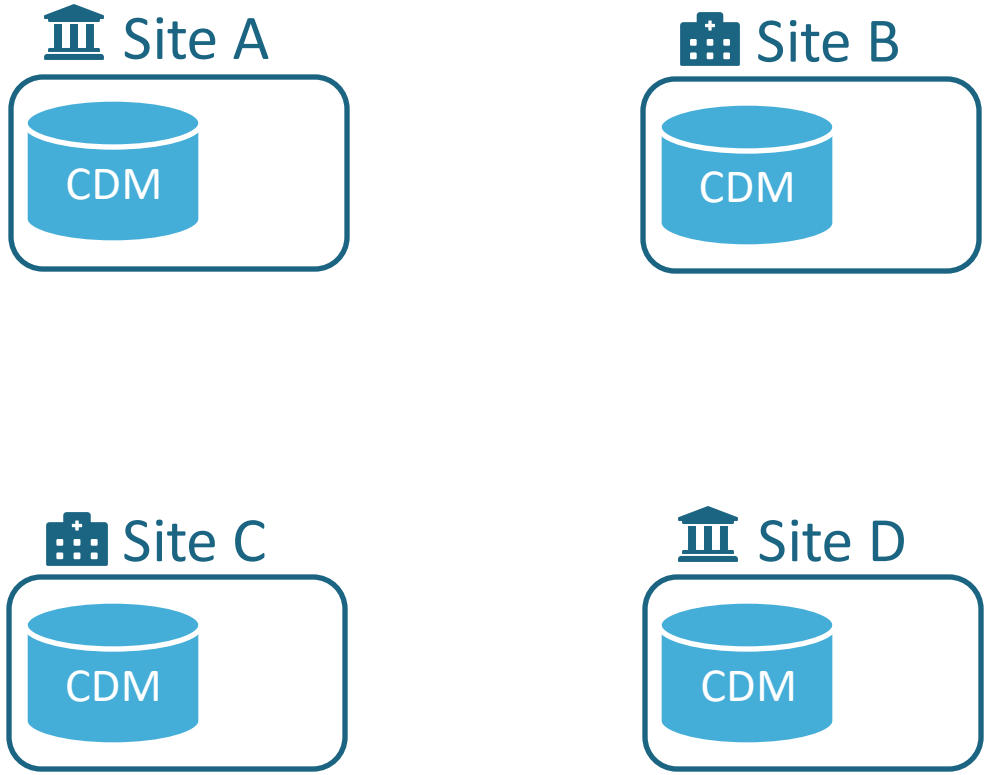
EvidenceSynthesis

combining causal effect estimates without sharing individual person data

Martijn Schuemie, Yong Chen

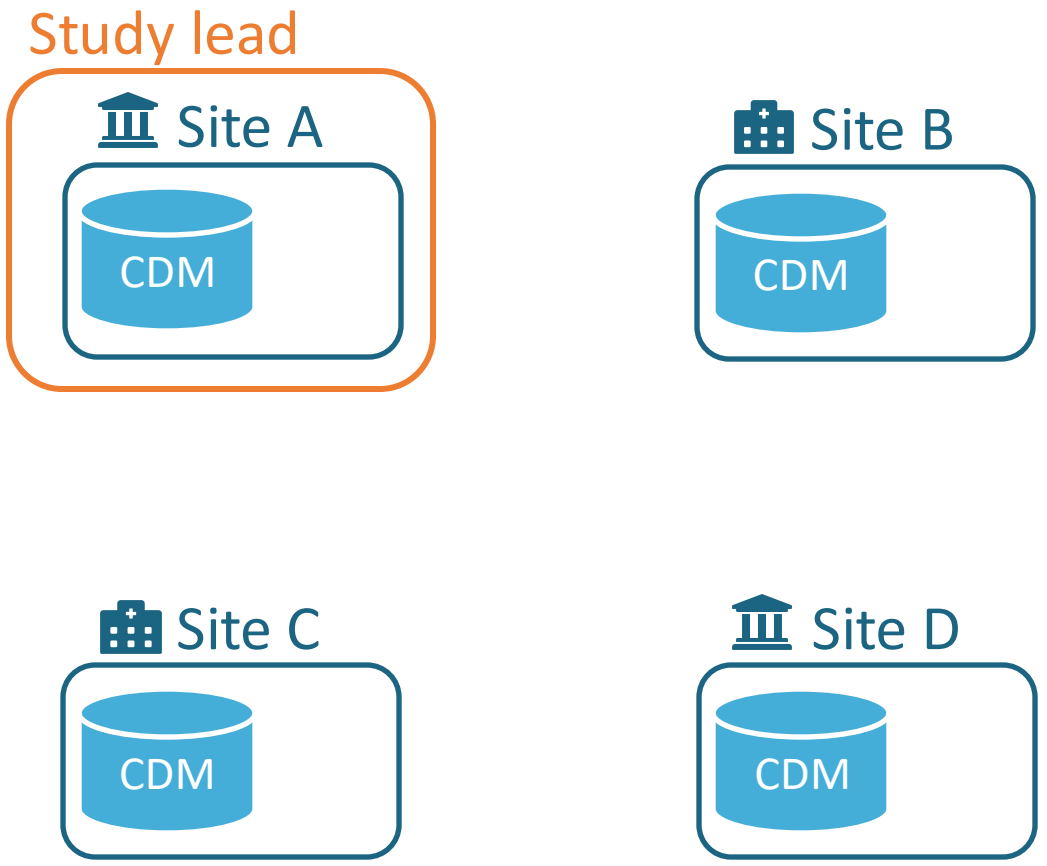
Distributed Research Network

- Multiple sites with data
  - Hospital EHRs (Electronic Health Records)
  - Administrative Claims
- Patient-level data cannot be shared
- Each site uses the Common Data Model (CDM)



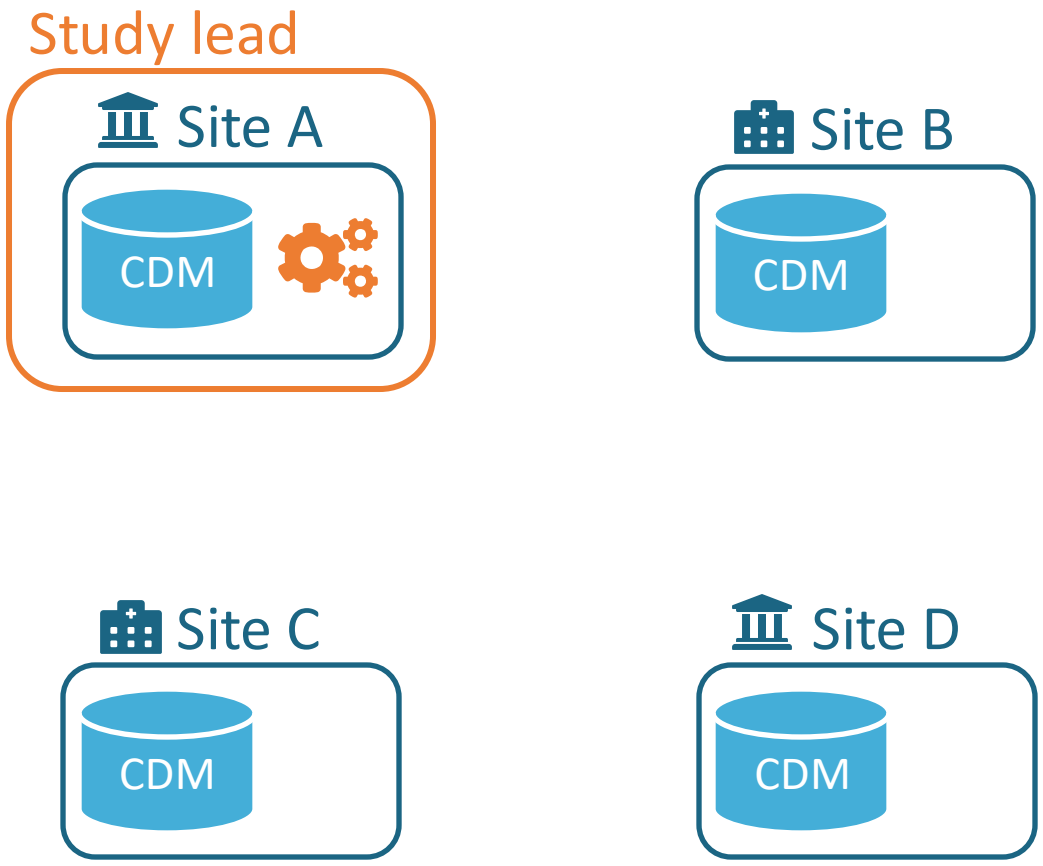
Distributed Research Network

- A site can lead a study



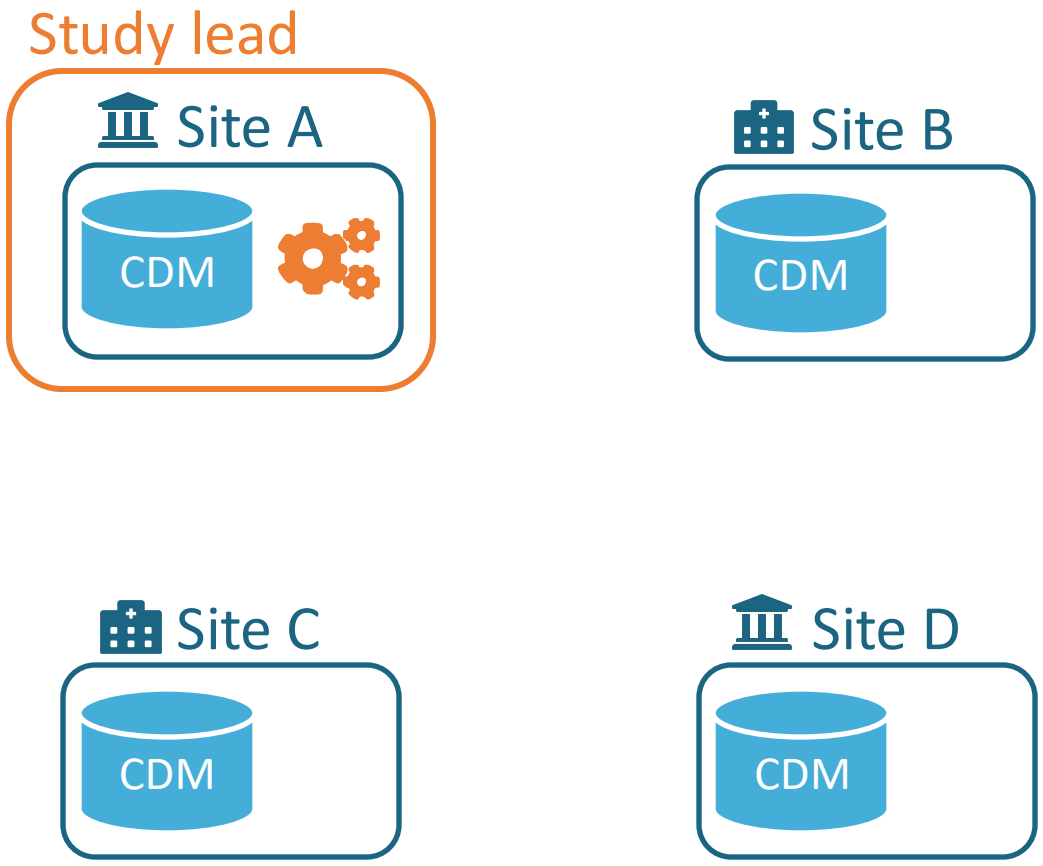
Distributed Research Network

- A site can lead a study
- Analysis code is developed locally



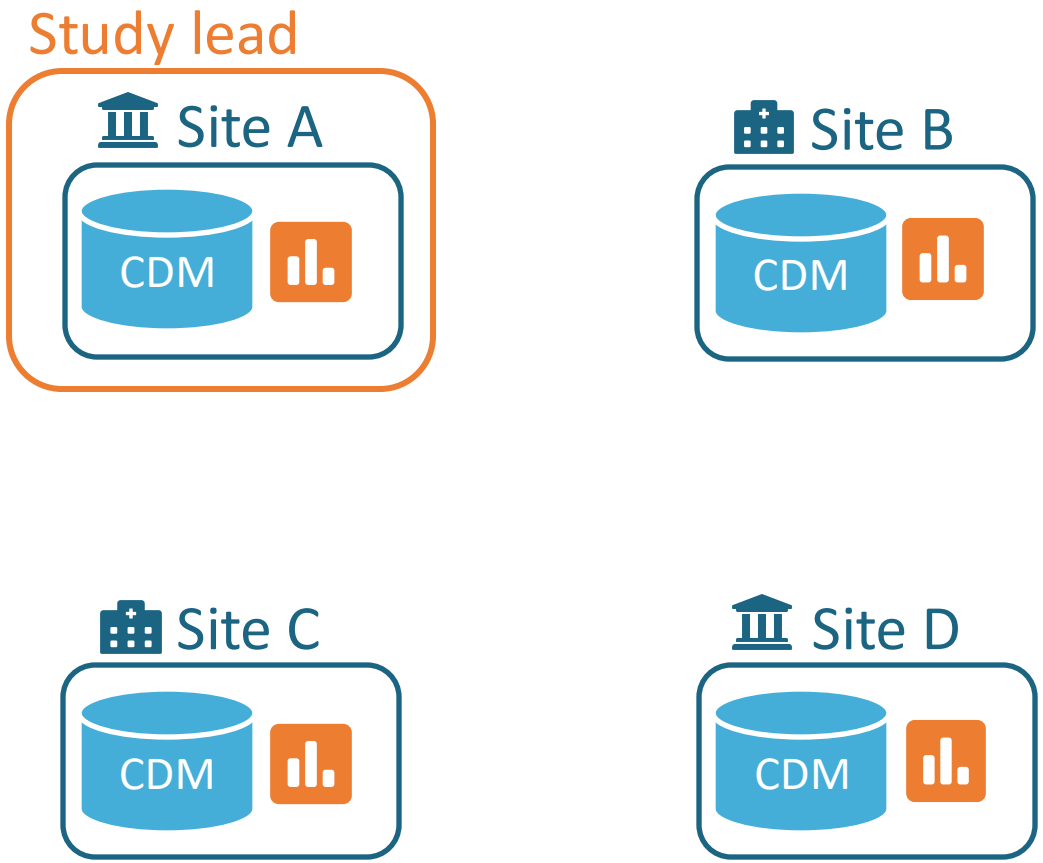
Distributed Research Network

- A site can lead a study
- Analysis code is developed locally
- Code is distributed to study participants



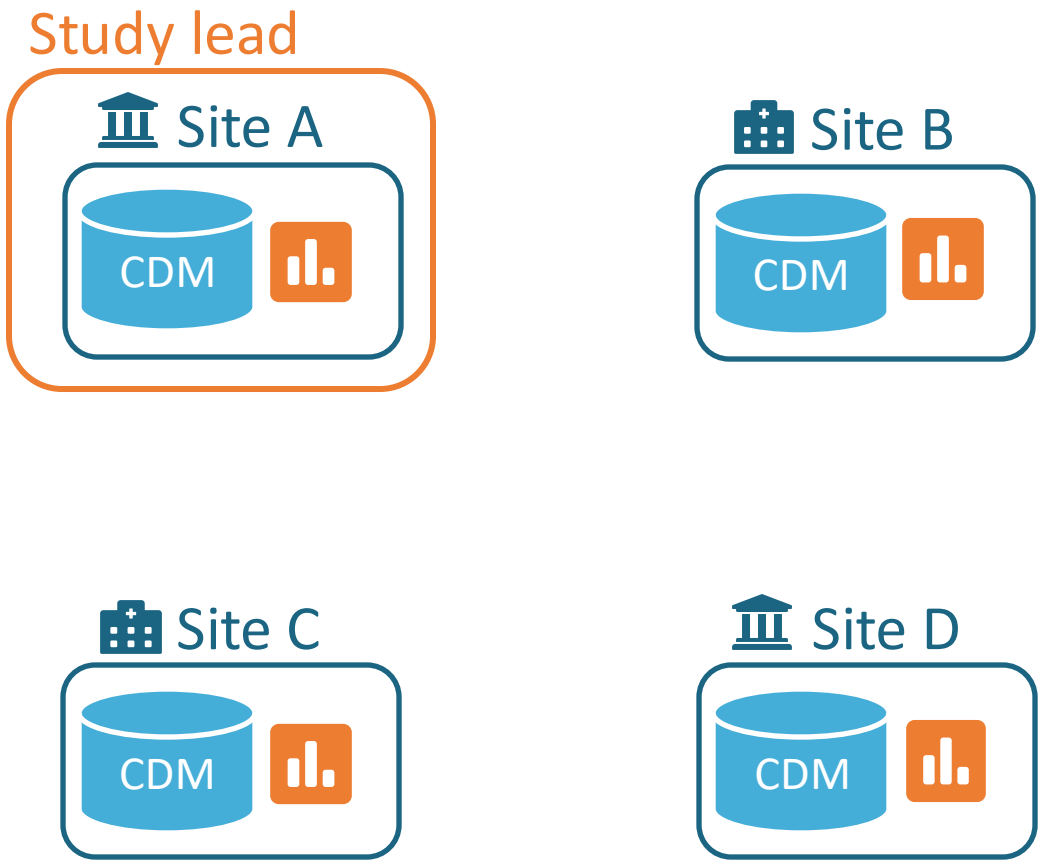
Distributed Research Network

- A site can lead a study
- Analysis code is developed locally
- Code is distributed to study participants
- Results are generated (aggregated statistics)



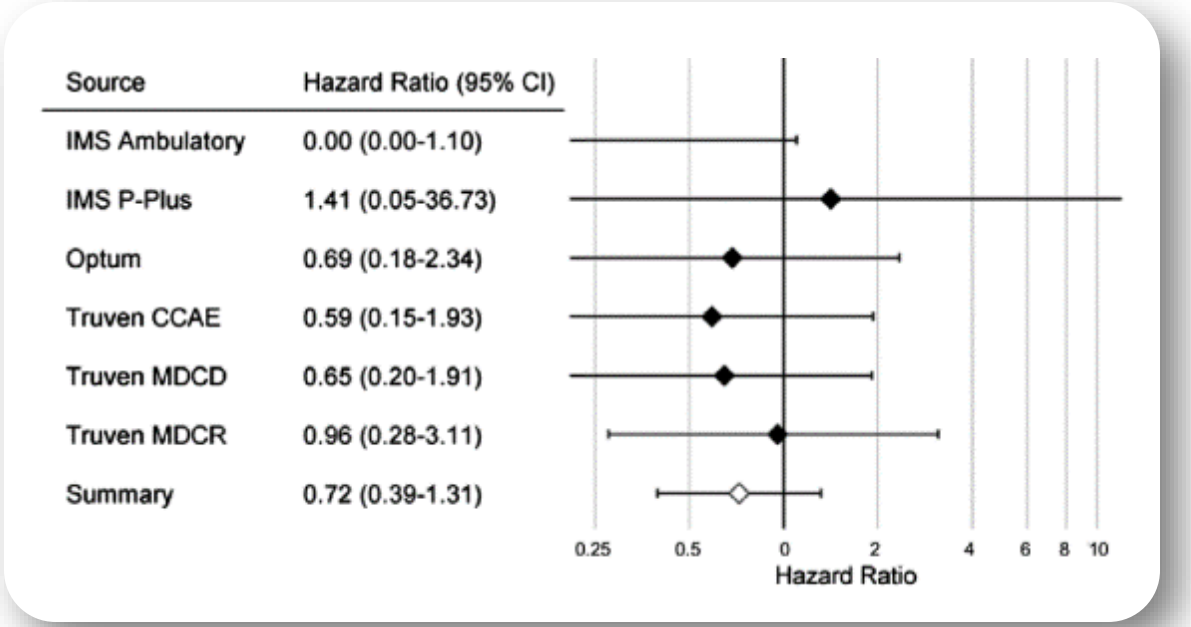
Distributed Research Network

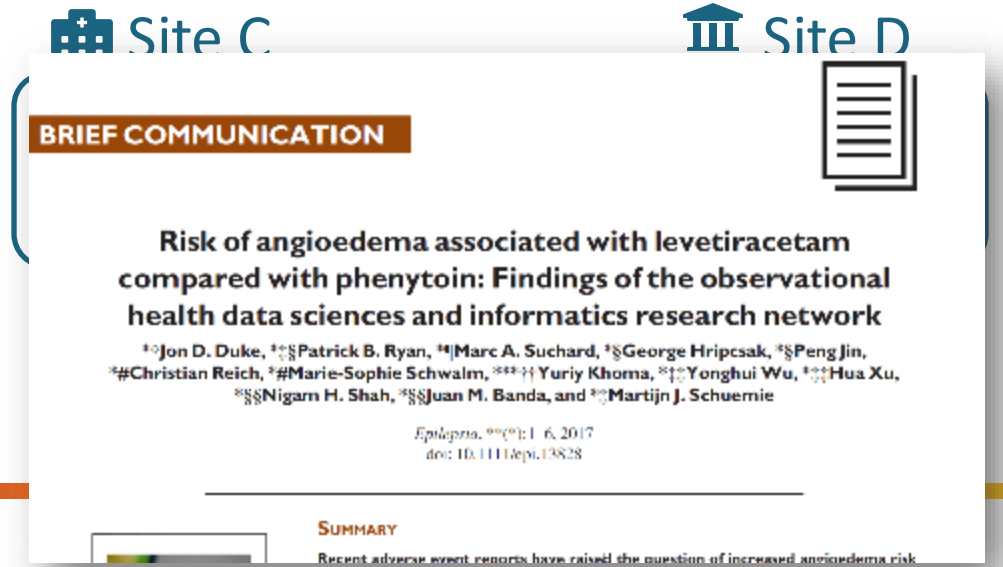
- A site can lead a study
- Analysis code is developed locally
- Code is distributed to study participants
- Results are generated (aggregated statistics)
- Results are send back to lead site



Distributed Research Network

- A site can lead a study
- Analysis code is developed locally





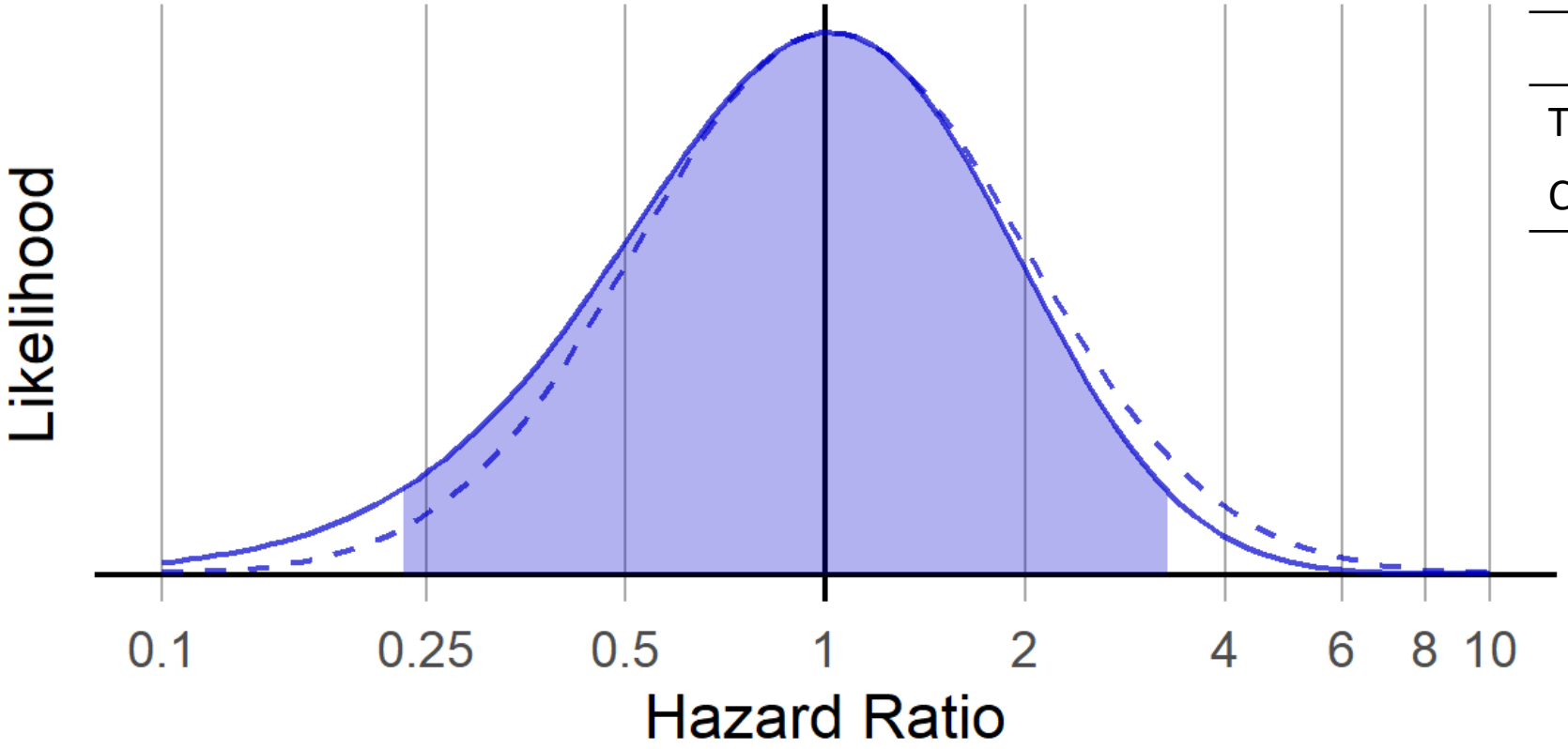
What to share when estimating causal effects?

- Cannot share patient-level data
- Usually **propensity-score stratified time-to-event** or **conditional Poisson regression**: no 2-by-2 tables
- Point-estimates + standard errors?

Normal assumption violated when counts are low

Hazard Ratio = 1.02 (0.27 – 3.78) Assuming normal distribution

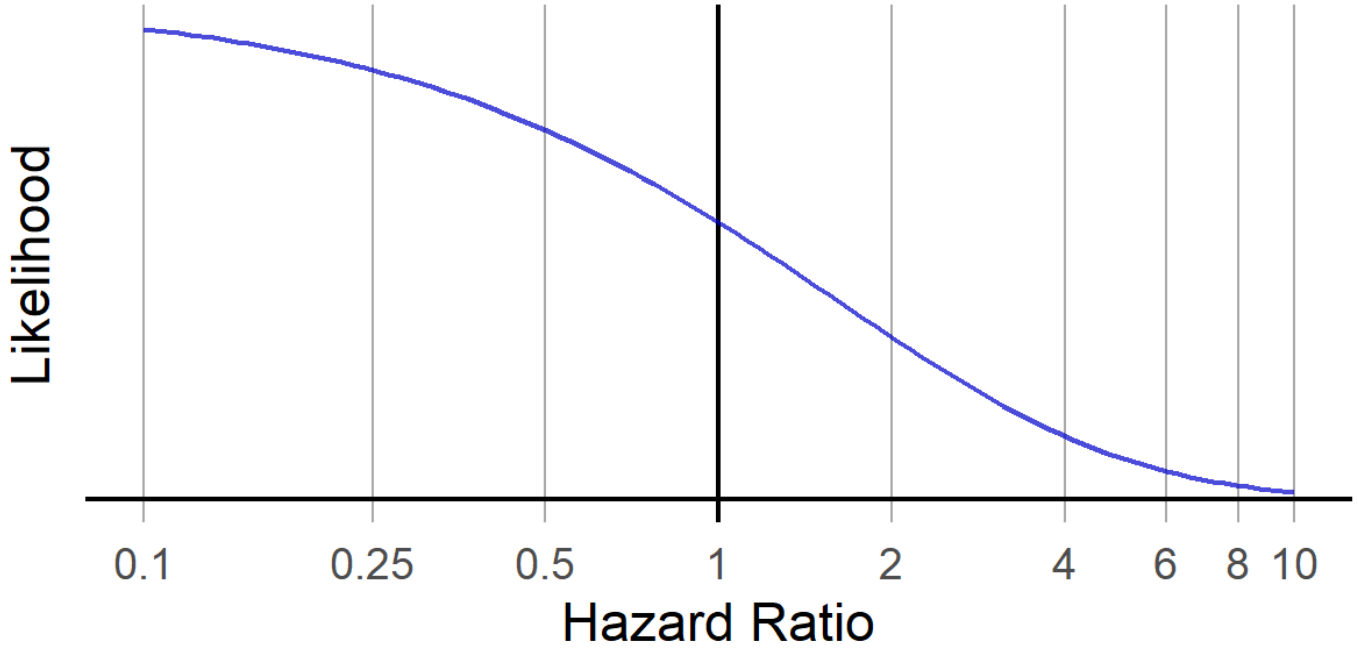
Hazard Ratio = 1.02 (0.22 – 3.31) No shape assumption



	Subjects	Outcomes
Target	22,002	3
Comparator	130,200	25

* Real data, no simulation

Even more when counts are 0

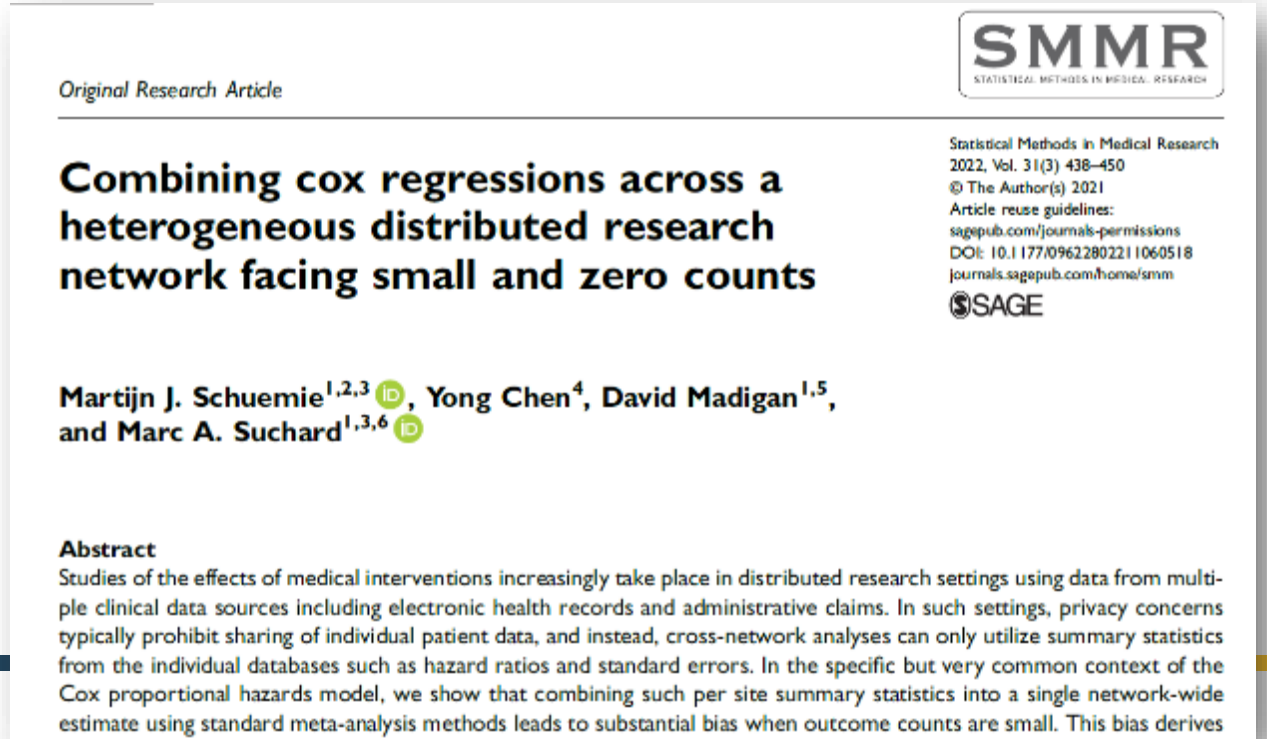


	Subjects	Outcomes
Target	2,834	0
Comparator	15,168	10

* Real data, no simulation

Solution

Data sites share actual shape of likelihood, instead of just the hazard ratio + confidence interval



Original Research Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2022, Vol. 31(3) 438–450
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09622802211060518
journals.sagepub.com/home/smm
SAGE

Combining cox regressions across a heterogeneous distributed research network facing small and zero counts

Martijn J. Schuemie[1,2,3], Yong Chen[4], David Madigan[1,5], and Marc A. Suchard[1,3,6]

Abstract

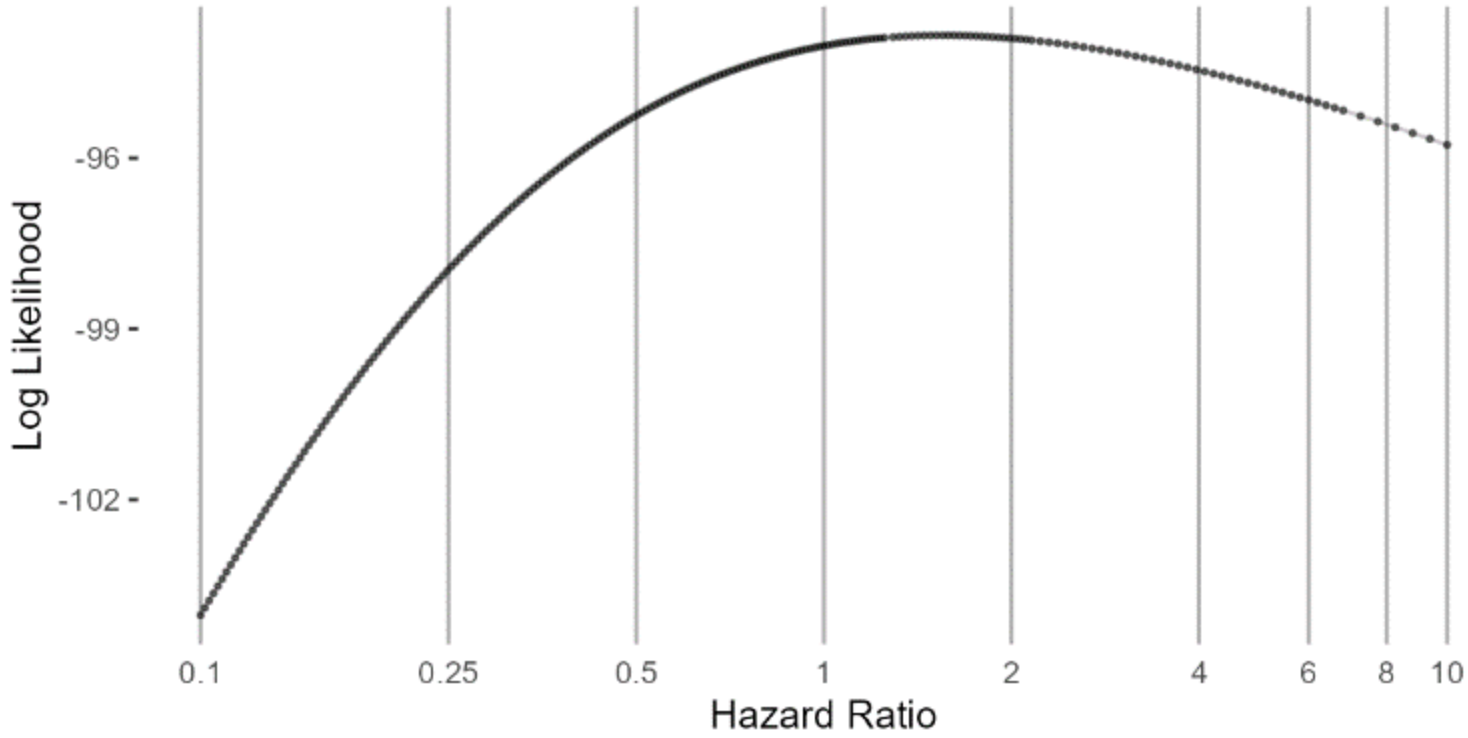
Studies of the effects of medical interventions increasingly take place in distributed research settings using data from multiple clinical data sources including electronic health records and administrative claims. In such settings, privacy concerns typically prohibit sharing of individual patient data, and instead, cross-network analyses can only utilize summary statistics from the individual databases such as hazard ratios and standard errors. In the specific but very common context of the Cox proportional hazards model, we show that combining such per site summary statistics into a single network-wide estimate using standard meta-analysis methods leads to substantial bias when outcome counts are small. This bias derives

Solution

Data sites share actual shape of likelihood, instead of just the hazard ratio + confidence interval

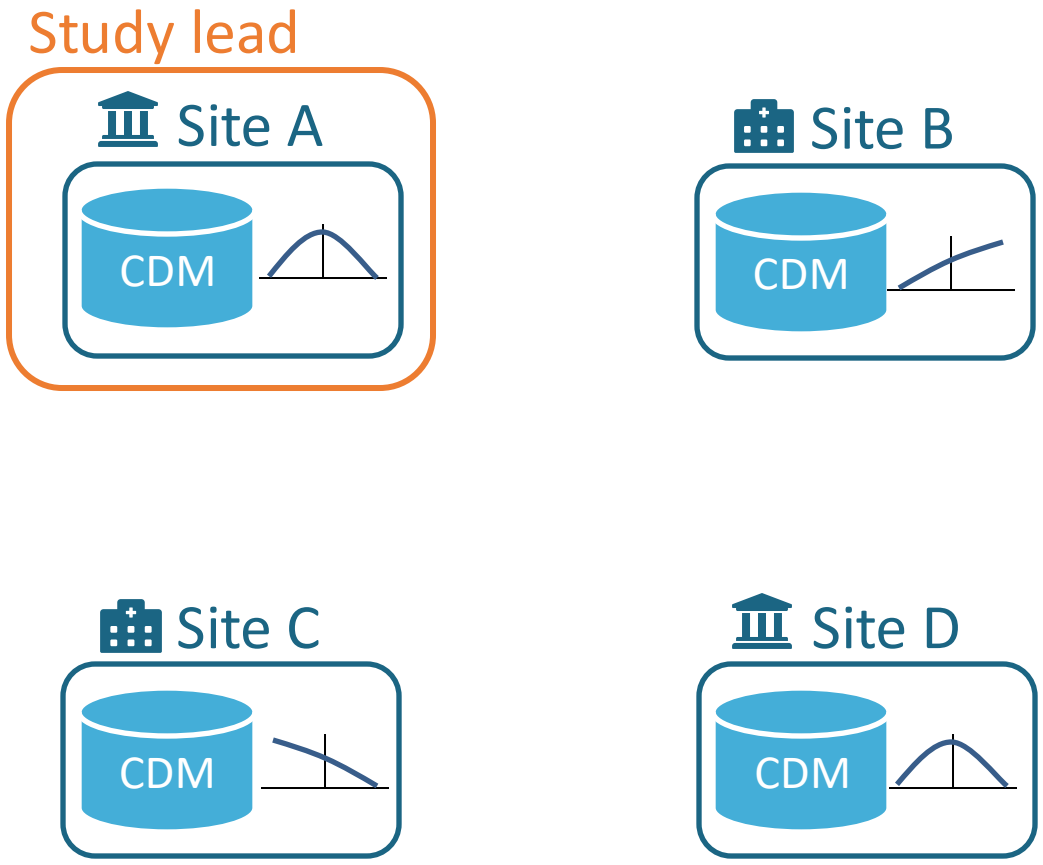
Current best-practice:

- Adaptive grid



Distributed Research Network

- Data sites share shape of likelihood



Evidence Synthesis

The **EvidenceSynthesis** package implements

- Fixed-effects model
- Random-effects model using a Bayesian approach.
  - Uses the BEAST MCMC engine.

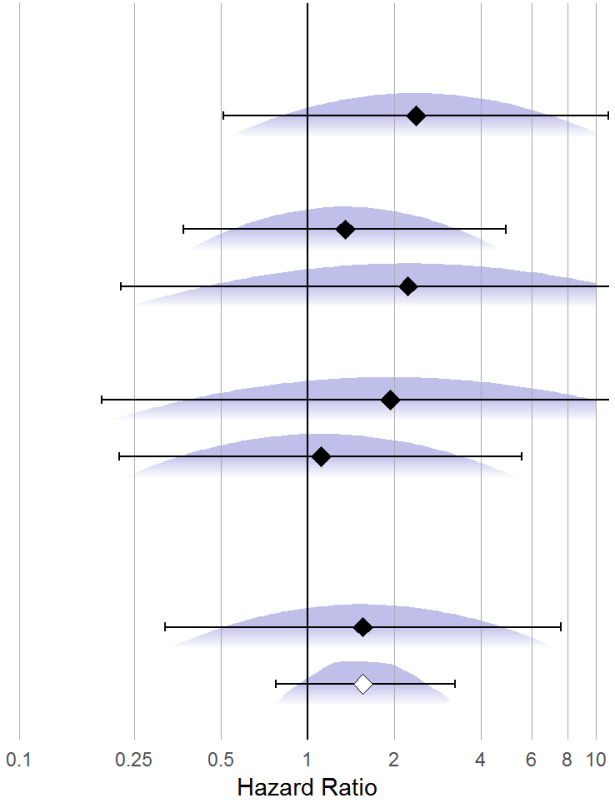
BEAST

Bayesian Evolutionary Analysis Sampling Trees

Visualization

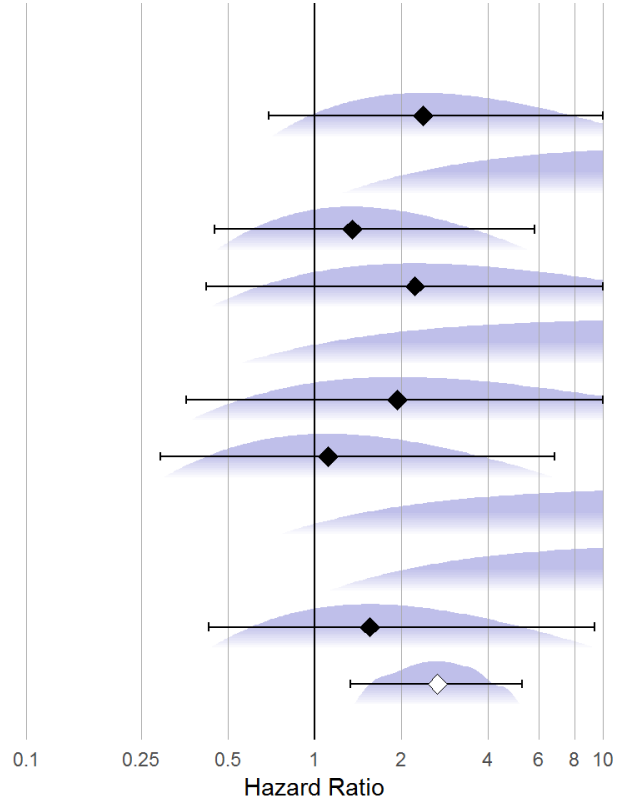
Normal approximation

Source	Hazard Ratio (95% CI)
Site 1	2.38 (0.51 - 11.08)
Site 2	-
Site 3	1.34 (0.37 - 4.87)
Site 4	2.23 (0.22 - 22.08)
Site 5	-
Site 6	1.93 (0.19 - 19.45)
Site 7	1.11 (0.22 - 5.54)
Site 8	-
Site 9	-
Site 10	1.55 (0.32 - 7.55)
Summary (tau = 0.25)	1.55 (0.77 - 3.25)

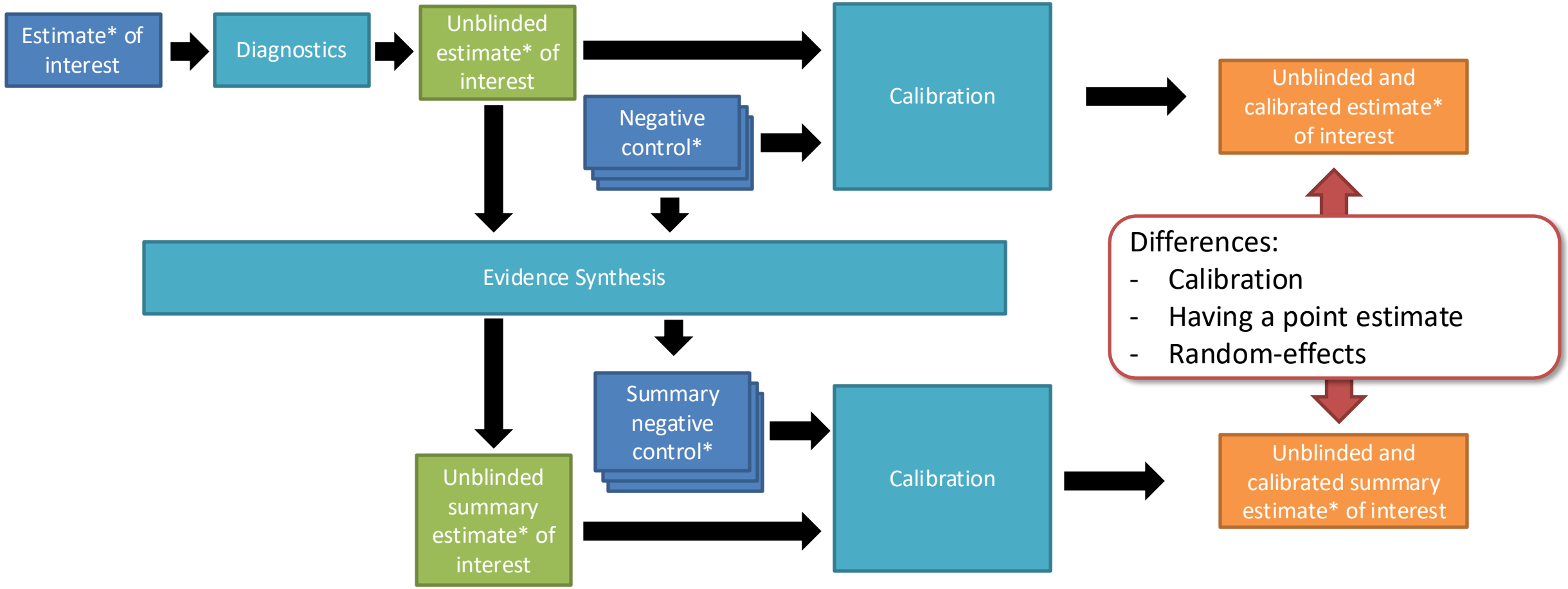


Adaptive grid approximation

Source	Hazard Ratio (95% CI)
Site 1	2.38 (0.69 - 10.00)
Site 2	-
Site 3	1.34 (0.45 - 5.81)
Site 4	2.23 (0.42 - 10.00)
Site 5	-
Site 6	1.93 (0.36 - 10.00)
Site 7	1.11 (0.29 - 6.81)
Site 8	-
Site 9	-
Site 10	1.55 (0.43 - 9.38)
Summary (tau = 0.28)	2.67 (1.33 - 5.26)

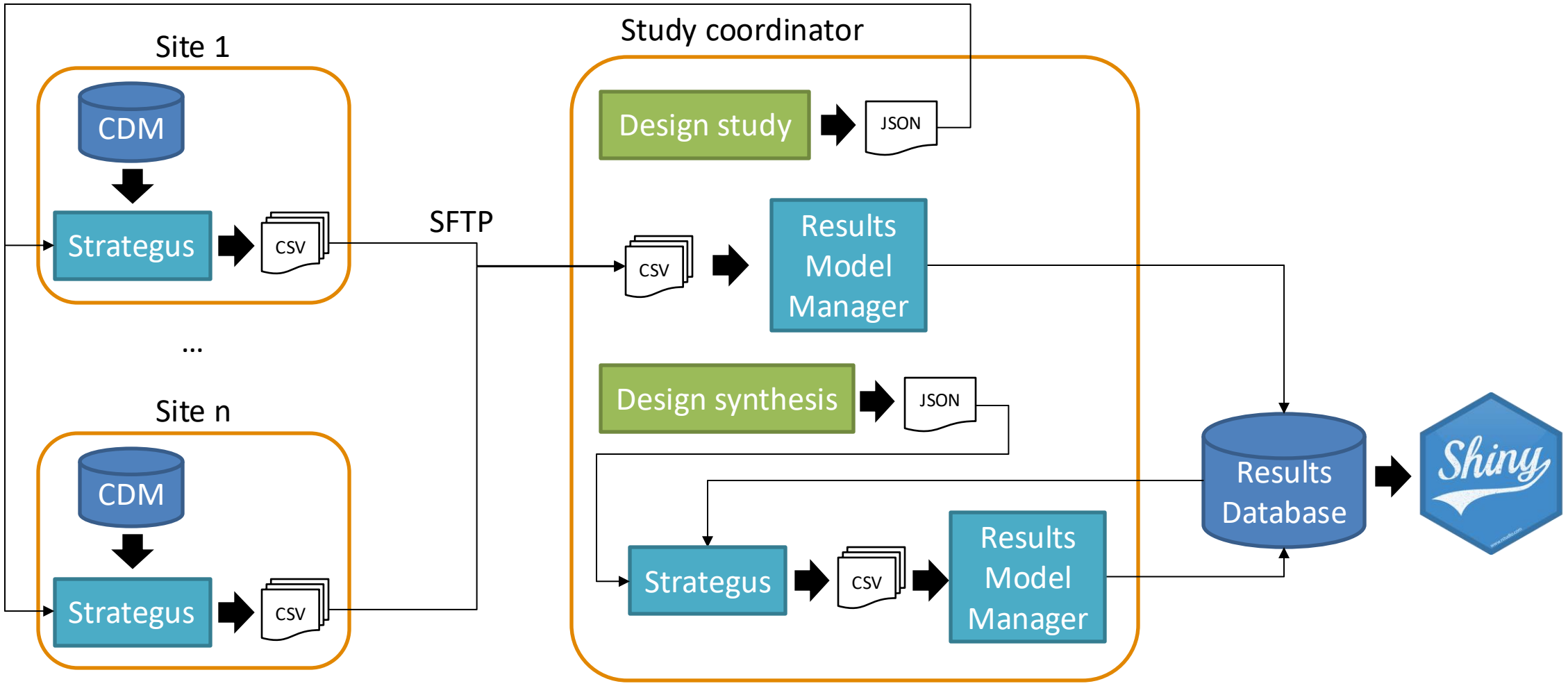


From per-database to summary estimate



* Estimate here means likelihood profile, which may not be expressed as an estimate

Overview


GitHub
Site 1
CDM
Strategus
CSV
...
Site n
CDM
Strategus
CSV
SFTP
Study coordinator
Design study
JSON
CSV
Results Model Manager
Design synthesis
JSON
Strategus
CSV
Results Model Manager
Results Database
Shiny
Firewall

Running EvidenceSynthesis in Strategus

https://github.com/ohdsi-studies/StrategusStudyRepoTemplate/blob/main/EvidenceSynthesis.R

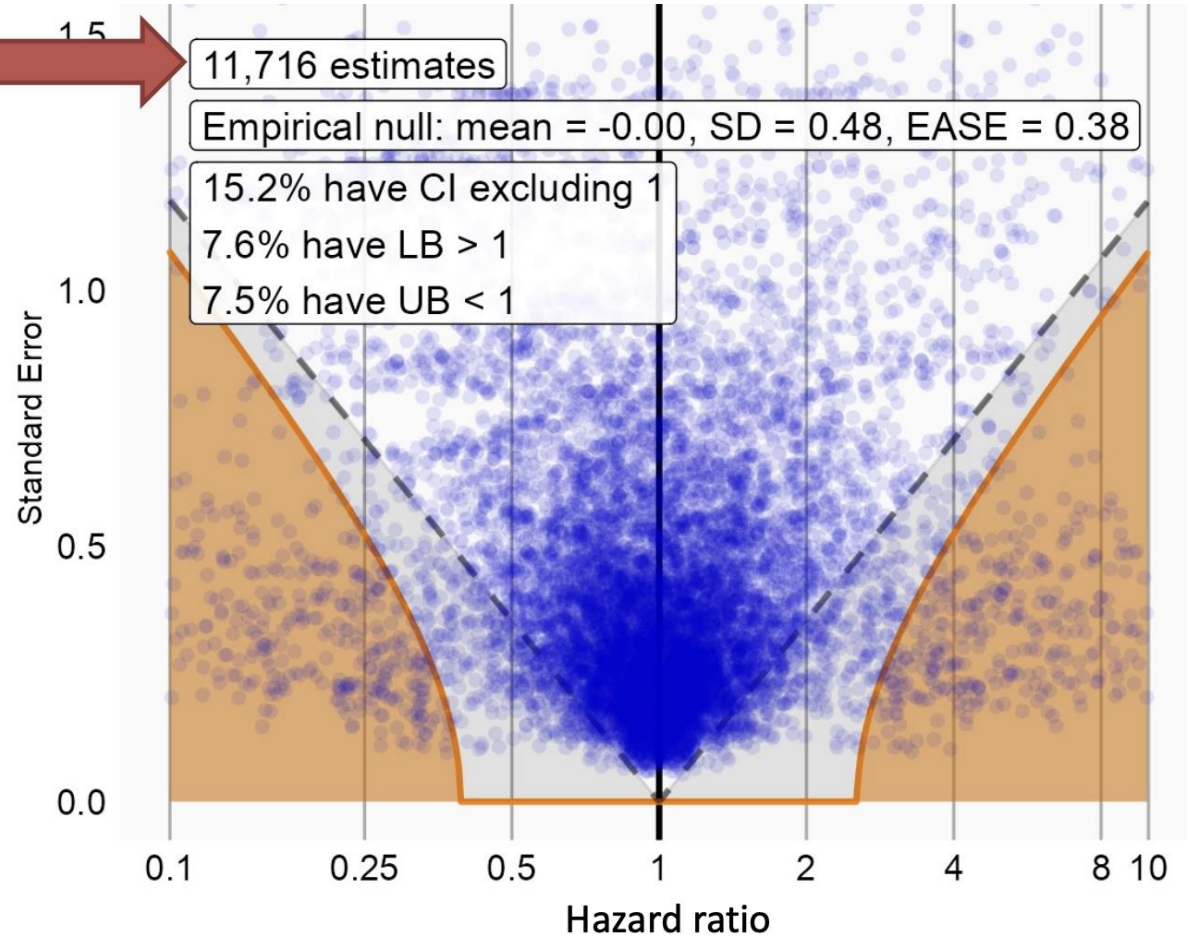
Evidence Synthesis Considerations

Martijn Schuemie

Yong Chen

Motivation -- Estimates Where Null is Likely True

Showing all 11,716 estimates from LEGEND Hypertension where believe the null to be true



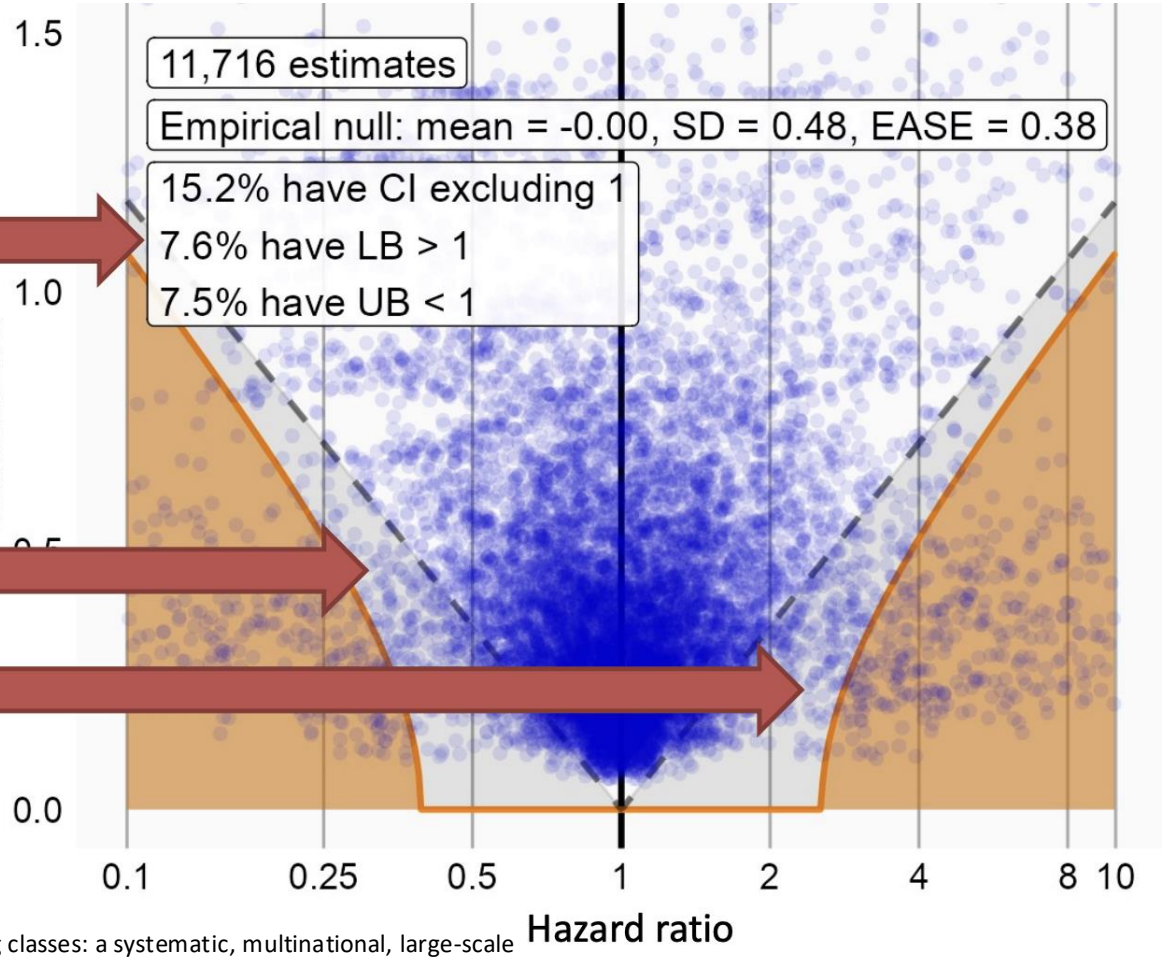
Suchard, Marc A., et al. "Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis." The Lancet 394.10211 (2019): 1816-1826.

Schuemie, Martijn J., et al. "Principles of large-scale evidence generation and evaluation across a network of databases (LEGEND)." Journal of the American Medical Informatics Association 27.8 (2020): 1331-1337.

Motivation -- Estimates Where Null is Likely True

If the null is true for all, we would expect 5% of CIs to not include 1

Estimates below the dashed line have confidence interval (CI) excluding 1



Suchard, Marc A., et al. "Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis." The Lancet 394.10211 (2019): 1816-1826.
Schuemie, Martijn J., et al. "Principles of large-scale evidence generation and evaluation across a network of databases (LEGEND)." Journal of the American Medical Informatics Association 27.8 (2020): 1331-1337.

Definition of Negative Control Pairs

- A negative control pair consists of two random variables (e.g., a drug and an outcome) where there is no causal relationship between the two

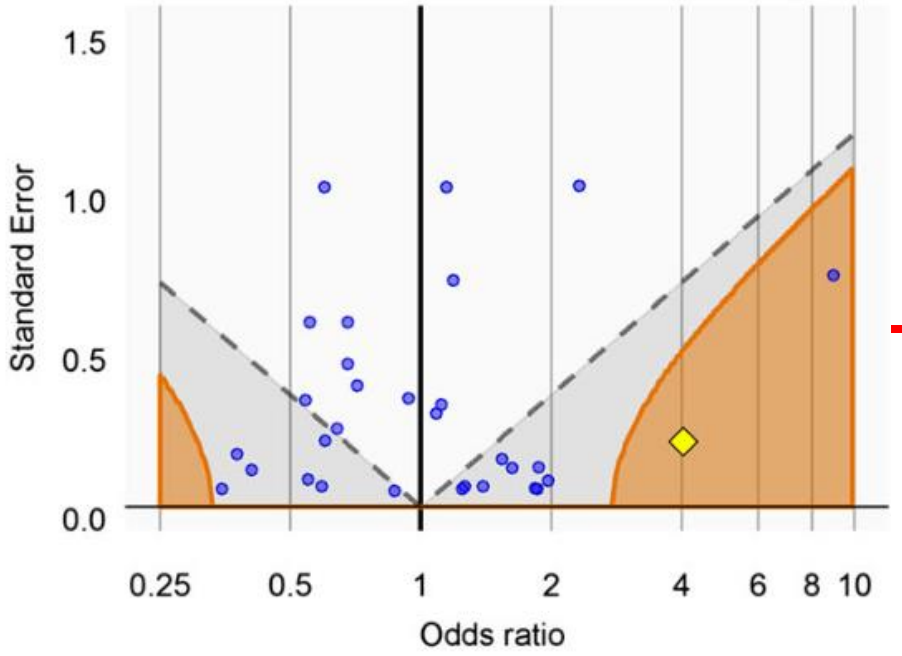
- Key Benefits:
  - Detecting Systematic Biases:
    - Negative controls help identify biases, such as confounding and selection bias, which can undermine the reliability and reproducibility of observational studies (Ioannidis, 2005)
  - Calibrating P-values:
    - Known negative control pairs can be used to adjust p-values, making the testing of causal relationships between new drug-outcome pairs more reliable (Schuemie et al., 2014)

Ioannidis, John PA. "Why most published research findings are false." PLoS Medicine 2, no. 8 (2005): e124.
Schuemie, Martijn J., Patrick B. Ryan, William DuMouchel, Marc A. Suchard, and David Madigan. "Interpreting observational studies: why empirical calibration is needed to correct p-values." Statistics in Medicine 33, no. 2 (2014): 209-218.

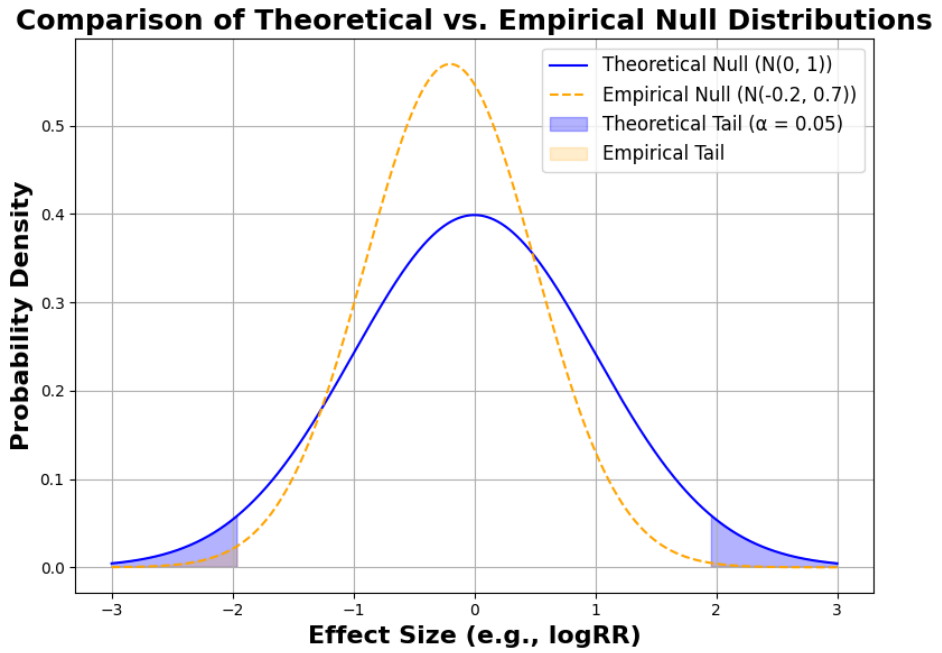
P-value Calibration from Negative Control Outcomes

- From an observational study with J known NCOs, let y_j denote the estimated log effect estimate (relative risk or incidence rate ratio) and σ_j be its estimated standard error with respect to the j-th NCO, $j = 1, \ldots, J$

Estimated negative control outcomes (y_j, σ_j^2)'s



Computed empirical null distribution $\theta_j \sim N(\mu, \tau^2)$



Schuemie, Martijn J., Patrick B. Ryan, William DuMouchel, Marc A. Suchard, and David Madigan. "Interpreting observational studies: why empirical calibration is needed to correct p-values." Statistics in Medicine 33, no. 2 (2014): 209-218.

P-value Calibration from Negative Control Outcomes

- From an observational study with J known NCOs, let y_j denote the estimated log effect estimate (relative risk or incidence rate ratio) and σ_j be its estimated standard error with respect to the j-th NCO, $j = 1, \ldots, J$
  - The marginal distribution of y_j is $N(\mu, \sigma_j^2 + \tau^2)$
  - Under the null hypothesis, the estimate of new outcome $y_{J+1} \sim N(\mu, \sigma_{J+1}^2 + \tau^2)$, where σ_{J+1} denotes the estimated standard error of y_{J+1}

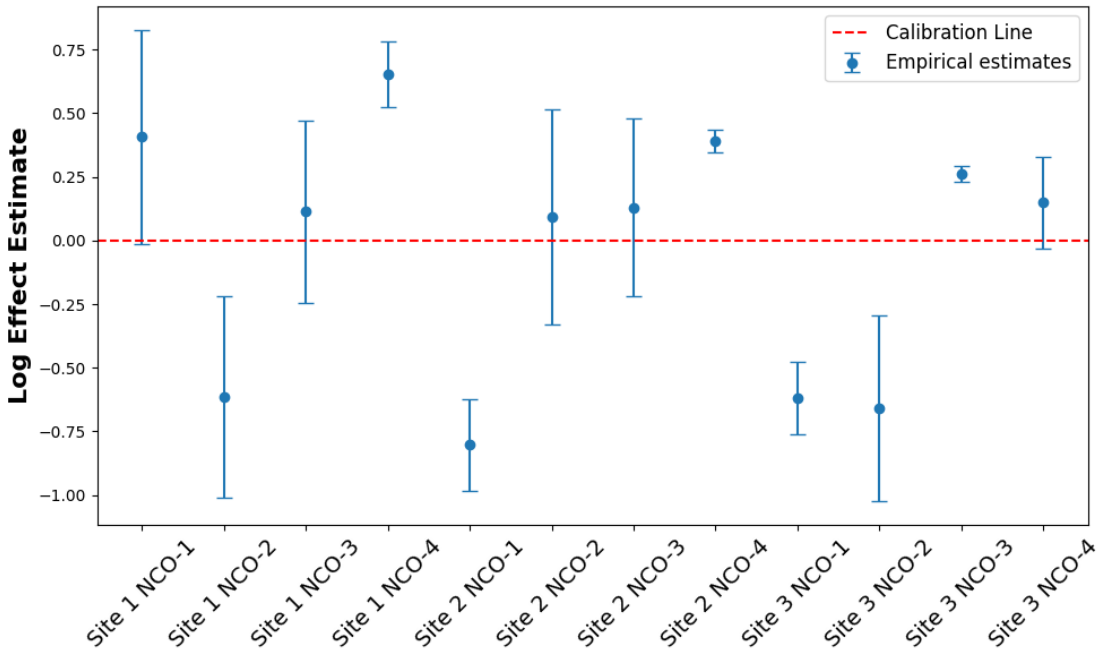
One-sided p-value for the new outcome is

$$\boldsymbol{\Phi}\left(\frac{\boldsymbol{y_{J+1}} - \boldsymbol{\hat{\mu}}}{\sqrt{\boldsymbol{\widehat{\sigma_{J+1}}^2} + \boldsymbol{\tau^2}}}\right)$$

Schuemie, Martijn J., Patrick B. Ryan, William DuMouchel, Marc A. Suchard, and David Madigan. "Interpreting observational studies: why empirical calibration is needed to correct p-values." Statistics in Medicine 33, no. 2 (2014): 209-218.

Bias Reduction using Multi-site Real-World Data

- Our research focuses on a distributed research network comprising a total of *K* sites, with each site contributing *J* uniformly selected Negative Control Outcomes (NCOs) with the interested exposure.



- For the *j*-th NCO at the *k*-th site, observe

$$\{(y_{k,j}, \sigma_{k,j}^2): k = 1, \cdots, K; j = 1, \cdots, J\}$$

where $y_{k,j}$ represents the estimated log effect estimate for the *j*-th NCO, and $\sigma_{k,j}^2$, *j* represents the associated asymptotic variances

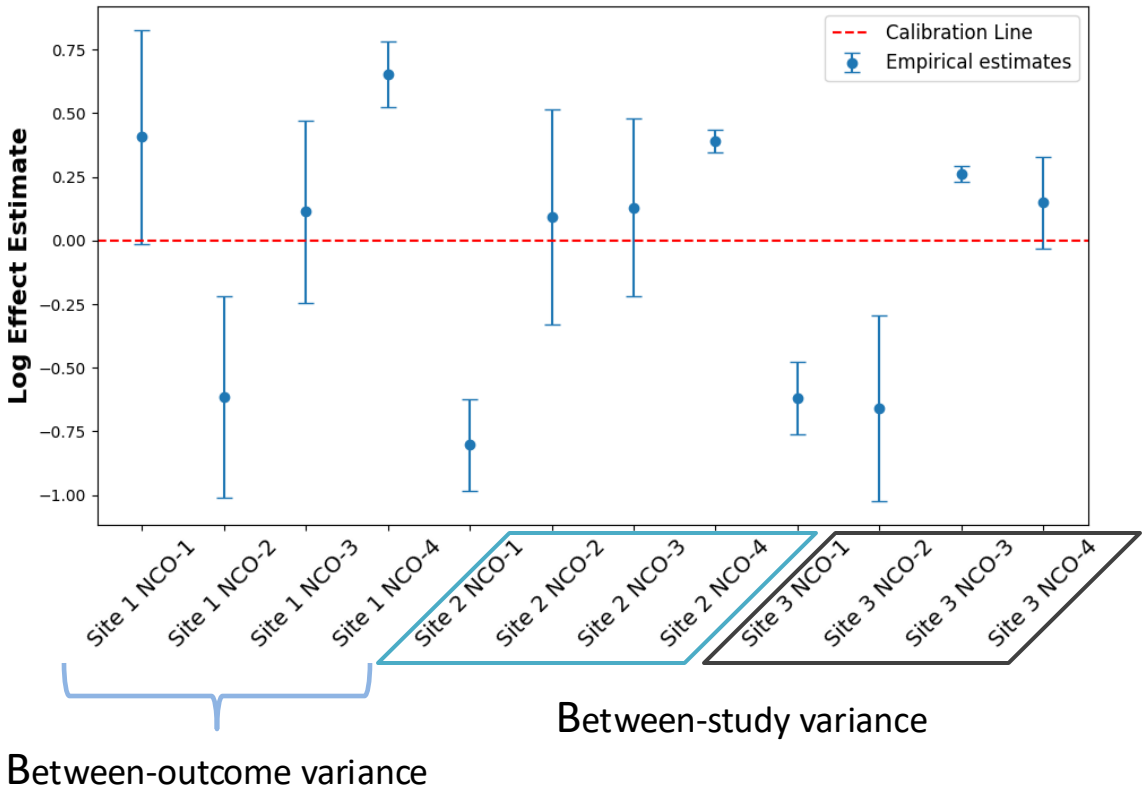
- Given a new drug-outcome pair

$$\{(y_{k,J+1}, \sigma_{k,J+1}^2): k = 1, \cdots, K\}$$

our objective is to utilize the previous estimates to conduct empirical calibration within distributed research networks

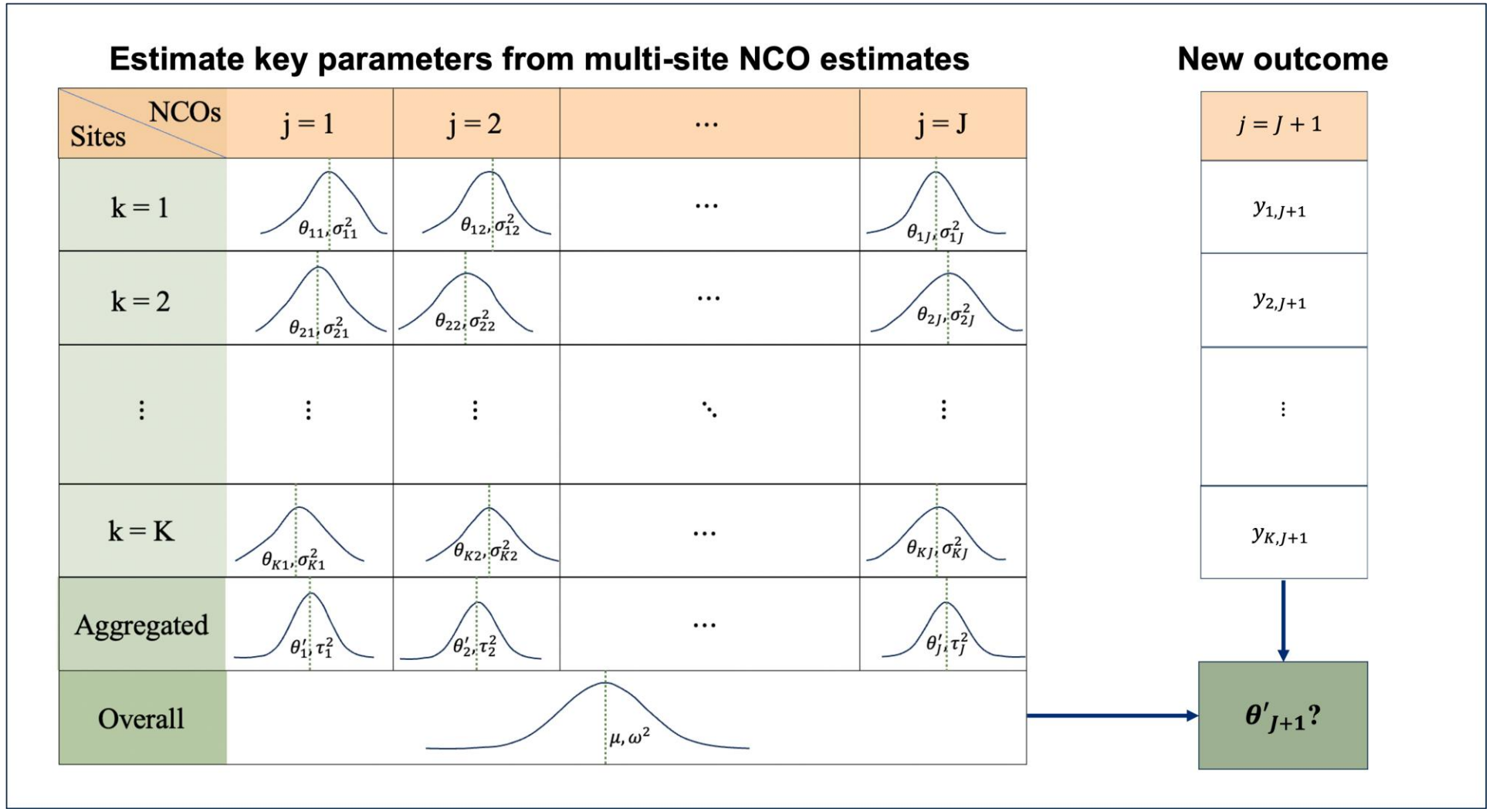
Bias Reduction using Multi-site Real-World Data

- Our research focuses on a distributed research network comprising a total of *K* sites, with each site contributing *J* uniformly selected Negative Control Outcomes (NCOs) with the interested exposure.



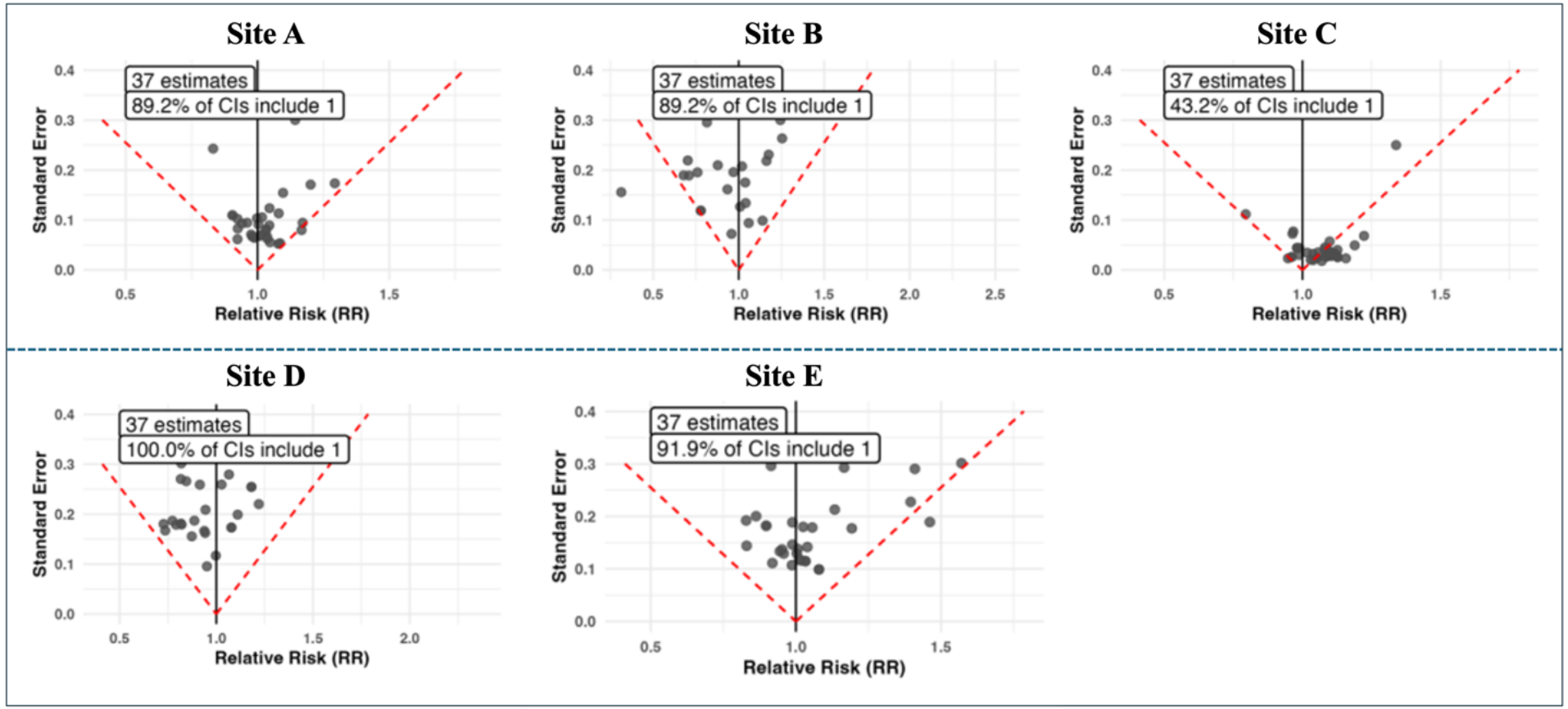
- A critical gap is that there is no method to address systematic bias specifically for multi-site settings
- In multi-site research networks, we assume that heterogeneity comes from two factors, between-study variance and between-outcome variance

Proposed Multi-site Bias-Reduction Method



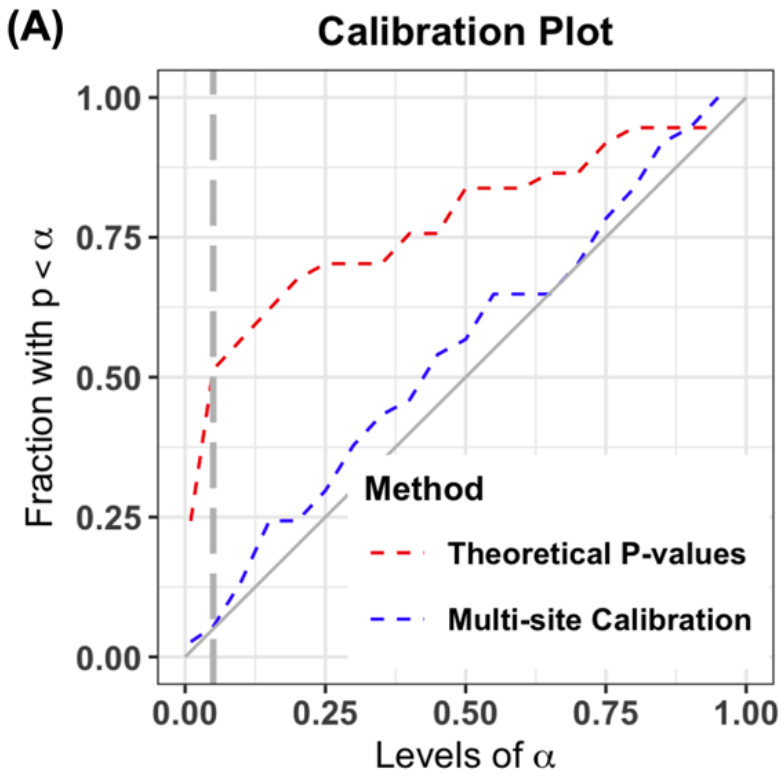
Real-world Application

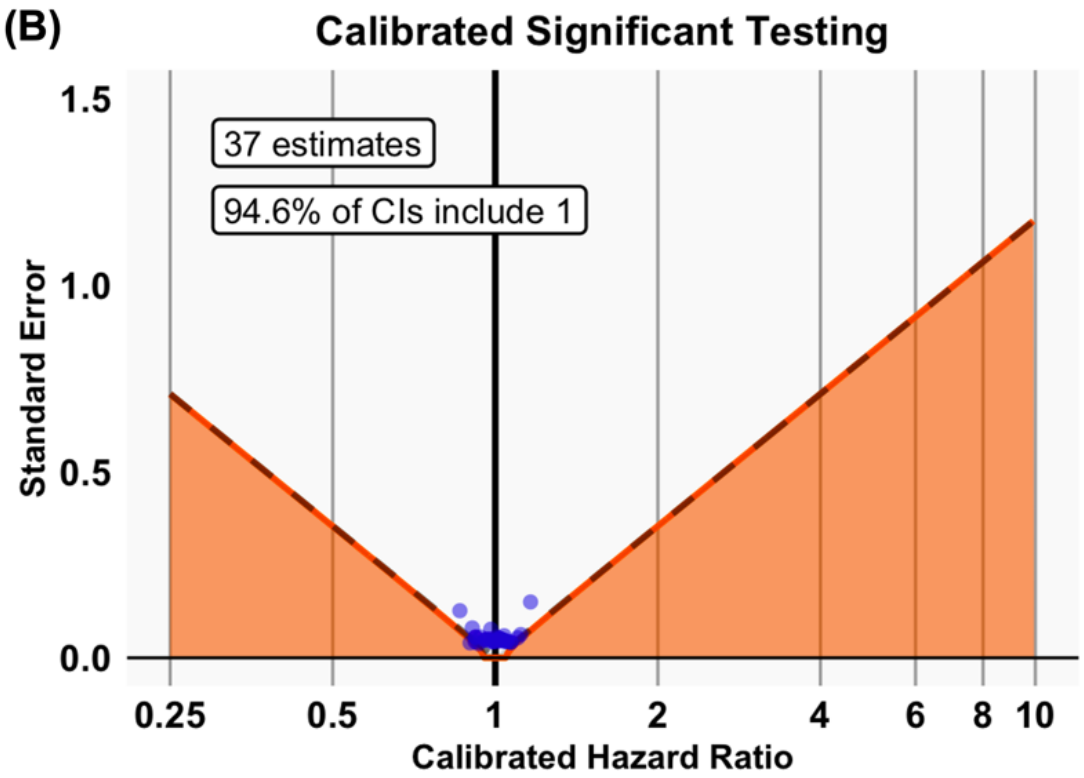
- Empirical Calibration via a list of 37 negative control outcomes from five federated institutions
  - Exposure: second-line antihyperglycemic agents in type 2 diabetes mellitus (T2DM)



Real-world Application

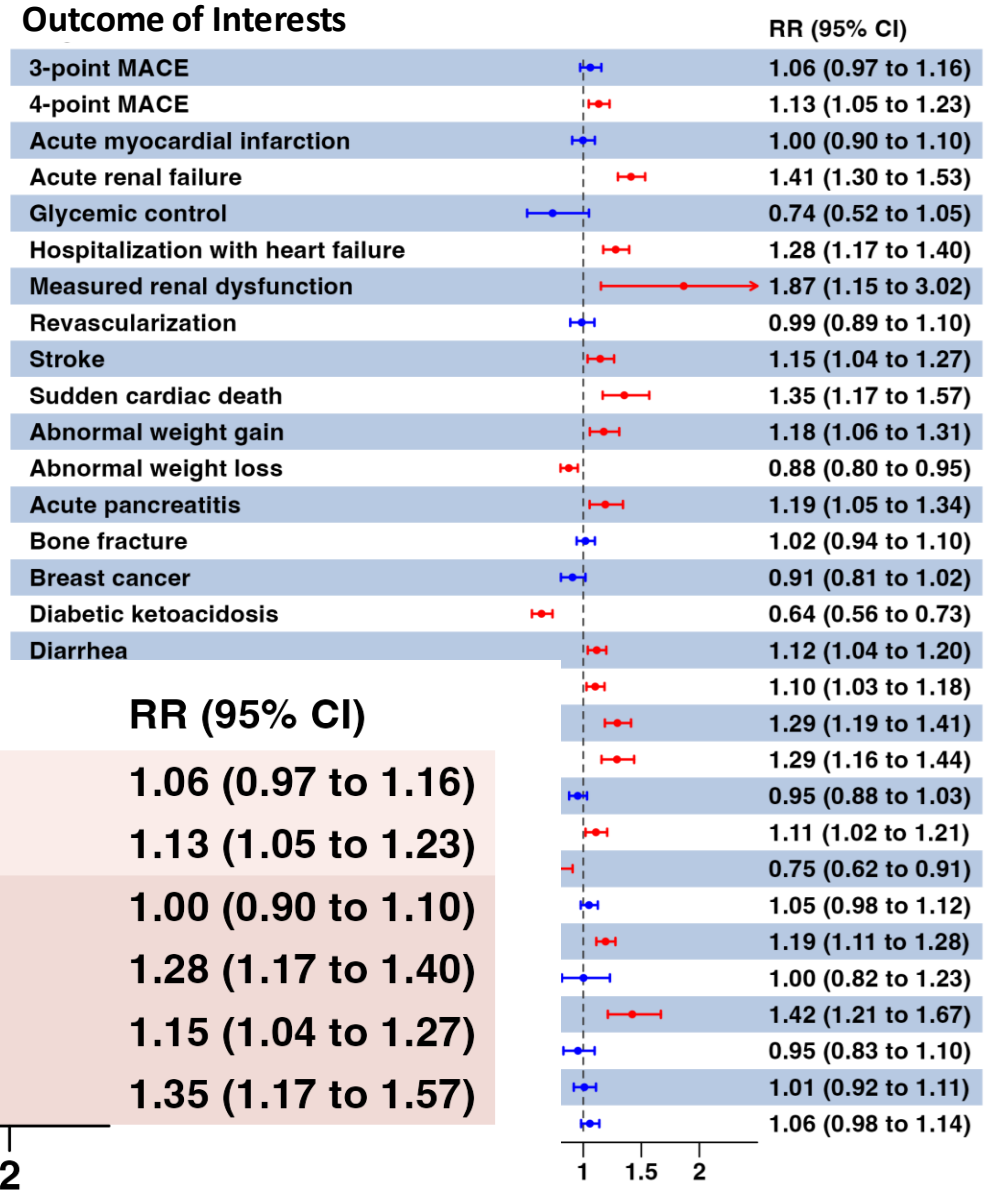
- Empirical Calibration via a list of 37 negative control outcomes from five federated institutions
  - Exposure: second-line antihyperglycemic agents in type 2 diabetes mellitus (T2DM)

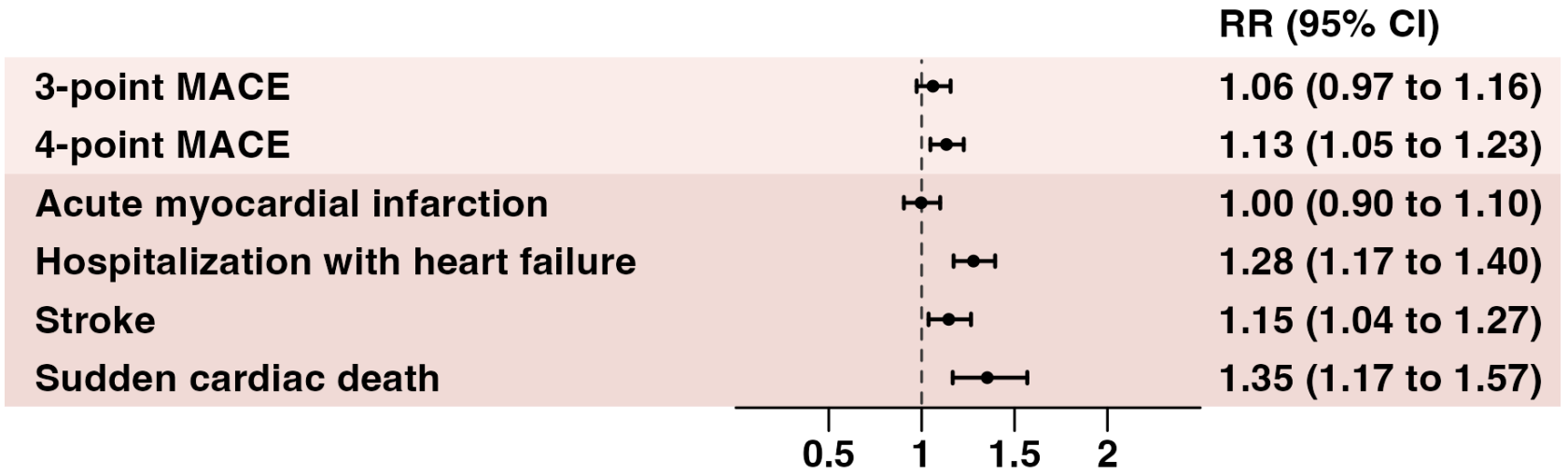




Real-world A

- Answer clinical question: compare the cardiovascular effectiveness of dipeptidyl peptidase-4 inhibitors (DPP4is) versus Sodium-glucose





View results

https://data.ohdsi.org/AntiVegfKidneyFailure/

Proposed Multi-site Bias-Reduction Method

- To account for both between-study variance and between-outcome variance, hereafter we describe our proposed method via a three-level hierarchical formulation,

$$y_{kj} | \theta_{kj} \sim N(\theta_{kj}, \sigma_{kj}^2), \theta_{kj} \sim N(\theta_j', \tau^2), \theta_j' \sim N(\mu, \omega^2)$$

for j $= 1, \dots, J, k = 1, \dots, K$

  - $y_{k,j}$ denotes the corresponding estimated log effect estimate
  - $\sigma_{k,j}^2$ denotes its observed estimated standard error
  - $\theta_{k,j}$ denotes the probabilistic limit of which is the underlying bias associated with the *j*-th drug-outcome pair in the *k*-th site
  - $\theta_{k,j}$ arises from a Gaussian distribution with mean θ_j' across sites and between-site variance τ^2
  - θ_j' generates from the Gaussian distribution with mean μ as the fixed effect of the *J* NCOs across *K* sites and the between-NCO variance ω^2