

**OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP**

Initial Findings

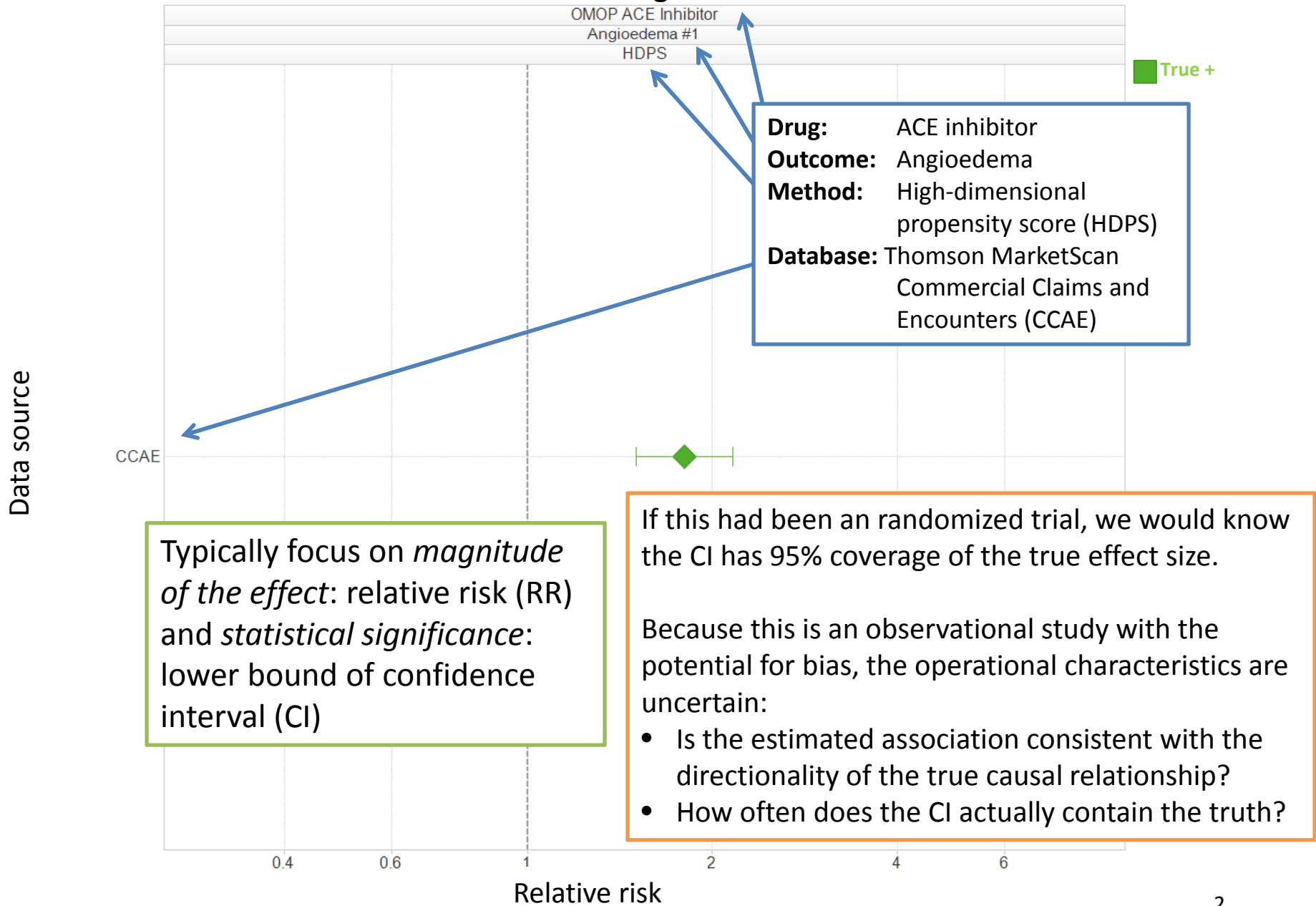
David Madigan

Columbia University

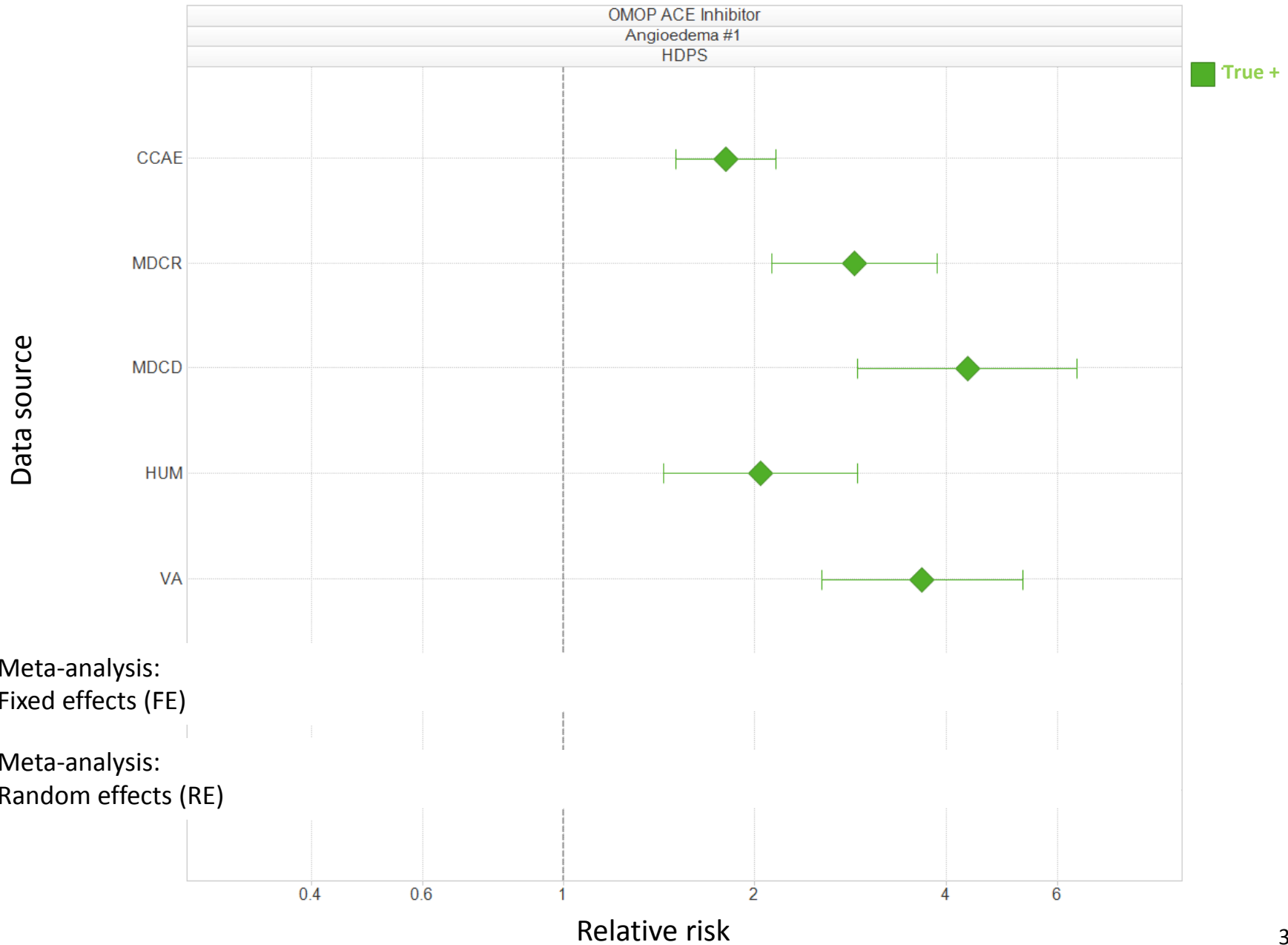
on behalf of OMOP Research Team

April 8, 2011

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



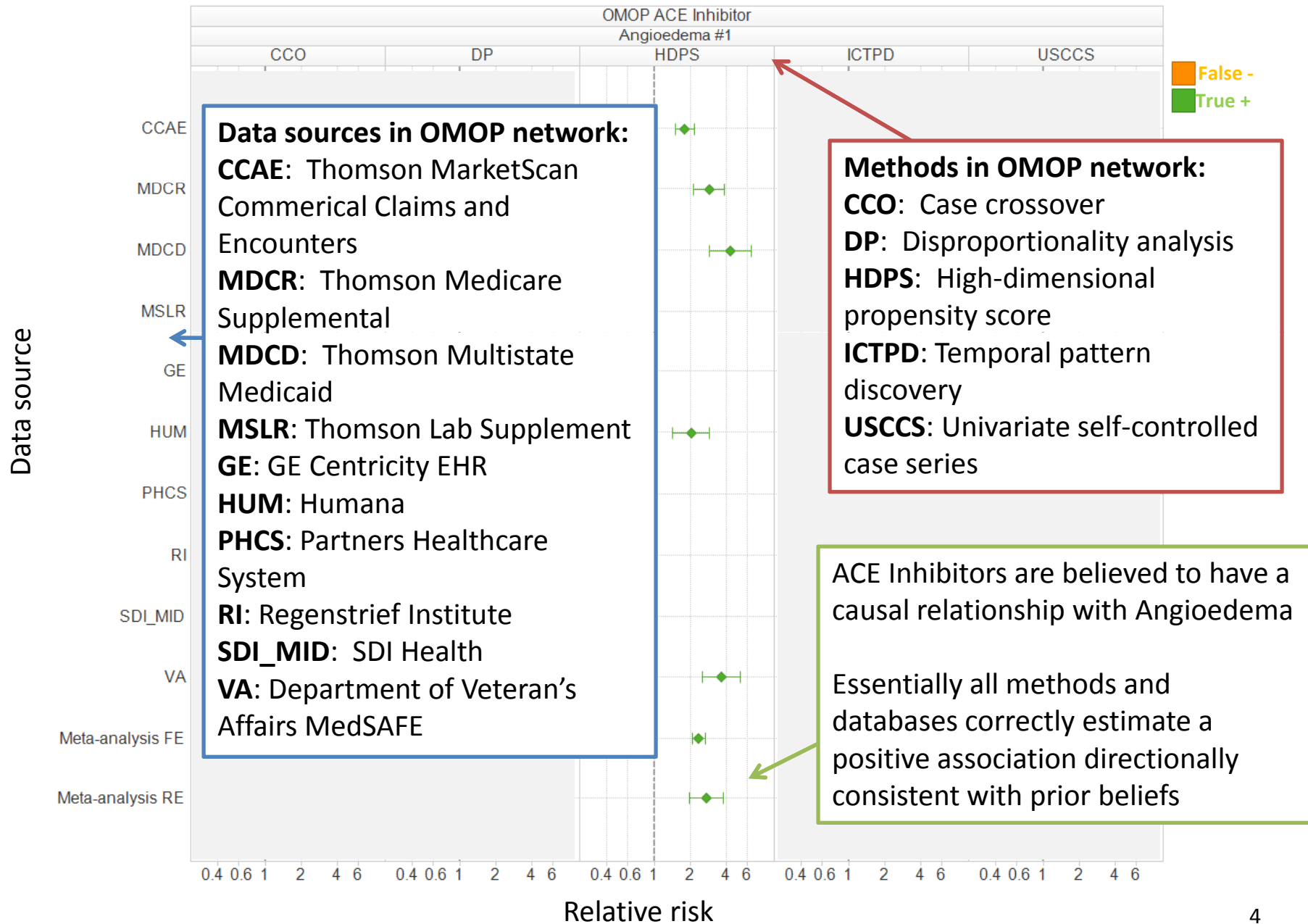
Opportunity for an active surveillance system: Pooling estimates across a network of disparate data sources



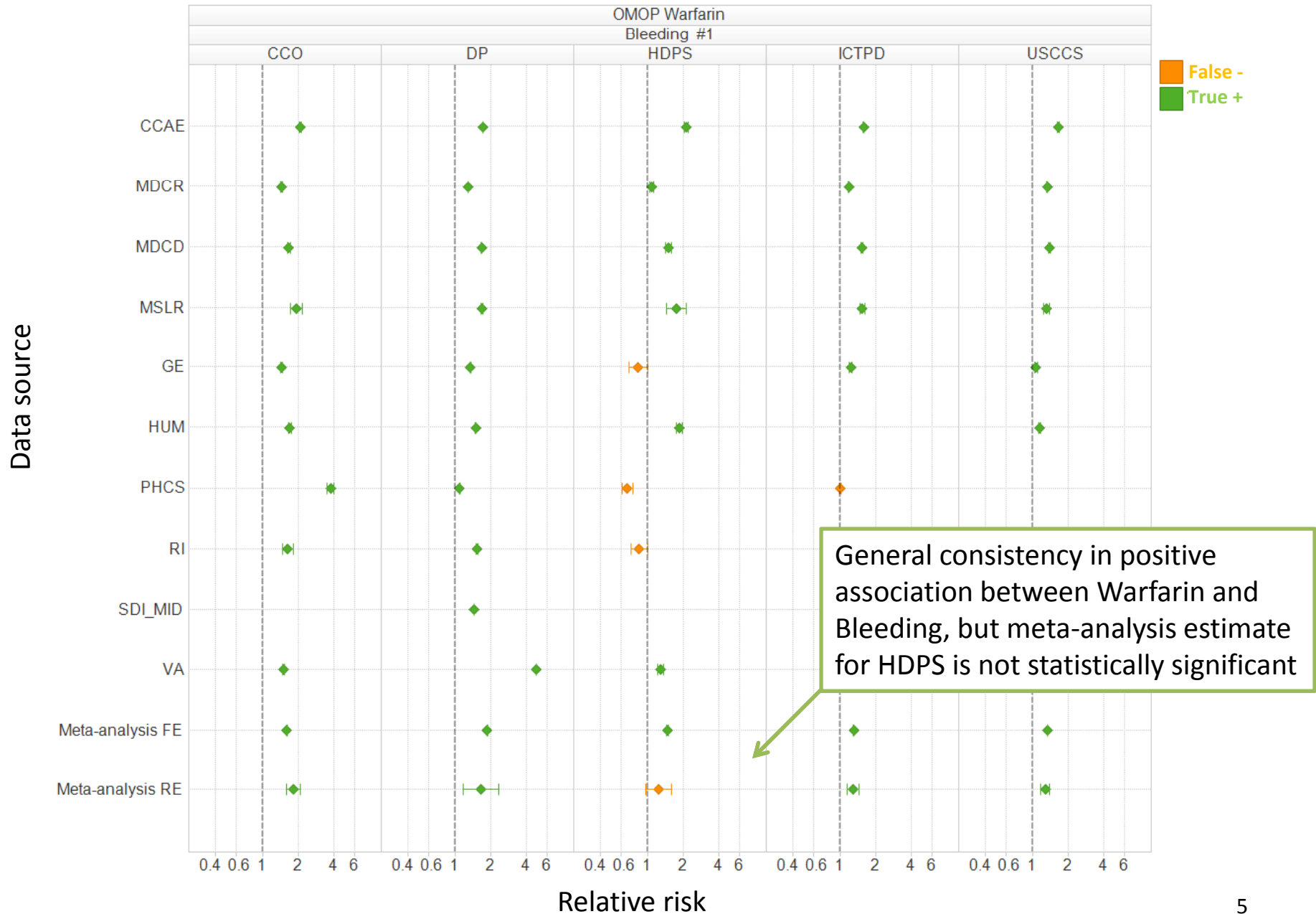
Meta-analysis:
Fixed effects (FE)

Meta-analysis:
Random effects (RE)

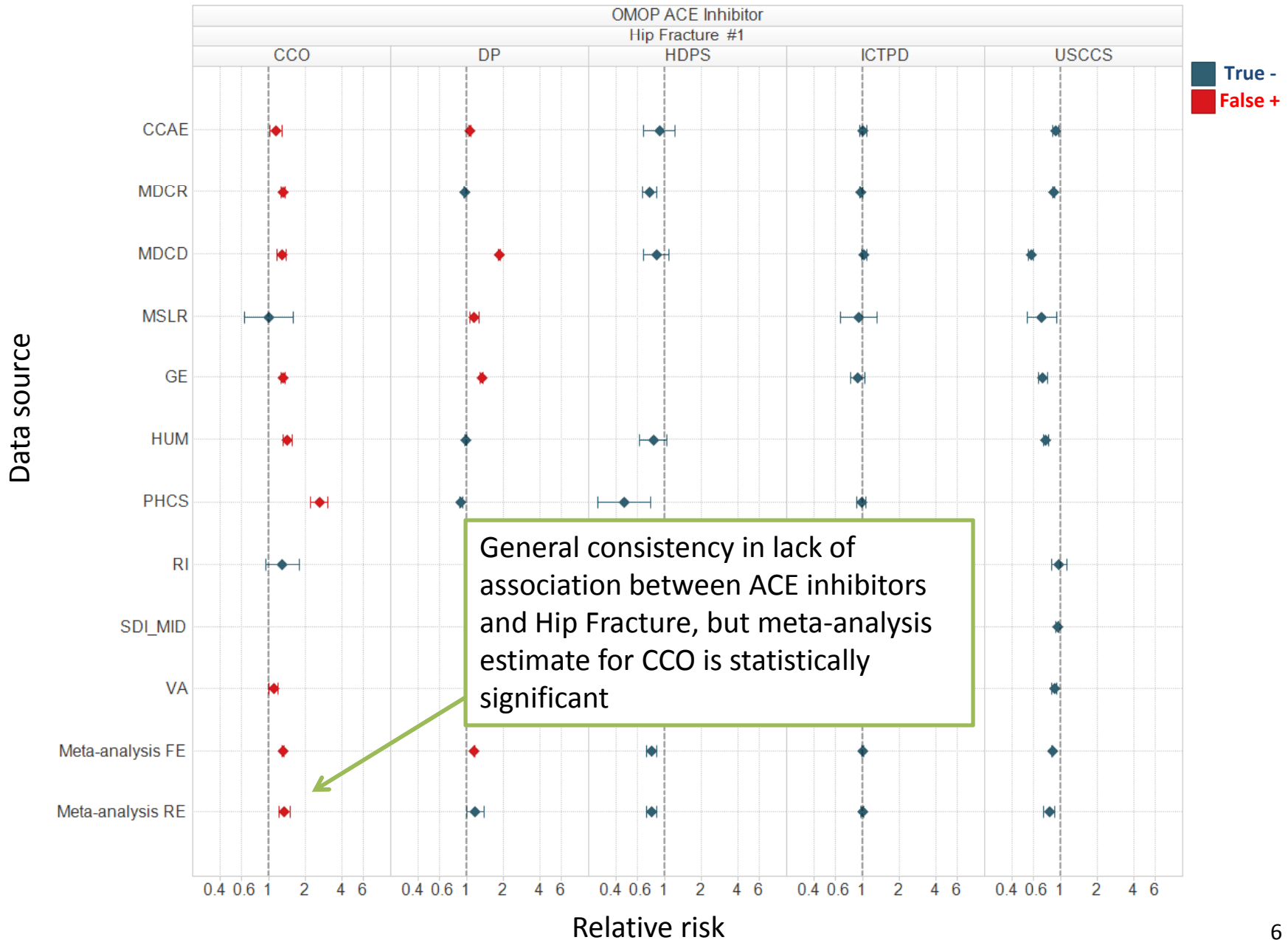
Systematic sensitivity analysis: Estimate the effect using multiple methods across the network of databases



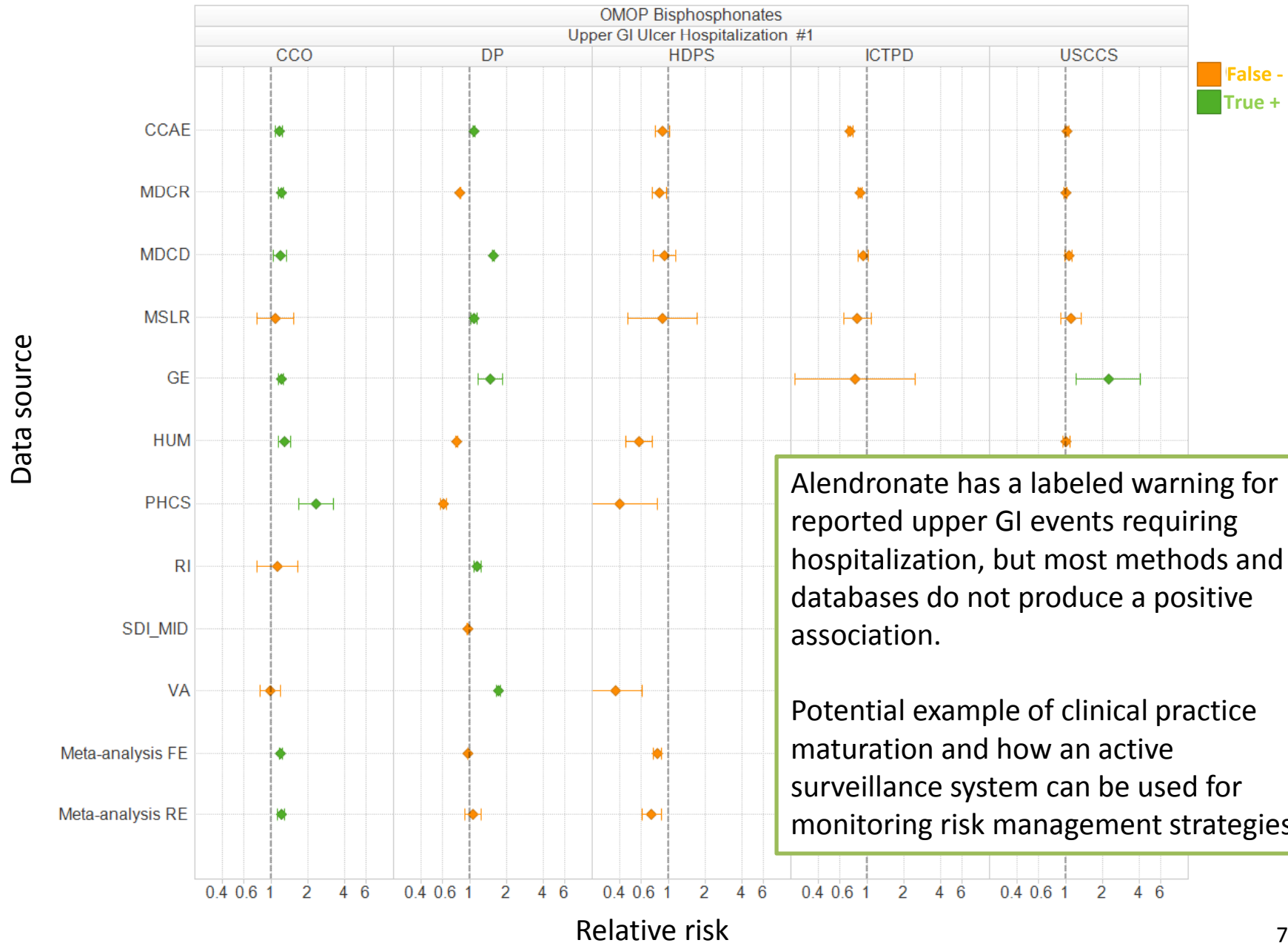
Evaluating the association between Warfarin and Bleeding



'True negative' observed for 'negative control' of ACE Inhibitors and Hip Fracture



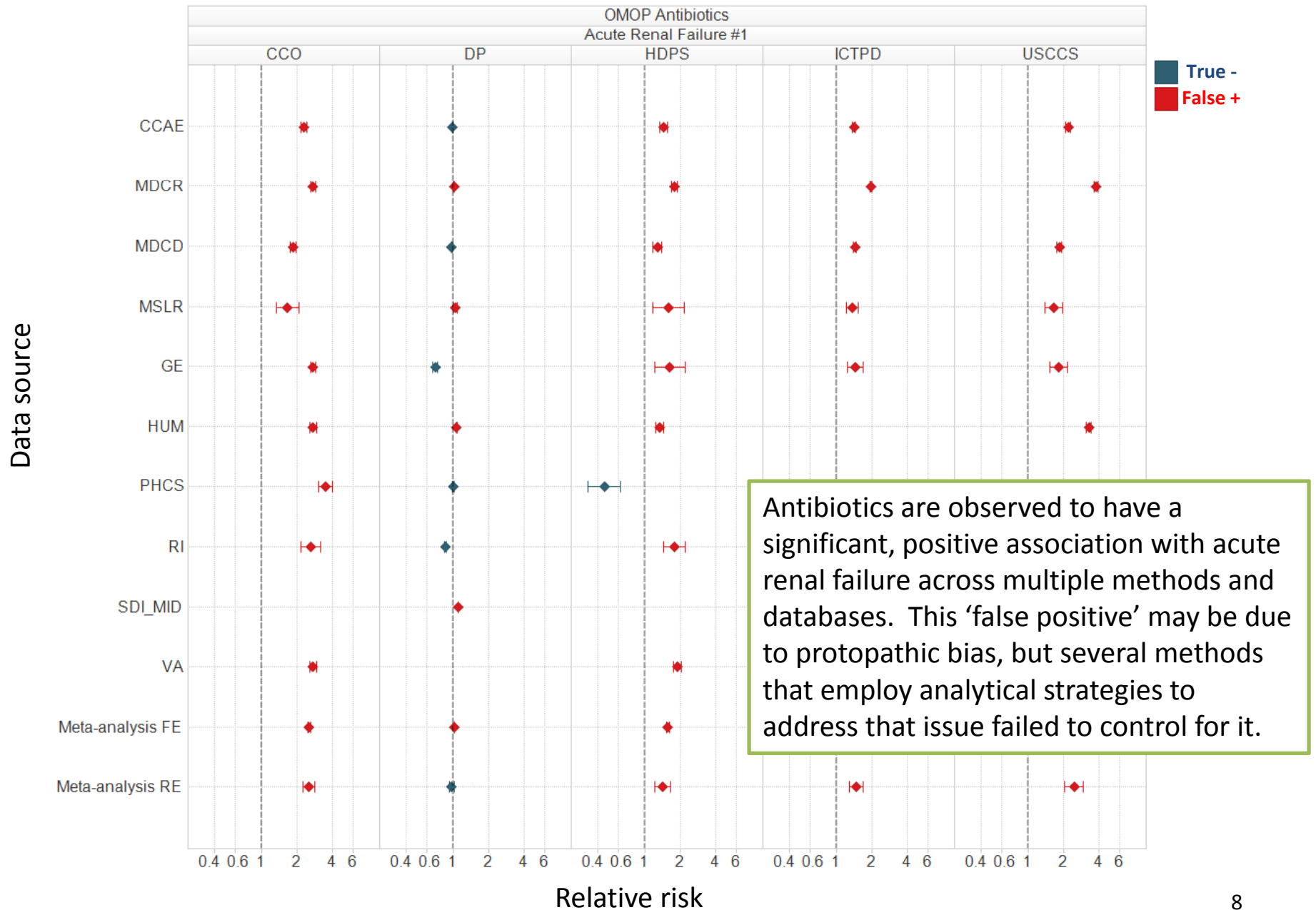
Consistent 'false negative' observed for 'true' association between Bisphosphonates and GI Ulcer Hospitalization



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.

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



Measuring method performance: Classifying the illustrative examples

Drug-condition association status

Y – ‘true association’,

N – ‘negative control’

Y

N

Method prediction:
 Drug-condition pair met a specific threshold:
 (LB 95% CI > 1)

Y

N

<p>True positives: ACE Inhibitors- Angioedema Warfarin- Bleeding</p>	<p>False positives: Typical antipsychotic- GI Ulcer hospitalization Antibiotics- Acute renal failure</p>
<p>False negatives: Bisphosphonates- GI Ulcer hospitalization</p>	<p>True negatives: ACE Inhibitors- Hip fracture</p>

Measuring method performance

Drug-condition association status

Y – ‘true association’,

N – ‘negative control’

Y

N

Method prediction:
Drug-condition
pair met a
specific
threshold

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?

'Ground truth' for Monitoring Health Outcomes of Interest

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

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

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’

Y

N

Method prediction:
Drug-condition pair met a specific threshold:
(LB 95% CI > 1)

Y

N

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

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

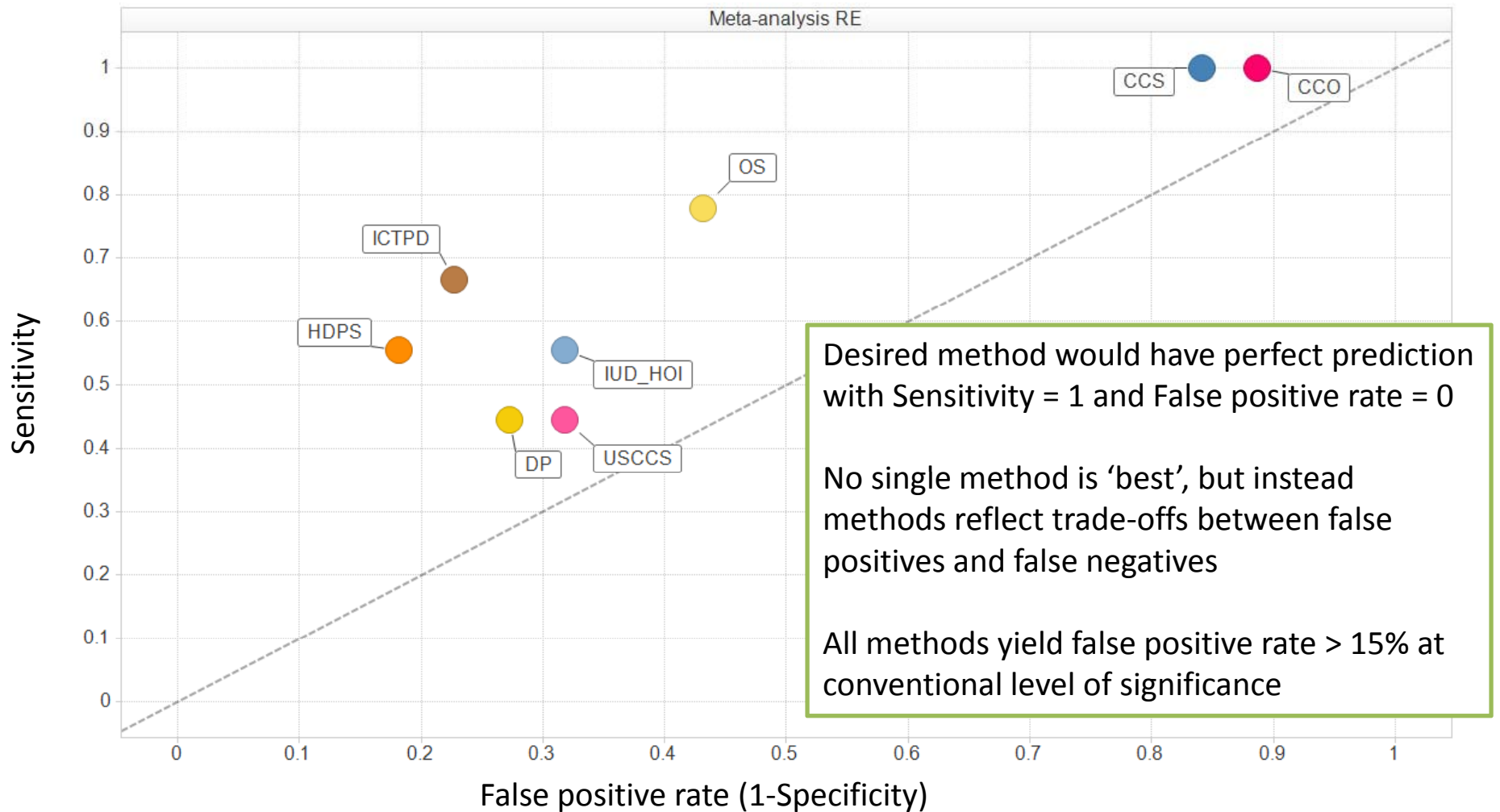
Negative predictive value
= $TN / (FN+TN)$
= $36 / (4+36) = 0.90$

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

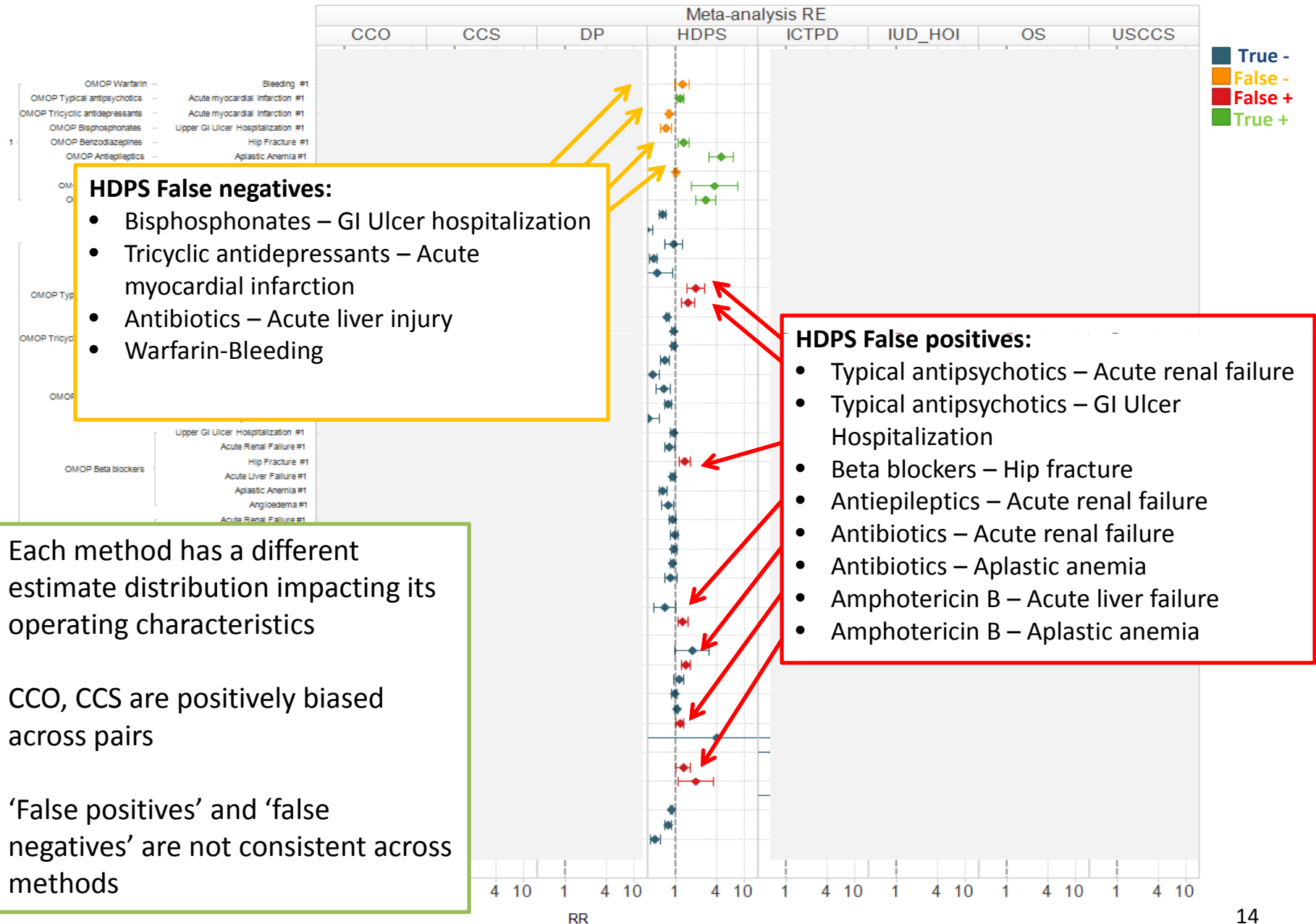
Specificity
= $TN / (FP+TN)$
= $36 / (8+36) = 0.82$
False positive rate
= $1 - 0.82 = 0.18$

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

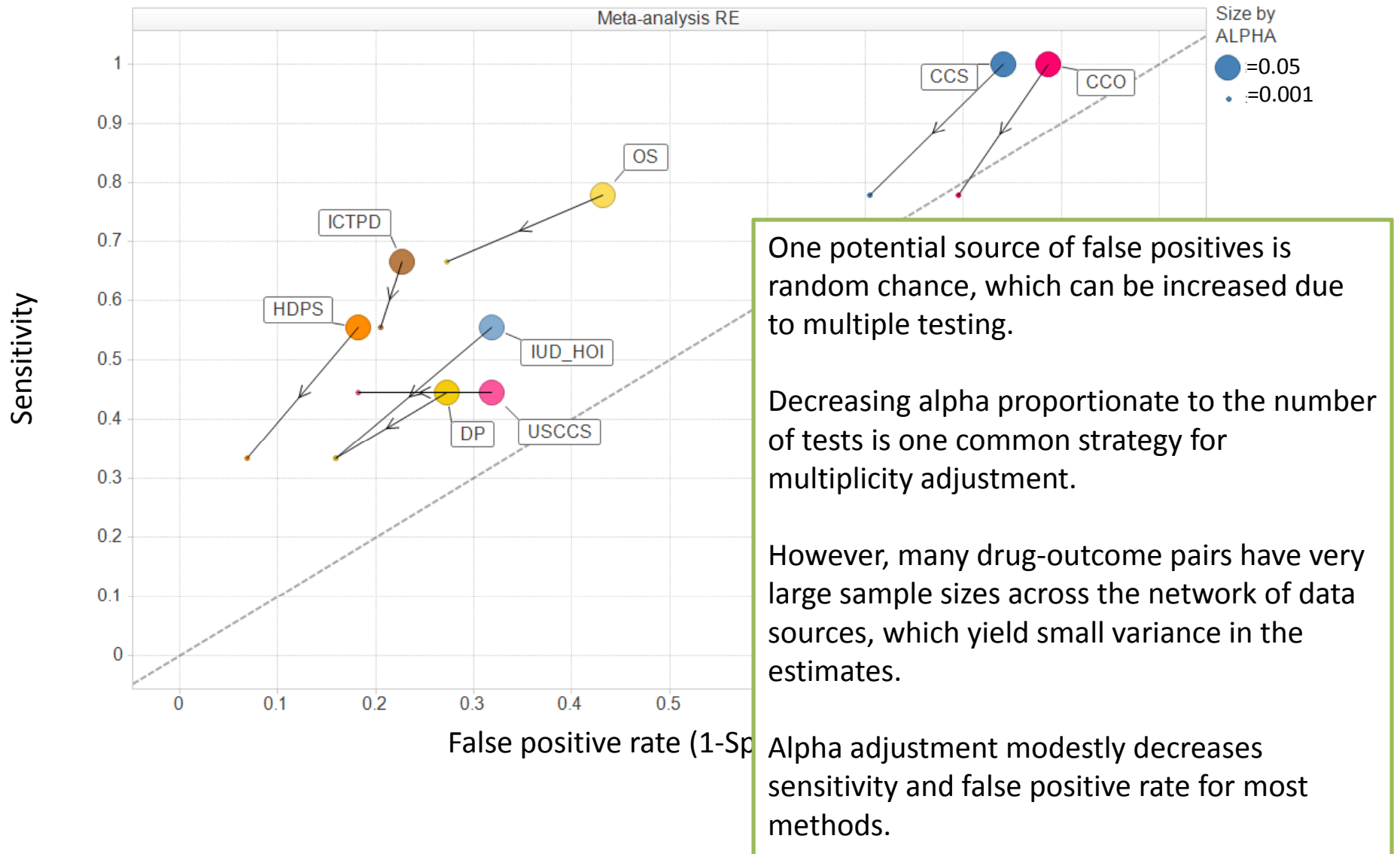
Comparing methods by sensitivity and specificity at alpha=0.05



Distribution of estimates across all drug-outcome pairs



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



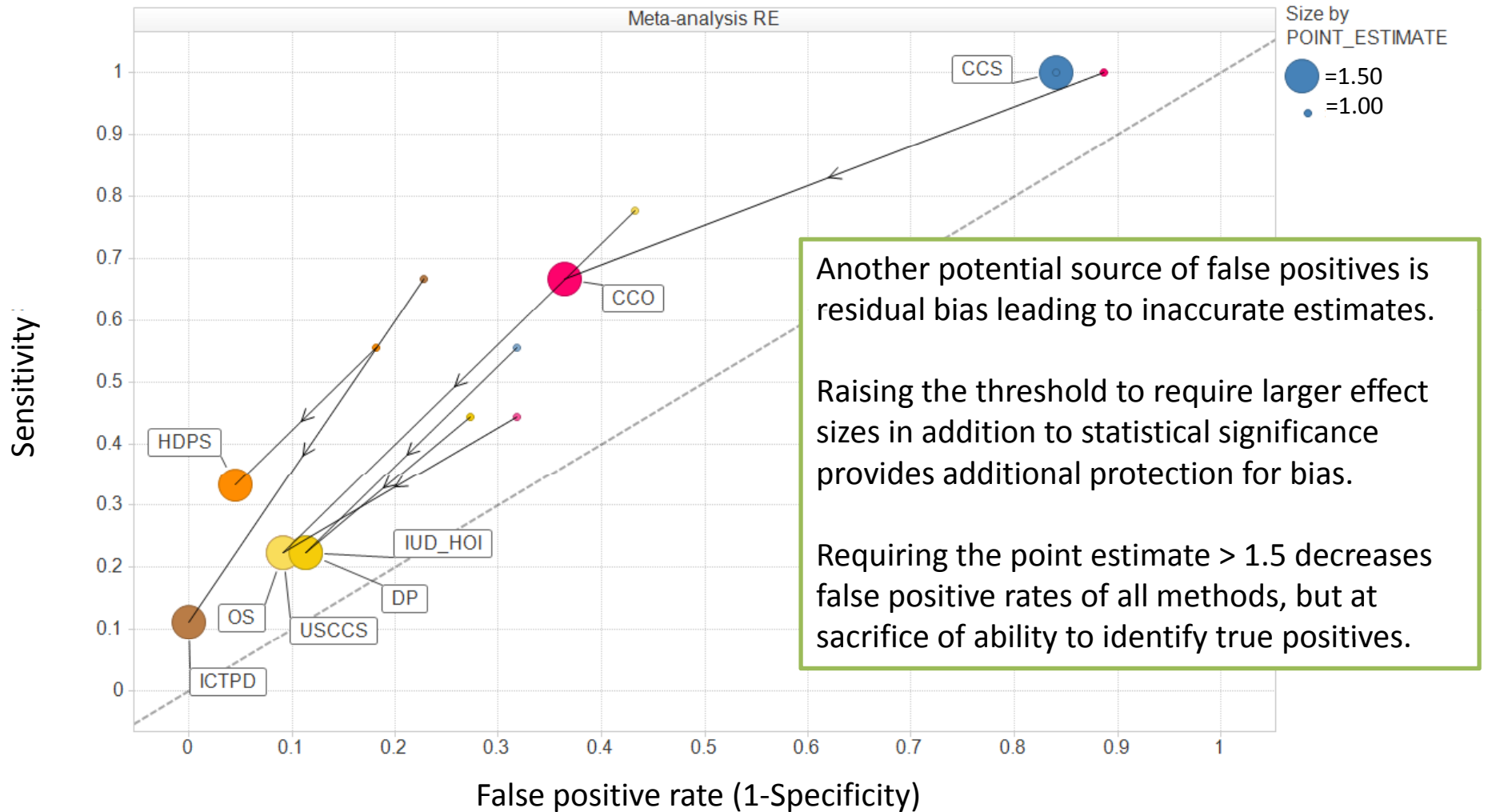
One potential source of false positives is random chance, which can be increased due to multiple testing.

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

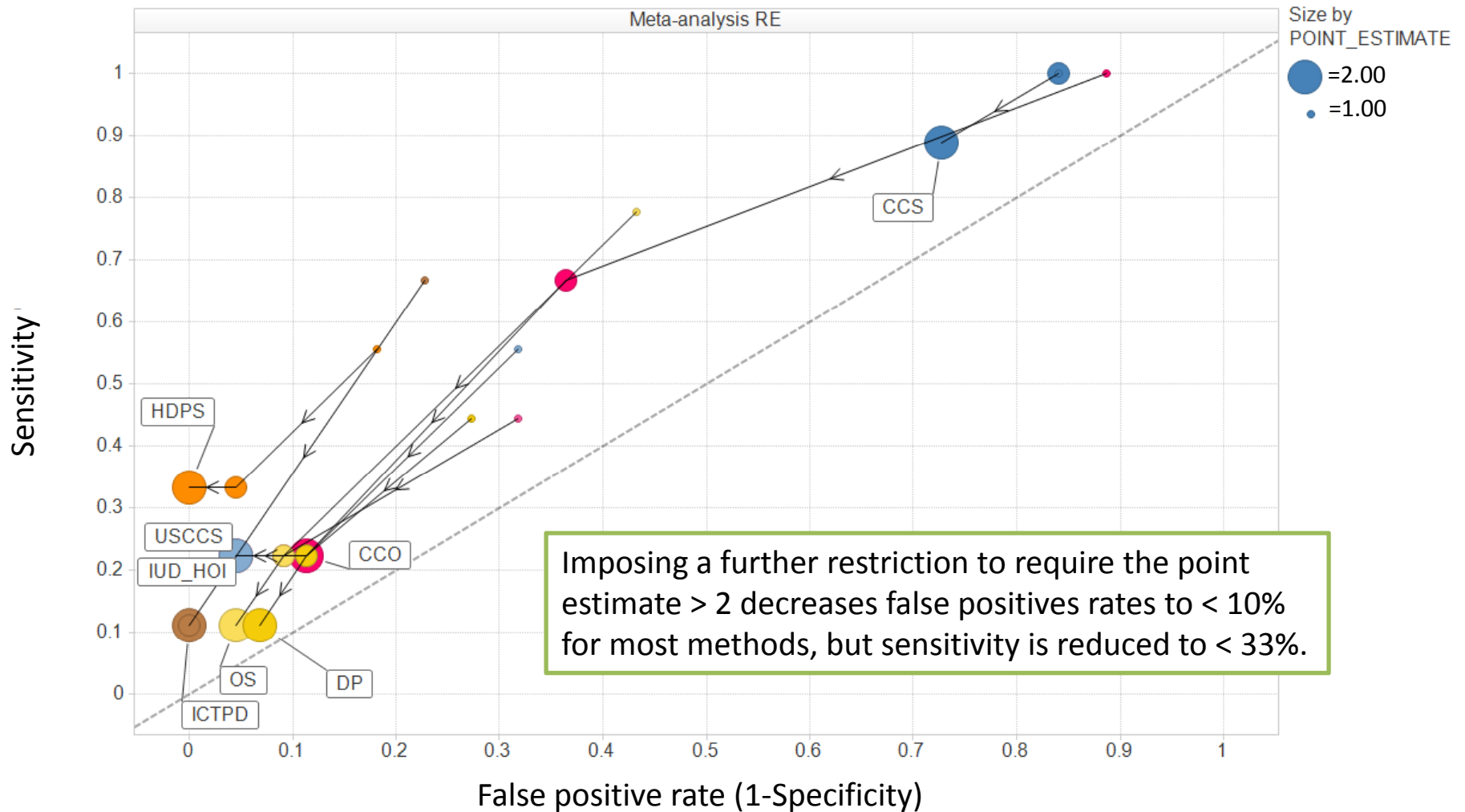
However, many drug-outcome pairs have very large sample sizes across the network of data sources, which yield small variance in the estimates.

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

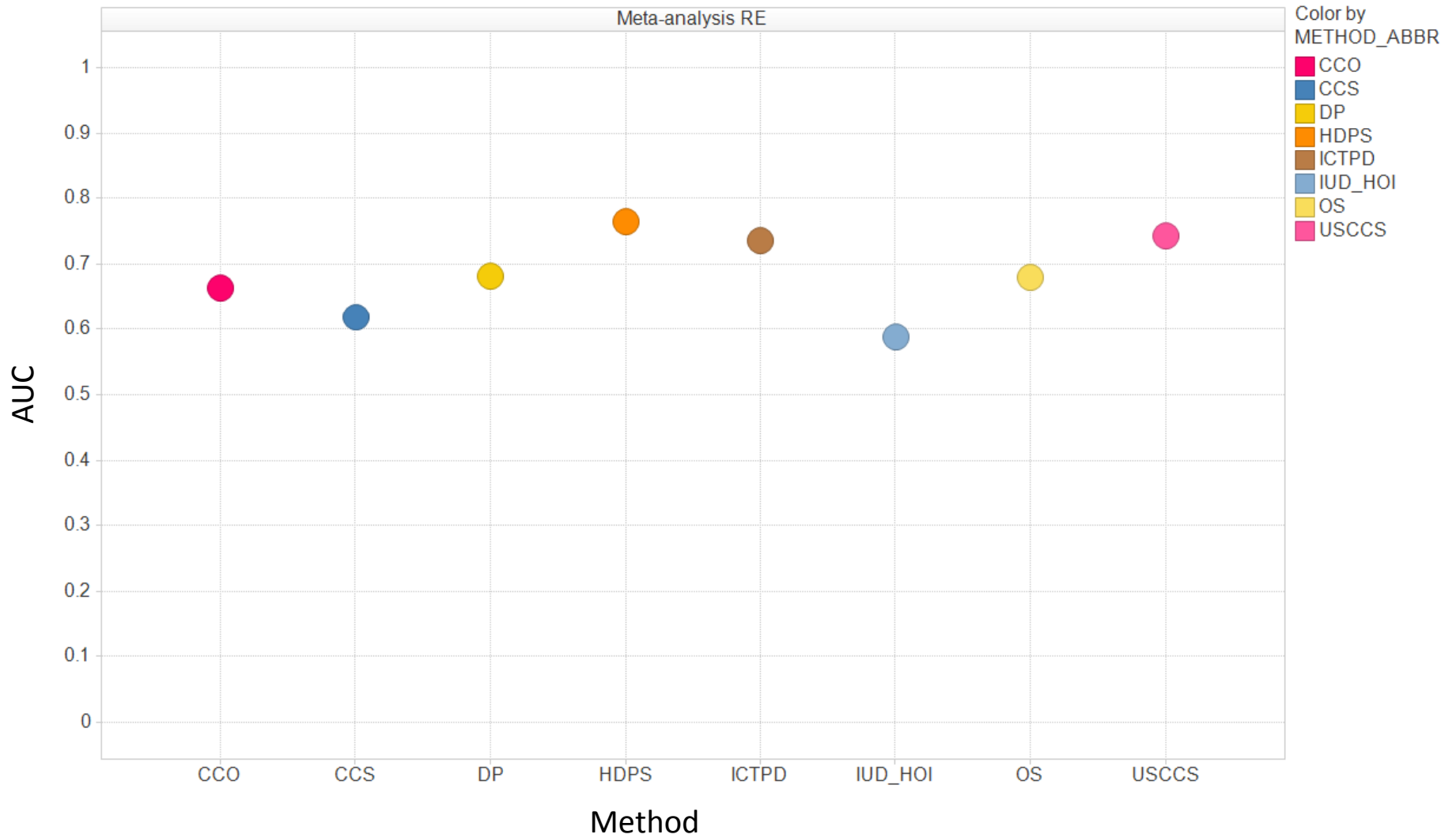
Change in sensitivity and specificity when raising threshold to require point estimate ≥ 1.5 (adjustment for bias)



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



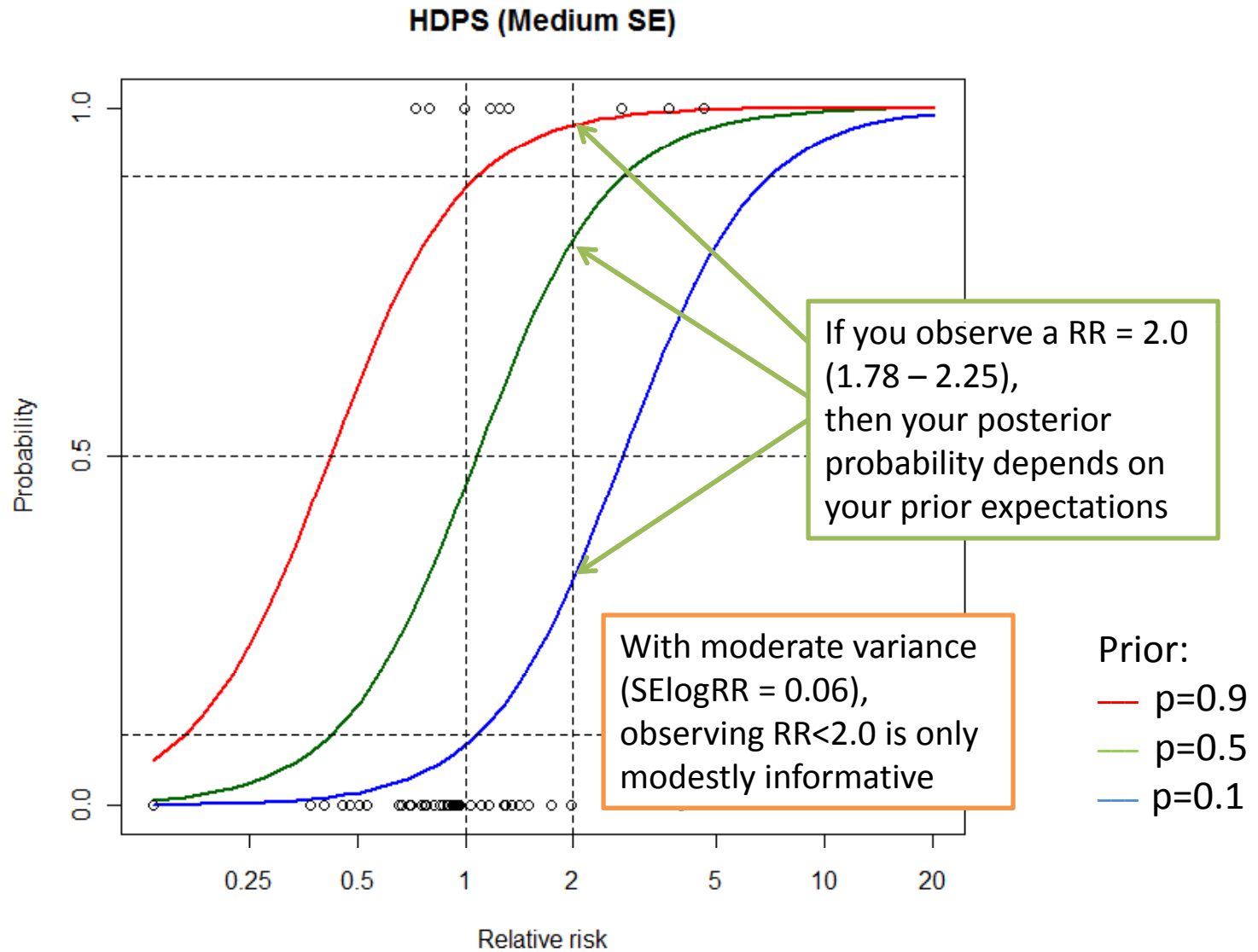
AUC performance by method



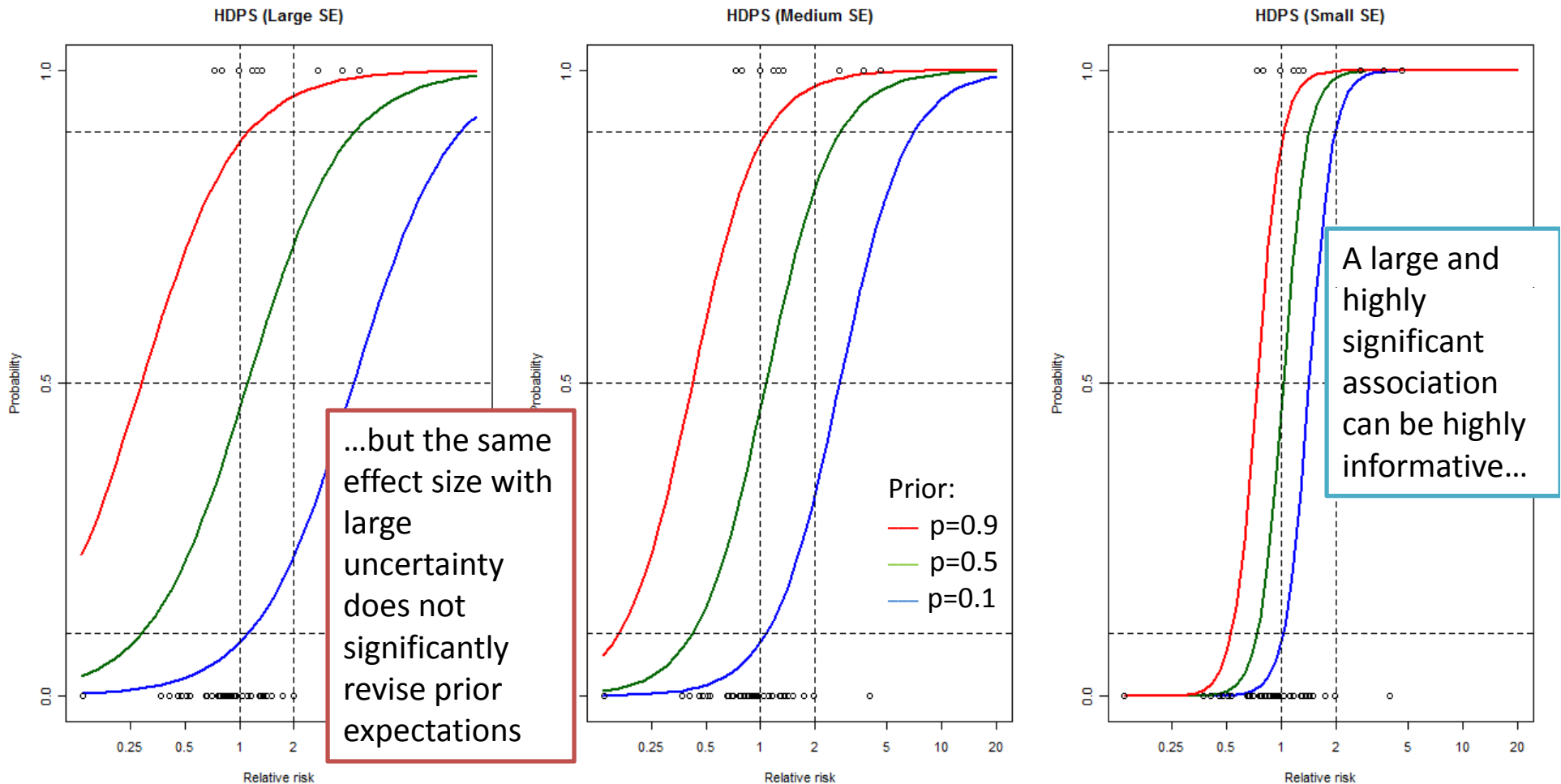
So given these operating characteristics, what can we expect to do in practice....

- 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...
 - $p(\text{true} | \text{RR}, \text{SE}) \sim p(\text{RR}, \text{SE} | \text{true}) * p(\text{true})$
 - $p(\text{true})$ is the prior probability of true association; consider a family of priors: skeptical (0.1), indifferent (0.5), enthusiastic (0.9)
 - $p(\text{RR}, \text{SE} | \text{true})$ can be estimated from empirical data (OMOP experimental results)

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

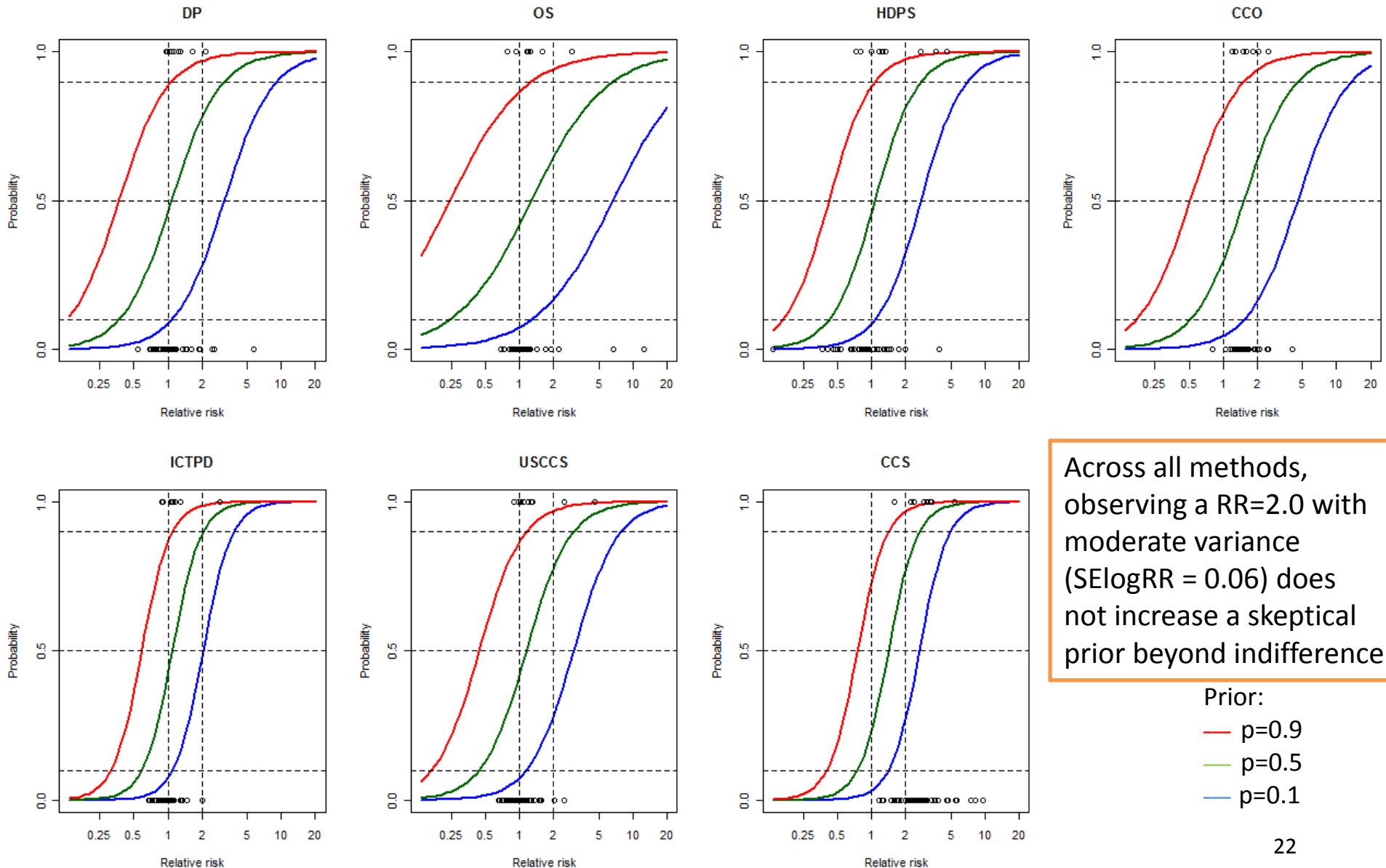


Revising prior expectations in light of new evidence from an active surveillance system: Impact of precision of observed estimates



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)

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



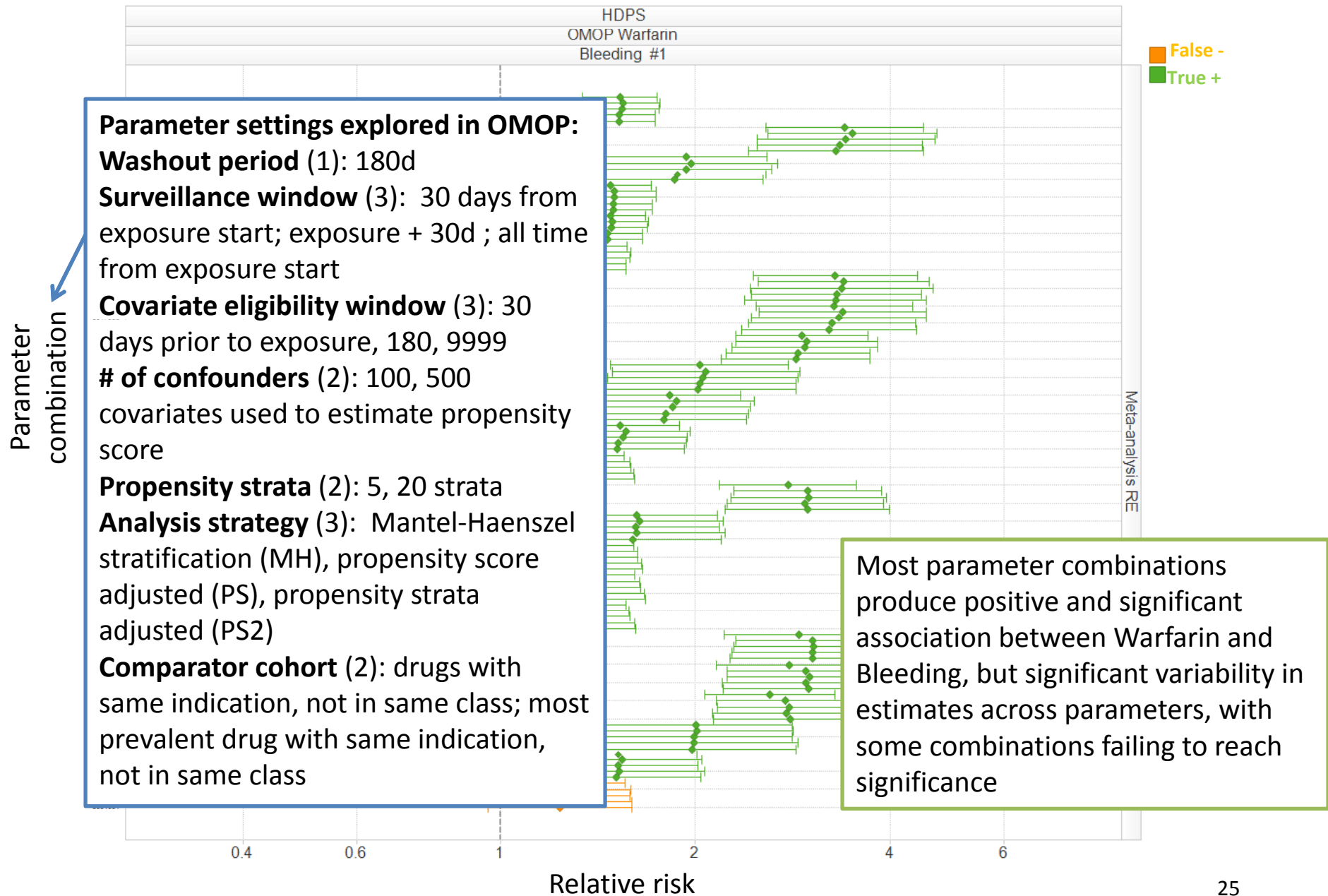
Concluding thoughts

- No one clear ‘best’ method, as it depends on tolerance for false positives vs. false negatives
- In this experiment, active surveillance methods achieved:
 - 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

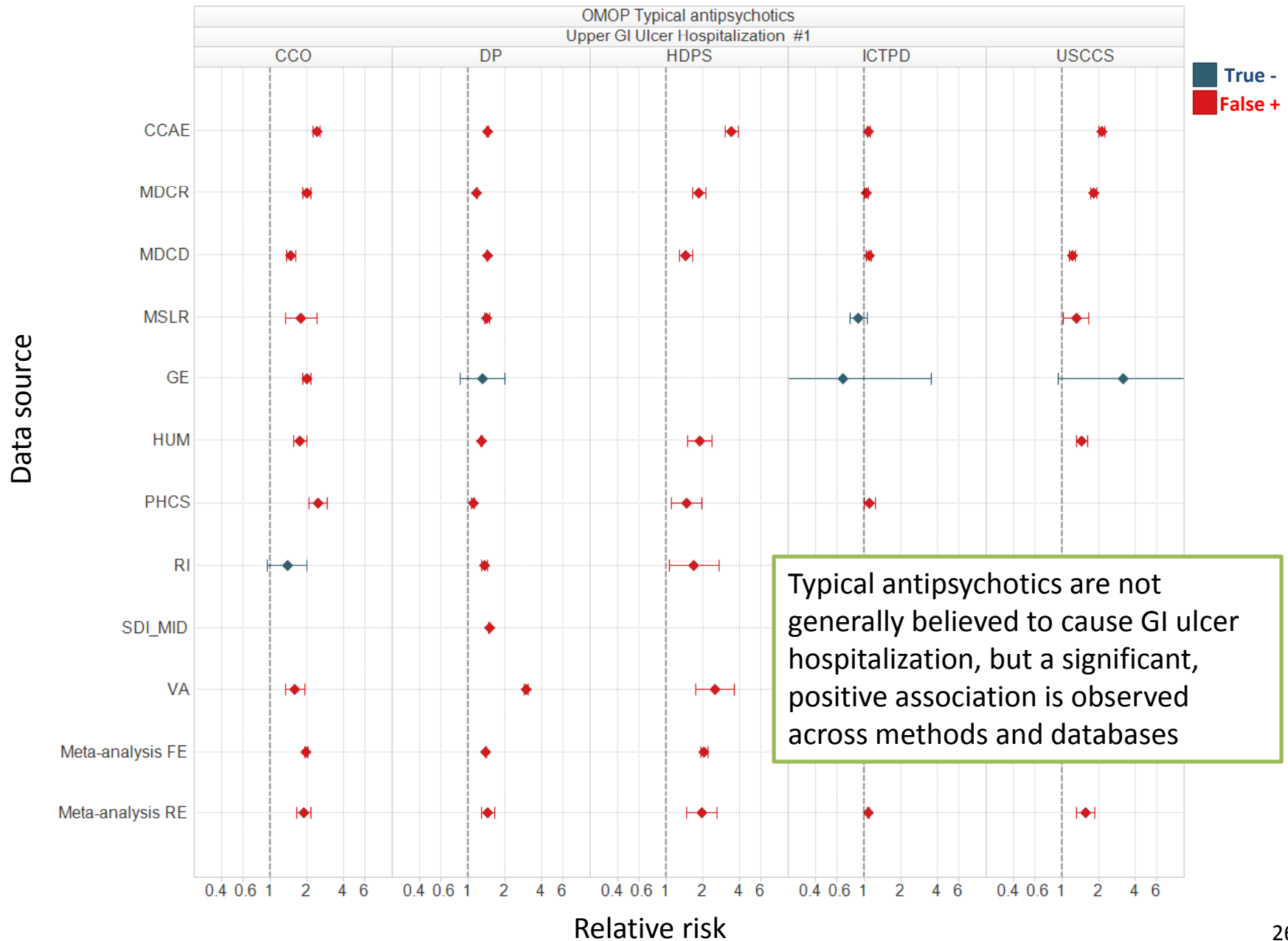
Acknowledgements

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

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



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



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

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