Negative controls in pharmacoepidemiology

ENCePP Plenary 2023

Presented by Daniel Morales on 1 December 2023
1 TDA/RWE Workstream – European Medicines Agency, 2 University of Dundee

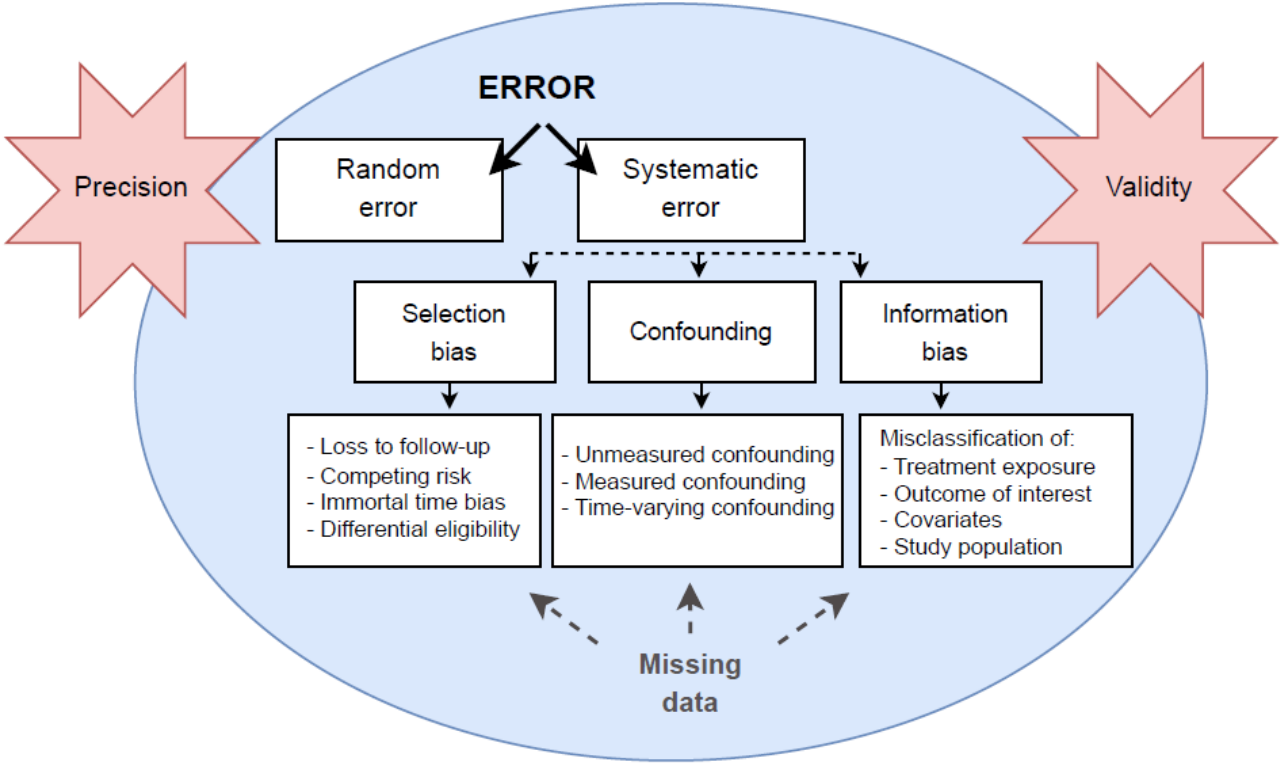


Figure adapted from FDA https://healthpolicy.duke.edu/sites/default/files/2023-03/NegativeControlWorkshopSlideDeck.pdf
Principles adapted from ROBINS-I domains doi: https://doi.org/10.1136/bmj.i4919

Strategies to address error

Design phase

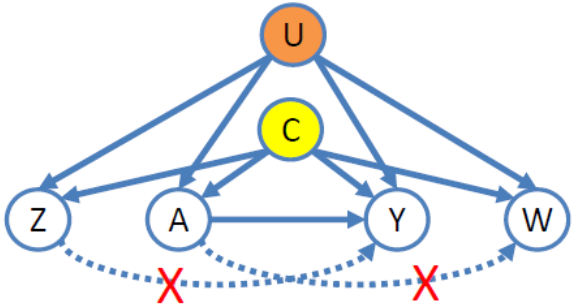
- Self-controlled case-only designs
- Instrumental variables
- Difference-in-difference
- Negative controls

Analysis phase

- Propensity scores
- Sensitivity analyses
- Quantitative bias analysis
- Negative controls?

Principles adapted from FDA https://healthpolicy.duke.edu/sites/default/files/2023-03/NegativeControlWorkshopSlideDeck.pdf and https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml

Negative controls



Negative control exposures (NCE, Z)

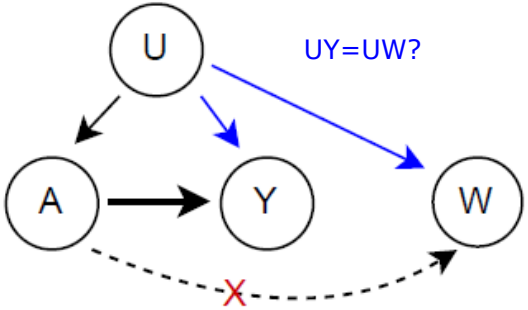
- A variable that is not causally related to outcome (Z does not cause Y)

Negative control outcomes (NCO, W)

- A variable that is not causally related to exposure (A does not cause W)

NCs identify unmeasured confounding (U) under the assumption all measured confounding (C) has been accounted for

NCs identify bias under the assumption they have similar underlying bias structure as the primary exposure-outcome association



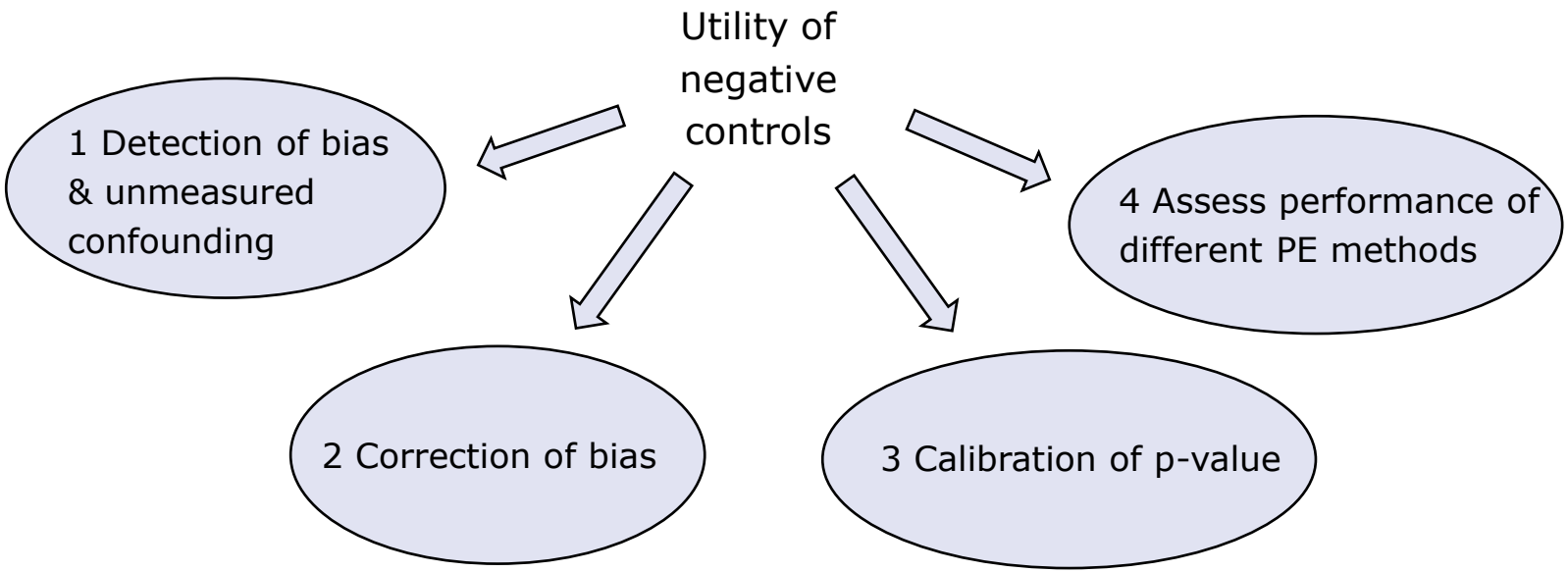
State of use of negative controls in pharmacoepidemiology

Systematic review including 184 papers up to Sep 2020 (Zafari et al. 2023)

- Cohort study 62.5%, case-control study 6.5%, self-controlled case series 4.9%
- Administrative data 62%, EHRs 33.2%
- NCO 49.5%, NCE 29.3%, NCO+NCE 18.5%
- Justification: unmeasured confounding 50.5%

doi.org/10.1093/aje/kwad201

State of use of negative controls in pharmacoepidemiology



Identified in doi.org/10.1093/aje/kwad201

Utility of Negative Controls

Detection of bias

Reject a null hypothesis if association was significant given a type I error of 5% ($p<0.05$)

Direction of the point estimate irrespective of statistical significance

Correct/reduce bias

Difference in difference method

Parametric empirical null distribution of bias method (multiple NCs)

Semiparametric/non-parametric models with multiply robust estimators (double NCs)

P-value calibration

Empirical calibration methods used for p-value calibration to correct for the type I error (multiple NCs)

Account for both random and systematic error

Assess performance of PE methods

NCs act as reference standard with known drug reactions

Early applications to signal detection methods

Identified in doi.org/10.1093/aje/kwad201

Examples

Detection of bias

Nested case control study examining the association between FQ and co-amoxiclav (NCE) with tendon rupture

Table 2 Incidence rate ratios for the association between tendon rupture and current systemic fluoroquinolone and co-amoxiclav exposure

Tendon rupture	Exposed cases/total	Exposed controls/total	Crude IRR	Adjusted IRR	Adjusted *p* value
Any tendon rupture					
Fluoroquinolones	111/4836	236/18,356	1.79 (1.41–2.27)	1.61 (1.25–2.09)	<0.001
Co-amoxiclav	98/4836	314/18,356	1.15 (0.90–1.45)	1.02 (0.79–1.31)	0.900
Achilles tendon rupture					
Fluoroquinolones	67/1577	82/6007	3.50 (2.45–5.02)	3.14 (2.11–4.65)	<0.001
Co-amoxiclav	38/1577	114/6007	1.19 (0.81–1.77)	1.00 (0.64–1.57)	0.989

https://doi.org/10.1007/s40261-018-0729-y

Examples

Correction of bias

Re-analysis of two randomized, Phase III, placebo-controlled, multicenter clinical trials of RV1 and RV5 to understand vaccine dose timing and severe rotavirus gastroenteritis incidence

Placebo arms used as a negative control to adjust for both confounding/administrative censoring

Estimates calibrated RDs (difference in difference) RRs (ratio of ratios)

Published in final edited form as:
Epidemiology. 2018 November ; 29(6): 867–875. doi:10.1097/EDE.0000000000000909.

Timing of Rotavirus Vaccine Doses and Severe Rotavirus Gastroenteritis among Vaccinated Infants in Low- and Middle-Income Countries

Joann F. Gruber[1], **Sylvia Becker-Dreps**[1,2], **Michael G. Hudgens**[3], **M. Alan Brookhart**[1], **James C. Thomas**[1,4], and **Michele Jonsson Funk**[1]

[1]Department of Epidemiology, University of North Carolina (UNC)-Chapel Hill, Chapel Hill, North Carolina, USA

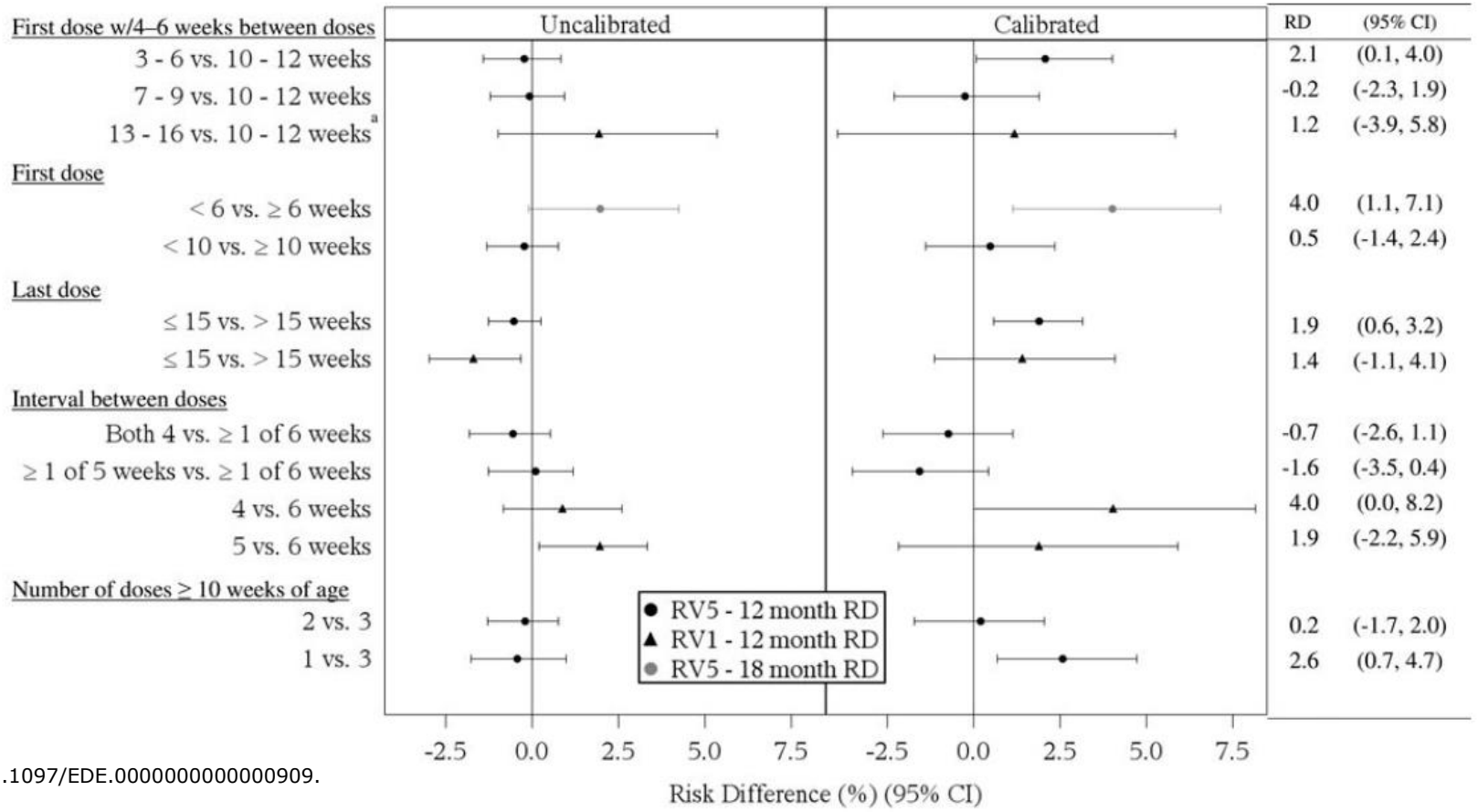
[2]Department of Family Medicine, UNC-Chapel Hill, Chapel Hill, North Carolina, USA

[3]Department of Biostatistics, UNC-Chapel Hill, Chapel Hill, North Carolina, USA

[4]MEASURE Evaluation, Carolina Population Center, UNC-Chapel Hill, NC, USA

Abstract

Background—Altering rotavirus vaccine schedules may improve vaccine performance in low- and middle-income countries. We analyzed data from clinical trials of the monovalent (RV1) and pentavalent (RV5) rotavirus vaccines in low- and middle-income countries to understand the association between vaccine dose timing and severe rotavirus gastroenteritis incidence.



doi: 10.1097/EDE.0000000000000909.
9

Strategies to validate negative control assumptions

Lack of causality

1. Logically implausible association
2. Clinically implausible association
3. No previous evidence of causality

Published literature/drug product labels

Spontaneous reporting systems

Expert opinion

Shared bias structure

1. Compare distributions of measured covariates/confounders
2. Account for different distributions of covariates/confounders

Example: study examining antihypertensive medication adherence and injurious falls re-weighted NCE samples with IPTW to adjust for different covariate distributions. doi:10.1136/bmjopen-2018-022927

Other perspectives: Double negative controls and DANCE

Proximal causal inference framework

- Assumption depends on accurately measuring covariates that may be imperfect proxies of confounders
- Leverage proxies to adjust for suspected unmeasured confounding

DNC combines information from NCE and an NCO nonparametrically to account for unmeasured confounders

Example use of DNCs in a test-negative design to evaluate vaccine effectiveness

Table 1. Estimated VE and 95% confidence intervals of the negative control estimator, logistic regression and IPTW estimator with the University of Michigan Health System data.

	Negative control	Logistic regression	IPTW
Pfizer-BioNTech	80.2% (78.3%, 81.9%)	74.1% (72.3%, 75.8%)	67.9% (66.0%, 69.6%)
Moderna	89.7% (88.1%, 91.1%)	78.8% (76.8%, 80.7%)	72.4% (69.5%, 75.1%)
Janssen (J & J)	65.8% (54.6%, 74.1%)	56.3% (48.4%, 62.9%)	48.2% (28.3%, 62.6%)

NCE=immunization visits before December 2020
NCO=having >=1 of the following conditions after April 5, 2021: arm/leg cellulitis, eye/ear disorder, gastroesophageal, disease, atopic dermatitis, injuries, and general adult examination visits.

Li 2023: DOI:10.1080/01621459.2023.2220935

Other perspectives: Double negative controls and DANCE

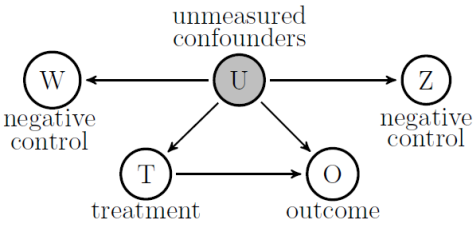
Proximal causal inference framework

- Assumption depends on accurately measuring covariates that may be imperfect proxies of confounders
- Leverage proxies to adjust for suspected unmeasured confounding

DNC combines information from NCE and an NCO to account for unmeasured confounders nonparametrically

Data-driven Automated Negative Control Estimation (DANCE)

- Data-driven method to potentially identify "disconnected" negative controls (DNC)
- Applies statistical test to validate whether NC meet assumptions for causal inference
- Estimate causal effect of treatment-outcome relationship



Kummerfeld 2022: arXiv:2210.00528

No standardized guidance

References

- FDA convened workshop 2023. Understanding the Use of Negative Controls to Assess the Validity of Non-Interventional Studies of Treatment Using Real-World Evidence. https://healthpolicy.duke.edu/events/understanding-use-negative-controls-assess-validity-non-interventional-studies-treatment
- Sterne JA, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355:i4919. doi: 10.1136/bmj.i4919.
- ENCePP methods guide revision 11 https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml
- Zafari Z, et al. The State of Use and Utility of Negative Controls in Pharmacoepidemiologic Studies. Am J Epidemiol. 2023 Oct 17:kwad201. doi: 10.1093/aje/kwad201.
- Morales DR, et al. Relative and Absolute Risk of Tendon Rupture with Fluoroquinolone and Concomitant Fluoroquinolone/Corticosteroid Therapy: Population-Based Nested Case-Control Study. Clin Drug Investig. 2019 Feb;39(2):205-213. doi: 10.1007/s40261-018-0729-y.
- Gruber JF, et al. Timing of Rotavirus Vaccine Doses and Severe Rotavirus Gastroenteritis Among Vaccinated Infants in Low- and Middle-income Countries. Epidemiology. 2018 Nov;29(6):867-875. doi: 10.1097/EDE.0000000000000909.
- Dillon P, et al. Association between gaps in antihypertensive medication adherence and injurious falls in older community-dwelling adults: a prospective cohort study. BMJ Open. 2019 Mar 4;9(3):e022927. doi: 10.1136/bmjopen-2018-022927.
- Li KQ, Shi X, Miao W, Tchetgen ET. Double Negative Control Inference in Test-Negative Design Studies of Vaccine Effectiveness. ArXiv [Preprint]. 2023 Mar 8:arXiv:2203.12509v4.
- Kummerfeld E, Shi X. Data-driven Automated Negative Control Estimation (DANCE): Search for, Validation of, and Causal Inference with Negative Controls. 2022. https://doi.org/10.48550/arXiv.2210.00528

Negative controls at scale

Marc A Suchard, MD PhD

Professor, Department of Biomathematics, University of California, Los Angeles

Research Investigator, VA Informatics and Computing Infrastructure, US Department of Veterans Affairs

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Disclosures:

- I am partially supported through a US Food & Drug Administration contract to evaluate methods for vaccine safety surveillance
- I also contract with Johnson & Johnson for work unrelated to this presentation

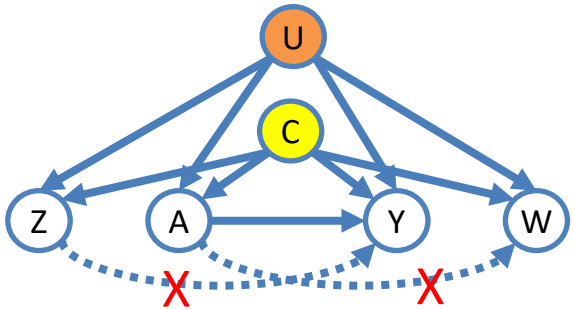
Conceptual model for negative controls

Negative control: an **exposure-outcome pair** with *a priori* **no causal relationship**

We typically define negative controls relative to a causal question of interest (e.g. 'does exposure A cause outcome Y'?):

- Negative exposure control: Z does not cause Y
- Negative outcome control: A does not cause W

Some add: Shares **exact same unmeasured confounding U**, while assuming **perfect control over measured confounders C**



Will discuss an approach to find error due to confounding, whether it was explicitly adjusted for or not

Negative Controls

A Tool for Detecting Confounding and Bias in Observational Studies

Lipsitch, Marc[a,b,c]; Tchetgen Tchetgen, Eric[a,c,d]; Cohen, Ted[a,c,e]

Author Information

Epidemiology 21(3):p 383-388, May 2010. | *DOI:* 10.1097/EDE.0b013e3181d61eeb

One vs many negative control experiments

If identical (unmeasured) confounding known: USE IT!

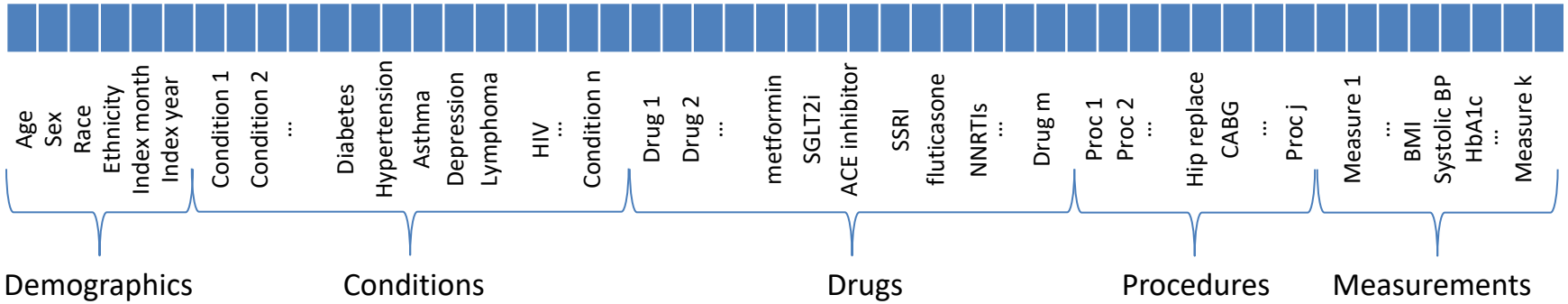
- Confounding structure is ***largely unknown*** and ***complex***
  - Any two experts will disagree, and both will likely miss important confounders
  - Seldom powered to measure it (outcome often too rare) (e.g. unlikely that DANCE will work)
- Unlikely that such a perfect negative control exists (my opinion)

Data-driven Automated Negative Control Estimation (DANCE): Search for, Validation of, and Causal Inference with Negative Controls

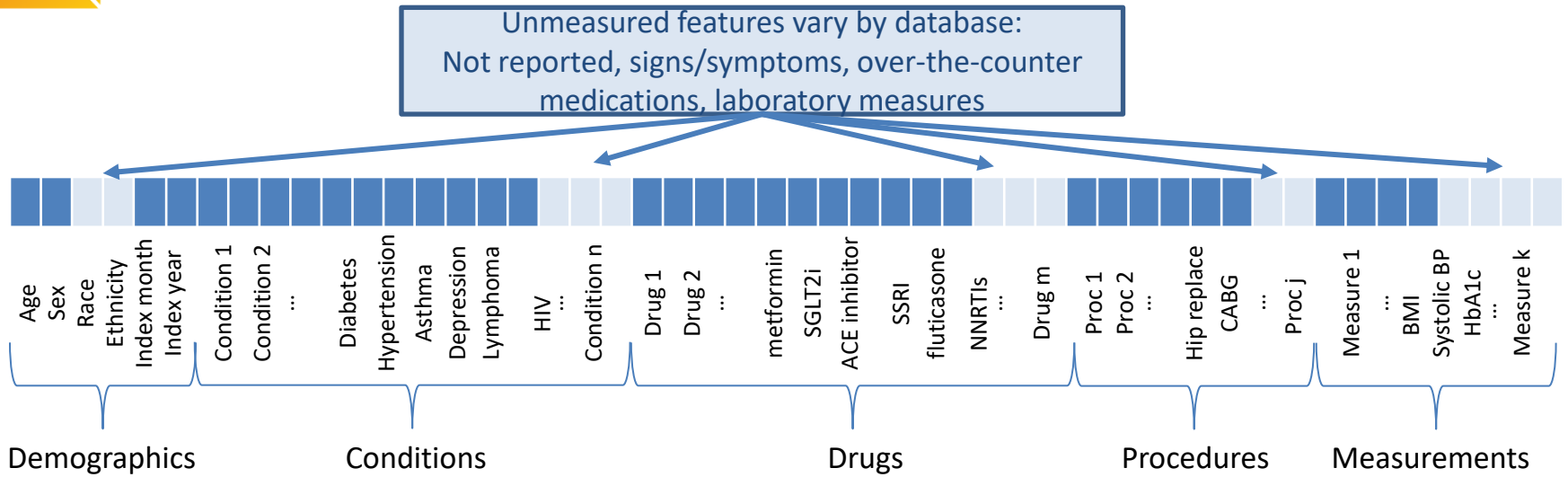
Erich Kummerfeld[1], Jaewon Lim[2], and Xu Shi[3]
[1]Institute for Health Informatics, University of Minnesota
[2]Department of Biostatistics, University of Washington
[3] Department of Biostatistics, University of Michigan

But there is a whole universe (population) of potential negative controls out there ….

Baseline characteristics (candidate confounders) in observational healthcare data



Not all baseline characteristics are measured in all data sources



- Most source have unmeasured features that are desired for any given study
- Measured features often total ***>10,000s*** in claims or EHR sources
- Many unmeasured features are ***indirectly measured*** if correlated with other measured features

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

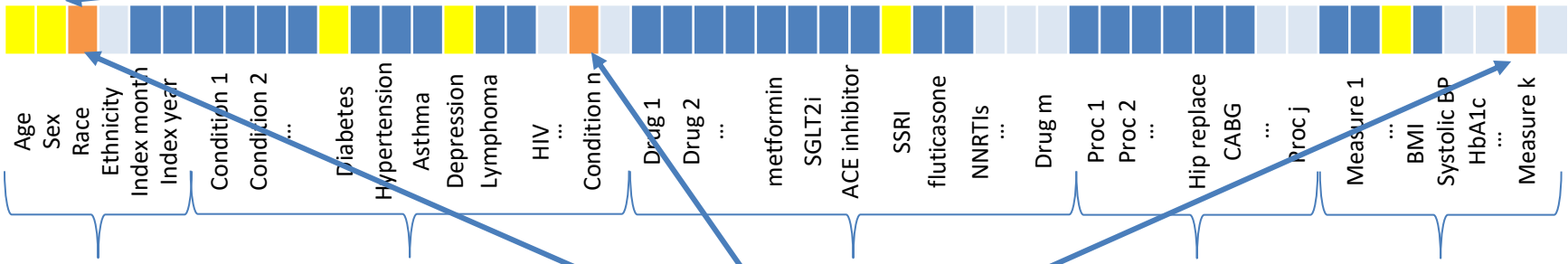
Adjusting for indirectly measured confounding using large-scale propensity score

Linying Zhang [a], Yixin Wang [b], Martijn J. Schuemie [c], David M. Blei [d,e], George Hripcsak [a,f,*]

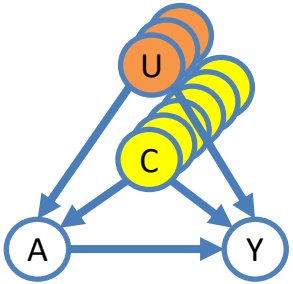
Confounders are baseline characteristics associated with exposure and outcome

C: Observed Confounders for Treatment A → Outcome Y

Treatment A → Outcome Y:



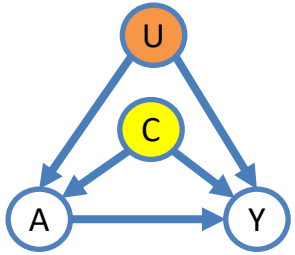
U: Unobserved Confounders for Treatment A → Outcome Y

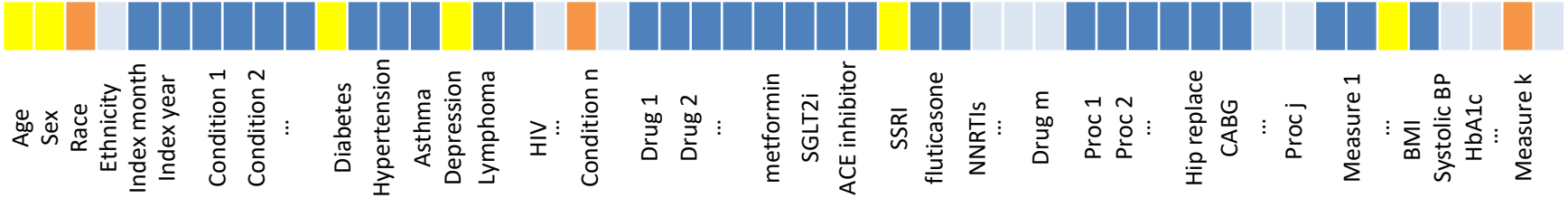


← really large DAG (but for simplicity shown with single U and C)

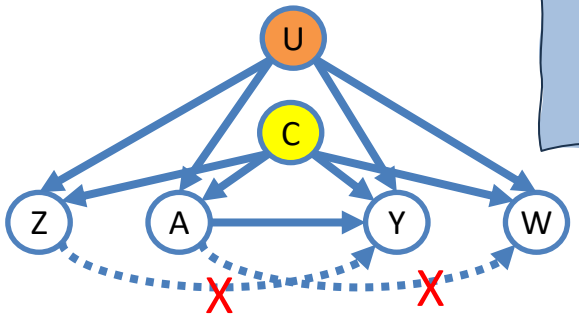
Desired exchangeability of confounding with negative controls

Treatment A → Outcome Y:





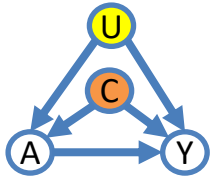
Negative control exposure Z → Outcome Y
OR
Treatment A → Negative control outcome W
OR
Negative control exposure Z → Negative control outcome W



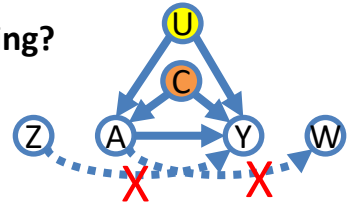
Unlikely to find at scale!

Relaxing assumption about confounding structure, we can still learn about the reliability of a study

Treatment A → Outcome Y:

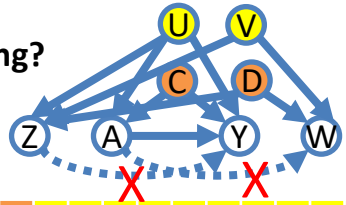


What if negative control has no confounding?



If a method produces a biased estimate in this case, would not one be concerned about the target estimate?

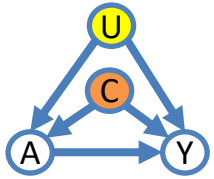
What if negative control has more confounding?



If a method produces an unbiased estimate in this case, would not one be reassured about the target estimate?

Value of using a large set of negative controls: confounding structure is unknown and can vary

Treatment A → Outcome Y:



Negative control exposure Z → Outcome Y OR Treatment A → Negative control outcome W:

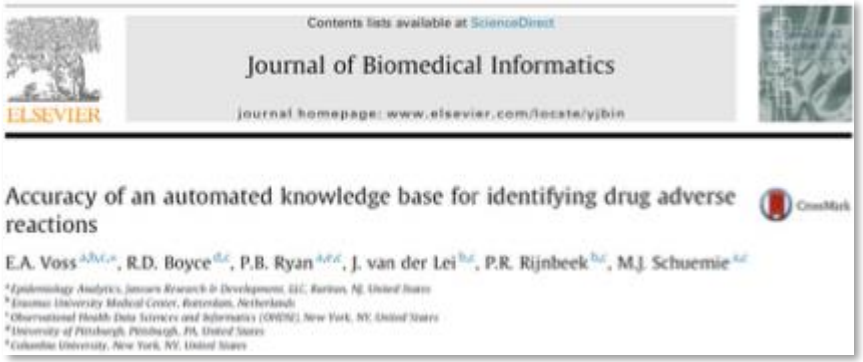
NC_1
NC_2
NC_3
NC_4
NC_5
⋮
NC_n

Estimated bias across a range of confounding structures characterizes the residual systematic error distribution expected for a study

Negative controls in a comparative cohort study

- If neither target nor comparator causes the outcome, the hazard ratio / incidence rate ratio / odds ratio should be 1
- Select 50-100 negative control outcomes per study
- OHDSI ***Common Evidence Model*** (CEM, in ATLAS) can help, using information from
  - Product labels
  - Scientific literature
  - Spontaneous reporting

Relatively easy when not requiring identical confounding


Contents lists available at ScienceDirect

Journal of Biomedical Informatics

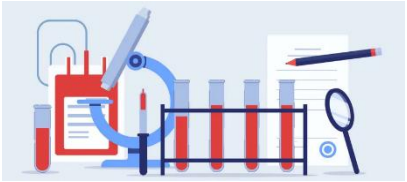
journal homepage: www.elsevier.com/locate/yjbin

Accuracy of an automated knowledge base for identifying drug adverse reactions

E.A. Voss, R.D. Boyce, P.B. Ryan, J. van der Lei, P.R. Rijnbeek, M.J. Schuemie

Some uses of (many) negative control experiments

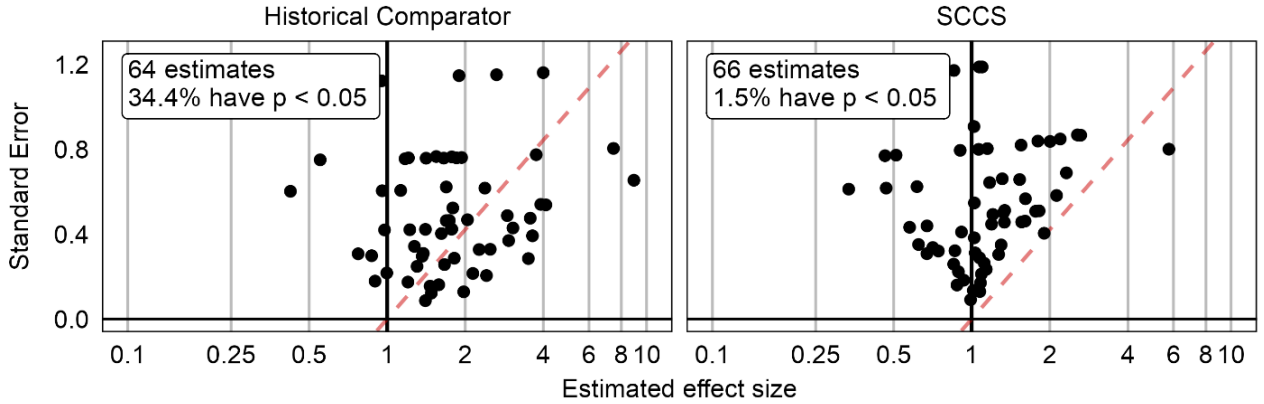
- Guide ***study design*** decisions
  - for both methods development and in practice
- Provide ***diagnostics*** to improve study reliability
  - no observed error is reassuring
  - large bias questions design, potentially stops analyses and keeps results blinded
- Help control for residual systematic error
  - ***empirical calibration*** (frequentist)
  - adaptive bias control (Bayesian)



Guiding study design decisions

Evaluating vaccine safety surveillance designs (in large EHR):

- Historical (rate) comparator
- Self-controlled case series



Should one trust a method that shows consistent bias on a large set of negative controls?

(even if confounding might not be the same)

Received: 8 July 2022 | Revised: 30 September 2022 | Accepted: 8 December 2022
DOI: 10.1002/sim.9631

RESEARCH ARTICLE

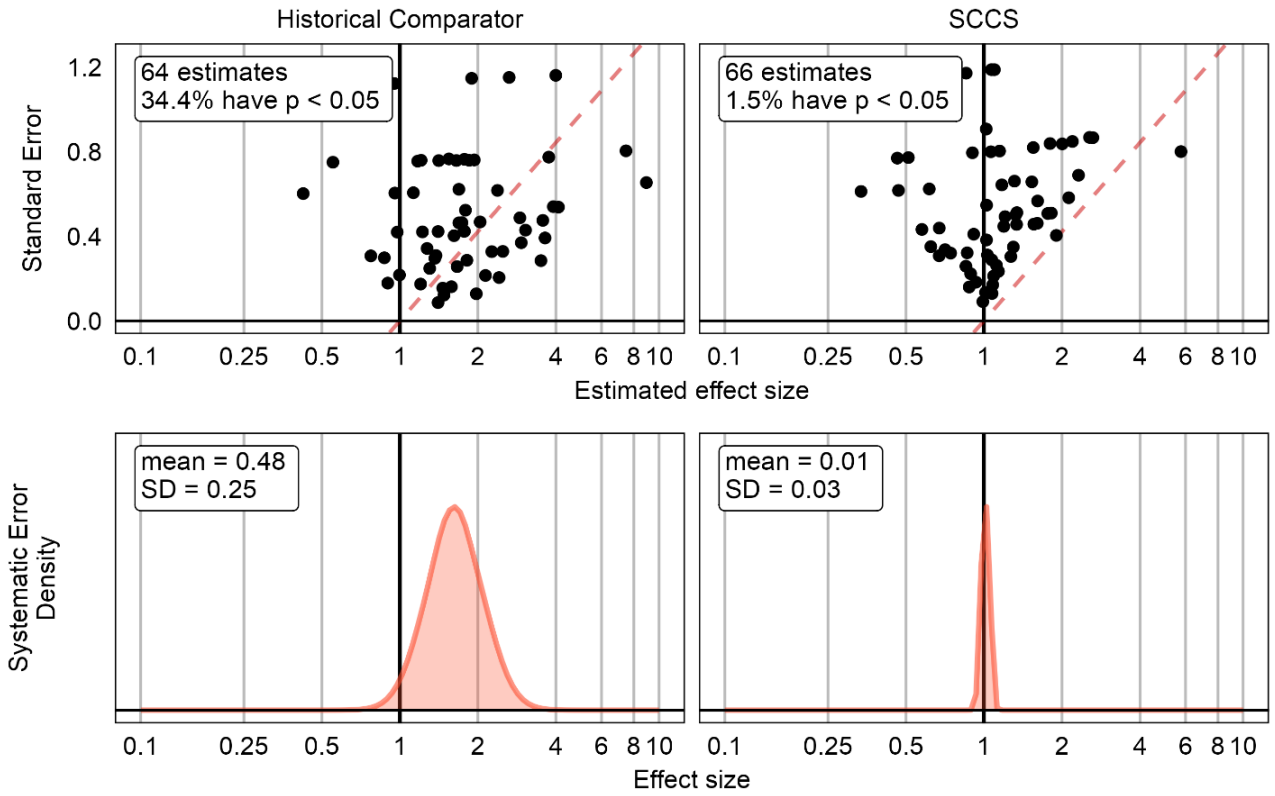
Statistics in Medicine

Adjusting for both sequential testing and systematic error in safety surveillance using observational data: Empirical calibration and MaxSPRT

Martijn J. Schuemie[1,2] | Fan Bu[2,3] | Akihiko Nishimura[4] | Marc A. Suchard[2,3,5]

Quantifying systematic error as a diagnostic

- One negative control estimate: unknown uncertainty
- Many negative controls: estimate of ***systematic error distribution***: spread of estimates not explained by random error alone

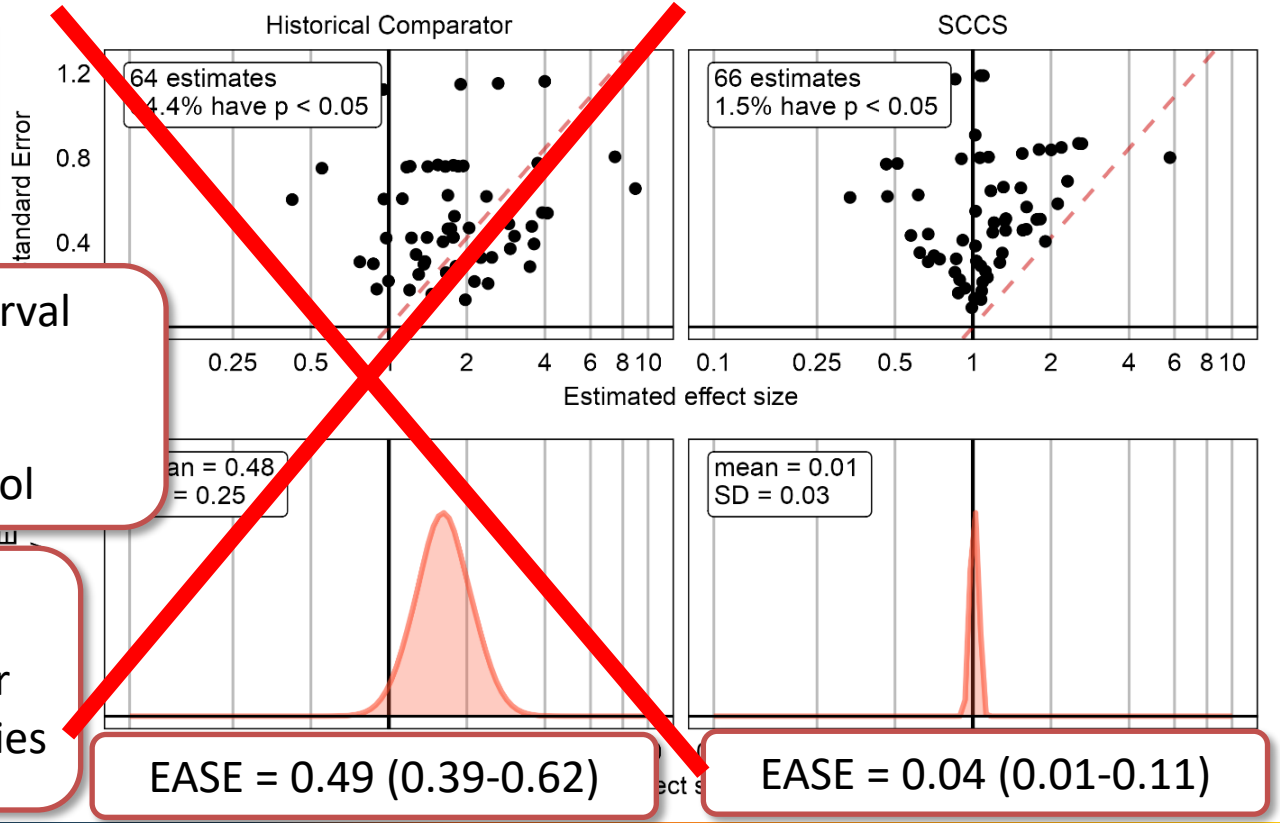


Quantifying systematic error as a diagnostic

Expected absolute systematic error (**EASE**) summarizes this distribution

Can compute a credible interval for EASE that is driven by
- # of negative controls
- power per negative control

Use a ***pre-specified*** EASE threshold (EASE < 0.25) for ***go/no-go*** decisions for studies



EASE = 0.49 (0.39-0.62)

EASE = 0.04 (0.01-0.11)

Empirical calibration

The systematic error distribution can be integrated in / used to help return ***nominal operating characteristics*** (like Type 1 Error rate) for

- *p*-values
- confidence intervals
- MaxSPRT critical values

TABLE 1 Type 1 error rates observed for negative control outcomes of the H1N1pdm vaccine with and without empirical calibration and sequential testing adjustment via MaxSPRT

	Type 1 error rate	
	Historical comparator	**SCCS**
Uncalibrated, no adjustment for sequential testing	28.0%	4.3%
Uncalibrated, MaxSPRT	18.3%	2.2%
Calibrated, no adjustment for sequential testing	10.8%	5.4%
Calibrated, MaxSPRT	5.4%	4.3%

Note: Nominal type 1 error rates should approach 5%.

Empirical calibration had larger effect than MaxSPRT in restoring Type 1 Error

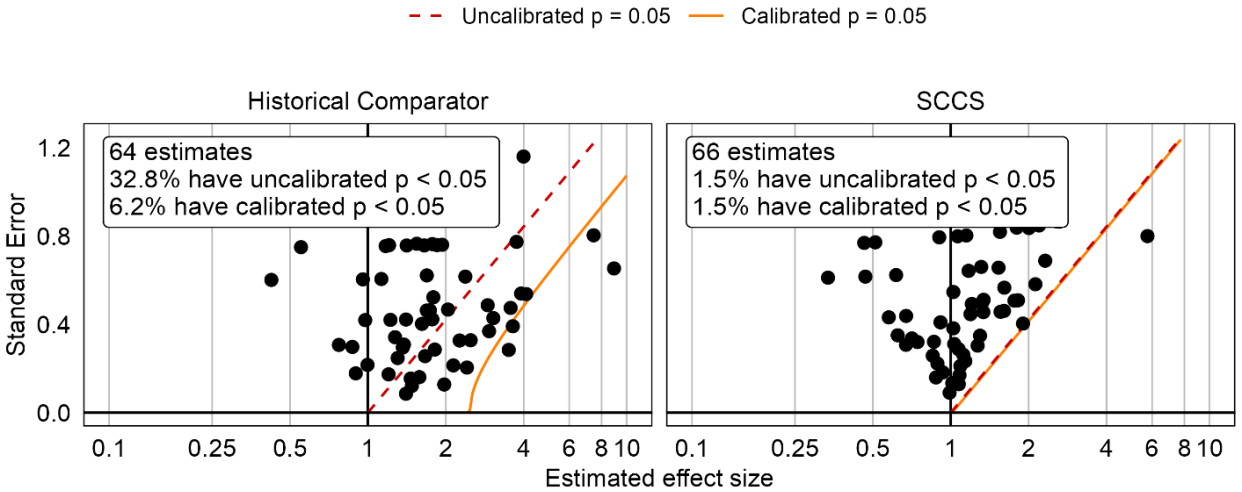
PNAS

Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data

Martijn J. Schuemie[a,b,1], George Hripcsak[a,c,d], Patrick B. Ryan[a,b,c], David Madigan[a,e], and Marc A. Suchard[a,f,g,h]

[a]Observational Health Data Sciences and Informatics, New York, NY 10032; [b]Epidemiology Analytics, Janssen Research & Development, Titusville, NJ 08560; [c]Department of Biomedical Informatics, Columbia University, New York, NY 10032; [d]Medical Informatics Services, New York–Presbyterian Hospital, New York, NY 10032; [e]Department of Statistics, Columbia University, New York, NY 10027; [f]Department of Biomathematics, University of California, Los Angeles, CA 90095; [g]Department of Biostatistics, University of California, Los Angeles, CA 90095; and [h]Department of Human Genetics, University of California, Los Angeles, CA 90095

To calibrate or not to calibrate?



Large systematic error: restoring Type 1 Error costs increased Type 2 Error

No systematic error: no Type 2 Error loss

- Pros: Favorable balance of Type 1 / 2 Error trade-off
  - claiming (& believing) $\alpha = 0.05$ but in fact $\alpha \gg 0.05$ is **not** best practice
- Cons: confidence interval calibration based on synthetic controls

Strong advice: use negative controls as **pre-specified go/no-go** diagnostics for any study

Questions?