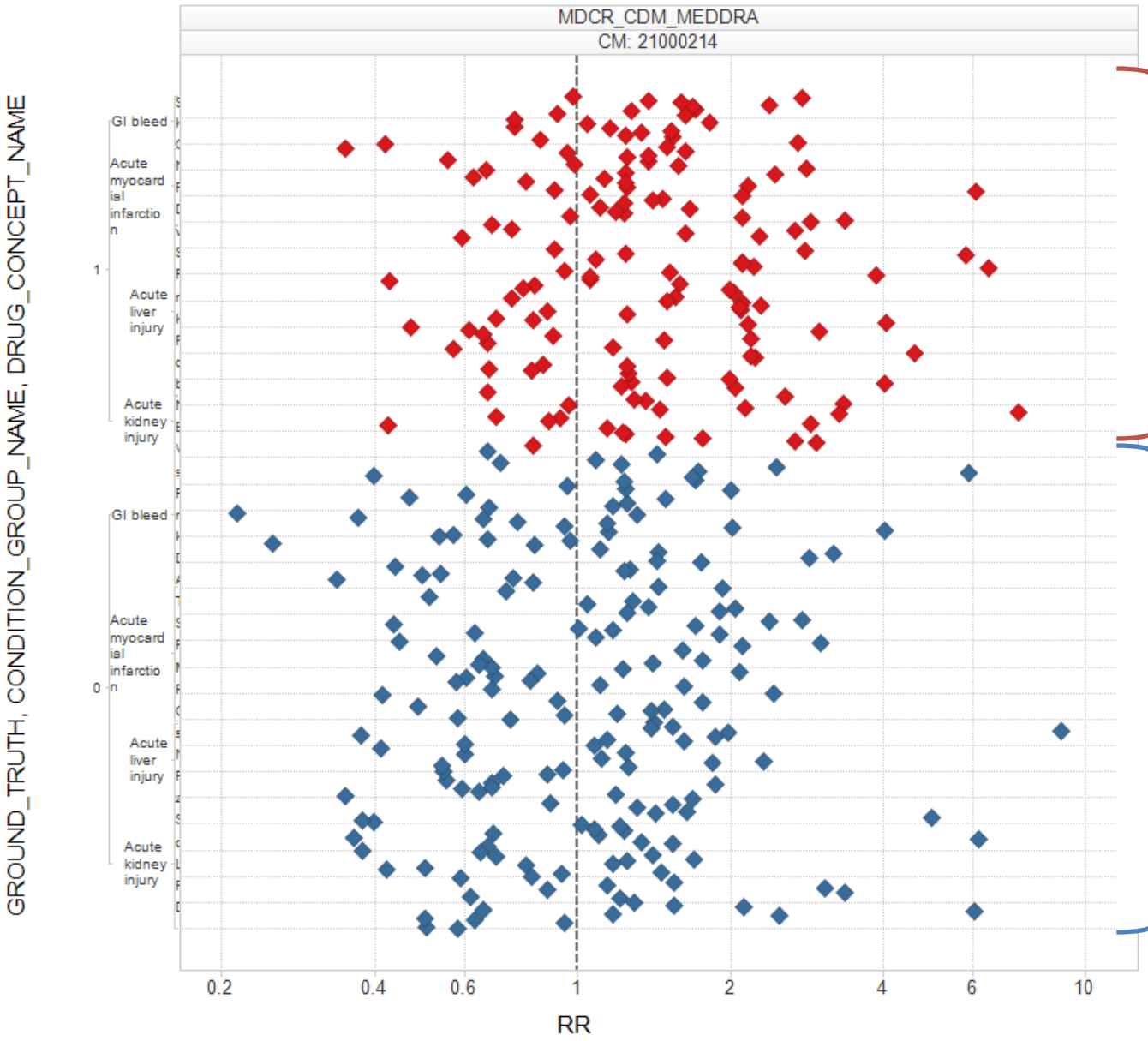
Open-source library of analytical methods for the research community to use, advance, and evaluate



http://omop.fnih.org/MethodsLibrary

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

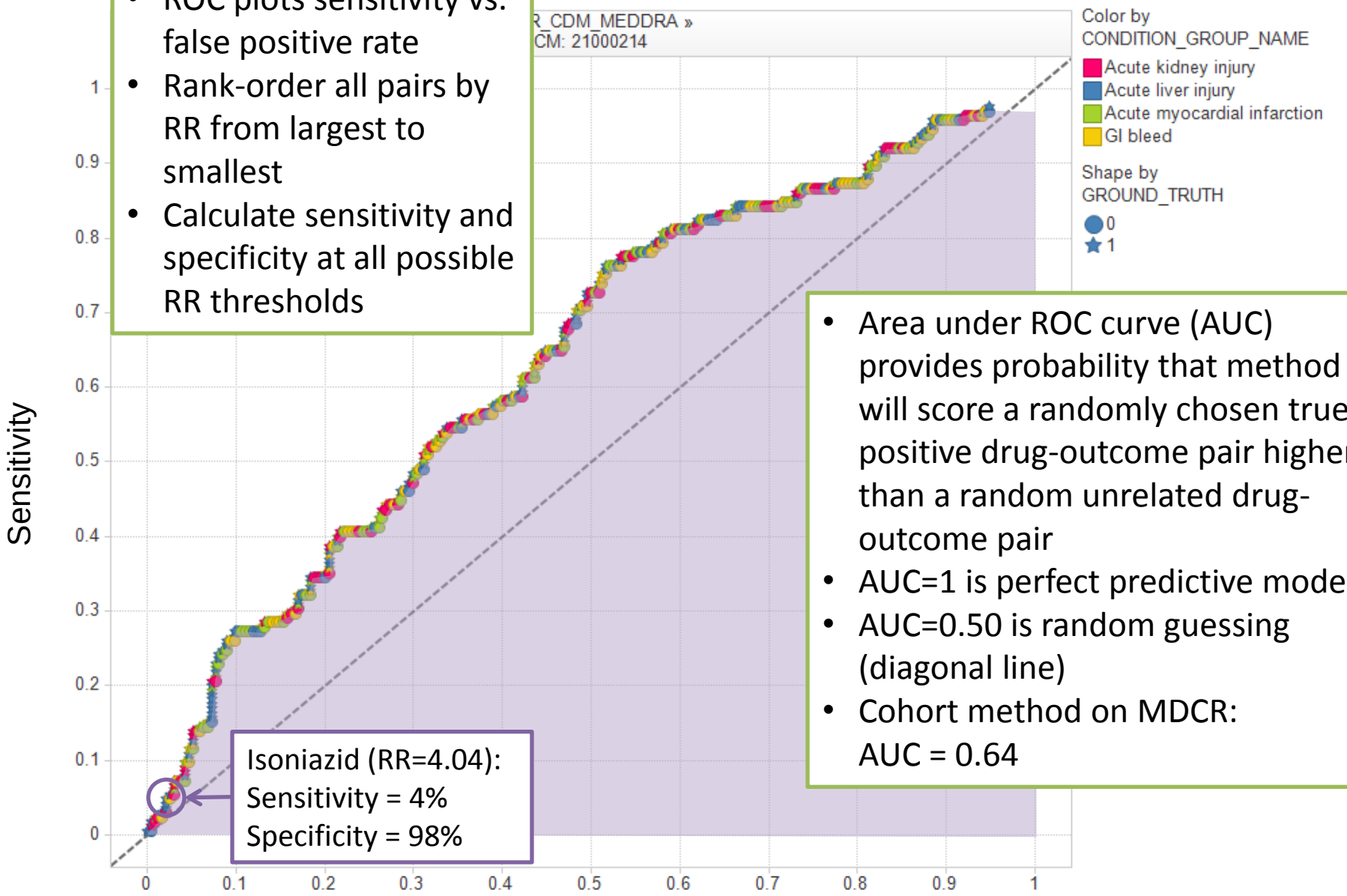
New user cohort design applied to all test cases



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNE

Receiver Operating Characteristic (ROC) curve

- ROC plots sensitivity vs. false positive rate
- Rank-order all pairs by RR from largest to smallest
- Calculate sensitivity and specificity at all possible RR thresholds



Isoniazid (RR=4.04):
Sensitivity = 4%
Specificity = 98%

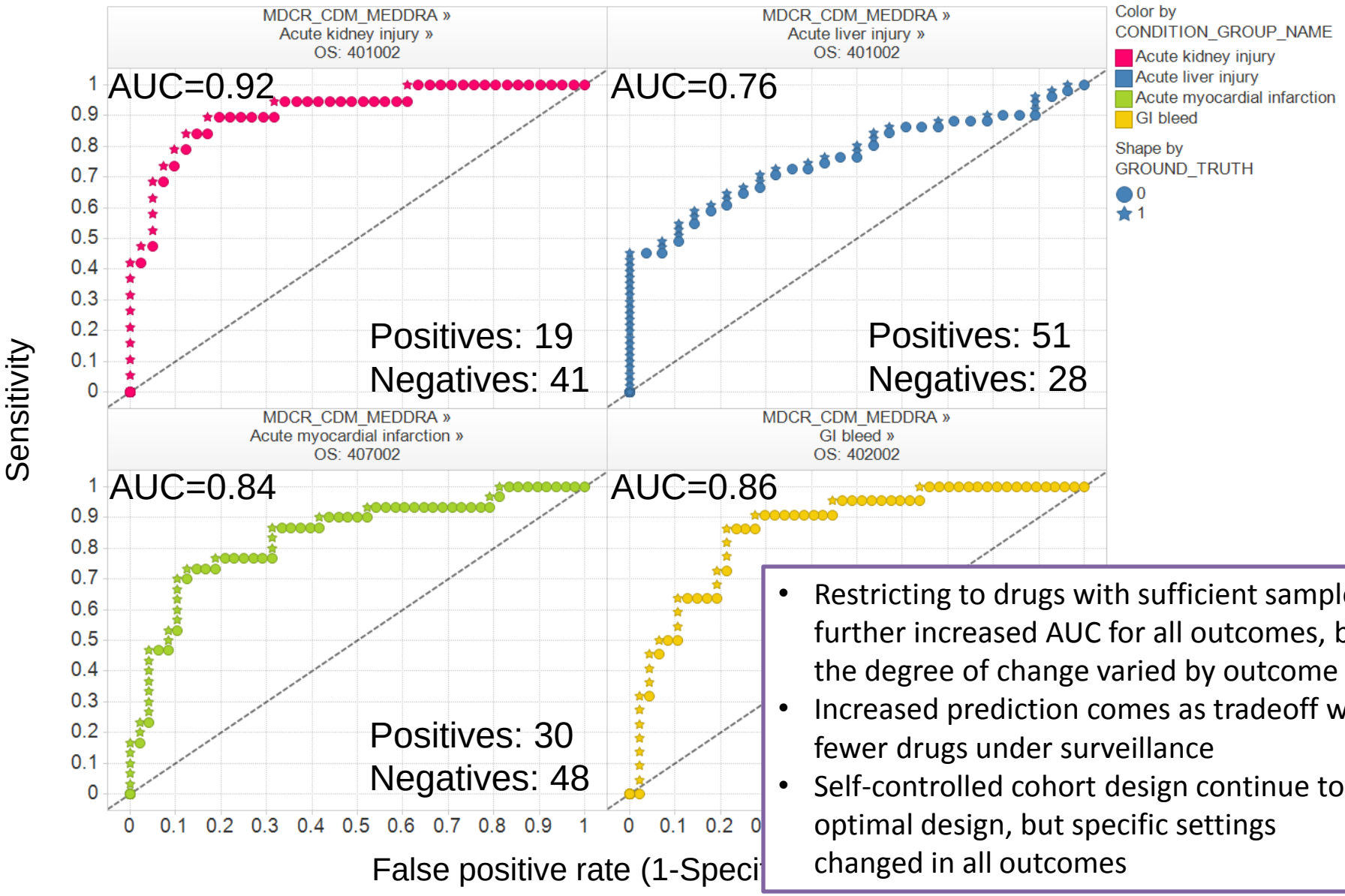
- Area under ROC curve (AUC) provides probability that method will score a randomly chosen true positive drug-outcome pair higher than a random unrelated drug-outcome pair
- AUC=1 is perfect predictive model
- AUC=0.50 is random guessing (diagonal line)
- Cohort method on MDCR: AUC = 0.64

Strategies to improve predictive accuracy

- Stratify results by outcome
- Tailor analysis to outcome
- Restrict to sufficient sample size
- Optimize analysis to the data source

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Performance after applying these strategies



- Restricting to drugs with sufficient sample further increased AUC for all outcomes, but the degree of change varied by outcome
- Increased prediction comes as tradeoff with fewer drugs under surveillance
- Self-controlled cohort design continue to be optimal design, but specific settings changed in all outcomes

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

To recap the improvements that could be achieved by following these ideas…

Before: One method applied to all test cases

If sensitivity = 50%:

Outcome	AUC	Threshold	Specificity
All	0.64	1.25	69%

After: Partitioning, tailoring, restriction

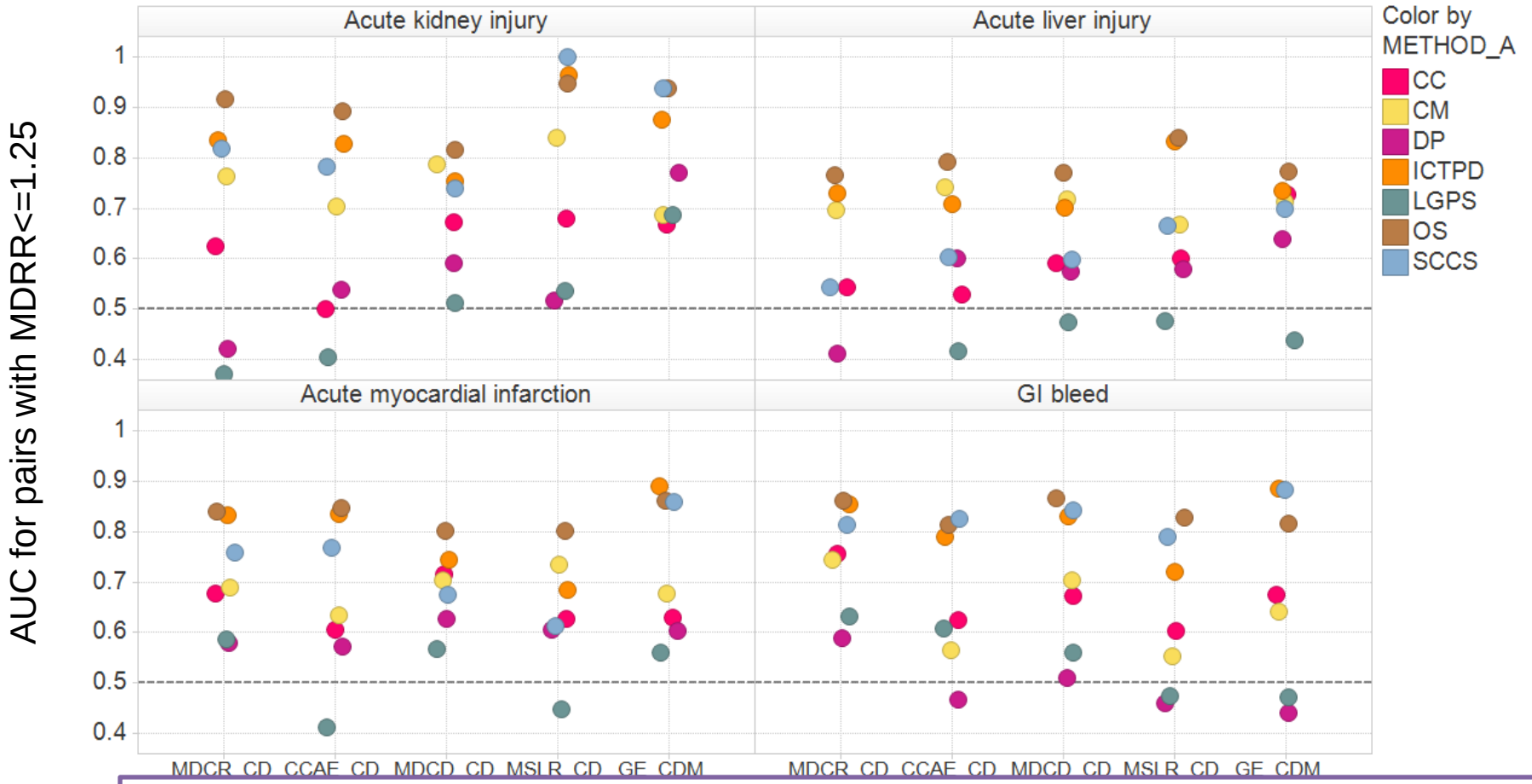
If sensitivity = 50%:

Outcome	AUC	Threshold	Specificity
Acute kidney injury	0.92	2.69	95%
Acute liver injury	0.76	1.51	89%
Acute myocardial infarction	0.84	1.59	92%
GI bleed	0.86	1.87	94%

In MDCR

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Performance across methods, by database

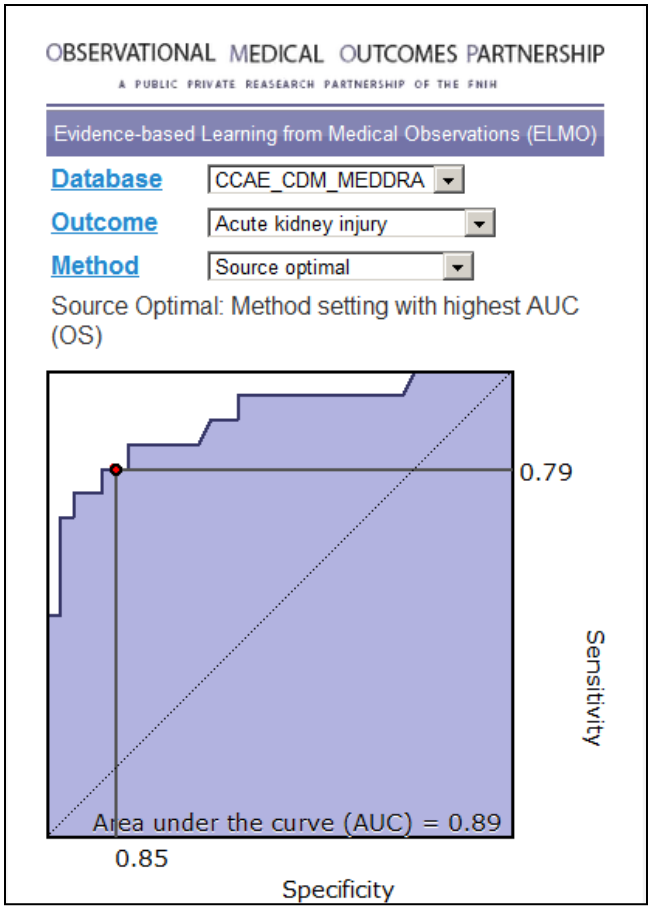


- All self-controlled designs (OS, ICTPD, SCCS) are consistently at or near the top of performance across all outcomes and sources
- Cohort and case-control designs have comparable performance, consistently lower than all self-controlled designs
- Substantial variability in performance across the optimal settings of each method

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Wow, that's really good performance…right?

- …it all depends on your tolerance of false positives and false negatives…
- …but we've created a tool to let you decide



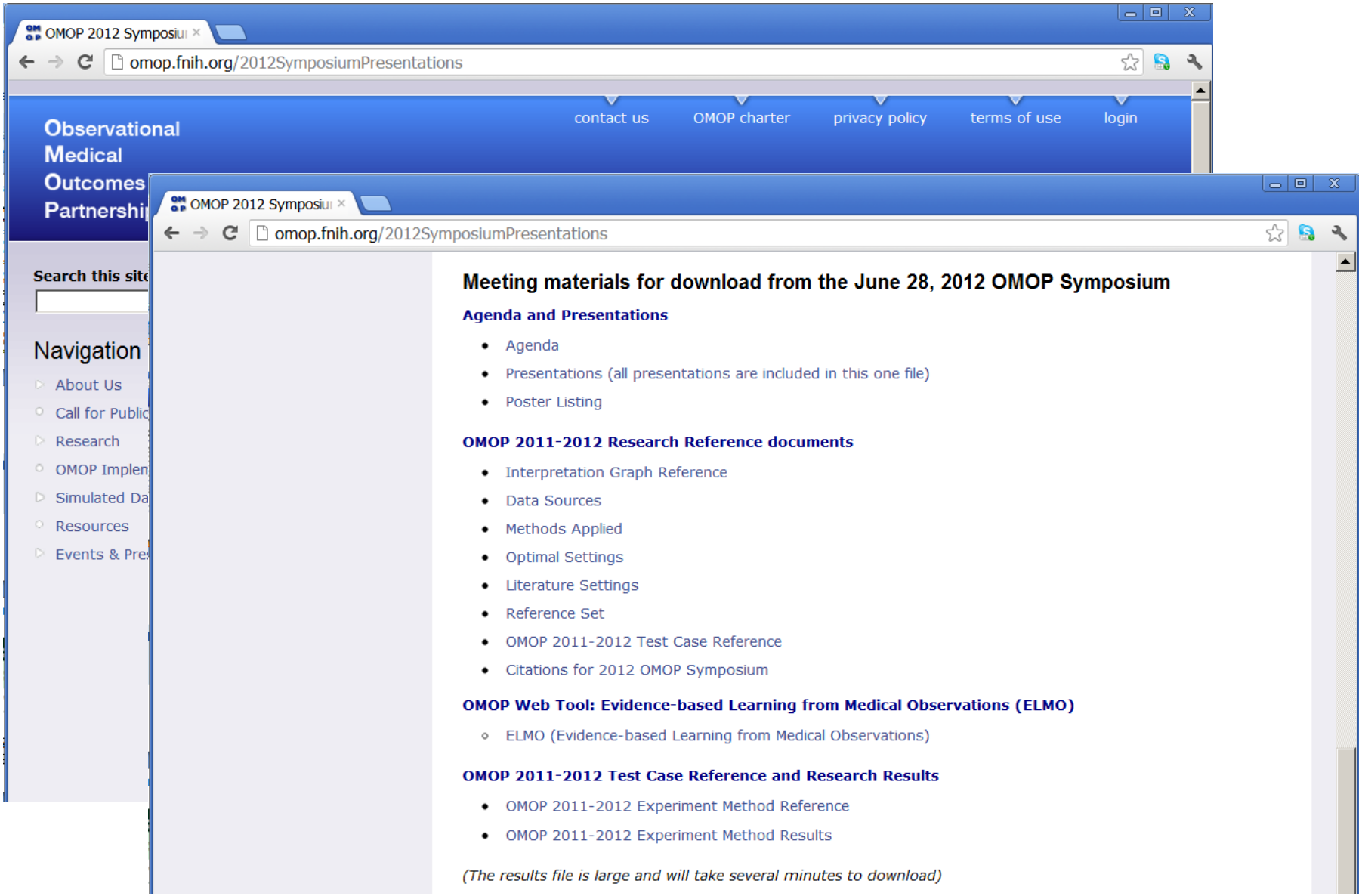
http://elmo.omop.org

Takeaways from insights about risk identification

- Performance of different methods
  - Self-controlled designs appear to consistently perform well
- Evaluating alternative HOI definitions
  - Broader definitions have better coverage and comparable performance to more specific definitions
- Performance across different signal sizes
  - A risk identification system should confidently discriminate positive effects with RR>2 from negative controls
- Data source heterogeneity
  - Substantial variation in estimates across sources suggest replication has value but may result in conflicting results
- Method parameter sensitivity
  - Each method has parameters that are expected to be more sensitive than others, but all parameters can substantially shift some drug-outcome estimates

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

All findings and results datasets are publicly available



http://omop.fnih.org/2012SymposiumPresentations

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

An empirical approach to null hypothesis testing

Revisiting clopidogrel & GI bleed (Opatrny, 2008)

Agent	Cases ($n = 4028$)	Controls ($n = 40\ 171$)	Crude rate ratio	Adjusted rate ratio*	95% confidence interval
Antidepressants					
SSRI	335 (8.3%)	1780 (4.4%)	1.97	1.33	1.09, 1.62
TCA	262 (6.5%)	1764 (4.4%)	1.52	1.04	0.83, 1.30
Venlafaxine	56 (1.4%)	229 (0.6%)	2.48	1.85	1.34, 2.55
Anticoagulant					
Warfarin	281 (7.0%)	1130 (2.8%)	2.64	2.17	1.82, 2.59
Clopidogrel	160 (4.0%)	532 (1.3%)	3.16	2.07	1.66, 2.58

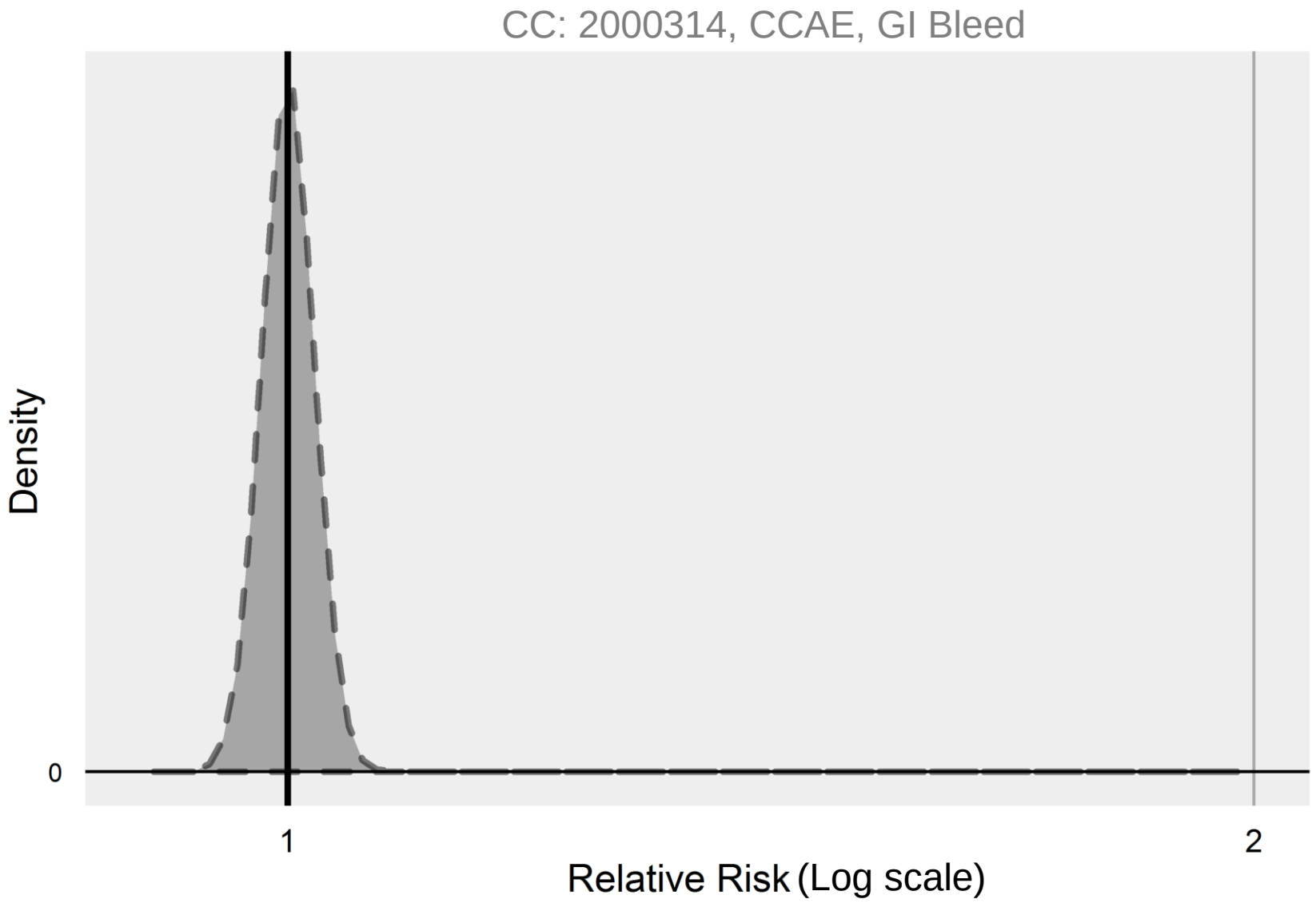
OMOP, 2012 (CC: 2000314, CCAE, GI Bleed)

Relative risk: 1.86, 95% CI: 1.79 – 1.93

Standard error: 0.02, p-value: <.001

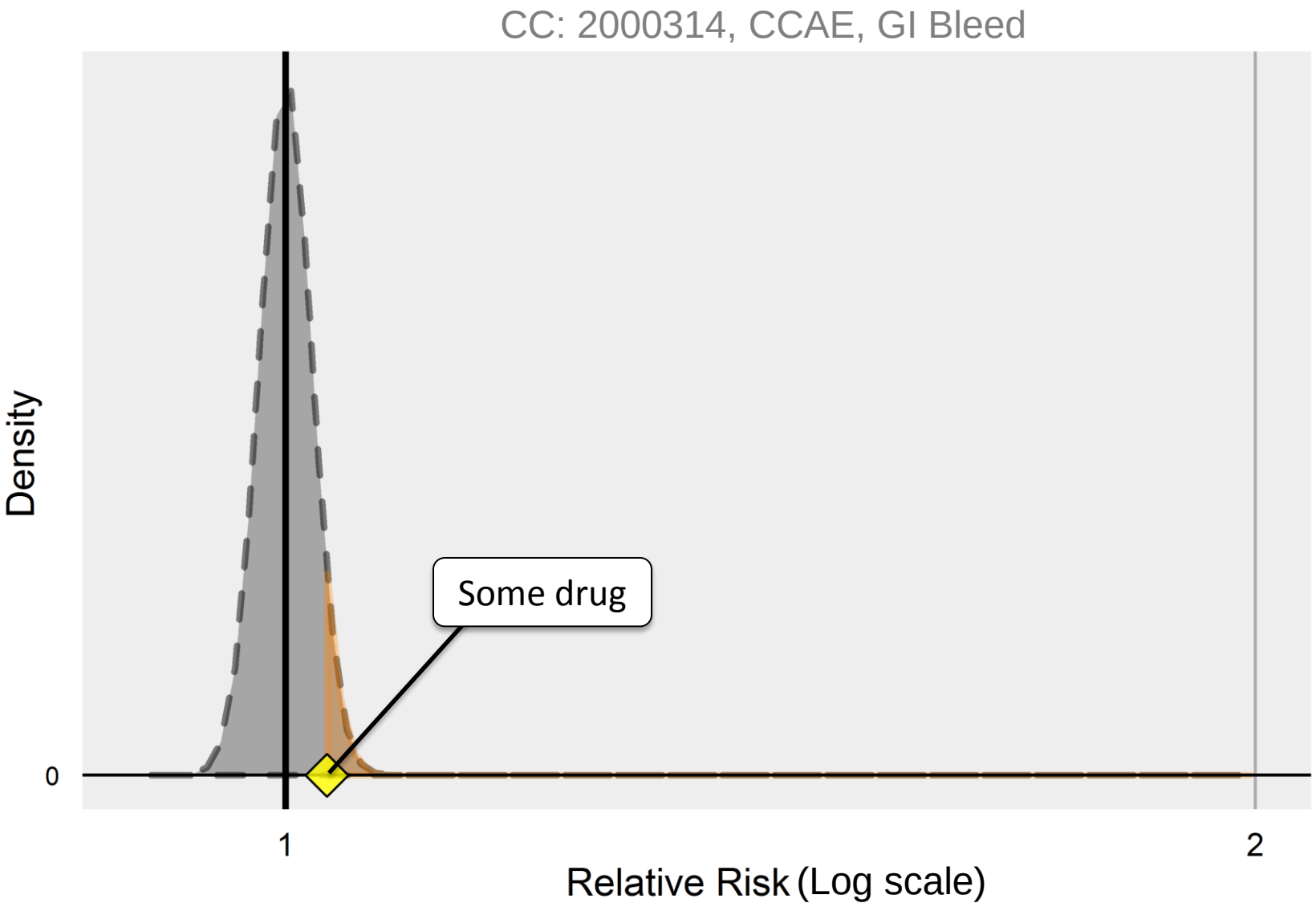
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Null distribution



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Null distribution



Evaluating the null distribution?

- Current p-value calculation assumes that you have an unbiased estimator (which means confounding either doesn't exist or has been fully corrected for)
- Traditionally, we reject the null hypothesis at $p<.05$ and we assume this threshold will incorrectly reject the null hypothesis 5% of time. Does this hold true in observational studies?
- We can test this using our negative controls

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Ground truth for OMOP 2011/2012 experiments

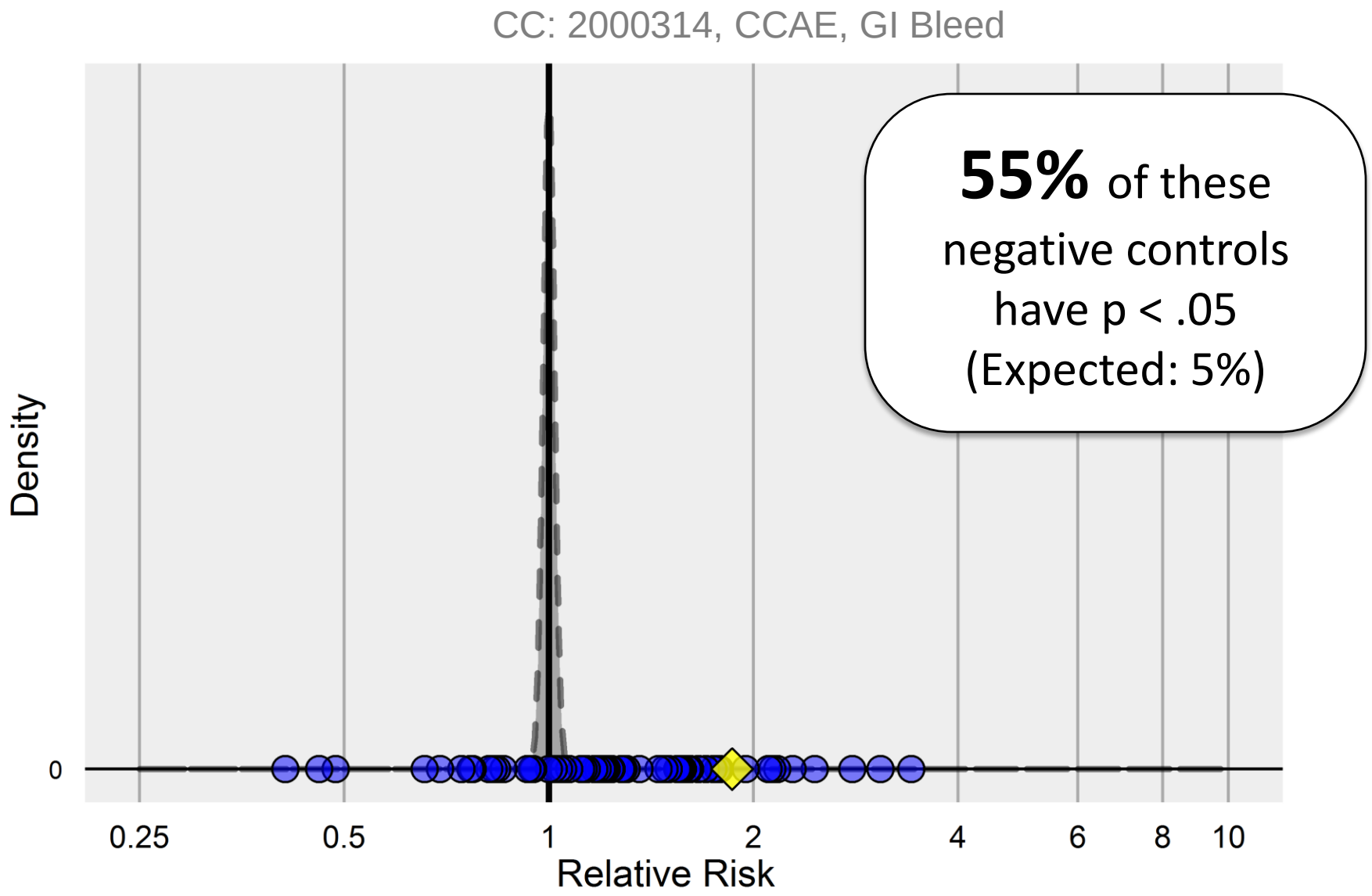
	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

Criteria for negative controls:

- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"
- Literature review identified no evidence of potential positive association

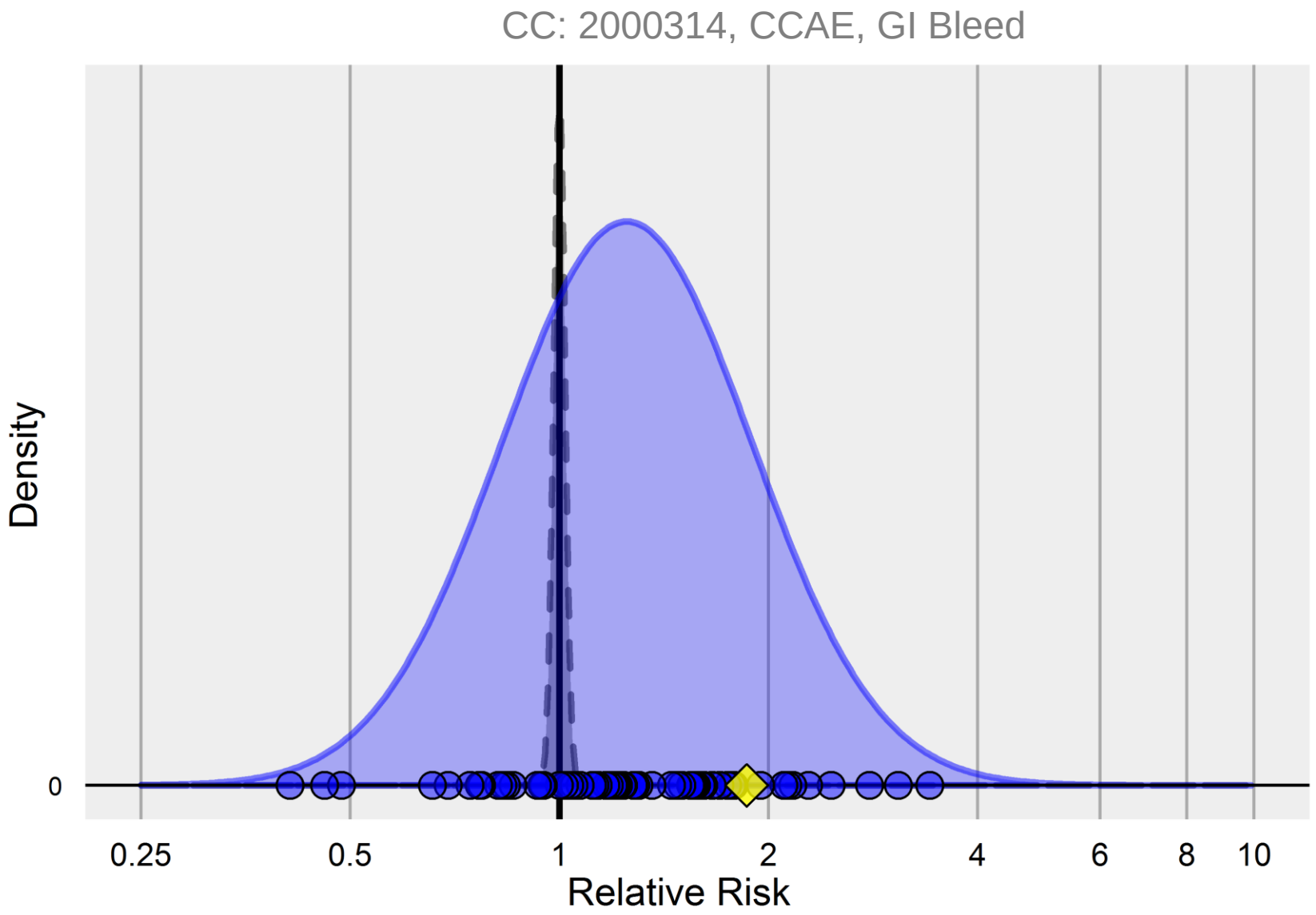
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution



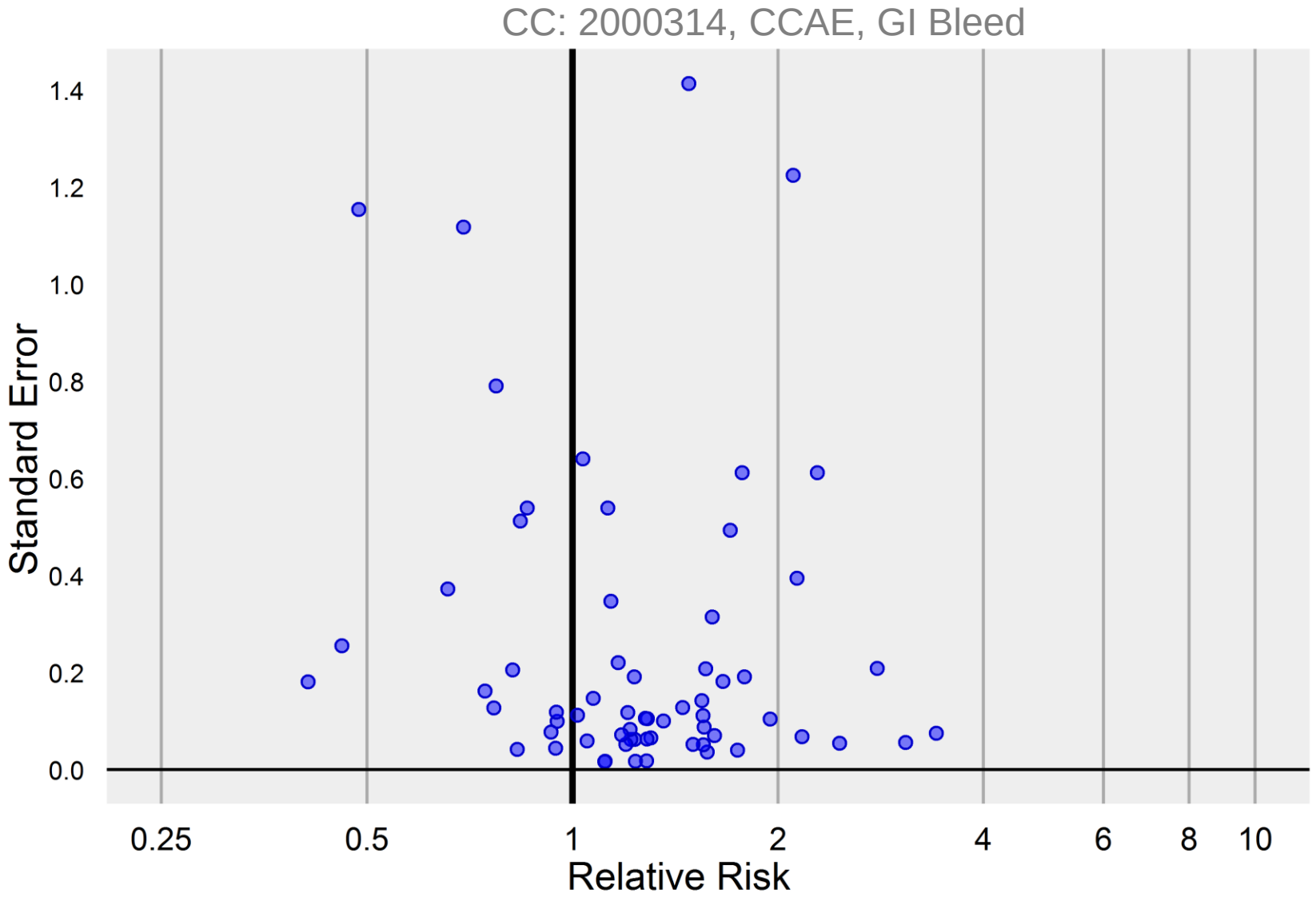
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution



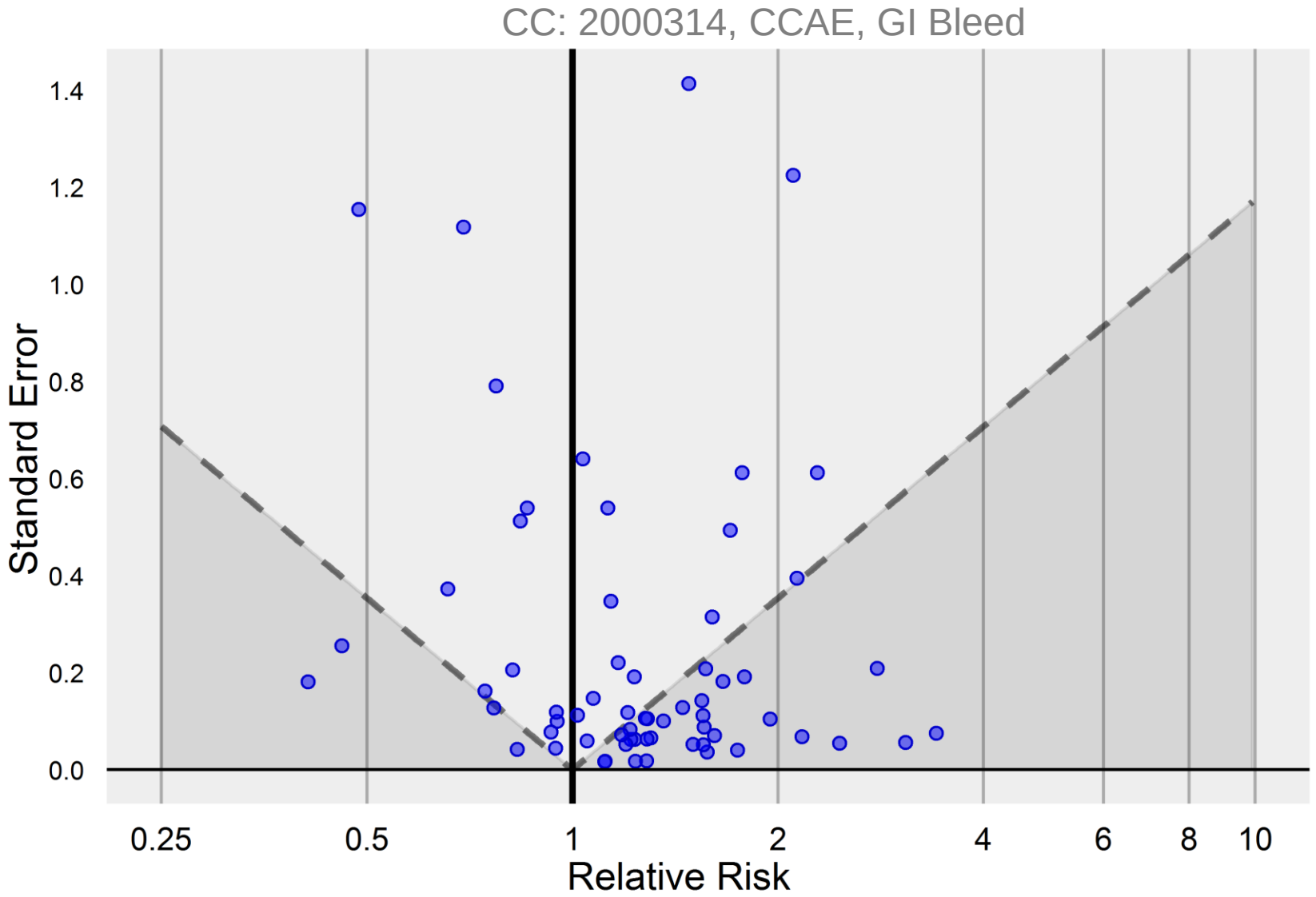
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot



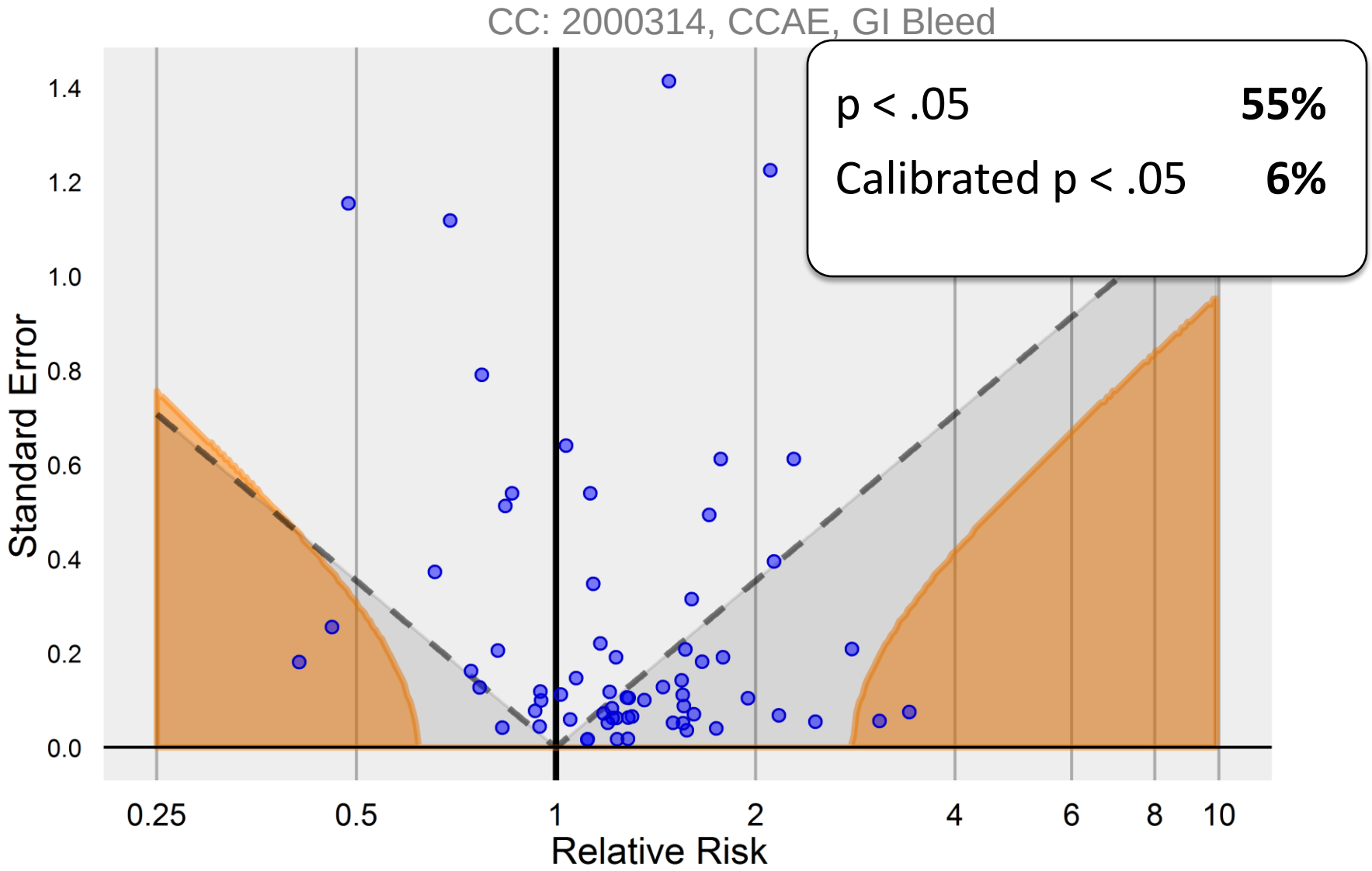
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot



Recap

- Traditional p-values are based on a theoretical null distribution assuming an unbiased estimator, but that assumption rarely holds in our examples
- One can estimate the empirical null distribution using negative controls
- Many observational study results with traditional $p < .05$ fail to reject the empirical null: we cannot distinguish them from negative controls
- Applying optimal methods, tailored to the outcome and database, can provide estimates that reject the null hypothesis for some of our positive controls
- Using adjusted p-values will provide a more calibrated assessment of whether an observed estimate is different from 'no effect'

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

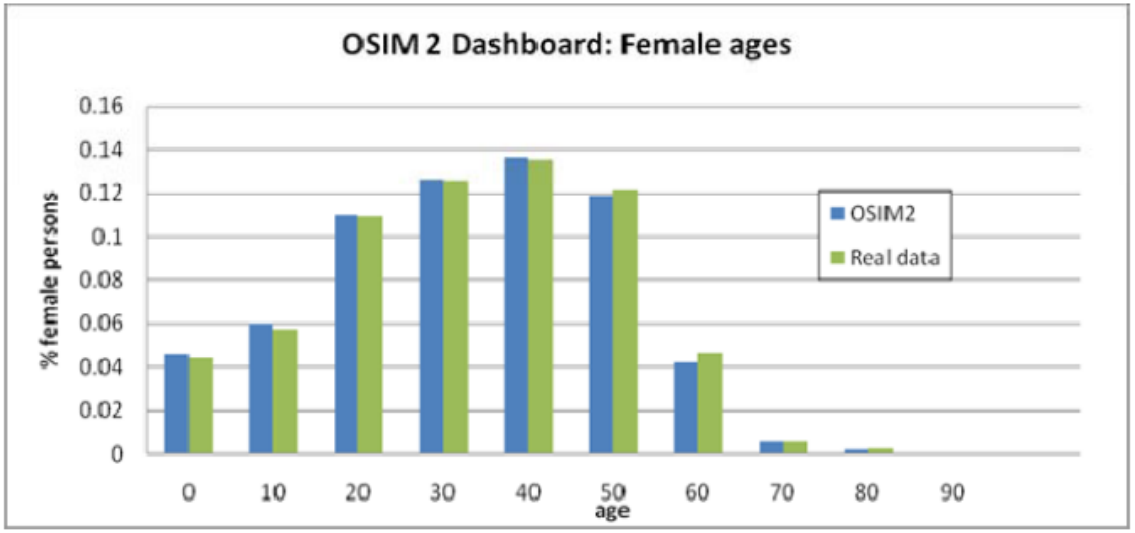
Simulating healthcare data

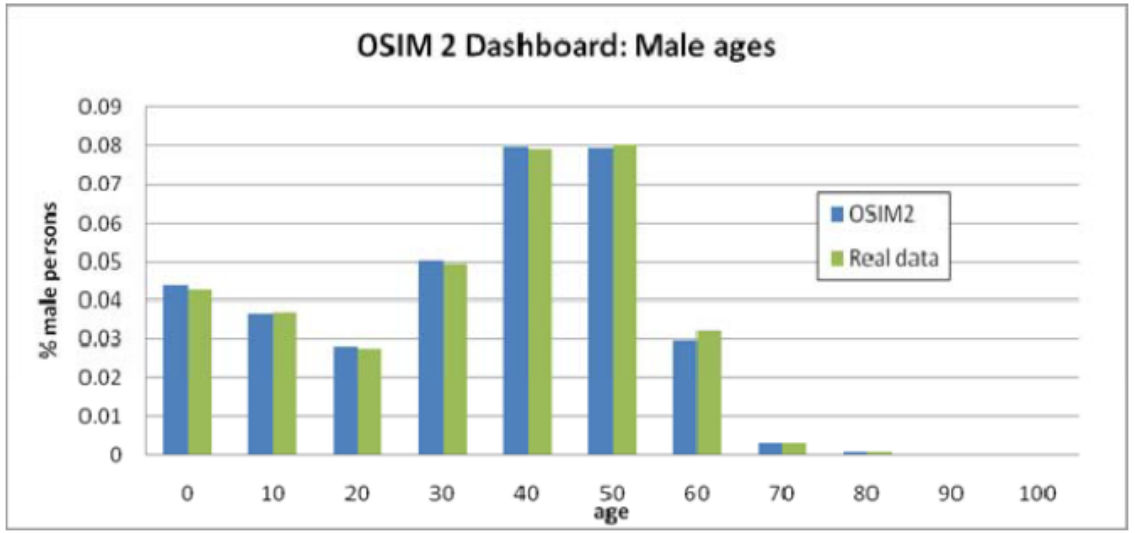
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OSIM2 approach to simulating real-world data

Step 1: Generate a population

Derive marginal statistics from a real database (here, MSLR) and apply them within the simulation model



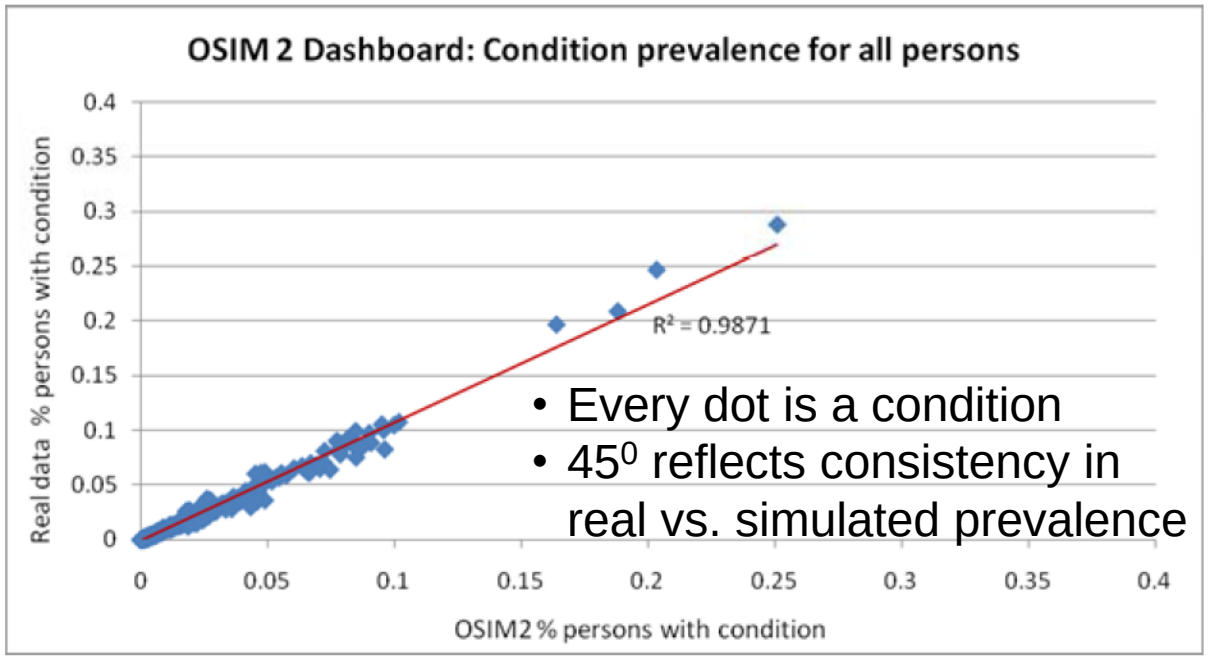


Implemented by UBC/ProSanos

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

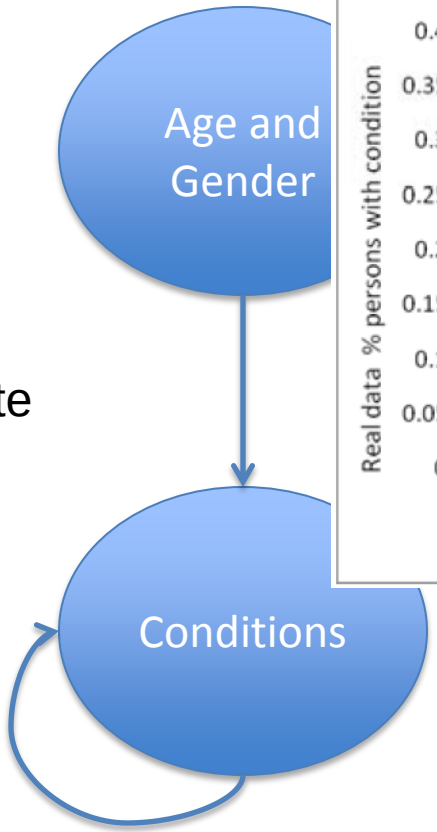
OSIM2 approach to simulating real-world data

Step 1: Generate a population



- Every dot is a condition
- 45^0 reflects consistency in real vs. simulated prevalence

Step 2: Create conditions for the simulated population



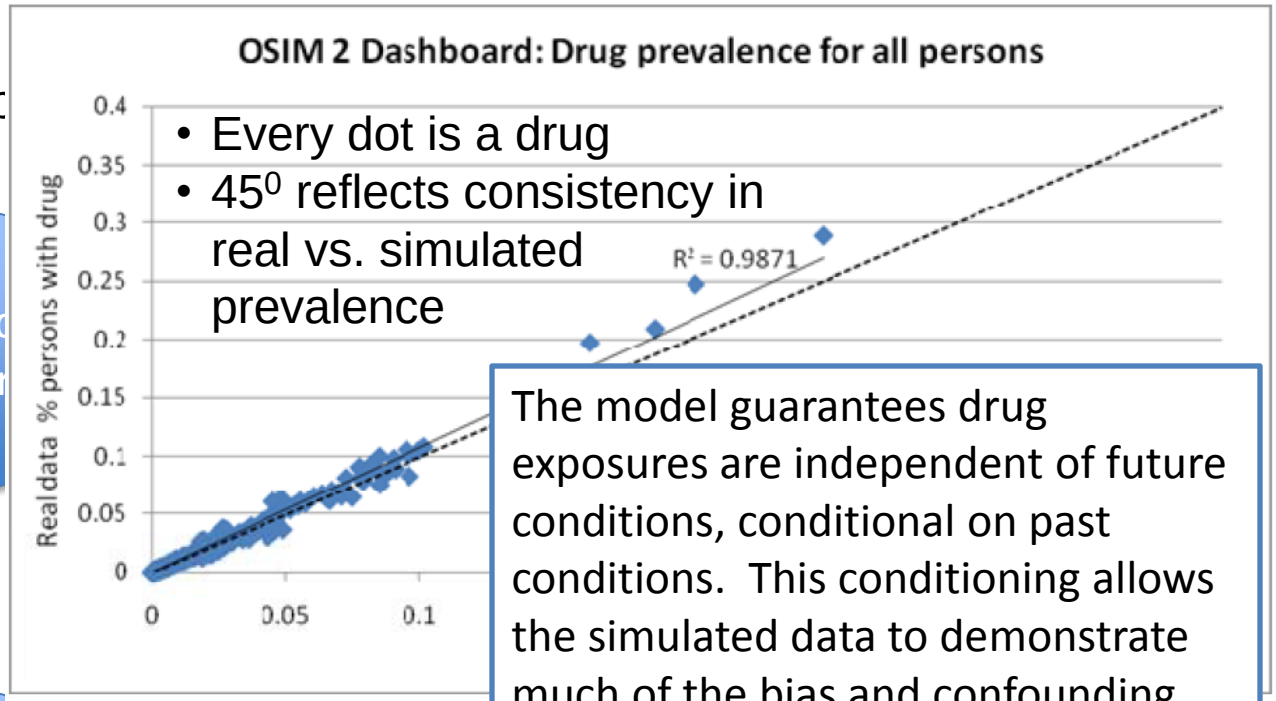
From the real data, we know the probability that a 50yo male will have diabetes….we use that probability to add diabetes to the simulated patients

From the real data, we know the probability that a 50yo male with diabetes will have an AMI….we use that probability to add AMI to simulated patients with diabetes

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

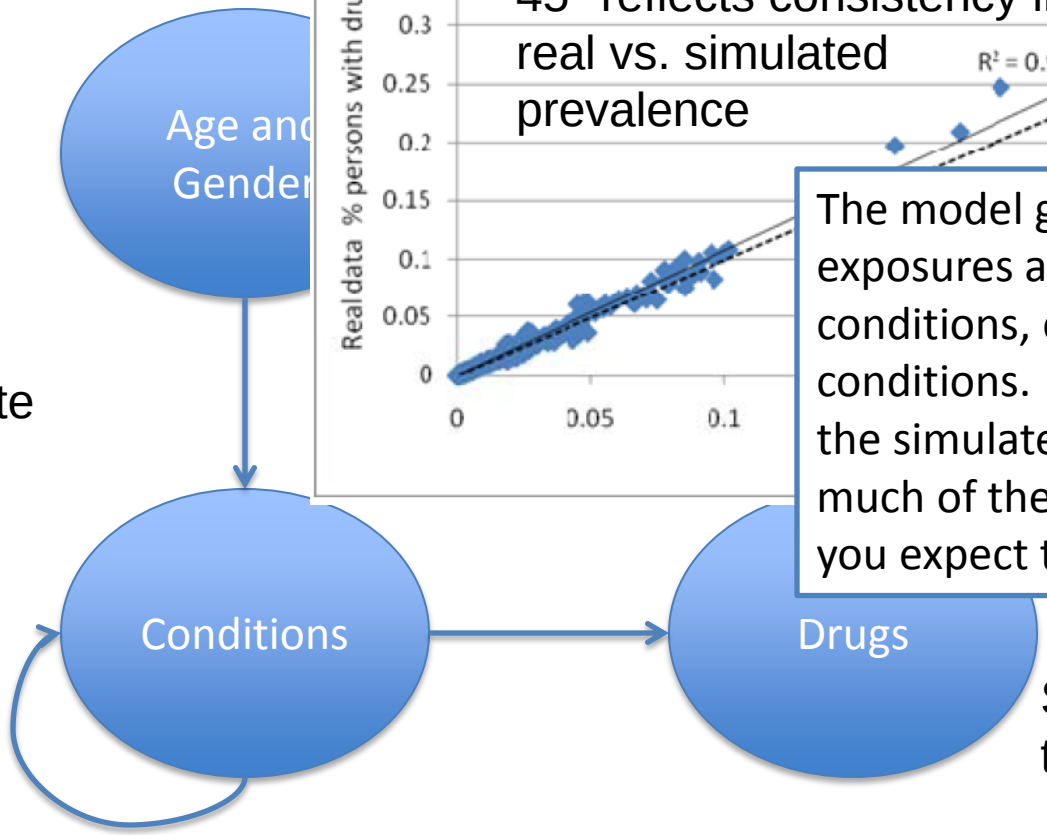
OSIM2 approach to simulating real-world data

Step 1: Generate a pop



The model guarantees drug exposures are independent of future conditions, conditional on past conditions. This conditioning allows the simulated data to demonstrate much of the bias and confounding you expect to observe in real data.

Step 2: Create conditions for the simulated population



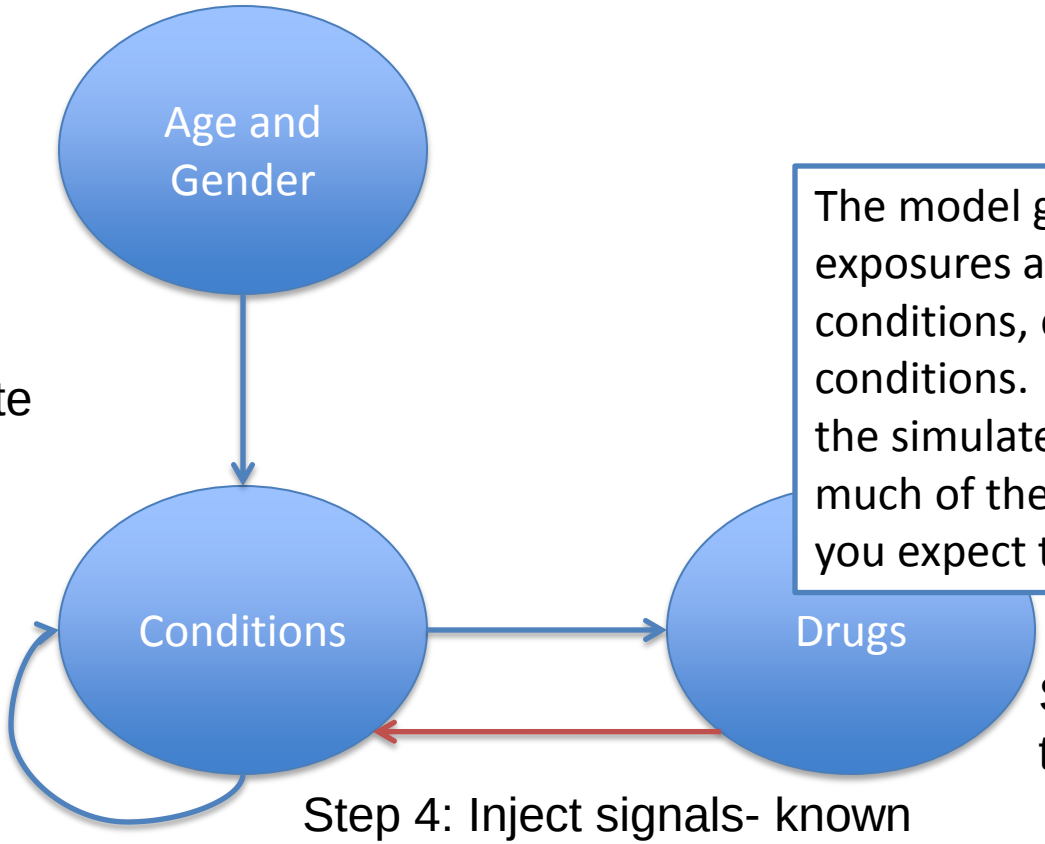
Step 3: Create drugs for the simulated population

From the real data, we know the probability that a 50yo male with diabetes will be dispensed a prescription for metformin….we use that probability to add metformin exposure to simulated patients with diabetes

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OSIM2 approach to simulating real-world data

Step 1: Generate a population



Step 2: Create conditions for the simulated population

The model guarantees drug exposures are independent of future conditions, conditional on past conditions. This conditioning allows the simulated data to demonstrate much of the bias and confounding you expect to observe in real data.

Step 3: Create drugs for the simulated population

Step 4: Inject signals- known causal effects of drugs- as additional conditions

Signals can be defined with specific effect sizes (RR) and types (acute, insidious, accumulative)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Interpreting effect sizes from confidence intervals

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

What have we learned so far?

Is there an effect?

- Can you reject the null hypothesis of no association between the drug and outcome at a given significance level (ex: p<.05)?
- What is the probability that the observed estimate is a positive association?

How big is the effect?

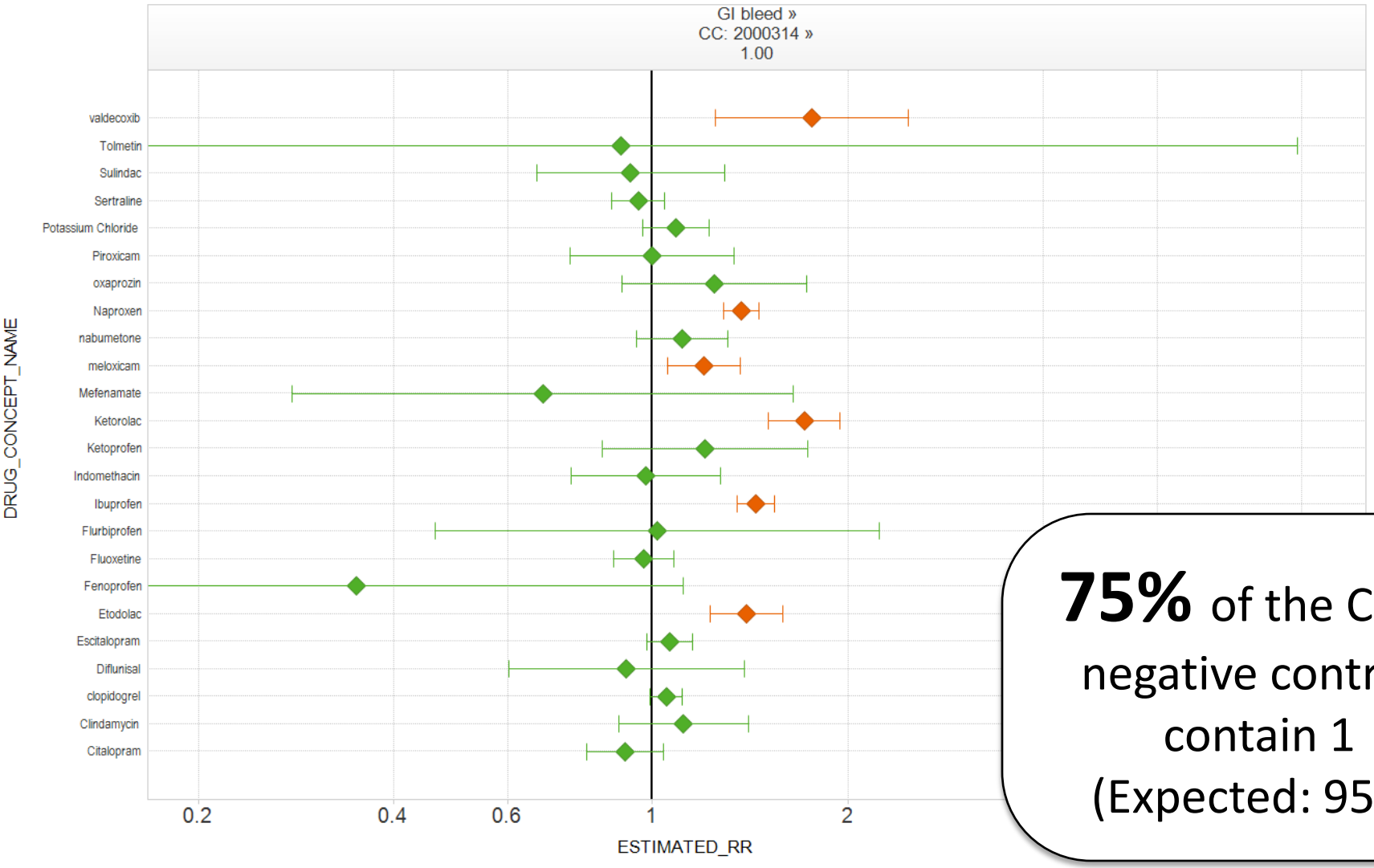
- New question: What is the probability that observed confidence interval contains the true effect size?

Estimating coverage probability

- What if a study design could be applied across a large sample of drug-outcome pairs for which we know the true effect?
- Coverage probability: the percentage of the test cases where the estimated confidence interval contains the true effect (LB 95 CI <= true effect <= UB 95 CI)
- Challenge: in real data, the 'true effect size' for negative controls can be assumed to be RR=1, but the RRs for positive controls are not known
- Opportunity: in simulated data (OSIM2), we can inject signals with known effect sizes (RR=1.25, 1.50, 2, 4, 10) across a sample of drug-outcome scenarios and estimate the coverage probability

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

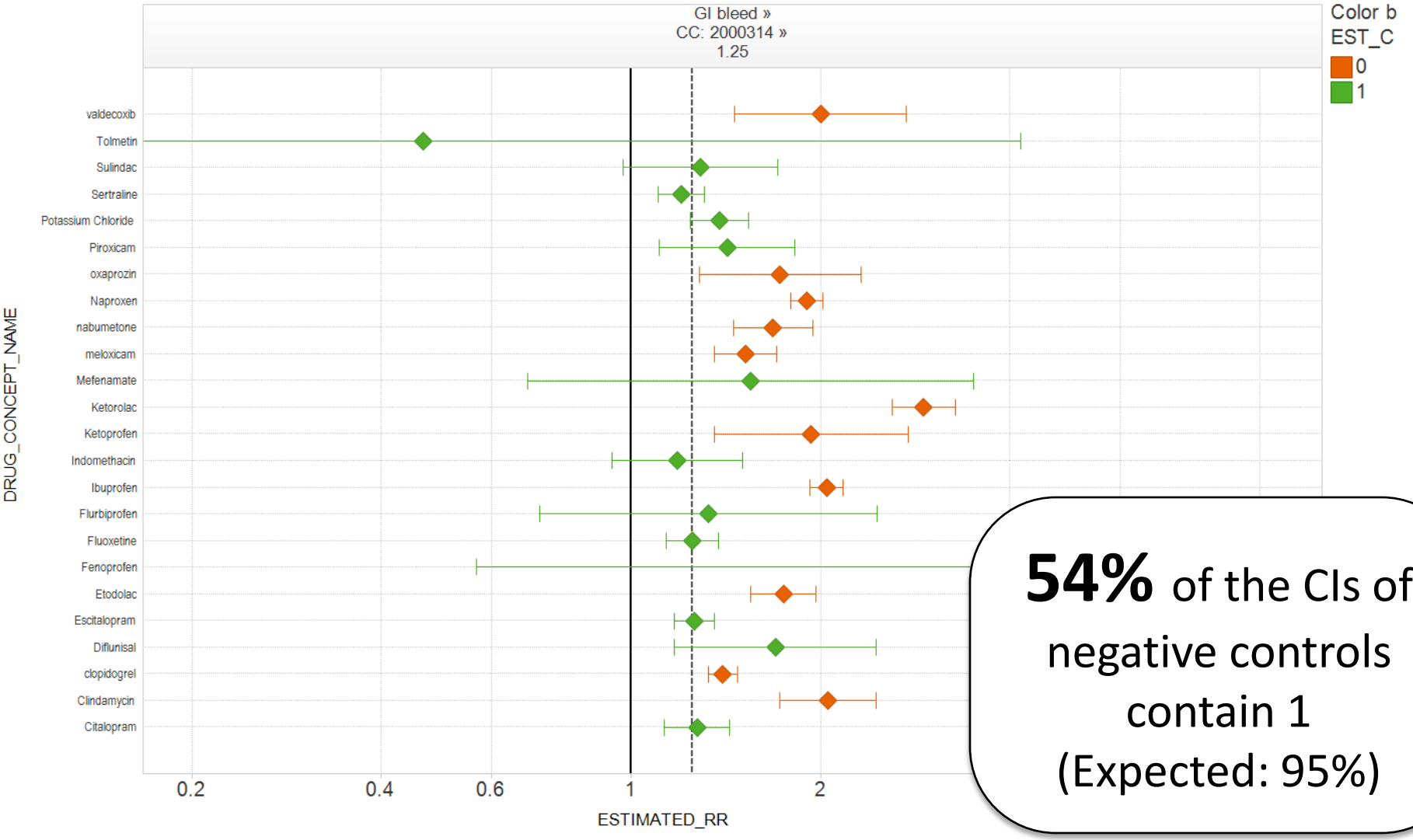
Applying case-control design to positive controls in simulated data, RR=1.0



75% of the CIs of negative controls contain 1 (Expected: 95%)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

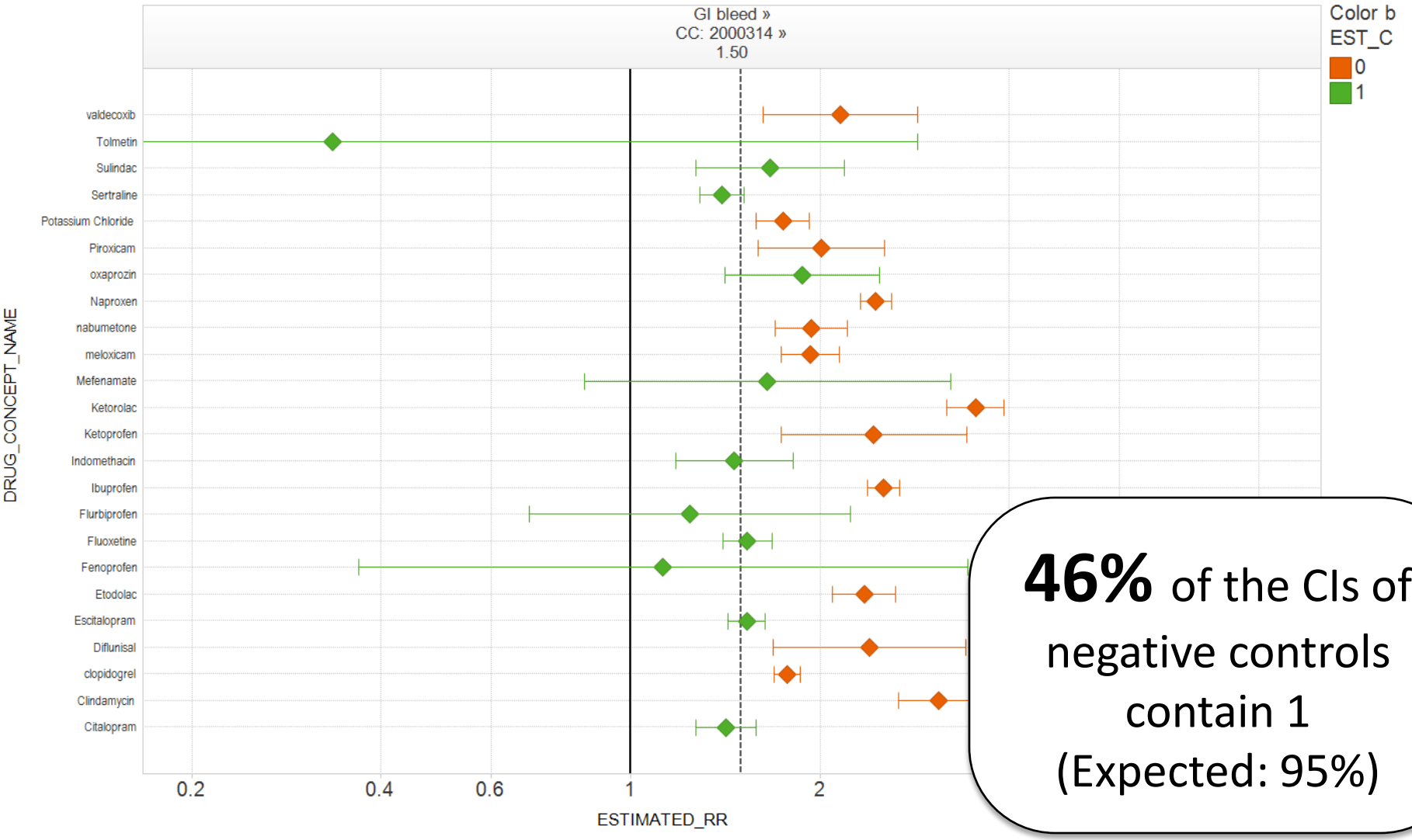
Applying case-control design to positive controls in simulated data, RR=1.25



54% of the CIs of negative controls contain 1 (Expected: 95%)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

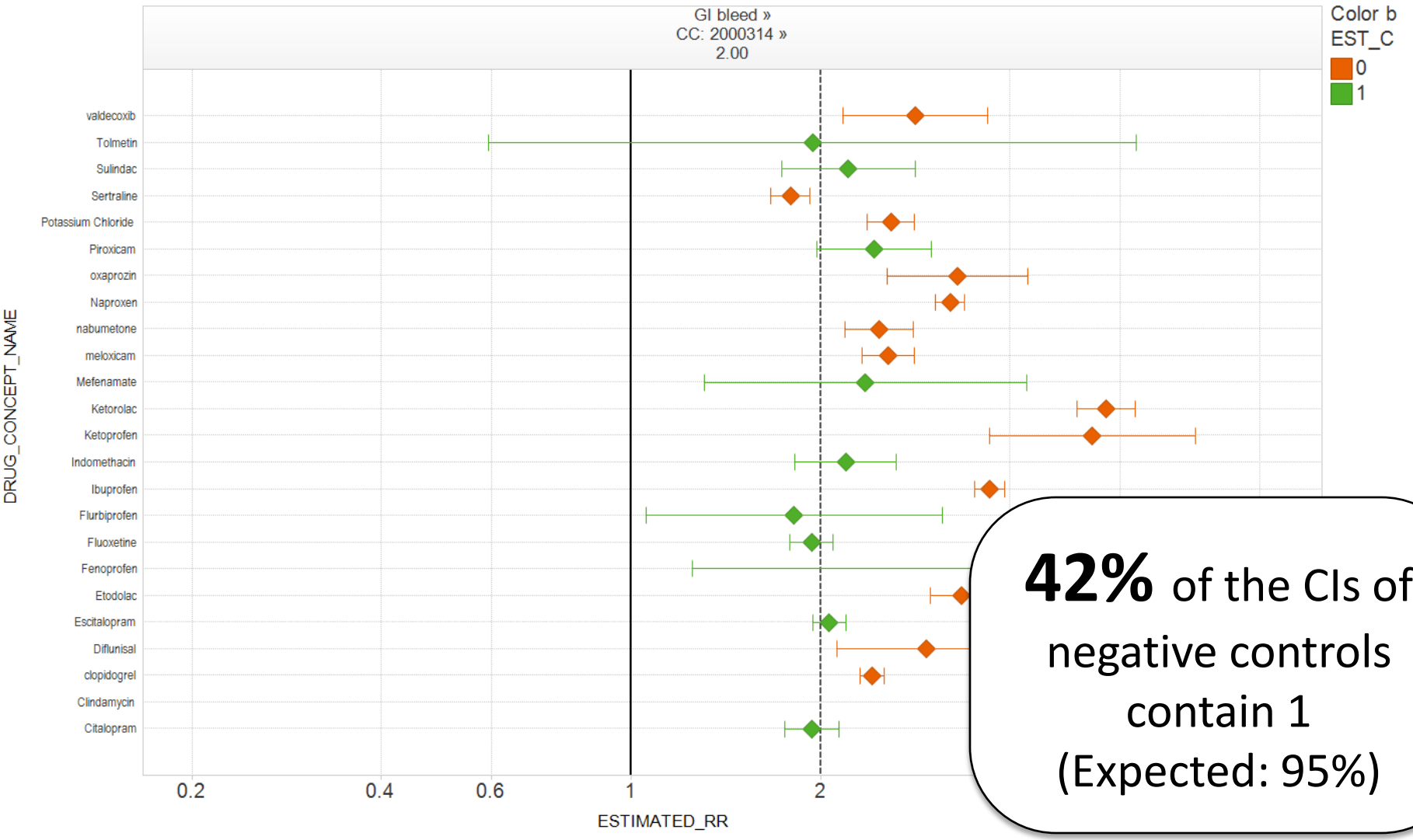
Applying case-control design to positive controls in simulated data, RR=1.50



46% of the CIs of negative controls contain 1 (Expected: 95%)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

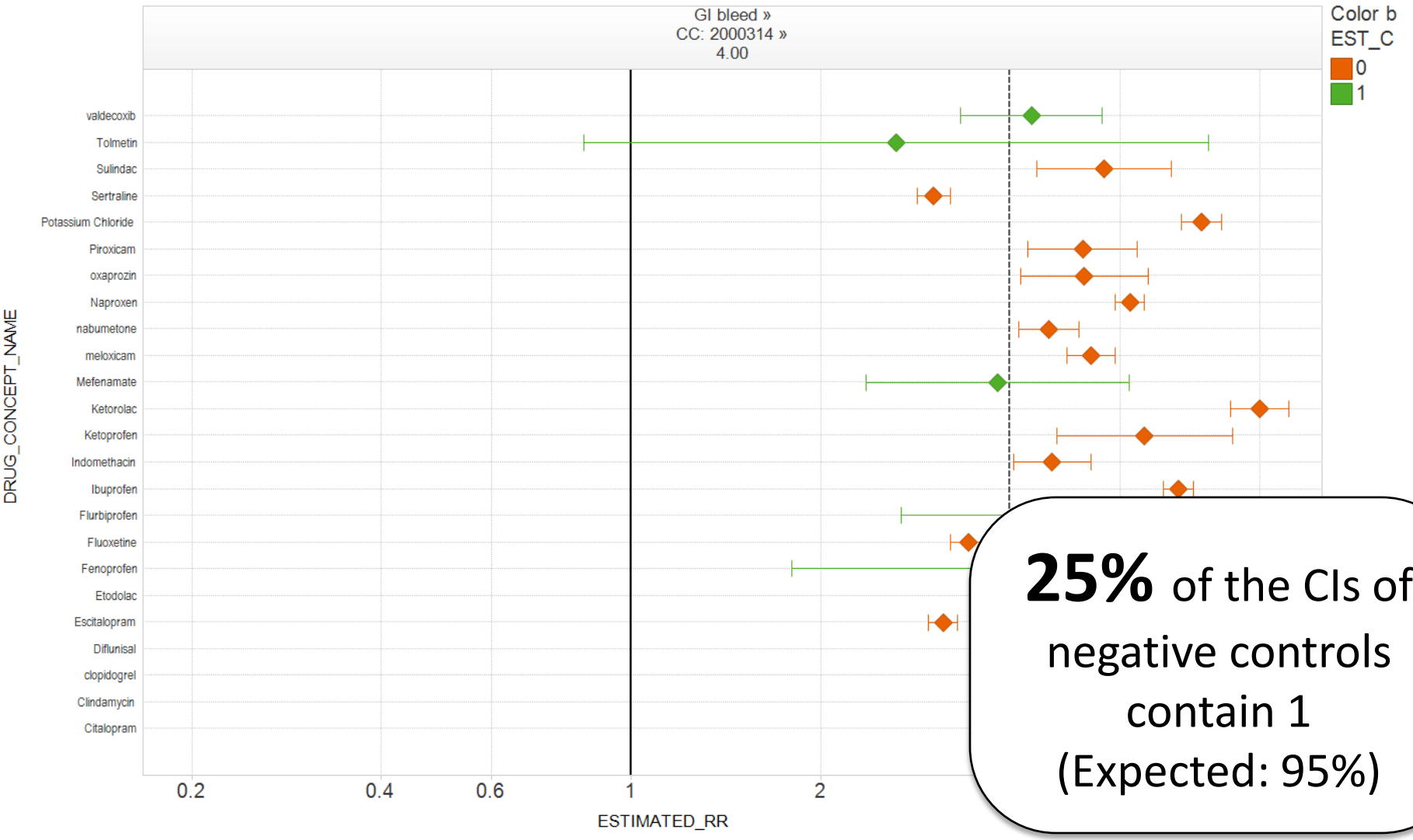
Applying case-control design to positive controls in simulated data, RR=2.00



42% of the CIs of negative controls contain 1 (Expected: 95%)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

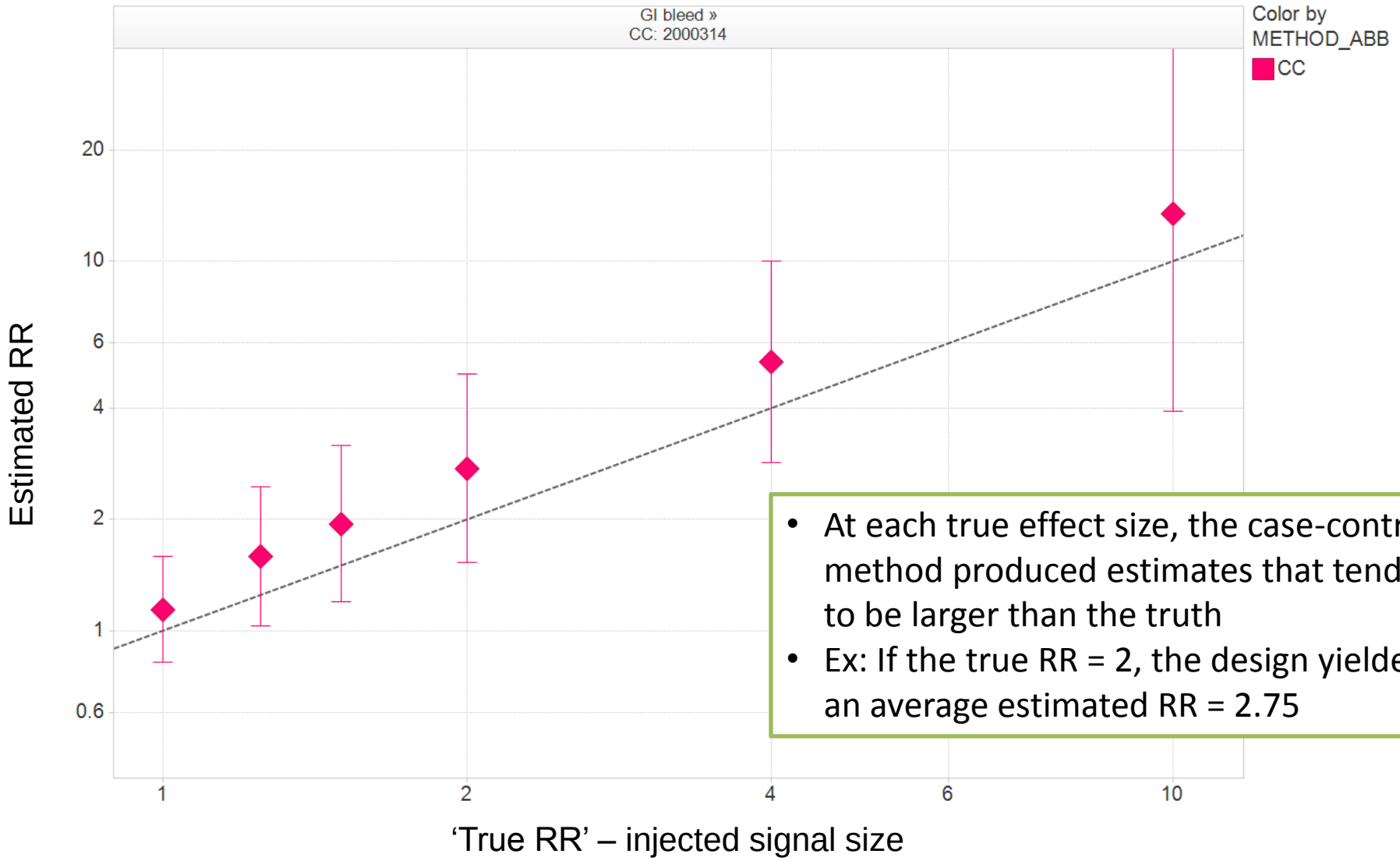
Applying case-control design to positive controls in simulated data, RR=4.00



25% of the CIs of negative controls contain 1 (Expected: 95%)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

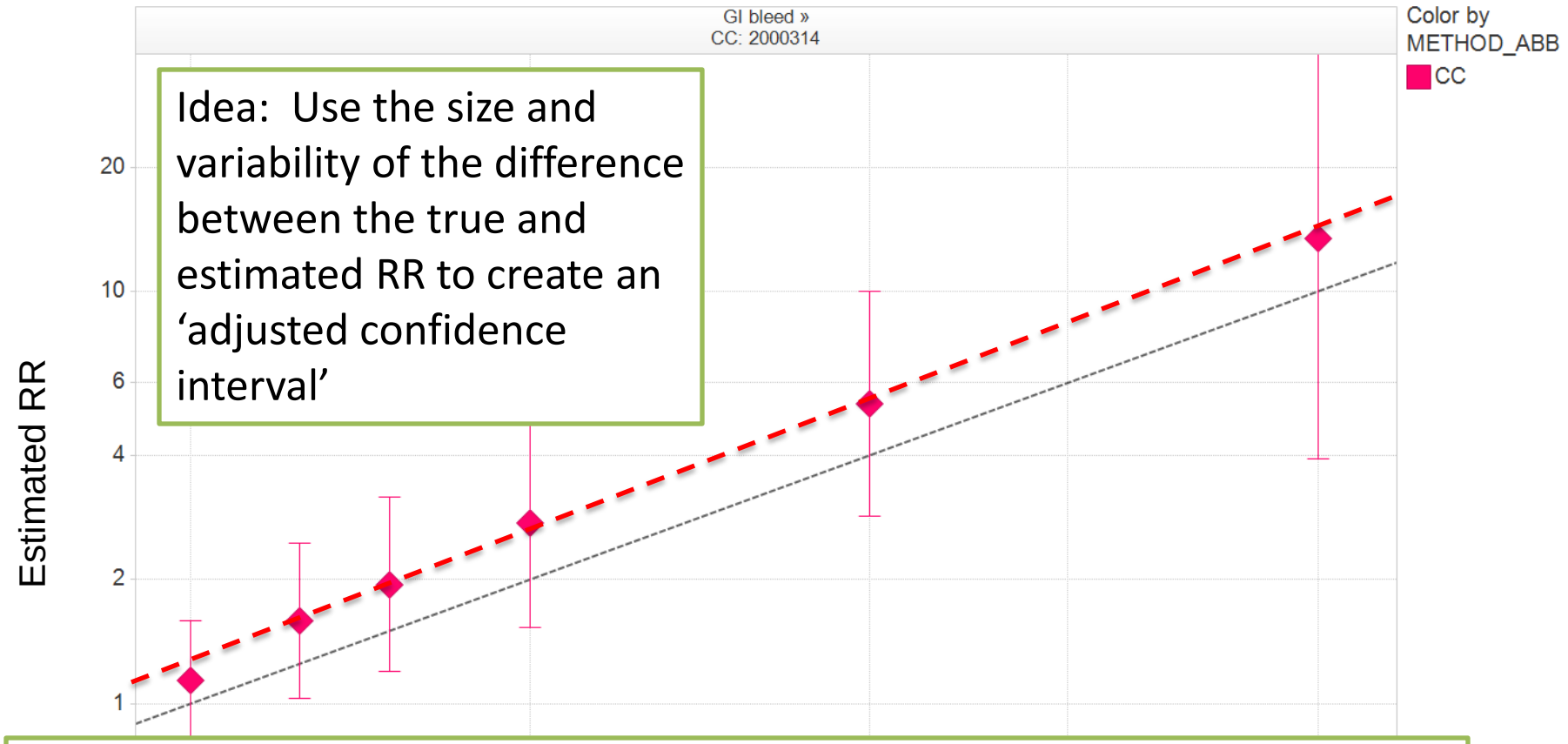
How far off were the case-control estimates from the truth?



- At each true effect size, the case-control method produced estimates that tended to be larger than the truth
- Ex: If the true RR = 2, the design yielded an average estimated RR = 2.75

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

How far off were the case-control estimates from the truth?



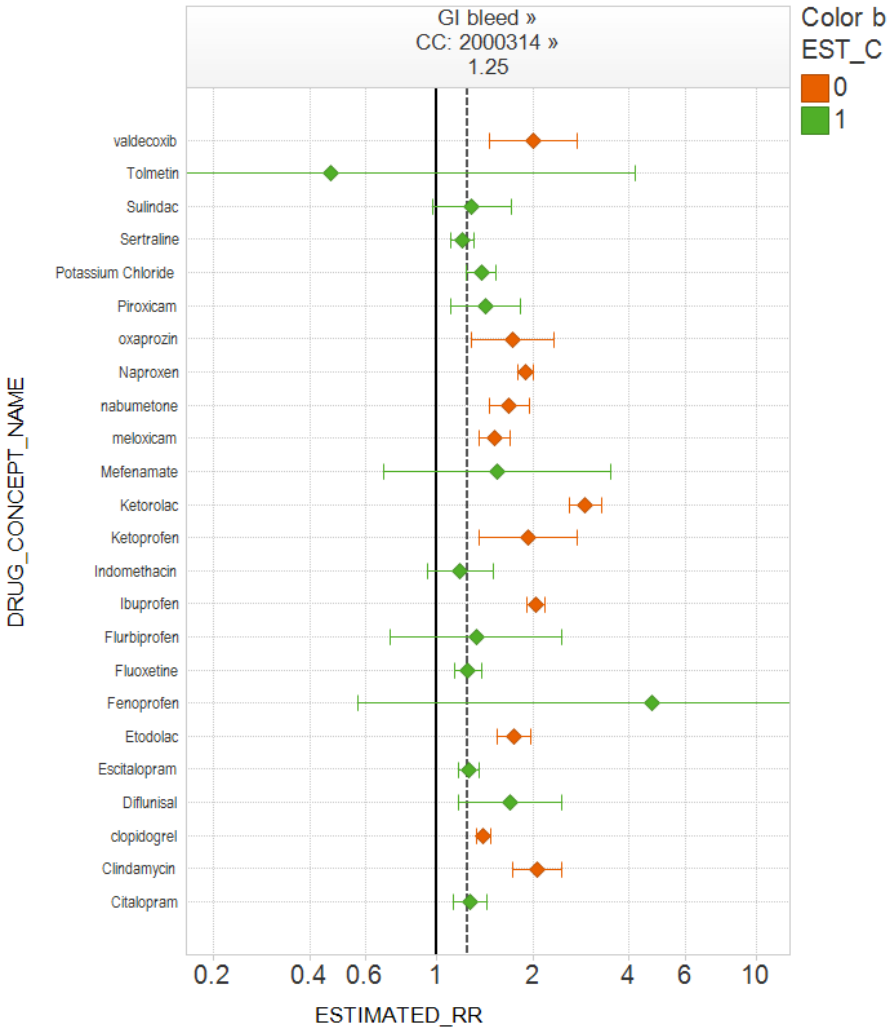
Idea: Use the size and variability of the difference between the true and estimated RR to create an 'adjusted confidence interval'

1. Model the distribution of estimates at each true RR ~N(μ, σ)
2. Fit a linear model to predict these distributions from the true RR values
3. Given a new estimated RR and SE, determine the 95% range of true RR values that have distributions from which the new estimate could have come from

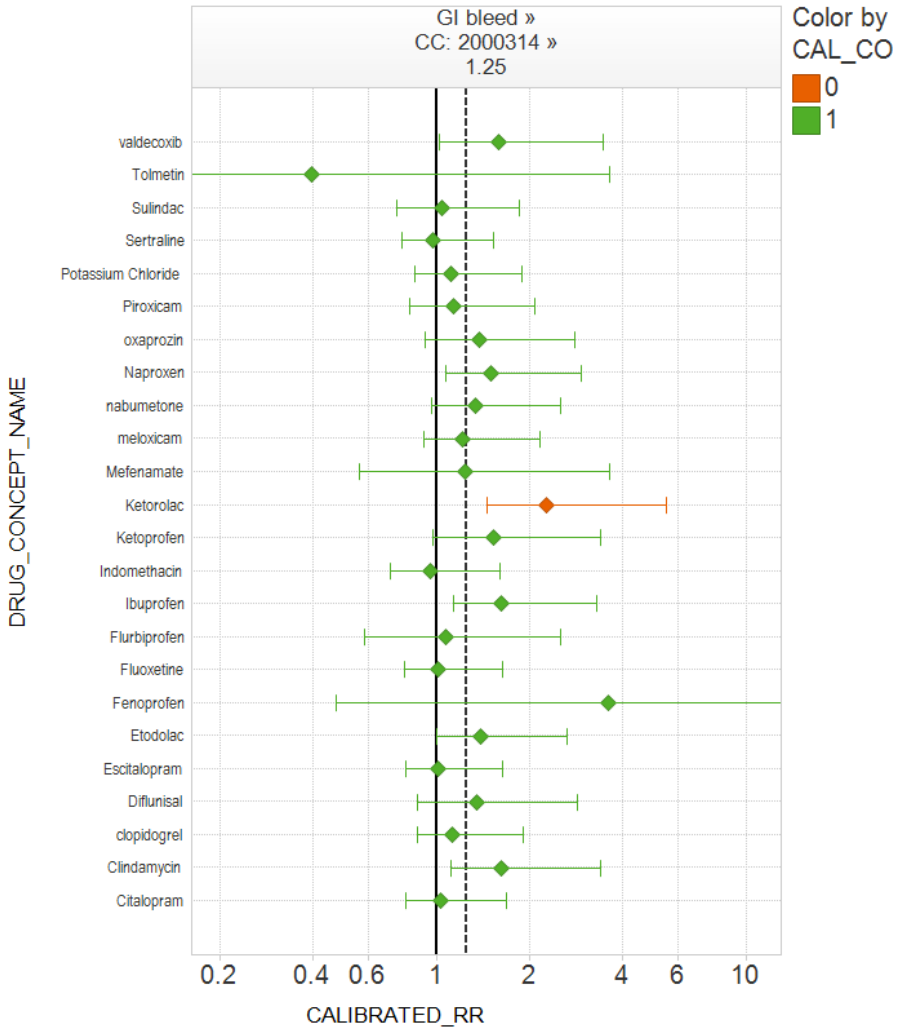
OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Applying case-control design and calibrating estimates of positive controls in simulated data, RR=1.25

Original estimated effects



Calibrated confidence intervals



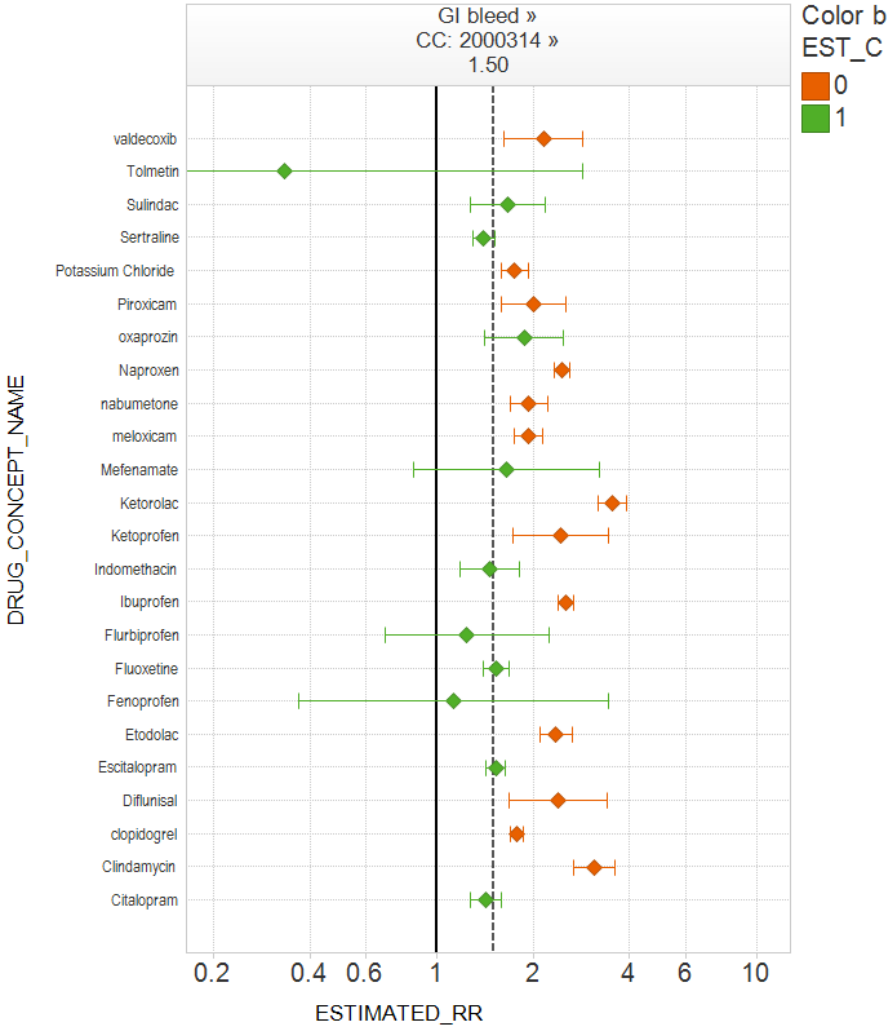
Original coverage probability = **54%**

Calibrated coverage probability = **96%**

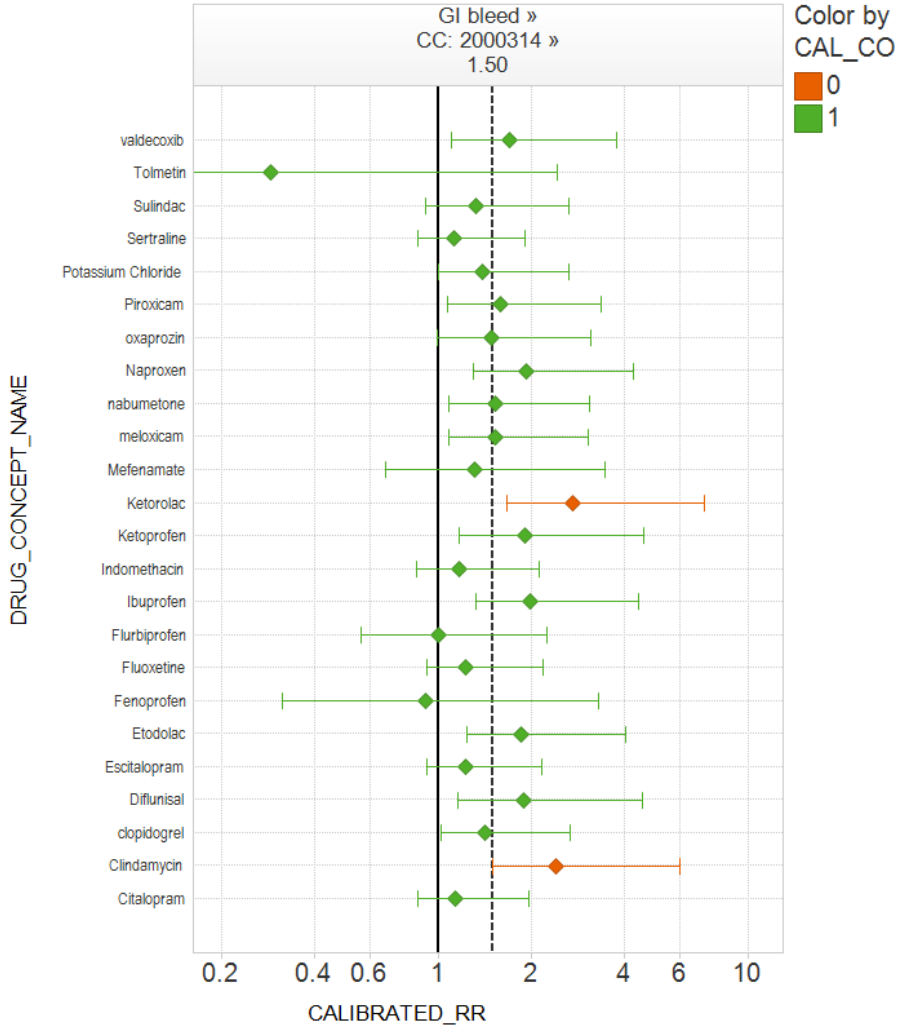
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Applying case-control design and calibrating estimates of positive controls in simulated data, RR=1.50

Original estimated effects



Calibrated confidence intervals



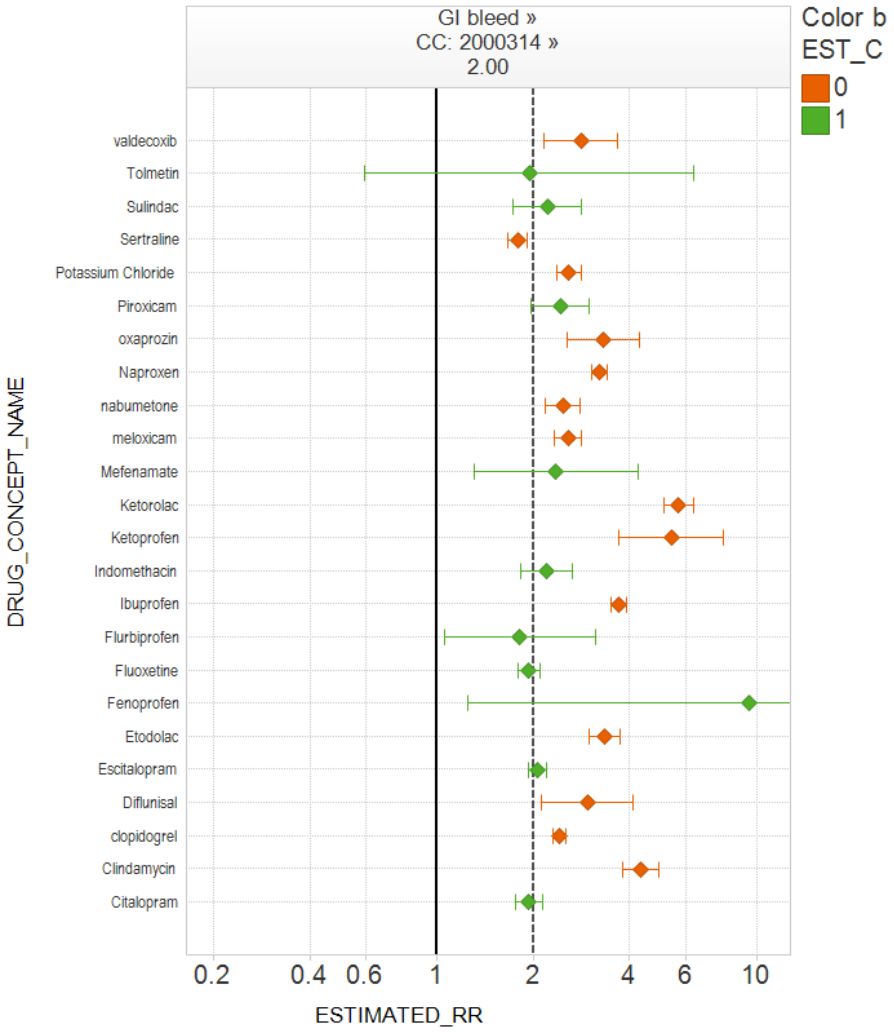
Original coverage probability = **46%**

Calibrated coverage probability = **92%**

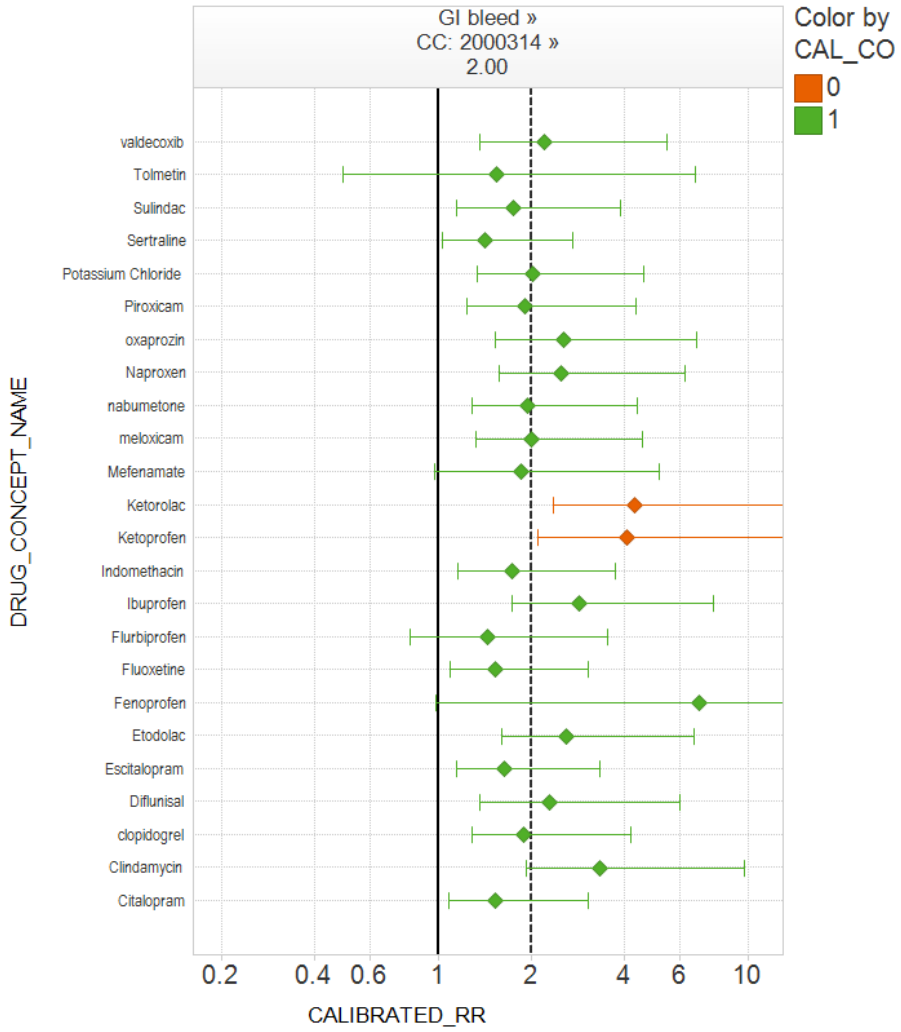
OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Applying case-control design and calibrating estimates of positive controls in simulated data, RR=2.00

Original estimated effects



Calibrated confidence intervals



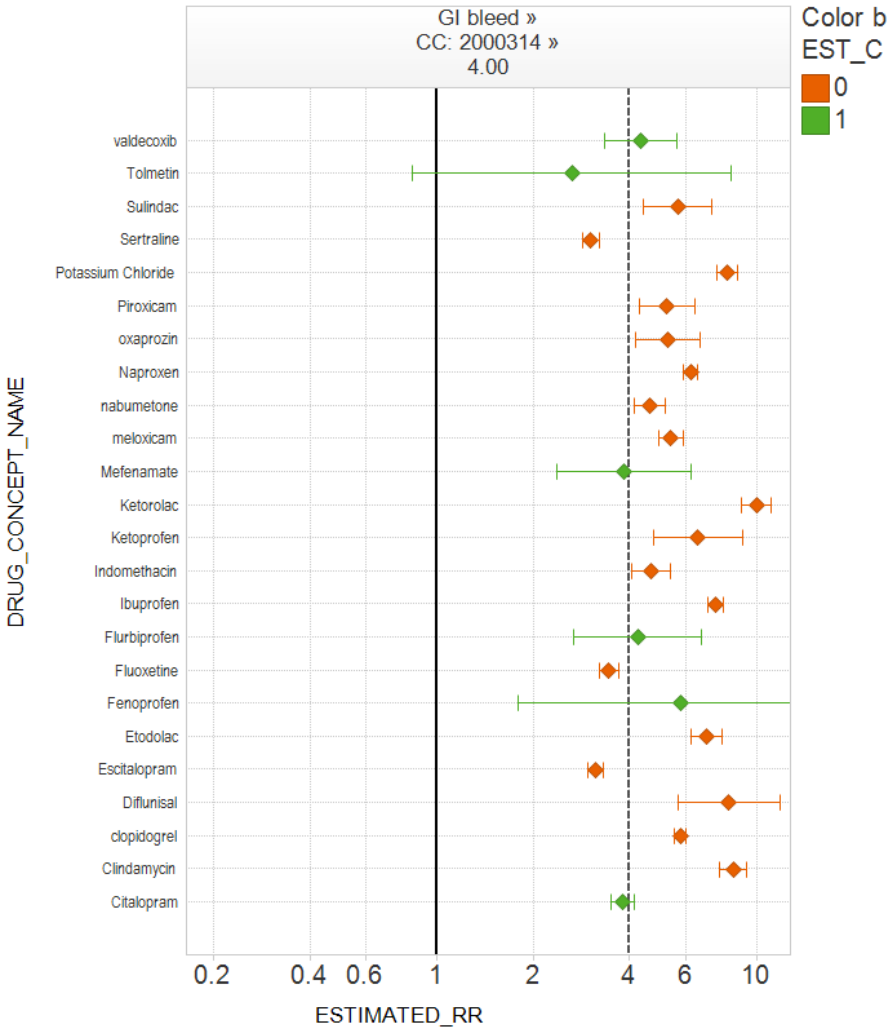
Original coverage probability = **42%**

Calibrated coverage probability = **92%**

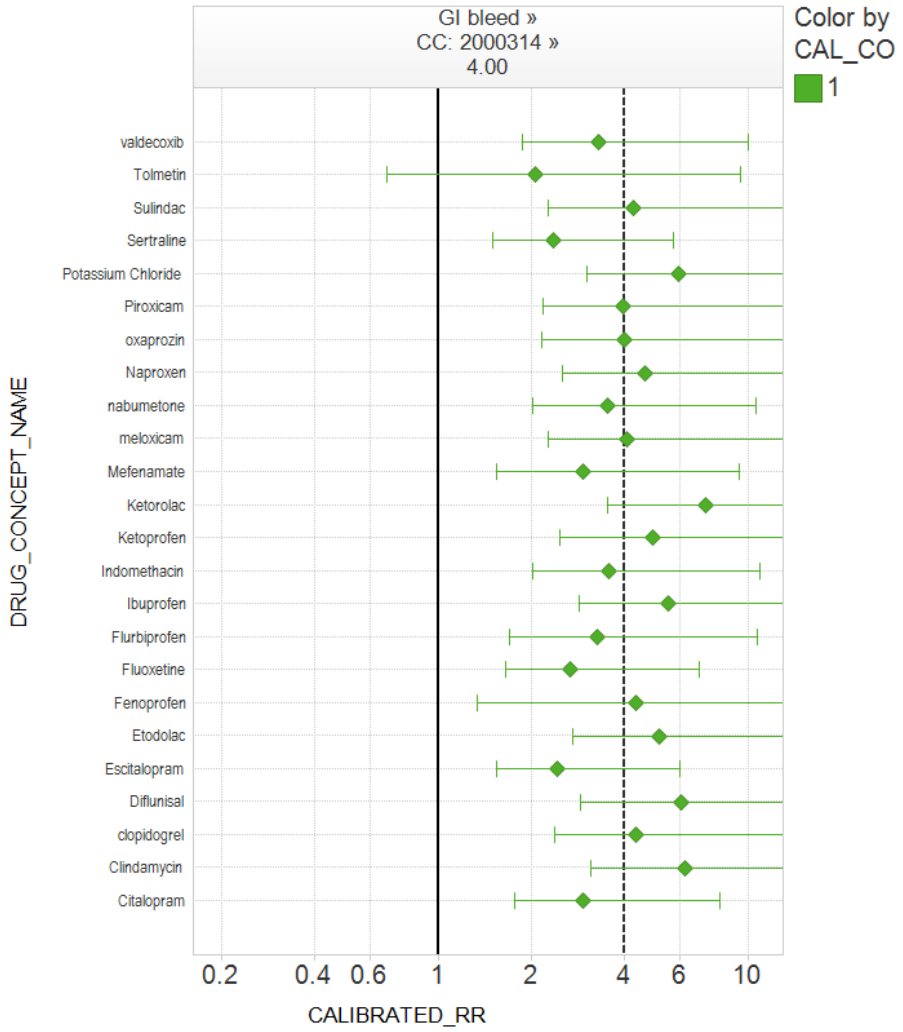
OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Applying case-control design and calibrating estimates of positive controls in simulated data, RR=4.00

Original estimated effects



Calibrated confidence intervals



Original coverage probability = **25%**

Calibrated coverage probability = **100%**

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Coverage probability by effect size

Recap

- Traditional interpretation of 95% confidence interval, that the CI covers the true effect size 95% of the time, may be misleading in the context of observational database studies
  - Coverage probability is much lower across all methods and all outcomes
  - Results were consistent across real data and simulated data
- Empirical adjustment of confidence intervals yields more robust coverage probabilities across most method-outcome scenarios
- Further research for developing heuristics to adjust confidence intervals could yield more reliable interpretation, but empirical approach would require confidence that simulated data adequately reflects the real world data

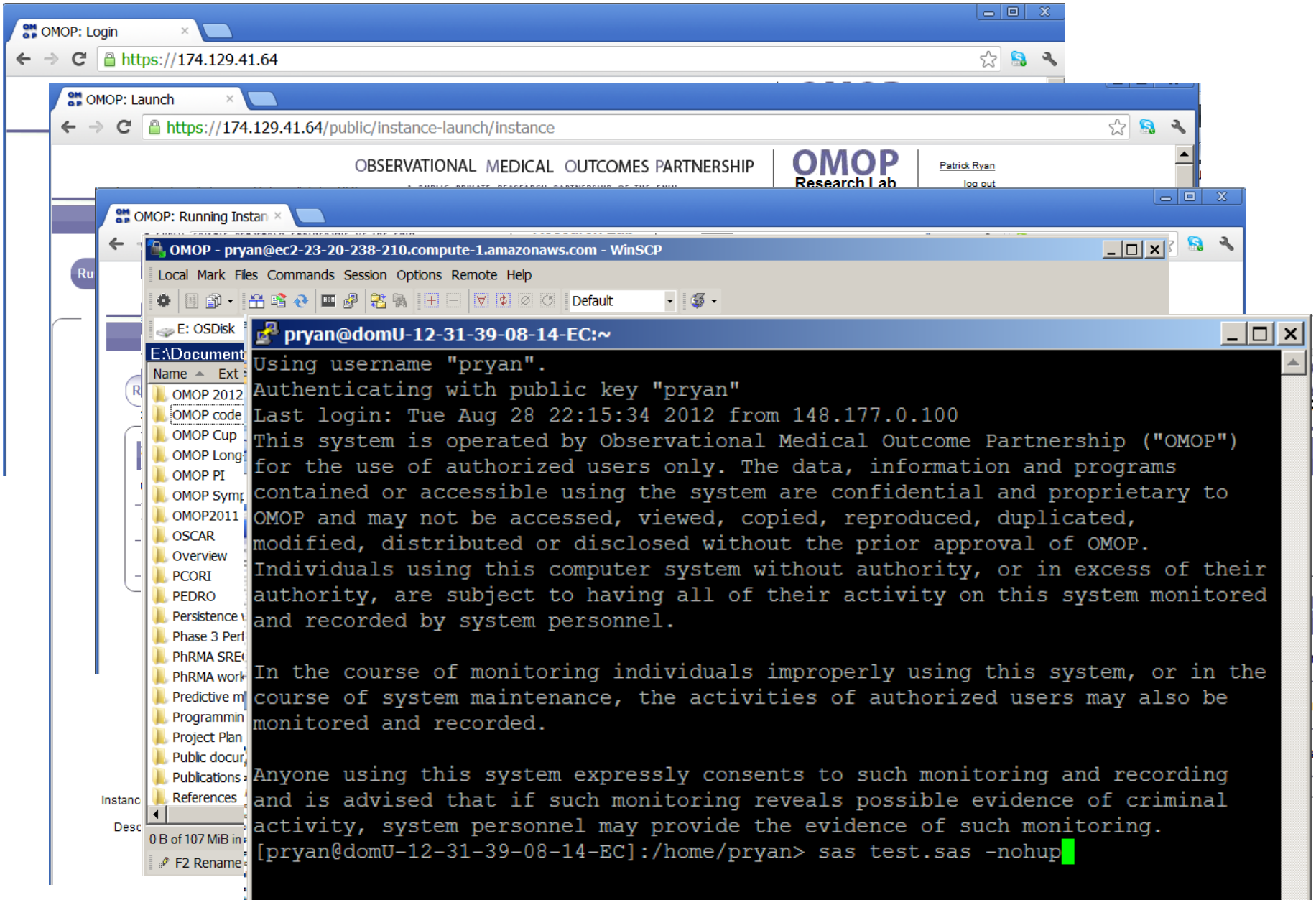
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Within OMOP, enough research areas for all backgrounds and interests

- Methods development
  - Estimate average treatment effects
  - Predict patient outcomes
- Methods implementation
  - Transform conceptual ideas into scalable computationally efficient applications
  - Contribute to open-source solutions within community of users to characterize, visualize, and analyze longitudinal observational data
- Methods evaluation
  - Measure and compare performance of different algorithms across an array of different databases, outcomes, exposures
  - Design and implement simulations to model real-world data and inject patterns of interest
  - Develop and apply metrics for empirical assessment of methods operating characteristics
- System optimization

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Working in the OMOP Research Lab



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Contact information

Patrick Ryan
Research Investigator
ryan@omop.org

Emily Welebob
Program Manager
welebob@omop.org

Thomas Scarnecchia
Executive Director
scarnecchia@omop.org

OMOP website: http://omop.fnih.org