

Best Practices for Secondary Use

Harold P Lehmann, MD PhD

Professor

Biomedical Informatics and Data Science

How



- Methodology
- Reporting
- National Efforts
- Computable

Methodological Considerations

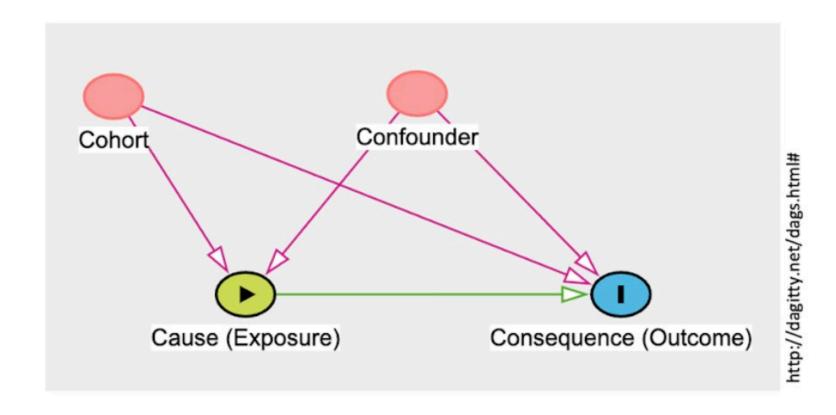


PRIMARY	SECONDARY (RWE)
Hypothesis-based	Hypothesis + Generative
Bespoke data collection	Ad hoc data collection
Pre-data collection analytic plan	Post-data collection analytic plan
Design minimizes biases	Data collection maximizes biases
Established rules	Evolving rules
∴ We know what is definitive	∴ We are not sure what is definitive



Pre-Data Collection Analytic Plan

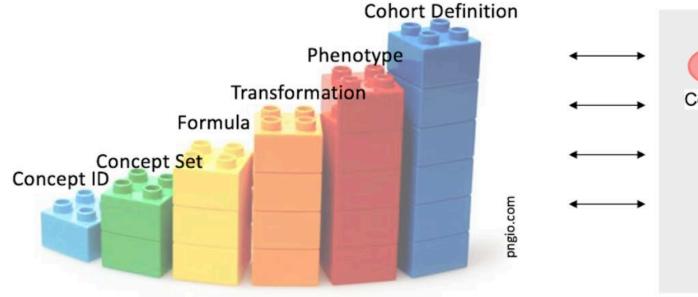
Analytic Variables



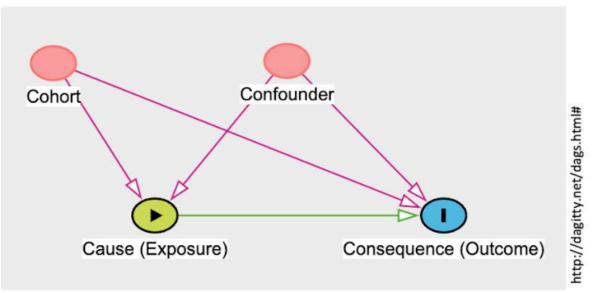
Post-Data Collection Analytic Plan



Data Constructs



Analytic Variables

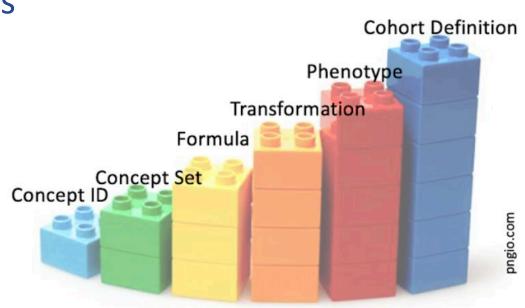


Data Constructs



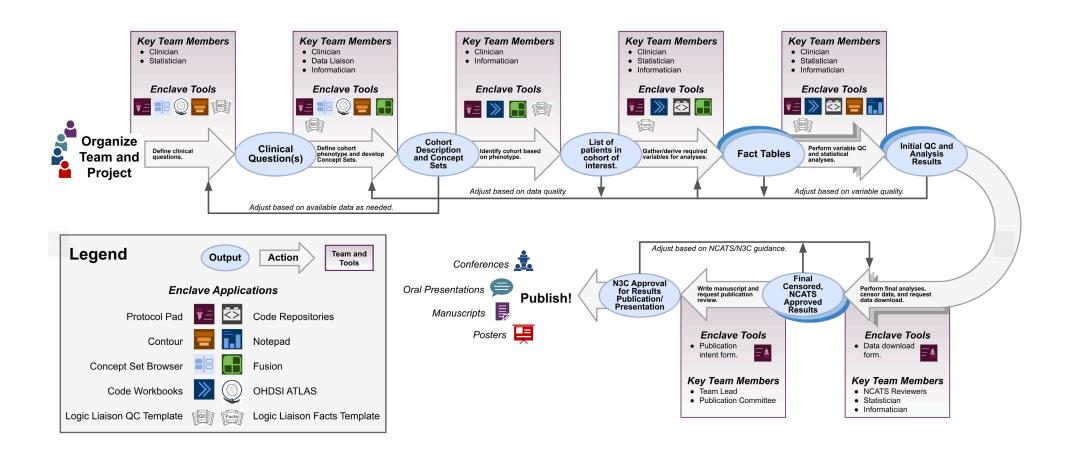
- Concept ID Standard code for a datum
- Concept Set Set of "synonymous" codes
- Formula E.g., BMI
- Transformation E.g., "inpatient stay"
- Phenotype E.g., "experienced MI"
- Cohort Definition Phenotype + Time

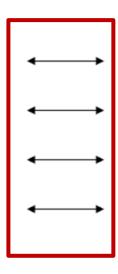
Data Constructs



"Data Cleaning" as First-Ranked Activity

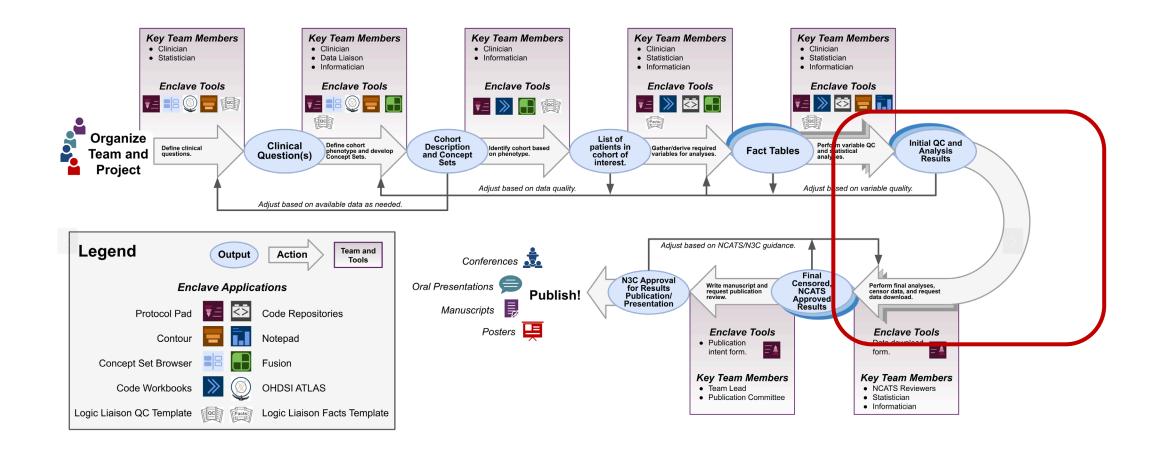
















Structured Template and Reporting Tool for Real World Evidence (STaRT-RWE)

Table 1. Administrative Information

Table 2. Version History

Figure 1. Design Diagram

Table 3. Summary of Analytic Study Population

A. Meta-data about data source and software

B. Index Date (day 0) defining criterion

C. Inclusion Criteria

D. Exclusion Criteria

E. Predefined Covariates

F. Empirically Defined Covariates

G. Outcome

H. Follow up

Table 4. Analysis Specification

Table 5. Sensitivity Analyses

Table 6. Attrition Table

Table 7. Power and Sample Size Calculation

Table 8. Glossary of Terminology

Table 9. Abbreviations

Also

STROBE

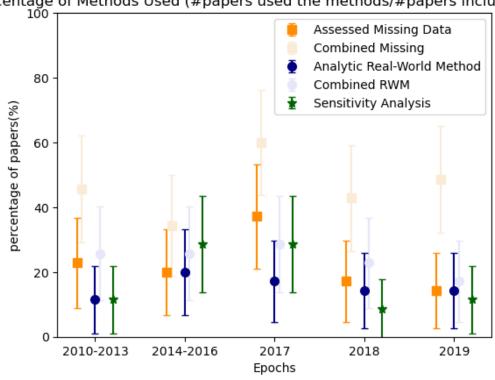
RECORD

(see equator-network)

Poor State of EHR-Based Research







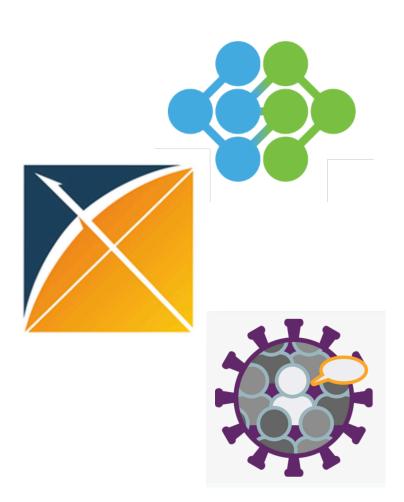
"The papers that addressed all 3 phases of of analysis numbered 5 (2.9%)."

Li C, Alsheikh AM, Robinson KR, Lehmann HP. Use of Recommended Real-World Methods for Electronic Health Record Data Analysis Has Not Improved Over 10 Years. https://www.medrxiv.org/content/10.1101/2023.06.21.23291706v1.

Examples from National Efforts

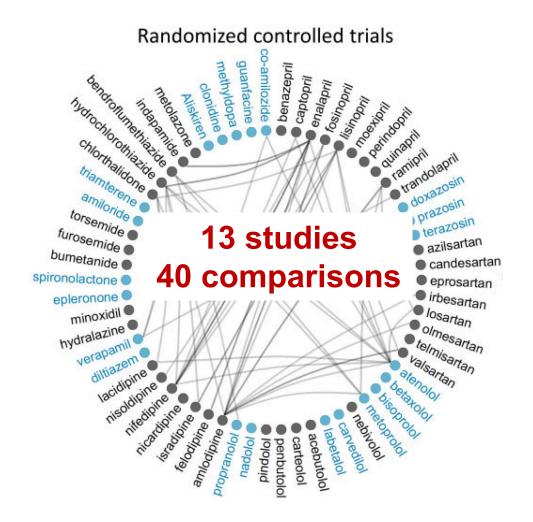


- PCORnet
 - 10 years, 9 Networks, > 800 Studies
- OHDSI
 - 14 years, > 200 Networks, > 575 Studies
- N3C
 - 3 years, > 70 sites (networks), >170 studies



Traditional Systematic Review



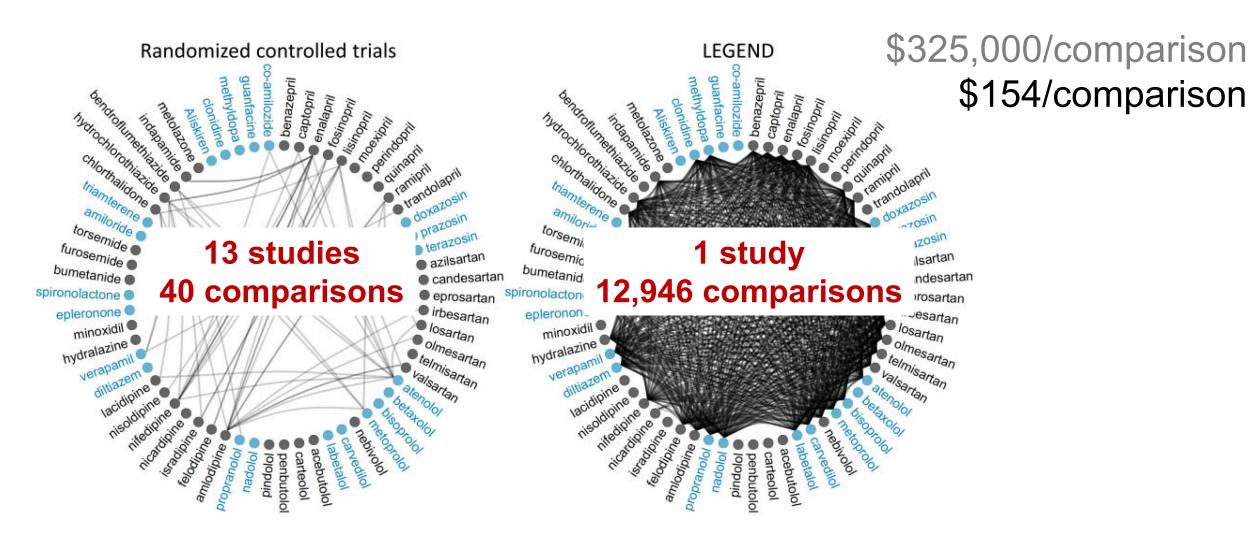


\$325,000/comparison

Source: Schuemie MJ, et al. Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study. J Am Med Inform Assoc. 2020 Aug 1;27(8):1268-1277. Erratum in: J Am Med Inform Assoc. 2021 Jan 15;28(1):196.

Massively Parallel Analysis





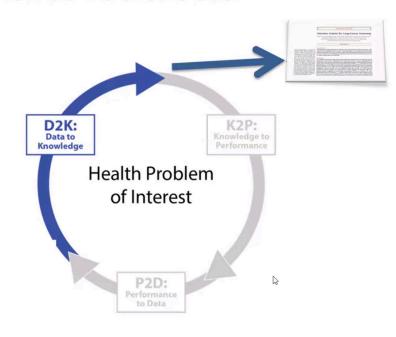
Source: Schuemie MJ, et al. Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study. J Am Med Inform Assoc. 2020 Aug 1;27(8):1268-1277. Erratum in: J Am Med Inform Assoc. 2021 Jan 15;28(1):196.

Communicating

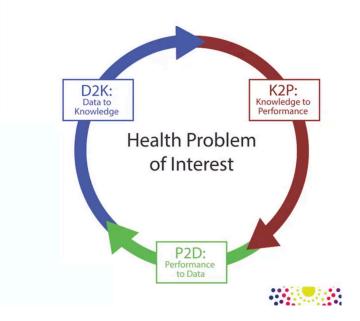


The Challenge...

• How do we evolve this:







FHIR:

Fast Health Interoperable Resources

Currently: Clinical Data

Future: Evidence EBMonFHIR

(Now: HEvKA:

Health Evidence

Knowledge Accelerator)

Courtesy: Chuck Friedman

Open For Involvement!



FEvIR Platform

