

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Empirical Performance of Large-scale Analysis Methods for Active Surveillance: Lessons from the Observational Medical Outcomes Partnership

Patrick Ryan
on behalf of OMOP Research Team
August 25, 2011

Full results and audio presentations from OMOP Symposium available at:
<http://omop.fnih.org/OMOP2011Symposium>

FDAAA call for establishing the Risk Identification and Analysis System

SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.

(a) IN GENERAL.—Subsection (k) of section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

“(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—

“(A) DEFINITION.—In this paragraph, the term ‘data’ refers to information with respect to a drug approved under this section or under section 351 of the Public Health Service Act, including claims data, patient survey data, standardized analytic files that allow for the pooling and analysis of data from disparate data environments, and any other data deemed appropriate by the Secretary.

“(B) DEVELOPMENT OF POSTMARKET RISK IDENTIFICATION AND ANALYSIS METHODS.—The Secretary shall, not later than 2 years after the date of the enactment of the Food and Drug Administration Amendments Act of 2007, in collaboration with public, academic, and private entities—

“(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);

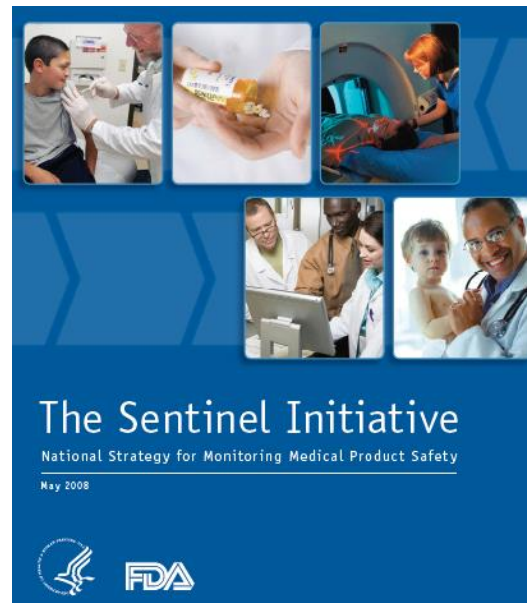
“(ii) develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

“(I) at least 25,000,000 patients by July 1, 2010; and

“(II) at least 100,000,000 patients by July 1, 2012; and

“(iii) convene a committee of experts, including individuals who are recognized in the field of protecting data privacy and security, to make recommendations to the Secretary on the development of tools and methods for the ethical and scientific uses for, and communication of, postmarketing data specified under subparagraph (C), including recommendations on the development of effective research methods for the study of drug safety questions.

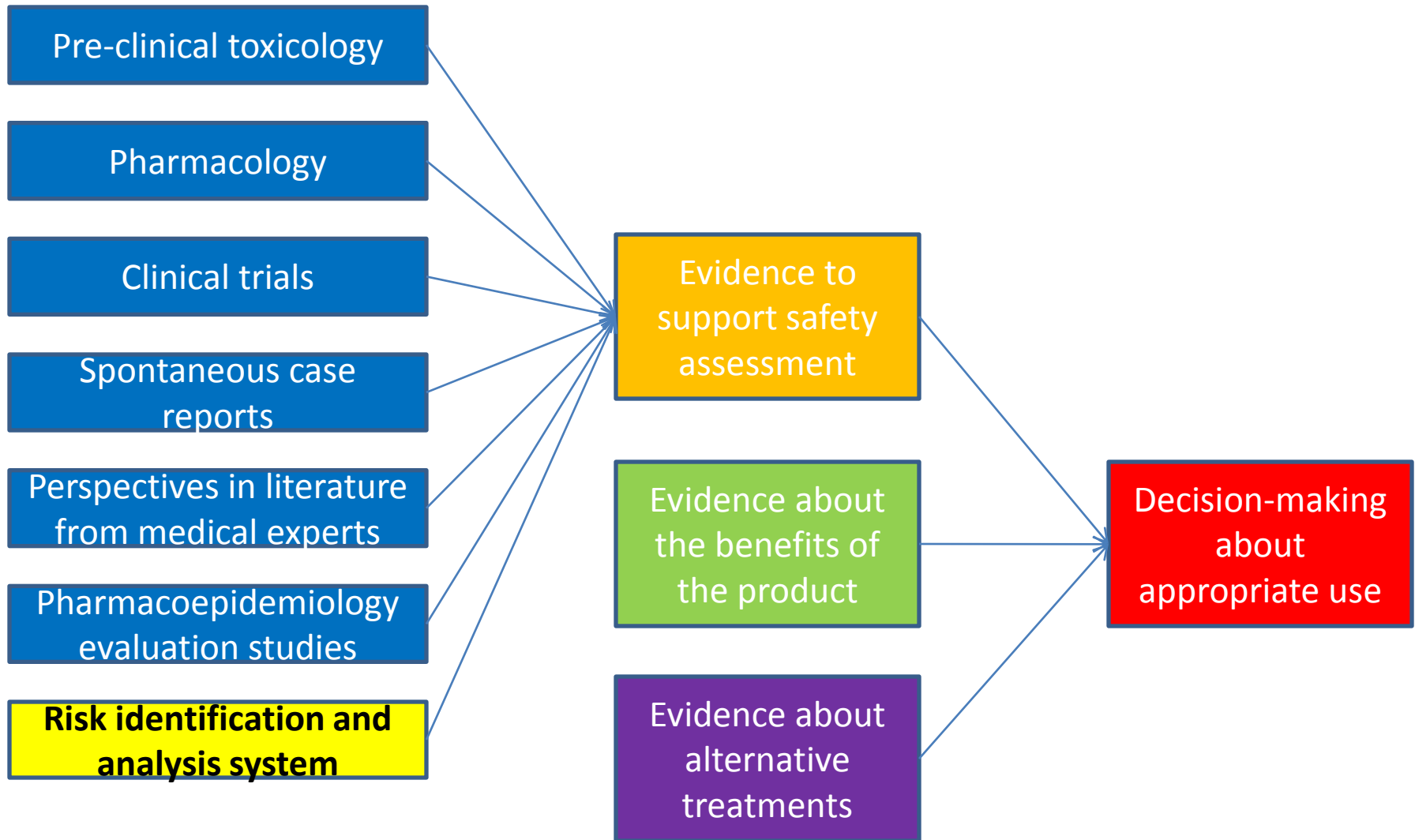
“(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTIFICATION AND ANALYSIS SYSTEM.—



Risk Identification and Analysis System:

a systematic and reproducible process to efficiently generate evidence to support the characterization of the potential effects of medical products from across a network of disparate observational healthcare data sources

Risk identification and analysis system: One additional piece of evidence to inform medical decision-making



Outstanding Questions For Active Surveillance

Governance

What are the keys to a successful public-private partnership?

- Transparency
- Stakeholder Access
- Sound Science

Data

Which types of data? administrative claims and/or electronic health records

Which sources? healthcare providers, insurers, data aggregators

What are viable data access models:

- centralized?
- distributed?

Performance

Architecture

Feasibility

What are appropriate analyses for:

- hypothesis generating?
- hypothesis strengthening?

What is the appropriate infrastructure:

- hardware?
- software?
- processes?
- policies?

How to maintain collaborations and engage research community?

Methods

Technology

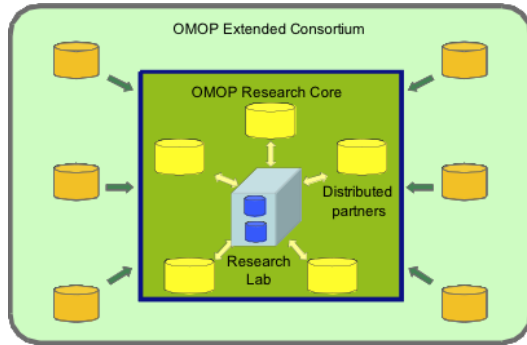
What are best practices for protecting data?

Observational Medical Outcomes Partnership

Public-Private Research Partnership established to inform the appropriate use of observational healthcare databases for studying the effects of medical products:

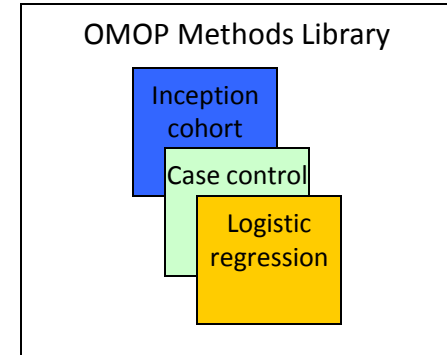
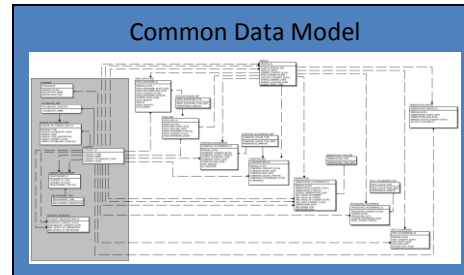
- Conducting methodological research to empirically evaluate the performance of alternative methods on their ability to identify true associations
- Developing tools and capabilities for transforming, characterizing, and analyzing disparate data sources across the health care delivery spectrum
- Establishing a shared resource so that the broader research community can collaboratively advance the science

OMOP Research Experiment



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

- Open-source
- Standards-based



- 14 methods
- Epidemiology designs
- Statistical approaches adapted for longitudinal data

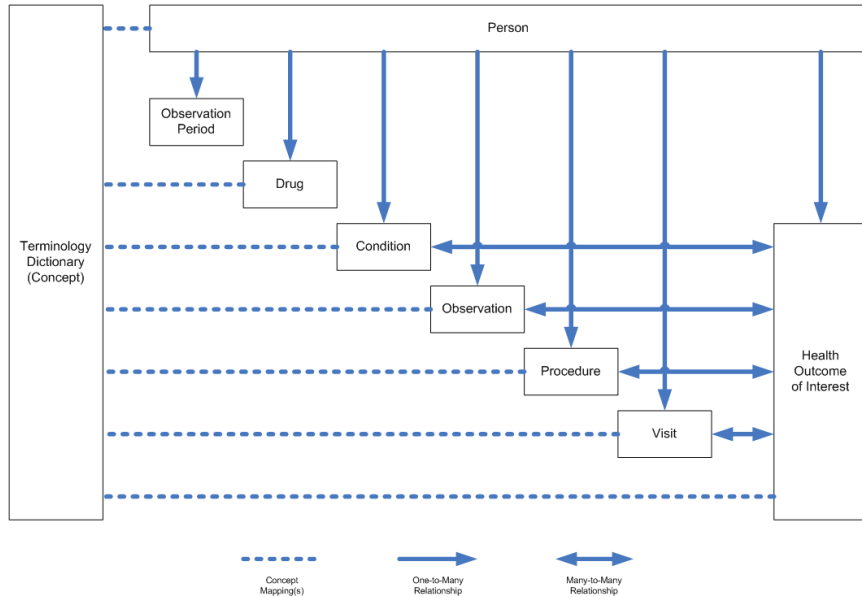


Outcome	ACE Inhibitors	Amphotericin B	Antibiotics: erythromycins, sulfonamides, tetracyclines	Antiepileptics: carbamazepine, phenytoin	Benzodiazepines	Beta blockers	Bisphosphonates: alendronate	Tricyclic antidepressants	Typical antipsychotics	Warfarin
Angioedema	Red	Blue	Blue	Blue	Blue	Blue	Blue	Blue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

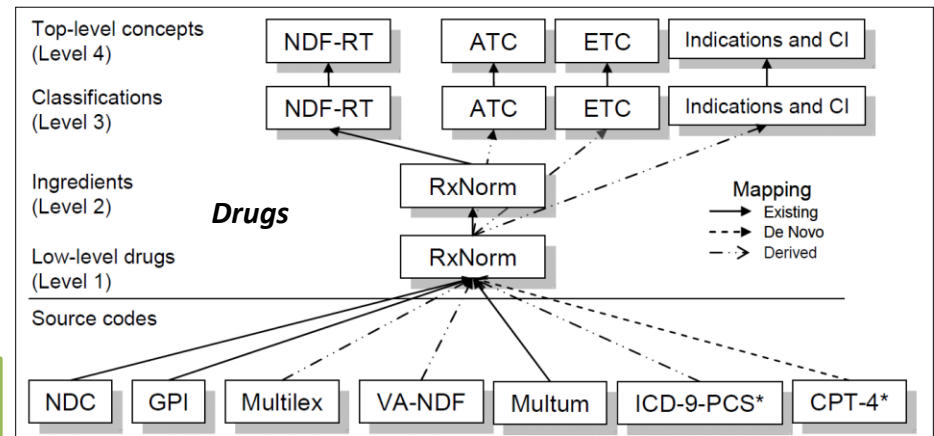
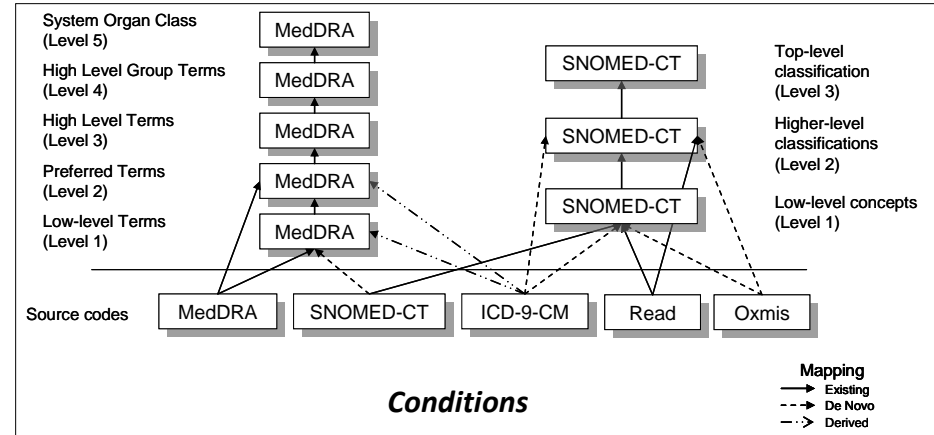
Common Framework

Accommodating Disparate Observational Data Sources

Common Data Model

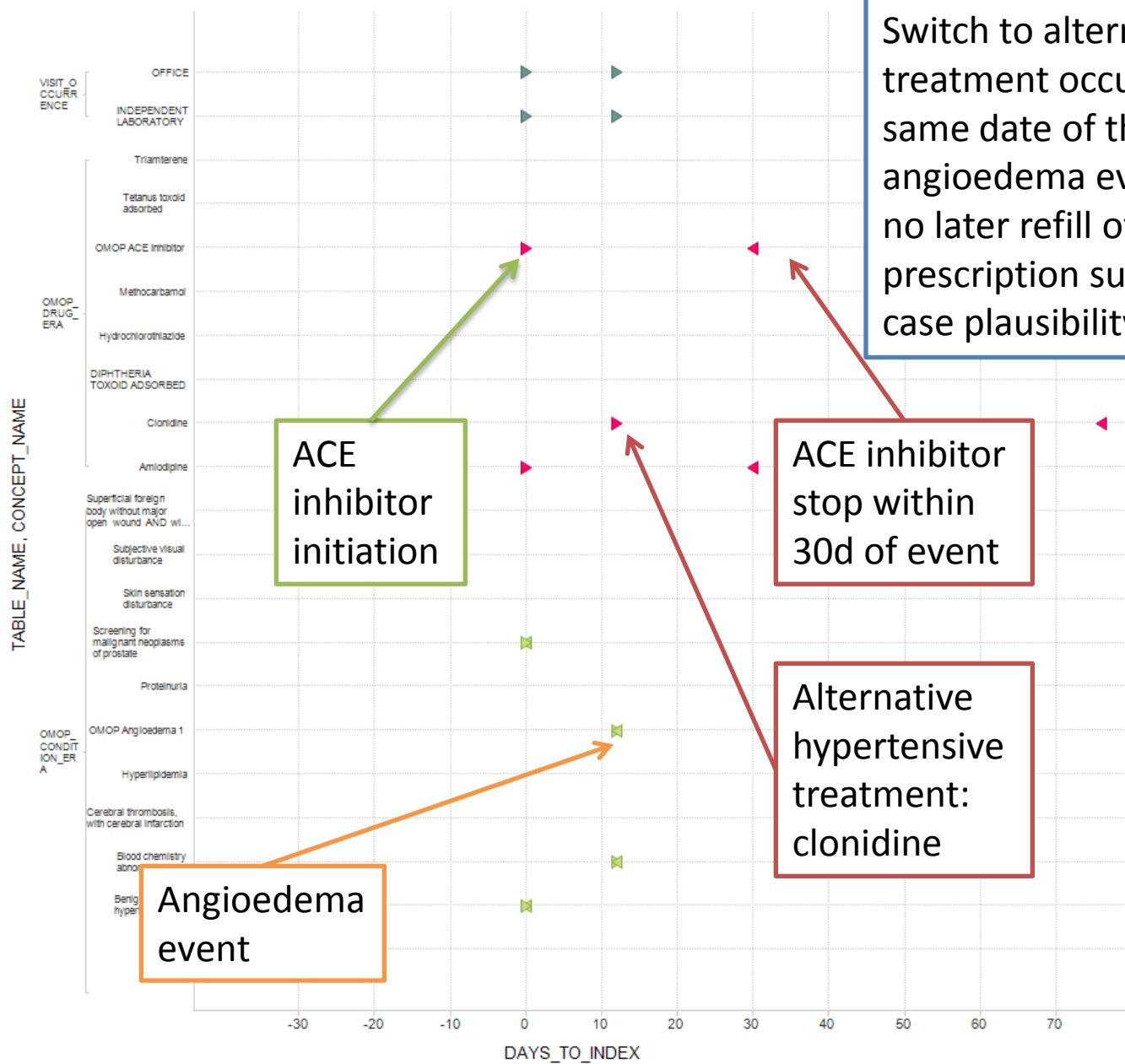


Standardized Terminologies



Mapping includes Vaccines coded in NDC, CPT4, HCPCS, ICD9, standardized to RxNorm

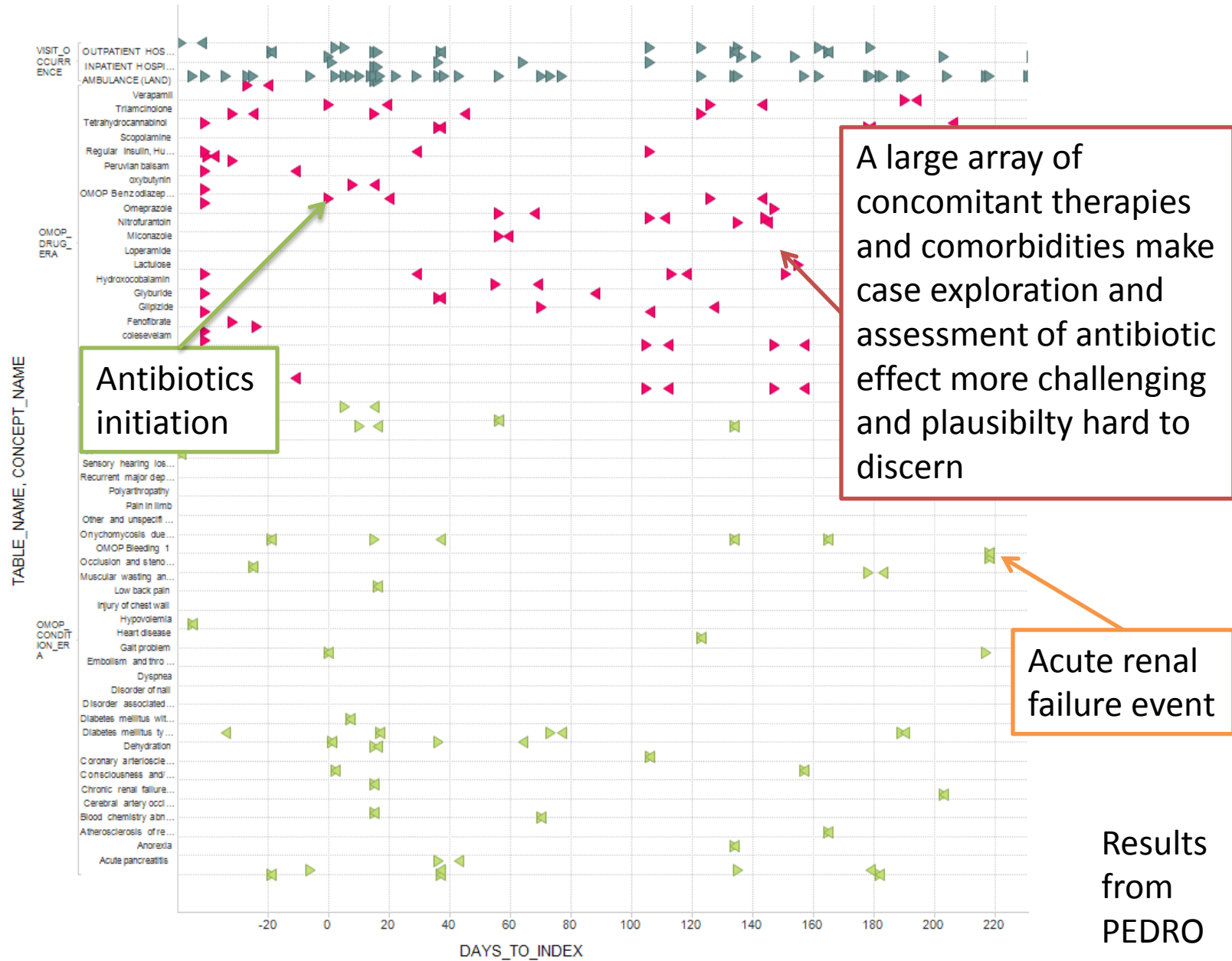
Plausibility: Ex 1: ACE inhibitor- Angioedema case 1



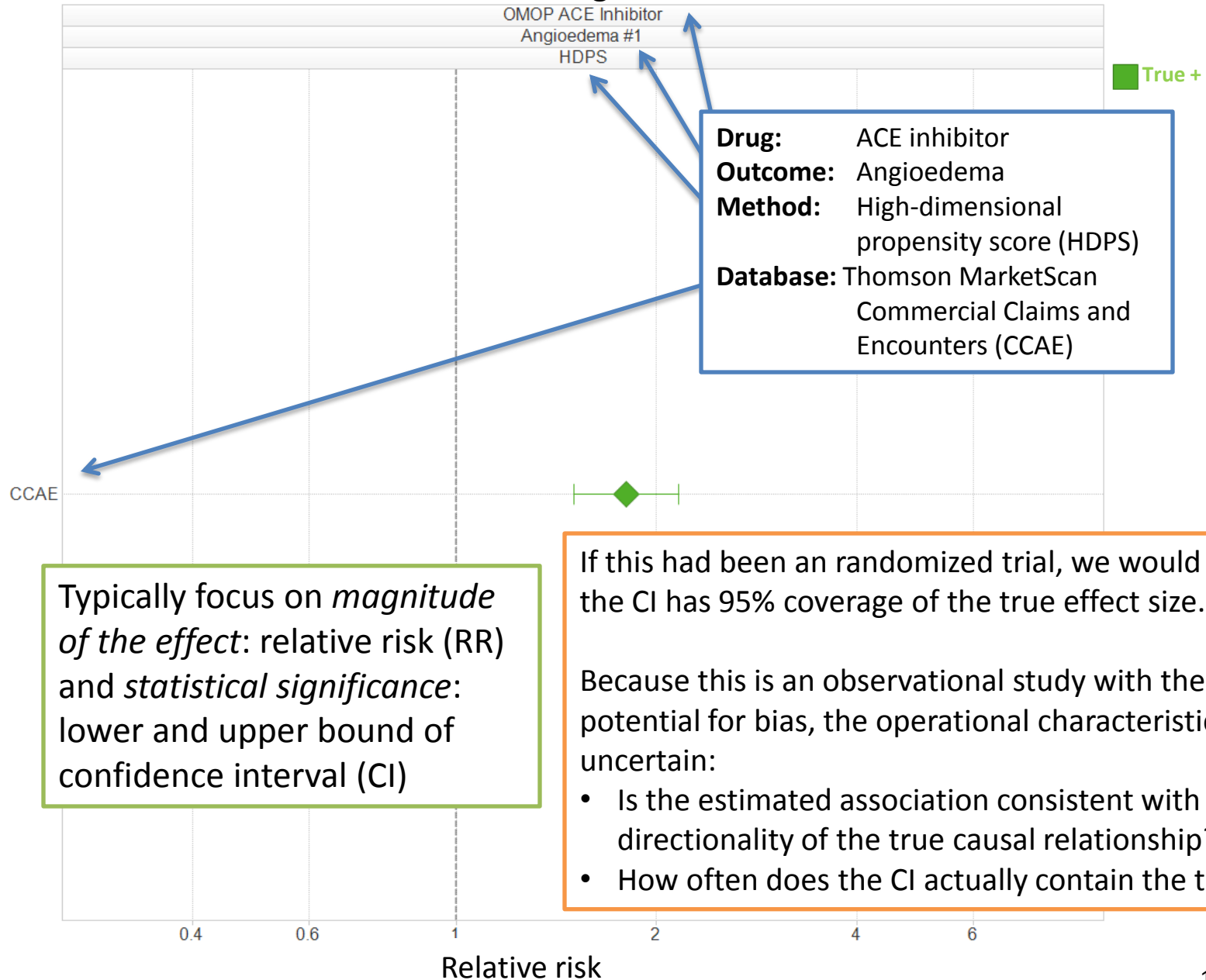
X-axis: Time, relative to index exposure of ACE inhibitor

Y-axis: all elements captured in database (visits, conditions, drugs)

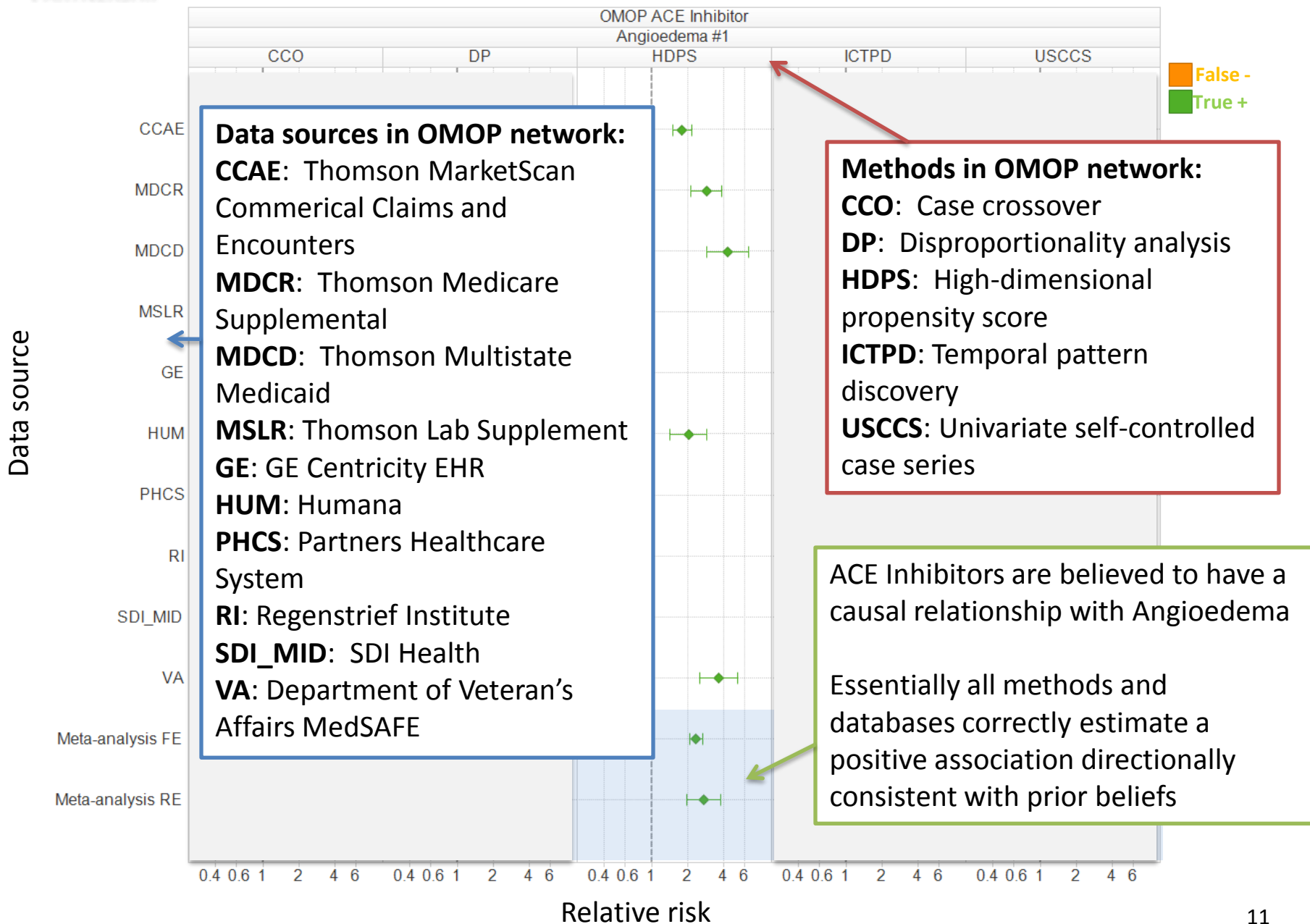
Plausibility: Ex 1: Antibiotics – Acute Renal Failure #4 case 1



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

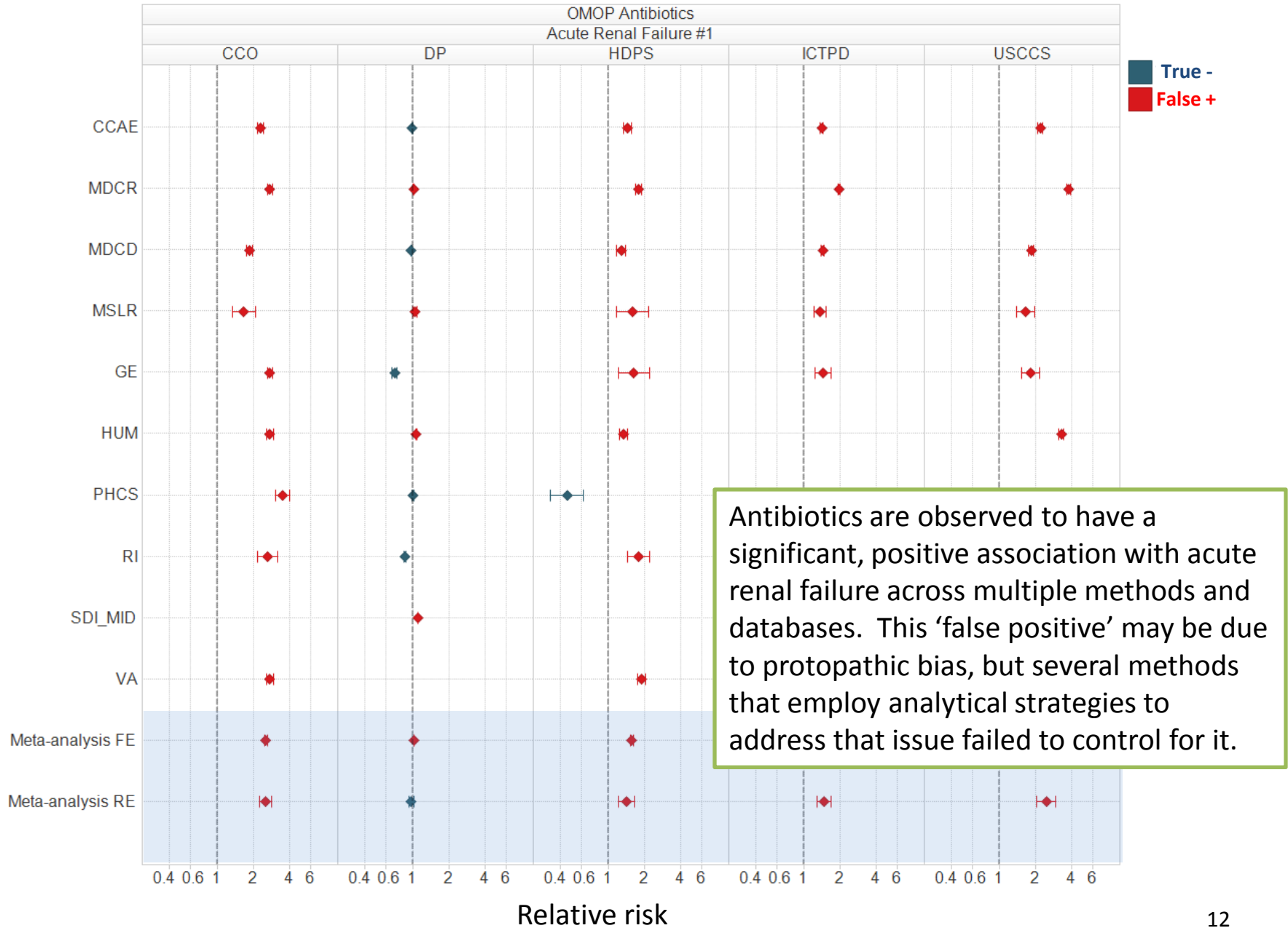


Systematic sensitivity analysis: Estimate the effect using multiple methods across the network of databases



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

Data source



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' assumed 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'
Aplastic Anemia	Negative control'		Negative control'	True positive' risk			Negative control'	Negative control'		Negative control'
Acute Liver Injury		Negative control'	True positive' risk			Negative control'				
Bleeding			Negative control'				Negative control'			True positive' risk
Hip Fracture	Negative control'	Negative control'			True positive' risk	Negative control'				Negative control'
Hospitalization	True positive' benefit									
Myocardial Infarction			Negative control'		Negative control'		Negative control'	True positive' risk	True positive' risk	
Mortality after MI		Negative control'		Negative control'		True positive' benefit				Negative control'
Renal Failure		True positive' risk	Negative control'			Negative control'	Negative control'	Negative control'		Negative control'
GI Ulcer Hospitalization	Negative control'			Negative control'			True positive' risk		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	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

= $(TP+TN) /$

$(TP+TN+FP+FN)$

= $(5+36) / (9+44) = 0.77$

Risk identification 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)	Columbia	6-Aug-10
Incident user design (IUD-HOI)	University of North Carolina	26-Oct-10

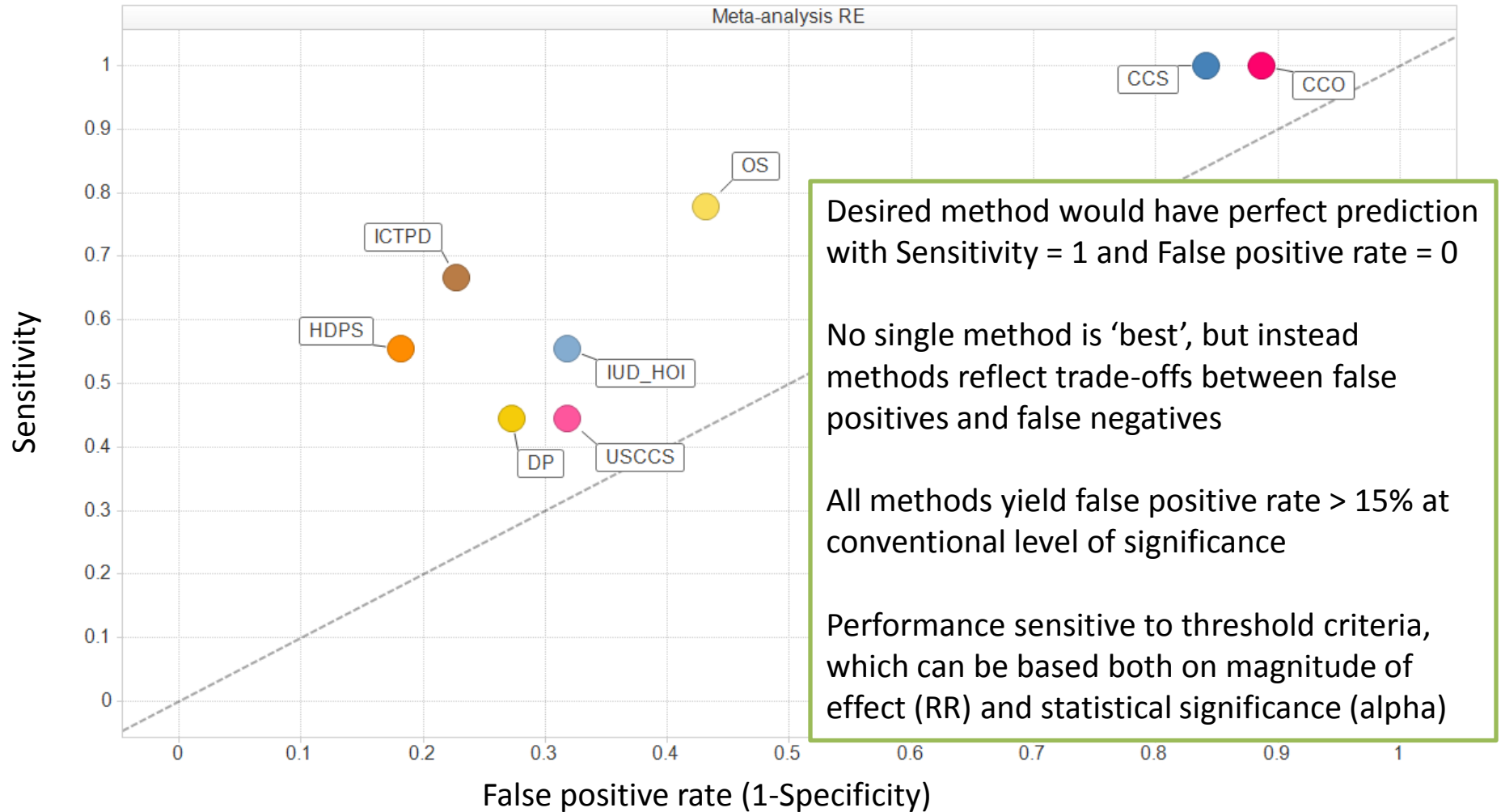
Sequential testing methods

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

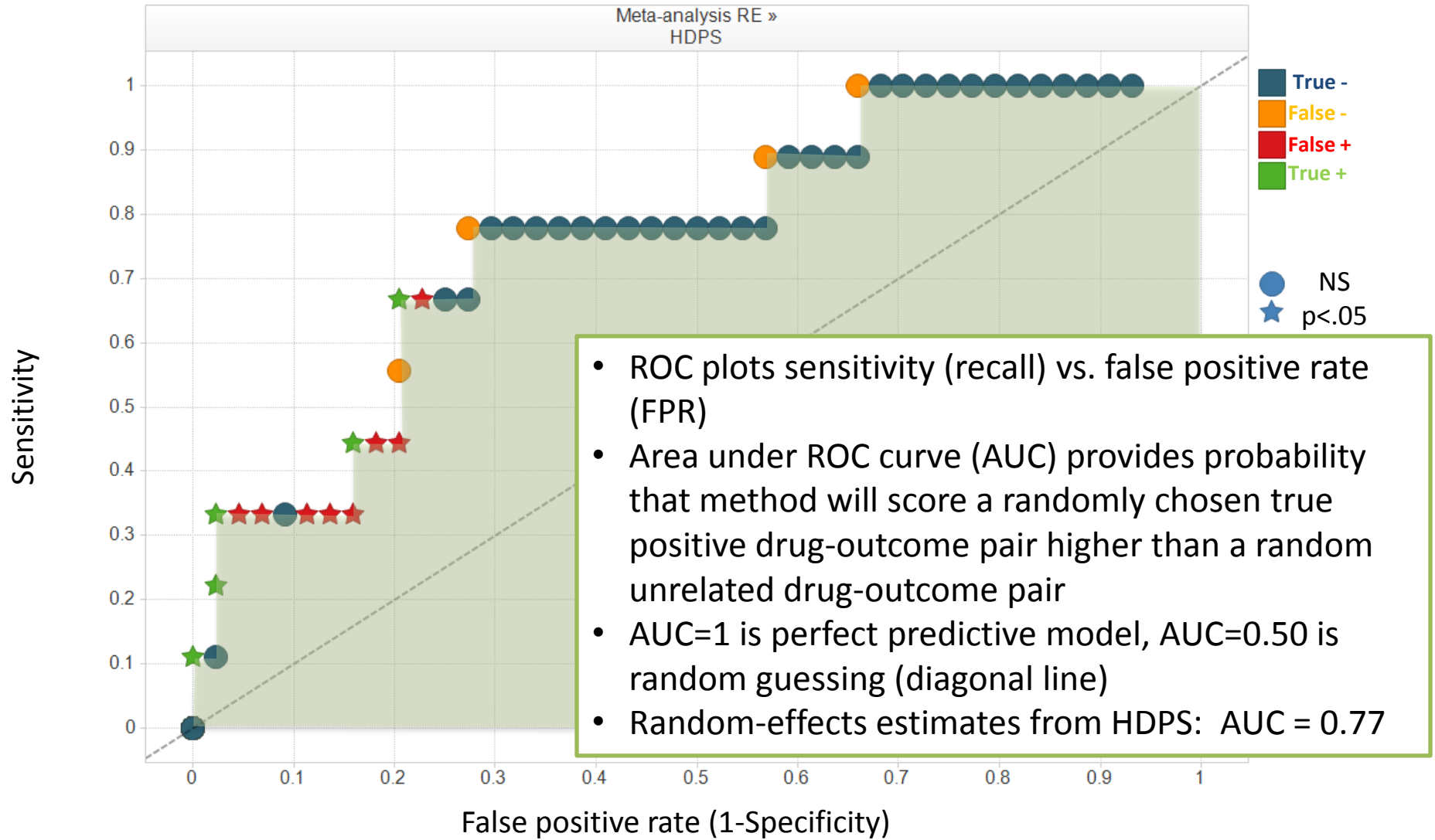
In what follows, we have chosen one parameter combination for each method that performs best for the meta-analysis estimates

<http://omop.fnih.org/MethodsLibrary>

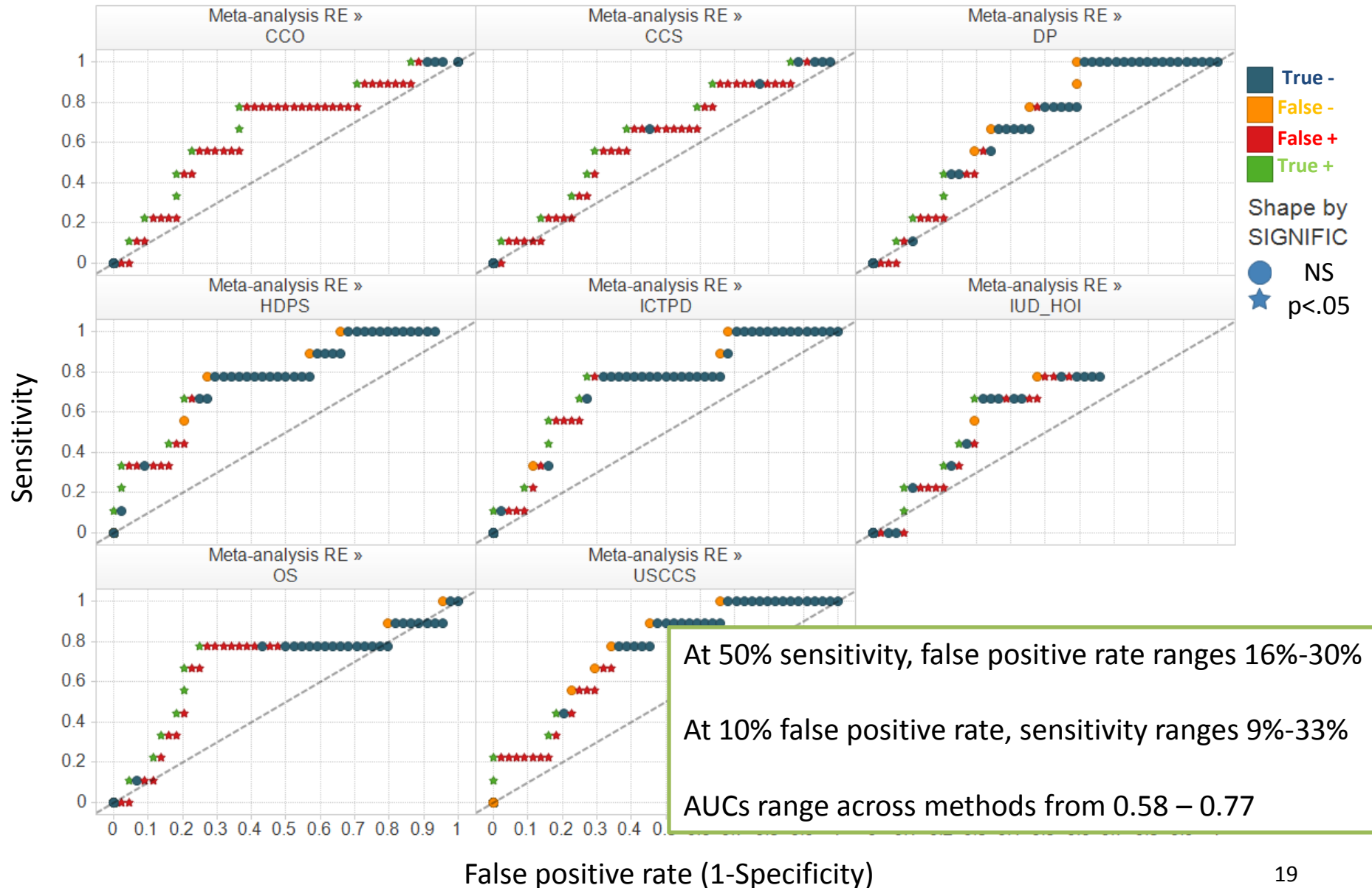
Comparing methods by sensitivity and specificity at alpha=0.05



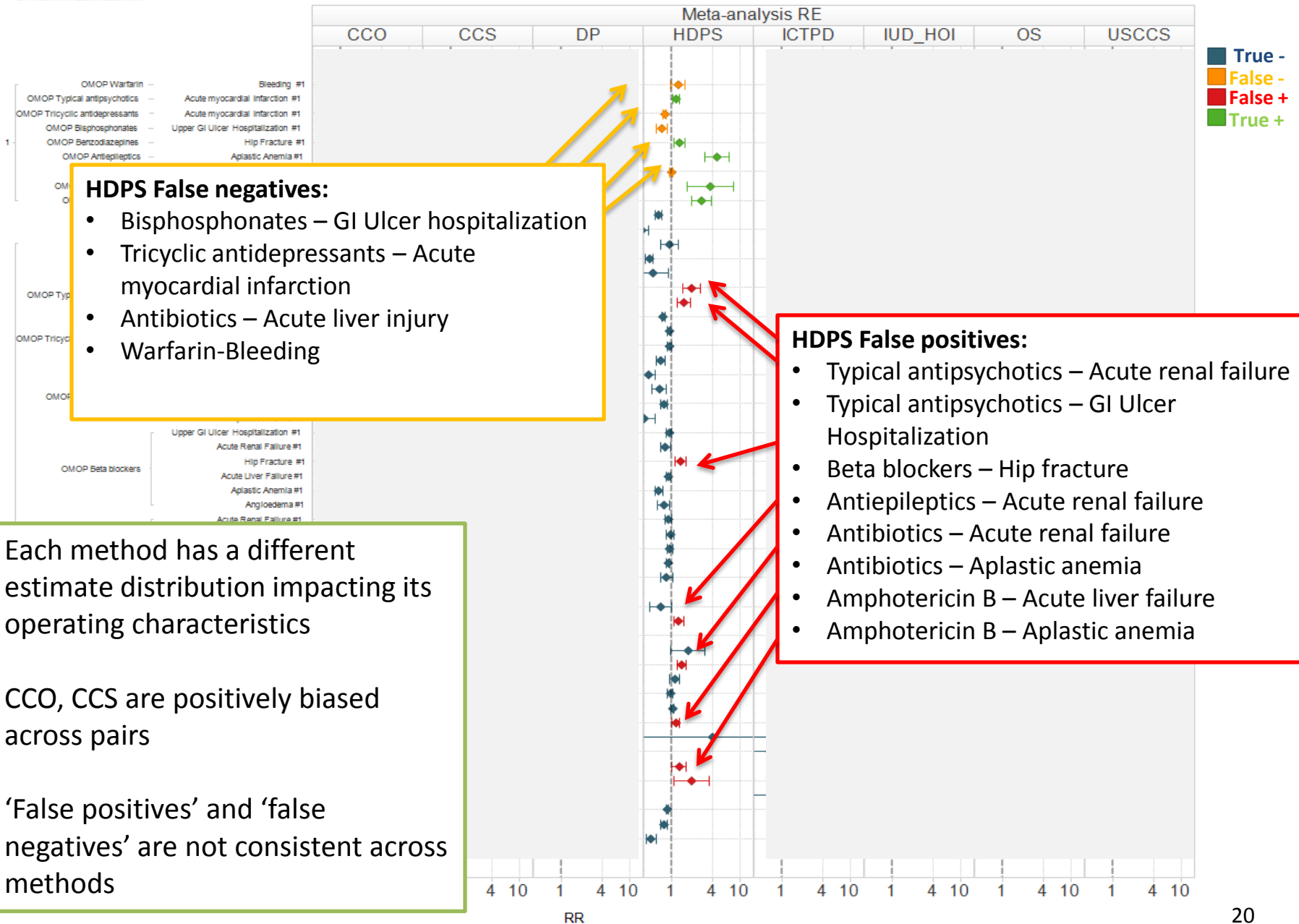
Receiver Operating Characteristic (ROC) curve



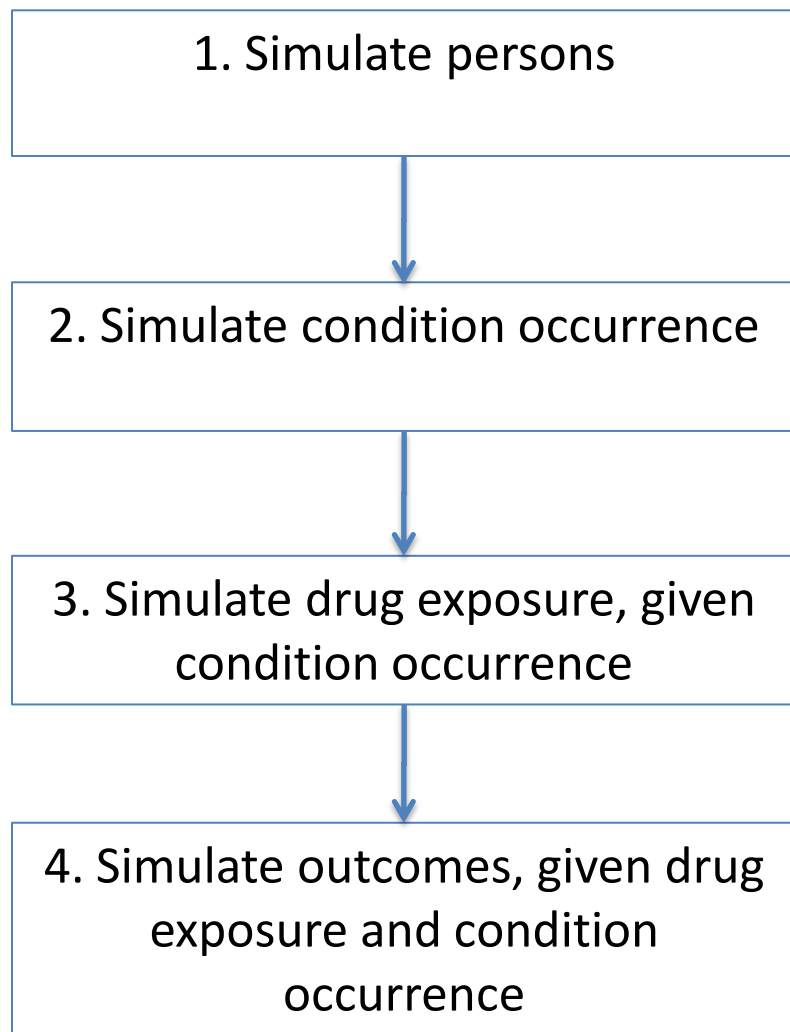
ROC curves of random-effects meta-analysis estimations for all methods



Distribution of estimates across all drug-outcome pairs



Creating Simulated Data



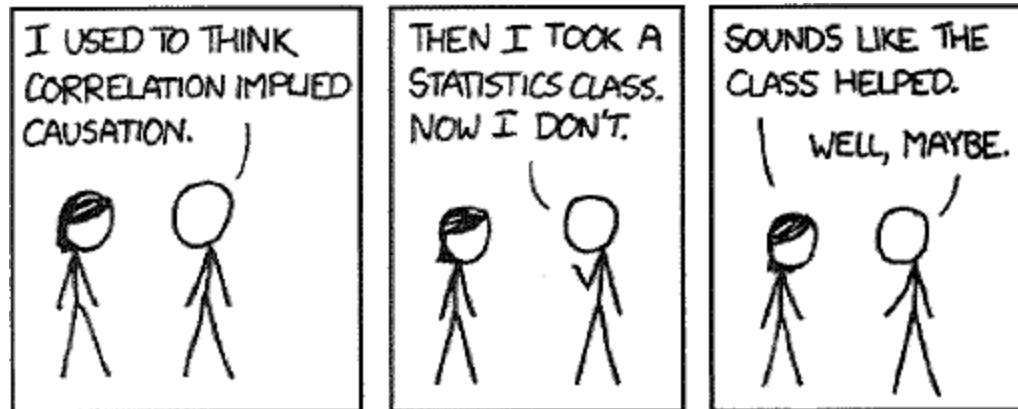
- OSIM2 can be applied to any real data source in the OMOP CDM format
- OSIM2 model preserves conditional independence between drugs and subsequent conditions
- Injecting known signals allows for precise measure of effect estimate error
- OMOP conducting experiments across an array of OSIM2 databases to complement real-world experiments

Concluding thoughts (1)

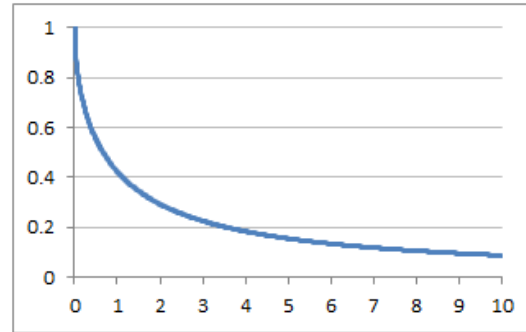
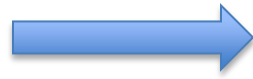
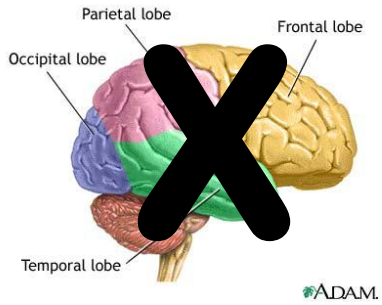
- A risk identification system can complement current practice by providing evidence to support a comprehensive safety assessment
- No one clear 'best' method, as it depends on tolerance for false positives vs. false negatives
- In this experiment, active surveillance methods achieved:
 - At 50% sensitivity, 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
- Further empirical research needed to have more complete understanding of operating characteristics before widespread adoption

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

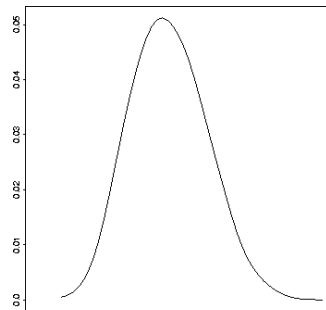
Integrating evidence from disparate sources: A learning paradigm for medical product safety



Bayesian Learning paradigm



prior



posterior

Bayes' Rule



Name	Thread pitch (mm)	Minor diameter tolerance	Nominal diameter (mm)	Head shape	Price for 50 screws	Available at factory outlet?	Number in stock	Flat or Phillips head?
M4	0.7	4g	4	Pan	\$10.08	Yes	276	Flat
M5	0.8	4g	5	Round	\$13.89	Yes	183	Both
M6	1	5g	6	Button	\$10.42	Yes	1043	Flat
M8	1.25	5g	8	Pan	\$11.98	No	298	Phillips
M10	1.5	6g	10	Round	\$16.74	Yes	488	Phillips
M12	1.75	7g	12	Pan	\$18.26	No	998	Flat
M14	2	7g	14	Round	\$21.19	No	235	Phillips
M16	2	8g	16	Button	\$23.57	Yes	292	Both
M18	2.1	8g	18	Button	\$25.87	No	664	Both
M20	2.4	8g	20	Pan	\$29.09	Yes	486	Both
M24	2.55	9g	24	Round	\$33.01	Yes	982	Phillips
M28	2.7	10g	28	Button	\$35.66	No	1067	Phillips
M36	3.2	12g	36	Pan	\$41.32	No	434	Both
M50	4.5	15g	50	Pan	\$44.72	No	740	Flat

data

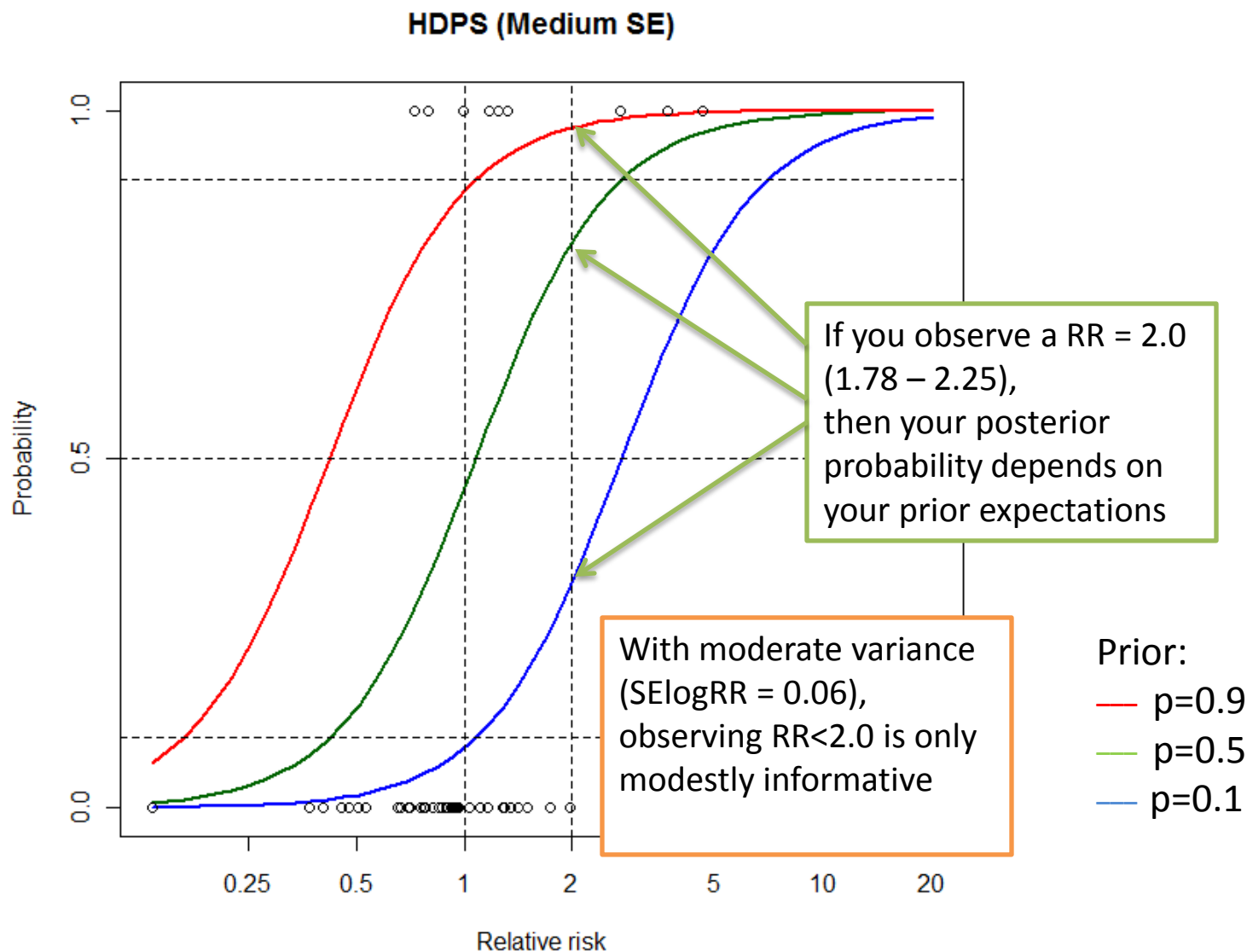


“Data”:

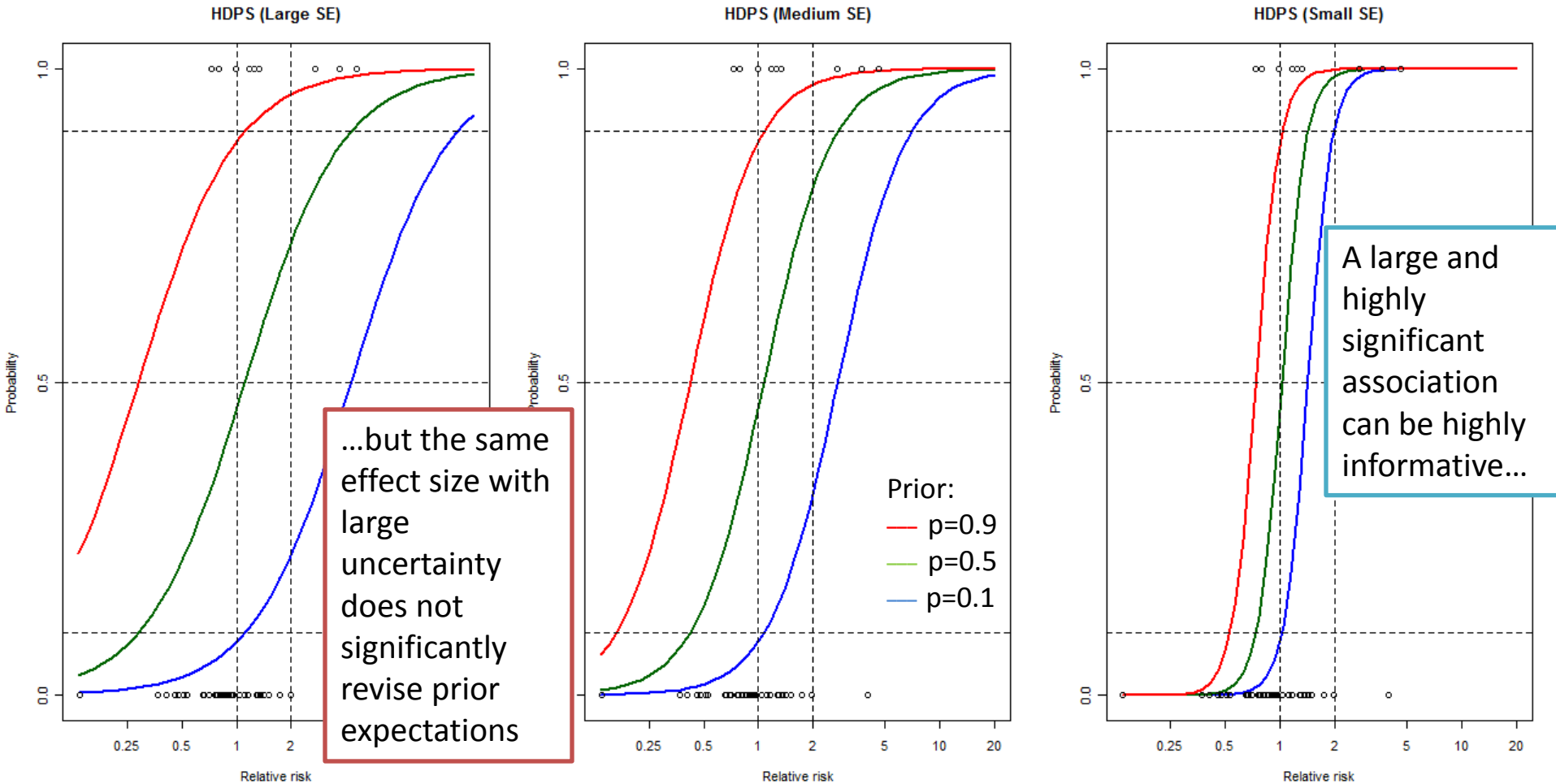
Effect estimates from one method against one database across an array of drug-outcome pairs

	Angioedema	Aplastic Anemia	Acute Liver Injury	Bleeding	Myocardial Infarction	Hip Fracture	Mortality after MI	Renal Failure	GI Ulcer Hospitalization
ACE Inhibitor	1.80 (0.15)	0.40 (0.05)				0.91 (0.12)			0.87 (0.03)
Amphotericin B	-	3.30 (0.99)	1.05 (0.24)			-	-	4.01 (0.99)	
Antibiotics		1.22 (0.08)	1.00 (0.01)	1.14 (0.01)	1.06 (0.03)	1.05 (0.09)		1.44 (0.05)	
Antiepileptics	1.74 (0.38)	4.60 (0.80)					-	1.63 (0.22)	0.54 (0.05)
Benzodiazepines	0.13 (0.01)	1.10 (0.06)	0.98 (0.01)	1.11 (0.01)	1.18 (0.03)	1.41 (0.12)		1.06 (0.04)	
Beta blockers	0.81 (0.07)	0.63 (0.06)	0.95 (0.01)			1.69 (0.18)		0.78 (0.04)	0.88 (0.03)
Bisphosphonates		0.27 (0.05)	0.85 (0.03)		0.82 (0.07)			0.40 (0.04)	0.90 (0.06)
Tricyclic antidepressants		0.63 (0.07)	1.02 (0.02)	0.96 (0.01)	0.80 (0.04)			0.82 (0.06)	
Typical antipsychotics					0.96 (0.07)			1.97 (0.16)	3.46 (0.21)
Warfarin	0.53 (0.11)	0.47 (0.04)		2.13 (0.04)		1.20 (0.09)	0.49 (0.07)	0.76 (0.05)	

Revising prior expectations in light of new evidence from a risk identification system



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



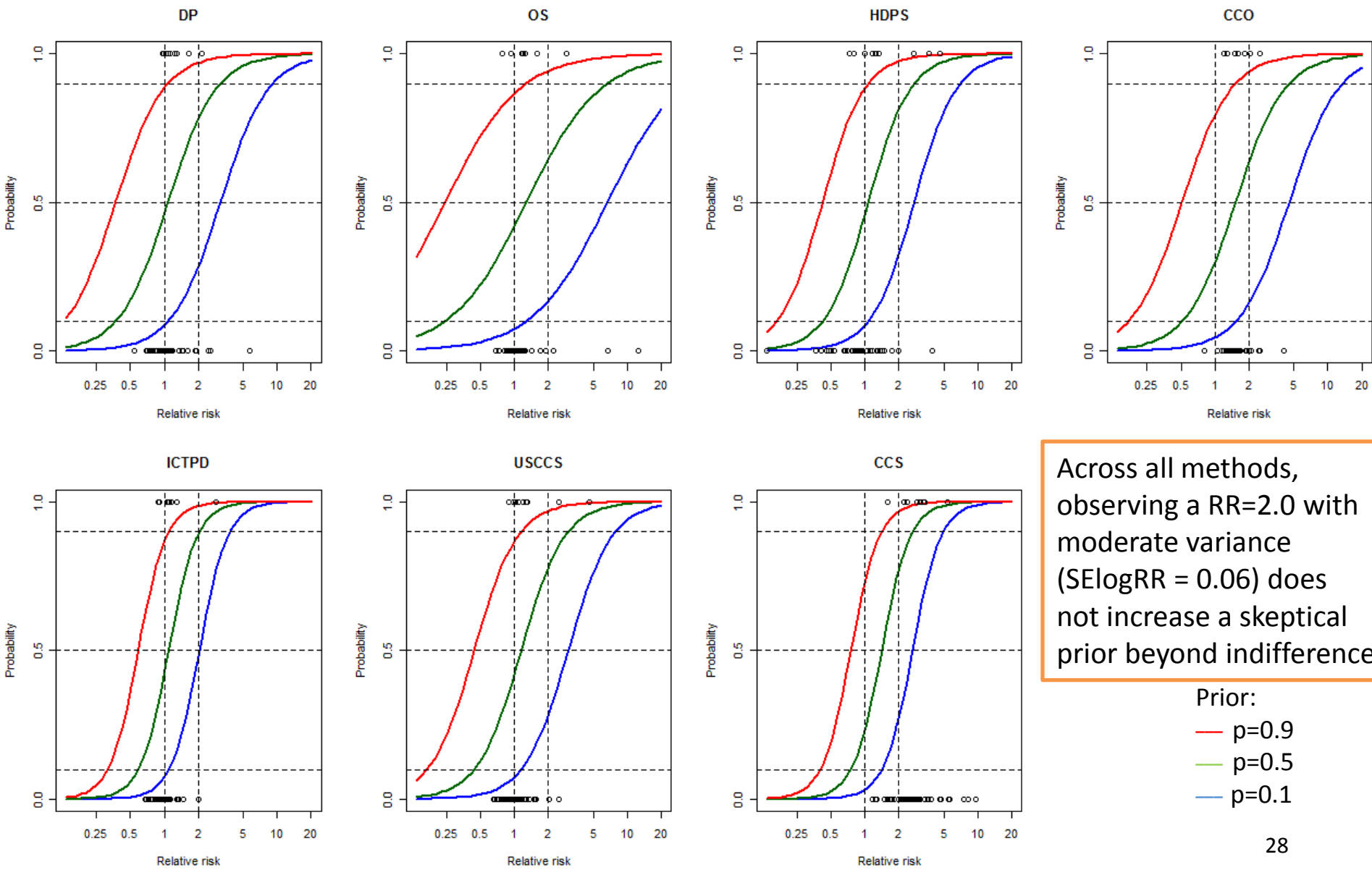
Scenarios: You observe RR=2.0 with confidence intervals based on standard error (SE):

Large SE: (1.01 – 3.97)

Medium SE: (1.78 – 2.25)

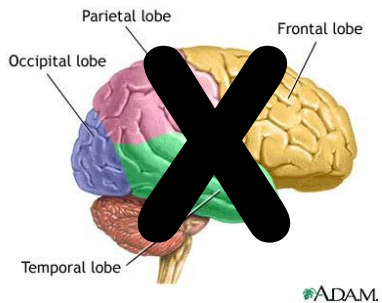
Small SE: (1.96 – 2.04)

Revising prior expectations in light of new evidence from a risk identification system: Impact of using estimates from different methods

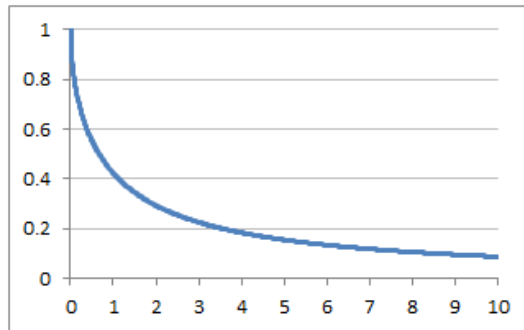


Across all methods, observing a RR=2.0 with moderate variance ($SE_{\log RR} = 0.06$) does not increase a skeptical prior beyond indifference

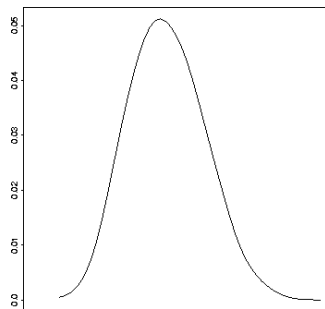
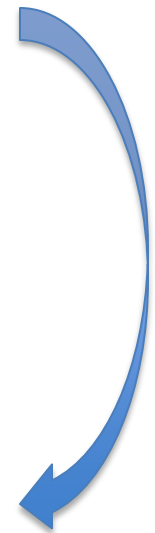
Bayesian Learning paradigm



ADAM



prior

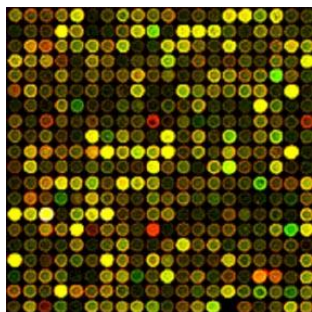


prior

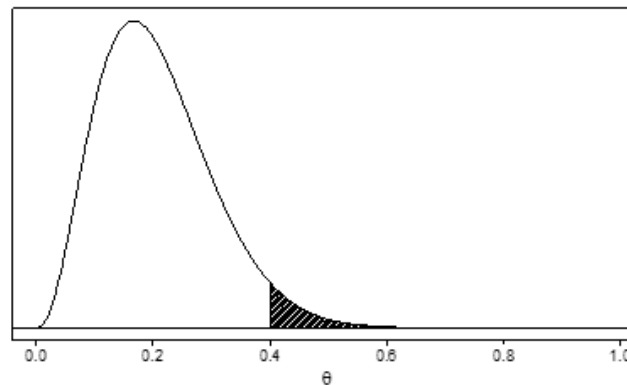
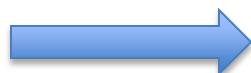
Bayes' Rule

Name	Thread pitch (mm)	Minor diameter tolerance	Nominal diameter (mm)	Head shape	Price for 50 screws	Available at factory outlet?	Number in stock	Flat or Phillips head?
M4	0.7	4g	4	Pan	\$10.08	Yes	276	Flat
M5	0.8	4g	5	Round	\$13.89	Yes	183	Both
M6	1	5g	6	Button	\$10.42	Yes	1043	Flat
M8	1.25	5g	8	Pan	\$11.98	No	298	Phillips
M10	1.5	6g	10	Round	\$16.74	Yes	488	Phillips
M12	1.75	7g	12	Pan	\$18.26	No	998	Flat
M14	2	7g	14	Round	\$21.19	No	235	Phillips
M16	2	8g	16	Button	\$23.57	Yes	292	Both
M18	2.1	8g	18	Button	\$25.87	No	664	Both
M20	2.4	8g	20	Pan	\$29.09	Yes	486	Both
M24	2.55	9g	24	Round	\$33.01	Yes	982	Phillips
M28	2.7	10g	28	Button	\$35.66	No	1067	Phillips
M36	3.2	12g	36	Pan	\$41.32	No	434	Both
M50	4.5	15g	50	Pan	\$44.72	No	740	Flat

data

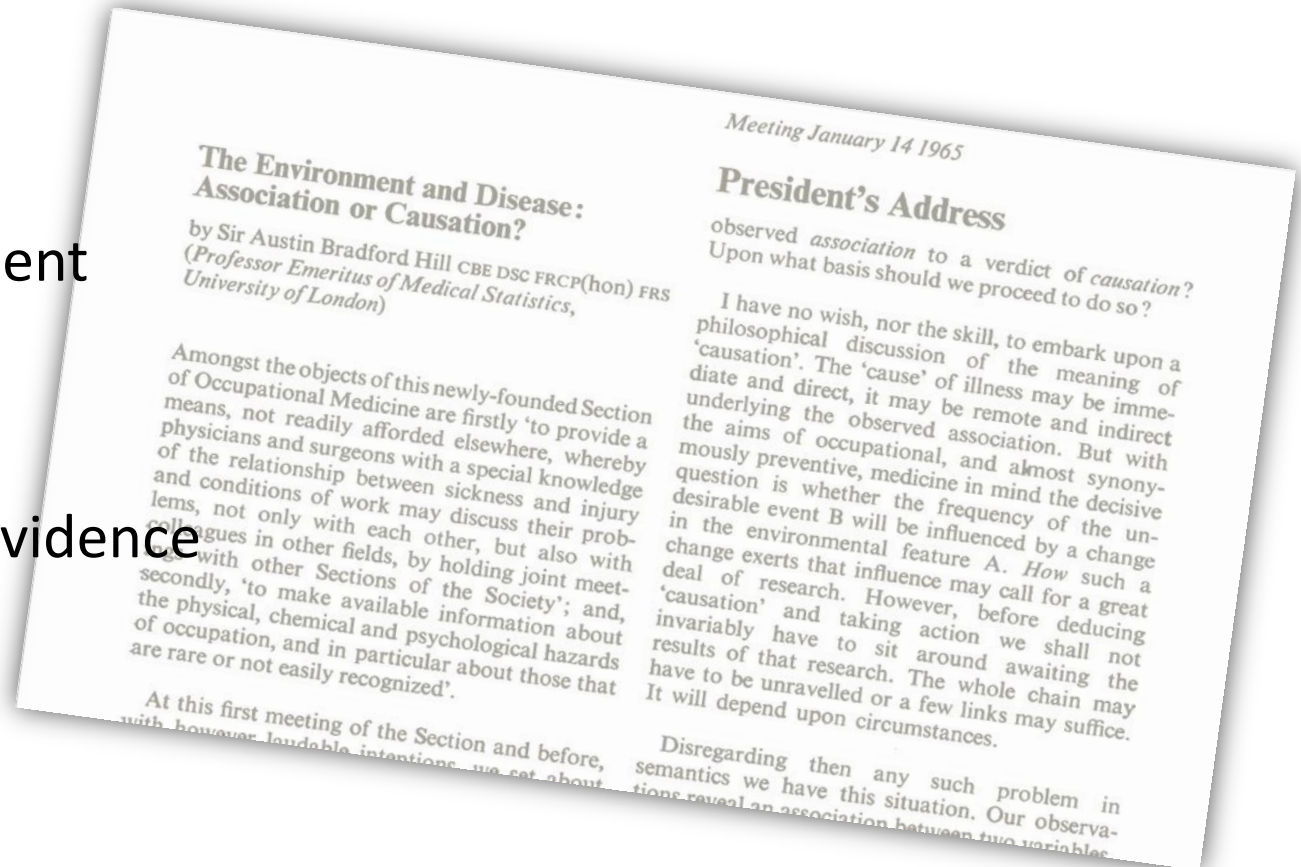


data



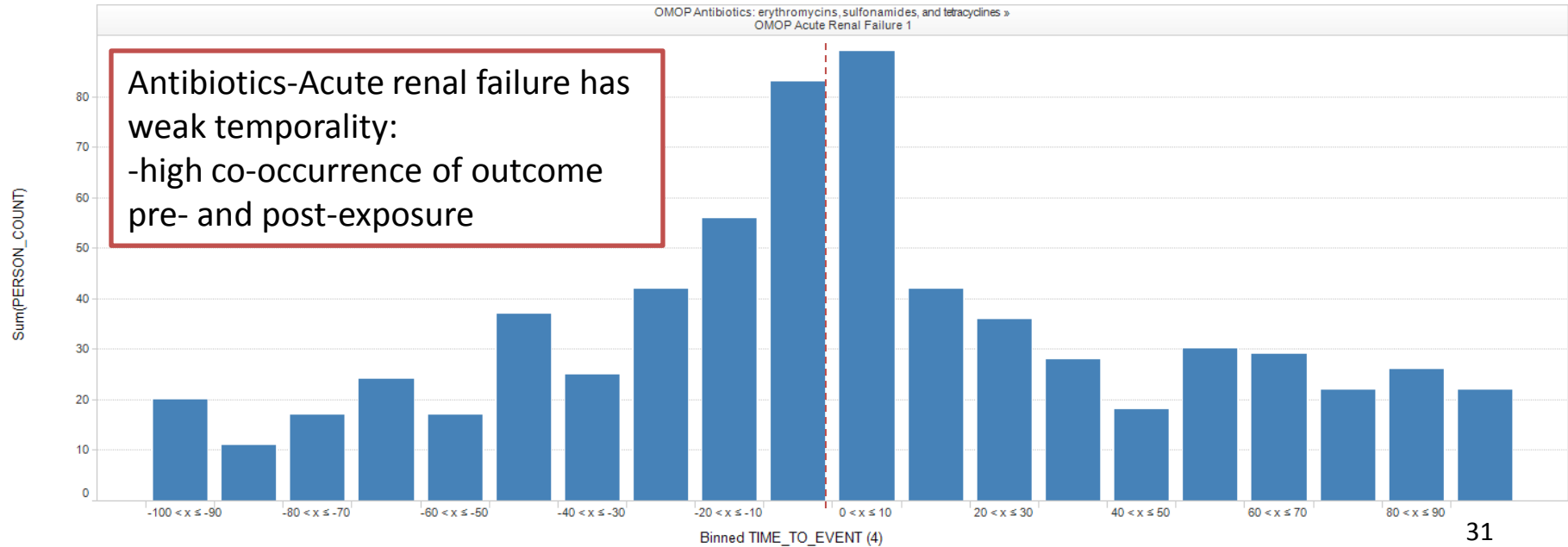
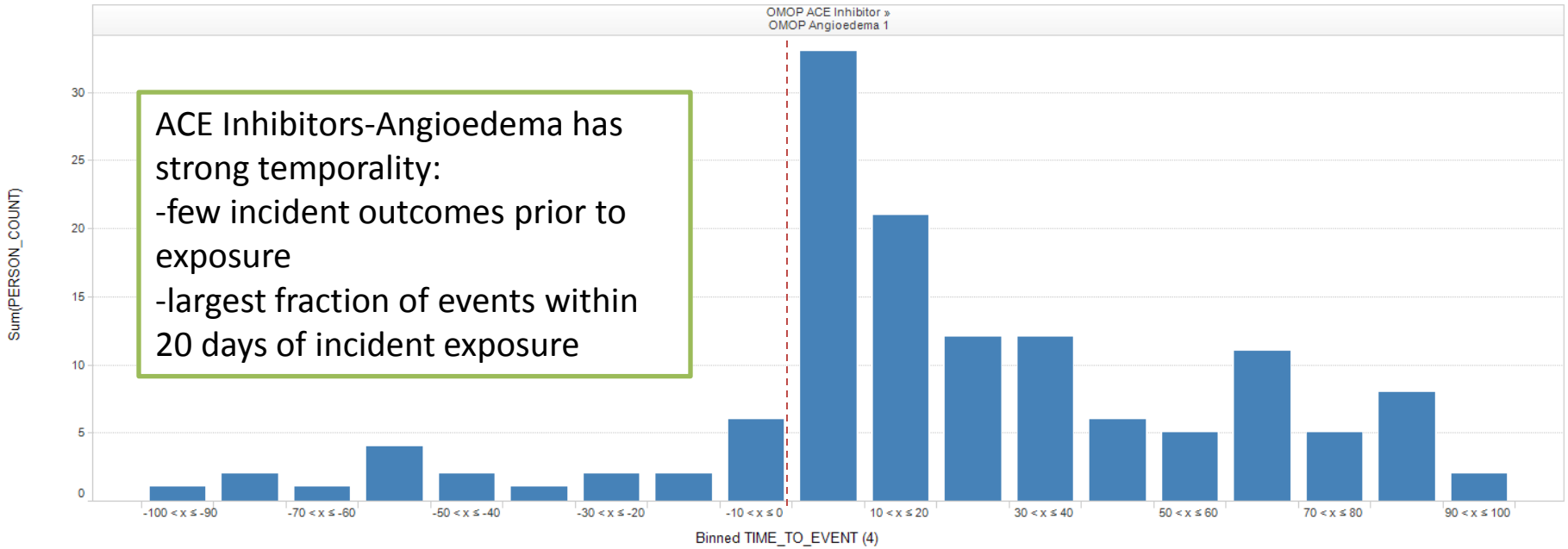
Hill's causality viewpoints

- Strength of association
- Consistency
- Specificity
- Temporality
- Biological gradient
- Plausibility
- Coherence
- Experimental evidence
- Analogy



Austin Bradford Hill, "The Environment and Disease: Association or Causation?," *Proceedings of the Royal Society of Medicine*, 58 (1965), 295-300.

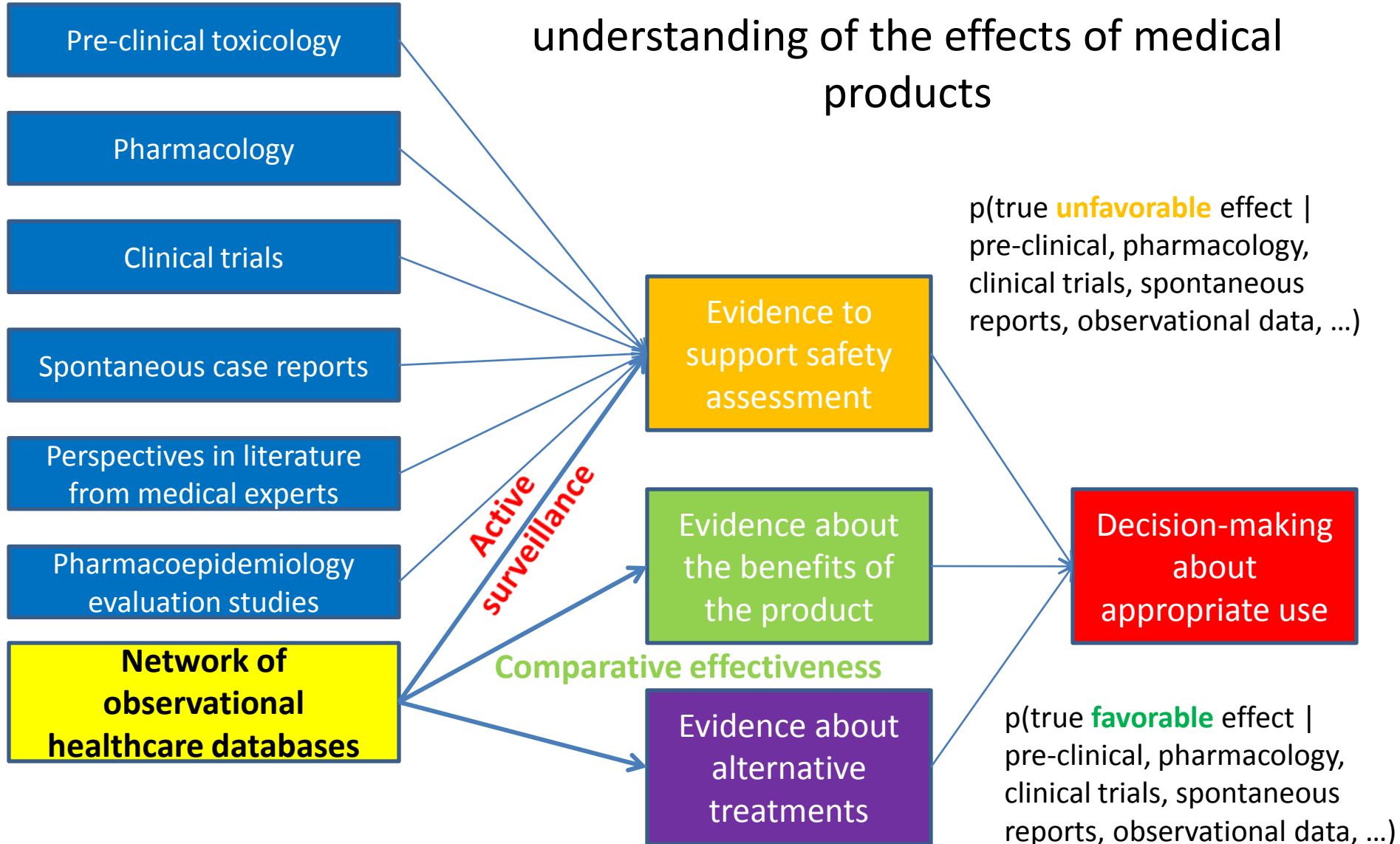
Temporality



Harnessing Hill

- Previously $p(\text{true} \mid \text{RR}, \text{SE})$
 - Logistic regression with 2 predictors
- Using Hill: $p(\text{true} \mid \text{RR}, \text{SE}, \text{temporality}, \text{coherence}, \text{consistency}, \text{plausibility}, \text{biological gradient}, \text{specificity}, \text{etc.})$
 - Logistic regression with many predictors
- Thus we have a framework to formally integrate diverse evidence into the causal judgment

Opportunities for a coordinated system that leverages a network of observational healthcare databases to enhance our understanding of the effects of medical products



Conclusions

- Observational healthcare data can be used to efficiently generate evidence about the potential effects of medical products
- The confidence in that evidence needs to be based on the operating characteristics of observational analyses
- The risk identification and analysis system will be only one piece of information that needs to be integrated with all other existing evidence to provide a more comprehensive safety assessment
- Safety assessments always need to be put into broader context with evidence about benefits and alternative treatments, incorporating stakeholder perspectives to guide medical decision-making