Opportunities and challenges for using networks of observational healthcare data for medical product safety surveillance

Jesse Berlin, ScD
VP, Epidemiology
Janssen Research & Development
A Johnson & Johnson Pharmaceutical Company

BASS Conference XIX
Savannah, GA
6 November 2012

With many thanks to:

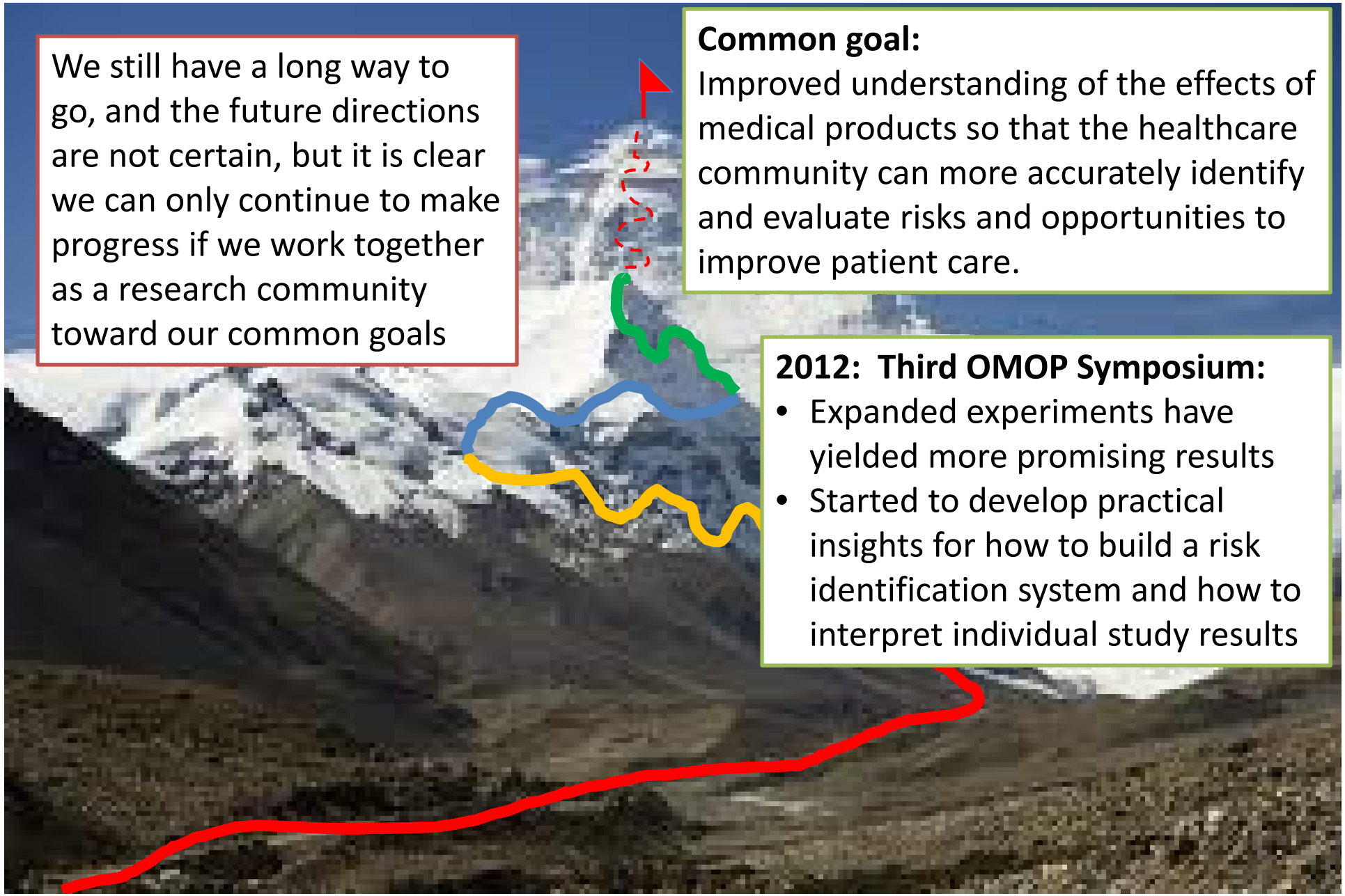
Patrick Ryan, Martijn Schuemie, and David Madigan
on behalf of the OMOP research team

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



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

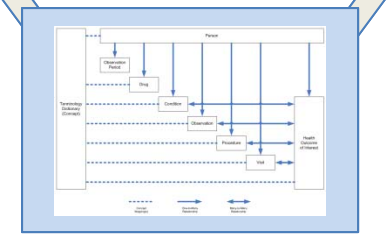
Drug-outcome pairs

	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

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

An empirical approach to null hypothesis testing

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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

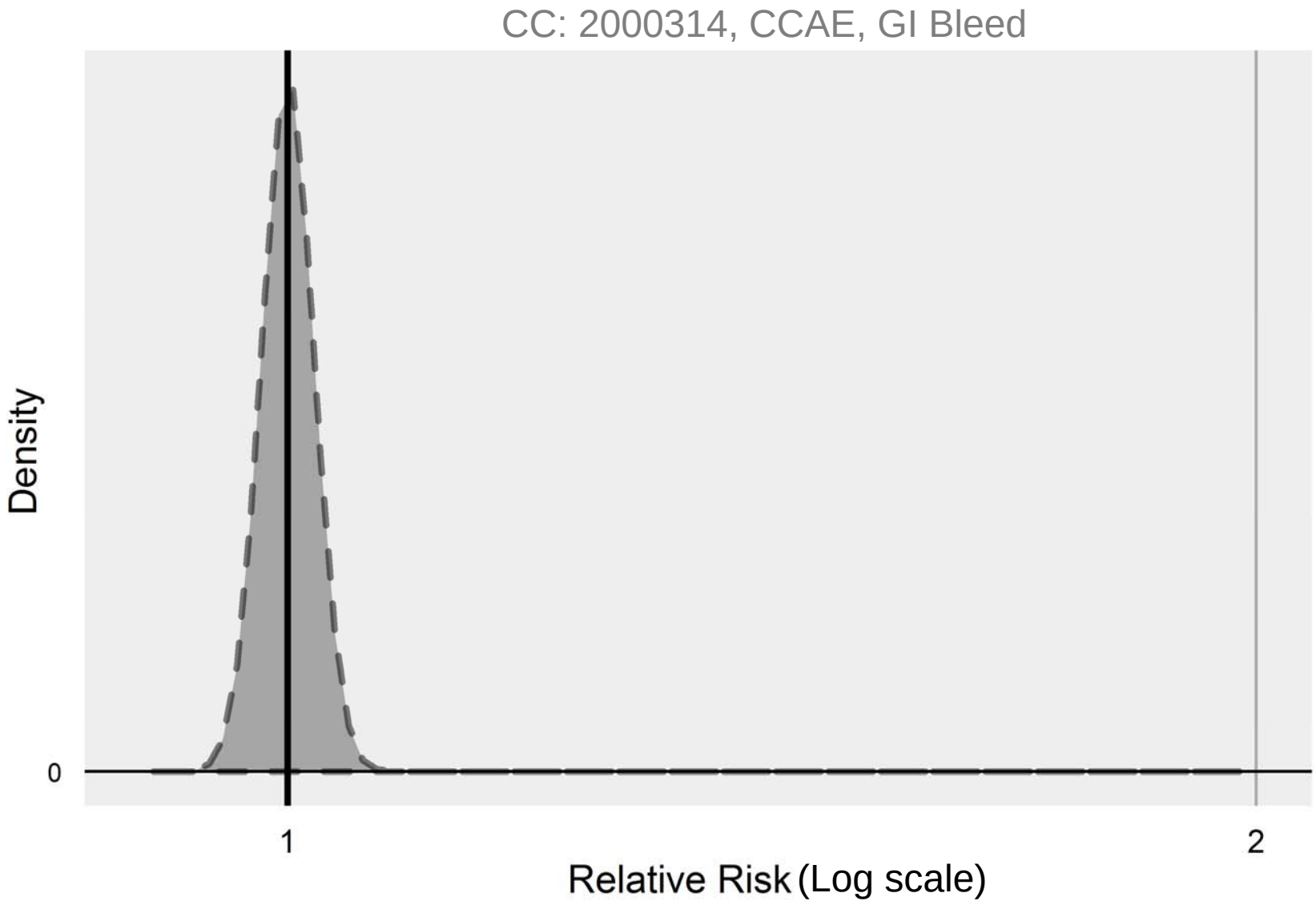
OMOP, 2012 (CC: 2000314, CCAE, GI Bleed)

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

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

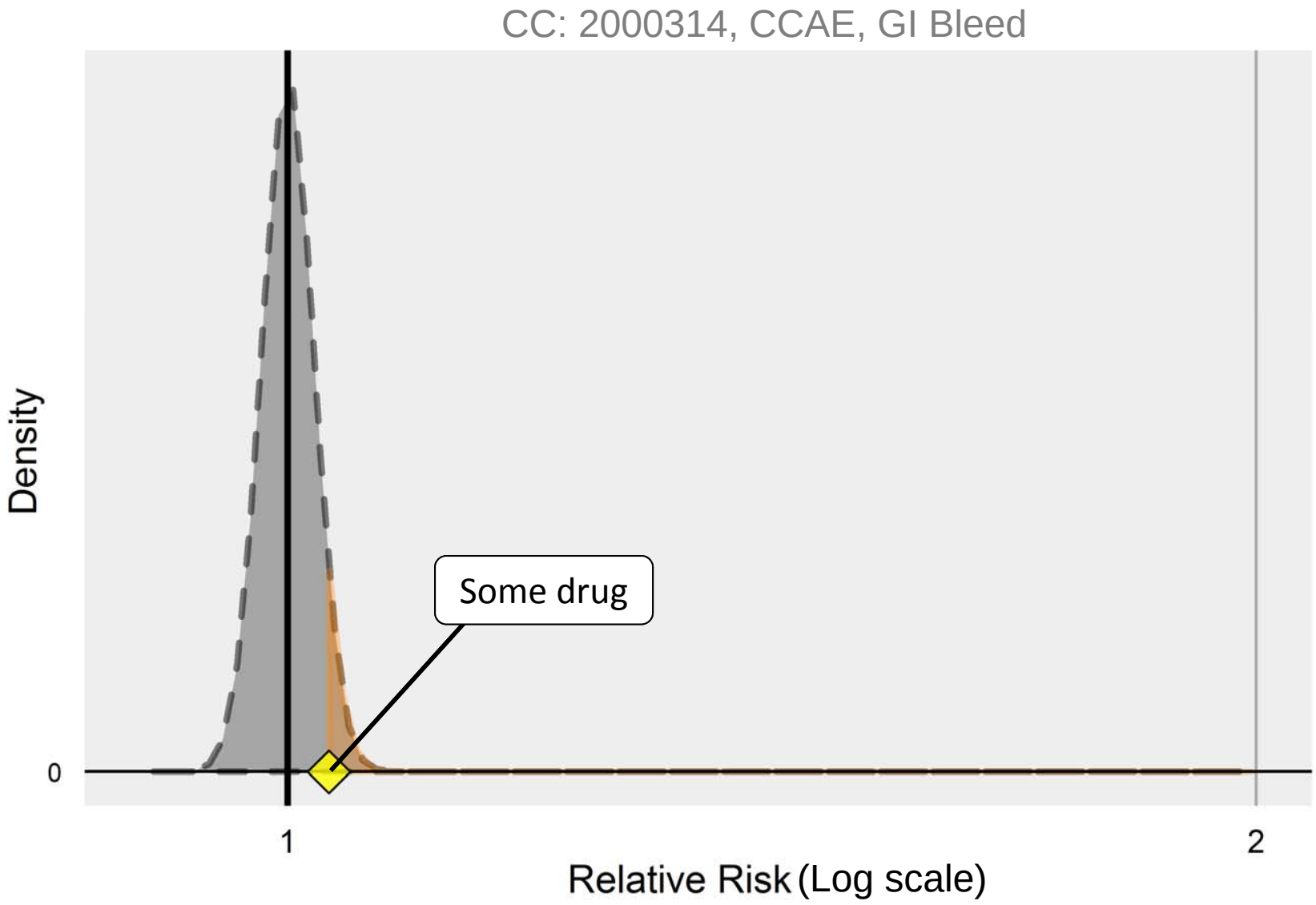
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Null distribution



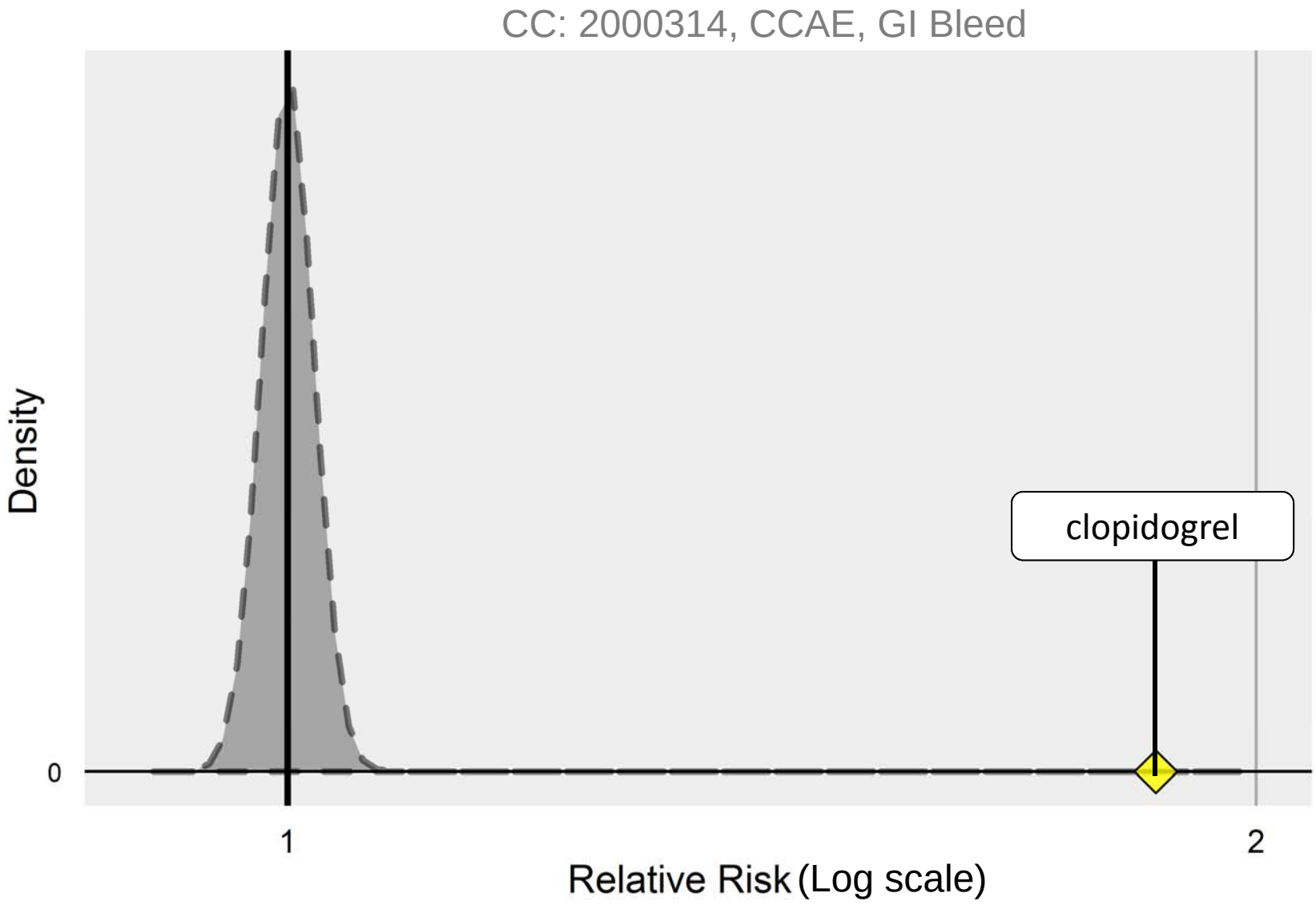
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Null distribution



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Null distribution



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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

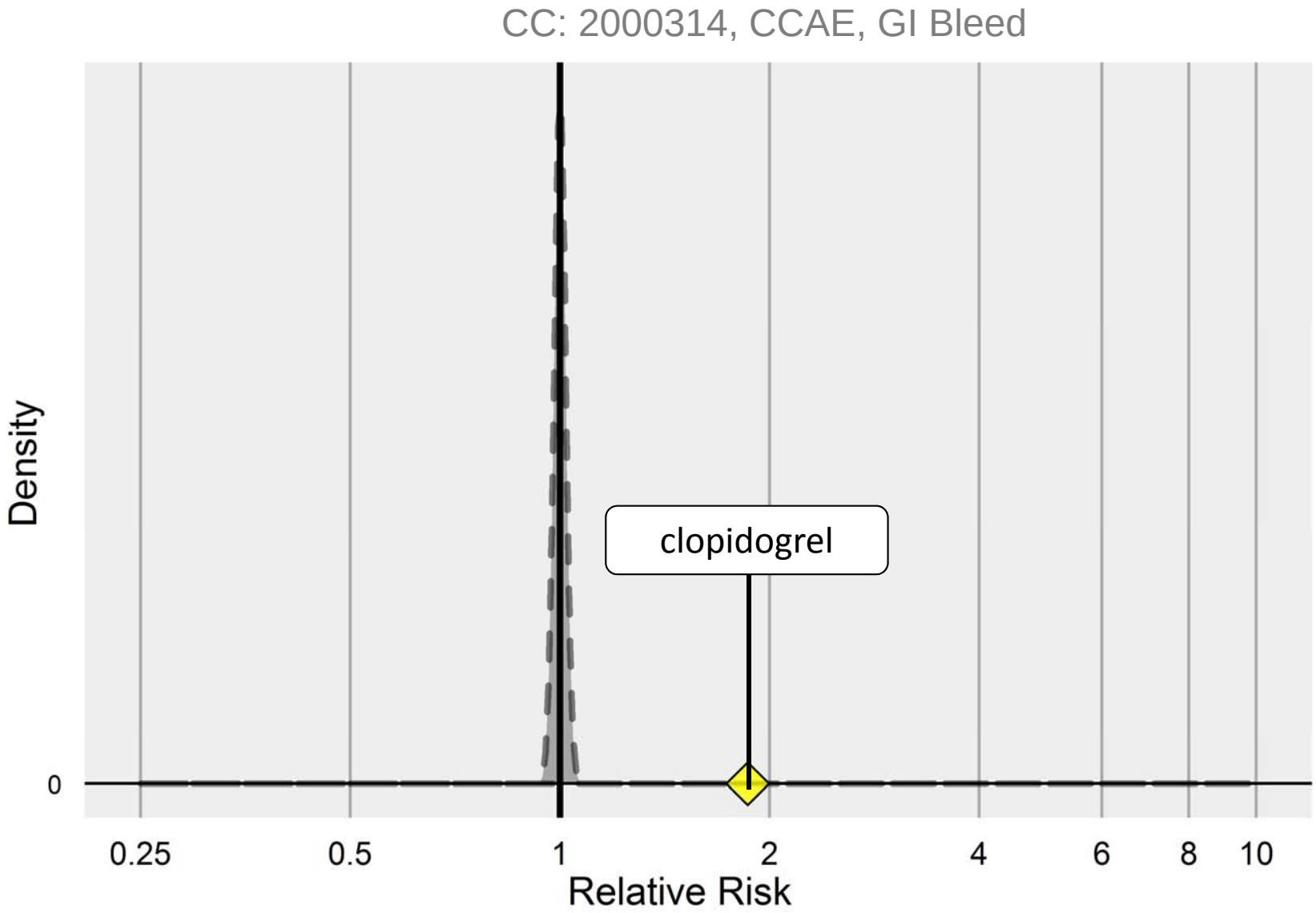
	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

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

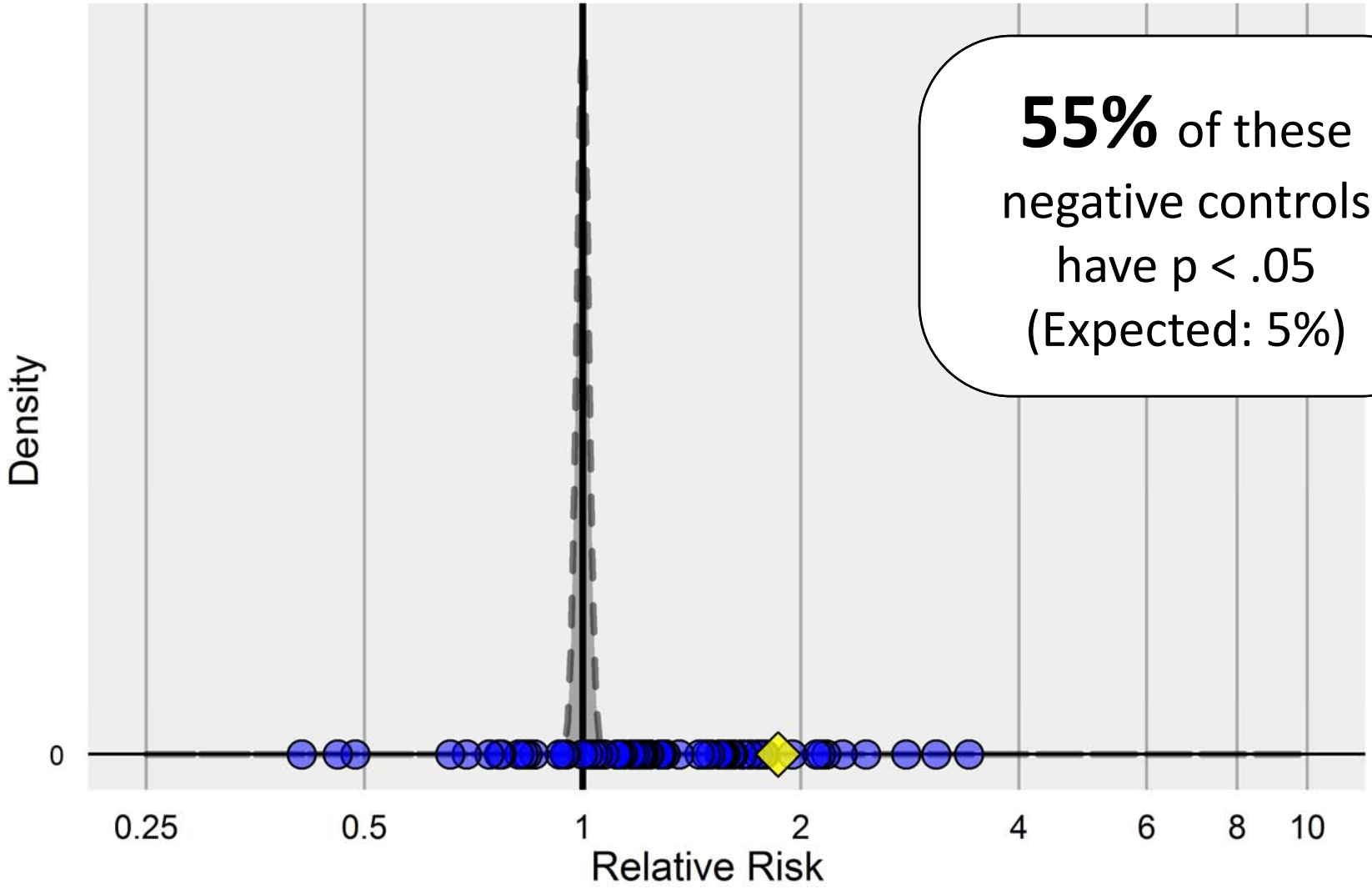
Negative controls & the null distribution



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution

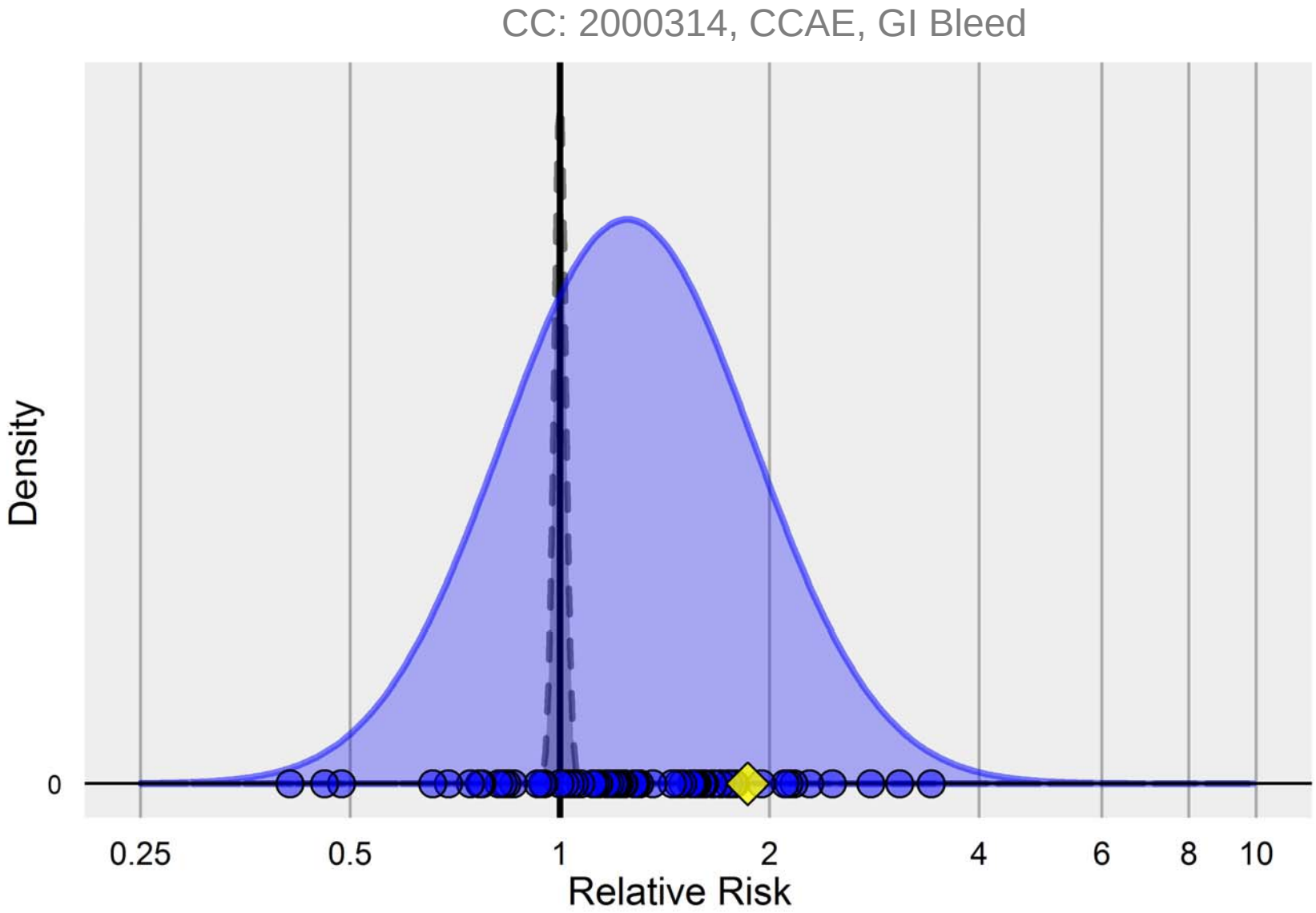
CC: 2000314, CCAE, GI Bleed



55% of these negative controls have p < .05 (Expected: 5%)

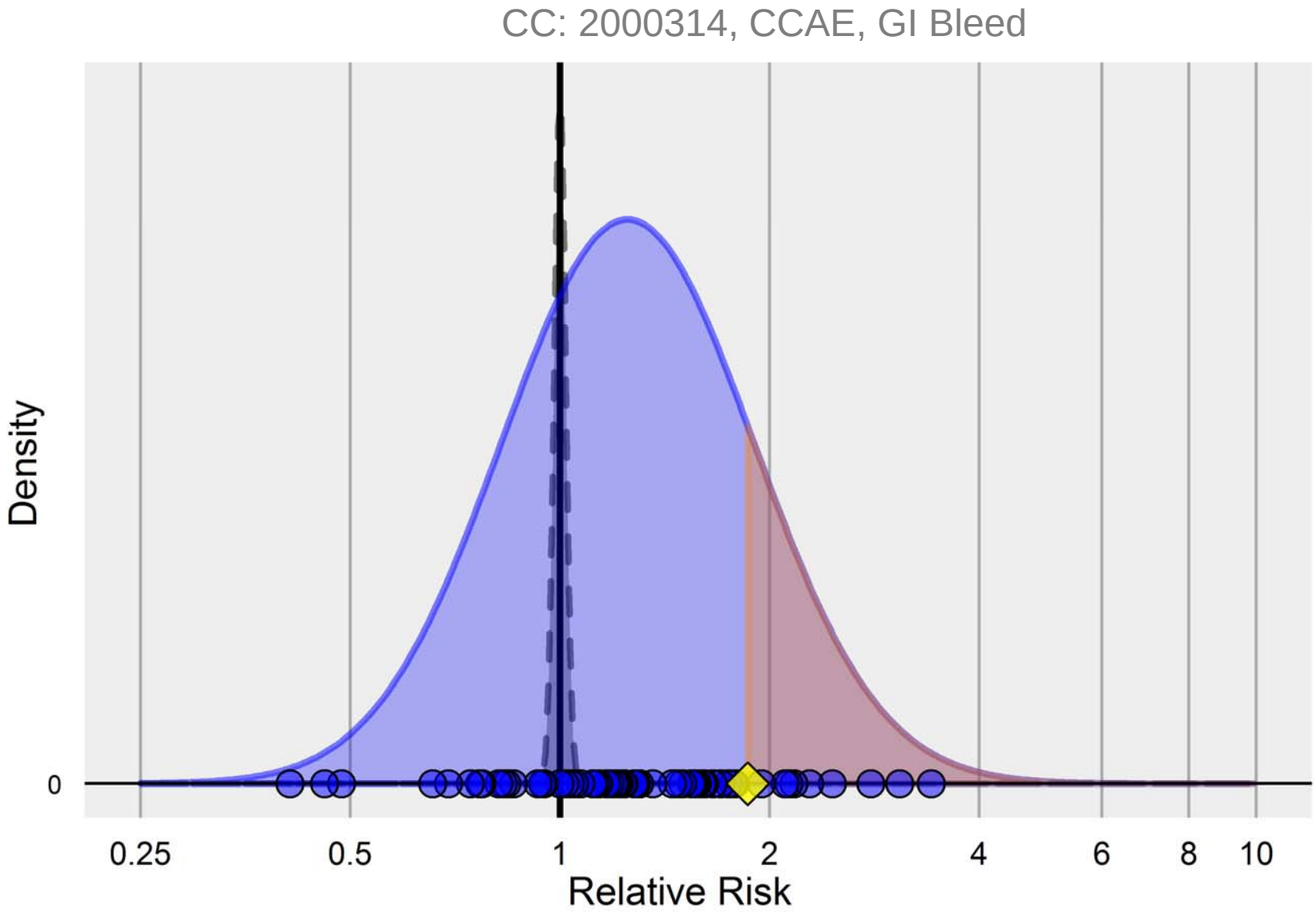
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution



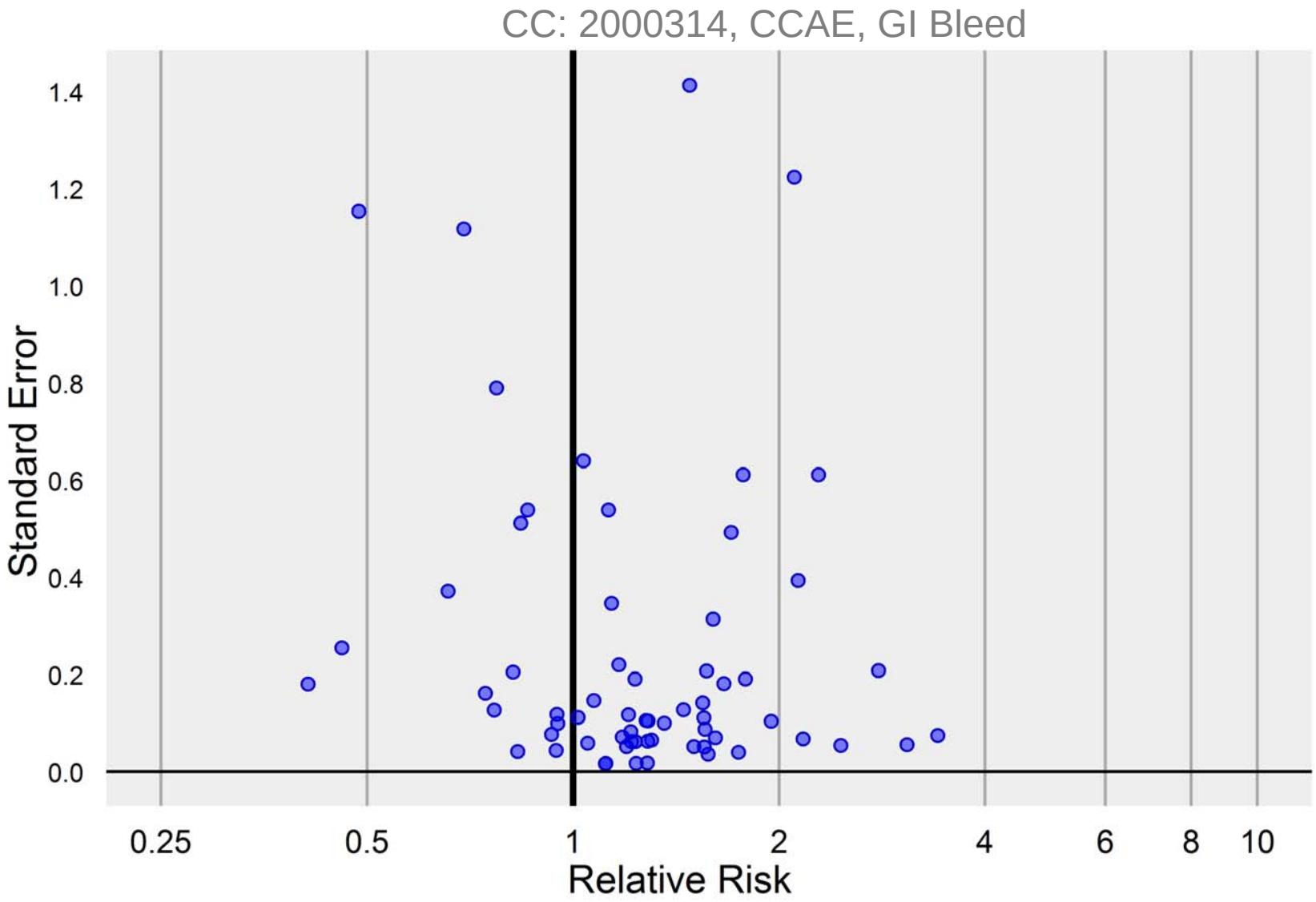
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Negative controls & the null distribution



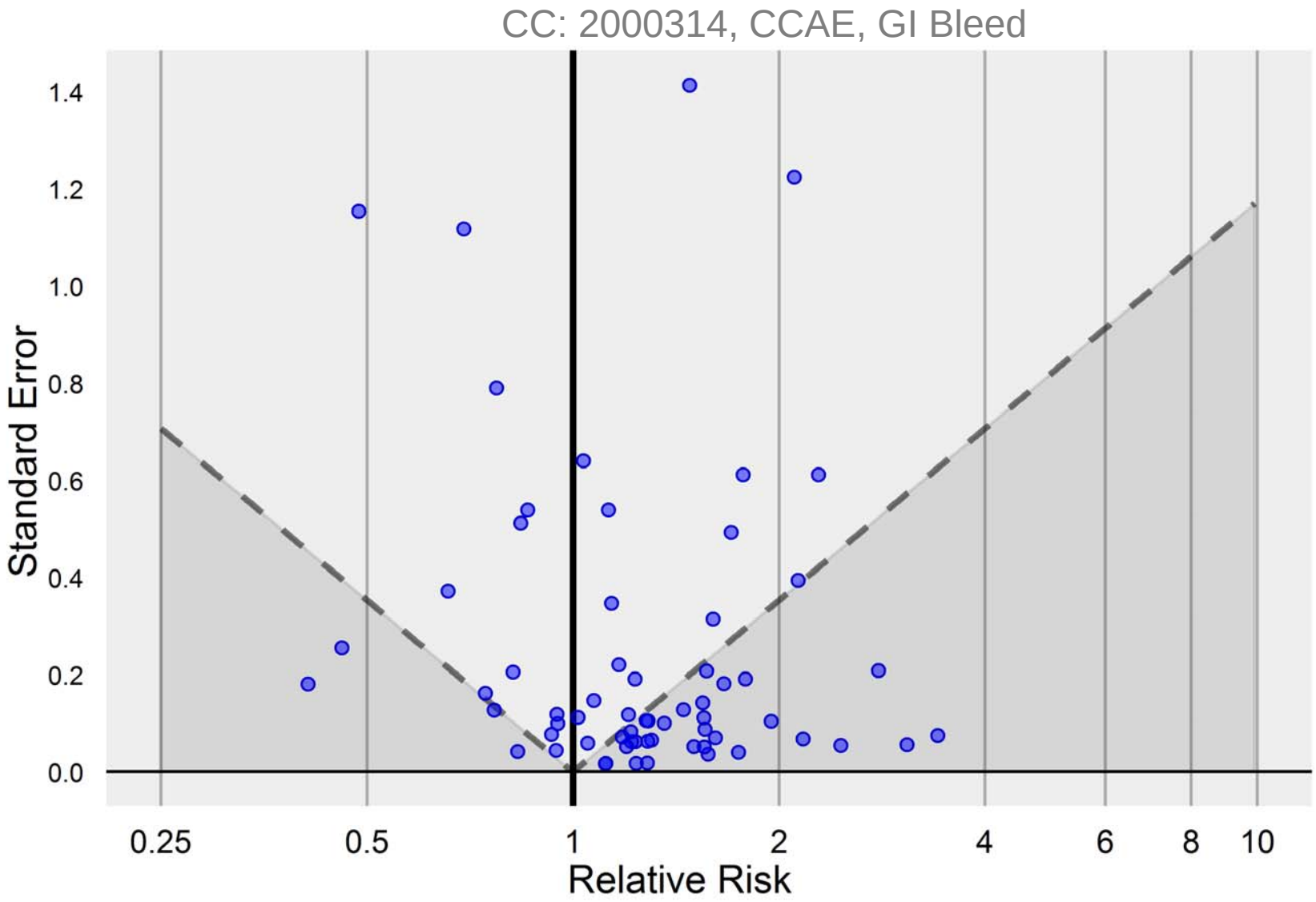
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot



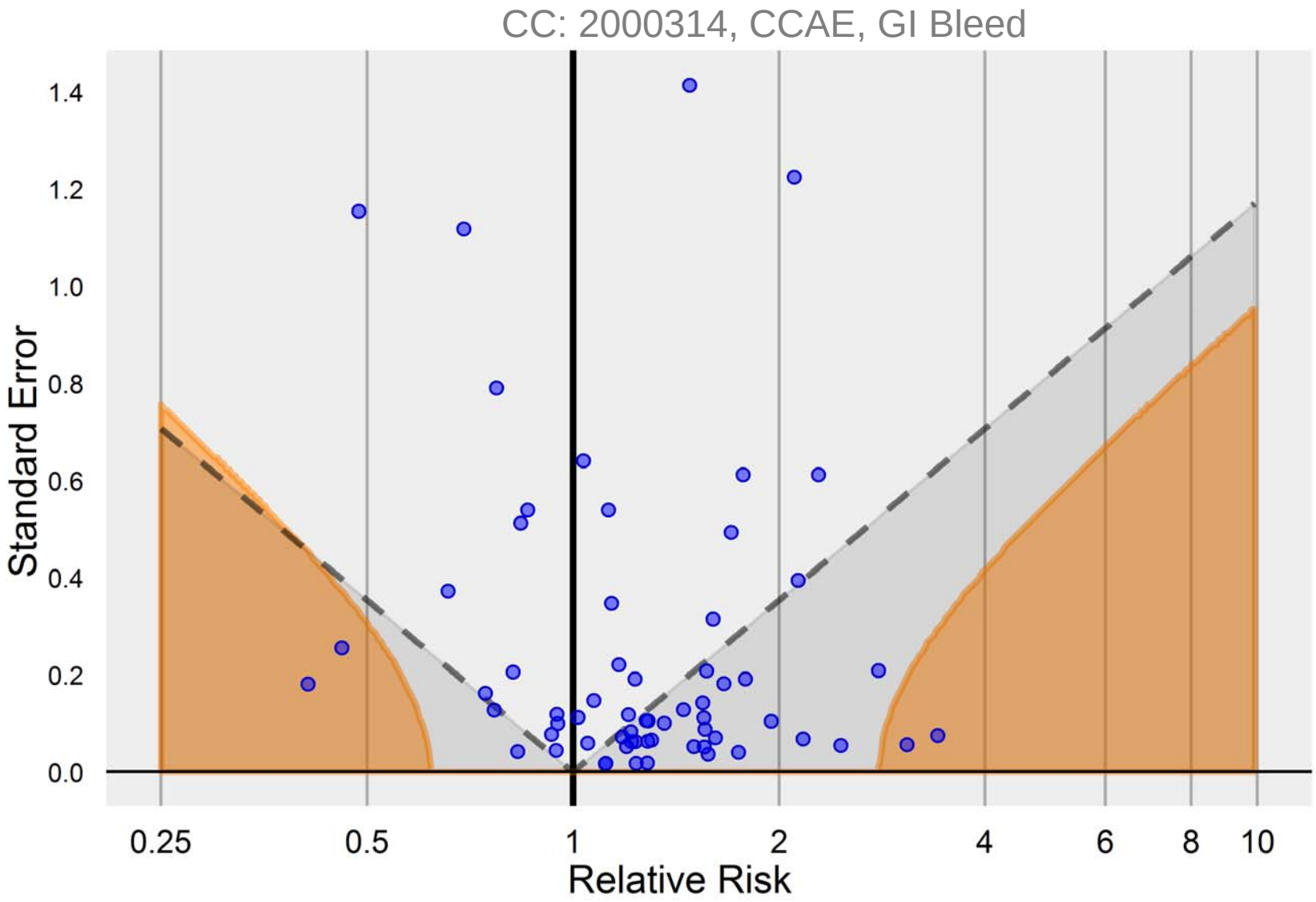
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

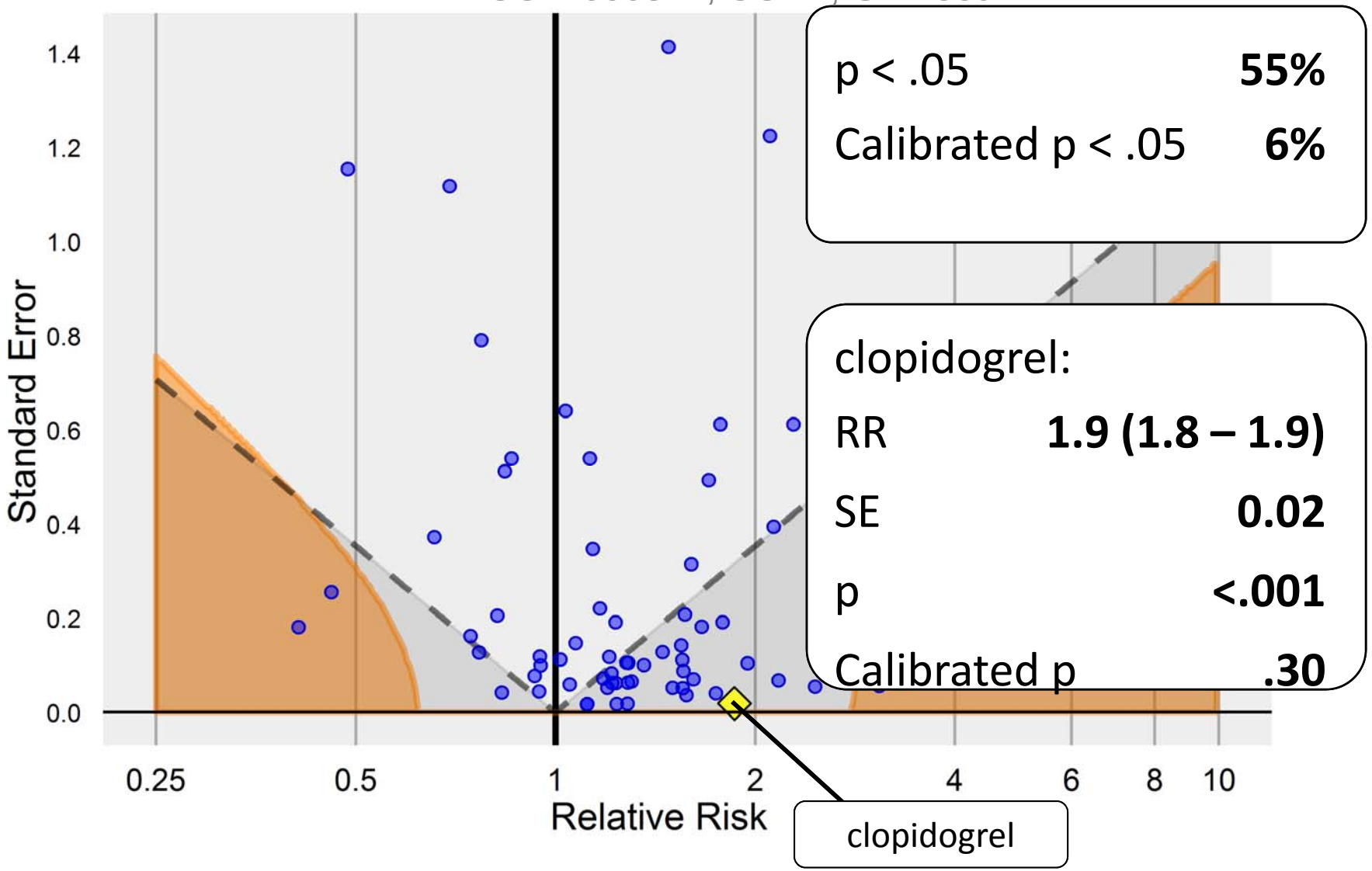
p-value calibration plot



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

p-value calibration plot

CC: 2000314, CCAE, GI Bleed



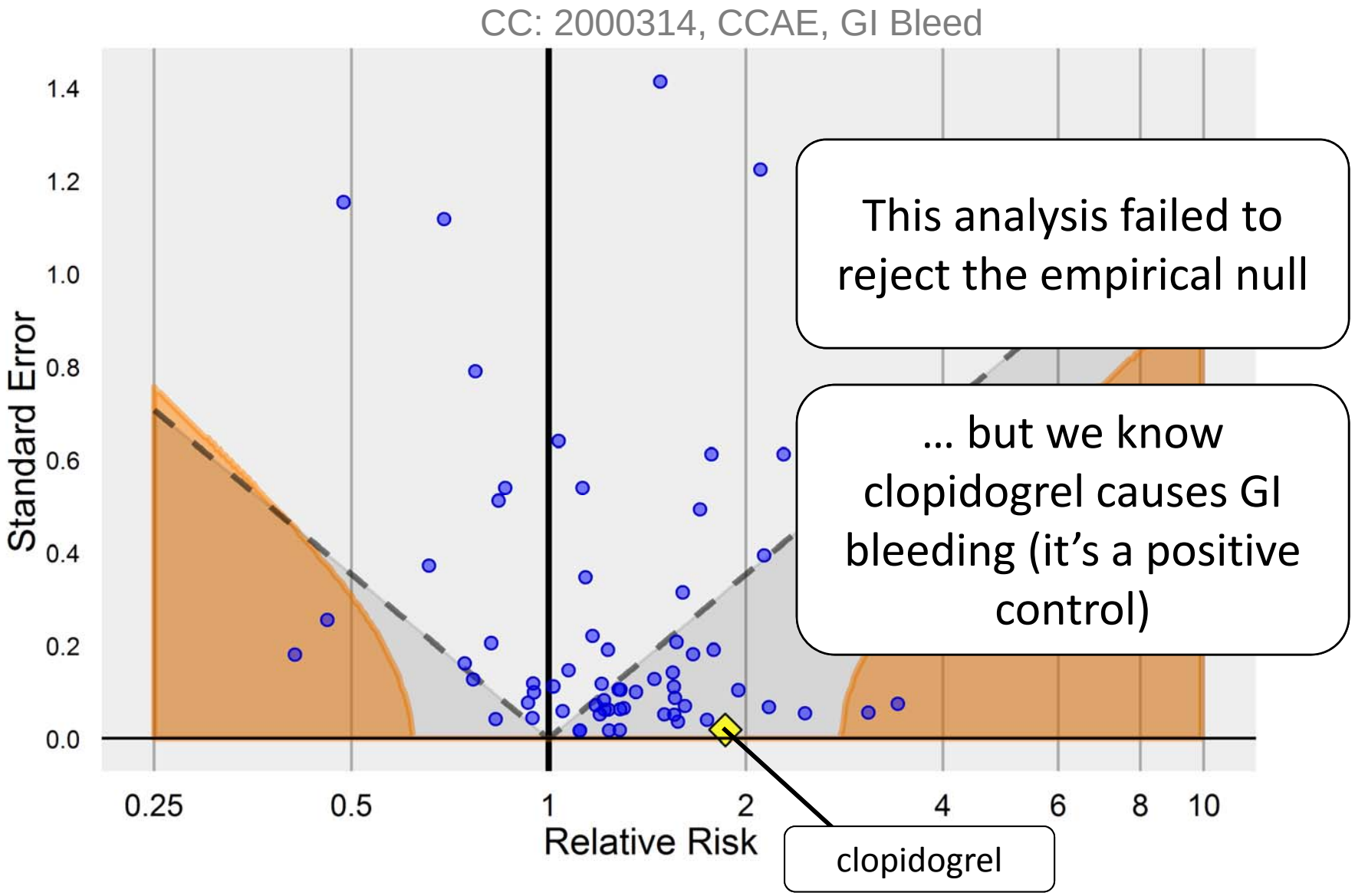
p < .05	**55%**
Calibrated p < .05	**6%**

clopidogrel:

RR	**1.9 (1.8 – 1.9)**
SE	**0.02**
p	**<.001**
Calibrated p	**.30**

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

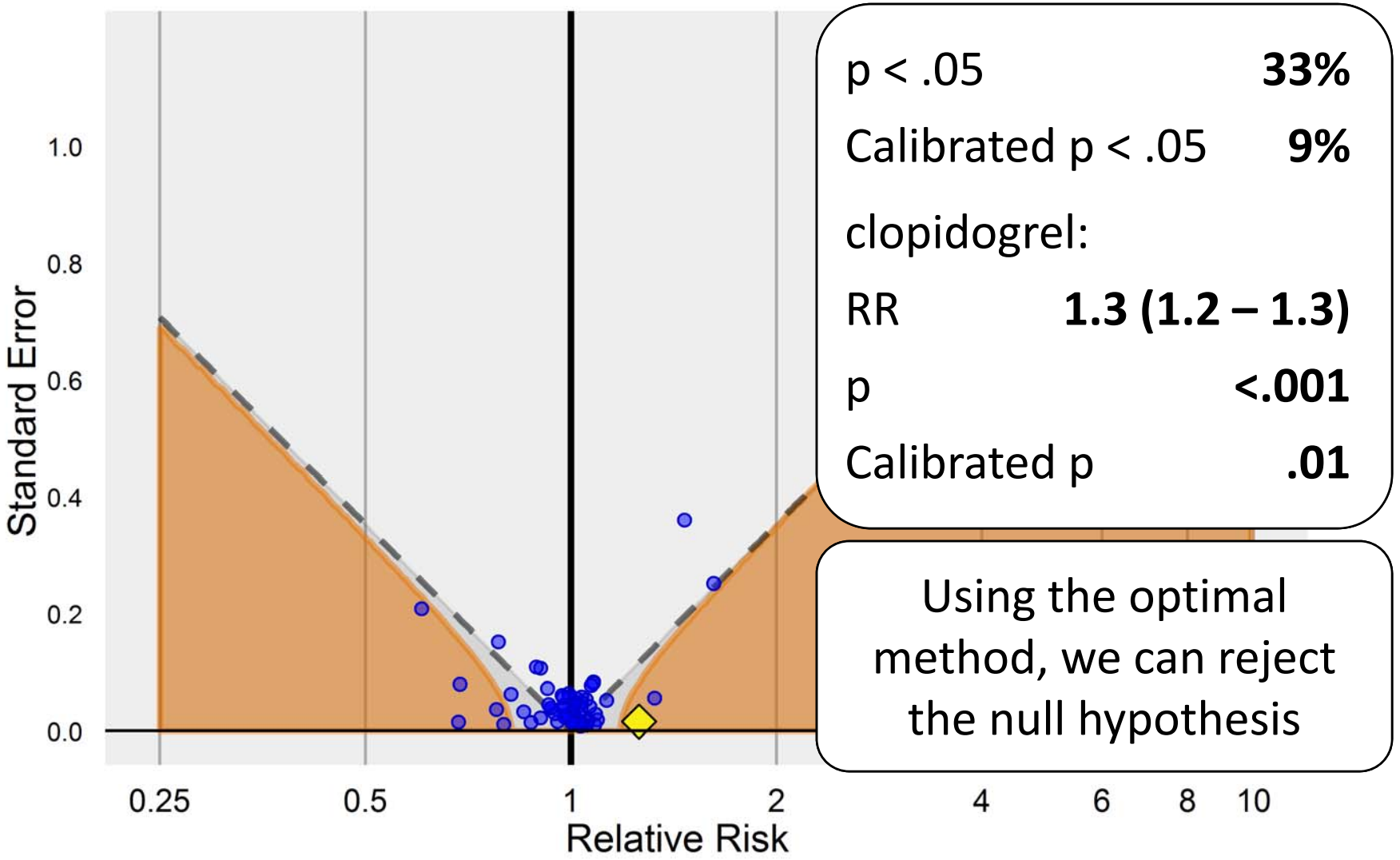
p-value calibration plot



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

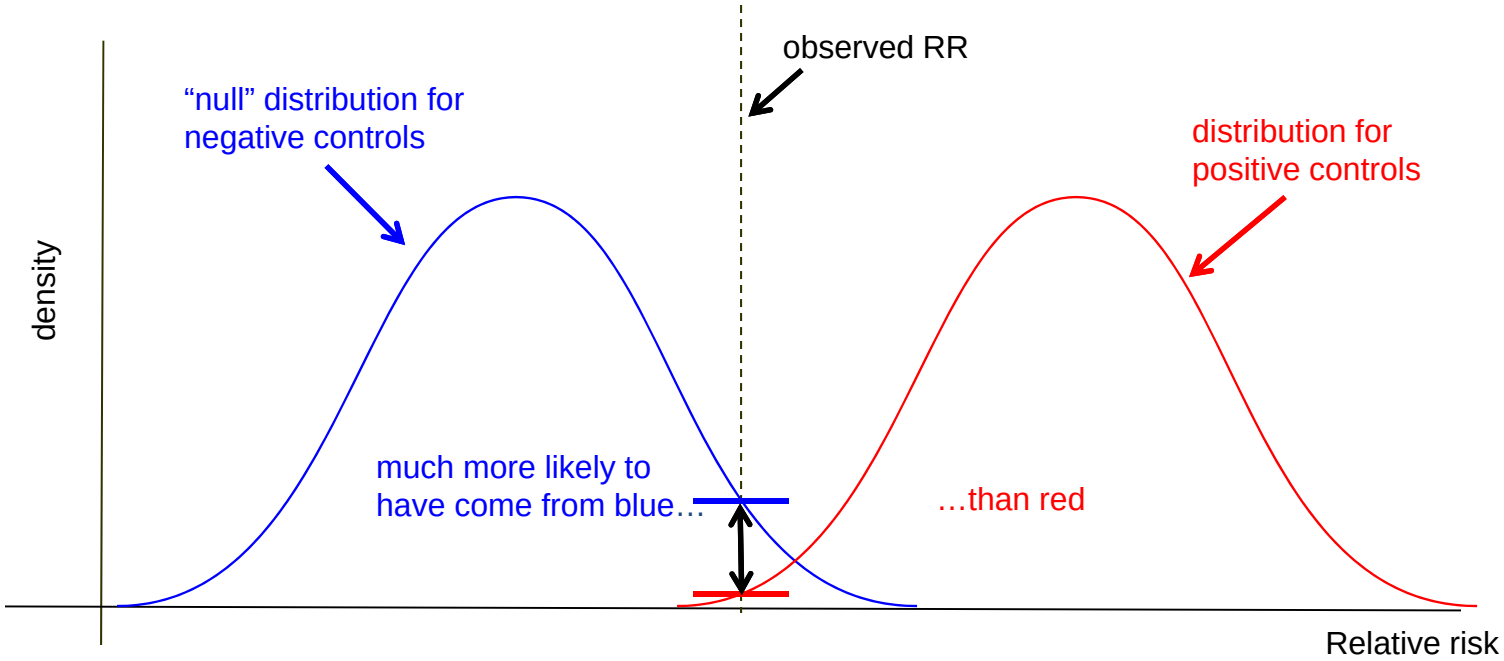
Recap

- 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'

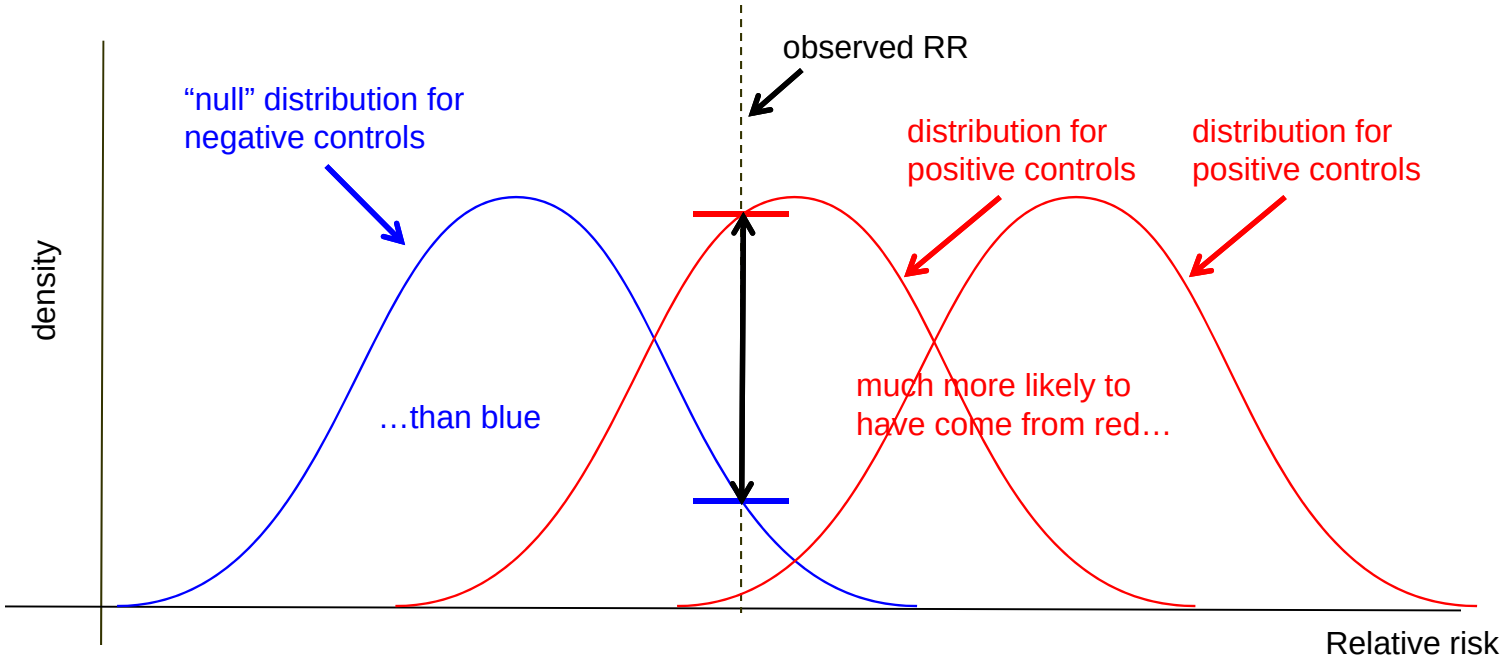
Beyond p-values: Computing the probability of a true association

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

We also have positive controls



But if AUC is small…



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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)

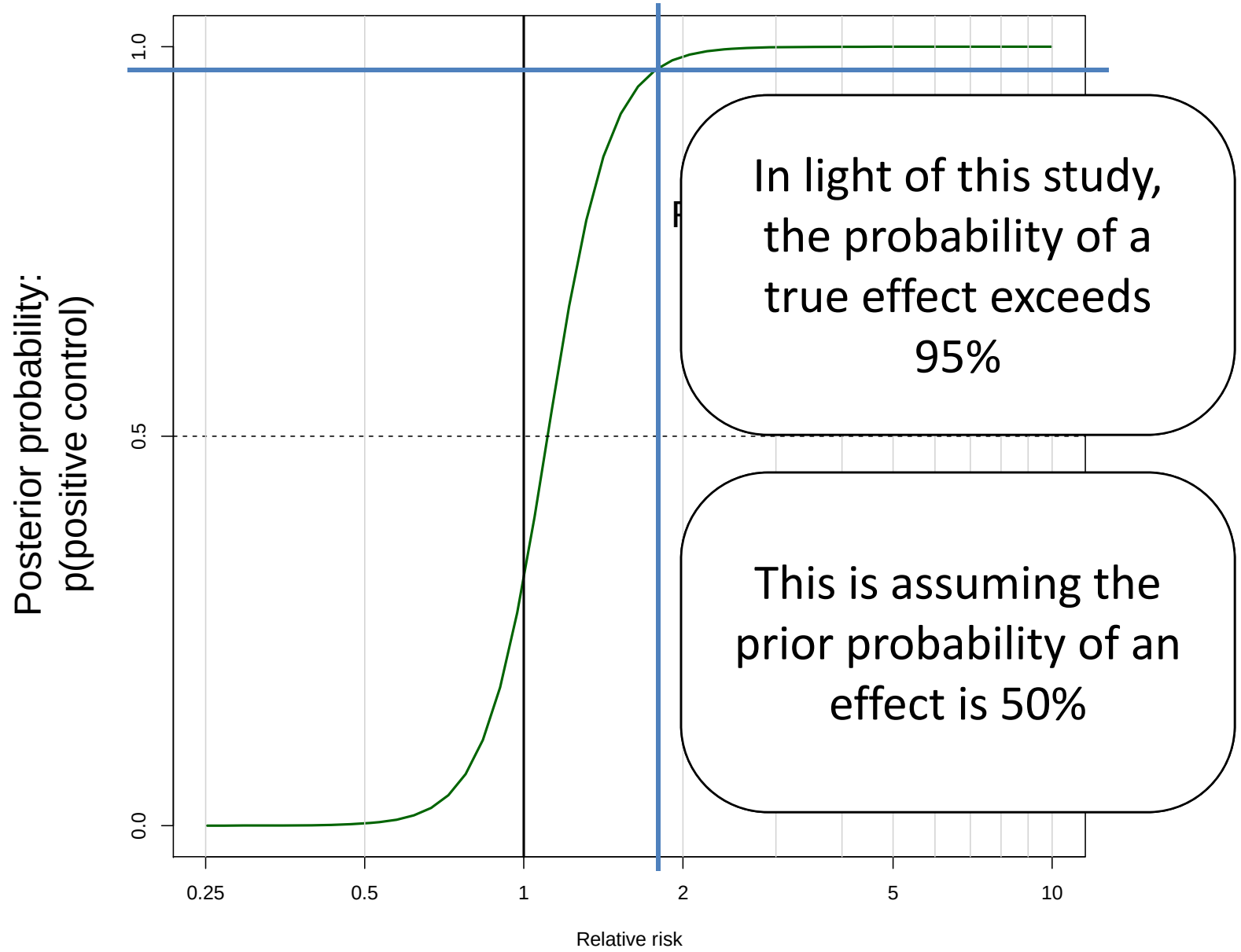
Relative risk: 1.86, 95% CI: 1.79 – 1.93

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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Clopidogrel – GI Bleed

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



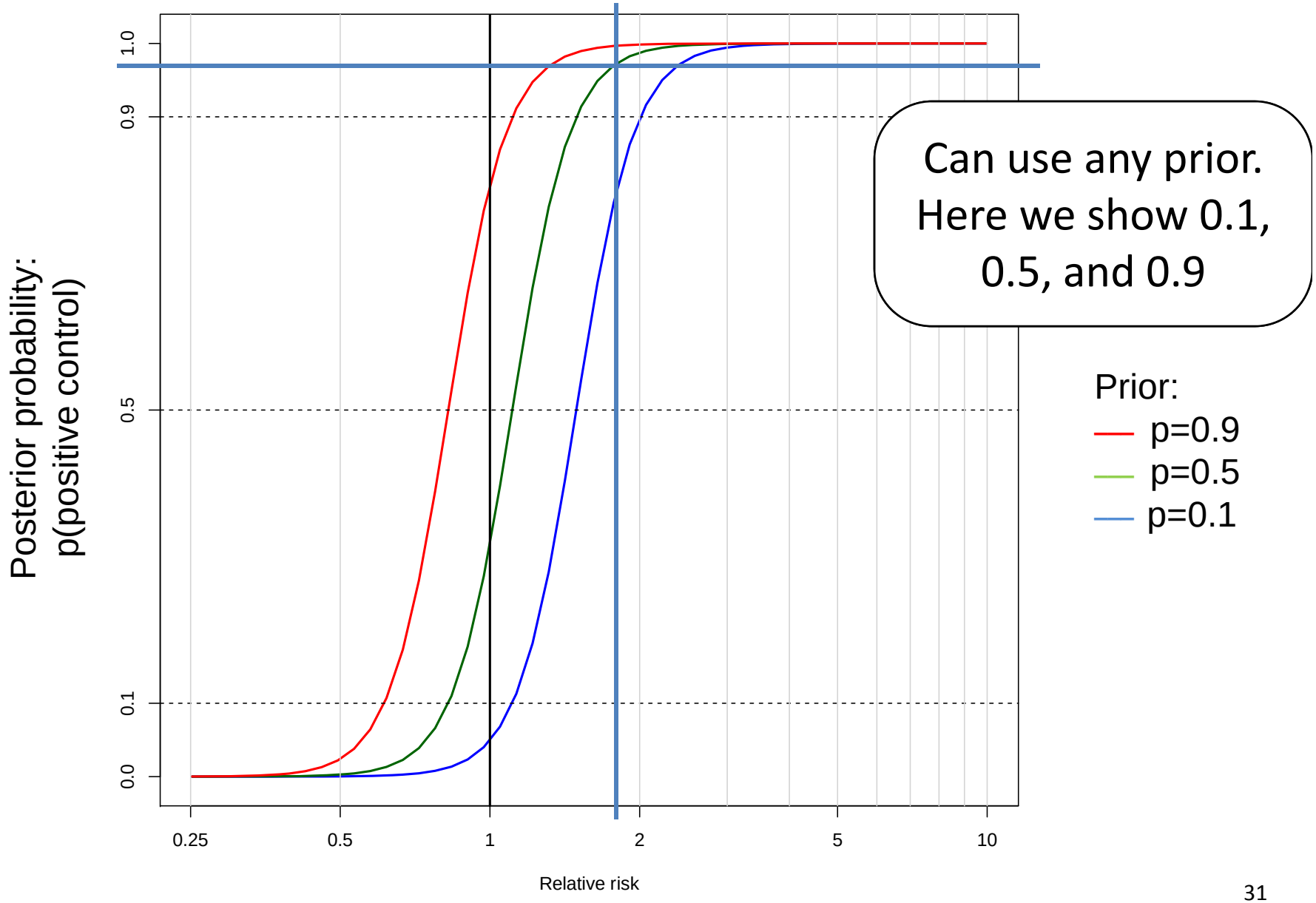
In light of this study, the probability of a true effect exceeds 95%

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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Clopidogrel – GI Bleed

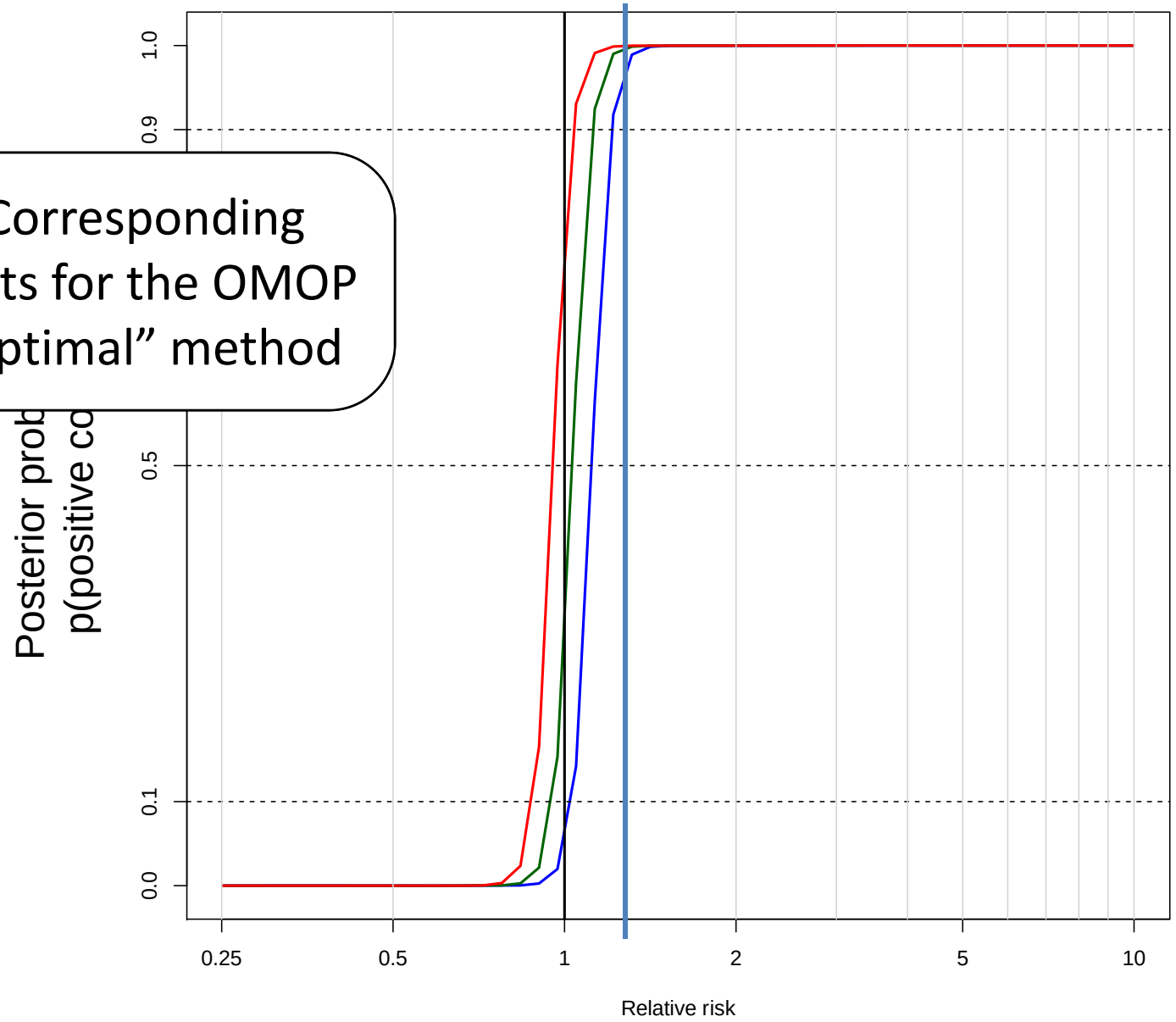
Method: CC−2000314, Source: CCAE, HOI: GI Bleed



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

Conclusions

- Calibration of p-values, using an empirical null distribution, in order to take into account the biases in database studies, may be feasible
- It is possible to calculate the posterior probability of an association, given a prior belief and the observed data