Large-scale **E**vidence **G**eneration and **E**valuation across a **N**etwork of **D**atabases

Principles

Martijn Schuemie

What is LEGEND?

- A group of OHDSI collaborators
- Goal: to generate evidence at large scale
- Have defined 10 guiding principles
- Have already published several articles following those principles


Research

JAMA Internal Medicine | Original Investigation

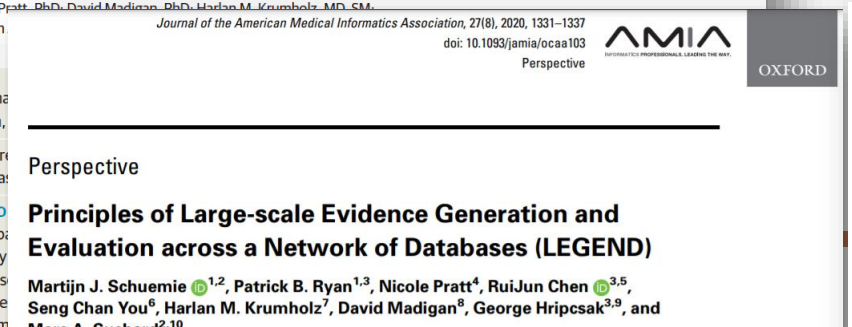
Comparison of Cardiovascular and Safety Outcomes of Chlorthalidone vs Hydrochlorothiazide to Treat Hypertension

George Hripcsak, MD, MS; Marc A. Suchard, MD, PhD; Steven Shea, MD; RuiJun Chen, MD; Seng Chan You, MD; Nicole Pratt, PhD; David Madigan, PhD; Harlan M. Krumholz, MD, SM; Patrick B. Ryan, PhD; Martijn

IMPORTANCE Chlortha to treat hypertension,

OBJECTIVE To compare hydrochlorothiazide as

DESIGN, SETTING, AND in a Network of Databa large-scale propensity calibration on databas inpatient care episode States based on 2 adm


Journal of the American Medical Informatics Association, 27(8), 2020, 1331–1337
doi: 10.1093/jamia/ocaa103
Perspective

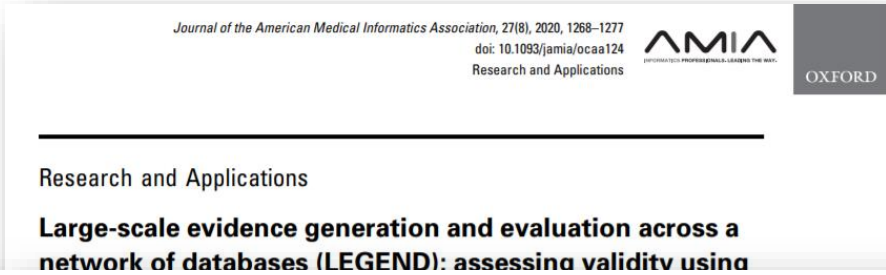
AMIA

OXFORD

Perspective

Principles of Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND)

Martijn J. Schuemie [1,2], Patrick B. Ryan[1,3], Nicole Pratt[4], RuiJun Chen [3,5], Seng Chan You[6], Harlan M. Krumholz[7], David Madigan[8], George Hripcsak[3,9], and

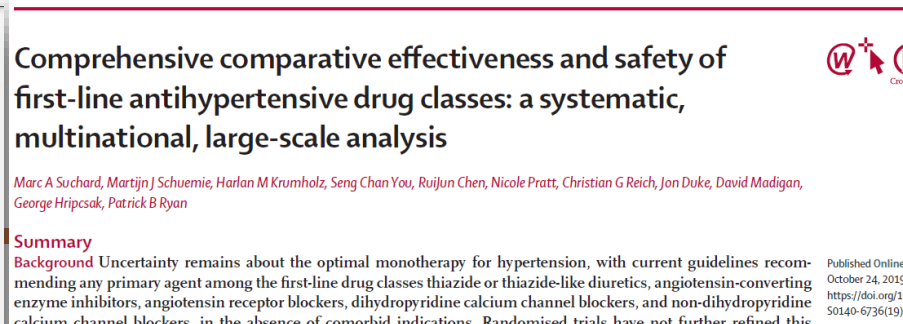

Journal of the American Medical Informatics Association, 27(8), 2020, 1268–1277
doi: 10.1093/jamia/ocaa124
Research and Applications

AMIA

OXFORD

Research and Applications

Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using


Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, RuiJun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripcsak, Patrick B Ryan

Summary

Background Uncertainty remains about the optimal monotherapy for hypertension, with current guidelines recommending any primary agent among the first-line drug classes thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydropyridine calcium channel blockers, in the absence of comorbid indications. Randomised trials have not further refined this

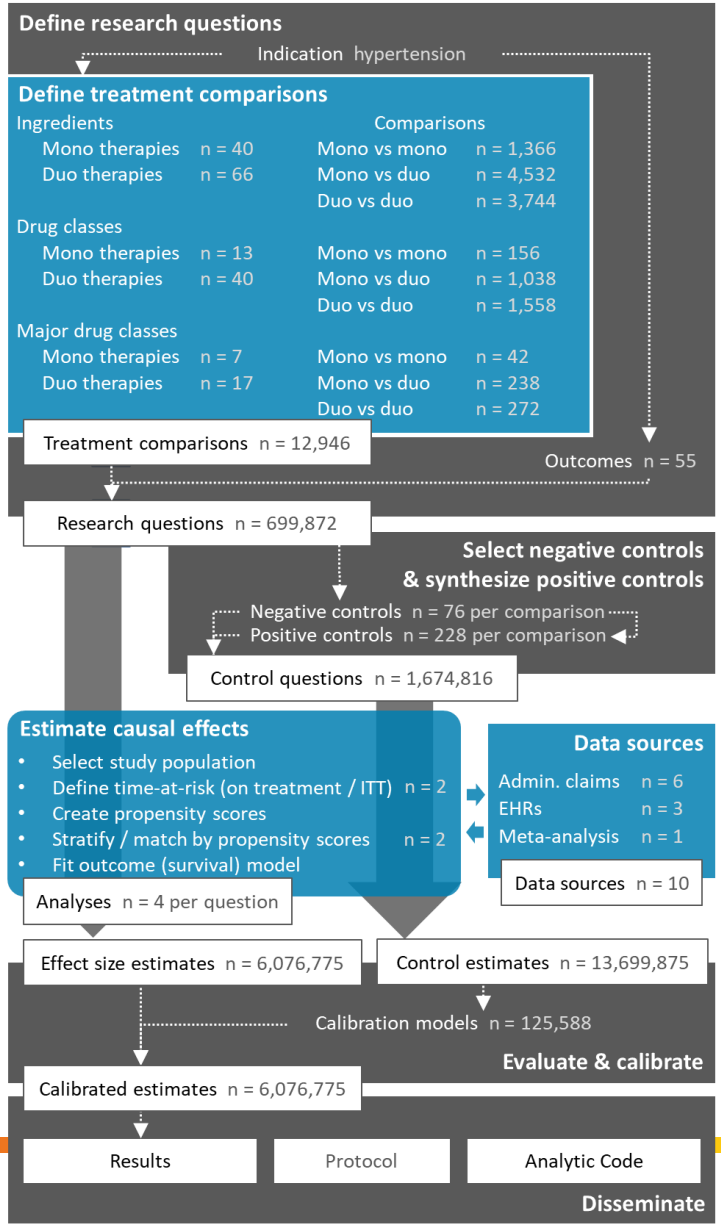
Published Online October 24, 2019 https://doi.org/10 S0140-6736(19)

LEGEND Guiding Principles

1. Evidence will be generated at **large-scale**.

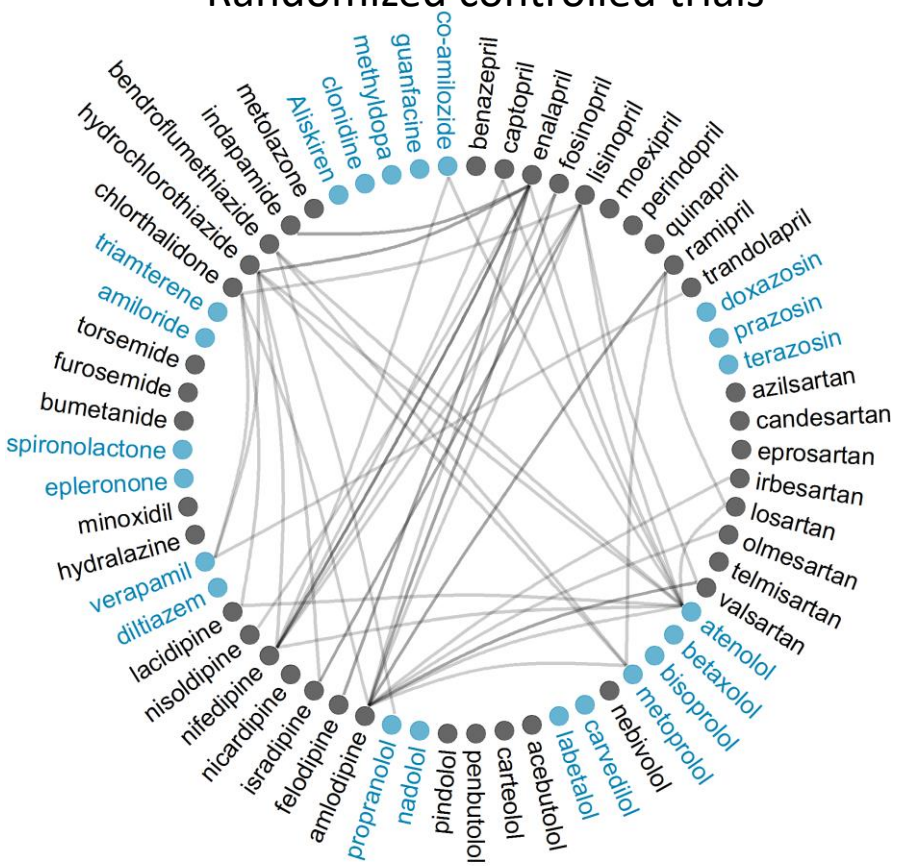
LEGEND hypertension study

- Compare all hypertension treatments
- For 55 outcomes
  - Safety
  - Effectiveness
- A total of 700k research questions

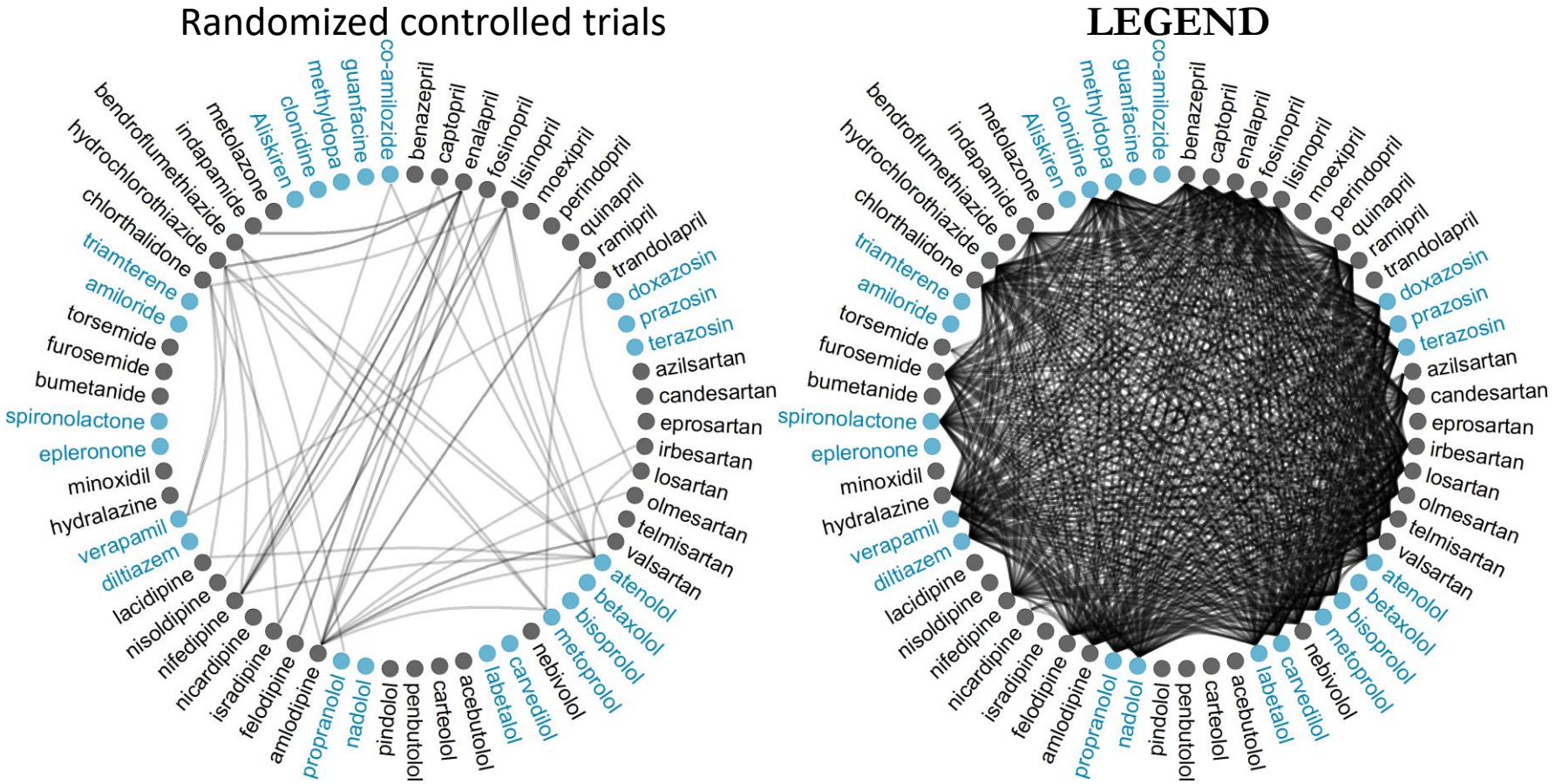


LEGEND hypertension study

Randomized controlled trials



LEGEND hypertension study



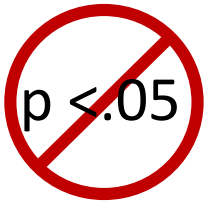
LEGEND Guiding Principles

1. Evidence will be generated at **large-scale**.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. The evidence will be generated using a **pre-specified** analysis design.
4. Evidence will be generated by consistently applying a **systematic approach** across all research questions.

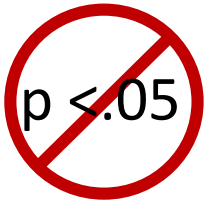
Publication bias

Idea | Perform study | Submit paper | Publication!







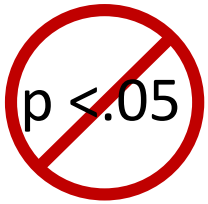


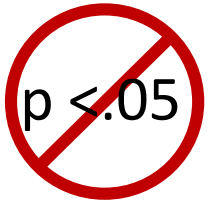

P-hacking

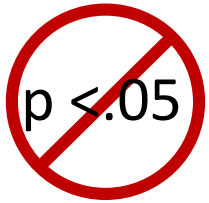
Idea

Perform study

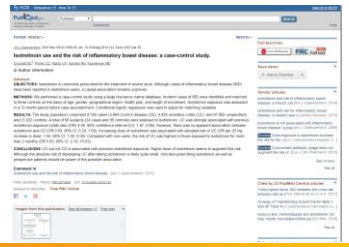
Submit paper





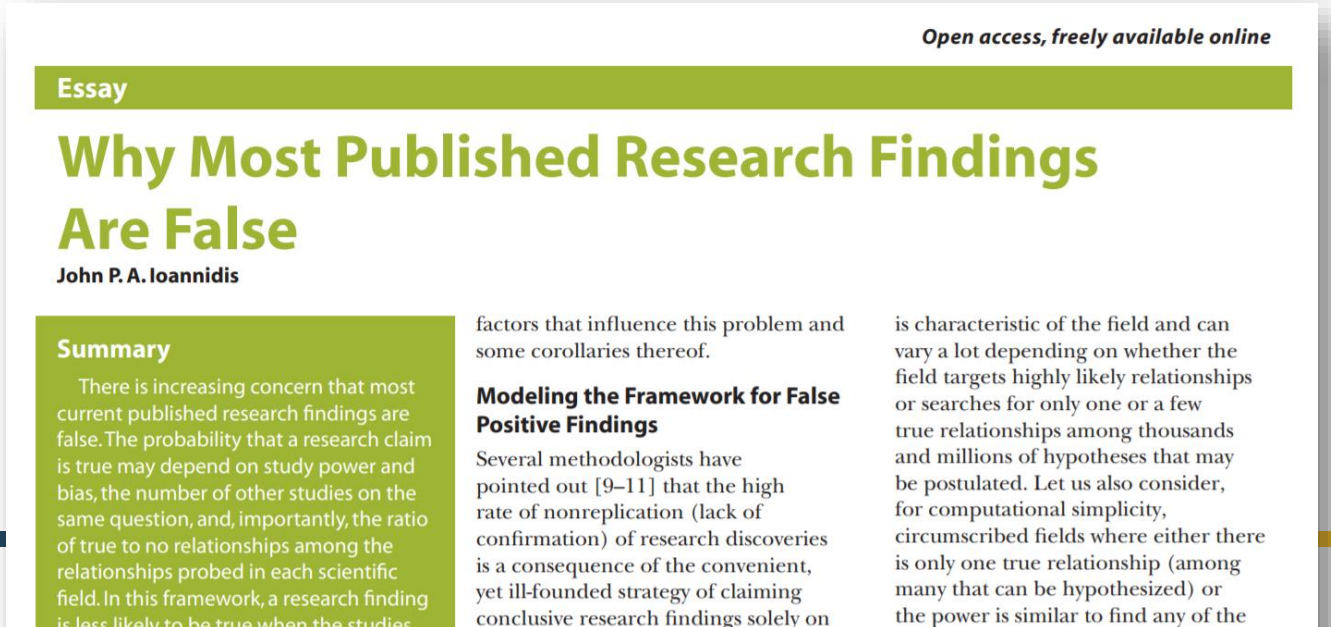


Publication!



Publication bias & p-hacking

- Publication bias and p-hacking result in
  - High false positive rate (most published results are wrong)
  - Lack of evidence on null and small effects



Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding

factors that influence this problem and some corollaries thereof.

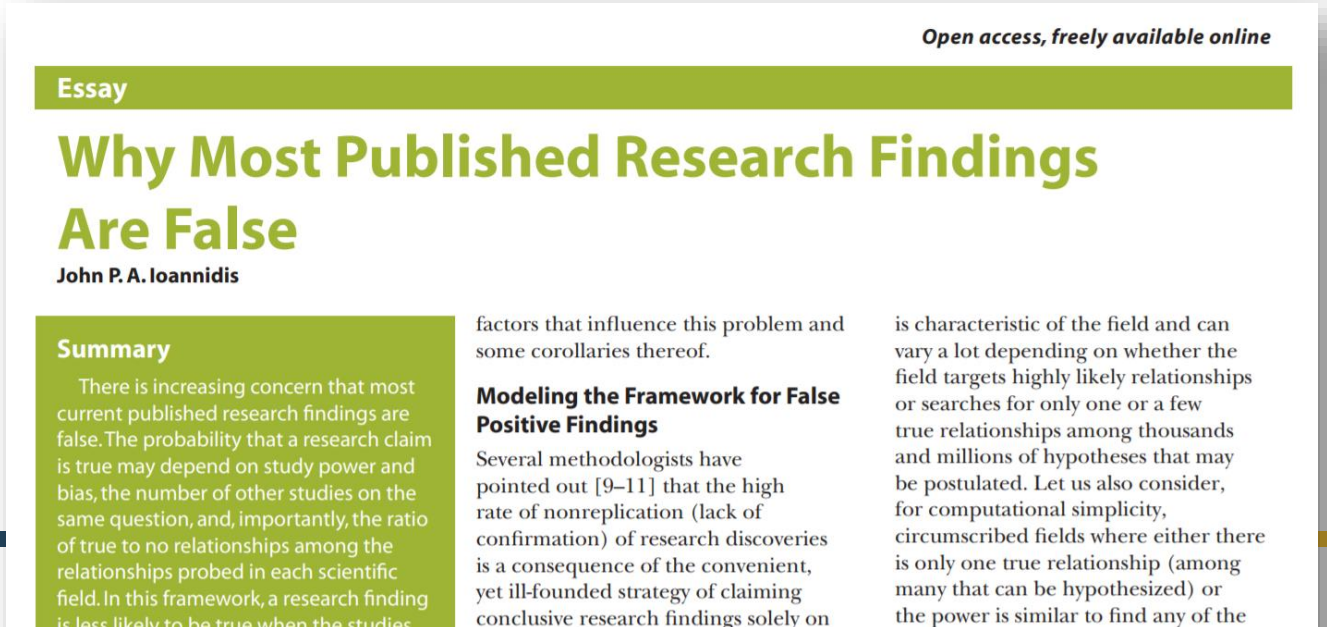
Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the

Publication bias & p-hacking

- Publication bias and p-hacking result in
  - High false positive rate (most published results are wrong)
  - Lack of evidence on null and small effects
- All LEGEND analysis are prespecified, and results are disseminated without filter



Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the

LEGEND Guiding Principles

1. Evidence will be generated at **large-scale**.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. The evidence will be generated using a **pre-specified** analysis design.
4. Evidence will be generated by consistently applying a **systematic approach** across all research questions.
5. The evidence will be generated using **best-practices**.

Advanced confounding adjustment

- Construct large generic set of covariates
  - 10,000 < n < 100,000
- Use regularized regression to fit propensity model
- Match or stratify on propensity score

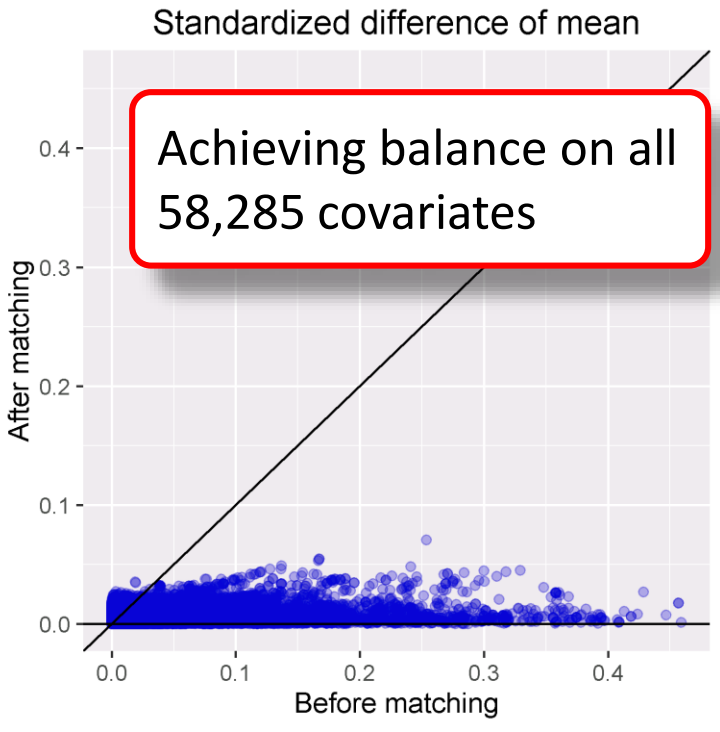
International Journal of Epidemiology, 2018, 1–10
doi: 10.1093/ije/dyy120
Original article

Original article

Evaluating large-scale propensity score performance through real-world and synthetic data experiments

Yuxi Tian,[1]* Martijn J Schuemie[2] and Marc A Suchard[1,3,4]

[1]Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA, [2]Epidemiology Department, Janssen Research and Development LLC, Titusville, NJ, USA, [3]Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA, USA and [4]Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA



LEGEND Guiding Principles

1. Evidence will be generated at **large-scale**.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. The evidence will be generated using a **pre-specified** analysis design.
4. Evidence will be generated by consistently applying a **systematic approach** across all research questions.
5. The evidence will be generated using **best-practices**.
6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.

Measuring residual bias

Control questions:

- exposure-outcome pairs with known effect size
- negative and positive controls

Empirical calibration:

- Adjust p-value and confidence interval using estimates for controls

PNAS

COLLOQUIUM PAPER

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

Edited by Victoria Stodden, University of Illinois at Urbana–Champaign, Champaign, IL, and accepted by Editorial Board Member Susan T. Fiske October 26, 2017 (received for review June 15, 2017)

Observational healthcare data, such as electronic health records and administrative claims, offer potential to estimate effects of medical products at scale. Observational studies have often been found to be nonreproducible, however, generating conflicting results even when using the same database to answer the

age treatment effect. Systematic error can manifest from multiple sources, including confounding, selection bias, and measurement error. While there is widespread awareness of the potential for systematic error in observational studies and a large body of research that examines how to diagnose and statistically adjust

LEGEND Guiding Principles

1. Evidence will be generated at **large-scale**.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. The evidence will be generated using a **pre-specified** analysis design.
4. Evidence will be generated by consistently applying a **systematic approach** across all research questions.
5. The evidence will be generated using **best-practices**.
6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.
7. The evidence will be generated using **open-source** software that is freely available to all.

Open-source software

- The LEGEND study package is available at https://github.com/OHDSI/Legend

- LEGEND relies on

https://ohdsi.github.io/Hades/

LEGEND Guiding Principles

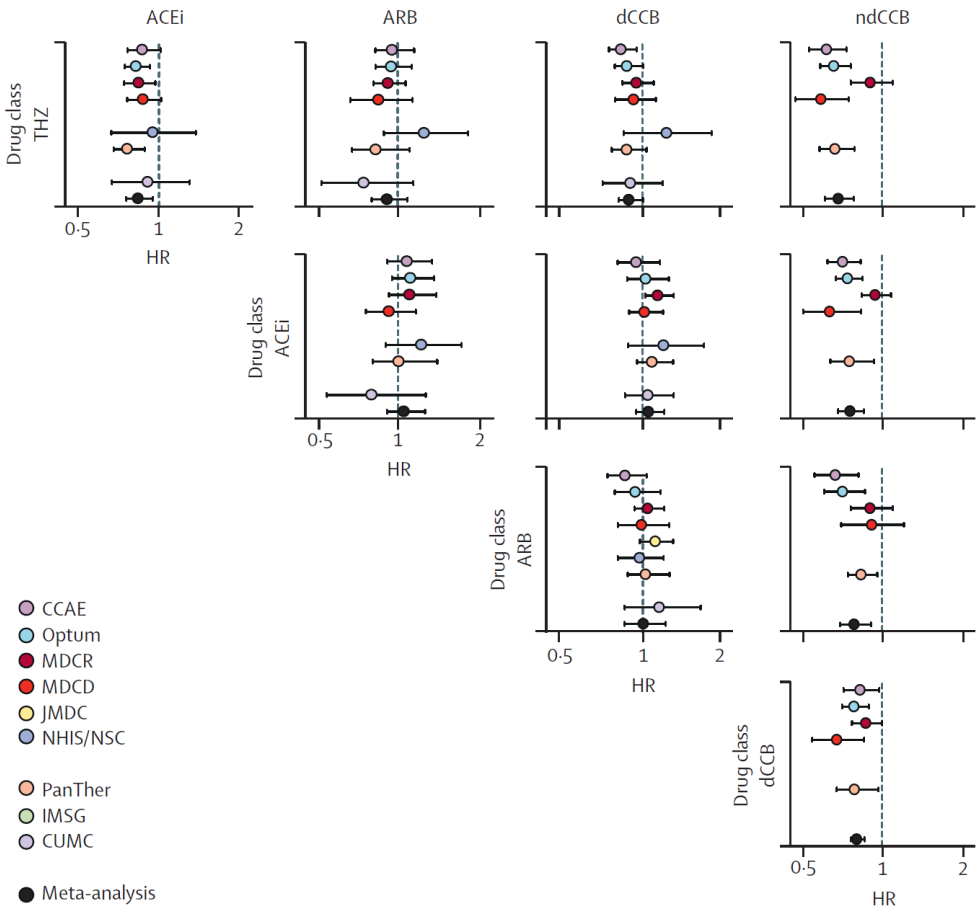
1. Evidence will be generated at **large-scale**.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. The evidence will be generated using a **pre-specified** analysis design.
4. Evidence will be generated by consistently applying a **systematic approach** across all research questions.
5. The evidence will be generated using **best-practices**.
6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.
7. The evidence will be generated using **open-source** software that is freely available to all.
8. LEGEND will **not** be used to **evaluate methods**.
9. LEGEND will generate evidence across a network of multiple databases
10. **No patient-level data** will be shared between sites in the network, only aggregated data.

LEGEND Guiding Principles

1. Evidence will be generated at **large-scale**.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. The evidence will be generated using a **pre-specified** analysis design.
4. Evidence will be generated by consistently applying a **systematic approach** across all research questions.
5. The evidence will be generated using **best-practices**.
6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.
7. The evidence will be generated using **open-source** software that is freely available to all.
8. LEGEND will **not** be used to **evaluate methods**.
9. LEGEND will generate evidence across a network of multiple databases
10. **No patient-level data** will be shared between sites in the network, only aggregated data.

Evidence from multiple databases

- Each study should be replicated across multiple databases
- More data: more statistical power
- Heterogeneity may cause doubt on the validity of the results

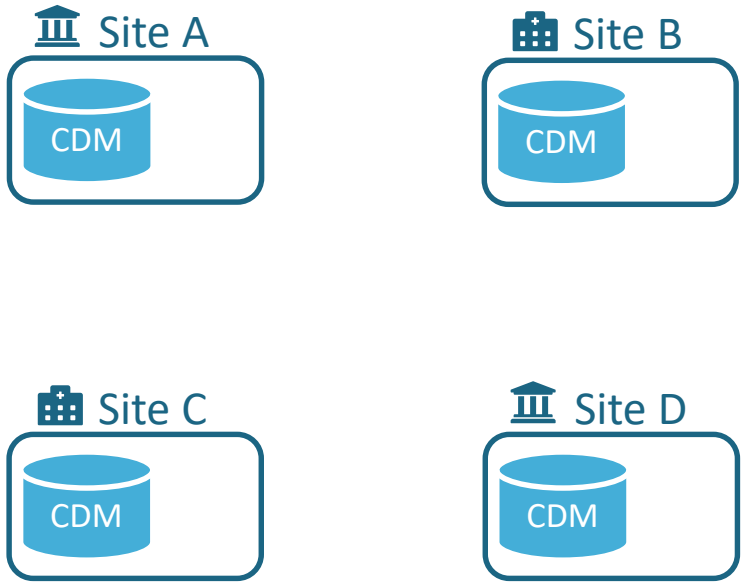


LEGEND Guiding Principles

1. Evidence will be generated at **large-scale**.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. The evidence will be generated using a **pre-specified** analysis design.
4. Evidence will be generated by consistently applying a **systematic approach** across all research questions.
5. The evidence will be generated using **best-practices**.
6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.
7. The evidence will be generated using **open-source** software that is freely available to all.
8. LEGEND will **not** be used to **evaluate methods**.
9. LEGEND will generate evidence across a network of multiple databases
10. **No patient-level data** will be shared between sites in the network, only aggregated data.

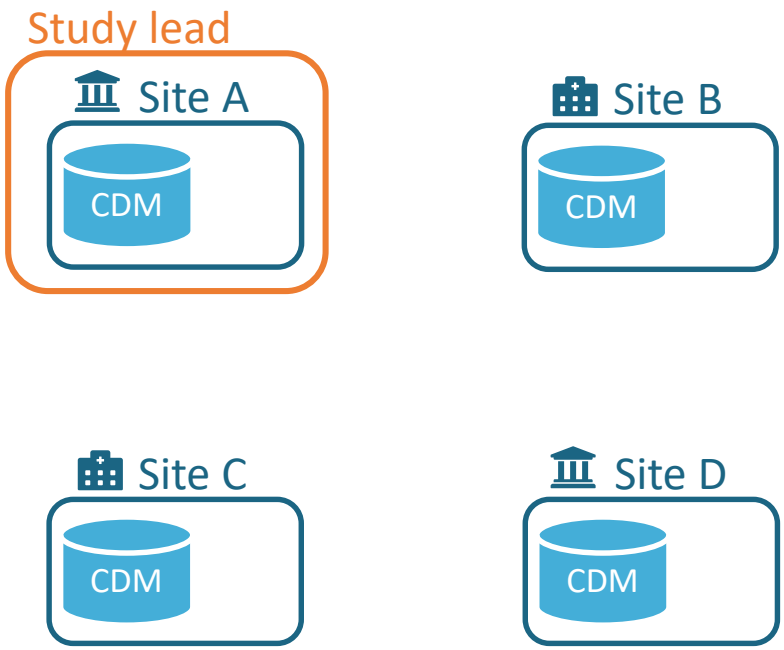
Distributed Research Network

- Multiple sites with data
  - Hospital EHRs
  - Administrative Claims
- Patient-level data cannot be shared



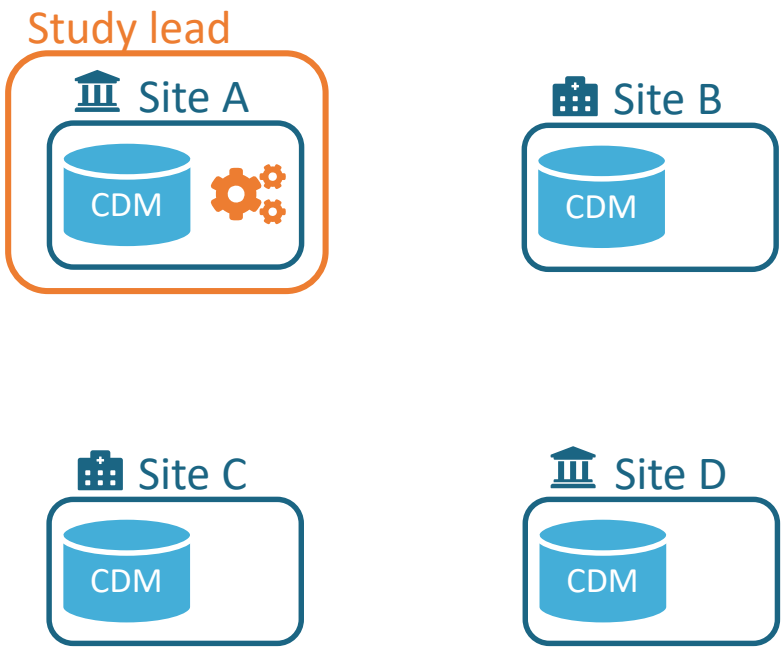
Distributed Research Network

- Any site can lead a study



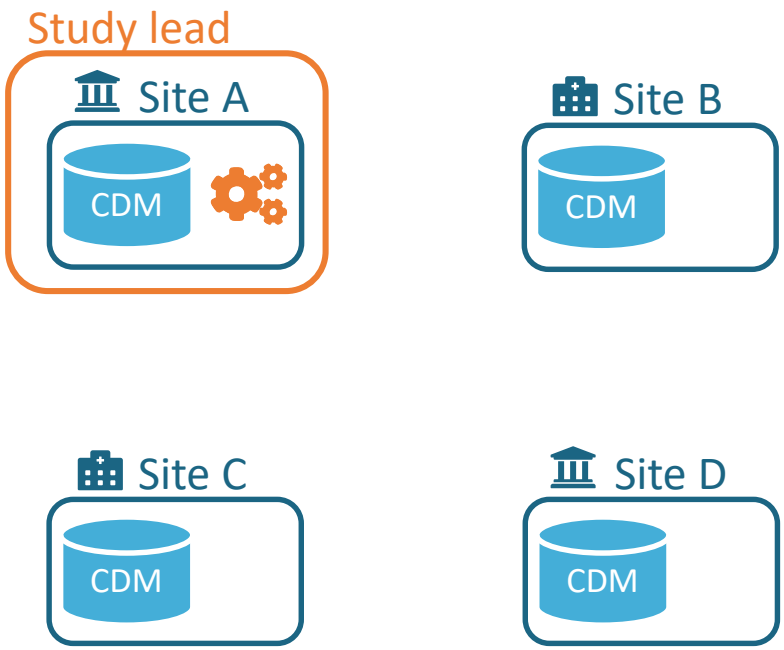
Distributed Research Network

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



Distributed Research Network

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



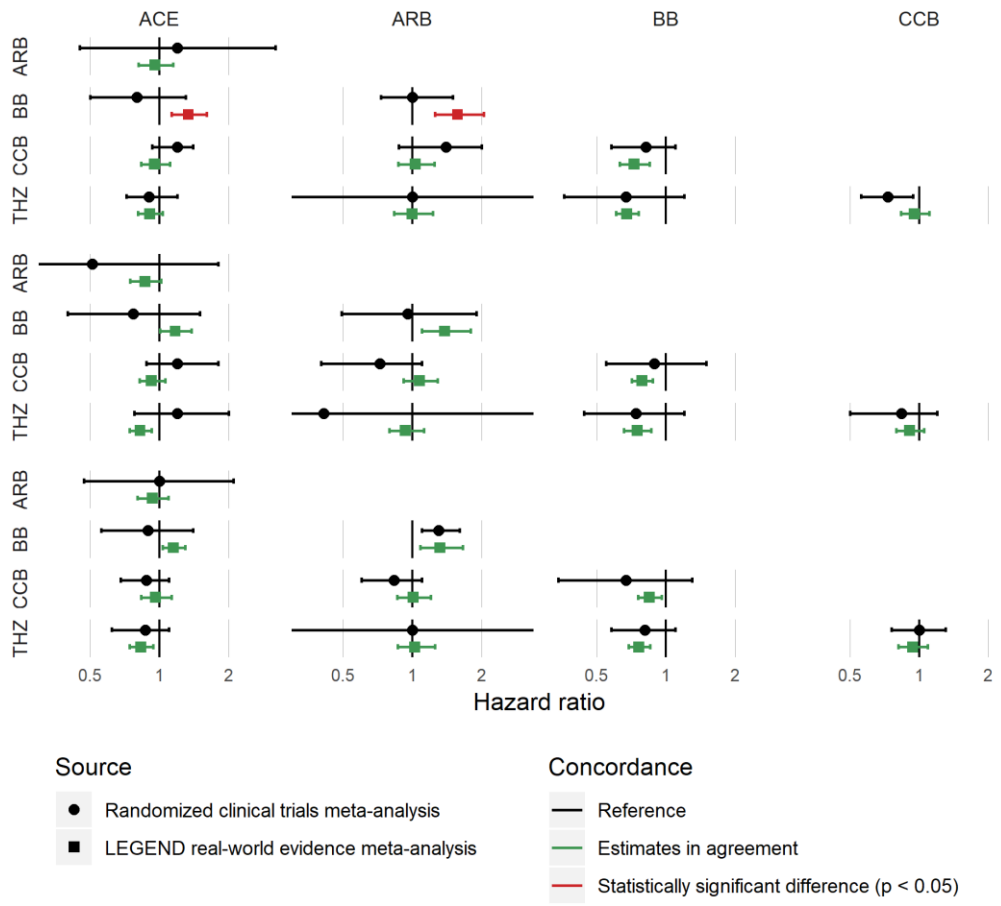
Distributed Research Network

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

Distributed Research Network

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

LEGEND vs RCTs



- Estimates were not statistically significantly different (more often than expected by chance)
- LEGEND estimates have much narrower confidence intervals
- Note: you could do almost as well by just always guessing ‘no effect’

In conclusion

- LEGEND principles aim to
  - Improve transparency
  - Ensure verification
- We hope more studies will follow these principles


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Thank you!

LEGEND Guiding Principles

1. Evidence will be generated at **large-scale**.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. The evidence will be generated using a **pre-specified** analysis design.
4. Evidence will be generated by consistently applying a **systematic approach** across all research questions.
5. The evidence will be generated using **best-practices**.
6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.
7. The evidence will be generated using **open-source** software that is freely available to all.
8. LEGEND will **not** be used to **evaluate methods**.
9. LEGEND will generate evidence across a network of multiple databases
10. **No patient-level data** will be shared between sites in the network, only aggregated data.