OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

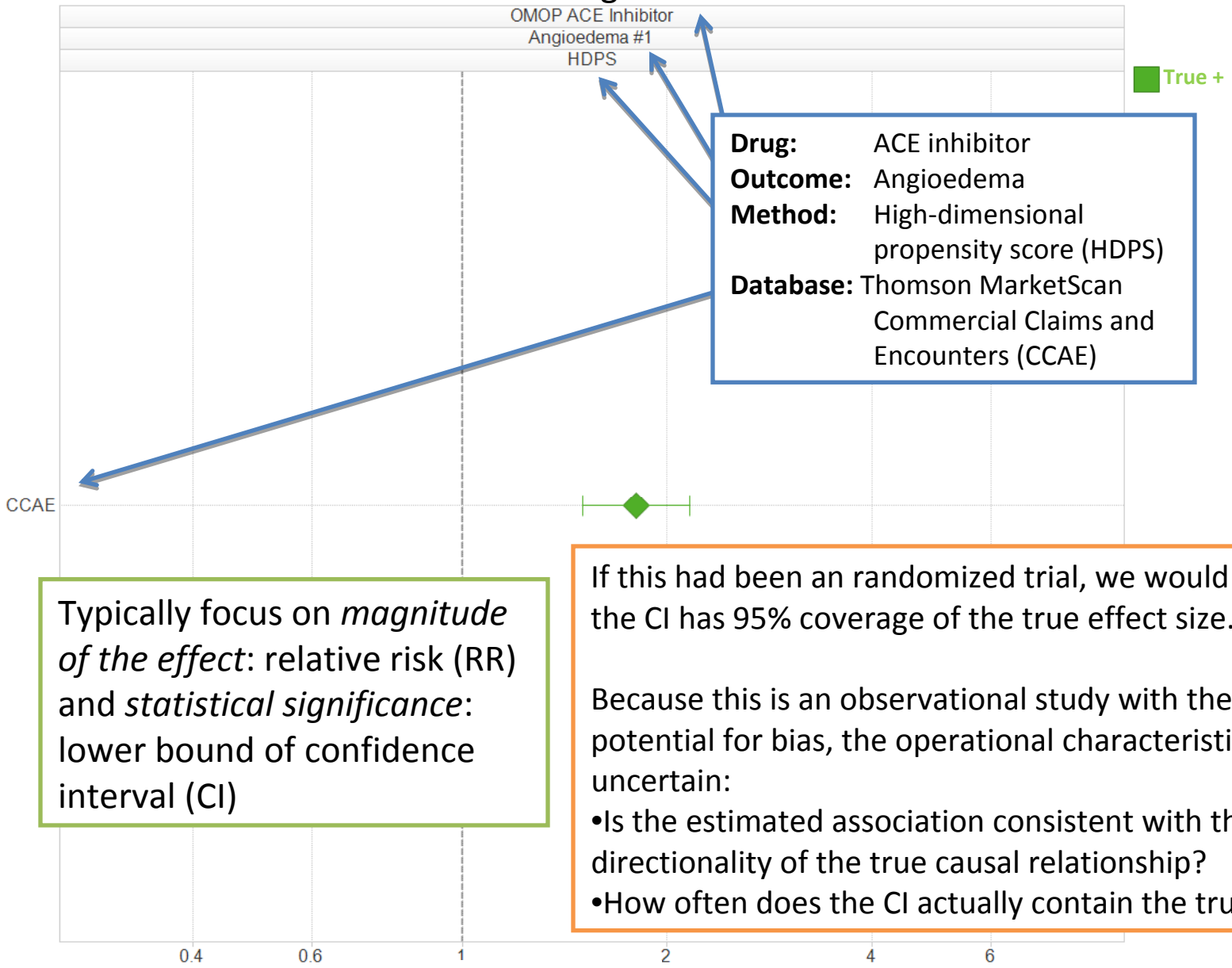
Initial Findings

David Madigan
Columbia University
on behalf of OMOP Research Team

April 8, 2011

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Typical scenario: Estimate the effect of one drug on one outcome using one method against one database



Drug: ACE inhibitor
Outcome: Angioedema
Method: High-dimensional propensity score (HDPS)
Database: Thomson MarketScan Commercial Claims and Encounters (CCAE)

Typically focus on *magnitude of the effect*: relative risk (RR) and *statistical significance*: lower bound of confidence interval (CI)

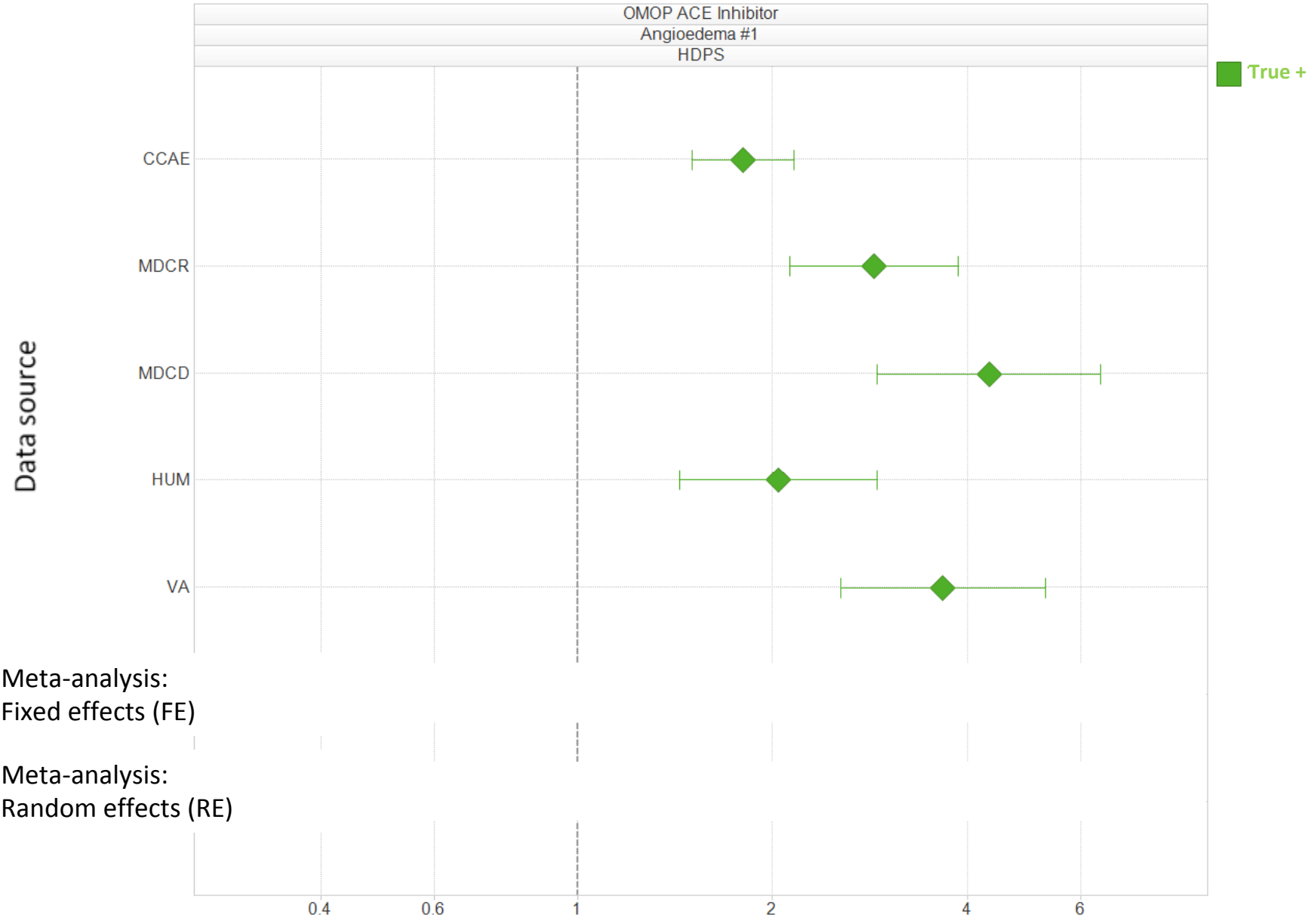
If this had been an randomized trial, we would know the CI has 95% coverage of the true effect size.

Because this is an observational study with the potential for bias, the operational characteristics are uncertain:

- Is the estimated association consistent with the directionality of the true causal relationship?
- How often does the CI actually contain the truth?

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Opportunity for an active surveillance system: Pooling estimates across a network of disparate data sources

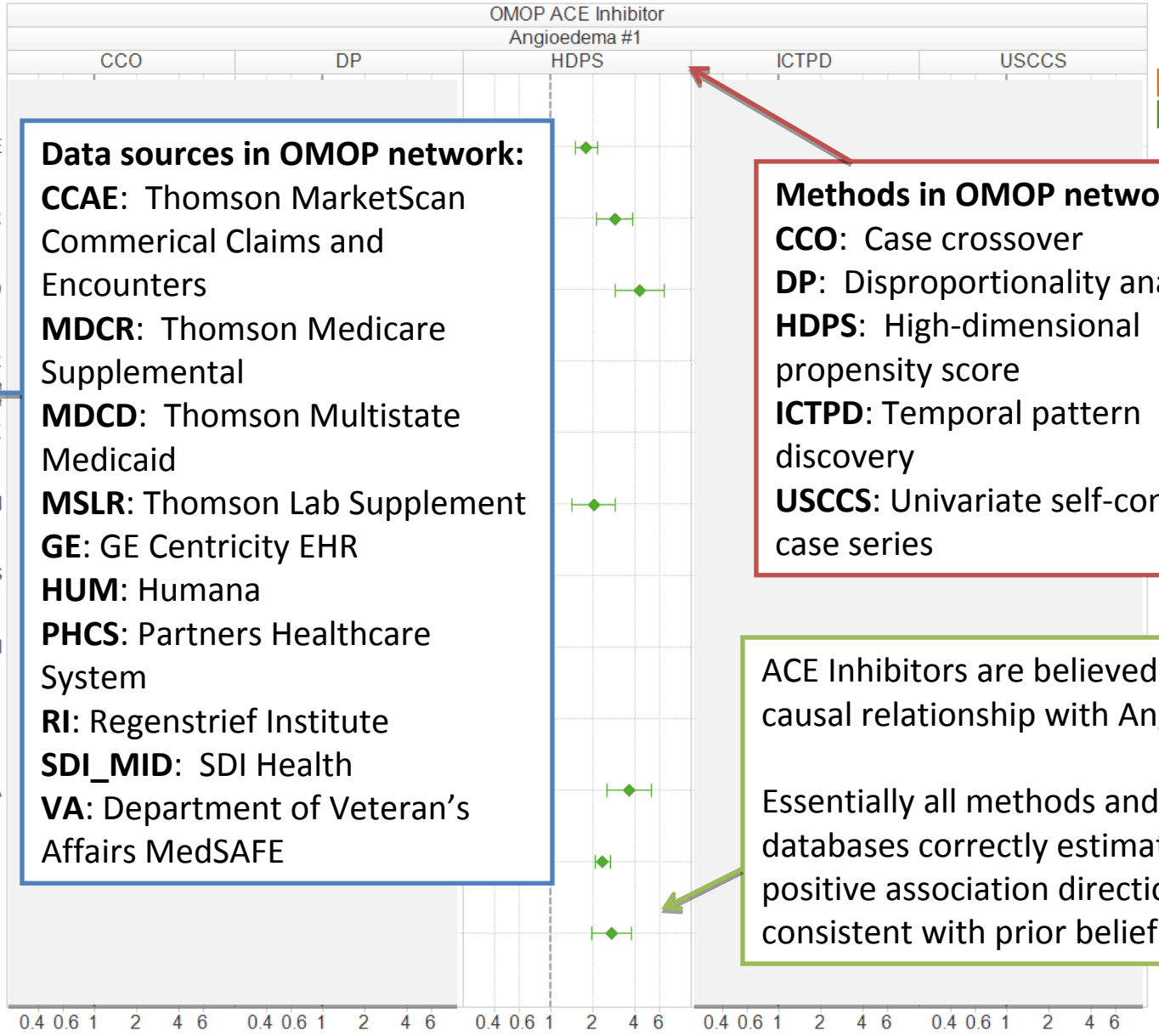


Meta-analysis:
Fixed effects (FE)

Meta-analysis:
Random effects (RE)

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Systematic sensitivity analysis: Estimate the effect using multiple methods across the network of databases



Data sources in OMOP network:

CCAE: Thomson MarketScan Commerical Claims and Encounters

MDCR: Thomson Medicare Supplemental

MDCD: Thomson Multistate Medicaid

MSLR: Thomson Lab Supplement

GE: GE Centricity EHR

HUM: Humana

PHCS: Partners Healthcare System

RI: Regenstrief Institute

SDI_MID: SDI Health

VA: Department of Veteran's Affairs MedSAFE

Methods in OMOP network:

CCO: Case crossover

DP: Disproportionality analysis

HDPS: High-dimensional propensity score

ICTPD: Temporal pattern discovery

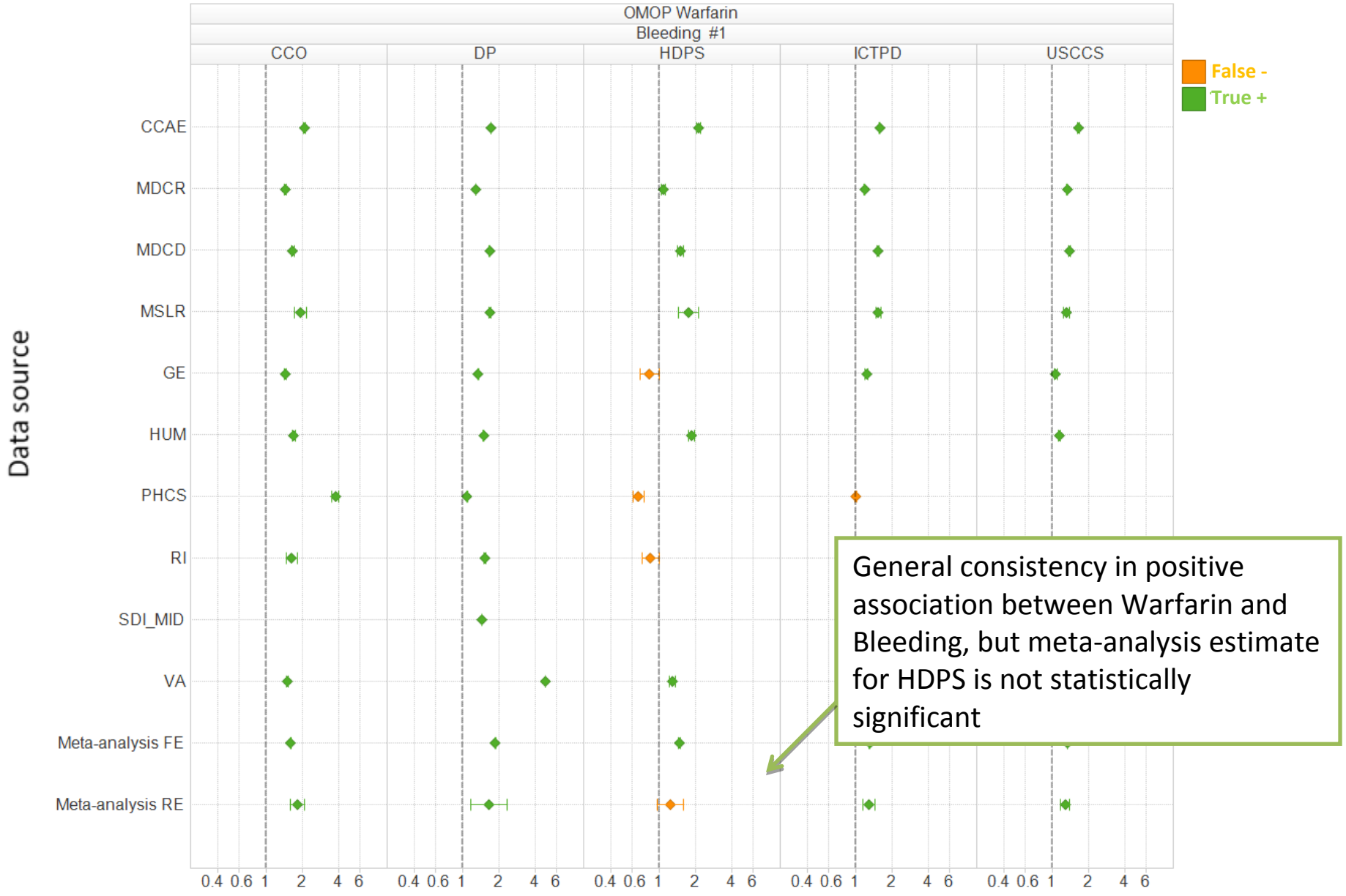
USCCS: Univariate self-controlled case series

ACE Inhibitors are believed to have a causal relationship with Angioedema

Essentially all methods and databases correctly estimate a positive association directionally consistent with prior beliefs

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

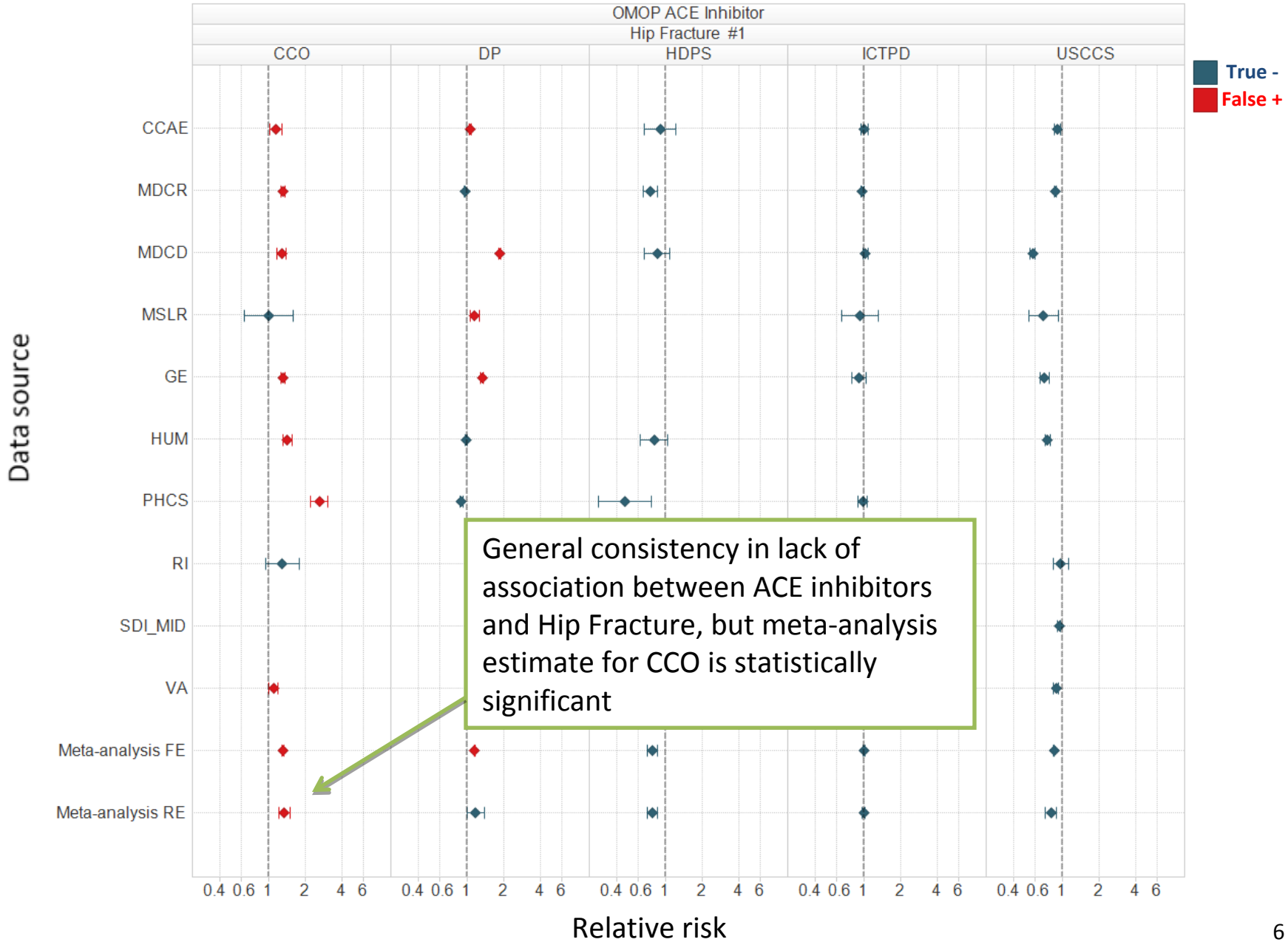
Evaluating the association between Warfarin and Bleeding



General consistency in positive association between Warfarin and Bleeding, but meta-analysis estimate for HDPS is not statistically significant

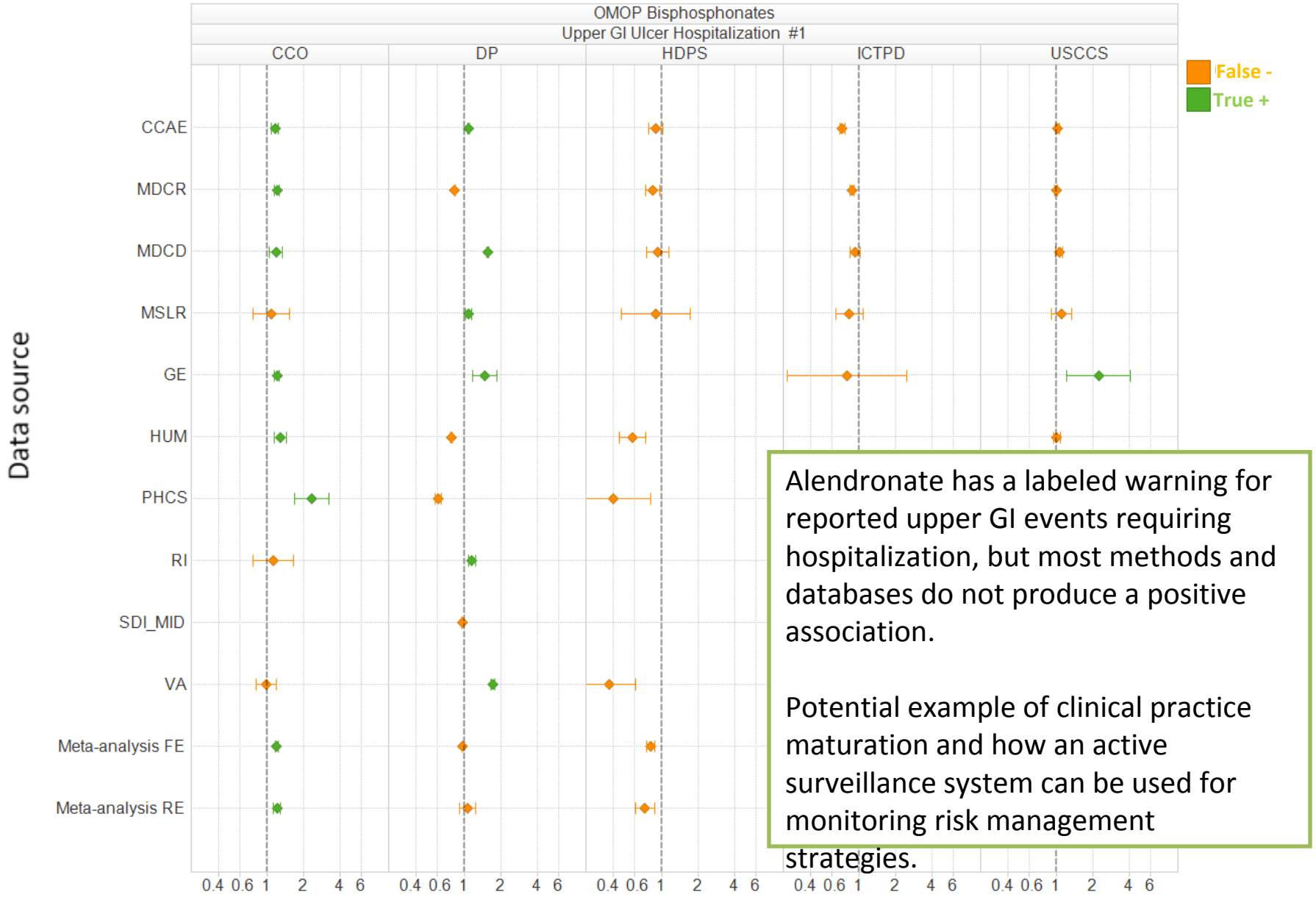
OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

‘True negative’ observed for ‘negative control’ of ACE Inhibitors and Hip Fracture



OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Consistent ‘false negative’ observed for ‘true’ association between Bisphosphonates and GI Ulcer Hospitalization

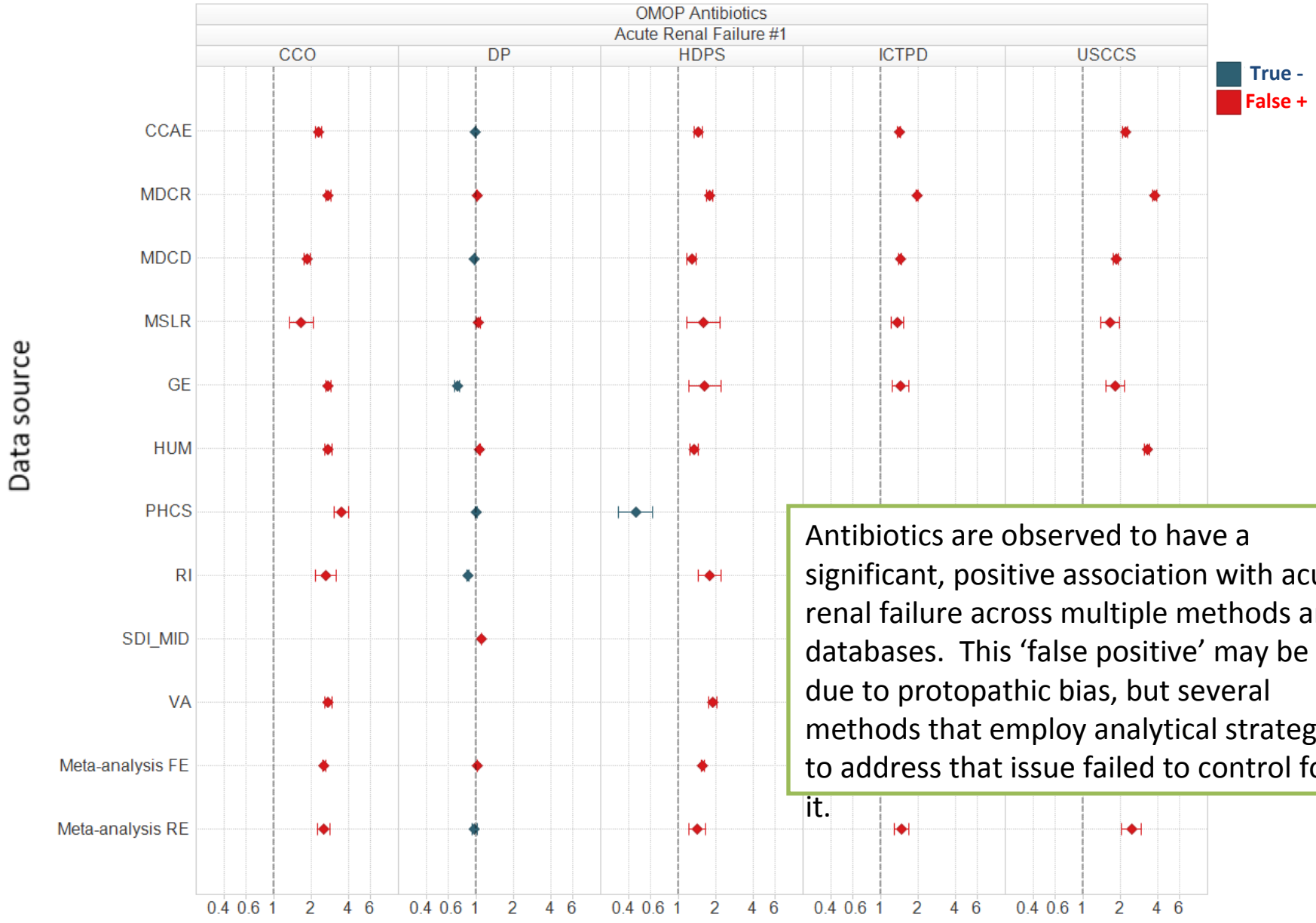


Alendronate has a labeled warning for reported upper GI events requiring hospitalization, but most methods and databases do not produce a positive association.

Potential example of clinical practice maturation and how an active surveillance system can be used for monitoring risk management strategies.

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Consistent ‘false positive’ observed for ‘negative control’ of Antibiotics and Acute Renal Failure



Antibiotics are observed to have a significant, positive association with acute renal failure across multiple methods and databases. This ‘false positive’ may be due to protopathic bias, but several methods that employ analytical strategies to address that issue failed to control for it.

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Measuring method performance: Classifying the illustrative examples

Drug-condition association status
Y – 'true association',
N – 'negative control'

Method prediction: Drug-condition pair met a specific threshold: (LB 95% CI > 1)	Drug-condition association status: Y	Drug-condition association status: N
Y	**True positives:** **ACE Inhibitors- Angioedema** **Warfarin- Bleeding**	**False positives:** **Typical antipsychotic- GI Ulcer hospitalization** **Antibiotics- Acute renal failure**
N	**False negatives:** **Bisphosphonates- GI Ulcer hospitalization**	**True negatives:** **ACE Inhibitors- Hip fracture**

Measuring method performance

Drug-condition association status
Y – 'true association',
N – 'negative control'

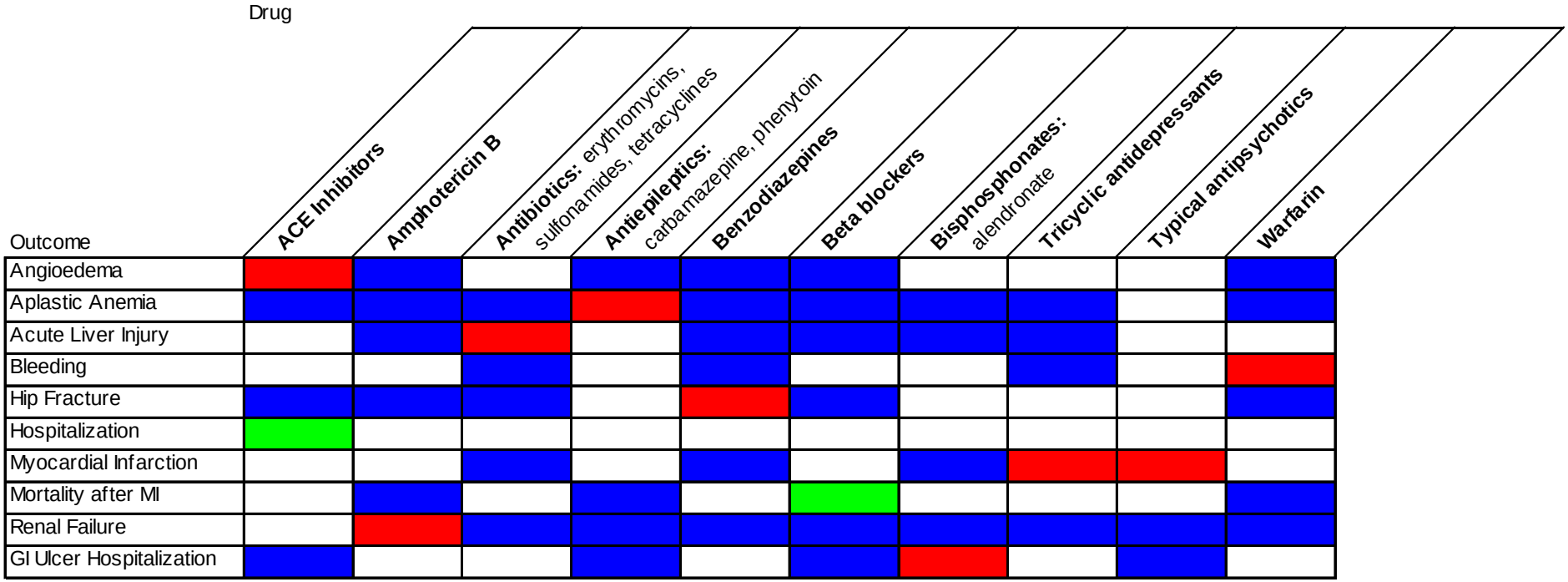
Method prediction: Drug-condition pair met a specific threshold

	Y	N
Y	**True positives**	**False positives**
N	**False negatives**	**True negatives**

Question: For any method applied to any data source, what are the expected operating characteristics?

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

‘Ground truth’ for Monitoring Health Outcomes of Interest



Legend	Total
True positive' benefit	2
True positive' risk	9
Negative control'	44

http://omop.fnih.org/OMOPWhitePapers

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Measuring method performance example: Random-effect meta-analysis of estimates from High-dimensional propensity score

Drug-condition association status
Y – 'true association',
N – 'negative control'

Method prediction: Drug-condition pair met a specific threshold:

(LB 95% CI > 1)

	Y	N
Y	True positives: 5	False positives: 8
N	False negatives: 4	True negatives: 36

Positive predictive value
= precision
= TP / (TP+FP)
= 5 / (5+8) = 0.38

Negative predictive value
= TN / (FN+TN)
= 36 / (4+36) = 0.90

Sensitivity
= Recall
= TP / (TP+FN)
= 5 / (5+4) = 0.56

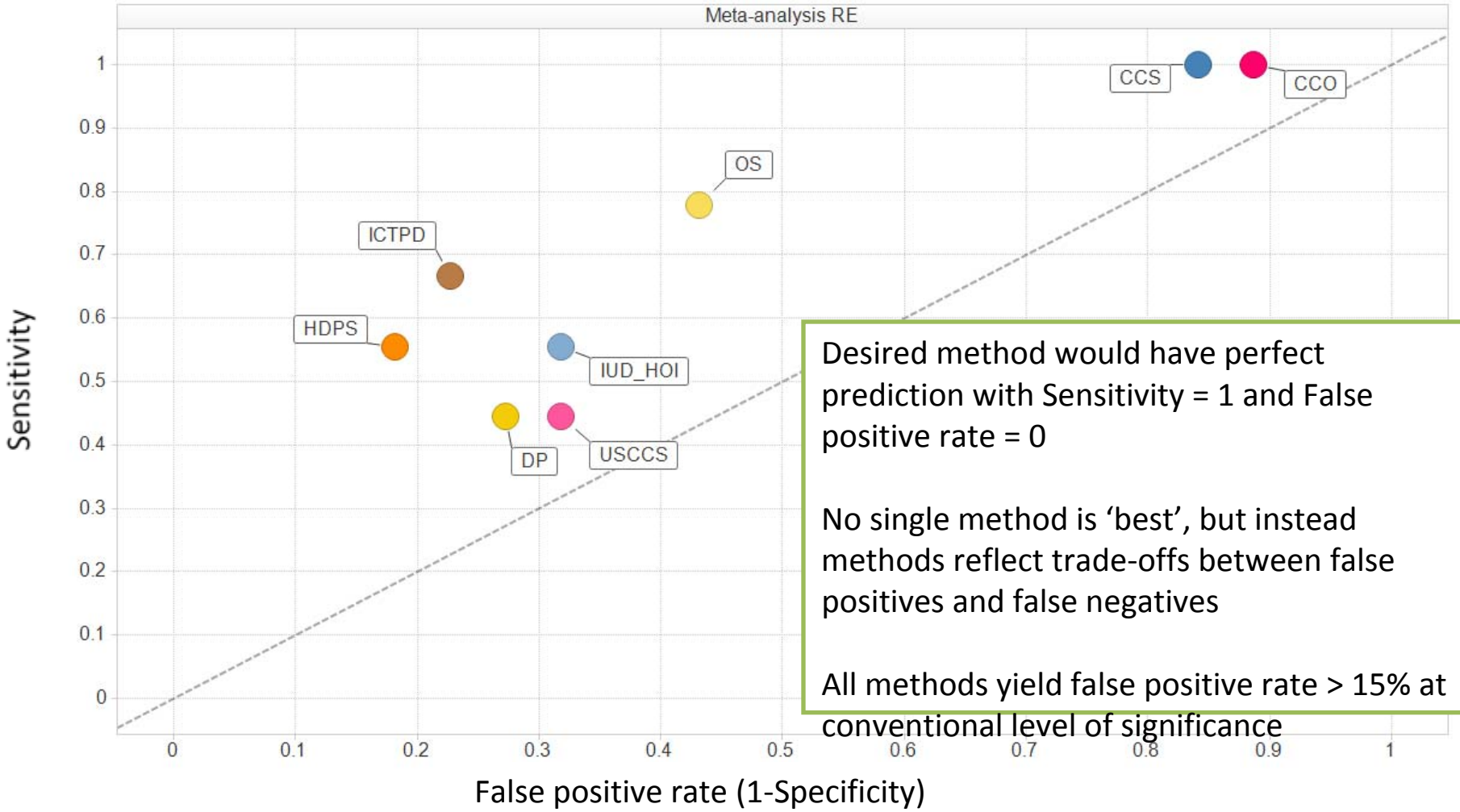
Specificity
= TN / (FP+TN)
= 36 /(8+36) = 0.82

False positive rate
= 1 – 0.82 = 0.18

Accuracy
= (TP+TN) /
(TP+TN+FP+FN)
=(5+36)/(9+44) = 0.77

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Comparing methods by sensitivity and specificity at alpha=0.05



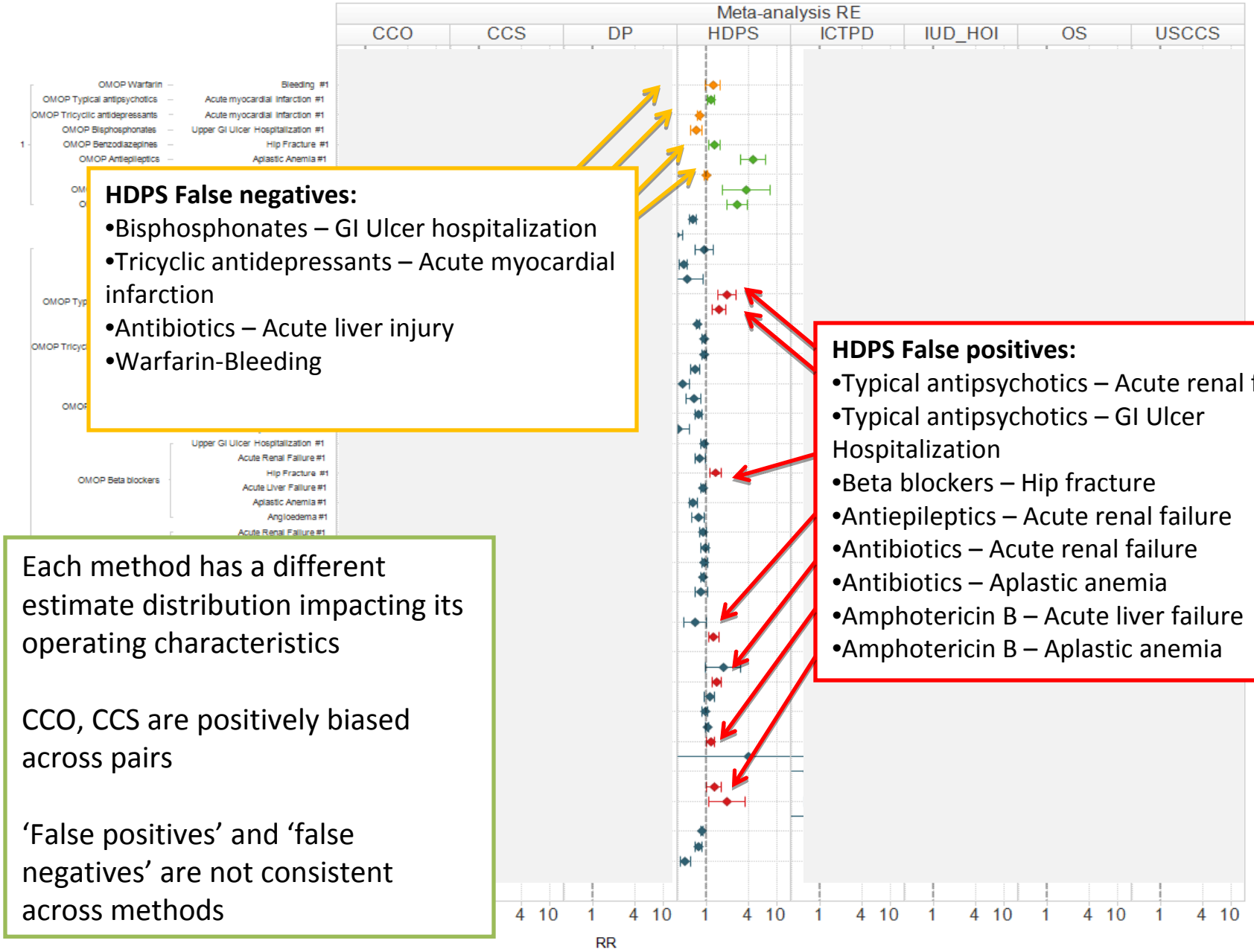
Desired method would have perfect prediction with Sensitivity = 1 and False positive rate = 0

No single method is 'best', but instead methods reflect trade-offs between false positives and false negatives

All methods yield false positive rate > 15% at conventional level of significance

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Distribution of estimates across all drug-outcome pairs



HDPS False negatives:

- Bisphosphonates – GI Ulcer hospitalization
- Tricyclic antidepressants – Acute myocardial infarction
- Antibiotics – Acute liver injury
- Warfarin-Bleeding

HDPS False positives:

- Typical antipsychotics – Acute renal failure
- Typical antipsychotics – GI Ulcer Hospitalization
- Beta blockers – Hip fracture
- Antiepileptics – Acute renal failure
- Antibiotics – Acute renal failure
- Antibiotics – Aplastic anemia
- Amphotericin B – Acute liver failure
- Amphotericin B – Aplastic anemia

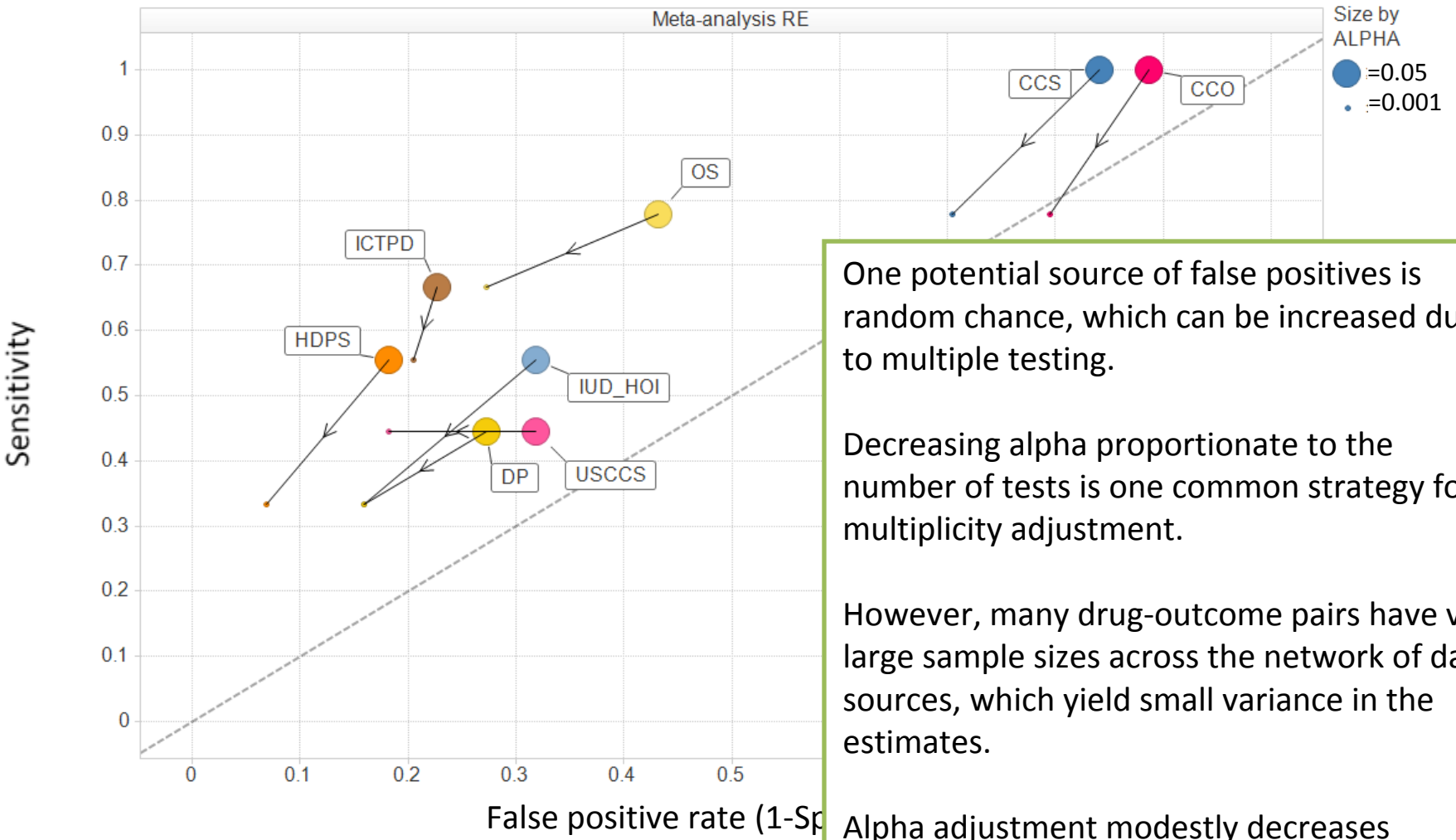
Each method has a different estimate distribution impacting its operating characteristics

CCO, CCS are positively biased across pairs

'False positives' and 'false negatives' are not consistent across methods

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Change in sensitivity and specificity when decreasing alpha=0.001 (adjustment for multiplicity)



One potential source of false positives is random chance, which can be increased due to multiple testing.

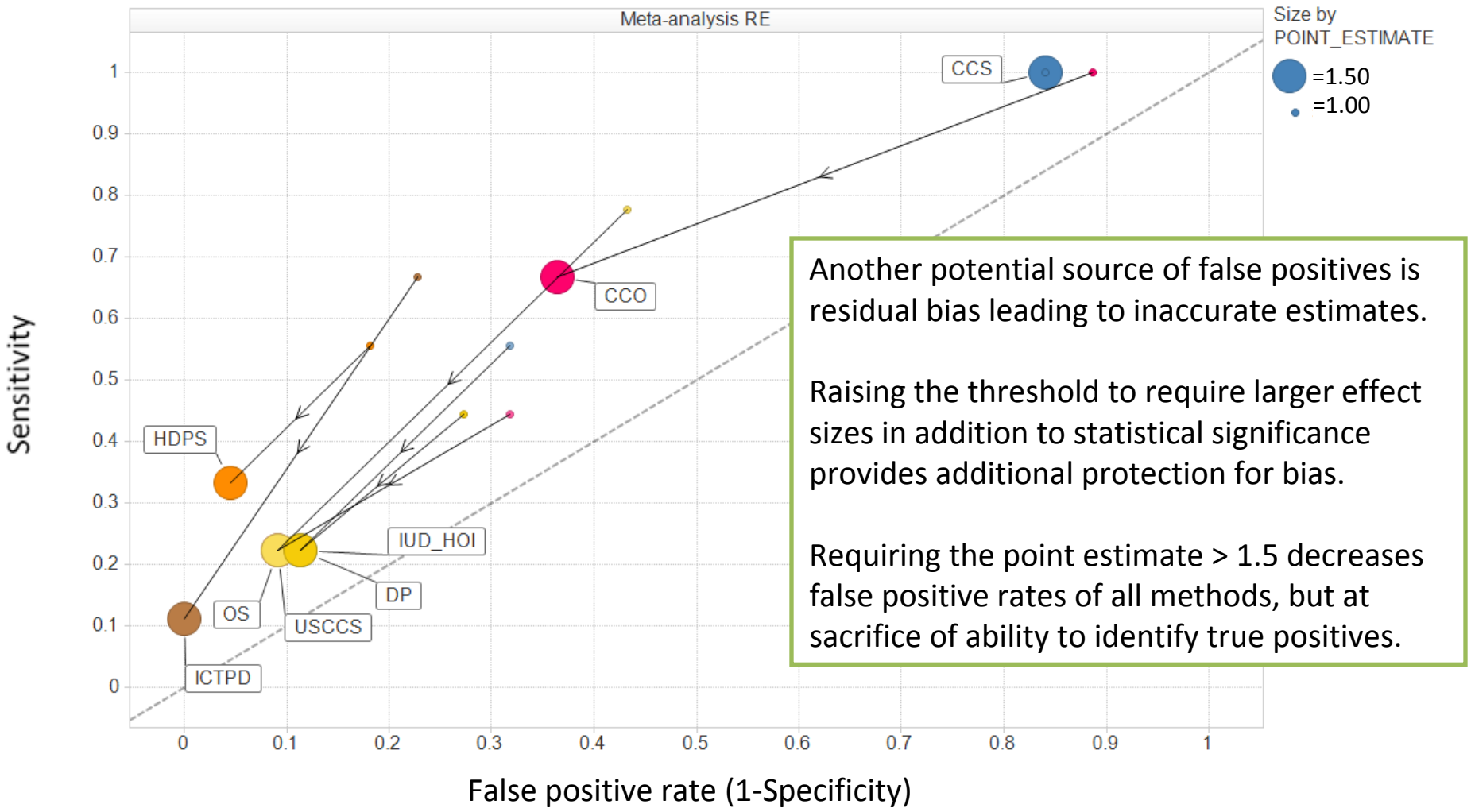
Decreasing alpha proportionate to the number of tests is one common strategy for multiplicity adjustment.

However, many drug-outcome pairs have very large sample sizes across the network of data sources, which yield small variance in the estimates.

Alpha adjustment modestly decreases sensitivity and false positive rate for most methods.

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Change in sensitivity and specificity when raising threshold to require point estimate >=1.5 (adjustment for bias)



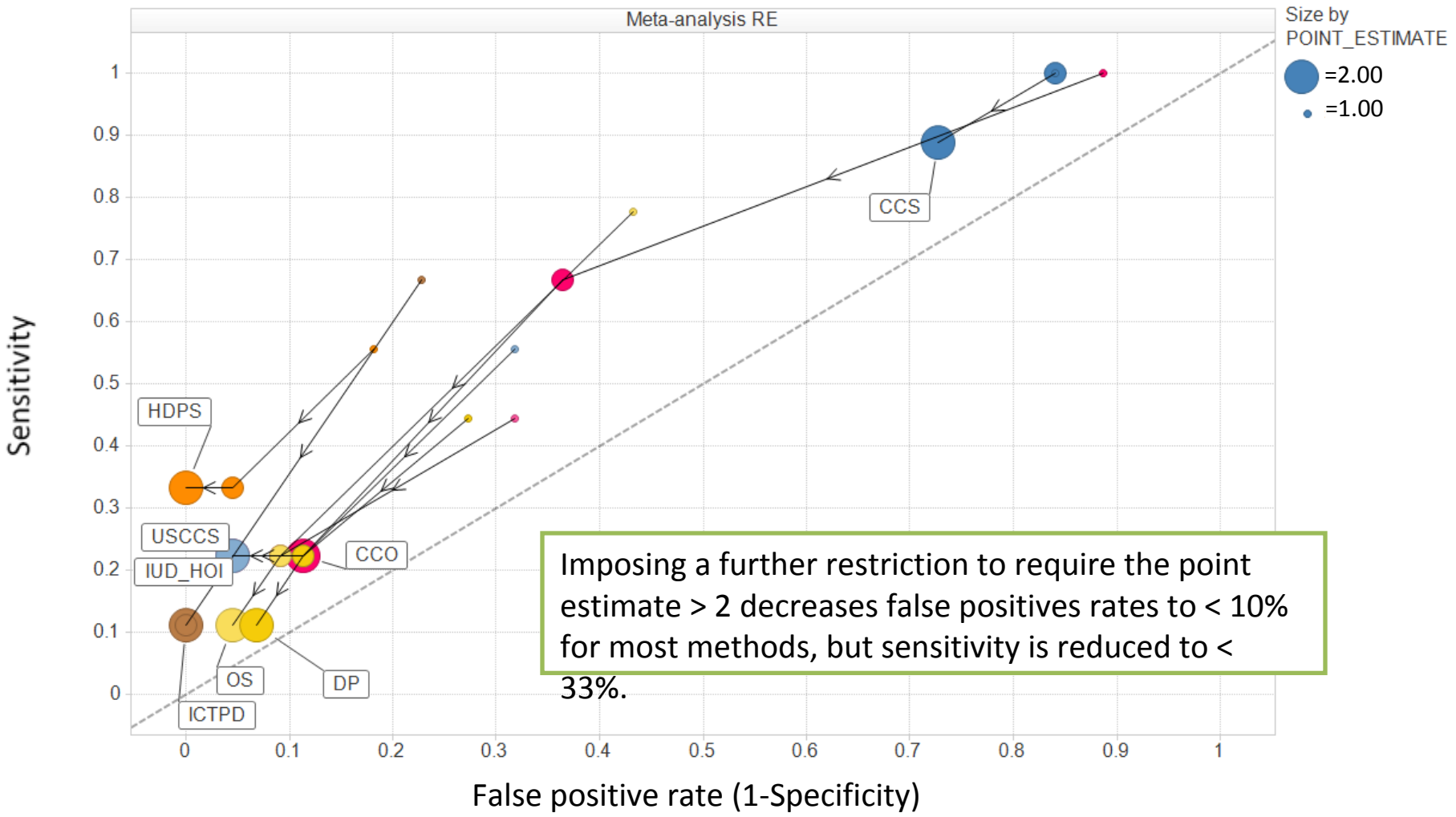
Another potential source of false positives is residual bias leading to inaccurate estimates.

Raising the threshold to require larger effect sizes in addition to statistical significance provides additional protection for bias.

Requiring the point estimate > 1.5 decreases false positive rates of all methods, but at sacrifice of ability to identify true positives.

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

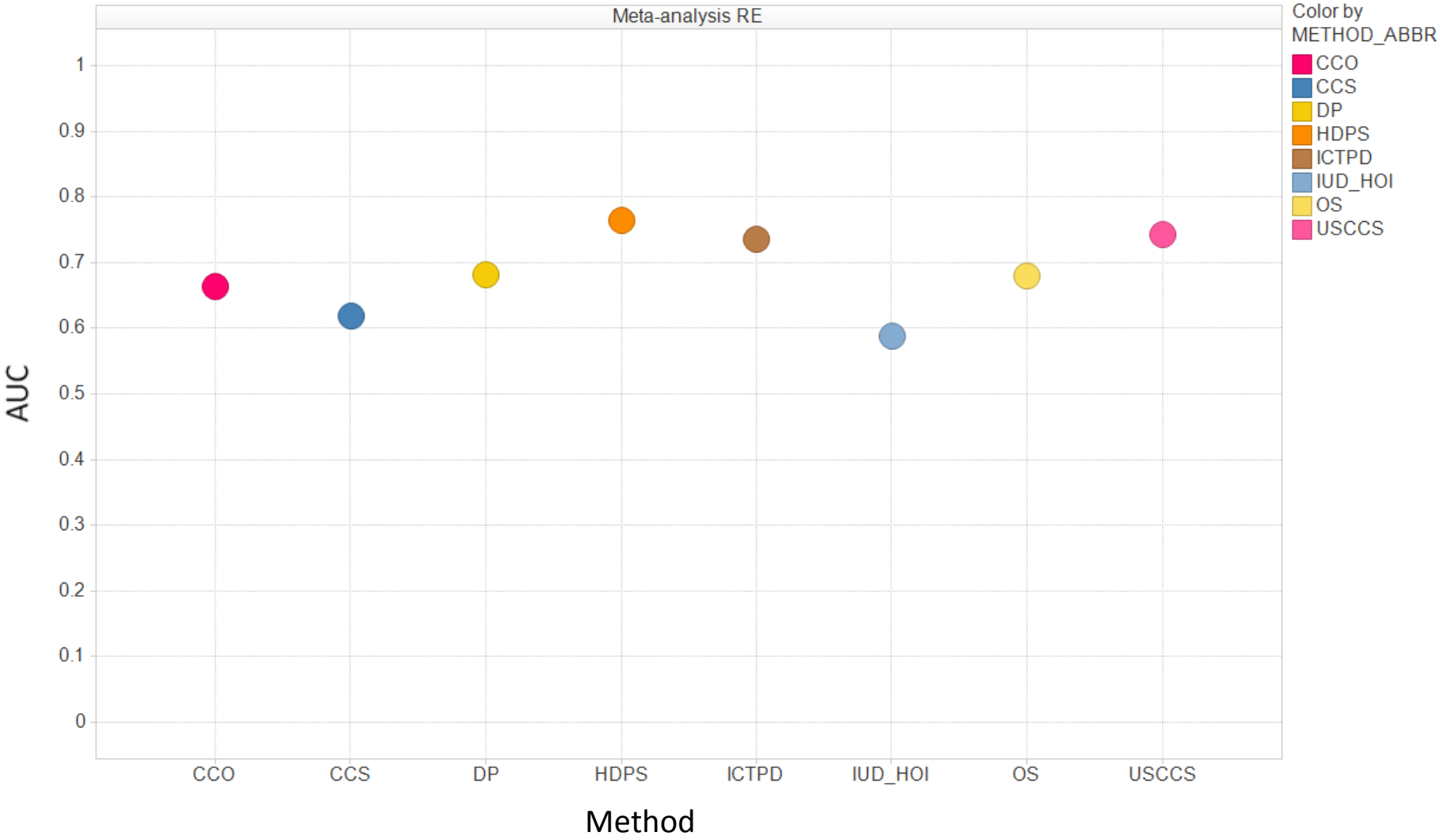
Change in sensitivity and specificity when raising threshold to require point estimate >=2.0 (adjustment for bias)



Imposing a further restriction to require the point estimate > 2 decreases false positives rates to < 10% for most methods, but sensitivity is reduced to < 33%.

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

AUC performance by method



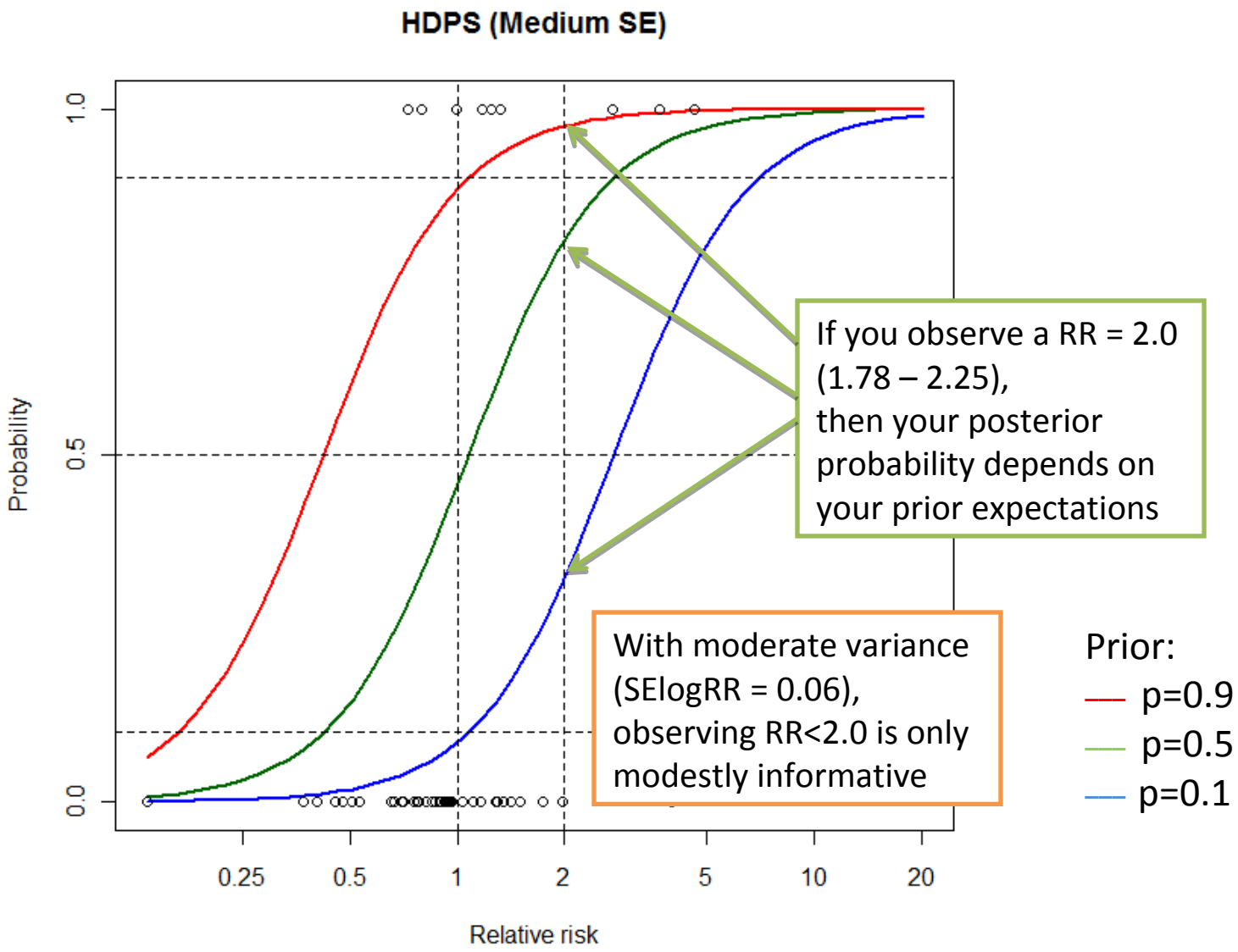
So given these operating characteristics, what can we expect to do in practice….

- Use case: An emerging safety concern is raised for a new medical product. The association between the drug and outcome could be estimated by running an OMOP active surveillance method across the network of observational databases
  - The method will produce a relative risk and standard error from each participating data source, which can then be pooled together in a meta-analytic framework
  - Hypothetical scenario: The random-effects meta-analysis yields an RR=2.0 with SE=0.06.
  - Question: what is the probability that there is a true causal relationship given this observed association?

- Bayes rule enables such calculation…
  - p(true|RR,SE) ~ p(RR,SE|true)*p(true)
  - p(true) is the prior probability of true association; consider a family of priors: skeptical (0.1), indifferent (0.5), enthusiastic (0.9)
  - p(RR,SE|true) can be estimated from empirical data (OMOP experimental results)

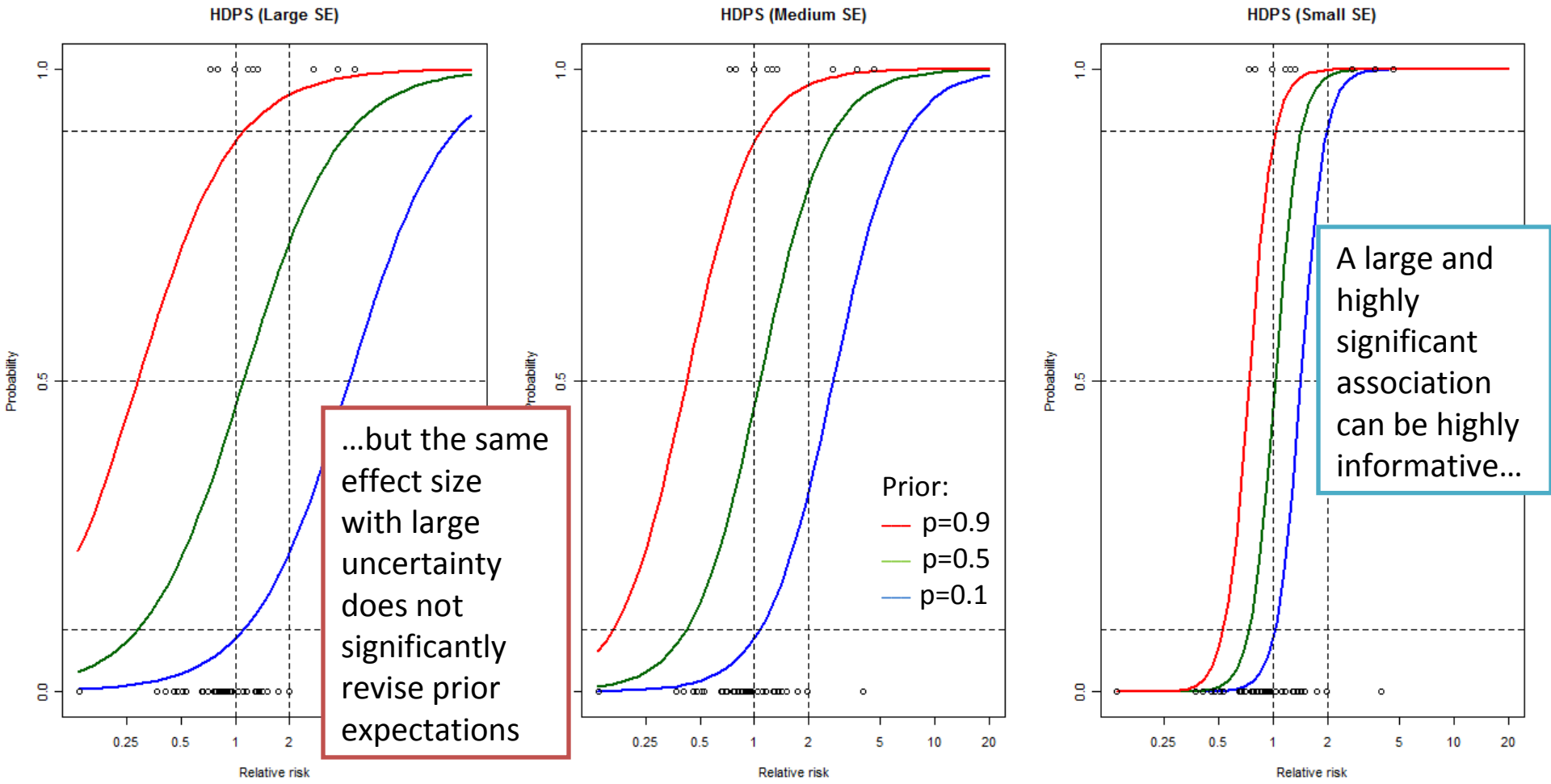
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Revising prior expectations in light of new evidence from an active surveillance system



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Revising prior expectations in light of new evidence from an active surveillance system: Impact of precision of observed estimates



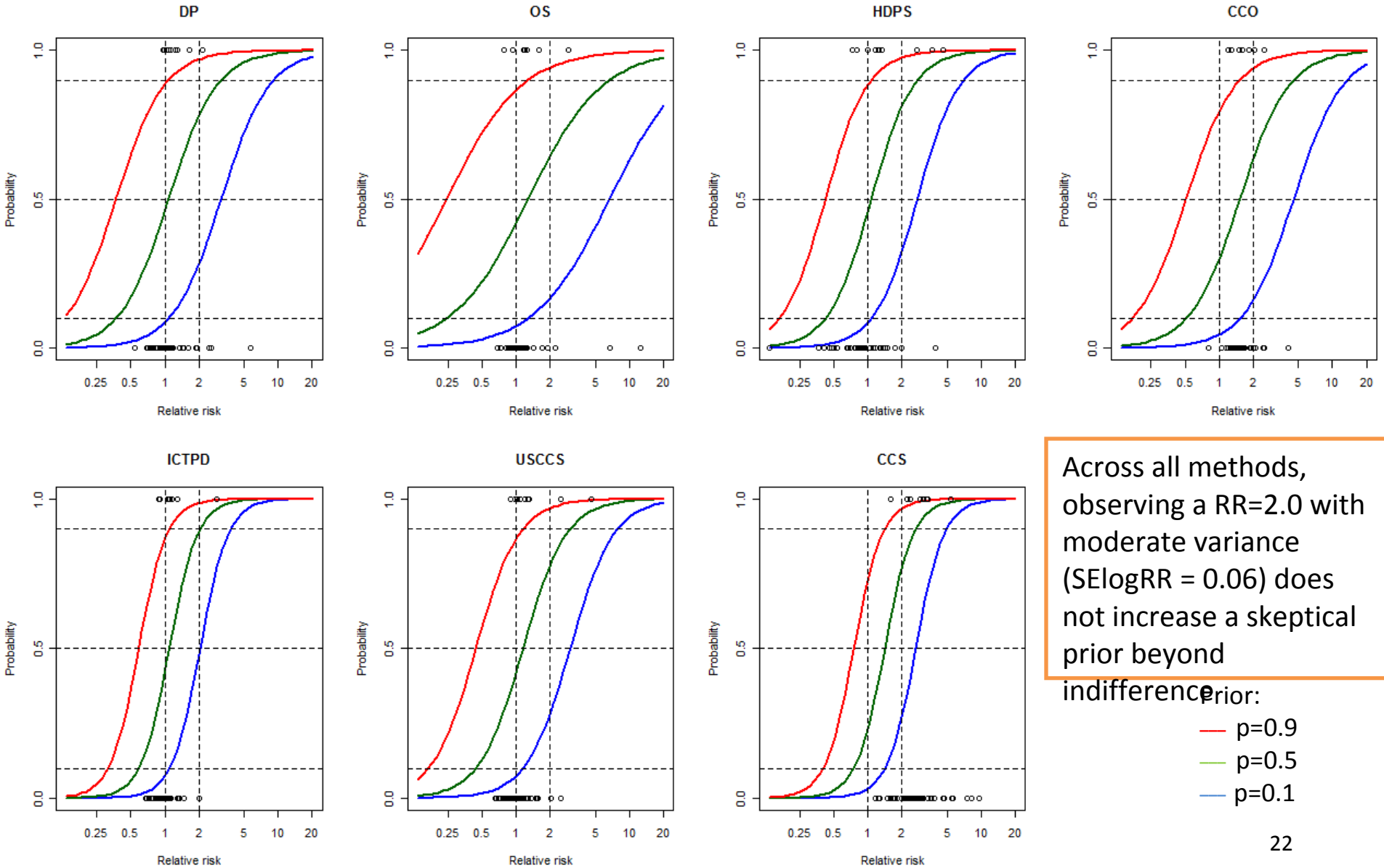
…but the same effect size with large uncertainty does not significantly revise prior expectations

A large and highly significant association can be highly informative…

Scenarios: You observe RR=2.0 with confidence intervals based on standard error (SE):
Large SE: (1.01 – 3.97) Medium SE: (1.78 – 2.25) Small SE: (1.96 – 2.04)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Revising prior expectations in light of new evidence from an active surveillance system: Impact of using estimates from different methods



Across all methods, observing a RR=2.0 with moderate variance (SElogRR = 0.06) does not increase a skeptical prior beyond indifference

Prior:
- p=0.9
- p=0.5
- p=0.1

Concluding thoughts

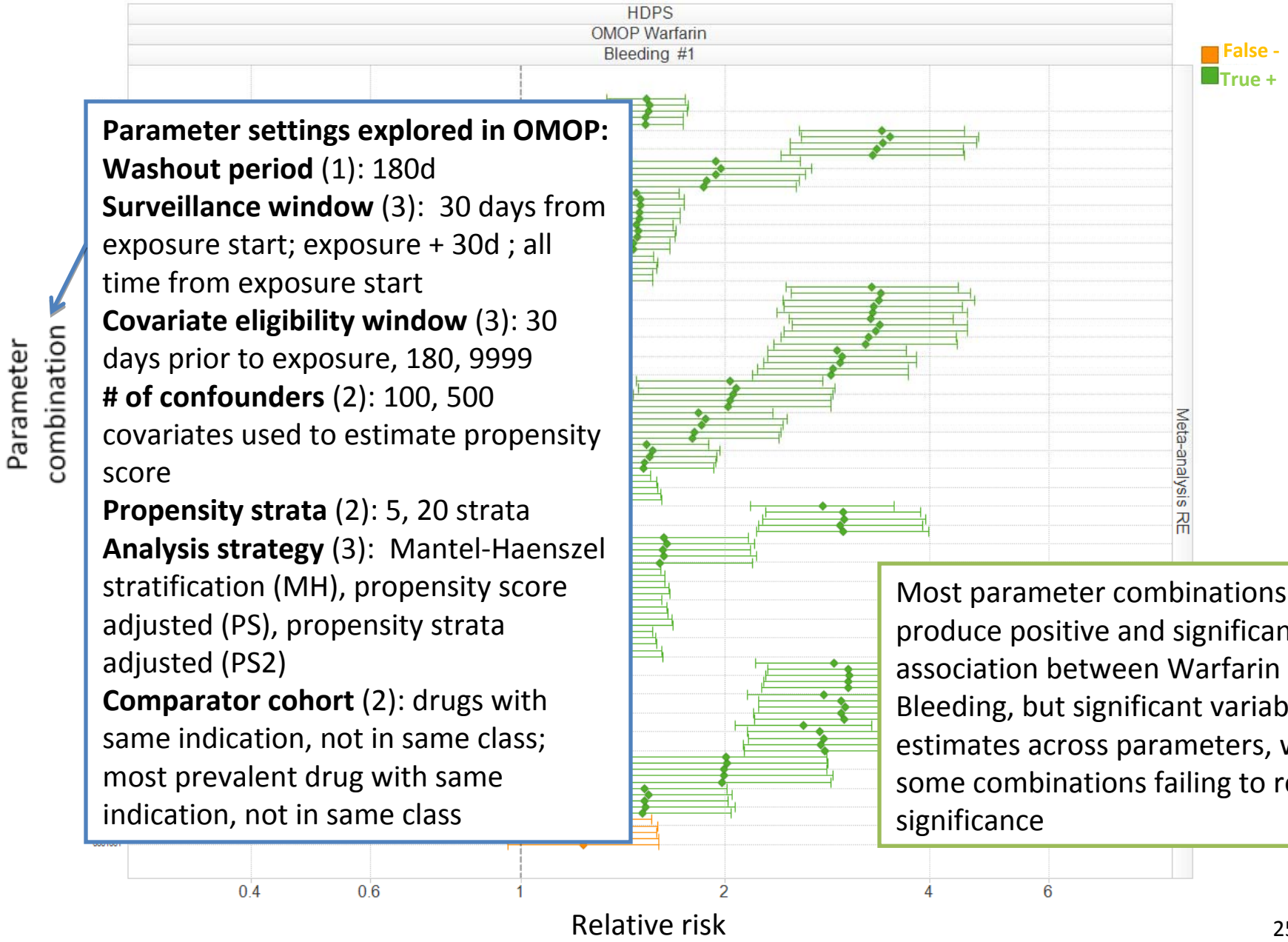
- No one clear 'best' method, as it depends on tolerance for false positives vs. false negatives
- Systematic pharmacoepidemiology can achieve:
  - At 50% sensitivity, false positive rate ranges 16%-30%
  - At 10% false positive rate, sensitivity ranges 9%-33%
- Need to be cautious in interpreting results from single method in single database
  - Replication does not necessarily provide complete confidence
- You need a relative risk > 2 to have confidence in result ….detecting effects smaller than 2 will incur higher cost of false positives

Acknowledgements

- Paul Stang
- Judy Racoosin
- Marc Overhage
- Bram Hartzema
- Christian Reich
- Emily Welebob
- Tom Scarnecchia
- Marc Suchard
- Jesse Berlin
- Ron Krall

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Evaluating the sensitivity of the estimated association between Warfarin and Bleeding when using HDPS



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, 9999

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

Propensity strata (2): 5, 20 strata

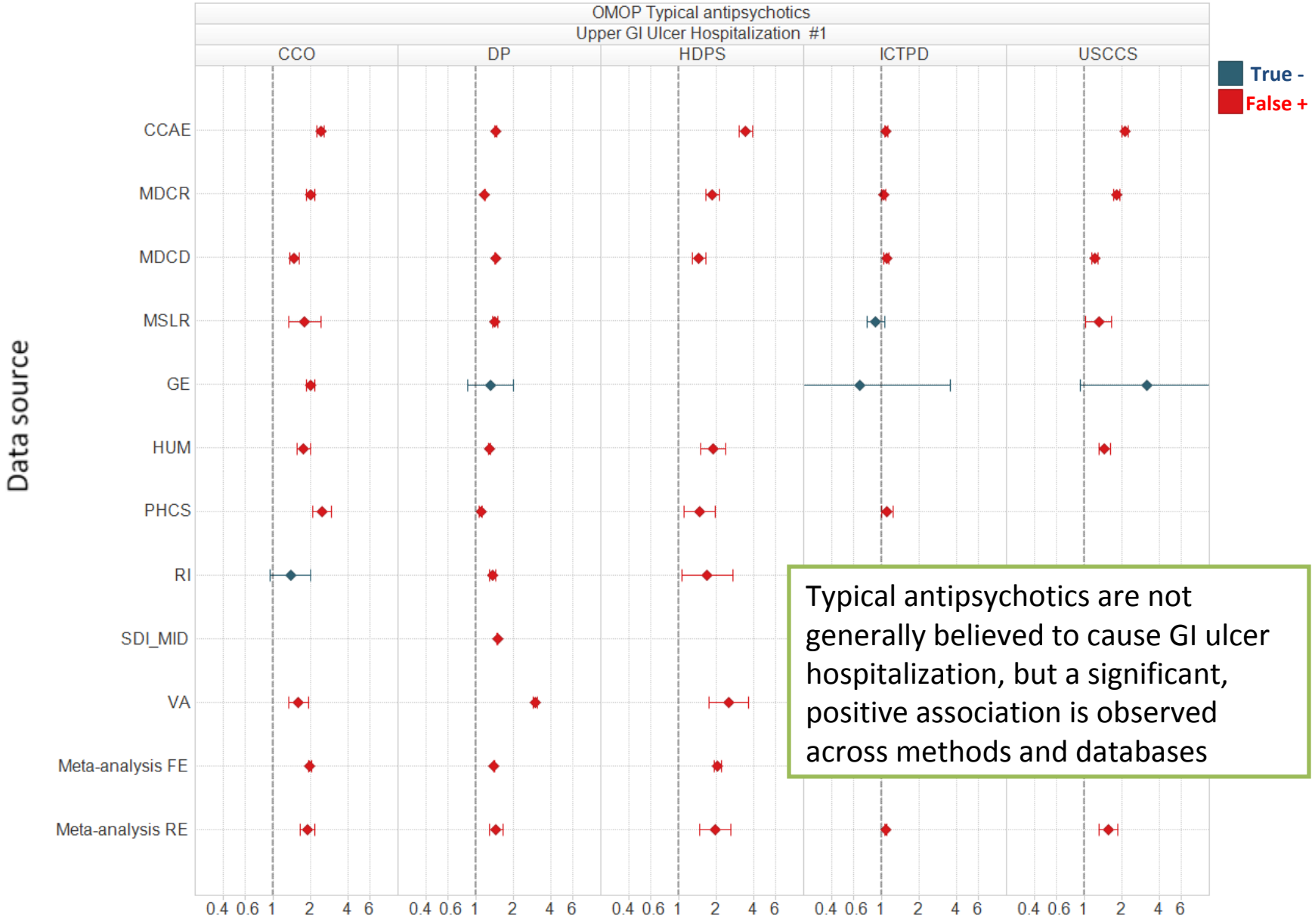
Analysis strategy (3): Mantel-Haenszel stratification (MH), propensity score adjusted (PS), propensity strata adjusted (PS2)

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

Most parameter combinations produce positive and significant association between Warfarin and Bleeding, but significant variability in estimates across parameters, with some combinations failing to reach significance

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Consistent 'false positive' observed for 'negative control' of Typical antipsychotics and GI Ulcer Hospitalization



Typical antipsychotics are not generally believed to cause GI ulcer hospitalization, but a significant, positive association is observed across methods and databases

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Active surveillance methods under evaluation in OMOP experiment

Method name	Contributor	Release date
Disproportionality analysis		
Disproportionality analysis (DP)	Columbia / Merck	15-Mar-10
IC Temporal Pattern Discovery (ICTPD)	Uppsala Monitoring Centre	23-May-10
HSIU cohort method (HSIU)	Regenstrief / Indiana University	8-Jun-10
Case-based methods		
Univariate self-controlled case series (USCCS)	Columbia	2-Apr-10
Multi-set case control estimation (MSCCE)	Columbia / GlaxoSmithKline	16-Apr-10
Bayesian logistic regression (BLR)	Rutgers / Columbia	21-Apr-10
Case-control surveillance (CCS)	Lilly	2-May-10
Case-crossover (CCO)	University of Utah	1-Jun-10
Exposure-based methods		
Observational screening (OS)	ProSanos / GlaxoSmithKline	8-Apr-10
High-dimensional propensity score (HDPS)	Harvard Medical School / Columbia	6-Aug-10
Incident user design (IUD-HOI)	University of North Carolina	26-Oct-10
Sequential testing methods		
Maximized Sequential Probability Ratio Test (MSPRT)	Harvard Pilgrim / Group Health	25-Jul-10
Conditional sequential sampling procedure (CSSP)	Harvard Pilgrim / Group Health	30-Aug-10

http://omop.fnih.org/MethodsLibrary

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Methodological considerations common across multiple approaches

- Exposure definition
  - Incident vs. prevalent exposure
  - Source of data capture
- Outcome definition
  - Incident vs. prevalent events
  - Diagnosis codes vs. HOI
- Defining temporal relationship
  - Time from exposure start
  - Time after exposure end
- Comparator selection
- Inclusion/exclusion criteria
  - Baseline history
  - Follow-up time
- Covariate selection and adjustment
  - Matching
  - Stratification
  - Multivariate modeling
- Output metric/statistic
  - Estimation vs. testing
  - Relative vs. attributable risk
  - Measure of uncertainty

In what follows, we have chosen one parameter combination for each method that performs best for the meta-analysis estimates