AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OHDSI RWE Revolution:

Igniting Data Modernization with Harmonized Standards for Cutting-Edge Health Research

11-Nov-2023

Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Introduction to OMOP and OHDSI

Or: What is it and why should you consider it?

11-Nov-2023

Disclosure

I disclose the following relevant relationship with commercial interests:

- CEO of Odysseus Data Services

The OMOP Experiment

79 of 4,000 Vioxx users suffered heart problems or died

Observational Medical Outcomes Partnership

FDAAA calls for establishing Risk Identification and Analysis System

SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.

(a) IN GENERAL.—Subsection (k) of section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

"(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—

"(A) DEFINITION.—In this paragraph, the term 'data' refers to information with respect to a drug approved under this section or under section 351 of the Public Health Service Act, including claims data, patient survey data, standardized analytic files that allow for the pooling and analysis of data from disparate data environments, and any other data deemed appropriate by the Secretary.

"(B) DEVELOPMENT OF POSTMARKET RISK IDENTIFICATION AND ANALYSIS METHODS.—The Secretary shall, not later than 2 years after the date of the enactment of the Food and Drug Administration Amendments Act of 2007, in collaboration with public, academic, and private entities—

"(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);

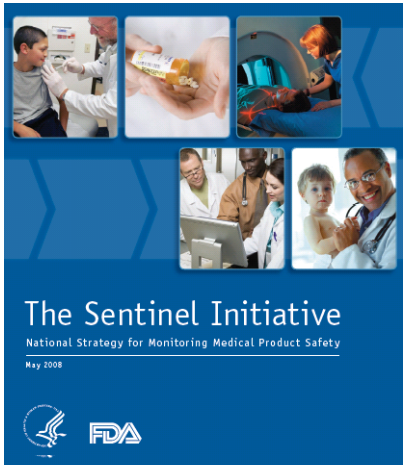
"(ii) develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

"(I) at least 25,000,000 patients by July 1, 2010; and

"(II) at least 100,000,000 patients by July 1, 2012; and

"(iii) convene a committee of experts, including individuals who are recognized in the field of protecting data privacy and security, to make recommendations to the Secretary on the development of tools and methods for the ethical and scientific uses for, and communication of, postmarketing data specified under subparagraph (C), including recommendations on the development of effective research methods for the study of drug safety questions.

"(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTIFICATION AND ANALYSIS SYSTEM.—

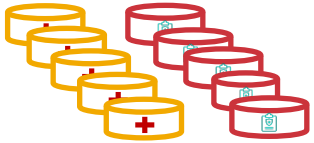


Risk Identification and Analysis System:

a systematic and reproducible process to efficiently generate evidence to support the characterization of the potential effects of medical products from across a network of disparate observational healthcare data sources

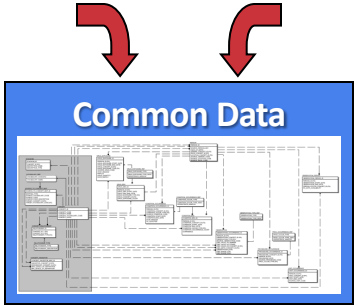
Observational Medical Outcomes Partnership

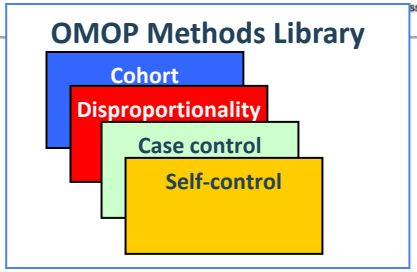
OMOP experiment



10 Databases

- Open-source
- Standards-based





- 14 methods * 70 settings = **1,000 SAS scripts**

10 Drugs

10 Outcomes

	ACE Inhibitors	Amphotericin B	Antibiotics: erythromycins, sulfonamides, tetracyclines	Antiepileptics: carbamazepine, phenytoin	Benzodiazepines	Beta blockers	Bisphosphonates: alendronate	Tricyclic antidepressants	Typical antipsychotics	Warfarin
Angioedema										
Aplastic Anemia										
Acute Liver Injury										
Bleeding										
Hip Fracture										
Hospitalization										
Myocardial Infarction										
Mortality after MI										
Renal Failure										
GI Ulcer Hospitalization										

OMOP to OHDSI

OMOP Investigators

OHDSI

Columbia University

The Observational Health Data Sciences and Informatics (OHDSI) program is a **multi-stakeholder, interdisciplinary collaborative** to create **open-source** solutions that bring out the value of observational health data through large-scale analytics

OHDSI has established an **international network of researchers and observational health databases** with a central coordinating centre housed at Columbia University

Public, open

Not pharma funded

International

OHDSI's mission

To improve health

by empowering a community

to collaboratively generate the evidence that promotes

better health decisions and better care

OHDSI's values

Innovation: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.

Reproducibility: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.

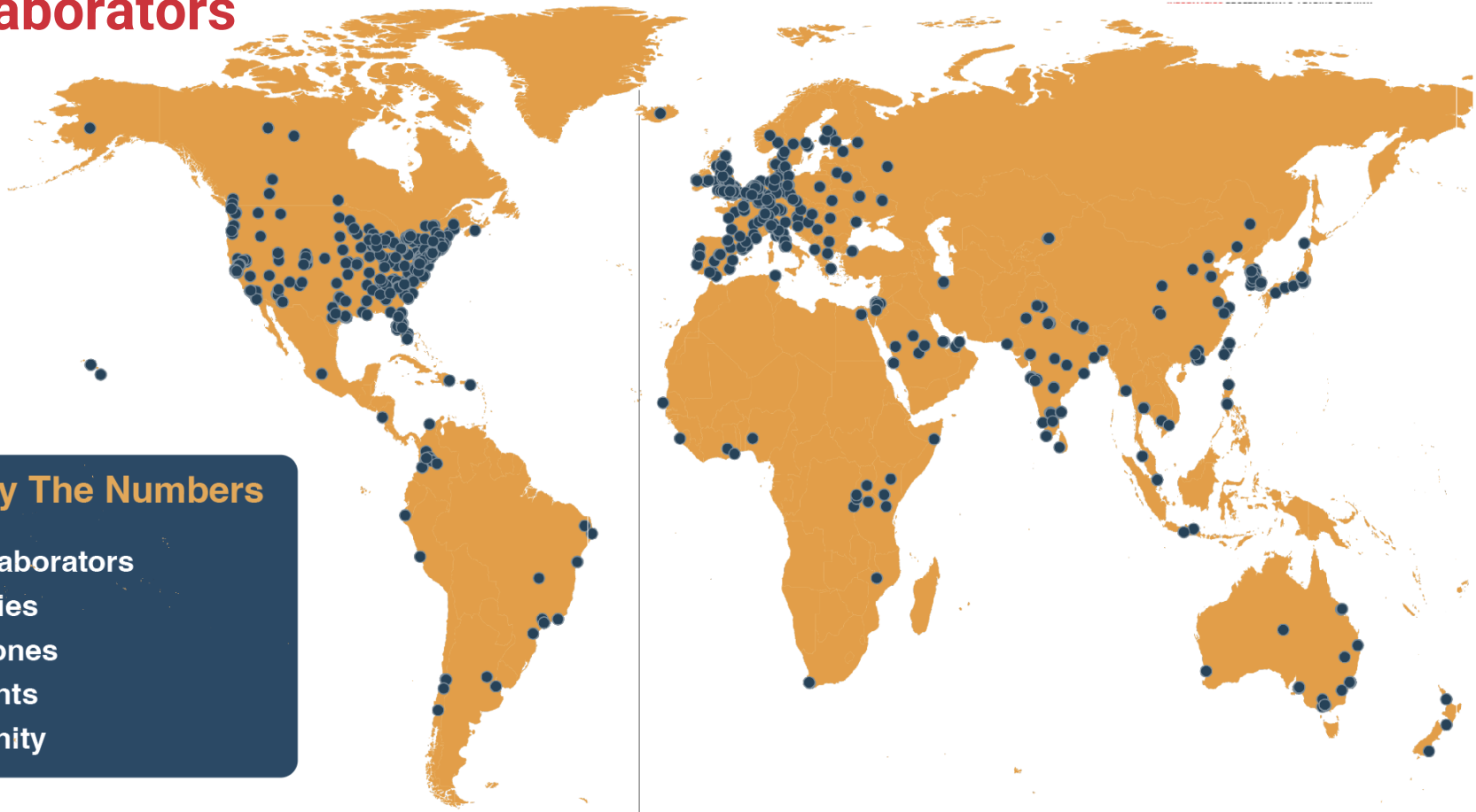
Community: Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.

Collaboration: We work collectively to prioritize and address the real world needs of our community's participants.

Openness: We strive to make all our community's proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.

Beneficence: We seek to protect the rights of individuals and organizations within our community at all times.

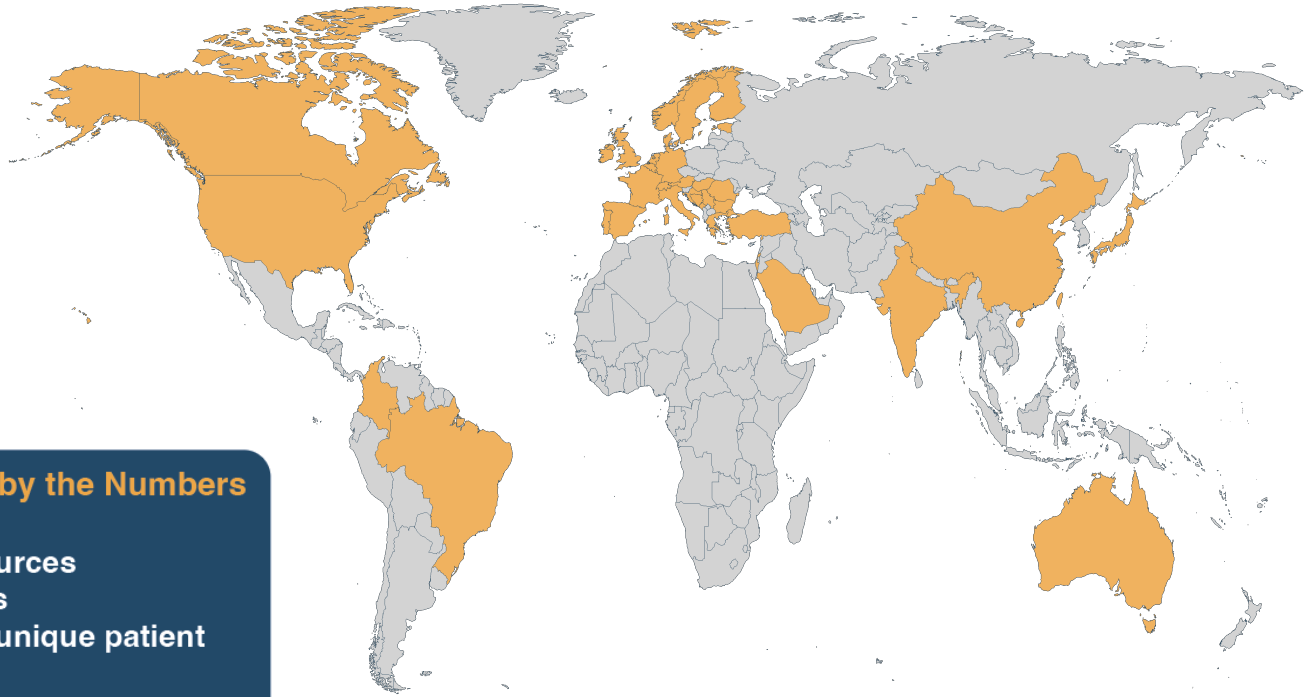
Collaborators



OHDSI By The Numbers

- 3,758 collaborators
- 83 countries
- 21 time zones
- 6 continents
- 1 community

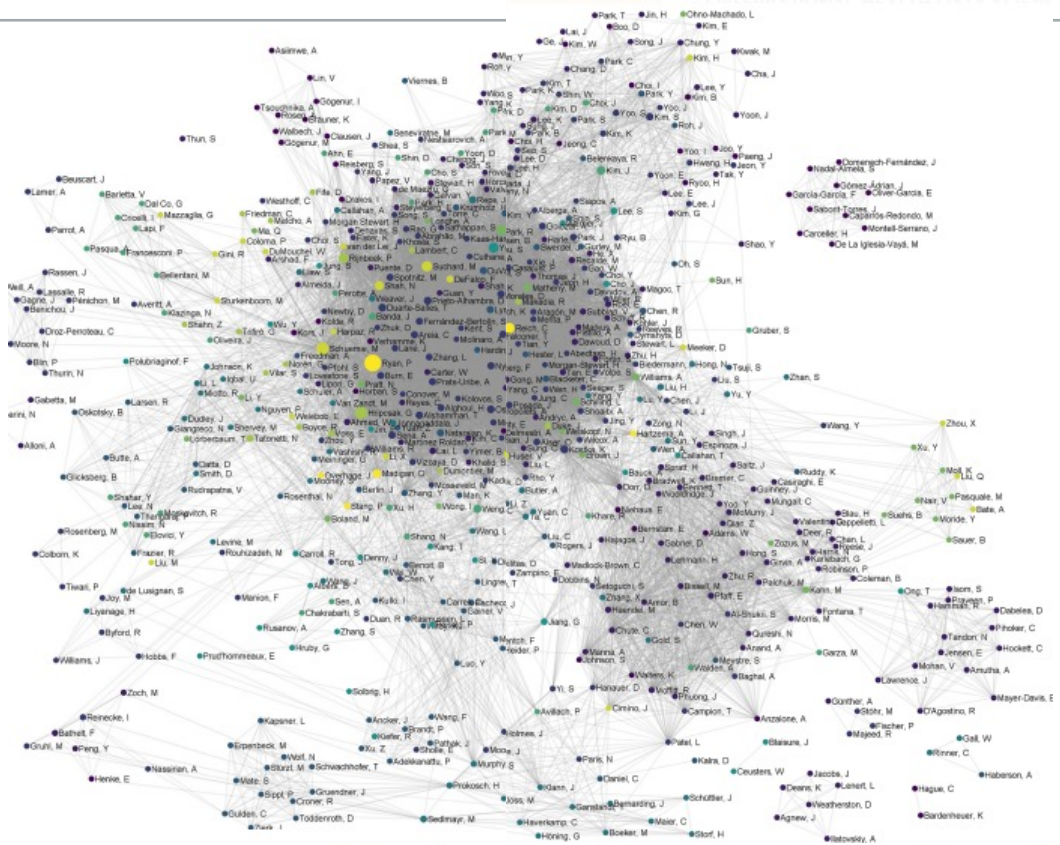
Data Partners



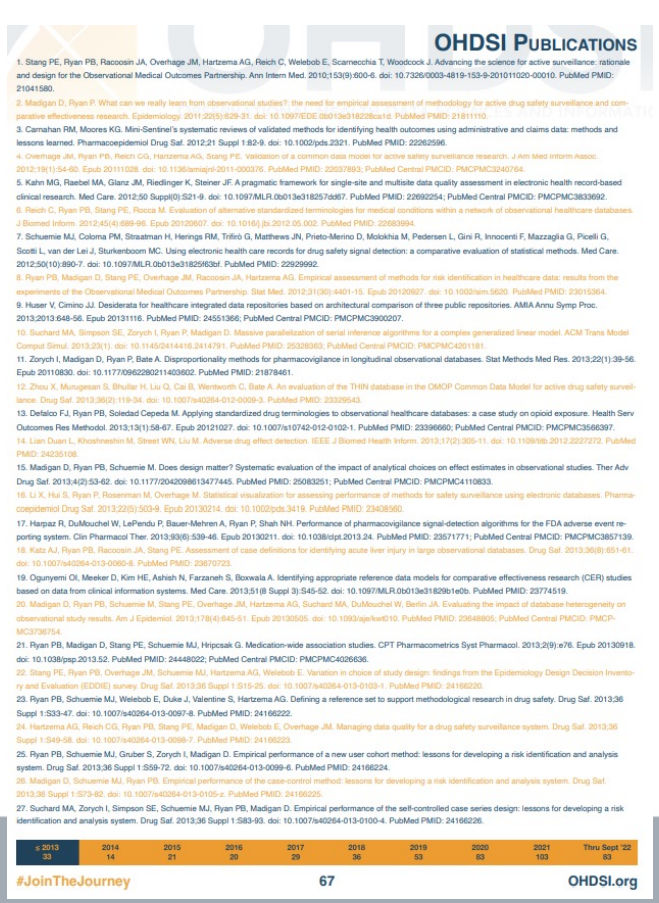
OMOP Data by the Numbers

- 534 data sources
- 49 countries
- 956 million unique patient records
- approximately 12% of the world's population

The Author Network



OHDSI Publications



OHDSI Publications

1. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, Welebob E, Scarnecchia T, Woodcock J. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann Intern Med. 2010;153(9):600-6. doi: 10.7326/0003-4819-153-9-201011020-00010. PubMed PMID: 21041580.

2. Madigan D, Ryan P. What can we really learn from observational studies? the need for empirical assessment of methodology for active drug safety surveillance and comparative effectiveness research. Epidemiology 2011;22(5):629-31. doi: 10.1097/EDE.0b013e318228cca1d. PubMed PMID: 21811110.

3. Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiol Drug Saf. 2012;21 Suppl 1:82-9. doi: 10.1002/pds.2321. PubMed PMID: 22262596.

4. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60. Epub 20111028. doi: 10.1136/amiajnl-2011-000376. PubMed PMID: 22037893; PubMed Central PMCID: PMCPMC3240764.

5. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Med Care. 2012;50 Suppl(0):S21-9. doi: 10.1097/MLR.0b013e318257dd67. PubMed PMID: 22692254; PubMed Central PMCID: PMCPMC3833692.

6. Reich C, Ryan PB, Stang PE, Rocca M. Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. J Biomed Inform. 2012;45(4):689-96. Epub 20120607. doi: 10.1016/j.jbi.2012.05.002. PubMed PMID: 22683994.

7. Schuemie MJ, Coloma PM, Straatman H, Herings RM, Trifirò G, Matthews JN, Prieto-Merino D, Molokhia M, Pedersen L, Gini R, Innocenti F, Mazzaglia G, Picelli G, Scotti L, van der Lei J, Sturkenboom MC. Using electronic health care records for drug safety signal detection: a comparative evaluation of statistical methods. Med Care. 2012;50(10):890-7. doi: 10.1097/MLR.0b013e31825f63bf. PubMed PMID: 22929992.

8. Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012;31(30):4401-15. Epub 20120927. doi: 10.1002/sim.5620. PubMed PMID: 23015364.

9. Huser V, Cimino JJ. Desiderata for healthcare integrated data repositories based on architectural comparison of three public repositories. AMIA Annu Symp Proc. 2013;2013:648-56. Epub 20131116. PubMed PMID: 24551366; PubMed Central PMCID: PMCPMC3900207.

10. Suchard MA, Simpson SE, Zorych I, Ryan P, Madigan D. Massive parallelization of serial inference algorithms for a complex generalized linear model. ACM Trans Model Comput Simul. 2013;23(1). doi: 10.1145/2414416.2414791. PubMed PMID: 25328363; PubMed Central PMCID: PMCPMC4201181.

11. Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. Stat Methods Med Res. 2013;22(1):39-56. Epub 20110830. doi: 10.1177/0962280211403602. PubMed PMID: 21878461.

12. Zhou X, Murugesan S, Bhullar H, Liu Q, Cai B, Wentworth C, Bate A. An evaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance. Drug Saf. 2013;36(2):119-34. doi: 10.1007/s40264-012-0009-3. PubMed PMID: 23329543.

13. Defalco FJ, Ryan PB, Soledad Cepeda M. Applying standardized drug terminologies to observational healthcare databases: a case study on opioid exposure. Health Serv Outcomes Res Methodol. 2013;13(1):58-67. Epub 20121027. doi: 10.1007/s10742-012-0102-1. PubMed PMID: 23396660; PubMed Central PMCID: PMCPMC3566397.

14. Lian Duan L, Khoshneshin M, Street WN, Liu M. Adverse drug effect detection. IEEE J Biomed Health Inform. 2013;17(2):305-11. doi: 10.1109/titb.2012.2227272. PubMed PMID: 24235109.

15. Madigan D, Ryan PB, Schuemie M. Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies. Ther Adv Drug Saf. 2013;4(2):53-62. doi: 10.1177/2042098613477445. PubMed PMID: 25083251; PubMed Central PMCID: PMCPMC4110833.

16. Li X, Hui S, Ryan P, Rosenman M, Overhage M. Statistical visualization for assessing performance of methods for safety surveillance using electronic databases. Pharmacoepidemiol Drug Saf. 2013;22(5):503-9. Epub 20130214. doi: 10.1002/pds.3419. PubMed PMID: 23408560.

17. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clin Pharmacol Ther. 2013;93(6):539-46. Epub 20130211. doi: 10.1038/clpt.2013.24. PubMed PMID: 23571771; PubMed Central PMCID: PMCPMC3857139.

18. Katz AJ, Ryan PB, Racoosin JA, Stang PE. Assessment of case definitions for identifying acute liver injury in large observational databases. Drug Saf. 2013;36(8):651-61. doi: 10.1007/s40264-013-0060-8. PubMed PMID: 23670723.

19. Ogunyemi OI, Meeker D, Kim HE, Ashish N, Farzaneh S, Boxwala A. Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems. Med Care. 2013;51(8 Suppl 3):S45-52. doi: 10.1097/MLR.0b013e31829b1e0b. PubMed PMID: 23774519.

20. Madigan D, Ryan PB, Schuemie M, Stang PE, Overhage JM, Hartzema AG, Suchard MA, DuMouchel W, Berlin JA. Evaluating the impact of database heterogeneity on observational study results. Am J Epidemiol. 2013;178(4):645-51. Epub 20130505. doi: 10.1093/aje/kwt010. PubMed PMID: 23648805; PubMed Central PMCID: PMCPMC3736754.

21. Ryan PB, Madigan D, Stang PE, Schuemie MJ, Hripcsak G. Medication-wide association studies. CPT Pharmacometrics Syst Pharmacol. 2013;2(9):e76. Epub 20130918. doi: 10.1038/psp.2013.52. PubMed PMID: 24448022; PubMed Central PMCID: PMCPMC4026636.

22. Stang PE, Ryan PB, Overhage JM, Schuemie MJ, Hartzema AG, Welebob E. Variation in choice of study design: findings from the Epidemiology Design Decision Inventory and Evaluation (EDDIE) survey. Drug Saf. 2013;36 Suppl 1:S15-25. doi: 10.1007/s40264-013-0103-1. PubMed PMID: 24166220.

23. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug Saf. 2013;36 Suppl 1:S33-47. doi: 10.1007/s40264-013-0097-8. PubMed PMID: 24166222.

24. Hartzema AG, Reich CG, Ryan PB, Stang PE, Madigan D, Welebob E, Overhage JM. Managing data quality for a drug safety surveillance system. Drug Saf. 2013;36 Suppl 1:S49-58. doi: 10.1007/s40264-013-0098-7. PubMed PMID: 24166223.

25. Ryan PB, Schuemie MJ, Gruber S, Zorych I, Madigan D. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. Drug Saf. 2013;36 Suppl 1:S59-72. doi: 10.1007/s40264-013-0099-6. PubMed PMID: 24166224.

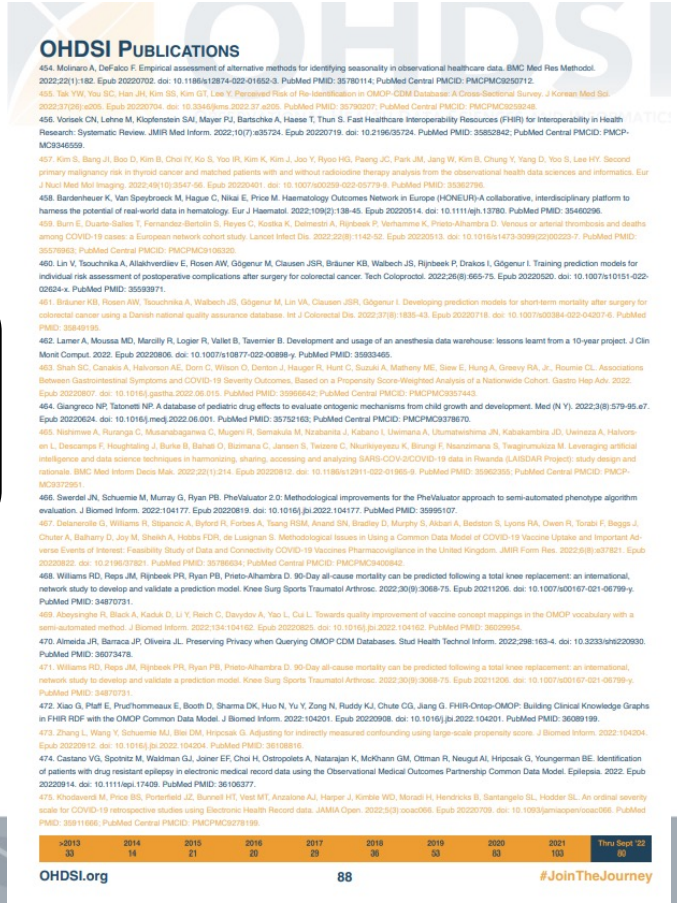
26. Madigan D, Schuemie MJ, Ryan PB. Empirical performance of the case-control method: lessons for developing a risk identification and analysis system. Drug Saf. 2013;36 Suppl 1:S73-82. doi: 10.1007/s40264-013-0105-z. PubMed PMID: 24166225.

27. Suchard MA, Zorych I, Simpson SE, Schuemie MJ, Ryan PB, Madigan D. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. Drug Saf. 2013;36 Suppl 1:S83-93. doi: 10.1007/s40264-013-0100-4. PubMed PMID: 24166226.

≤ 2013	2014	2015	2016	2017	2018	2019	2020	2021	Thru Sept '22
33	14	21	20	29	36	53	63	103	63

#JoinTheJourney 67 OHDSI.org

22 pages highlighting the **475** publications from our community



OHDSI Publications

454. Molinaro A, DeFalco F. Empirical assessment of alternative methods for identifying seasonality in observational healthcare data. BMC Med Res Methodol. 2022;22(1):182. Epub 20220702. doi: 10.1186/s12874-022-01652-3. PubMed PMID: 35780114; PubMed Central PMCID: PMCPMC9250712.

455. Tak YW, You SC, Han JH, Kim SS, Kim GT, Lee Y. Perceived Risk of Re-Identification in OMOP-CDM Database: A Cross-Sectional Survey. J Korean Med Sci. 2022;37(26):e205. Epub 20220704. doi: 10.3346/jkms.2022.37.e205. PubMed PMID: 35793027; PubMed Central PMCID: PMCPMC9259248.

456. Vorisek CN, Lehne M, Klopfenstein SAI, Mayer PJ, Bartschke A, Haese T, Thun S. Fast Healthcare Interoperability Resources (FHIR) for Interoperability in Health Research: Systematic Review. JMIR Med Inform. 2022;10(7):e35724. Epub 20220719. doi: 10.2196/35724. PubMed PMID: 35852842; PubMed Central PMCID: PMCPMC9346559.

457. Kim S, Bang JI, Boo D, Kim B, Choi IY, Ko S, Yoo IR, Kim K, Kim J, Joo Y, Ryoo HG, Paeng JC, Park JM, Jang W, Kim B, Chung Y, Yang D, Yoo S, Lee HY. Second primary malignancy risk in thyroid cancer and matched patients with and without radioiodine therapy analysis from the observational health data sciences and informatics. Eur J Nucl Med Mol Imaging. 2022;49(10):3547-56. Epub 20220401. doi: 10.1007/s00259-022-05779-8. PubMed PMID: 35362796.

458. Bardenheuer K, Van Speybroeck M, Hague C, Nikai E, Price M. Haematology Outcomes Network in Europe (HONEUR)-A collaborative, interdisciplinary platform to harness the potential of real-world data in hematology. Eur J Haematol. 2022;109(2):138-45. Epub 20220514. doi: 10.1111/ejh.13780. PubMed PMID: 35460296.

459. Burn E, Duarte-Salles T, Fernandez-Bertolin S, Reyes C, Kostka K, Delmestri A, Rijnbeek P, Verhamme K, Prieto-Alhambra D. Venous or arterial thrombosis and deaths among COVID-19 cases: a European network cohort study. Lancet Infect Dis. 2022;22(8):1142-52. Epub 20220513. doi: 10.1016/s1473-3099(22)00223-7. PubMed PMID: 35576963; PubMed Central PMCID: PMCPMC9106320.

460. Lin V, Tsouchnika A, Allakhverdiev E, Rosen AW, Gögenur M, Clausen JSR, Bräuner KB, Walbech JS, Rijnbeek P, Drakos I, Gögenur I. Training prediction models for individual risk assessment of postoperative complications after surgery for colorectal cancer. Tech Coloproctol. 2022;26(8):665-75. Epub 20220520. doi: 10.1007/s10151-022-02624-x. PubMed PMID: 35593971.

461. Bräuner KB, Rosen AW, Tsouchnika A, Walbech JS, Gögenur M, Lin VA, Clausen JSR, Gögenur I. Developing prediction models for short-term mortality after surgery for colorectal cancer using a Danish national quality assurance database. Int J Colorectal Dis. 2022;37(8):1835-43. Epub 20220718. doi: 10.1007/s00384-022-04207-6. PubMed PMID: 35849193.

462. Lamer A, Moussa MD, Marcilly R, Logier R, Vallet B, Tavernier B. Development and usage of an anesthesia data warehouse: lessons learnt from a 10-year project. J Clin Monit Comput. 2022. Epub 20220806. doi: 10.1007/s10877-022-00898-y. PubMed PMID: 35933465.

463. Shah SC, Canakis A, Halvorson AE, Burns G, Wilson O, Denton J, Haeger R, Hunt C, Suzuki A, Matheny ME, Siew E, Hung A, Greevy RA, Jr., Roumie CL. Associations Between Gastrointestinal Symptoms and COVID-19 Severity Outcomes, Based on a Propensity Score-Weighted Analysis of a Nationwide Cohort. Gastro Hep Adv. 2022. Epub 20220807. doi: 10.1016/j.gastha.2022.06.015. PubMed PMID: 35966642; PubMed Central PMCID: PMCPMC9357443.

464. Giangreco NP, Tatonetti NP. A database of pediatric drug effects to evaluate ontogenic mechanisms from child growth and development. Med (N Y). 2022;3(8):579-95.e7. Epub 20220624. doi: 10.1016/j.medj.2022.06.001. PubMed PMID: 35752163; PubMed Central PMCID: PMCPMC9378670.

465. Nishimwe A, Ruranga C, Musanabaganwa C, Mugeni R, Semakula M, Nzabanita J, Kabano I, Uwimana A, Utumatwishima JN, Kabakambira JD, Uwineza A, Halvorsen L, Descamps F, Houghtaling J, Burke B, Bahati O, Bizimana C, Jansen S, Twizere C, Nkurikiyeyezu K, Birungi F, Nsanzimana S, Twagirumukiza M. Leveraging artificial intelligence and data science techniques in harmonizing, sharing, accessing and analyzing SARS-COV-2/COVID-19 data in Rwanda (LAISDAR Project): study design and rationale. BMC Med Inform Decis Mak. 2022;22(1):214. Epub 20220812. doi: 10.1186/s12911-022-01965-9. PubMed PMID: 35962355; PubMed Central PMCID: PMCPMC9372951.

466. Swerdel JN, Schuemie M, Murray G, Ryan PB. PheValuator 2.0: Methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation. J Biomed Inform. 2022:104177. Epub 20220819. doi: 10.1016/j.jbi.2022.104177. PubMed PMID: 35995107.

467. Delanerolle G, Williams R, Stipancic A, Byford R, Forbes A, Tsang RSM, Anand SN, Bradley D, Murphy S, Akbari A, Bedston S, Lyons RA, Owen R, Torabi F, Beggs J, Chuter A, Balharry D, Joy M, Sheikh A, Hobbs FDR, de Lusignan S. Methodological Issues in Using a Common Data Model of COVID-19 Vaccine Uptake and Important Adverse Events of Interest: Feasibility Study of Data and Connectivity COVID-19 Vaccines Pharmacovigilance in the United Kingdom. JMIR Form Res. 2022;6(8):e37821. Epub 20220822. doi: 10.2196/37821. PubMed PMID: 35786694; PubMed Central PMCID: PMCPMC9400849.

468. Williams RD, Reps JM, Rijnbeek PR, Ryan PB, Prieto-Alhambra D. 90-Day all-cause mortality can be predicted following a total knee replacement: an international, network study to develop and validate a prediction model. Knee Surg Sports Traumatol Arthrosc. 2022;30(9):3068-75. Epub 20211206. doi: 10.1007/s00167-021-06799-y. PubMed PMID: 34870731.

469. Abeysinghe R, Black A, Kaduk D, Li Y, Reich C, Davydov A, Yao L, Cui L. Towards quality improvement of vaccine concept mappings in the OMOP vocabulary with a semi-automated method. J Biomed Inform. 2022:134:104162. Epub 20220825. doi: 10.1016/j.jbi.2022.104162. PubMed PMID: 36029954.

470. Almeida JR, Barraca JP, Oliveira JL. Preserving Privacy when Querying OMOP CDM Databases. Stud Health Technol Inform. 2022;298:163-4. doi: 10.3233/shti220930. PubMed PMID: 36073478.

471. Williams RD, Reps JM, Rijnbeek PR, Ryan PB, Prieto-Alhambra D. 90-Day all-cause mortality can be predicted following a total knee replacement: an international, network study to develop and validate a prediction model. Knee Surg Sports Traumatol Arthrosc. 2022;30(9):3068-75. Epub 20211206. doi: 10.1007/s00167-021-06799-y. PubMed PMID: 34870731.

472. Xiao G, Pfaff E, Prud'hommeaux E, Booth D, Sharma DK, Huo N, Yu Y, Zong N, Ruddy KJ, Chute CG, Jiang G. FHIR-Ontop-OMOP: Building Clinical Knowledge Graphs in FHIR RDF with the OMOP Common Data Model. J Biomed Inform. 2022:104201. Epub 20220908. doi: 10.1016/j.jbi.2022.104201. PubMed PMID: 36089199.

473. Zhang L, Wang Y, Schuemie MJ, Blei DM, Hripcsak G. Adjusting for indirectly measured confounding using large-scale propensity score. J Biomed Inform. 2022:104204. Epub 20220912. doi: 10.1016/j.jbi.2022.104204. PubMed PMID: 36108816.

474. Castano VG, Spotnitz M, Waldman GJ, Joiner EF, Choi H, Ostropolets A, Natarajan K, McKhann GM, Ottman R, Neugut AI, Hripcsak G, Youngerman BE. Identification of patients with drug resistant epilepsy in electronic medical record data using the Observational Medical Outcomes Partnership Common Data Model. Epilepsia. 2022. Epub 20220914. doi: 10.1111/epi.17409. PubMed PMID: 36106377.

475. Khodaverdi M, Price BS, Porterfield JZ, Bunnell HT, Vest MT, Anzalone AJ, Harper J, Kimble WD, Moradi H, Hendricks B, Santangelo SL, Hodder SL. An ordinal severity scale for COVID-19 retrospective studies using Electronic Health Record data. JAMIA Open. 2022;5(3):ooac066. Epub 20220709. doi: 10.1093/jamiaopen/ooac066. PubMed PMID: 35911666; PubMed Central PMCID: PMCPMC9278199.

>2013	2014	2015	2016	2017	2018	2019	2020	2021	Thru Sept '22
33	14	21	20	29	36	53	63	103	80

OHDSI.org 88 #JoinTheJourney

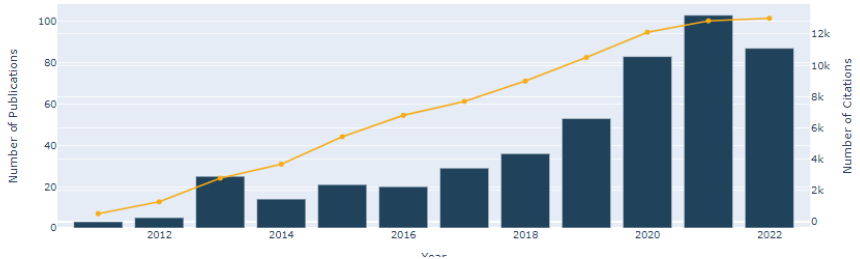
OHDSI Publications

OHDSI Community Dashboard Dashboards

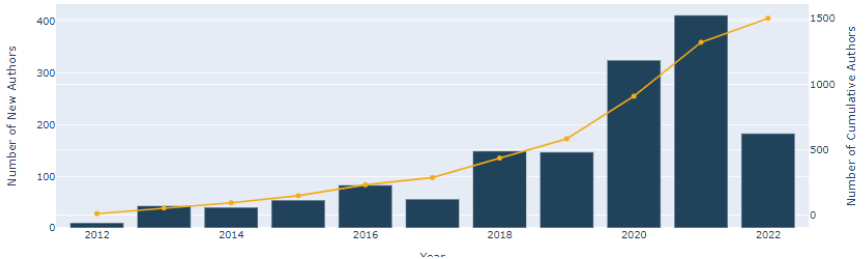
Publication Analysis

PubMed Publication Tracking highlights scholarship generated using the OMOP Common Data Model, OHDSI tools, or the OHDSI network. These publications represent scientific accomplishments across areas of data standards, methodological research, open-source development, and clinical applications. We provide the resource to search and browse the catalogue of OHDSI-related publications by date, author, title, journal, and SNOMED terms. We monitor the impact of our community using summary statistics (number of publications and citations), and the growth and diversity of our community with the number of distinct authors. Searches for new papers are performed daily, and citation counts are updated monthly.

OHDSI Publications & Cumulative Citations



New and Cumulative OHDSI Researchers



Explore our community progress:
http://dash.ohdsi.org

The Secret Sources

Open-Source

Community

Standardization

Reproducibility

Open-Science

Scale

Summary and Segue

- OHDSI – largest and fastest growing community for RWE
- Because of
  - Standardization
  - Open Source
  - Open Science
  - Reproducibility
  - Community
  - Scale
- You should join, too

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Join the Journey!

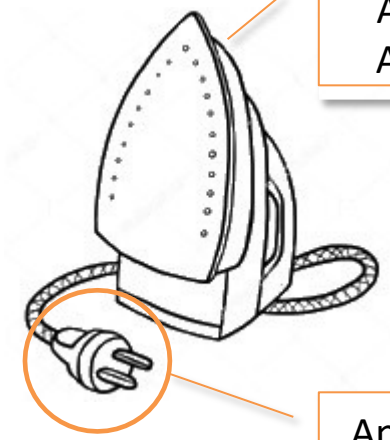
Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

OHDSI Standardization of Evidence Generation

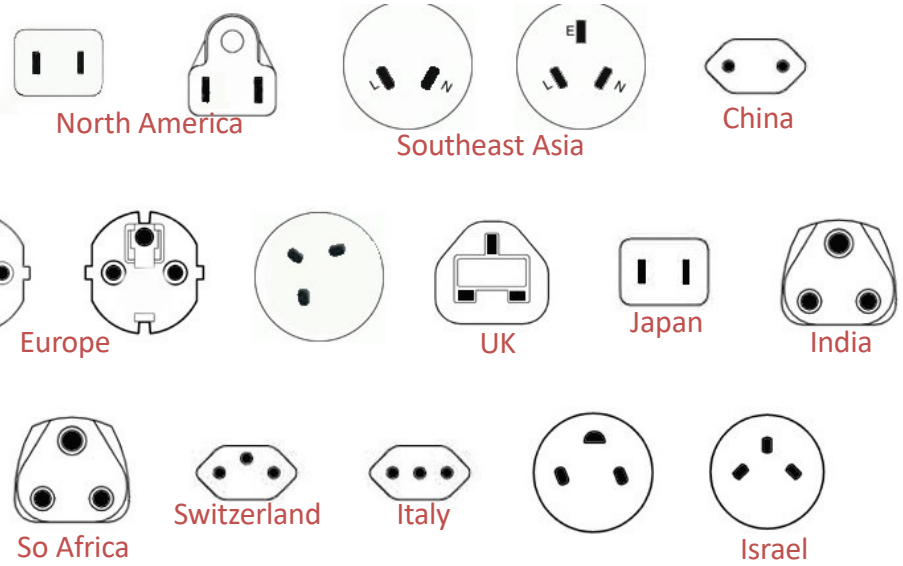
Current Approach: “One Study – One Script”

"What's the adherence to my drug in the data assets I own?"

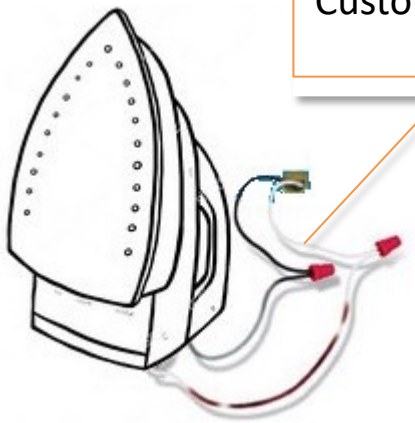


Analytical method: Adherence to Drug

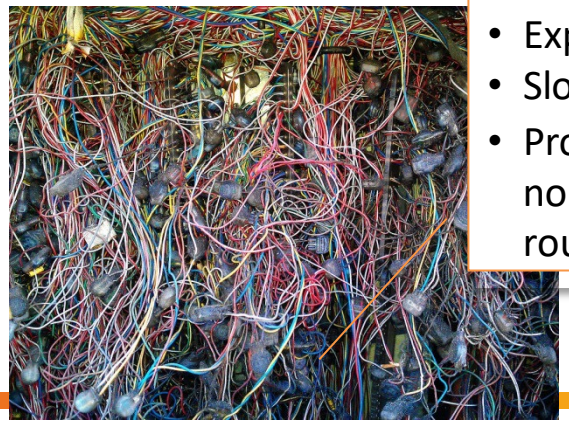
Application to data



Current solution:

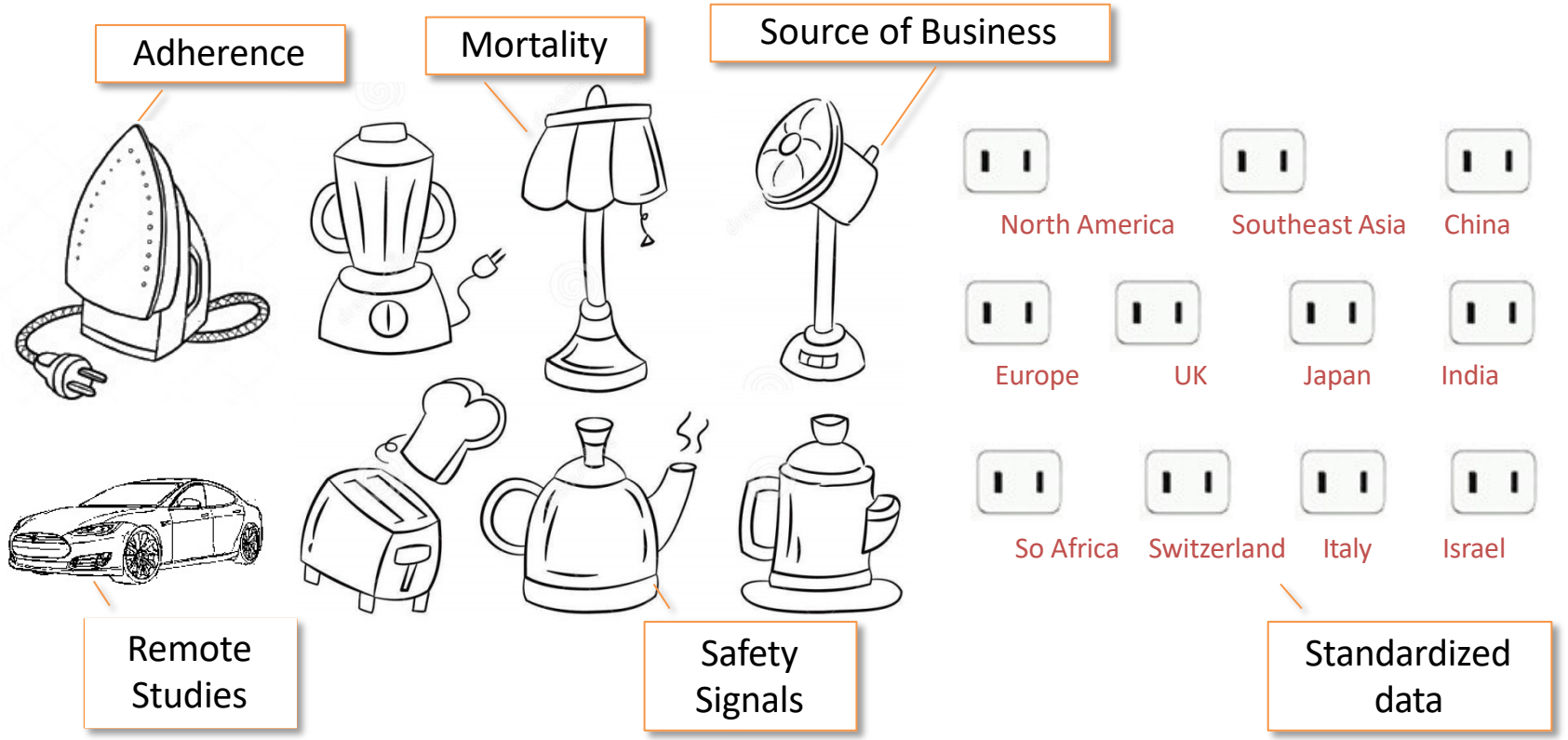


Custom script for each study



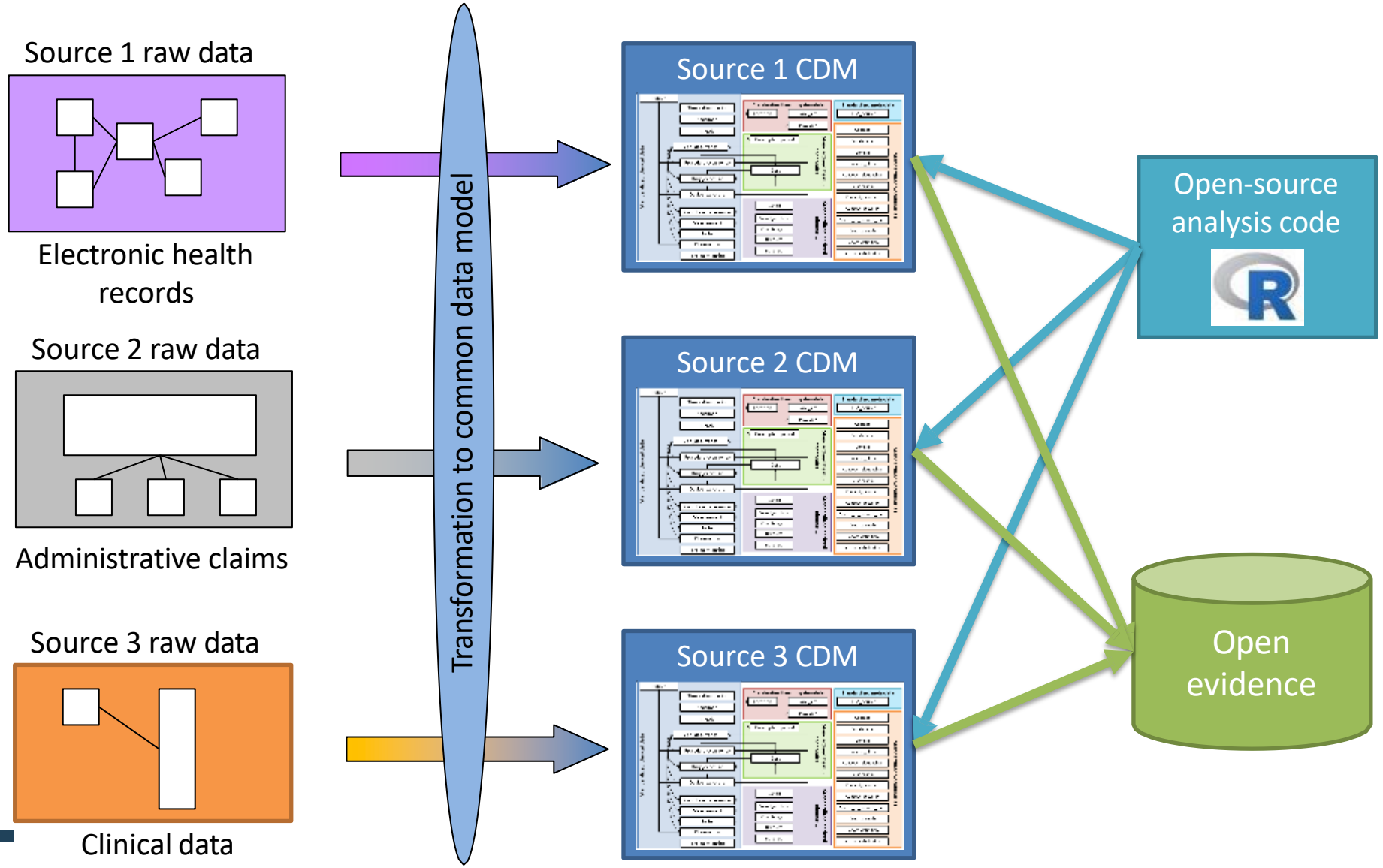
- Not scalable
- Expensive
- Slow
- Prohibitive to non-expert routine use

Solution: Standardized Data and Analytics



1. ATLAS, Remote Studies
   - Standard Cohorts
   - Standardized Analytics
2. OMOP CDM
   - Standardized Format
   - Standardized Coding

Common data model can enable standardized analytics across a distributed data network



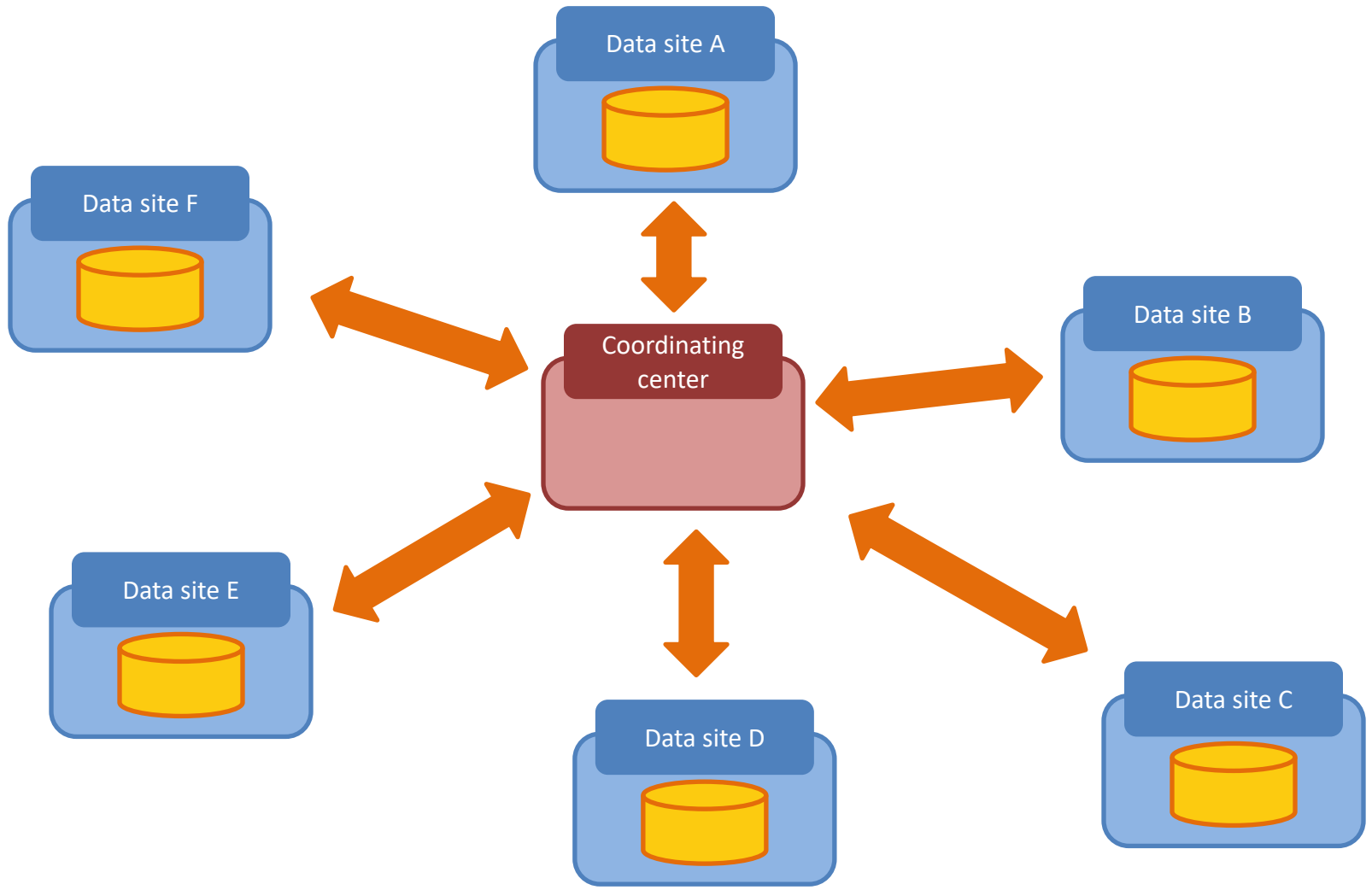
Research Across Distributed Research Networks

Traditional way:

1. Share data
2. Harmonize data
3. Then analyze

OHDSI way:

1. Leave data where it is
2. Harmonize each site's data to OMOP CDM
3. Share aggregated statistics for analysis



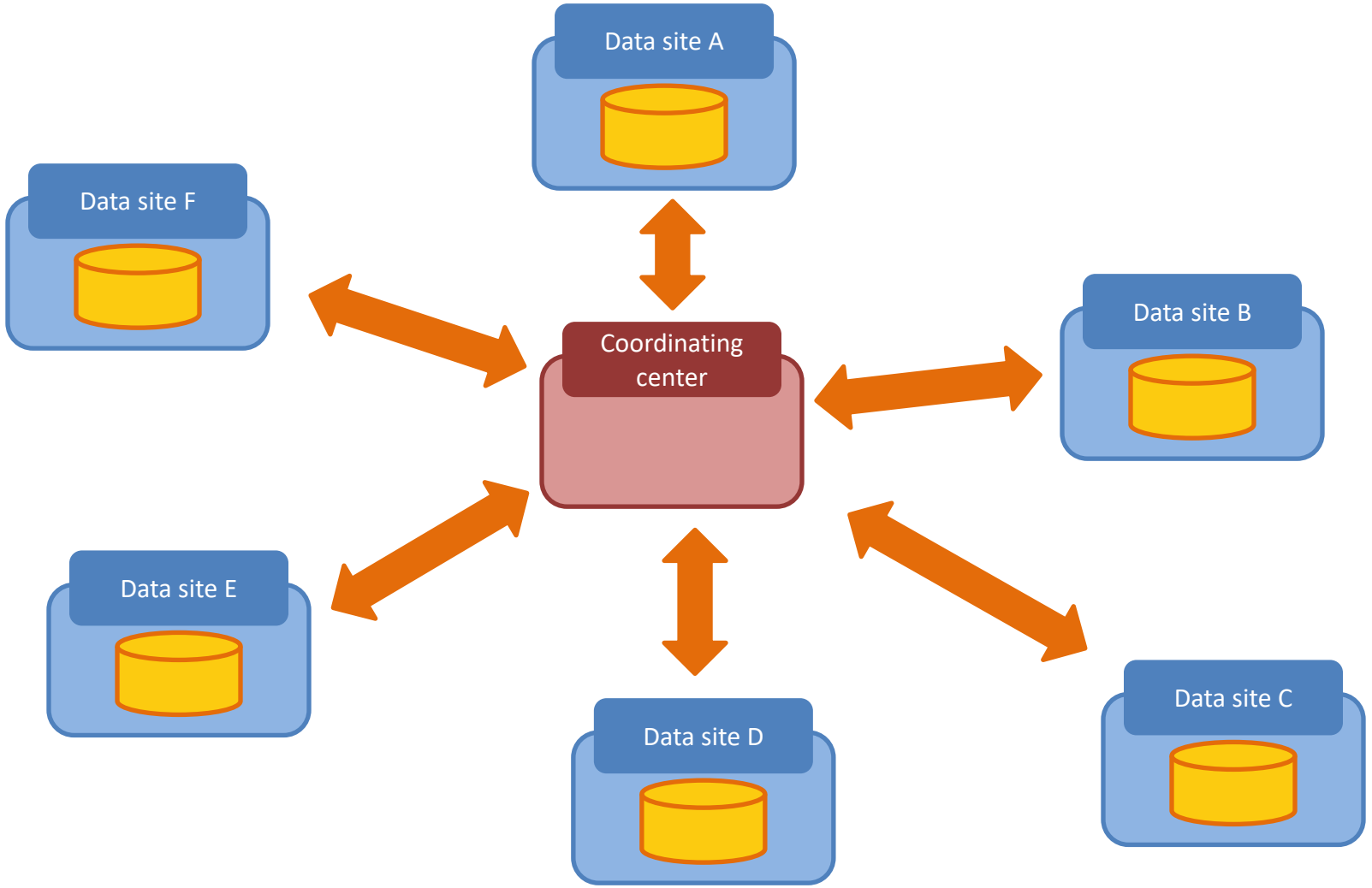
Research Across Distributed Research Networks

Same across sites:

- Common Data Model
- Standardized Vocabularies
- Phenotypes/Cohorts
- Analysis/Methods
- Evidence generation

Different across sites:

- Health care system
- Data capture process
- Source coding systems
- ETL
- Database platform

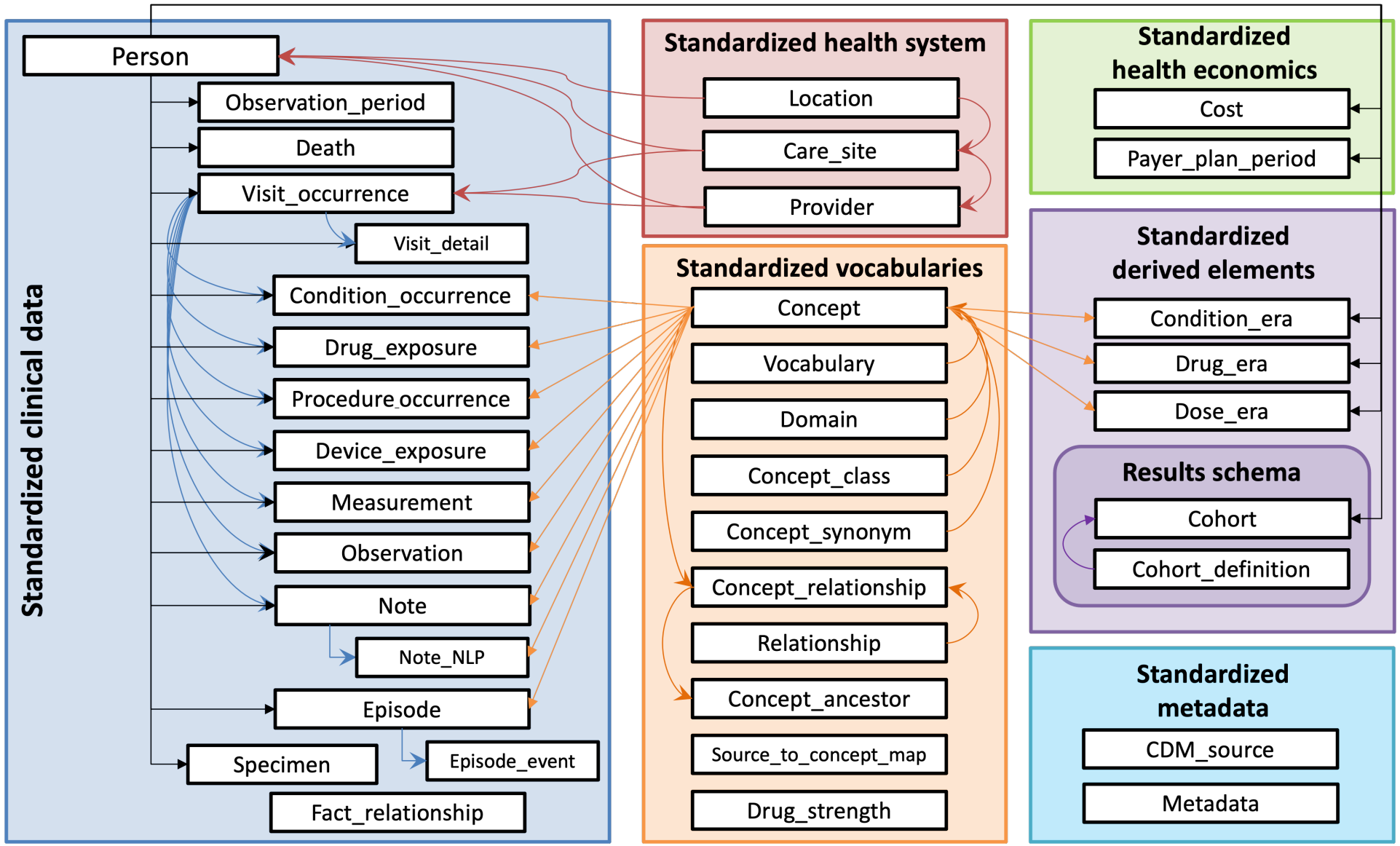


OMOP Common Data Model

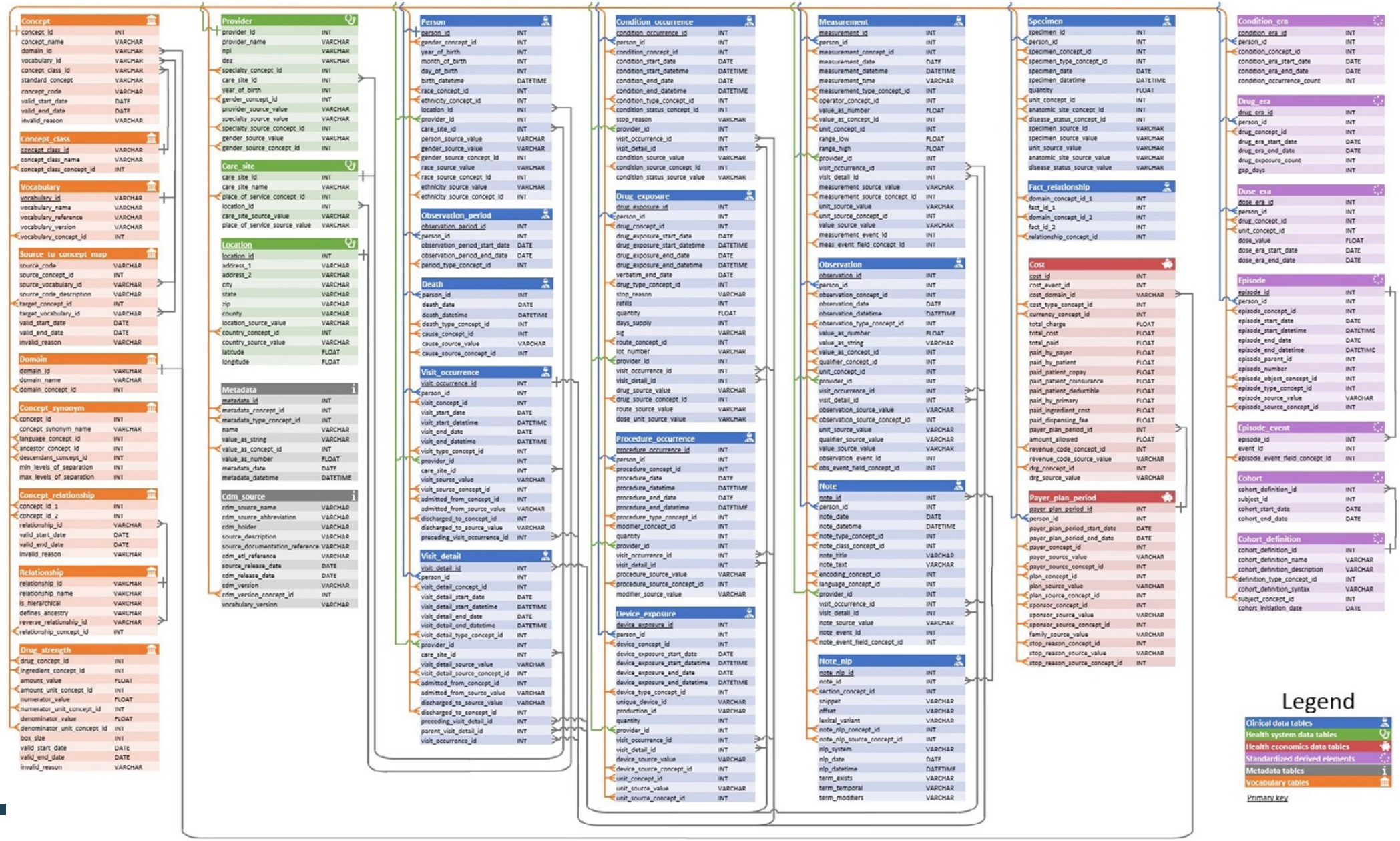
- Components
  - **Schema** – tables where you put data
  - **Vocabulary** – what codes go in the table
  - **Conventions** – how to store data

- Open committee structure to govern it
  - Contracted vocabulary maintenance

OMOP Common Data Model

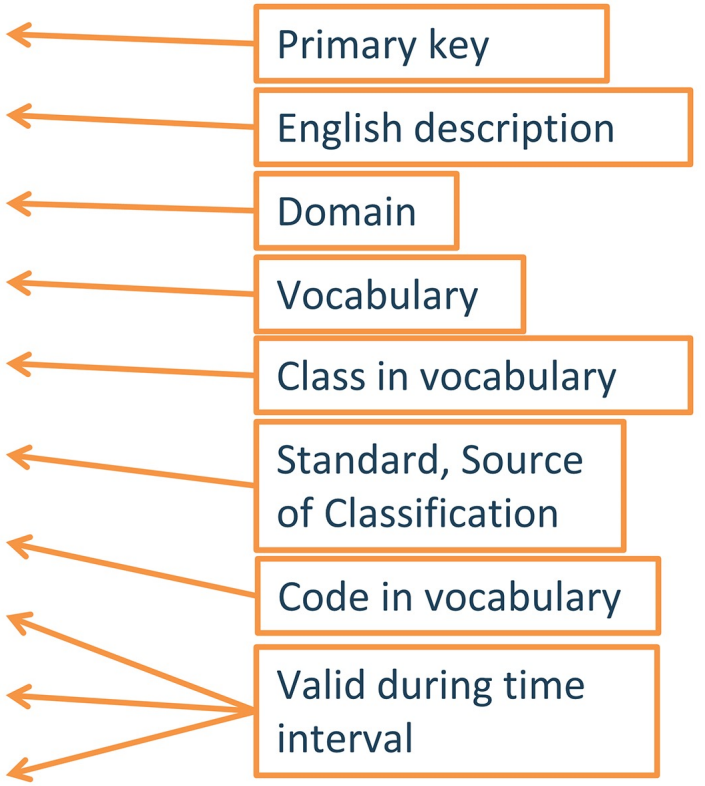


OMOP Common Data Model v5.4



OHDSI's Standardized Vocabularies

CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	



Standard representation of vocabulary concepts in the OMOP CDM. The example provided is the CONCEPT table record for the SNOMED code for Atrial Fibrillation.

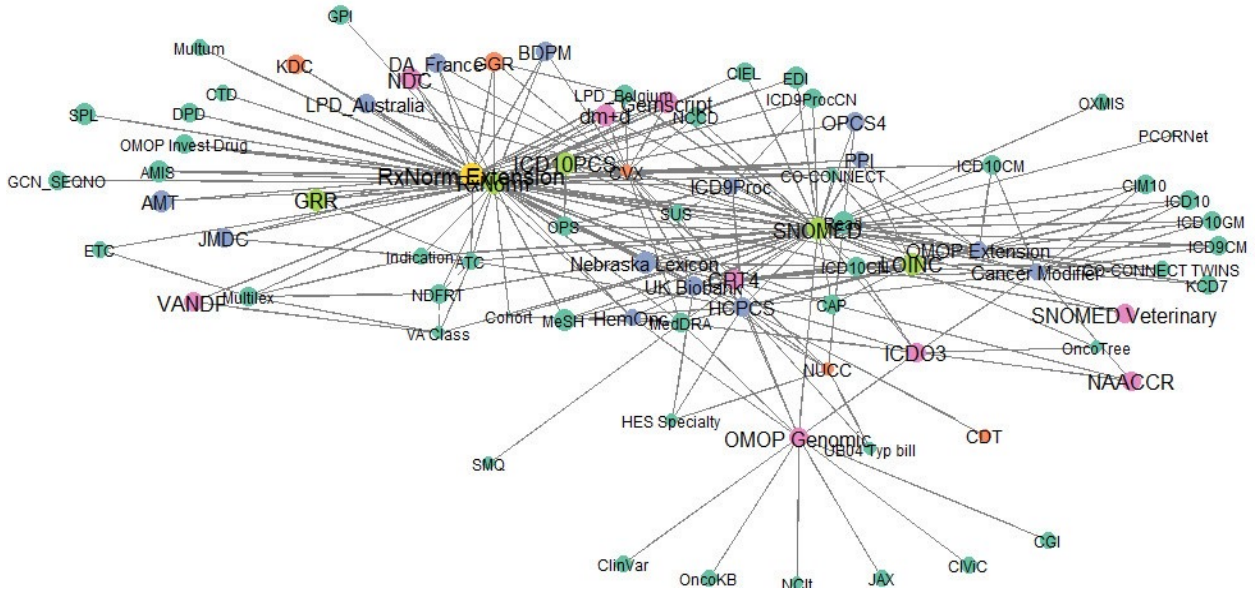
OHDSI's Standardized Vocabularies

- 142 Vocabularies across 44 domains
  - MU3 standards: SNOMED, RxNorm, LOINC
  - Disparate sources: ICD9CM, ICD10(CM), Read, NDC, Gemscript, CPT4, HCPCS…
- >11 million concepts
  - 3.6 million standard concepts
  - 5.1 million source codes
  - 847k classification concepts
- 82 million concept relationships
- 88 million ancestral relationships

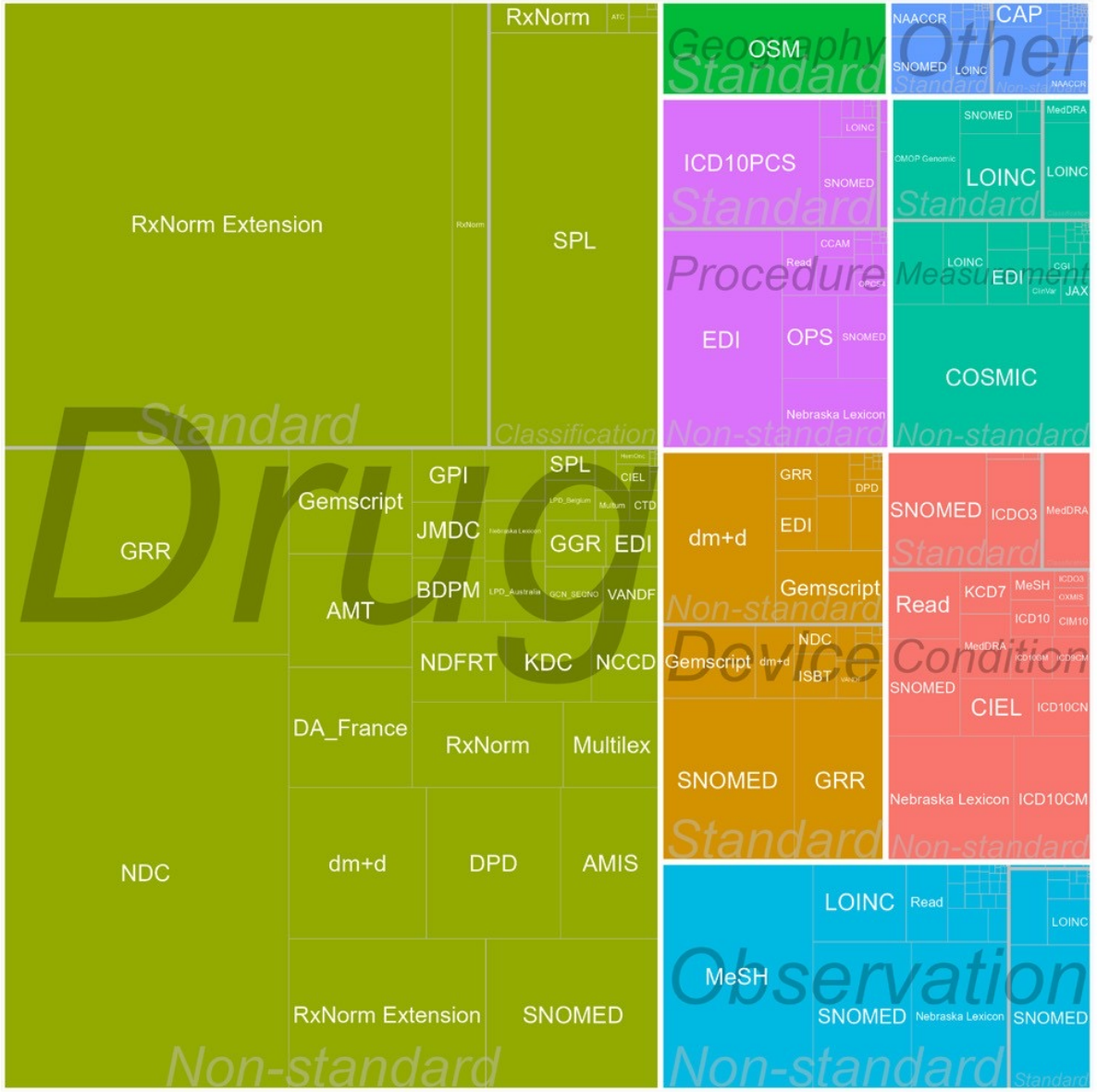
Often referred to as "The Vocabulary"

Publicly available at:
https://athena.ohdsi.org/

OHDSI Standardized Vocabularies



This network diagram shows the relationships between vocabularies. Nodes are vocabularies, sized by the number of concepts. Edges show connections between concepts within vocabularies.



This treemap shows all concepts in the OHDSI vocabularies, organized by domain (color) and vocabularies (boxes sized by the number of concepts).

Standardized Analysis

OHDSI standardized analytics use cases:

- Characterization
- Population-level estimation
- Patient-level prediction

HADES

HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

HADES (formally known as the OHDSI Methods Library) is a set of open-source R packages for large scale analytics, designed specifically for direct interaction with the OMOP CDM.

Standardized Analysis

HADES

HADES is a set of open source R packages for large scale analytics, including population characterization, population-level causal effect estimation, and patient- level prediction.

The packages offer R functions that together can be used to perform an observational study through the full journey from data to evidence, including data manipulation, statistical modeling, and results generation with supporting statistics, tables and figures.

Each package includes functions for specifying and subsequently executing multiple analyses efficiently. HADES supports best practices for use of observational data as learned from previous and ongoing research, such as transparency, reproducibility, as well as measuring of the operating characteristics of methods in a particular context and subsequent empirical calibration of estimates produced by the methods.

Population-Level Estimation

CohortMethod

CohortMethod is an R package for performing new-user cohort studies in an observational database in the OMOP Common Data Model.

SelfControlledCaseSeries

SelfControlledCaseSeries is an R package for performing Self-Controlled Case Series (SCCS) analyses in an observational database in the OMOP Common Data Model.

EvidenceSynthesis

This R package contains routines for combining causal effect estimates and study diagnostics across multiple data sites in a distributed study. This includes functions for performing meta-analysis and forest plots.

SelfControlledCohort

This package provides a method to estimate risk by comparing time exposed with time unexposed among the exposed cohort.

Patient-Level Prediction/Characterization

PatientLevelPrediction

PatientLevelPrediction is an R package for building and validating patient-level predictive models using data in the OMOP Common Data Model format.

EnsemblePatientLevelPrediction

EnsemblePatientLevelPrediction is an R package for building and validating ensemble patient-level predictive models using data in the OMOP Common Data Model format. The package expands the OHDSI R PatientLevelPrediction package to enable ensemble learning.

DeepPatientLevelPrediction

DeepPatientLevelPrediction is an R package for building and validating deep learning patient-level predictive models using data in the OMOP Common Data Model format and OHDSI PatientLevelPrediction framework.

Characterization

Characterization is an R package for performing characterization of a target and a comparator cohort.

Standardized Analysis

Cohort Construction

CAPR

The goal of Capr, pronounced 'kay-pr' like the edible flower, is to provide a language for expressing OHDSI Cohort definitions in R code. OHDSI defines a cohort as "a set of persons who satisfy one or more inclusion criteria for a duration of time" and provides a standardized approach for defining them (Circe-be). Capr exposes the standardized approach to cohort building through a programmatic interface in R which is particularly helpful when creating a large number of similar cohorts. Capr version 2 introduces a new user interface designed for readability with the goal that Capr code being a human readable description of a cohort while also being executable on an OMOP Common Data Model.

CirceR

A R-wrapper for Circe, a library for creating queries for the OMOP Common Data Model. These queries are used in cohort definitions (CohortExpression) as well as custom features (CriteriaFeature). This package provides convenient wrappers for Circe functions, and includes the necessary Java dependencies.

CohortDiagnostics

CohortDiagnostics is an R utility package for the development and evaluation of phenotype algorithms for OMOP CDM compliant data sets. This package provides a standard, end to end, set of analytics for understanding patient capture including data generation and result exploration through an R Shiny interface. Analytics computed include cohort characteristics, record counts, index event misclassification, captured observation windows and basic incidence proportions for age, gender and calendar year. Through the identification of errors, CohortDiagnostics enables the comparison of multiple candidate cohort definitions across one or more data sources, facilitating reproducible research.

CohortExplorer

This software tool is designed to extract data from a randomized subset of individuals within a cohort and make it available for exploration in a 'Shiny' application environment. It retrieves date-stamped, event-level records from one or more data sources that represent patient data in the Observational Medical Outcomes Partnership (OMOP) data model format. This tool features a user-friendly interface that enables users to efficiently explore the extracted profiles, thereby facilitating applications, such as reviewing structured profiles. The output of this R-package is a self-contained R shiny that contains person-level data for review.

CohortGenerator

This R package contains functions for generating cohorts using data in the CDM.

PheValuator

The goal of PheValuator is to produce a large cohort of subjects each with a predicted probability for a specified health outcome of interest (HOI). This is achieved by developing a diagnostic predictive model for the HOI using the PatientLevelPrediction (PLP) R package and applying the model to a large, randomly selected population. These subjects can be used to test one or more phenotype algorithms.

PhenotypeLibrary

The OHDSI community has developed a publicly accessible, version-controlled Phenotype Library to guide real-world evidence towards the FAIR principles: Findability, Accessibility, Reproducibility, and Interoperability. This library aims to foster the submission and retrieval of high-quality cohort definitions, cataloging of metadata, attribution and promotion of discovery and reuse in scientific research. Within the OHDSI Phenotype Library (OHDSI PL), each entry represents a unique cohort definition identifiable by a stable, externally referenceable ID. Comprehensive metadata about each cohort definition is cataloged and made searchable for researchers.Content in the library is subject to version control, with each version is assigned a specific DOI.

Evidence Quality

Achilles

Automated Characterization of Health Information at Large-Scale Longitudinal Evidence Systems (ACHILLES) Achilles provides descriptive statistics on an OMOP CDM database. ACHILLES currently supports CDM version 5.3 and 5.4.

Data Quality Dashboard

DataQualityDashboard (DQD) is an R package for exposing and evaluating observational data quality. This package runs a series of data quality checks against an OMOP CDM instance. It systematically runs the checks, evaluates each check against a pre-specified threshold, and then communicates what was done in a transparent and easily understandable way.

Evidence Quality

EmpiricalCalibration

This R package contains routines for performing empirical calibration of observational study estimates. By using a set of negative control hypotheses we can estimate the empirical null distribution of a particular observational study setup. This empirical null distribution can be used to compute a calibrated p-value, which reflects the probability of observing an estimated effect size when the null hypothesis is true taking both random and systematic error into account, as described in the paper *Interpreting observational studies*: why empirical calibration is needed to correct p-values.
Also supported is empirical calibration of confidence intervals, based on the results for a set of negative and positive controls, as described in the paper *Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data*.

Method Evaluation

This R package contains resources for the evaluation of the performance of methods that aim to estimate the magnitude (relative risk) of the effect of a drug on an outcome. These resources include reference sets for evaluating methods on real data, as well as functions for inserting simulated effects in real data based on negative control drug-outcome pairs. Further included are functions for the computation of the minimum detectable relative risks and functions for computing performance statistics such as predictive accuracy, error and bias.

Supporting Packages

Andromeda

AsynchroNous Disk-based Representation of MassivE DAta (ANDROMEDA): An R package for storing large data objects. Andromeda allow storing data objects on a local drive, while still making it possible to manipulate the data in an efficient manner.

BigKNN

An R package implementing a large scale k-nearest neighbor (KNN) classifier using the Lucene search engine.

BrokenAdaptiveRidge

BrokenAdaptiveRidge is an R package for performing L_0-based regressions using Cyclops.

Cyclops

Cyclops (Cyclic coordinate descent for logistic, Poisson and survival analysis) is an R package for performing large scale regularized regressions.

DatabaseConnector

This R package provides function for connecting to various DBMSs. Together with the SqlRender package, the main goal of DatabaseConnector is to provide a uniform interface across database platforms: the same code should run and produce equivalent results, regardless of the database back end.

Eunomia

Eunomia is a standard dataset in the OMOP (Observational Medical Outcomes Partnership) Common Data Model (CDM) for testing and demonstration purposes. Eunomia is used for many of the exercises in the Book of OHDSI. For functions that require schema name, use 'main'.

FeatureExtraction

An R package for generating features (covariates) for a cohort using data in the Common Data Model.

Hydra

An R package and Java library for hydrating package skeletons into executable R study packages based on specifications in JSON format.

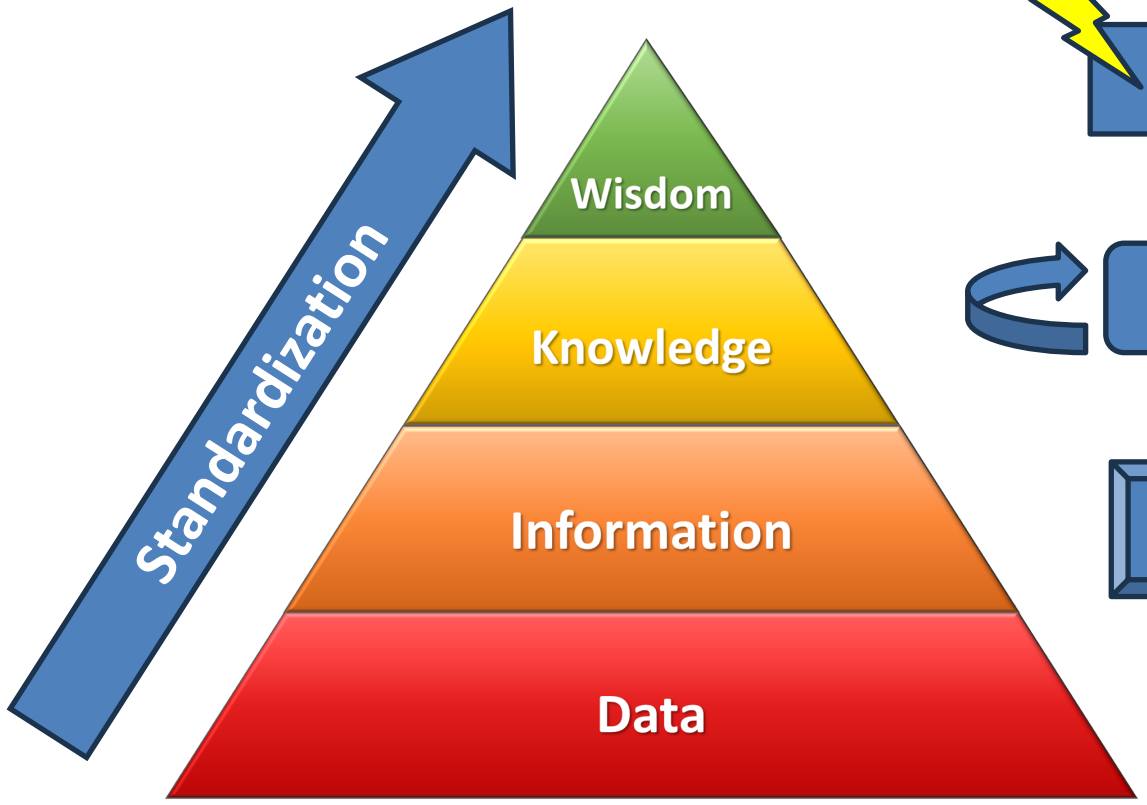
IterativeHardThresholding

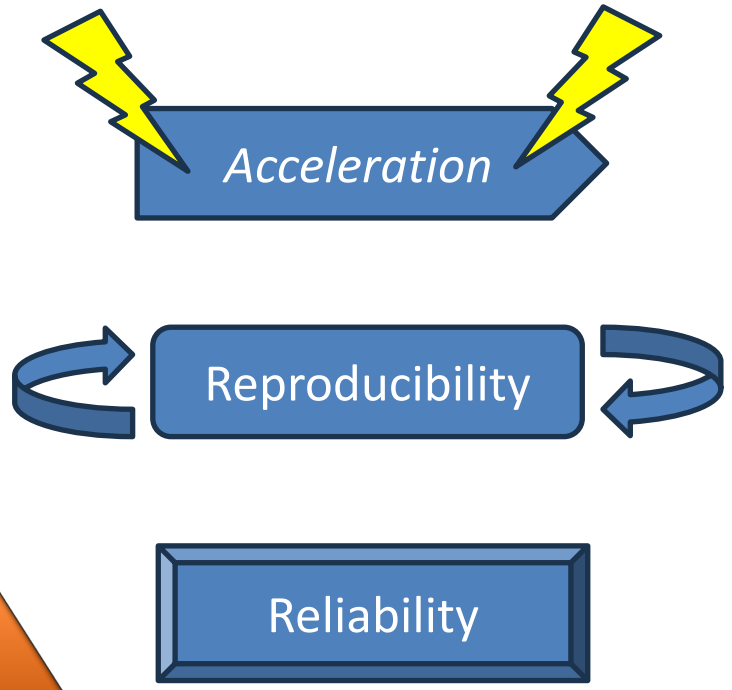
IterativeHardThresholding is an R package for performing L_0-based regressions using Cyclops.

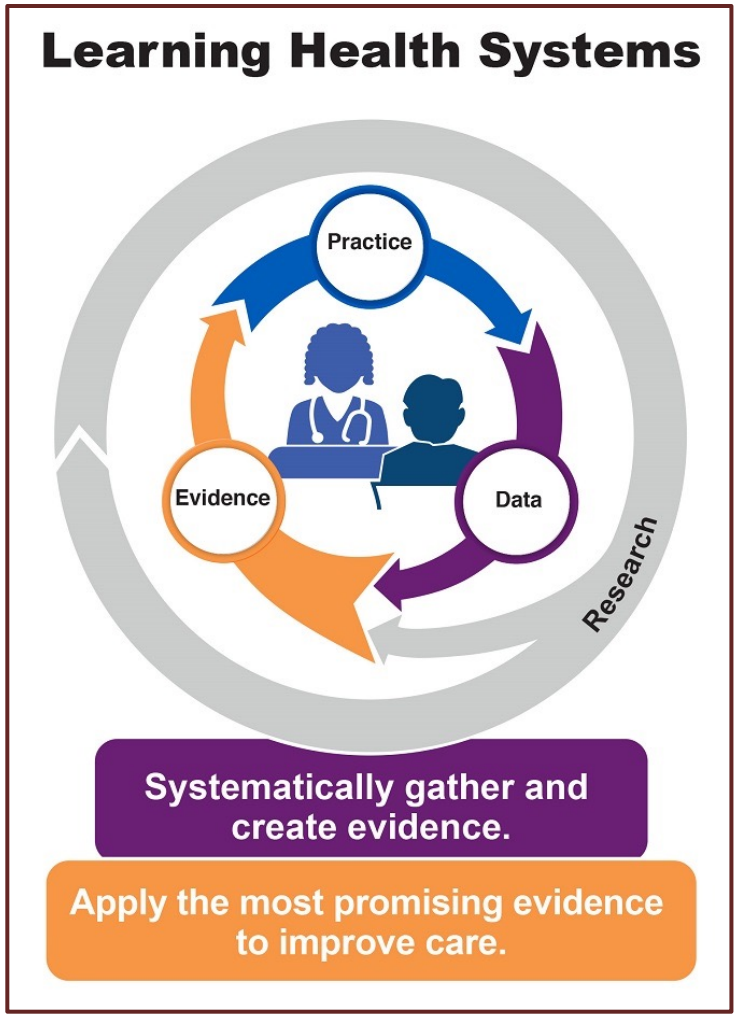
OhdsiSharing

This is an R package for sharing data between OHDSI partners.

Standardization of evidence generation → acceleration, reproducibility, reliability







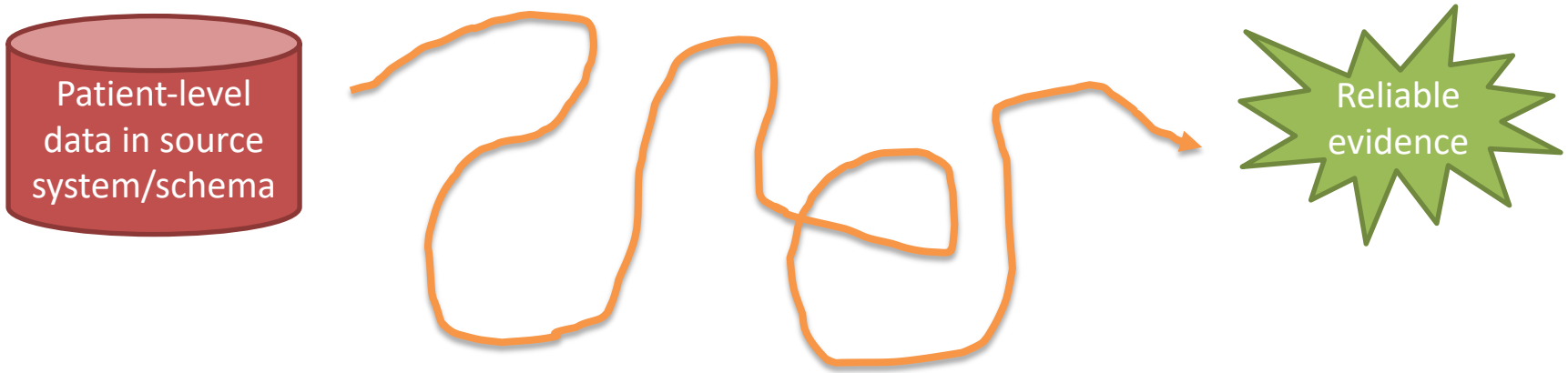
Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

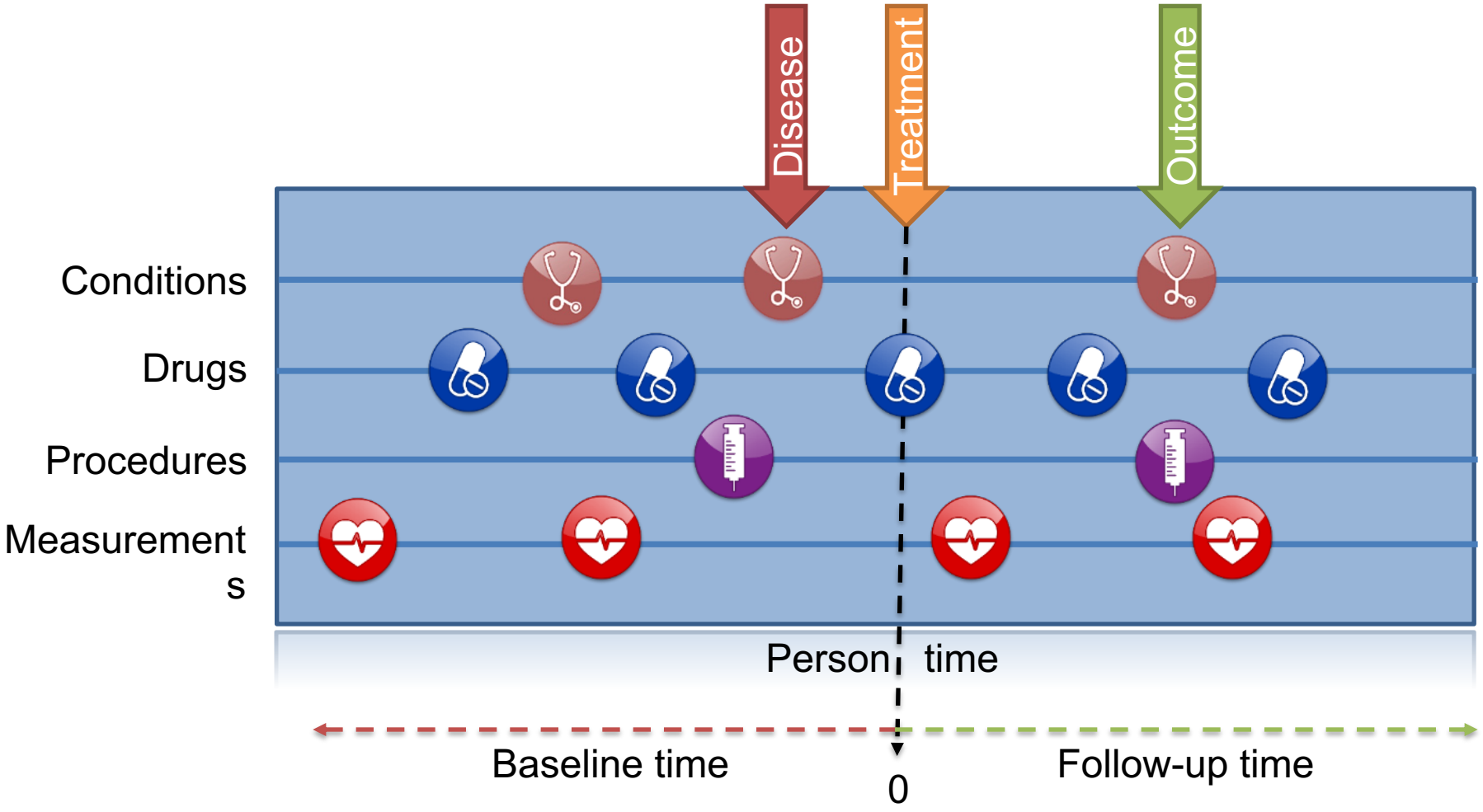
Evidence Generation

Large Scale Observational Research Preparation

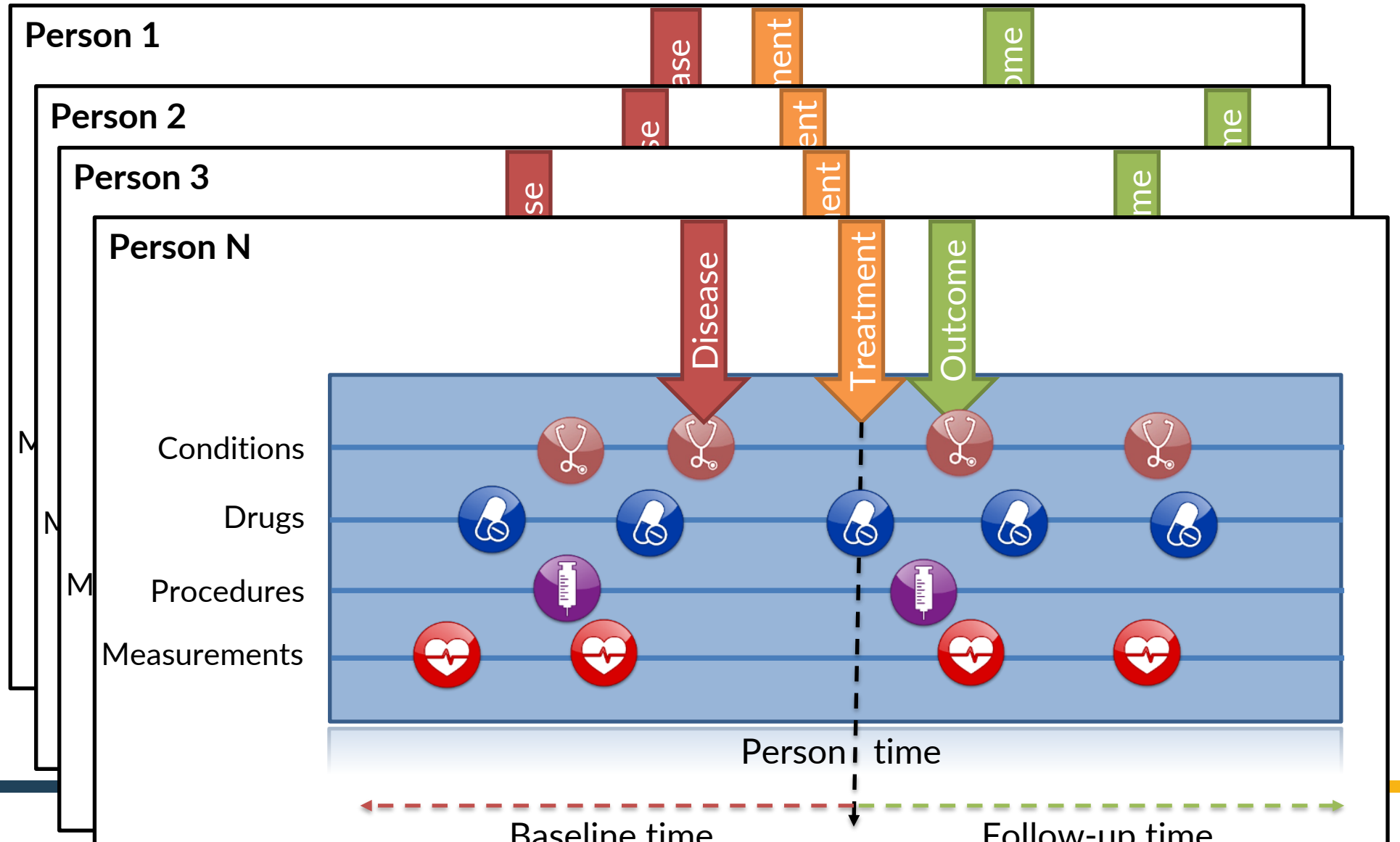
The journey to real-world evidence



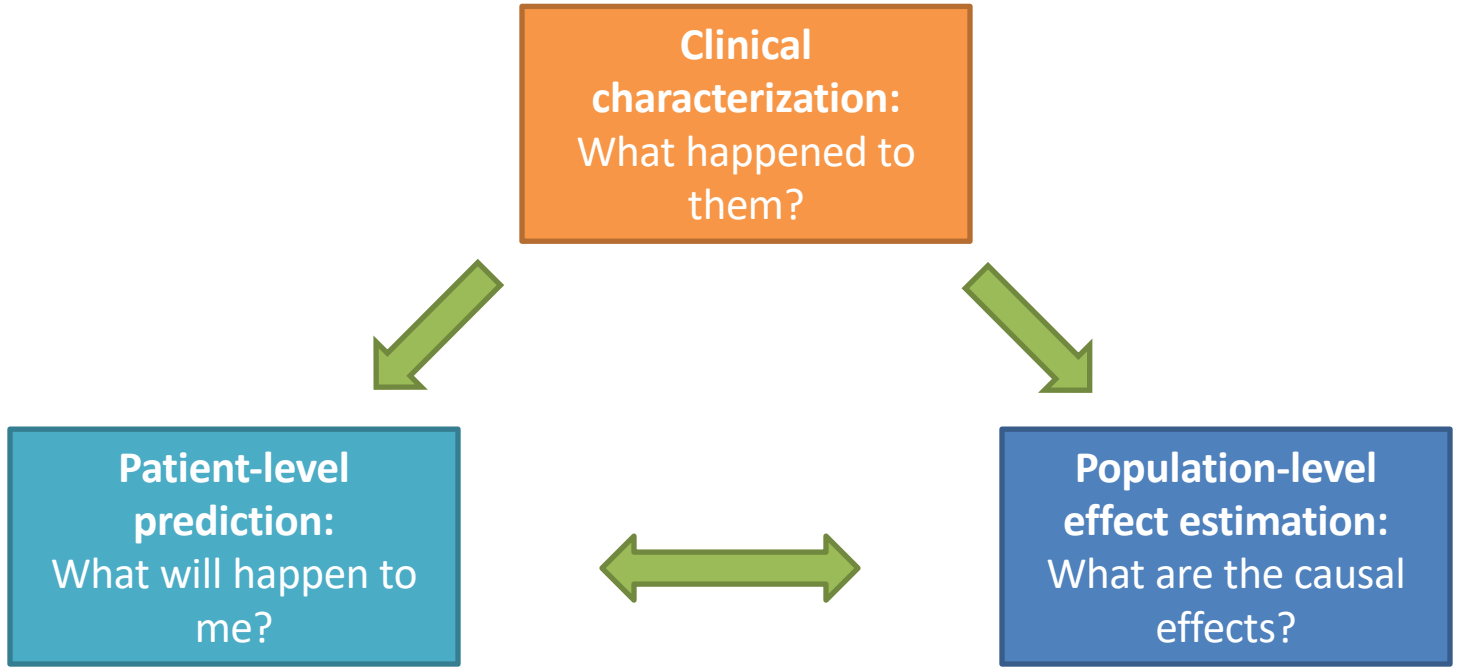
A Caricature of The Patient Journey



Each Observational Database Is Just an (Incomplete) Compilation of Patient Journeys



Complementary evidence to inform the patient journey



Analytic use case	Type	Structure	Example
Clinical characterization	**Disease Natural History**	Amongst patients who are diagnosed with **<insert your favorite disease>**, what are the patient's characteristics from their medical history?	Amongst patients with **rheumatoid arthritis**, what are their demographics (age, gender), prior conditions, medications, and health service utilization behaviors?
	Treatment utilization	Amongst patients who have **<insert your favorite disease>**, which treatments were patients exposed to amongst **<list of treatments for disease>** and in which sequence?	Amongst patients with **depression**, which treatments were patients exposed to **SSRI, SNRI, TCA, bupropion, esketamine** and in which sequence?
	Outcome incidence	Amongst patients who are new users of **<insert your favorite drug>,** how many patients experienced **<insert your favorite known adverse event from the drug profile>** within **<time horizon following exposure start>**?	Amongst patients who are new users of **methylphenidate,** how many patients experienced **psychosis** within **1 year of initiating treatment**?
Patient level prediction	**Disease onset and progression**	For a given patient who is diagnosed with **<insert your favorite disease>**, what is the probability that they will go on to have **<another disease or related complication>** within **<time horizon from diagnosis>**?	For a given patient who is **newly diagnosed with atrial fibrillation**, what is the probability that they will go onto to have **ischemic stroke** in **next 3 years**?
	Treatment response	For a given patient who is a new user of **<insert your favorite chronically-used drug>**, what is the probability that they will **<insert desired effect>** in **<time window>**?	For a given patient **with T2DM who start on metformin** , what is the probability that they will **maintain HbA1C<6.5%** after **3 years**?
	Treatment safety	For a given patient who is a new user of **<insert your favorite drug>,** what is the probability that they will experience **<insert adverse event >** within **<time horizon following exposure>**?	For a given patients who is a **new user of warfarin**, what is the probability that they will have **GI bleed** in **1 year**?
Population-level effect estimation	**Safety surveillance**	Does exposure to **<insert your favorite drug>** increase the risk of experiencing **<insert an adverse event>** within **<time horizon following exposure start>**?	Does exposure to **ACE inhibitor** increase the risk of experiencing **Angioedema** within **1 month after exposure start**?
	Comparative effectiveness	Does exposure to **<insert your favorite drug>** have a different risk of experiencing **<insert any outcome (safety or benefit) >** within **<time horizon following exposure start>**, relative to **<insert your comparator treatment>**?	Does exposure to **ACE inhibitor** have a different risk of experiencing **acute myocardial infarction** while **on treatment**, relative to **thiazide diuretic**?

Let’s Dissect One Research Question

- How does the rate of side effects / local problems (including secondary / palliative treatments needed) compare between treatments (open, laparoscopic, robot surgery, with or without lymph node dissection; brachytherapy, different forms of external beam radiation therapy), and which patient specific factors are associated with these adverse secondary endpoints?

Let's Dissect One Research Question

- How does the rate of side effects / local problems (including secondary / palliative treatments needed) compare between treatments (open, laparoscopic, robot surgery, with or without lymph node dissection; brachytherapy, different forms of external beam radiation therapy), and which patient specific factors are associated with these adverse secondary endpoints?

- Characterization study: incidence rate

Amongst patients with **prostate cancer receiving different treatments**, how many patients experienced **side effects/local problems** within <time horizon >?

Let's Dissect One Research Question

- How does the rate of side effects / local problems (including secondary / palliative treatments needed) compare between treatments (open, laparoscopic, robot surgery, with or without lymph node dissection; brachytherapy, different forms of external beam radiation therapy), and which patient specific factors are associated with these adverse secondary endpoints?

- Population level estimation: comparative effectiveness

Comparative effectiveness: Does exposure to **treatment A** have a different risk of experiencing **side effects/local problems** within <time horizon > , relative to **treatment B**?

Let’s Dissect One Research Question

- How does the rate of side effects / local problems (including secondary / palliative treatments needed) compare between treatments (open, laparoscopic, robot surgery, with or without lymph node dissection; brachytherapy, different forms of external beam radiation therapy), and which patient specific factors are associated with these adverse secondary endpoints?

- Characterization study: natural history

Amongst patients with **prostate cancer receiving different treatment A-Z**, what are the patient’s characteristics from their medical history?

Let’s dissect one Research Question

- RQ5. Which specific patient groups benefit most of upfront chemotherapy? What are the side effects and What is impact on quality of life in real-life practice of chemotherapy in this setting? the benefit of potentially toxic upfront chemotherapy appears to be highly individual. Other factors to predict who would benefit most are needed. the benefit of chemotherapy in the subgroup patients who have recurrence after primary treatment is not known.

- Study
- **Target Cohorts:**
- **Comparator Cohorts:**
- **Outcome Cohorts:**

Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

A collaborative open-science community transforming clinical research with real world evidence.

Paul Nagy, PhD, FSIIM
Johns Hopkins University

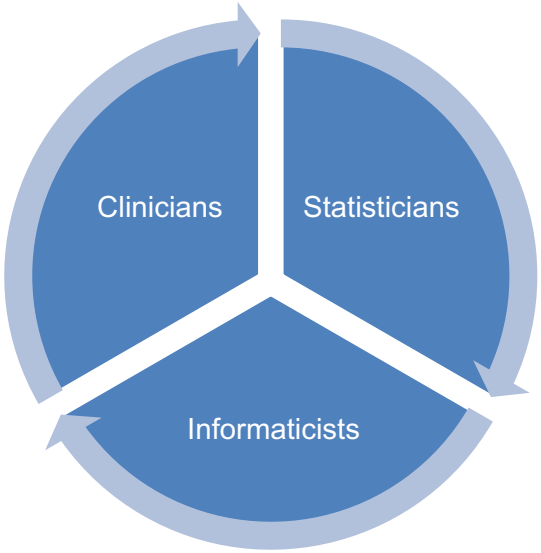
OHDSI

"You come for the data model, but you stay for the community"

Multi-disciplinary Innovation with Open Science at Scale

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

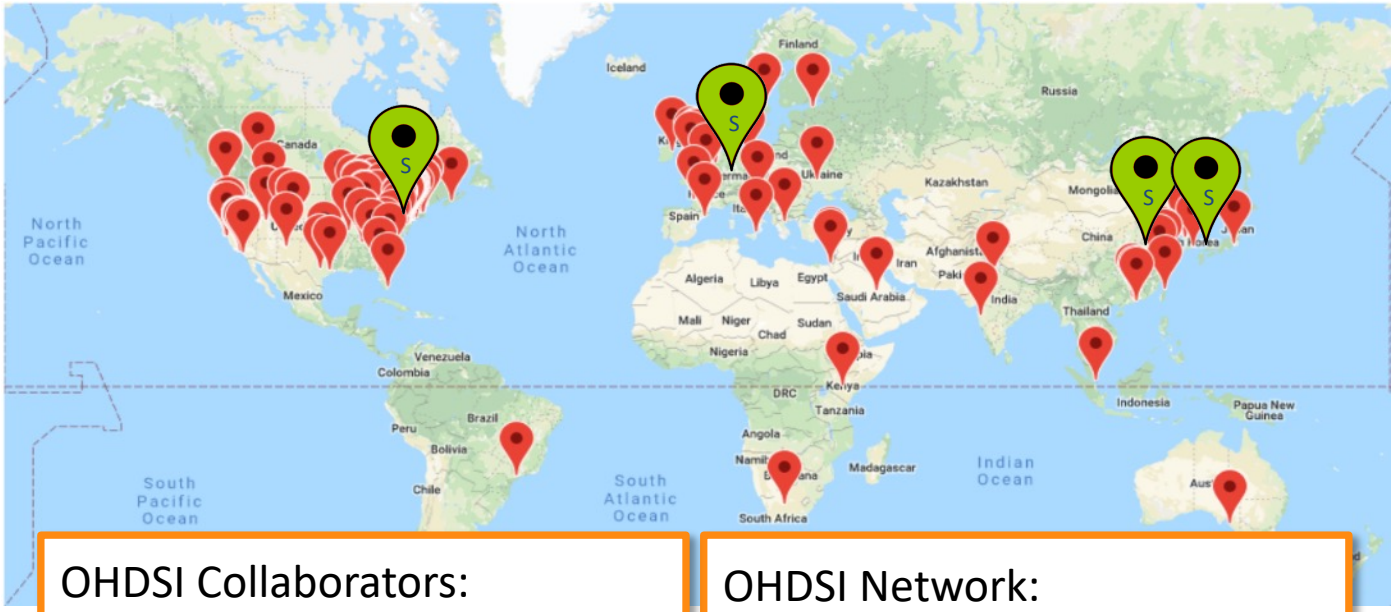
Translate EMR and claims data into a common data model tied to standard terminologies.



Validated statistical tools on observational health data

Open source community and research network with 900+ Million unique patients

OHDSI Is a Highly Active Global Community



OHDSI Collaborators:

- **3,997** users
- **30+** workgroups
- **5** regional chapters
- **25,117** posts with **4,705,098** page views

OHDSI Network:

- **>450** databases
- **34** countries
- **3.6B** patient records (**960M** unique)
  - **>250m** in US, **>290m** in Europe, **>169m** in South America, **>65m** in Asia

OHDSI is a vibrant multi-speciality open science community

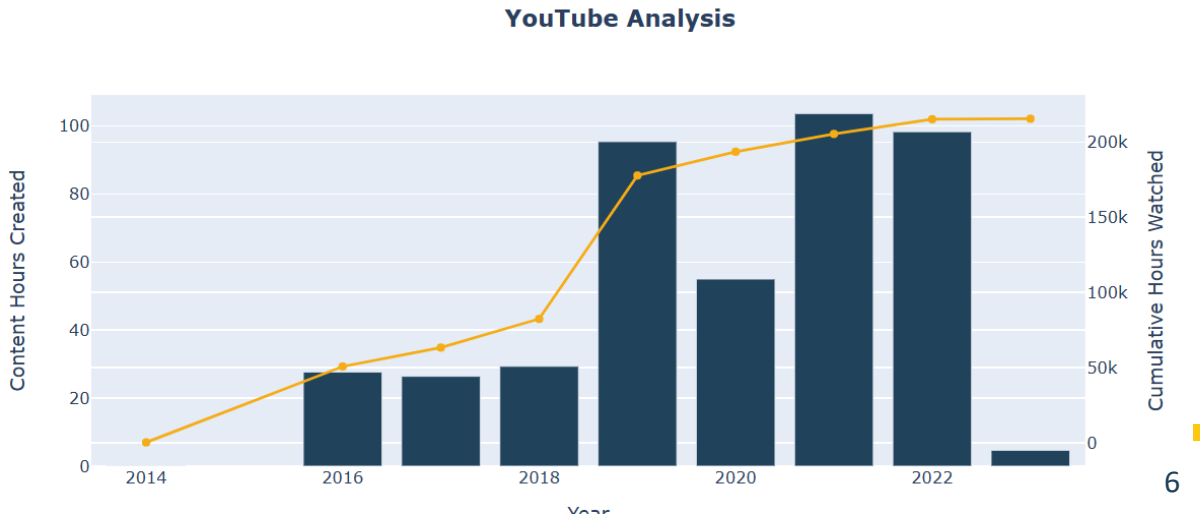
- Innovation
- Reproducibility
- Community
- Collaboration
- Openness
- Beneficence

Our Mission

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

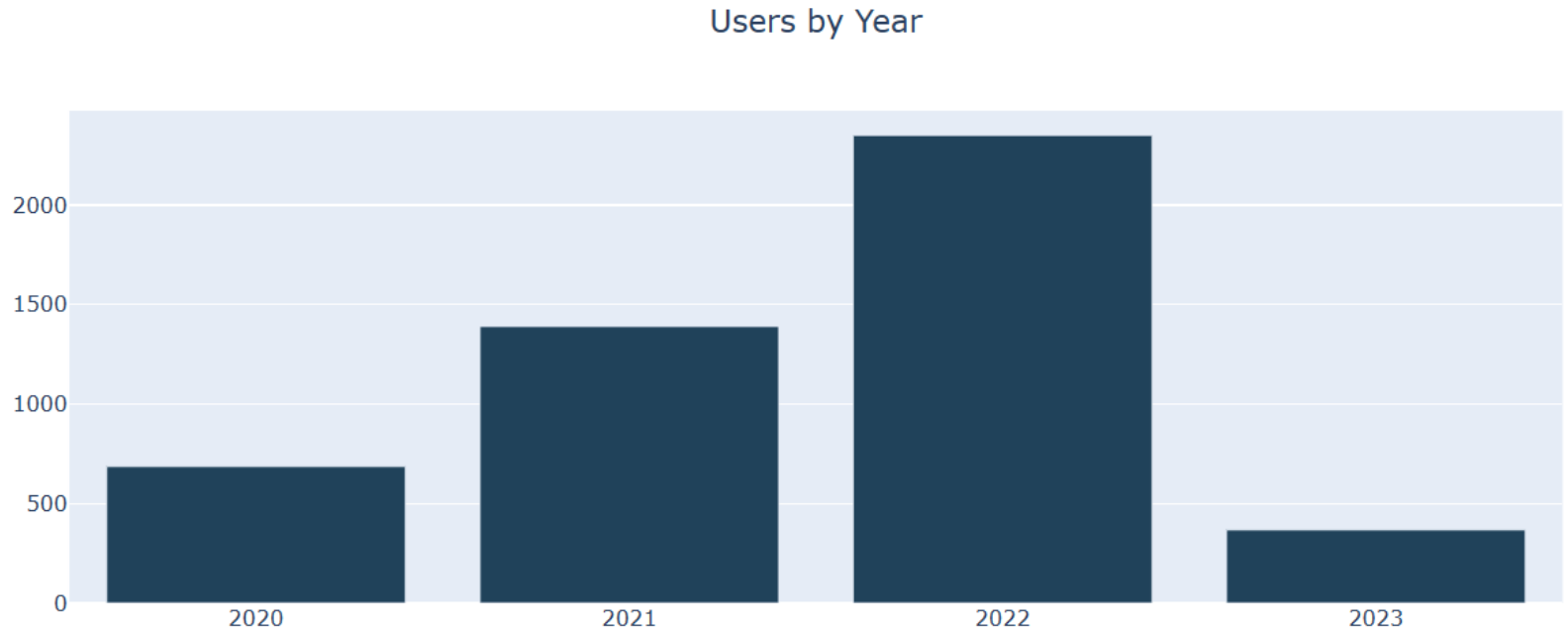
Education

- Weekly Community Calls
- Phenotype Phebruary
- Save our Sysphus Challenge
- Open source developers conference
- Ehden Acdemy online learning mgmt. system
- Book of OHDSI



https://ohdsi.github.io/TheBookOfOhdsi/

Ehden Academy



- Free online LMS based on Moodle
- 20+ self paced courses on OHDSI
- https://academy.ehden.eu/

OHDSI Working Groups

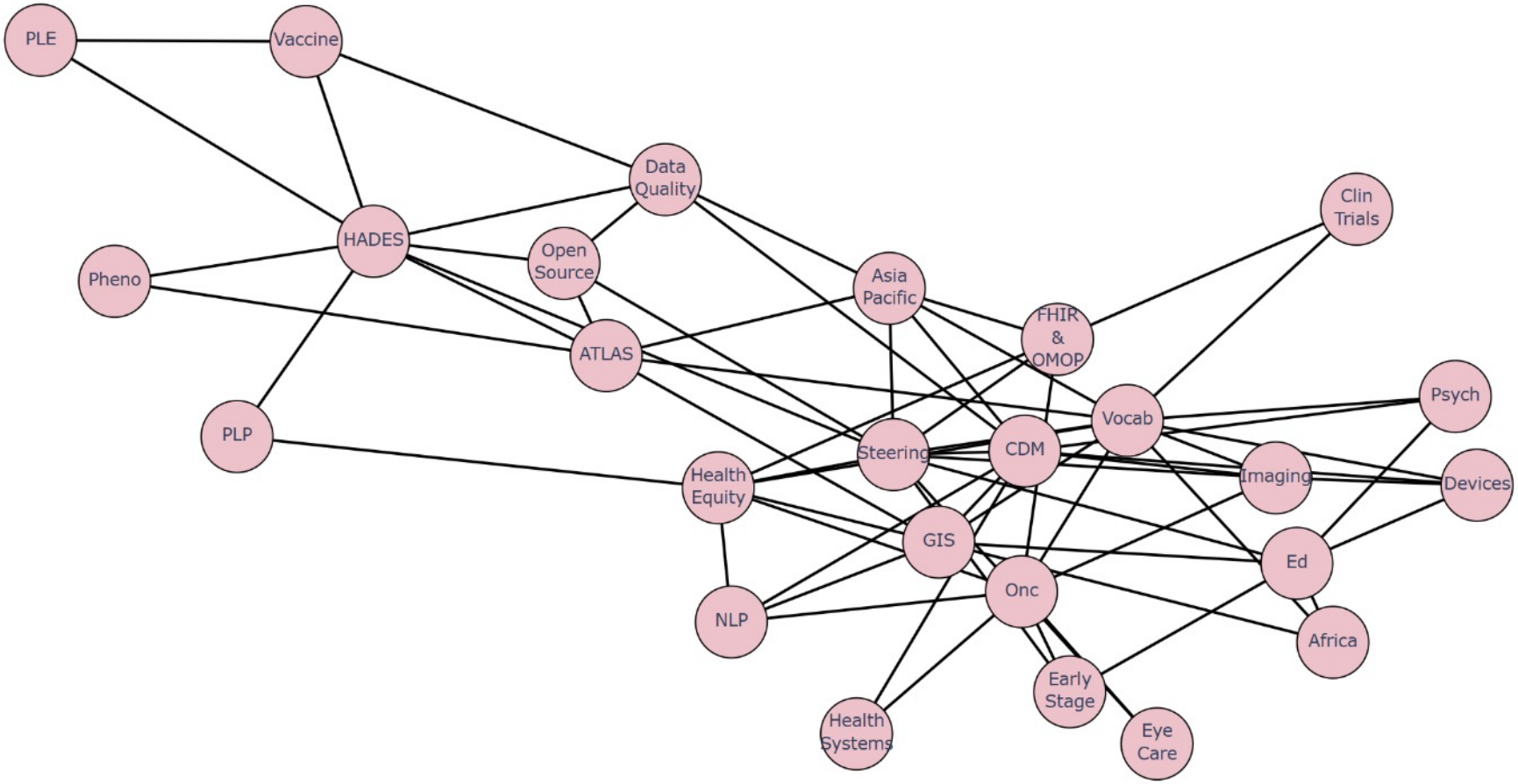
		Workgroup
Core	Data model	Vocabulary
		Common Data Model
	Methods	Patient-Level Prediction
		Population-Level Estimation
	Tools	HADES
		Data Quality Dashboard
		Phenotype Development & Evaluation
		ATLAS/WebAPI
Domain	Data source	Clinical Trials
		FHIR & OMOP
		Geographic Information Systems
		Medical Devices
		Medical Imaging
		Natural Language Processing
		Registry
		Vaccine Vocabulary
	Exposure	Health Equity
		Surgery and Perioperative Medicine

Domain	Disease	Oncology
		Psychiatry
		Eye Care
		Dentistry
Support	Regional	Africa Chapter
		Asia-Pacific Chapter
		Latin America Chapter
	Community segment	Early-Stage Researchers
		Open-Source Community
		Technical Advisory board
		Perseus Uses Group
		Databricks Users Group
		Healthcare Systems
	Broad	Education
		Steering Group

Working Groups

- Any one can become an OHDSI member at no cost. OHDSI is an open inclusive community.
- Any OHDSI member is welcome to join any working group.
- Most working groups are 2x/Month, some are weekly.
- Working group meetings are often recorded and if of educational nature will be uploaded to YouTube.

Working group collaboration



In Person Conferences

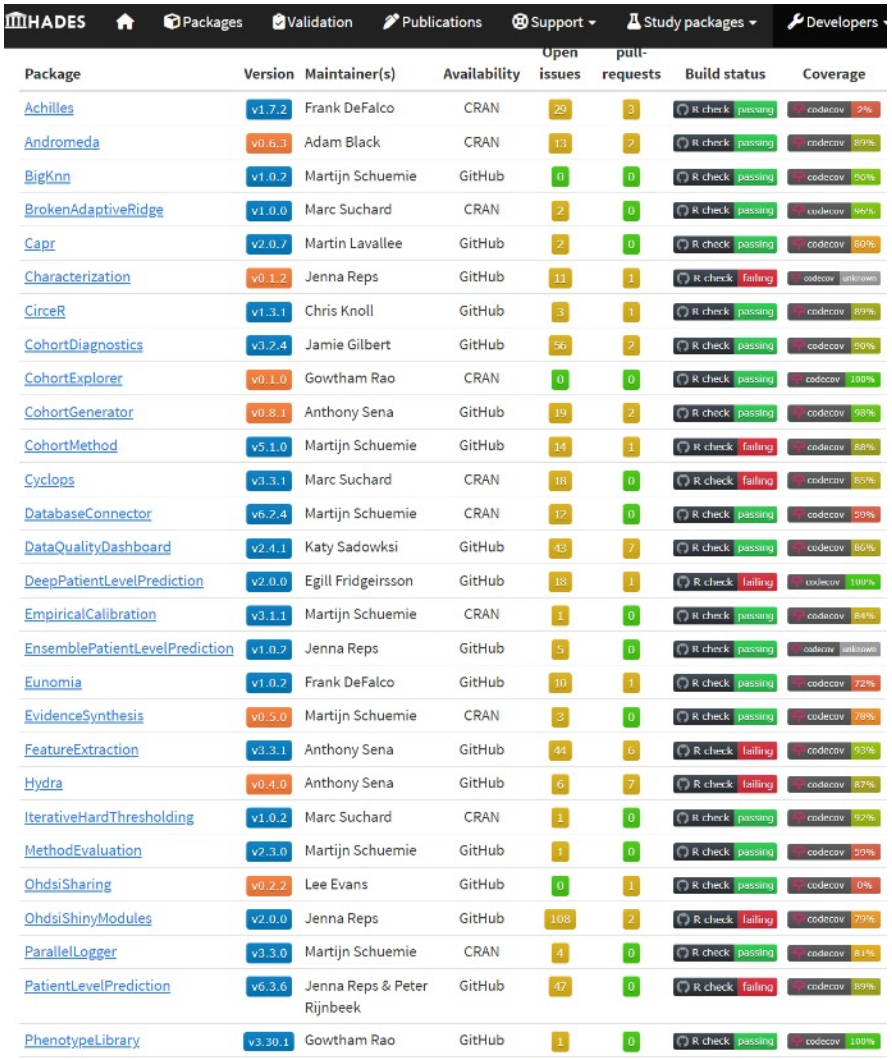
OHDSI Asia Pacific

OHDSI Europe

OHDSI Global Symposium

OHDSI Open Source Community

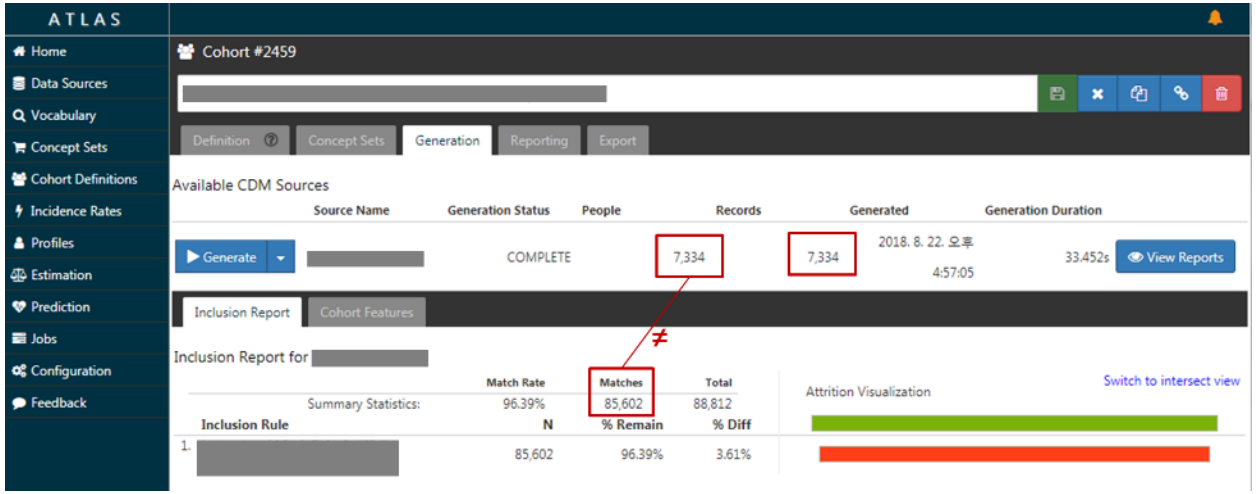
- 262 Repositories
- 30 M+ lines of code
- 681 Developers
- 31 organizations
- 47,672 commits
- 2,838 GitHub Forks
- 4,168 GitHub Stars
- 5,547 GitHub Subscribers

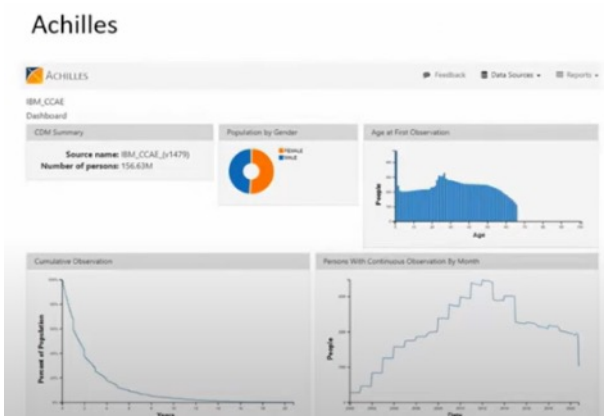


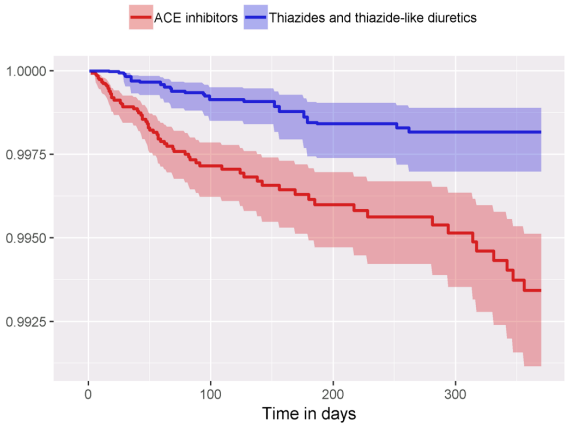
Package	Version	Maintainer(s)	Availability	Open issues	pull-requests	Build status	Coverage
Achilles	v1.7.2	Frank DeFalco	CRAN	29	3	R check passing	codecov 2%
Andromeda	v0.6.3	Adam Black	CRAN	13	2	R check passing	codecov 89%
BigKnn	v1.0.2	Martijn Schuemie	GitHub	0	0	R check passing	codecov 56%
BrokenAdaptiveRidge	v1.0.0	Marc Suchard	CRAN	2	0	R check passing	codecov 96%
Capr	v2.0.7	Martin Lavallee	GitHub	2	0	R check passing	codecov 80%
Characterization	v0.1.2	Jenna Reps	GitHub	11	1	R check failing	codecov unknown
CirceR	v1.3.1	Chris Knoll	GitHub	3	1	R check passing	codecov 89%
CohortDiagnostics	v3.2.4	Jamie Gilbert	GitHub	56	2	R check passing	codecov 90%
CohortExplorer	v0.1.0	Gowtham Rao	CRAN	0	0	R check passing	codecov 100%
CohortGenerator	v0.8.1	Anthony Sena	GitHub	19	2	R check passing	codecov 98%
CohortMethod	v5.1.0	Martijn Schuemie	GitHub	14	1	R check failing	codecov 88%
Cyclops	v3.3.1	Marc Suchard	CRAN	18	0	R check failing	codecov 85%
DatabaseConnector	v6.2.4	Martijn Schuemie	CRAN	12	0	R check passing	codecov 59%
DataQualityDashboard	v2.4.1	Katy Sadowksi	GitHub	43	7	R check passing	codecov 66%
DeepPatientLevelPrediction	v2.0.0	Egill Fridgeirsson	GitHub	18	1	R check failing	codecov 100%
EmpiricalCalibration	v3.1.1	Martijn Schuemie	CRAN	1	0	R check passing	codecov 84%
EnsemblePatientLevelPrediction	v1.0.2	Jenna Reps	GitHub	5	0	R check passing	codecov unknown
Eunomia	v1.0.2	Frank DeFalco	GitHub	10	1	R check passing	codecov 72%
EvidenceSynthesis	v0.5.0	Martijn Schuemie	CRAN	3	0	R check passing	codecov 78%
FeatureExtraction	v3.3.1	Anthony Sena	GitHub	44	6	R check failing	codecov 93%
Hydra	v0.4.0	Anthony Sena	GitHub	6	7	R check failing	codecov 87%
IterativeHardThresholding	v1.0.2	Marc Suchard	CRAN	1	0	R check passing	codecov 92%
MethodEvaluation	v2.3.0	Martijn Schuemie	GitHub	1	0	R check passing	codecov 59%
OhdsiSharing	v0.2.2	Lee Evans	GitHub	0	1	R check passing	codecov 0%
OhdsiShinyModules	v2.0.0	Jenna Reps	GitHub	108	2	R check failing	codecov 79%
ParallelLogger	v3.3.0	Martijn Schuemie	CRAN	4	0	R check passing	codecov 81%
PatientLevelPrediction	v6.3.6	Jenna Reps & Peter Rijnbeek	GitHub	47	0	R check failing	codecov 89%
PhenotypeLibrary	v3.30.1	Gowtham Rao	GitHub	1	0	R check passing	codecov 100%

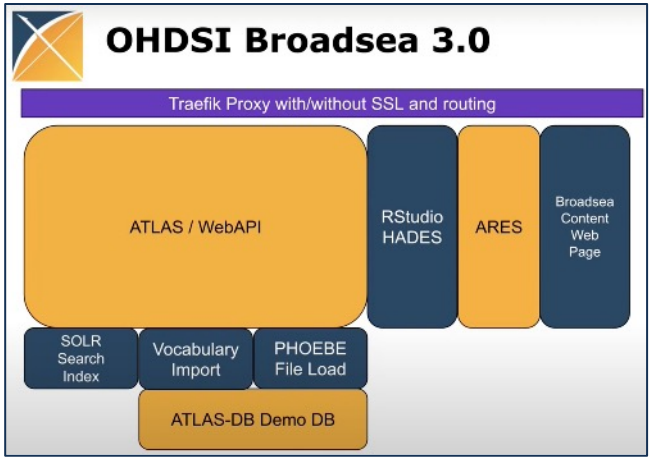
https://ohdsi.github.io/Hades/packageStatuses.html

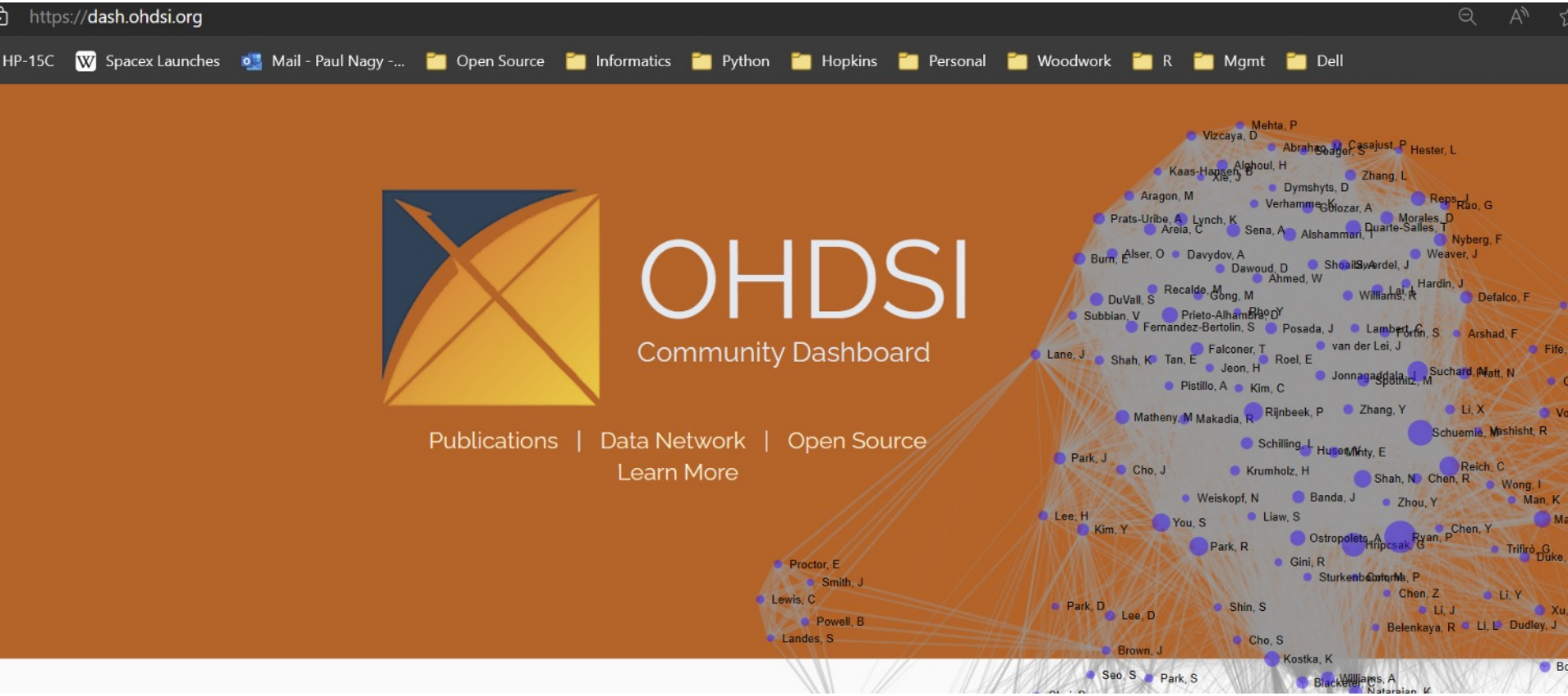
Data Science Applications





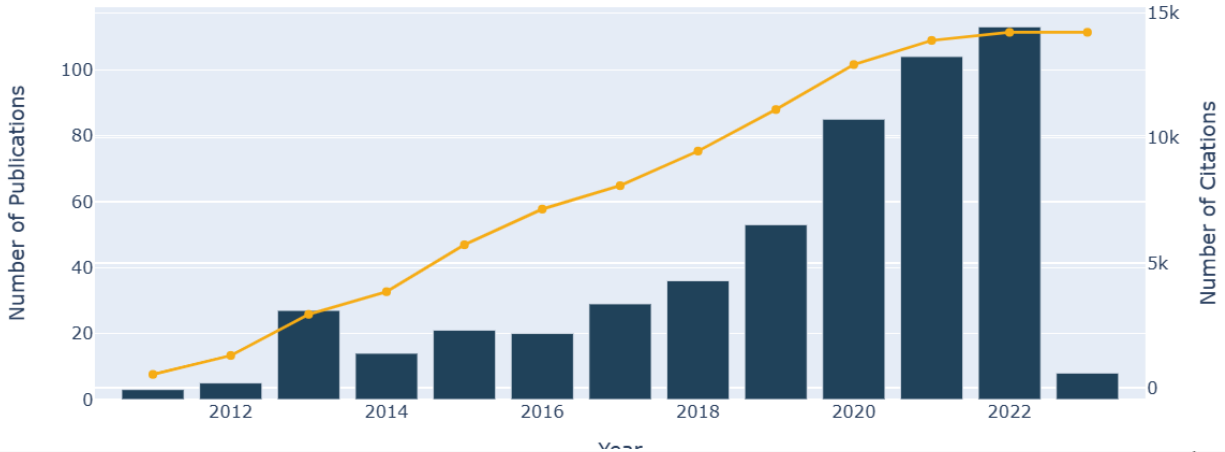





https://dash.ohdsi.org
HP-15C
Spacex Launches
Mail - Paul Nagy -...
Open Source
Informatics
Python
Hopkins
Personal
Woodwork
R
Mgmt
Dell
OHDSI
Community Dashboard
Publications | Data Network | Open Source
Learn More

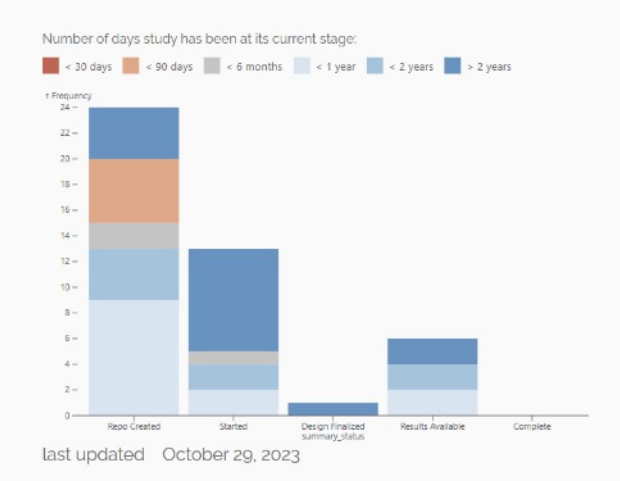
Publications

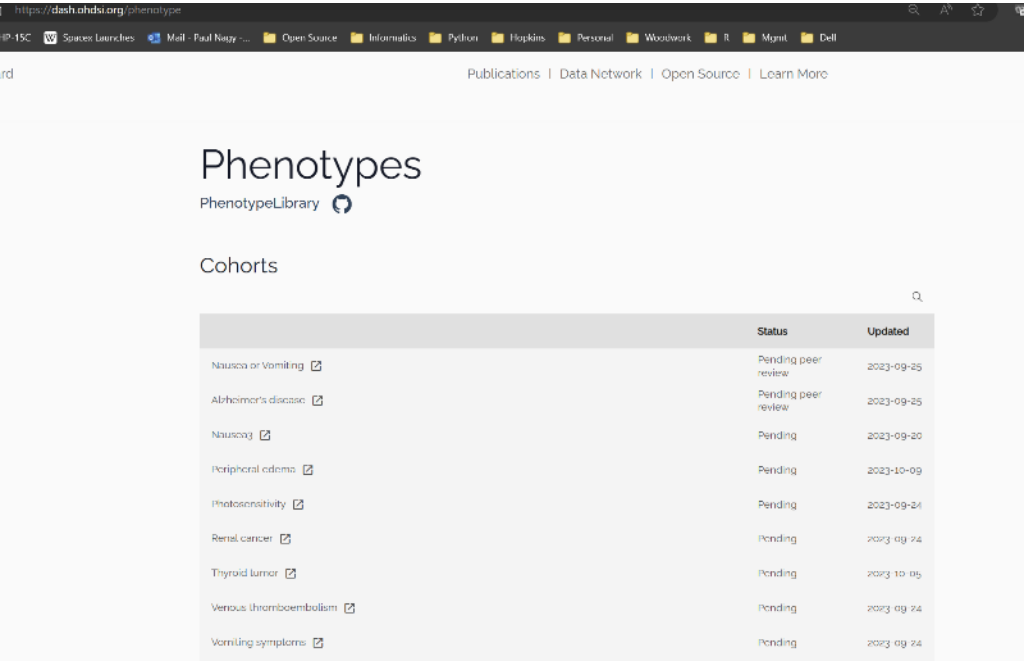
OHDSI Publications & Cumulative Citations

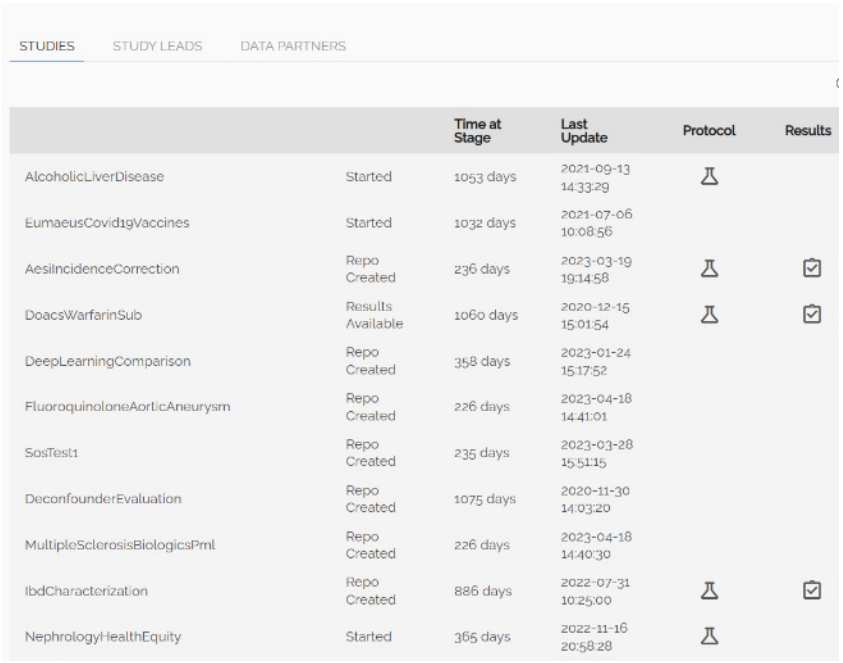


	Journal	Creation Date	Authors
Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD.	Pharmacoepidemiology and drug safety	2010/10/15 06:00	Schuemie, Martijn J
Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership.	Annals of internal medicine	2010/11/03 06:00	Stang, Paul E | Ryan, Patrick B | Racoosin, Judith A | Overhage, J Marc | Hartzema, Abraham G | Reich, Christian | Welebob, Emily | Scarnecchia, Thomas | Woodcock, Janet
Validation of a common data model for active safety surveillance research.	Journal of the American Medical Informatics Association : JAMIA	2011/11/01 06:00	Overhage, J Marc | Ryan, Patrick B | Reich, Christian G | Hartzema, Abraham G | Stang, Paul E
Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned.	Pharmacoepidemiology and drug safety	2012/01/21 06:00	Carnahan, Ryan M | Moores, Kevin G
Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases.	Journal of biomedical informatics	2012/06/12 06:00	Reich, Christian | Ryan, Patrick B | Stang, Paul E | Rocca, Mitra

Creating Evidence







https://ohdsi.org

Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

AMIA

INFORMATICS PROFESSIONALS. LEADING THE WAY.

Stump the Experts

Panel Session

Panelists

Mui Van Zandt
VP/Global Head Data Strategy, Access & Enablement - GM Inteliquet
IQVIA

Paul Nagy
Program Director for Graduate Training in Biomedical Informatics and Data Science,
Deputy Director of the Johns Hopkins Medicine Technology Innovation Center
Johns Hopkins University

Christian Reich
Professor of PracticeProfessor of Practice; Northeastern University,
CEO; Odysseus Data Services,

Format

45 min – moderated questions

15 min – Audience questions

Challenge our OHDSI panelists at AMIA!

Submit your most intriguing questions and be a part of our 'Stump the Experts' session!

Join by Web

PollEv.com/clt3

Join by Text

Send **clt3** and your message to **37607**

Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OHDSI RWE Revolution:

Igniting Data Modernization with Harmonized Standards for Cutting-Edge Health Research

11-Nov-2023

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Demo 1

Atlas 1: Building cohorts

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Demo 2

Atlas 2: Characterization and visualization

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Hands-on Session

Atlas 3: Group Exercise

Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OHDSI Methods and Research

OHDSI RWE Revolution: Igniting Data Modernization with Harmonized Standards for Cutting-Edge Health Research

Gowtham A Rao
Johnson and Johnson
Connect with me rao@ohdsi.org
#AMIA2023

Disclosure

I disclose the following relevant relationship with commercial interests:

- Employee of Johnson and Johnson
- Spouse is employee of Johnson and Johnson

Much of the slides and content borrowed with permission from other OHDSI collaborators

- Martijn Schuemie, Patrick Ryan, Marc Suchard, Anthony Sena etc.

Standard Framework for Research Questions: TCIO-TAR inputs

What are the standardized inputs?

Target (T): The exposure of interest

Comparator (C): A suitable comparator

Indication (I): Ensure prior membership in an underlying disease cohort (optional)

Outcome (O): Includes primary and secondary health status of interest either from an efficacy or safety perspective

Time at Risk (TAR): The a priori determined period of time upon which the outcome is assessed

Example Research: TCIO-TAR for Diabetes Mellitus

T: SGLT2i

C: GLP-1RA

I: T2D

O: MACE, HHF, DKA, genital infections, fractures, LLA, AKI, UTI, mortality

TAR: On treatment

Age: >66

Characterization: Differences in baseline characteristics between T and C.

Estimation: Difference in risk between T and C for the O in the TAR.

Prediction: Occurrence of O among T within TAR

Comparative Effectiveness and Safety of Sodium–Glucose Cotransporter 2 Inhibitors Versus Glucagon-Like Peptide 1 Receptor Agonists in Older Adults

Elisabetta Patorno,[1] Ajinkya Pawar,[1] Lily G. Bessette,[1] Dae H. Kim,[1,2,3] Chintan Dave,[1,4] Robert J. Glynn,[1] Medha N. Munshi,[2,5] Sebastian Schneeweiss,[1] Deborah J. Wexler,[6] and Seoyoung C. Kim[1]

Diabetes Care 2021;44:826–835 | https://doi.org/10.2337/dc20-1464

OBJECTIVE

Both sodium–glucose cotransporter 2 inhibitors (SGLT2i) and glucagon-like peptide 1 receptor agonists (GLP-1RA) demonstrated cardiovascular benefits in randomized controlled trials of patients with type 2 diabetes (T2D) generally <65 years old and mostly with cardiovascular disease. We aimed to evaluate the comparative effectiveness and safety of SGLT2i and GLP-1RA among real-world older adults.

RESEARCH DESIGN AND METHODS

Using Medicare data (April 2013–December 2016), we identified 90,094 propensity score–matched (1:1) T2D patients ≥66 years old initiating SGLT2i or GLP-1RA. Primary outcomes were major adverse cardiovascular events (MACE) (i.e., myocardial infarction, stroke, or cardiovascular death) and hospitalization for heart failure (HHF). Other outcomes included diabetic ketoacidosis (DKA), genital infections, fractures, lower-limb amputations (LLA), acute kidney injury (AKI), severe urinary tract infections, and overall mortality. We estimated hazard ratios (HRs) and rate differences (RDs) per 1,000 person-years, controlling for 140 baseline covariates.

Example Research: TCIO-TAR for Hormone Replacement Therapy

T: CE/BZA

C: EP

I: none

O: endometrial cancer, endometrial hyperplasia, and breast cancer (and others in the methods section)

TAR: On treatment

Characterization: Differences in baseline characteristics between T and C.

Estimation: Difference in risk between T and C for the O in the TAR.

Prediction: Occurrence of O among T within TAR

ORIGINAL STUDY

Comparative safety of conjugated estrogens/bazedoxifene versus estrogen/progestin combination hormone therapy among women in the United States: a multidatabase cohort study

Hoffman, Sarah R. MS, MPH, PhD[1]; Governor, Samuel MD, MPH[1]; Daniels, Kimberly PhD, MS[1]; Seals, Ryan M. MPH, ScD[2]; Ziyadeh, Najat J. MA, MPH[2]; Wang, Florence T. ScD[2]; Dai, Dingwei MD, PhD[3]; Mcmahill-Walraven, Cheryl N. MSW, PhD[3]; Shuminski, Patty AS[3]; Frajzyngier, Vera PhD[4]; Zhou, Xiaofeng PhD[4]; Shen, Rongjun MS[4]; Garg, Renu K. PhD, MPH[4]; Fournakis, Nicole MPH[1]; Lanes, Stephan PhD[1]; Beachler, Daniel C. MHS, PhD[1]

Author Information

Menopause 30(8):p 824-830, August 2023. | **DOI:** 10.1097/GME.0000000000002217

OPEN SDC

Metrics

Abstract | In Brief

Objective

To assess the risk of select safety outcomes including endometrial cancer, endometrial hyperplasia, and breast cancer among women using conjugated estrogens/bazedoxifene (CE/BZA) as compared with estrogen/progestin combination hormone therapy (EP).

Methods

We conducted a new-user cohort study in five US healthcare claims databases representing more than 92 million women. We included CE/BZA or EP new users from May 1, 2014, to August 30, 2019. EP users were propensity score (PS) matched to users of CE/BZA. Incidence of endometrial cancer, endometrial hyperplasia, breast cancer, and eight additional cancer and cardiovascular outcomes were ascertained using claims-based algorithms. Rate ratios (RR) and differences pooled across databases were estimated using random-effects models.

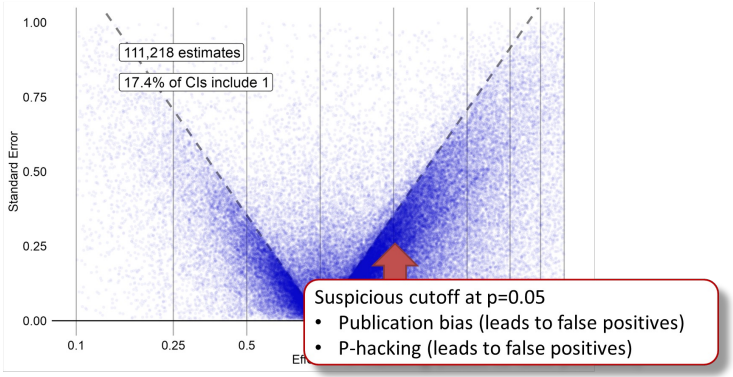
Value Proposition: Problem of 'TRUST'

We can answer a large set of questions using TCIO framework – but are our results trustable?

Problem: Long-standing issues of lack of TRUST in real-world evidence to guide clinical practice.

- Observational study bias
- Publication bias
- P-hacking
- Misleading estimates due to study design and analytical choices

Published observational study results



How do we earn and prevent erosion of TRUST in our science?

Three ideas

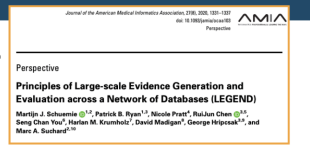
1. LEGEND Principles

METHODS RESEARCH

LEGEND in Principle

LEGEND (Large-scale Evidence Generation and Evaluation across a Network of Databases) applies high-level analytics to perform observational research on hundreds of millions of patient records within OHDSI's international database network.

LEGEND is based on 10 guiding principles that were published in JAMIA (August, 2020) and are listed below.


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

Perspective

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

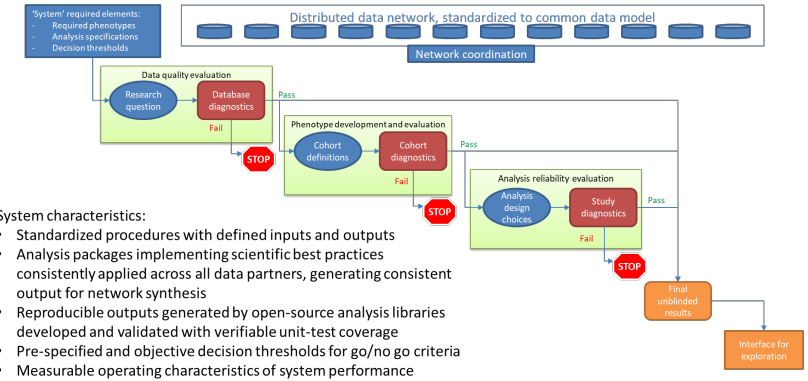
Martijn J. Schuemie[1,2], Patrick B. Ryan[1,3], Nicole Pratt[4], Rui Jun Chen[5,6], Seng Chan You[7], Harlan M. Krumholz[1], David Madigan[8], George Hripcsak[3,9], and Marc A. Suchard[2,10]

1. **LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). Aim: Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2. **Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. Aim: Avoids publication bias and enhances transparency.
3. **LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. Aim: Avoids P hacking.
4. **LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. Aim: Avoids P hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5. **LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. Aim: Minimizes bias.
6. **LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our P values, effect estimates, and confidence intervals using empirical calibration. [7,8] Aim: Enhances transparency on the uncertainty due to residual bias.
7. **LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. Aim: Enhances transparency and allows replication.
8. **LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). Aim: Avoids bias and improves interpretability.
9. **LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. Aim: Enhances generalizability and uncovers potential between-site heterogeneity.
10. **LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. Aim: Privacy.

#JoinTheJourney 47 OHDSI.org

2. Objective Diagnostics

Engineering open science systems that build trust into the real-world evidence generation and dissemination process



System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

3. Standardized software

HADES HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

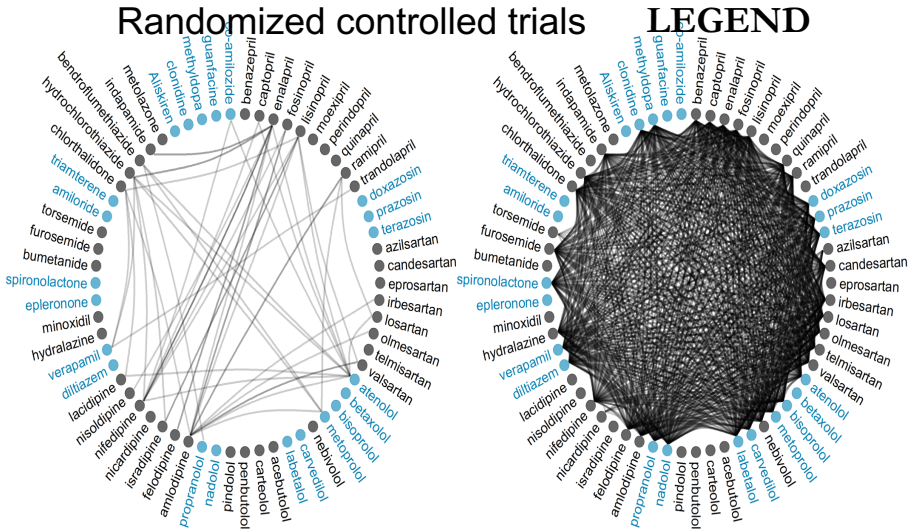
Open-source software

TRUST 1: LEGEND principles

Table 1: Guiding principles of the Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) initiative.

1	**LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). **Aim:** Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2	**Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. **Aim:** Avoids publication bias and enhances transparency.
3	**LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. **Aim:** Avoids *P* hacking.
4	**LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. **Aim:** Avoids *P* hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5	**LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. **Aim:** Minimizes bias.
6	**LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our *P* values, effect estimates, and confidence intervals using empirical calibration.[7,8] **Aim:** Enhances transparency on the uncertainty due to residual bias.
7	**LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. **Aim:** Enhances transparency and allows replication.
8	**LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). **Aim:** Avoids bias and improves interpretability.
9	**LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. **Aim:** Enhances generalizability and uncovers potential between-site heterogeneity.
10	**LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. **Aim:** Privacy.

Note: LEGEND: Large-scale Evidence Generation and Evaluation across a Network of Databases.



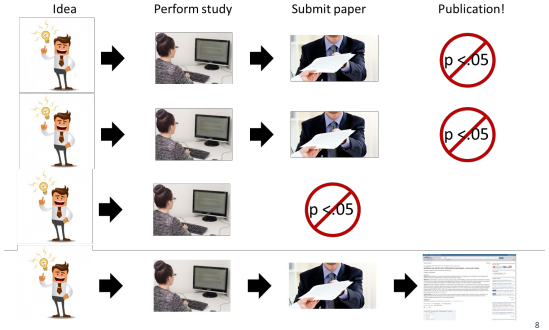
TRUST 1: LEGEND principles

Table 1: Guiding principles of the Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) initiative.

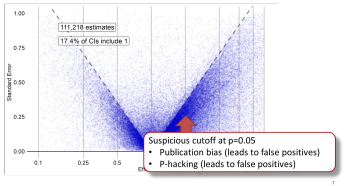
1	**LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). **Aim:** Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2	**Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. **Aim:** Avoids publication bias and enhances transparency.
3	**LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. **Aim:** Avoids *P* hacking.
4	**LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. **Aim:** Avoids *P* hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5	**LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. **Aim:** Minimizes bias.
6	**LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our *P* values, effect estimates, and confidence intervals using empirical calibration.[7,8] **Aim:** Enhances transparency on the uncertainty due to residual bias.
7	**LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. **Aim:** Enhances transparency and allows replication.
8	**LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). **Aim:** Avoids bias and improves interpretability.
9	**LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. **Aim:** Enhances generalizability and uncovers potential between-site heterogeneity.
10	**LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. **Aim:** Privacy.

Note: LEGEND: Large-scale Evidence Generation and Evaluation across a Network of Databases.

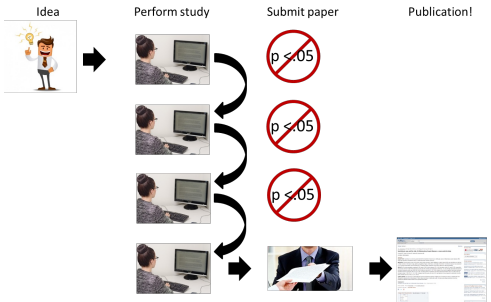
Publication bias



Published observational study results



P-hacking

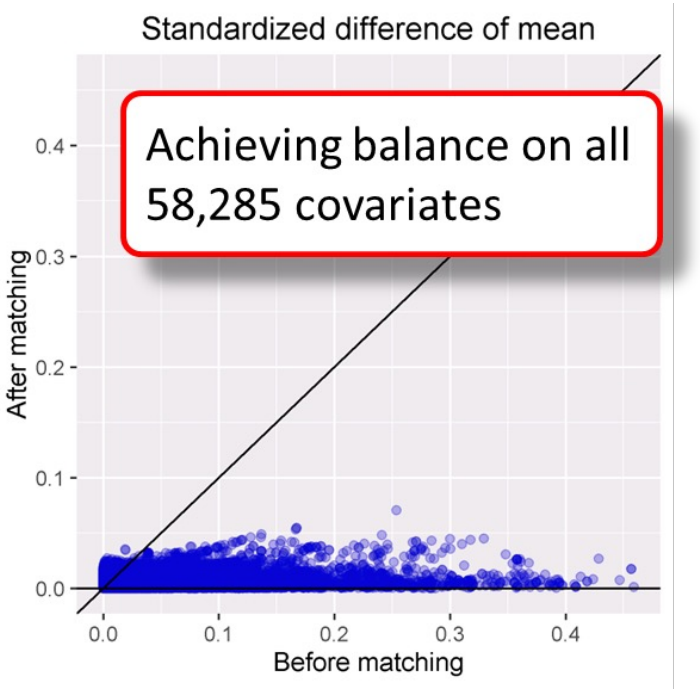


TRUST 1: LEGEND principles

Table 1: Guiding principles of the Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) initiative.

1	**LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). **Aim:** Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2	**Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. **Aim:** Avoids publication bias and enhances transparency.
3	**LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. **Aim:** Avoids *P* hacking.
4	**LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. **Aim:** Avoids *P* hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5	**LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. **Aim:** Minimizes bias.
6	**LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our *P* values, effect estimates, and confidence intervals using empirical calibration.[7,8] **Aim:** Enhances transparency on the uncertainty due to residual bias.
7	**LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. **Aim:** Enhances transparency and allows replication.
8	**LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). **Aim:** Avoids bias and improves interpretability.
9	**LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. **Aim:** Enhances generalizability and uncovers potential between-site heterogeneity.
10	**LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. **Aim:** Privacy.

Note: LEGEND: Large-scale Evidence Generation and Evaluation across a Network of Databases.



TRUST 1: LEGEND principles

Table 1: Guiding principles of the Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) initiative.

1	**LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). **Aim:** Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2	**Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. **Aim:** Avoids publication bias and enhances transparency.
3	**LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. **Aim:** Avoids *P* hacking.
4	**LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. **Aim:** Avoids *P* hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5	**LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. **Aim:** Minimizes bias.
6	**LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our *P* values, effect estimates, and confidence intervals using empirical calibration.[7,8] **Aim:** Enhances transparency on the uncertainty due to residual bias.
7	**LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. **Aim:** Enhances transparency and allows replication.
8	**LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). **Aim:** Avoids bias and improves interpretability.
9	**LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. **Aim:** Enhances generalizability and uncovers potential between-site heterogeneity.
10	**LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. **Aim:** Privacy.

Note: LEGEND: Large-scale Evidence Generation and Evaluation across a Network of Databases.

Measuring residual bias

TRUST 1: LEGEND principles

Table 1: Guiding principles of the Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) initiative.

1	**LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). **Aim:** Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2	**Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. **Aim:** Avoids publication bias and enhances transparency.
3	**LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. **Aim:** Avoids *P* hacking.
4	**LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. **Aim:** Avoids *P* hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5	**LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. **Aim:** Minimizes bias.
6	**LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our *P* values, effect estimates, and confidence intervals using empirical calibration.[7,8] **Aim:** Enhances transparency on the uncertainty due to residual bias.
7	**LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. **Aim:** Enhances transparency and allows replication.
8	**LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). **Aim:** Avoids bias and improves interpretability.
9	**LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. **Aim:** Enhances generalizability and uncovers potential between-site heterogeneity.
10	**LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. **Aim:** Privacy.

Note: LEGEND: Large-scale Evidence Generation and Evaluation across a Network of Databases.

HADES

HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

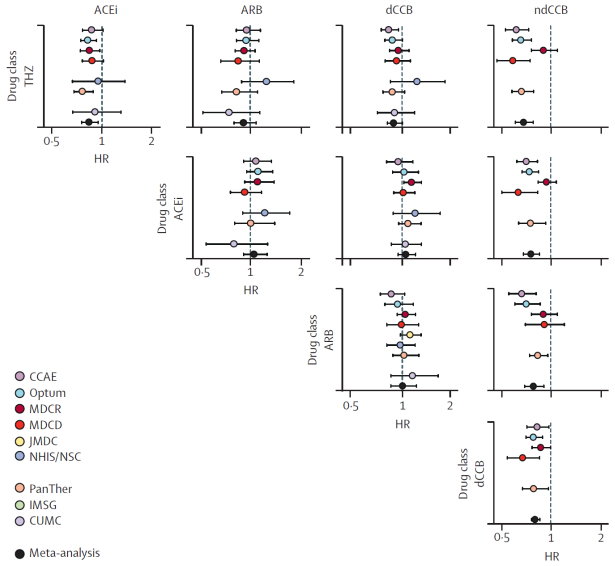
Open-source software

TRUST 1: LEGEND principles

Table 1: Guiding principles of the Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) initiative.

1	**LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). **Aim:** Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2	**Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. **Aim:** Avoids publication bias and enhances transparency.
3	**LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. **Aim:** Avoids *P* hacking.
4	**LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. **Aim:** Avoids *P* hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5	**LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. **Aim:** Minimizes bias.
6	**LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our *P* values, effect estimates, and confidence intervals using empirical calibration.[7,8] **Aim:** Enhances transparency on the uncertainty due to residual bias.
7	**LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. **Aim:** Enhances transparency and allows replication.
8	**LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). **Aim:** Avoids bias and improves interpretability.
9	**LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. **Aim:** Enhances generalizability and uncovers potential between-site heterogeneity.
10	**LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. **Aim:** Privacy.

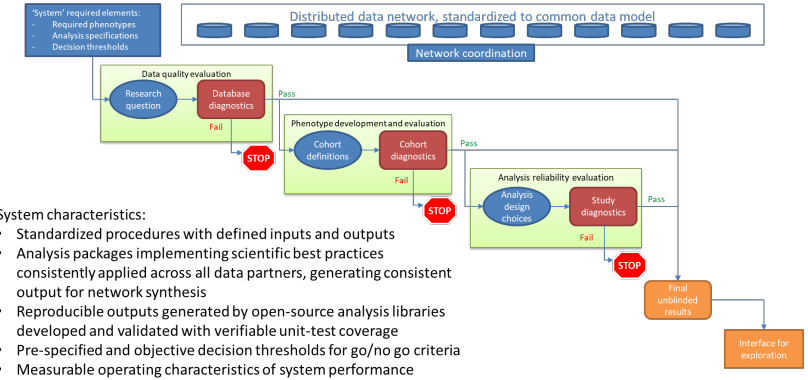
Note: LEGEND: Large-scale Evidence Generation and Evaluation across a Network of Databases.



TRUST 2: Objective Diagnostics

2. Objective Diagnostics

Engineering open science systems that build trust into the real-world evidence generation and dissemination process



System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

Statistical power Rule: Minimum Detectable Relative Risk (MDRR) < 10

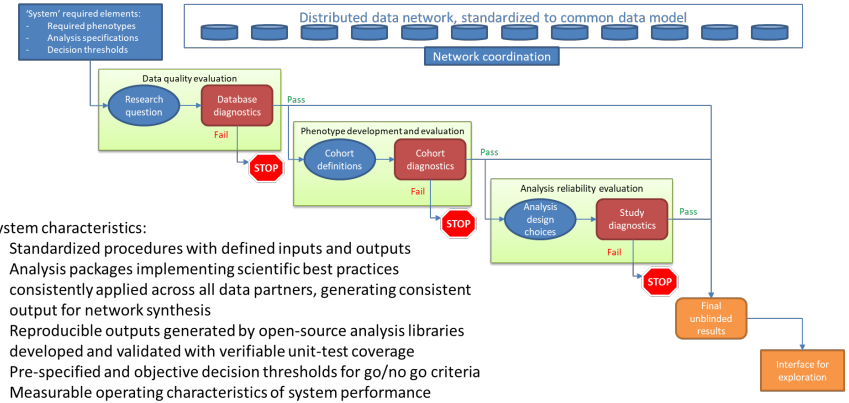
Comparability Rule: Equipoise > 0.5

Covariate Balance Rule: Max standardized difference of mean (SDM) < 0.1

Generalizability Rule: Max SDM between analytic cohort and target cohort < 0.25

Residual systematic error Rule: Expected Absolute Systematic Error (EASE) < 0.25

TRUST 2: Objective Diagnostics

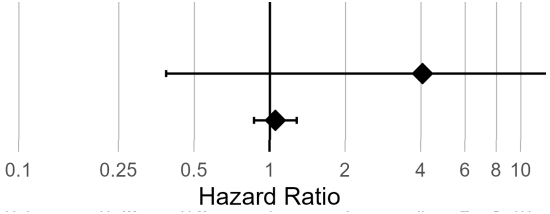


System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

Statistical power rule

Hazard Ratio (95% CI)
4.06 (0.39 - 42.60)
1.05 (0.86 - 1.28)

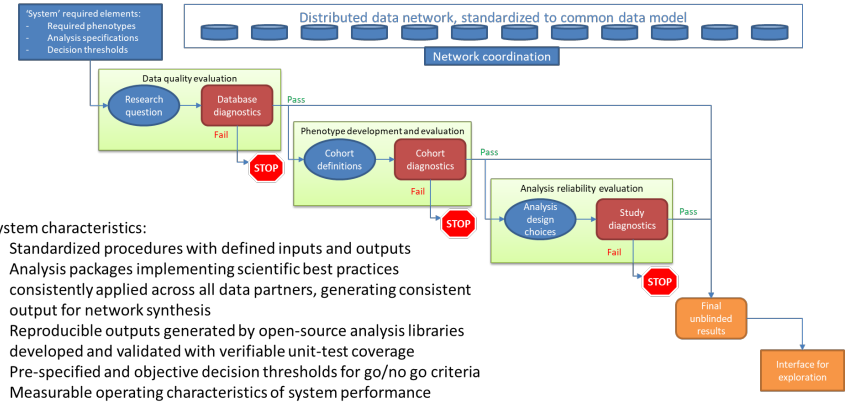


Rule: Minimum Detectable Relative Risk (MDRR) < 10

Reasoning: Even low-power estimate (wide CI) could be helpful, but we want to avoid misinterpreting grossly underpowered studies

Minimum Detectable Risk Ratio (MDRR) is a term used to describe the smallest relative risk that a study with a given power is capable of detecting.

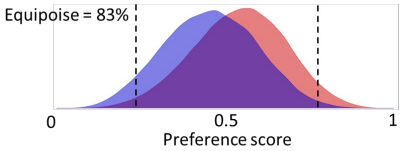
TRUST 2: Objective Diagnostics

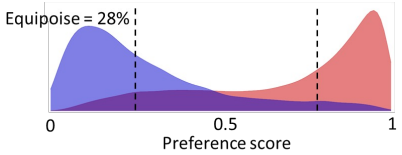


System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

Comparability Rule (Equipoise)



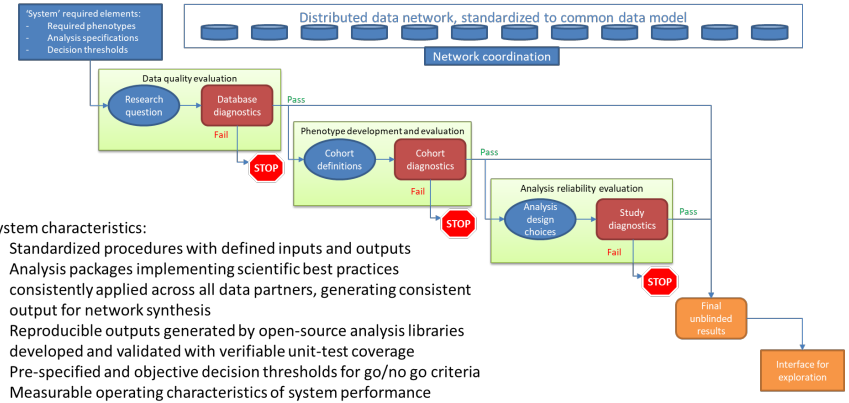


Rule: Equipoise > 0.5
(Equipoise is percent of population that has 0.3 < preference score < 0.7)

Reasoning: If equipoise is low, the populations are too incomparable, and we probably shouldn't trust our ability to make them comparable.

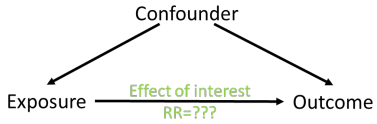
Preference = probability of patient choosing target vs. comparator treatment, given baseline features

TRUST 2: Objective Diagnostics

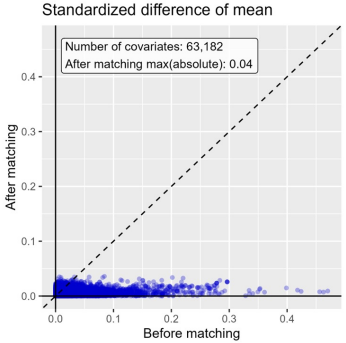


System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance



Covariate balance rule

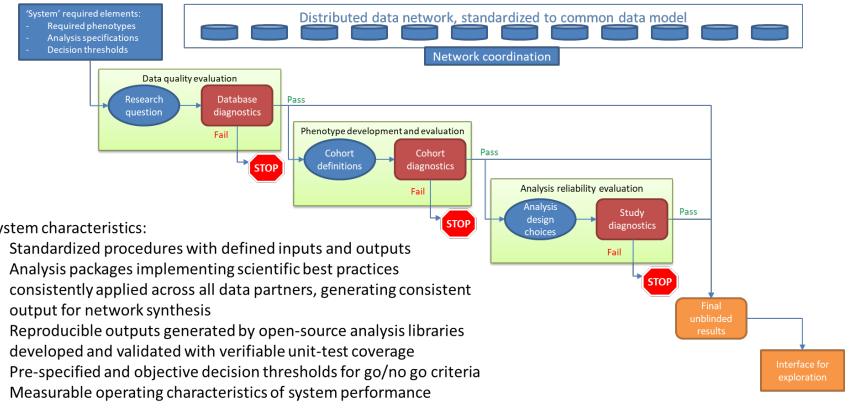


Rule: Max standardized difference of mean (SDM) < 0.1 (no covariate may have a SDM >= 0.1 after PS adjustment)

Reasoning: If covariates are unbalanced there may be confounding.

Confounding variables associated with both exposure and outcome can bias effect estimates if not properly addressed

TRUST 2: Objective Diagnostics



System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

Generalizability rule

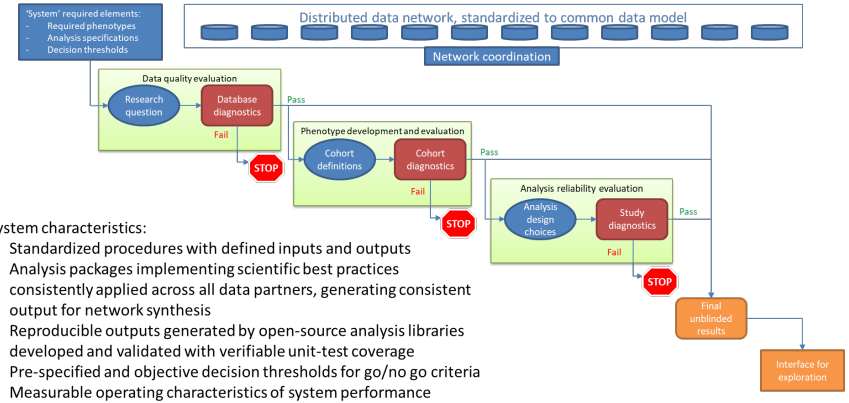
Strategies employed to reduce confounding (e.g. PS matching) can shift the composition of the analytic cohort away from the original target

Rule: Max SDM between analytic cohort and target cohort < 0.25
(target cohort: the cohort we started with (those exposed))
(analytic cohort: the cohort after all adjustments)

Reasoning: Estimate may not generalize to our target population if differences are too great.

Generalizability is the extent to which a study result can be applied to a target population of interest

TRUST 2: Objective Diagnostics

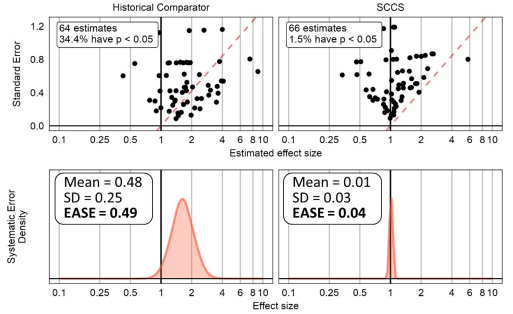


System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

Bias – expected value of systematic error – can be estimated using negative control experiments in which estimates can be compared with known truth

Residual systematic error rule (EASE statistic)



Rule: Expected Absolute Systematic Error (EASE) < 0.25 (EASE is the expected abs(log(estimated RR) – log(true RR)), based on negative control estimates)

Reasoning: Even though we can and should empirically calibrate to account for residual error, readers may not trust results if calibration shifts the estimates too much.

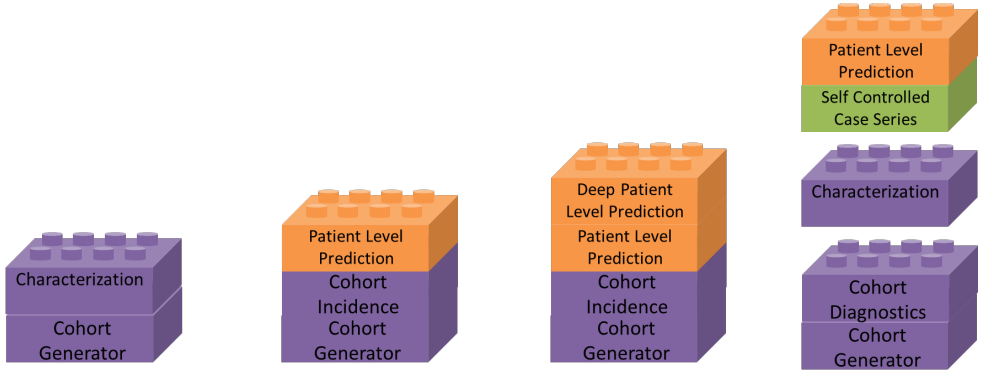
Residual systematic error can still exist due to model misspecification inherent to analysis or data

TRUST 3: Standardized software

3. Standardized software

HADES
HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Open-source software



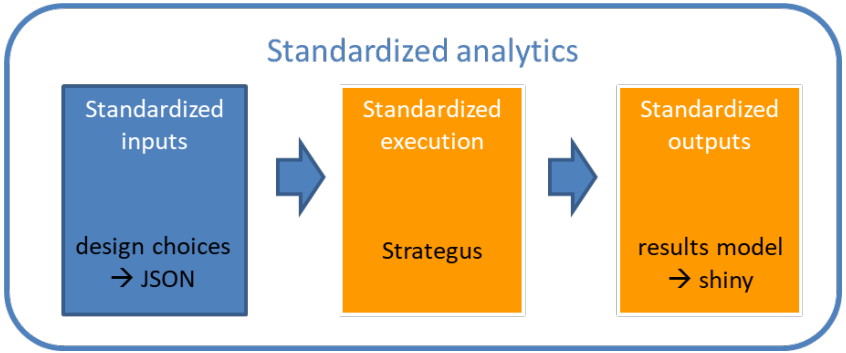
Cohort Generator: R package for generating cohorts using data in the CDM
Cohort Diagnostics: Evaluation of phenotype algorithms for OMOP CDM
Cohort Incidence: Performs incidence calculations on a CDM
Characterization: Performs characterization on target and comparator cohort
Cohort Method: performs new-user cohort studies in the OMOP CDM
Self Controlled Method: Performs Self-Controlled Case Series (SCCS) analyses in the OMOP CDM
Patient-Level Prediction: Performs patient level prediction in the OMOP CDM
Evidence Synthesis: R package for combining evidence from multiple sources

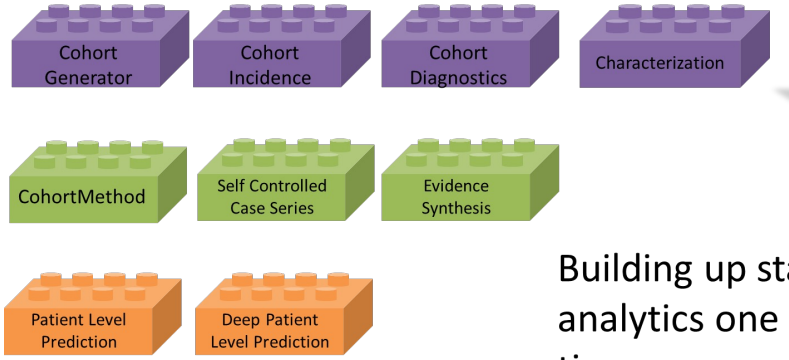
TRUST 3: Standardized software

What is Health Analytics Data to Evidence Suite (HADES)

- Open-source R packages for execution on OMOP CDM
- Principled software, grounded in methods research
- Documented, maintained, tested, empirically validated software
- Facilitates multi-question
- Facilitates large-scale analytics (big data)
- Distributed data network support: Enables federated analyses
- Platform independent: Compatible with diverse technical infrastructures

TRUST 3: Standardized software





Building up standardized analytics one Lego at a time.

Conclusion: TRUST → TCIO-TAR

Three ideas

1. LEGEND Principles

METHODS RESEARCH

LEGEND in Principle

LEGEND (Large-scale Evidence Generation and Evaluation across a Network of Databases) applies high-level analytics to perform observational research on hundreds of millions of patient records within OHDSI's international database network.

LEGEND is based on 10 guiding principles that were published in JAMIA (August, 2020) and are listed below.

Journal of the American Medical Informatics Association, 27(8), 2020, 1331–1337

Perspective

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

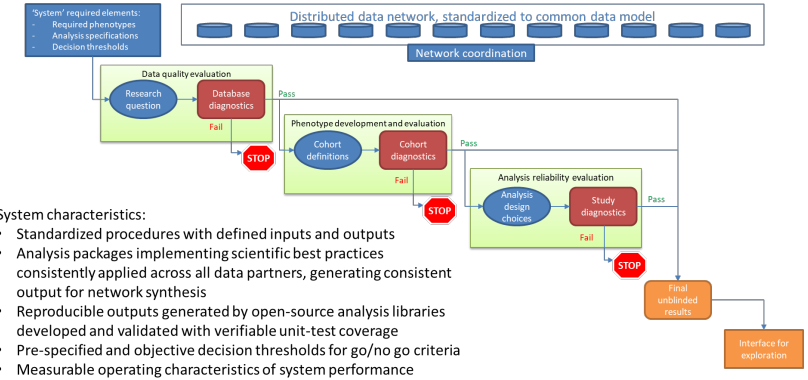
Martijn J. Schuemie, Patrick B. Ryan, Nicole Pratt, Rui Jun Chen, Seng Chan You, Harlan M. Krumholz, David Madigan, George Hripcsak, and Marc A. Suchard

1. **LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). Aim: Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2. **Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. Aim: Avoids publication bias and enhances transparency.
3. **LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. Aim: Avoids P hacking.
4. **LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. Aim: Avoids P hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5. **LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. Aim: Minimizes bias.
6. **LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our P values, effect estimates, and confidence intervals using empirical calibration. [7,8] Aim: Enhances transparency on the uncertainty due to residual bias.
7. **LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. Aim: Enhances transparency and allows replication.
8. **LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). Aim: Avoids bias and improves interpretability.
9. **LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. Aim: Enhances generalizability and uncovers potential between-site heterogeneity.
10. **LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. Aim: Privacy.

#JoinTheJourney 47 OHDSI.org

2. Objective Diagnostics

Engineering open science systems that build trust into the real-world evidence generation and dissemination process



System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

3. Standardized software

HADES HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Open-source software

Thank you!

Email me at:
rao@ohdsi.org

Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

Reproducibility and Trust

Ross D. Williams, PhD
Erasmus University Medical Center
Darwin EU® Analytics Team Lead
r.williams@erasmusmc.nl

Erasmus MC
University Medical Center Rotterdam

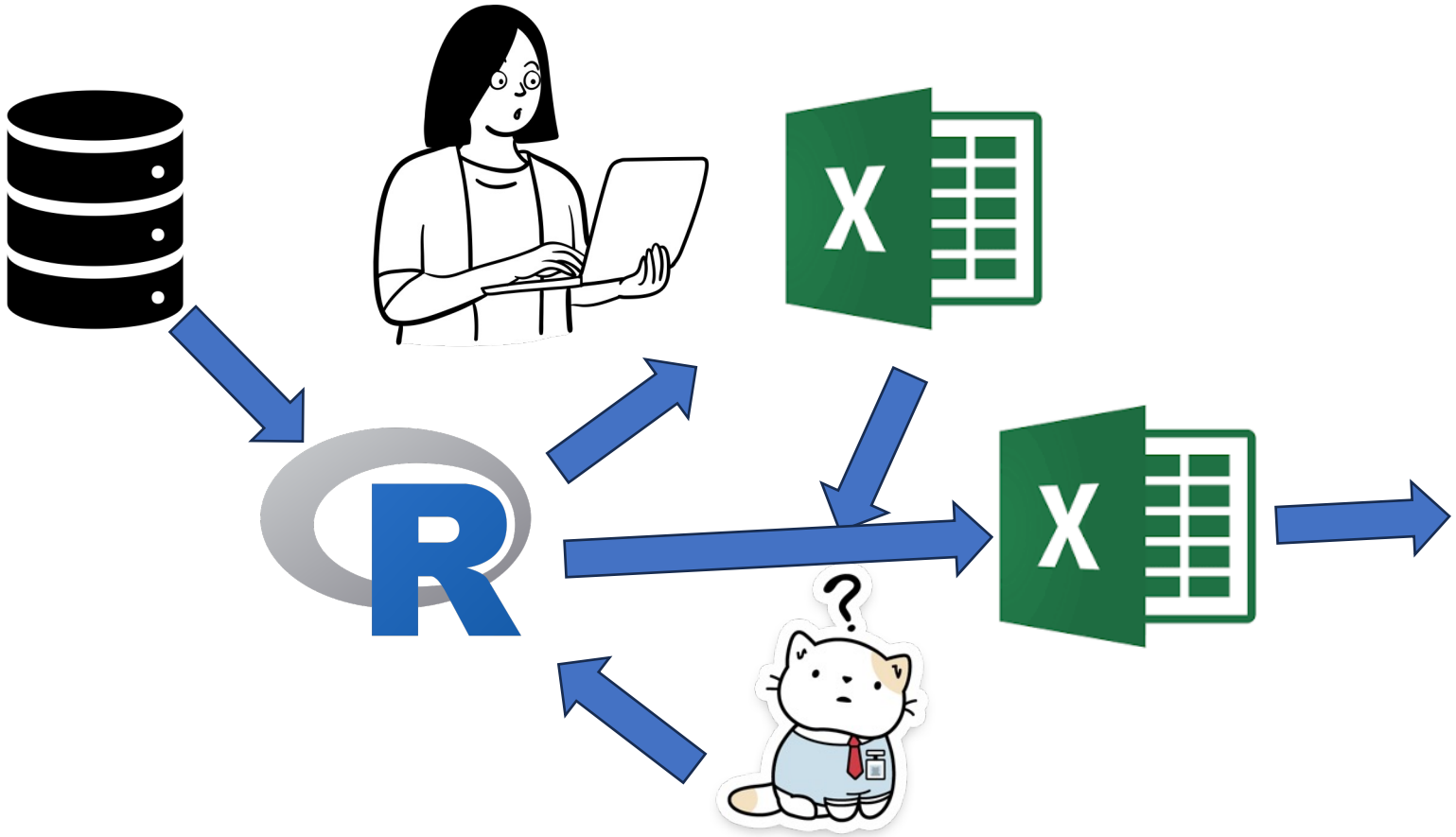
Disclosure

This presentation represents the views of the DARWIN EU® Coordination Centre only and cannot be interpreted as reflecting those of the European Medicines Agency or the European Medicines Regulatory Network.

What do we mean by reproducibility?

Desired attribute	Question	Researcher	Data	Analysis		Result
Repeatable	Identical	Identical	Identical	Identical	=	Identical
Reproducible	Identical	Different	Identical	Identical	=	Identical
Replicable	Identical	Same or different	Similar	Identical	=	Similar
Generalizable	Identical	Same or different	Different	Identical	=	Similar
Robust	Identical	Same or different	Same or different	Different	=	Similar
Calibrated	Similar (controls)	Identical	Identical	Identical	=	Statistically consistent

What does a traditional epi study look like?



Drug Safety (2022) 45:563–570
https://doi.org/10.1007/s40264-022-01161-8

ORIGINAL RESEARCH ARTICLE

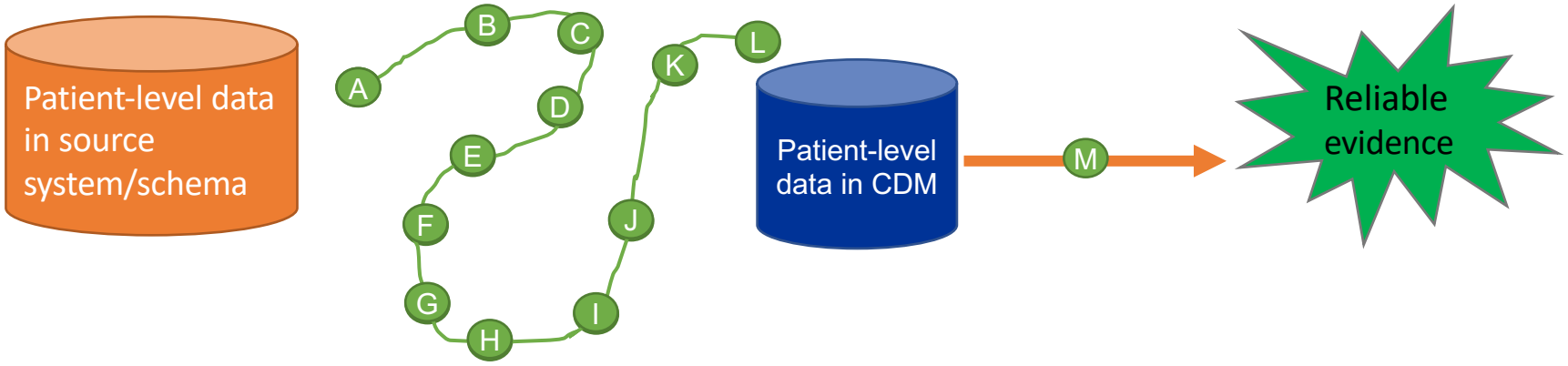
Using Iterative Pairwise External Validation to Contextualize Prediction Model Performance: A Use Case Predicting 1-Year Heart Failure Risk in Patients with Diabetes Across Five Data Sources

Ross D. Williams[1] · Jenna M. Reps[2] · Jan A. Kors[1] · Patrick B. Ryan[2] · Ewout Steyerberg[3] · Katia M. Verhamme[1] · Peter R. Rijnbeek[1]

Accepted: 9 February 2022

Generating Reliable Evidence using the OMOP Common Data Model

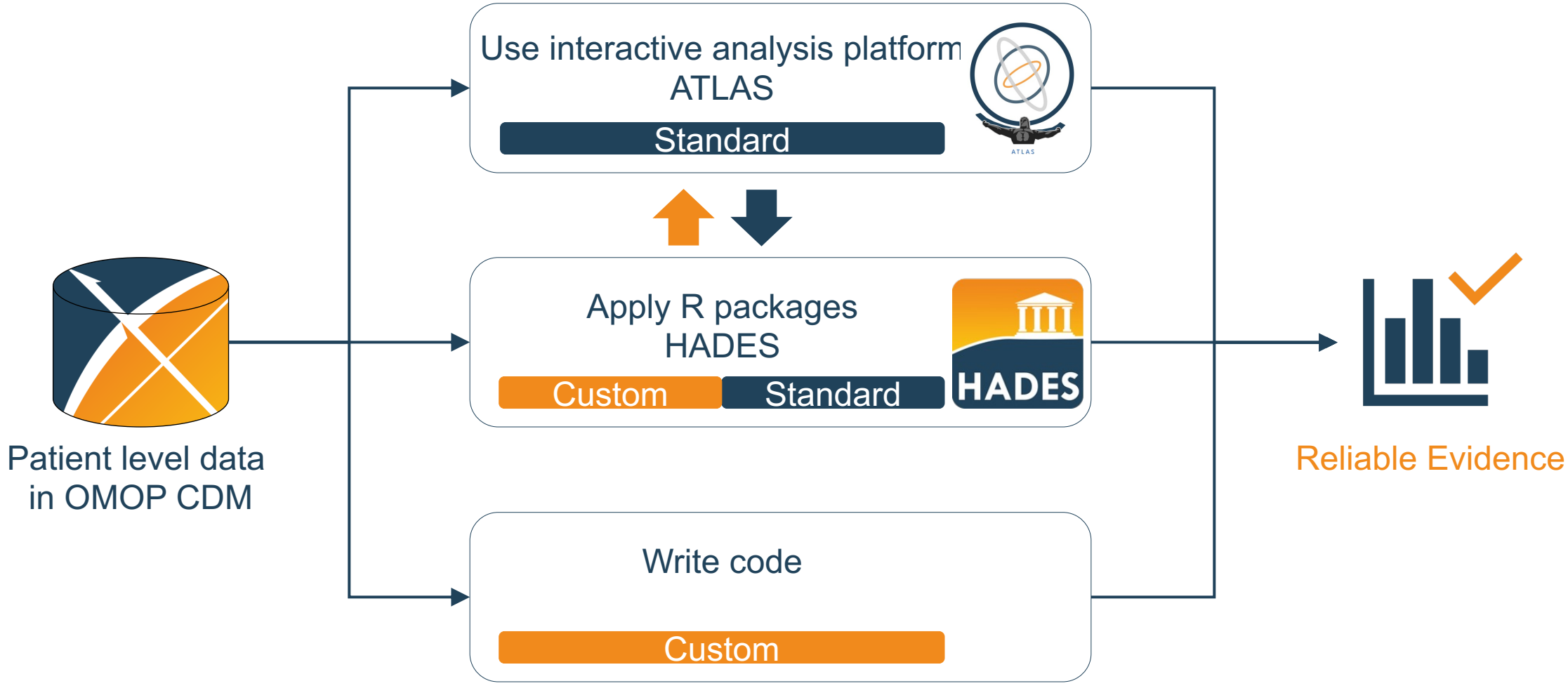
We need to make studies repeatable, reproducible, replicable, generalisable, and robust



A Common Data Model enables standardised analytics to generate reliable evidence.

Erasmus MC

OHDSI roads to reliable evidence



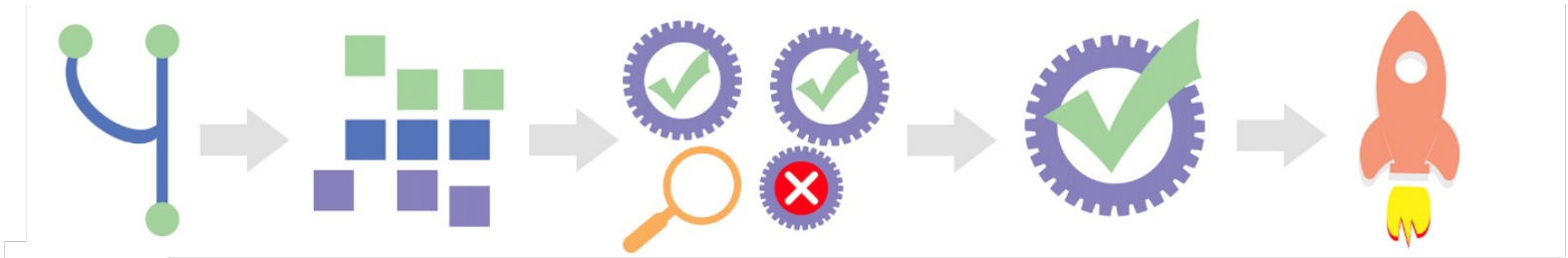
Diagnostics

- By standardizing the elements of studies (Characterisation, Estimation, Prediction)
  - We can standardize diagnostics

Darwin EU® perspective

- Standardise the questions, this allows for standardised software
- For the standard questions have standard software
  - Test this software
  - When errors are discovered, create a test, fix error
- This produces better software, more reliable answers and the research is reproducible
- Use renv to increase the likelihood of reproducing estimates at a later point

Software Review Process



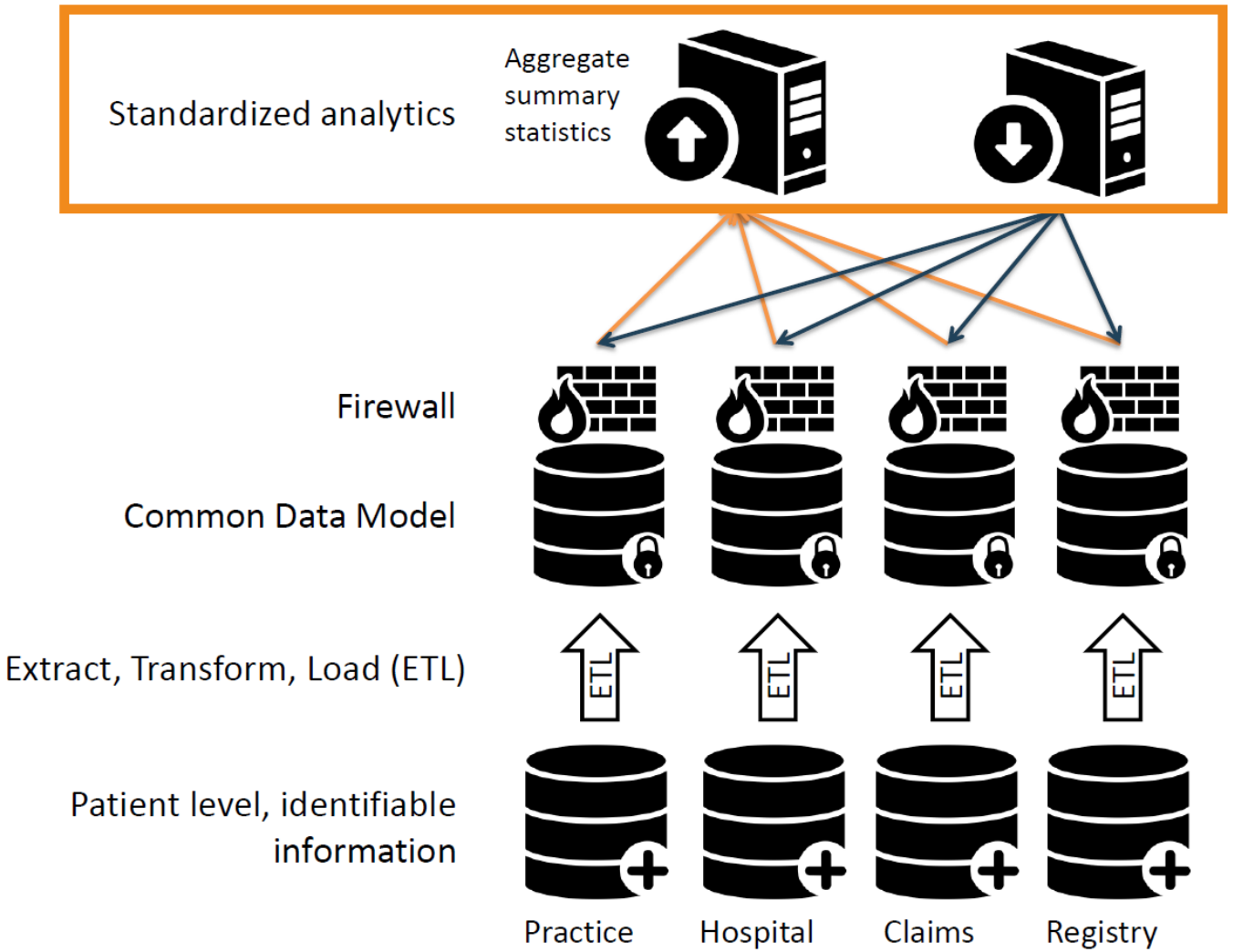
Clarity and openness

We have to respect patient privacy, and we must be open with analysis

Protocols, standard software, clear decision making

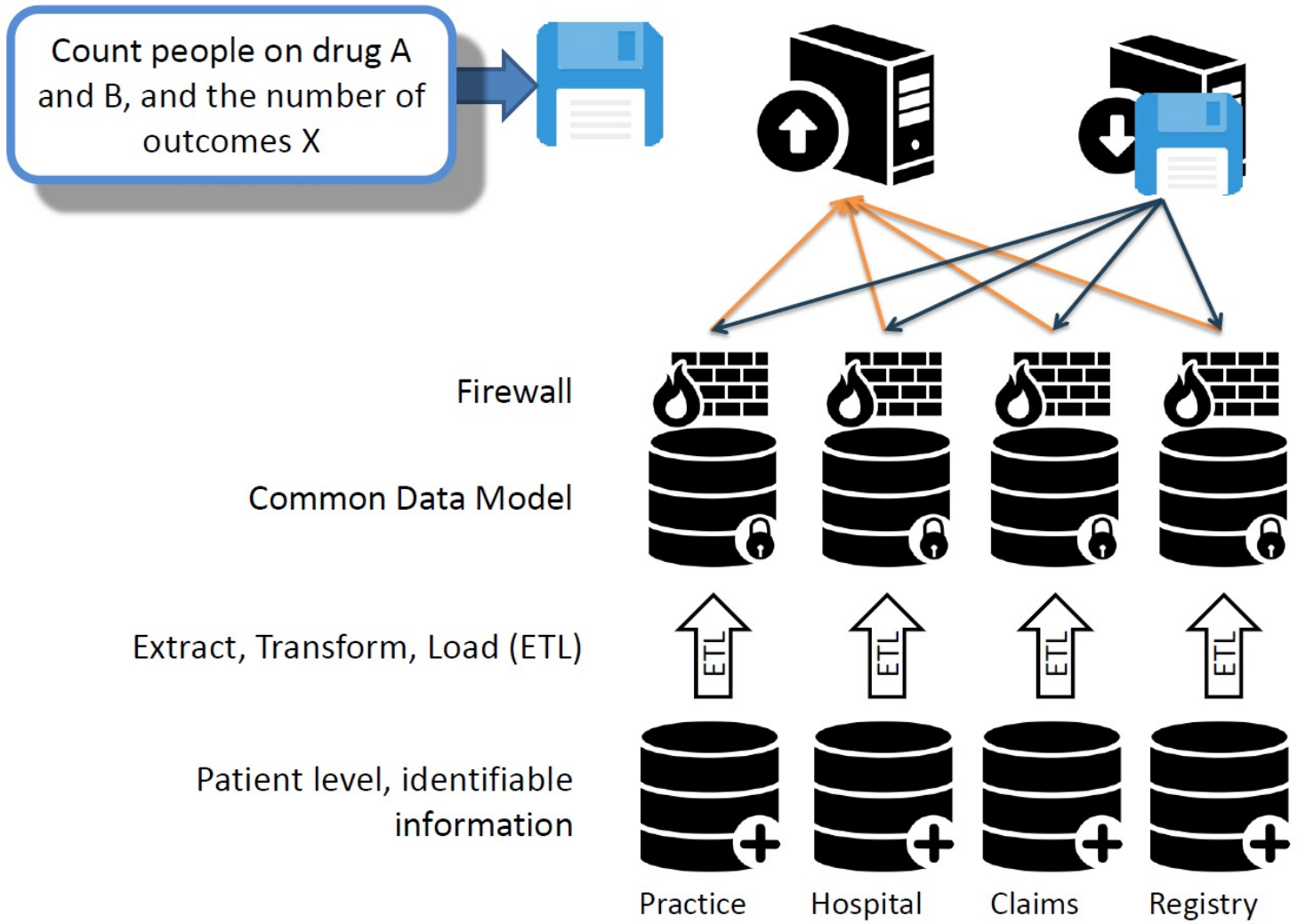
Erasmus MC

Federated analyses in OHDSI network



Schuemie MJ. How to extract transform and load observational data. https://www.ohdsi.org/wp-content/uploads/2014/07/Beijing2015.pdf?_ga=2.178811554.749634320.1678273784-1300990784.1664885317 Last accessed 09-MAR-23

Federated analyses in OHDSI network



Why standardisation makes research more trustworthy

Standardised pipelines can be incrementally improved over time

Flexibility can be improved through user interaction and development cycles

Moving to a standard design approach

Thanks for your attention

r.williams@erasmusmc.nl

Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

Erasmus MC

INFORMATICS PROFESSIONALS. LEADING THE WAY.

Evidence generation at scale

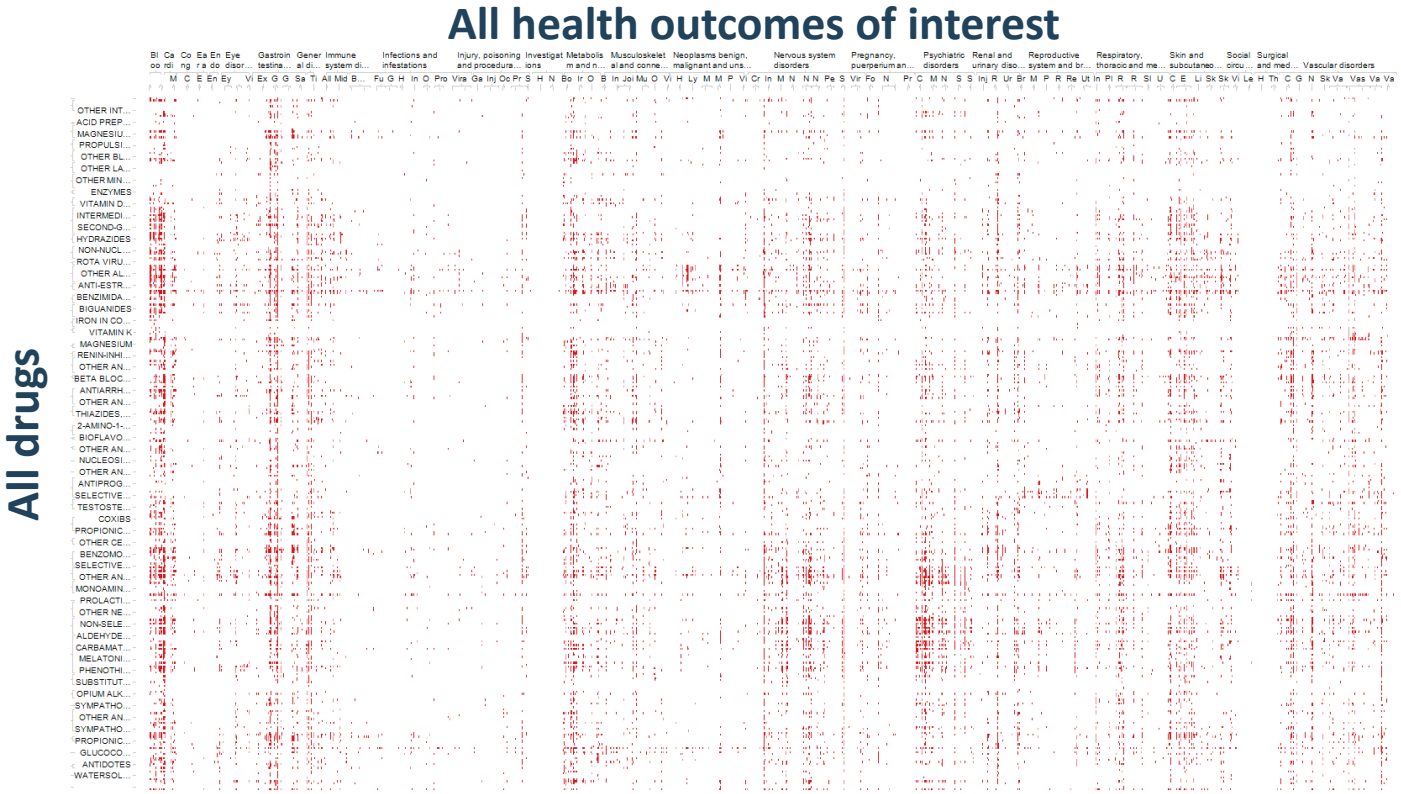
11-Nov-2023

Disclosure

I disclose the following relevant relationship with commercial interests:

- VP, Head of Data Science at Odysseus Data Services

Current pace of evidence generation in healthcare



Current evidence base for hypertension

Head-to-head antihypertensive drug comparisons

- Driven primarily by one clinical study ALLHAT- only 3 individual drugs
- Focus: mostly on efficacy

Can we provide

1. **reliable** – concordant w/ RCTs
2. **rich** – across "all" comparators
3. **relevant** – inform practice evidence?

Typical observational study

HEALTH CARE REFORM

Comparative Effectiveness of 2 β-Blockers in Hypertensive Patients

Emily D. Parker, MPH, PhD; Karen L. Margolis, MD, MPH; Nicole K. Trower, BS; David. J. Magid, MD, MPH; Heather M. Tavel, BS; Susan M. Shetterly, MS; P. Michael Ho, MD, PhD; Bix E. Swain, MS; Patrick J. O'Connor, MD, MPH

Two targets: atenolol and metoprolol
Three outcomes:

- Acute myocardial infarction
- Stroke
- Heart failure

ARCH INTERN MED/ VOL 172 (NO. 18)

Creating an evidence base for hypertension

Single ingredient comparisons

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296

Creating an evidence base for hypertension

+ Single drug classes comparisons

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156

Creating an evidence base for hypertension

+ single vs dual ingredient comparisons

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810

Creating an evidence base for hypertension

+ dual classes comparisons

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32

Creating an evidence base for hypertension

+ single vs dual class comparisons

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832

Creating an evidence base for hypertension

+ dual vs duo drugs comparisons

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784

Creating an evidence base for hypertension

+ dual vs duo class comparisons

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992

Creating an evidence base for hypertension

+ dual vs duo class comparisons

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278

Creating an evidence base for hypertension

+ expert curated outcomes

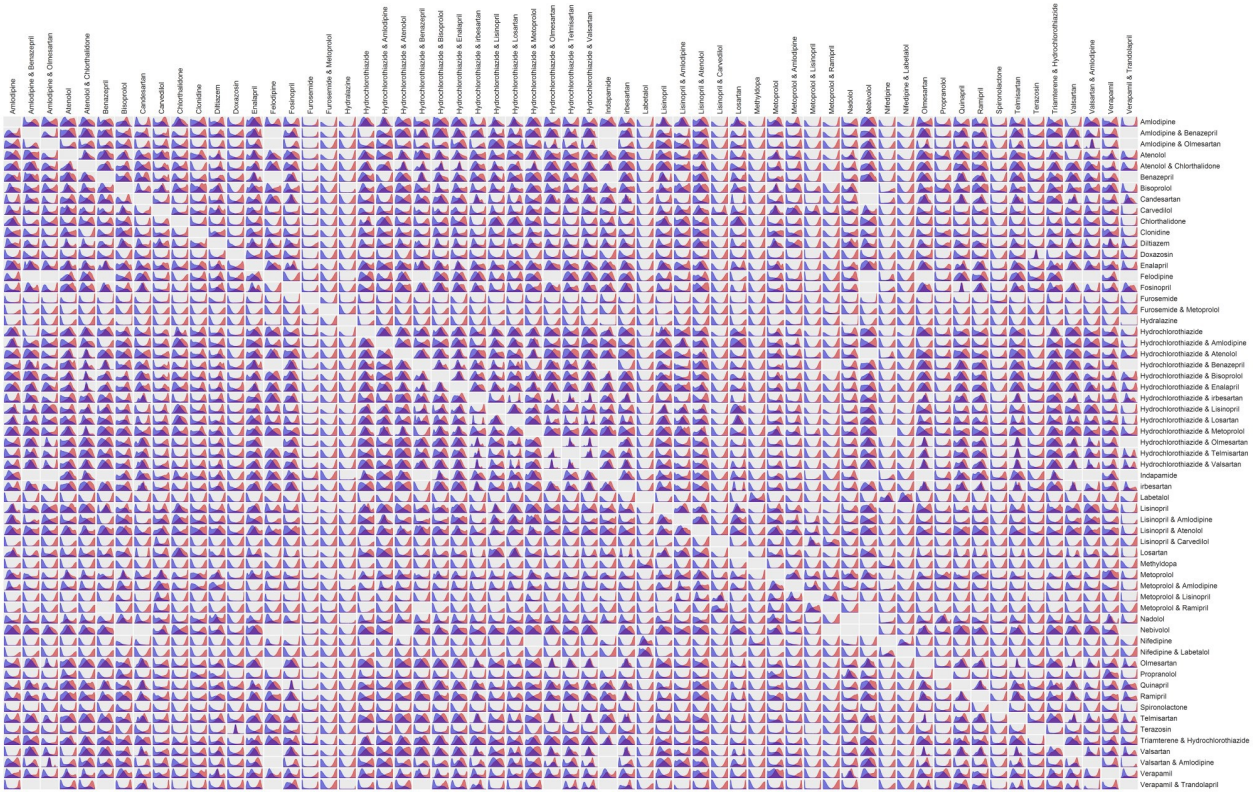
	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020

Creating an evidence base for hypertension

+ Diagnostics

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020
Diagnostics	164,908,500	587,020

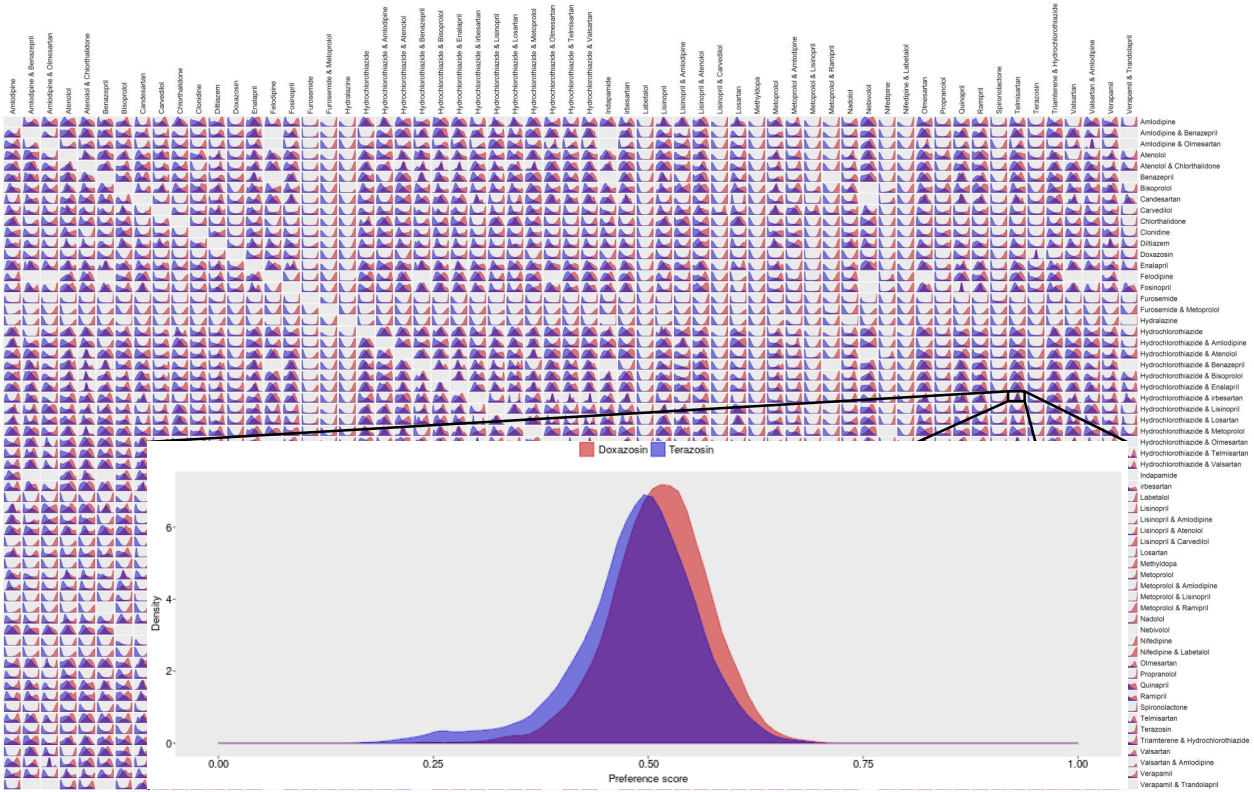
Best-practices: systematic large-scale PS



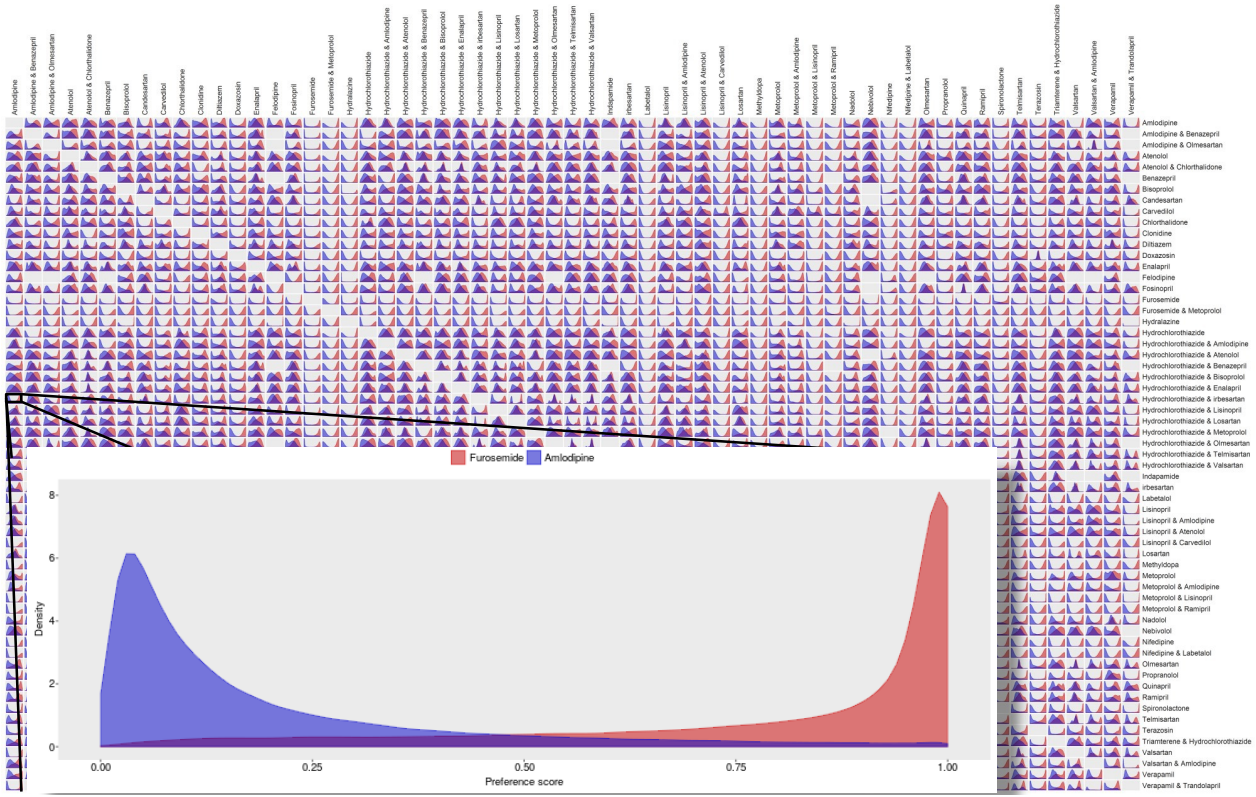
- >8,000 (regularized) baseline patient characteristics (all dx, rx, tx)
- Address observed (and some unobserved – BP control) confounding

Tian et al, 2019, IJE

Not all comparisons are valid



Not all comparisons are valid



Creating an evidence base for hypertension

+ negative controls

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020
Negative control outcomes	76	76
Target-comparator-neg controls	2,843,250 * 76=216,087,000	769,476

Creating an evidence base for hypertension

+ positive controls

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020
Negative control outcomes	76	76
Target-comparator-neg controls	2,843,250 * 76=216,087,000	769,476
Positive control outcomes	76 * 3=228	228
Target-comparator-pos controls	2,843,250 * 228=648,261,00	662,484
Total comparisons	864,348,000	1,431,960

Creating an evidence base for hypertension

+ positive controls

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020
Negative control outcomes	76	76
Target-comparator-neg controls	2,843,250 * 76=216,087,000	769,476
Positive control outcomes	76 * 3=228	228
Target-comparator-pos controls	2,843,250 * 228=648,261,00	662,484
Total comparisons	864,348,000	1,431,960
Total	864,348,000 * 9= 7,779,132,000	1,431,960 * 9=12,887,640

US Insurance databases

- IBM® MarketScan® CCAE
- IBM® MarketScan® MDCD
- IBM® MarketScan® MDCR
- Optum© Clinformatics®

Japanese insurance databases

- Japan Medical Data Center

Korean national insurance databases

- NHIS-NSC

US EHR databases

- Columbia University Medical Center
- Optum© PANTHER®

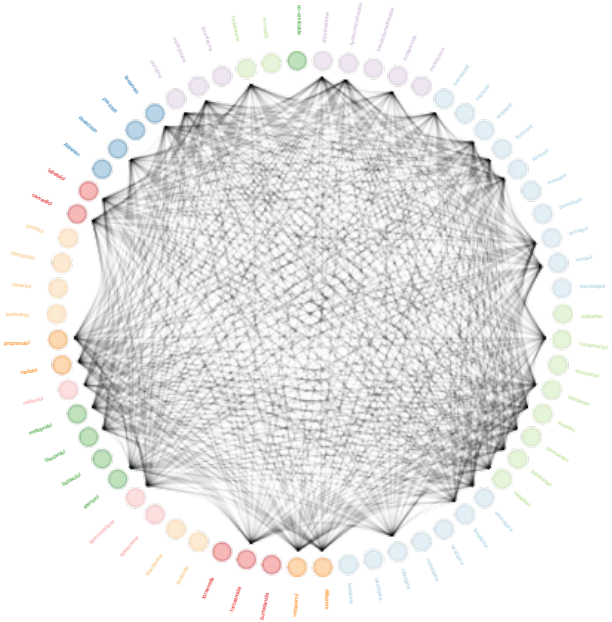
German EHR databases

- QuintilesIMS Disease Analyzer (DA) Germany

LEGEND knowledge base for hypertension

Head-to-head HTN drug comparisons

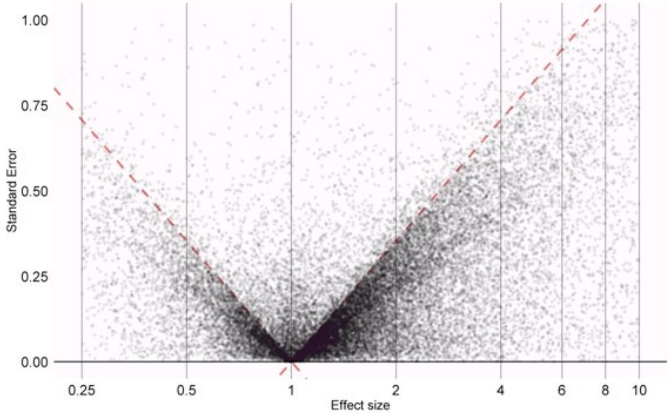
Trials: 40



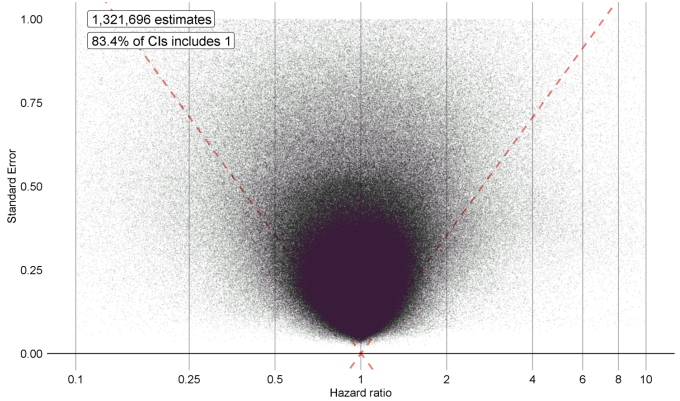
Comparisons: 10, 278

How does LEGEND perform?

Literature



LEGEND



- Best-practices **systematic design**, **evaluation** and empirical **calibration** return near nominal performance
- Provide a more complete and reliable evidence basis

Summary

OHDSI has created the know how, people and the technical stack to make evidence generation an industrial process

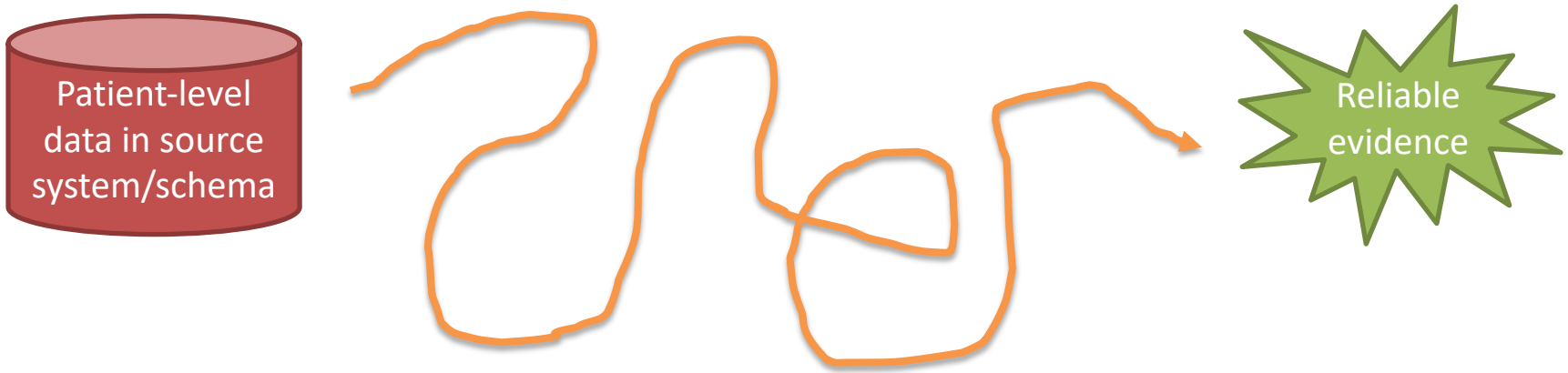
Agenda

8:30 – 8:50	Intro to OHDSI
8:50 – 9:10	Value Prop 1: Standardization
9:10 – 9:30	How to frame an RWE research question
9:30 – 9:50	Value Prop 2 : Community, Open-source
9:50 – 10:10	BREAK
10:10 – 11:10	Panel: Stump the Experts
11:10 – 11:30	Atlas 1: Building Cohorts
11:30 – 11:50	Value Prop 3: Methods and Research
11:50 – 12:10	Atlas 2: Characterization
12:10 – 13:10	LUNCH
13:10 – 13:30	Value Prop 4: Reproducibility and Trust
13:30 – 14:30	Atlas 3: Group Exercise
14:30 – 14:50	BREAK
14:50 – 15:10	RWE Evidence at Scale
15:10 – 15:30	Take Home Message

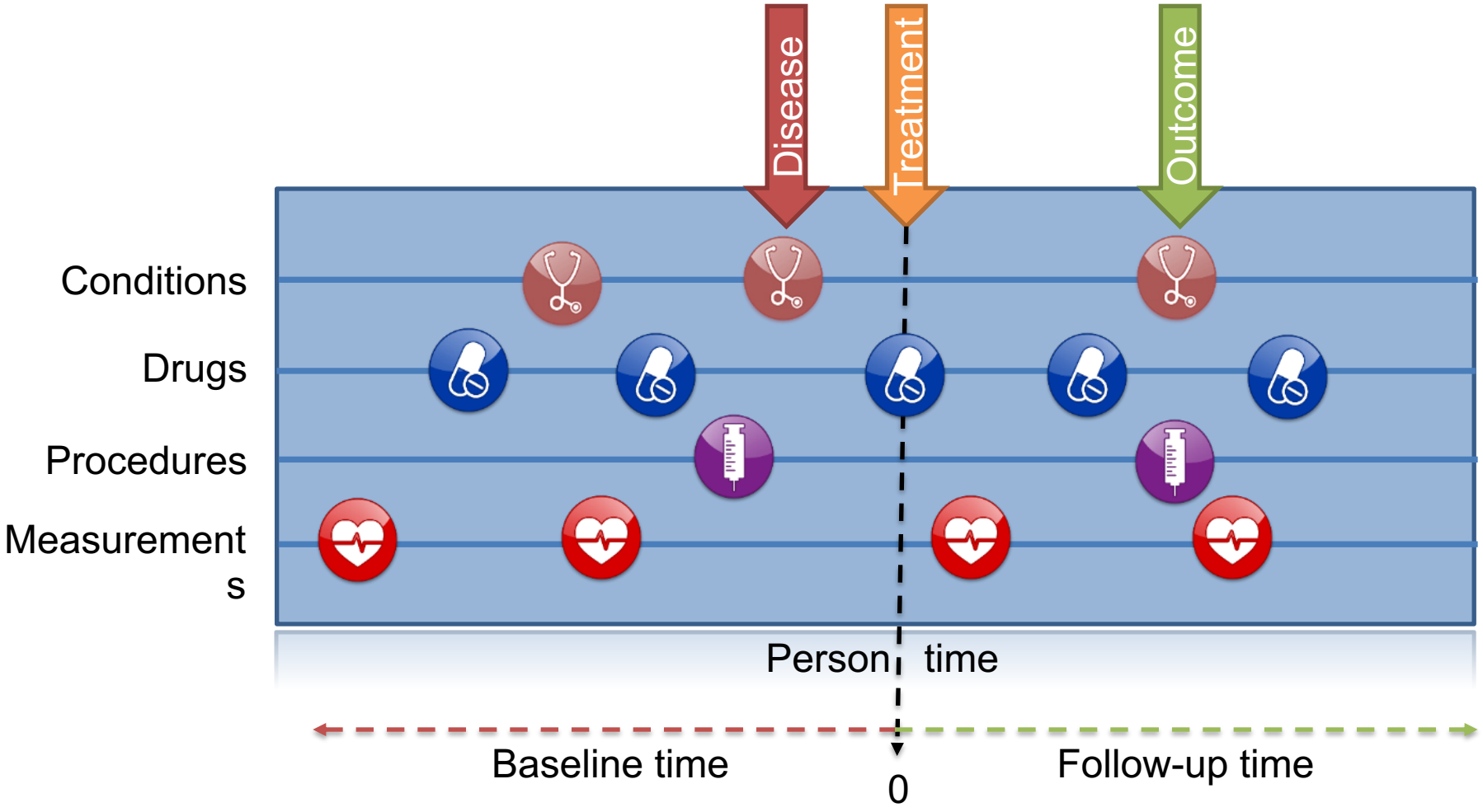
Evidence Generation

ME.250.961
Large Scale Observational Research Preparation

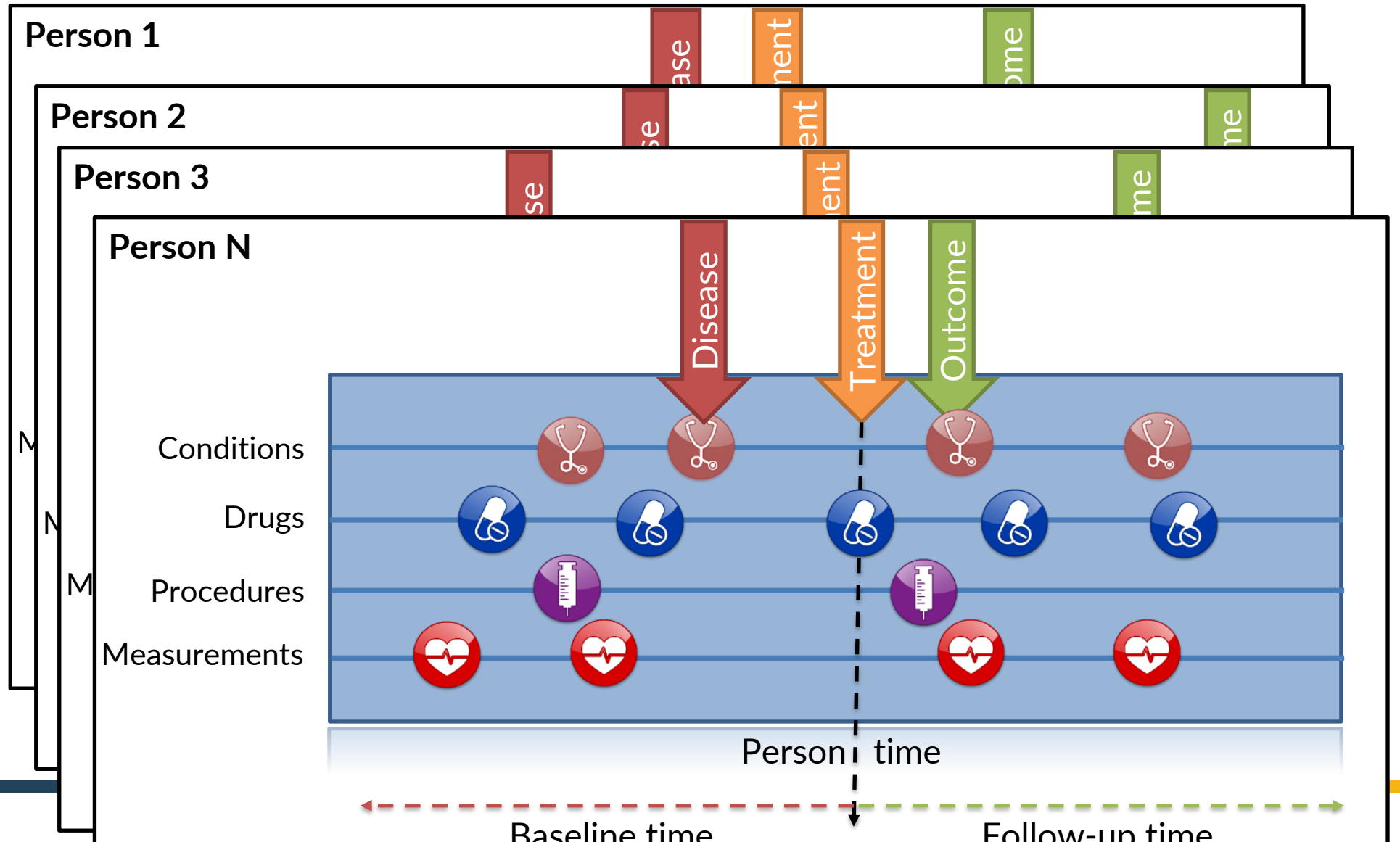
The journey to real-world evidence



A Caricature of The Patient Journey

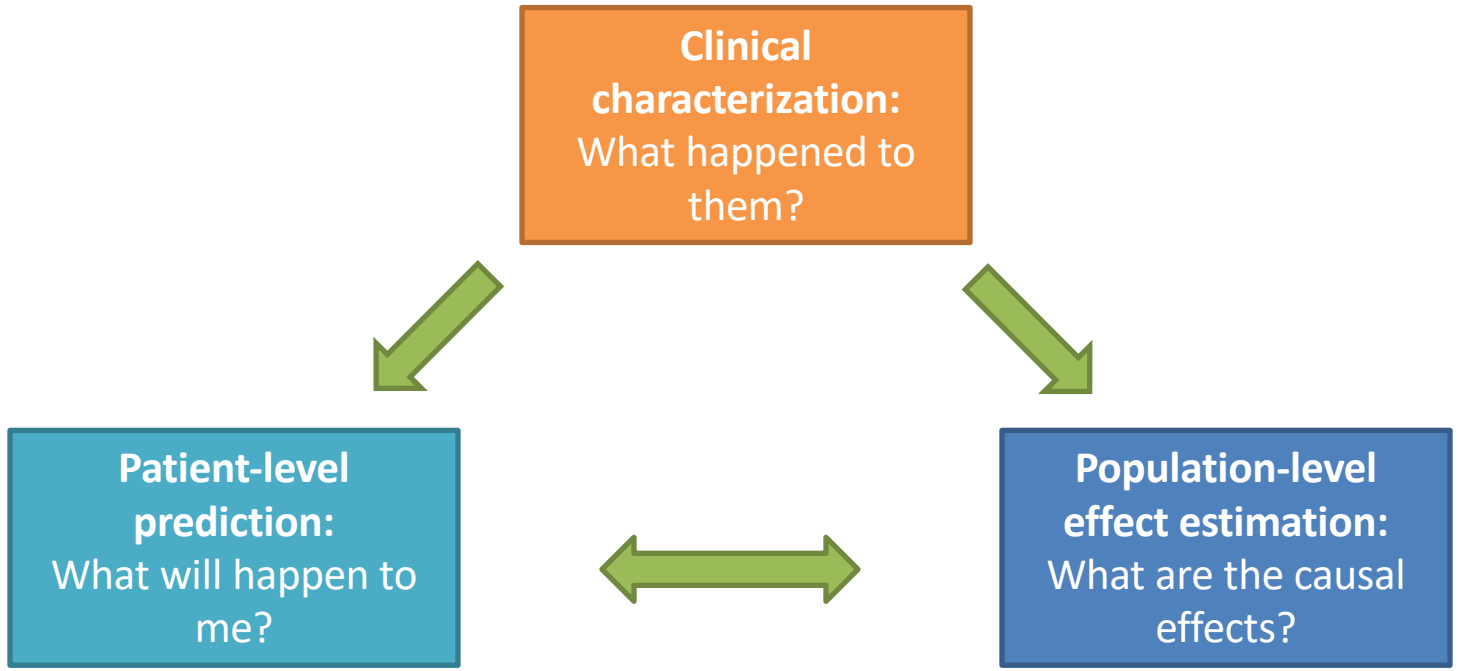


Each Observational Database Is Just an (Incomplete) Compilation of Patient Journeys



Complementary evidence to inform the patient journey

Evidence use cases:



Analytic use case	Type	Structure	Example
Clinical characterization	**Disease Natural History**	Amongst patients who are diagnosed with **<insert your favorite disease>**, what are the patient's characteristics from their medical history?	Amongst patients with **rheumatoid arthritis**, what are their demographics (age, gender), prior conditions, medications, and health service utilization behaviors?
	Treatment utilization	Amongst patients who have **<insert your favorite disease>**, which treatments were patients exposed to amongst **<list of treatments for disease>** and in which sequence?	Amongst patients with **depression**, which treatments were patients exposed to **SSRI, SNRI, TCA, bupropion, esketamine** and in which sequence?
	Outcome incidence	Amongst patients who are new users of **<insert your favorite drug>,** how many patients experienced **<insert your favorite known adverse event from the drug profile>** within **<time horizon following exposure start>**?	Amongst patients who are new users of **methylphenidate,** how many patients experienced **psychosis** within **1 year of initiating treatment**?
Patient level prediction	**Disease onset and progression**	For a given patient who is diagnosed with **<insert your favorite disease>**, what is the probability that they will go on to have **<another disease or related complication>** within **<time horizon from diagnosis>**?	For a given patient who is **newly diagnosed with atrial fibrillation**, what is the probability that they will go onto to have **ischemic stroke** in **next 3 years**?
	Treatment response	For a given patient who is a new user of **<insert your favorite chronically-used drug>**, what is the probability that they will **<insert desired effect>** in **<time window>**?	For a given patient **with T2DM who start on metformin** , what is the probability that they will **maintain HbA1C<6.5%** after **3 years**?
	Treatment safety	For a given patient who is a new user of **<insert your favorite drug>,** what is the probability that they will experience **<insert adverse event >** within **<time horizon following exposure>**?	For a given patients who is a **new user of warfarin**, what is the probability that they will have **GI bleed** in **1 year**?
Population-level effect estimation	**Safety surveillance**	Does exposure to **<insert your favorite drug>** increase the risk of experiencing **<insert an adverse event>** within **<time horizon following exposure start>**?	Does exposure to **ACE inhibitor** increase the risk of experiencing **Angioedema** within **1 month after exposure start**?
	Comparative effectiveness	Does exposure to **<insert your favorite drug>** have a different risk of experiencing **<insert any outcome (safety or benefit) >** within **<time horizon following exposure start>**, relative to **<insert your comparator treatment>**?	Does exposure to **ACE inhibitor** have a different risk of experiencing **acute myocardial infarction** while **on treatment**, relative to **thiazide diuretic**?

Let’s Dissect One Research Question

- How does the rate of side effects / local problems (including secondary / palliative treatments needed) compare between treatments (open, laparoscopic, robot surgery, with or without lymph node dissection; brachytherapy, different forms of external beam radiation therapy), and which patient specific factors are associated with these adverse secondary endpoints?

Let's Dissect One Research Question

- How does the rate of side effects / local problems (including secondary / palliative treatments needed) compare between treatments (open, laparoscopic, robot surgery, with or without lymph node dissection; brachytherapy, different forms of external beam radiation therapy), and which patient specific factors are associated with these adverse secondary endpoints?

- Characterization study: incidence rate

Amongst patients with **prostate cancer receiving different treatments**, how many patients experienced **side effects/local problems** within <time horizon >?

Let's Dissect One Research Question

- How does the rate of side effects / local problems (including secondary / palliative treatments needed) compare between treatments (open, laparoscopic, robot surgery, with or without lymph node dissection; brachytherapy, different forms of external beam radiation therapy), and which patient specific factors are associated with these adverse secondary endpoints?

- Population level estimation: comparative effectiveness

Comparative effectiveness: Does exposure to **treatment A** have a different risk of experiencing **side effects/local problems** within <time horizon > , relative to **treatment B**?

Let’s Dissect One Research Question

- How does the rate of side effects / local problems (including secondary / palliative treatments needed) compare between treatments (open, laparoscopic, robot surgery, with or without lymph node dissection; brachytherapy, different forms of external beam radiation therapy), and which patient specific factors are associated with these adverse secondary endpoints?

- Characterization study: natural history

Amongst patients with **prostate cancer receiving different treatment A-Z**, what are the patient’s characteristics from their medical history?

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OMOP and OHDSI

What have we learned?

10-Nov-2023

OHDSI's mission

To improve health

through community

and evidence

OHDSI's values

Innovation: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.

Reproducibility: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.

Community: Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.

Collaboration: We work collectively to prioritize and address the real world needs of our community's participants.

Openness: We strive to make all our community's proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.

Beneficence: We seek to protect the rights of individuals and organizations within our community at all times.

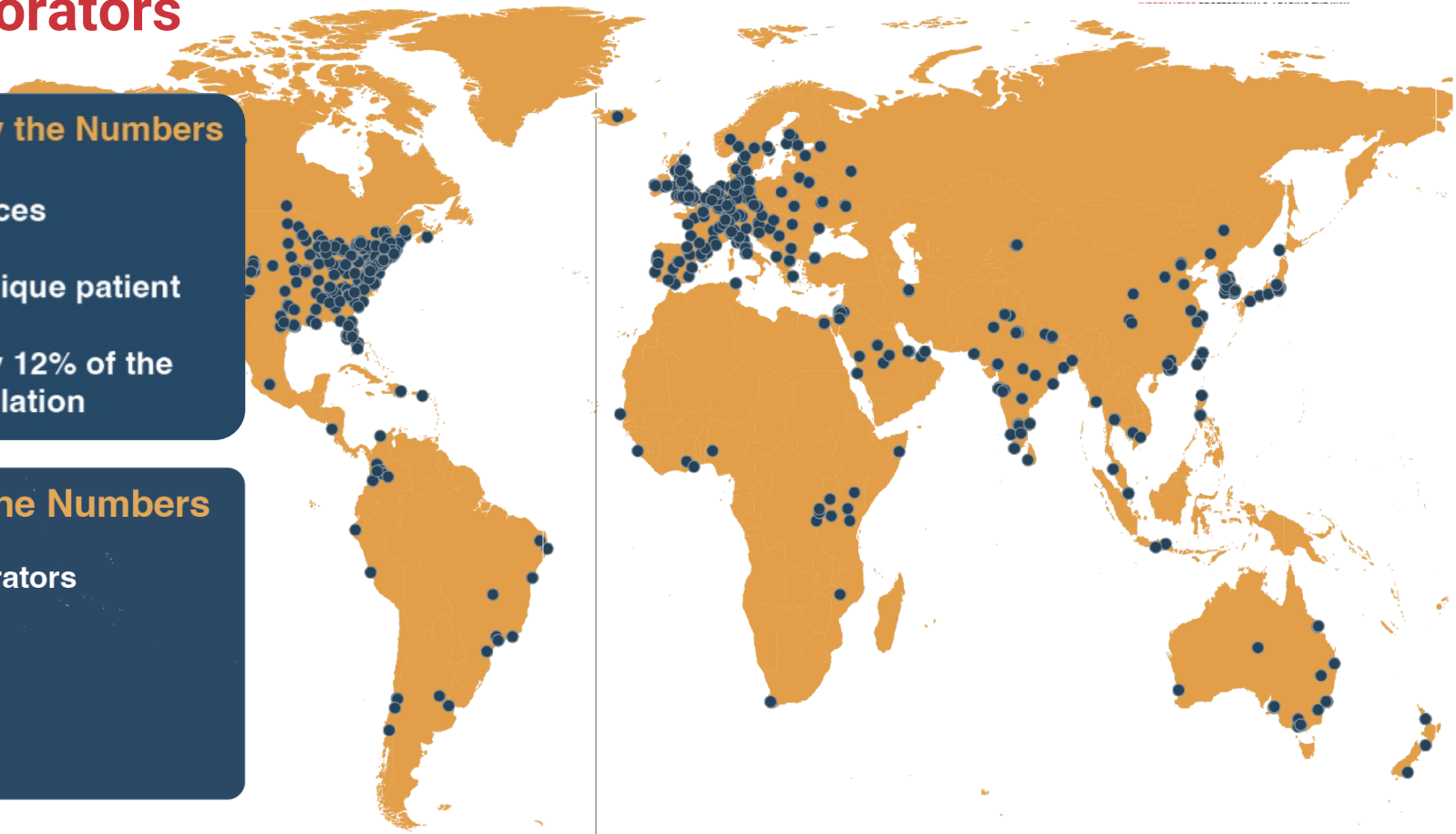
Collaborators

OMOP Data by the Numbers

- 534 data sources
- 49 countries
- 956 million unique patient records
- approximately 12% of the world's population

OHDSI By The Numbers

- 3,758 collaborators
- 83 countries
- 21 time zones
- 6 continents
- 1 community



The Secret Sources

Open-Source

Community

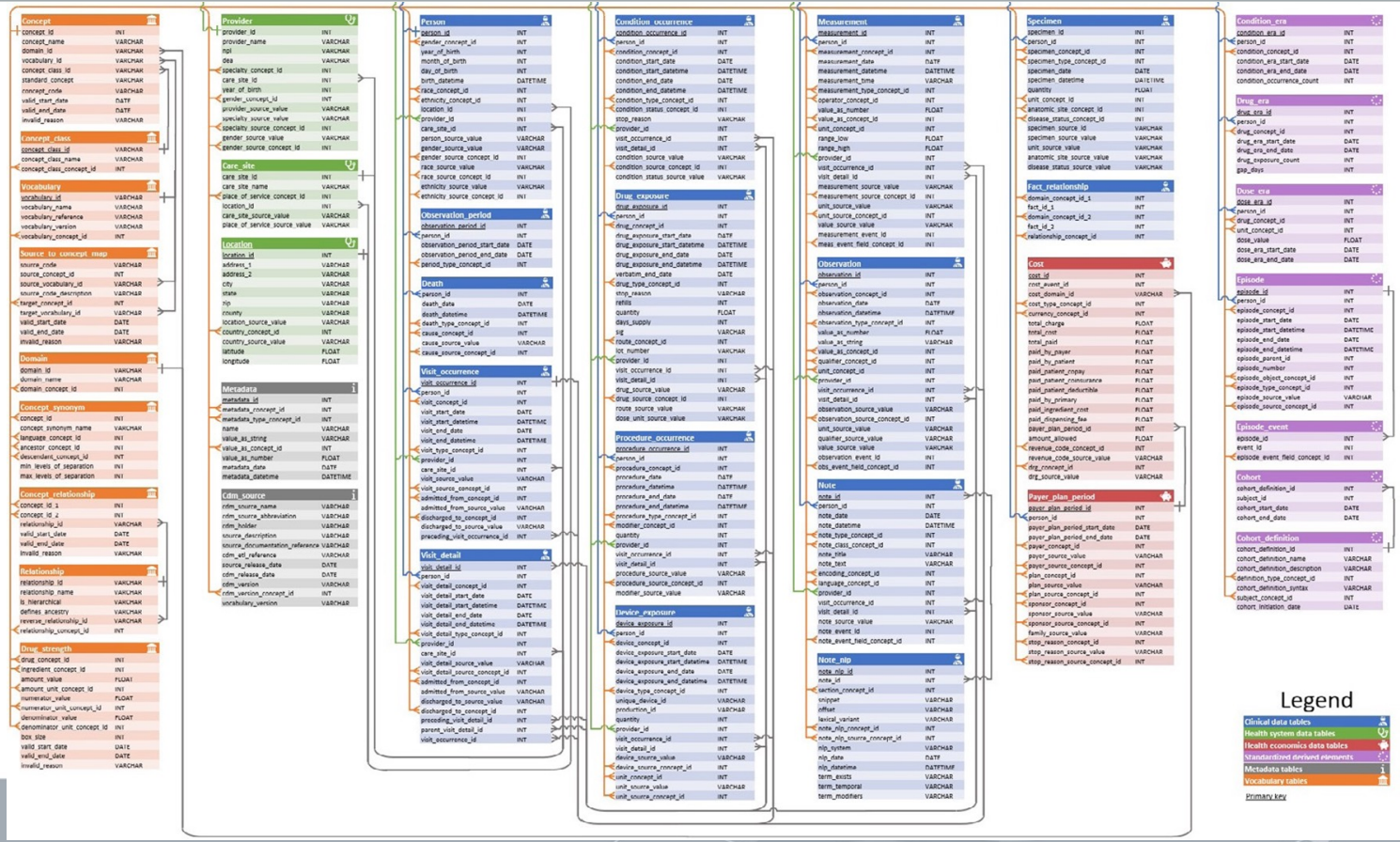
Standardization

Reproducibility

Open-Science

Scale

Data Standardization: OMOP CDM



OHDSI Scales

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020
Negative control outcomes	76	76
Target-comparator-neg controls	2,843,250 * 76=216,087,000	769,476
Positive control outcomes	76 * 3=228	228
Target-comparator-pos controls	2,843,250 * 228=648,261,00	662,484
Total comparisons	864,348,000	1,431,960
Total	864,348,000 * 9= 7,779,132,000	1,431,960 * 9=12,887,640

US Insurance databases

- IBM® MarketScan® CCAE
- IBM® MarketScan® MDCD
- IBM® MarketScan® MDCR
- Optum© Clinformatics®

Japanese insurance databases

- Japan Medical Data Center

Korean national insurance databases

- NHIS-NSC

US EHR databases

- Columbia University Medical Center
- Optum© PANTHER®

German EHR databases

- QuintilesIMS Disease Analyzer (DA) Germany

Open Source Community

262 Repositories

30 M+ lines of code

681 Developers

31 organizations

47,672 commits

2,838 GitHub Forks

4,168 GitHub Stars

5,547 GitHub Subscribers

HADES | Packages | Validation | Publications | Support | Study packages | Developers

Package	Version	Maintainer(s)	Availability	Open issues	pull-requests	Build status	Coverage
Achilles	v1.7.2	Frank DeFalco	CRAN	29	3	R check passing	codecov 2%
Andromeda	v0.6.3	Adam Black	CRAN	13	2	R check passing	codecov 89%
BigKnn	v1.0.2	Martijn Schuemie	GitHub	0	0	R check passing	codecov 96%
BrokenAdaptiveRidge	v1.0.0	Marc Suchard	CRAN	2	0	R check passing	codecov 96%
Capr	v2.0.7	Martin Lavallee	GitHub	2	0	R check passing	codecov 80%
Characterization	v0.1.2	Jenna Reps	GitHub	11	1	R check failing	codecov unknown
CirceR	v1.3.1	Chris Knoll	GitHub	3	1	R check passing	codecov 89%
CohortDiagnostics	v3.2.4	Jamie Gilbert	GitHub	56	2	R check passing	codecov 90%
CohortExplorer	v0.1.0	Gowtham Rao	CRAN	0	0	R check passing	codecov 100%
CohortGenerator	v0.8.1	Anthony Sena	GitHub	19	2	R check passing	codecov 98%
CohortMethod	v5.1.0	Martijn Schuemie	GitHub	14	1	R check failing	codecov 88%
Cyclops	v3.3.1	Marc Suchard	CRAN	18	0	R check failing	codecov 85%
DatabaseConnector	v6.2.4	Martijn Schuemie	CRAN	12	0	R check passing	codecov 59%
DataQualityDashboard	v2.4.1	Katy Sadowksi	GitHub	43	7	R check passing	codecov 86%
DeepPatientLevelPrediction	v2.0.0	Egill Fridgeirsson	GitHub	18	1	R check failing	codecov 100%
EmpiricalCalibration	v3.1.1	Martijn Schuemie	CRAN	1	0	R check passing	codecov 84%
EnsemblePatientLevelPrediction	v1.0.2	Jenna Reps	GitHub	5	0	R check passing	codecov unknown
Eunomia	v1.0.2	Frank DeFalco	GitHub	10	1	R check passing	codecov 72%
EvidenceSynthesis	v0.5.0	Martijn Schuemie	CRAN	3	0	R check passing	codecov 78%
FeatureExtraction	v3.3.1	Anthony Sena	GitHub	44	6	R check failing	codecov 93%
Hydra	v0.4.0	Anthony Sena	GitHub	6	7	R check failing	codecov 87%
IterativeHardThresholding	v1.0.2	Marc Suchard	CRAN	1	0	R check passing	codecov 92%
MethodEvaluation	v2.3.0	Martijn Schuemie	GitHub	1	0	R check passing	codecov 59%
OhdsiSharing	v0.2.2	Lee Evans	GitHub	0	1	R check passing	codecov 0%
OhdsiShinyModules	v2.0.0	Jenna Reps	GitHub	108	2	R check failing	codecov 79%
ParallelLogger	v3.3.0	Martijn Schuemie	CRAN	4	0	R check passing	codecov 81%
PatientLevelPrediction	v6.3.6	Jenna Reps & Peter Rijnbeek	GitHub	47	0	R check failing	codecov 89%
PhenotypeLibrary	v1.30.1	Gowtham Rao	GitHub	1	0	R check passing	codecov 100%

OHDSI Methods and Research

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Three ideas

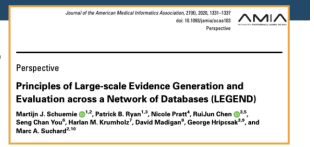
1. LEGEND Principles

METHODS RESEARCH

LEGEND in Principle

LEGEND (Large-scale Evidence Generation and Evaluation across a Network of Databases) applies high-level analytics to perform observational research on hundreds of millions of patient records within OHDSI's international database network.

LEGEND is based on 10 guiding principles that were published in JAMIA (August, 2020) and are listed below.


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

AMIA

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], Rui Jun Chen[5,6], Seng Chan You[7], Harlan M. Krumholz[1], David Madigan[8], George Hripcsak[3,9], and Marc A. Suchard[2,10]

1. **LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). Aim: Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.

2. **Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. Aim: Avoids publication bias and enhances transparency.

3. **LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. Aim: Avoids P hacking.

4. **LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. Aim: Avoids P hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).

5. **LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. Aim: Minimizes bias.

6. **LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our P values, effect estimates, and confidence intervals using empirical calibration. [7,8] Aim: Enhances transparency on the uncertainty due to residual bias.

7. **LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. Aim: Enhances transparency and allows replication.

8. **LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). Aim: Avoids bias and improves interpretability.

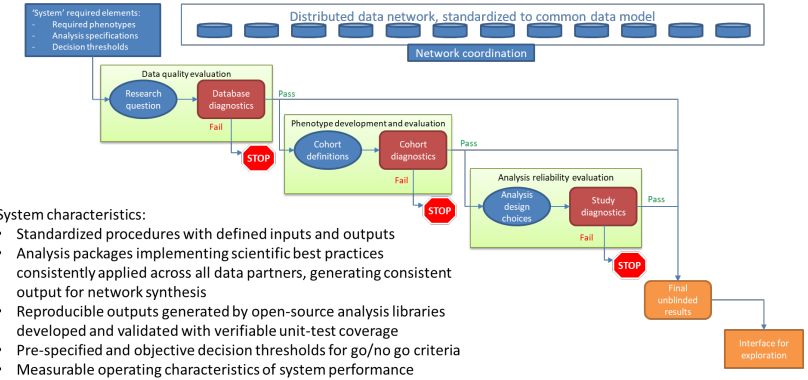
9. **LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. Aim: Enhances generalizability and uncovers potential between-site heterogeneity.

10. **LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. Aim: Privacy.

#JoinTheJourney 47 OHDSI.org

2. Objective Diagnostics

Engineering open science systems that build trust into the real-world evidence generation and dissemination process



System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

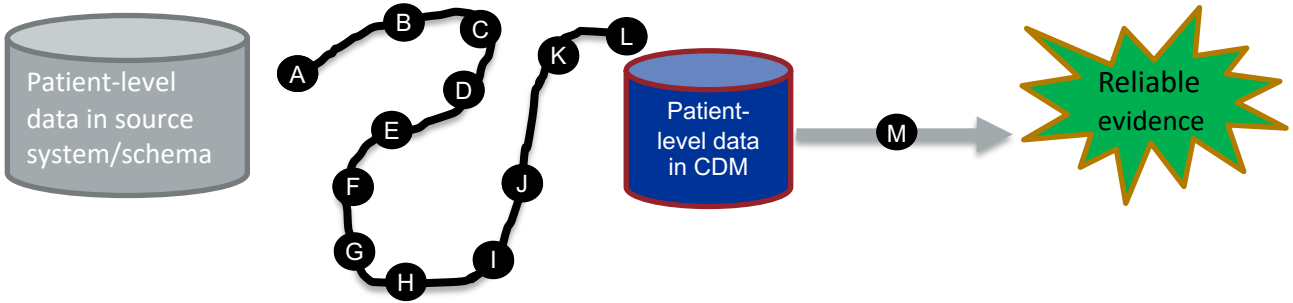
3. Standardized software

HADES
HADES HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Open-source software

Reproducibility and Trust

We need to make studies repeatable, reproducible, replicable, generalisable, and robust



A Common Data Model enables standardised analytics to generate reliable evidence.

Where do you find us?

Home: https://www.ohdsi.org

Book of OHDSI: https://book.ohdsi.org

Methods and Tools: https://github.com/OHDSI

Common Data Model: https://ohdsi.github.io/CommonDataModel

Vocabularies: https://athena.ohdsi.org

Studies: https://github.com/OHDSI/ohdsistudies

Workgroups: https://www.ohdsi.org/workgroups

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Join the Journey!