Large-scale Evidence Generation and Evaluation across a Network of Databases

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
1. Evidence will be generated at large-scale.
• Compare all hypertension treatments
• For 55 outcomes
  – Safety
  – Effectiveness
• A total of 700k research questions
LEGEND hypertension study

Randomized controlled trials

-beindroflumethiazide
-clonidine
-delapril
-foxapril
-hydralazine
-losartan
-metoprolol
-metolazone
-mexalazine
-nesidipine

-amiodine
-torsemide
-furosemide
-burnetanide
-spirobolactone
-epidrone

-minoxidil
-hydralazine
-verapamil
-diltiazem
-lacidipine
-nisoldipine
-nicardipine
-isradipine
-felodipine
-amiodipine
-propidipine
-pimidipine

-betaxolol
-bisoprolol
-metoprolol
-cavendolol
-labelolol

-doxazosin
-prazosin
-terazosin
-azilsartan
-candesartan
-eprisartan
-irbesartan
-losartan
-olmesartan
-telmisartan
-valsartan

-atenolol
-propranolol
-pindolol
-nadolol
-propranolol
-atenolol
-betaxolol
-bisoprolol
-metoprolol
-cavendolol
-labelolol

-amilodipine
-clonidine
-delapril
-foxapril
-hydralazine
-losartan
-metoprolol
-metolazone
-mexalazine
-nesidipine

-beindroflumethiazide
-clonidine
-delapril
-foxapril
-hydralazine
-losartan
-metoprolol
-metolazone
-mexalazine
-nesidipine
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.
P-hacking

Idea → Perform study → Submit paper → Publication!

$p < .05$

$p < .05$

$p < .05$
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
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

---

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 is less likely to be true when the study 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 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

---

Open access, freely available online
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

Achieving balance on all 58,285 covariates
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

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

Martijn J. Schuemie\(^{a,b,1}\), George Hripcsak\(^{a,d,1}\), Patrick B. Ryan\(^{b,d}\), David Madigan\(^{a,2}\), and Marc A. Suchard\(^{a,4,a,}\)

\(^{a}\)Observational Health Data Sciences and Informatics, New York, NY 10022; \(^{b}\)Epidemiology Analytics, Janssen Research & Development, Titusville, NJ 08560; \(^{c}\)Department of Biomedical Informatics, Columbia University, New York, NY 10032; \(^{d}\)Medical Informatics Service, New York Presbyterian Hospital, New York, NY 10032; \(^{e}\)Department of Bioinformatics, Columbia University, New York, NY 10027; \(^{f}\)Department of Biostatistics, 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 same question. One cause of discrepancy is some both confounding and measurement error. 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 for specific sources of bias, there has been comparatively little work addressing the impact of measurement error in making treatment effect estimates.
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.
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

Study lead

- Site A
- CDM

- Site B
- CDM

- Site C
- CDM

- Site D
- CDM
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
Thank you!
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.