



11/11/2025

# A CRASH COURSE IN OHDSI

---

Renske Los, PhD

Deputy Head of Medical Informatics, Erasmus MC

National Nodes Coordinator, OHDSI Europe

Co-Lead OHDSI.NL



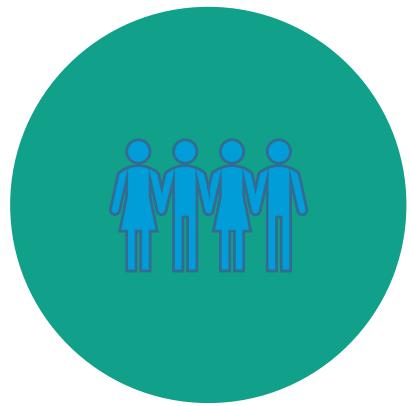
# IN THE NEXT 20 MINUTES...



WHAT IS OHDSI?  
WHAT IS OMOP?



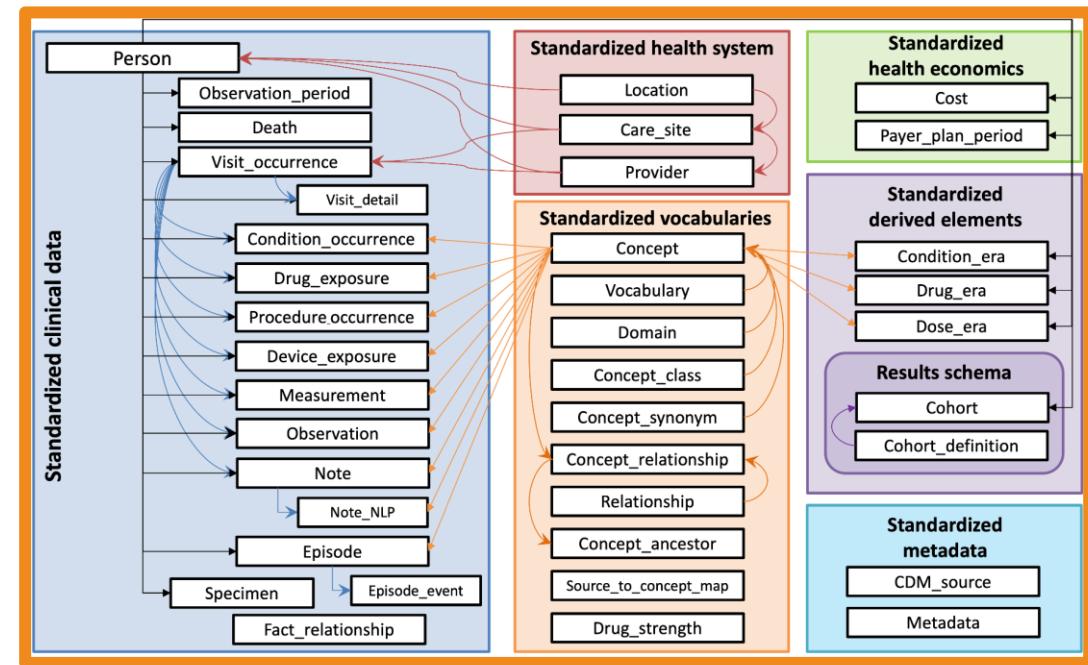
WHAT'S IN IT FOR  
YOU?



HOW TO GET  
INVOLVED



# WHAT IS OHDSI?



OMOP Common Data Model

Standard analytics

A large blue question mark icon is positioned above two rectangular boxes. The left box contains the ATLAS logo, which features a stylized figure with arms raised inside a circle, with the word "ATLAS" below it. The right box contains the HADES logo, which features a yellow and orange gradient background with a classical building silhouette and the word "HADES" in white capital letters.

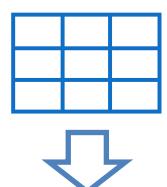
A large group photograph of a diverse community of people, mostly young adults, gathered in a large auditorium or conference room. They are standing in multiple rows, smiling for a group photo. The background shows rows of red theater-style seats and some equipment on stage.

Active global community

*OHDSI is a fun way to collaborate with an amazing community across the globe to collectively advance science and improve the lives of patients around the world.*



# IF YOU DO NOT HAVE OMOP, RESEARCH BE LIKE...



**Why Most Published Research Findings Are False**

**Abstract** The probability that a research claim is true may depend on factors such as the ratio of true to false hypotheses being tested. This ratio is commonly unacceptably high in many active fields. Therefore, it is likely that most current published results are false.

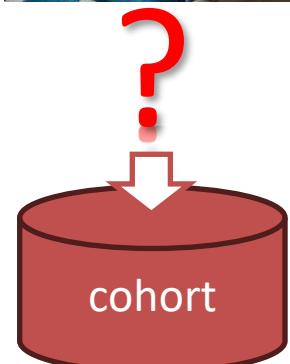
**Introduction** In this paper, I will argue that most current published results are false. My argument is based on the ratio of true to false hypotheses that are typically tested. I will show that this ratio is often unacceptably high in many active fields. As a result, it is likely that most current published results are false.

**1. Introduction** In this paper, I will argue that most current published results are false. My argument is based on the ratio of true to false hypotheses that are typically tested. I will show that this ratio is often unacceptably high in many active fields. As a result, it is likely that most current published results are false.

**2. The Problem** The problem is that most current published results are false. This is because the ratio of true to false hypotheses that are typically tested is often unacceptably high in many active fields. As a result, it is likely that most current published results are false.

**3. The Solution** The solution is to change the way we do science. We need to change the way we do science so that the ratio of true to false hypotheses that are typically tested is acceptably low. This will help ensure that most current published results are true.

**4. Conclusion** In conclusion, most current published results are false. This is because the ratio of true to false hypotheses that are typically tested is often unacceptably high in many active fields. As a result, it is likely that most current published results are false.



**Why Most Published Research Findings Are False**

John P. A. Ioannidis  
Stanford University

**Abstract**

Most current published research results are likely to be true findings for no other reason than that studies are far more likely to be true than false. The probability that a research claim is true may depend on study power and important effect size, but may also depend on study design and analysis. In this framework, an analysis of current research performance shows that true new research claims are less likely to be made in this sample of studies. This is particularly so when the studies are less well designed and powered, and when effect sizes are smaller. Thus, the current situation is not sustainable and major changes are likely.

**Keywords:** scientific method; scientific inference; scientific validity; scientific reliability; scientific reproducibility; scientific credibility; scientific validity; scientific reliability; scientific reproducibility; scientific credibility

**Introduction**

It is common for researchers to claim that their results are true. However, it is not always clear what they mean by this claim. In this article, I will argue that most published research results are likely to be true findings for no other reason than that studies are far more likely to be true than false. The probability that a research claim is true may depend on study power and important effect size, but may also depend on study design and analysis. In this framework, an analysis of current research performance shows that true new research claims are less likely to be made in this sample of studies. This is particularly so when the studies are less well designed and powered, and when effect sizes are smaller. Thus, the current situation is not sustainable and major changes are likely.

**1. Introduction**

It is common for researchers to claim that their results are true. However, it is not always clear what they mean by this claim. In this article, I will argue that most published research results are likely to be true findings for no other reason than that studies are far more likely to be true than false. The probability that a research claim is true may depend on study power and important effect size, but may also depend on study design and analysis. In this framework, an analysis of current research performance shows that true new research claims are less likely to be made in this sample of studies. This is particularly so when the studies are less well designed and powered, and when effect sizes are smaller. Thus, the current situation is not sustainable and major changes are likely.

**2. The Probability That a Research Claim Is True**

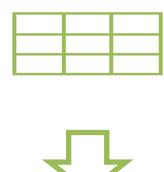
It has been argued that most claimed research results are true. *1* This is not because most studies are well designed and well analyzed. Rather, it is because it is far more likely for a true new finding to be true than for a false new finding to be false. In this framework, the probability that a research claim is true may depend on study power and important effect size, but may also depend on study design and analysis.

**3. Current Research Performance**

An analysis of current research performance shows that true new research claims are less likely to be made in this sample of studies. This is particularly so when the studies are less well designed and powered, and when effect sizes are smaller. Thus, the current situation is not sustainable and major changes are likely.

**4. Conclusions**

The current situation is not sustainable and major changes are likely.



**Why Most Published Research Findings Are False**

John P. A. Ioannidis

**Abstract** The probability that a research claim is true may depend on study power and number of other studies on the same question. This framework is used to discuss the current situation in most of the current large-scale multiple testing settings in biostatistics, medical research, and other sciences. It is shown that for most current large-scale studies, there is a very low probability that a research claim is true. The expected proportion of true null hypotheses may be low, as well as the fraction of true discoveries. This is the result of low study power and the relative emphasis on P values rather than effect sizes and other statistical markers of true effects. This framework is used to discuss current validation studies and the need for a paradigm shift in the validation of new biomarkers and other research claims. © 2005 by John P. A. Ioannidis

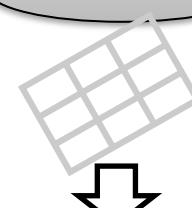
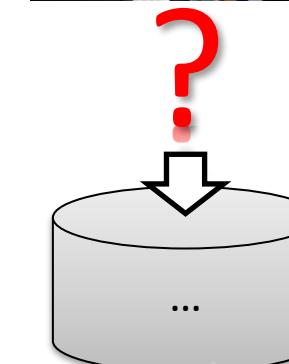
**Keywords:** multiple comparisons; P value; study power; validation

**Introduction** In this article, I will argue that most current published research results are likely to be false. The probability that a research claim is true may depend on study power and number of other studies on the same question. This framework is used to discuss the current situation in most of the current large-scale multiple testing settings in biostatistics, medical research, and other sciences.

**Current Situation in Large-Scale Studies** The situation in most current large-scale studies is that many claims are made on the basis of studies that are underpowered to detect the true effects. This is the case in genome-wide association studies, in proteomic studies, in studies of gene expression, in pharmacogenomics, in studies of complex diseases, and in other areas of large-scale studies. The problem is that the number of studies on the same question is often large, and the number of true null hypotheses is also often large. This leads to a situation where it is very difficult to identify true effects. The situation is even worse if the true effects are small.

**Validation Studies** Validation studies are often used to validate new biomarkers or other research claims. However, the validation of new biomarkers or other research claims is often based on studies that are underpowered to detect the true effects. This is the case in many validation studies. The problem is that the number of studies on the same question is often large, and the number of true null hypotheses is also often large. This leads to a situation where it is very difficult to identify true effects.

**Conclusion** The current situation in most of the current large-scale studies is that many claims are made on the basis of studies that are underpowered to detect the true effects. This is the case in genome-wide association studies, in proteomic studies, in studies of gene expression, in pharmacogenomics, in studies of complex diseases, and in other areas of large-scale studies. The problem is that the number of studies on the same question is often large, and the number of true null hypotheses is also often large. This leads to a situation where it is very difficult to identify true effects.



**Editor's Note**

## Why Most Published Research Findings Are False

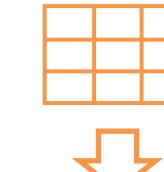
John P. Ioannidis

**Abstract**

Most current published research results are likely to be true findings for no other reason than their random occurrence. The probability that a research claim is true may depend on study power and importance of the effect size, as well as on ratio of true to no relationships probed in each individual study. In this framework, many current publication practices are likely to result in most current published research results being false. The probability that a research claim is true may depend on study power and importance of the effect size, as well as on ratio of true to no relationships probed in each individual study. In this framework, many current publication practices are likely to result in most current published research results being false.

**Keywords:** scientific method; hypothesis testing; statistical power; multiple comparisons; effect size; research validity

**Journal of the American Statistical Association**, Vol. 100, No. 470, pp. 427–436, June 2005  
© 2005 American Statistical Association  
DOI 10.1198/016214504178818835



**Why Most Published Research Findings Are False**

John P. Ioannidis

**Abstract** *Most current published research results are likely true for no more than half of the studies. The probability that a research claim is true may depend on study power and important effect size, as well as on bias factors such as multiple comparisons and publication bias. In this framework, an analysis of 49 pion-*

**Introduction** *Most current published research results are likely true for no more than half of the studies. The probability that a research claim is true may depend on study power and important effect size, as well as on bias factors such as multiple comparisons and publication bias. In this framework, an analysis of 49 pion-*

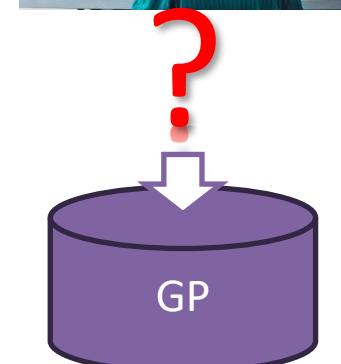
**Keywords:** *multiple comparisons, publication bias, study power, effect size, false positive rate*

**Author's Note:** *This work was partially funded by National Institutes of Health grants R01GM070332 and R01GM070332-02S1.*

**Address reprint requests to:** *John P. Ioannidis, Department of Health Research and Policy, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94301-5822. E-mail: jioannid@stanford.edu*

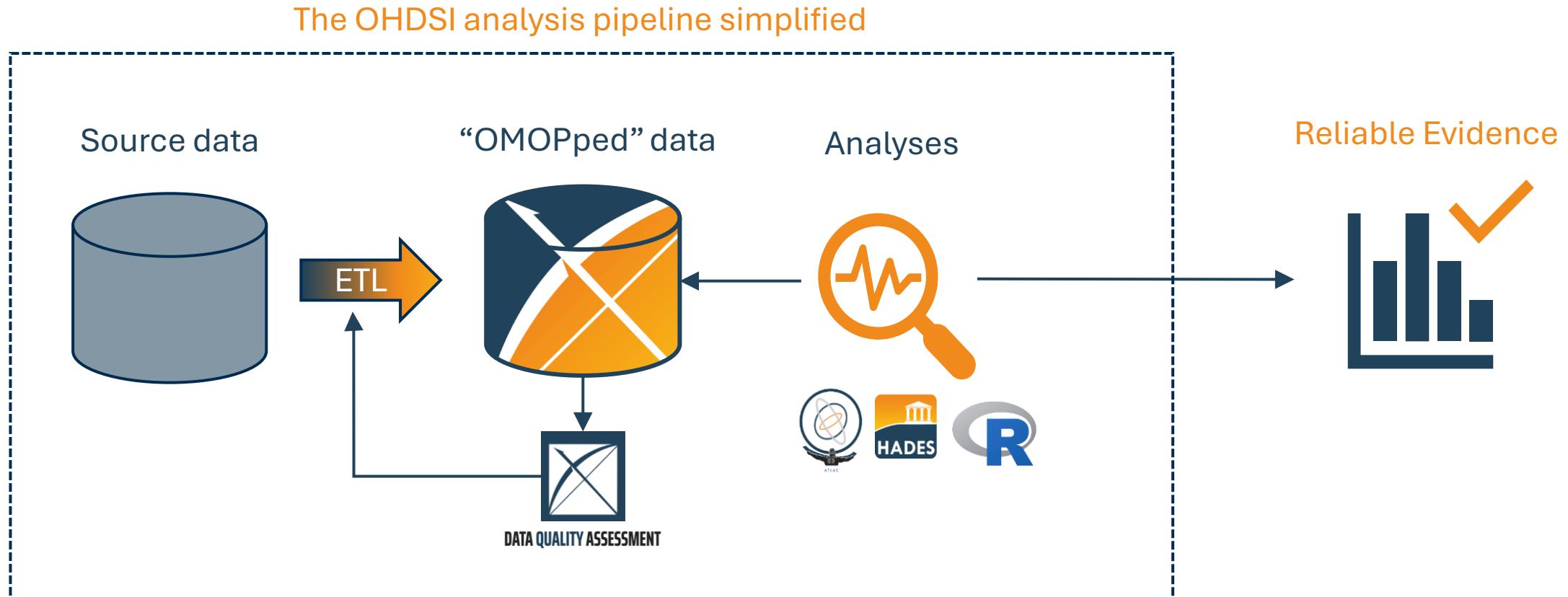
**Received June 20, 2005; accepted July 20, 2005.**

**Copyright © 2005, American Psychological Association or Society for Psychophysiology and Behavioral Neuroscience. 0882-7959/05/\$12.00 DOI: 10.1037/0882-7959.41.3.333**





# IF YOU HAVE OMOP, RESEARCH BE LIKE...





# REAL-WORLD EVIDENCE GENERATION USING OMOP CDM

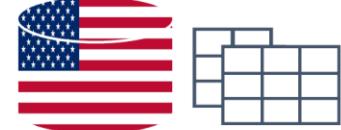
Different structure  
& terminology



e.g. IPCI



e.g. CPRD



e.g. CCAE

Standardized data



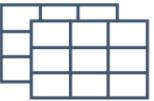
ETL



Share aggregated  
results



ETL



ETL



Conversion to  
OMOP CDM



Run standard  
analytic tools locally

**Standardized data enables standardized analytics!**

Common format  
& terminology  
Single curation,  
multiple use

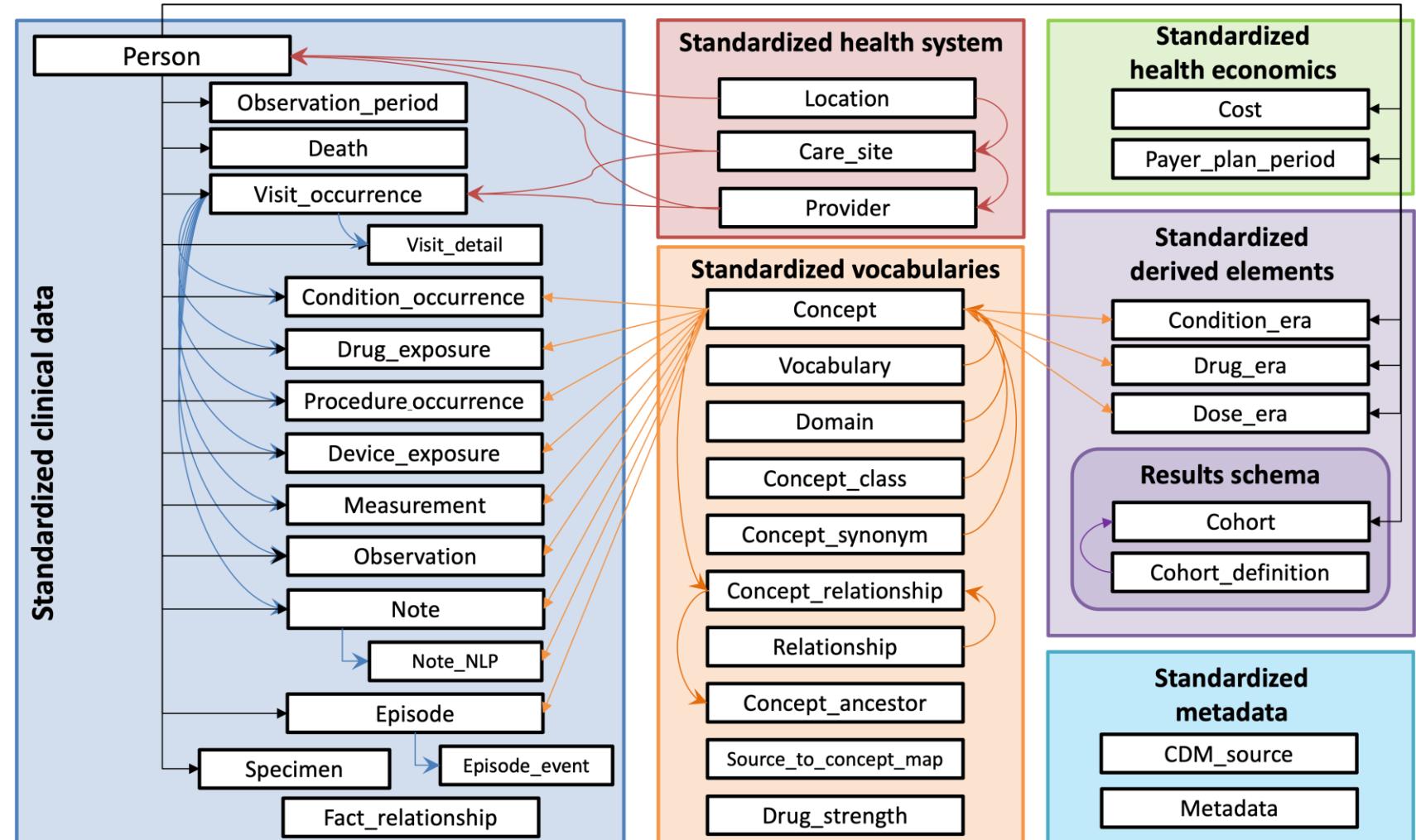
Standard input & output  
Parametrized choices

# OMOP COMMON DATA MODEL (CDM) v5.4



OMOP CDM

Syntactic  
interoperability



Semantic  
interoperability



# BENEFITS OF STANDARDIZED DATA & ANALYTICS

## 1. Large-scale evidence

Federated analysis allows multi-database studies without compromising privacy

## 2. Rapid response

Large time savings through data readiness and pre-developed tooling

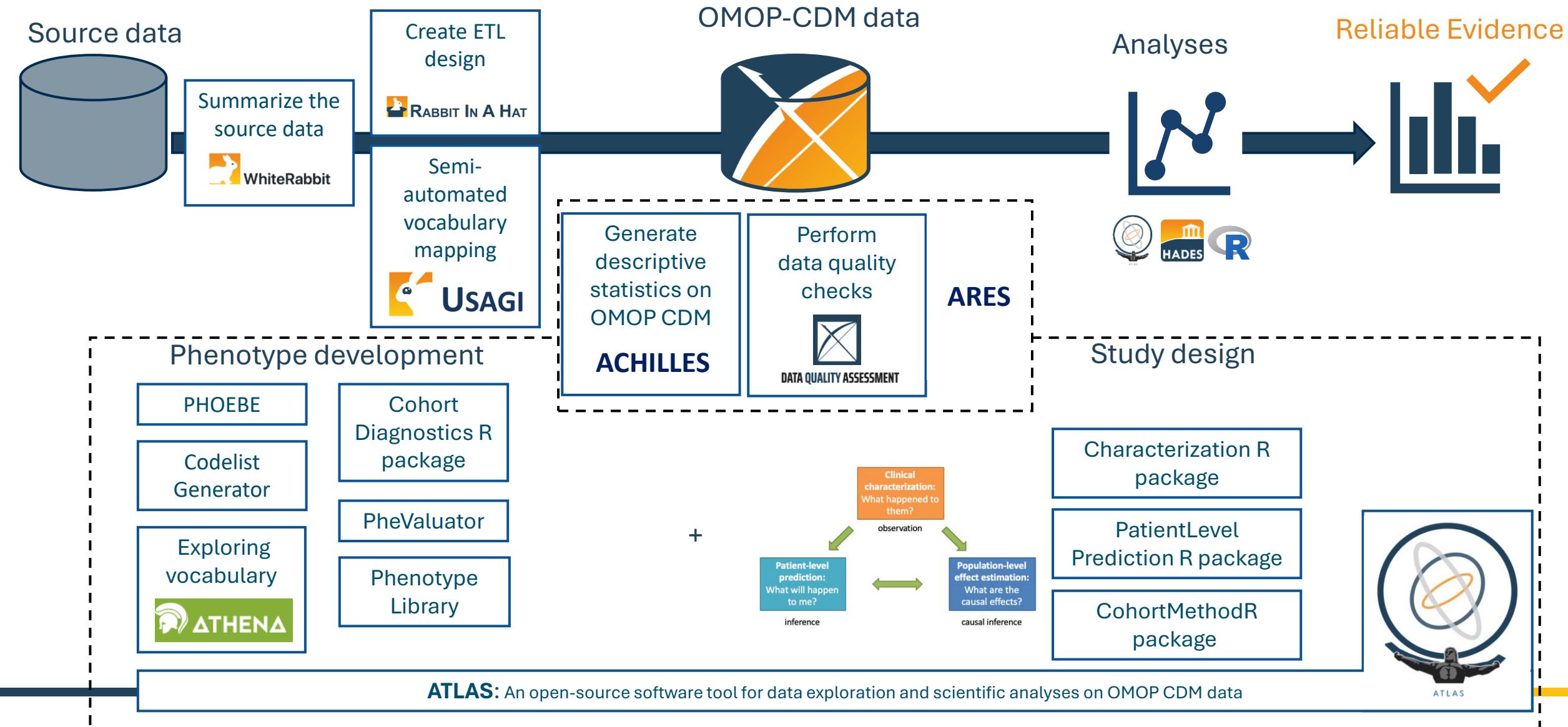
## 3. Scalable & reproducible

Tooling can be reused across databases and disease domains

→ Reliable evidence



# OHDSI TOOLS (OPEN-SOURCE)





**WHAT'S IN IT FOR YOU?**

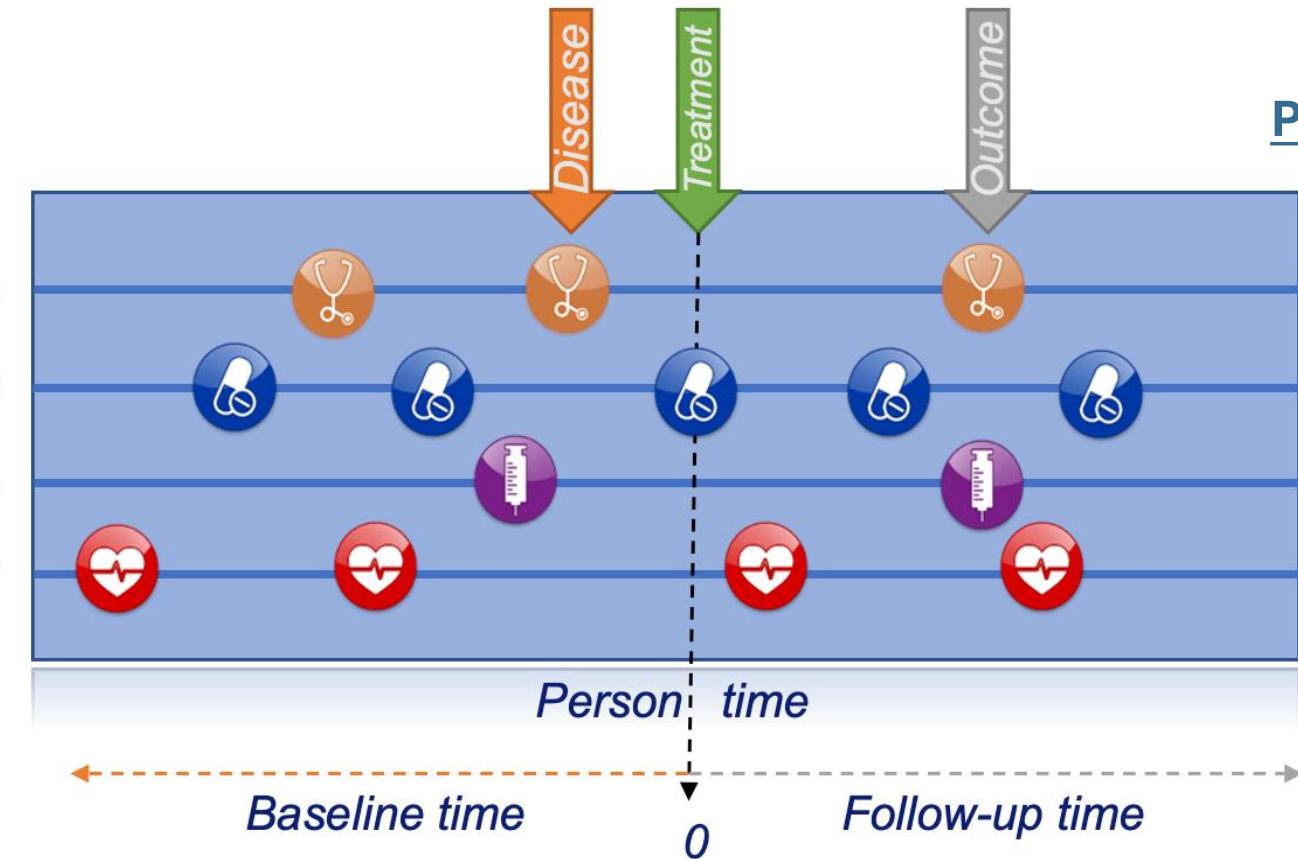
# QUESTIONS ASKED ACROSS THE PATIENT JOURNEY

## Clinical characterization

Which patients take which treatments?

How many patients experienced complications?

Conditions  
Drugs  
Procedures  
Measurements



## Patient-level prediction

What is the probability that I will develop a given disease?

What is the probability that I will experience an (adverse) outcome?

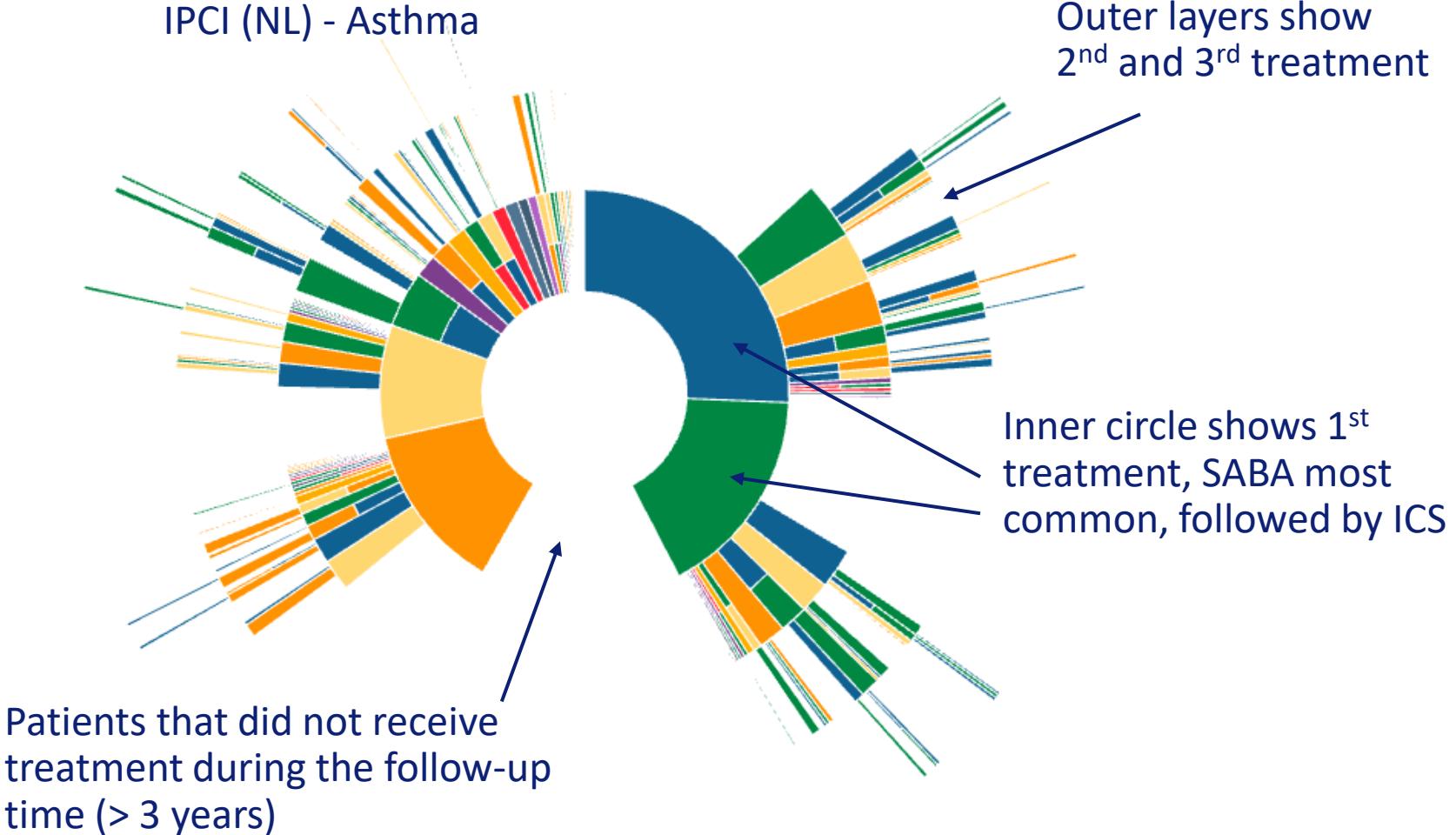
## Population-level effect estimation

Does treatment cause an outcome?

Does one treatment lead to a better outcome than the alternative?



# TREATMENT PATTERNS



ICS
LABA
LABA&ICS
LABA&LAMA
LABA&LAMA&ICS
LAMA
LTRA
SABA
SABA&SAMA
SAMA
Systemic glucocorticoids (acute)
Systemic glucocorticoids (therapy)
Xanthines
Anti IgE
Anti IL5
Systemic B2 agonist
PDE4
Other combinations



# TREATMENT PATTERNS ACROSS DISEASE DOMAINS

- Growing number of studies (also used in DARWIN EU® now)

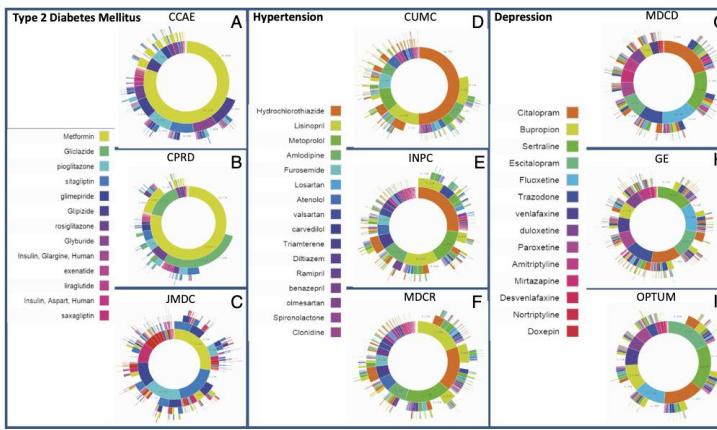


Fig. 3. For each disease, diabetes (A-C), hypertension (D-F), and depression (G-I), the inner circle shows the first relevant medication that the patient took, the second circle shows the second medication, and so forth. Three data sources are shown for each disease; the data source abbreviations are defined in Table 2.

Hripcak et al. (2016)

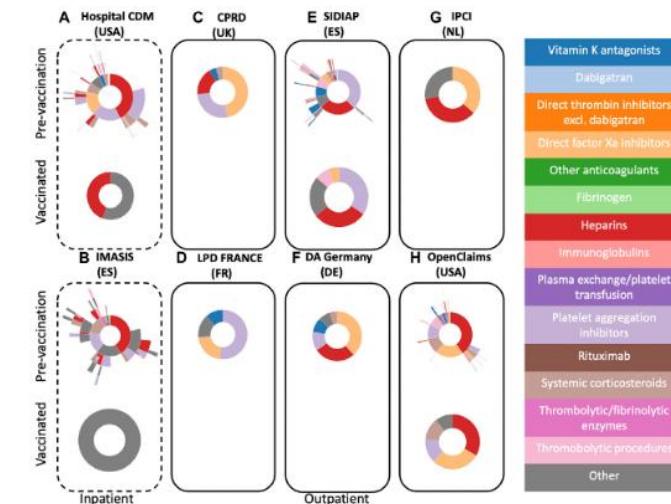
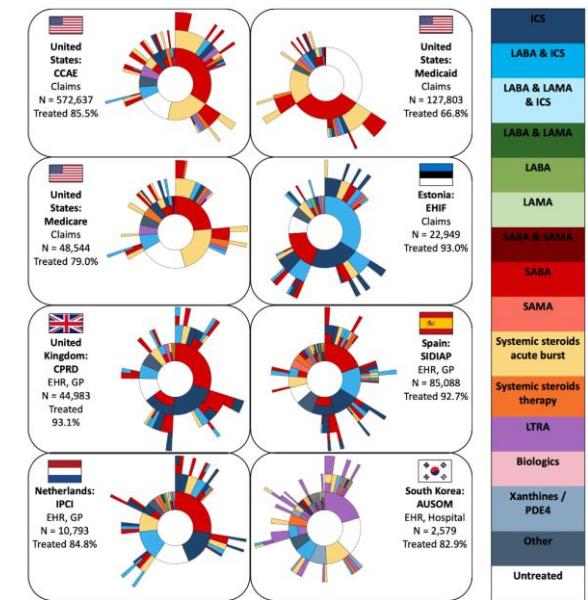


FIGURE 1  
Sunburst plots visualizing treatment pathways for TTS patients in pre-vaccination cohorts (top) versus vaccinated cohorts (bottom). Inpatient databases are depicted with a dashed line frame, whilst outpatient ones have a solid frame.

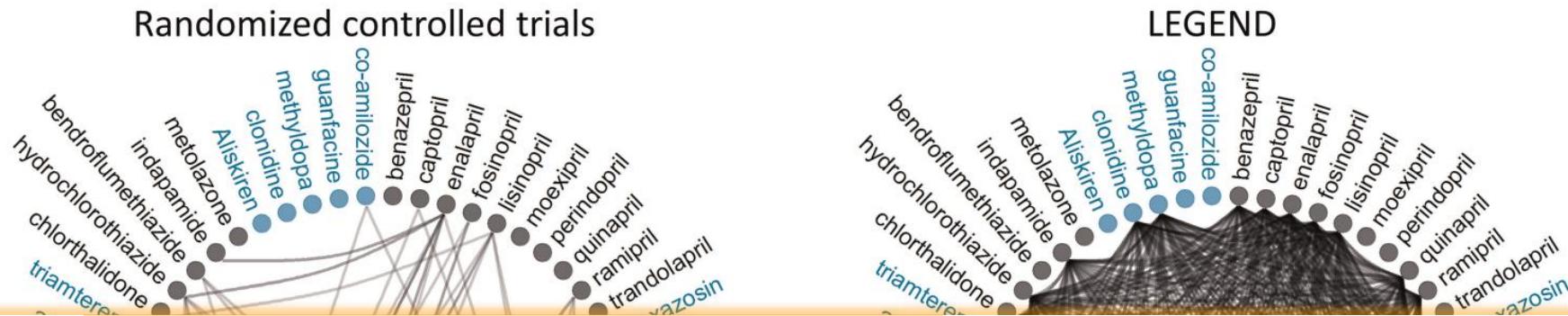
Markus et al. (2023)



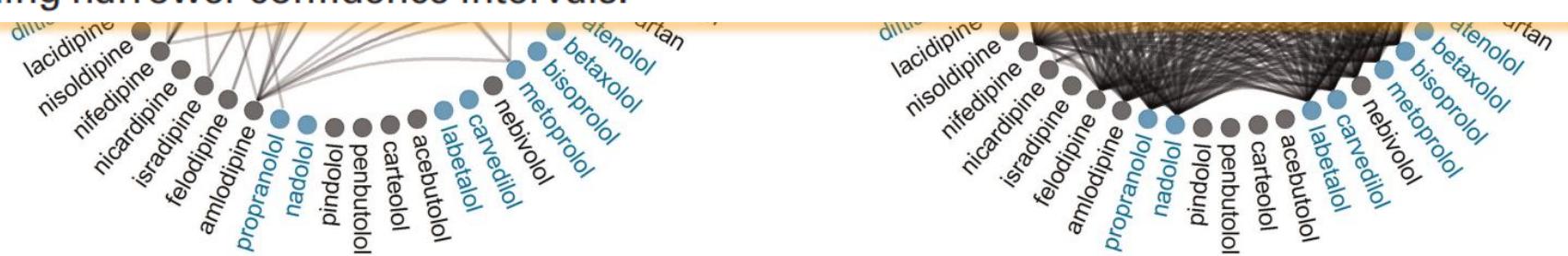
Markus et al. (2024)

- **A lot of potential:** use the OHDSI data network to get high-level insight in current treatment practices across different *patient populations, care settings, and countries* in Europe (and globally)

# OHDSI LEGEND HYPERTENSION STUDY



**Results:** From 21.6 million unique antihypertensive new users, we generate 6 076 775 effect size estimates for 699 872 research questions on 12 946 treatment comparisons. Through propensity score matching, we achieve balance on all baseline patient characteristics for 75% of estimates, observe 95.7% coverage in our effect-estimate 95% confidence intervals, find high between-database consistency, and achieve transitivity in 84.8% of triplet hypotheses. Compared with meta-analyses of RCTs, our results are consistent with 28 of 30 comparisons while providing narrower confidence intervals.



**Figure 3.** Comparisons of single-drug hypertension treatments in randomized controlled trials (left) and in LEGEND (right). Each circle represents an ingredient. Color groupings indicate drug classes. A line between circles indicates the 2 drugs are compared in at least 1 study.



# A GLOBAL NETWORK

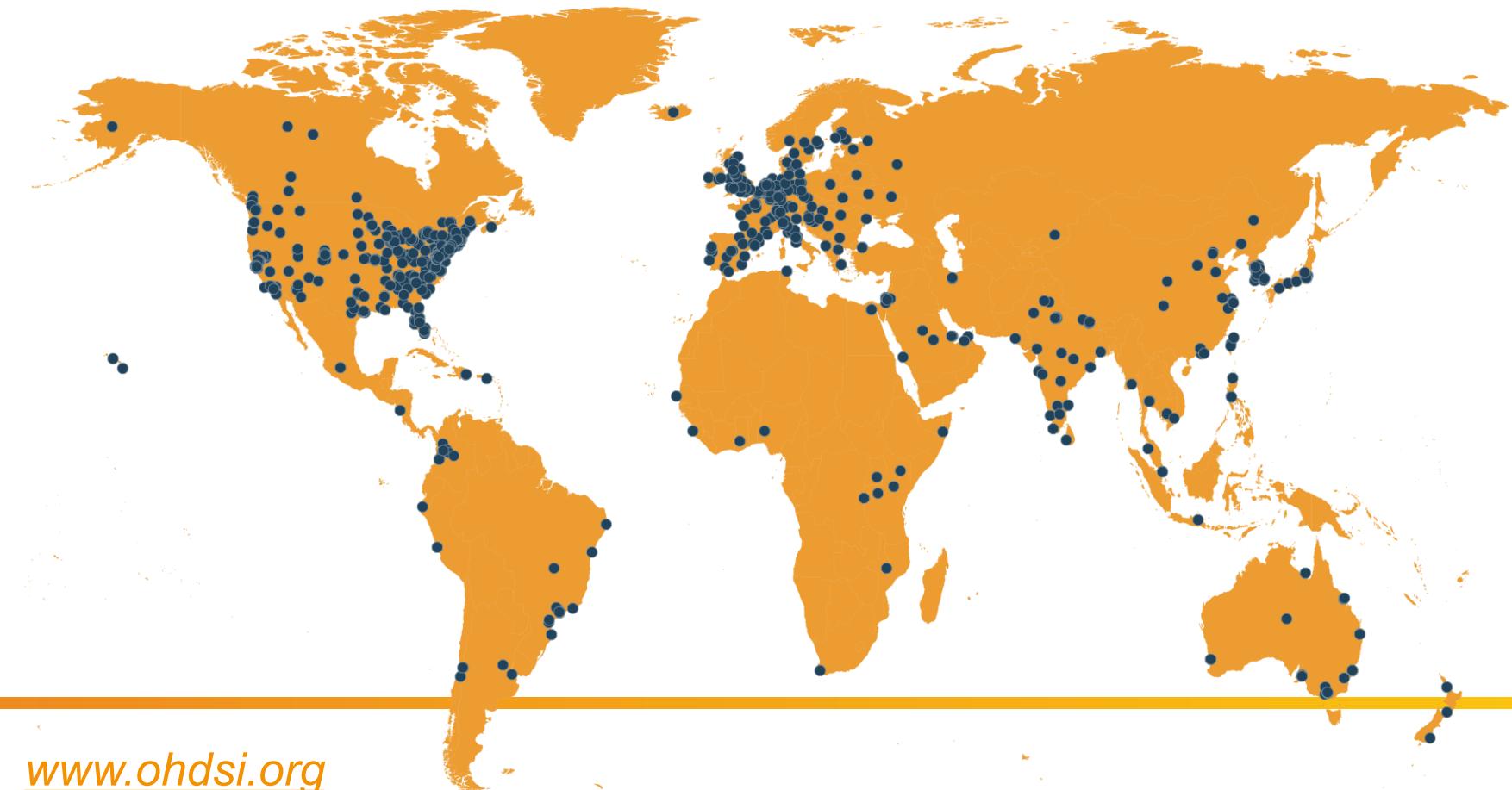
*To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care*

## OHDSI Collaborators

- 4,294 collaborators
- 83 countries
- 21 time zones
- 6 continents

## OMOP CDM Users By The Numbers

- 544 data sources
- 54 countries
- 974 million unique patient records  
(12% of world's population)

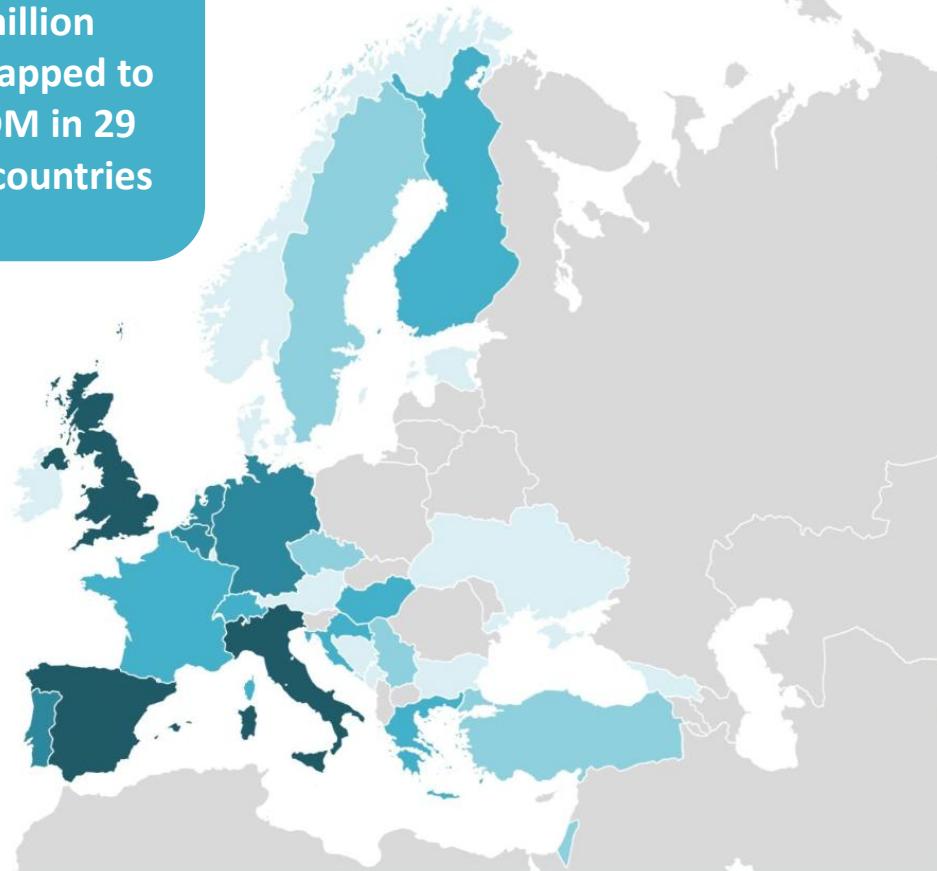




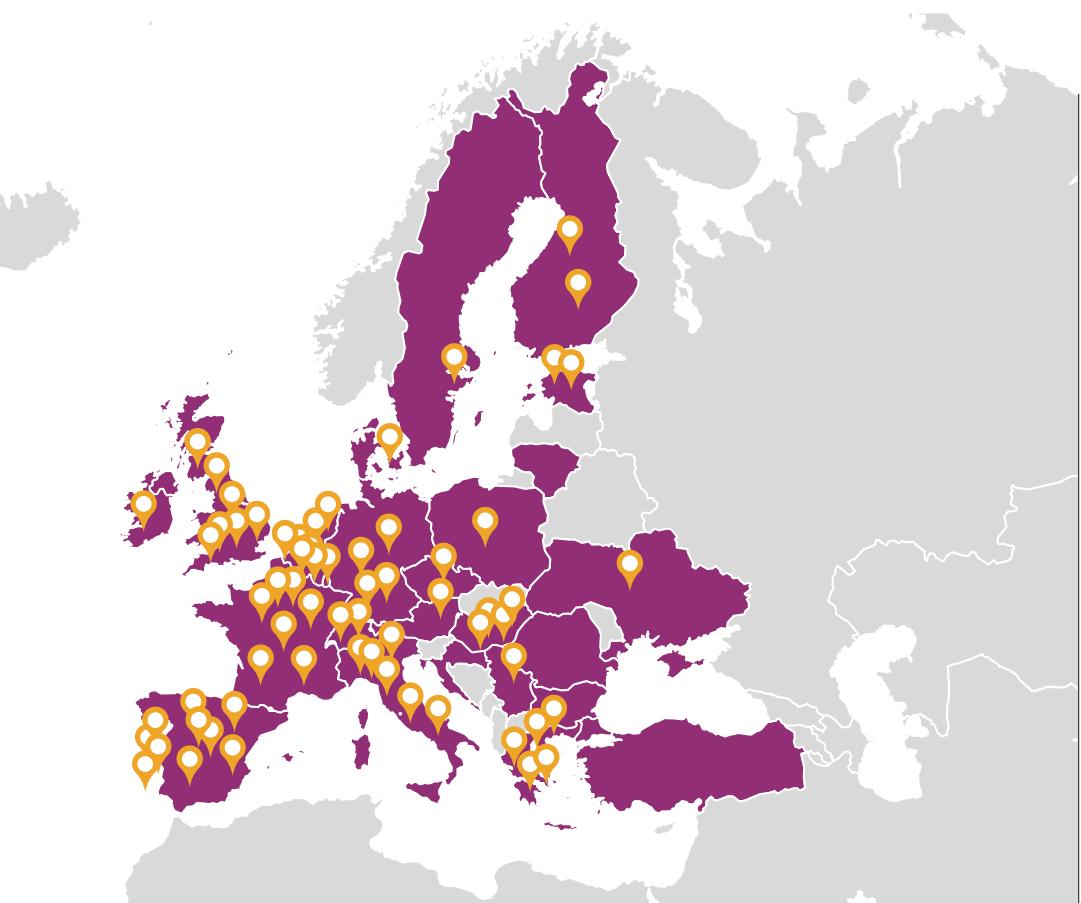
# NETWORK DATA PARTNERS & SERVICE PROVIDERS

187 Data partners

~ 370 million  
patients mapped to  
OMOP CDM in 29  
European countries



64 'EHDEN-certified' SMEs across Europe



<https://www.ehdenn.eu/business-directory/>

# HOW TO GET INVOLVED OR LEARN MORE



# EU NATIONAL Nodes

An OHDSI Europe National Node is a collection of organizations within a member country.

The Node builds on the strengths of the **stakeholders and scientific communities** of that country.

The goal of national nodes is to facilitate national and international collaborations.

National nodes were started as part of the EHDEN project and in collaboration with the OHDSI Europe Chapter.

<https://ohdsi-europe.org/index.php/national-nodes>

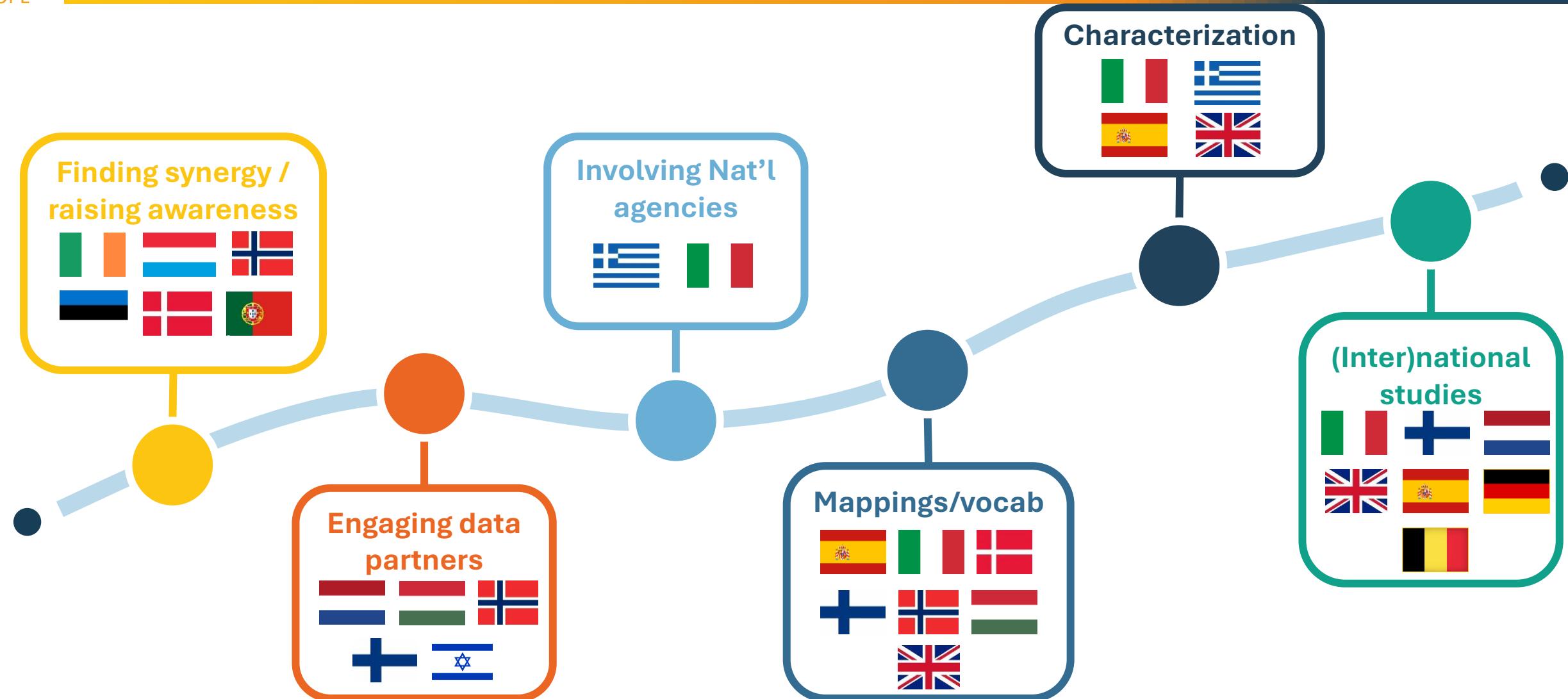


© GeoNames, Microsoft, Open Places, OpenStreetMap, TomTom





# WHATS GOING ON IN THE NODES





# OHDSI WORKGROUPS



# OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Who We Are ▾ Updates & News ▾ Standards Software Tools ▾ Network Studies ▾ Community Forums ▾ Education

Community Calls ▾ Past Events ▾ Workgroups ▾ 2024 'Our Journey' Annual Report Current Events ▾ Support & Sponsorships

2025 Europe Symposium 2025 Global Symposium Learn About Our Workgroups Opportunities GitHub YouTube Twitter LinkedIn

Join Our Teams Environment  
Join Our Workgroups  
Workgroup Call Schedule  
Workgroup Tips  
Best Practices in MS Teams

See an area where you want to contribute? Please Join The Journey!

Join A Workgroup

Meeting Schedule

Workgroup Tips

## Get to Know the OHDSI Workgroups

[Africa Chapter](#)

[APAC](#)

[ATLAS/WebAPI](#)

[Clinical Trials](#)

[Common Data Model](#)

[CDM Survey Subgroup](#)

[CDM Vocabulary Subgroup](#)

[Dentistry](#)

[Early-Stage Researchers](#)

[Electronic Animal Health Records](#)

[Eye Care & Vision Research](#)

[FHIR and OMOP](#)

[Generative AI & Analytics in Healthcare \(GAIA\)](#)

[GIS – Geographic Information System](#)

[HADES](#)

[Health Equity](#)

[Healthcare Systems](#)

[Industry](#)

[Latin America](#)

[Medical Devices](#)

[Medical Imaging](#)

[Methods Research](#)

[Natural Language Processing](#)

[Network Data Quality](#)

[Oncology](#)

[Open-Source Community](#)

[Patient-Level Prediction](#)

[Perinatal and Reproductive Health](#)

[Phenotype Development & Evaluation](#)

[Psychiatry](#)

[Rehabilitation](#)

[Registry](#)

[Steering Group](#)

[Surgery and Perioperative Medicine](#)

[Themis](#)

[Vaccine Vocabulary](#)



# COMMUNITY CALLS

Online

Global: Every Tuesday @ 17:00 CET

Europe: Every 2<sup>nd</sup> Thursday @ 13:00 CET



## Upcoming Europe Community Calls

Date	Topic
Sept. 11	Europe Community Call Introduction / DARWIN EU Update
Oct. 9	Meet the New National Nodes
Nov. 13	Patient-Reported Outcome Measures (PROMs)
Dec. 11	Vocabularies in Europe



# LEARNING MORE – EHDEN ACADEMY

- Free online educational resource
- More than 5,350 active learners across 108 countries



Getting Started	EHDEN Foundation	Patient Organizations: Real World Data and Real World Research	OMOP CDM and Standardized Vocabularies	ATLAS	Infrastructure
Extract, Transform and Load	Introduction to Usagi & Code Mappings for an ETL	OHDSI-in-a-Box	ETL Learning Pathway: Data Partner & SME Real World Use Cases	10 Minute Tutorial: PheEvaluator	10 Minute Tutorial: ATHENA
Open Science & FAIR Principles	Introduction to Data Quality	Phenotype Definition, Characterisation and Evaluation	Population-level Effect Estimation	Patient-Level Prediction	R for Patient-level Prediction
Applied Cost-Effectiveness Modeling with R	Assessing healthcare using outcomes that matter to patients	OHDSI2022 Tutorial - Creating Cohort Definitions	OHDSI2022 Tutorial - OMOP Common Data Model/Vocabulary	One hour of your time: The Phenotyping Problem	Health Technology Assessment

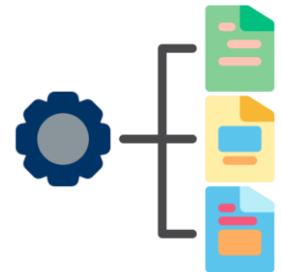


# LEARNING MORE – THE BOOK OF OHDSI

A comprehensive guide to



The OHDSI  
Community



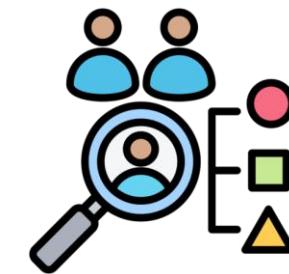
Uniform Data  
Representation



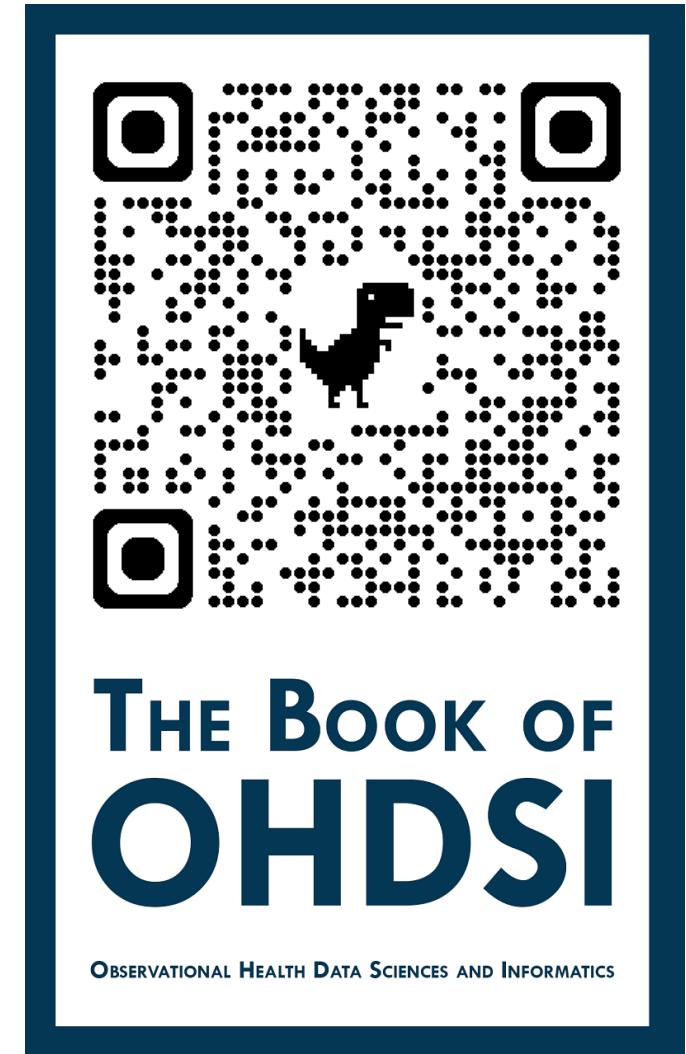
Data Analytics



Evidence Quality



OHDSI studies



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



# LEARNING MORE – YOUTUBE CHANNEL



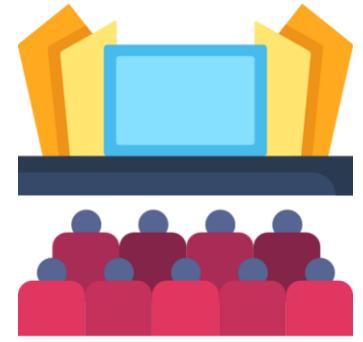
Community call  
meetings



Tool tutorials



Workshops



Previous events



# YouTube

<https://www.youtube.com/@OHDSI>

