OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Are Observational Studies Any Good?

David Madigan, Columbia University
on behalf of the OMOP research team

“The sole cause and root of almost every defect in the sciences is this: that whilst we falsely admire and extol the powers of the human mind, we do not search for its real helps.”

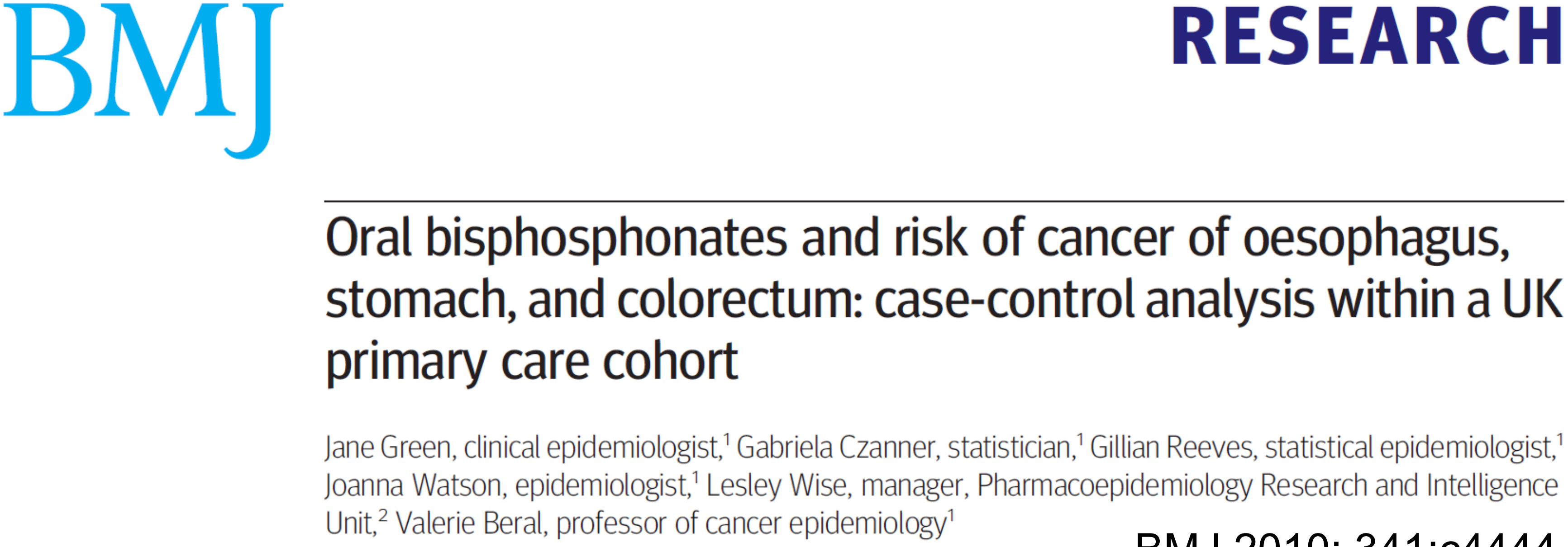
— Novum Organum: Aphorisms [Book One], 1620, Sir Francis Bacon

Observational Studies

- A empirical study in which:

> “The objective is to elucidate cause-and-effect relationships in which it is not feasible to use controlled experimentation”

- Examples:
  - smoking and heart disease
  - vitamin C and cancer survival
  - DES and vaginal cancer
  - aspirin and mortality
  - cocaine and birthweight
  - diet and mortality

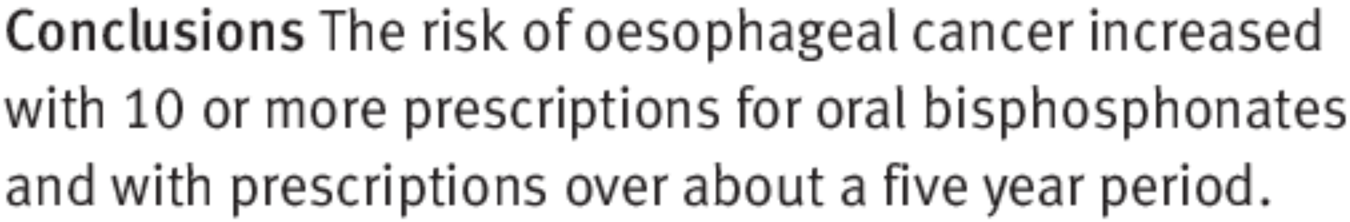

BMJ

RESEARCH

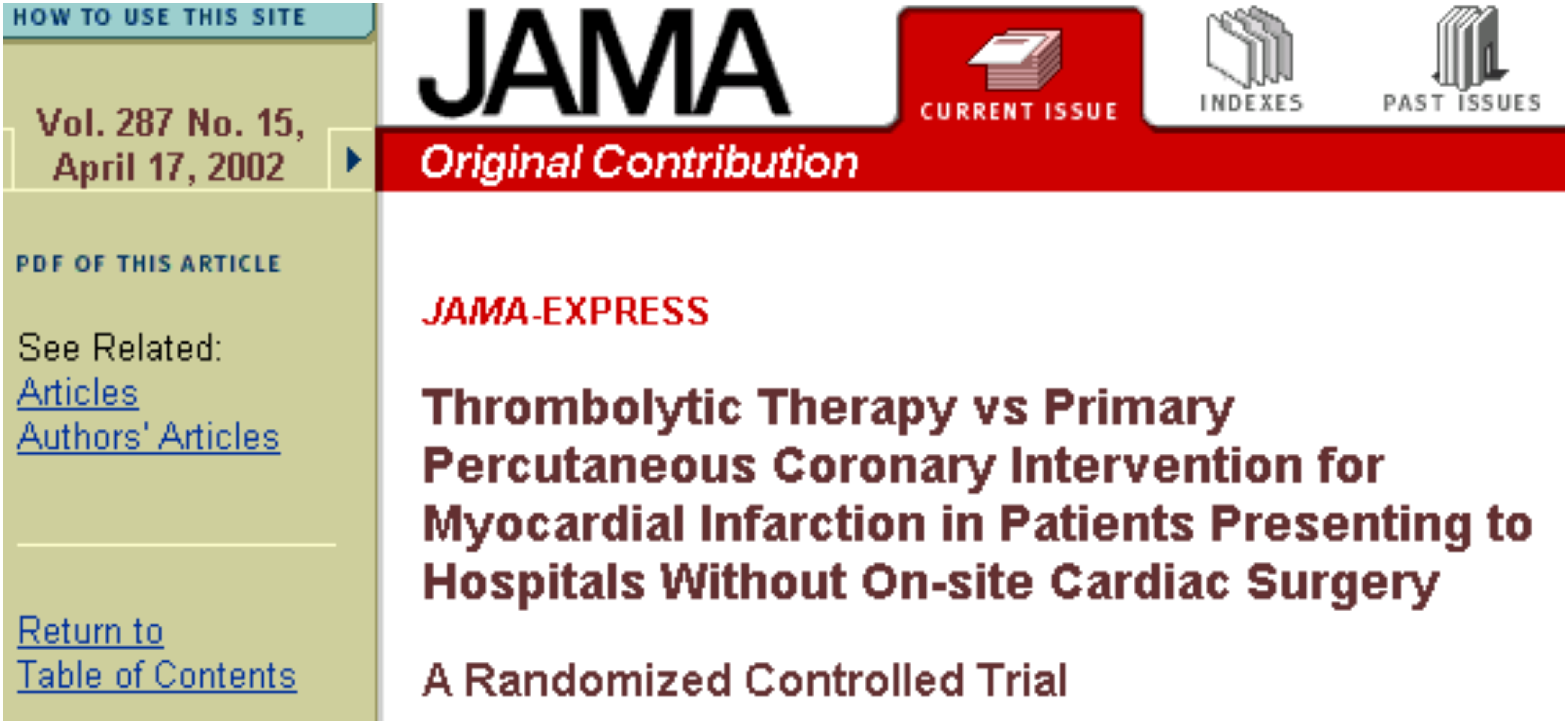
Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort

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]

BMJ 2010; 341:c4444


Conclusions The risk of oesophageal cancer increased with 10 or more prescriptions for oral bisphosphonates and with prescriptions over about a five year period.

Why does randomization work?



Vol. 287 No. 15,
April 17, 2002

Original Contribution

JAMA-EXPRESS

Thrombolytic Therapy vs Primary Percutaneous Coronary Intervention for Myocardial Infarction in Patients Presenting to Hospitals Without On-site Cardiac Surgery

A Randomized Controlled Trial

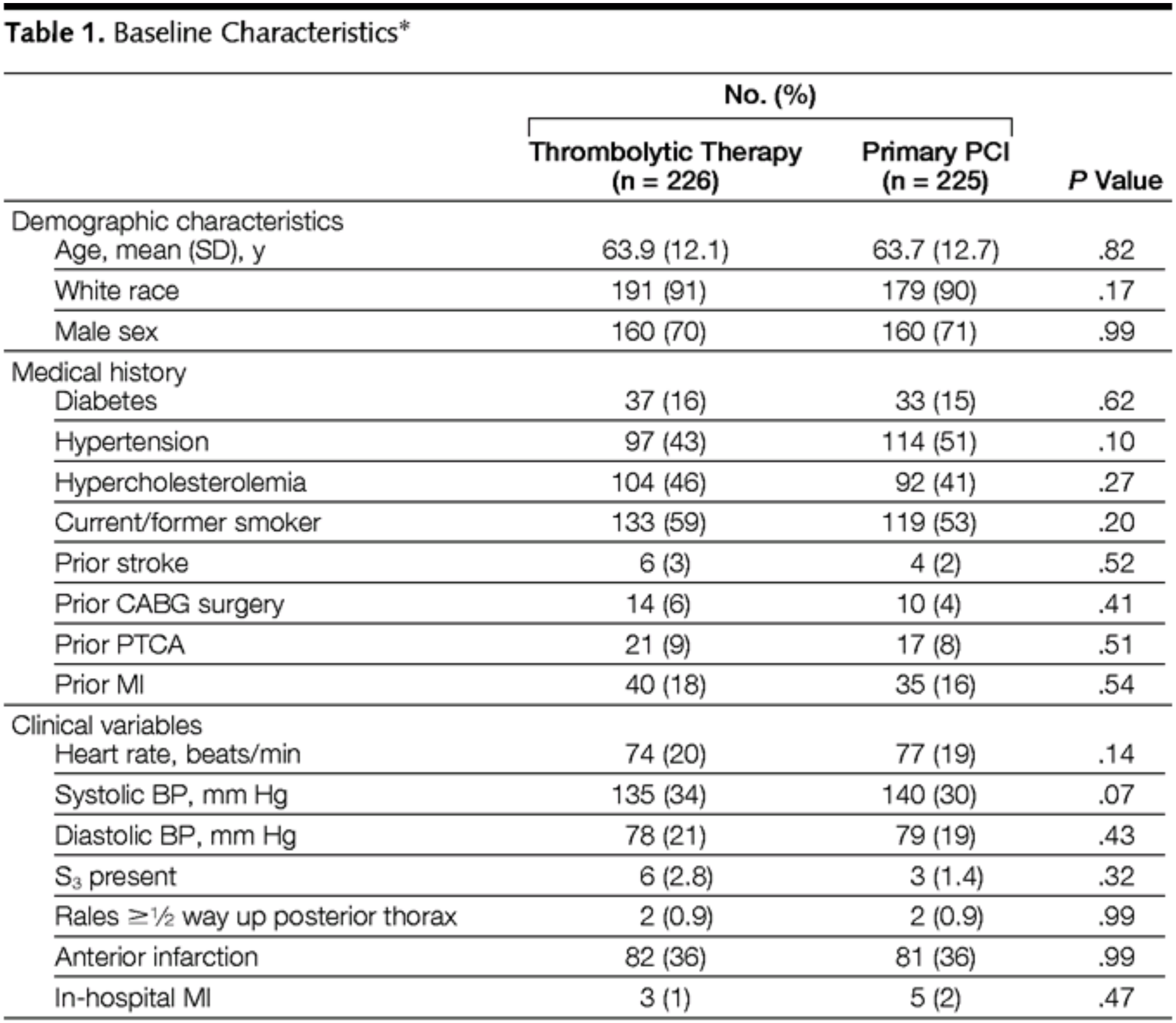
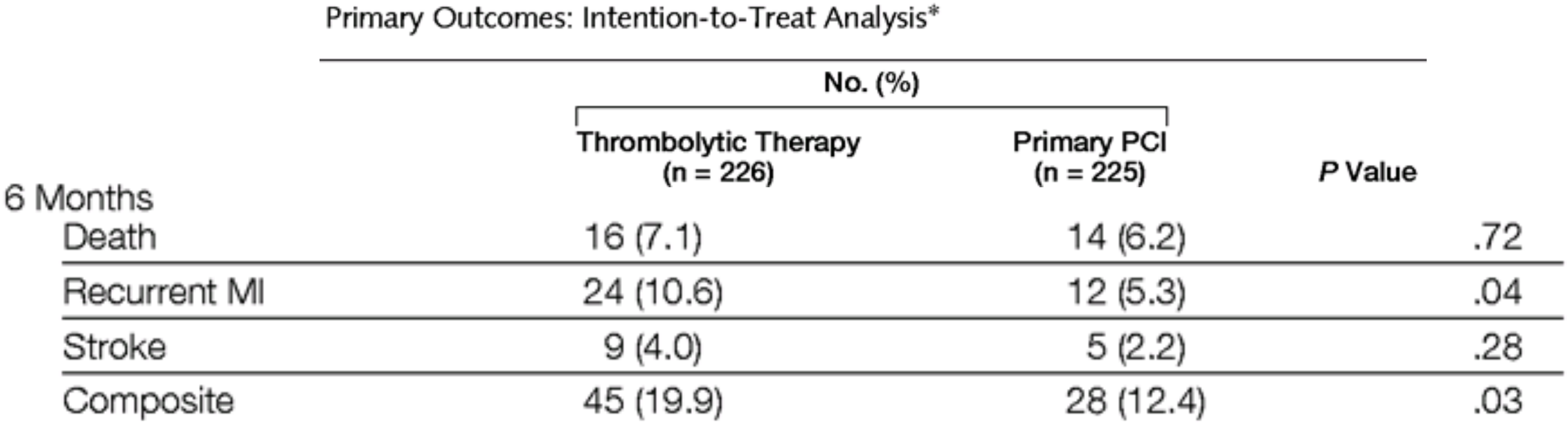


Table 1. Baseline Characteristics*

	No. (%)		
	Thrombolytic Therapy (n = 226)	**Primary PCI (n = 225)**	***P* Value**
Demographic characteristics			
Age, mean (SD), y	63.9 (12.1)	63.7 (12.7)	.82
White race	191 (91)	179 (90)	.17
Male sex	160 (70)	160 (71)	.99
Medical history			
Diabetes	37 (16)	33 (15)	.62
Hypertension	97 (43)	114 (51)	.10
Hypercholesterolemia	104 (46)	92 (41)	.27
Current/former smoker	133 (59)	119 (53)	.20
Prior stroke	6 (3)	4 (2)	.52
Prior CABG surgery	14 (6)	10 (4)	.41
Prior PTCA	21 (9)	17 (8)	.51
Prior MI	40 (18)	35 (16)	.54
Clinical variables			
Heart rate, beats/min	74 (20)	77 (19)	.14
Systolic BP, mm Hg	135 (34)	140 (30)	.07
Diastolic BP, mm Hg	78 (21)	79 (19)	.43
S_3 present	6 (2.8)	3 (1.4)	.32
Rales ≥½ way up posterior thorax	2 (0.9)	2 (0.9)	.99
Anterior infarction	82 (36)	81 (36)	.99
In-hospital MI	3 (1)	5 (2)	.47



Primary Outcomes: Intention-to-Treat Analysis*

	No. (%)		
	Thrombolytic Therapy (n = 226)	Primary PCI (n = 225)	*P* Value
6 Months			
Death	16 (7.1)	14 (6.2)	.72
Recurrent MI	24 (10.6)	12 (5.3)	.04
Stroke	9 (4.0)	5 (2.2)	.28
Composite	45 (19.9)	28 (12.4)	.03

•The two groups are comparable at baseline

•Could do a better job manually matching patients on 18 characteristics listed, but no guarantees for other characteristics

•Randomization did a good job without being told what the 18 characteristics were

•Chance assignment could create some imbalances but the statistical methods account for this properly

The Hypothesis of No Treatment Effect

- In a randomized experiment, can test this hypothesis essentially without making any assumptions at all
- “no effect” formally means for each patient the outcome would have been the same regardless of treatment assignment
- Test statistic, e.g., proportion (D|TT)-proportion(D|PCI)

TT	D
TT	D
PCI	L
PCI	L

observed

TT	D
PCI	D
TT	L
PCI	L

TT	D
PCI	D
PCI	L
TT	L

PCI	D
TT	D
TT	L
PCI	L

PCI	D
TT	D
PCI	L
TT	L

PCI	D
PCI	D
TT	L
TT	L

P=1/6

Causal Inference View

- Rubin causal model
  - Potential outcomes

    Factual outcome

      I am a smoker and I get lung cancer

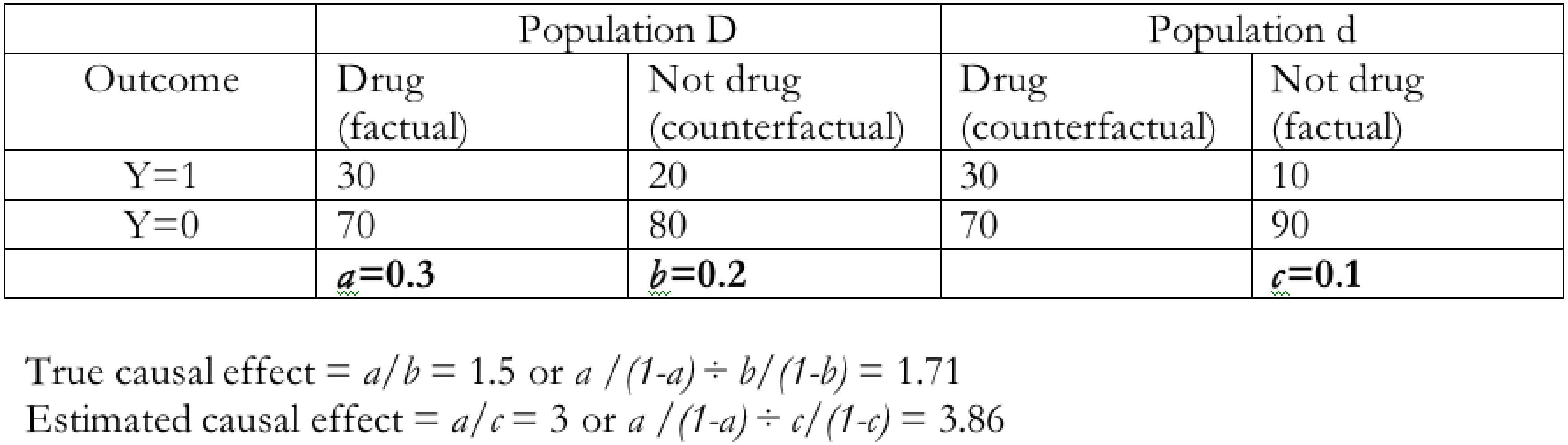
    Counterfactual outcome

      If I had not been a smoker, I would not have gotten lung cancer

- Define:
  - Z_i : treatment applied to unit i (0=control, 1=treat)
  - $Y_i(0)$: response for unit *i* if $Z_i = 0$
  - $Y_i(1)$: response for unit *i* if $Z_i = 1$
  - Unit level causal effect: $Y_i(1) - Y_i(0)$
  - Fundamental problem: only see one of these!

Confounding and Causality

- Confounding is a causal concept



	Population D		Population d	
Outcome	Drug (factual)	Not drug (counterfactual)	Drug (counterfactual)	Not drug (factual)
Y=1	30	20	30	10
Y=0	70	80	70	90
	a=0.3	**b=0.2**		**c=0.1**

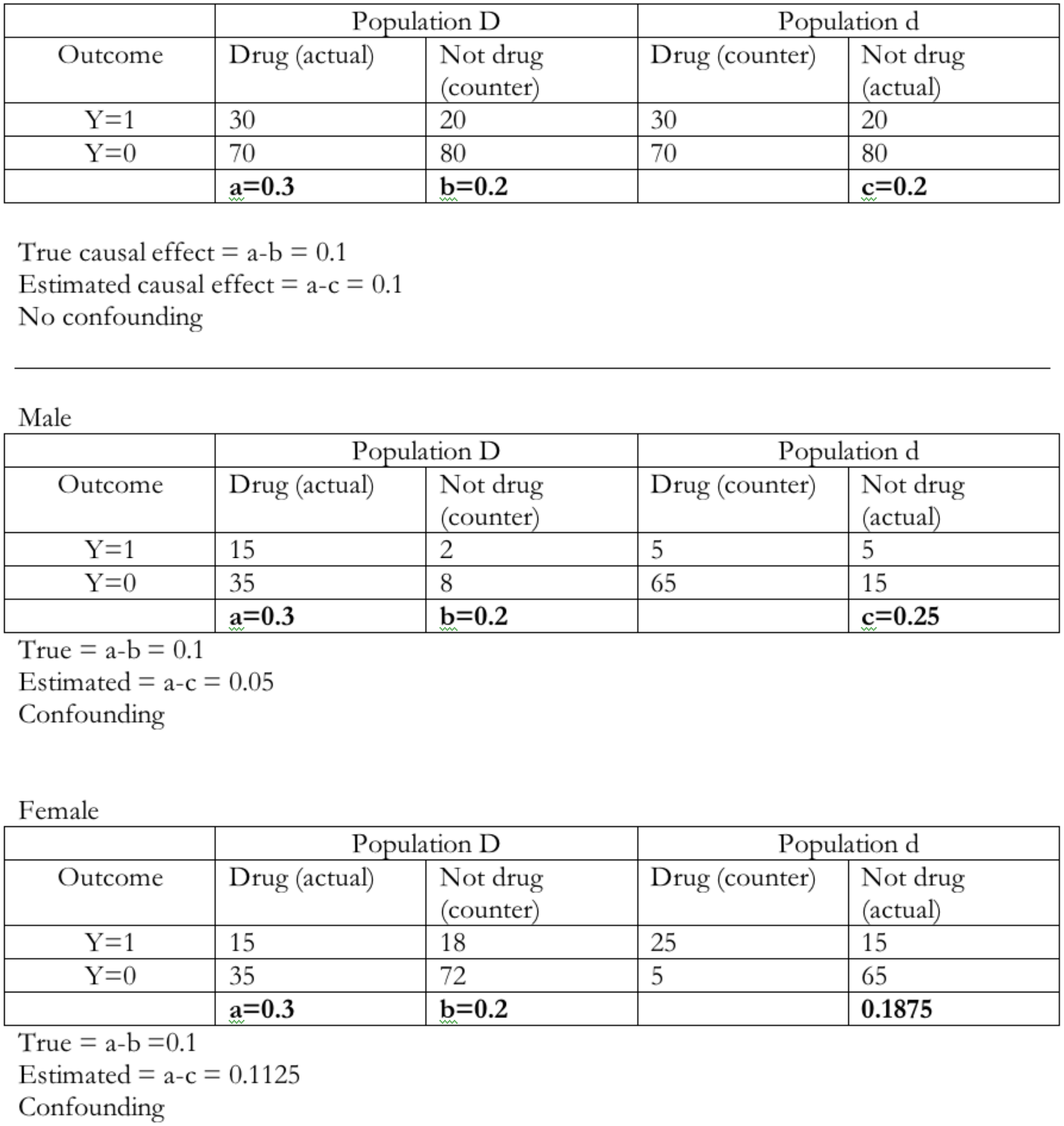
True causal effect = a/b = 1.5 or $a\,/(1\text{-}a) \div b/(1\text{-}b)$ = 1.71
Estimated causal effect = a/c = 3 or $a\,/(1\text{-}a) \div c/(1\text{-}c)$ = 3.86

- “The association in the combined D+d populations is confounded for the effect in population D”

Why does this happen?

- For confounding to occur there must be some characteristics/covariates/conditions that distinguish D from d.
- However, the existence of such factors does not in and of itself imply confounding.
- For example, D could be males and d females but it could still be the case that b=c.

Stratification can introduce confounding



	Population D		Population d	
Outcome	Drug (actual)	Not drug (counter)	Drug (counter)	Not drug (actual)
Y=1	30	20	30	20
Y=0	70	80	70	80
	a=0.3	**b=0.2**		**c=0.2**

True causal effect = a-b = 0.1
Estimated causal effect = a-c = 0.1
No confounding

Male

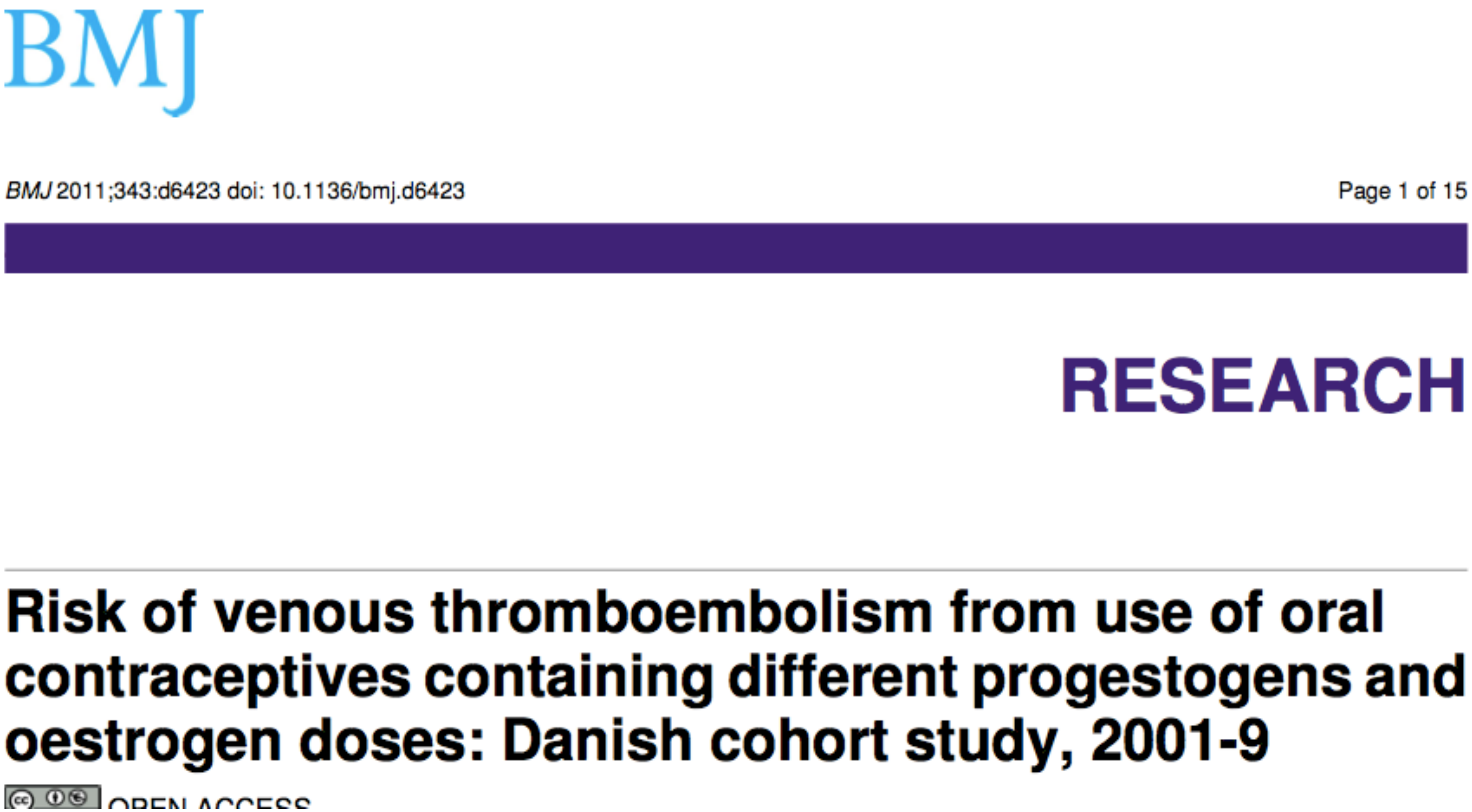
	Population D		Population d	
Outcome	Drug (actual)	Not drug (counter)	Drug (counter)	Not drug (actual)
Y=1	15	2	5	5
Y=0	35	8	65	15
	a=0.3	**b=0.2**		**c=0.25**

True = a-b = 0.1
Estimated = a-c = 0.05
Confounding

Female

	Population D		Population d	
Outcome	Drug (actual)	Not drug (counter)	Drug (counter)	Not drug (actual)
Y=1	15	18	25	15
Y=0	35	72	5	65
	a=0.3	**b=0.2**		**0.1875**

True = a-b =0.1
Estimated = a-c = 0.1125
Confounding

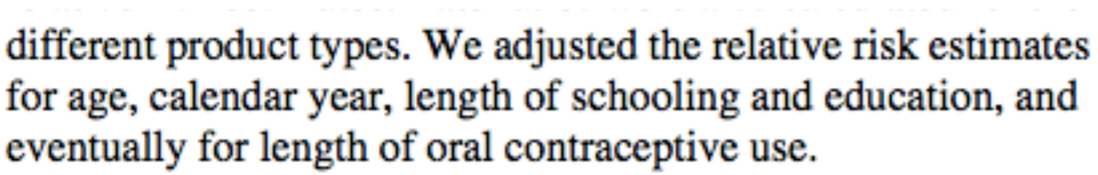


BMJ

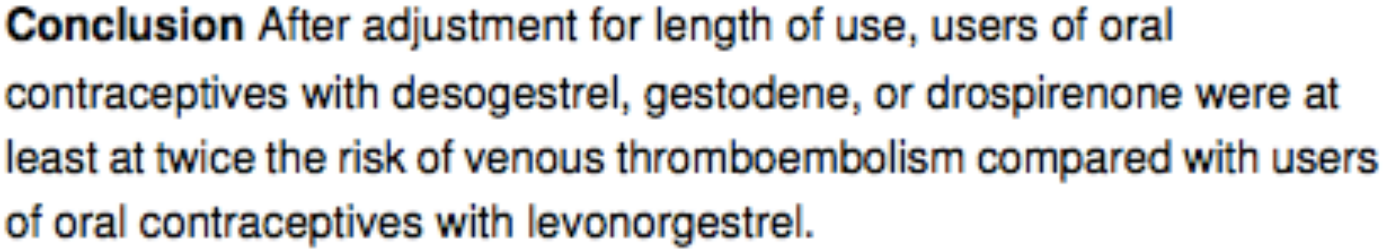
BMJ 2011;343:d6423 doi: 10.1136/bmj.d6423

RESEARCH

Risk of venous thromboembolism from use of oral contraceptives containing different progestogens and oestrogen doses: Danish cohort study, 2001-9



different product types. We adjusted the relative risk estimates for age, calendar year, length of schooling and education, and eventually for length of oral contraceptive use.



Conclusion After adjustment for length of use, users of oral contraceptives with desogestrel, gestodene, or drospirenone were at least at twice the risk of venous thromboembolism compared with users of oral contraceptives with levonorgestrel.

ELSEVIER

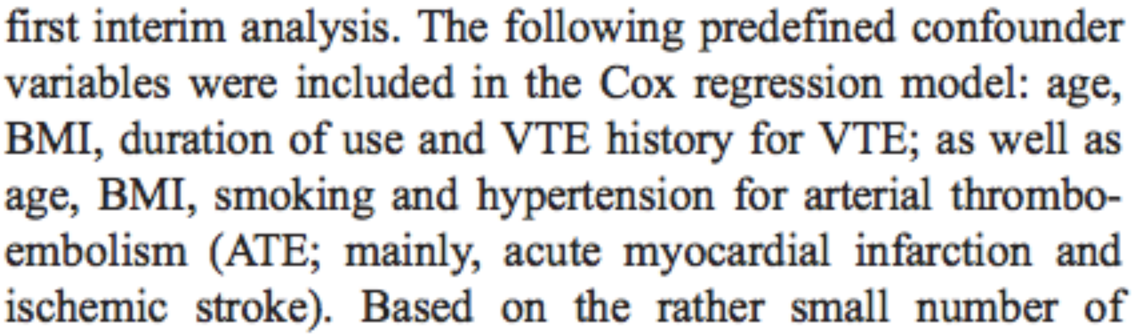
Contraception 75 (2007) 344–354

Contraception

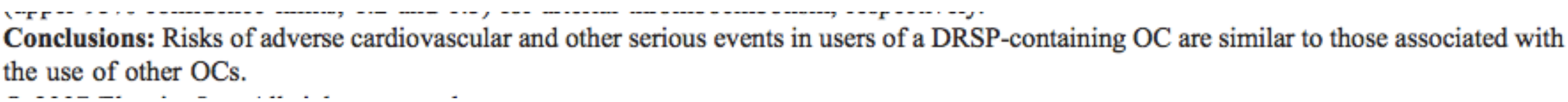
Original research article

The safety of a drospirenone-containing oral contraceptive: final results from the European Active Surveillance study on Oral Contraceptives based on 142,475 women-years of observation

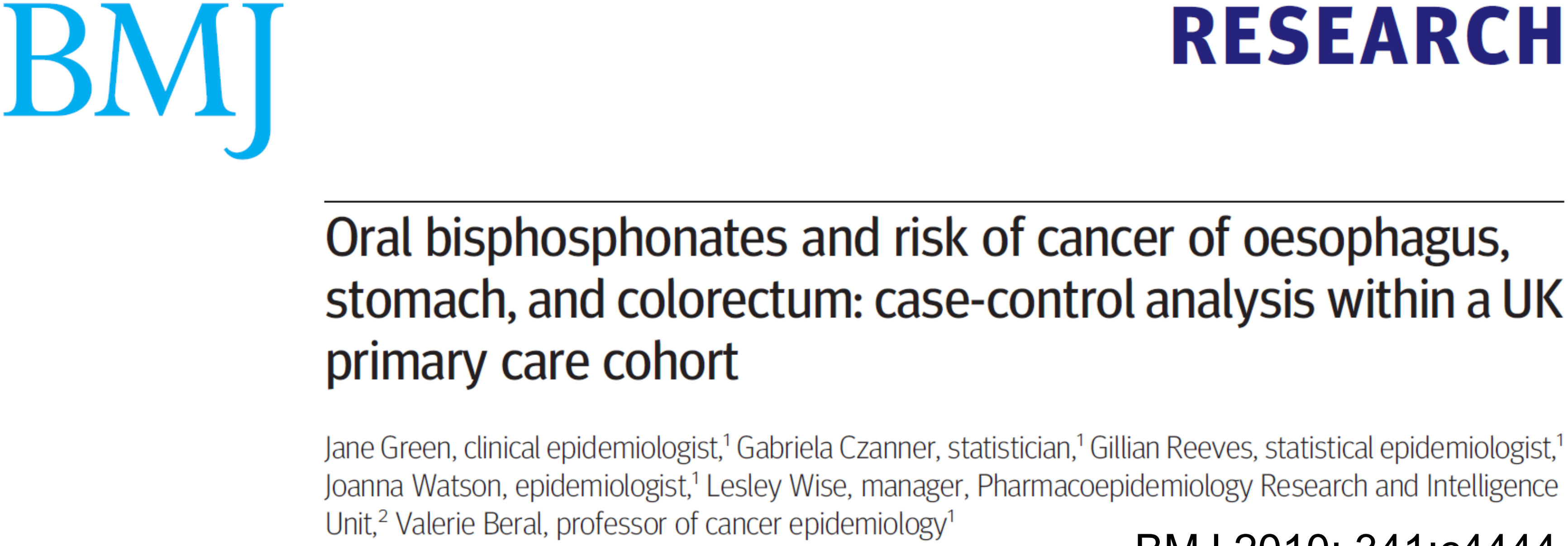
Jürgen C. Dinger[a,*], Lothar A.J. Heinemann[a], Dörthe Kühl-Habich[b]



first interim analysis. The following predefined confounder variables were included in the Cox regression model: age, BMI, duration of use and VTE history for VTE; as well as age, BMI, smoking and hypertension for arterial thromboembolism (ATE; mainly, acute myocardial infarction and ischemic stroke). Based on the rather small number of



Conclusions: Risks of adverse cardiovascular and other serious events in users of a DRSP-containing OC are similar to those associated with the use of other OCs.



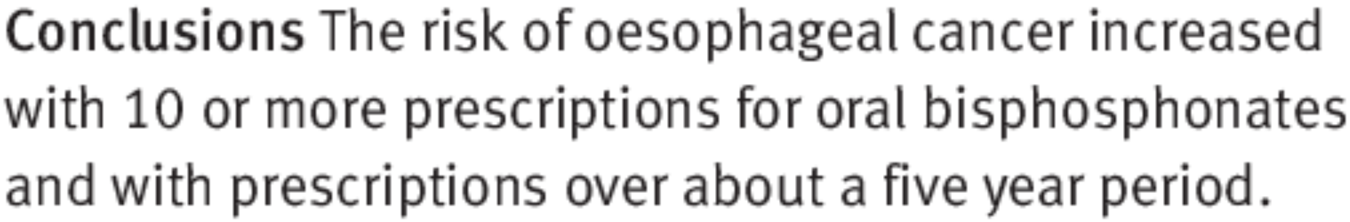
BMJ

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]

BMJ 2010; 341:c4444



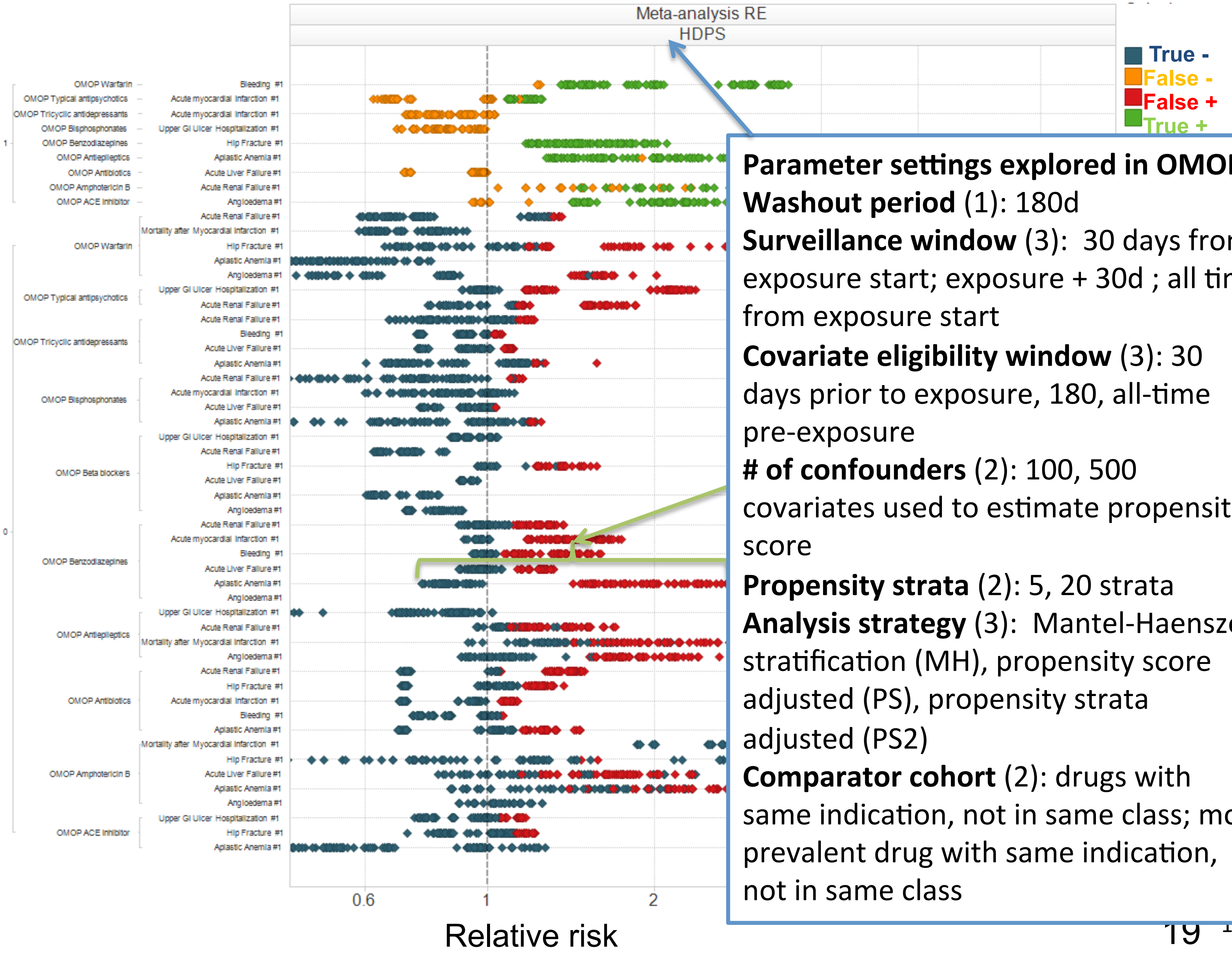
Conclusions The risk of oesophageal cancer increased with 10 or more prescriptions for oral bisphosphonates and with prescriptions over about a five year period.

BMJ study design choices

- Data source: General Practice Research Database
- Study design: Nested case-control
- Inclusion criteria: Age > 40
- Case: cancer diagnosis between 1995-2005 with 12-months of follow-up pre-diagnosis
- 5 controls per case
- Matched on age at index date, sex, practice, observation period prior to index
- Exposure definition: >=1 prescription during observation period
- "RR" estimated with conditional logistic regression
- Covariates: smoking, alcohol, BMI before *outcome* index date
- Sensitivity analyses:
  - exposure = 2+ prescriptions
  - covariates not missing
  - time-at-risk = >1 yr post-exposure
- Subgroup analyses:
  - Short vs. long exposure duration
  - Age, Sex, smoking, alcohol, BMI
  - Osteoporosis or osteopenia
  - Fracture pre-exposure
  - Prior diagnosis of Upper GI dx pre-exposure
  - NSAID, corticosteroid, H2blocker, PPI utilization pre-exposure

Do these choices matter?

Range of estimates across high-dimensional propensity score inception cohort (HDPS) parameter settings



Parameter settings explored in OMOP:

Washout period (1): 180d

Surveillance window (3): 30 days from exposure start; exposure + 30d ; all time from exposure start

Covariate eligibility window (3): 30 days prior to exposure, 180, all-time pre-exposure

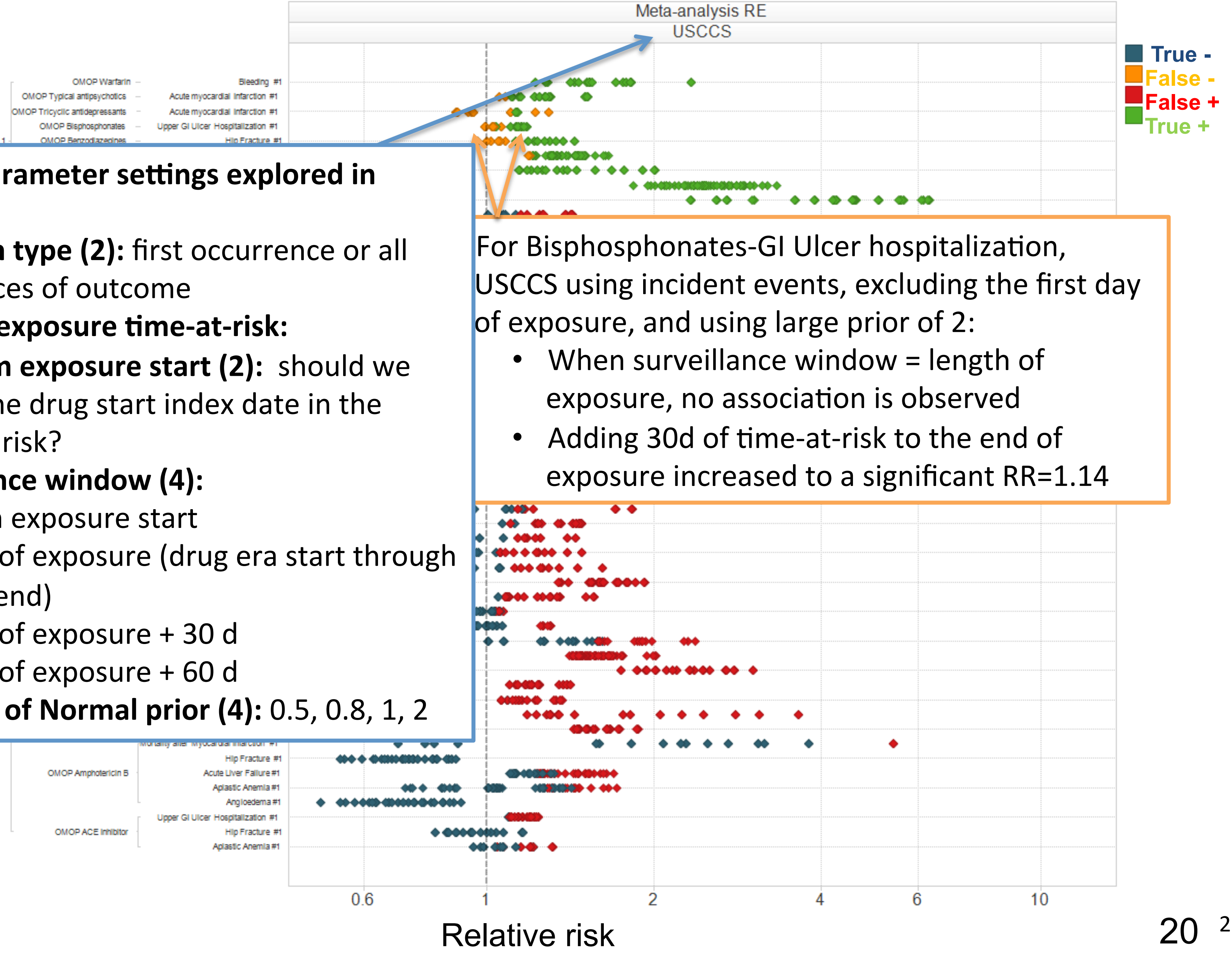
# of confounders (2): 100, 500 covariates used to estimate propensity score

Propensity strata (2): 5, 20 strata

Analysis strategy (3): Mantel-Haenszel stratification (MH), propensity score adjusted (PS), propensity strata adjusted (PS2)

Comparator cohort (2): drugs with same indication, not in same class; most prevalent drug with same indication, not in same class

Range of estimates across univariate self-controlled case series (USCCS) parameter settings



USCCS Parameter settings explored in OMOP:

Condition type (2): first occurrence or all occurrences of outcome

Defining exposure time-at-risk:

Days from exposure start (2): should we include the drug start index date in the period at risk?

Surveillance window (4):

30 d from exposure start

Duration of exposure (drug era start through drug era end)

Duration of exposure + 30 d

Duration of exposure + 60 d

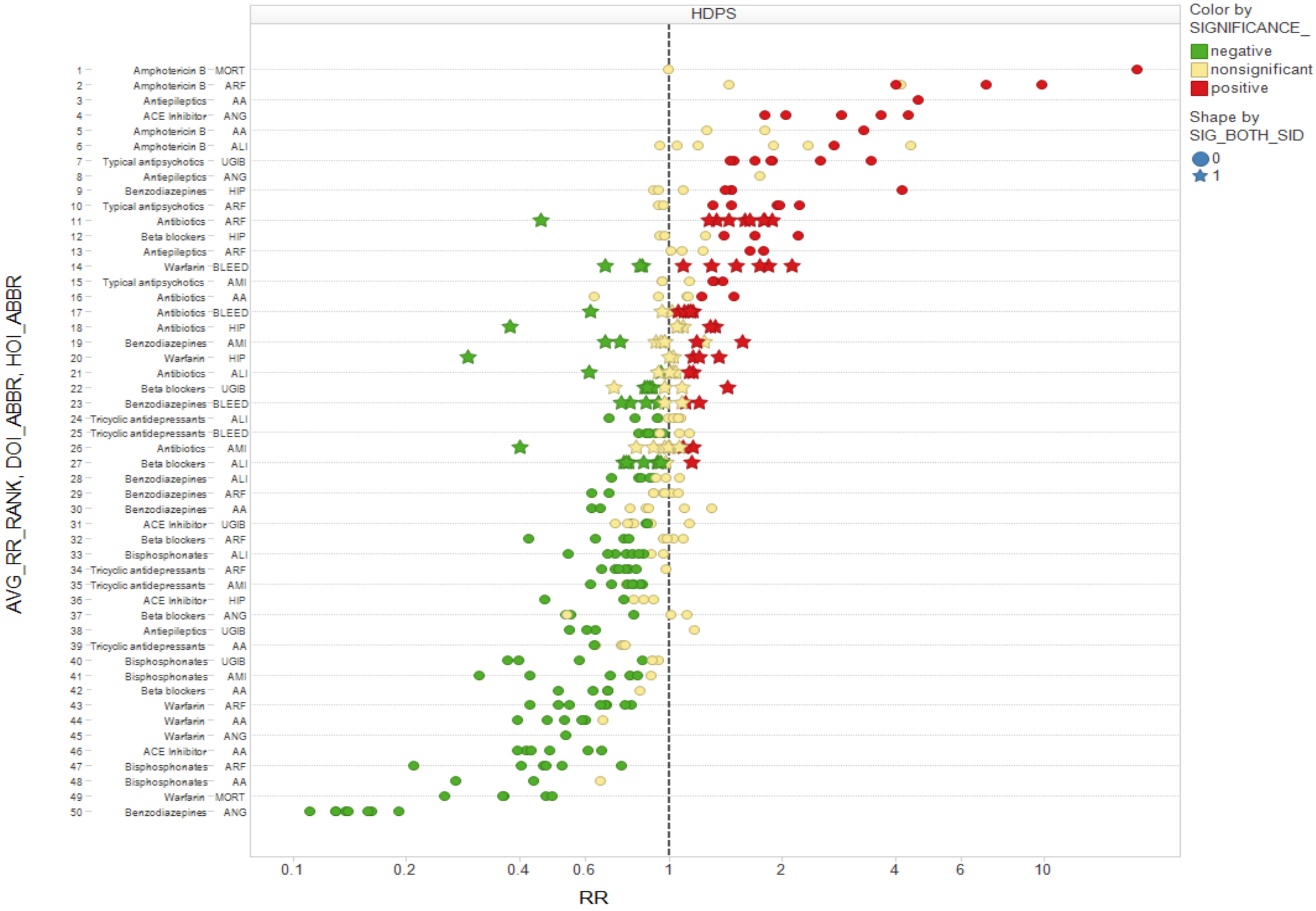
Precision of Normal prior (4): 0.5, 0.8, 1, 2

For Bisphosphonates-GI Ulcer hospitalization, USCCS using incident events, excluding the first day of exposure, and using large prior of 2:

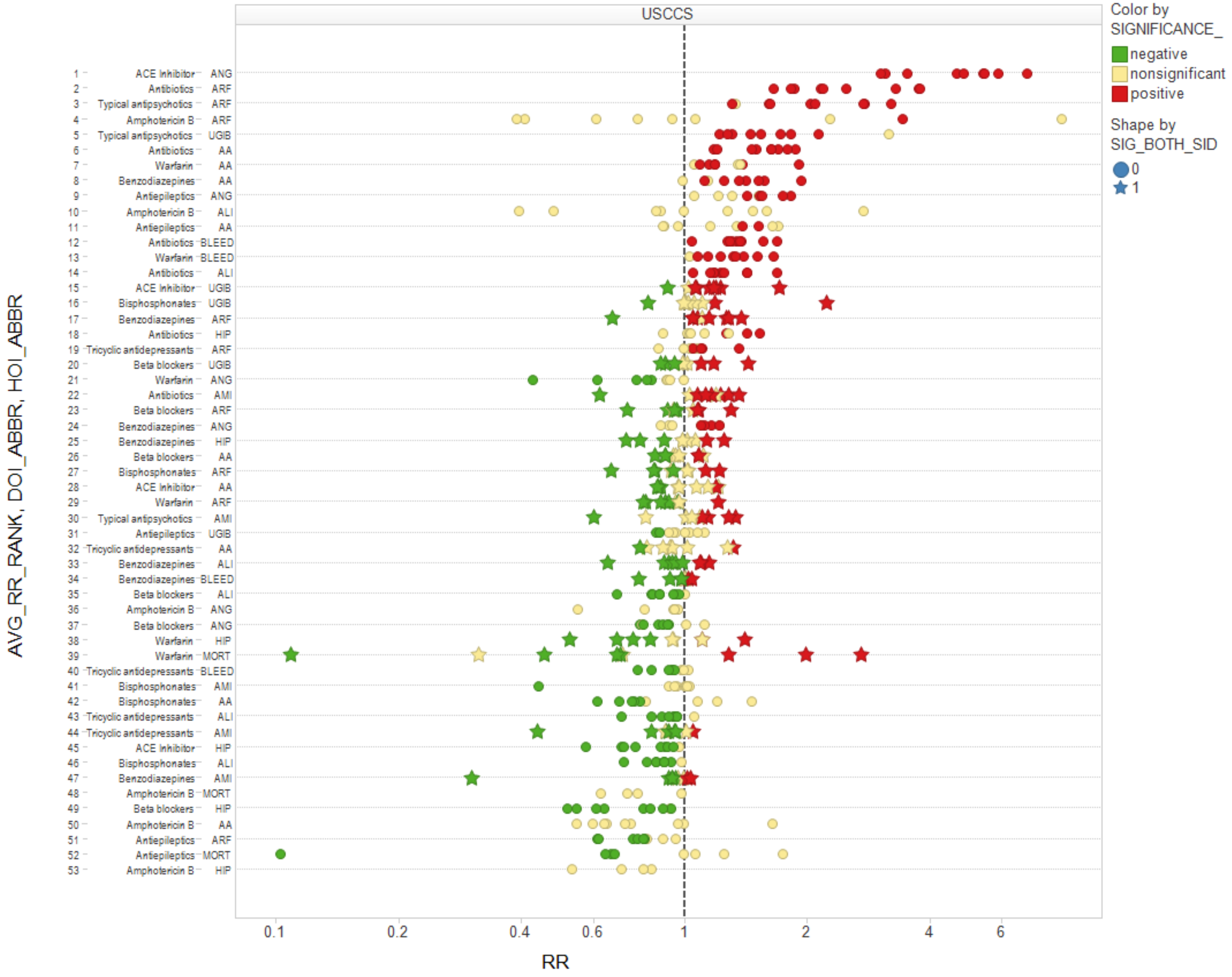
- When surveillance window = length of exposure, no association is observed
- Adding 30d of time-at-risk to the end of exposure increased to a significant RR=1.14

Fix everything *except* the database…

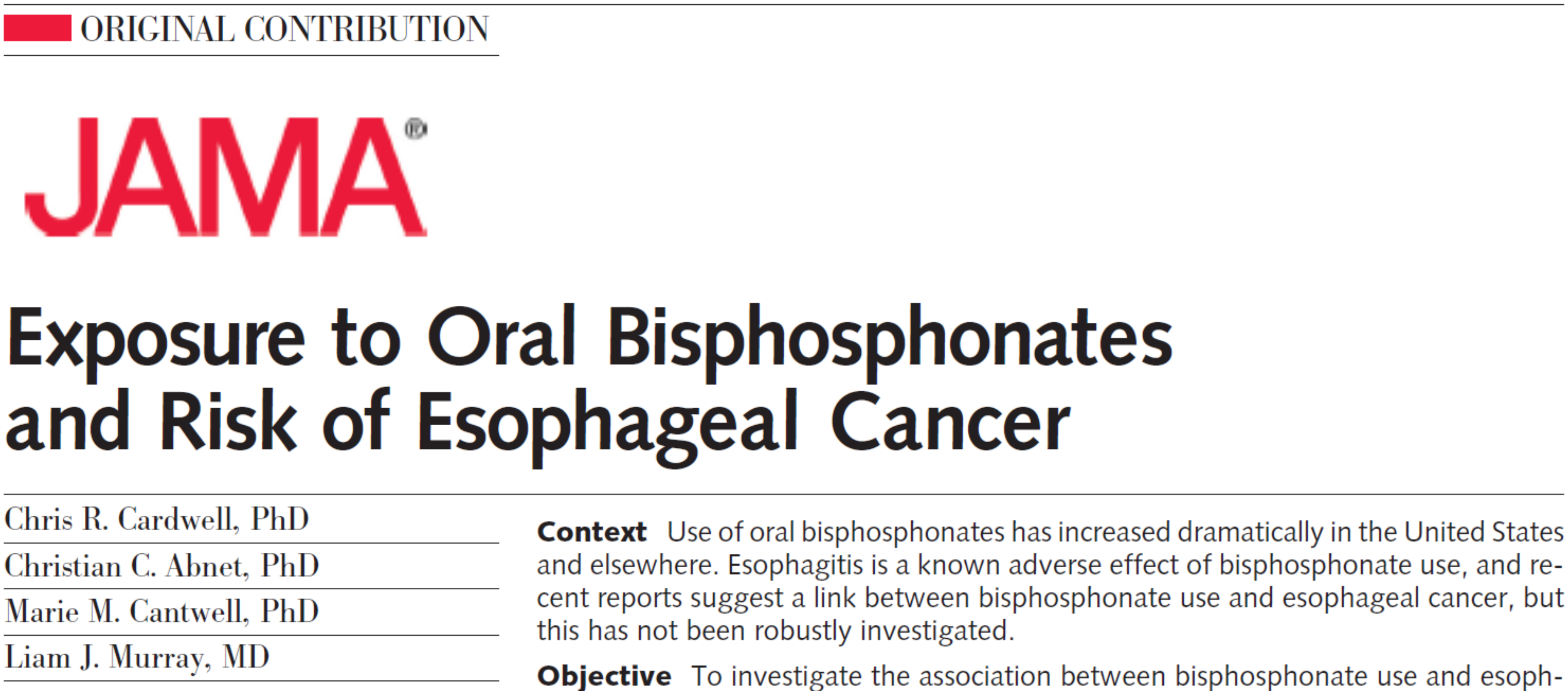
Cohort


HDPS
AVG_RR_RANK, DOI_ABBR, HOI_ABBR
1 Amphotericin B MORT
2 Amphotericin B ARF
3 Antiepileptics AA
4 ACE Inhibitor ANG
5 Amphotericin B AA
6 Amphotericin B ALI
7 Typical antipsychotics UGIB
8 Antiepileptics ANG
9 Benzodiazepines HIP
10 Typical antipsychotics ARF
11 Antibiotics ARF
12 Beta blockers HIP
13 Antiepileptics ARF
14 Warfarin BLEED
15 Typical antipsychotics AMI
16 Antibiotics AA
17 Antibiotics BLEED
18 Antibiotics HIP
19 Benzodiazepines AMI
20 Warfarin HIP
21 Antibiotics ALI
22 Beta blockers UGIB
23 Benzodiazepines BLEED
24 Tricyclic antidepressants ALI
25 Tricyclic antidepressants BLEED
26 Antibiotics AMI
27 Beta blockers ALI
28 Benzodiazepines ALI
29 Benzodiazepines ARF
30 Benzodiazepines AA
31 ACE Inhibitor UGIB
32 Beta blockers ARF
33 Bisphosphonates ALI
34 Tricyclic antidepressants ARF
35 Tricyclic antidepressants AMI
36 ACE Inhibitor HIP
37 Beta blockers ANG
38 Antiepileptics UGIB
39 Tricyclic antidepressants AA
40 Bisphosphonates UGIB
41 Bisphosphonates AMI
42 Beta blockers AA
43 Warfarin ARF
44 Warfarin AA
45 Warfarin ANG
46 ACE Inhibitor AA
47 Bisphosphonates ARF
48 Bisphosphonates AA
49 Warfarin MORT
50 Benzodiazepines ANG
0.1
0.2
0.4
0.6
1
2
4
6
10
RR
Color by
SIGNIFICANCE_
🟩 negative
🟨 nonsignificant
🟥 positive
Shape by
SIG_BOTH_SID
● 0
★ 1

SCCS



USCCS
AVG_RR_RANK, DOI_ABBR, HOI_ABBR
1 ACE Inhibitor ANG
2 Antibiotics ARF
3 Typical antipsychotics ARF
4 Amphotericin B ARF
5 Typical antipsychotics UGIB
6 Antibiotics AA
7 Warfarin AA
8 Benzodiazepines AA
9 Antiepileptics ANG
10 Amphotericin B ALI
11 Antiepileptics AA
12 Antibiotics BLEED
13 Warfarin BLEED
14 Antibiotics ALI
15 ACE Inhibitor UGIB
16 Bisphosphonates UGIB
17 Benzodiazepines ARF
18 Antibiotics HIP
19 Tricyclic antidepressants ARF
20 Beta blockers UGIB
21 Warfarin ANG
22 Antibiotics AMI
23 Beta blockers ARF
24 Benzodiazepines ANG
25 Benzodiazepines HIP
26 Beta blockers AA
27 Bisphosphonates ARF
28 ACE Inhibitor AA
29 Warfarin ARF
30 Typical antipsychotics AMI
31 Antiepileptics UGIB
32 Tricyclic antidepressants AA
33 Benzodiazepines ALI
34 Benzodiazepines BLEED
35 Beta blockers ALI
36 Amphotericin B ANG
37 Beta blockers ANG
38 Warfarin HIP
39 Warfarin MORT
40 Tricyclic antidepressants BLEED
41 Bisphosphonates AMI
42 Bisphosphonates AA
43 Tricyclic antidepressants ALI
44 Tricyclic antidepressants AMI
45 ACE Inhibitor HIP
46 Bisphosphonates ALI
47 Benzodiazepines AMI
48 Amphotericin B MORT
49 Beta blockers HIP
50 Amphotericin B AA
51 Antiepileptics ARF
52 Antiepileptics MORT
53 Amphotericin B HIP
0.1
0.2
0.4
0.6
1
2
4
6
RR
Color by SIGNIFICANCE_
🟩 negative
🟨 nonsignificant
🟥 positive
Shape by SIG_BOTH_SID
● 0
★ 1



ORIGINAL CONTRIBUTION

JAMA®

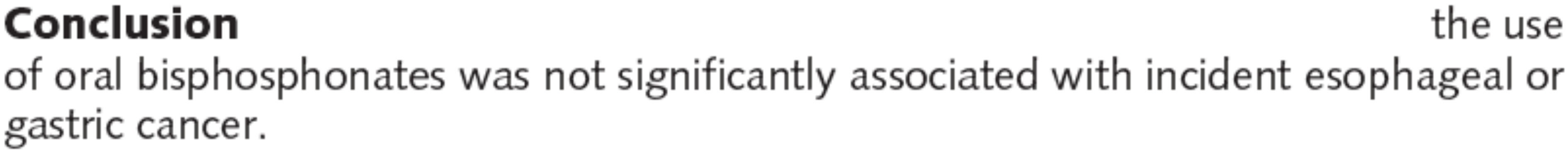
Exposure to Oral Bisphosphonates and Risk of Esophageal Cancer

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-

JAMA 2010; 304(6): 657-663



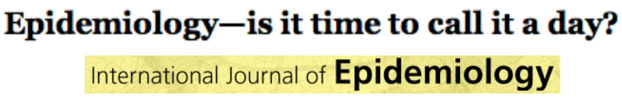
Conclusion the use of oral bisphosphonates was not significantly associated with incident esophageal or gastric cancer.

Does this stuff work at all?

Why Most Published Research Findings Are False

John P. A. Ioannidis

PLoS Medicine | www.plosmedicine.org

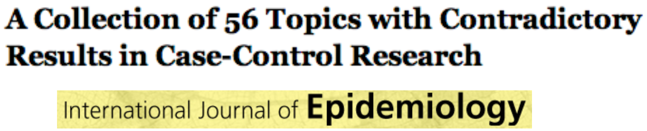


Epidemiology—is it time to call it a day?

International Journal of **Epidemiology**

Microarrays and molecular research: noise discovery?

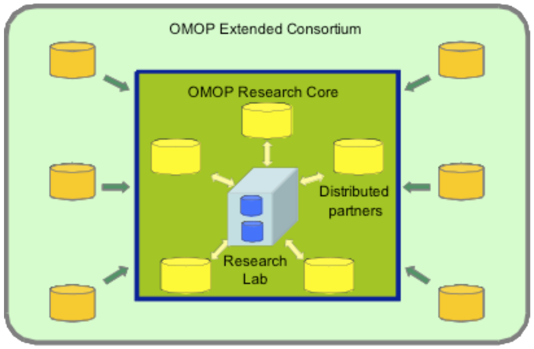
THE LANCET



A Collection of 56 Topics with Contradictory Results in Case-Control Research

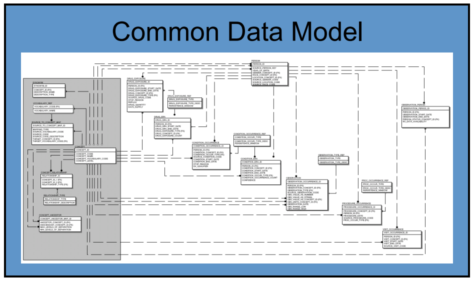
International Journal of **Epidemiology**

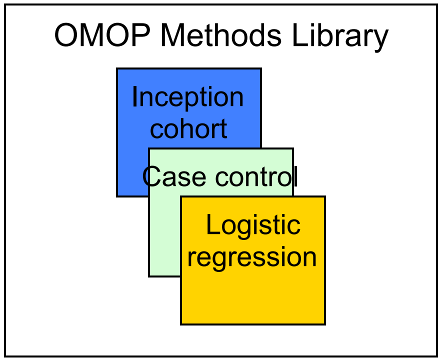
OMOP 2010/2011 Research Experiment



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

- 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

OMOP 2011/2012 Research

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

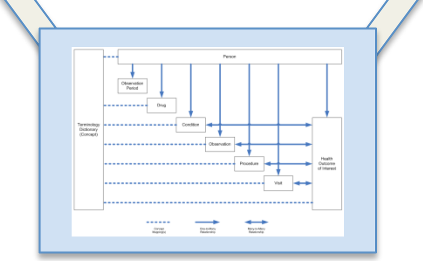
- *Multivariate self-controlled case series*

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:

Thomson MarketScan

GE

+ OMOP Distributed Partners
+ EU-ADR network

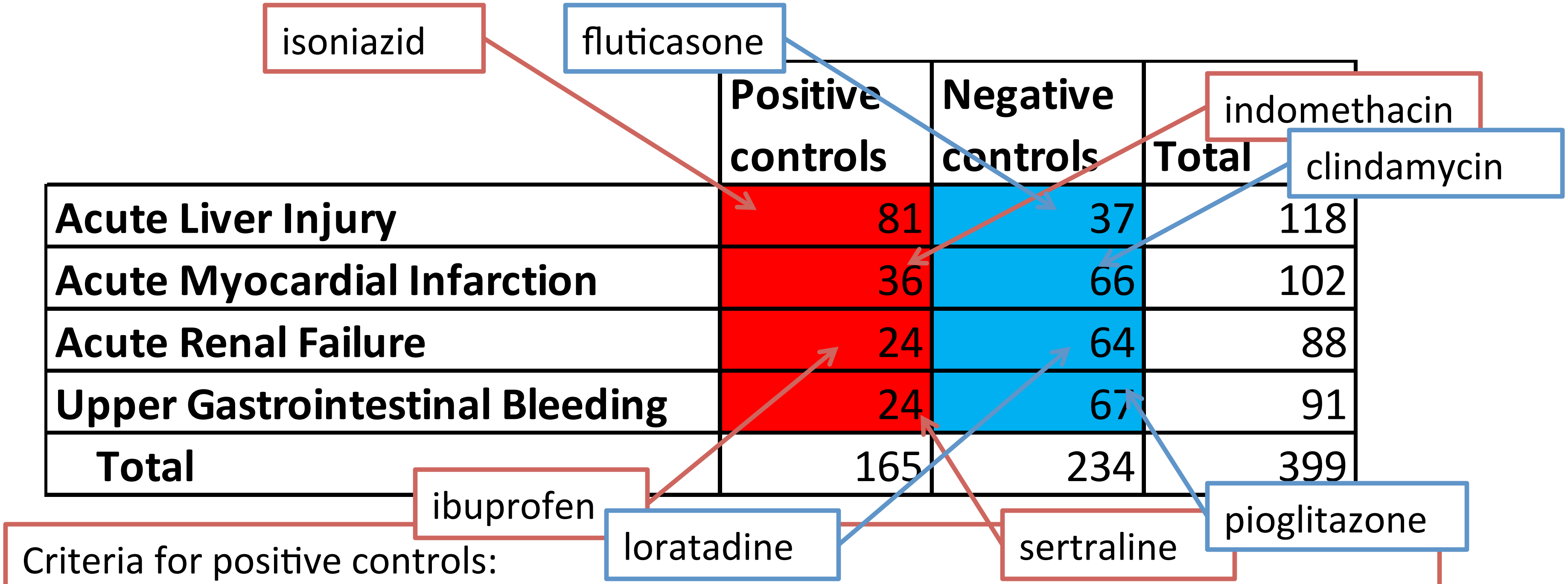
Simulated data:

signal

OSIM2

- Strength (RR)
- Type (timing)

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

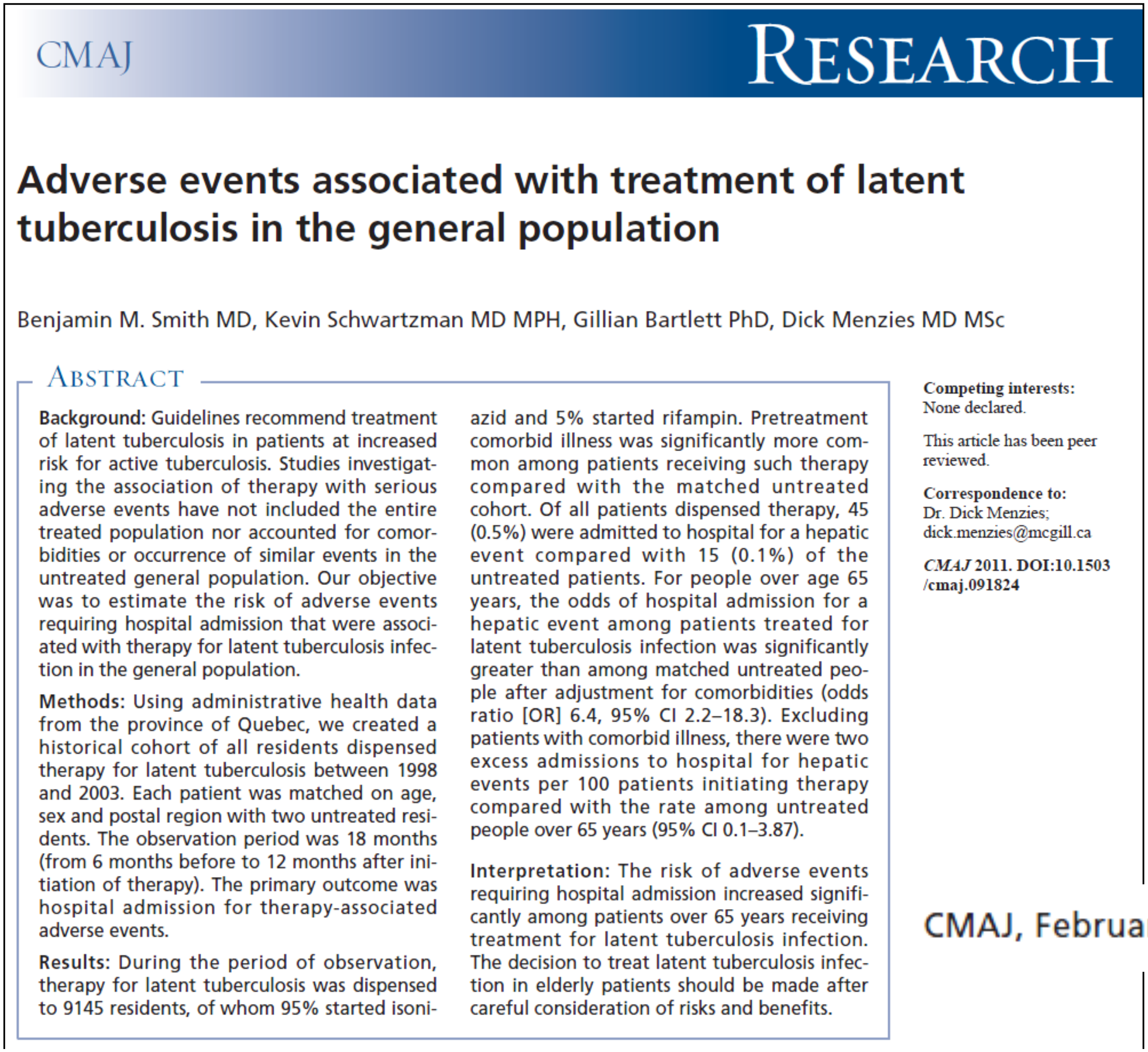
Criteria for positive controls:

- 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

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)

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

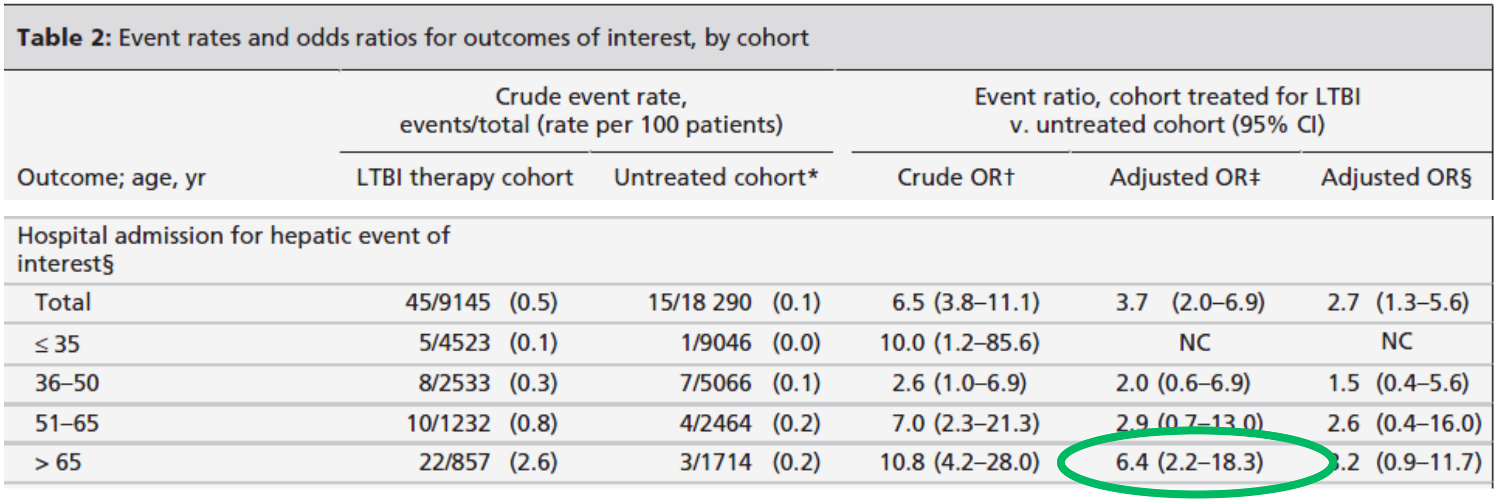
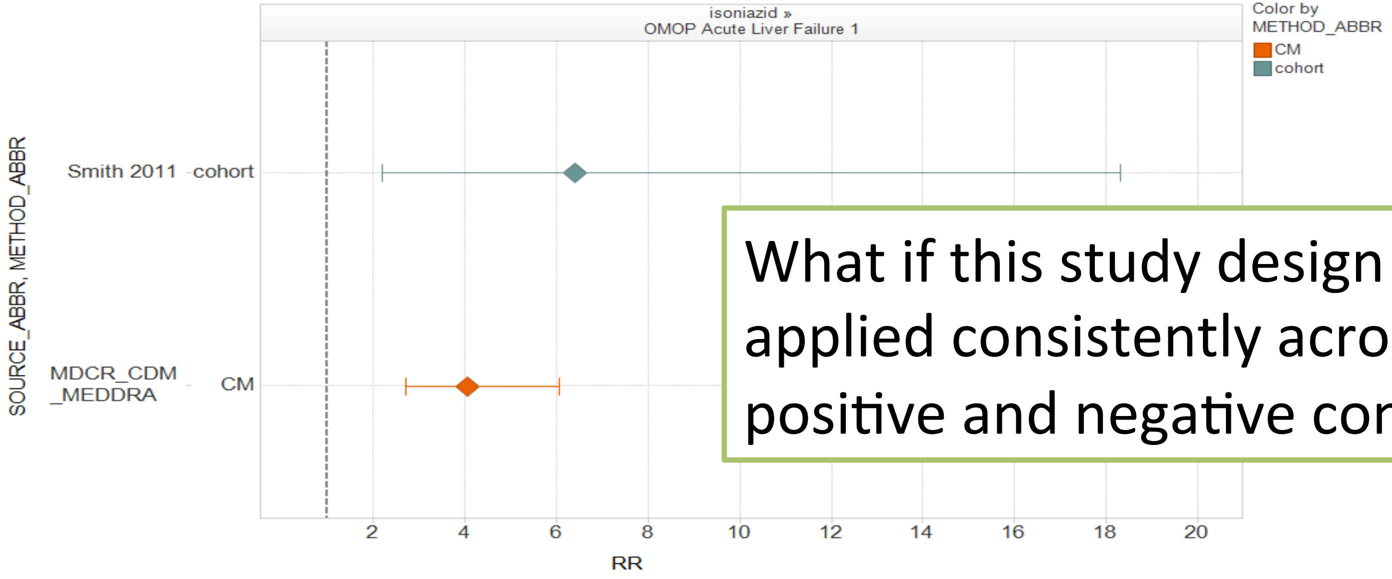


Table 2: Event rates and odds ratios for outcomes of interest, by cohort

	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)

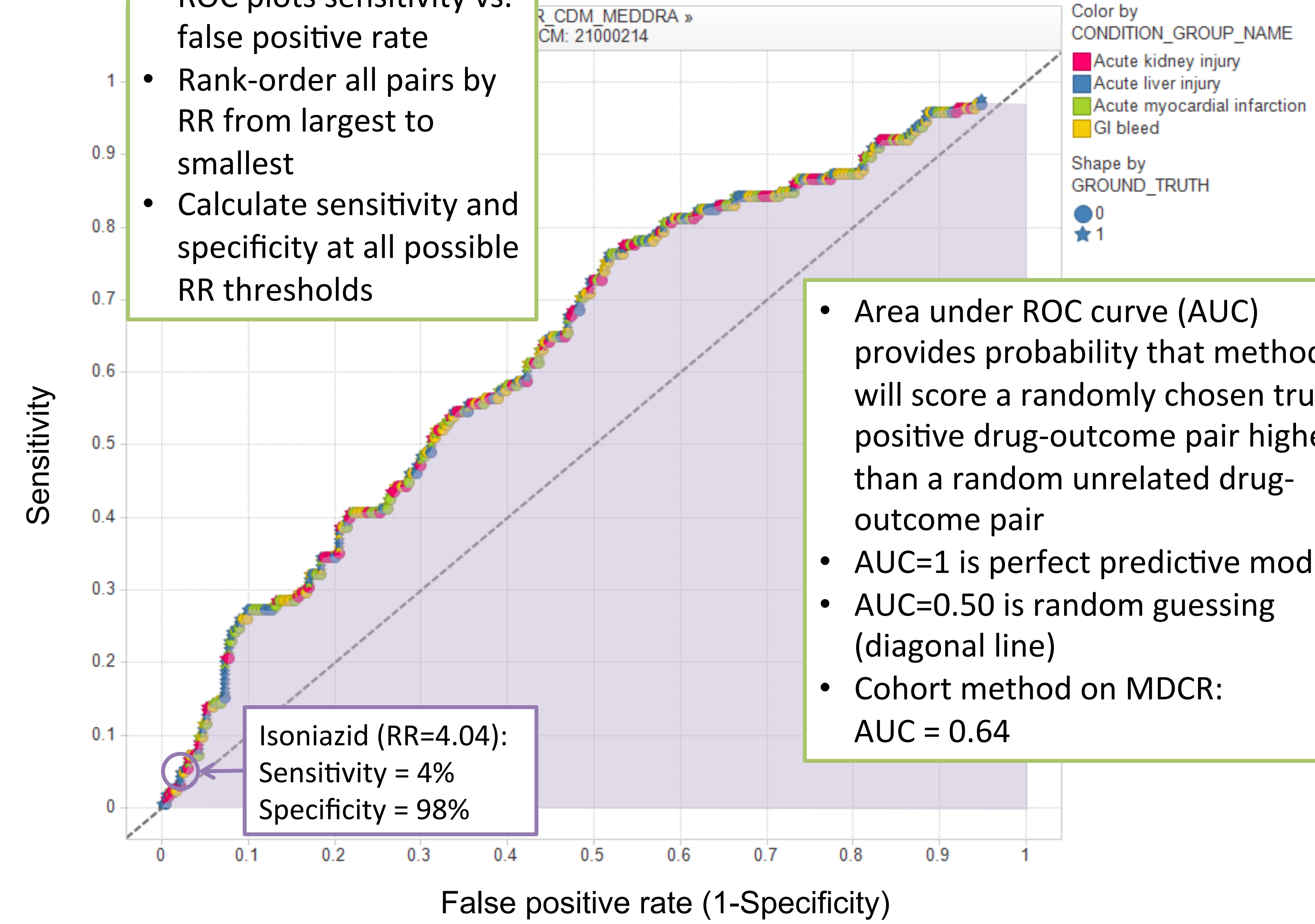
Revisiting the isoniazid – acute liver injury example

- 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?

Receiver Operating Characteristic (ROC) curve

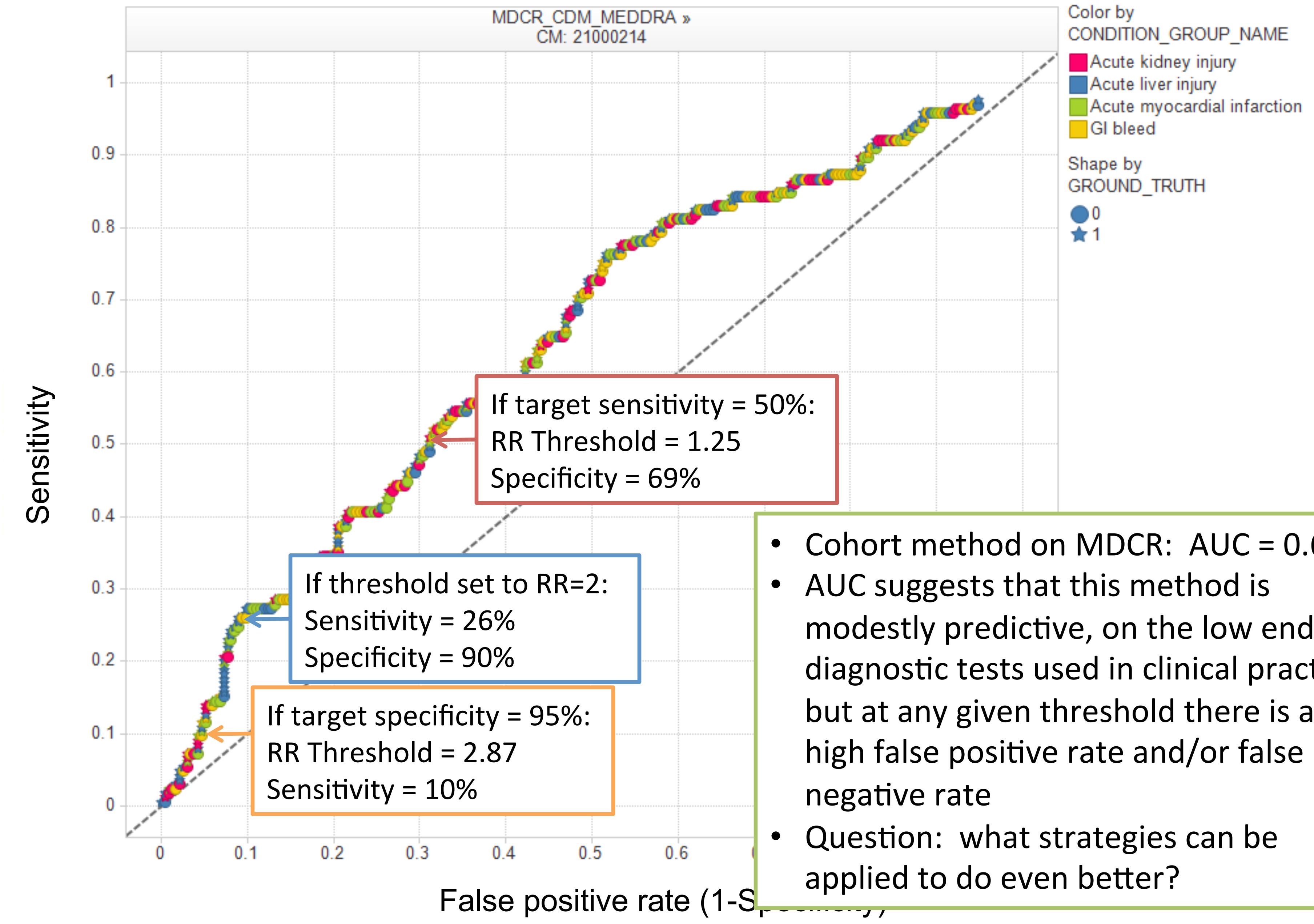


- ROC plots sensitivity vs. false positive rate
- Rank-order all pairs by RR from largest to smallest
- Calculate sensitivity and specificity at all possible RR thresholds

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

Setting thresholds from an ROC curve



If target sensitivity = 50%:
RR Threshold = 1.25
Specificity = 69%

If threshold set to RR=2:
Sensitivity = 26%
Specificity = 90%

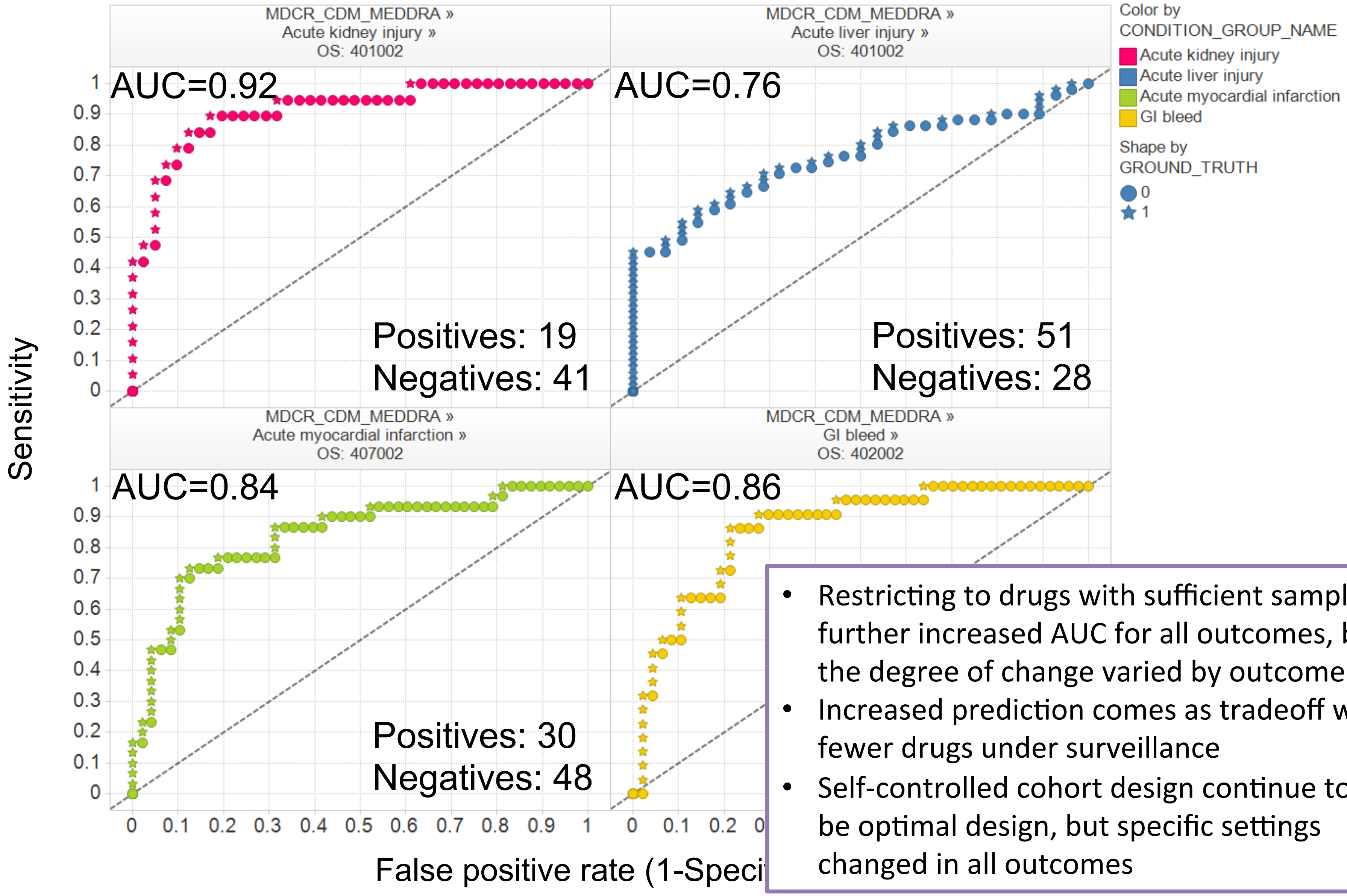
If target specificity = 95%:
RR Threshold = 2.87
Sensitivity = 10%

- 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

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

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

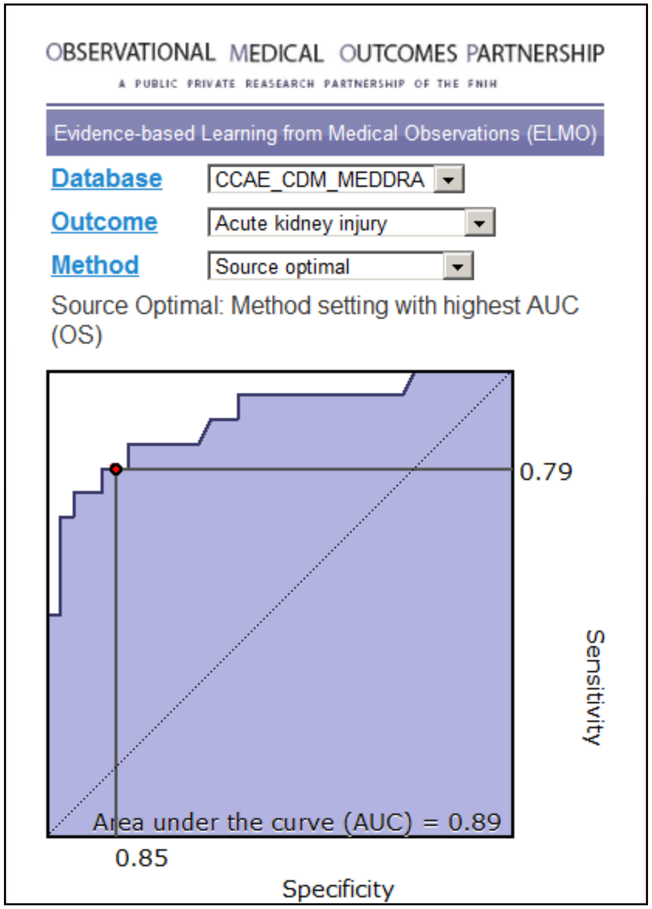
Optimal methods (AUC) by outcome and data source

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

Good performance?

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



OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

A PUBLIC PRIVATE REASEARCH PARTNERSHIP OF THE FNIH

Evidence-based Learning from Medical Observations (ELMO)

Database: CCAE_CDM_MEDDRA

Outcome: Acute kidney injury

Method: Source optimal

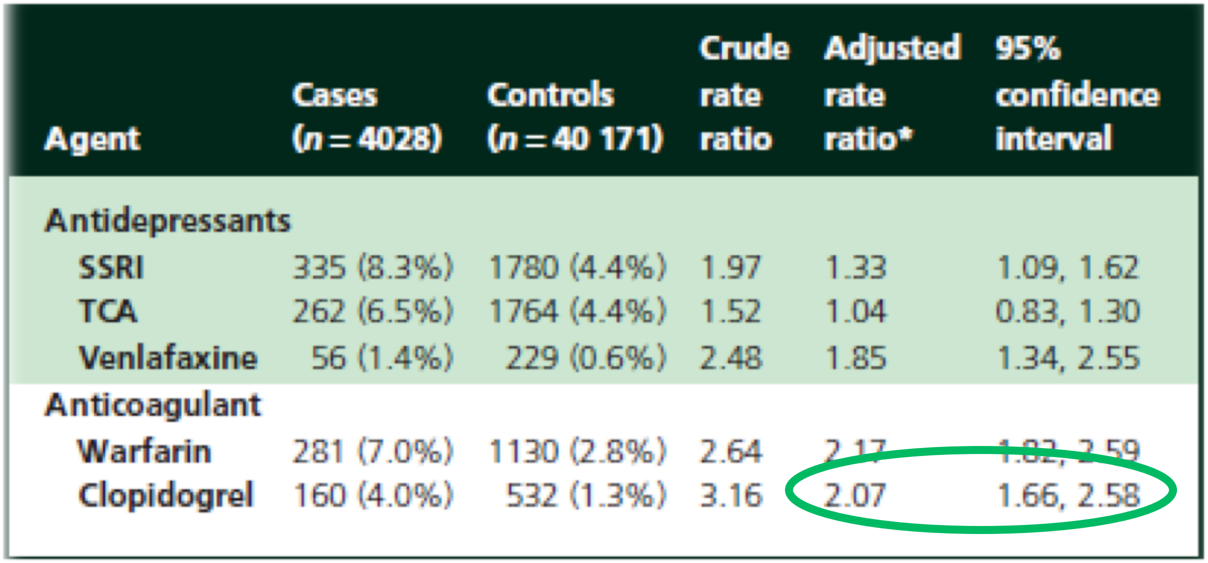
Source Optimal: Method setting with highest AUC (OS)

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

Revisiting clopidogrel & GI bleed (Opatrny, 2008)



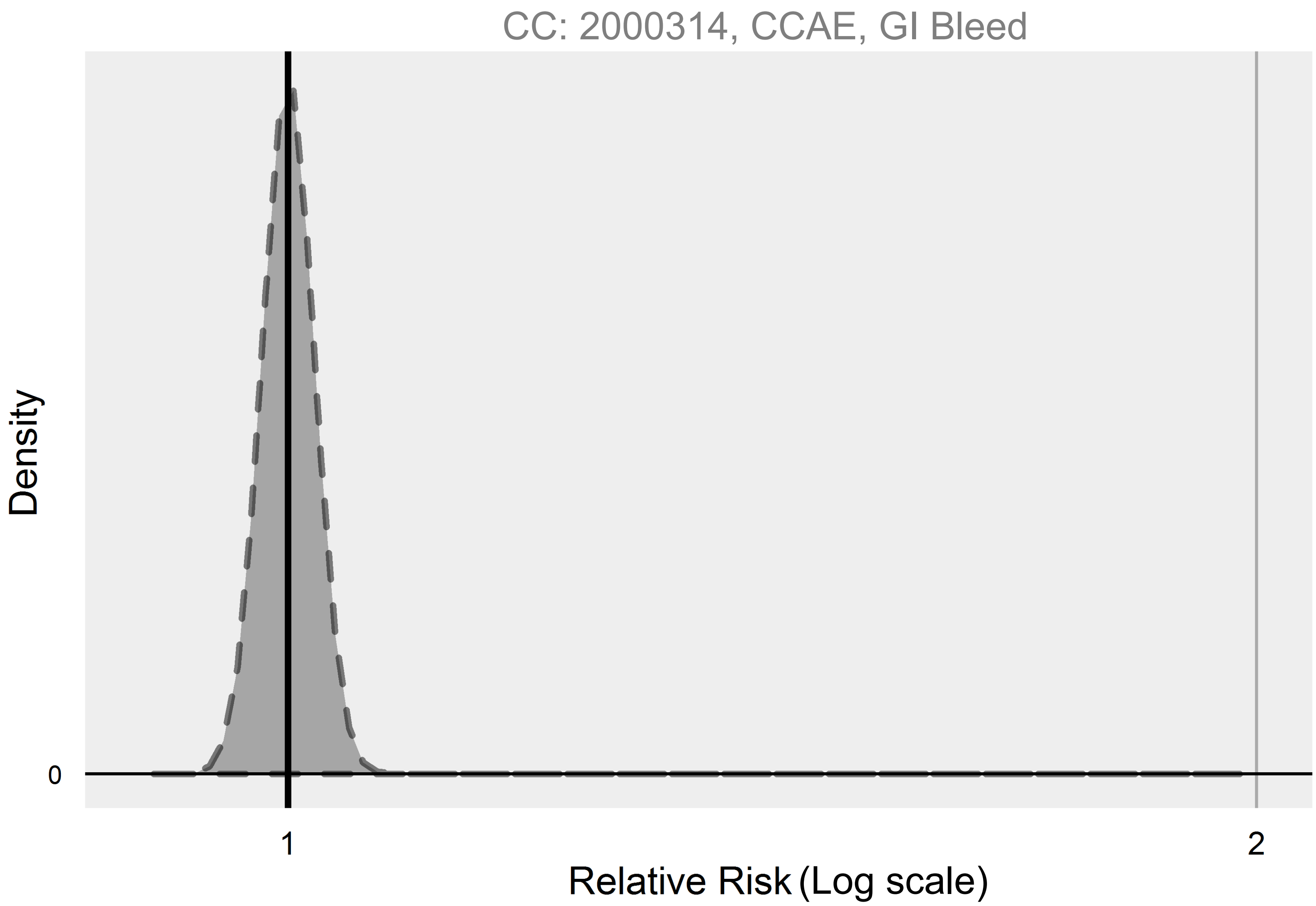
Agent	Cases (*n* = 4028)	Controls (*n* = 40 171)	Crude rate ratio	Adjusted rate ratio*	95% confidence interval
Antidepressants					
SSRI	335 (8.3%)	1780 (4.4%)	1.97	1.33	1.09, 1.62
TCA	262 (6.5%)	1764 (4.4%)	1.52	1.04	0.83, 1.30
Venlafaxine	56 (1.4%)	229 (0.6%)	2.48	1.85	1.34, 2.55
Anticoagulant					
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)

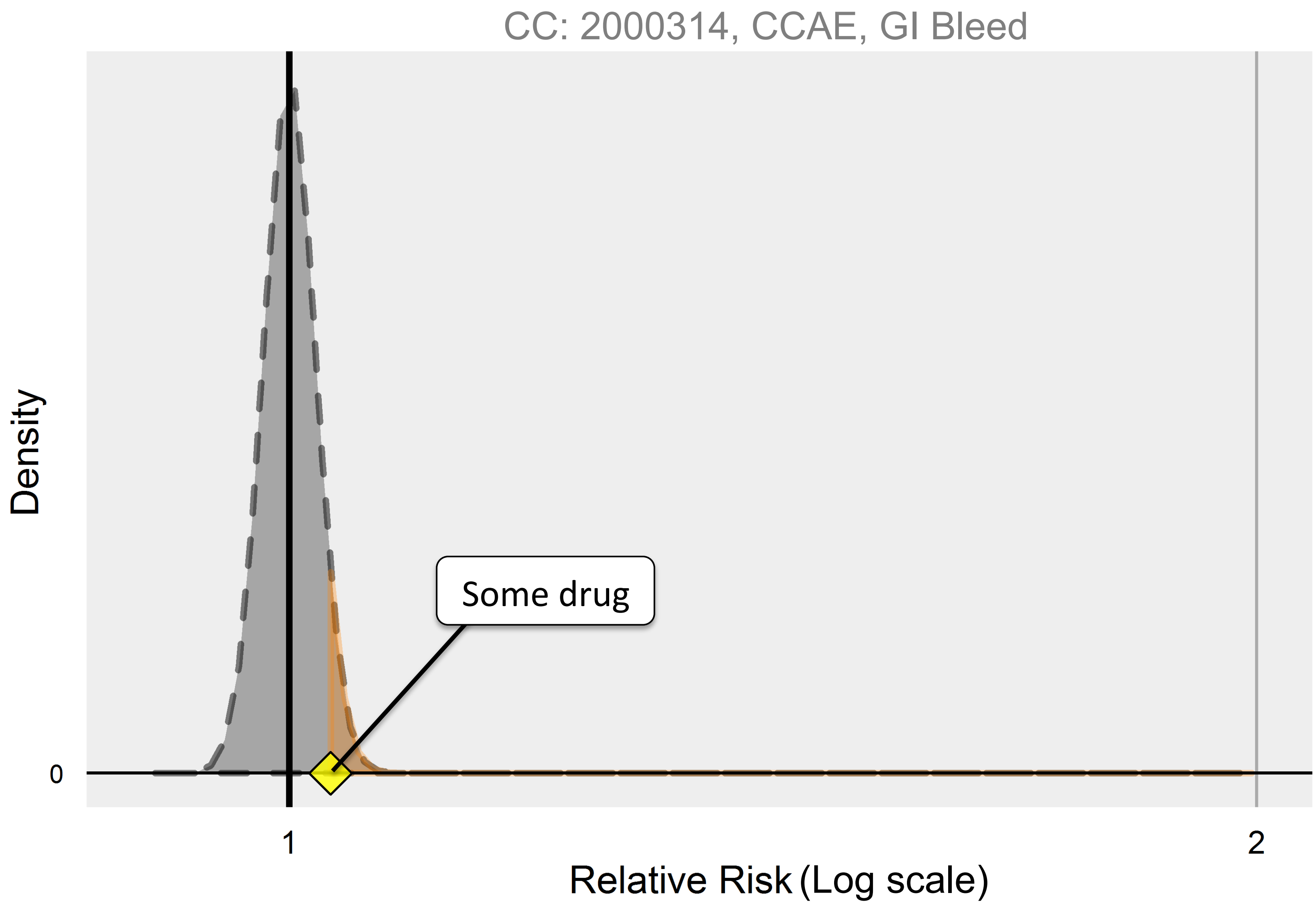
Relative risk: 1.86, 95% CI: 1.79 – 1.93

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

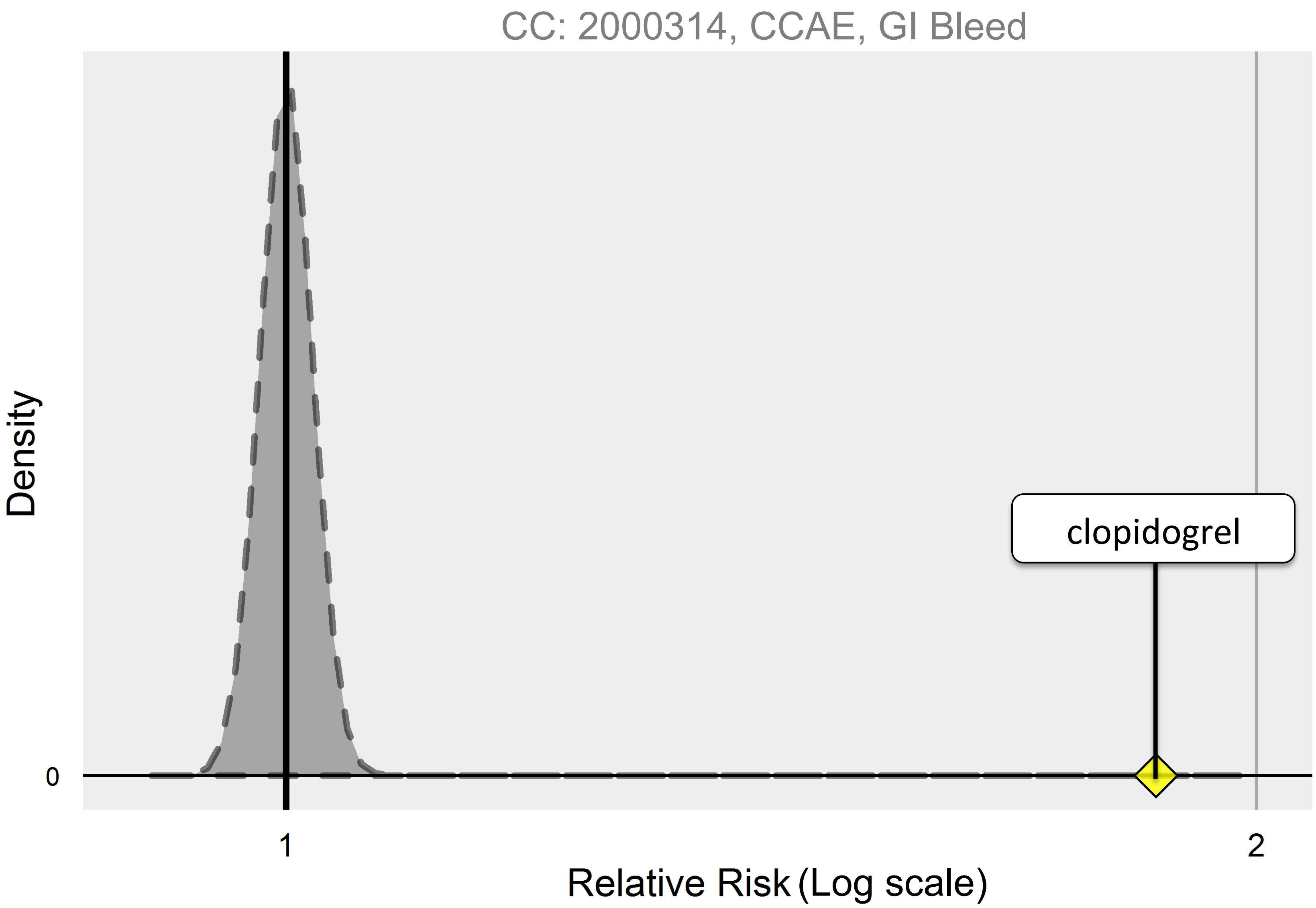
Null distribution



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

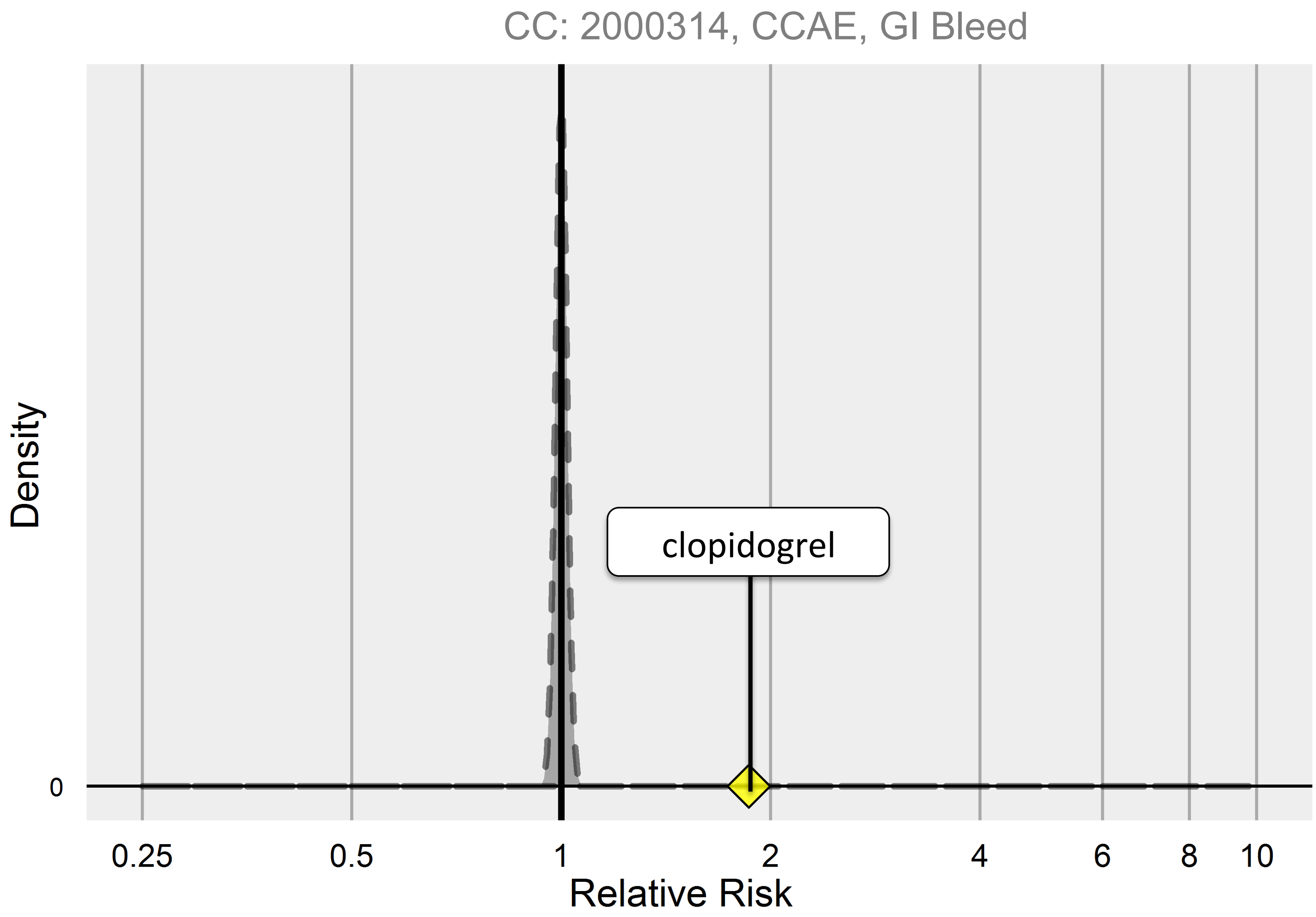
Ground truth for OMOP 2011/2012 experiments

	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

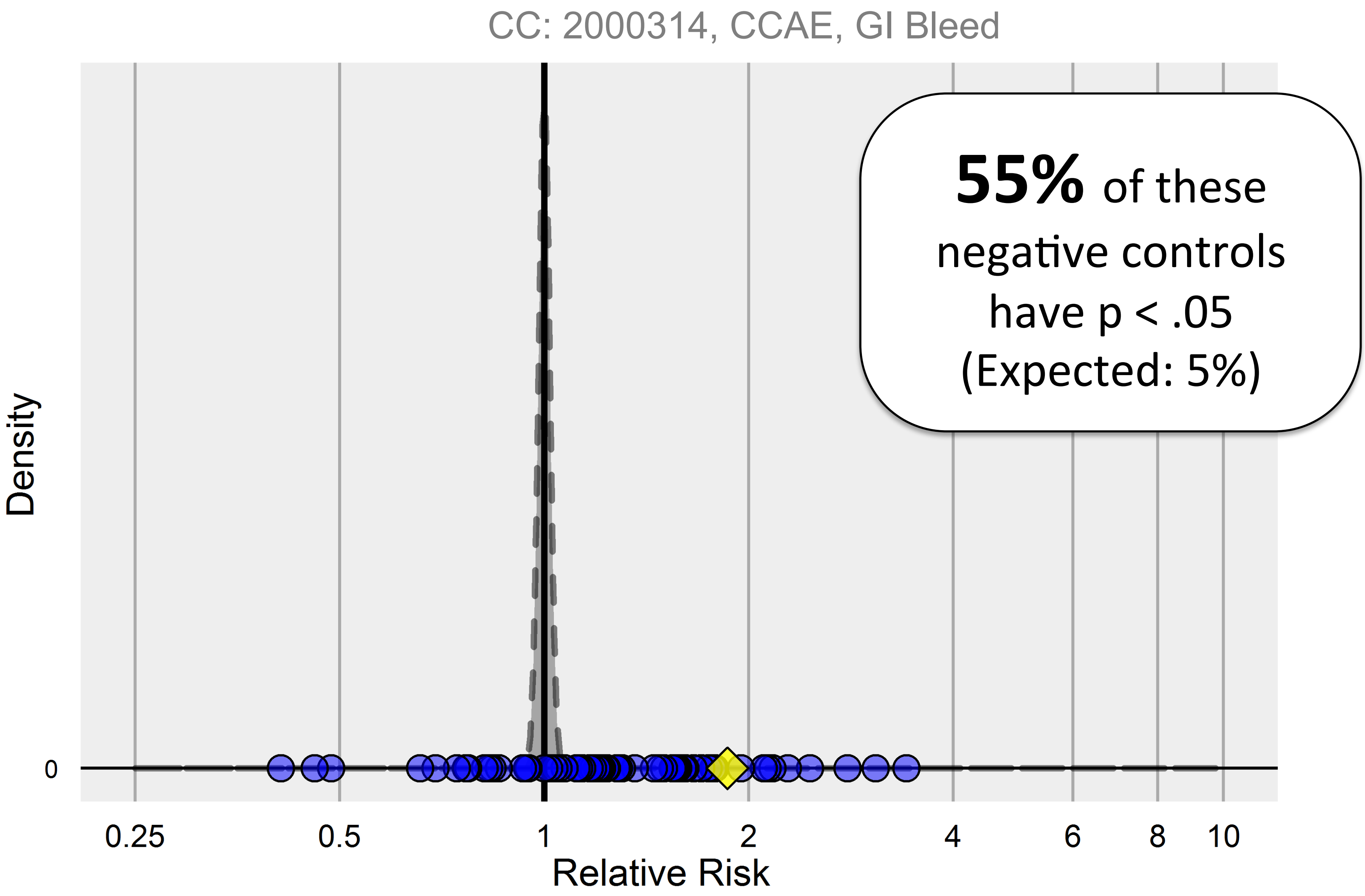
Criteria for negative controls:

- 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

Negative controls & the null distribution

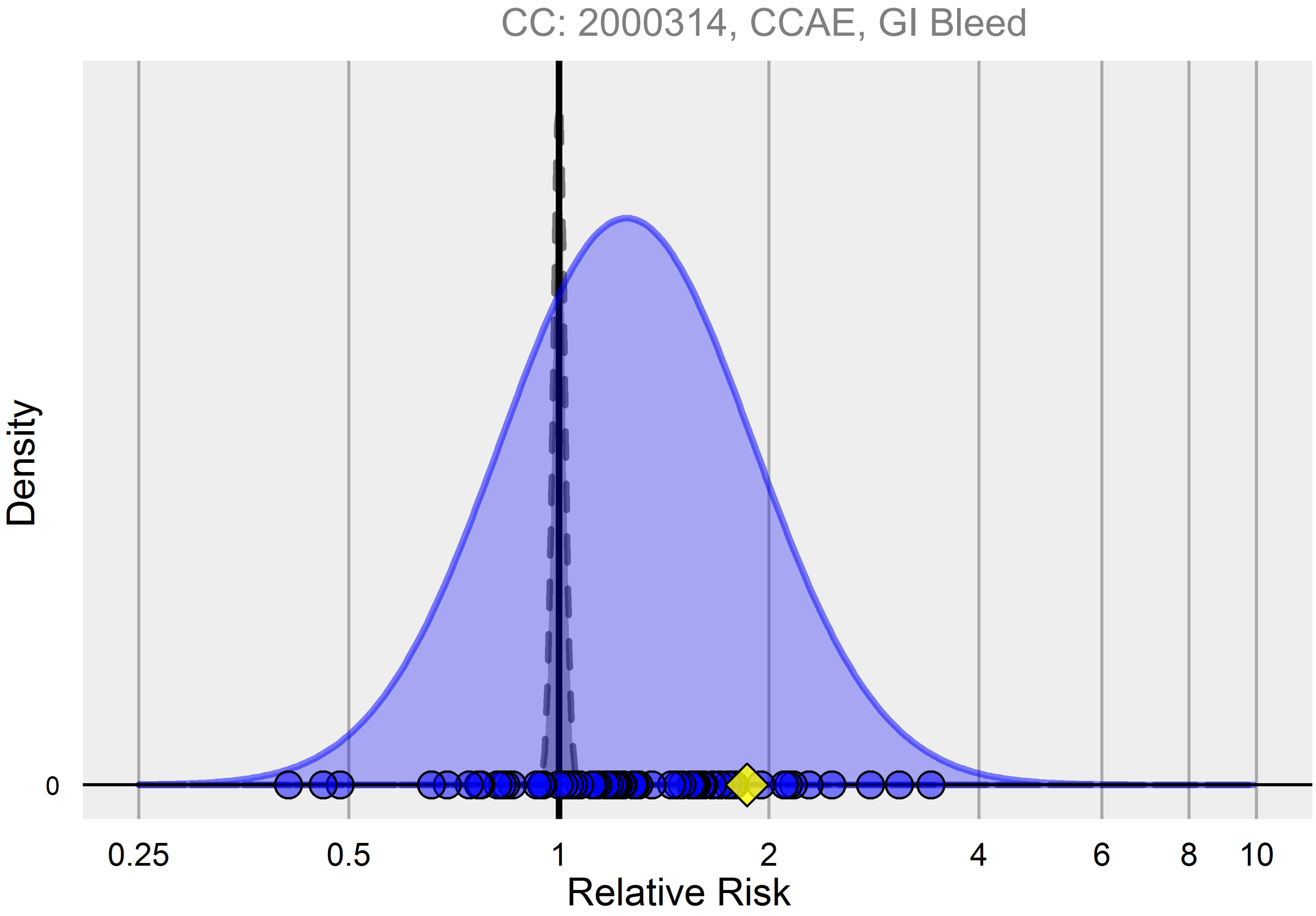


Negative controls & the null distribution

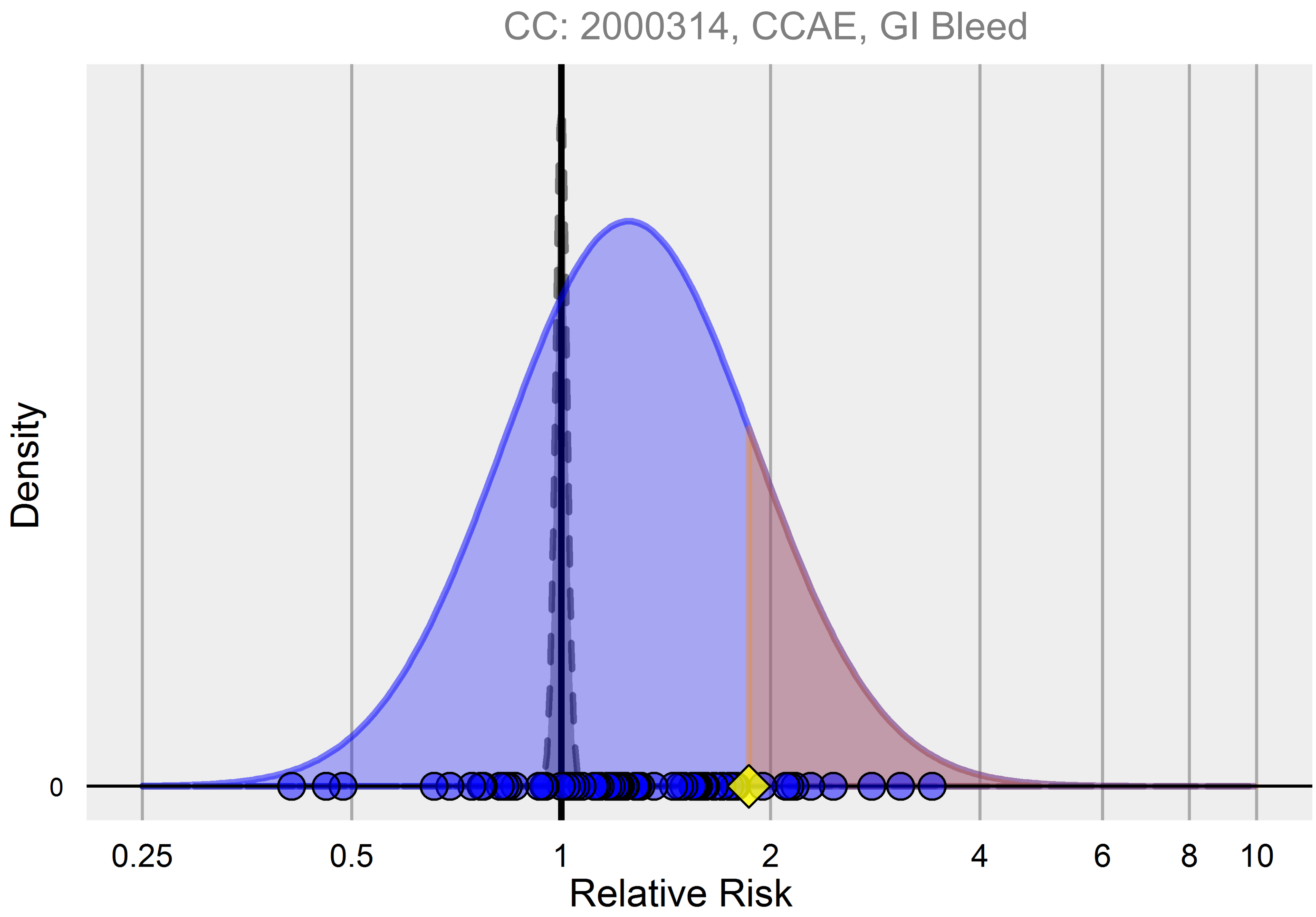


55% of these negative controls have p < .05 (Expected: 5%)

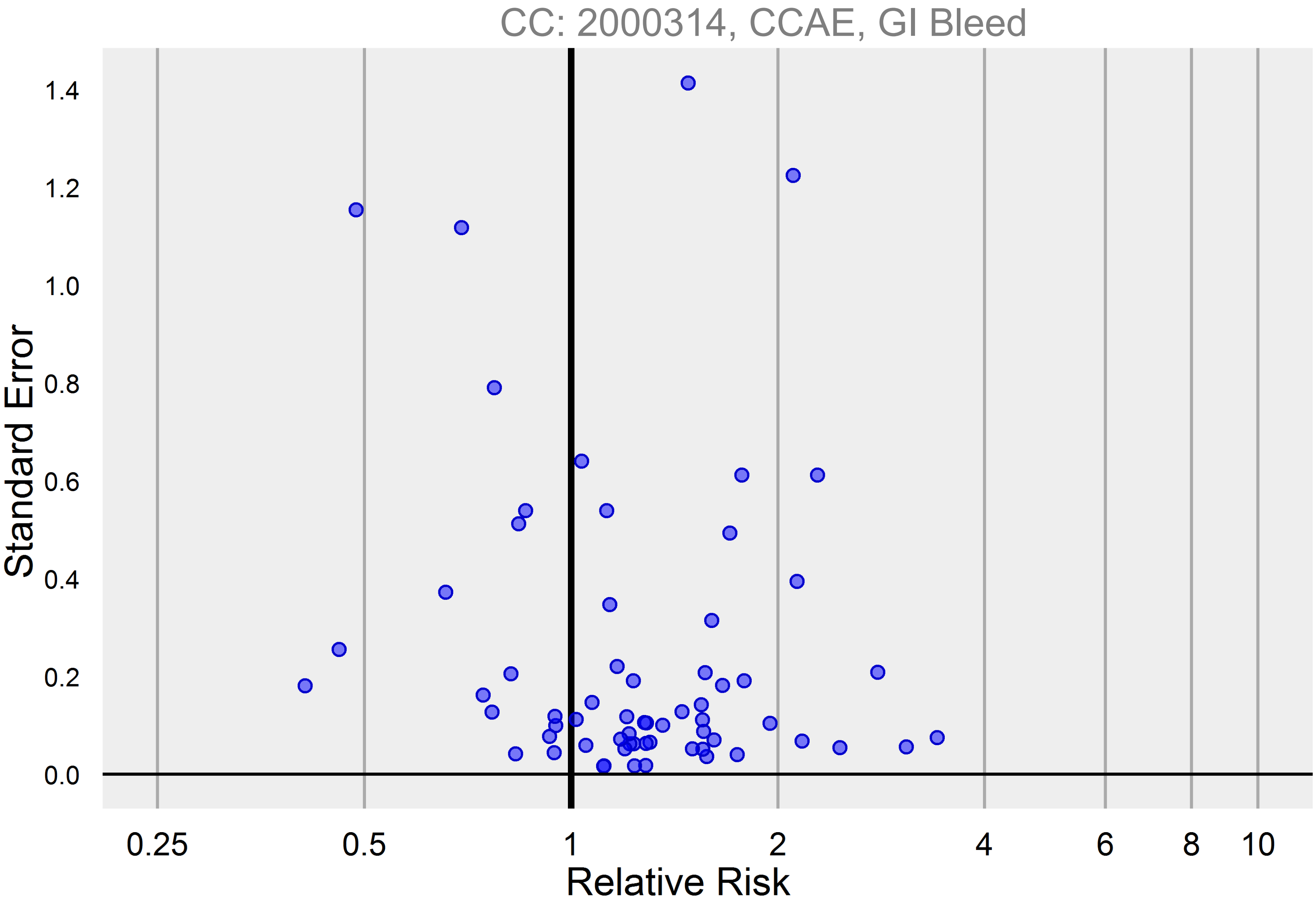
Negative controls & the null distribution



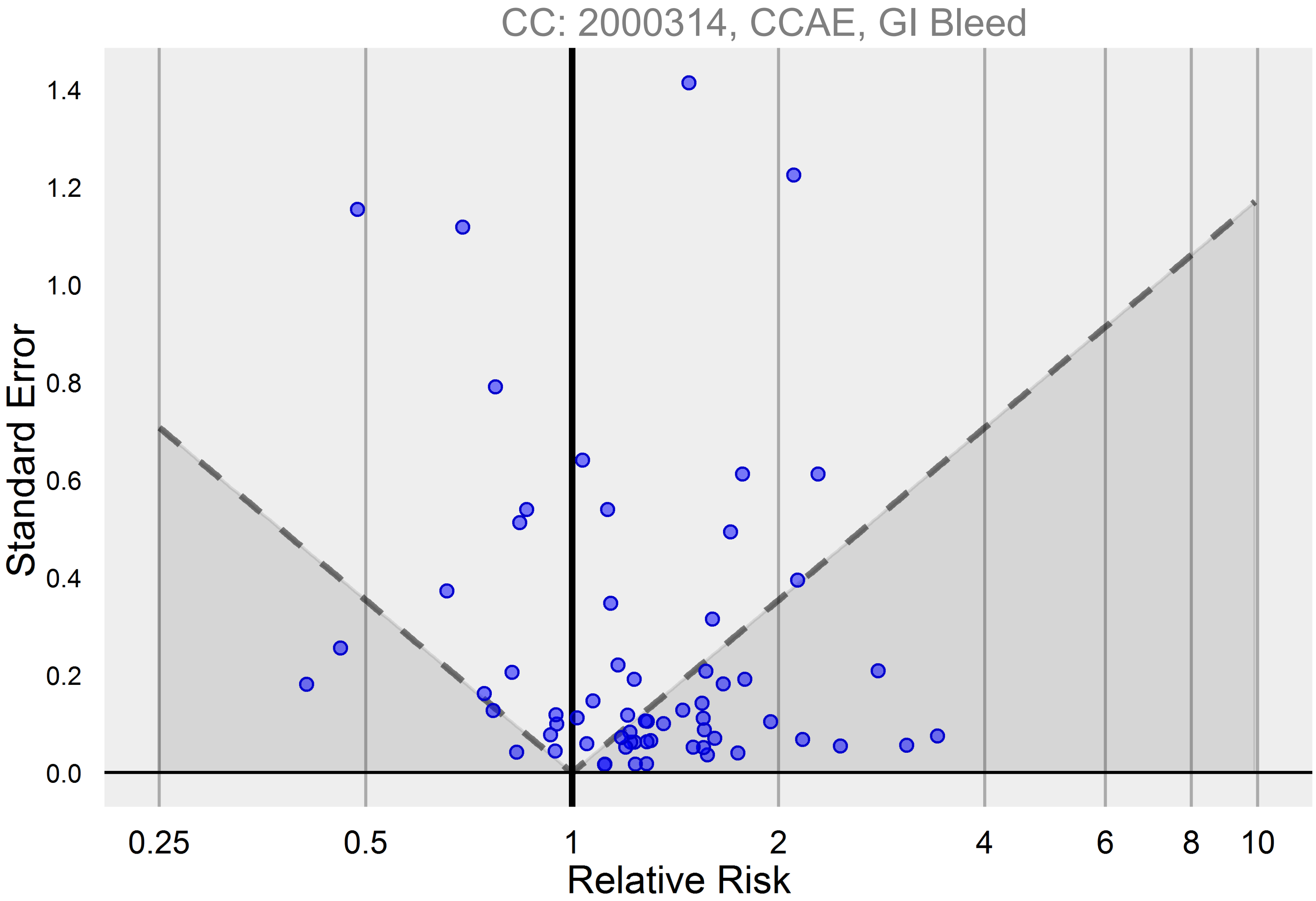
Negative controls & the null distribution



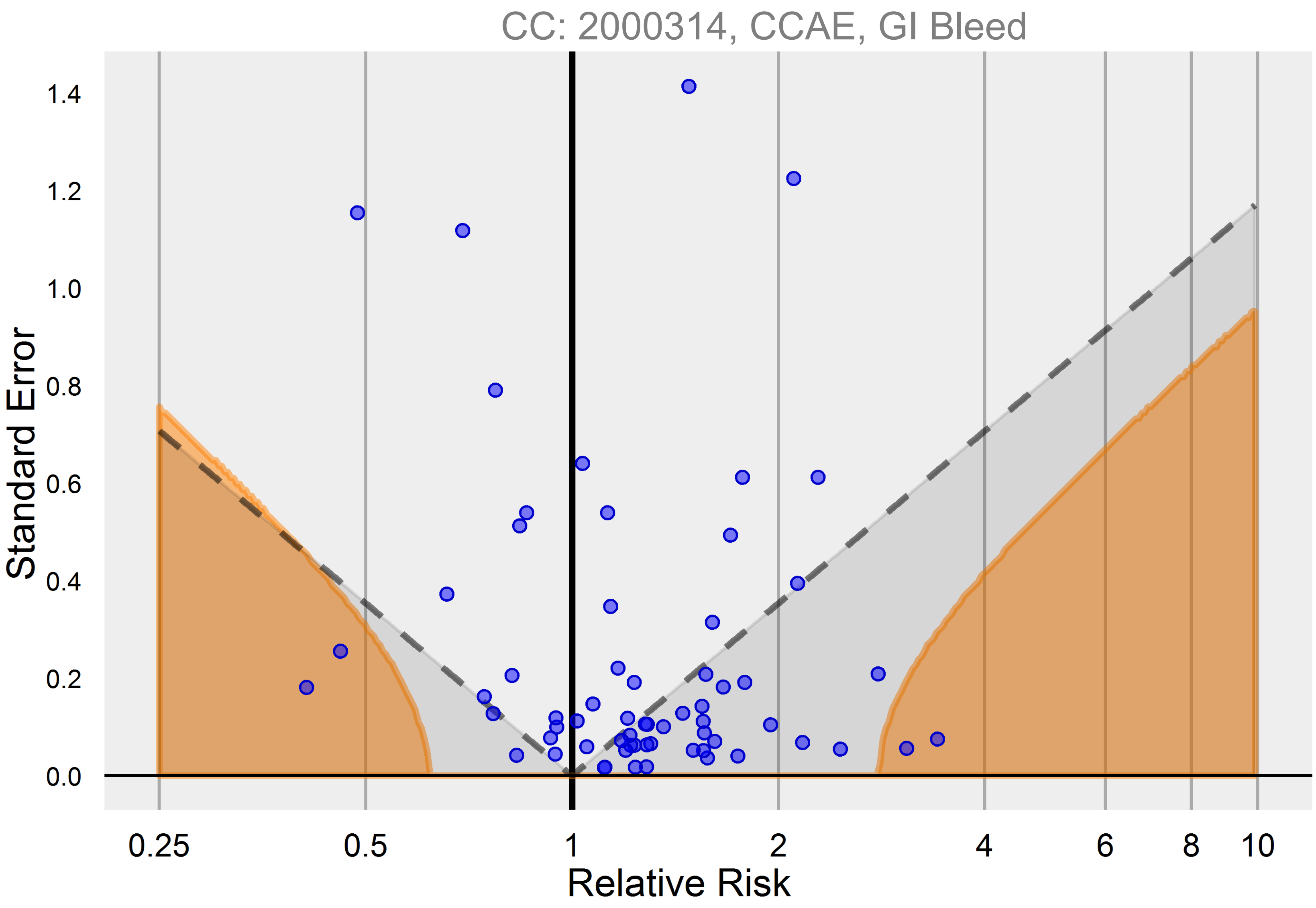
p-value calibration plot



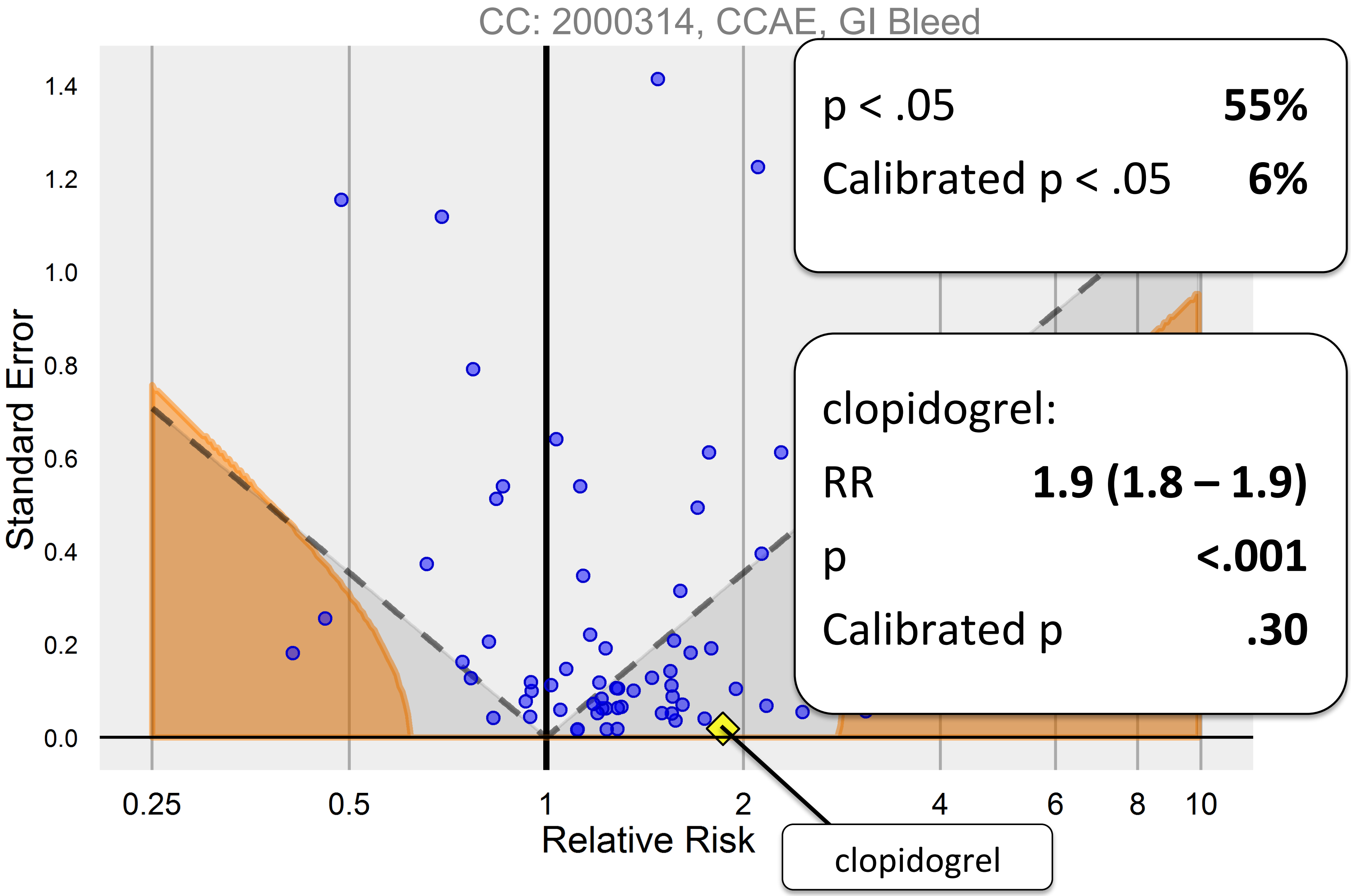
p-value calibration plot



p-value calibration plot



p-value calibration plot

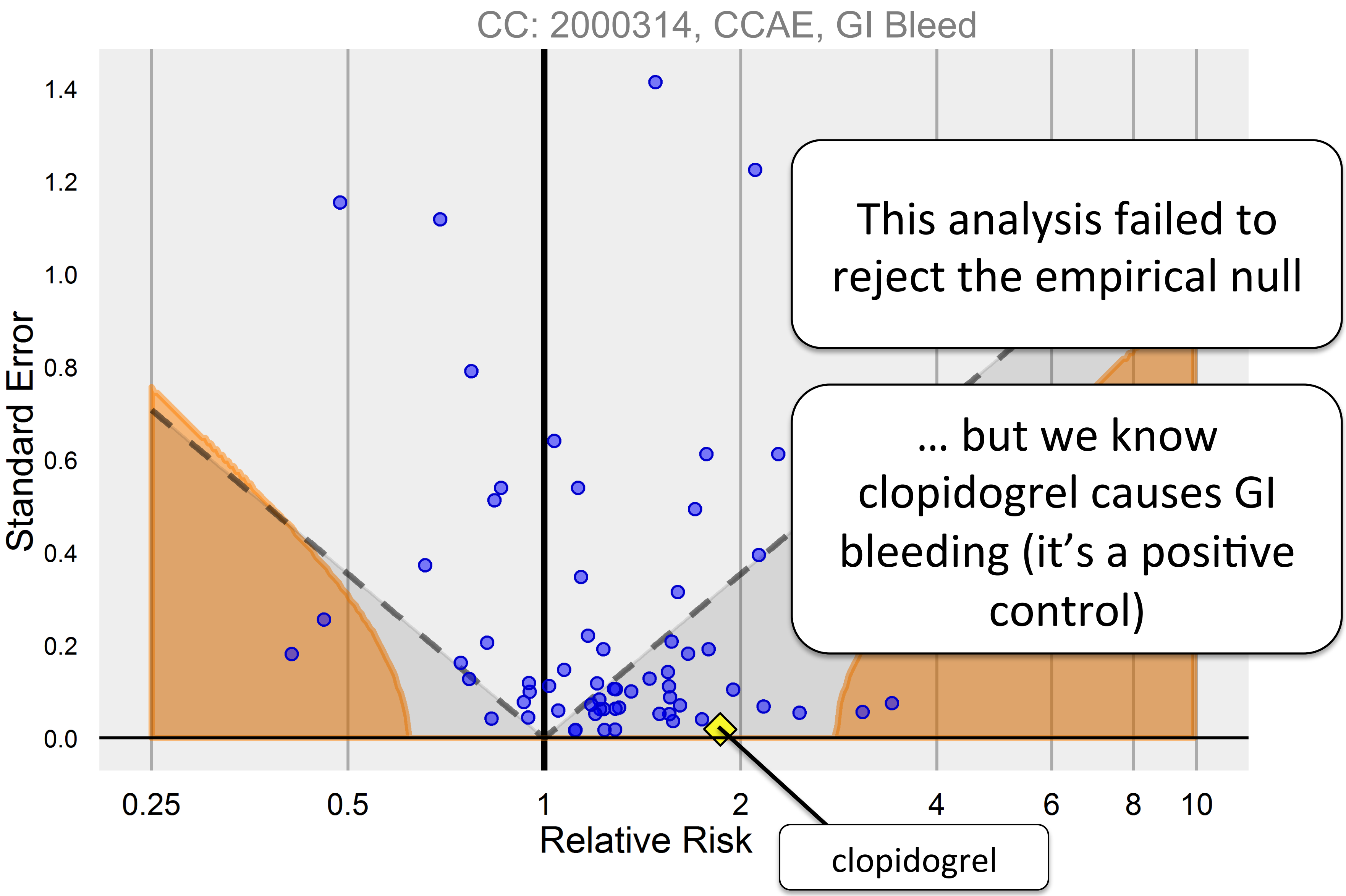


p < .05	**55%**
Calibrated p < .05	**6%**

clopidogrel:

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

p-value calibration plot

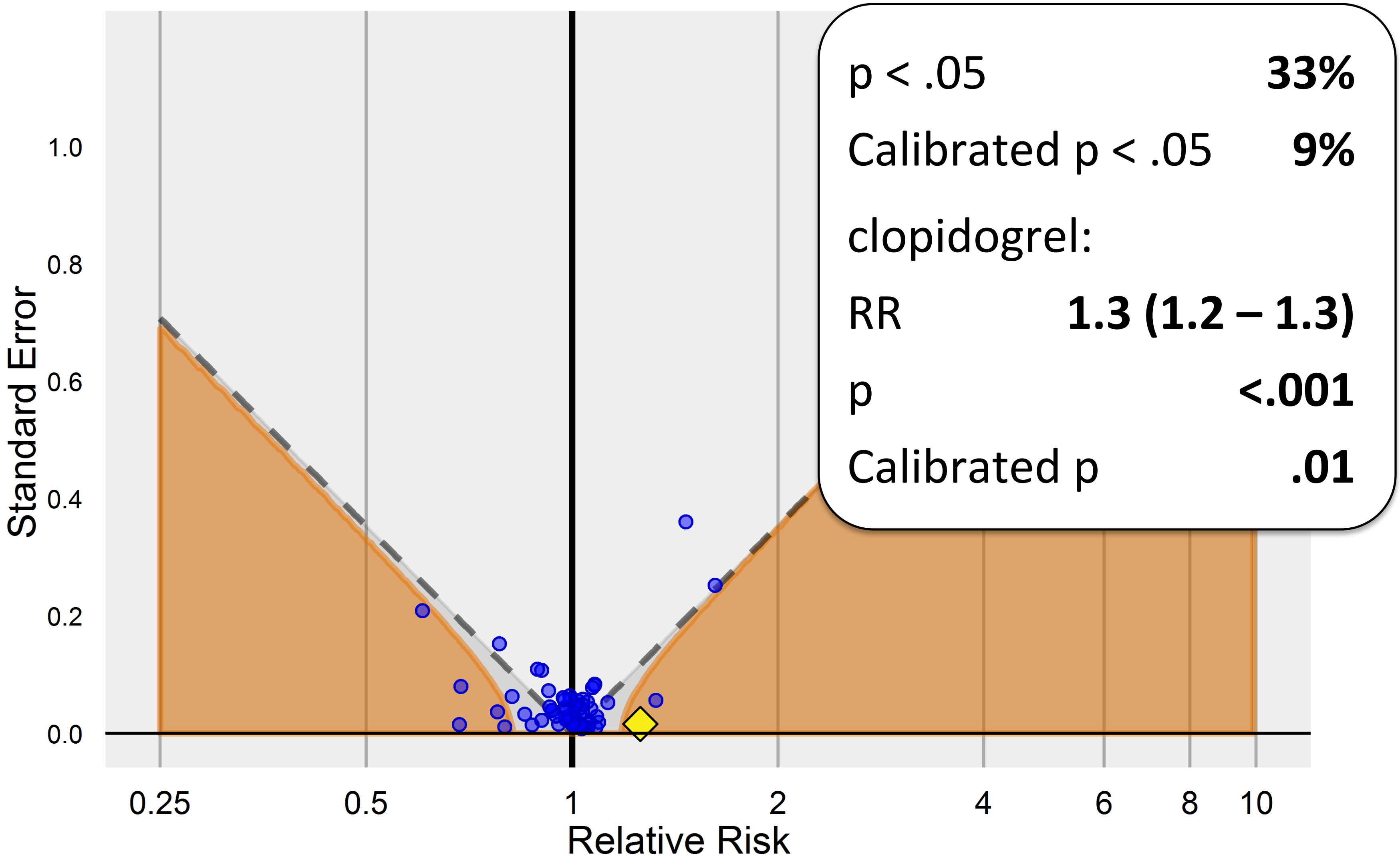


This analysis failed to reject the empirical null

… but we know clopidogrel causes GI bleeding (it's a positive control)

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**

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'

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)?

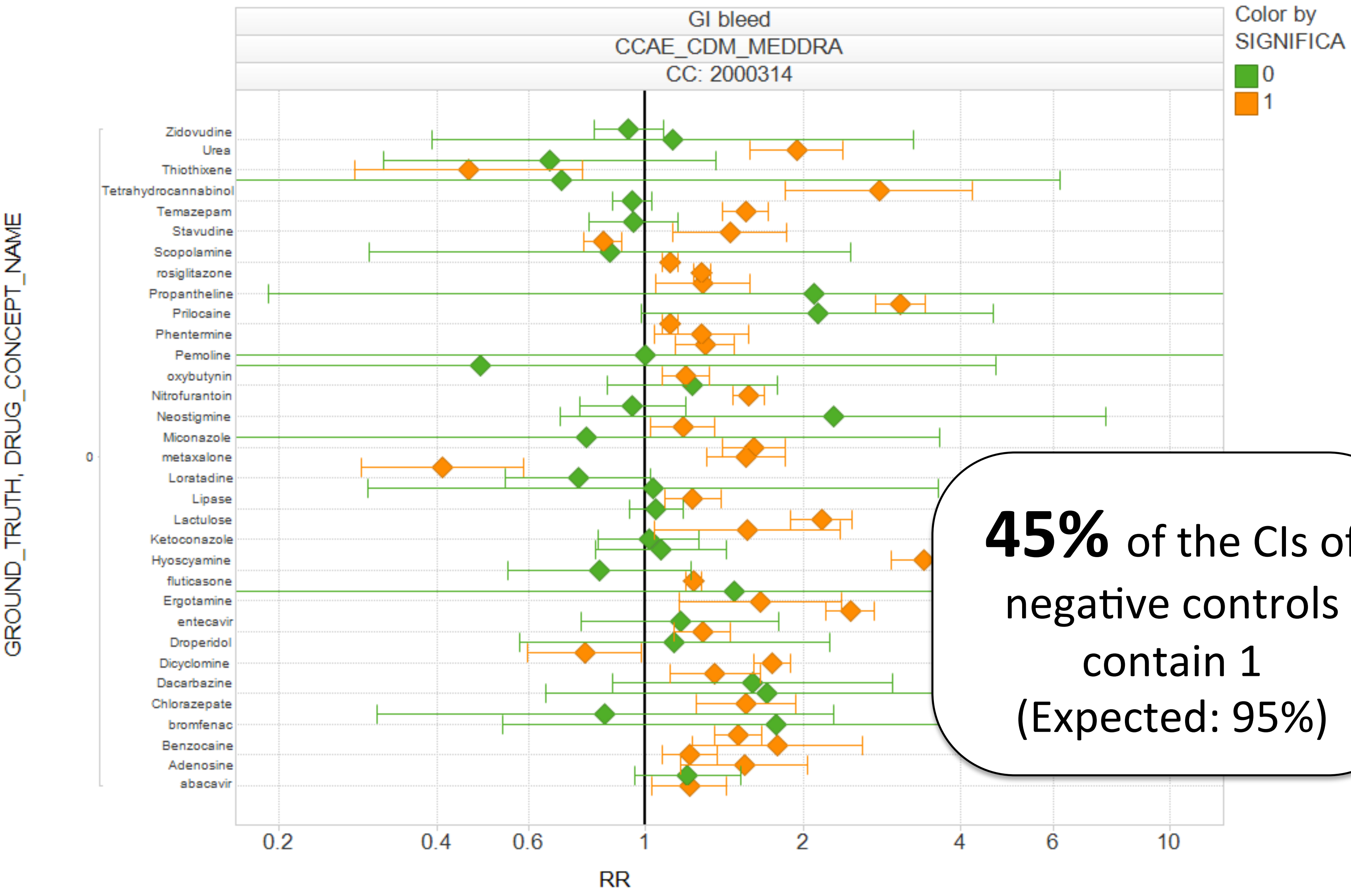
How big is the effect?

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

Estimating coverage probability

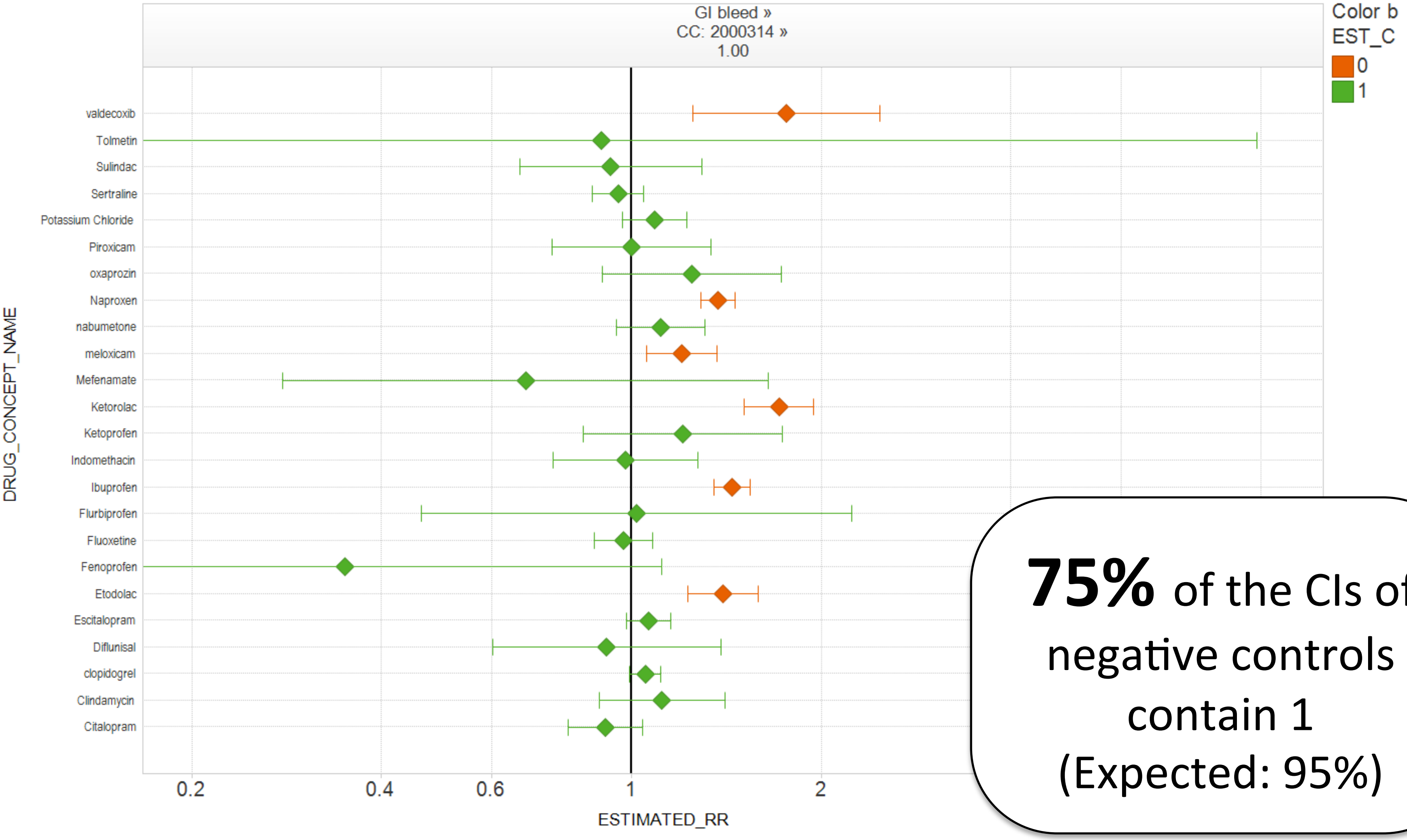
- What if a study design could be applied across a large sample of drug-outcome pairs for which we know the true effect?
- Coverage probability: the percentage of the test cases where the estimated confidence interval contains the true effect (LB 95 CI <= true effect <= UB 95 CI)
- Challenge: in real data, the 'true effect size' for negative controls can be assumed to be RR=1, but the RRs for positive controls are not known
- 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

Applying case-control design to negative controls in real data



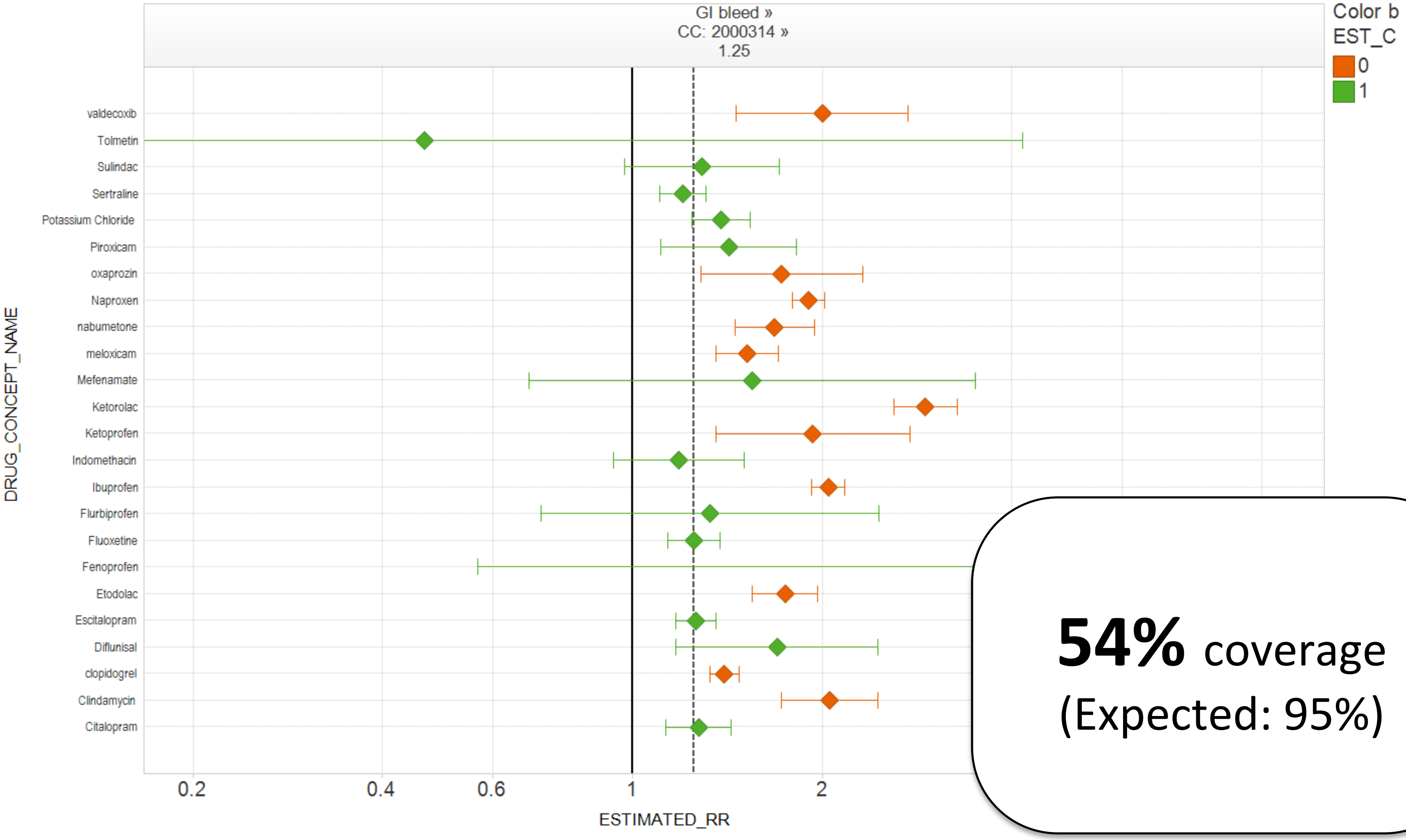
45% of the CIs of negative controls contain 1 (Expected: 95%)

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



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

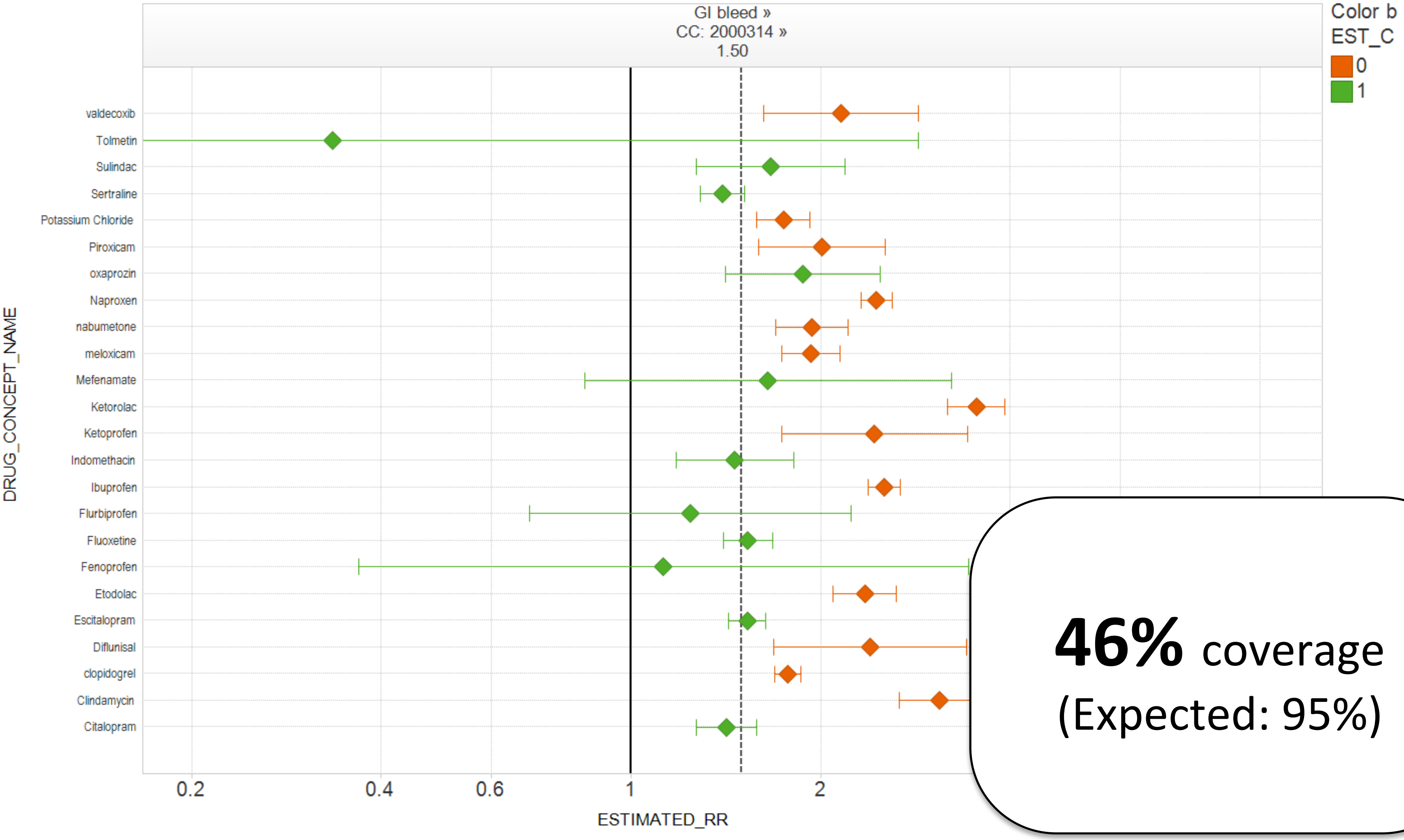
Applying case-control design to positive controls in simulated data, RR=1.25



54% coverage

(Expected: 95%)

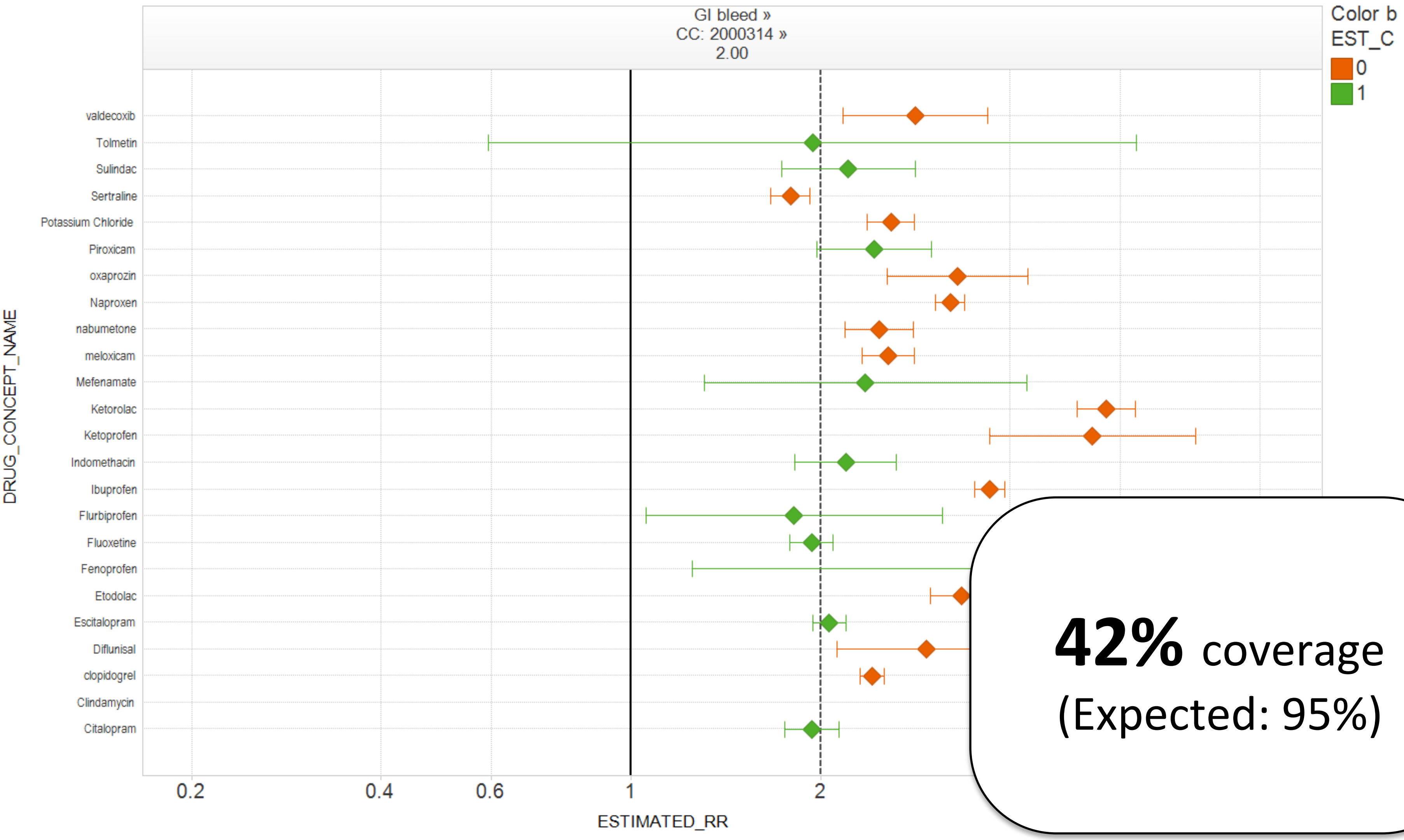
Applying case-control design to positive controls in simulated data, RR=1.50



46% coverage

(Expected: 95%)

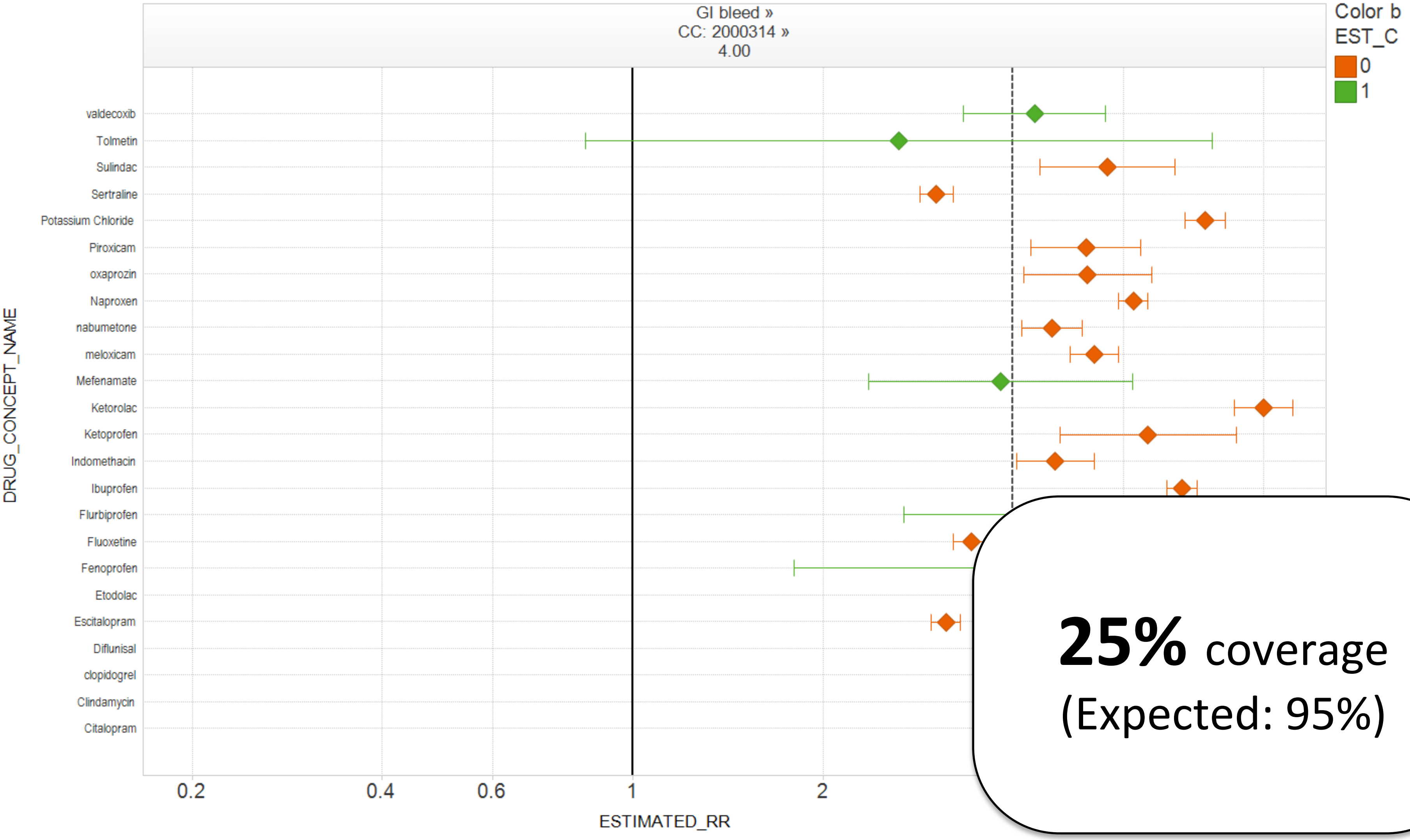
Applying case-control design to positive controls in simulated data, RR=2.00



42% coverage

(Expected: 95%)

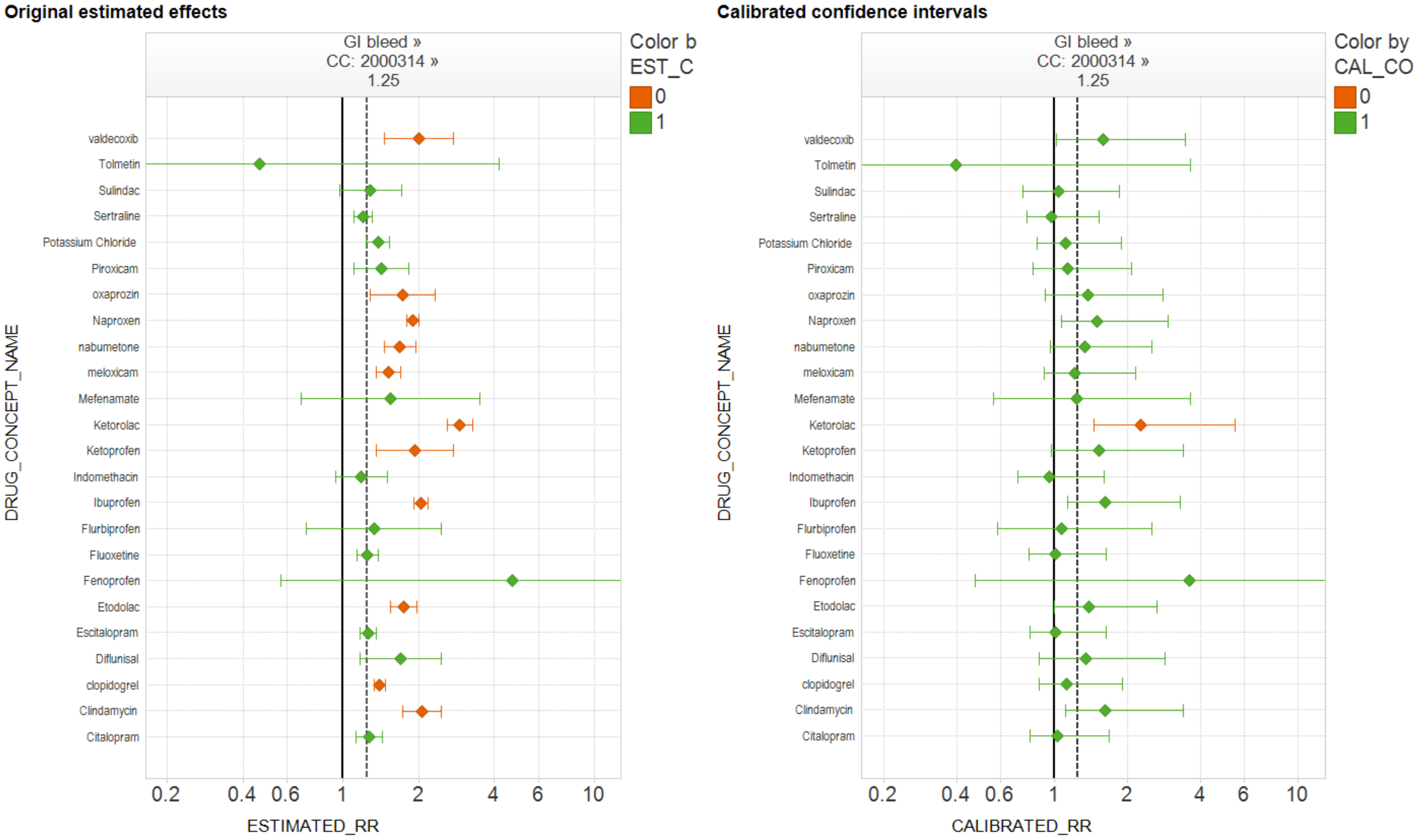
Applying case-control design to positive controls in simulated data, RR=4.00



25% coverage

(Expected: 95%)

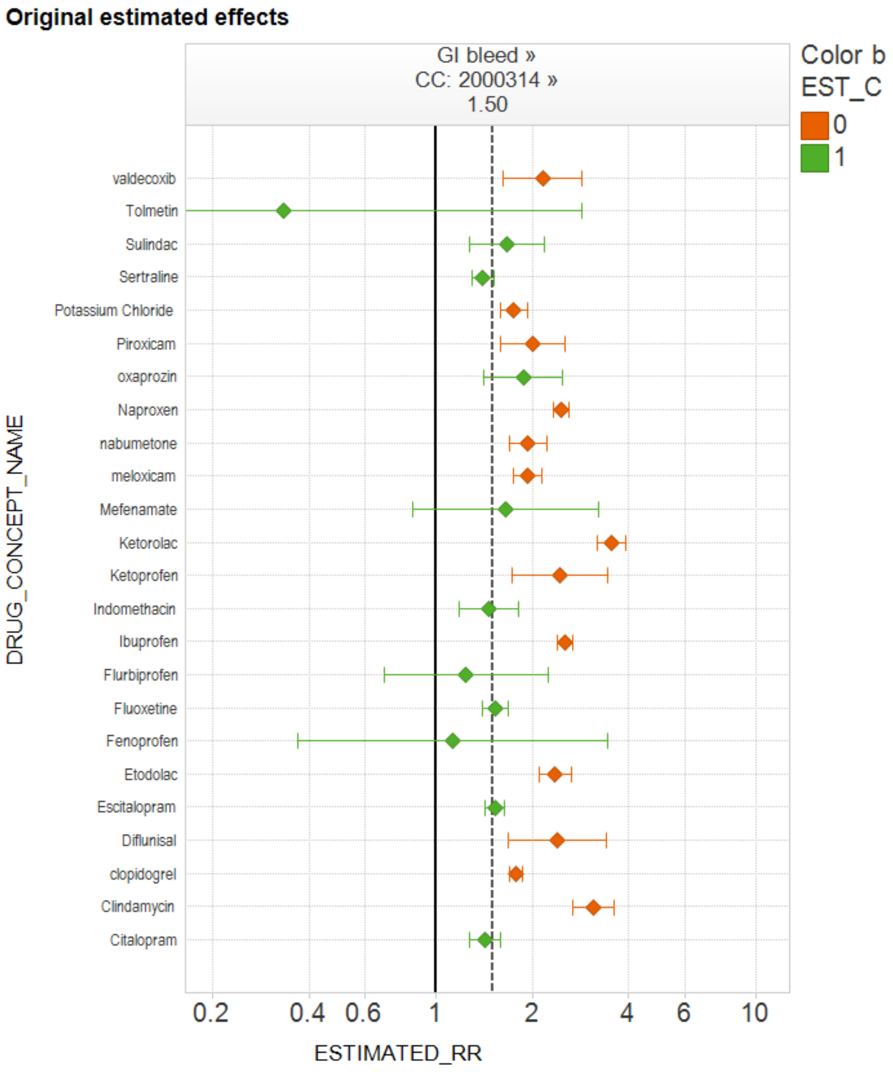
Applying case-control design and calibrating estimates of positive controls in simulated data, RR=1.25

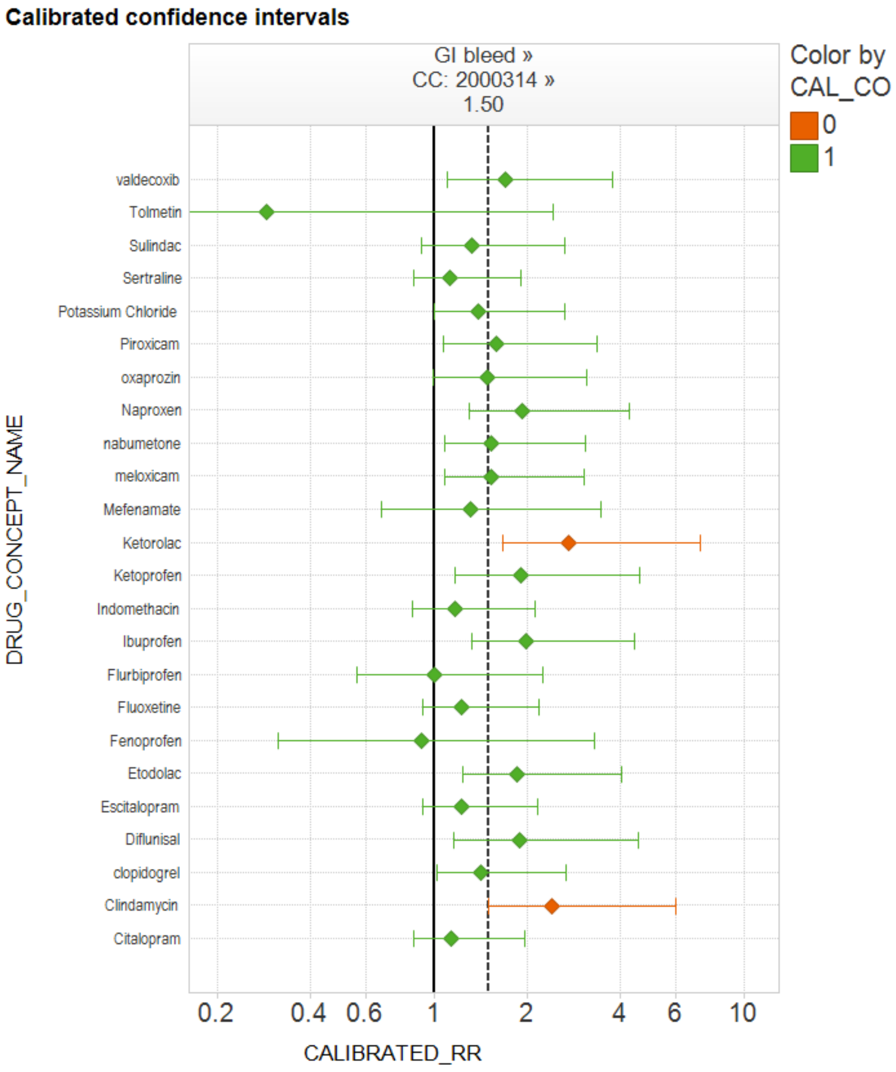


Original coverage probability = **54%**

Calibrated coverage probability = **96%**

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



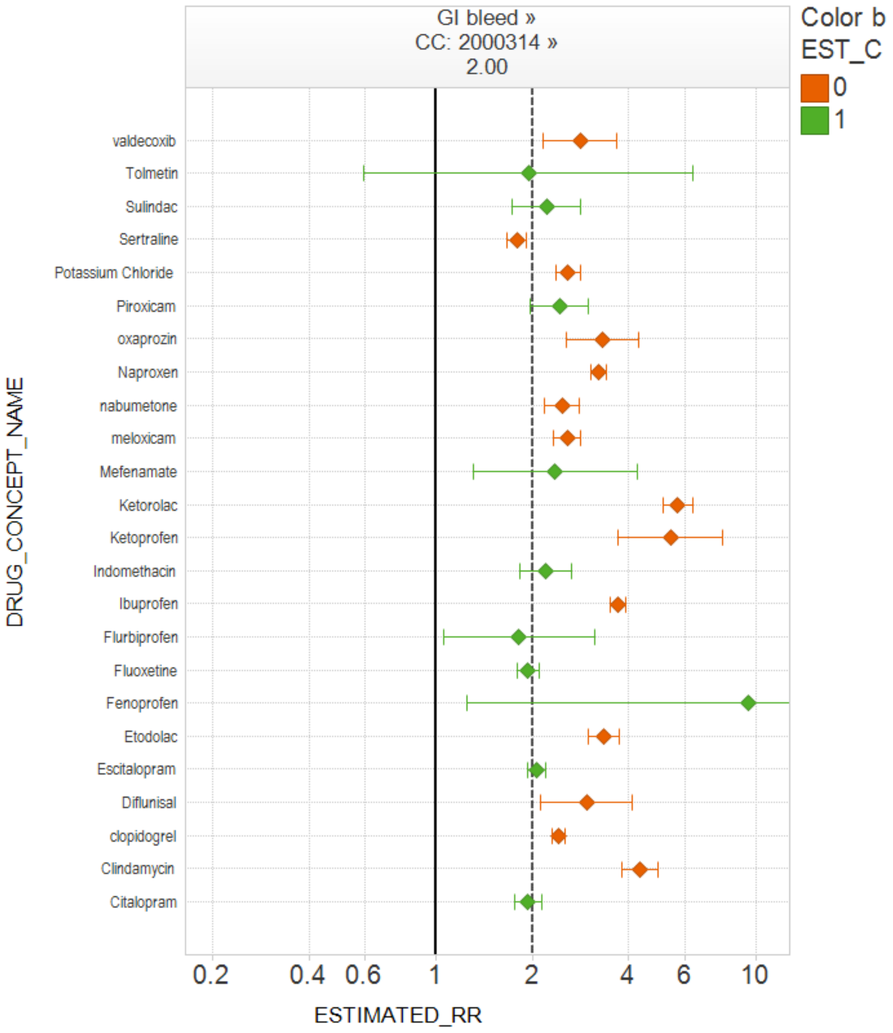


Original coverage probability = **46%**

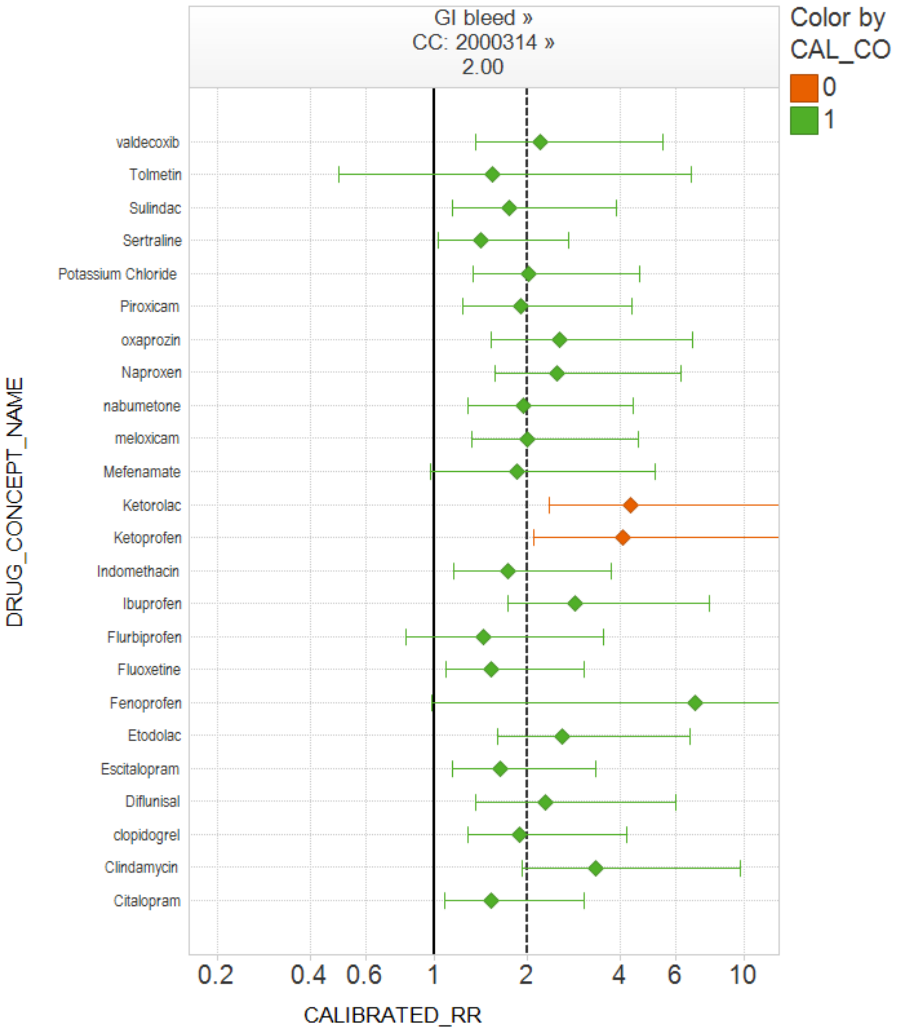
Calibrated coverage probability = **92%**

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

Original estimated effects



Calibrated confidence intervals

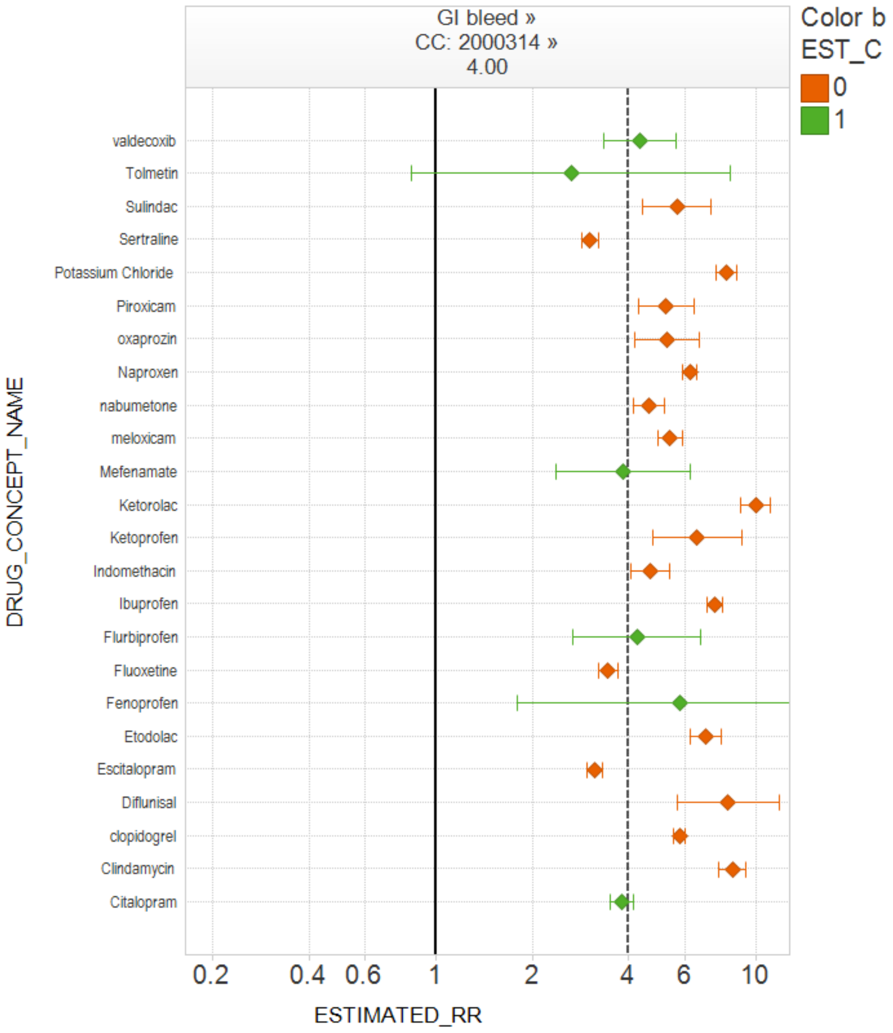


Original coverage probability = **42%**

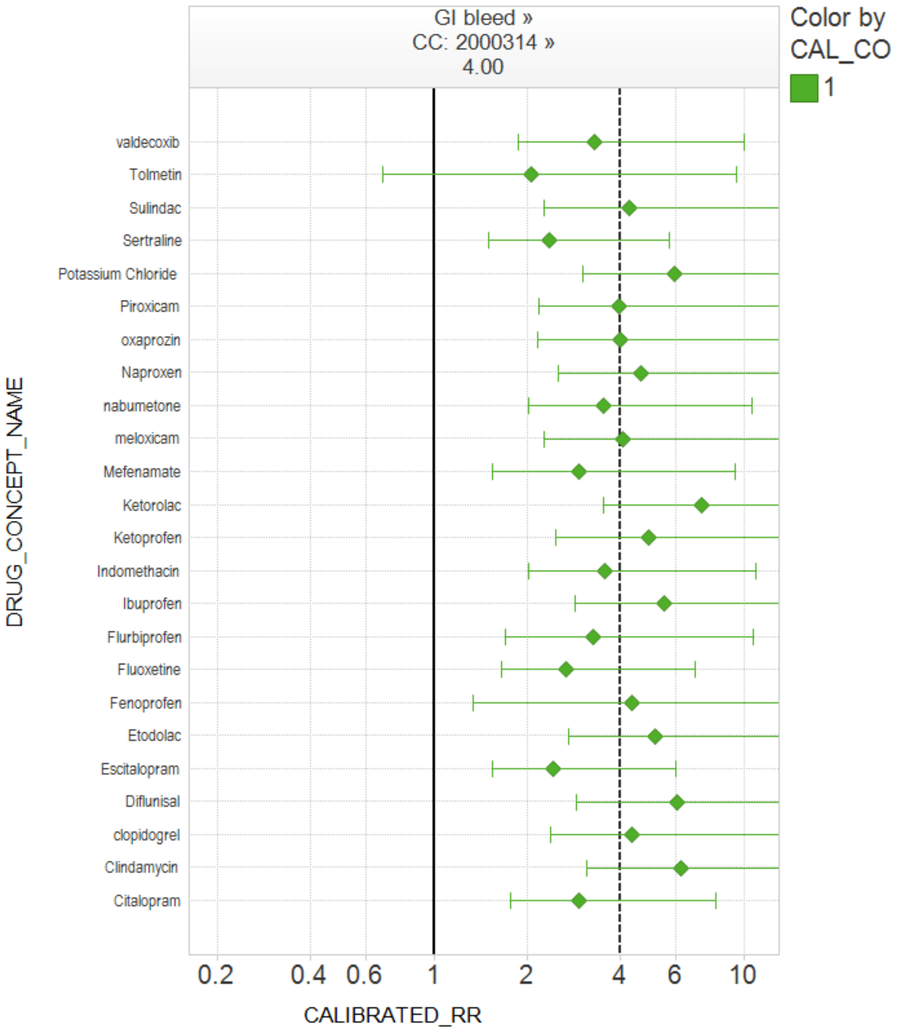
Calibrated coverage probability = **92%**

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

Original estimated effects



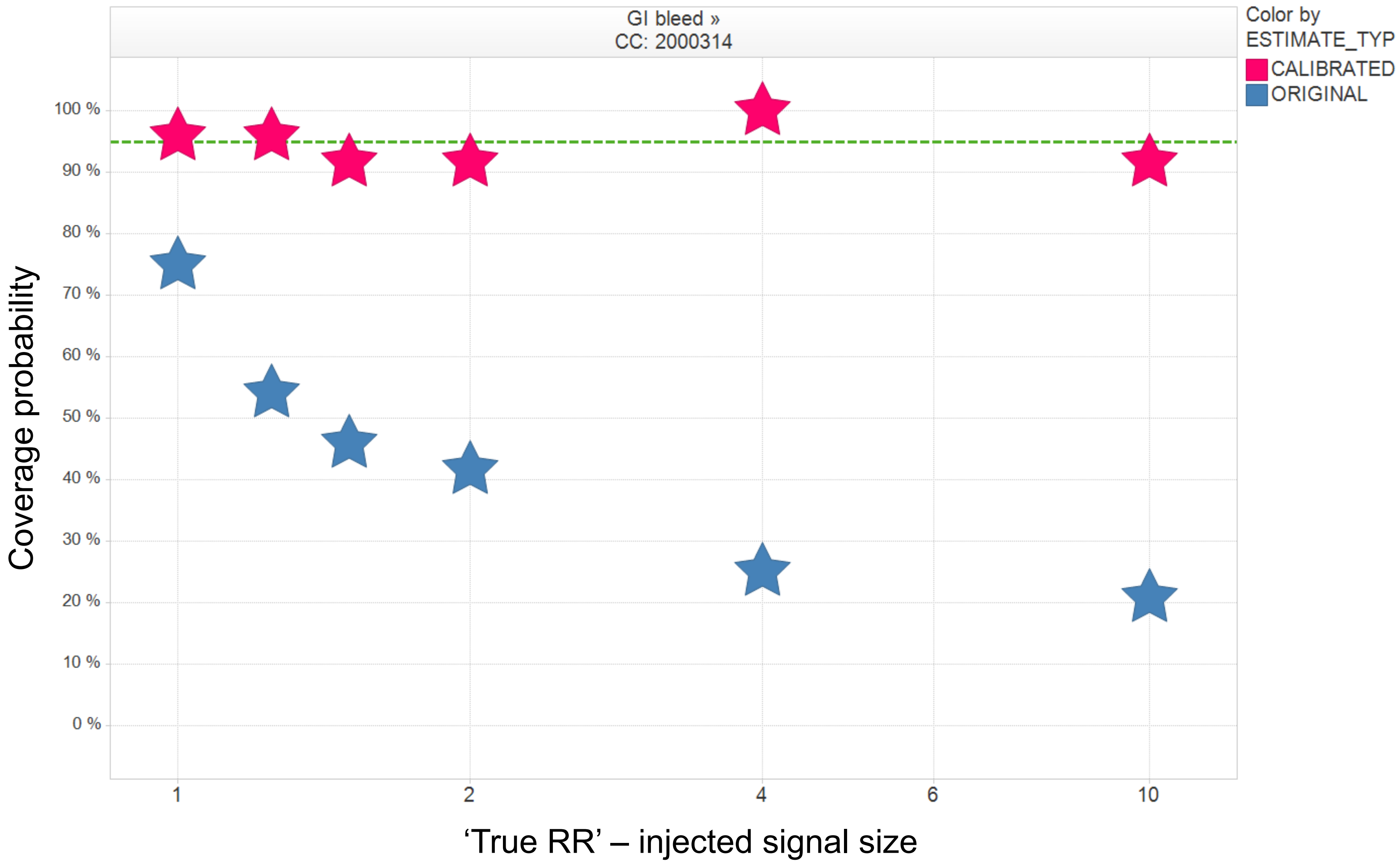
Calibrated confidence intervals



Original coverage probability = **25%**

Calibrated coverage probability = **100%**

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

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

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

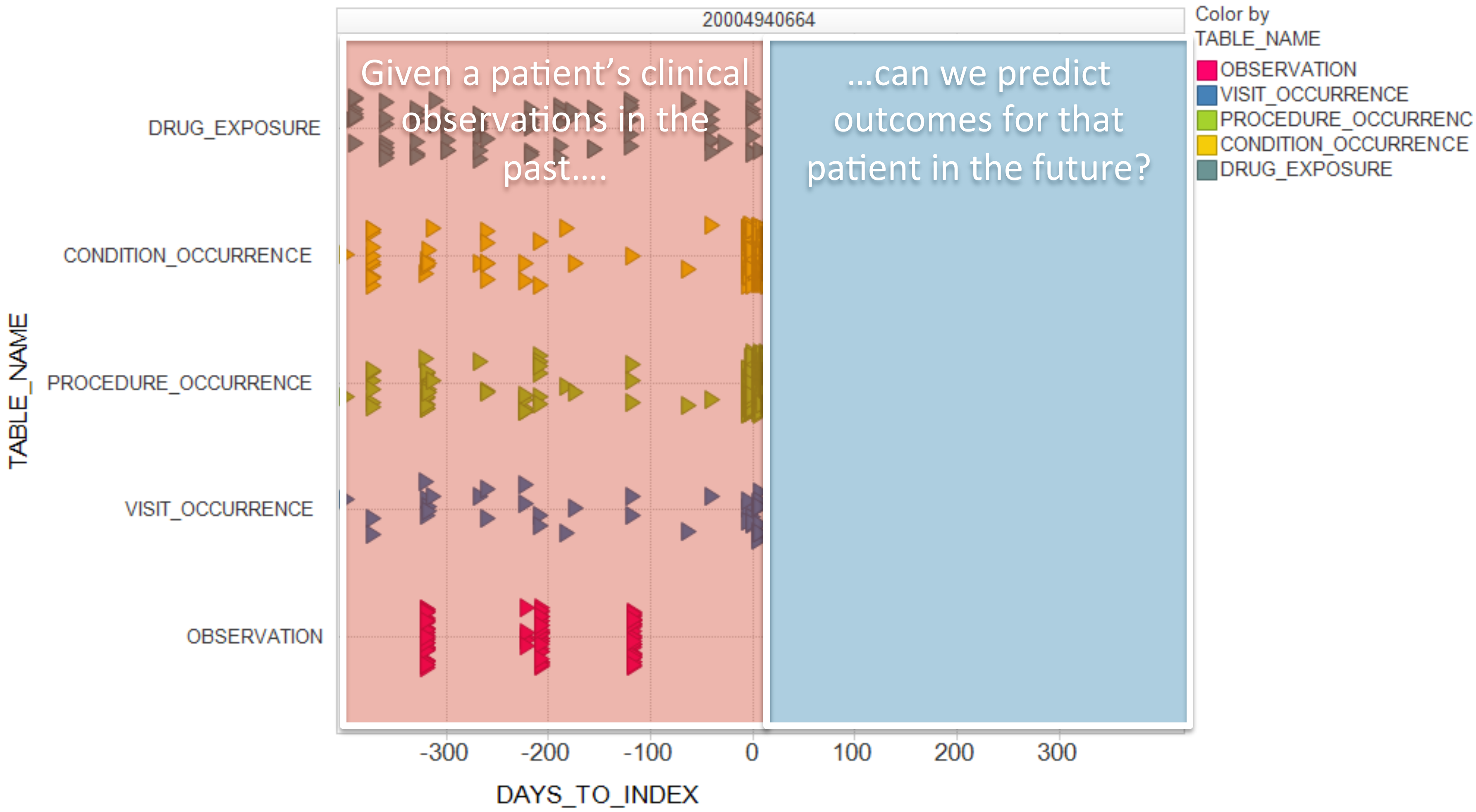
[1] Trifiro et al, PDS 2009

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

Predictive Modeling

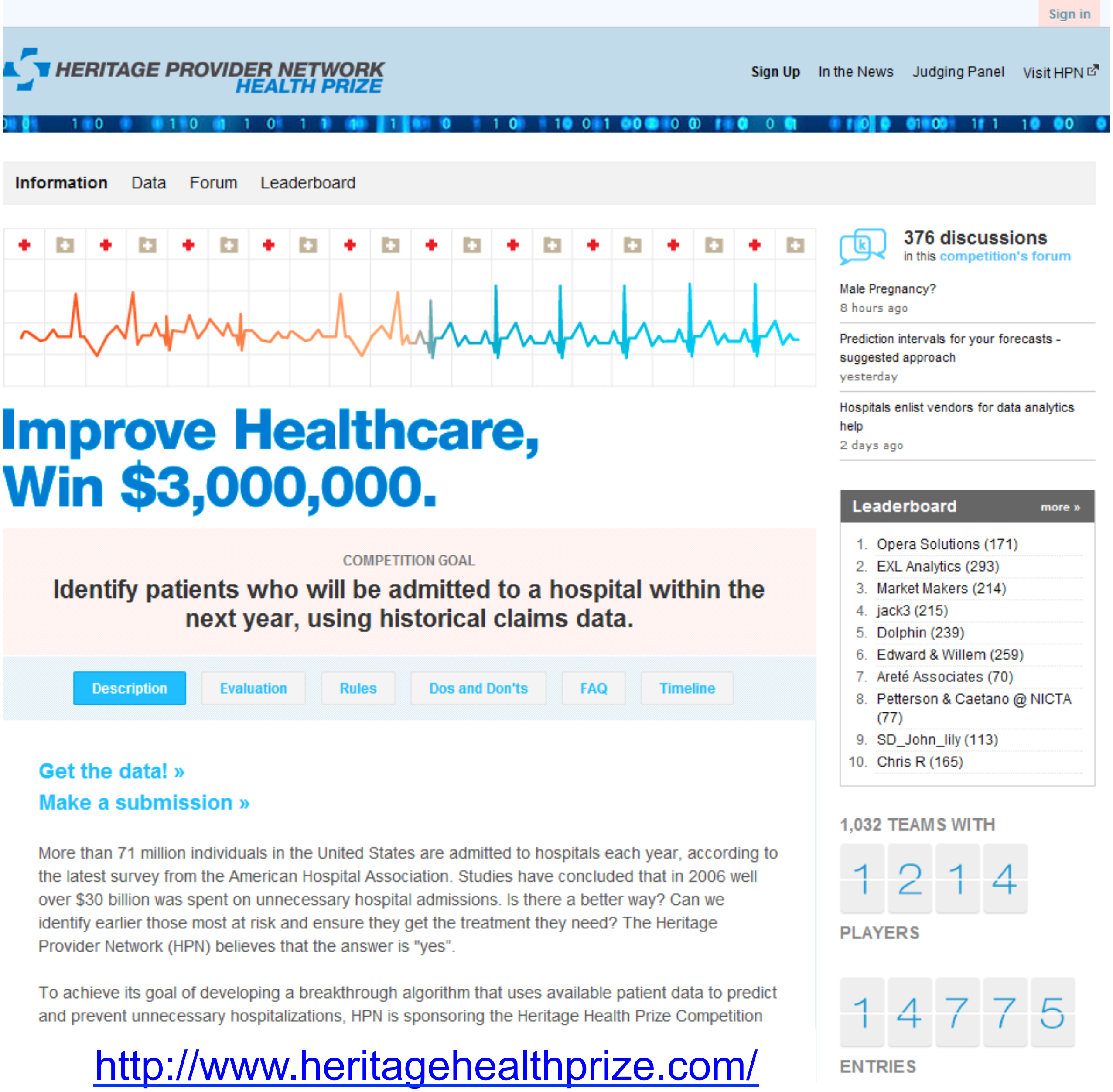
New Focus…



Given a patient's clinical observations in the past….

…can we predict outcomes for that patient in the future?

Patient-centered predictive modeling on big data has big value and big interest



Sign in

HERITAGE PROVIDER NETWORK
HEALTH PRIZE

Sign Up In the News Judging Panel Visit HPN

Information Data Forum Leaderboard

Improve Healthcare, Win $3,000,000.

COMPETITION GOAL

Identify patients who will be admitted to a hospital within the next year, using historical claims data.

Description Evaluation Rules Dos and Don'ts FAQ Timeline

Get the data! »

Make a submission »

More than 71 million individuals in the United States are admitted to hospitals each year, according to the latest survey from the American Hospital Association. Studies have concluded that in 2006 well over $30 billion was spent on unnecessary hospital admissions. Is there a better way? Can we identify earlier those most at risk and ensure they get the treatment they need? The Heritage Provider Network (HPN) believes that the answer is "yes".

To achieve its goal of developing a breakthrough algorithm that uses available patient data to predict and prevent unnecessary hospitalizations, HPN is sponsoring the Heritage Health Prize Competition

376 discussions
in this competition's forum

Male Pregnancy?
8 hours ago

Prediction intervals for your forecasts - suggested approach
yesterday

Hospitals enlist vendors for data analytics help
2 days ago

Leaderboard more »

1. Opera Solutions (171)
2. EXL Analytics (293)
3. Market Makers (214)
4. jack3 (215)
5. Dolphin (239)
6. Edward & Willem (259)
7. Areté Associates (70)
8. Petterson & Caetano @ NICTA (77)
9. SD_John_lily (113)
10. Chris R (165)

1,032 TEAMS WITH

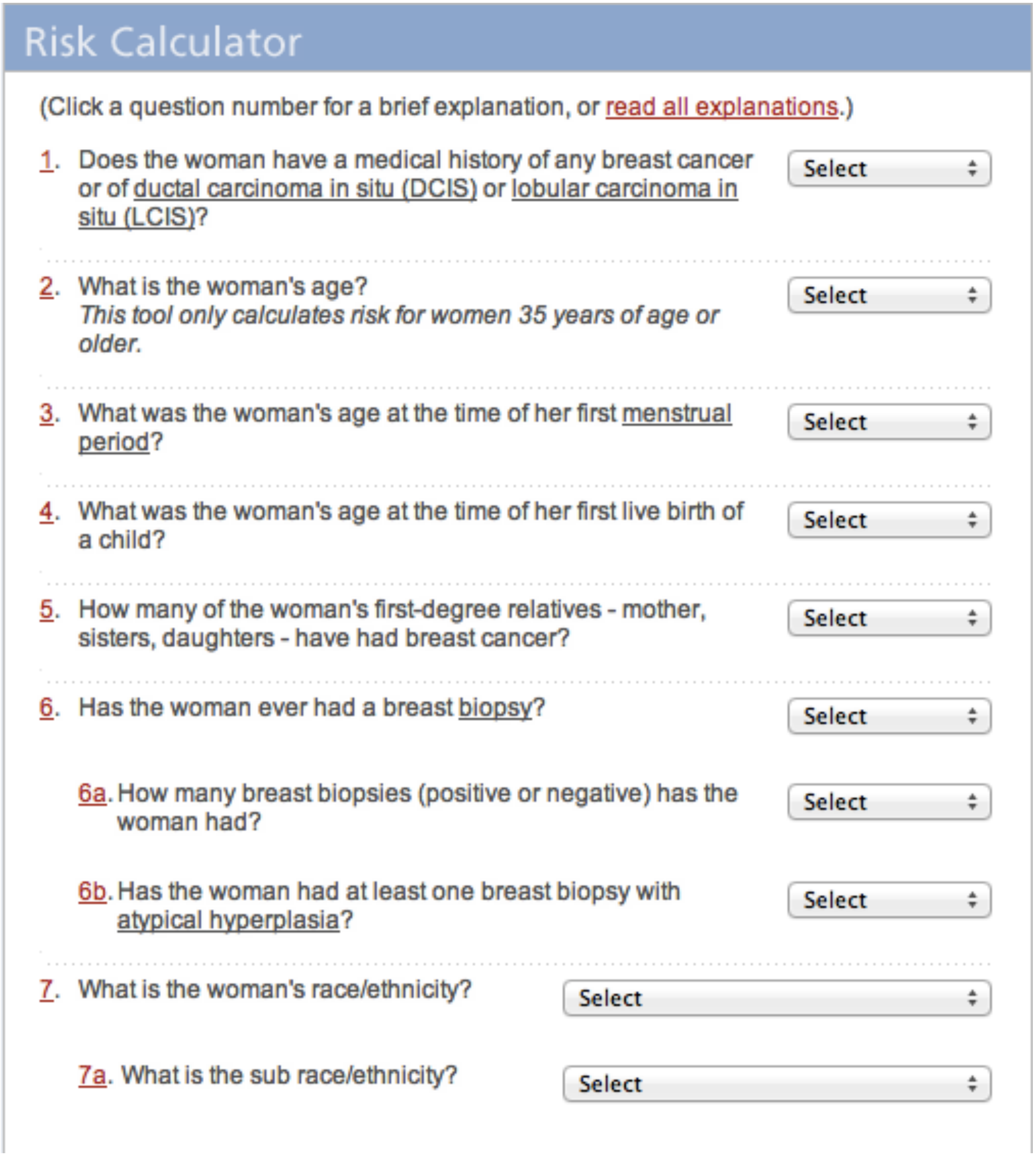
1214

PLAYERS

14775

ENTRIES

http://www.heritagehealthprize.com/



Risk Calculator

(Click a question number for a brief explanation, or read all explanations.)

1. Does the woman have a medical history of any breast cancer or of ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS)?

2. What is the woman's age?
This tool only calculates risk for women 35 years of age or older.

3. What was the woman's age at the time of her first menstrual period?

4. What was the woman's age at the time of her first live birth of a child?

5. How many of the woman's first-degree relatives - mother, sisters, daughters - have had breast cancer?

6. Has the woman ever had a breast biopsy?

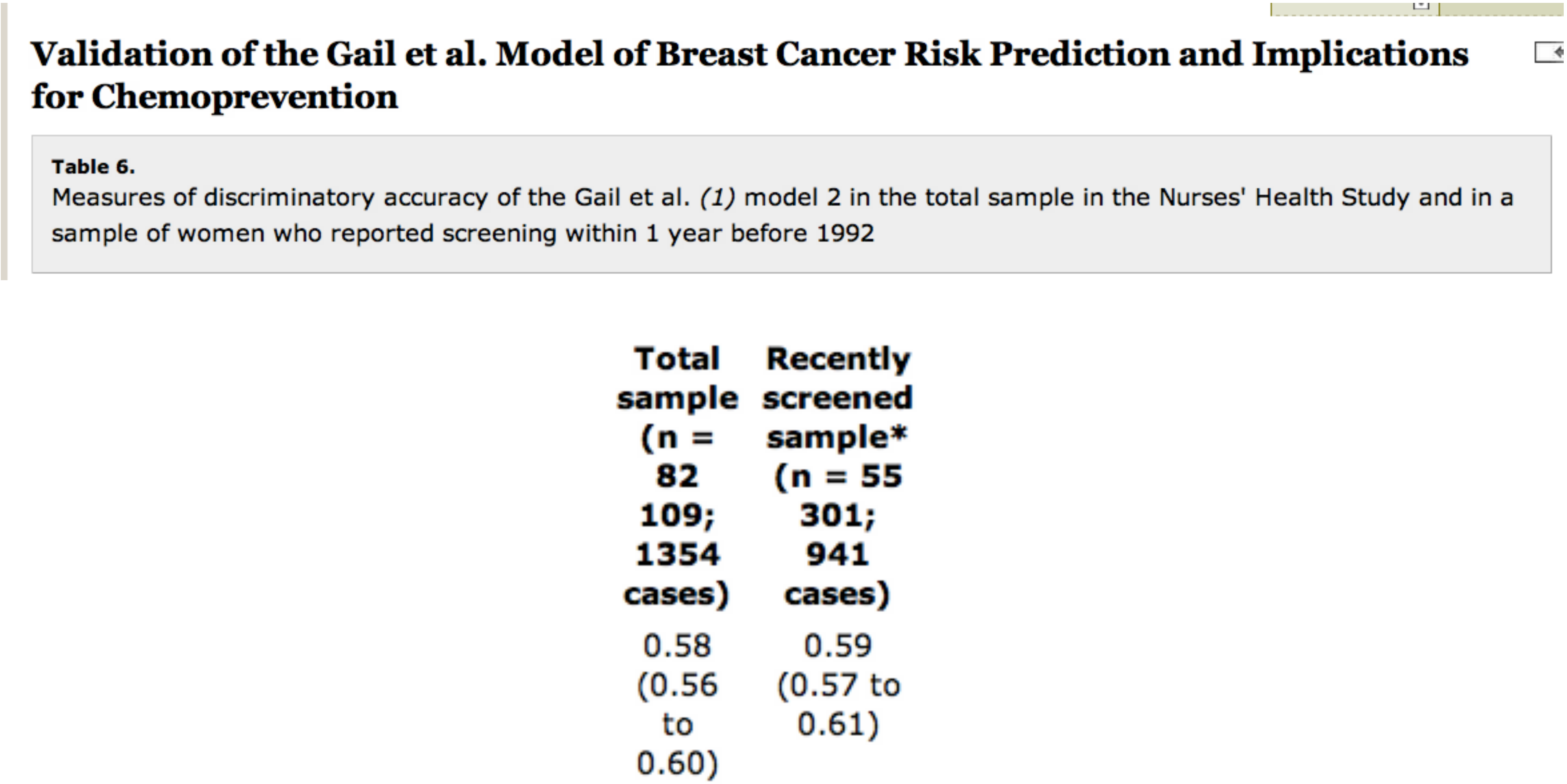
6a. How many breast biopsies (positive or negative) has the woman had?

6b. Has the woman had at least one breast biopsy with atypical hyperplasia?

7. What is the woman's race/ethnicity?

7a. What is the sub race/ethnicity?

Gail Breast Cancer Model



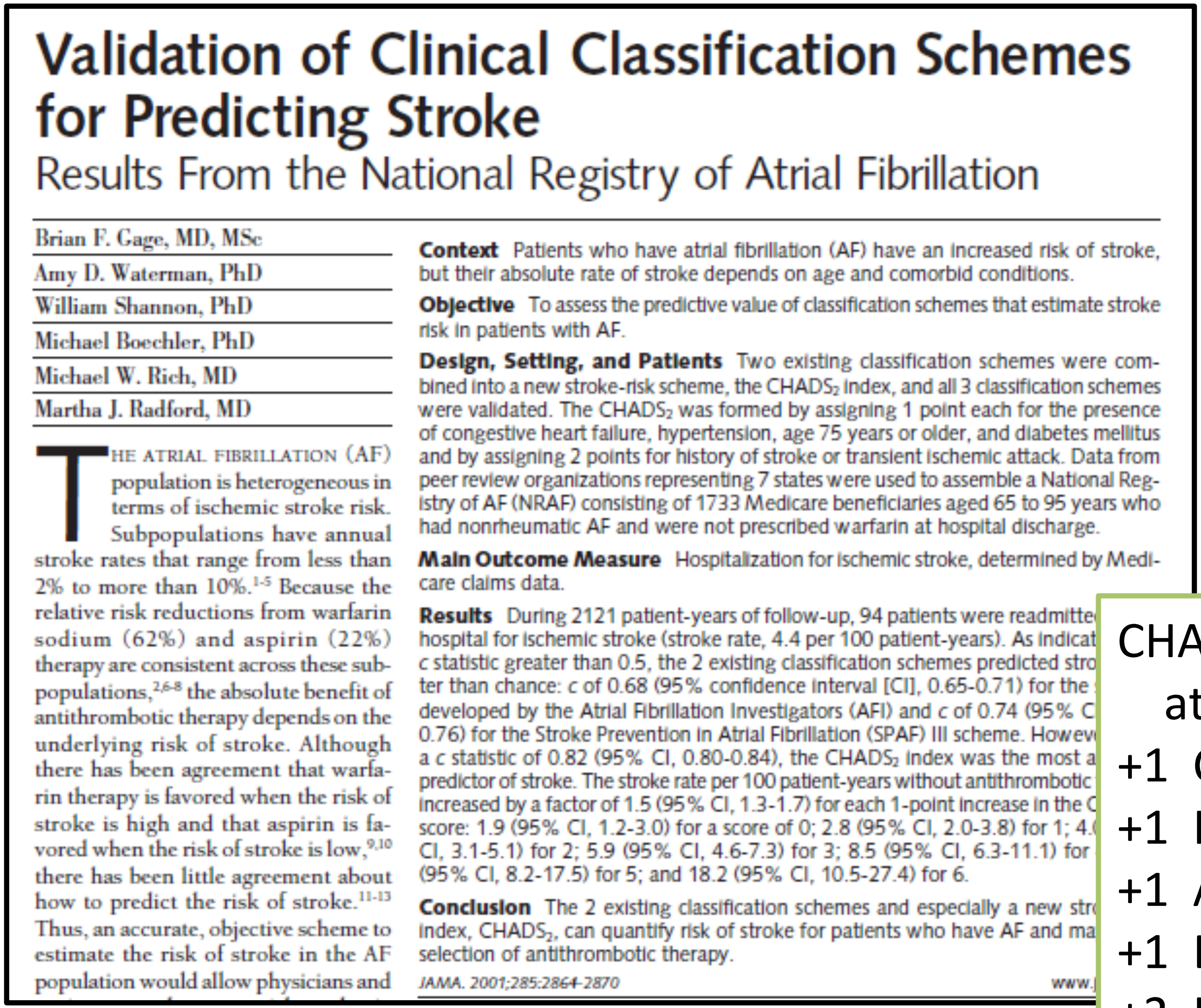
Validation of the Gail et al. Model of Breast Cancer Risk Prediction and Implications for Chemoprevention

Table 6.
Measures of discriminatory accuracy of the Gail et al. *(1)* model 2 in the total sample in the Nurses' Health Study and in a sample of women who reported screening within 1 year before 1992

Total sample (n = 82 109; 1354 cases)	Recently screened sample* (n = 55 301; 941 cases)
0.58 (0.56 to 0.60)	0.59 (0.57 to 0.61)

concordance coefficient

Patient-centered predictive models are already in clinical practice



Validation of Clinical Classification Schemes for Predicting Stroke

Results From the National Registry of Atrial Fibrillation

Brian F. Gage, MD, MSc
Amy D. Waterman, PhD
William Shannon, PhD
Michael Boechler, PhD
Michael W. Rich, MD
Martha J. Radford, MD

Context Patients who have atrial fibrillation (AF) have an increased risk of stroke, but their absolute rate of stroke depends on age and comorbid conditions.

Objective To assess the predictive value of classification schemes that estimate stroke risk in patients with AF.

Design, Setting, and Patients Two existing classification schemes were combined into a new stroke-risk scheme, the $CHADS_2$ index, and all 3 classification schemes were validated. The $CHADS_2$ was formed by assigning 1 point each for the presence of congestive heart failure, hypertension, age 75 years or older, and diabetes mellitus and by assigning 2 points for history of stroke or transient ischemic attack. Data from peer review organizations representing 7 states were used to assemble a National Registry of AF (NRAF) consisting of 1733 Medicare beneficiaries aged 65 to 95 years who had nonrheumatic AF and were not prescribed warfarin at hospital discharge.

Main Outcome Measure Hospitalization for ischemic stroke, determined by Medicare claims data.

Results During 2121 patient-years of follow-up, 94 patients were readmitte
hospital for ischemic stroke (stroke rate, 4.4 per 100 patient-years). As indicat
c statistic greater than 0.5, the 2 existing classification schemes predicted stro
ter than chance: *c* of 0.68 (95% confidence interval [CI], 0.65-0.71) for the
developed by the Atrial Fibrillation Investigators (AFI) and *c* of 0.74 (95% C
0.76) for the Stroke Prevention in Atrial Fibrillation (SPAF) III scheme. Howev
a *c* statistic of 0.82 (95% CI, 0.80-0.84), the $CHADS_2$ index was the most a
predictor of stroke. The stroke rate per 100 patient-years without antithrombotic
increased by a factor of 1.5 (95% CI, 1.3-1.7) for each 1-point increase in the C
score: 1.9 (95% CI, 1.2-3.0) for a score of 0; 2.8 (95% CI, 2.0-3.8) for 1; 4.
CI, 3.1-5.1) for 2; 5.9 (95% CI, 4.6-7.3) for 3; 8.5 (95% CI, 6.3-11.1) for
(95% CI, 8.2-17.5) for 5; and 18.2 (95% CI, 10.5-27.4) for 6.

Conclusion The 2 existing classification schemes and especially a new str
index, $CHADS_2$, can quantify risk of stroke for patients who have AF and ma
selection of antithrombotic therapy.

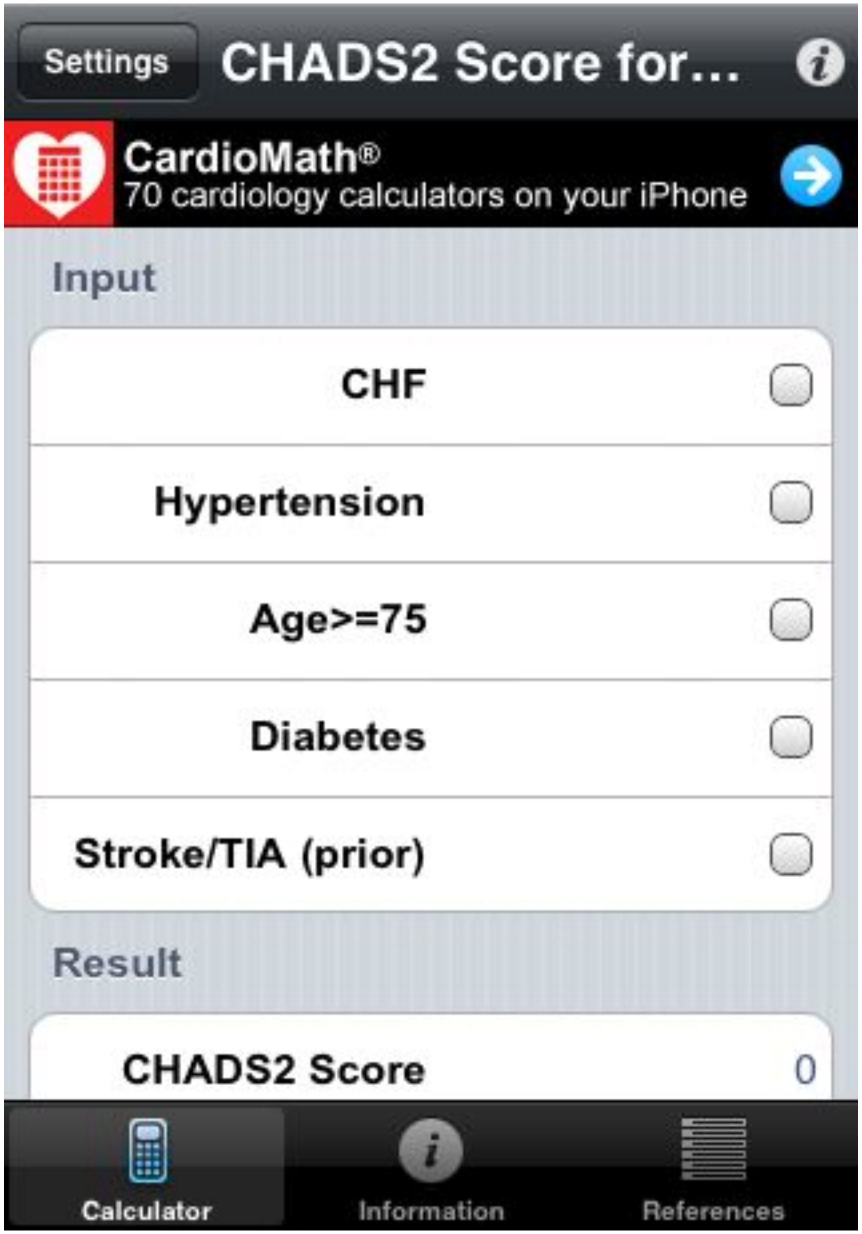
JAMA. 2001;285:2864-2870 www.j

THE ATRIAL FIBRILLATION (AF) population is heterogeneous in terms of ischemic stroke risk. Subpopulations have annual stroke rates that range from less than 2% to more than 10%.[1-5] Because the relative risk reductions from warfarin sodium (62%) and aspirin (22%) therapy are consistent across these subpopulations,[2,6-8] the absolute benefit of antithrombotic therapy depends on the underlying risk of stroke. Although there has been agreement that warfarin therapy is favored when the risk of stroke is high and that aspirin is favored when the risk of stroke is low,[9,10] there has been little agreement about how to predict the risk of stroke.[11-13] Thus, an accurate, objective scheme to estimate the risk of stroke in the AF population would allow physicians and

CHADS2 for patients with atrial fibrillation:

+1 Congestive heart failure
+1 Hypertension
+1 Age >= 75
+1 Diabetes mellitus
+2 History of transient ischemic attack

JAMA, 2001; 285: 2864-2870



CHADS2 Score for...

Input

CHF ☐

Hypertension ☐

Age>=75 ☐

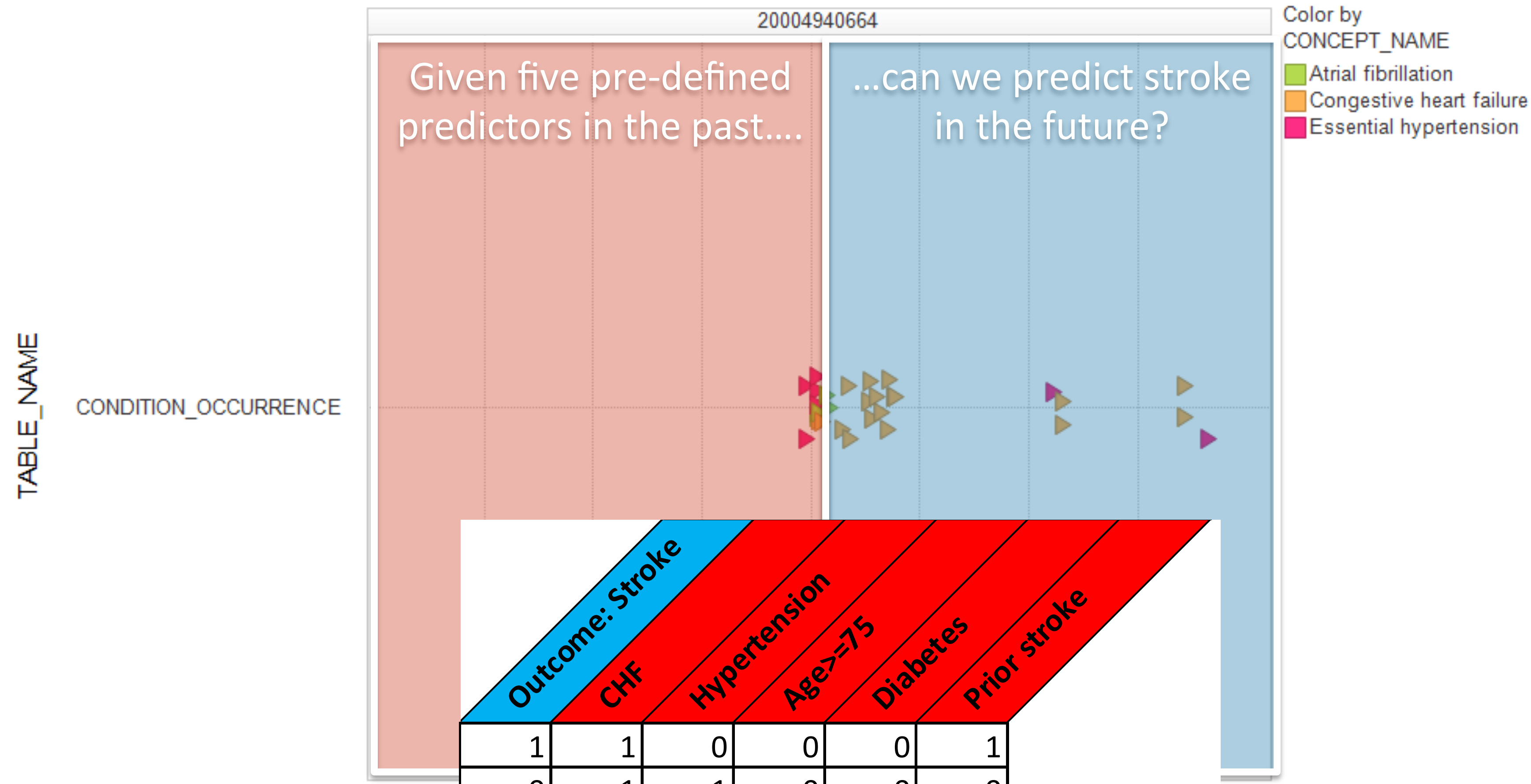
Diabetes ☐

Stroke/TIA (prior) ☐

Result

CHADS2 Score 0

Applying CHADS2 to a patient



Given five pre-defined predictors in the past….

…can we predict stroke in the future?

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

81

Evaluating the predictive accuracy of CHADS2

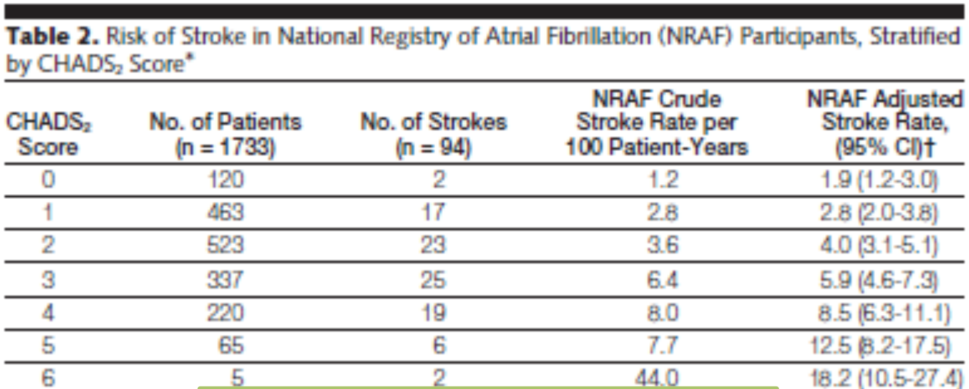
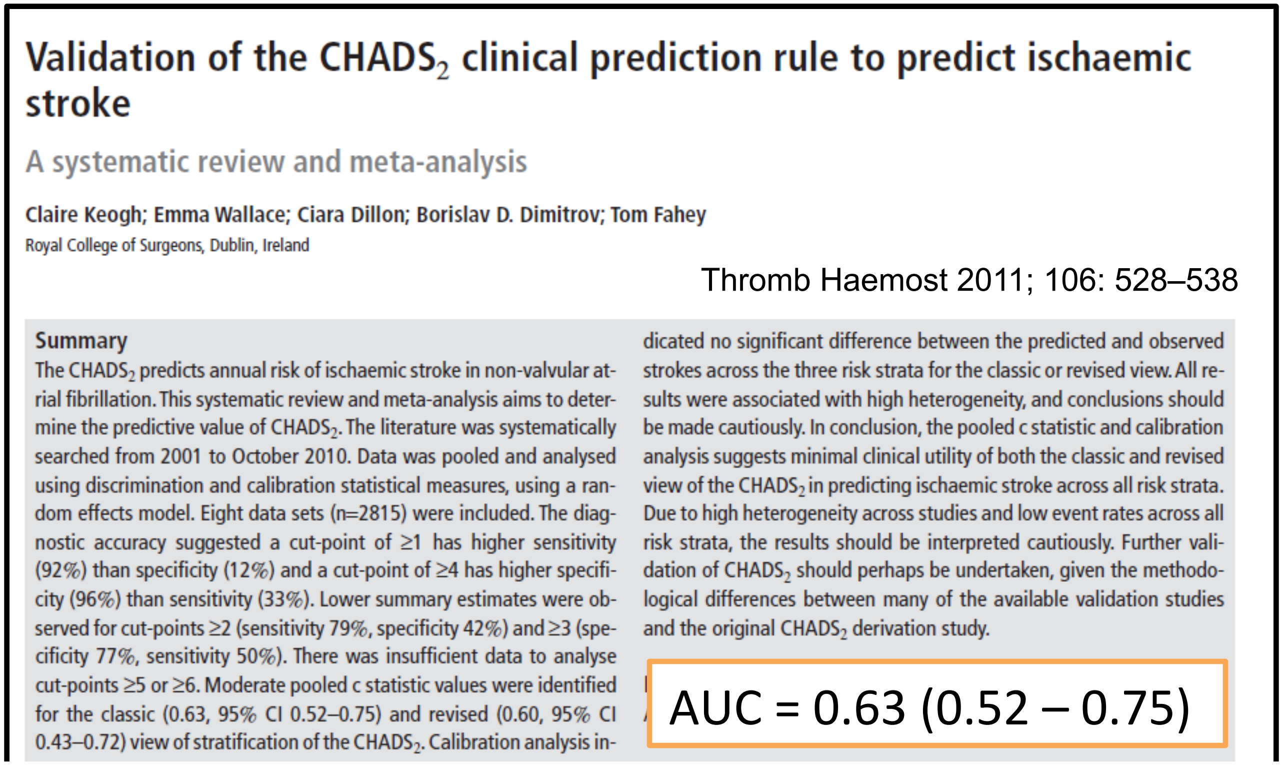


Table 2. Risk of Stroke in National Registry of Atrial Fibrillation (NRAF) Participants, Stratified by CHADS$_2$ Score*

CHADS$_2$ Score	No. of Patients (n = 1733)	No. of Strokes (n = 94)	NRAF Crude Stroke Rate per 100 Patient-Years	NRAF Adjusted Stroke Rate, (95% CI)†
0	120	2	1.2	1.9 (1.2-3.0)
1	463	17	2.8	2.8 (2.0-3.8)
2	523	23	3.6	4.0 (3.1-5.1)
3	337	25	6.4	5.9 (4.6-7.3)
4	220	19	8.0	8.5 (6.3-11.1)
5	65	6	7.7	12.5 (8.2-17.5)
6	5	2	44.0	18.2 (10.5-27.4)

JAMA, 2001; 285: 2864-2870

AUC = 0.82 (0.80 – 0.84)



Validation of the CHADS$_2$ clinical prediction rule to predict ischaemic stroke

A systematic review and meta-analysis

Claire Keogh; Emma Wallace; Ciara Dillon; Borislav D. Dimitrov; Tom Fahey

Royal College of Surgeons, Dublin, Ireland

Thromb Haemost 2011; 106: 528–538

Summary

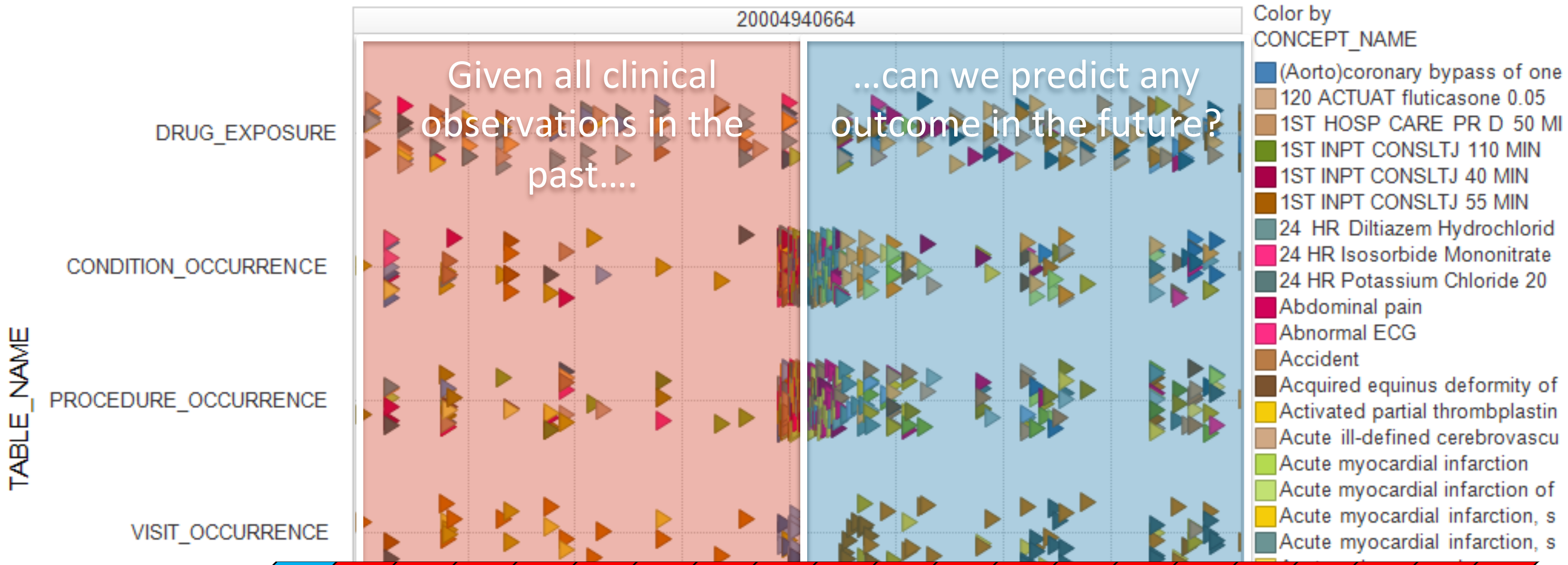
The CHADS$_2$ predicts annual risk of ischaemic stroke in non-valvular atrial fibrillation. This systematic review and meta-analysis aims to determine the predictive value of CHADS$_2$. The literature was systematically searched from 2001 to October 2010. Data was pooled and analysed using discrimination and calibration statistical measures, using a random effects model. Eight data sets (n=2815) were included. The diagnostic accuracy suggested a cut-point of ≥1 has higher sensitivity (92%) than specificity (12%) and a cut-point of ≥4 has higher specificity (96%) than sensitivity (33%). Lower summary estimates were observed for cut-points ≥2 (sensitivity 79%, specificity 42%) and ≥3 (specificity 77%, sensitivity 50%). There was insufficient data to analyse cut-points ≥5 or ≥6. Moderate pooled c statistic values were identified for the classic (0.63, 95% CI 0.52–0.75) and revised (0.60, 95% CI 0.43–0.72) view of stratification of the CHADS$_2$. Calibration analysis indicated no significant difference between the predicted and observed strokes across the three risk strata for the classic or revised view. All results were associated with high heterogeneity, and conclusions should be made cautiously. In conclusion, the pooled c statistic and calibration analysis suggests minimal clinical utility of both the classic and revised view of the CHADS$_2$ in predicting ischaemic stroke across all risk strata. Due to high heterogeneity across studies and low event rates across all risk strata, the results should be interpreted cautiously. Further validation of CHADS$_2$ should perhaps be undertaken, given the methodological differences between many of the available validation studies and the original CHADS$_2$ derivation study.

AUC = 0.63 (0.52 – 0.75)

Is CHADS2 as good as we can do?

- What about other measures of CHADS2 predictors?
  - Disease severity and progression
  - Medication adherence
  - Health service utilization
- What about other known risk factors?
  - Hypercholesterolemia
  - Atherosclerosis
  - Anticoagulant exposure
  - Tobacco use
  - Alcohol use
  - Obesity
  - Family history of stroke
- What about other unknown risk factors?

High-dimensional analytics can help reframe the prediction problem


20004940664
TABLE_NAME
DRUG_EXPOSURE
CONDITION_OCCURRENCE
PROCEDURE_OCCURRENCE
VISIT_OCCURRENCE
Color by CONCEPT_NAME
(Aorto)coronary bypass of one
120 ACTUAT fluticasone 0.05
1ST HOSP CARE PR D 50 MI
1ST INPT CONSLTJ 110 MIN
1ST INPT CONSLTJ 40 MIN
1ST INPT CONSLTJ 55 MIN
24 HR Diltiazem Hydrochlorid
24 HR Isosorbide Mononitrate
24 HR Potassium Chloride 20
Abdominal pain
Abnormal ECG
Accident
Acquired equinus deformity of
Activated partial thromboplastin
Acute ill-defined cerebrovascu
Acute myocardial infarction
Acute myocardial infarction of
Acute myocardial infarction, s
Acute myocardial infarction, s

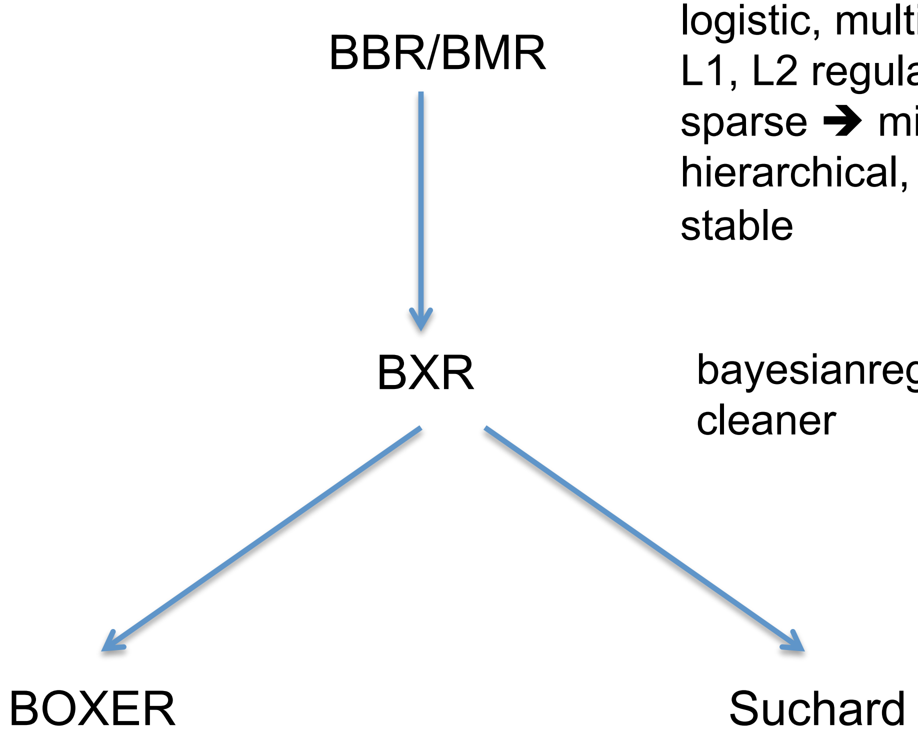
Given all clinical observations in the past….

…can we predict any outcome in the future?

Outcome: Stroke	Age	Gender	Race	Location	Drug 1	Drug 2	…	Drug n	Condition 1	Condition 2	…	Condition n	Procedure 1	Procedure 2	…	Procedure n	Lab 1	Lab 2	…	Lab n
0	76	M	B	441	0	0	1	1	1	1										
1	77	F	W	521	1	0	0	1	0	0										
1	96	F	B	215	1	1	1	0	1	0										
1	76	F	B	646	0	1	0	0	1	0										
0	64	M	B	379	0	0	1	1	1	1										
1	74	M	W	627	0	1	1	1	0	0										
1	68	M	B	348	0	0	0	1	0	0										
	Demographics				All drugs				All conditions				All procedures				All lab values			

Modern predictive modeling techniques, such as Bayesian logistic regression, can handle millions of covariates. The challenge is creating covariates that might be meaningful for the outcome of interest

Tools for Large-Scale Regression



BBR/BMR

bayesianregression.org
logistic, multinomial
L1, L2 regularization
sparse → millions of predictors
hierarchical, priors, autosearch
stable

BXR

bayesianregression.org
cleaner

BOXER

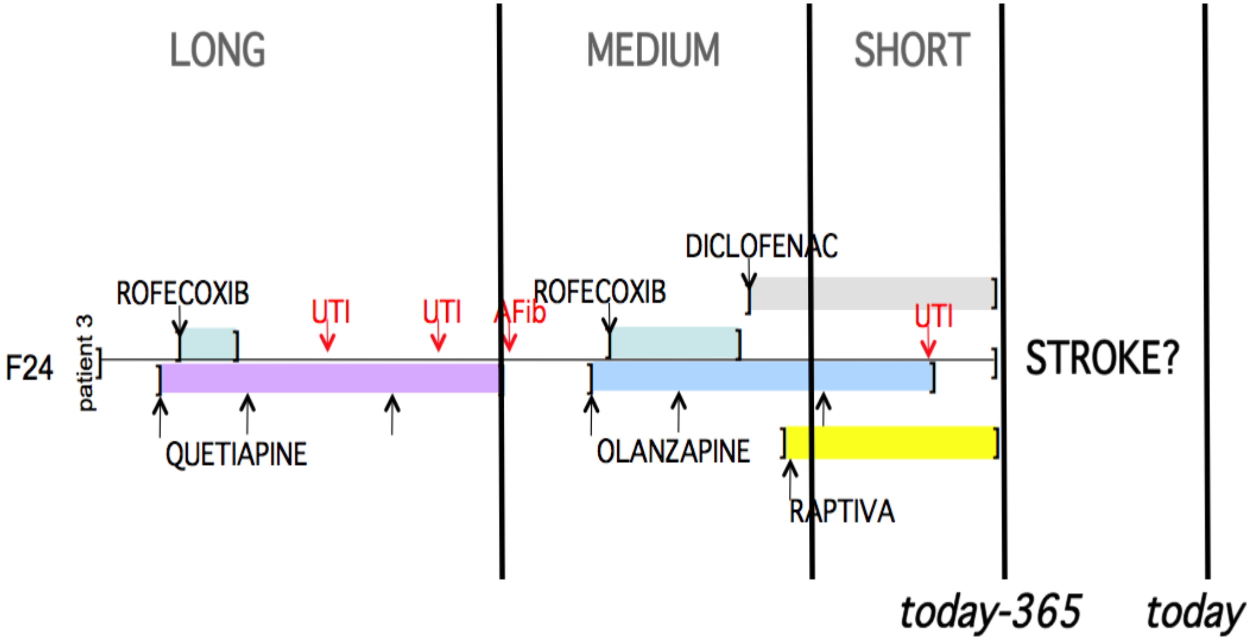
online logistic regression

Suchard

bsccs.googlecode.com
logistic, conditional logistic,
multinomial, Poisson, Cox,
ParamSurv, least squares
L1, L2 regularization
sparse → millions of predictors
imputation
CPU, GPU

Full Bayes?

Methodological Challenges

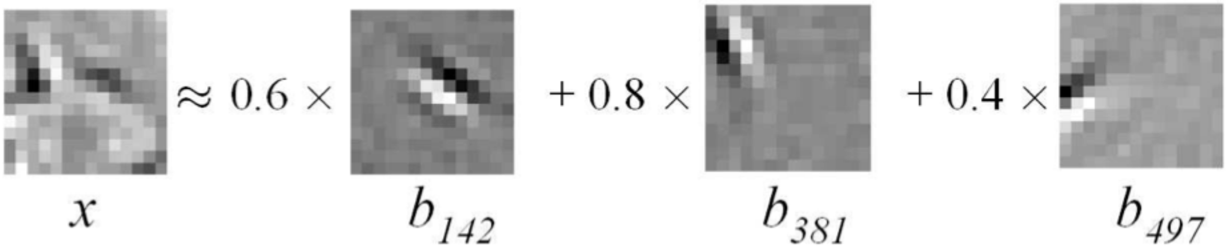


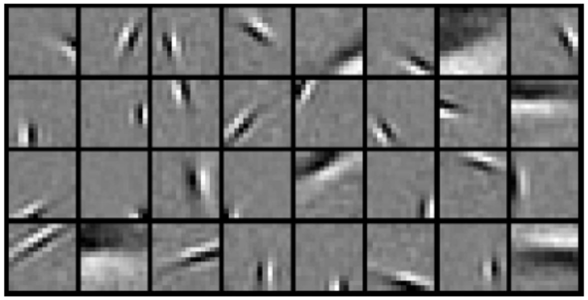
Central challenge: how to extract features from a longitudinal health record?

Sparse Coding: Learning Good Features

- Express each input vector as a linear combination of basis vectors
- Learn the basis *and* the weights:

$$\underset{a,b}{\operatorname{argmin}} \sum_i \left\| x^i - \sum_j a_j^i b_j \right\|_2^2 + \beta \left\| a^i \right\|_1 \text{ such that } \left\| b_j \right\|_2 \leq 1,\ j = 1, \dots, s, i = 1, \dots, n.$$





- Supervised sparse coding

Decision Tree Approach

(>-30, appendectomy, Y/N):
in the last 30 days, did the patient have an appendectomy?

(<0, max(SBP), 140):
at any time in the past did the patient's systolic blood pressure exceed 140 mmHg?

(<-90, rofecoxib, Y/N):
in the time period up to 90 days ago, did the patient have a prescription for rofecoxib?

(>-7, fever, Y/N):
in the last week, did the patient have a fever?

Rule Mining

McCormick, Rudin, Madigan

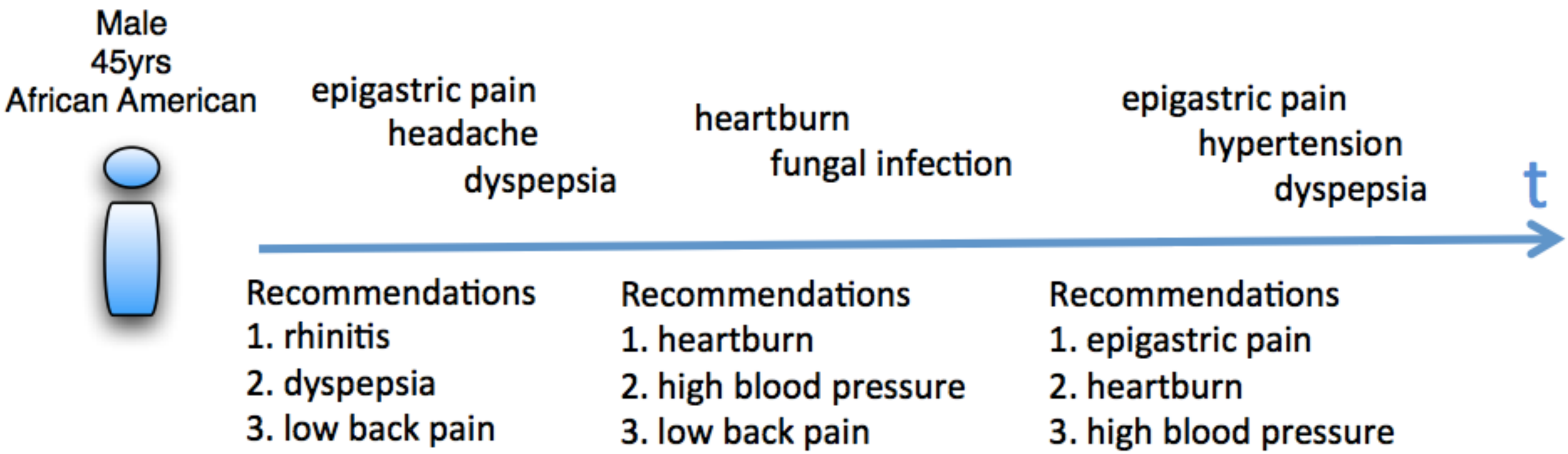
- Goal: Predict next event in current sequence given sequence database

- Association Rules:
  - item 1 and item 2 ➔ item 3
  - Recommender systems
  - Built-in explanation

- (Bayesian) Hierarchical Association Rule Mining

Predicting Medical Conditions

- Patients visit providers periodically
- Report time-stamped series of conditions since last encounter
- Predict next condition given past sequences



Male
45yrs
African American

epigastric pain
headache
dyspepsia

heartburn
fungal infection

epigastric pain
hypertension
dyspepsia

t

Recommendations
1. rhinitis
2. dyspepsia
3. low back pain

Recommendations
1. heartburn
2. high blood pressure
3. low back pain

Recommendations
1. epigastric pain
2. heartburn
3. high blood pressure

- Observe y_{ir} co-occurrences (support for lhs $\cup$ rhs) for patient i and rule r
- n_{ir} encounters that include the lhs
- Hierarchical Association Rule Model (HARM)

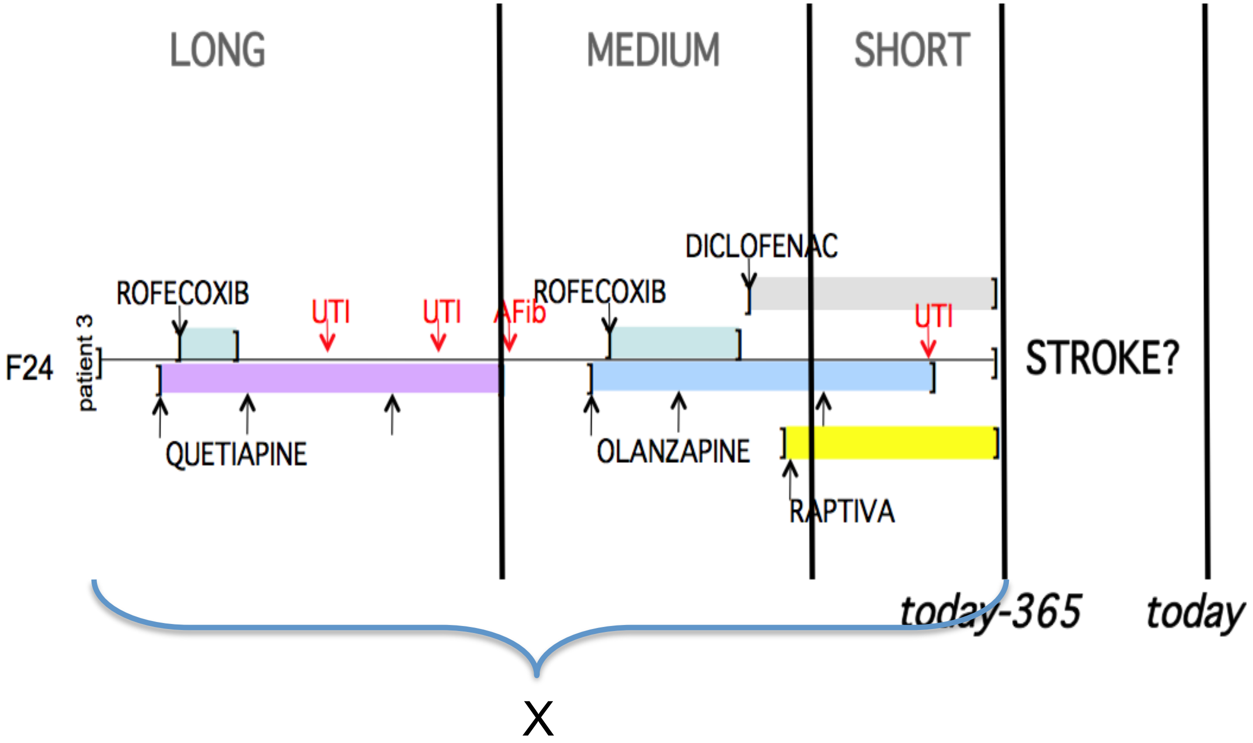
$$\begin{aligned} y_{ir} &\sim \text{Binomial}(n_{ir}, p_{ir}) \\ p_{ir} &\sim \text{Beta}(\pi_{ir}, \tau_i) \end{aligned}$$

- Model π_{ir} hierarchically

$$\pi_{ir} = \exp(\mathbf{M}_i'\beta_r + \gamma_i)$$

- $\mathbf{M}$ is matrix of patient characteristics, γ_i is patient-specific variation

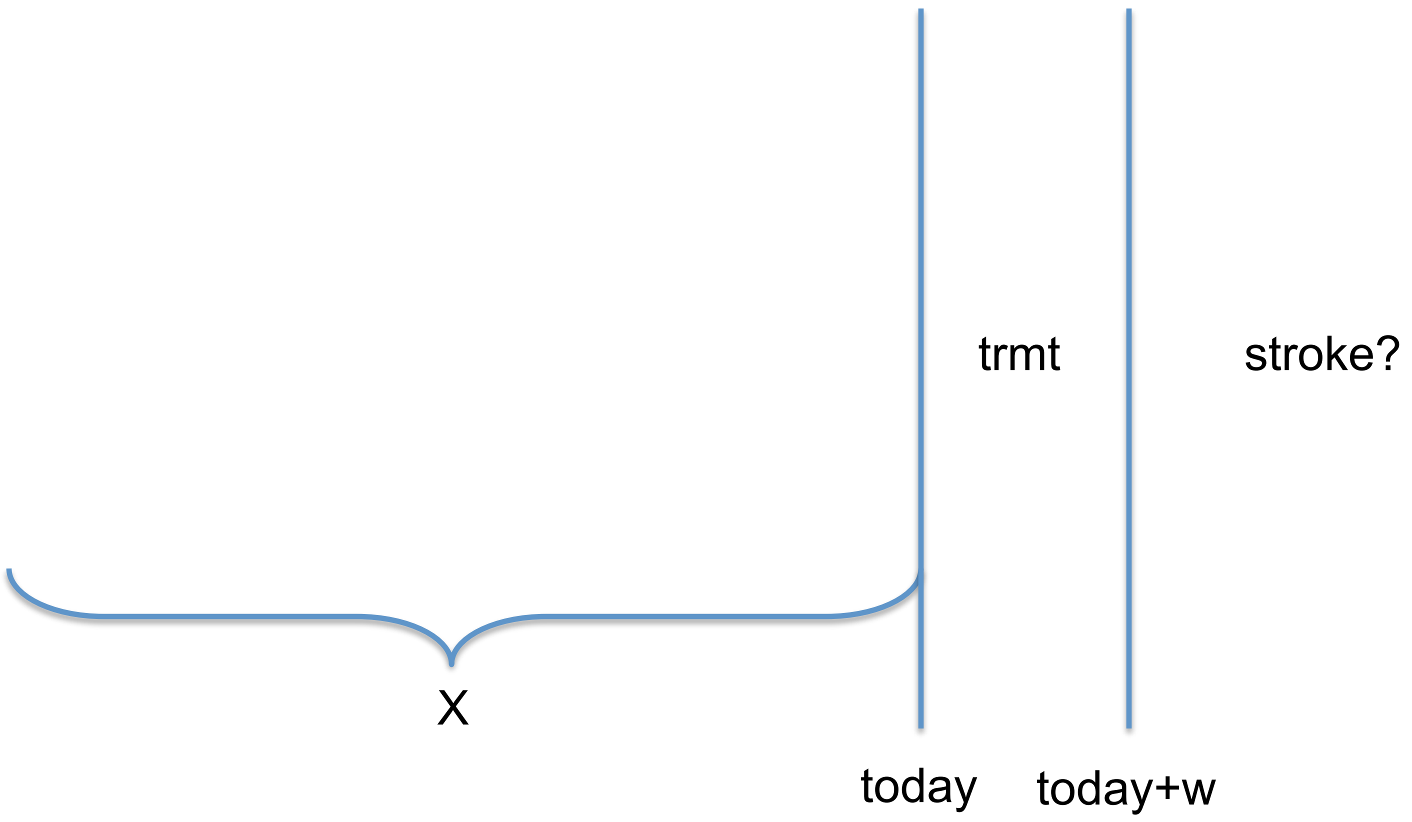
Methodological Challenges



Pr(Stroke | X) = Σ Pr(Stroke | X, t) Pr(X | t)

where the summation is over all possible treatment plans t

Methodological Challenges



$$\Pr(\text{Stroke} \mid X) = \Sigma \Pr(\text{Stroke} \mid X, t) \Pr(X \mid t)$$

where the summation is over all possible treatment plans t

93

Primarily Interested in Pr(Stroke | X, t)

- Pr(Stroke | X, t=1) - Pr(Stroke | X, t=0) is a causal effect

- There is no escape!

- For a given X=x', there is a concern that either X=x', t=1 or X=x', t=0 has poor support; standard error of prediction should account for this

- Bias due to unmeasured confounders is a different matter

Why patient-centered analytics holds promise

Average treatment effects:

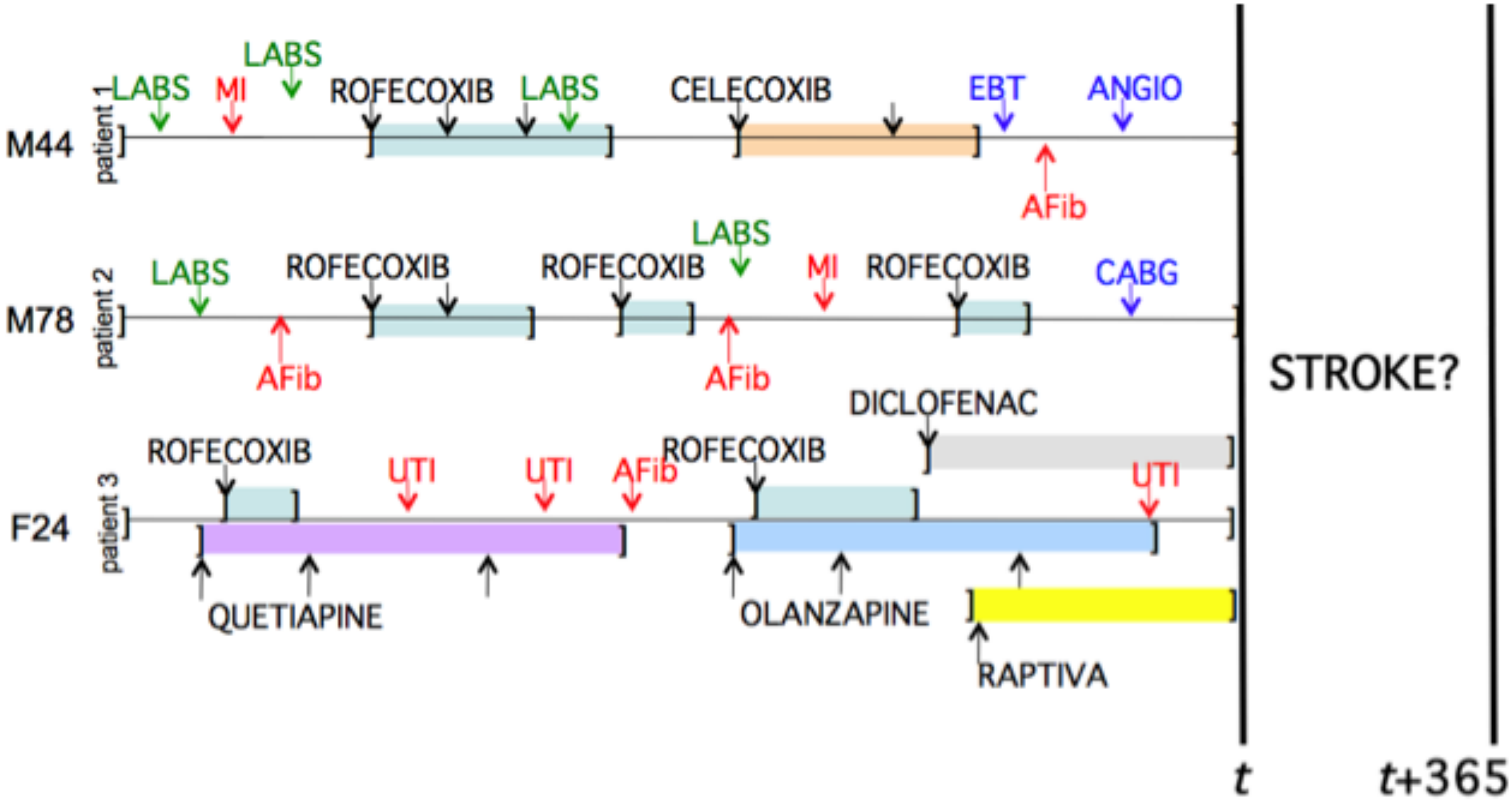
- Hundreds of drug-outcome pairs
- Unsatisfactory ground truth:
  - how confident are we that drug is associated with outcome?
  - What is ‘true’ effect size?
- Questionable generalizability: who does the average treatment effect apply to?
- Final answer often insufficient:
  - Need to drilldown to explore treatment heterogeneity
  - Truth about ‘causality’ is largely unobtainable

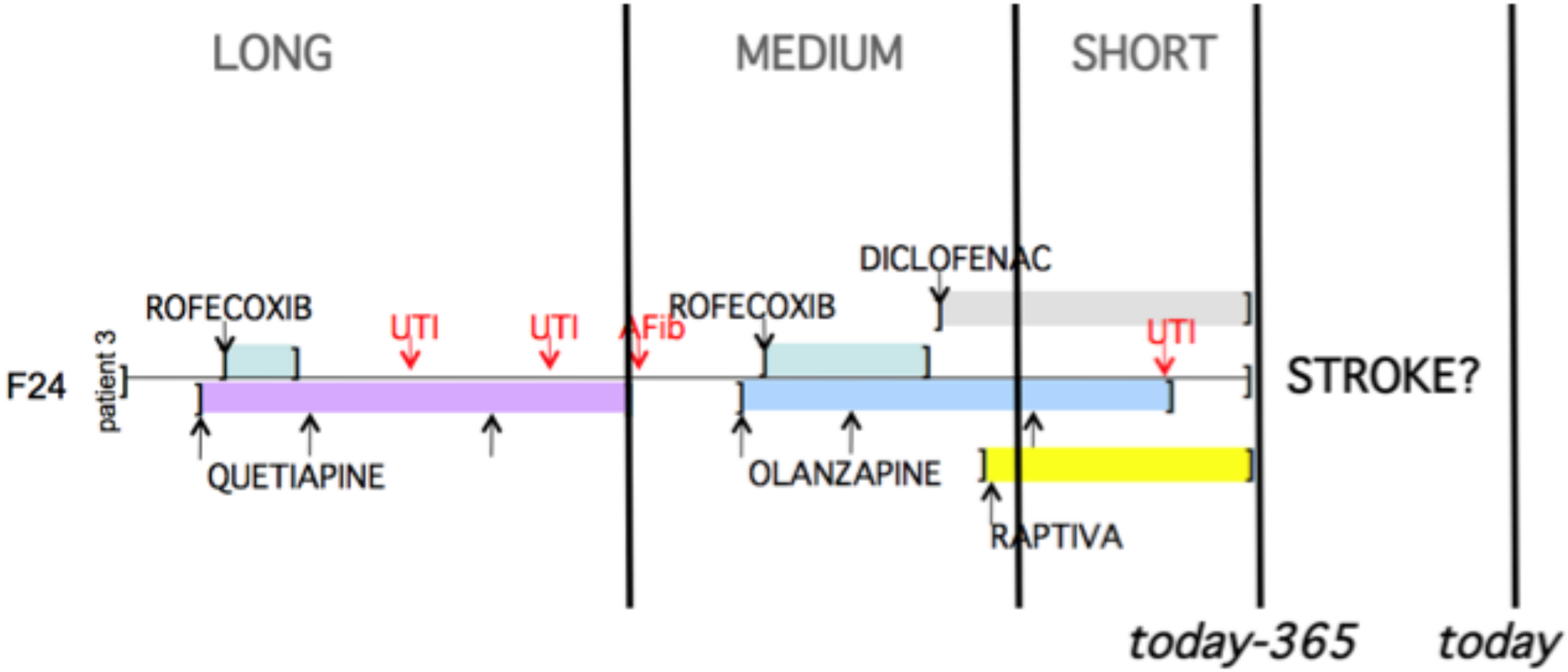
Patient-centered predictions:

- Millions of patients
- Explicit ground truth
  - Each patient did or did not have the outcome within the defined time interval
- Direct applicability: model computes probability for each individual
- Final model can address broader questions:
  - Which patients are most at risk?
  - What factors are most predictive of outcome?
  - How much would change in health behaviors impact risk?
  - What is the average treatment effect?

Concluding thoughts

- Not all patients are created equally…
  - Average treatment effects are commonly estimated from observational databases, but the validity and utility of these estimates remains undetermined
  - Patient-centered predictive modeling offers a complementary perspective for evaluating treatments and understanding disease
- …but all patients can equally benefit from the potential of predictive modeling in observational data
  - Clinical judgment may be useful, but selecting of a handful of predictors is unlikely to maximize the use of the data
  - High-dimensional analytics can enable exploration of high-dimensional data, but further research and evaluation is needed
  - Empirical question still to be answered: Which outcomes can be reliably predicted using which models from which data?


M44
patient 1
LABS
MI
LABS
ROFECOXIB
LABS
CELECOXIB
EBT
ANGIO
AFib
M78
patient 2
LABS
ROFECOXIB
ROFECOXIB
LABS
MI
ROFECOXIB
CABG
AFib
AFib
F24
patient 3
DICLOFENAC
ROFECOXIB
UTI
UTI
AFib
ROFECOXIB
UTI
QUETIAPINE
OLANZAPINE
RAPTIVA
STROKE?
t
t+365


LONG
MEDIUM
SHORT
F24
patient 3
ROFECOXIB
UTI
UTI
AFib
ROFECOXIB
DICLOFENAC
UTI
STROKE?
QUETIAPINE
OLANZAPINE
RAPTIVA
today-365
today