Establishing scientific best practices for real-world analysis

David Madigan PhD
Professor of Statistics at Columbia University

Martijn Schuemie PhD
Senior Director and Research Fellow of Epidemiology Analytics at Janssen Research & Development
Visiting Scholar, Biostatistics, UCLA

How to achieve reliable evidence?

- **Transparency**: others should be able to reproduce your study in every detail using the information you provide.

- **Prespecify** what you're going to estimate and how, to avoid hidden multiple testing (fishing expeditions, p-value hacking). Run your analysis only once.

- **Validation of your analysis**: you should have evidence that your analysis does what you say it does, and that what it does is the right way to answer the question.

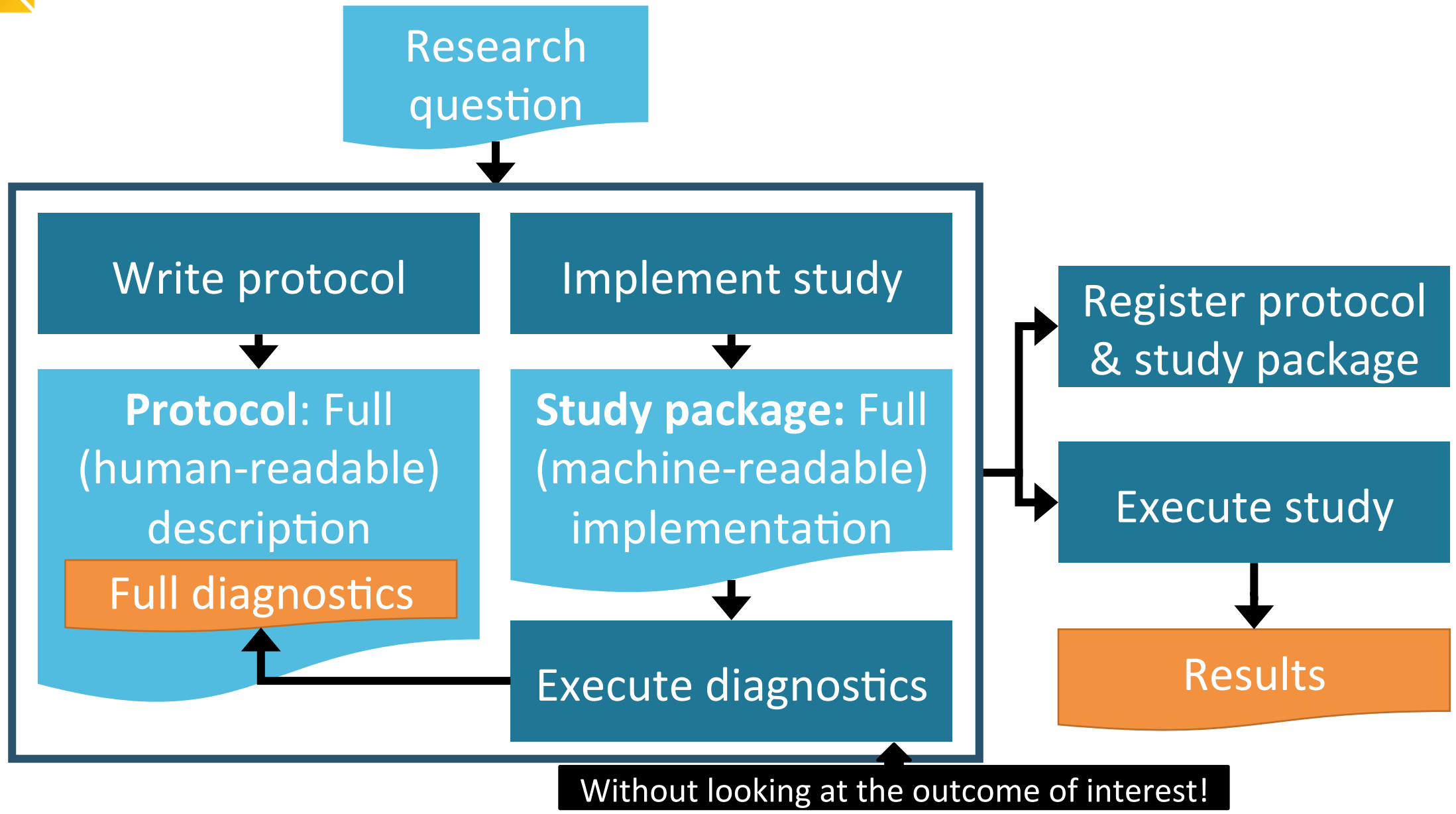
How to achieve reliable evidence?

- **Transparency**: others should be able to reproduce your study in every detail using the information you provide.

- **Prespecify** what you're going to estimate and how, to avoid hidden multiple testing (fishing expeditions, p-value hacking). Run your analysis only once.

- **Validation of your analysis**: you should have evidence that your analysis does what you say it does, and that what it does is the right way to answer the question.

Transparency & Prespecification



How to achieve reliable evidence?

- **Transparency**: others should be able to reproduce your study in every detail using the information you provide.

- **Prespecify** what you're going to estimate and how, to avoid hidden multiple testing (fishing expeditions, p-value hacking). Run your analysis only once.

- **Validation of your analysis**: you should have evidence that your analysis does what you say it does, and that what it does is the right way to answer the question.

Study validity

Evidence Quality

Data Quality: Are the data of sufficient quality for our research purposes?

Clinical Validity: To what extent does the analysis conducted match the clinical intention?

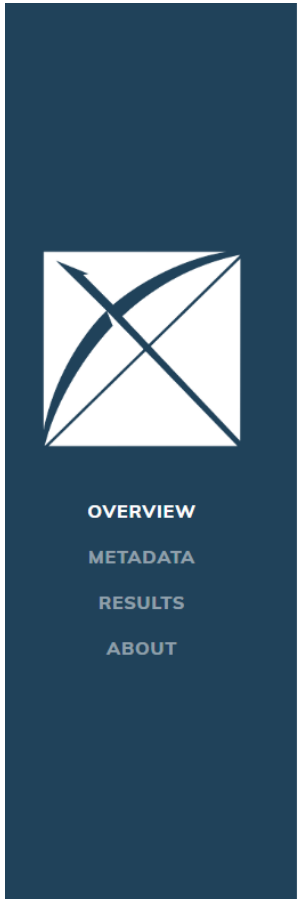
Software Validity: Does the software do what it is expected to do?

Method Validity: Is this method valid for answering this question?

Data Quality

Are the data of sufficient quality for our research purposes?

Just discussed by Clair and Andrew:



DATA QUALITY ASSESSMENT

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-21 18:08:28 in 13 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
Total	1165	72	1237	94%	392	10	402	98%	1557	82	1639	**95%**

Data Quality

Are the data of sufficient quality for our research purposes?

Just discussed by Clair and Andrew:

DATA QUALITY ASSESSMENT

DATABASE

3:28 in 13 mins

	Verification			Validation				Total			
il	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
1	180	88%	283	0	283	100%	442	21	463	95%
4	671	95%	104	0	104	100%	741	34	775	96%
7	386	96%	5	10	15	33%	374	27	401	93%
2	1237	94%	392	10	402	98%	1557	82	1639	**95%**

EDM Forum

EDM Forum Community

eGEMs (Generating Evidence & Methods to improve patient outcomes) — Publish

9-11-2016

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

Michael G. Kahn
University of Colorado Anschutz Medical Campus, michael.kahn@ucdenver.edu

Clinical Validity

To what extent does the analysis conducted match the clinical intention?

For example, will we answer

“Do ACE inhibitors cause angioedema”

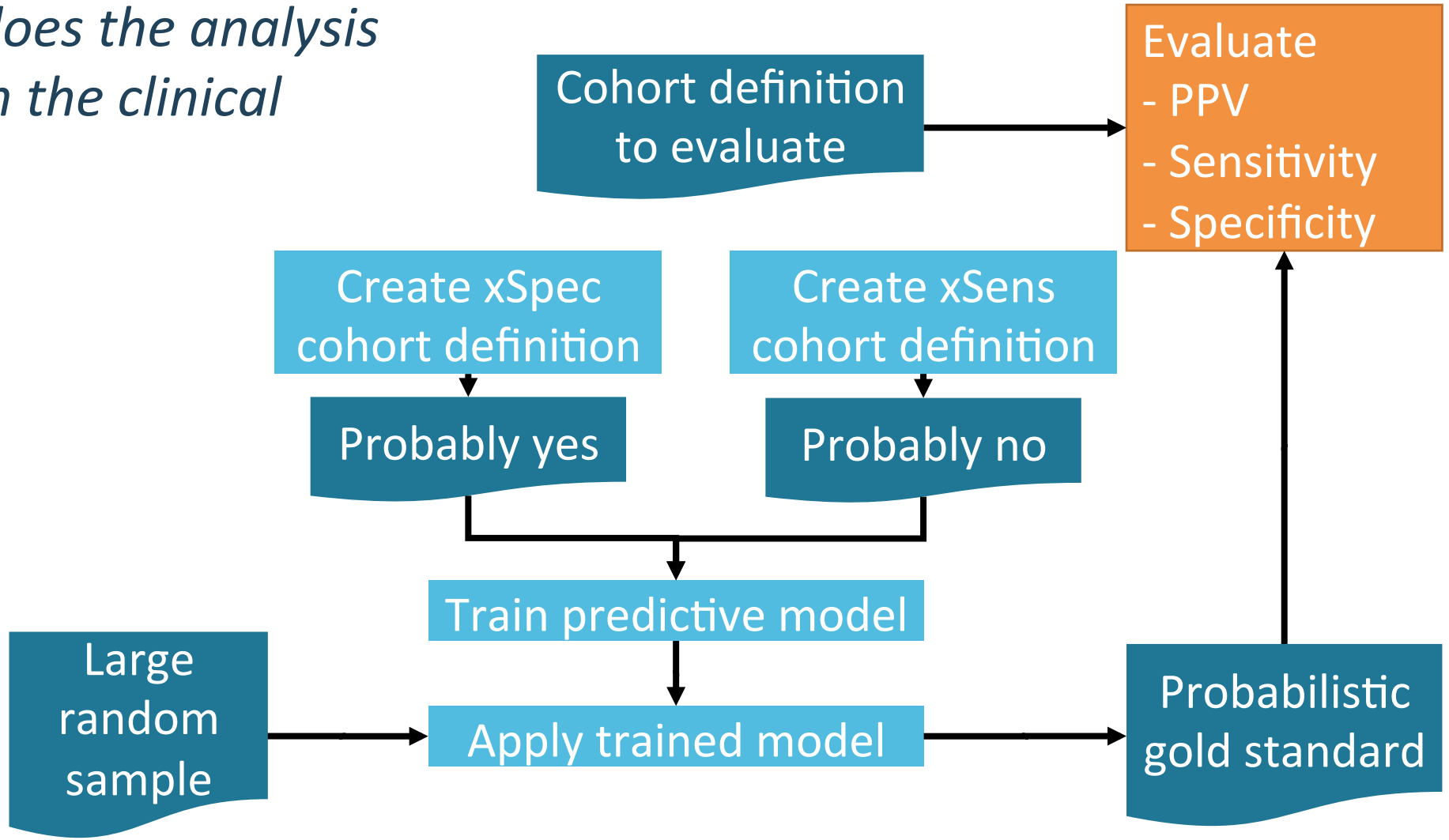
or

“Do ACE inhibitors cause **suspicion of** angioedema”

Clinical Validity

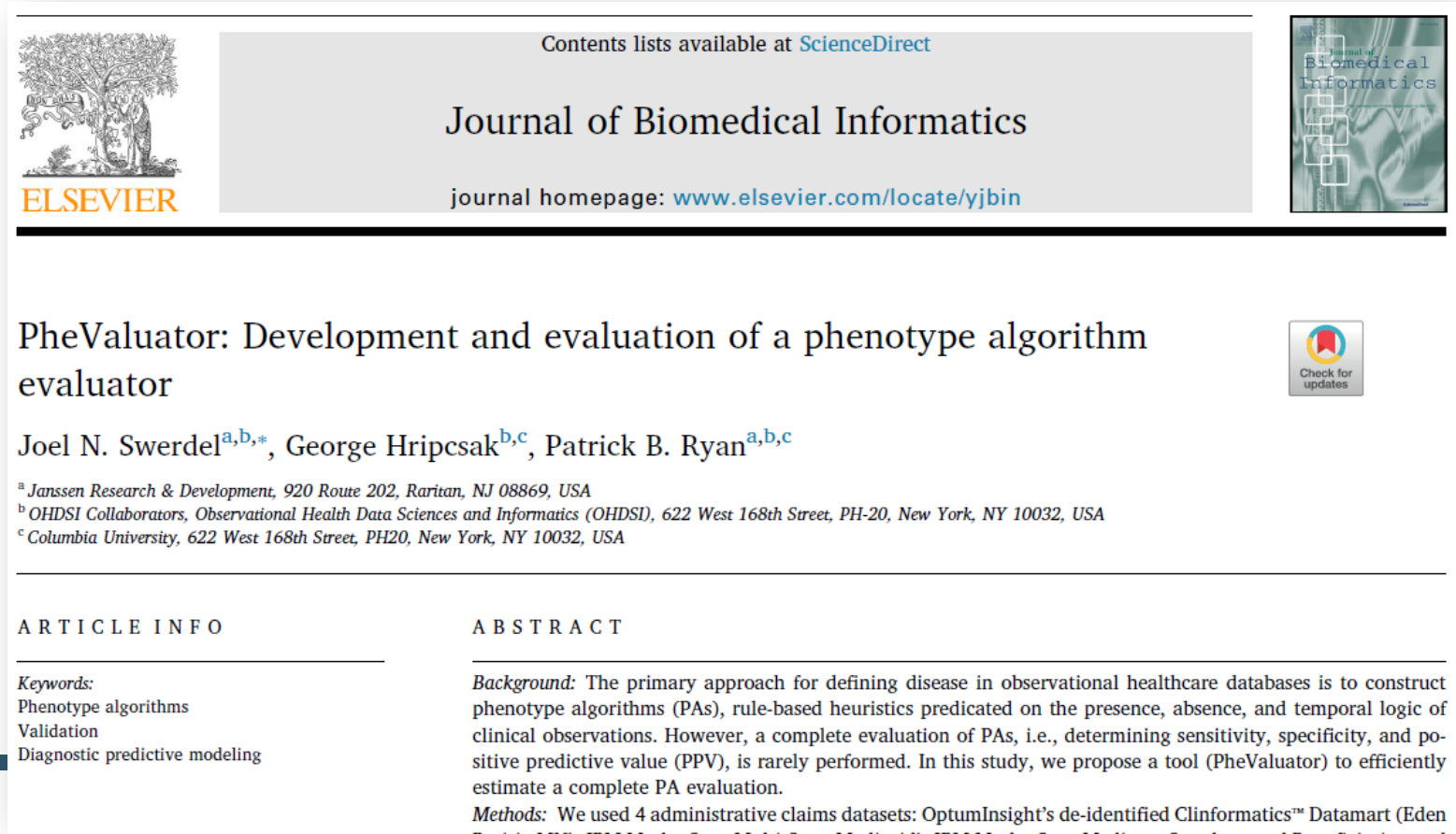
To what extent does the analysis conducted match the clinical intention?

- Chart review
- PheValuator



Clinical Validity

To what extent does the analysis conducted match the clinical intention?



Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

PheValuator: Development and evaluation of a phenotype algorithm evaluator

Joel N. Swerdel[a,b,*], George Hripcsak[b,c], Patrick B. Ryan[a,b,c]

[a] Janssen Research & Development, 920 Route 202, Raritan, NJ 08869, USA
[b] OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), 622 West 168th Street, PH-20, New York, NY 10032, USA
[c] Columbia University, 622 West 168th Street, PH20, New York, NY 10032, USA

ARTICLE INFO

Keywords:
Phenotype algorithms
Validation
Diagnostic predictive modeling

ABSTRACT

Background: The primary approach for defining disease in observational healthcare databases is to construct phenotype algorithms (PAs), rule-based heuristics predicated on the presence, absence, and temporal logic of clinical observations. However, a complete evaluation of PAs, i.e., determining sensitivity, specificity, and positive predictive value (PPV), is rarely performed. In this study, we propose a tool (PheValuator) to efficiently estimate a complete PA evaluation.

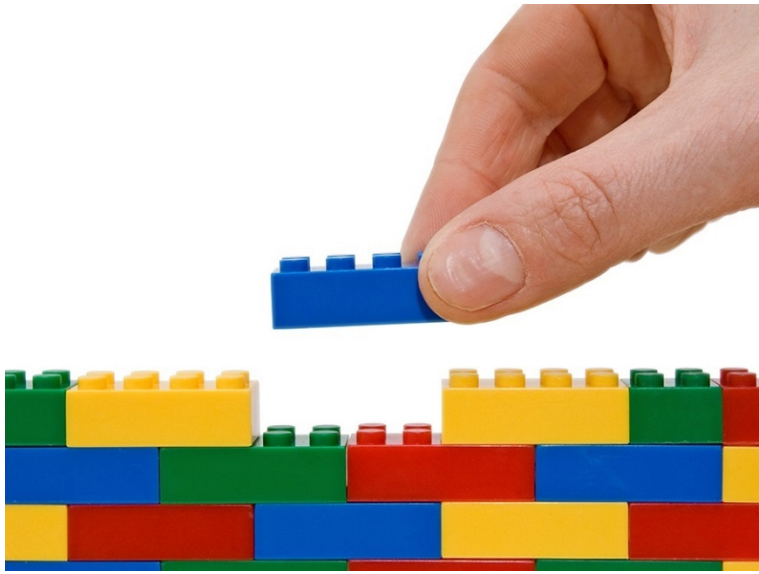
Methods: We used 4 administrative claims datasets: OptumInsight's de-identified Clinformatics™ Datamart (Eden

Software Validity

Does the software do what it is expected to do?

- Building blocks: OHDSI Standardized Tools
  - Unit tests
  - Simulations

- Configuration of blocks: Study package
  - Code review
  - Double coding



Method Validity

Is this method valid for answering this question?

Method Validity

Is this method valid for answering this question?

This requires a method that can be validated!

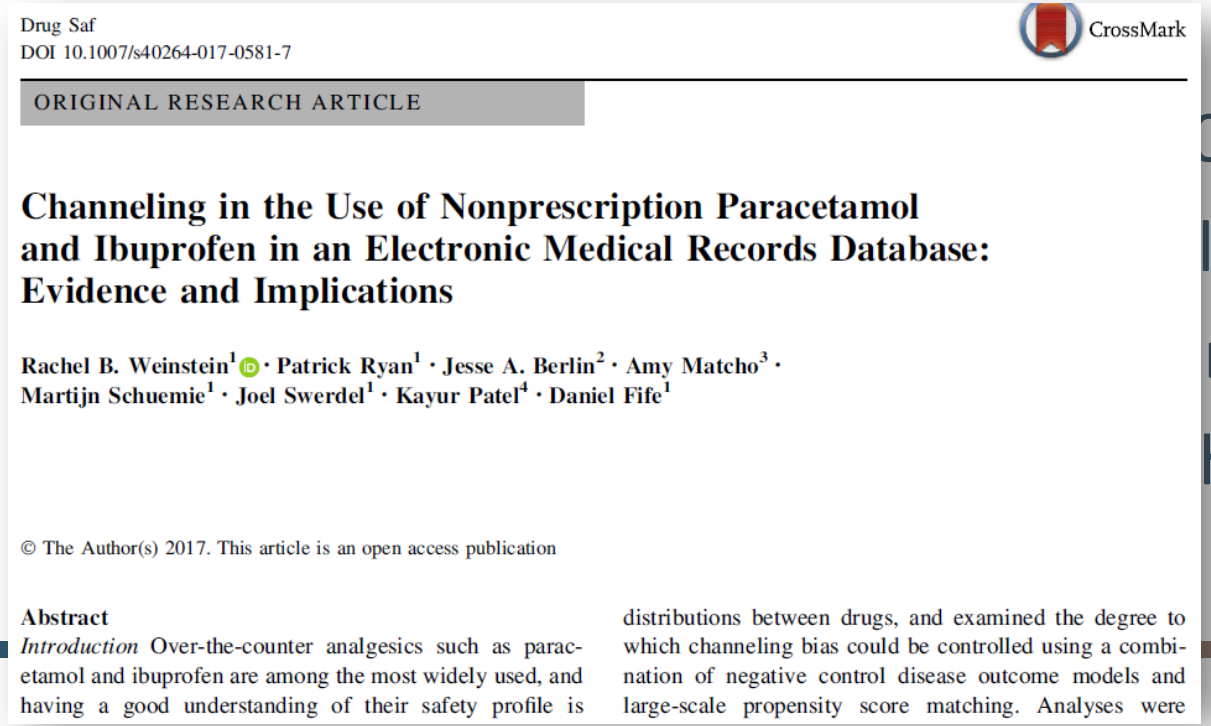
A **systematic** approach, for example our large-scale propensity models:

- Create a large set of covariates (10,000 < n 100,000)
- Use LASSO to fit propensity model
- Match / stratify on propensity score
- Check that covariate balance is achieve on all observed variables

Method Validity

Is this method valid for answering this question?

This requires a method that can be validated!



Drug Saf
DOI 10.1007/s40264-017-0581-7

CrossMark

ORIGINAL RESEARCH ARTICLE

Channeling in the Use of Nonprescription Paracetamol and Ibuprofen in an Electronic Medical Records Database: Evidence and Implications

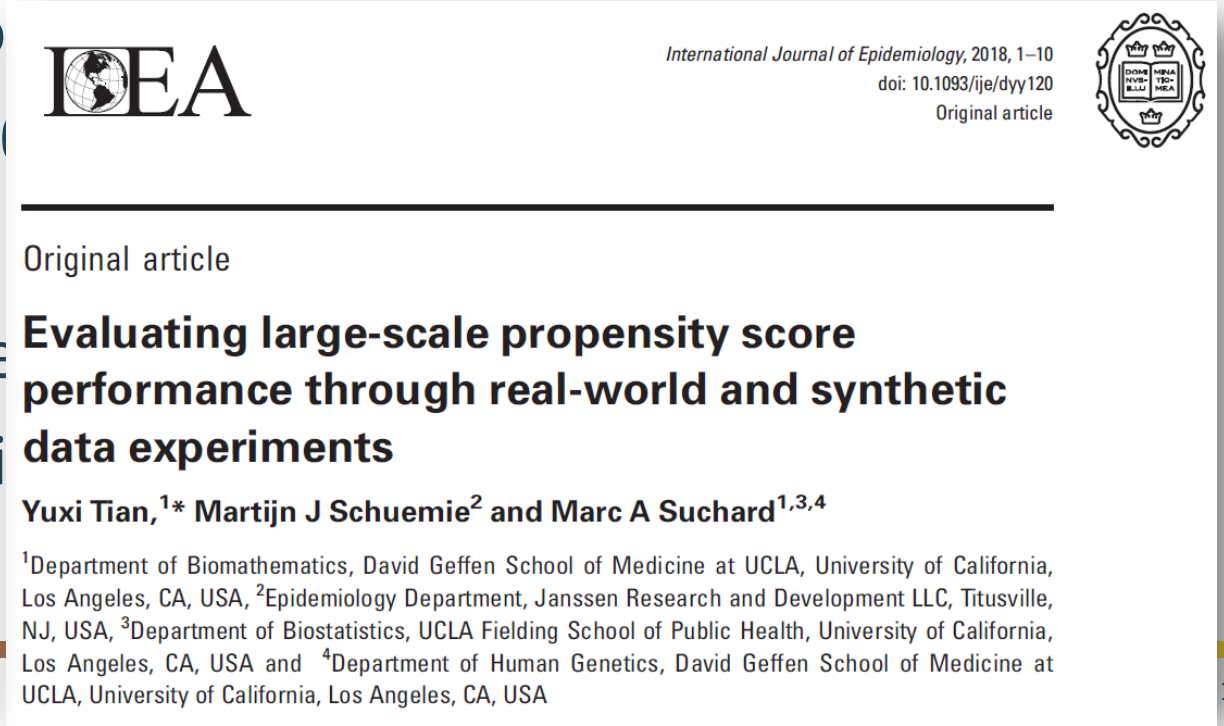
Rachel B. Weinstein[1] · Patrick Ryan[1] · Jesse A. Berlin[2] · Amy Matcho[3] · Martijn Schuemie[1] · Joel Swerdel[1] · Kayur Patel[4] · Daniel Fife[1]

© The Author(s) 2017. This article is an open access publication

Abstract

Introduction Over-the-counter analgesics such as paracetamol and ibuprofen are among the most widely used, and having a good understanding of their safety profile is

distributions between drugs, and examined the degree to which channeling bias could be controlled using a combination of negative control disease outcome models and large-scale propensity score matching. Analyses were



IEA

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

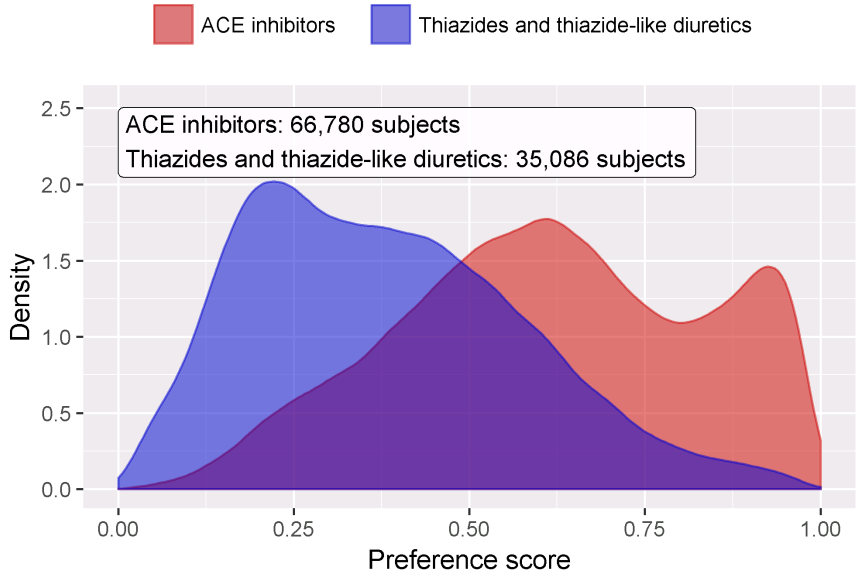
Yuxi Tian,[1]* Martijn J Schuemie[2] and Marc A Suchard[1,3,4]

[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

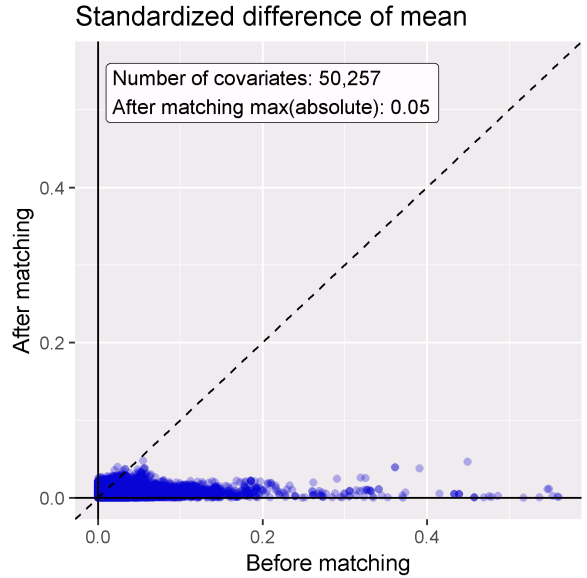
Method Validity

Is this method valid for answering this question?

- Do we pass our (design-specific) **study diagnostics**? E.g.



Are the exposure groups sufficiently comparable?

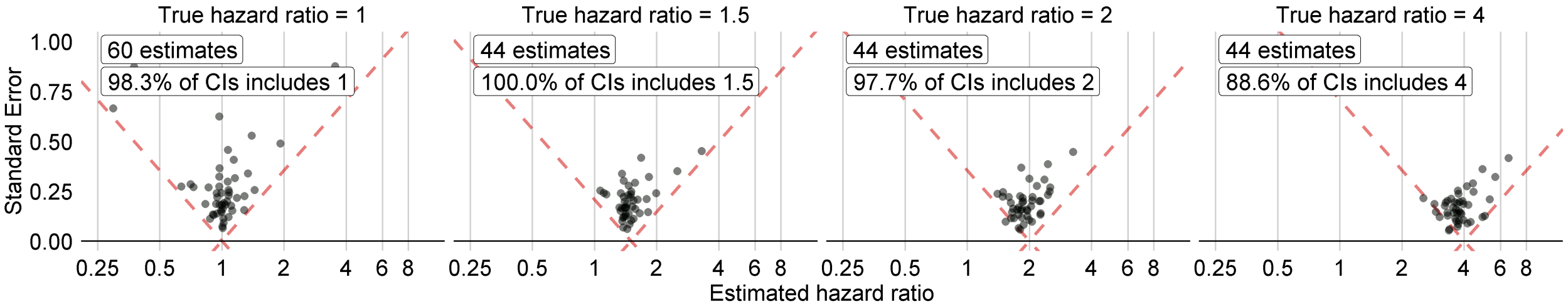


Do we balance on all observed variables?

Method Validity

Is this method valid for answering this question?

- Do we pass our (design-specific) **study diagnostics**?
- Do **negative** and **positive controls** show nominal operating characteristics? (after empirical calibration)

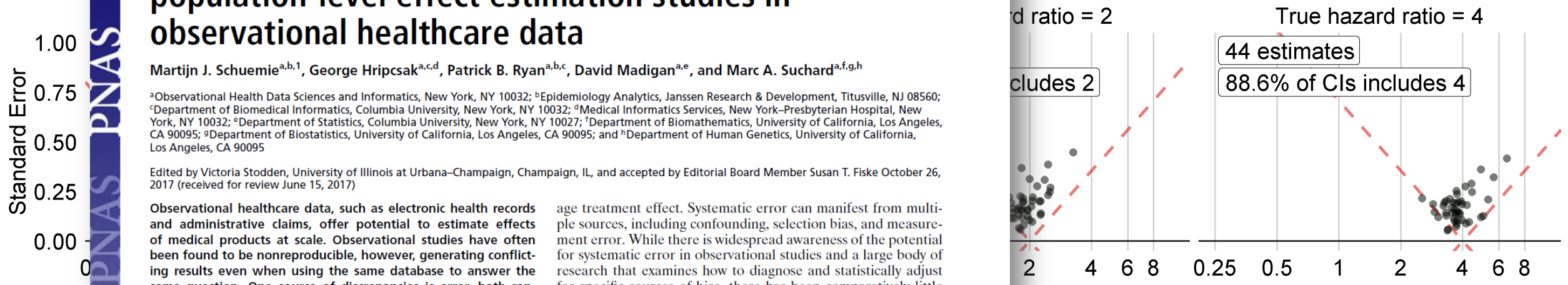


Method Validity

Is this method valid for answering this question?

- Do we pass our (design-specific) **study diagnostics**?

...nal operating



COLLOQUIUM PAPER

STATISTICS

Check for updates

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 same question. One source of discrepancies is error, both random caused by sampling variability and systematic (for example, because of confounding, selection bias, and measurement error). Only random error is typically quantified but converges to zero as databases become larger, whereas systematic error persists independent from sample size and therefore, increases in relative importance. Negative controls are exposure–outcome pairs,

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 for specific sources of bias, there has been comparatively little work in devising approaches to empirically estimate the magnitude of systematic error or clinical applications that show how to integrate this error into effect estimation methods.

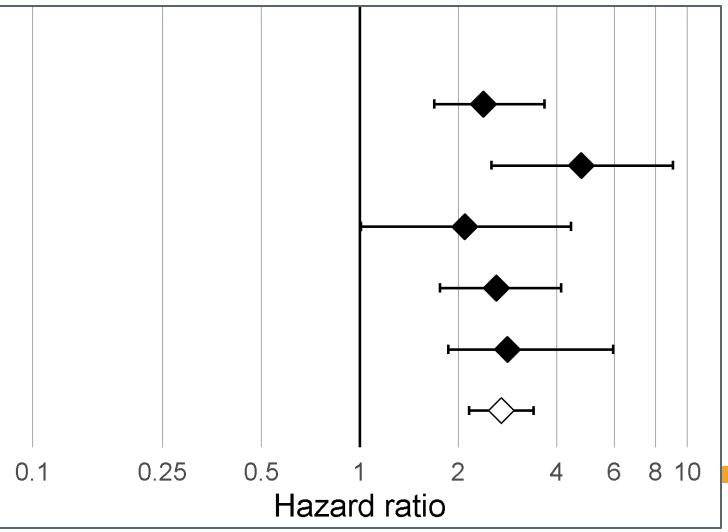
The acuity of this problem is only exacerbated as the size of observational databases grow: random error (the only component that is typically quantified) converges to zero as sample

Method Validity

Is this method valid for answering this question?

- Do we pass our (design-specific) **study diagnostics**?
- Do **negative** and **positive controls** show nominal operating characteristics?
- Do we observe **between-database heterogeneity**?

Source	Hazard ratio (95% CI)
CCAE	2.39 (1.69-3.66)
MDCD	4.75 (2.52-9.04)
MDCR	2.09 (1.01-4.41)
Optum	2.62 (1.75-4.11)
Panther	2.83 (1.86-5.94)
Summary (I^2 = 0.00)	2.71 (2.16-3.40)



Method Validity

Is this method valid for answering this question?

- Do we pass our (design-specific) **study diagnostics**?

minal operating

eneity?

American Journal of Epidemiology
© The Author 2013. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
Vol. 178, No. 4
DOI: 10.1093/aje/kwt010
Advance Access publication: May 5, 2013

Practice of Epidemiology

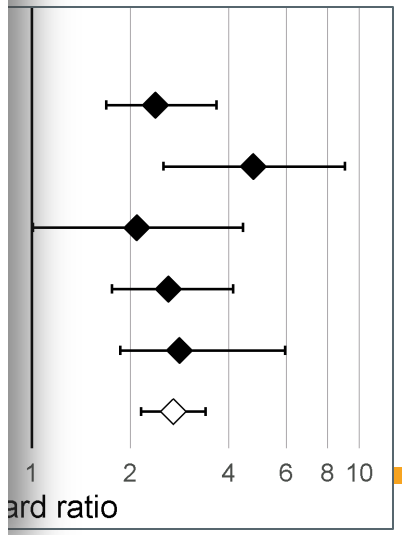
Evaluating the Impact of Database Heterogeneity on Observational Study Results

David Madigan*, Patrick B. Ryan, Martijn Schuemie, Paul E. Stang, J. Marc Overhage, Abraham G. Hartzema, Marc A. Suchard, William DuMouchel, and Jesse A. Berlin

* Correspondence to Dr. David Madigan, Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027 (e-mail: david.madigan@columbia.edu).

Initially submitted November 11, 2012; accepted for publication January 17, 2013.

Clinical studies that use observational databases to evaluate the effects of medical products have become commonplace. Such studies begin by selecting a particular database, a decision that published papers invariably report but do not discuss. Studies of the same issue in different databases, however, can and do generate different results, sometimes with strikingly different clinical implications. In this paper, we systematically study heterogene-



ard ratio

Best practices checklist

- Fully **prespecify** your study
  - Protocol
  - Study package

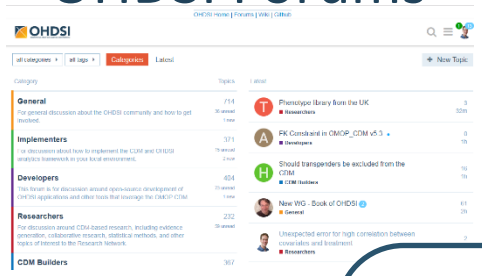
  Finalized when you **pass all study diagnostics**

  Must be **posted publicly**
- Establish **data quality** using the **Data Quality Dashboard** and other checks
- Establish **clinical validity**, e.g. using **PheValuator**
- Establish **software validity** by
  - Using the **OHDSI's standardized analytics tools** for the standardized parts, and
  - Using **code review** or **double coding** for study-specific parts
- Establish **method validity** by
  - Study diagnostics
  - Using negative and positive controls
  - Executing on multiple databases

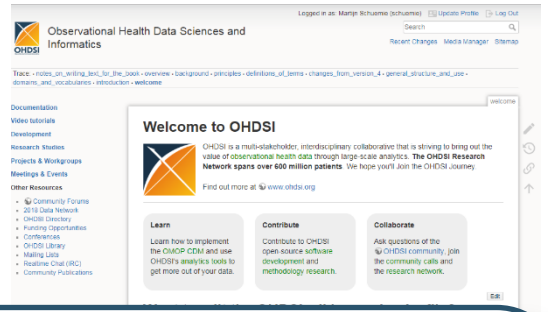
How can you follow these best practices yourself?

Where can I learn how to implement best practices?

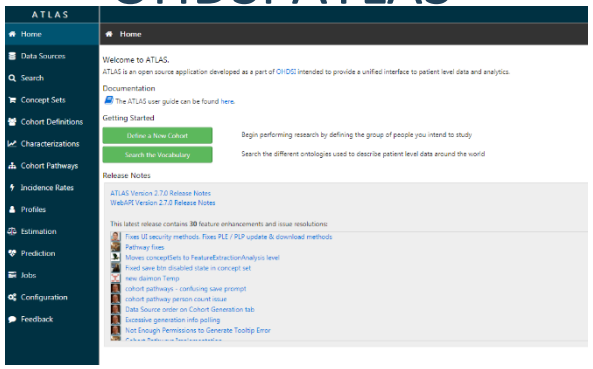
OHDSI Forums



OHDSI Wiki



OHDSI ATLAS

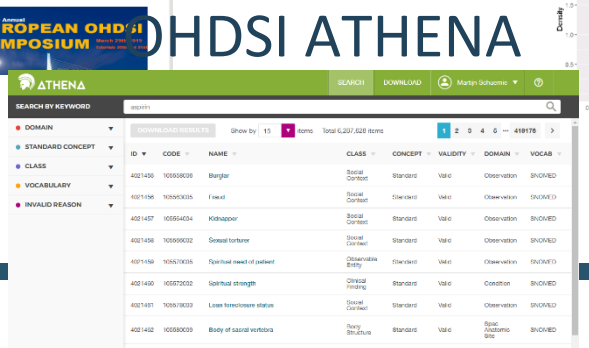


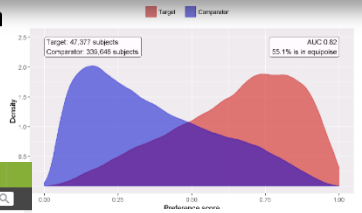
A lot of information is available

Just not in one place

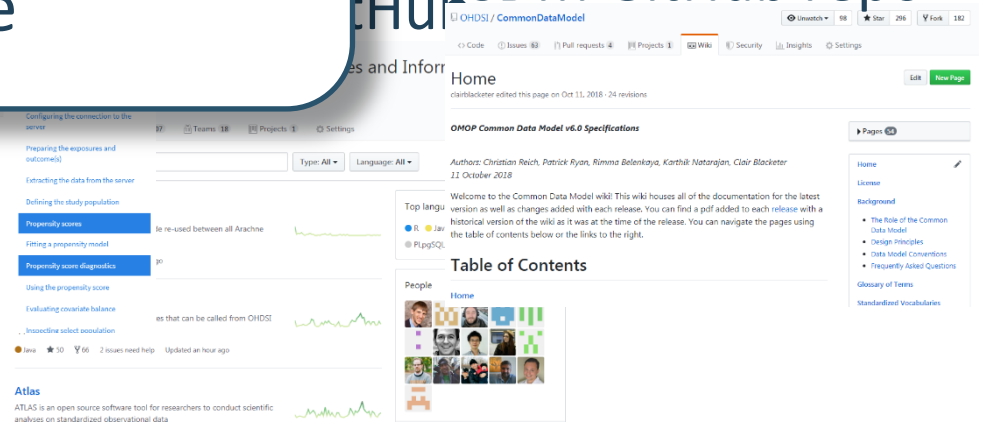
OHDSI Websit

OHDSI ATHENA

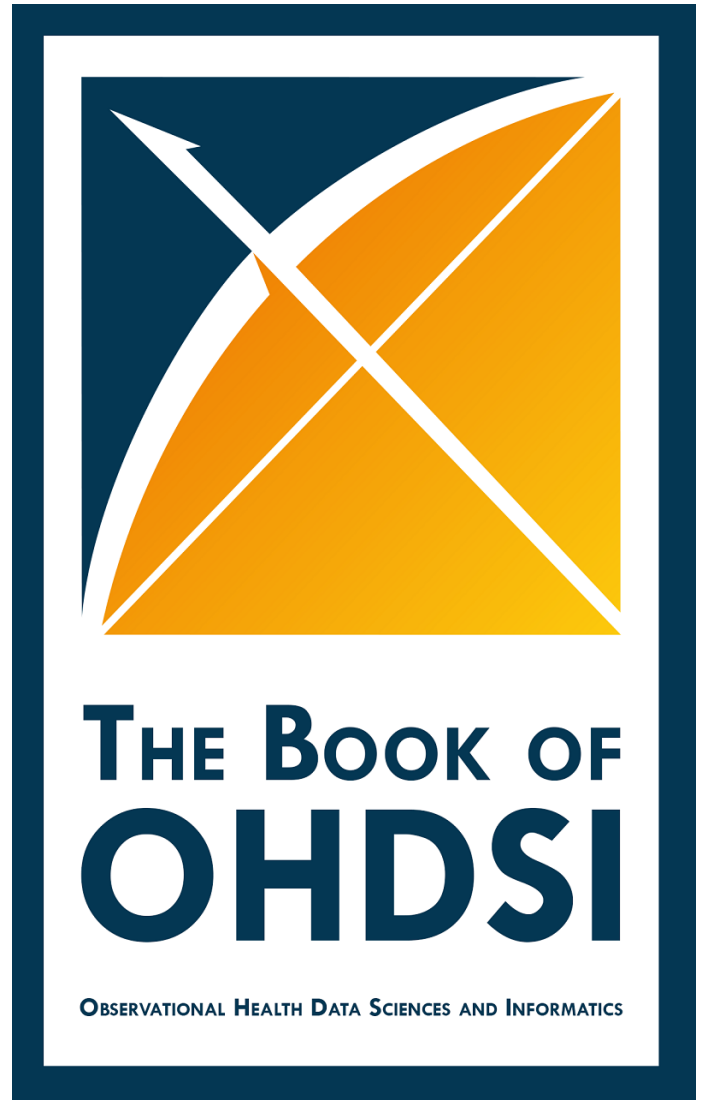




CDM GitHub repo - Wiki



A central knowledge repository…


THE BOOK OF
OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Contents of the Book of OHDSI

Parts:

I. The OHDSI Community
II. Uniform Data Representation
III. Data Analytics
IV. Evidence Quality
V. OHDSI Studies

Contents of the Book of OHDSI

Parts:

I. The OHDSI Community

II. Uniform Data Representation

III. Data Analytics

IV. Evidence Quality

V. OHDSI Studies

1. The OHDSI Community
2. Where to Begin
3. Open Science

Contents of the Book of OHDSI

Parts:

I. The OHDSI Community

II. Uniform Data Representation

III. Data Analytics

IV. Evidence Quality

V. OHDSI Studies

4. The Common Data Model
5. Standardized Vocabularies
6. Extract Transform Load

Contents of the Book of OHDSI

Parts:

I. The OHDSI Community

II. Uniform Data Representation

III. Data Analytics

IV. Evidence Quality

V. OHDSI Studies

7. Data Analytics Use Cases
8. OHDSI Analytics Tools
9. SQL and R
10. Defining Cohorts
11. Characterization
12. Population-Level Estimation
13. Patient-Level Prediction

Contents of the Book of OHDSI

Parts:

I. The OHDSI Community

II. Uniform Data Representation

III. Data Analytics

IV. Evidence Quality

V. OHDSI Studies

14. Evidence Quality
15. Data Quality
16. Clinical Validity
17. Software Validity
18. Method Validity

Contents of the Book of OHDSI

Parts:

I. The OHDSI Community

II. Uniform Data Representation

III. Data Analytics

IV. Evidence Quality

V. OHDSI Studies

19.Study steps

20.OHDSI Network Research

Best Practices & the Book of OHDSI

Best practices:

- Fully **prespecify** your study
- Establish **data quality**
- Establish **clinical validity**
- Establish **software validity**
- Establish **method validity**

Book chapters:

1. The OHDSI Community
2. Where to Begin
3. Open Science
4. The Common Data Model
5. Standardized Vocabularies
6. Extract Transform Load
7. Data Analytics Use Cases
8. OHDSI Analytics Tools
9. SQL and R
10. Defining Cohorts
11. Characterization
12. Population-Level Estimation
13. Patient Level Prediction
14. Evidence Quality
15. Data Quality
16. Clinical Validity
17. Software Validity
18. Method Validity
19. Study steps
20. OHDSI Network Research

Best Practices & the Book of OHDSI

Best practices:

- Fully **prespecify** your study
- Establish **data quality**
- Establish **clinical validity**
- Establish **software validity**
- Establish **method validity**

Book chapters:

1. The OHDSI Community
2. Where to Begin
3. Open Science
4. The Common Data Model
5. Standardized Vocabularies
6. Extract Transform Load
7. Data Analytics Use Cases
8. OHDSI Analytics Tools
9. SQL and R
10. Defining Cohorts
11. Characterization
12. Population-Level Estimation
13. Patient Level Prediction
14. Evidence Quality
15. Data Quality
16. Clinical Validity
17. Software Validity
18. Method Validity
19. Study steps
20. OHDSI Network Research

Best Practices & the Book of OHDSI

Best practices:

- Fully **prespecify** your study
- Establish **data quality**
- Establish **clinical validity**
- Establish **software validity**
- Establish **method validity**

Book chapters:

1. The OHDSI Community
2. Where to Begin
3. Open Science
4. The Common Data Model
5. Standardized Vocabularies
6. Extract Transform Load
7. Data Analytics Use Cases
8. OHDSI Analytics Tools
9. SQL and R
10. Defining Cohorts
11. Characterization
12. Population-Level Estimation
13. Patient Level Prediction
14. Evidence Quality
15. Data Quality
16. Clinical Validity
17. Software Validity
18. Method Validity
19. Study steps
20. OHDSI Network Research

Best Practices & the Book of OHDSI

Best practices:

- Fully **prespecify** your study
- Establish **data quality**
- Establish **clinical validity**
- Establish **software validity**
- Establish **method validity**

Book chapters:

1. The OHDSI Community
2. Where to Begin
3. Open Science
4. The Common Data Model
5. Standardized Vocabularies
6. Extract Transform Load
7. Data Analytics Use Cases
8. OHDSI Analytics Tools
9. SQL and R
10. Defining Cohorts
11. Characterization
12. Population-Level Estimation
13. Patient Level Prediction
14. Evidence Quality
15. Data Quality
16. Clinical Validity
17. Software Validity
18. Method Validity
19. Study steps
20. OHDSI Network Research

Best Practices & the Book of OHDSI

Best practices:

- Fully **prespecify** your study
- Establish **data quality**
- Establish **clinical validity**
- Establish **software validity**
- Establish **method validity**

Book chapters:

1. The OHDSI Community
2. Where to Begin
3. Open Science
4. The Common Data Model
5. Standardized Vocabularies
6. Extract Transform Load
7. Data Analytics Use Cases
8. OHDSI Analytics Tools
9. SQL and R
10. Defining Cohorts
11. Characterization
12. Population-Level Estimation
13. Patient Level Prediction
14. Evidence Quality
15. Data Quality
16. Clinical Validity
17. Software Validity
18. Method Validity
19. Study steps
20. OHDSI Network Research

Best Practices & the Book of OHDSI

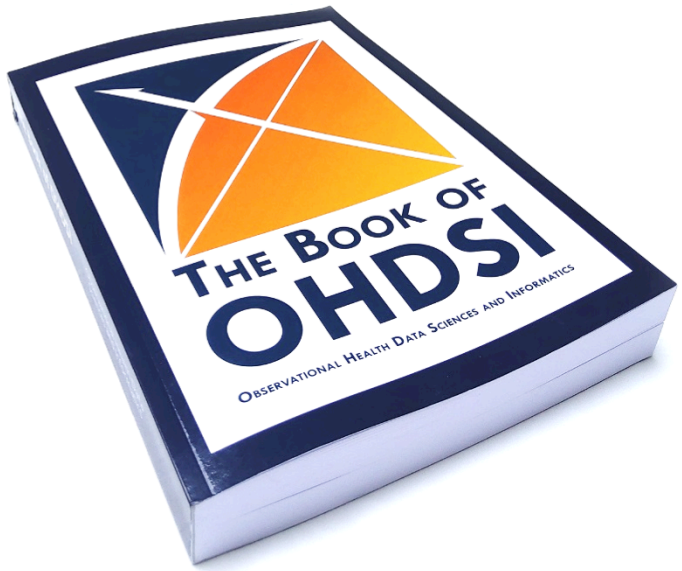
Best practices:

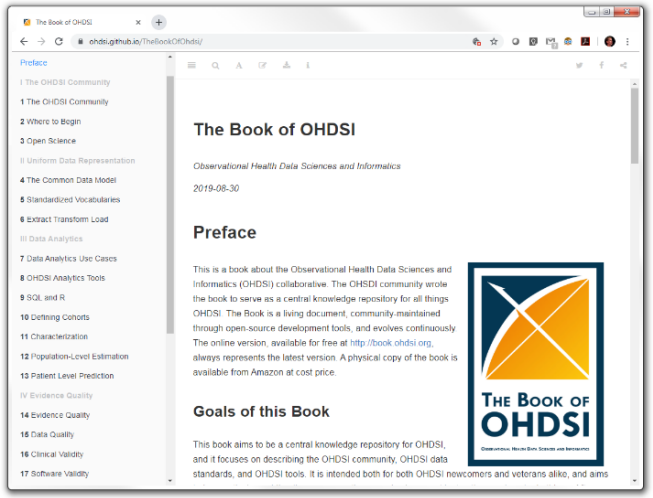
- Fully **prespecify** your study
- Establish **data quality**
- Establish **clinical validity**
- Establish **software validity**
- Establish **method validity**

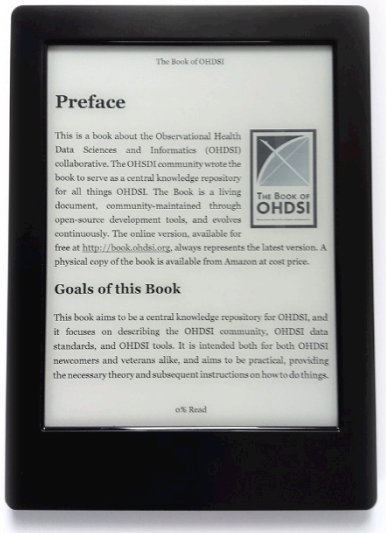
Book chapters:

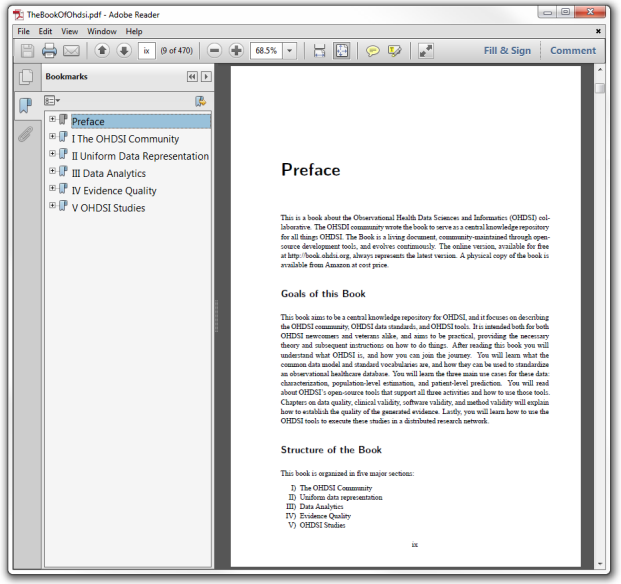
1. The OHDSI Community
2. Where to Begin
3. Open Science
4. The Common Data Model
5. Standardized Vocabularies
6. Extract Transform Load
7. Data Analytics Use Cases
8. OHDSI Analytics Tools
9. SQL and R
10. Defining Cohorts
11. Characterization
12. Population-Level Estimation
13. Patient Level Prediction
14. Evidence Quality
15. Data Quality
16. Clinical Validity
17. Software Validity
18. Method Validity
19. Study steps
20. OHDSI Network Research

Available in Different Formats









Paperback

HTML

EPUB

PDF

amazon.com

At cost price

book.ohdsi.org

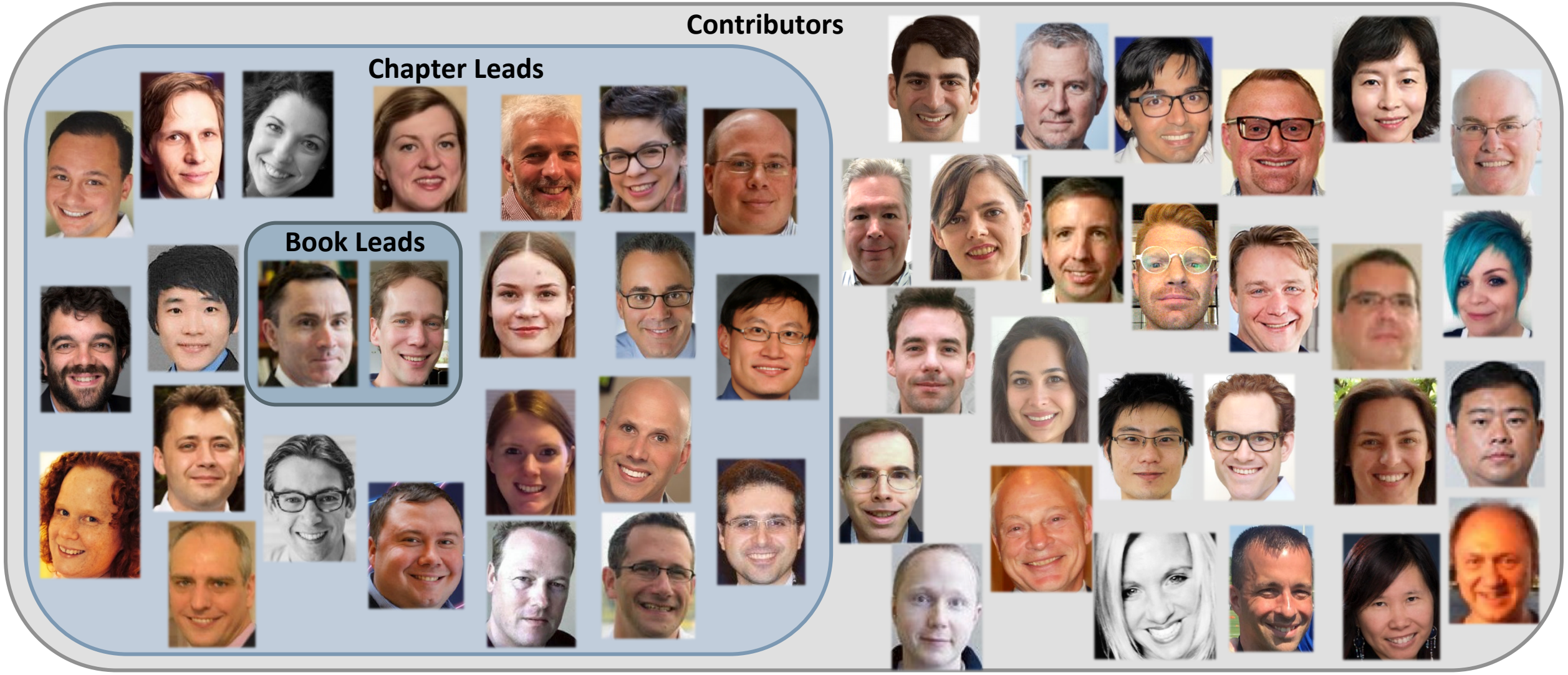
Free

Writing a Book OHDSI-Style!

Requirements:

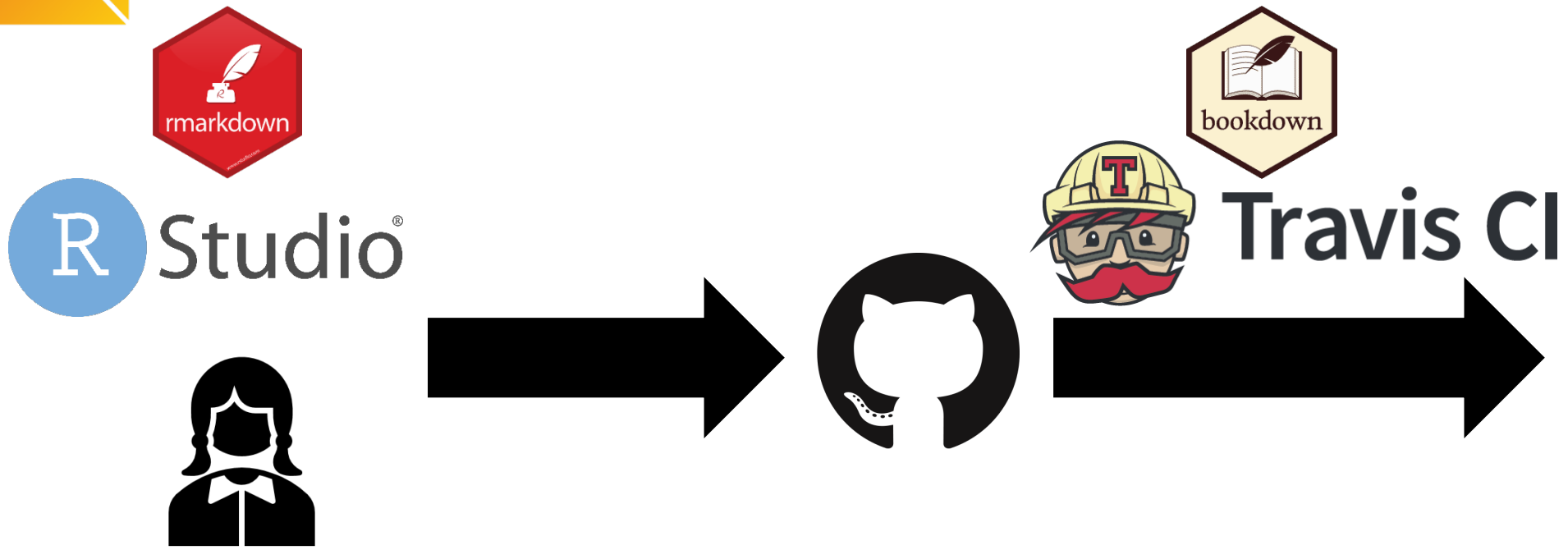
- Community effort
- Completely open and transparent

It takes a village…



Document-a-thon in Cleveland

Open and Transparent Process



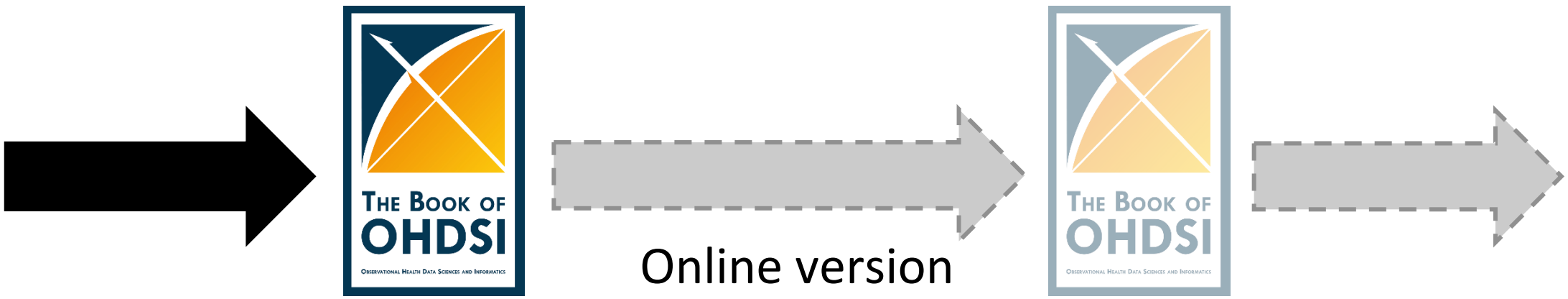
book.ohdsi.org
- HTML
- EPUB
- PDF

Contributor writes content in rmarkdown using RStudio

Pushes to open source GitHub repository

Automatically rendered to the Book

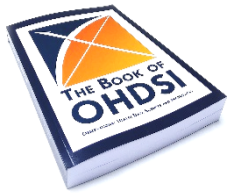
Living Document



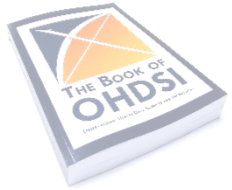
1st edition

Online version will continue to be updated

2nd edition



Paperback on amazon



Paperback on amazon

Thank you

Don’t forget your free copy of the book!