ENGELBERG CENTER for Health Care Reform at BROOKINGS

Highlights from the Observational Medical Outcomes Partnership's (OMOP) Annual Symposium

Patrick Ryan, Director, Epidemiology Analytics, Janssen Research and Development, and Research Investigator, Observational Medical Outcomes Partnership

Martijn Schuemie, Assistant Professor, Erasmus University Medical Center, Rotterdam, Netherlands, and Research Investigator, Observational Medical Outcomes Partnership

August 8, 2012

Brookings Roundtable on Active Medical Product Surveillance

Some Initial Housekeeping

- To minimize feedback, please confirm that the microphone on your telephone is muted.
- To mute your phone, press the mute button or '*6'. (To unmute, press '*7' as well.)
- **There will be opportunities for questions and discussion at the end of today's presentations. Please use the chat box at the right side of your screen to submit your questions into the queue at any point and we will call upon you to state your question.**
- We will open up the lines for questions from those participating only by phone at the end of each Q&A session.
- Call the WebEx help line at 1-866-229-3239 with technical problems.

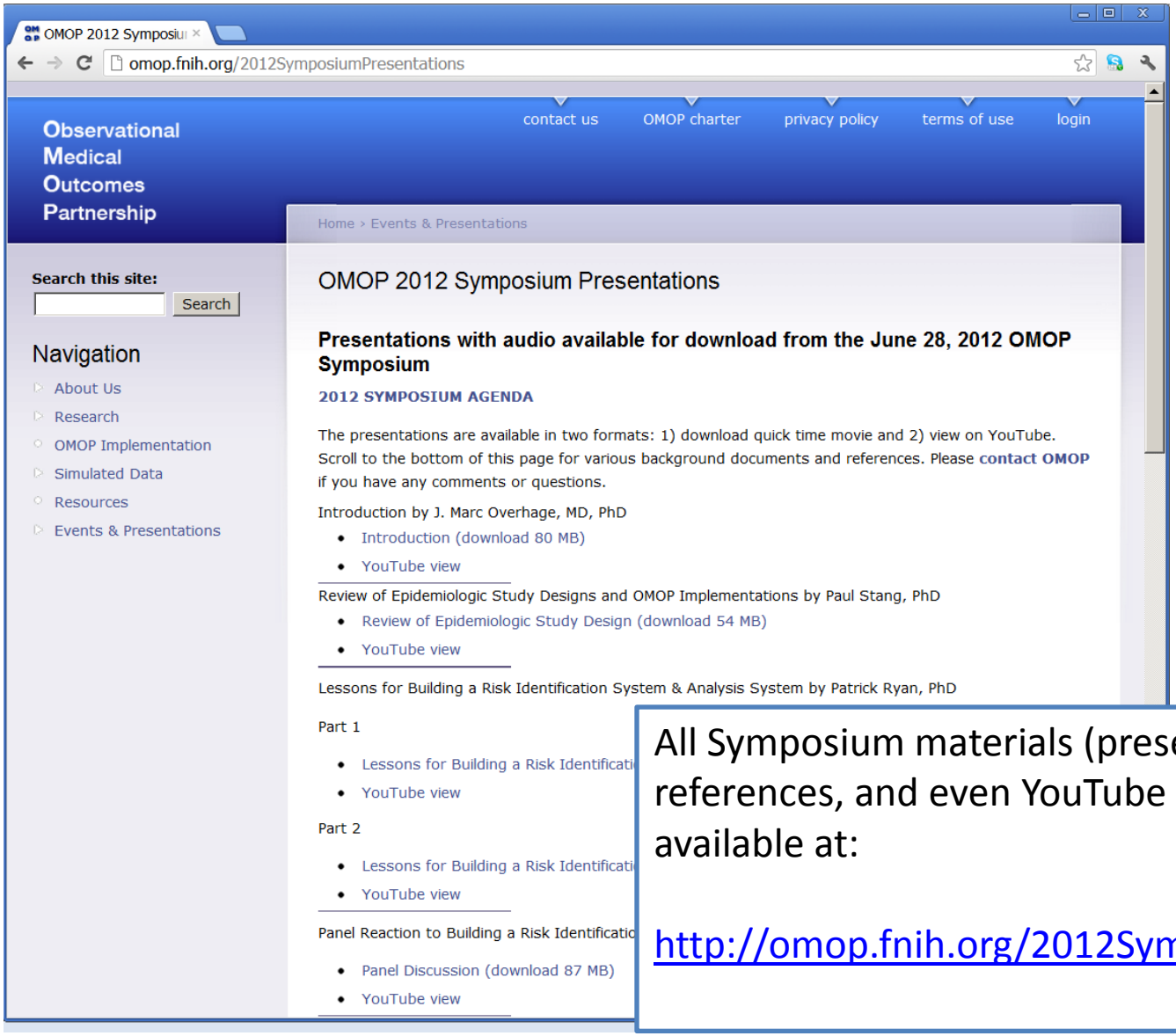
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Whirlwind tour through the 2012 OMOP Symposium

Patrick Ryan, Martijn Schuemie
on behalf of the OMOP research team

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

A brief summary from a long journey



All Symposium materials (presentations, handouts, references, and even YouTube videos) are all publicly available at:

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

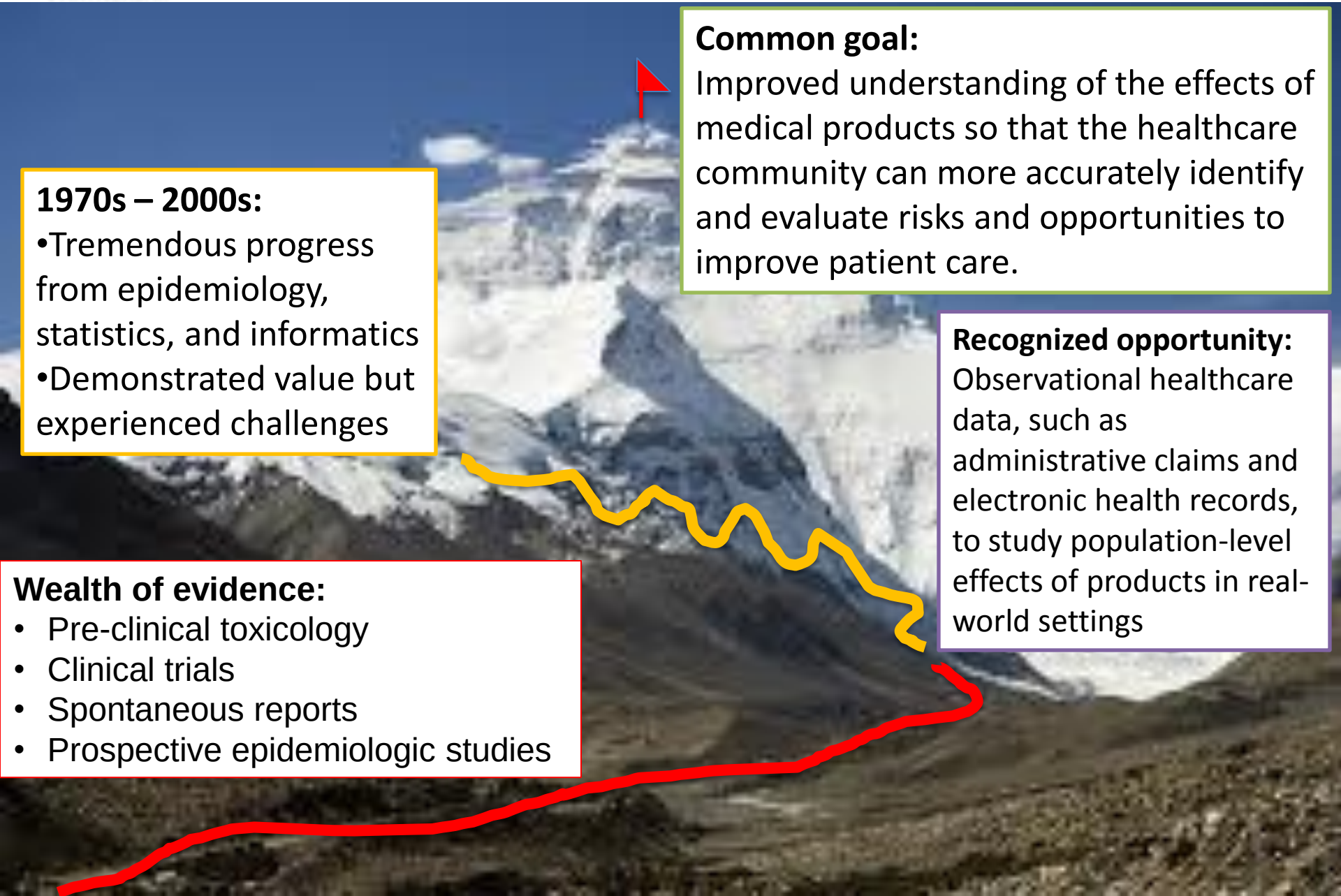
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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.

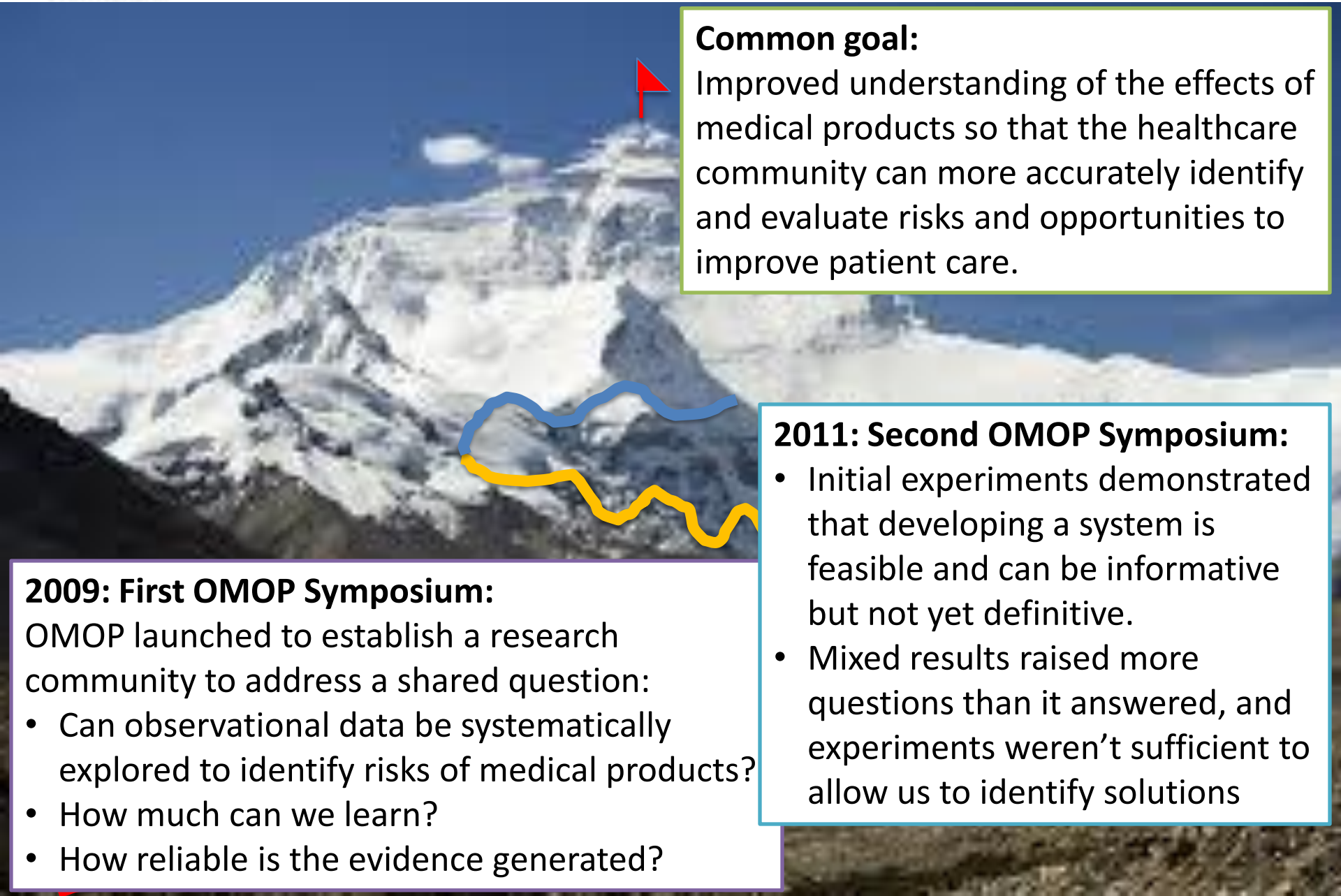
1970s – 2000s:
- Tremendous progress from epidemiology, statistics, and informatics
- Demonstrated value but experienced challenges

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:

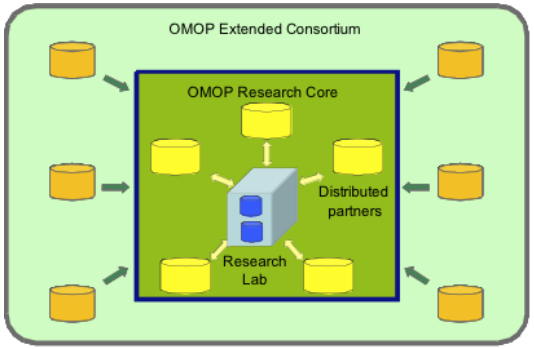
- Can observational data be systematically explored to identify risks of medical products?
- How much can we learn?
- How reliable is the evidence generated?

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

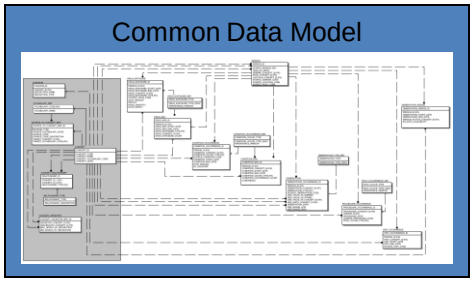
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

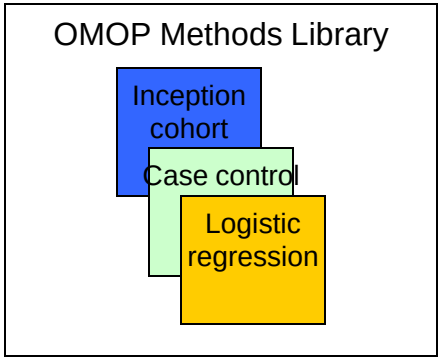
OMOP 2010/2011 Research Experiment



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

- Open-source
- Standards-based





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

Drug

Outcome	ACE Inhibitors	Amphotericin B	Antibiotics: erythromycins, sulfonamides, tetracyclines	Antiepileptics: carbamazepine, phenytoin	Benzodiazepines	Beta blockers	Bisphosphonates: alendronate	Tricyclic antidepressants	Typical antipsychotics	Warfarin
Angioedema										
Aplastic Anemia										
Acute Liver Injury										
Bleeding										
Hip Fracture										
Hospitalization										
Myocardial Infarction										
Mortality after MI										
Renal Failure										
GI Ulcer Hospitalization										

Positives: 9
Negatives: 44

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 2011/2012 Research Agenda

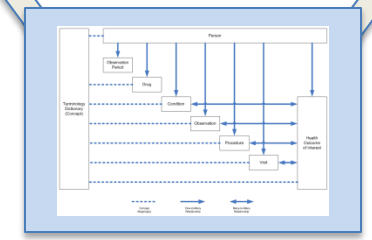
Drug-outcome pairs

	Positives	Negatives
Total	165	234
Myocardial Infarction	36	66
Upper GI Bleed	24	67
Acute Liver Injury	81	37
Acute Renal Failure	24	64

+ EU-ADR replication

- Improve HOI definitions
- Explore false positives

- Evaluate study design decisions (EDDIE)



Methods development

Methods enhancements
- *Multivariate self-controlled case series*

Increased parameterization
- *Case-control, new user cohort designs*

Application of existing tools
- *ICTPD, OS, LGPS, DP*

- Expand CDM for additional use cases

Observational data

Real-world performance:

+ OMOP Distributed Partners
+ EU-ADR network

Simulated data:

- Strength (RR)
- Type (timing)

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Ground truth for OMOP 2011/2012 experiments

isoniazid

fluticasone

indomethacin

clindamycin

	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

ibuprofen

loratadine

sertraline

pioglitazone

Criteria for positive controls:

- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label
- Drug listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"
- Literature review identified no powered studies with refuting evidence of effect

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 powered studies with evidence of potential positive association

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Exploring isoniazid and acute liver injury

CMAJ RESEARCH

Adverse events associated with treatment of latent tuberculosis in the general population

Benjamin M. Smith MD, Kevin Schwartzman MD MPH, Gillian Bartlett PhD, Dick Menzies MD MSc

ABSTRACT

Background: Guidelines recommend treatment of latent tuberculosis in patients at increased risk for active tuberculosis. Studies investigating the association of therapy with serious adverse events have not included the entire treated population nor accounted for comorbidities or occurrence of similar events in the untreated general population. Our objective was to estimate the risk of adverse events requiring hospital admission that were associated with therapy for latent tuberculosis infection in the general population.

Methods: Using administrative health data from the province of Quebec, we created a historical cohort of all residents dispensed therapy for latent tuberculosis between 1998 and 2003. Each patient was matched on age, sex and postal region with two untreated residents. The observation period was 18 months (from 6 months before to 12 months after initiation of therapy). The primary outcome was hospital admission for therapy-associated adverse events.

Results: During the period of observation, therapy for latent tuberculosis was dispensed to 9145 residents, of whom 95% started isoniazid and 5% started rifampin. Pretreatment comorbid illness was significantly more common among patients receiving such therapy compared with the matched untreated cohort. Of all patients dispensed therapy, 45 (0.5%) were admitted to hospital for a hepatic event compared with 15 (0.1%) of the untreated patients. For people over age 65 years, the odds of hospital admission for a hepatic event among patients treated for latent tuberculosis infection was significantly greater than among matched untreated people after adjustment for comorbidities (odds ratio [OR] 6.4, 95% CI 2.2–18.3). Excluding patients with comorbid illness, there were two excess admissions to hospital for hepatic events per 100 patients initiating therapy compared with the rate among untreated people over 65 years (95% CI 0.1–3.87).

Interpretation: The risk of adverse events requiring hospital admission increased significantly among patients over 65 years receiving treatment for latent tuberculosis infection. The decision to treat latent tuberculosis infection in elderly patients should be made after careful consideration of risks and benefits.

Competing interests:
None declared.

This article has been peer reviewed.

Correspondence to:
Dr. Dick Menzies;
dick.menzies@mcgill.ca

CMAJ 2011. DOI:10.1503/cmaj.091824

CMAJ, February 22, 2011, 183(3)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Smith et al. 2011 study design and results

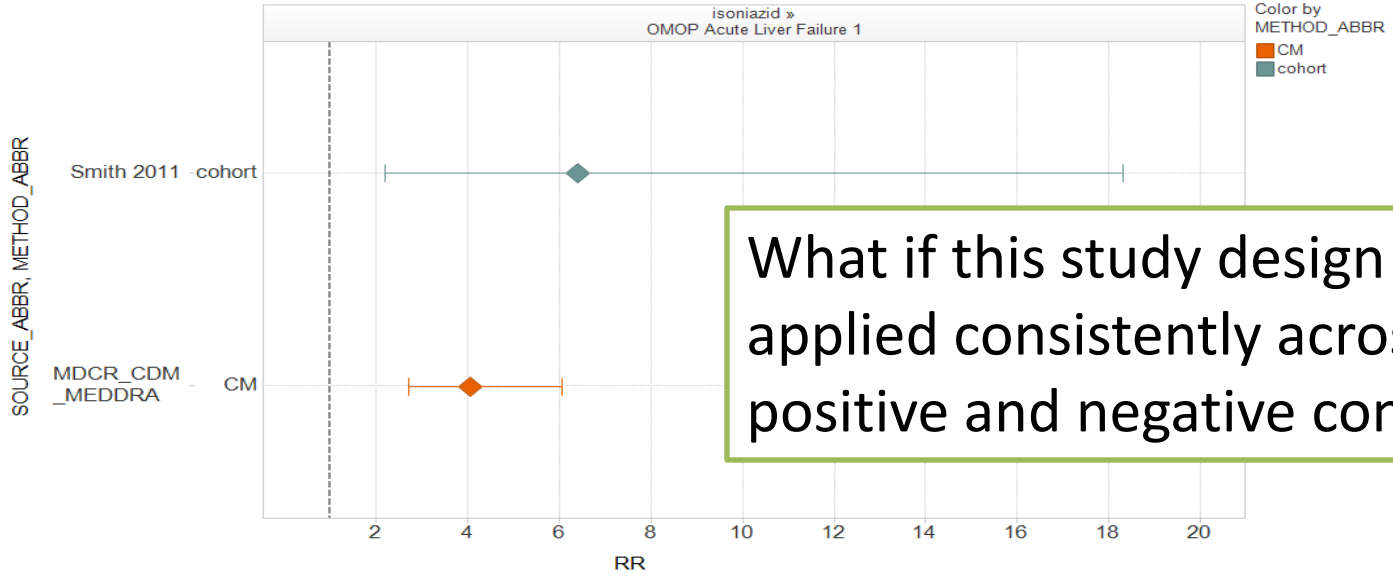
- Data source: Administrative claims from health insurance board of Quebec
- Study design: Cohort
- Exposure: all patients dispensed >=30d of therapy, 180d washout
- Unexposed cohort: 2 patients per exposed, matched by age, gender, and region, with no tuberculosis therapy
- Time-at-risk: Length of exposure + 60 days
- Events: Incident hospital admission for noninfectious or toxic hepatitis
- "Event ratio" estimated with conditional logistic regression
- Covariates: prior hospitalization, Charlson score, comorbidities

Table 2: Event rates and odds ratios for outcomes of interest, by cohort

	Crude event rate, events/total (rate per 100 patients)		Event ratio, cohort treated for LTBI v. untreated cohort (95% CI)		
Outcome; age, yr	LTBI therapy cohort	Untreated cohort*	Crude OR†	Adjusted OR‡	Adjusted OR§
Hospital admission for hepatic event of interest§					
Total	45/9145 (0.5)	15/18 290 (0.1)	6.5 (3.8–11.1)	3.7 (2.0–6.9)	2.7 (1.3–5.6)
≤ 35	5/4523 (0.1)	1/9046 (0.0)	10.0 (1.2–85.6)	NC	NC
36–50	8/2533 (0.3)	7/5066 (0.1)	2.6 (1.0–6.9)	2.0 (0.6–6.9)	1.5 (0.4–5.6)
51–65	10/1232 (0.8)	4/2464 (0.2)	7.0 (2.3–21.3)	2.9 (0.7–13.0)	2.6 (0.4–16.0)
> 65	22/857 (2.6)	3/1714 (0.2)	10.8 (4.2–28.0)	6.4 (2.2–18.3)	3.2 (0.9–11.7)

Revisiting the isoniazid – acute liver injury example

- Data source: MarketScan Medicare Beneficiaries (MDCR)
- Study design: Cohort
- Exposure: all patients dispensed new use of isoniazid, 180d washout
- Unexposed cohort: Patient with indicated diagnosis (e.g. pulmonary tuberculosis) but no exposure to isoniazid; negative control drug referents
- Time-at-risk: Length of exposure + 30 days, censored at incident events
- Covariates: age, sex, index year, Charlson score, number of prior visits, all prior medications, all comorbidities, all priority procedures
- "Odds ratio" estimated through propensity score stratification (20 strata)

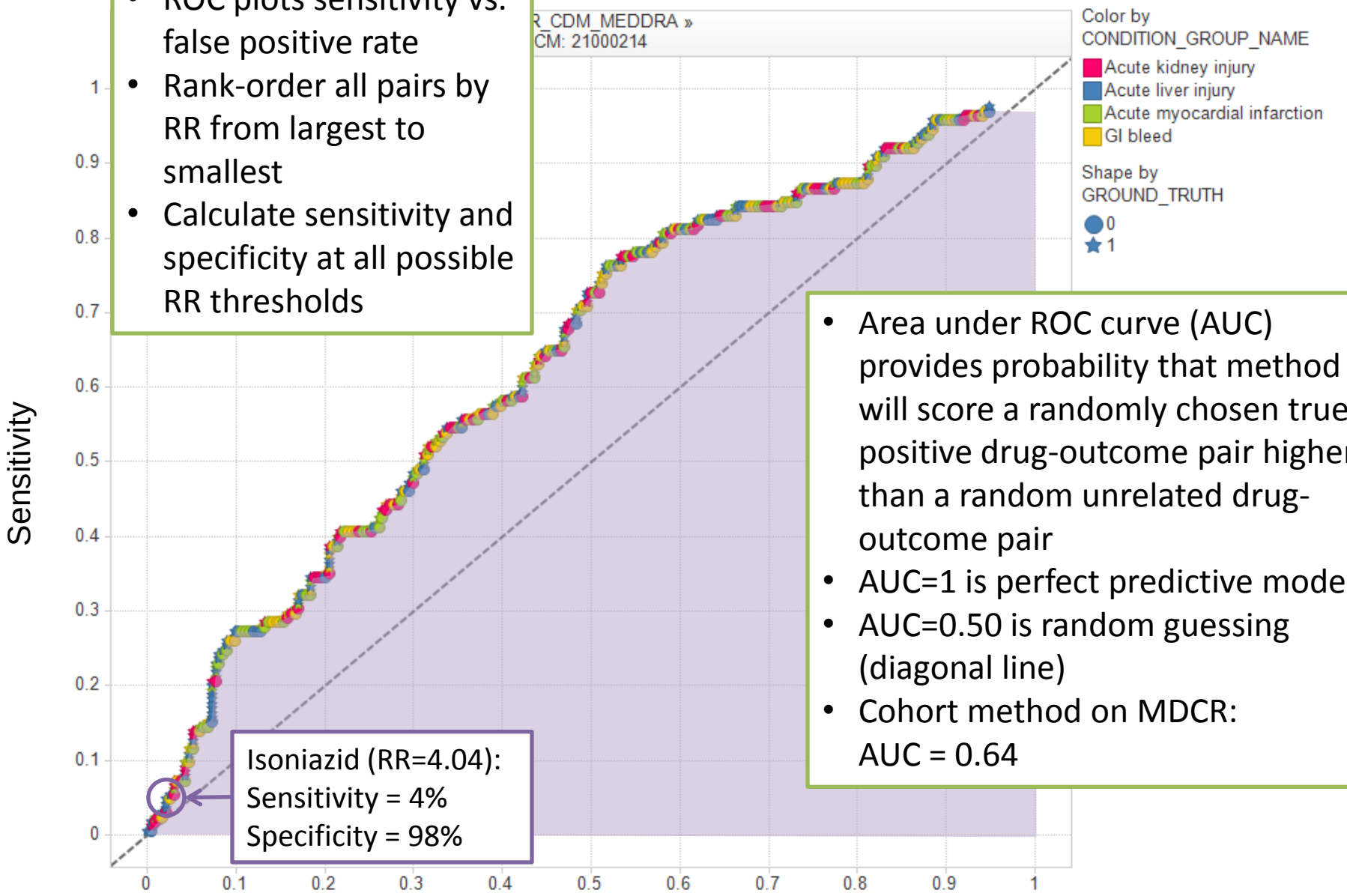


What if this study design were applied consistently across all the positive and negative controls?

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

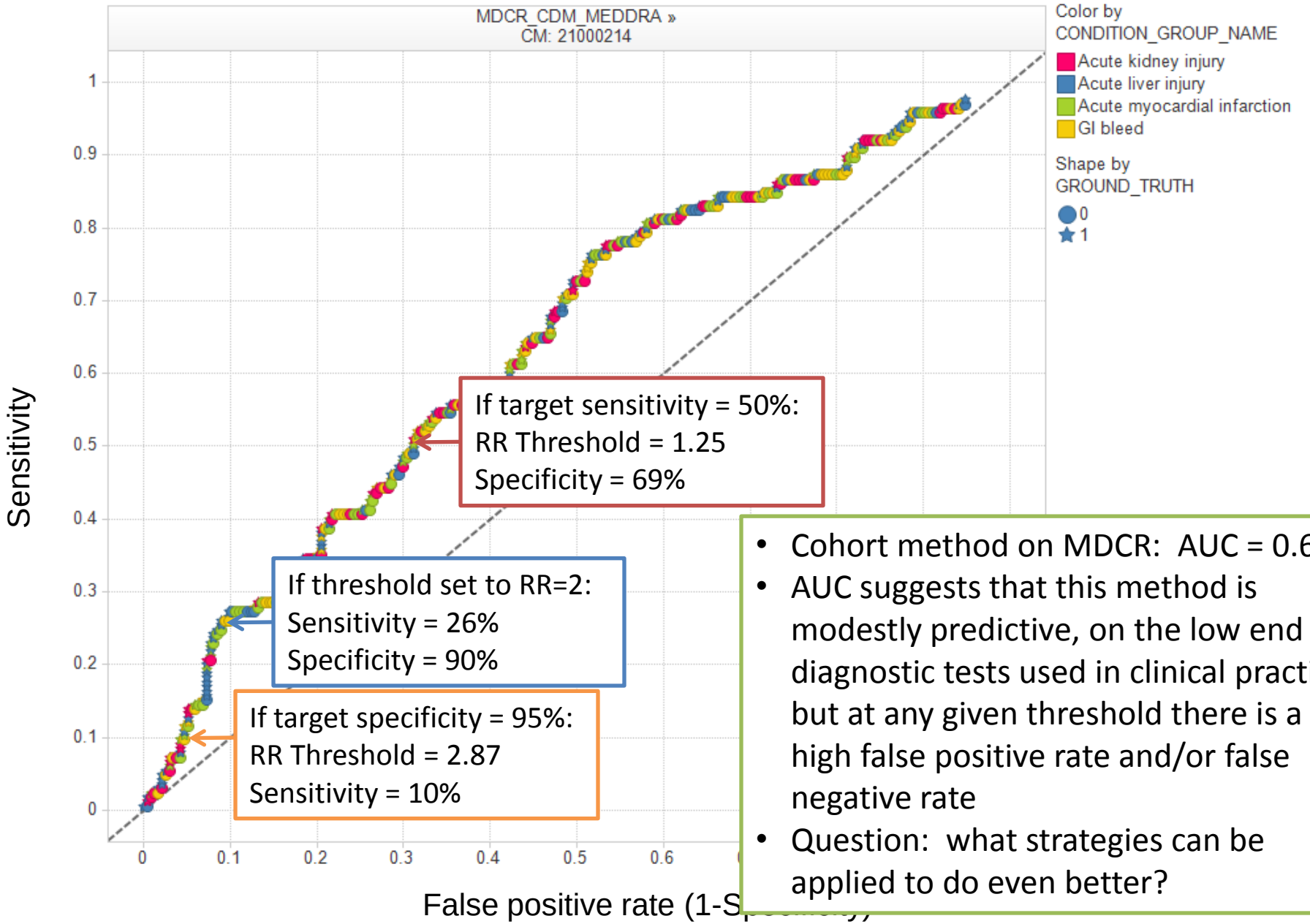


Isoniazid (RR=4.04):
Sensitivity = 4%
Specificity = 98%

- 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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Setting thresholds from an ROC curve



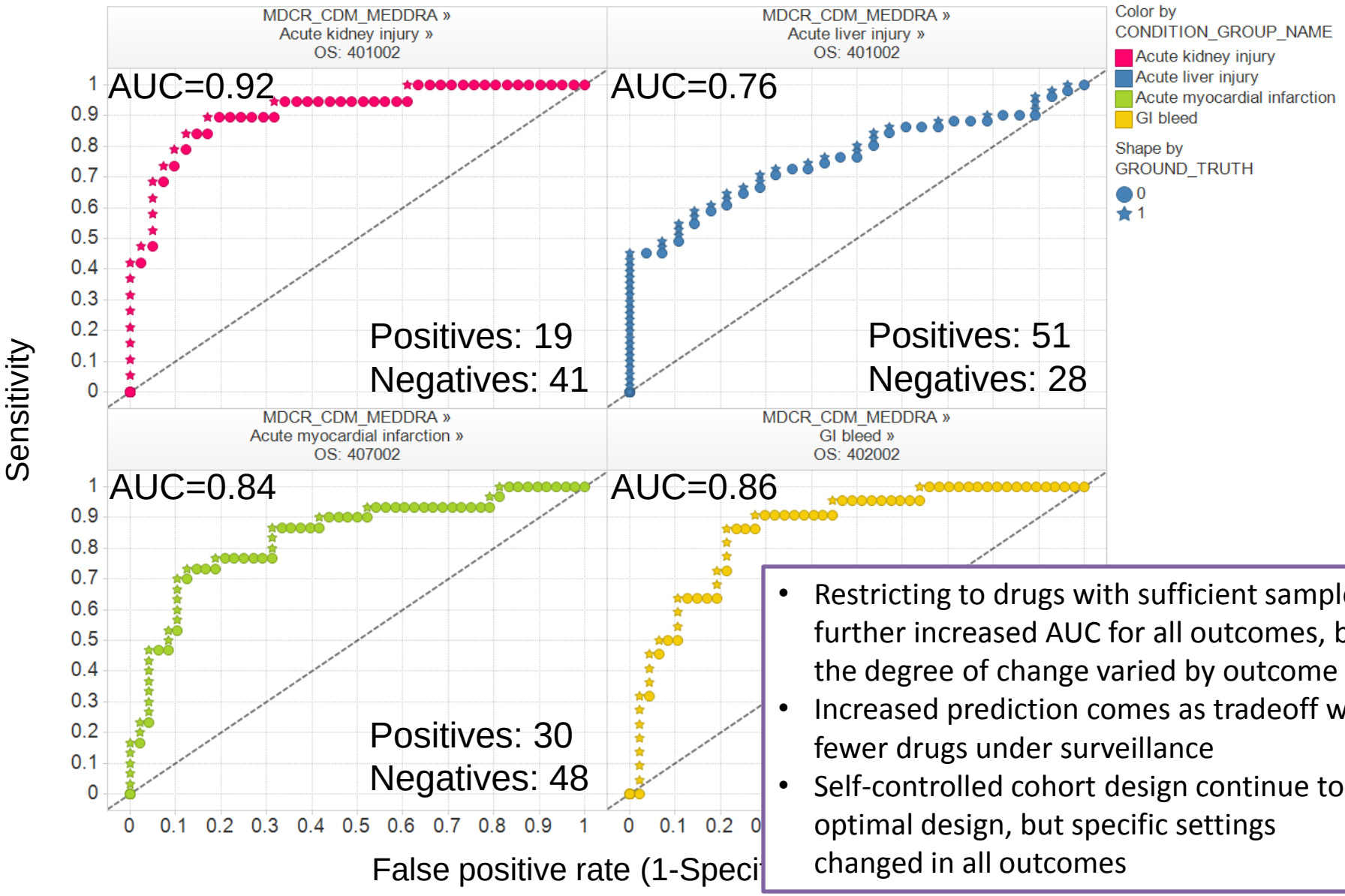
- Cohort method on MDCR: AUC = 0.64
- AUC suggests that this method is modestly predictive, on the low end of diagnostic tests used in clinical practice, but at any given threshold there is a high false positive rate and/or false negative rate
- Question: what strategies can be applied to do even better?

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

Optimal methods (AUC) by outcome and data source

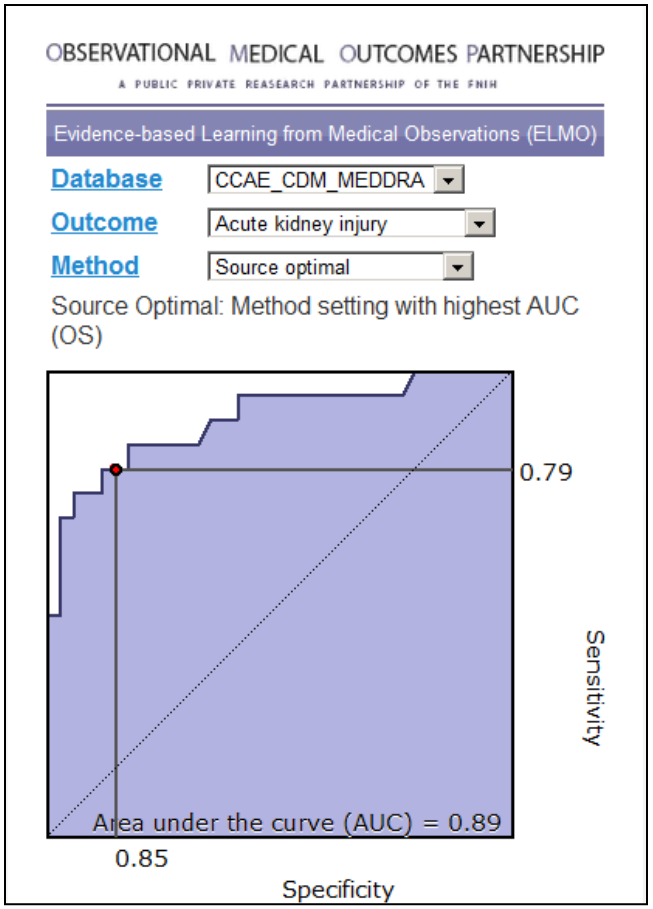
Data source	Acute kidney injury	Acute liver injury	Acute myocardial infarction	GI bleed
MDCR	OS: 401002 (0.92)	OS: 401002 (0.76)	OS: 407002 (0.84)	OS: 402002 (0.86)
CCAE	OS: 404002 (0.89)	OS: 403002 (0.79)	OS: 408013 (0.85)	SCCS: 1931010 (0.82)
MDCD	OS: 408013 (0.82)	OS: 409013 (0.77)	OS: 407004 (0.80)	OS: 401004 (0.87)
MSLR	SCCS: 1939009 (1.00)	OS: 406002 (0.84)	OS: 403002 (0.80)	OS: 403002 (0.83)
GE	SCCS: 1949010 (0.94)	OS: 409002 (0.77)	ICTPD: 3016001 (0.89)	ICTPD: 3034001 (0.89)

- Self-controlled designs are optimal across all outcomes and all sources, but the specific settings are different in each scenario
- AUC > 0.80 in all sources for acute kidney injury, acute MI, and GI bleed
- Acute liver injury has consistently lower predictive accuracy
- No evidence that any data source is consistently better or worse than others

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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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

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

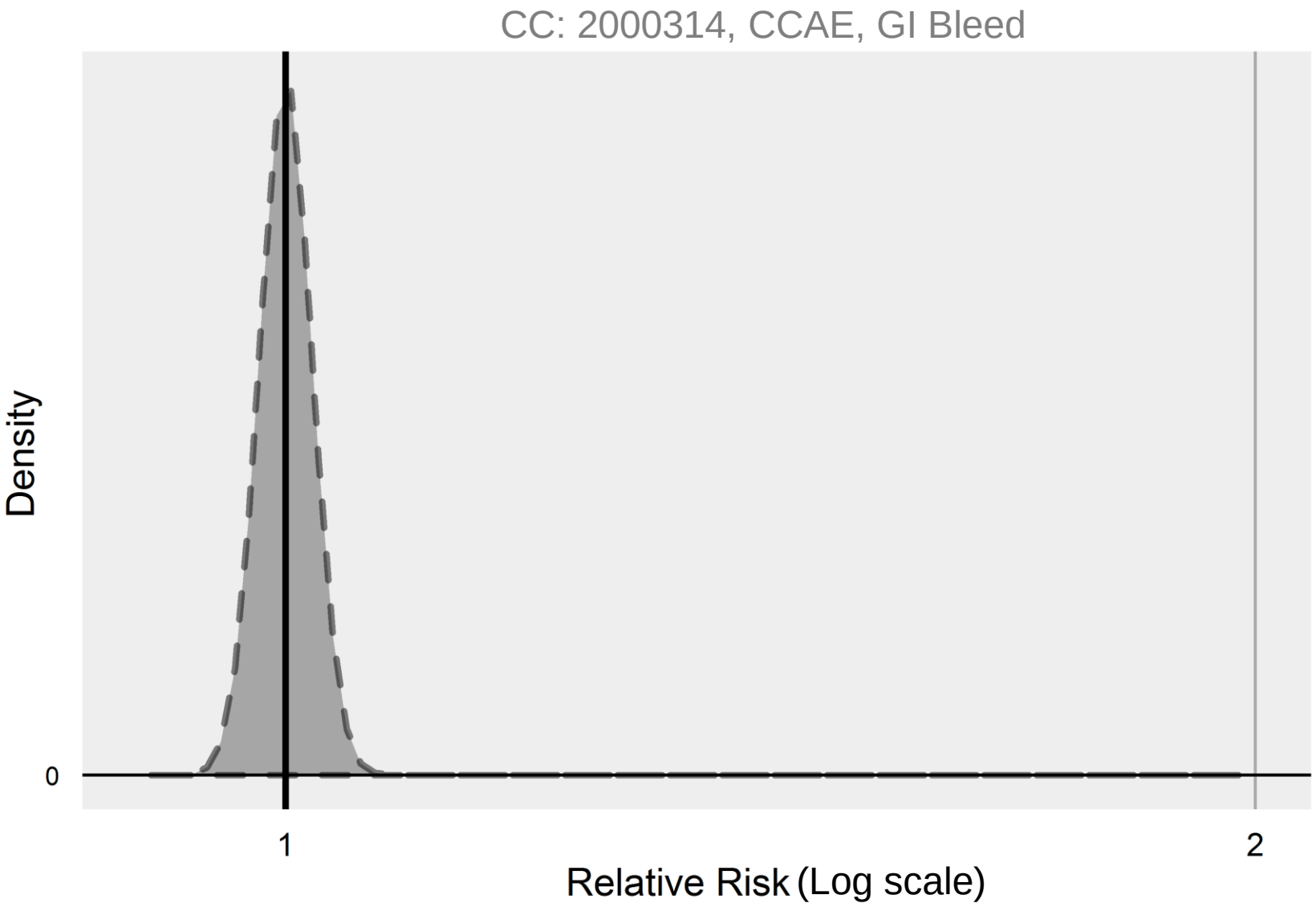
OMOP, 2012 (CC: 2000314, CCAE, GI Bleed)

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

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

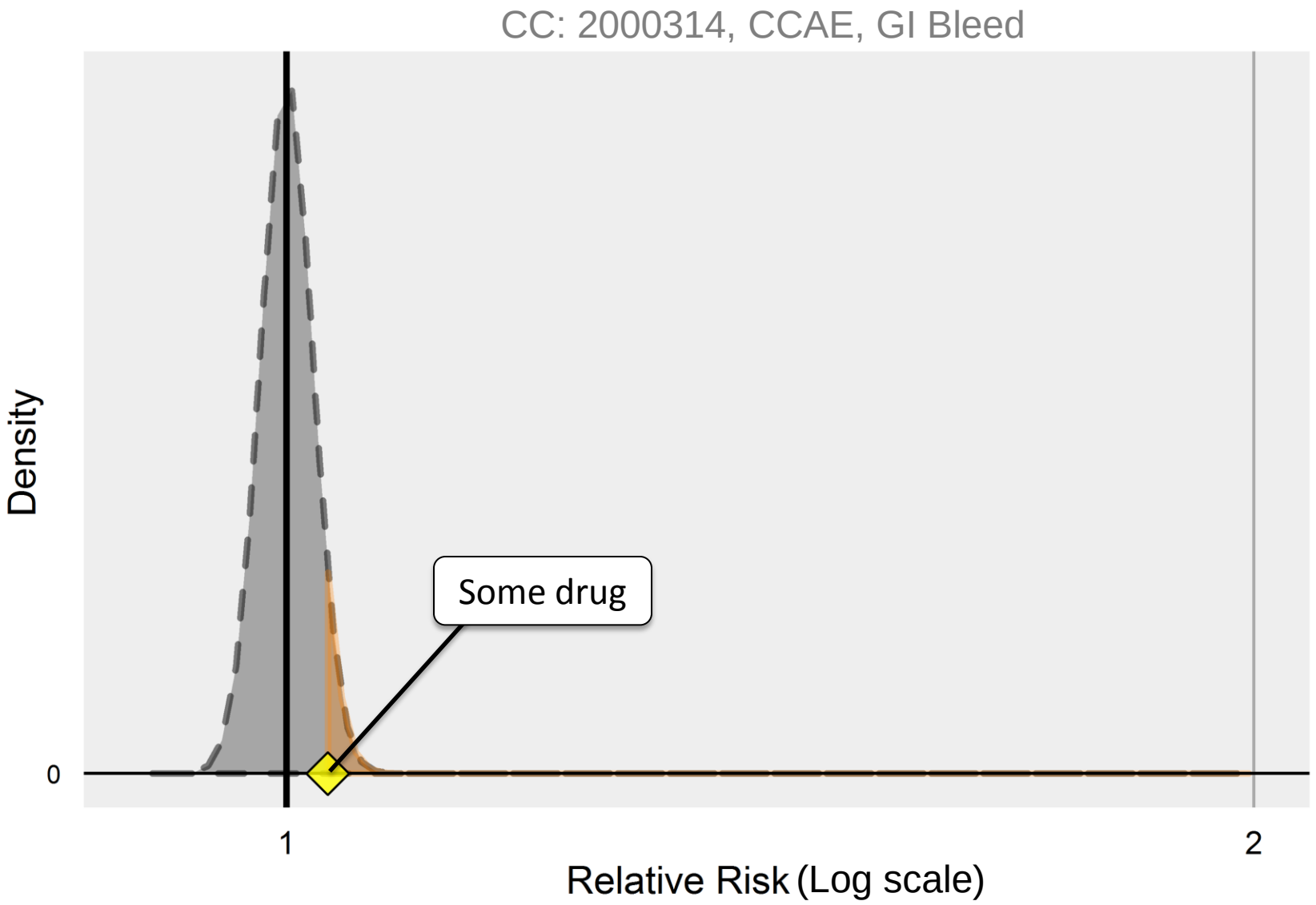
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Null distribution

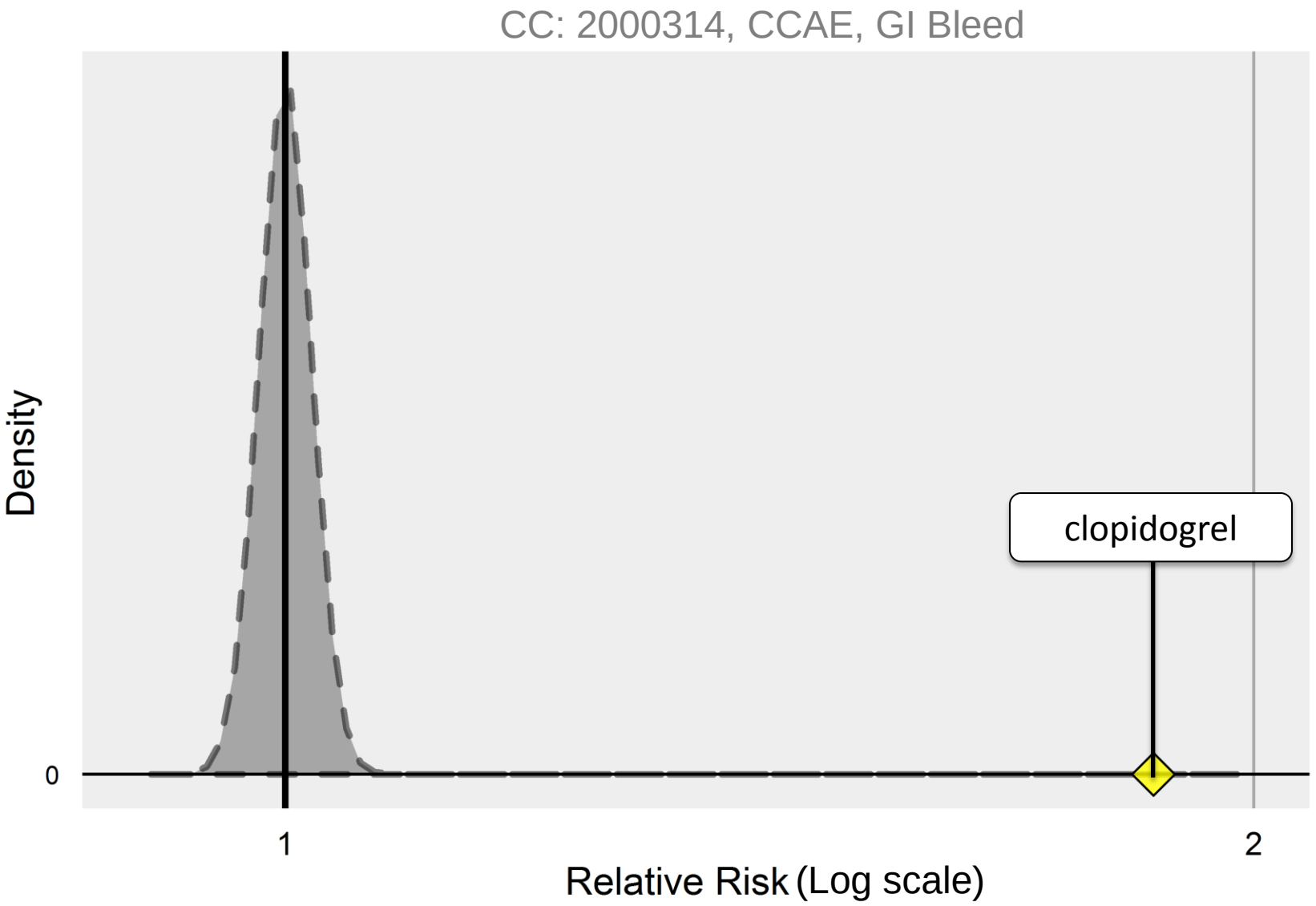


OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Null distribution



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

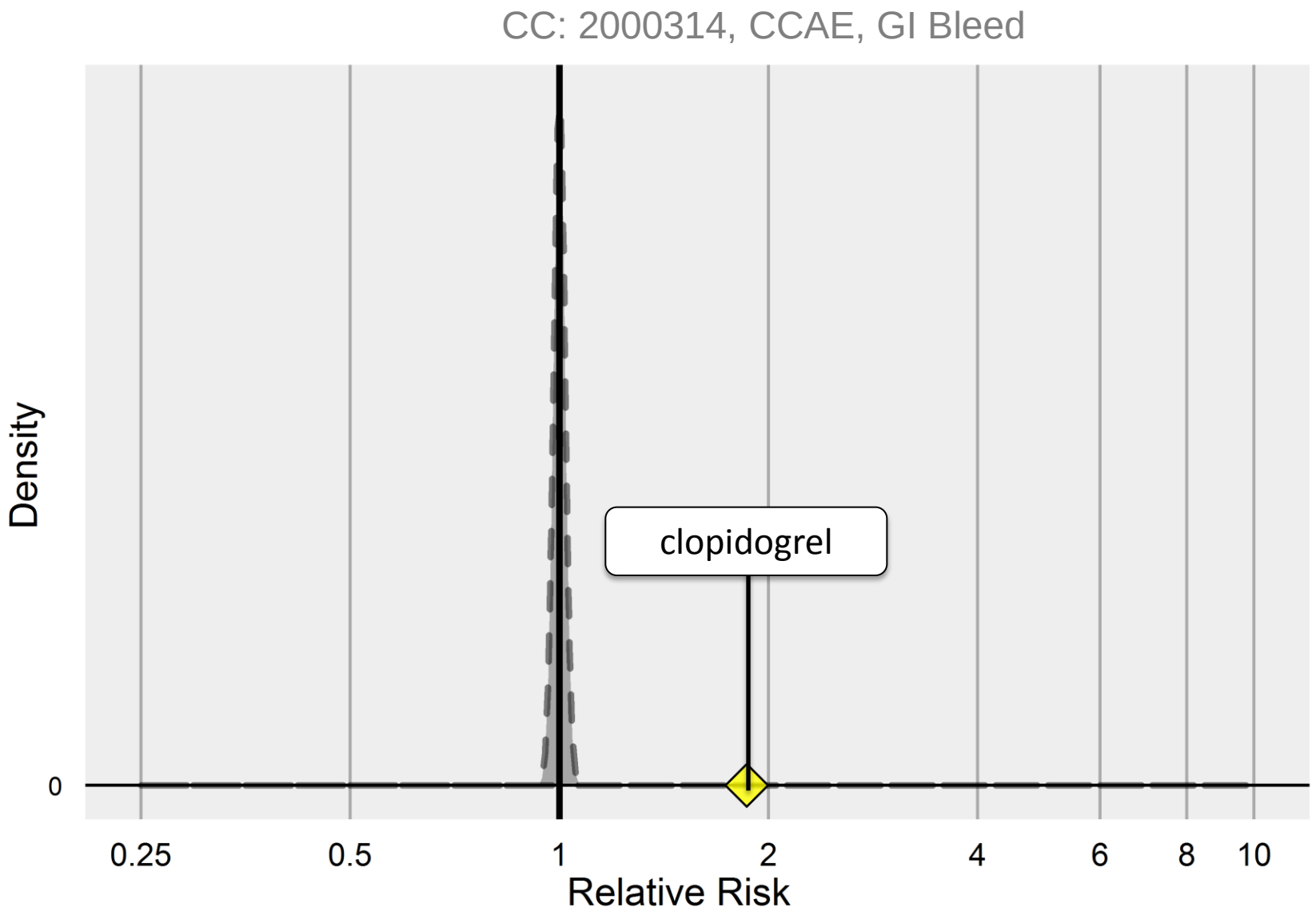
	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

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

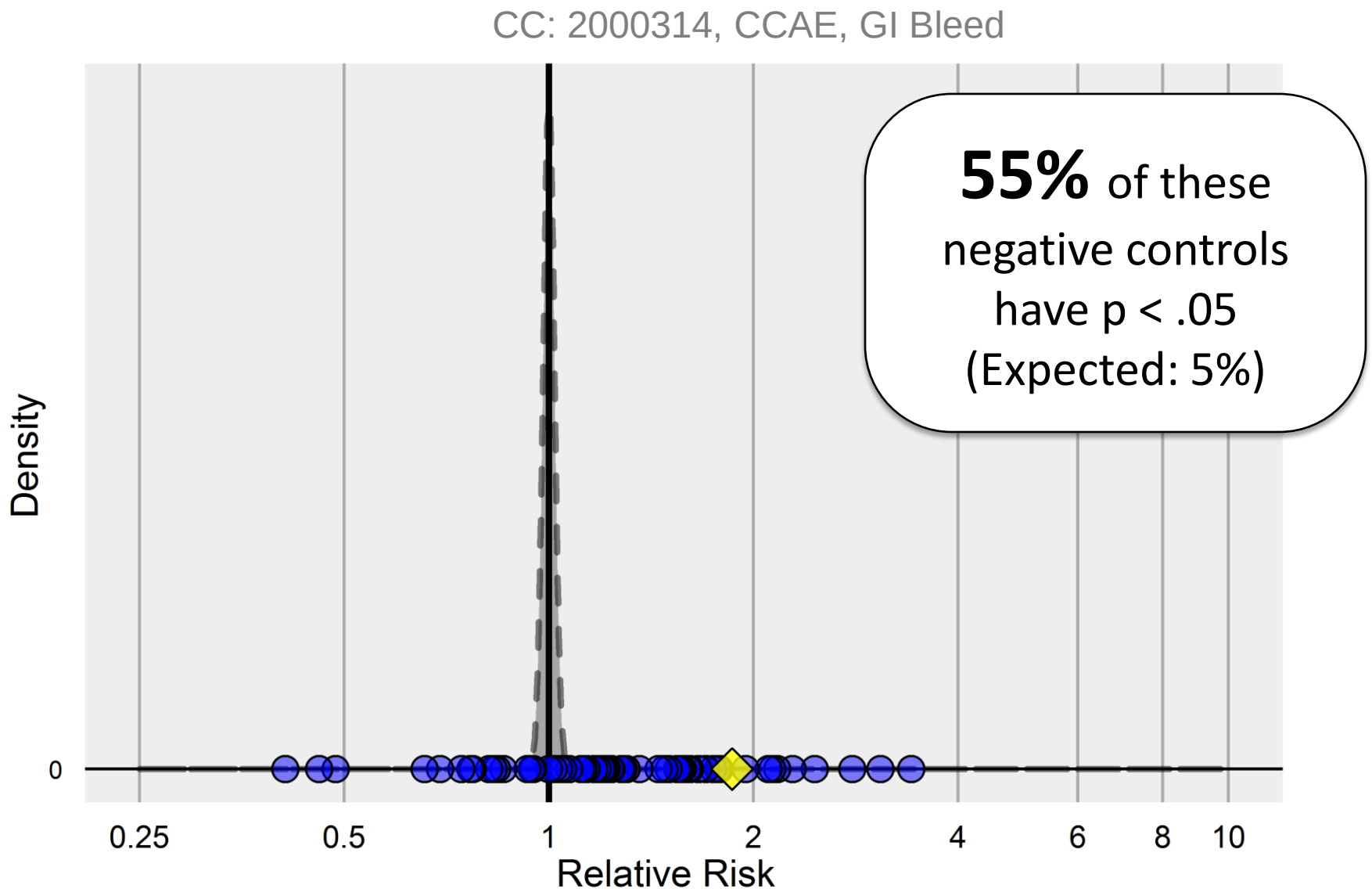
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution



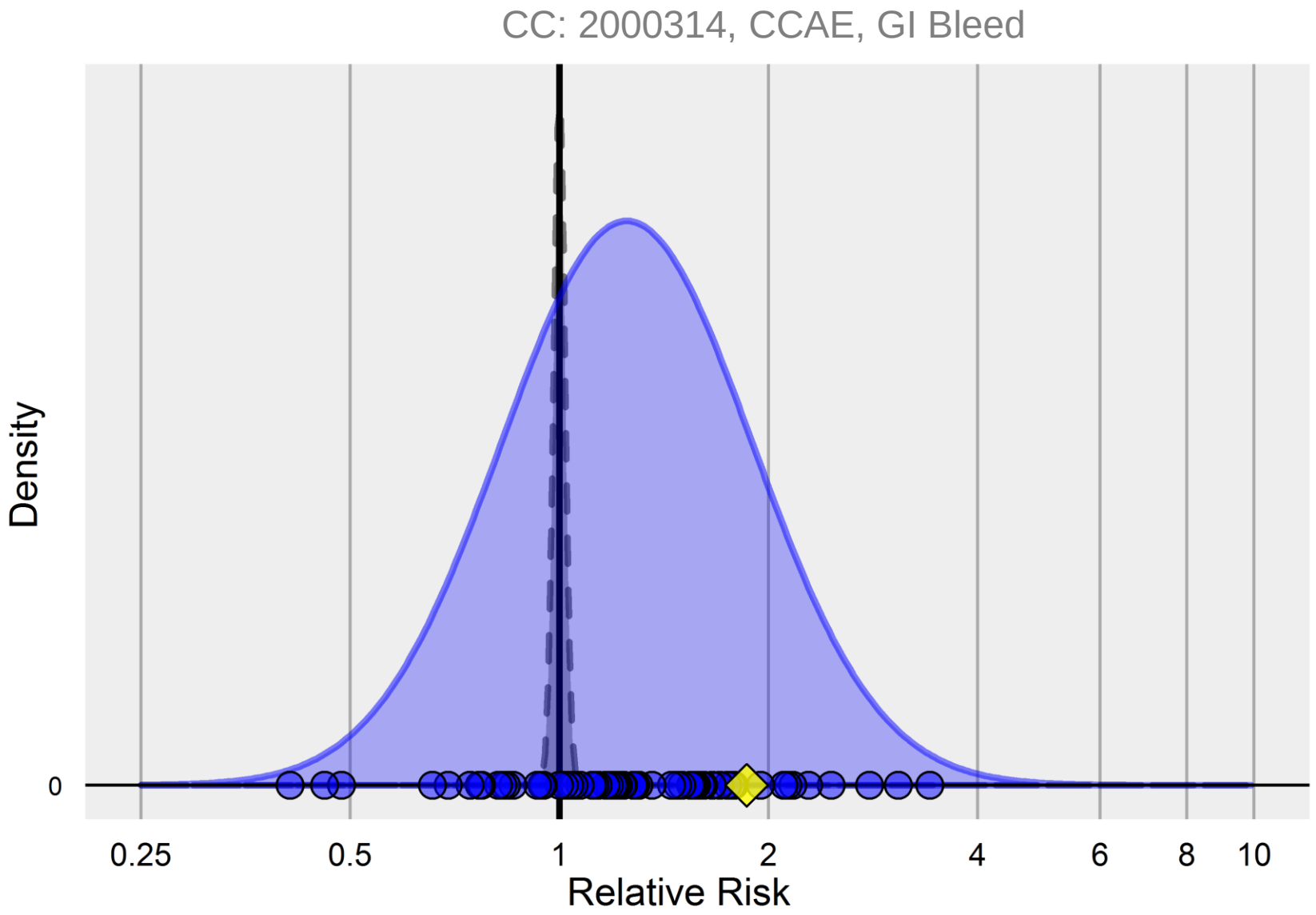
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution



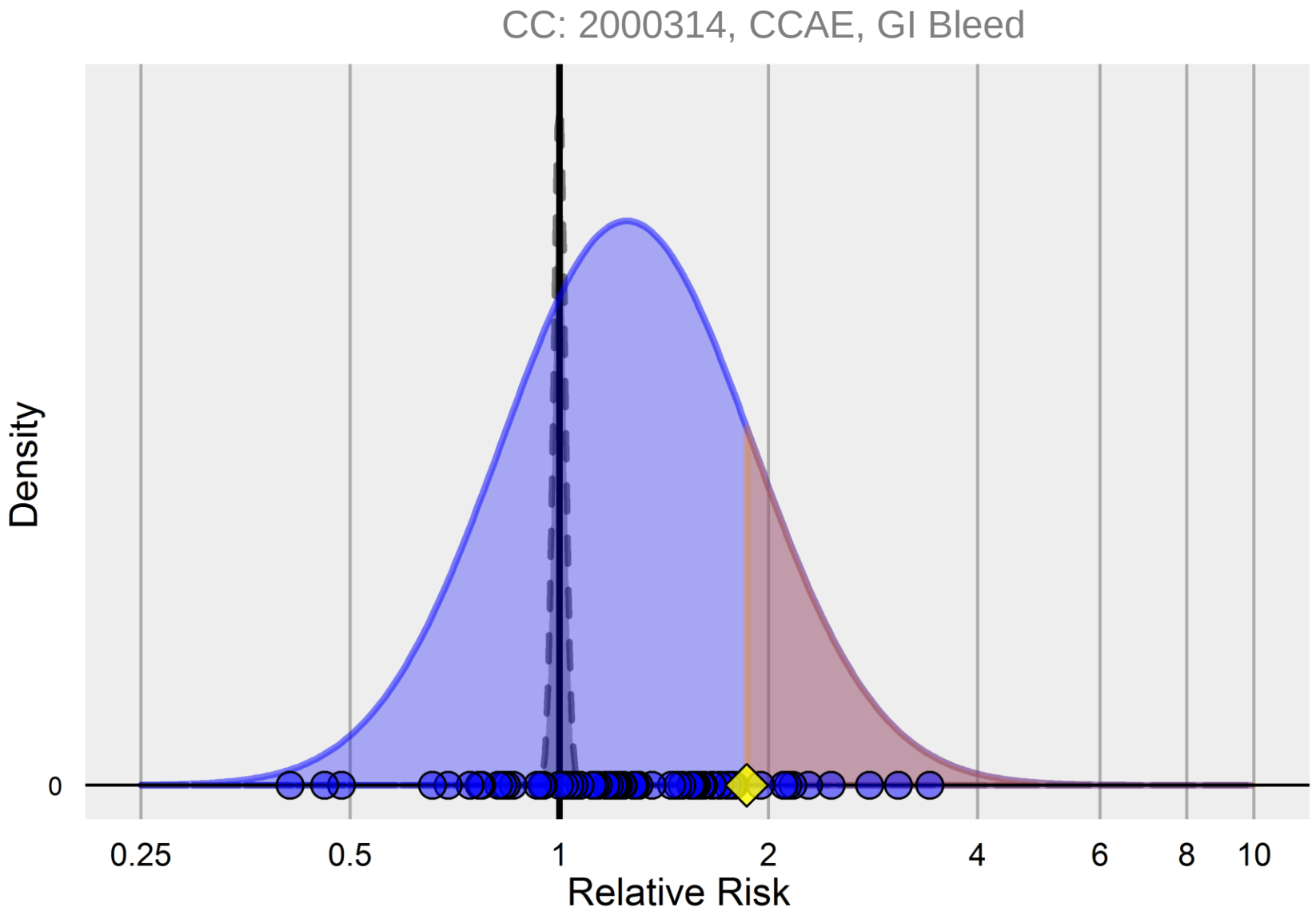
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution



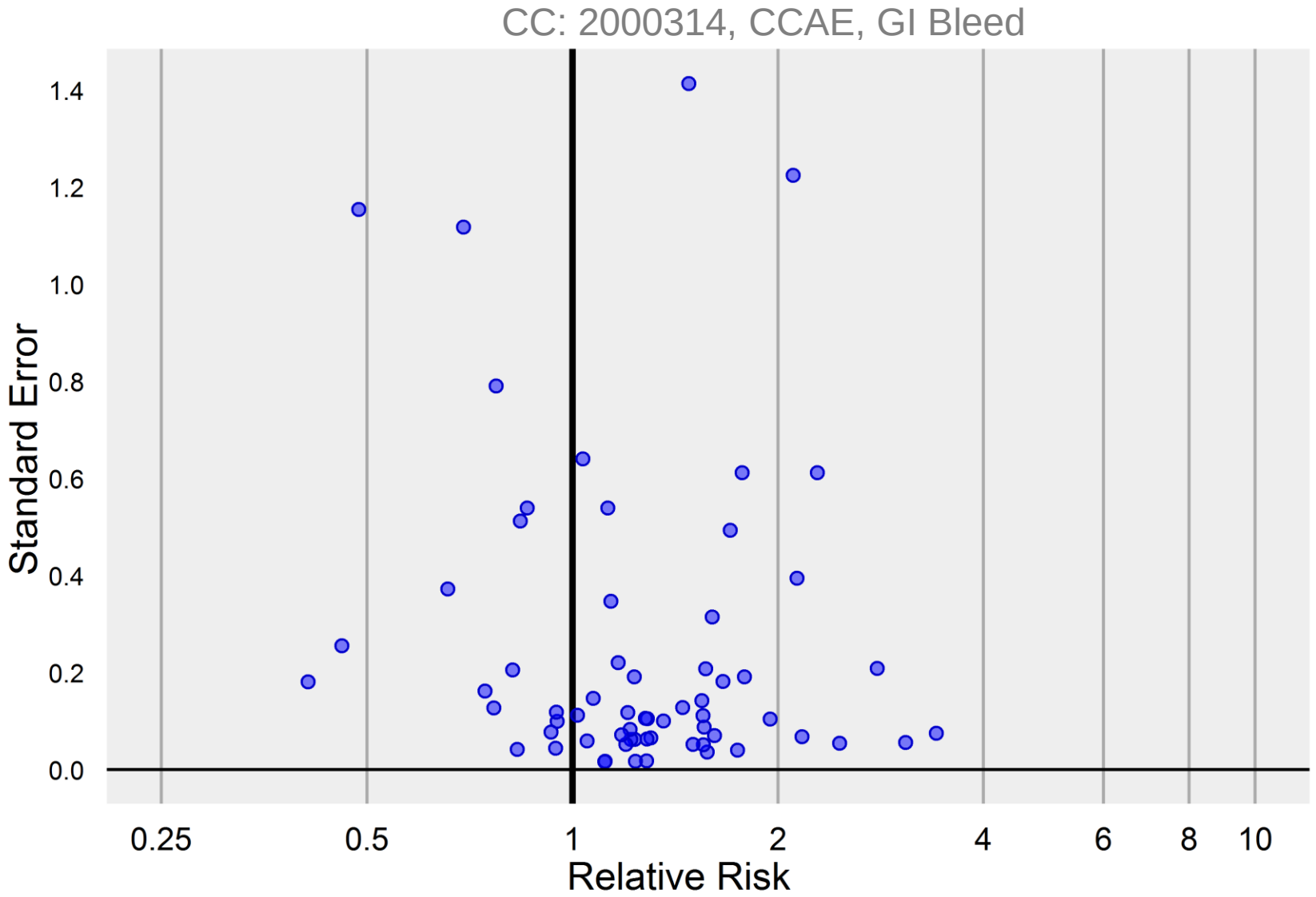
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution



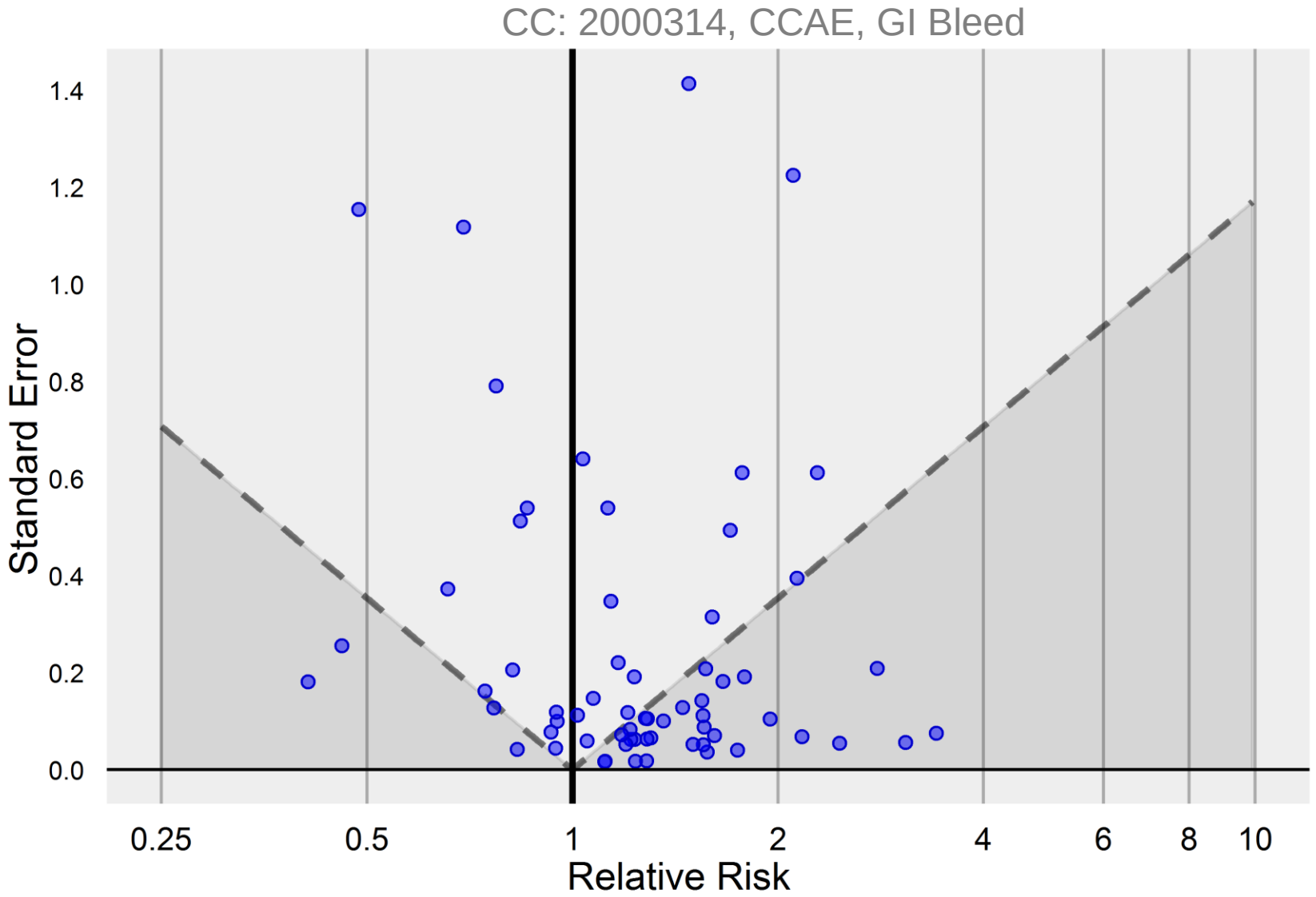
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot



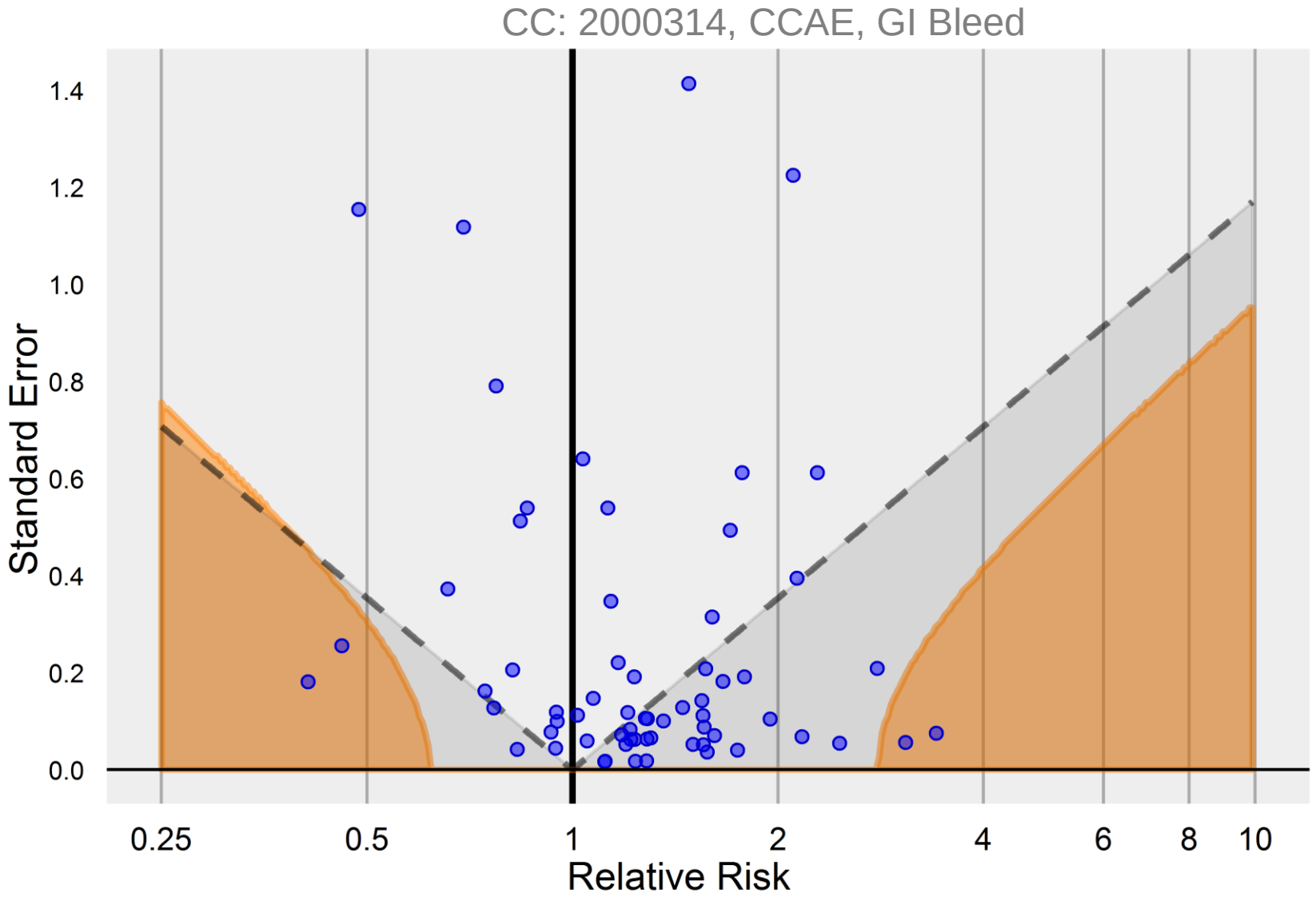
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

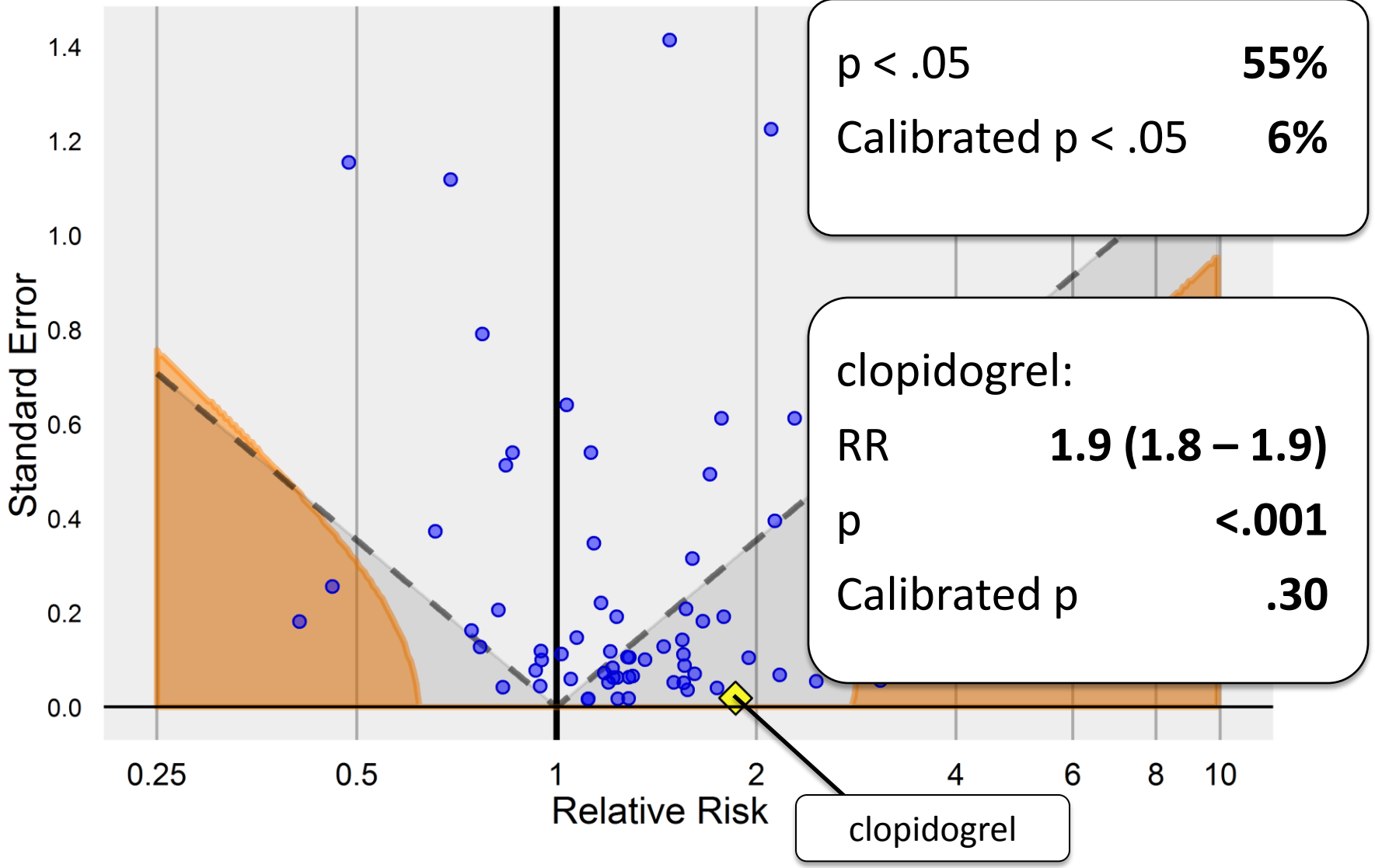
p-value calibration plot



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot

CC: 2000314, CCAE, GI Bleed



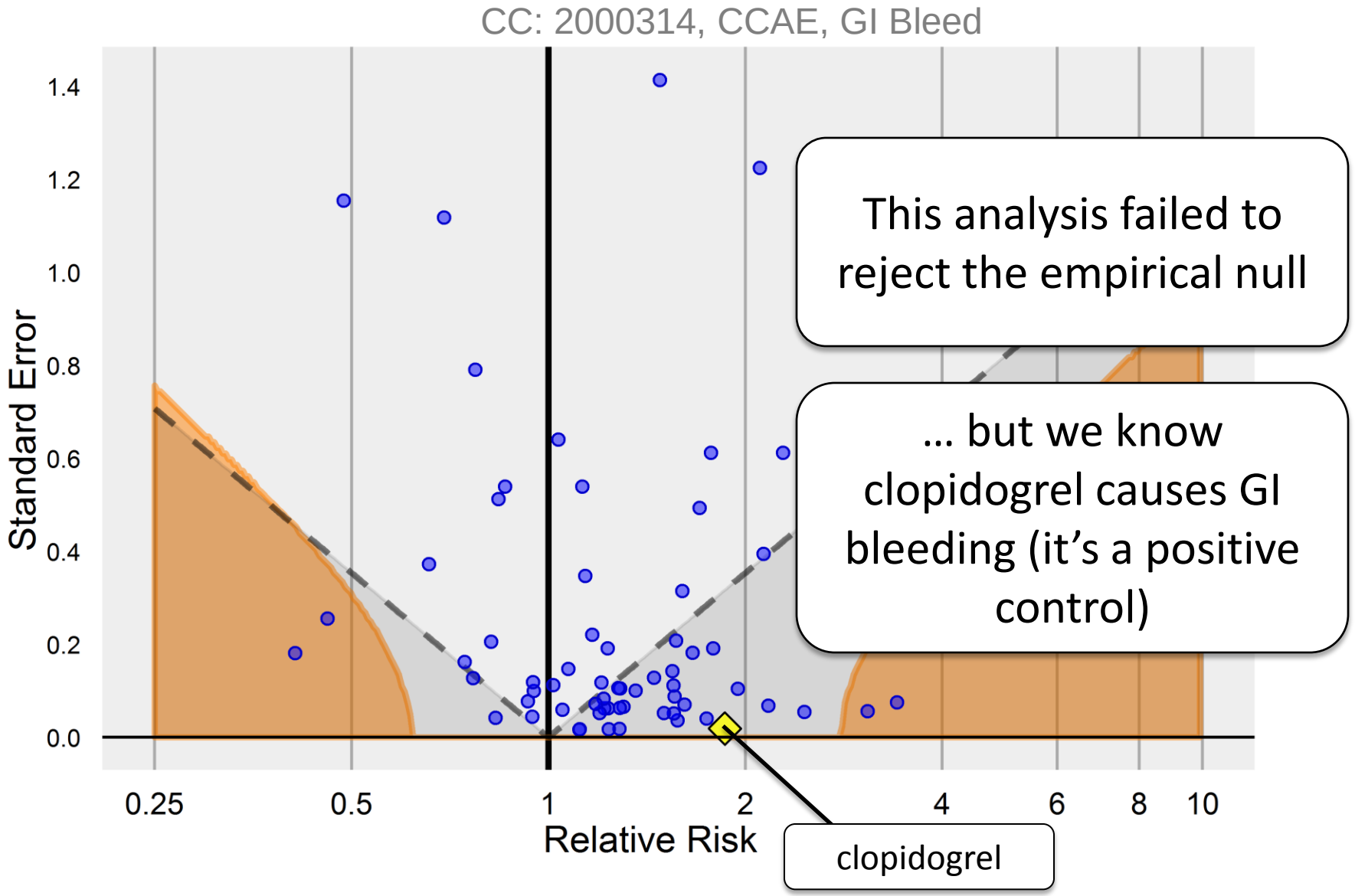
p < .05	**55%**
Calibrated p < .05	**6%**

clopidogrel:

RR	**1.9 (1.8 – 1.9)**
p	**<.001**
Calibrated p	**.30**

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

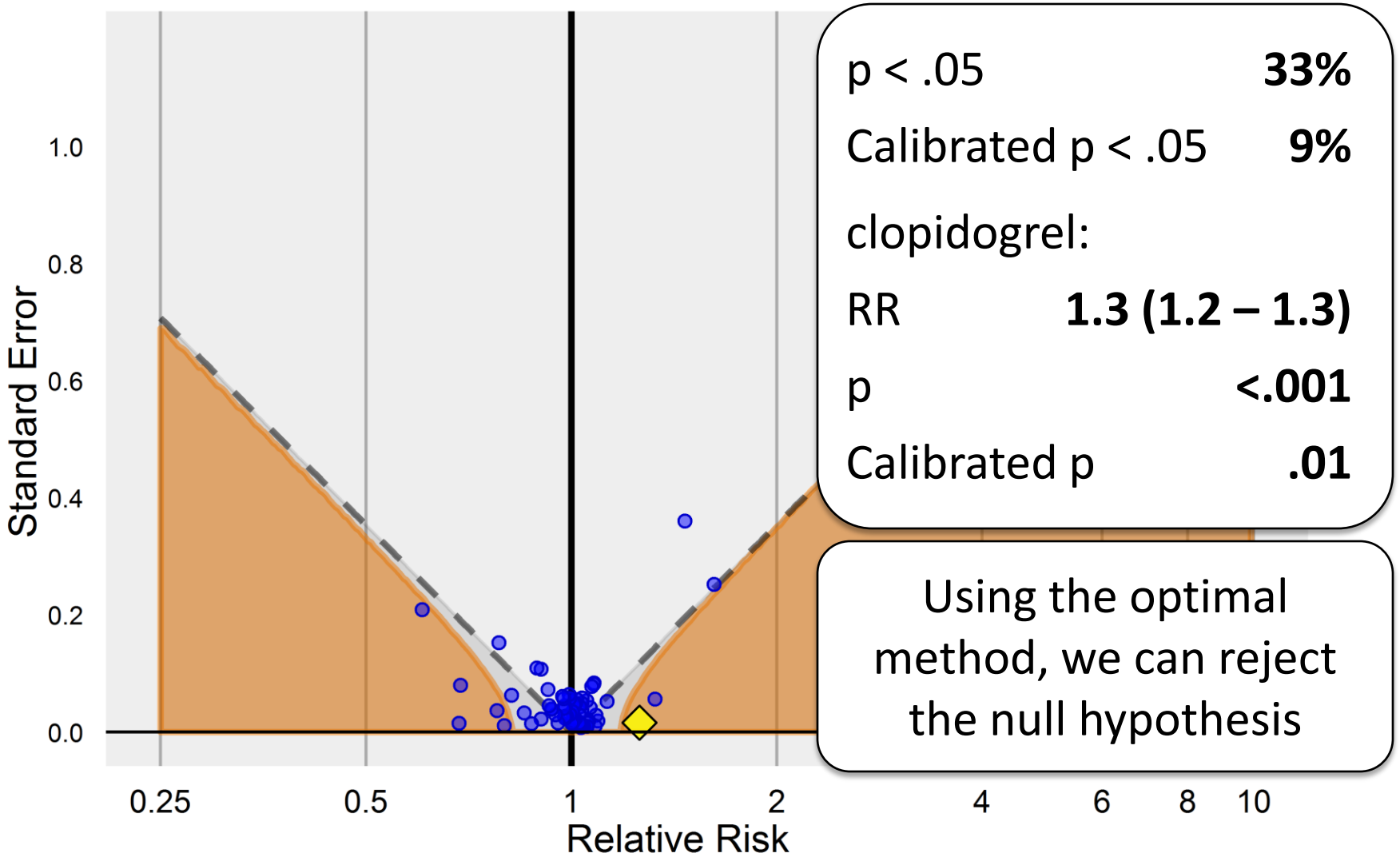
p-value calibration plot



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot

Optimal method: SCCS:1931010, CCAE, GI Bleed



p < .05	**33%**
Calibrated p < .05	**9%**
clopidogrel:	
RR	**1.3 (1.2 – 1.3)**
p	**<.001**
Calibrated p	**.01**

Using the optimal method, we can reject the null hypothesis

Recap

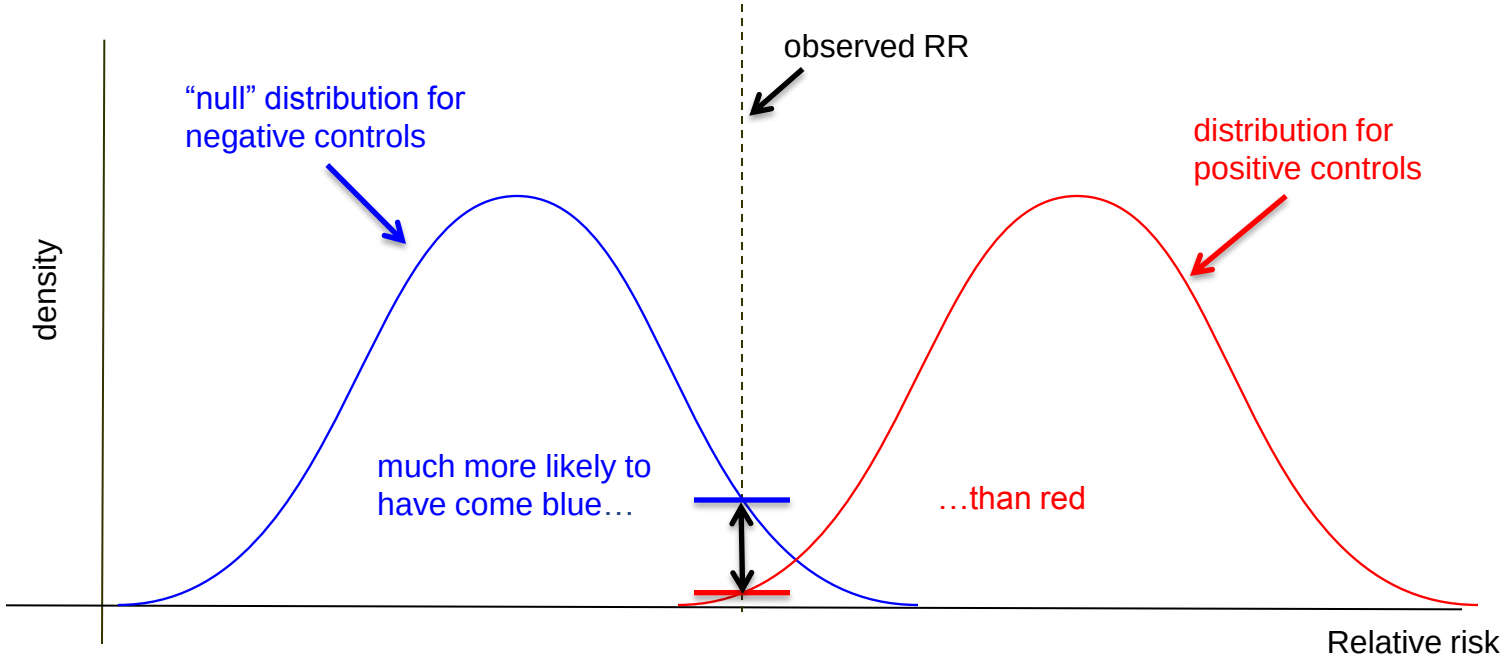
- Traditional p-values are based on a theoretical null distribution assuming an unbiased estimator, but that assumption rarely holds in our examples
- One can estimate the empirical null distribution using negative controls
- Many observational study results with traditional $p < .05$ fail to reject 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

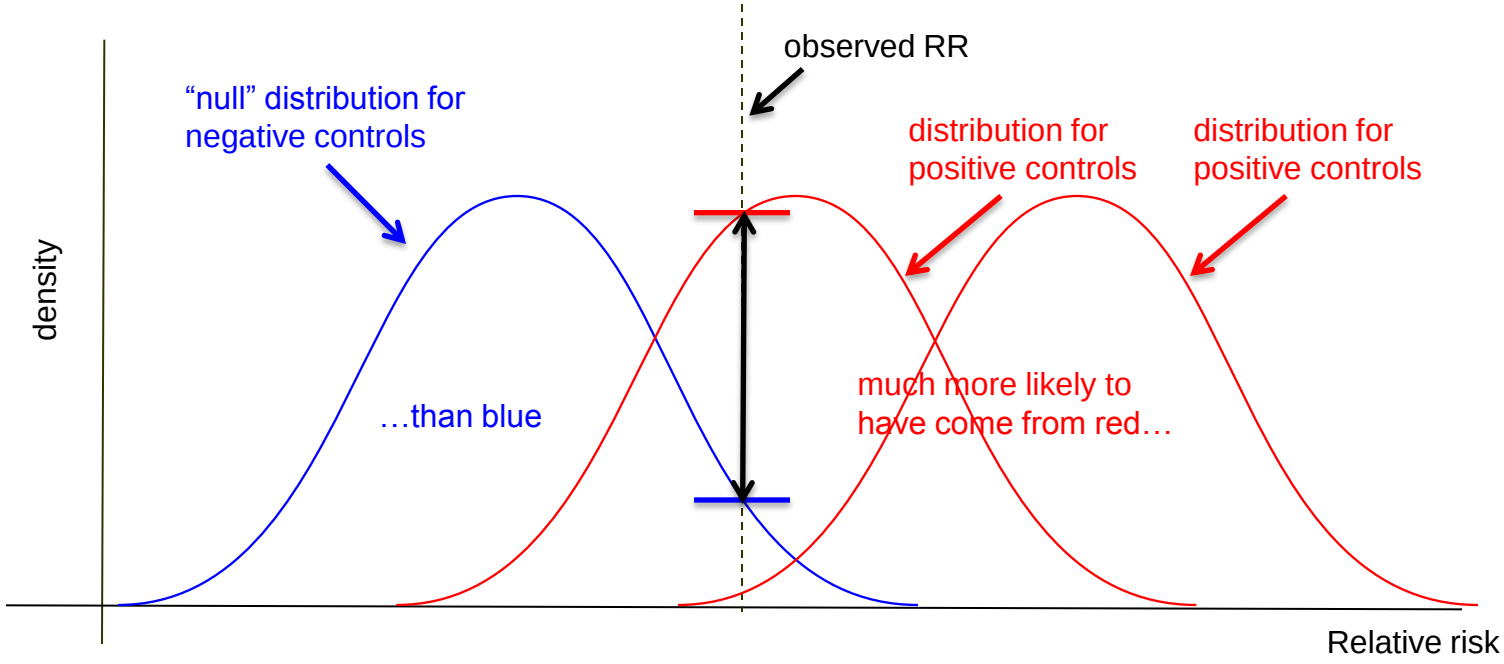
Beyond p-values: Computing the probability of a true association

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

We also have positive controls



But if AUC is small…



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)

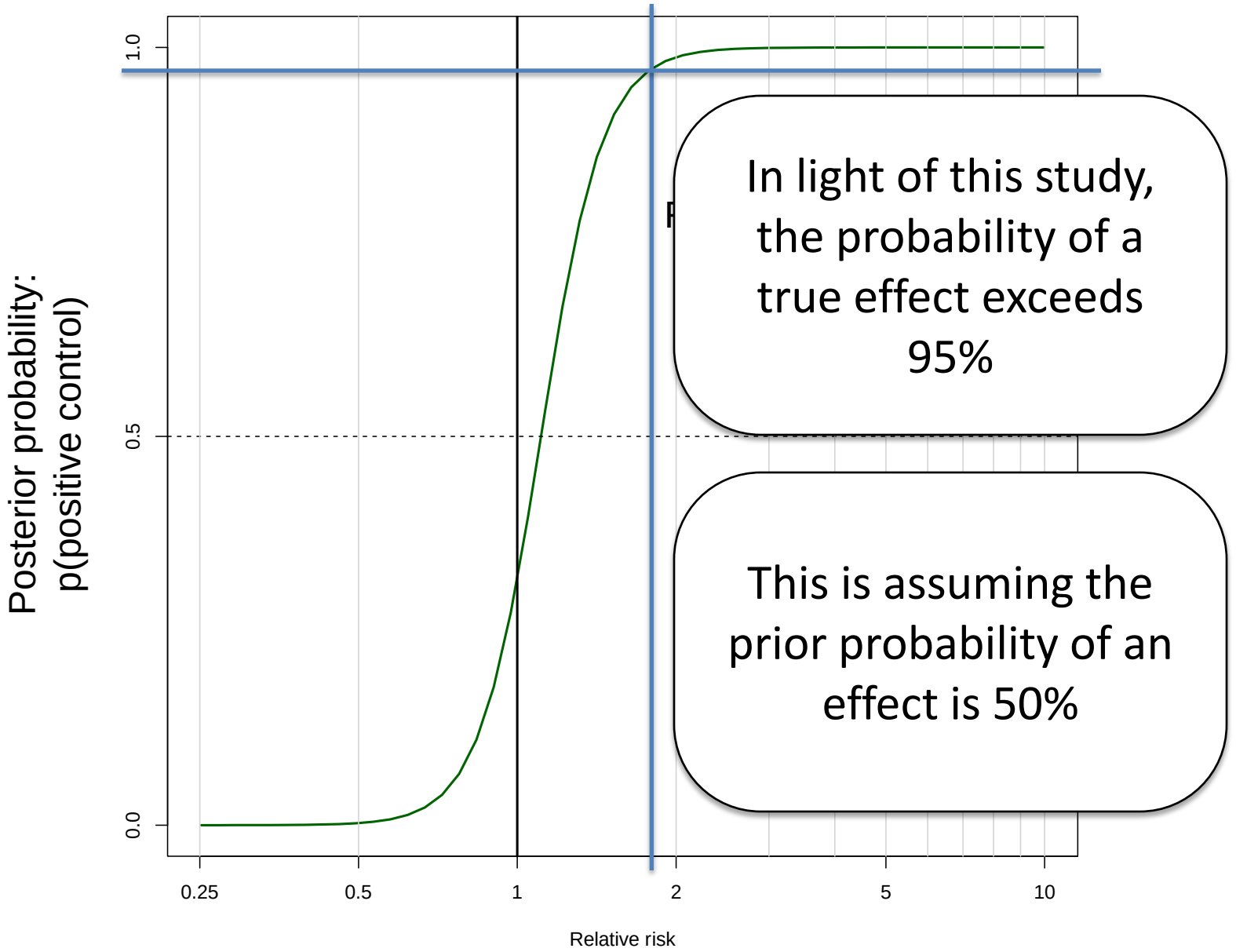
Relative risk: 1.86, 95% CI: 1.79 – 1.93

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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Clopidogrel – GI Bleed

Method: CC−2000314, Source: CCAE, HOI: GI Bleed



In light of this study, the probability of a true effect exceeds 95%

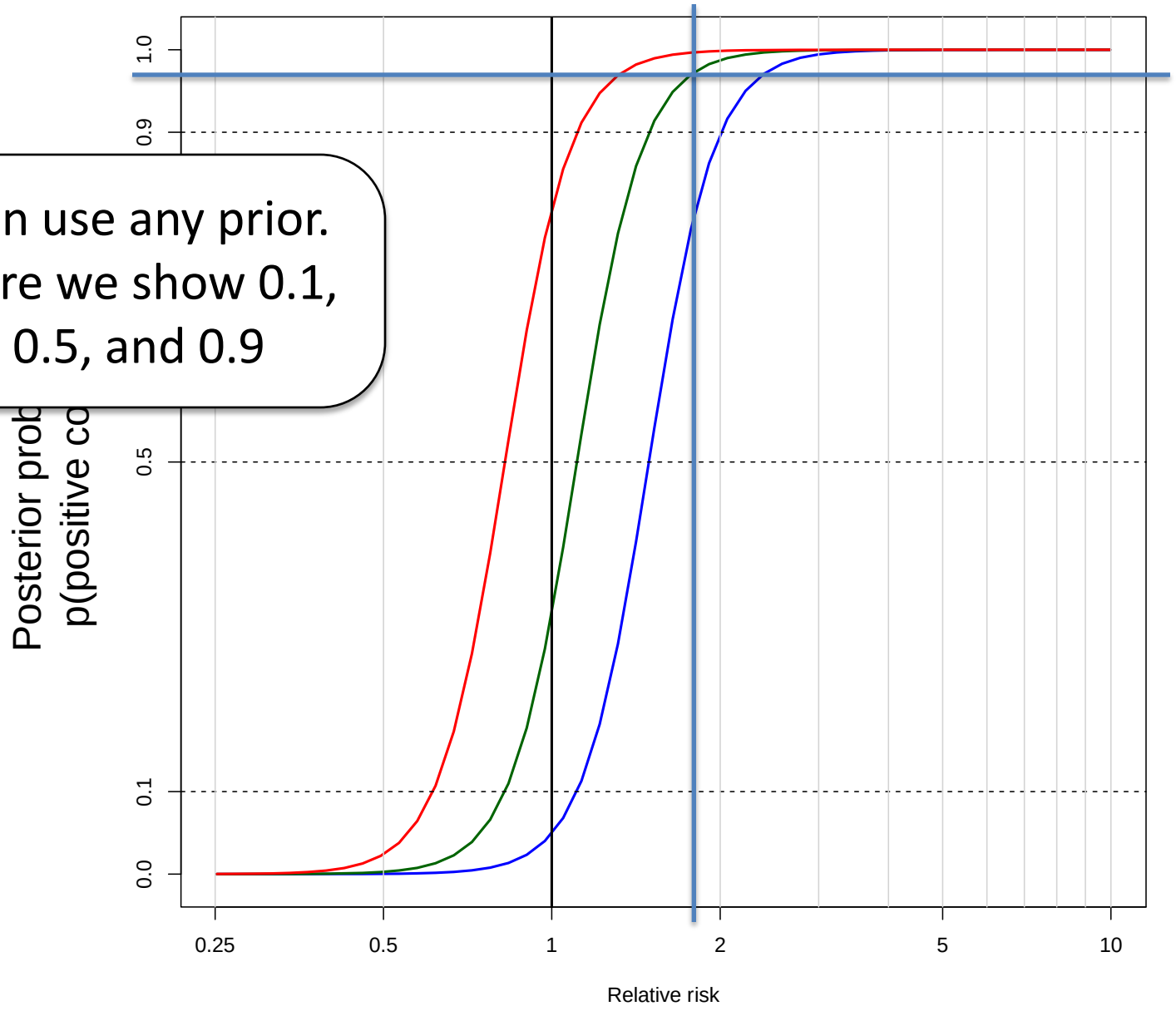
This is assuming the prior probability of an effect is 50%

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Clopidogrel – GI Bleed

Method: CC−2000314, Source: CCAE, HOI: GI Bleed

Can use any prior. Here we show 0.1, 0.5, and 0.9

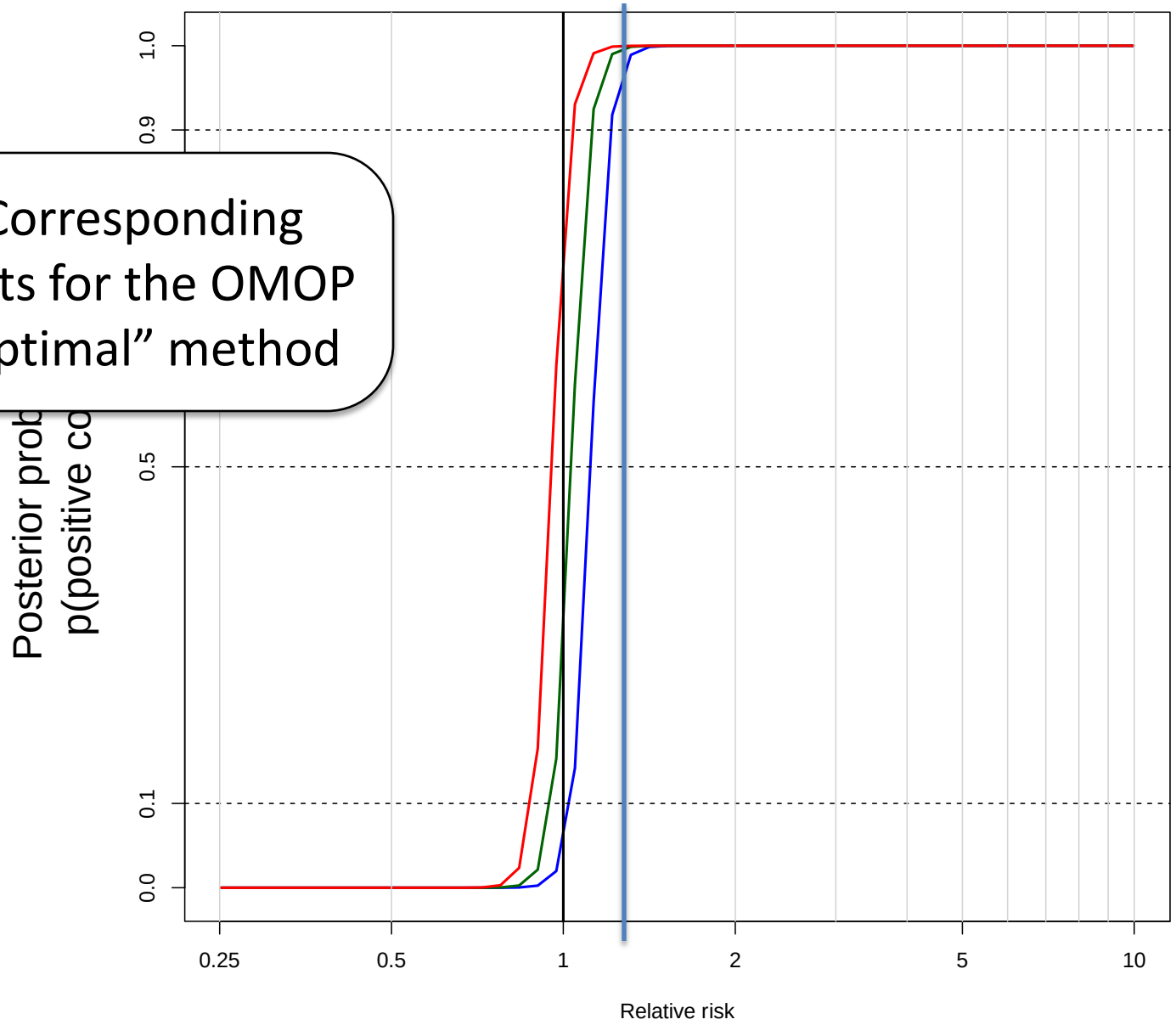


OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Clopidogrel – GI Bleed

Method: SCCS−1931010, Source: CCAE, HOI: GI Bleed

Corresponding plots for the OMOP "optimal" method



RR:1.26
(1.22, 1.30)
SE: 0.02

Prior:
p=0.9
p=0.5
p=0.1

β_1= 2.67
β_2= 0.41

Recap

- We have developed an empirical approach to quantifying the posterior probability of a true effect, given an observed estimate and prior beliefs

- Comparing the distribution of negative controls with the distribution of positive controls provides complementary information beyond the p-value
  - p<0.05 doesn't guarantee a true effect exists
  - p>0.05 doesn't guarantee no effect is present

Recap (continued)

- For each outcome, different methods may provide different weights of evidence
  - Some methods have greater discrimination and are more informative for interpreting a new estimate
  - Sometimes prior beliefs will drive the revised understanding
  - Other times, evidence will be sufficiently compelling that everyone, with different prior beliefs, should reach similar conclusions

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?

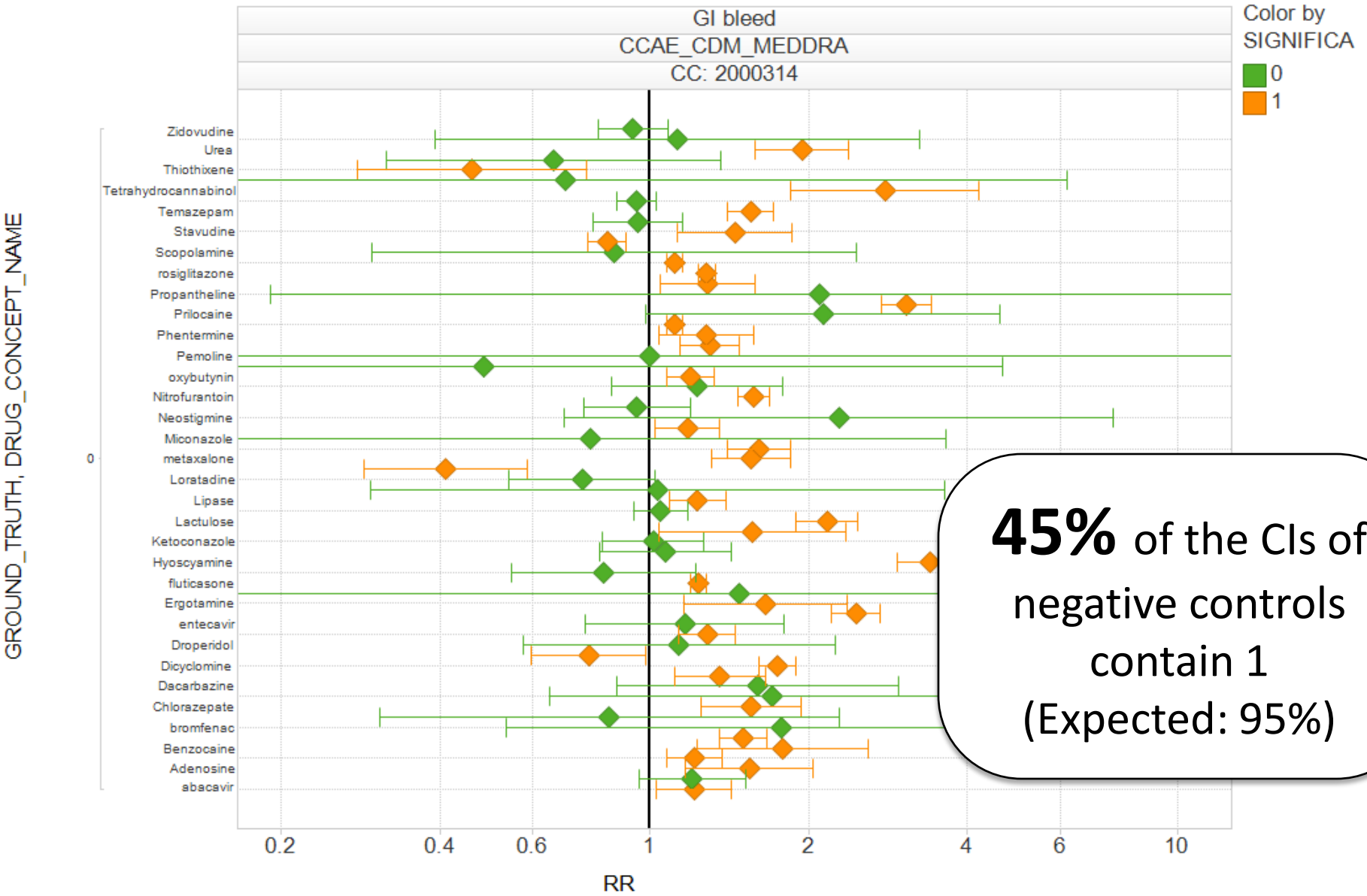
- New question: What is the probability that observed confidence interval contains the true effect size?

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

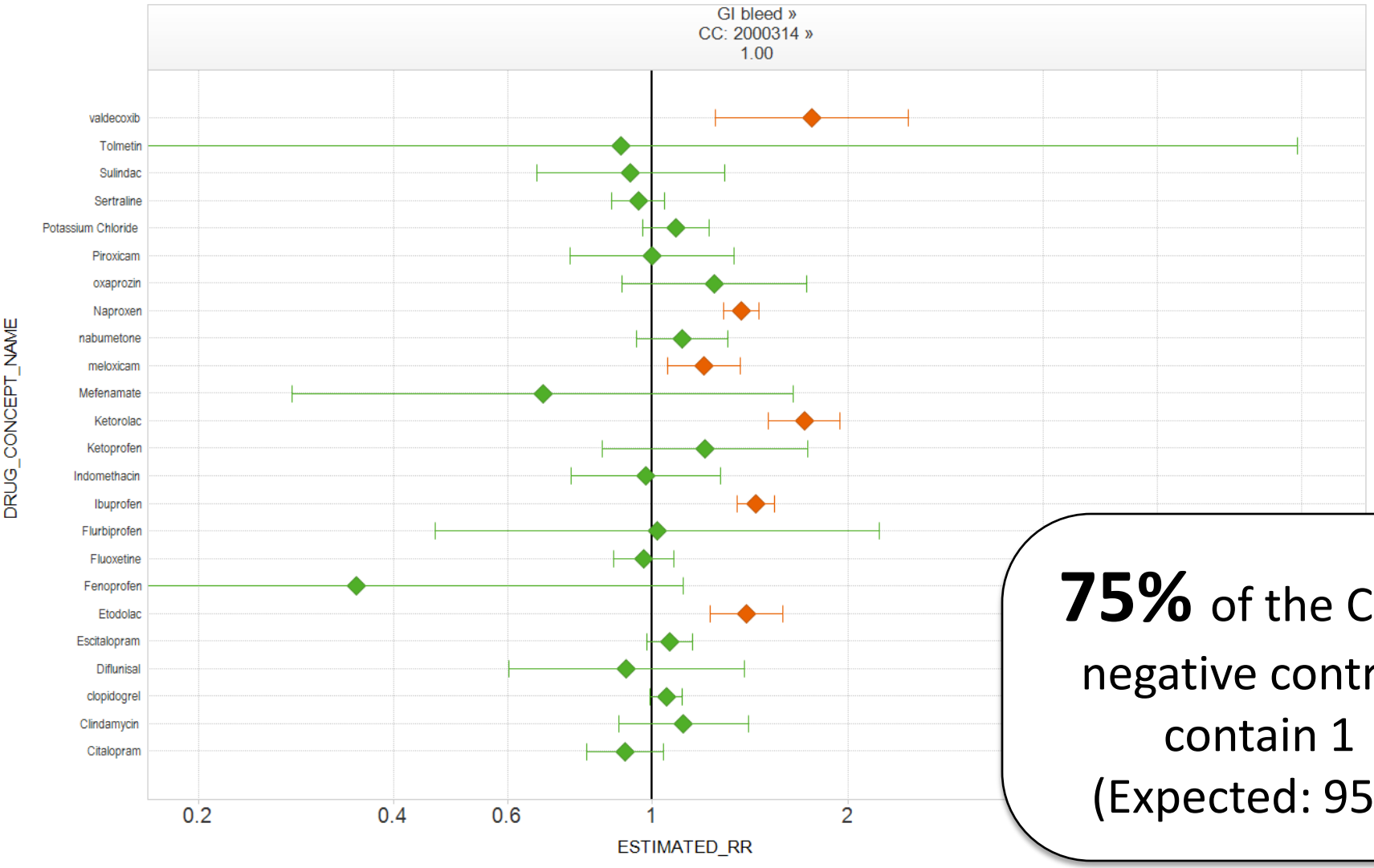
Applying case-control design to negative controls in real data, RR=1.25



45% of the CIs of negative controls contain 1 (Expected: 95%)

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

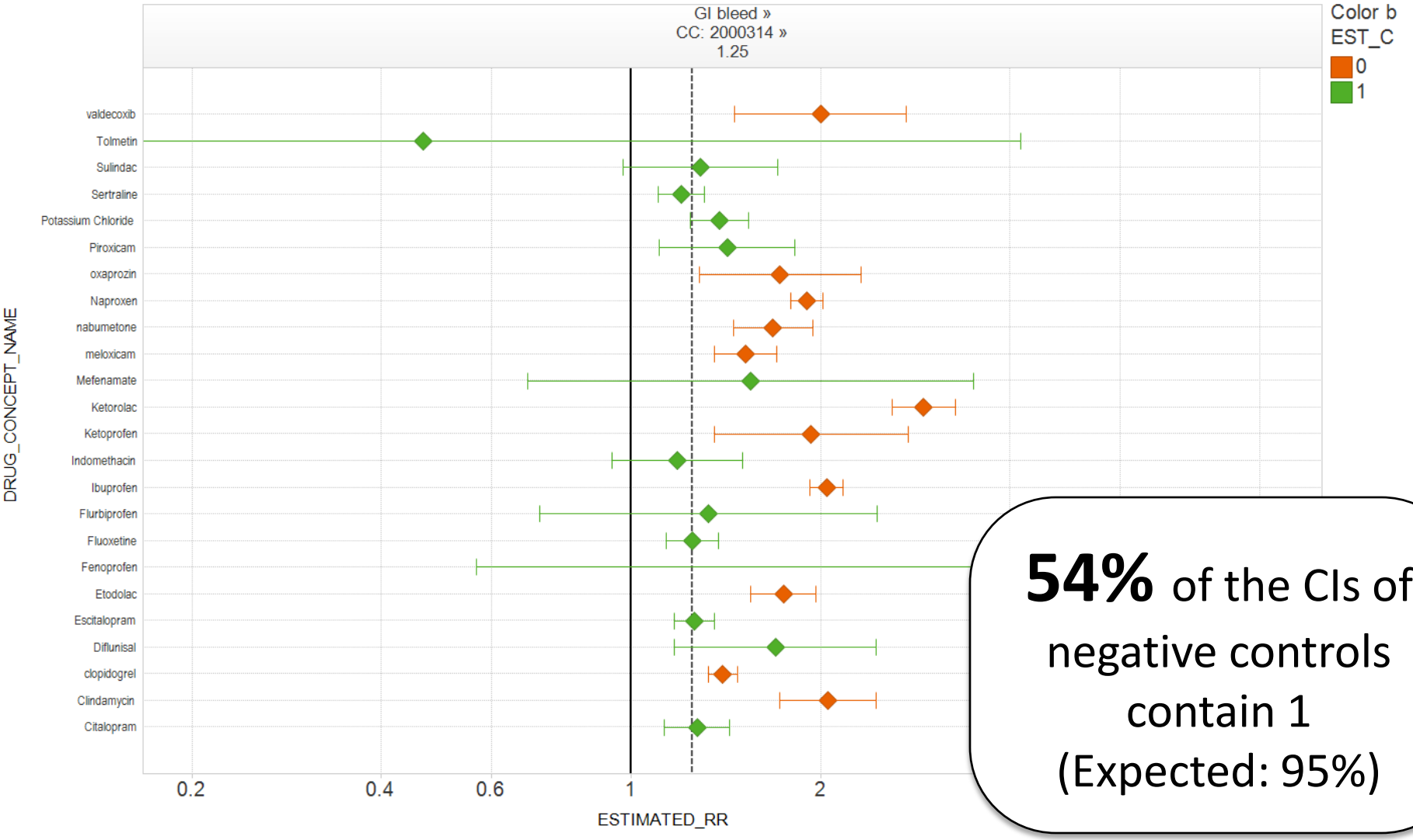
Applying case-control design to positive controls in simulated data, RR=1.0



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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

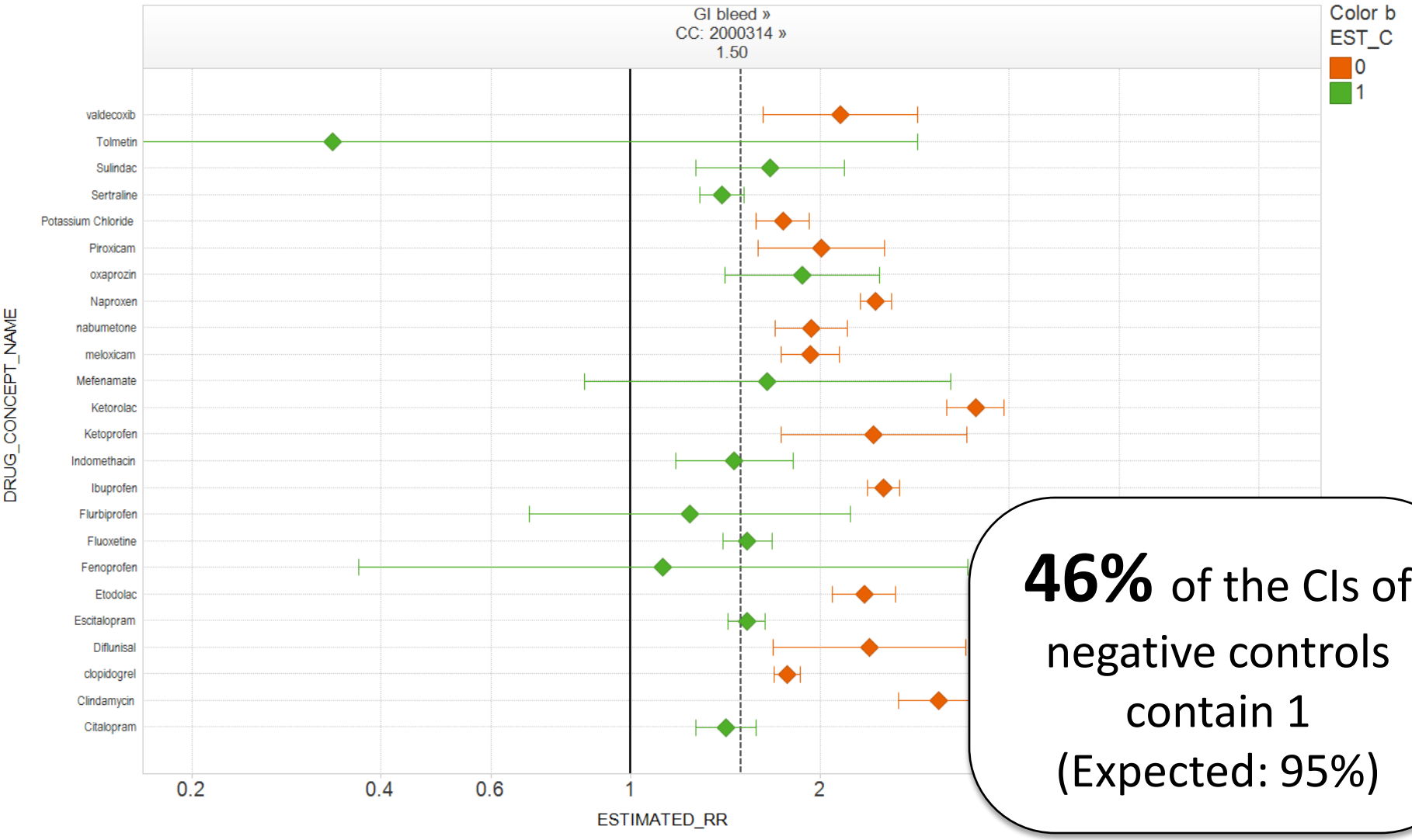
Applying case-control design to positive controls in simulated data, RR=1.25



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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

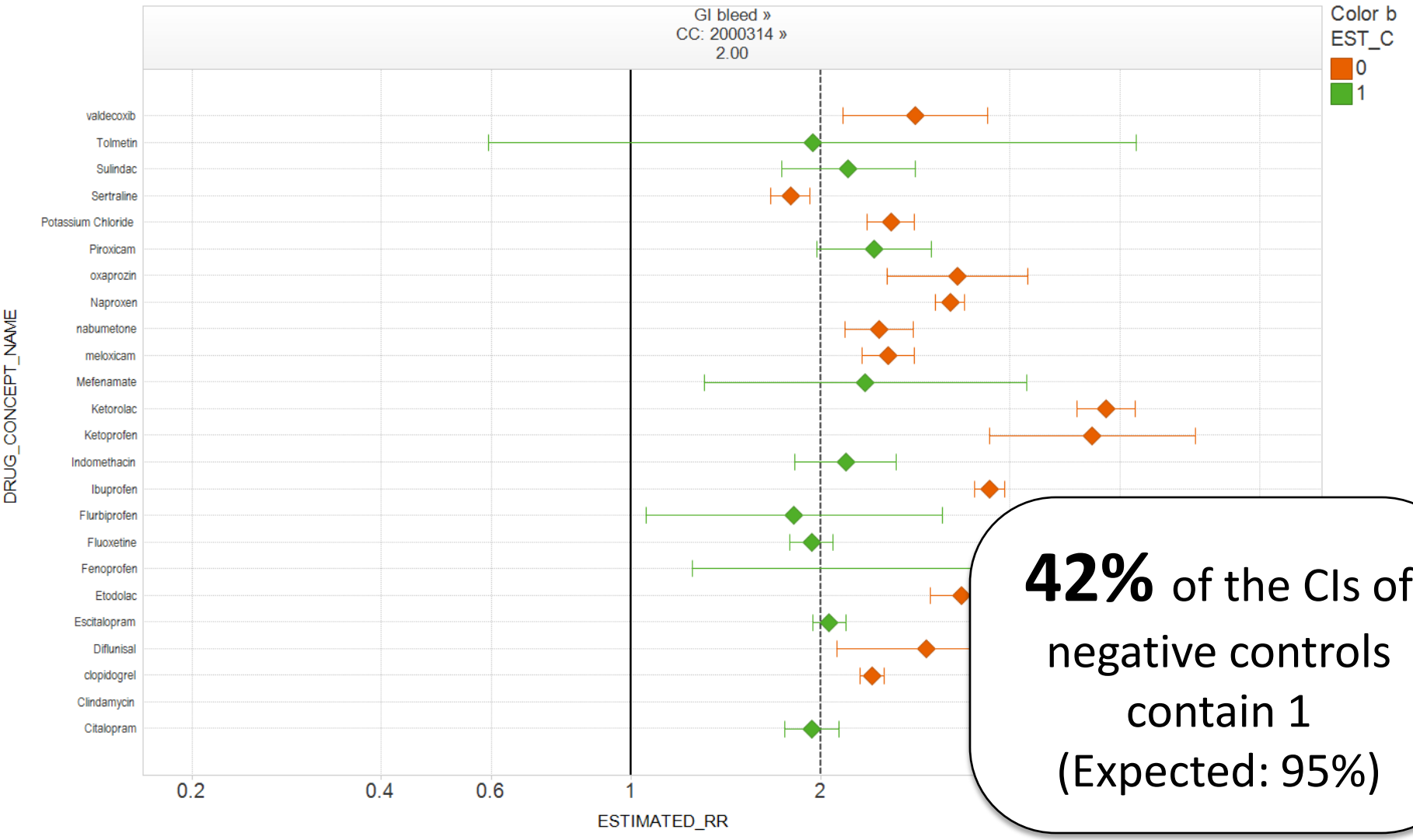
Applying case-control design to positive controls in simulated data, RR=1.50



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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

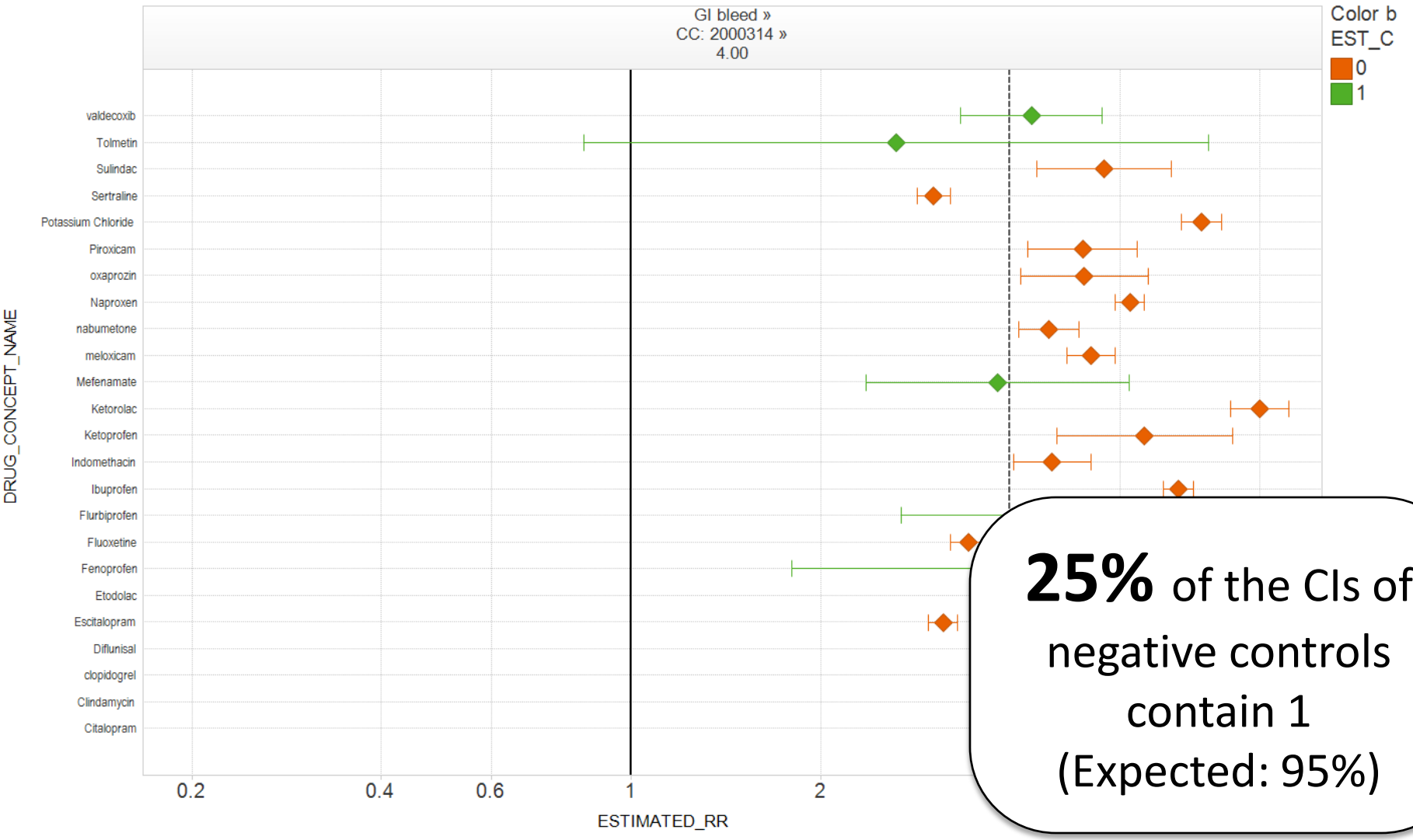
Applying case-control design to positive controls in simulated data, RR=2.00



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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

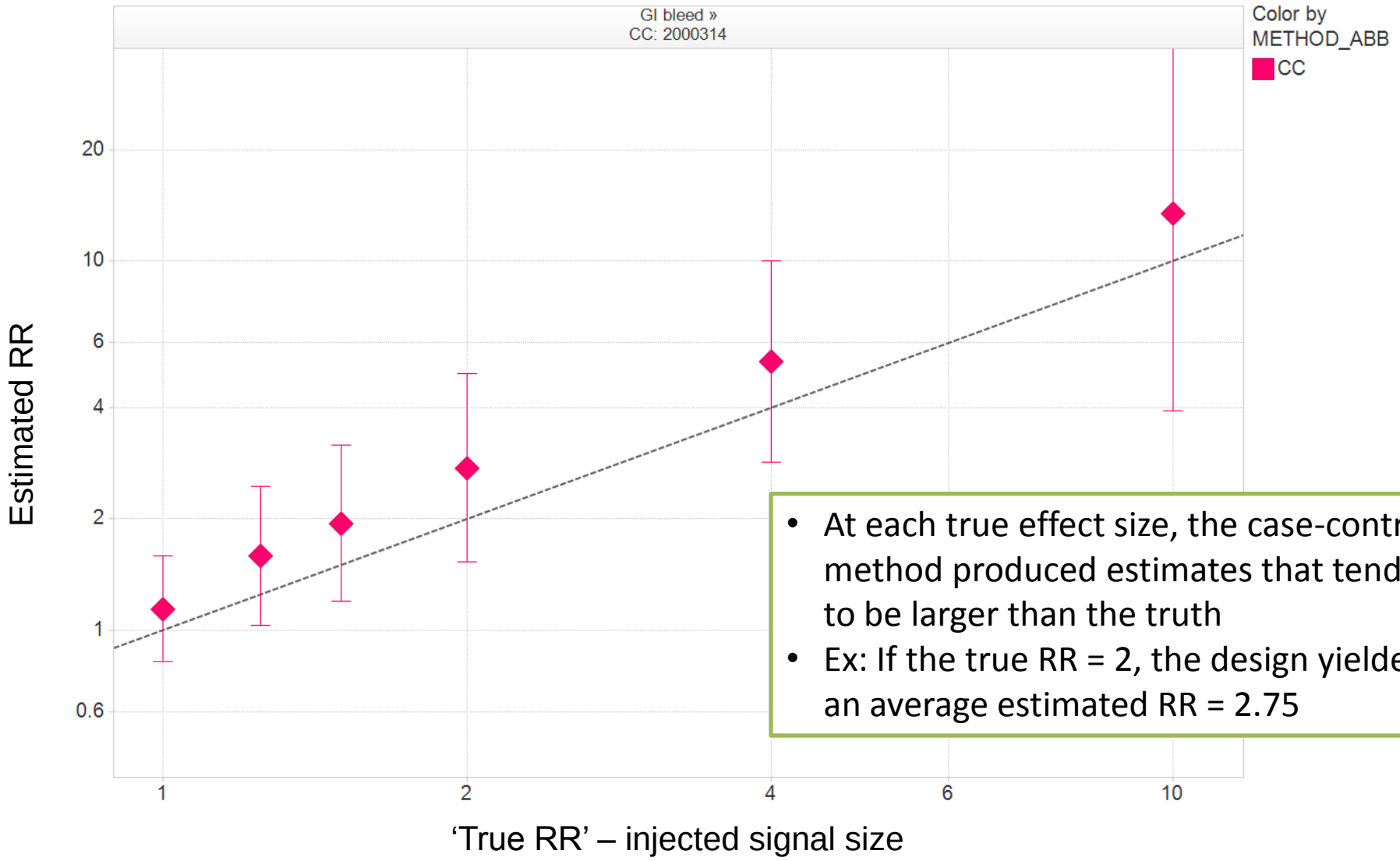
Applying case-control design to positive controls in simulated data, RR=4.00



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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

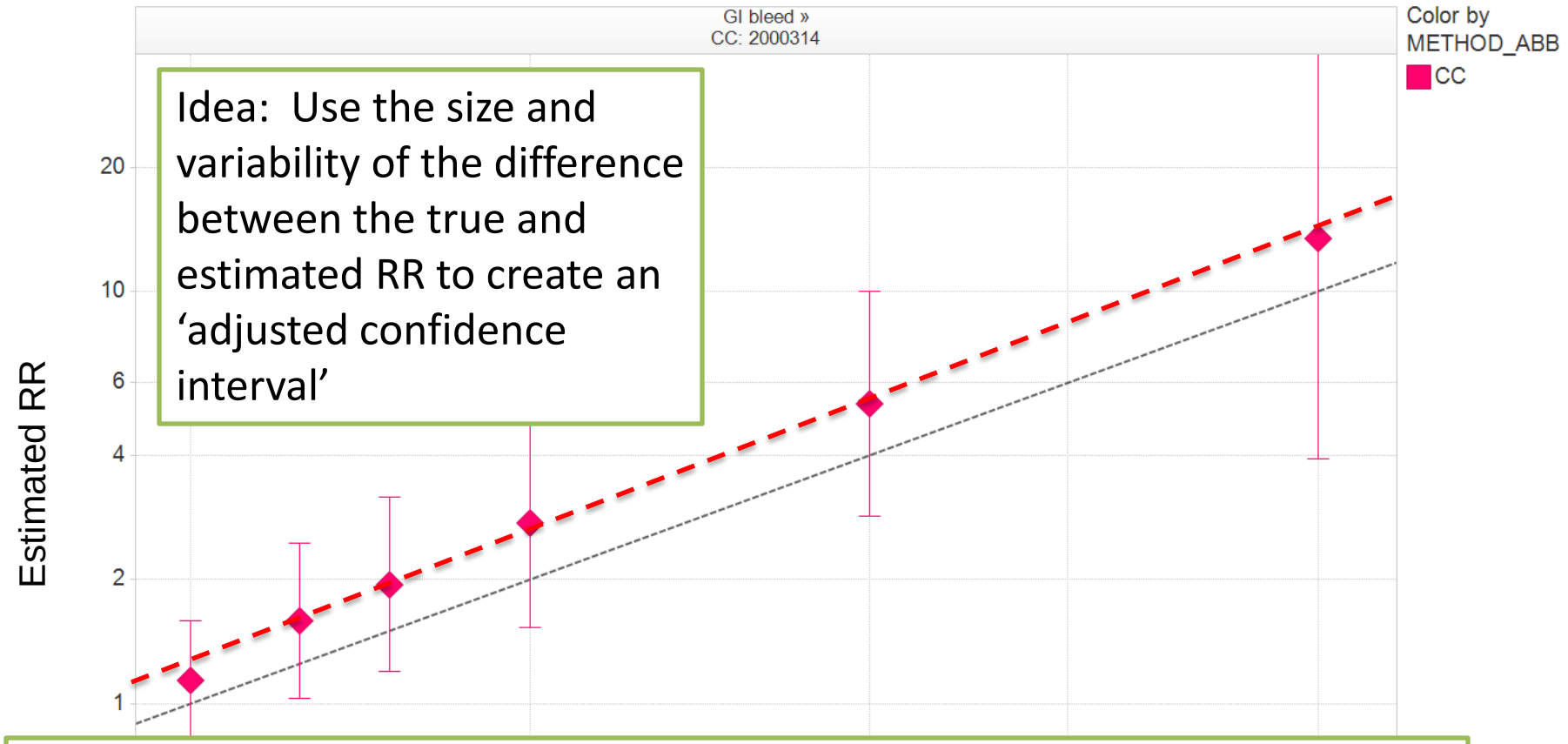
How far off were the case-control estimates from the truth?



- 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?



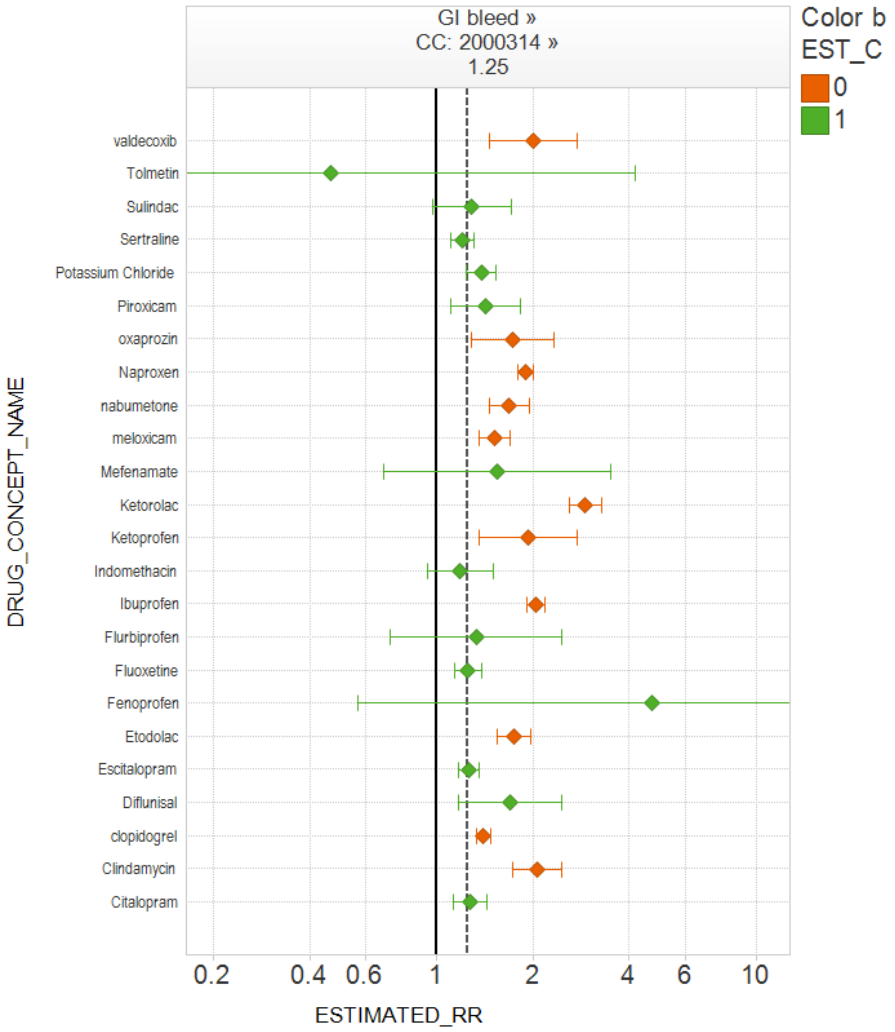
Idea: Use the size and variability of the difference between the true and estimated RR to create an 'adjusted confidence interval'

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

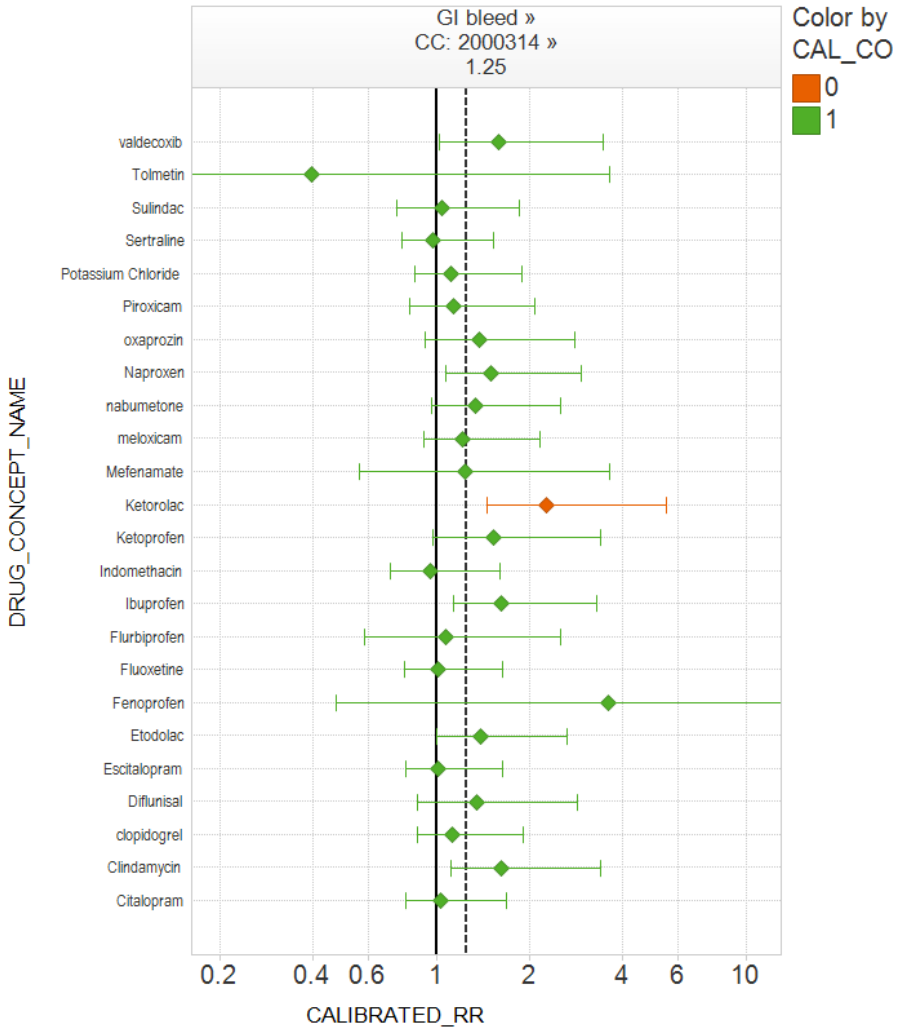
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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

Original estimated effects



Calibrated confidence intervals



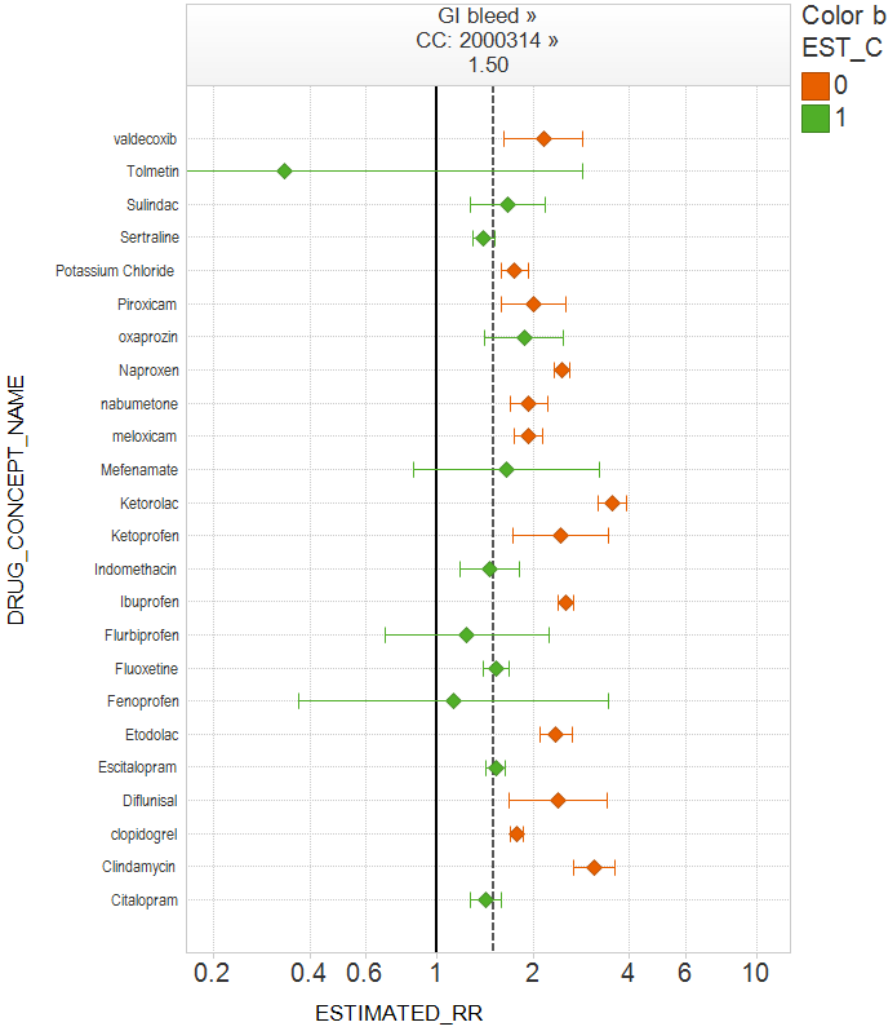
Original coverage probability = **54%**

Calibrated coverage probability = **96%**

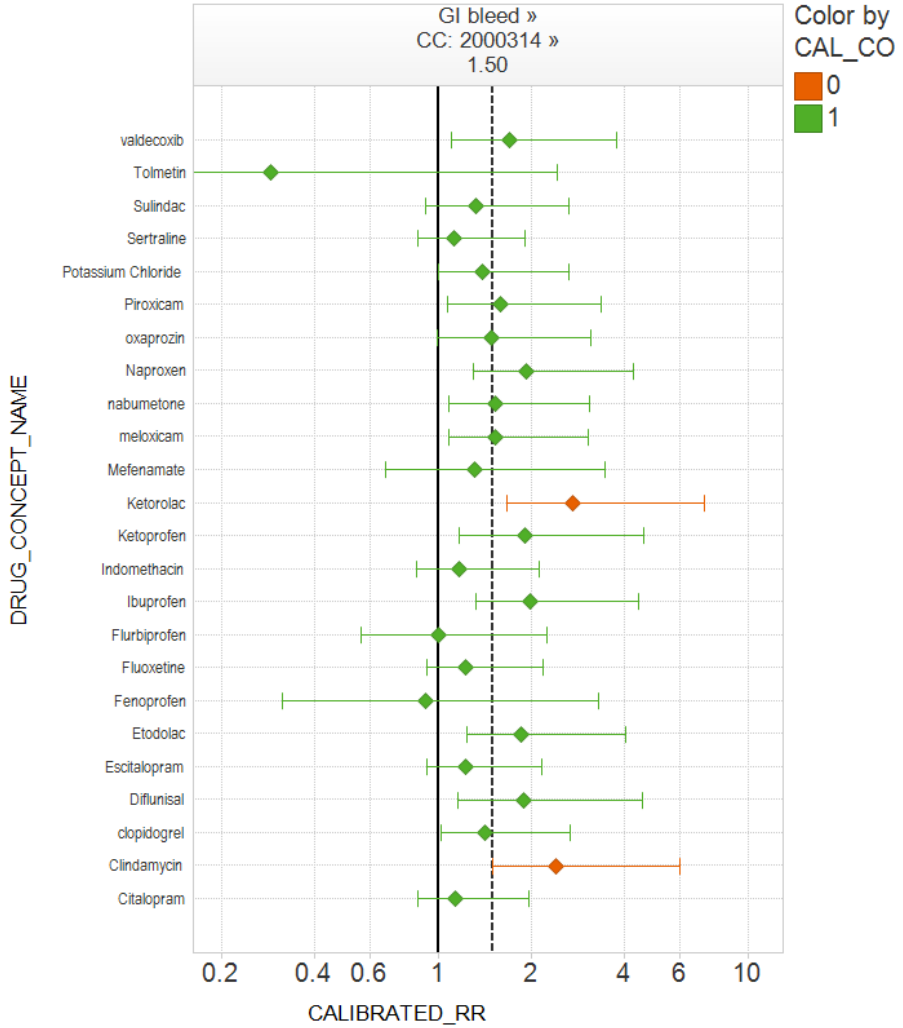
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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

Original estimated effects



Calibrated confidence intervals



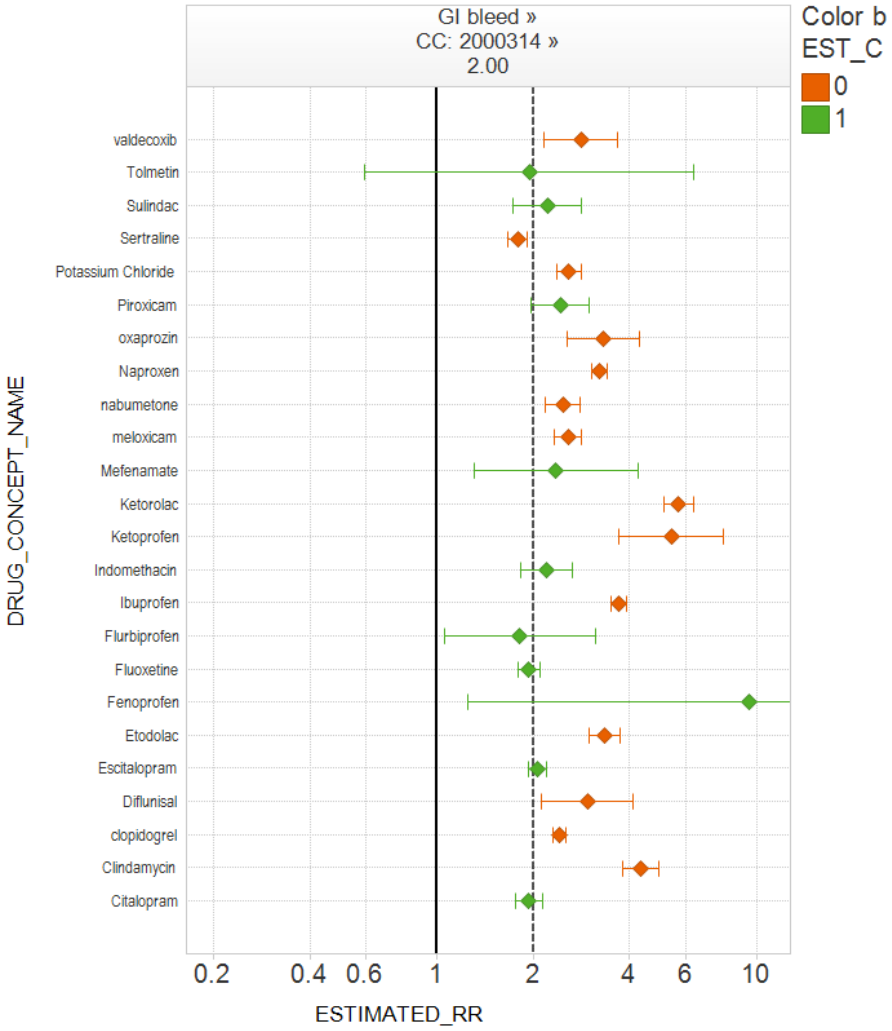
Original coverage probability = **46%**

Calibrated coverage probability = **92%**

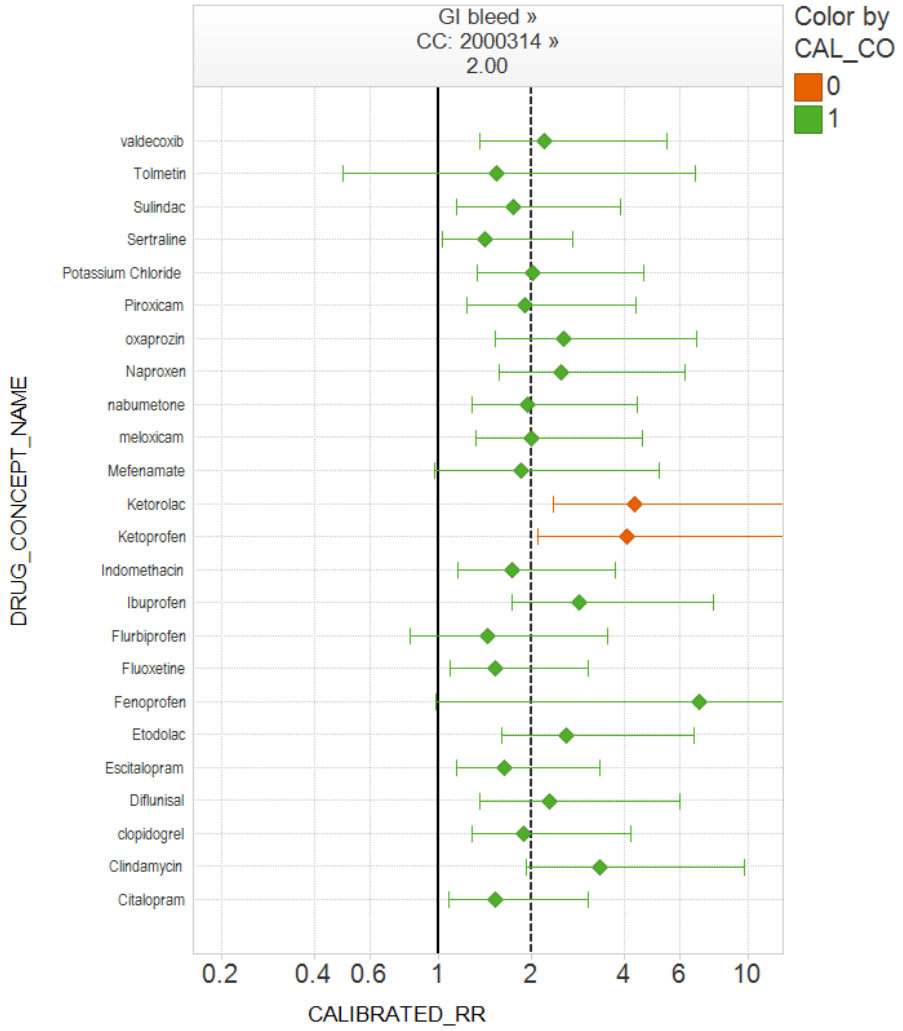
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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

Original estimated effects



Calibrated confidence intervals



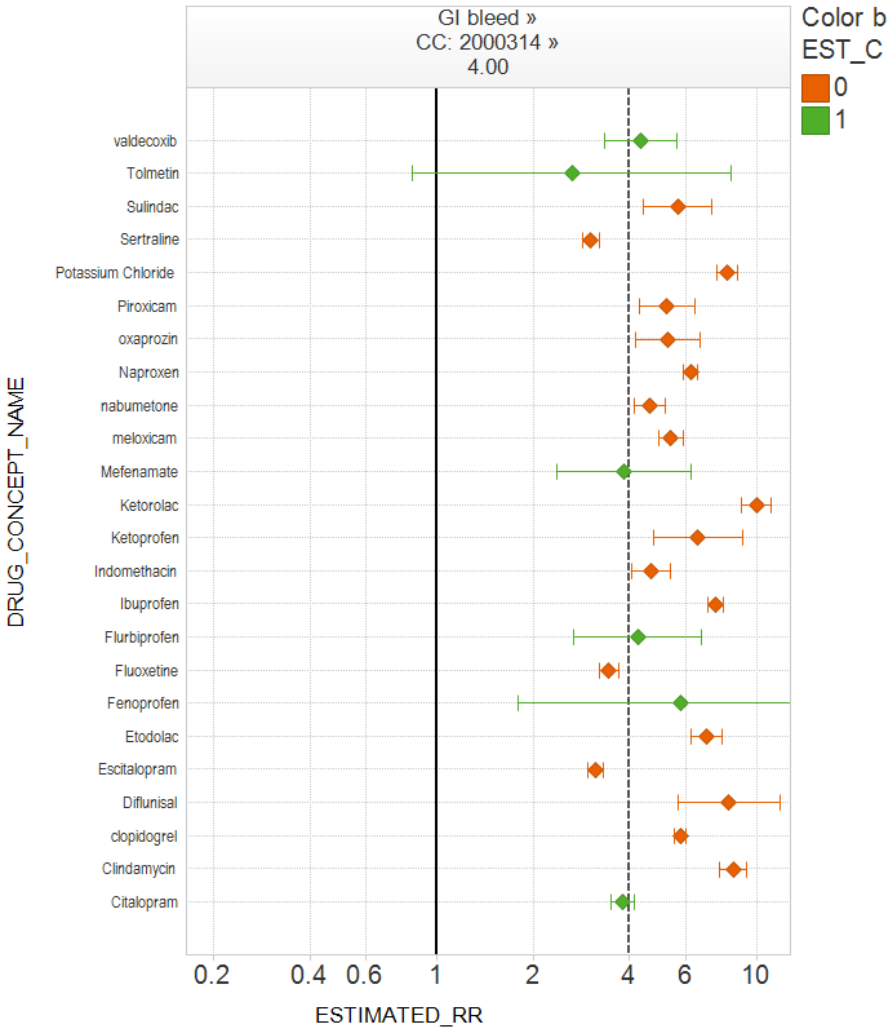
Original coverage probability = **42%**

Calibrated coverage probability = **92%**

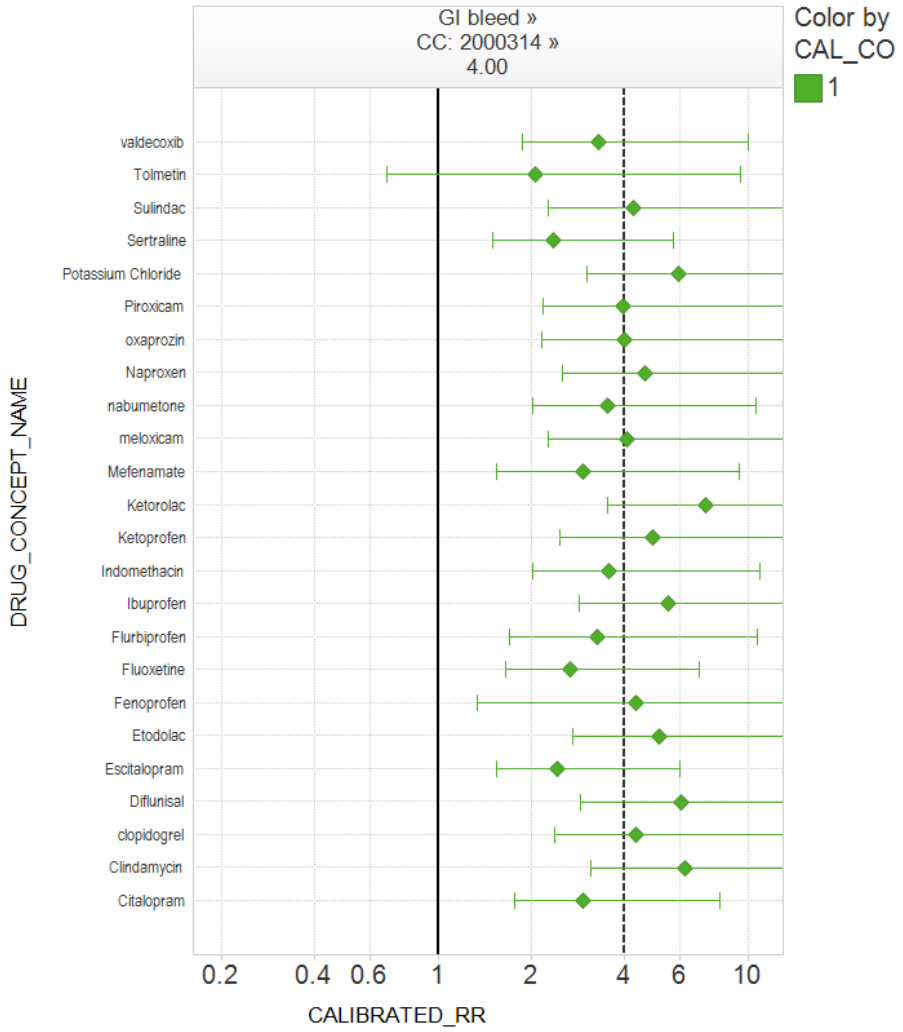
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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

Concluding thoughts

Lessons for building a risk identification system

- Strategies to improve performance:
  - Partition results by outcome
  - Tailor analysis to outcome
  - Restrict to sufficient sample size
  - Optimize analysis to the data source
- OMOP's experimental evidence suggests that following these strategies may yield predictive accuracy at or better than most clinical screening tools used in standard practice

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Lessons for building a risk identification system

- Where we are now:
  - Given the diversity in performance and heterogeneity in estimates, we caution against generalizing these results to other outcomes or other data sources
  - If you want to apply risk identification to different outcomes and/or different data sources, we suggest performing an empirical assessment to establish best practice and benchmark performance
- Potential next step:
  - conduct similar experiment for additional 19 outcomes identified by EUADR[1] as high-priority safety issues
  - Once 23 HOIs complete, re-assess whether patterns emerge that would allow generalization to other outcomes

[1]Trifiro et al, PDS 2009

Conclusions

- Using the OMOP approach, a risk identification system can perform at AUC>0.80
- Traditional p-values and confidence intervals require empirical calibration to account for bias in observational studies
- Advancing the science of observational research requires an empirical and reproducible approach to methodology and systematic application

ENGELBERG CENTER for Health Care Reform at BROOKINGS

Roundtable Discussion and Questions

View this and past Active Medical Product Surveillance webinars at:

http://www.brookings.edu/health/Projects/surveillance/roundtables.aspx