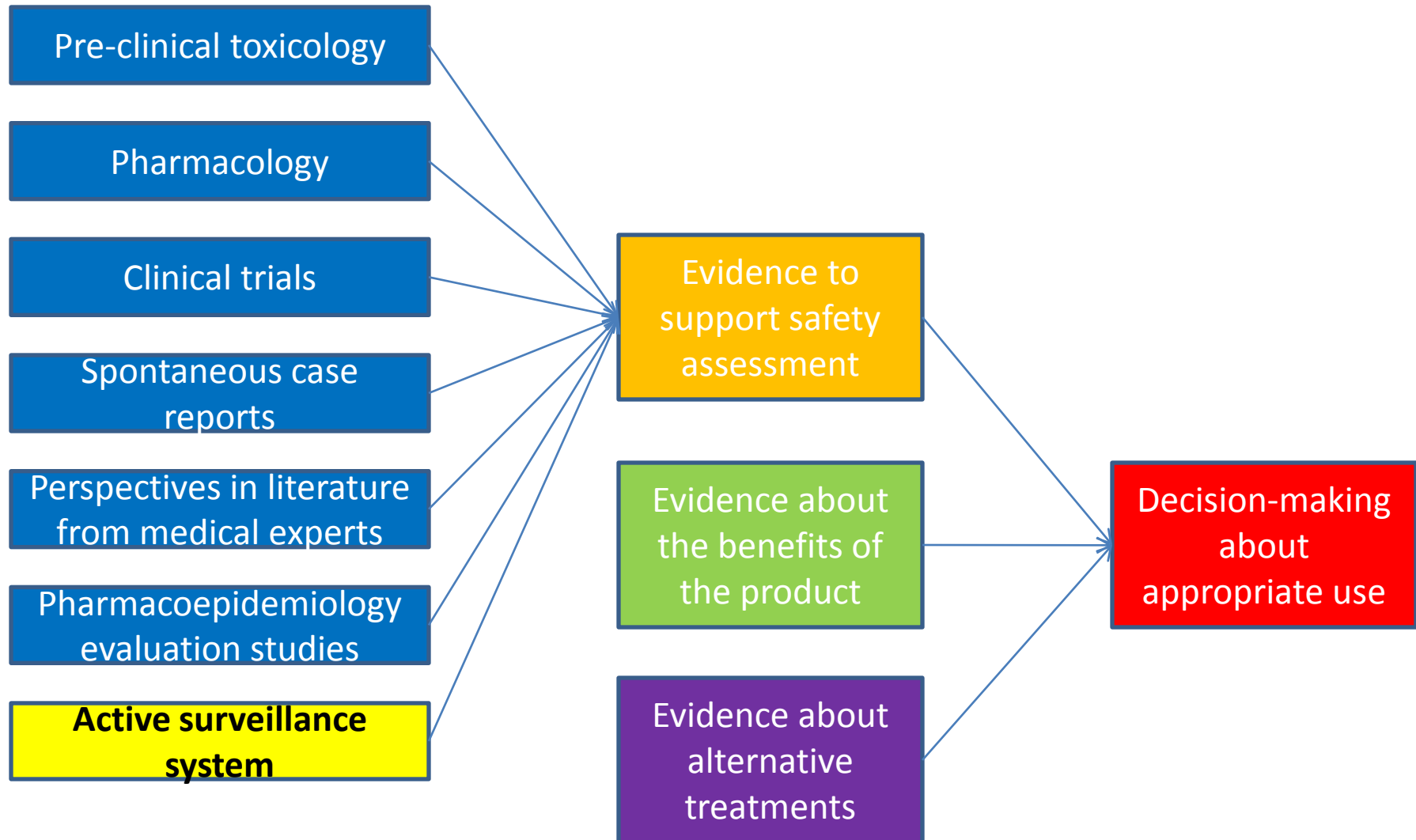


**OBSERVATIONAL  
MEDICAL  
OUTCOMES  
PARTNERSHIP**

## **OMOP Methods Evaluation**

Patrick Ryan, David Madigan  
on behalf of OMOP Research Team  
January 11, 2011

# Active surveillance: One additional piece of evidence to inform medical decision-making



## Concluding thoughts, part 1

- An active surveillance system can successfully complement current practice by providing useful evidence to support a comprehensive safety assessment
- 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 rates range from 16%-30%
  - At 10% false positive rate, sensitivities range from 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

## Additional questions warranting exploration

- Is performance differentiated by specific drugs or outcomes?
- How do alternative outcome definitions impact estimation?
- How does methods performance vary by data sources?
- Can method parameter settings be further tailored to improve performance?

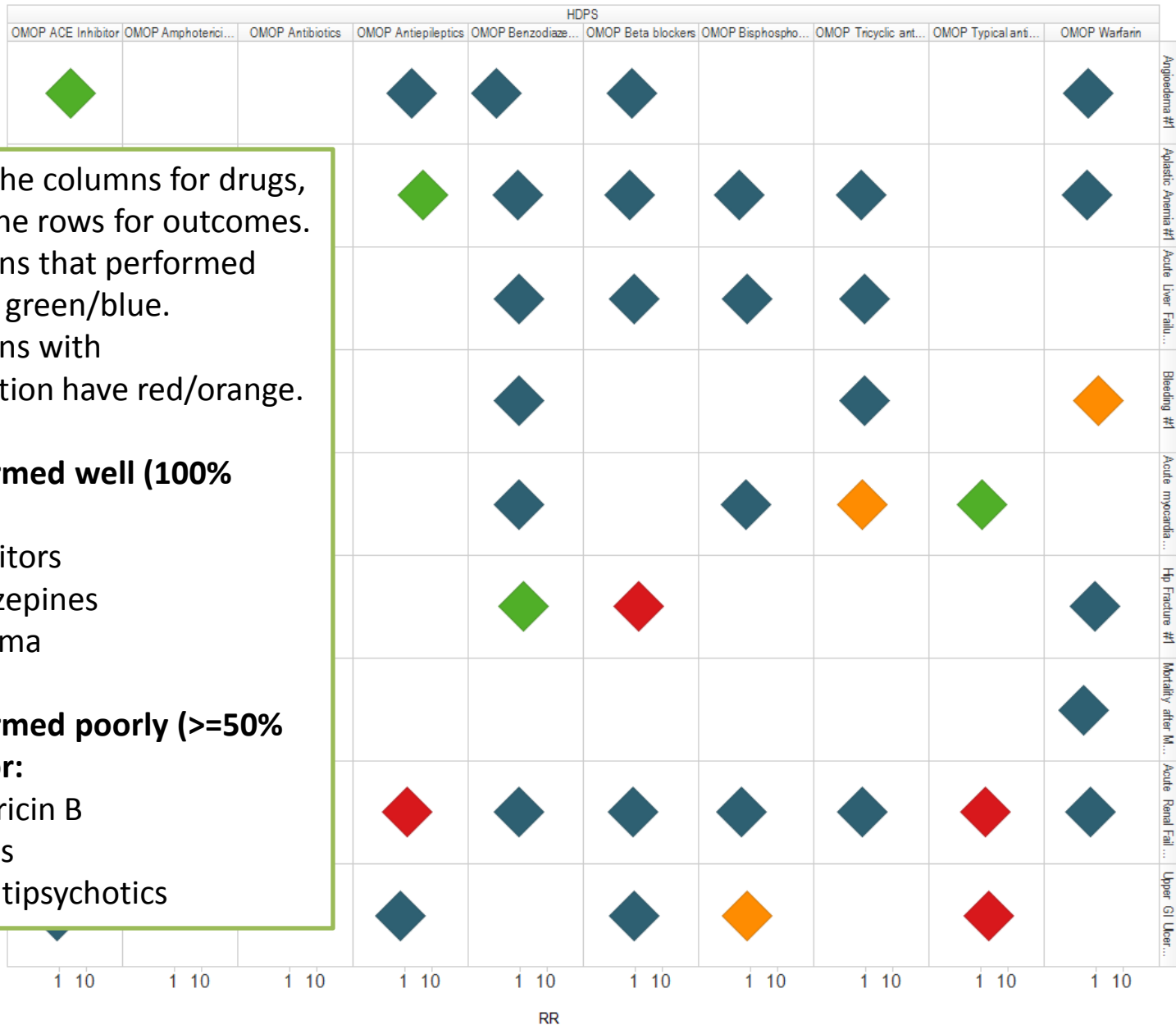
## Additional questions warranting exploration

- **Is performance differentiated by specific drugs or outcomes?**
- How do alternative outcome definitions impact estimation?
- How does methods performance vary by data sources?
- Can method parameter settings be further tailored to improve performance?

# Drug and outcome differentiation analysis

- Analyses thus far have focused on methods performance across the entire array of drug-outcome pairs
- Question: are there specific drugs or outcomes for which we observe differential performance?
  - Are there scenarios where methods perform well by classifying all true positives and true negatives correctly?
  - Are there scenarios where methods perform badly by misclassifying negative controls as false positives and true associations as false negatives?

# HDPS performance by drug-outcome pair



Look down the columns for drugs, and across the rows for outcomes. Rows/columns that performed well have all green/blue. Rows/columns with misclassification have red/orange.

**HDPS performed well (100% correct) for:**

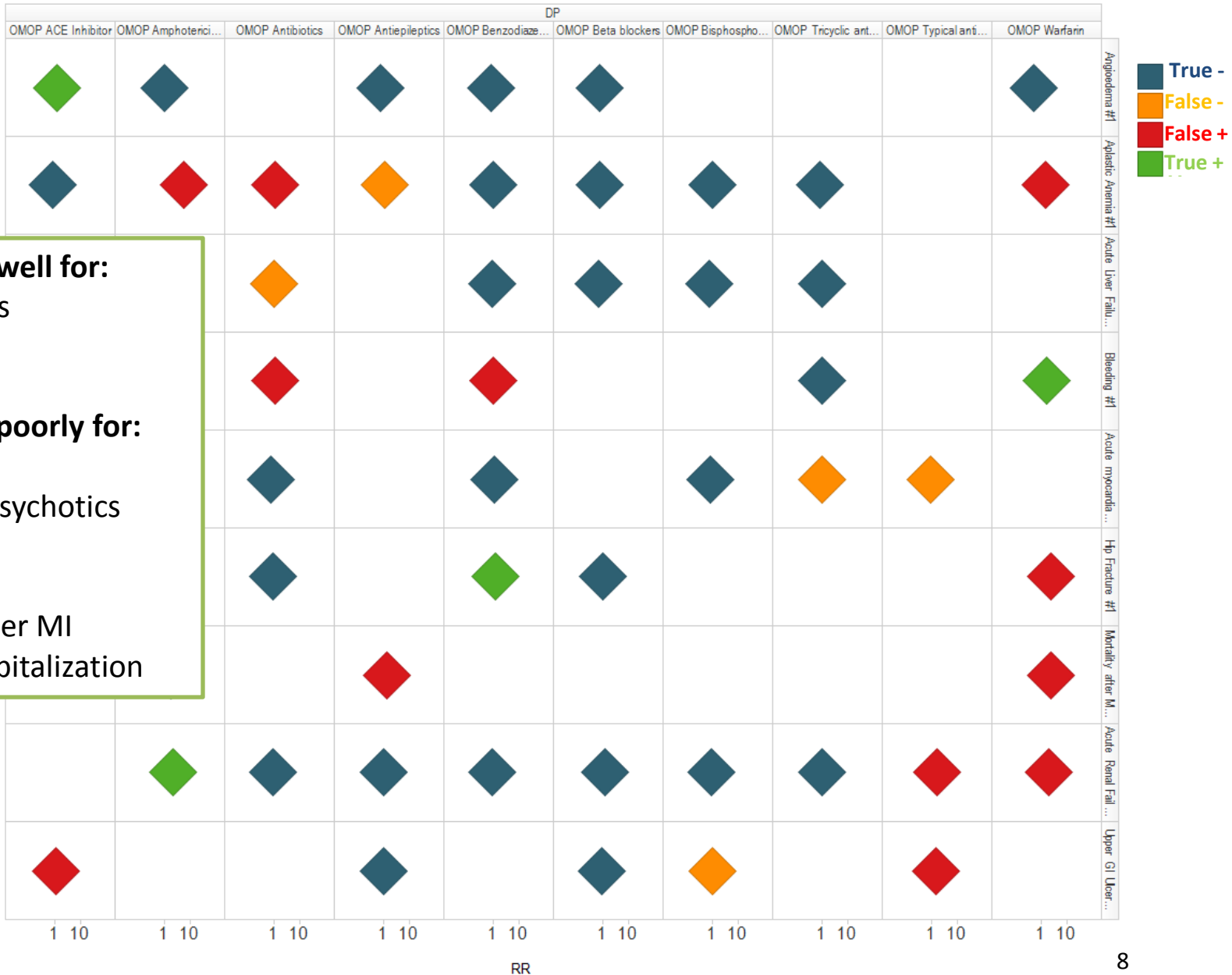
- ACE Inhibitors
- Benzodiazepines
- Angioedema

**HDPS performed poorly (>=50% incorrect) for:**

- Amphotericin B
- Antibiotics
- Typical antipsychotics

True -  
False -  
False +  
True +

# DP performance by drug-outcome pair



**DP performed well for:**

- Beta blockers
- Angioedema

**DP performed poorly for:**

- Antibiotics
- Typical antipsychotics
- Warfarin
- Bleeding
- Mortality after MI
- GI Ulcer hospitalization

# ICTPD performance by drug-outcome pair



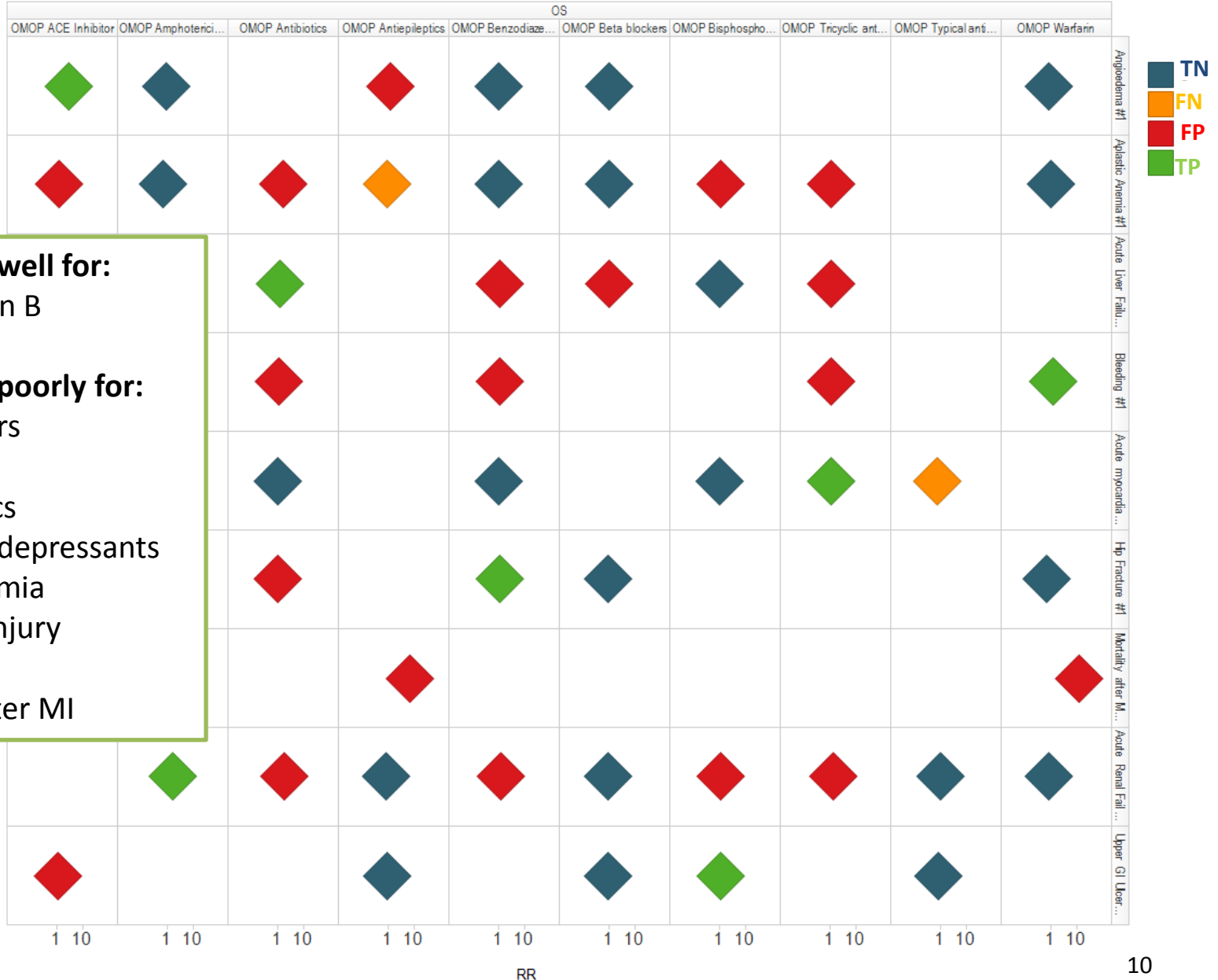
**ICTPD performed well for:**

- ACE inhibitors
- Tricyclic antidepressants
- Warfarin
- Hip fracture

**ICTPD performed poorly for:**

- Antibiotics
- Antiepileptics
- Typical antipsychotics

# OS performance by drug-outcome pair



**OS performed well for:**

- Amphotericin B

**OS performed poorly for:**

- ACE inhibitors
- Antibiotics
- Antiepileptics
- Tricyclic antidepressants
- Aplastic anemia
- Acute liver injury
- Bleeding
- Mortality after MI

# USCCS performance by drug-outcome pair



- ◆ True -
- ◆ False -
- ◆ False +
- ◆ True +

**USCCS performed well for:**

- Amphotericin B
- Beta blockers
- Acute liver injury
- Mortality after MI

**USCCS performed poorly for:**

- Antibiotics
- Typical antipsychotics
- Acute myocardial infarction
- GI Ulcer hospitalization

## Drug and outcome differentiation analysis summary

- Multiple methods performed well and no method performed poorly for:
  - Beta blockers
  - Angioedema
- Multiple methods fared poorly but no method performed well for:
  - Antibiotics
  - Typical antipsychotics
  - GI Ulcer hospitalization
- Further research needed with additional test cases to draw generalizable conclusions about the scenarios where we should expect differential performance from an active surveillance system

## Additional questions warranting exploration

- Is performance differentiated by specific drugs or outcomes?
- **How do alternative outcome definitions impact estimation?**
- How does methods performance vary by data sources?
- Can method parameter settings be further tailored to improve performance?

# Acute renal failure definitions

1. Occurrence of at least one diagnosis code  
ICD9 584\*

2. Occurrence of at least one diagnosis code  
AND treatment procedure for acute dialysis  $\geq 60$ d after  
EXCLUDING diagnosis code for chronic dialysis status

3. An increase in serum creatinine level (LOINC 2160-0) of  $\geq 0.5$   
mg/dl for patients with a baseline serum creatinine level of  
 $\leq 1.9$  mg/dl,  $\geq 1.0$  mg/dl for patients with a baseline level of  
2.0–4.9 mg/dl, and  $\geq 1.5$  mg/dl for patients with a baseline  
level  $\geq 5.0$  mg/dl



# Acute liver injury definitions

1. Occurrence of at least one broad diagnosis code

2. Occurrence of at least one narrow diagnosis code

3. Occurrence of at least one narrow diagnosis code  
AND (diagnostic procedure  $\leq 30$ d before  
OR treatment procedure  $\geq 60$ d after)

4. Occurrence of at least one narrow diagnosis code  
AND (diagnostic procedure  $\leq 30$ d before  
OR treatment procedure  $\geq 60$ d after)  
AND laboratory results indicative of Hy's law:  
ALT  $\geq 3$ xULN AND AST  $\geq 3$ xULN AND Bilirubin  $\geq 2$ xULN  
within 7 days

5. Laboratory results indicative of Hy's law:  
ALT  $\geq 3$ xULN AND AST  $\geq 3$ xULN AND Bilirubin  
 $\geq 2$ xULN  
within 7 days

6. Laboratory results strongly indicative of Hy's law:  
ALT  $\geq 10$ xULN AND AST  $\geq 10$ xULN AND Bilirubin  $\geq 2$ xULN  
within 7 days



# Aplastic anemia definitions

1. Occurrence of at least one broad diagnosis code: ICD9:  
284.0\* or 284.8\* or 284.9

2. Occurrence of at least one broad diagnosis code  
AND diagnostic procedure for bone marrow aspiration or  
biopsy (CPT 38220, 38221)  $\leq$  60d before

3. Occurrence of at least one narrow diagnosis code  
(ICD9 284.8\*) AND diagnostic procedure for bone marrow  
aspiration or biopsy (CPT 38220, 38221)  $\leq$  60d before

5. #3 OR #4

4. Occurrence of at least two or more of the following laboratory values:

- White blood count (LOINC 26464-8, 6690-2, 804-5)  $</$   $3.5 \times 10^9/L$
- platelet count (LOINC 26515-7, 777-3, 778-1)  $\leq 50 \times 10^9 /L$
- hemoglobin (LOINC 718-7)  $\leq 100 \text{ g/L}$



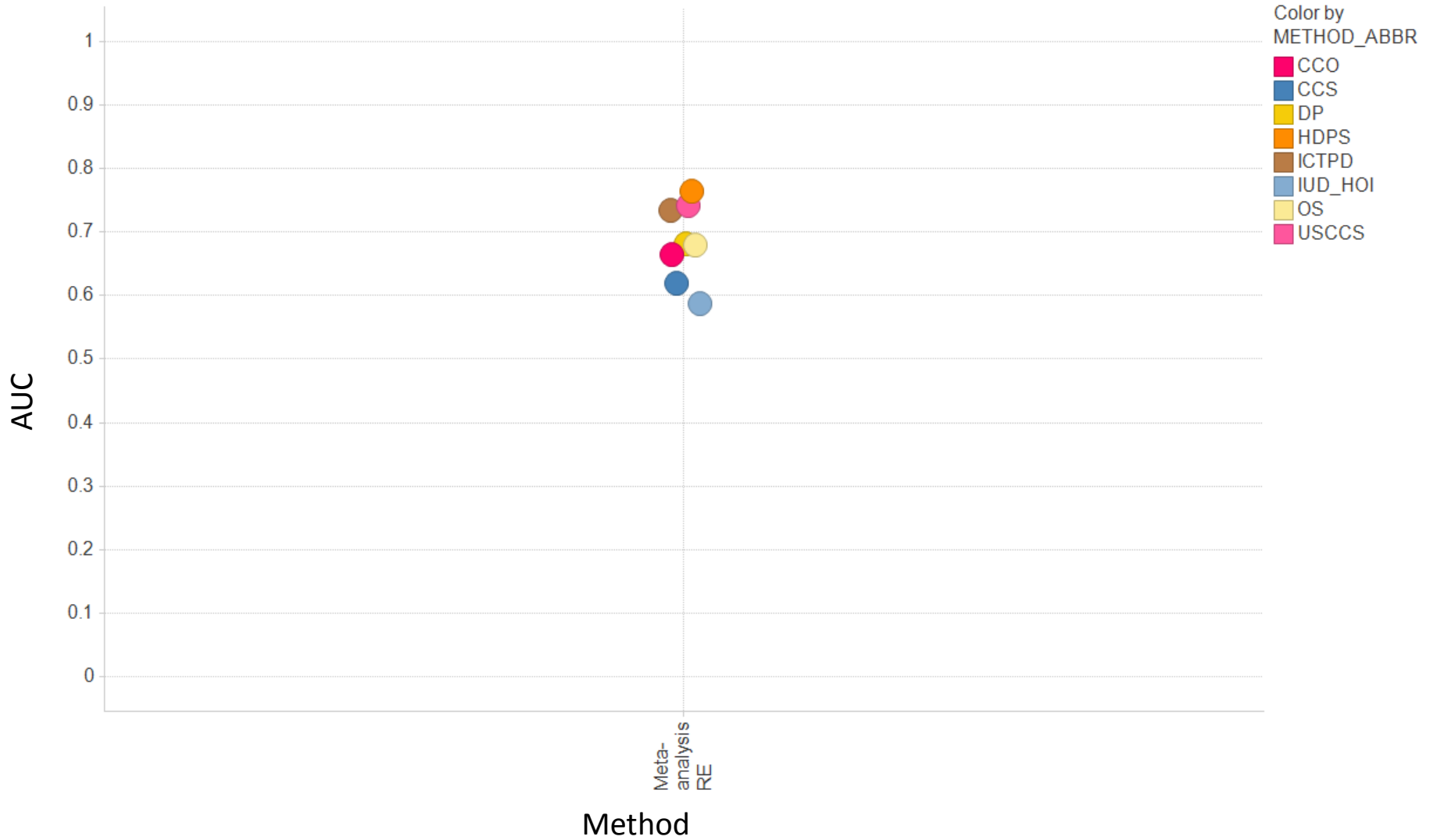
## HOI definition sensitivity

- No definitive patterns about best practice to apply broad vs. narrow definitions across conditions
- Statistical significance of method estimation can be sensitive to outcome definition due to varying number of observed cases
- Laboratory value-based definitions provide a complementary perspective to diagnostic codes with comparable or better predictive value, but performance varies by method

## Additional questions warranting exploration

- Is performance differentiated by specific drugs or outcomes?
- How do alternative outcome definitions impact estimation?
- **How does methods performance vary by data sources?**
- Can method parameter settings be further tailored to improve performance?

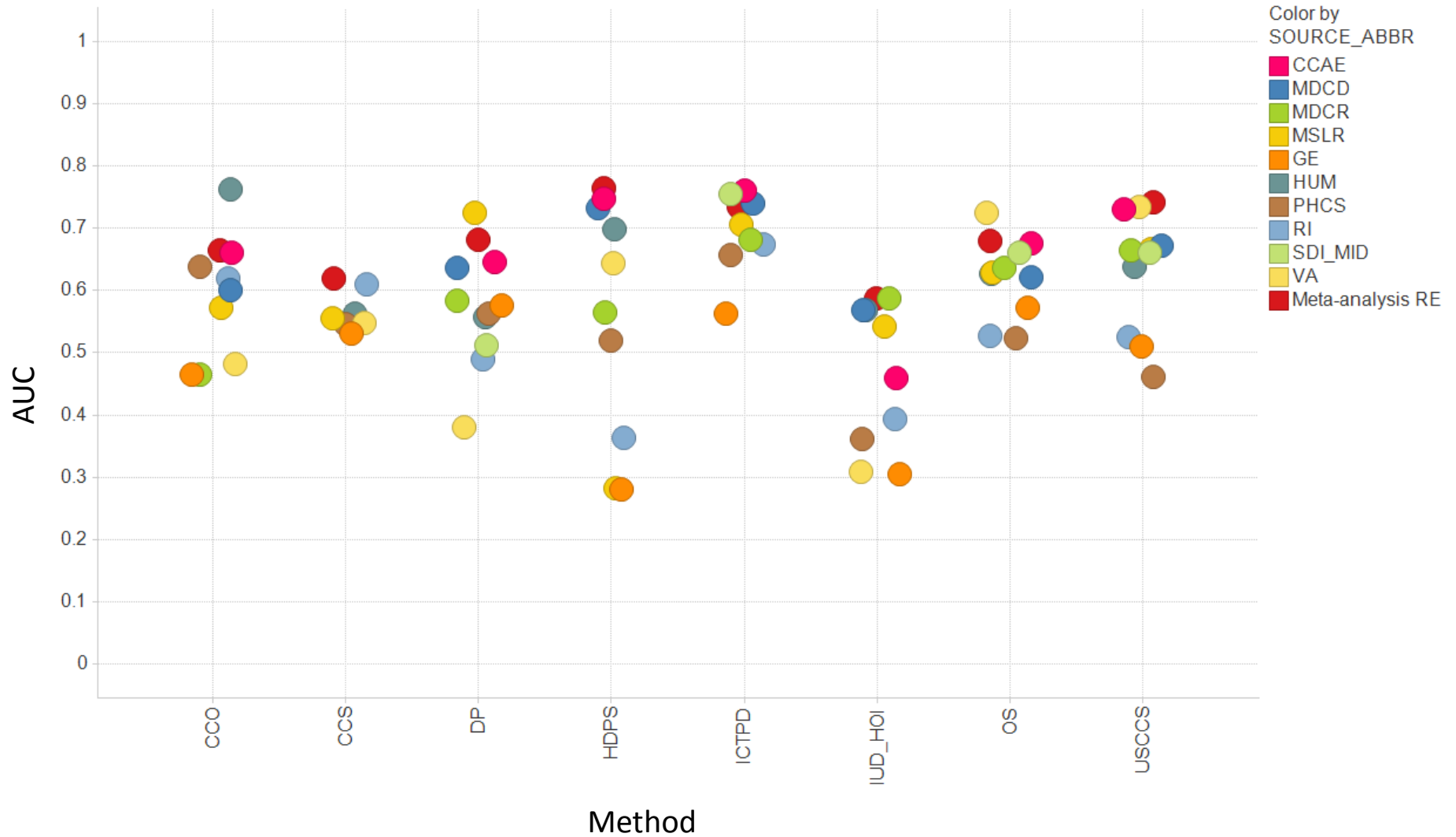
# AUC performance by method



# AUC performance by source



# AUC performance by method



# Tuning method parameters for data sources

Community optimal:

Same parameter settings used for all sources and all drug-outcome pairs

Selection: setting with maximum AUC from random-effects meta-analysis estimates

Source optimal:

Parameters tailored to each data source, but applied consistently

Selection: setting with maximum AUC within each source

**CAUTION:** all of these approaches result in model overfitting, so consider performance metrics to be optimistic upper bounds of what can be achieved

# AUC performance changes from community optimal to source optimal by method

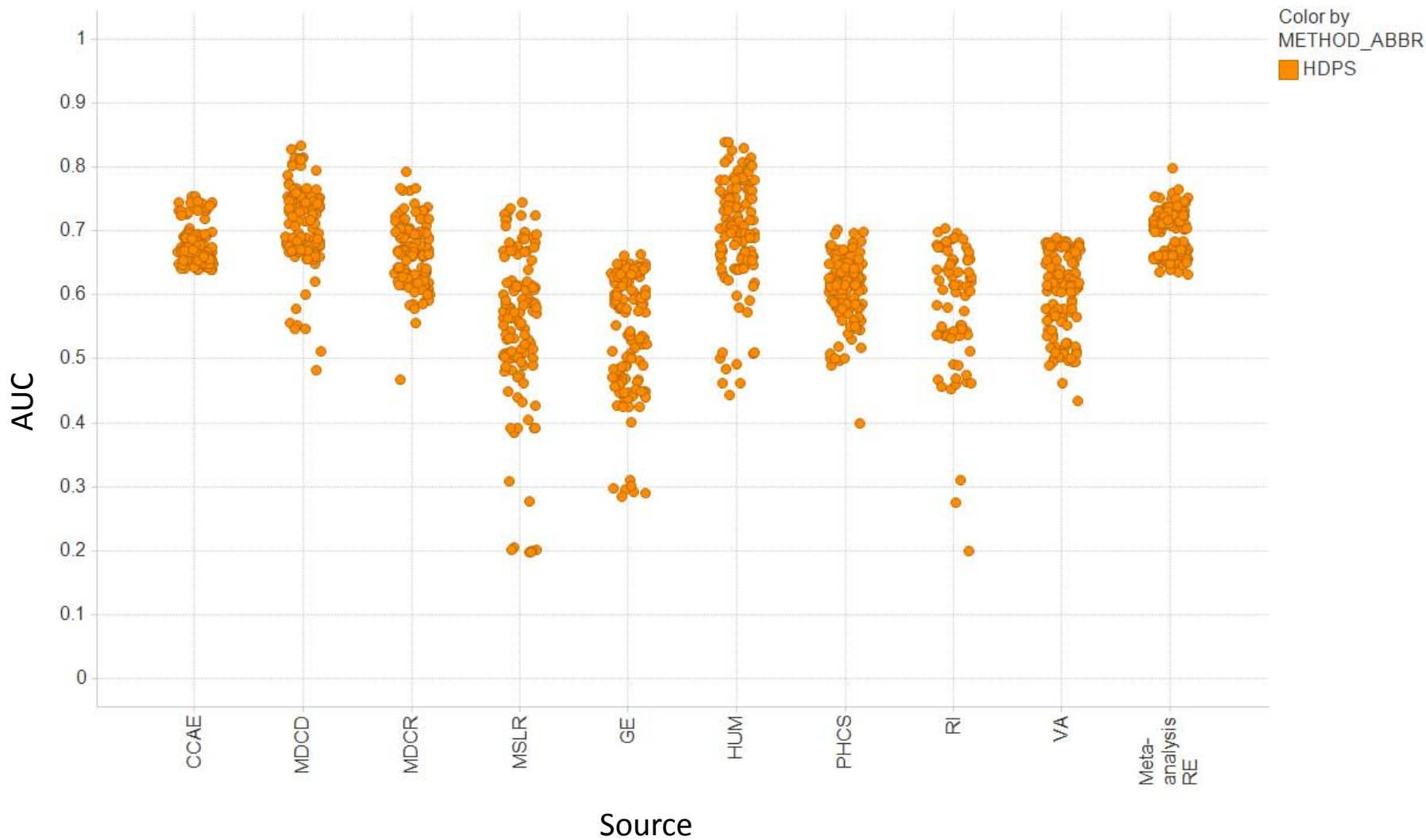


# Optimal parameters for HDPS vary by source

Source	AUC	Parameter settings					Analysis strategy	Comparator cohort
		Surveillance window (days from exposure start)	Covariate eligibility window (days prior to exposure)	# of confounders				
Meta-analysis RE	0.77	30d	30d	100		Strata (n=20) adjusted	Drug group	
HUM	0.84	All	All	100		Score adjusted	One drug	
RI	0.70	All	All	100		Strata (n=5) adjusted	Drug group	
PHCS	0.70	exposure + 30d	All	100		Strata (n=5) adjusted	Drug group	
VA	0.69	exposure + 30d	All	100		Strata (n=5) adjusted	One drug	
MSLR	0.74	30d	180d	500		Score adjusted	One drug	
MDCD	0.83	30d	30d	100		Strata (n=5) adjusted	Drug group	
MDCR	0.79	All	All	100		Score adjusted	Drug group	
CCAE	0.76	30d	180d	500		Strata (n=5) adjusted	Drug group	
GE	0.66	30d	180d	500		Mantel-Haenszel strata (n=5)	One drug	

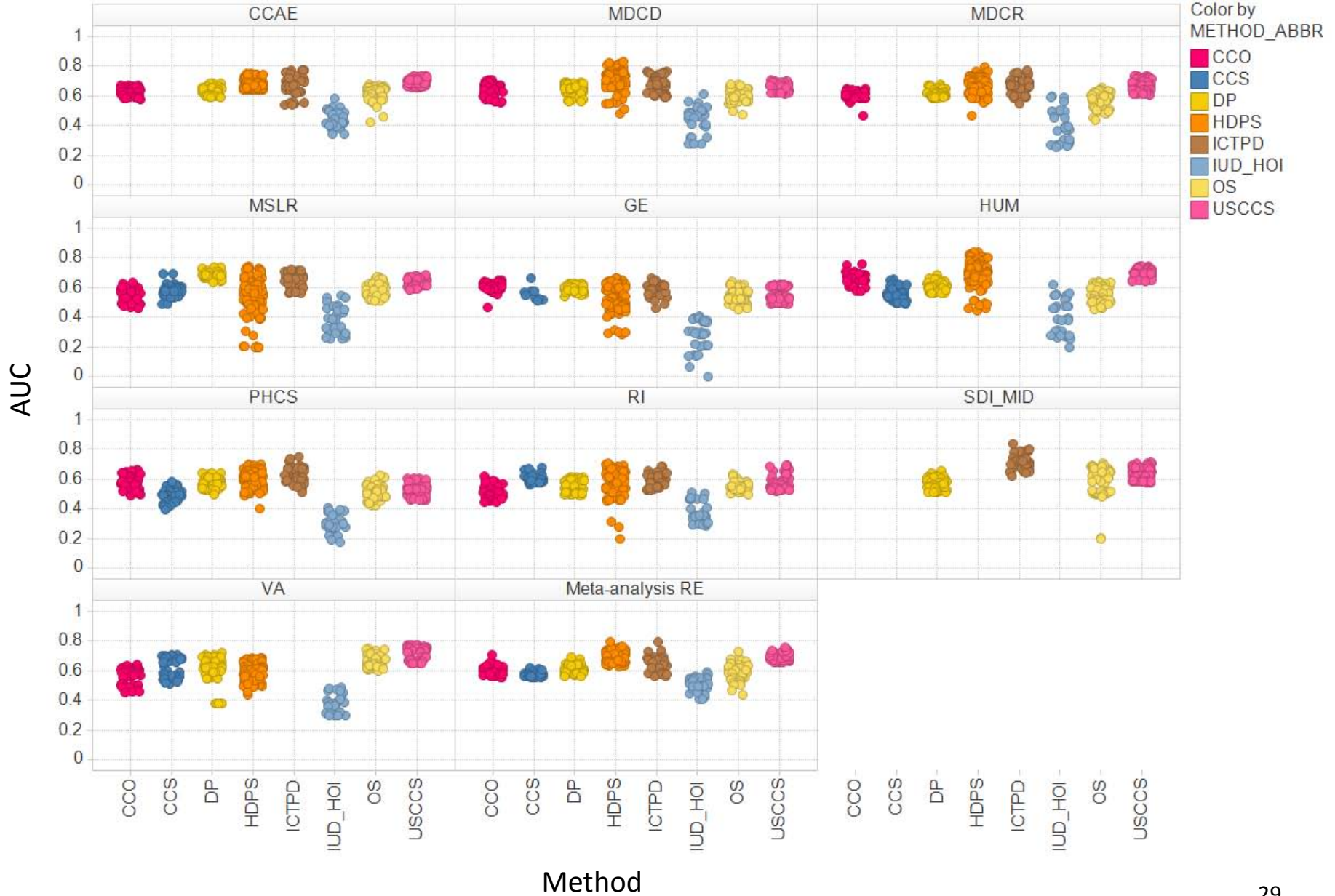
# AUC performance by source for all HDPS parameter combinations

OBSERVATIONAL  
MEDICAL  
OUTCOMES  
PARTNERSHIP



OBSERVATIONAL  
 MEDICAL  
 OUTCOMES  
 PARTNERSHIP

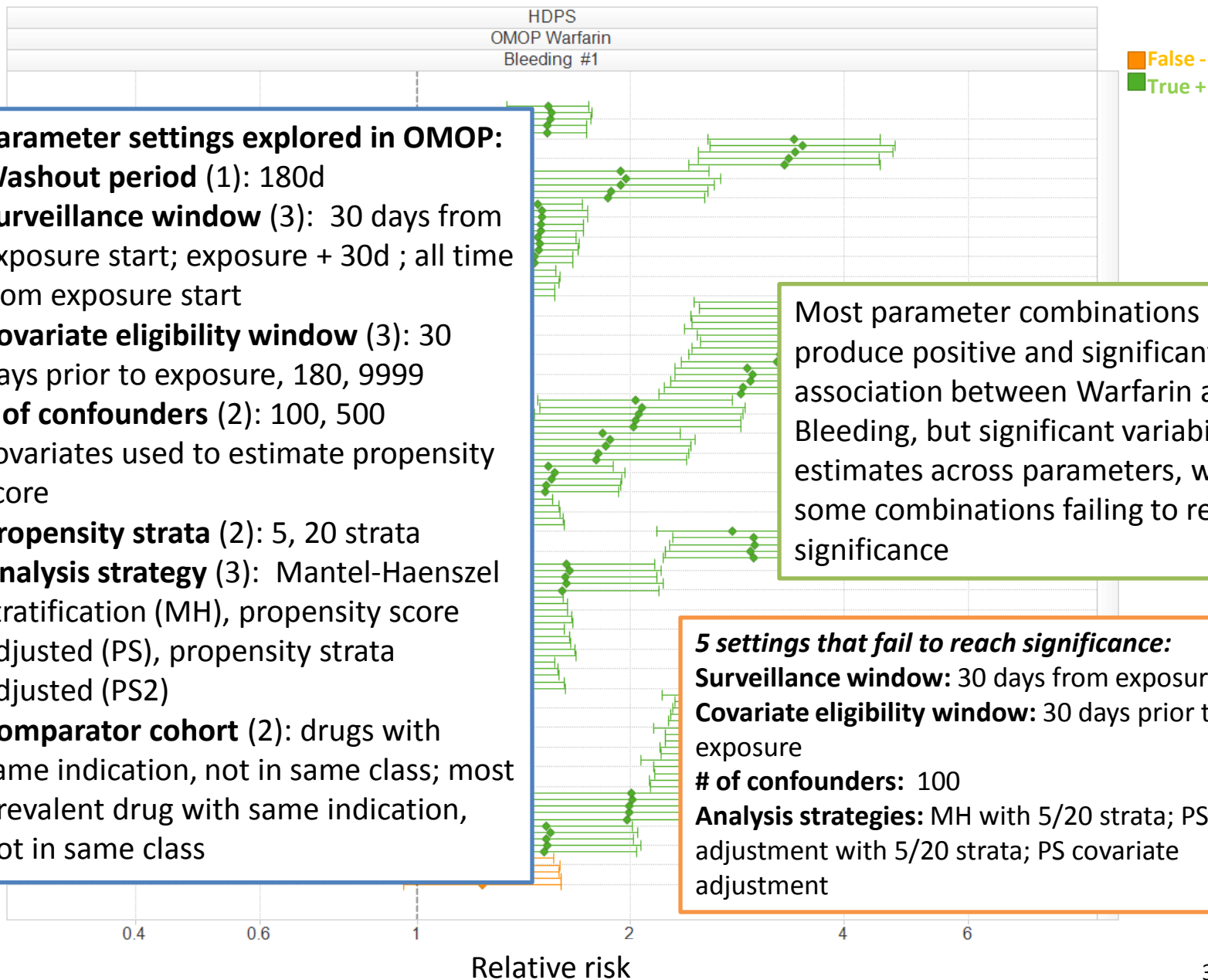
# AUC performance by method for all parameter combinations



## Additional questions warranting exploration

- Is performance differentiated by specific drugs or outcomes?
- How do alternative outcome definitions impact estimation?
- How does methods performance vary by data sources?
- **Can method parameter settings be further tailored to improve performance?**

# 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

**5 settings that fail to reach significance:**  
**Surveillance window:** 30 days from exposure start  
**Covariate eligibility window:** 30 days prior to exposure  
**# of confounders:** 100  
**Analysis strategies:** MH with 5/20 strata; PS adjustment with 5/20 strata; PS covariate adjustment

# Range of estimates across HDPS parameter settings

‘Consistent’ findings across parameters:

TP:

1. Benzodiazepine-Hip fracture

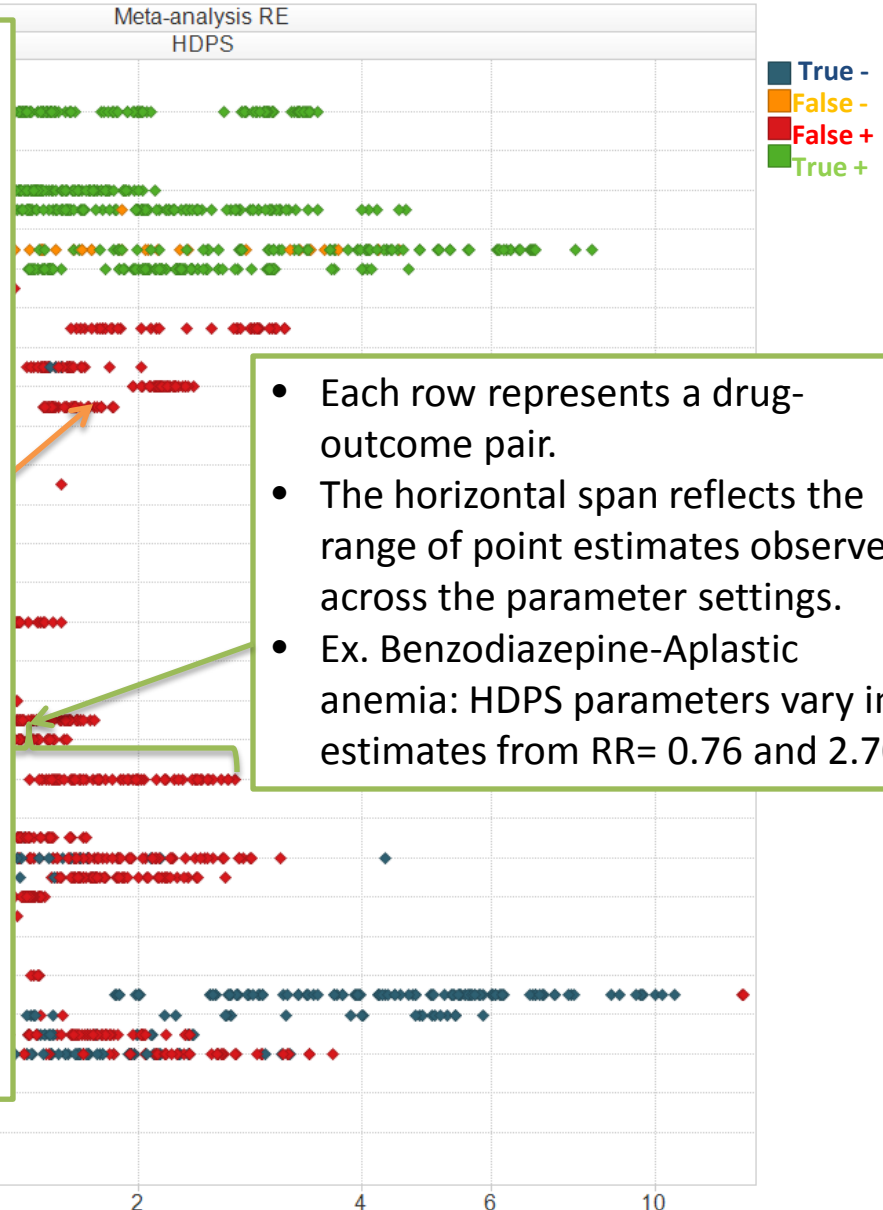
FN:

1. Tricyclic antidepressants-Acute myocardial infarction
2. Bisphosphonates-GI Ulcer hospitalization
3. Antibiotics-Acute liver injury

TN:

1. ACE inhibitors - Aplastic anemia
2. Amphotericin B - Angioedema
3. Antiepileptics - GI Ulcer hospitalization
4. Beta blockers - Acute liver injury
5. Beta blockers – Angioedema
6. Beta blockers – Acute renal failure
7. Beta blockers – GI Ulcer hospitalization
8. Beta blockers – Aplastic anemia
9. Benzodiazepines- Angioedema
10. Warfarin-Mortality after MI
11. Warfarin-Aplastic anemia
12. Bisphosphonates – Acute myocardial infarction

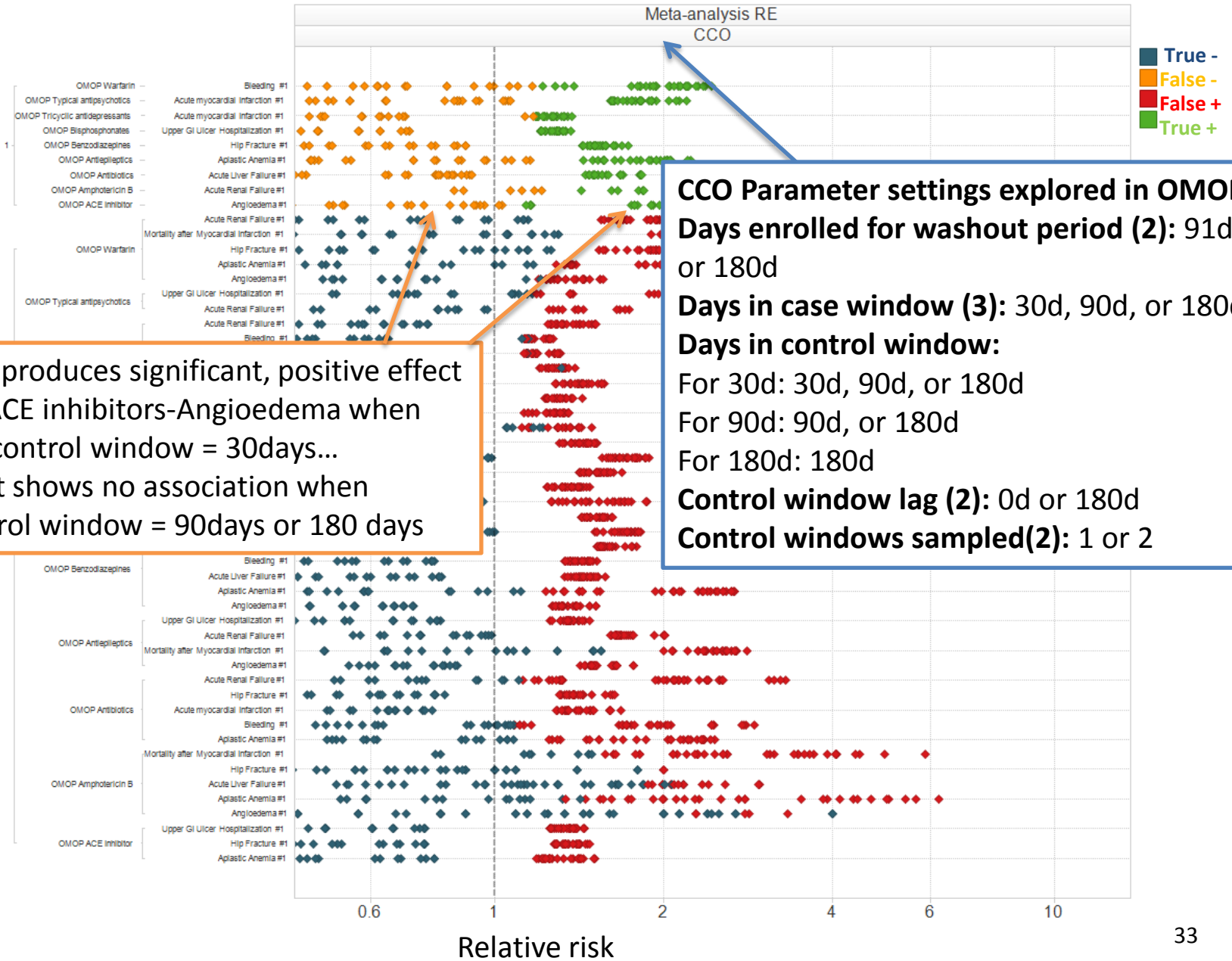
‘Inconsistent’ estimates: 37



- Each row represents a drug-outcome pair.
- The horizontal span reflects the range of point estimates observed across the parameter settings.
- Ex. Benzodiazepine-Aplastic anemia: HDPS parameters vary in estimates from RR= 0.76 and 2.70



# Range of estimates across CCO parameter settings



- CCO produces significant, positive effect for ACE inhibitors-Angioedema when the control window = 30days...
- ...but shows no association when control window = 90days or 180 days

**CCO Parameter settings explored in OMOP:**

**Days enrolled for washout period (2):** 91d, or 180d

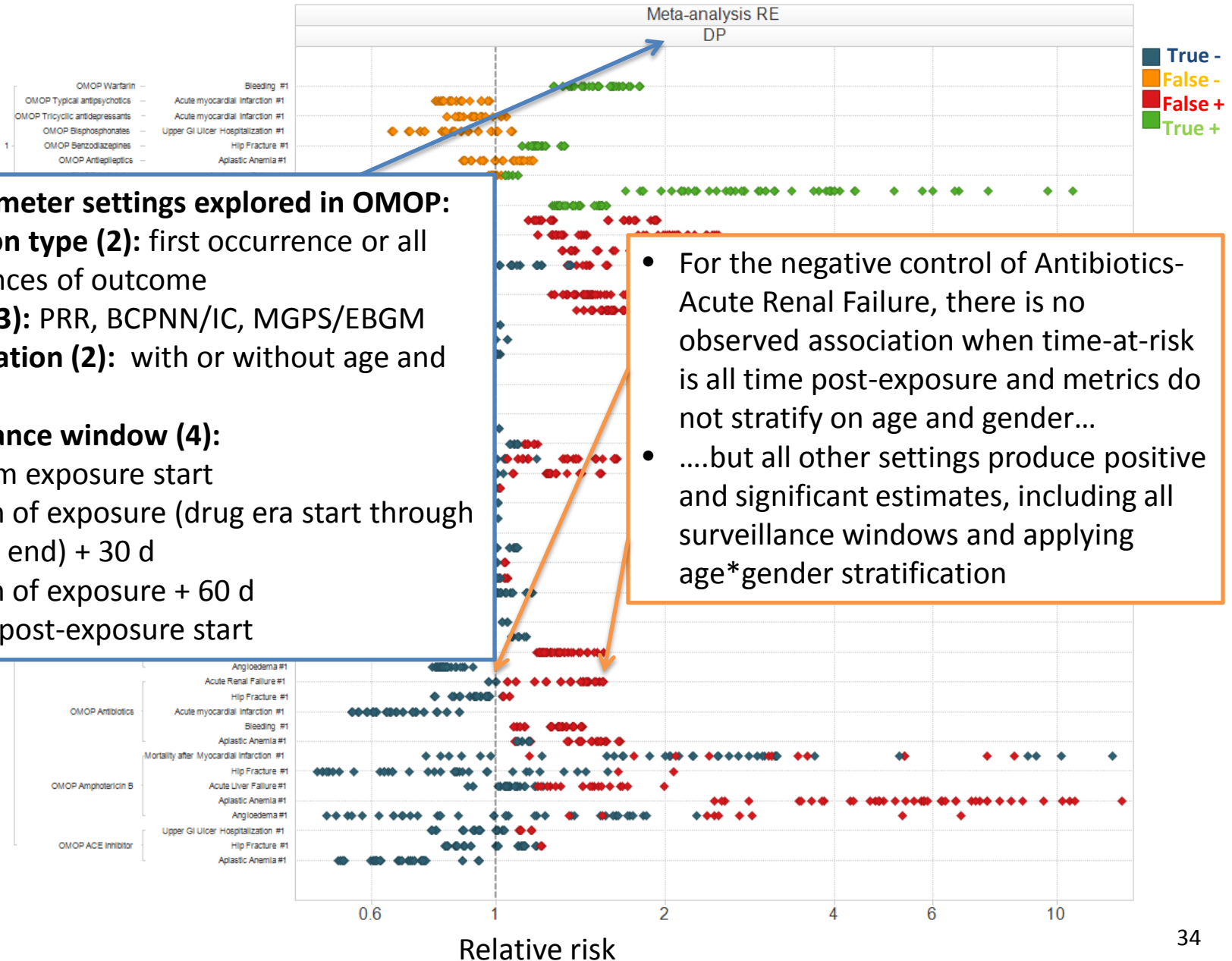
**Days in case window (3):** 30d, 90d, or 180d

**Days in control window:**  
For 30d: 30d, 90d, or 180d  
For 90d: 90d, or 180d  
For 180d: 180d

**Control window lag (2):** 0d or 180d

**Control windows sampled(2):** 1 or 2

# Range of estimates across DP parameter settings



# Range of estimates across ICTPD parameter settings

**ICTPD Parameter settings explored in OMOP:**

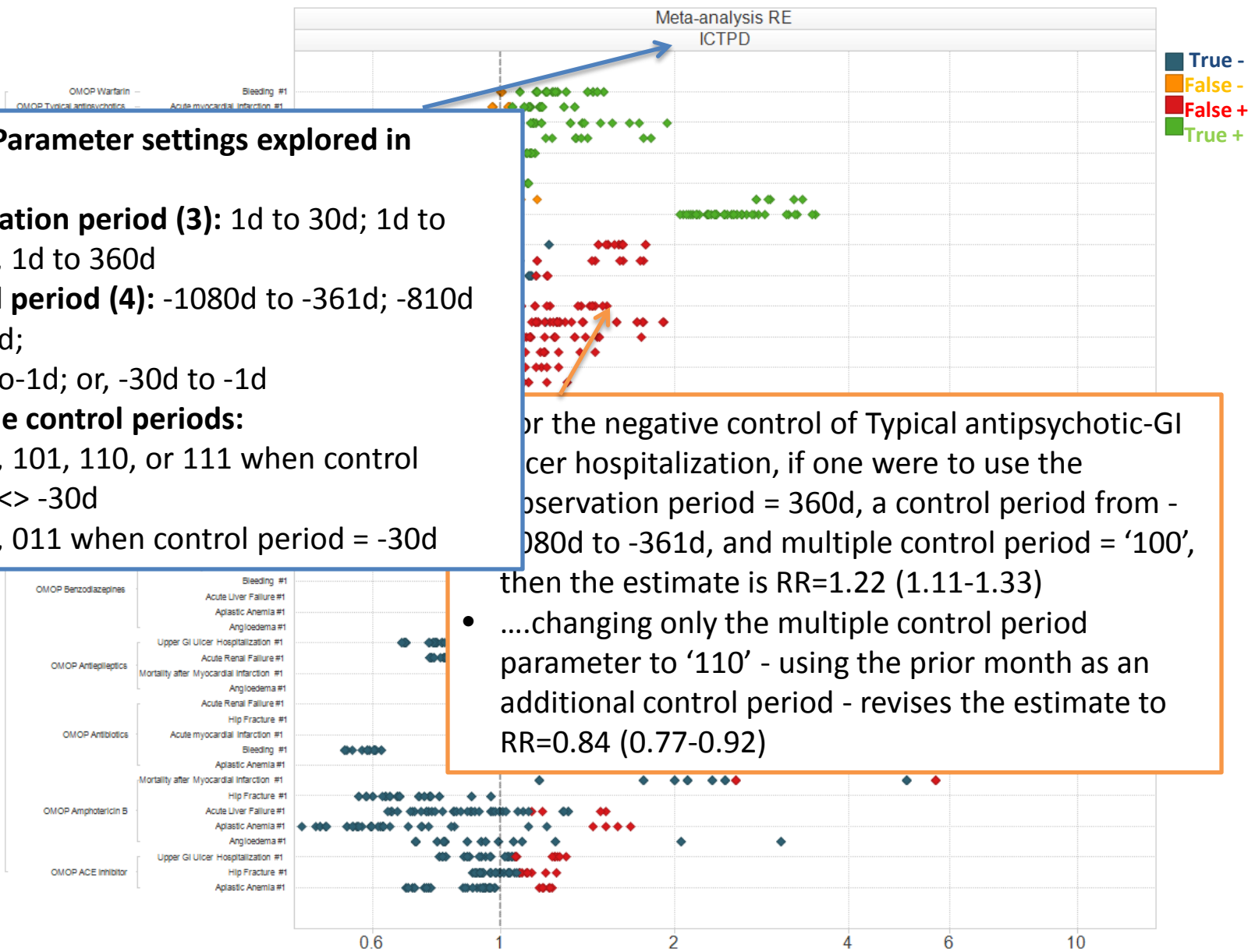
**Observation period (3):** 1d to 30d; 1d to 60d; or, 1d to 360d

**Control period (4):** -1080d to -361d; -810d to -361d; -180d to -1d; or, -30d to -1d

**Multiple control periods:**

(4) 100, 101, 110, or 111 when control period  $\neq$  -30d

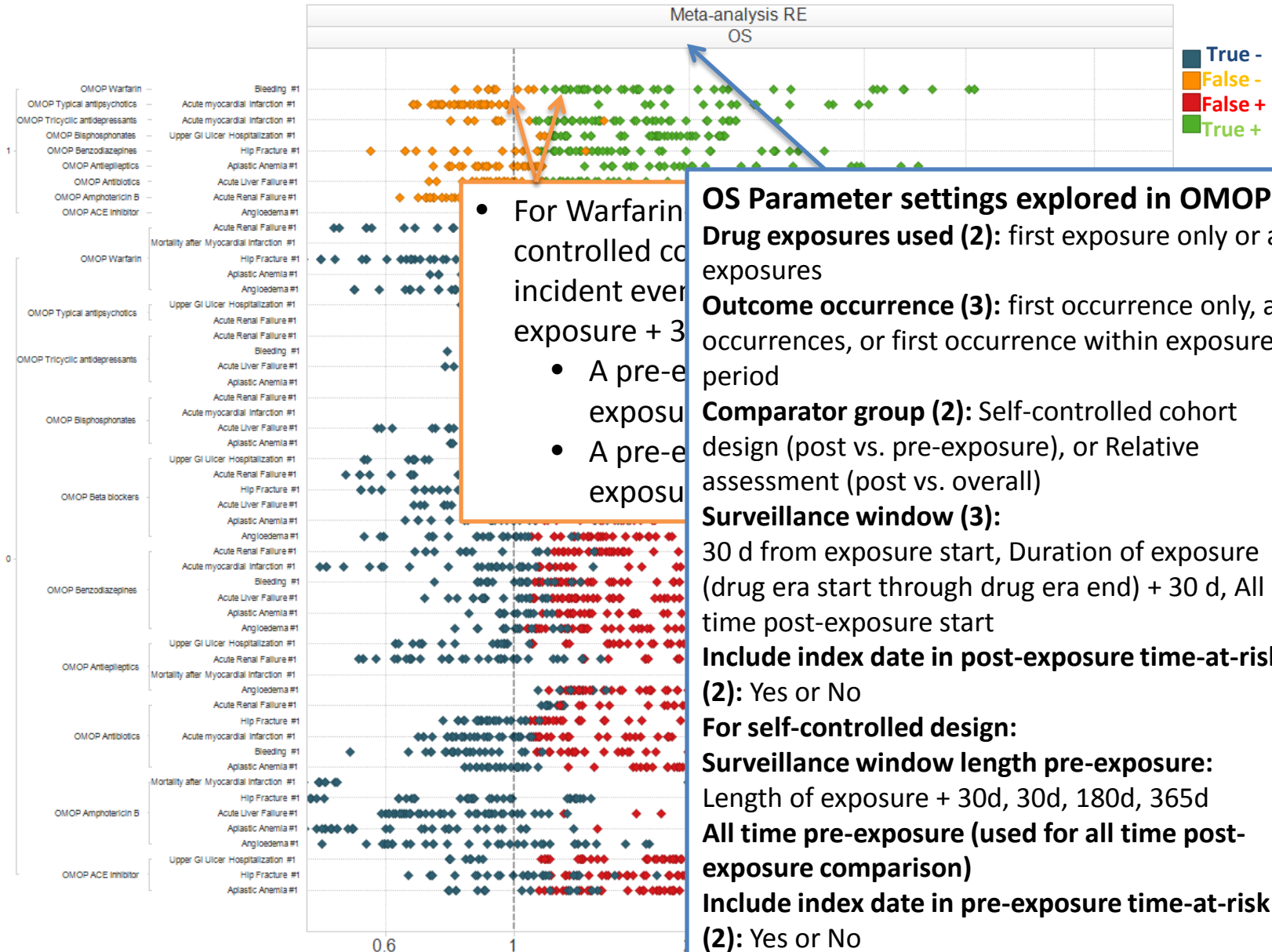
(2) 010, 011 when control period = -30d



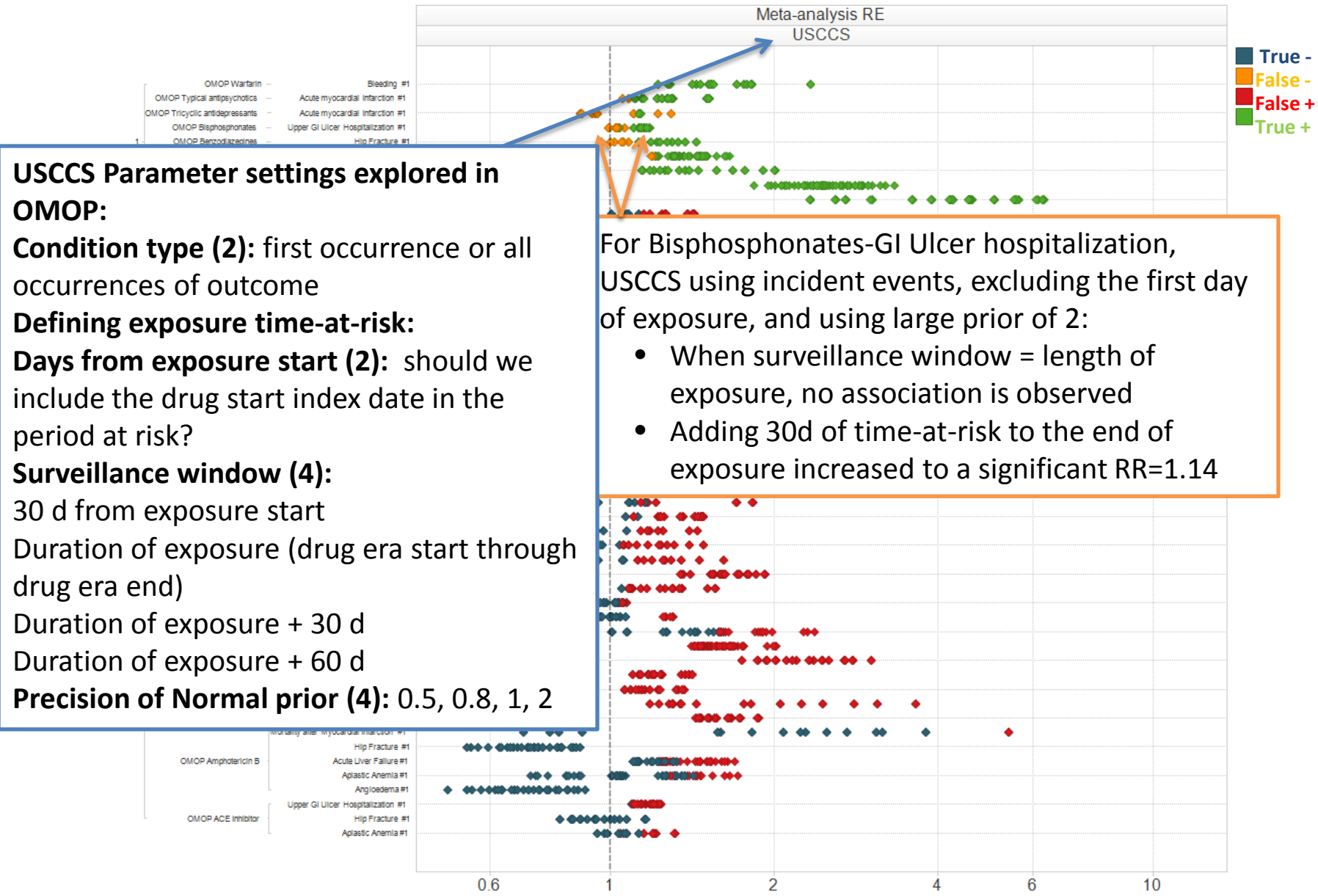
...for the negative control of Typical antipsychotic-GI ulcer hospitalization, if one were to use the observation period = 360d, a control period from -1080d to -361d, and multiple control period = '100', then the estimate is RR=1.22 (1.11-1.33)

- ...changing only the multiple control period parameter to '110' - using the prior month as an additional control period - revises the estimate to RR=0.84 (0.77-0.92)

# Range of estimates across OS parameter settings



# Range of estimates across 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

# Method consistency across parameters

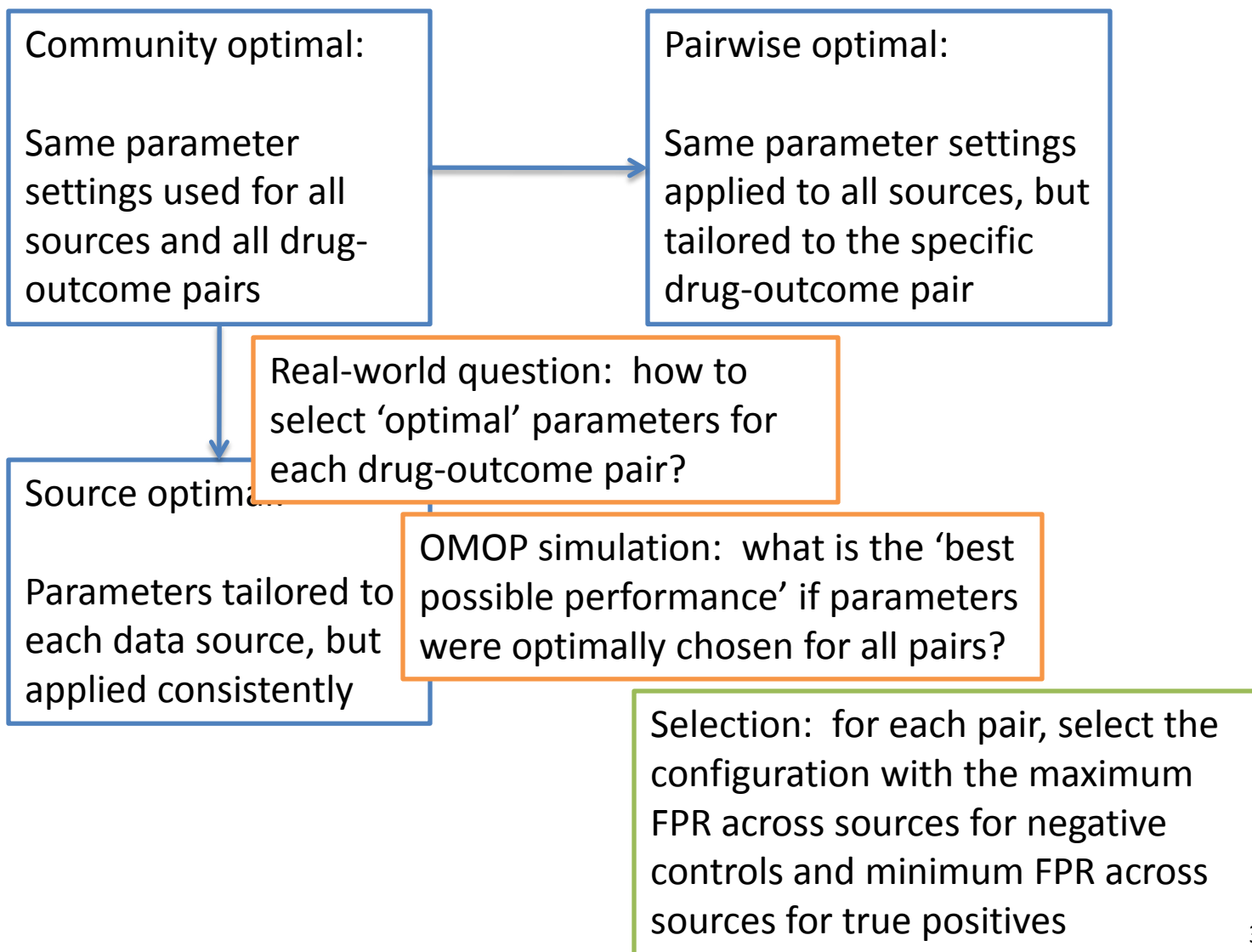
Among the pairs for which consistent estimates are observed across all parameter settings, what fraction of the cases are the consistent estimates correct?

Method consistency across parameter settings

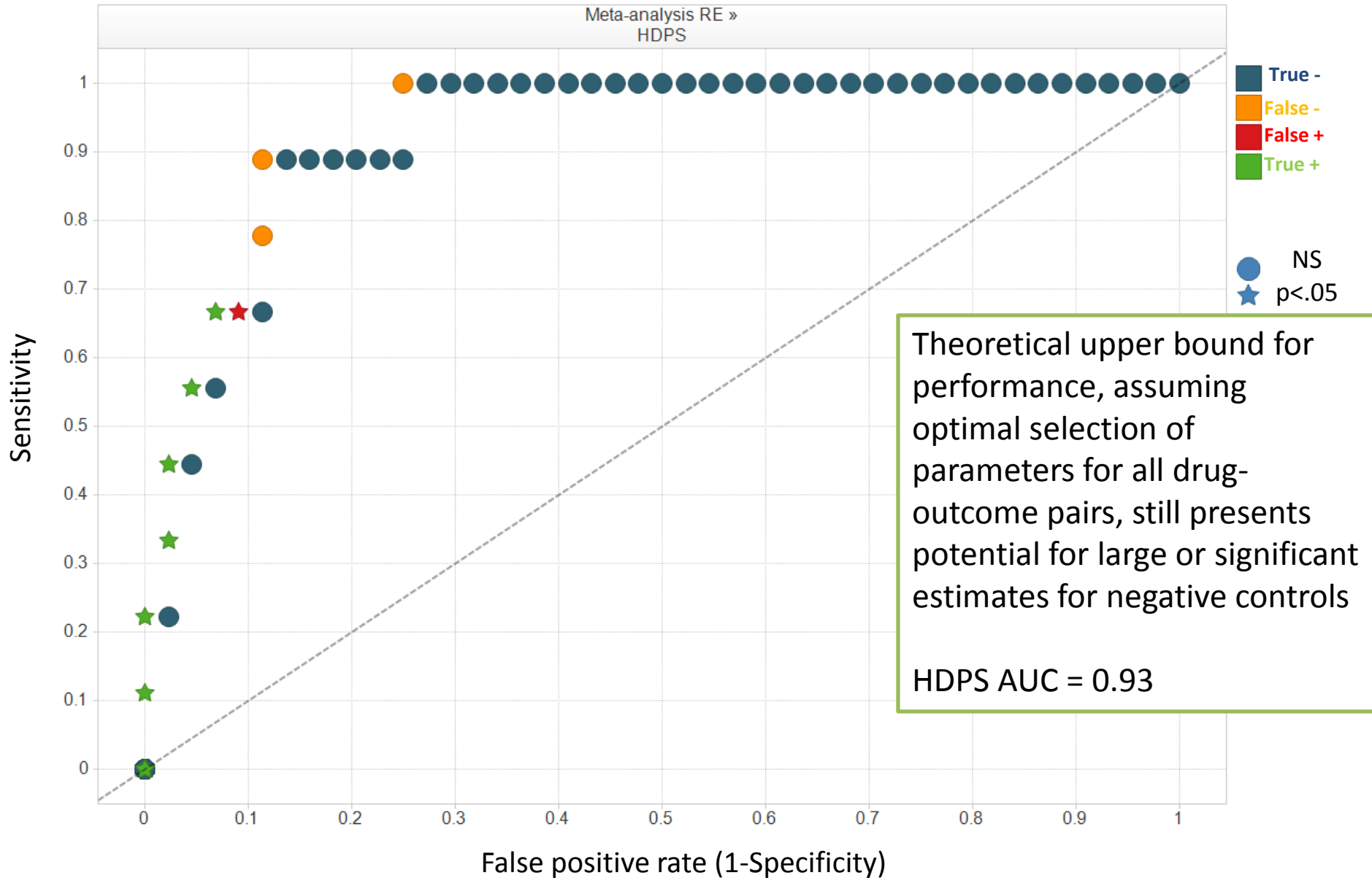
Method	Consistent				Total	
	TP	FP	FN	TN	Consistent	Inconsistent
CCS	8	32	0	1	41	12
DP	4	8	4	20	36	17
USCCS	4	9	0	5	18	35
HDPS	1	0	3	12	16	37
IUD_HOI	1	2	0	13	16	31
ICTPD	1	0	1	9	11	42
OS	1	0	0	1	2	51
CCO	0	0	0	0	0	53

Most drug-outcome pair estimates are sensitive to method parameter settings such that both significance/insignificance can be observed

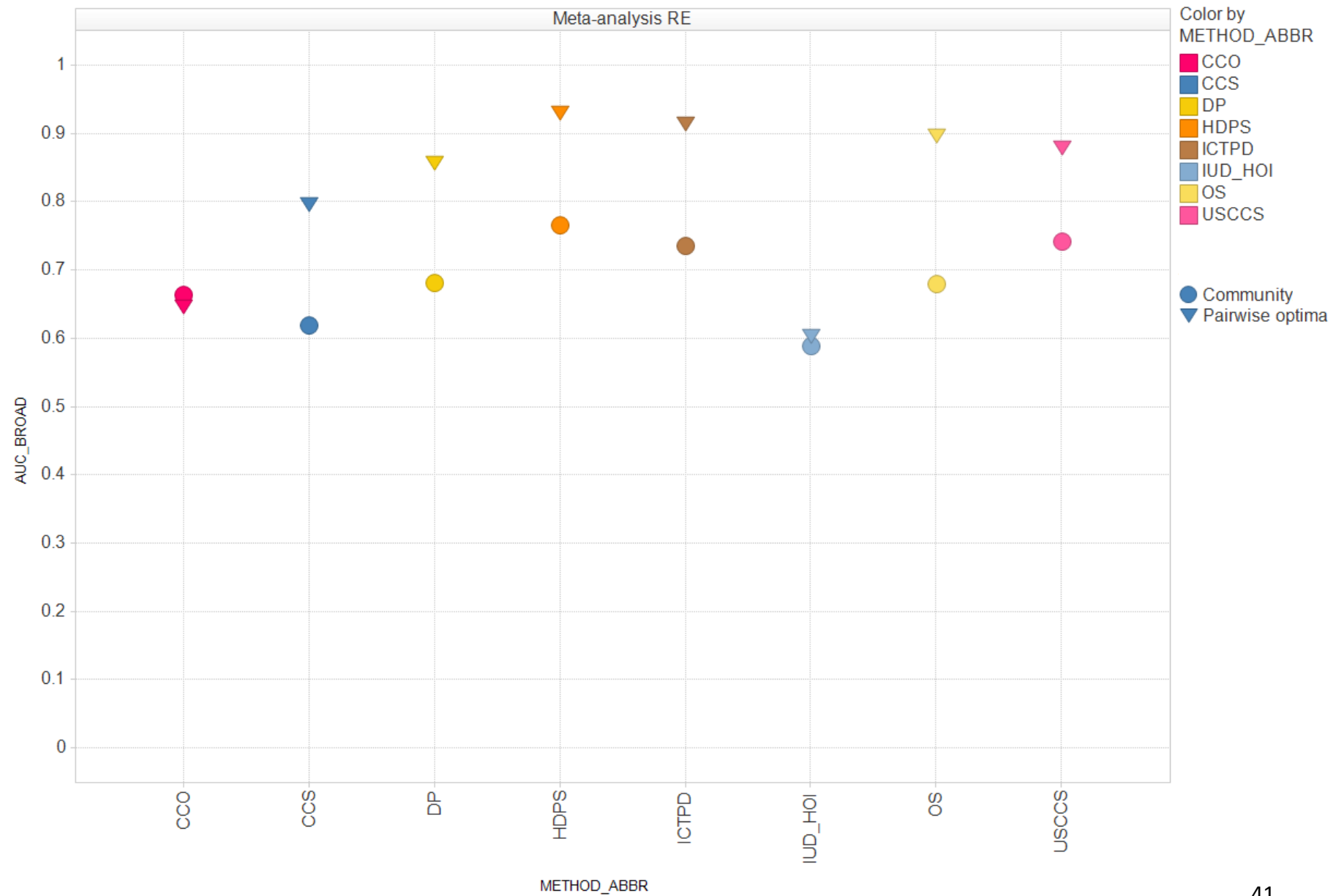
# Tuning method parameters for drug-outcome pairs



# ROC for HDPS pairwise optimal on meta-analysis estimates



# Upper bound of potential AUC performance based on pairwise optimal



## Concluding thoughts

- Method performance can vary by data source, drug, and outcome
- Method estimates are sensitive to outcome definitions and parameter settings
- Need to be cautious in interpreting results from single method in single database
  - Replication does not necessarily provide complete confidence
- Need to develop strategies for principled parameter selection and implement comprehensive sensitivity analyses for evaluating the robustness of any findings across:
  - Data source and target populations
  - Method and parameter settings
  - Outcome and exposure definitions
- Additional research across a broader array of test cases is needed to fully characterize expected method behavior to improve confidence in the results that are obtained

## Panel Discussion: Method Performance Results from the Health Outcomes of Interest Experiment

- **Paul Stang**, PhD, Senior Director, Epidemiology, Johnson & Johnson Pharmaceutical Research and Development, OMOP Research Investigator
- **Shawn Murphy** MD, PhD, Medical Director of Research Computing and Informatics, Partners Healthcare Research Computing; Assistant Professor, Harvard Medical School
- **Niklas Norén**, PhD, Manager, Research Department, Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring
- **Ram Tiwari**, PhD, Associate Director, Office of Biostatistics, CDER/FDA
- **Judy A. Staffa**, PhD, R.Ph, FDA/CDER/OSE, Associate Director for Regulatory Research

## Panel Discussion: Lessons Learned from Systematic Observational Analysis

- **Marc Overhage**, MD, PhD, Director, Medical Informatics and Research Scientist, Regenstrief Institute, Inc.; Regenstrief Professor of Medical Informatics, Indiana University School of Medicine CEO and President Indiana Health Information Exchange; OMOP Research Investigator
- **Joshua S. Benner**, Pharm.D., Sc.D. Research Director, Engelberg Center for Health Care Reform, Brookings Institution.
- **Abraham G. Hartzema** PharmD, MSPH, PhD, FISPE, Professor and Eminent Scholar, Perry A. Foote Chair in Health Outcomes and Pharmacoeconomics; Professor, Department of Epidemiology and Biostatistics, College of Public Health and Health Professions, University of Florida; OMOP Research Investigator
- **William D. Marder**, PhD, Senior Vice President and General Manager, Healthcare & Science Thomson Reuters
- **Richard Platt**, MD, MSc, Professor and Chair of the Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care
- **Mitra Rocca**, Associate Director, Medical Informatics, FDA

# Question and Answer Session