INTRODUCTION TO OHDSI

Laura Verbeij, Fleur Vereijken & Julia Kurps

OHDSI Europe Symposium 2026

WELCOME!

How to start your OHDSI journey & Patient Reported Outcomes

Laura

OHDSI tools, prediction modelling & individual prediction instability

Fleur

OHDSI community, All things OMOP CDM

Julia

THIS MORNING…

Topic	Presenter
History & philosophy behind OHDSI	Laura
How does the community work?	Julia
Where and how can you learn more about OHDSI?	Laura
Q&A	
All things OMOP CDM (model, ETL, DQ, conventions)	Julia
What can be done currently with the OHDSI tools?	Fleur
What does it take to be able to use the tools?	Fleur
Q&A	

More info?

What is OHDSI?

2025

EUROPE

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

THIS IS OHDSI!

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

HISTORY

- Since 2000 rapidly increasing number of drug side effects

- FDA Amendments Act → Risk Identification and Analysis System

Sentinel

OMOP

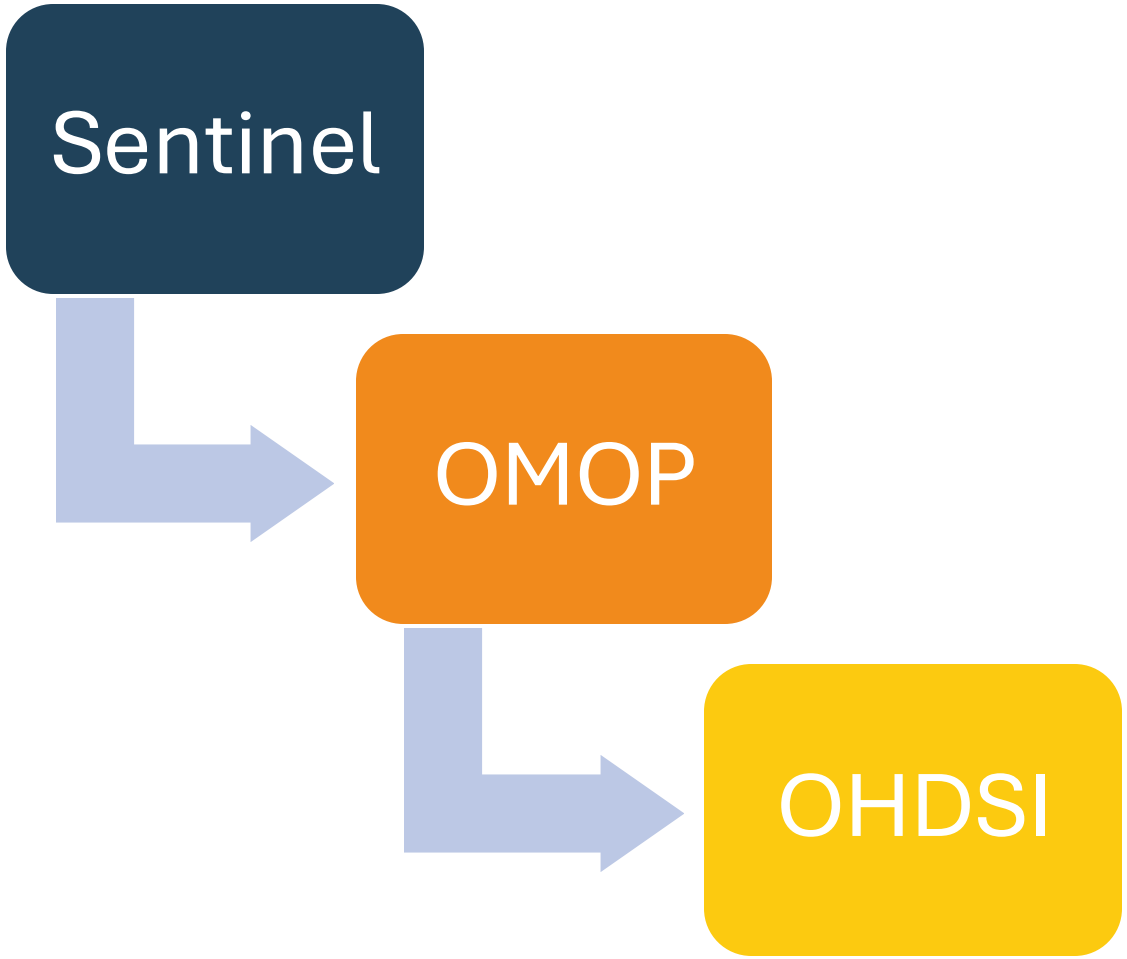
OMOP EXPERIMENT (2009 →)

- Empirically reproduce a set of known (medical) facts using positive and negative test cases. The assumption was that this could be achieved with the best database and the best methods.
- 10 databases; 14 methods

→ Different database, same method, potentially different outcom

→ Analysis choices are crucial

HISTORY



OHDSI (EST. 2014)

EUROPE

Vision

A world in which observational research produces a comprehensive understanding of health and disease.

Mission

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

PHILOSOPHY

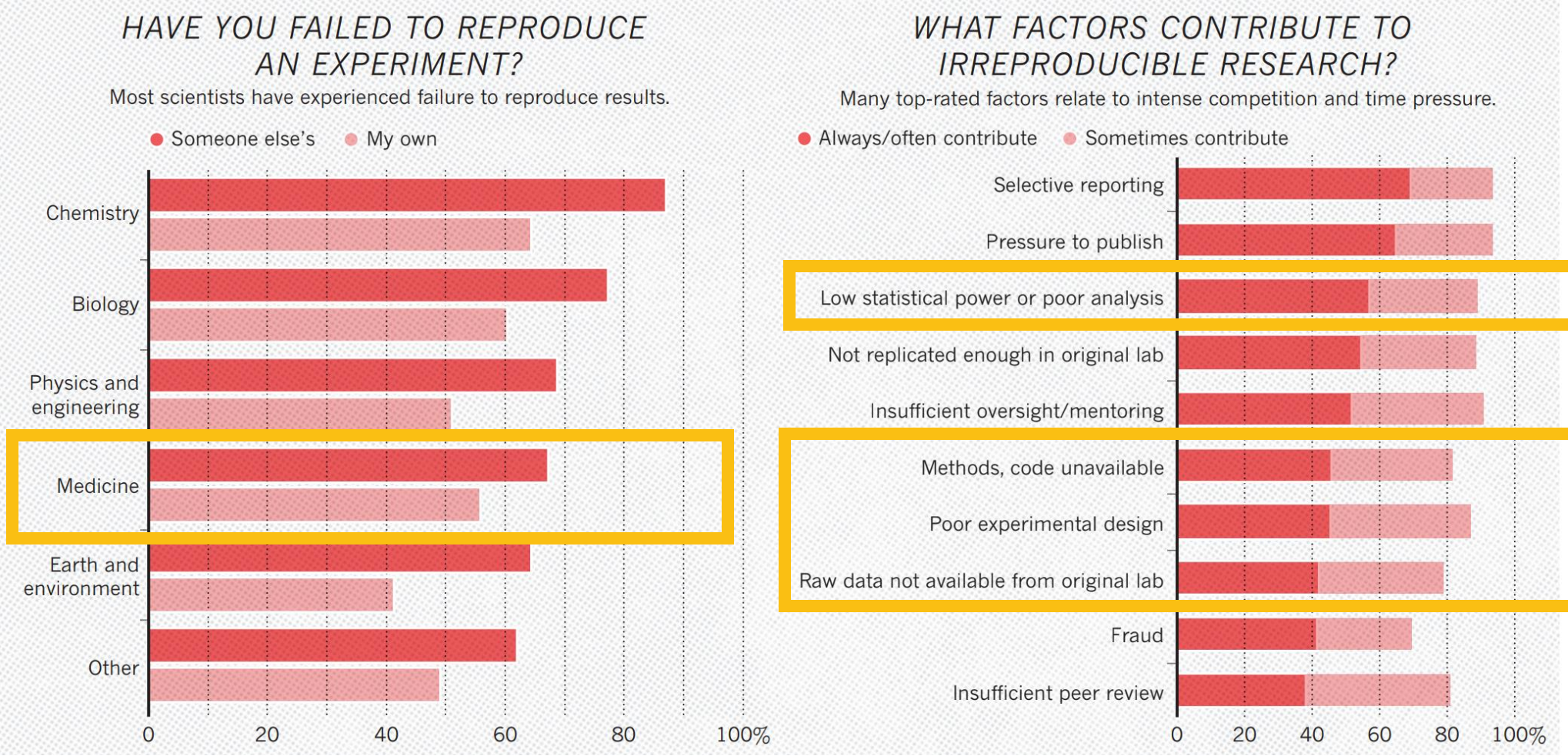
- Interdisciplinary teams
- International collaboration
- All tools should remain public and freely available

More info?
- **The Book of OHDSI** chapter 1.2
- **EHDEN academy course** "OMOP CDM and Standardised Vocabularies"

VALUES

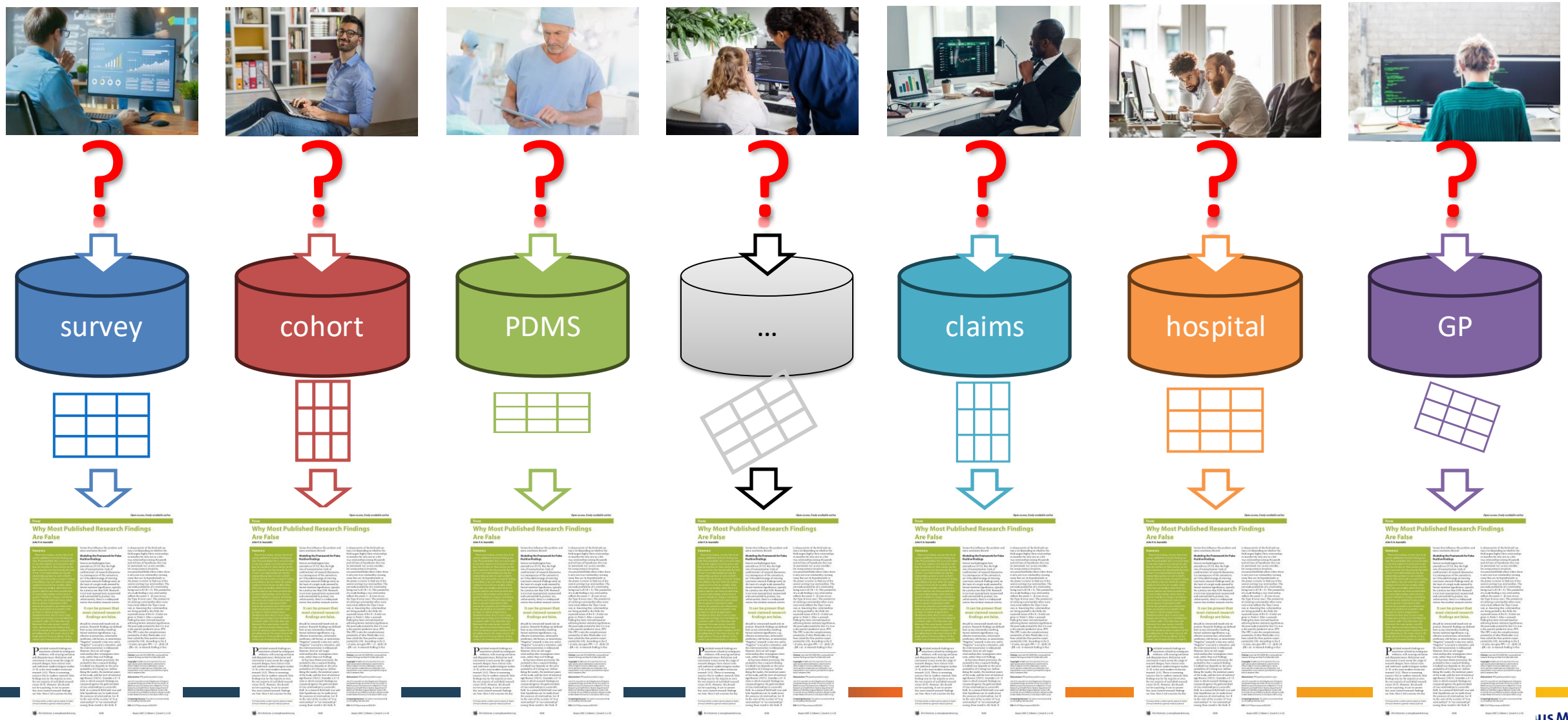
- **Innovation**: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.
- **Reproducibility**: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.
- **Community**: 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.

REPRODUCIBILITY CRISIS (NATURE 2016)

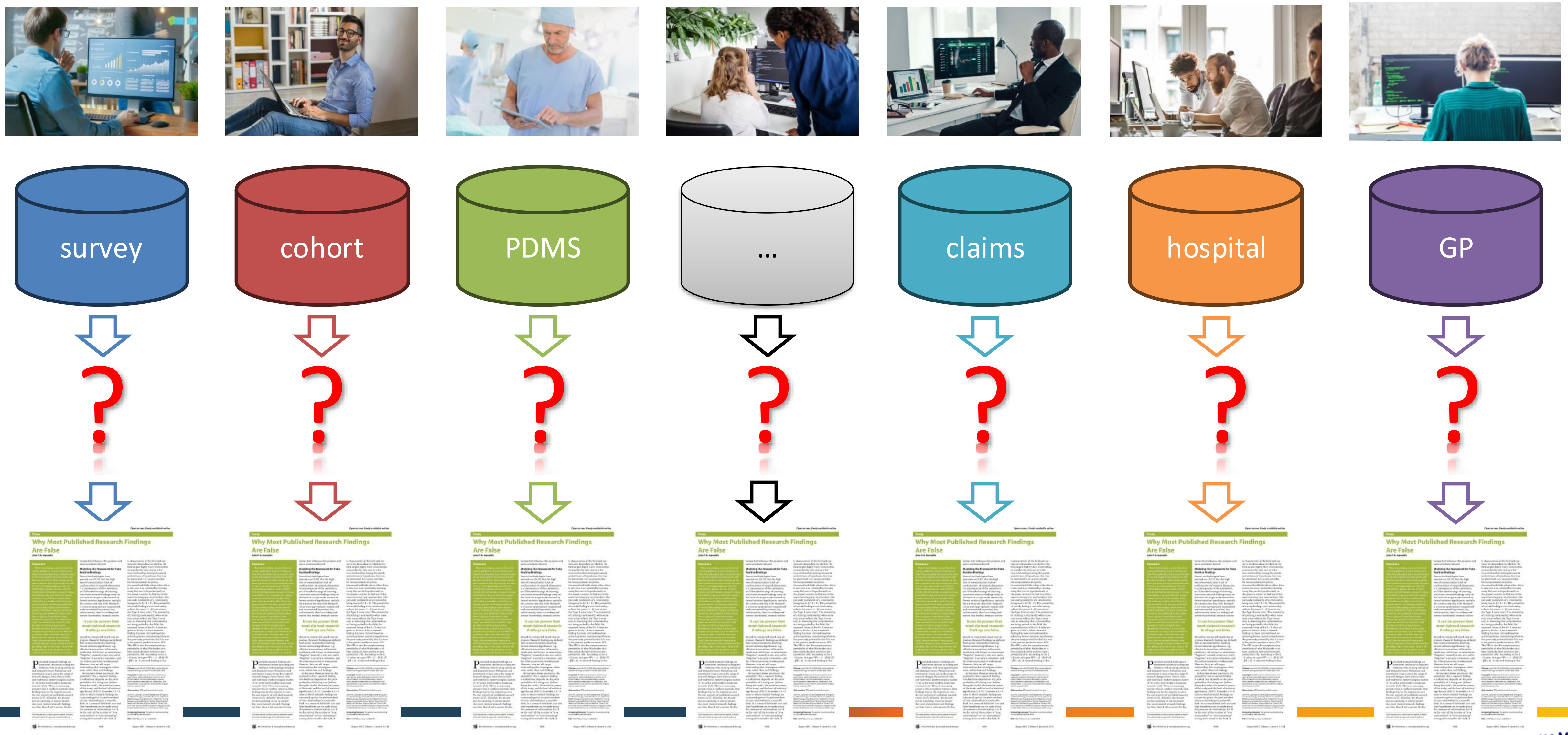


Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016)

OLD-SCHOOL OBSERVATIONAL RESEARCH

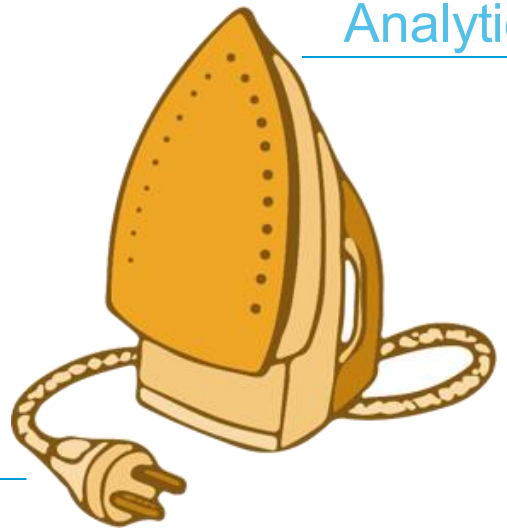


OLD-SCHOOL RESEARCH GONE WRONG



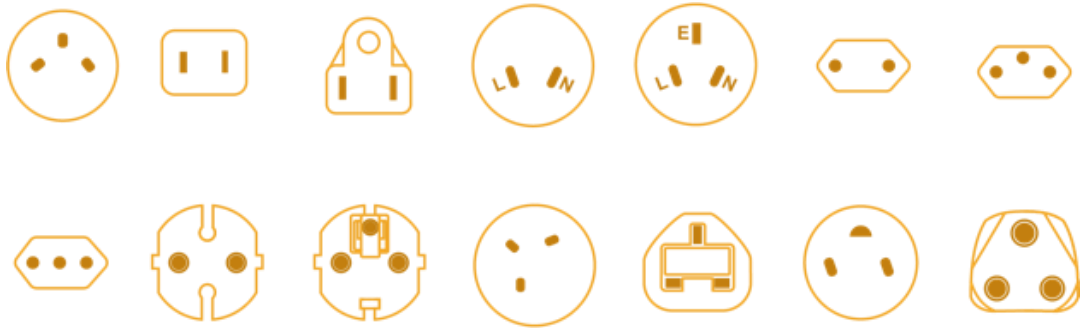
THE CHALLENGES OF REAL-WORLD DATA

Analytical method



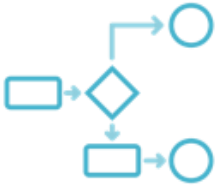
Link to data

The data…



What will it require?

Data interoperability



Standardized analytics

Data network

Strong community

OBSERVATIONAL RESEARCH OHDSI STYLE

EUROPE

Strong community

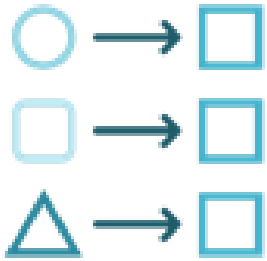
Standardized analytics

survey | cohort | PDMS | ... | claims | hospital | GP

Data aggregation & analysis

Erasmus MC
University Medical Center Rotterdam

Data Interoperability



Data interoperability

OMOP CDM

Syntactic

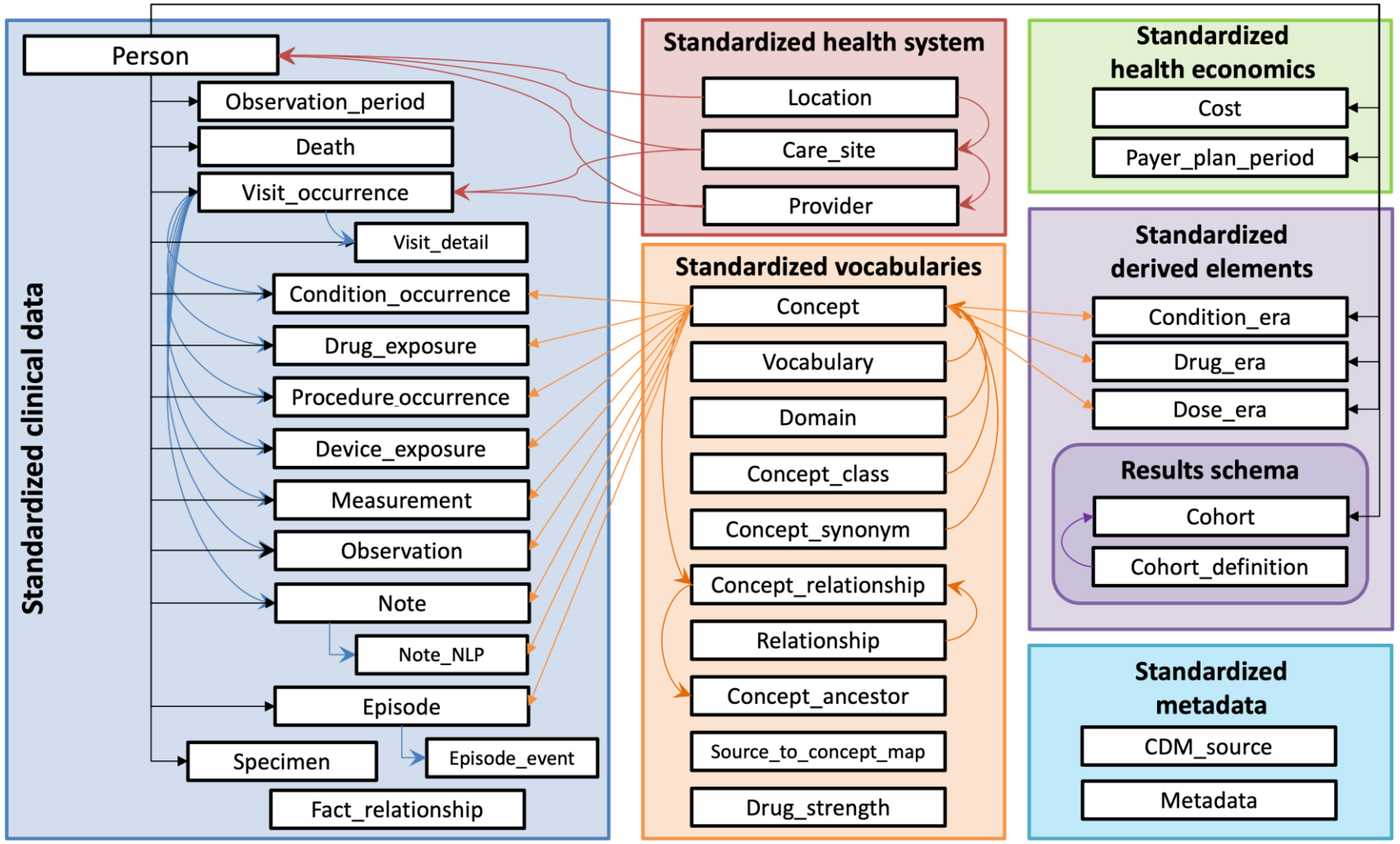
Interoperability

Semantic

OMOP Common Data Model (CDM) v5.4

EUROPE

Syntactic interoperability



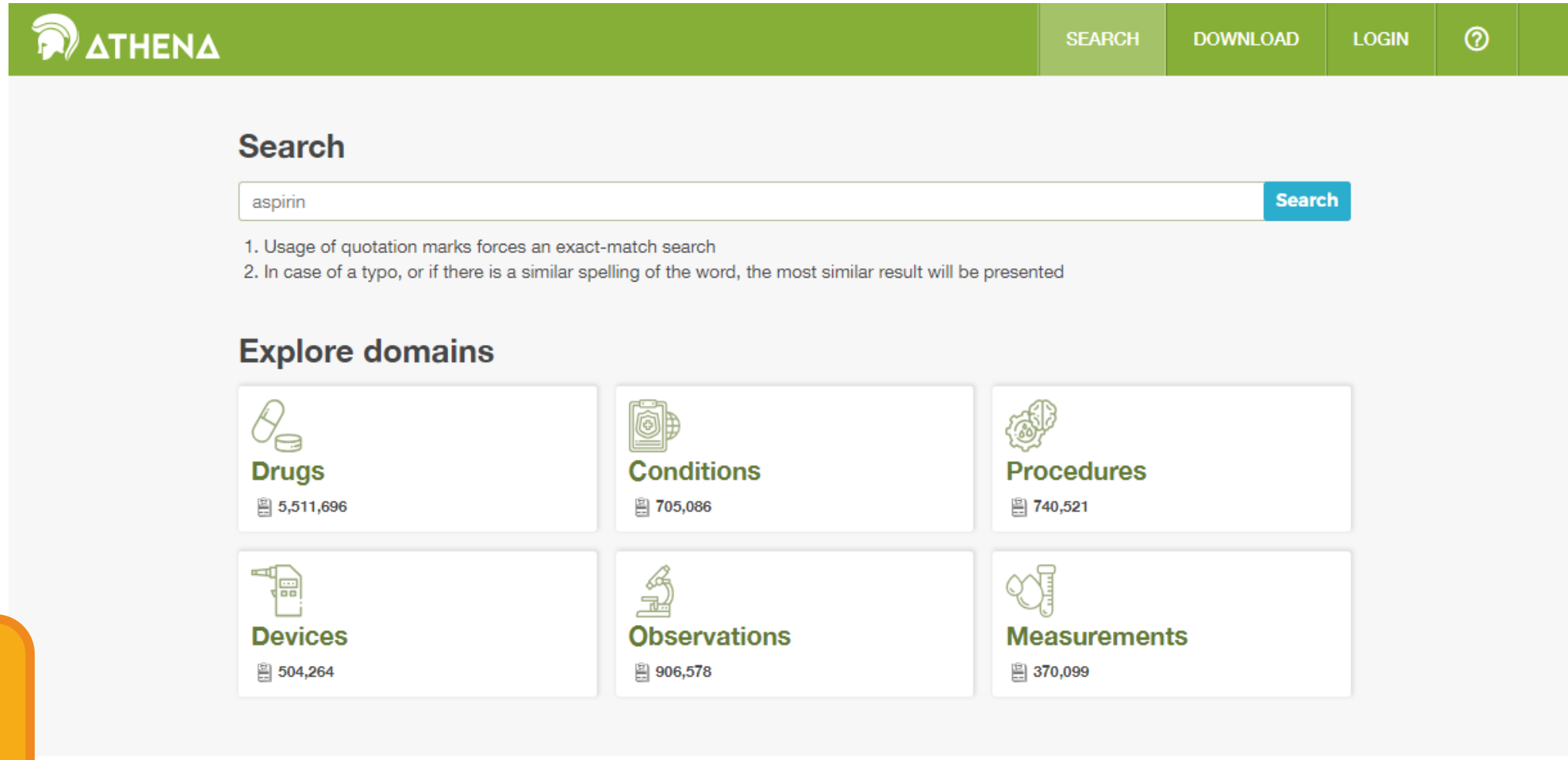
More info?
- **The Book of OHDSI** chapter 4

Erasmus MC University Medical Center Rotterdam

(STANDARDIZED) VOCABULARIES

EUROPE

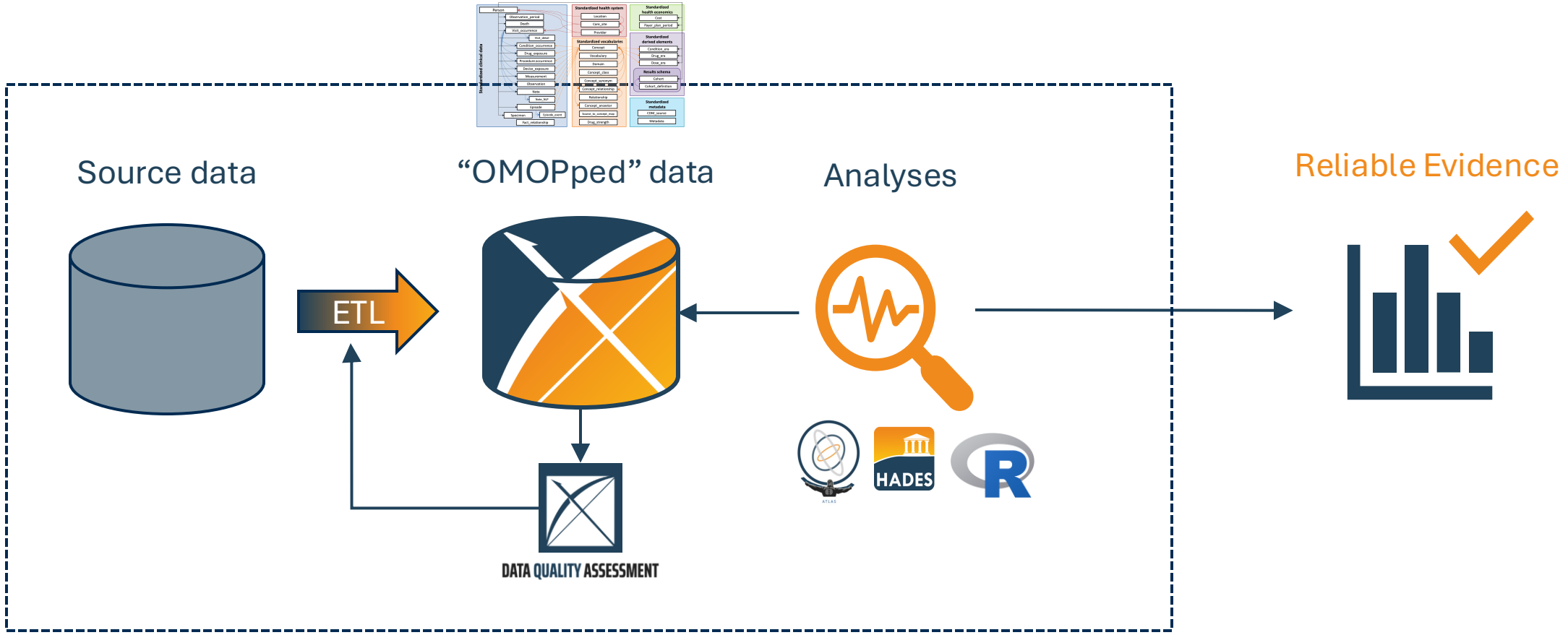
Semantic interoperability



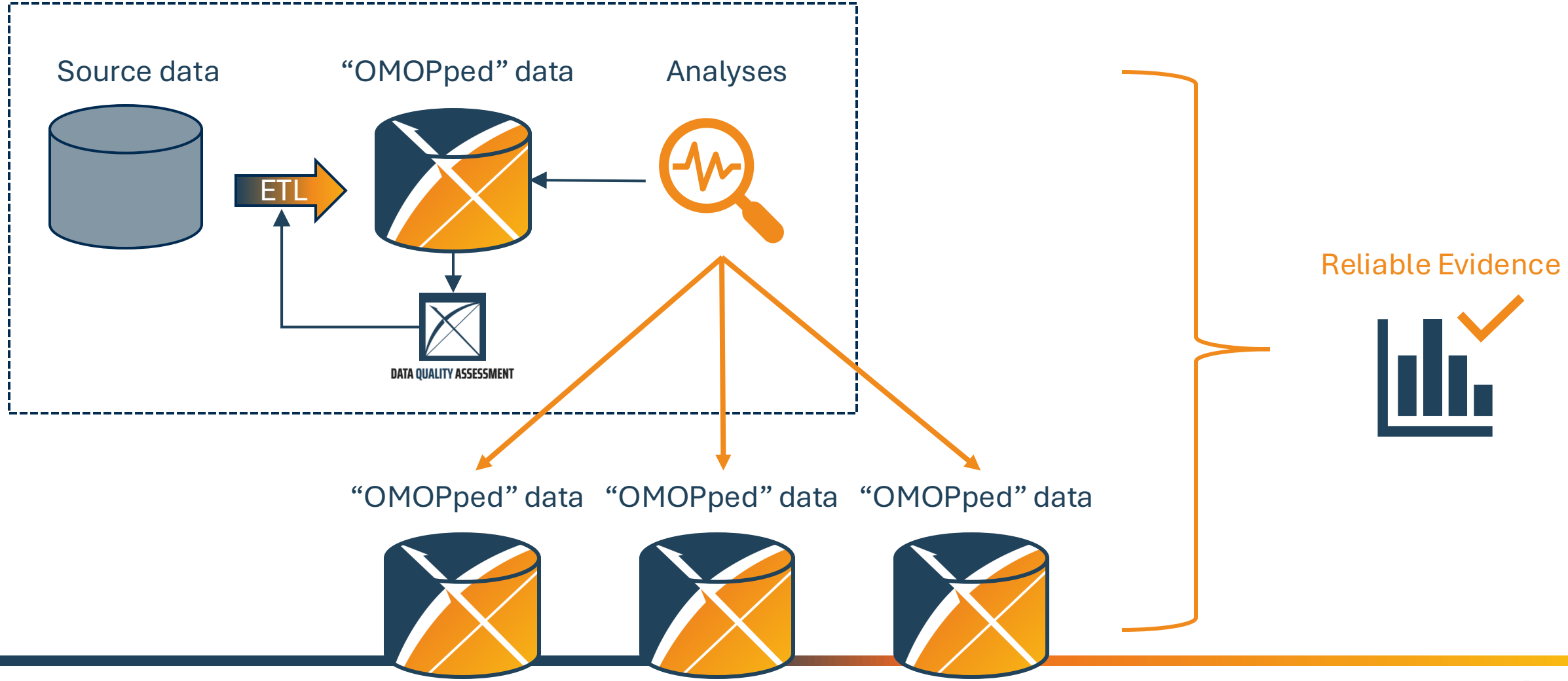
More info?
- **The Book of OHDSI** chapter 5
- athena.ohdsi.org

Erasmus MC University Medical Center Rotterdam

FROM DATA TO EVIDENCE: THE OHDSI PIPELINE SIMPLIFIED

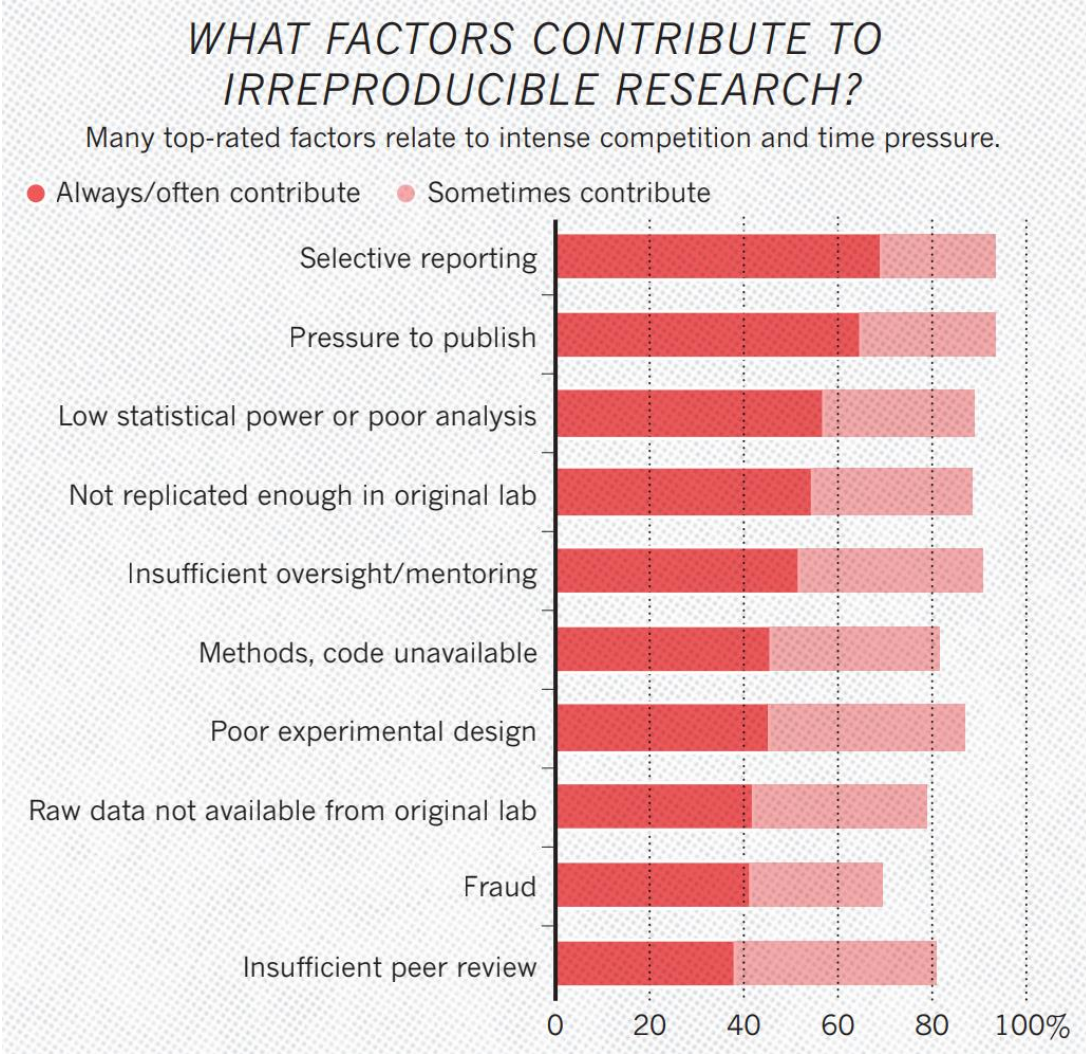


STANDARDIZED ANALYTICS CAN BE REUSED OVER DATA SOURCES



In summary, what is OHDSI?

- An awesome community
- A common data model
- Standardized open-source tools
- A gateway to a large international data network



- Take a few minutes to get to know your neighbour, perhaps you could talk about:
  - **What do you do?**
  - **What do you hope that OMOP/OHDSI will bring you?**
  - **What are you hoping to find at the symposium?**

How does the community work?


ROTTERDAM

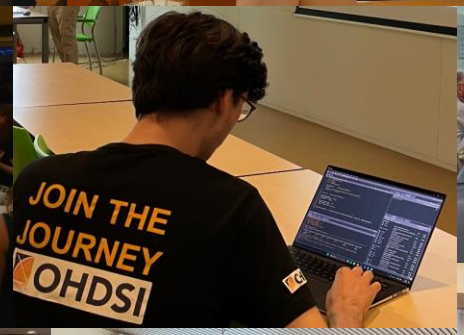






OHDSI





JOIN THE
JOURNEY
OHDSI





What does it take to be able to run the tools


OHDSI GLOBAL

EUROPE

OHDSI COLLABORATORS

Map of Collaborators

The OHDSI community brings together volunteers from around the world to establish open community data standards, develop open-source software, conduct methodological research, and apply scientific best practices to answer public health questions by generating reliable clinical evidence.

OHDSI COLLABORATORS

Our community is ALWAYS seeking new collaborators. Do you want to focus on data standards or methodological research? Are you passionate about open-source development or clinical applications? Do you have data that you want to be part of global network studies? Do you want to join a global community that truly values the benefits of open science? Add a dot to the map below and JOIN THE JOURNEY!

OHDSI By The Numbers

- 4,751 collaborators
- 88 countries
- 9,004 followers on LinkedIn
- 21 time zones
- 6 continents
- 1 community

Erasmus MC University Medical Center Rotterdam

LARGE NETWORK OF DATA SOURCES

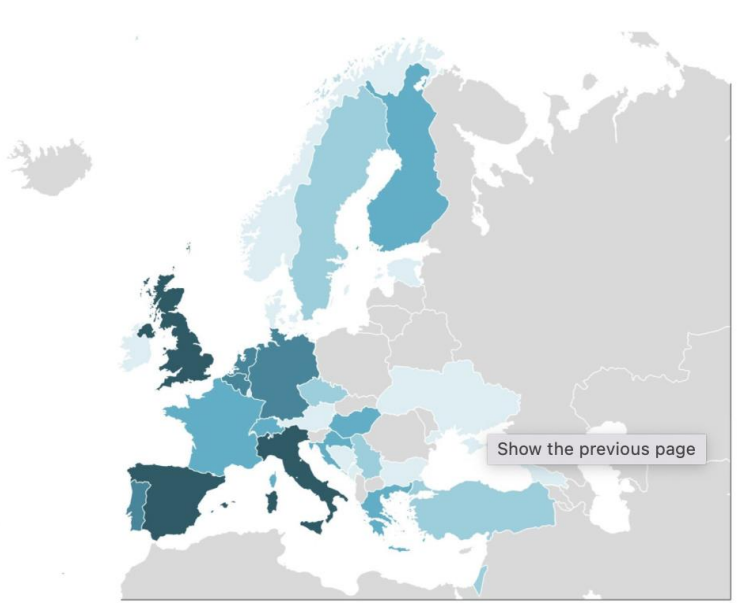
544 database from 54 countries

More than 974 million unique patient records

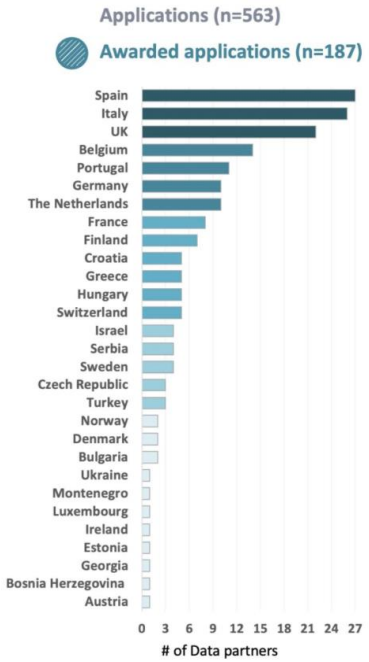
~ 12% of the global population

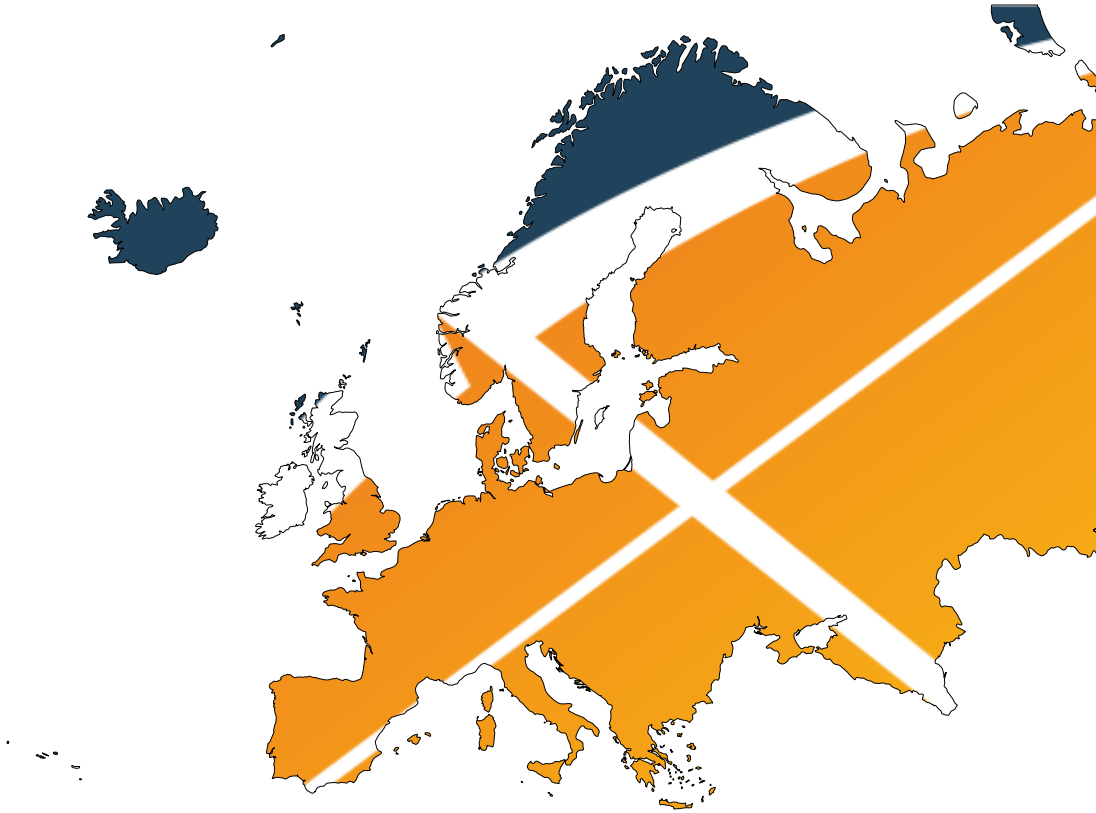
OHDSI EUROPE

From EHDEN to a network of national nodes



Geographic spread of data partners. The shade of blue indicates the # of data partners in that country (darker = more)





EUROPEAN HEALTH DATA & EVIDENCE NETWORK
EHDEN Foundation

207 databases from 30 countries
More than 455 million patients

Erasmus MC University Medical Center Rotterdam

COMMUNITY CALLS

EUROPE

Upcoming Community Calls

Date	Topic
Apr. 14	Phenotype Aphril, Week 2: Phenotype Validation Using KEEPER
Apr. 21	NO MEETING / EUROPE SYMPOSIUM
Apr. 28	Phenotype Aphril, Week 4: Final Evaluation and Learnings
May 5	Europe Symposium Review/Phenotype Aphril Finale
May 12	Collaborator Showcase Brainstorm (Submission Deadline is June 5)
May 19	MEDS (Medical Event Data Standard) & Potential Collaborations with OHDSI

#JoinTheJourney www.ohdsi.org

Weekly community calls

Monthly Europe community calls

Upcoming Europe Community Calls

Date	Topic
May 7	Recap OHDSI EU Symposium / Welcome Newcomers
June 11	Oncology Studies
July 9	Deep Dive Topics OH!-Factor

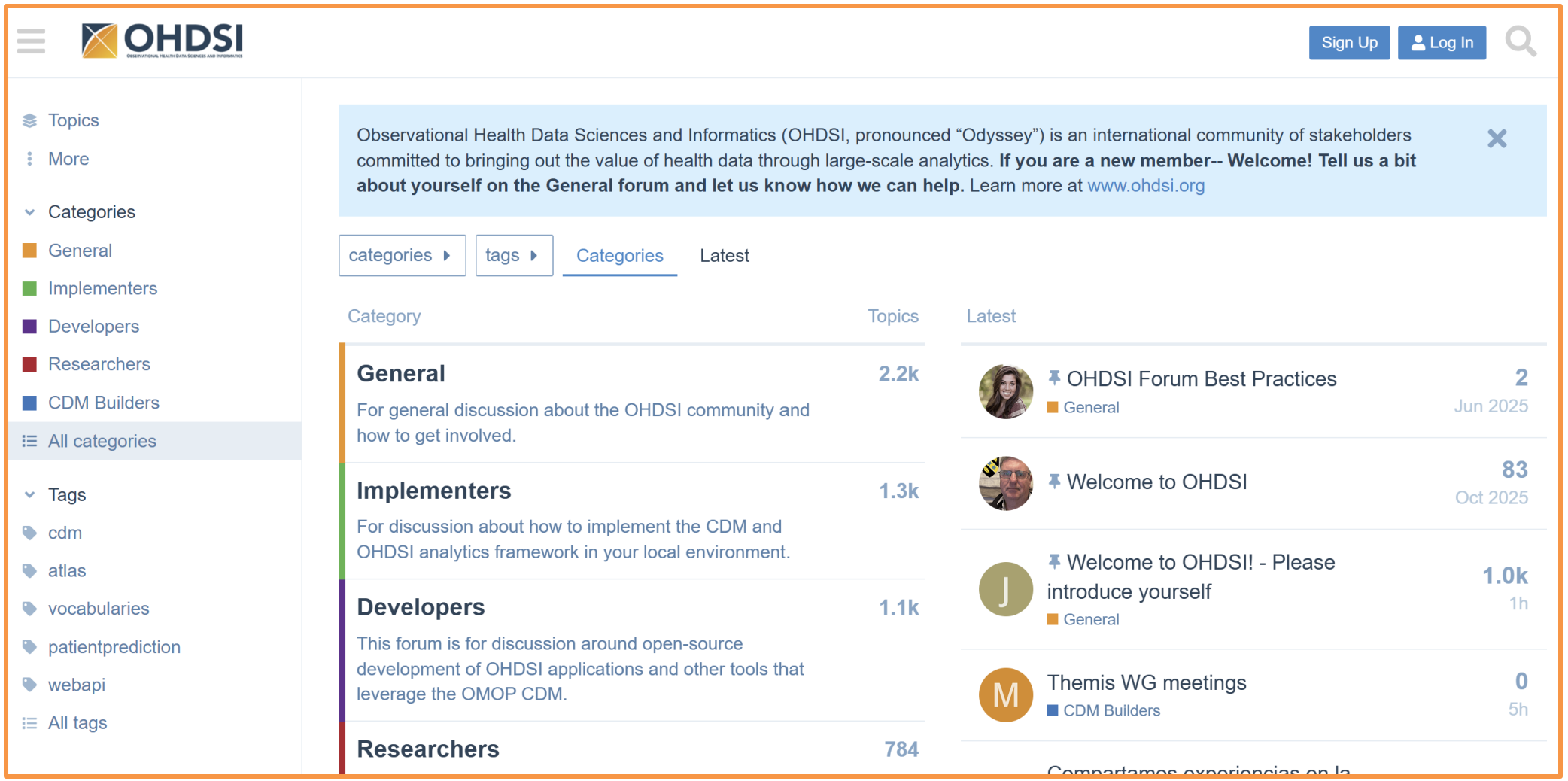
#JoinTheJourney www.ohdsi.org

Erasmus MC University Medical Center Rotterdam

JUST ASK @ OHDSI FORUM!!

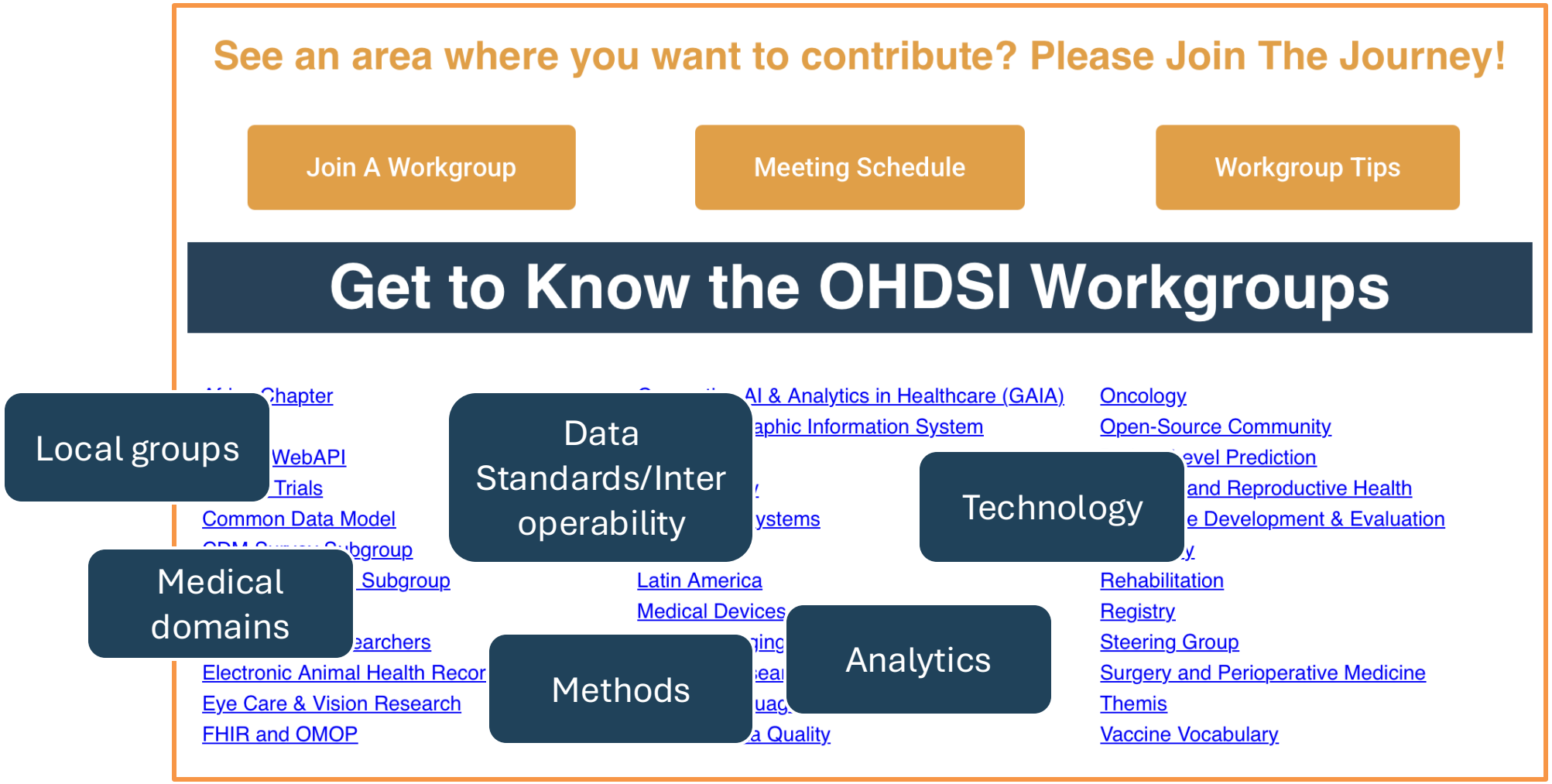
Huge corpus of background information

No barriers, everyone can sign-up & ask a question



OHDSI WORKGROUPS

EUROPE



COLLABORATIVE EVENTS

Community-wide events:

- **Phenotype Phebruary (Aphril)**

Focus events:

- **Studyathons**
- **Datathons**

2022 Korea Datathon


2025 Finland Studyathon: Exploring the Real-World Treatment Landscape of mNSCLC

EVENTS

EUROPE

Registration & Call for Participation are OPEN

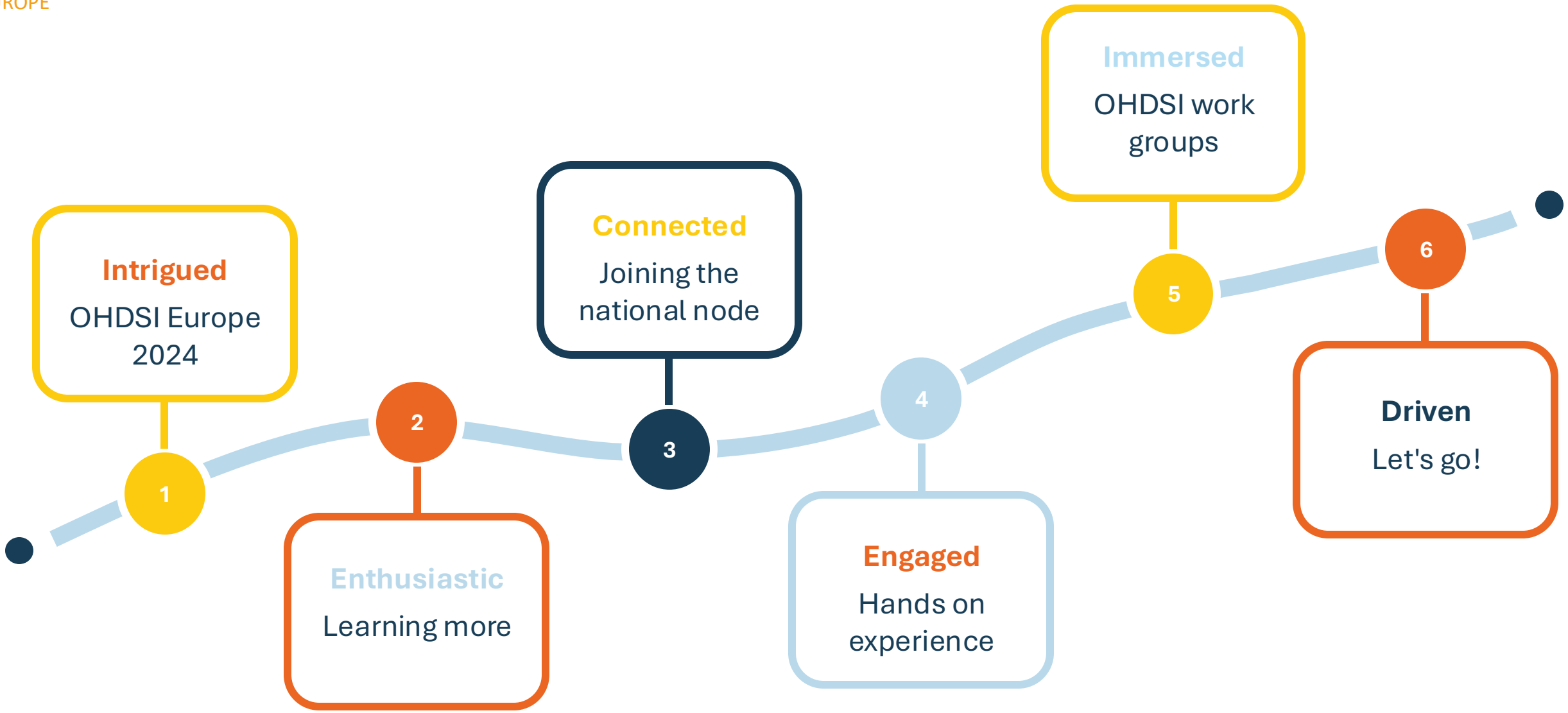
2026 OHDSI Global Symposium

Oct. 20-22 • New Brunswick, N.J. • Hyatt Regency Hotel

Erasmus MC University Medical Center Rotterdam

HOW CAN YOU LEARN MORE ABOUT OHDSI

MY FIRST YEAR OHDSI VOYAGE



OHDSI EUROPE 2024

New job & starting PhD

Invitation to join OHDSI Europe symposium

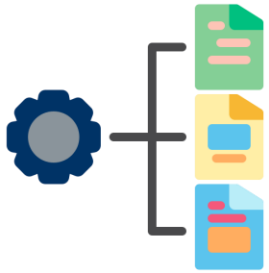
First impression

Introduction to OHDSI

LEARNING MORE – THE BOOK OF OHDSI

A comprehensive guide to

The OHDSI Community

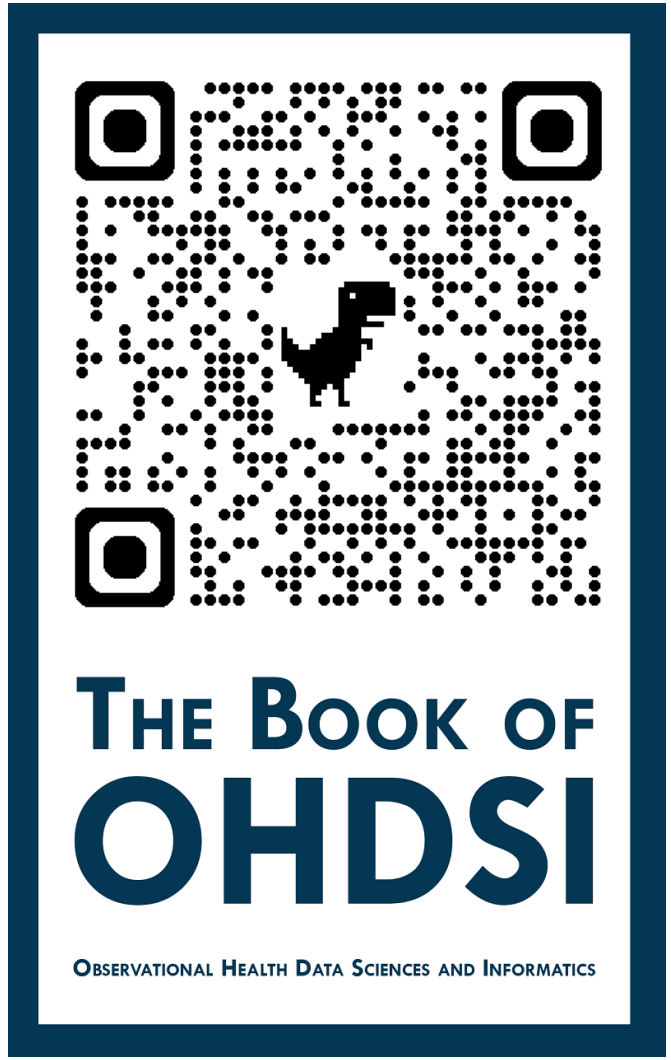


Uniform Data Representation

Data Analytics

Evidence Quality

OHDSI studies



LEARNING MORE – EHDEN ACADEMY

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

Getting Started	EHDEN Foundation	Patient Organisations: Real World Data and Real World Research	OMOP CDM and Standardised 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: PheValuator	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

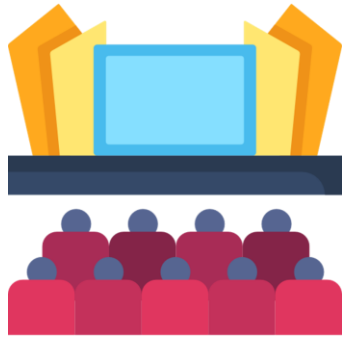
https://academy.ehden.eu

LEARNING MORE – YOUTUBE CHANNEL

Community call meetings

Tool tutorials

Workshops



Previous events

YouTube

JOINING THE NATIONAL NODE

- Meeting great people with lots of humor, knowledge and good ideas
- Joining national and international collaborations
- Getting involved in studies
  - Study-a-thon
- Showcasing work to a group spanning much of the healthcare system

CONNECT TO YOUR NATIONAL NODE!

EUROPE

1. Belgium
2. Denmark
3. Estonia
4. Germany
5. Greece
6. Israel
7. Italy
8. Luxemburg
9. Netherlands
10. Norway
11. Portugal
12. Spain
13. United Kingdom
14. Finland
15. Hungary
16. Ireland
17. Sweden
18. Austria

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

Erasmus MC
University Medical Center Rotterdam

HANDS-ON EXPERIENCE

Pan-European Health Outcomes Observatory
- Alignment (common approach) on data and outcome standards (what should be measured)
- Regular publication of aggregated results
- Technological governance for interoperability and reproducibility

Pan-European Observatory

National H2O observatories, in four countries for the three disease areas (diabetes, IBD, cancer)
- Separate legal entities
- Mapped to the common data model (OMOP)
- Providing regular publication of aggregated results

National Observatories

Austria Germany Netherlands Spain

Other data users (academic and industry researchers, health authorities)

Data access as appropriate according to legal/ethical governance, reports, decision support

Anonymized patient-level data

Identifiable patient-level data

Patients and healthcare professionals

Data collection layer

Protected under local/national legal and ethical frameworks

Patient-reported data e.g. PROs

Provider-reported data: clinical health records

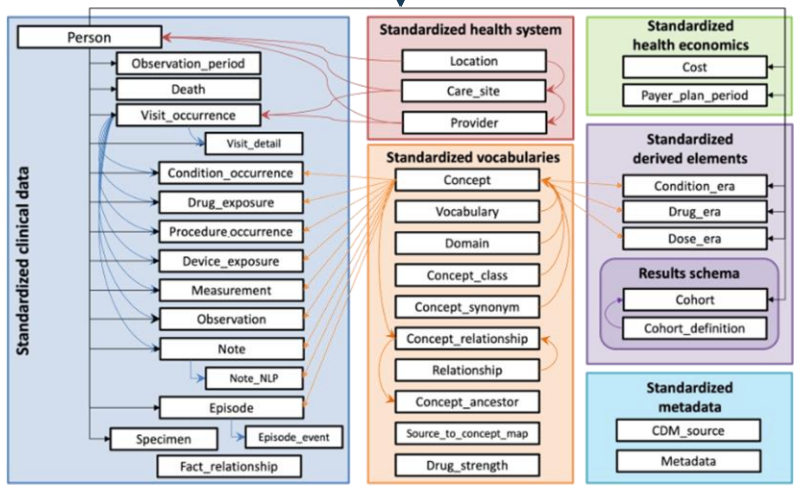
Other data: registries, biobanks

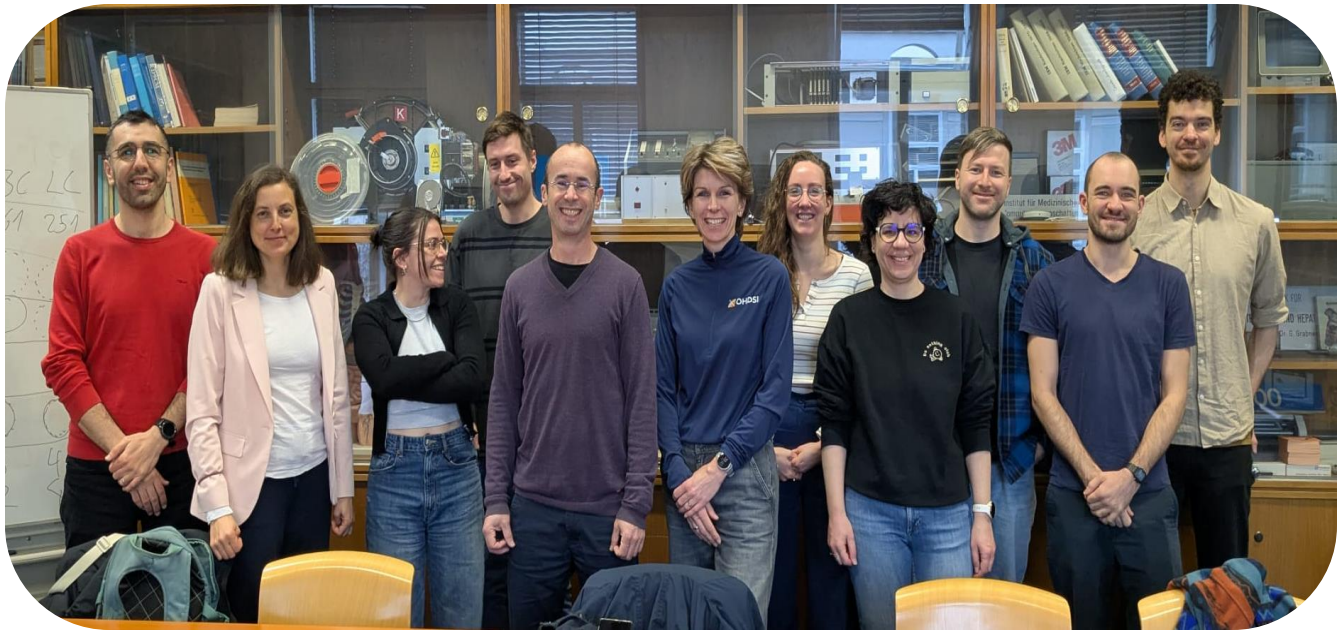
HANDS-ON EXPERIENCE

Patient-reported data e.g. PROs

Provider-reported data: clinical health records

Other data: registries, biobanks



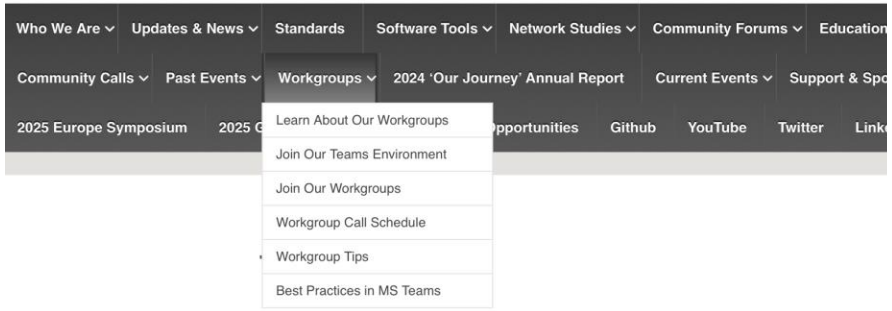


H2O HEALTH OUTCOMES OBSERVATORY

OHDSI WORKGROUPS

EUROPE

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INF



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

Erasmus MC University Medical Center Rotterdam

OHDSI Meet-Ups

In-person

2025 OHDSI Global Symposium

Oct. 7-9 · New Brunswick, N.J. · Hyatt Regency Hotel

2024 OHDSI APAC Symposium

December 4-8 · Marina Bay Sands & National University of Singapore (NUS)

Online

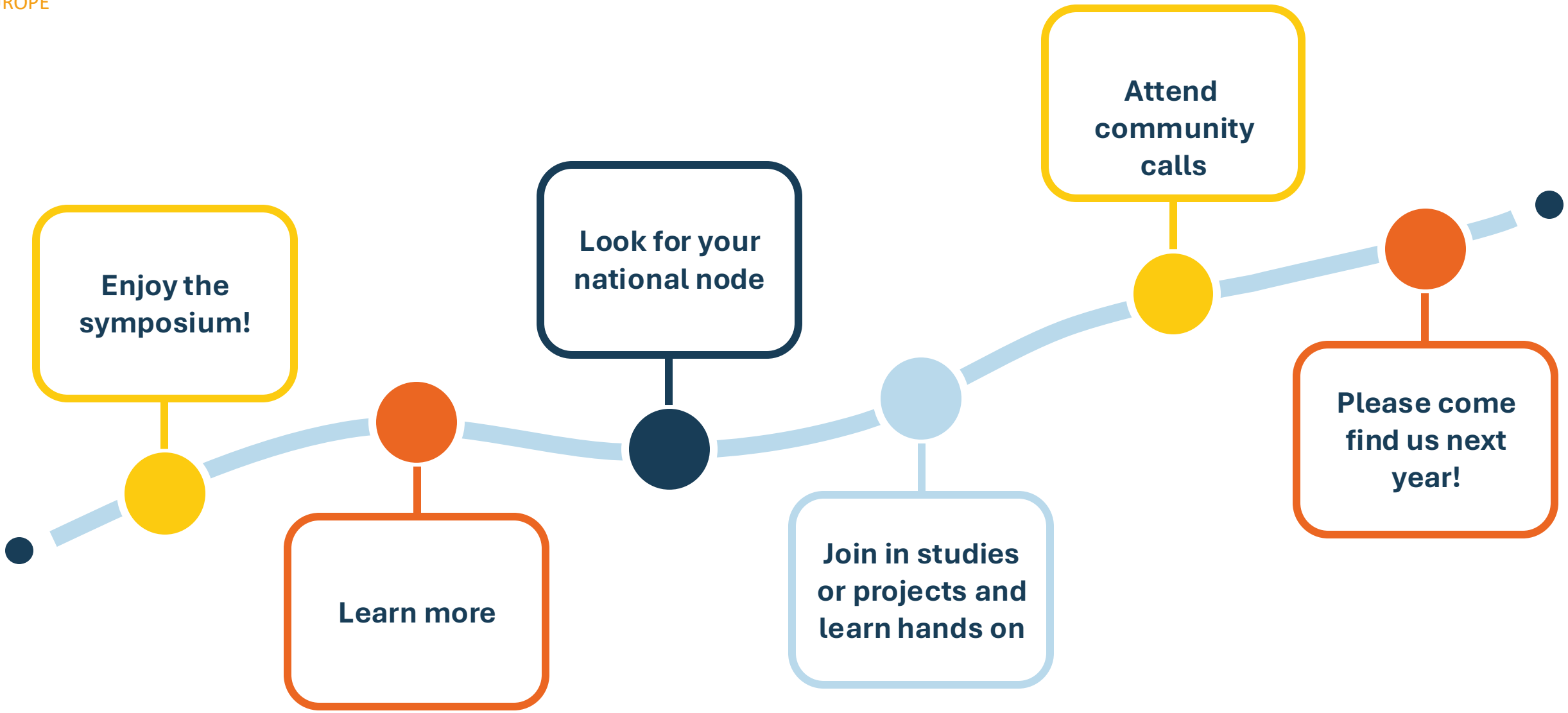
Every Tuesday @ 17:00

Upcoming Community Calls

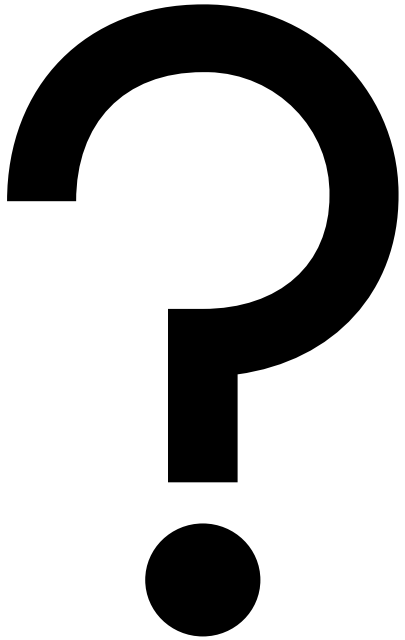
Date	Topic
June 24	ATLAS Deepdive: Characterization, Cohort Pathways, Incidence
July 1	ATLAS Deepdive: Technical and Administrative Capabilities
July 8	No Meeting – Europe Symposium
July 15	Europe Symposium Review
July 22	OMOP/OHDSI Research Spotlight
July 29	Asia-Pacific Regional Updates
Aug. 5	No Meeting
Aug. 12	Newcomer Introductions

@OHDSI www.ohdsi.org #JoinTheJourney ohdsi

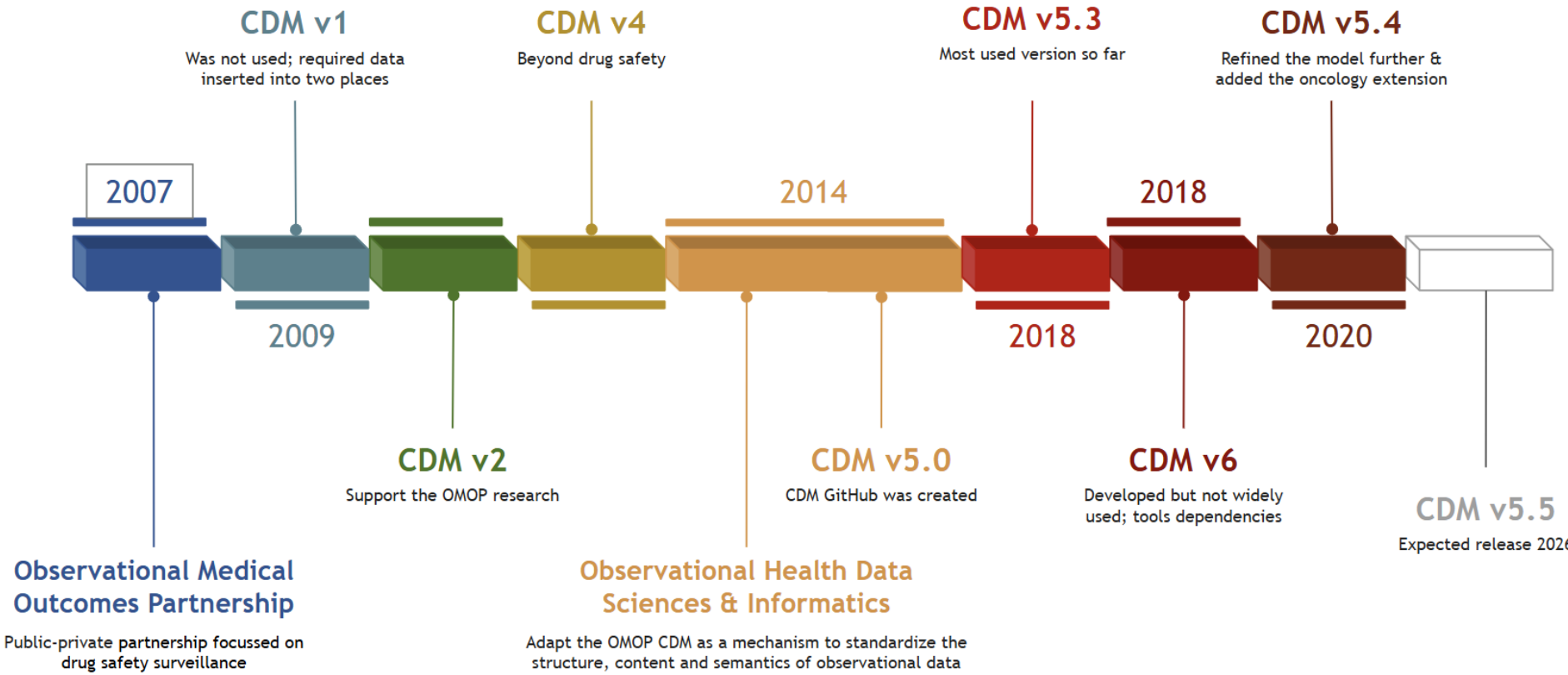
YOUR FIRST YEAR OHDSI VOYAGE



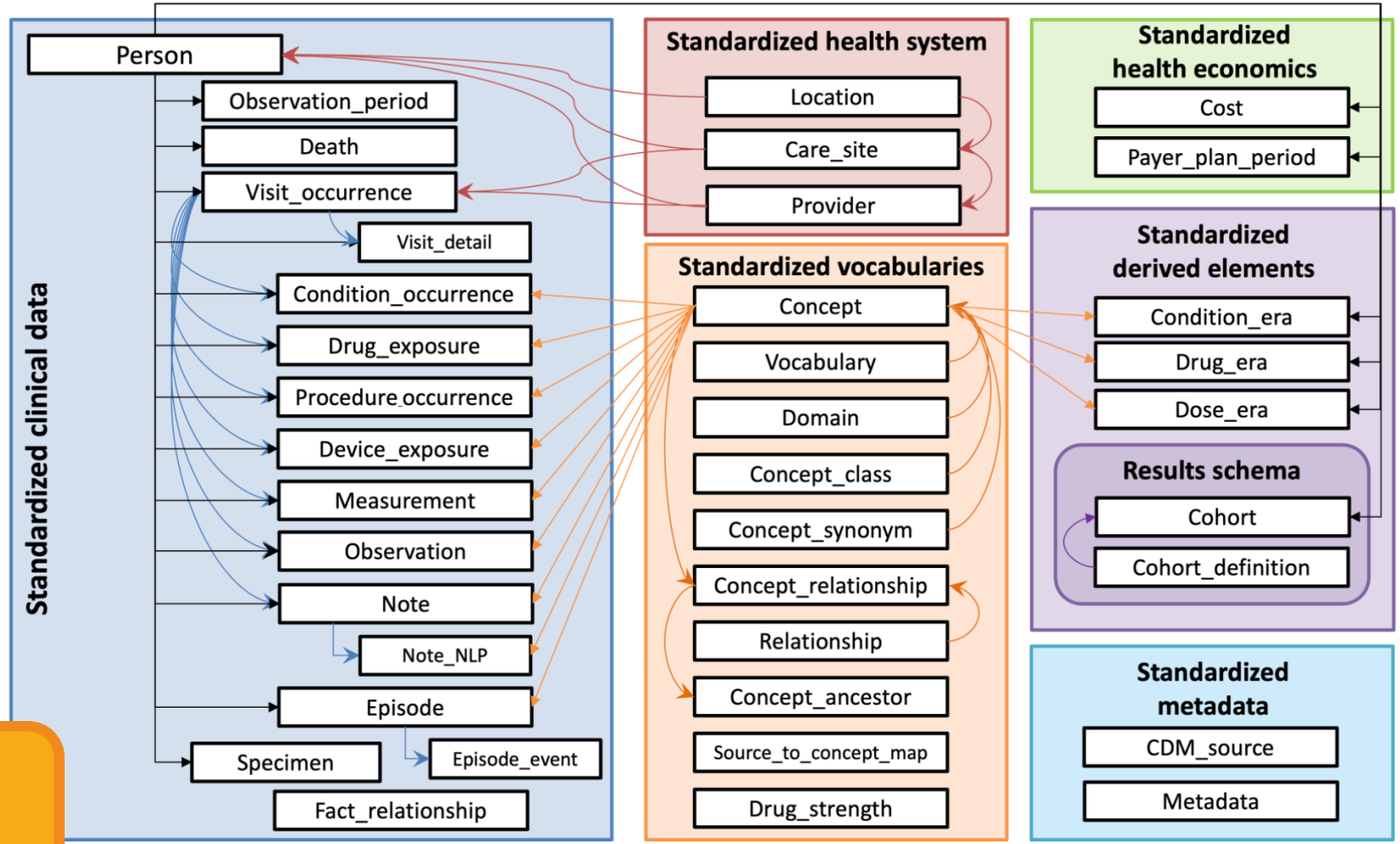
Q + A



HISTORY OF THE OMOP CDM



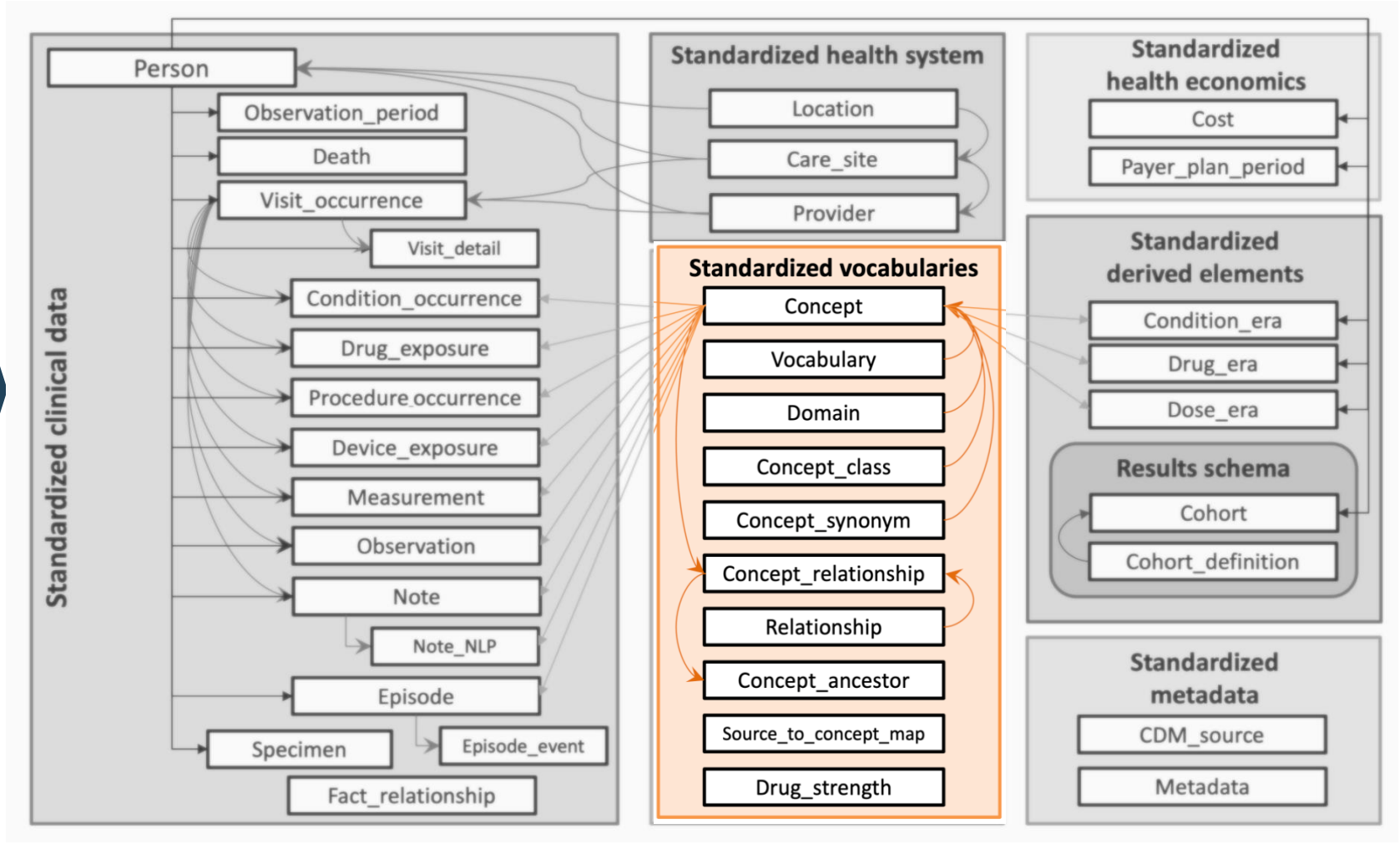
CDM v5.4



More info?
- **The Book of OHDSI** chapter 4

EUROPE

OHDSI STANDARDIZED VOCABULARIES



Erasmus MC University Medical Center Rotterdam

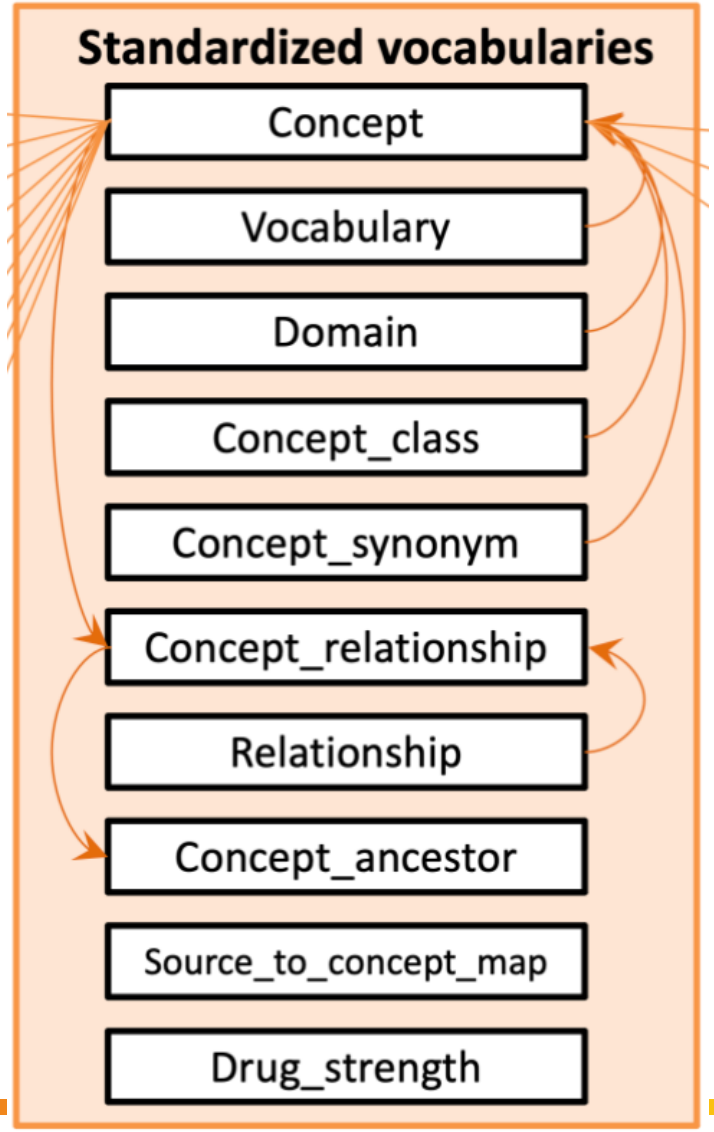
STANDARDIZED VOCABULARIES

Foundational part of OMOP OHDSI for standardizing data

>100 vocabularies harmonized to a compiled standard

Continuous maintenance and improvements

Updates in **February and August** – Roadmap Vocabulary WG



STANDARDIZED VOCABULARIES

- Existing vocabularies (e.g. SNOMED, ICD-O-3, etc.)
- Internally produced *administrative* vocabularies (eg. Episode Type)

- **Standard concept:** One concept representing the meaning of each clinical event
- **Non-standard concepts:** rest of the concepts of the standardized vocabularies which represent the same clinical event

- **CONCEPT_RELATIONSHIP table**: Mapping between non-standard and standard concepts

STANDARDIZED VOCABULARIES

Regional enteritis of unspecified site

Concept_id: 44822028
Vocabulary_id: **ICD9CM**
Concept_code: 555.9
Concept: Non-standard

Crohn's disease

Concept_id: 201606
Vocabulary_id: **SNOMED**
Concept_code: 34000006
Concept: **Standard**

Crohn disease, unspecified

Concept_id: 45562490
Vocabulary_id: **ICD10**
Concept_code: K50.9
Concept: Non-standard

Crohn's disease NOS

Concept_id: 45446786
Vocabulary_id: **Read**
Concept_code: J40z.11
Concept: Non-standard

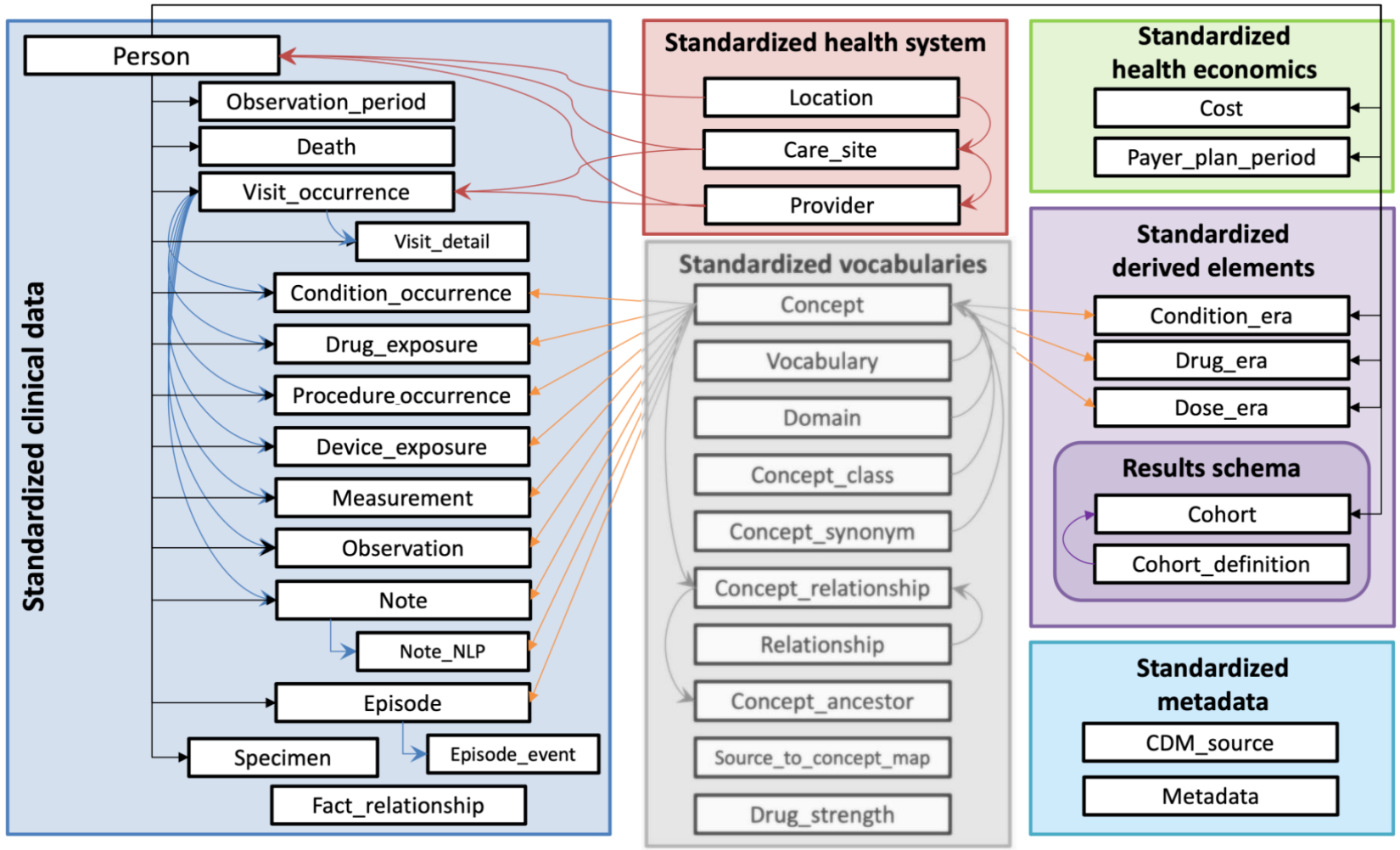
Crohn's disease of the small bowel NOS

Concept_id: 4055022
Vocabulary_id: **SNOMED**
Concept_code: 196980005
Concept: Non-standard

IF YOU WANT TO DISCUSS VOCABS …

Sunday	**OHDSI office hours: Q&A with experts** *Open Q&A with community experts to discuss methods, best practices, and real-world experiences across topics such as ETL, data mapping, study design, and OHDSI research workflows.*			
15:30 – 17:00	**ETL / Mapping** Anne van Winzum & Stefan Payralbe (The Hyve)	**OHDSI Standardized Vocabularies** Anna Ostropolets	**OHDSI studies** Ross Williams	**European Node leads meeting*** Renske Los **(only for NN leads)*

CDM v5.4

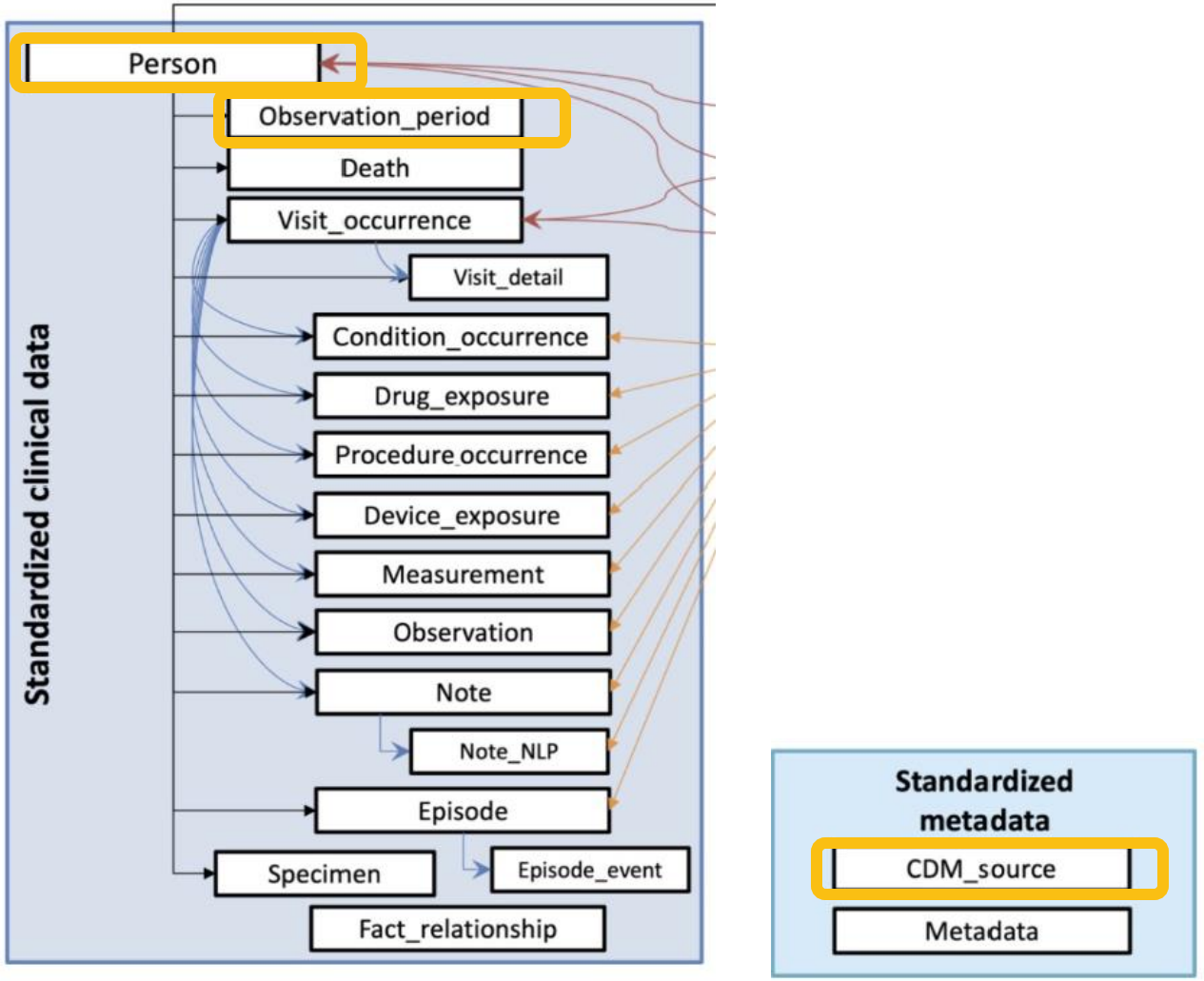

Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Specimen
Episode_event
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata

CORE TABLES

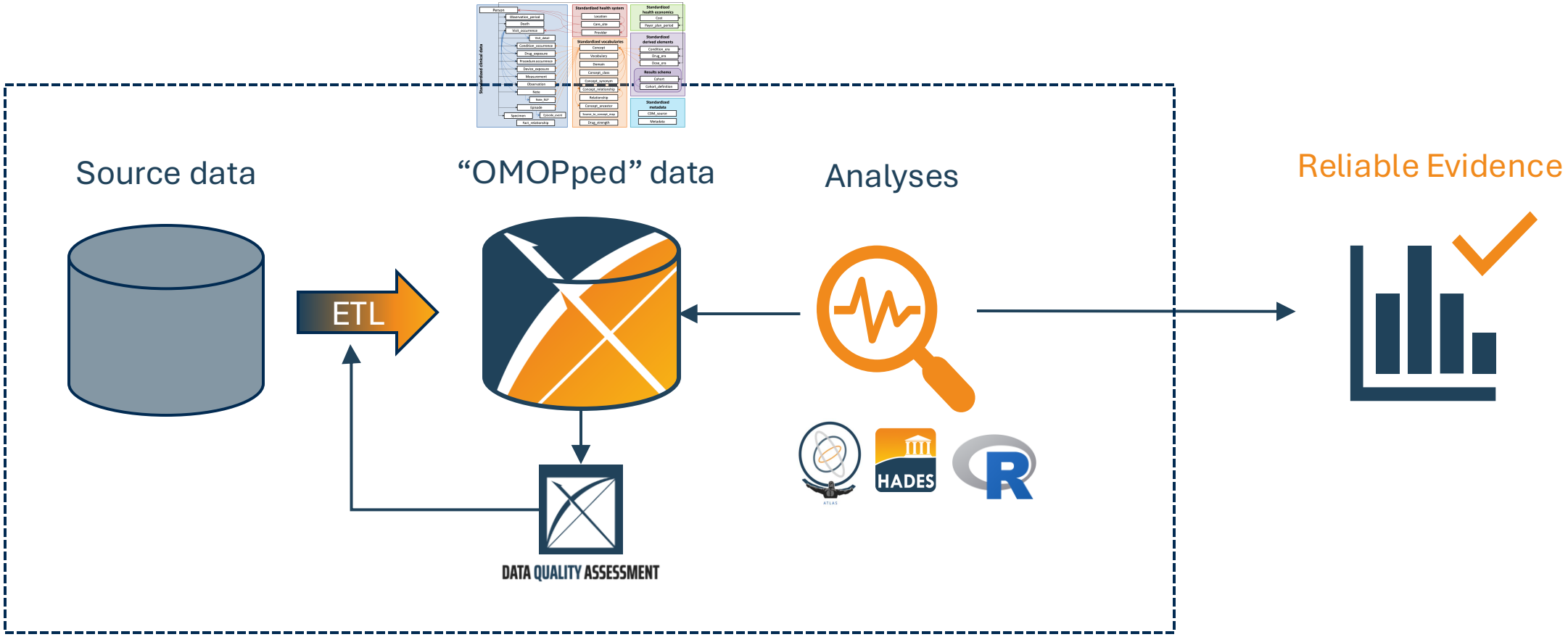
Person (mandatory)
Observation period (mandatory)
CDM source (mandatory)

Medical Events

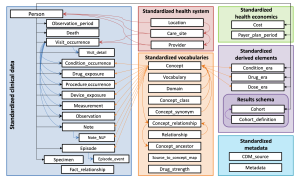
- Death
- Visit occurrence
- Clinical event domains (8 tables)



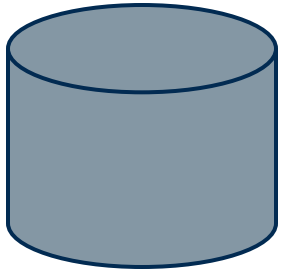
How to get there – ETL tooling



How to get there – ETL tooling



Source data



ETL

"OMOPped" data

DATA QUALITY ASSESSMENT

Analyses

Reliable Evidence

ETL TOOLING – RABBIT SUITE

White Rabbit (WR)

Source data profiling

- Run on relational database or delimited text files

Output summary statistics

- Overview of all tables and variable names
- Variable sparsity, most frequently occurring values

Synthetic data generation

ETL TOOLING – RABBIT SUITE

White Rabbit (WR) is an application that is used to profile and summarize the source data

Accepts **csv, SQL databases, SnowFlake** etc. as input

Output is a **Scan Report** containing descriptive statistics on table- and field level:

- Data type (e.g. string/numeric)
- Max. length per field
- Emptiness per field
- No. of unique values

	A	B	C	D	E	F	G	H	I	J
1	Table	Field	Descriptio	Type	Max lengt	N rows	N rows ch	Fraction e	N unique values	
2	Tabel A	Veld 1		string	32	2.24E+09	100000	0.0%	12263	
3	Tabel A	Veld 2		string	32	2.24E+09	100000	0.0%	86623	
4	Tabel A	Veld 3		string	3	2.24E+09	100000	60.5%	44	
5										
6	Tabel B	Veld 11		string	32	2.24E+09	100000	0.0%	99675	
7	Tabel B	Veld 12		string	32	2.24E+09	100000	0.0%	99675	
8	Tabel B	Veld 13		string	1	2.24E+09	100000	0.0%	2	
9										

Field Overview | Table Overview

Ready | Accessibility: Good to go | 100%

	A	B	C	D	E	F	G	H
1	patient_id	Frequency	geboortedatum	Frequency	seks	Frequency	plaats	Frequency
2	List truncated...		1998-08-01	5	0	900	"Heerlen"	1565
3			1953-12-15	5	1	1100	"Maastricht"	400
4			List truncated...				"Venlo"	10
5							"Sittard"	5
6							List truncated...	

ETL TOOLING – RABBIT SUITE

Rabbit-in-a-Hat (RiaH)

Define source data to OMOP mapping

- Created with White Rabbit scan as input
- Table to table and column to column

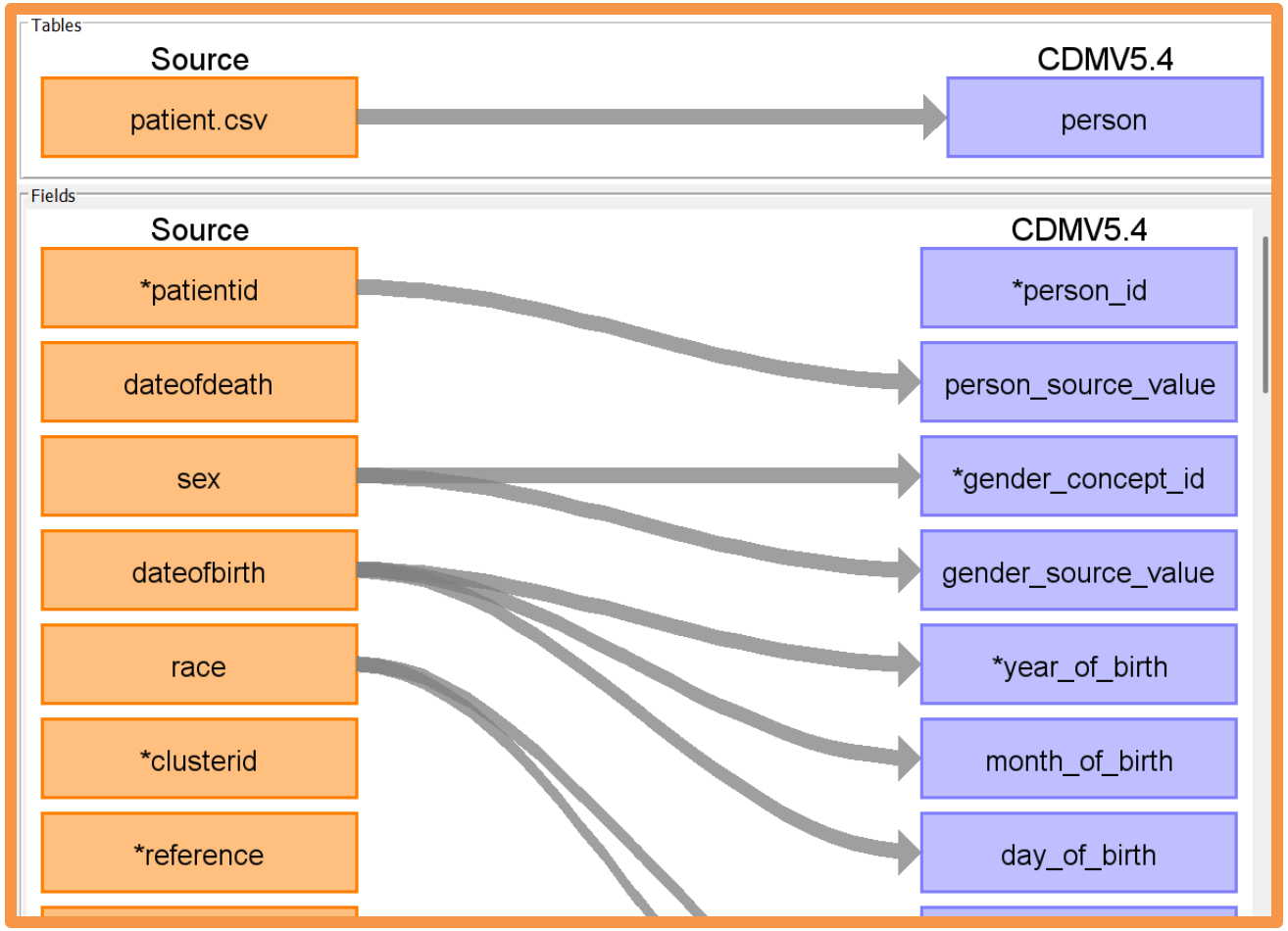
Outputs the foundations for the ETL

- Document describing the needed transformations
- Test framework for ETL verification

ETL TOOLING – RABBIT SUITE

Rabbit-in-a-Hat is a tool used for syntactically mapping source data to target (OMOP) tables and fields

- Input to RiaH is a WR Scan Report
- Output is syntactic mapping documentation
- Enables documenting ETL logic in words (not in screenshot)
- Allows for generation of unit testing framework



ETL TOOLING – RABBIT SUITE

Usagi

Map source codes/values to OMOP standards

- Usagi will run term similarity approach
- May require domain expertise

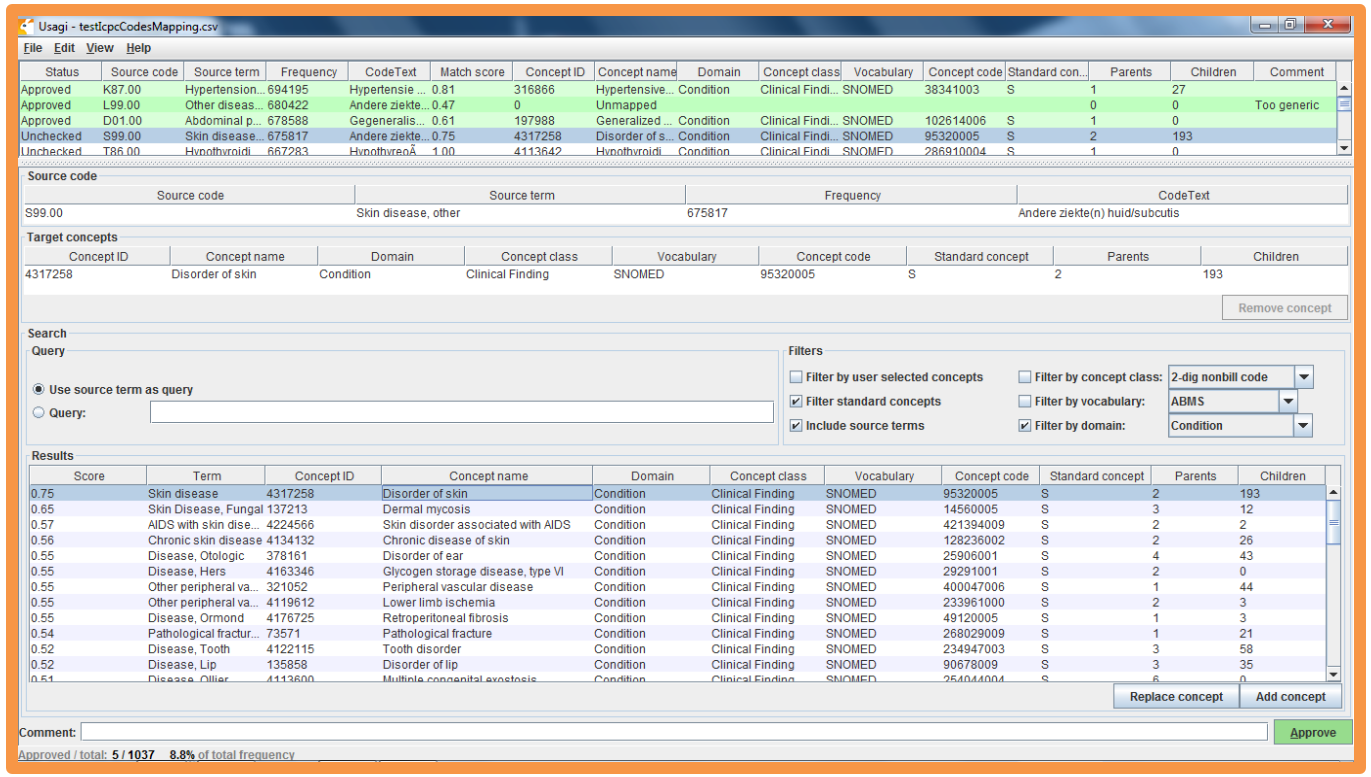
Export into mapping table used in ETL

- Does not have to cover 100% of source values

ETL TOOLING – RABBIT SUITE

Usagi is a helper tool that facilitates **semantically mapping** source values to OMOP concepts, saving a lot of manual work

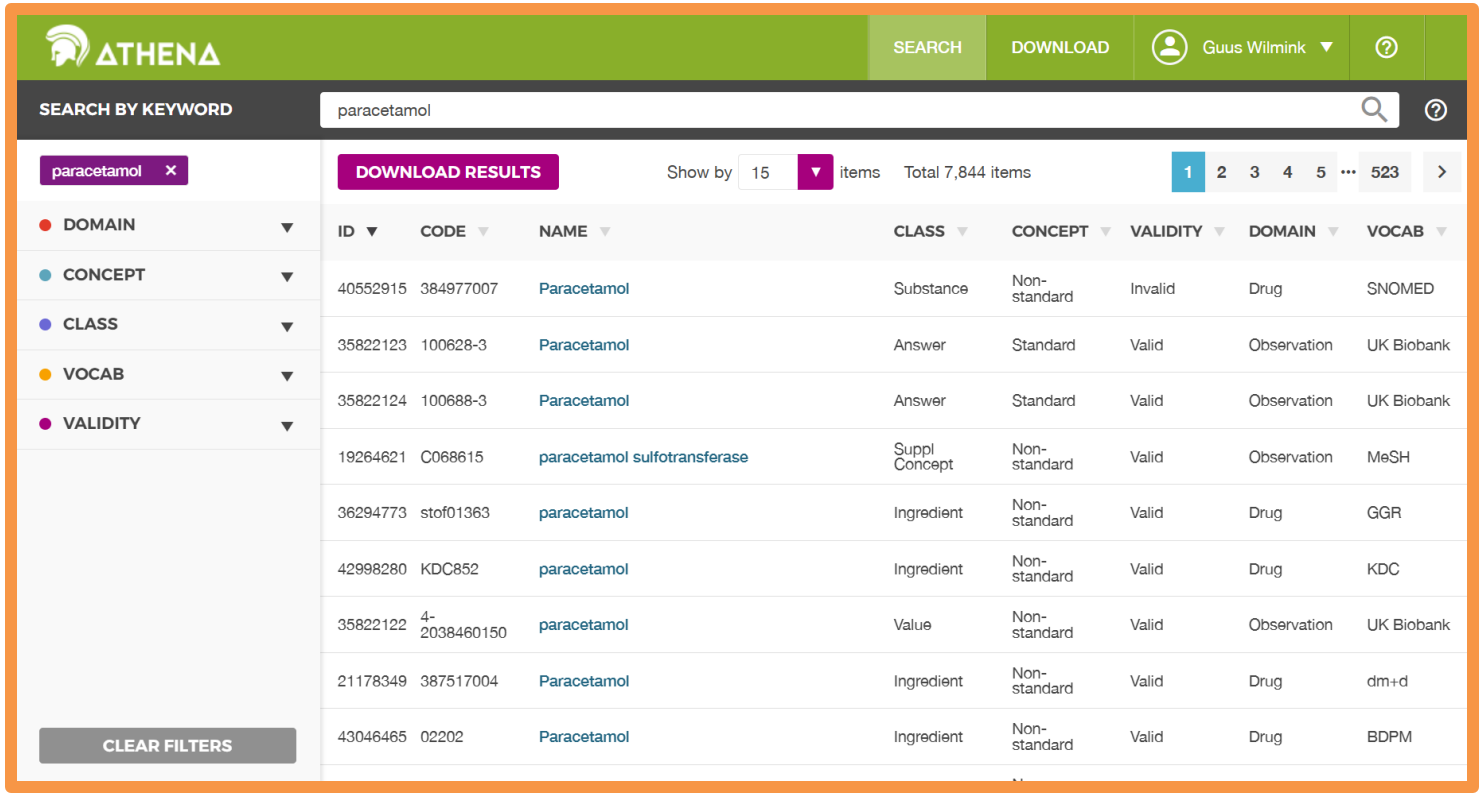
- Requires loading of vocabularies, to be selected by the user
- Will select most similar concept from loaded vocabularies to map from source term
- Initial mapping can be changed using the `query` function
- Facilitated querying using filters
- Collaboration: allows for approval of certain mappings and flagging of uncertain ones



ETL TOOLING – RABBIT SUITE

Athena is a **web application** used for **querying the OMOP vocabularies**

- No option to load source terms; only manual querying possible
- Allows for many filters on OMOP domains, vocabularies, classes, standardness, etc.
- OMOP vocabularies can be accessed and downloaded from Athena
- Required for Usagi usage



More info?
- **The Book of OHDSI** chapter 5
- athena.ohdsi.org

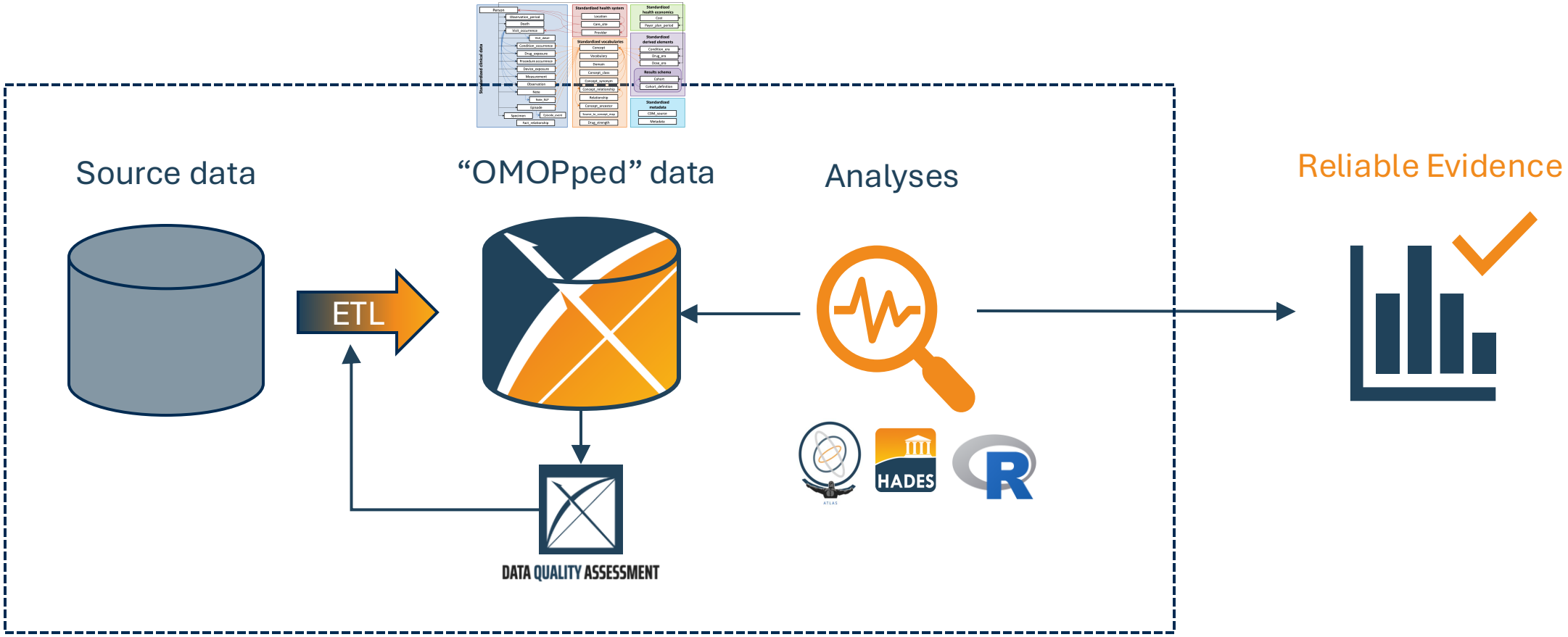
I NEED TO KNOW MORE …

Session
ETL development and updating

After lunch!

13:00 – 15:00		**ETL development and updating** Maxim Moinat (EMC), Anne van Winzum, Stefan Payralbe (The Hyve) Description • Short introduction to OMOP ETL conventions and ETL implementation examples. • Latest developments in available tooling to assist with ETL/mappingTools • Importance and best practices for maintaining and updating ETL/mapping after initial conversion Target audience New and current data holders, ETL developers, data engineers responsible for OMOP CDM conversions and ongoing maintenance.

BUT IS IT GOOD...?



DATA QUALITY ASSESSMENT

EUROPE

SYNTHEA SYNTHETIC HEALTH DATABASE

OVERVIEW

METADATA

RESULTS

ABOUT

DATA QUALITY ASSESSMENT

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-22 14:15:06 in 29 mins

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

DQ Assessment – customizable checks & Thresholds

EUROPE

% of records that violate check

0% = any violation → FAIL
100% = always PASS

Check Category	Check Description	Check Result	Decision Threshold	Pass / Fail
Verification - Plausibility	The number and percent of records with a value in the YEAR_OF_BIRTH field of the PERSON table less than 1850.	0%	0%	PASS
Verification - Plausibility	The number and percent of records with a value in the DAYS_SUPPLY field of the DRUG_EXPOSURE table less than 0.	0%	1%	PASS
Verification - Plausibility	For Hemoglobin A1c percent, the number and percent of records with a value in the VALUE_AS_NUMBER field of the MEASUREMENT table less than 4.	0.01%	5%	PASS
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field CONDITION_CONCEPT_ID in the CONDITION_OCCURRENCE table.	0.02%	5%	PASS
Verification - Completeness	The number and percent of records with a value of 0 in the standard concept field UNIT_CONCEPT_ID in the MEASUREMENT table.	93.66%	5%	FAIL

Erasmus MC
University Medical Center Rotterdam

CDMONBOARDING PACKAGE

EUROPE

DARWIN EU

- R package
- Data table counts
- Vocabulary counts
- Technical Infrastructure Chec
- Creates report

3. Clinical data

3.1. Record counts per OMOP CDM

Table 2. The number of records in all clinical da

Table
measurement
drug_exposure
procedure_occurrence
visit_detail
visit_occurrence
observation
condition_occurrence
condition_era
drug_era
person
observation_period
provider
cdm_source
episode_event
fact_relationship
episode
cost
death
device_exposure
dose_era
note
payer_plan_period
specimen
care_site
location
note_nlp
metadata

Query executed in 3.69 seconds

4.4. Mapped Codes

Table 25. Top 25 mapped drugs. Counts are rounded up to the nearest hundred. Values with a record count <=5 are omitted.

#	Concept id	Concept Name	#Records	%Records
1	19080128	lisinopril 10 MG Oral Tablet	228,200	15.4%
2	19078106	hydrochlorothiazide 25 MG Oral Tablet	188,600	12.7%
3	19073094	amlodipine 2.5 MG Oral Tablet	179,000	12.1%
4	19009384	insulin isophane, human 70 UNT/ML / insulin, regular, human 30 UNT/ML Injectable Suspension [Humulin]	120,600	8.1%
5	19030765	1 ML epoetin alfa 4000 UNT/ML Injection [Epogen]	88,100	5.9%
6	1511248	NDA020503 200 ACTUAT albuterol 0.09 MG/ACTUAT Metered Dose Inhaler	51,500	3.5%
7	19126352	nitroglycerin 0.4 MG/ACTUAT Mucosal Spray	41,900	2.8%
8	1539463	simvastatin 10 MG Oral Tablet	39,800	2.7%
9	1539411	simvastatin 20 MG Oral Tablet	36,600	2.5%
10	40163924	24 HR metformin hydrochloride 500 MG Extended Release Oral Tablet	35,600	2.4%
11	1332419	amlodipine 5 MG Oral Tablet	34,700	2.3%
12	19075601	clopidogrel 75 MG Oral Tablet	32,900	2.2%
13	40163554	warfarin sodium 5 MG Oral Tablet	25,900	1.7%
14	19018935	digoxin 0.125 MG Oral Tablet	25,900	1.7%
15	40171902	verapamil hydrochloride 40 MG	25,100	1.7%
16	40173590	alendronic acid 10 MG Oral Tablet	20,300	1.4%
17	40169217	120 ACTUAT fluticasone propionate 0.044 MG/ACTUAT Metered Dose Inhaler [Flovent]	16,500	1.1%
18	1154380	albuterol 5 MG/ML Inhalation Solution	15,900	1.1%
19	40169281	60 ACTUAT fluticasone propionate 0.25 MG/ACTUAT / salmeterol 0.05 MG/ACTUAT Dry Powder Inhaler	15,700	1.1%

Erasmus MC University Medical Center Rotterdam

DATA NETWORK QUALITY

Individual OMOP database

Network of OMOP databases

Deep Dive into DQ

15:30
–
17:30

Data Quality Assessment Framework & Tools

Clair Blacketer, Anthony Sena (J&J)

Description

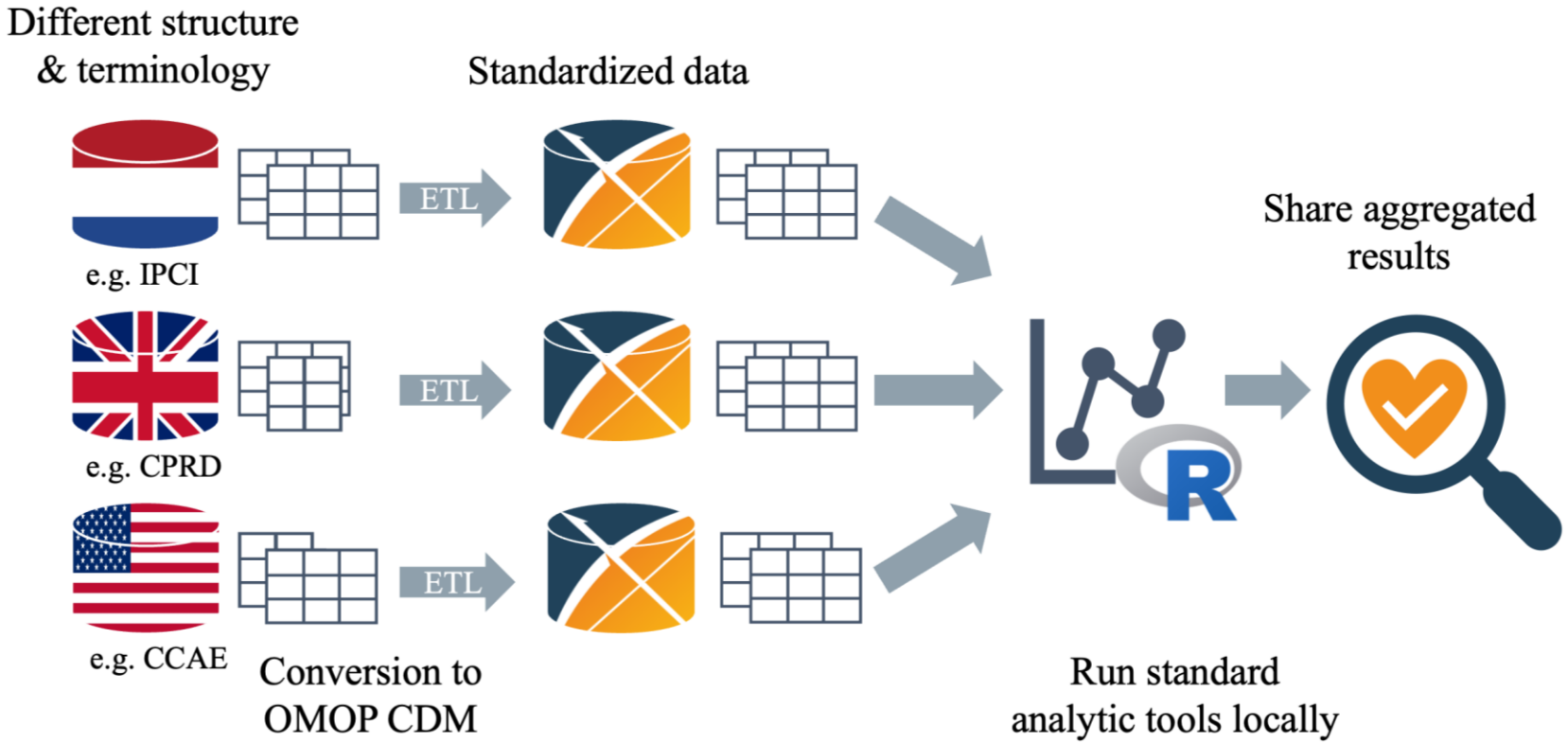
- Data Quality Dashboard (DQD) and other recent developments for tools to assess data quality
- Hands-on exercise for running DQD to identify and address ETL conversion issues
- Data quality considerations for network studies

Target audience

Anyone responsible for assessing and improving the quality of OMOP CDM ETL or data quality for study

What can (currently) be done with the OHDSI tools?

REAL-WORLD EVIDENCE GENERATION USING OMOP CDM



Standardized data enables standardized analytics!

Common format & terminology

Standard input & output
Parametrized choices

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**

ATTRIBUTES OF RELIABLE EVIDENCE

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

Source: https://ohdsi.github.io/TheBookOfOhdsi/EvidenceQuality.html

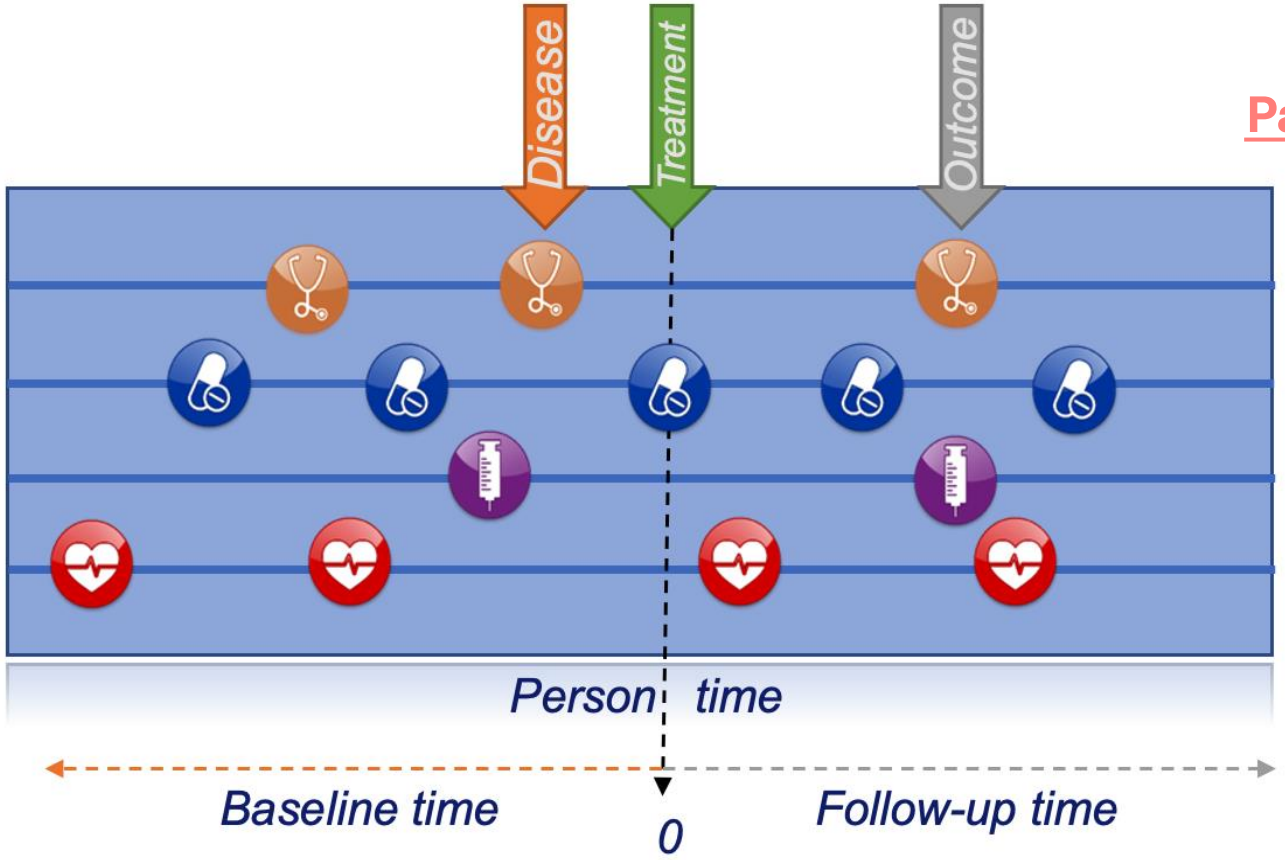
QUESTIONS ASKED ACROSS THE PATIENT JOURNEY

EUROPE

Clinical characterization

Which patients take which treatments?

How many patients experienced complications?



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?

More info?
- **The Book of OHDSI** chapter 7

Population-level effect estimation

Does treatment cause an outcome?

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

Erasmus MC University Medical Center Rotterdam

WHAT DO YOU DO?

Population-Level Effect Estimation

- What are the causal effects?

Patient-Level Prediction

- What will happen to me?

Clinical Characterisation

- What happened to the patients?

Hernán MA, Hsu J, Healy B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. CHANCE 2019;32(1): 42-49, DOI: 10.1080/09332480.2019.1579578.

TREATMENT PATTERNS ACROSS DISEASE DOMAINS *

= LARGE-SCALE EVIDENCE (CHARACTERIZATION)

EUROPE

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

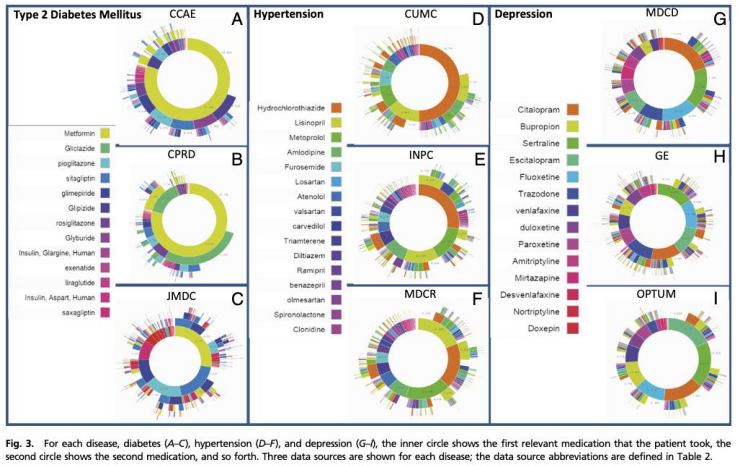


Fig. 3. For each disease, diabetes (*A–C*), hypertension (*D–F*), and depression (*G–I*), the inner circle shows the first relevant medication that the patient took, the second circle shows the second medication, and so forth. Three data sources are shown for each disease; the data source abbreviations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

Hripcsak et al. (2016)

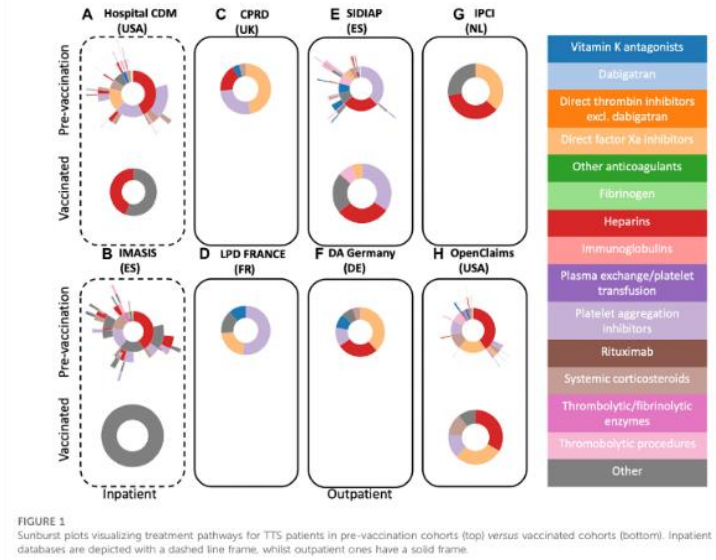
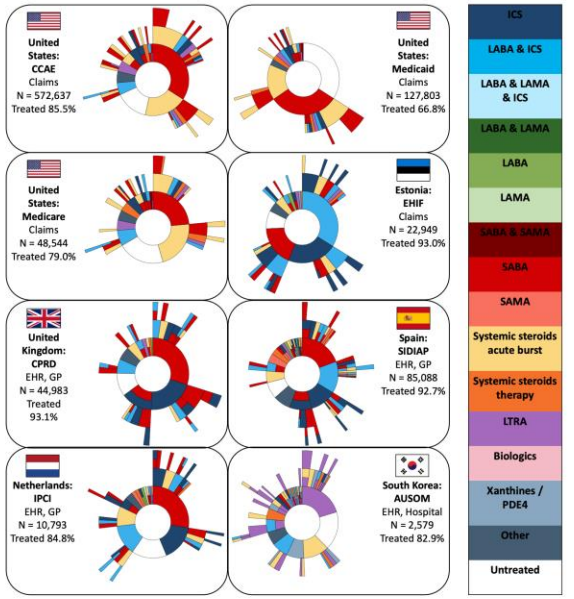


FIGURE 1
Sunburst plots visualizing treatment pathways for TTS patients in pre-vaccination cohorts (top) *versus* vaccinated cohorts (bottom). Inpatient databases are depicted with a dashed line frame, whilst outpatient ones have a solid frame.

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)

Erasmus MC University Medical Center Rotterdam

TREATMENT PATTERNS ACROSS DISEASE DOMAINS *

= LARGE-SCALE EVIDENCE (CHARACTERIZATION)

Computer Methods and Programs in Biomedicine
Volume 225, October 2022, 107081

TreatmentPatterns: An R package to facilitate the standardized development and analysis of treatment patterns across disease domains

Aniek F. Markus [a], Katia M.C. Verhamme [a b], Jan A. Kors [a], Peter R. Rijnbeek [a]

Show more

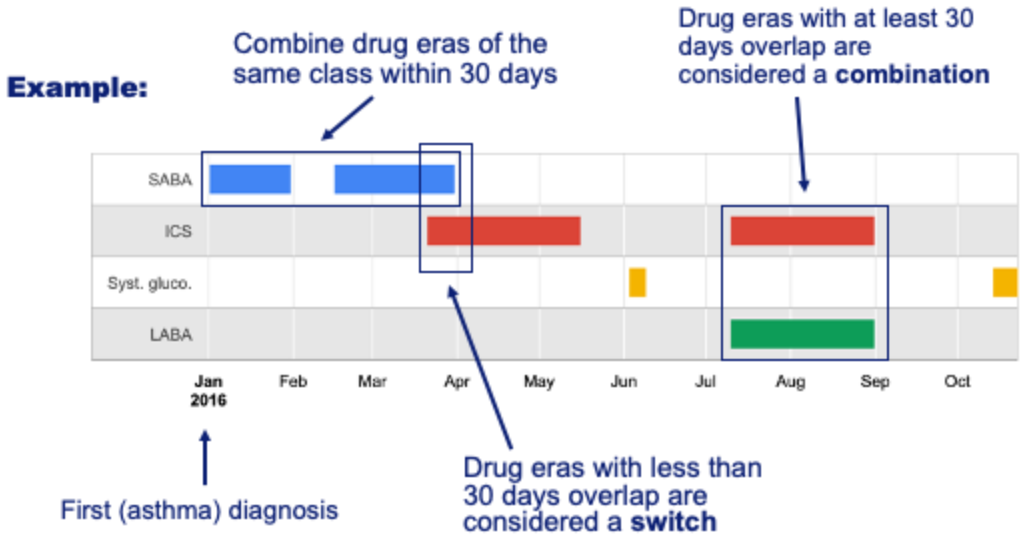
Add to Mendeley Share Cite

https://doi.org/10.1016/j.cmpb.2022.107081

Get rights and content

Highlights

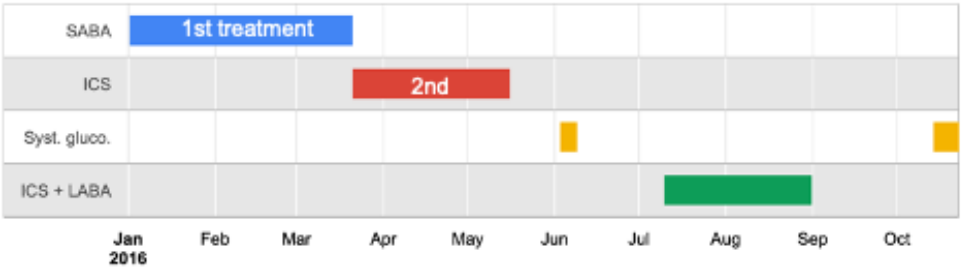
- TreatmentPatterns R package for development and analysis of treatment patterns.
- Formally defines and implements the process of constructing treatment pathways.
- Accessible, standardized, and interpretation friendly tool for a broad audience.
- Supports the accumulation of knowledge on treatment patterns across disease domains.
- Example study in Dutch IPCI database to demonstrate functionalities of the package.



More info?

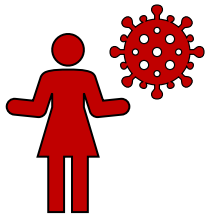
- **The Book of OHDSI** chapter 11.3 + 11.9
- **EHDEN academy course** in progress

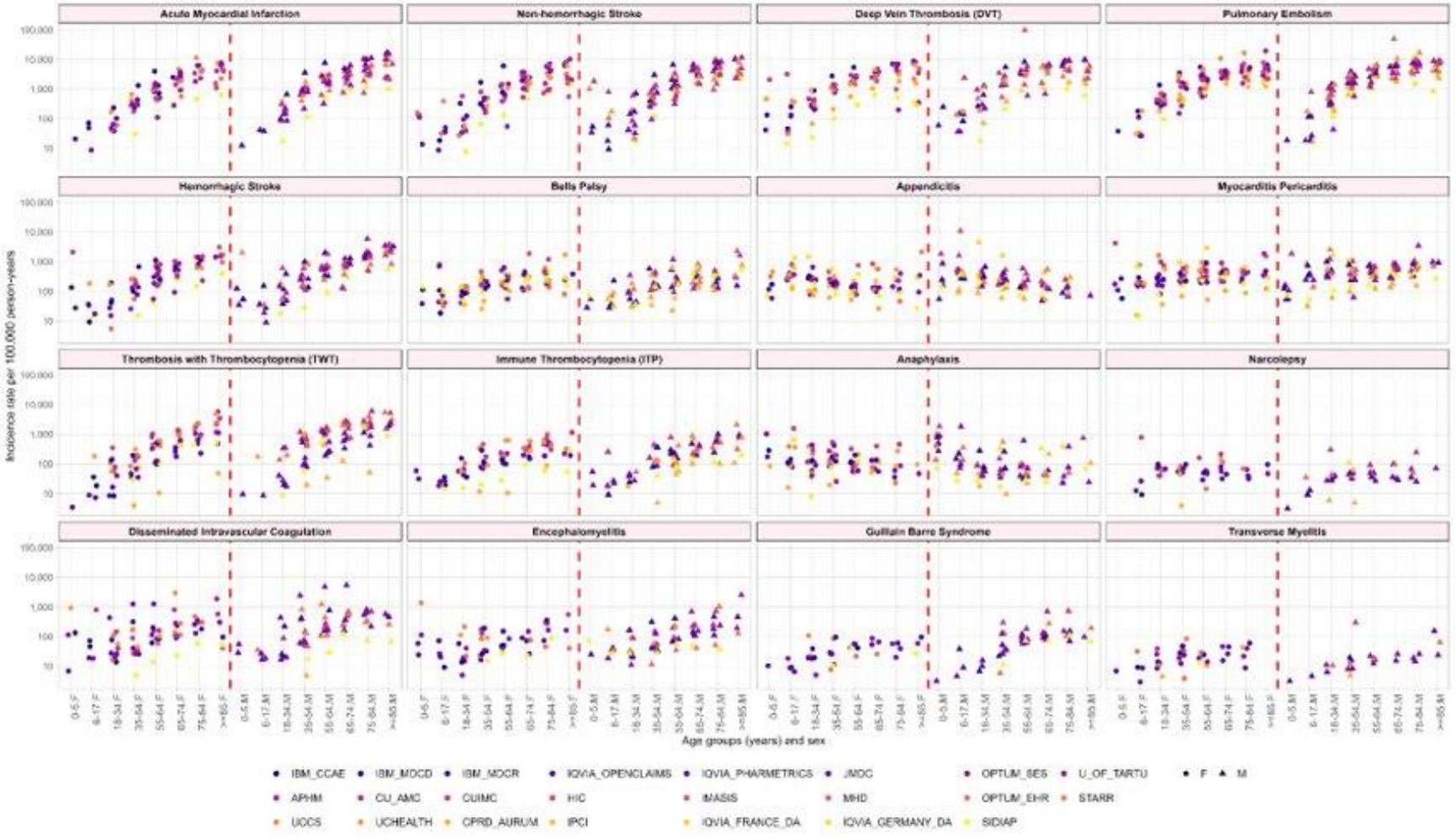
Example: SABA – ICS – Systemic glucocorticoids – ICS + LABA



https://darwin-eu.github.io/TreatmentPatterns/

CHARACTERIZING VACCINE ADVERSE EVENTS IN COVID-19 PATIENTS ACROSS THE OHDSI NETWORK * = LARGE-SCALE EVIDENCE



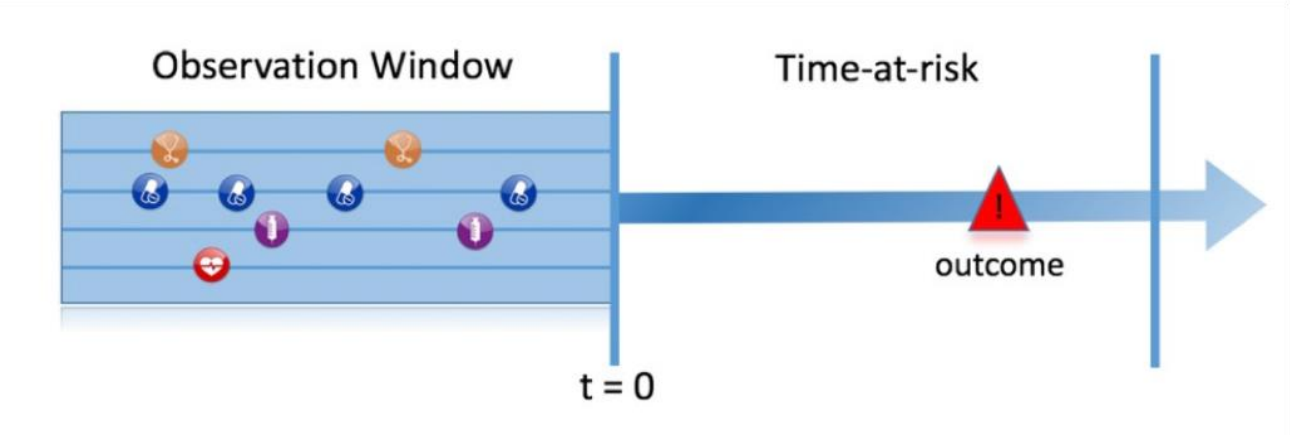


Voss et al. (2023)

26 databases
11 countries

PATIENT LEVEL PREDICTION

- What will happen to me?



Patient-level prediction

PatientLevelPrediction

Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.

Learn more…

DeepPatientLevelPrediction

Performing patient level prediction using deep learning

Learn more…

EnsemblePatientLevelPrediction

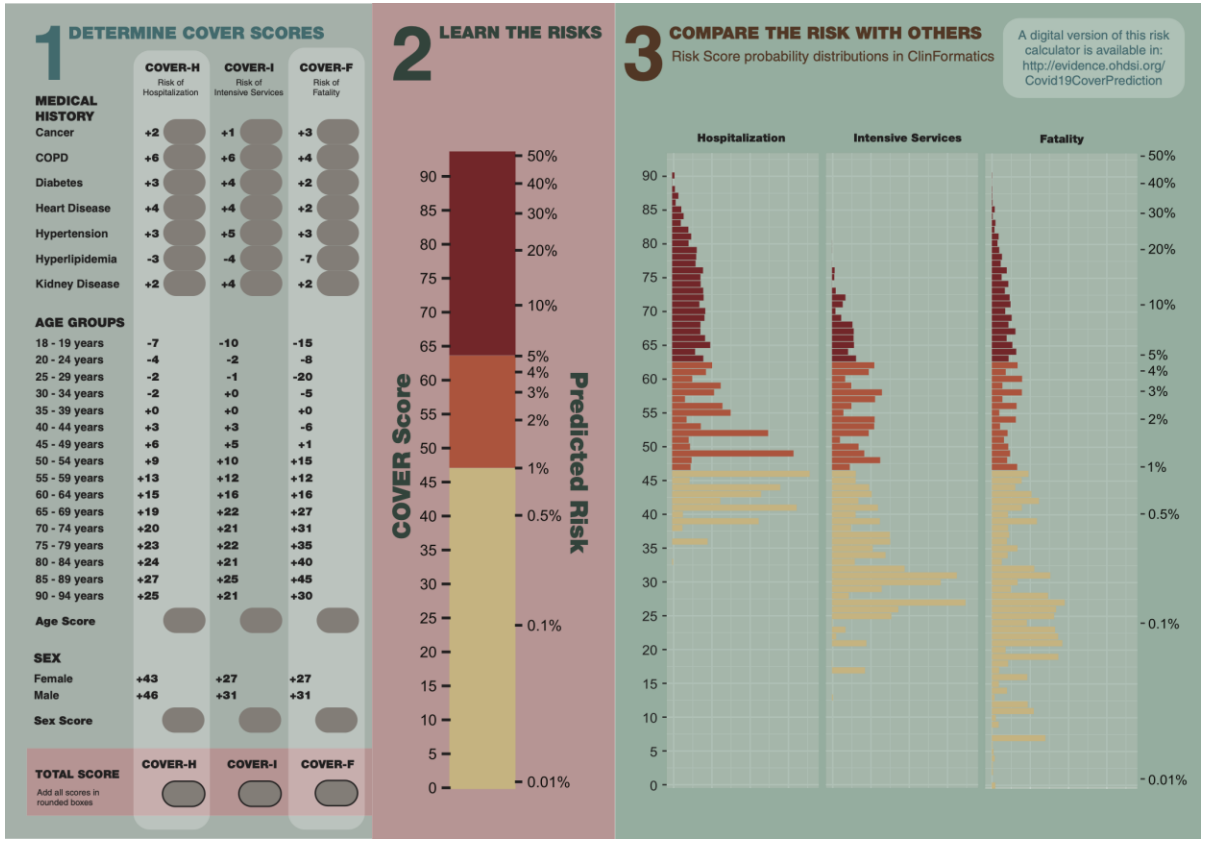
Building and validating ensemble patient-level predictive models.

Learn more…

SEEK COVER * = RAPID RESPONSE

Objective: develop and externally validate **COV**ID-19 **E**stimated **R**isk scores that quantify a patient's risk of hospital admission, hospitalization requiring intensive services or fatality.

- 14 data sources from 6 countries
- Externally validated in 44,507 COVID cases from 5 data sources in South Korea, Spain, USA



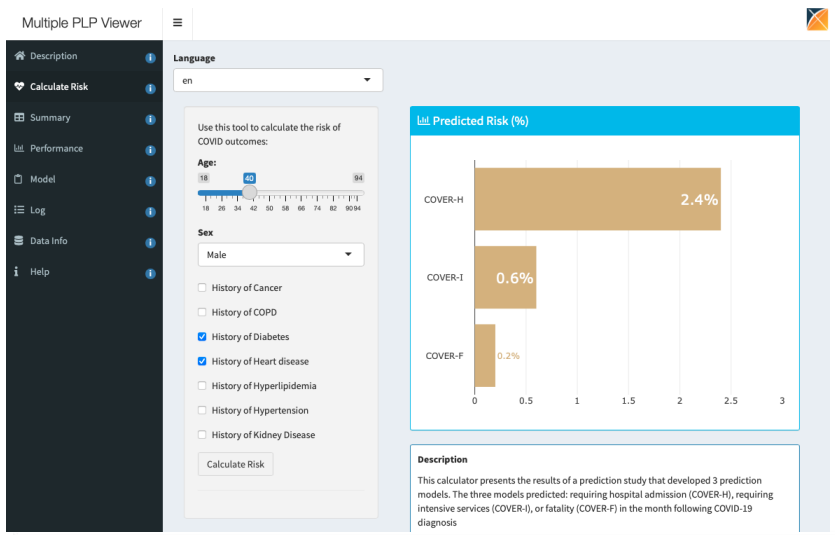
https://data.ohdsi.org/Covid19CoverPrediction/

Williams, R.D., Markus, A.F., Yang, C. et al. Seek COVER: using a disease proxy to rapidly develop and validate a personalized risk calculator for COVID-19 outcomes in an international network. BMC Med Res Methodol 22, 35 (2022). https://doi.org/10.1186/s12874-022-01505-z

SEEK COVER * = RAPID RESPONSE

- COVER interactive website to provide risk scores.

- **Impact**: Health minister of Catalonia Spain explicitly mentioned the COVER index as one of the indicators they used to measure the impact of a given outbreak.



LEARNING AND ESTABLISHING BEST PRACTICES (PLP) *

ELSEVIER

International Journal of Medical Informatics

Volume 163, July 2022, 104762

Logistic regression models for patient-level prediction based on massive observational data: Do we need all data?

Luis H. John [a], Jan A. Kors [a], Jenna M. Reps [b], Patrick B. Ryan [b], Peter R. Rijnbeek [a]

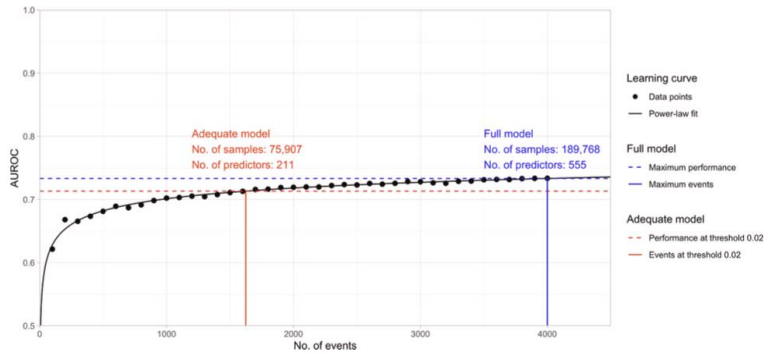


Fig. 1. Learning curve for the prediction of venous thromboembolic events in patients with hypertension using data from MDCR. The horizontal lines indicate the maximum performance of the fitted curve (blue) and the performance at a threshold of 0.02 (red). The vertical lines denote the maximum number of events (blue) and the adequate number of events (red). Number of samples and predictors shown for the adequate model pertain to the model at the closest data point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

"Our results suggest that in most cases **only a fraction of the available data was sufficient** to produce a model close to the performance of one developed on the full data set, but with a substantially reduced model complexity."

Reps et al. BMC Med Inform Decis Mak (2021) 21:43
https://doi.org/10.1186/s12911-021-01408-x

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

An empirical analysis of dealing with patients who are lost to follow-up when developing prognostic models using a cohort design

Jenna M. Reps[1*], Peter Rijnbeek[2], Alana Cuthbert[3], Patrick B. Ryan[1], Nicole Pratt[4] and Martijn Schuemie[1]

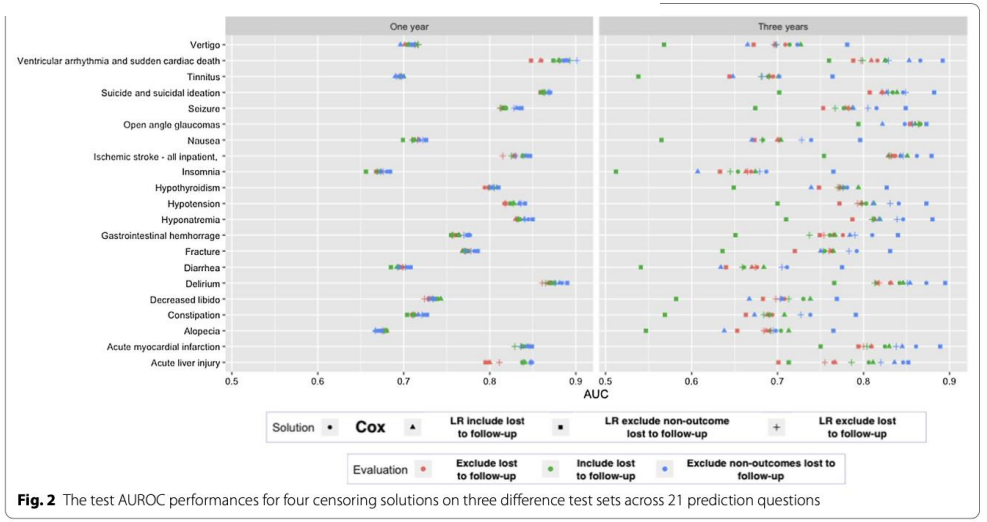


Fig. 2 The test AUROC performances for four censoring solutions on three difference test sets across 21 prediction questions

"Based on this research it appears that it is best to develop models using data that **includes patients that are lost to follow-up**. However, recalibration is likely to be required as this strategy appears to result in models that under-estimate risk."

LEARNING AND ESTABLISHING BEST PRACTICES (PLP) *

EUROPE

Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data

Jenna M Reps ✉, Martijn J Schuemie, Marc A Suchard, Patrick B Ryan, Peter R Rijnbeek

Journal of the American Medical Informatics Association, Volume 25, Issue 8, August 2018, Pages 969–975, https://doi.org/10.1093/jamia/ocy032
Published: 27 April 2018 **Article history** ▾

PDF | Split View | Cite | Permissions | Share ▾

Abstract

Objective

To develop a conceptual prediction model framework containing standardized steps and describe the corresponding open-source software developed to consistently implement the framework across computational environments and observational healthcare databases to enable model sharing and reproducibility.

PatientLevelPrediction 6.3.2 | Get started | Videos | Reference | Articles | Tutorial | Best Practices | Changelog | HADES

- Adding Custom Feature Engineering Functions
- Adding Custom Patient-Level Prediction Algorithms
- Adding Custom Sampling Functions
- Adding Custom Data Splitting Functions
- Best Practice Research
- Automatically Build Multiple Patient-Level Predictive Models
- Building patient-level predictive models
- Creating Learning Curves
- Making patient-level predictive network study packages
- Patient-Level Prediction Installation Guide
- Demo Videos

PatientLevelPrediction

R-CMD-check passing | codecov 90%

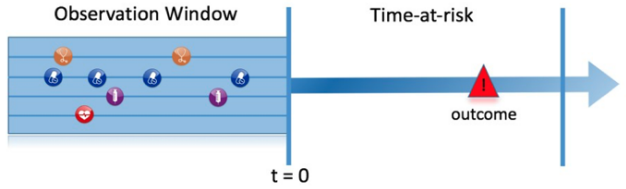
PatientLevelPrediction is part of HADES.

Introduction

PatientLevelPrediction is an R package for building and validating patient-level predictive
Data Model format.

Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018;25(8):969-975.

The figure below illustrates the prediction problem we address. Among a population at risk, we aim to predict which patients at a defined moment in time (t = 0) will experience some outcome during a time-at-risk. Prediction is done using only information about the patients in an observation window prior to that moment in time.



Developers

Jenna Reps, Author, maintainer
Martijn Schuemie, Author
Marc Suchard, Author
Patrick Ryan, Author
Peter Rijnbeek, Author
Egill Fridgeirsson, Author

Best practice publications using the OHDSI PatientLevelPrediction framework

Topic	Research Summary	Link
Problem Specification	When is prediction suitable in observational data?	Guidelines needed
Data Creation	Comparison of cohort vs case-control design	Journal of Big Data
Data Creation	Addressing loss to follow-up (right censoring)	BMC medical informatics and decision makingk
Data Creation	Investigating how to address left censoring in features construction	BMC Medical Research Methodology
Data Creation	Impact of over/under-sampling	Paper under review
Data Creation	Impact of phenotypes	Study Done - Paper submitted
Model development	How much data do we need for prediction - Learning curves at scale	International Journal of Medical Informatics
Model development	What impact does test/train/validation design have on model performance	BMJ Open
Model development	What is the impact of the classifier	JAMIA
Model development	Can we find hyper-parameter combinations per classifier that consistently lead to good performing models when using claims/EHR data?	Study needs to be done
Model development	Can we use ensembles to combine different algorithm models within a database to improve models transportability?	Study Complete
Model development	Can we use ensembles to combine models developed using different databases to improve models transportability?	BMC Medical Informatics and Decision Making
Evaluation	How should we present model performance? (e.g., new visualizations)	JAMIA Open

More info?
- **The Book of OHDSI** chapter 13
- **EHDEN academy course Patient Level Prediction + R for Patient Level Prediction**

R package: https://ohdsi.github.io/PatientLevelPrediction/

Erasmus MC University Medical Center Rotterdam

OHDSI LEGEND STUDIES * = SCALABLE AND REPRODUCIBLE

Journal of the American Medical Informatics Association, 0(0), 2020, 1268–1277
doi: 10.1093/jamia/ocaa124
Research and Applications

Research and Applications

Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study

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

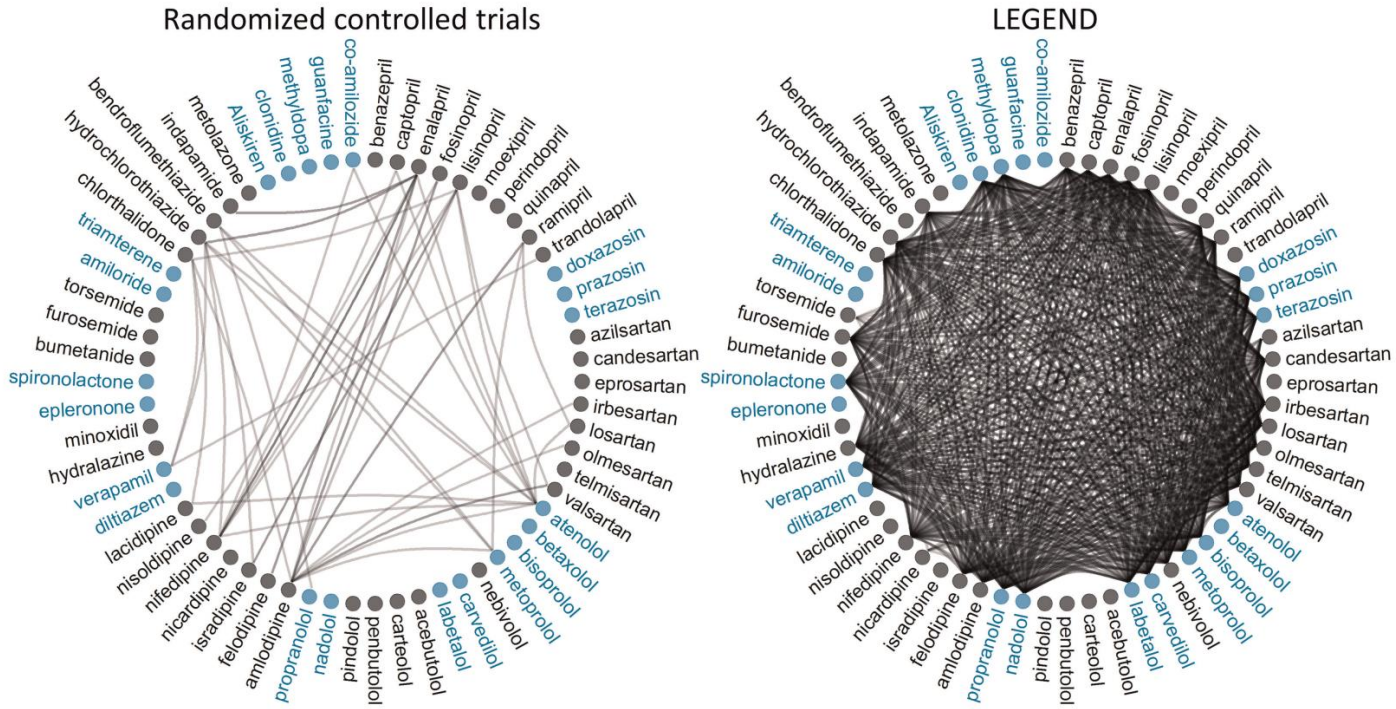
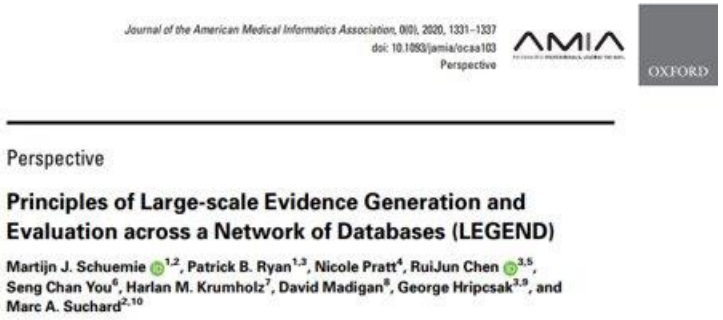


Figure 3. Comparisons of single-drug hypertension treatments in randomized controlled trials (left) and in LEGEND (right). Each circle represents an ingredient. Color groupings indicate drug classes. A line between circles indicates the 2 drugs are compared in at least 1 study.

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.

OHDSI LEGEND STUDIES * = SCALABLE AND REPRODUCIBLE

EUROPE



Journal of the American Medical Informatics Association, 0(0), 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], RuiJun Chen[3,5], Seng Chan You[6], Harlan M. Krumholz[7], David Madigan[8], George Hripcsak[3,9], and Marc A. Suchard[2,10]

Towards Reliable Evidence …

Some current practices across the broader research community

- Examining one target-comparator pair at a time
- Not using appropriate methods to control for bias
- Modify the design until significant results are found

Large-scale evidence generation across a network of databases (LEGEND)

- * Pre-specified fixed design and dissemination of the results regardless of the estimates (avoid publication bias)
- * Systematic process across all research questions
- * Large-scale: looking at thousands of target-comparator pairs at a time
- * Use of best practices: LSPS, extensive diagnostics, negative and positive controls

More info?

- **OHDSI website**
https://www.ohdsi.org/legend-oct2021-update/
- **Various publications**
e.g. 10.1093/jamia/ocaa103

Erasmus MC University Medical Center Rotterdam

OHDSI Tools (Open-source)

EUROPE

Source data

Summarize the source data

WhiteRabbit

Create ETL design

Rabbit In A Hat

Semi-automated vocabulary mapping

Usagi

OMOP-CDM data

Analyses

Reliable Evidence

HADES

R

ARES

Generate descriptive statistics on OMOP CDM

ACHILLES

Perform data quality checks

DATA QUALITY ASSESSMENT

Phenotype development

PHOEBE

Codelist Generator

Exploring vocabulary

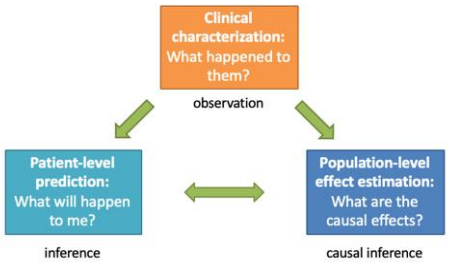
ATHENA

Cohort Diagnostics R package

PheValuator

Phenotype Library

+



Study design

Characterization R package

PatientLevel Prediction R package

CohortMethodR package

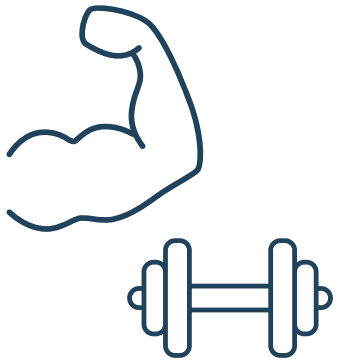
ATLAS

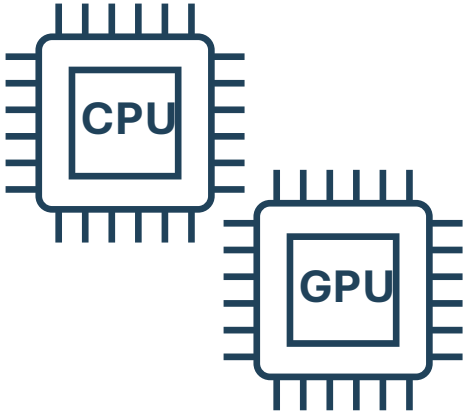
ATLAS: An open-source software tool for data exploration and scientific analyses on OMOP CDMdata

- Given what you know now, discuss with your neighbour:
  - **Which opportunities do you see for using the OHDSI tools?**
  - **Which tools can you see yourself using?**
  - **Are you ready to start using the tools? Why (not)?**

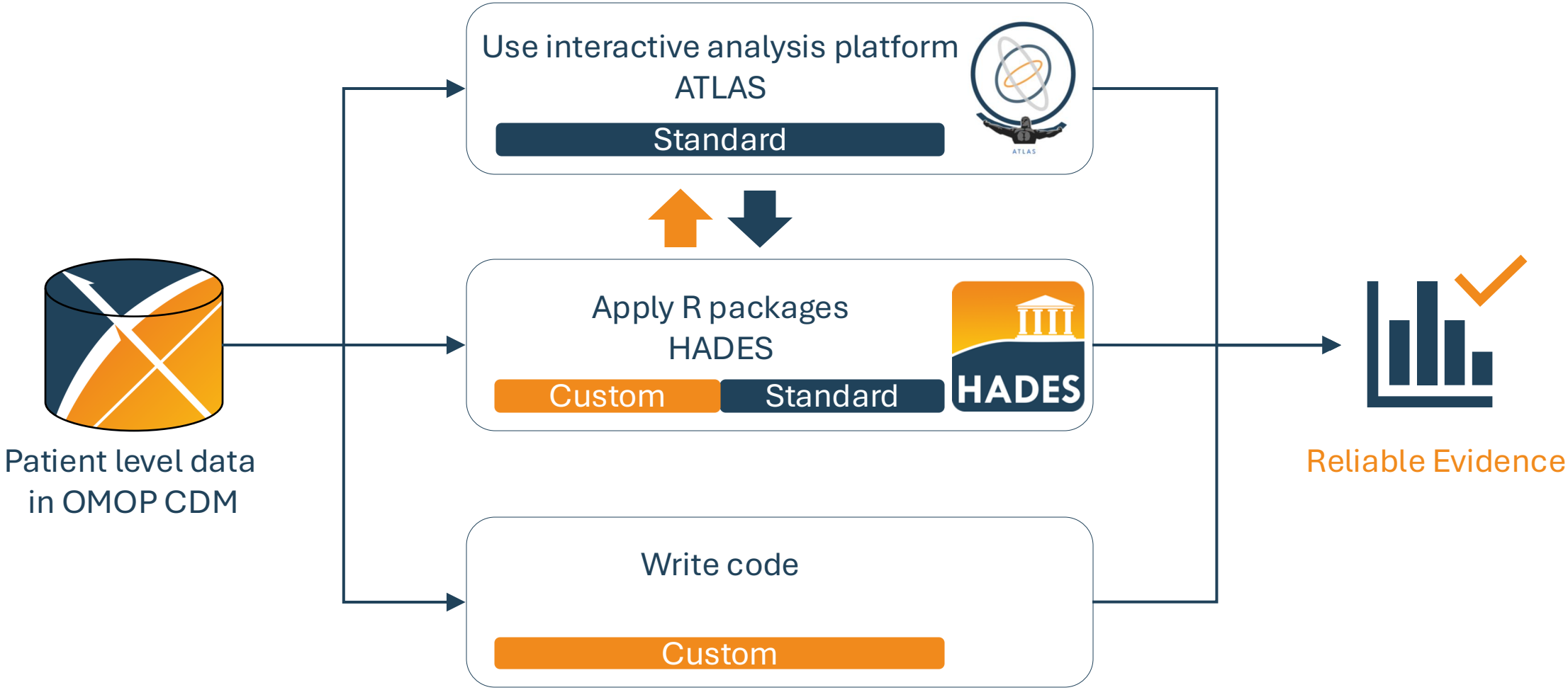
What does it take to be able to use the tools?

WHAT DOES IT TAKE TO BE ABLE TO USE THE TOOLS?





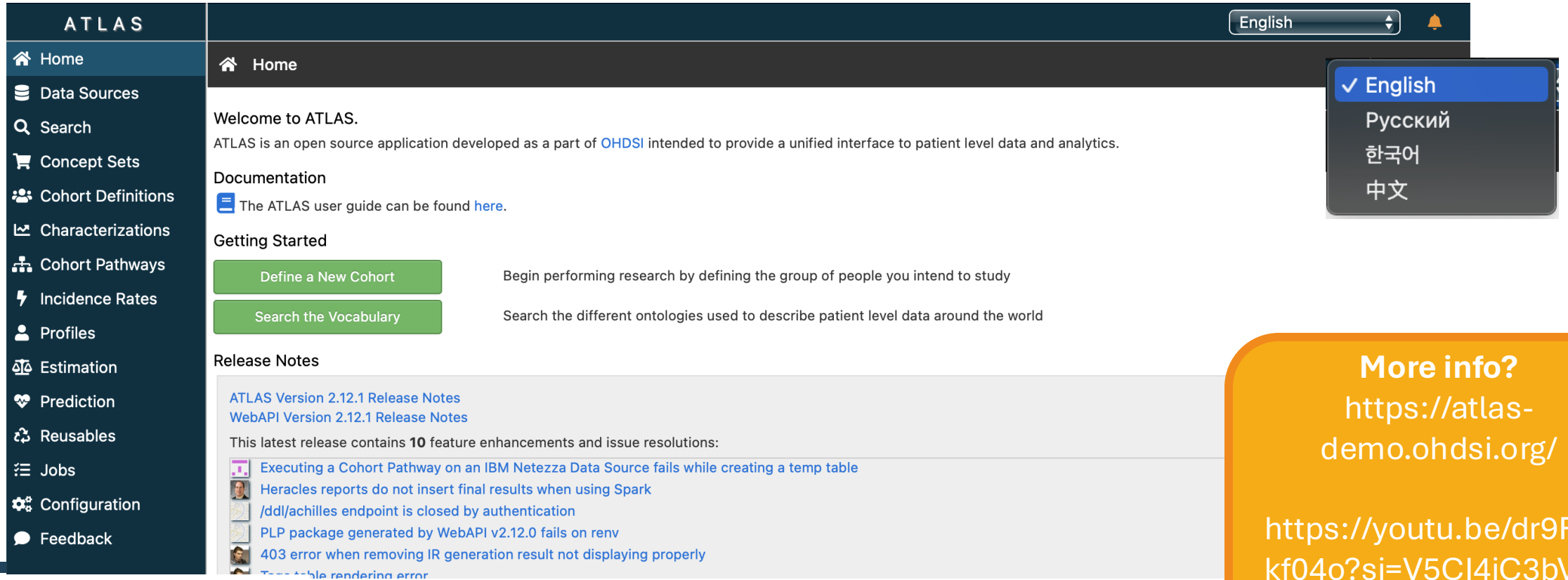
OHDSI Roads to Reliable Evidence



ATLAS

EUROPE

ATLAS is a free, publicly available web-based, open-source software application developed by the OHDSI community to support the design and execution of observational analyses to generate real world evidence from patient level observational data. Atlas is an open science analytics platform that can be installed locally within your institution to perform analyses across one or more observational databases which have been standardized to the OMOP Common Data Model V5 and can facilitate exchange of analysis designs with any other organizations across the OHDSI community who have adopted the same open science community standards and tools.

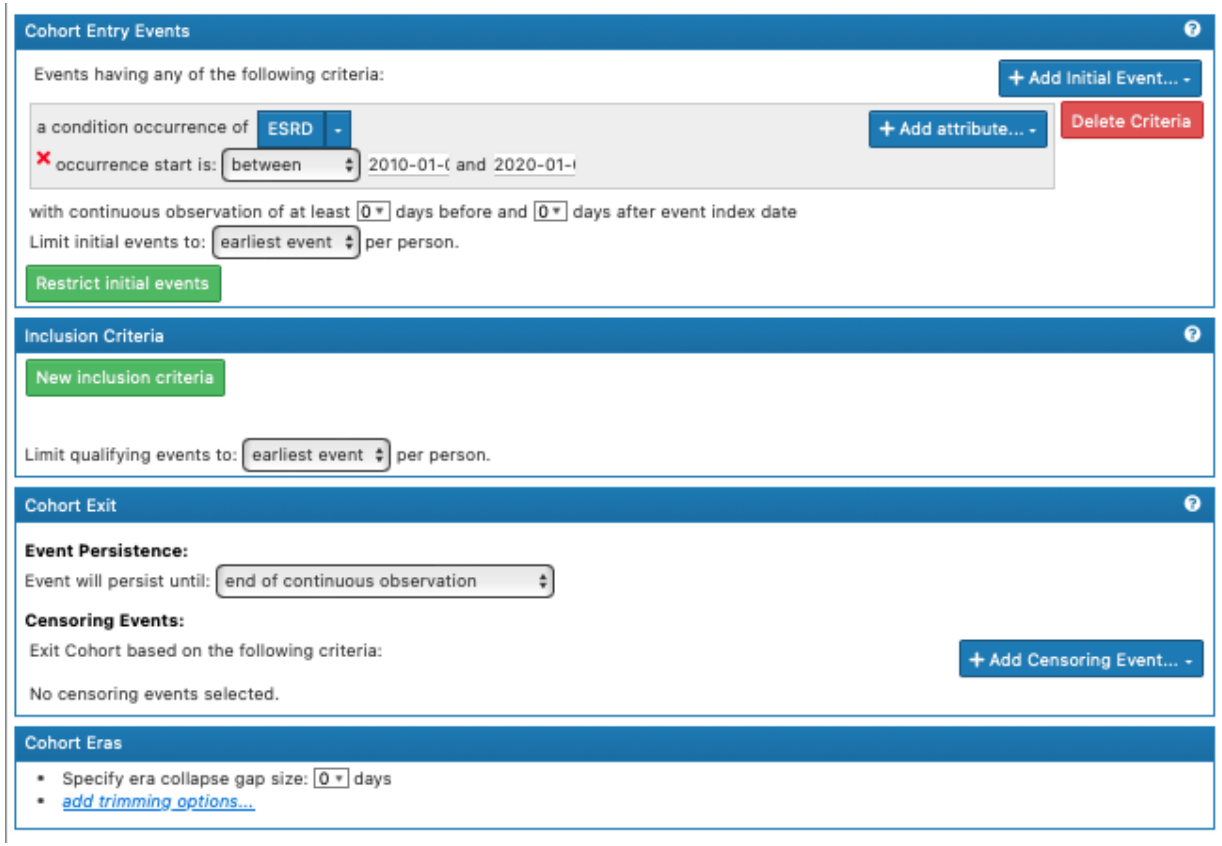


More info?
https://atlas-demo.ohdsi.org/

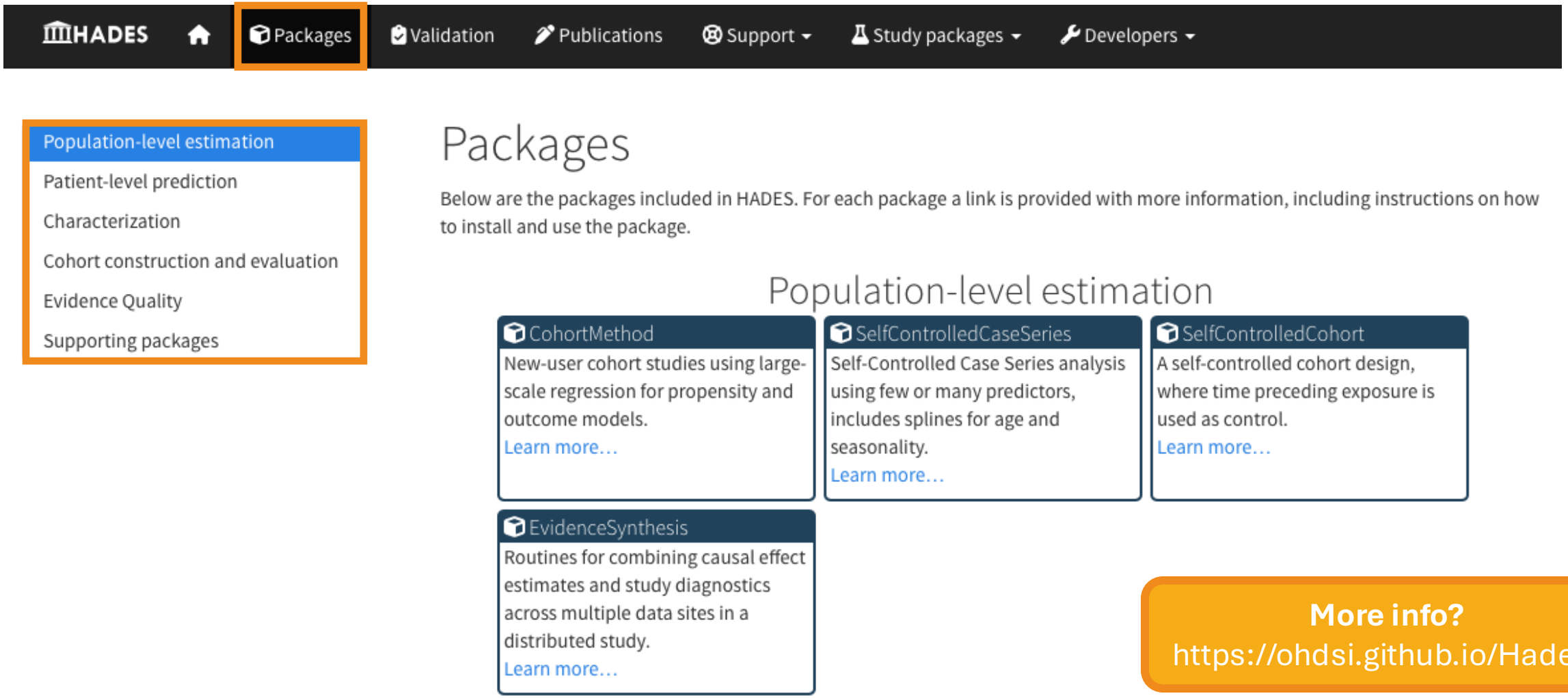
https://youtu.be/dr9FhEkf04o?si=V5CI4jC3bVLbnFi3

EXAMPLE: DEFINING A COHORT (E.G. STUDY POPULATION)

More info?
- **The Book of OHDSI** chapter 10

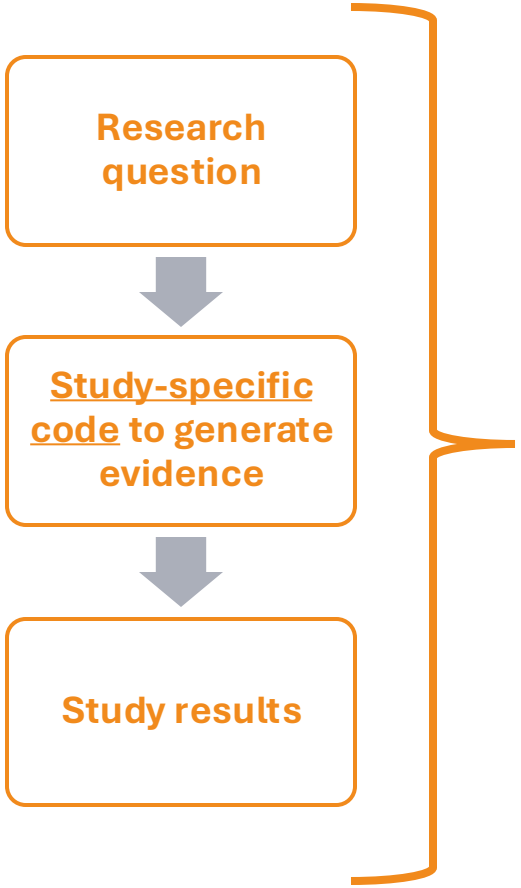


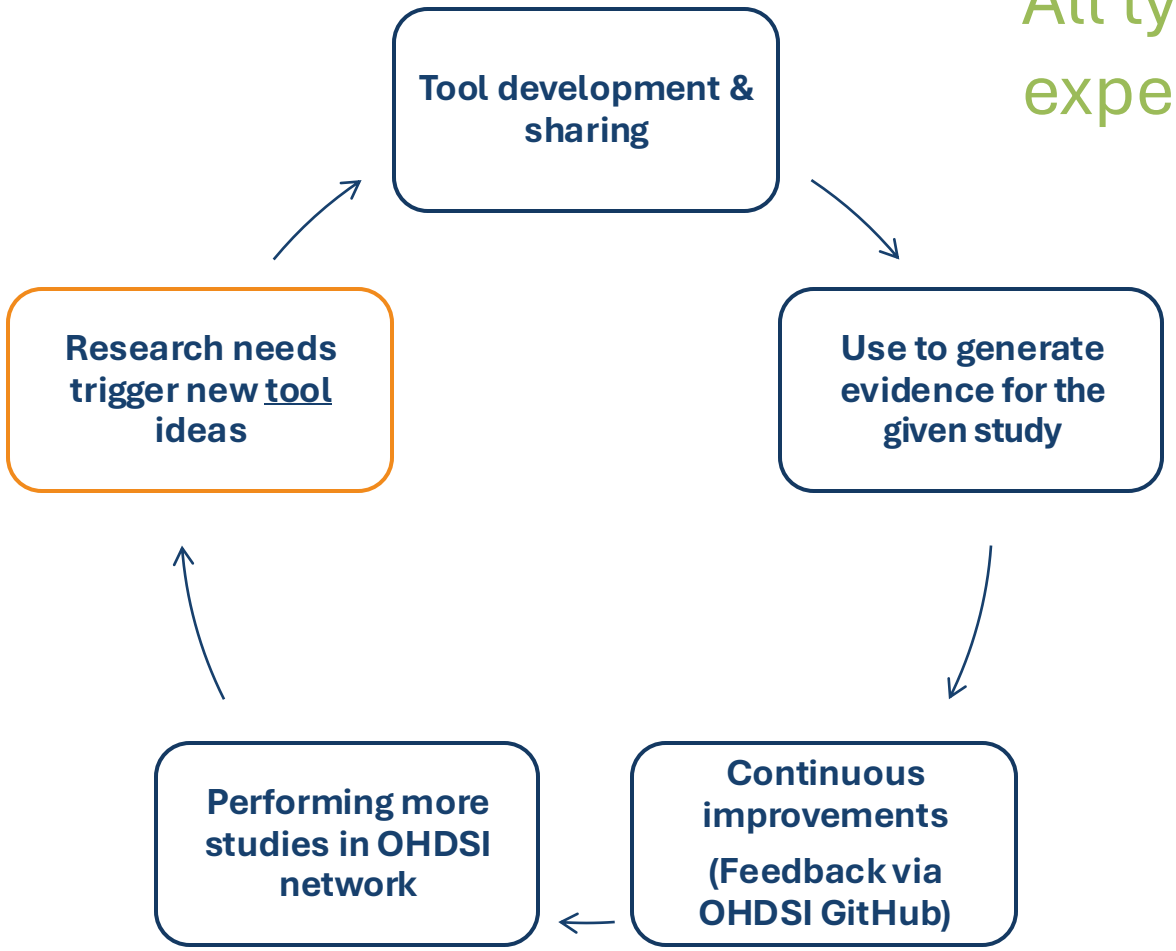
HADES: Open-source R Packages



More info?
https://ohdsi.github.io/Hades/

OHDSI = TEAMWORK





All types of expertise needed!

WRAPPING UP

- Any questions for us?

  Tackle us, or our colleagues in the next few days

TIPS FOR OHDSI 2026 SYMPOSIUM

- Visit your national node

(poster area on Monday)

- Mingle! (the unscheduled events are often most valuable ;-))

OHDSI EUROPE 2026 – WHAT'S AFTER LUNCH?

13:00 – 15:00		**ETL development and updating** Maxim Moinat (EMC), Anne van Winzum, Stefan Payralbe (The Hyve) Description • Short introduction to OMOP ETL conventions and ETL implementation examples. • Latest developments in available tooling to assist with ETL/mappingTools • Importance and best practices for maintaining and updating ETL/mapping after initial conversion Target audience New and current data holders, ETL developers, data engineers responsible for OMOP CDM conversions and ongoing maintenance.	**Phenotyping in OHDSI: pipelines, steps & tools** Anna Ostropolets (J&J), Maria Khitrun, Azza Shoaibi, Dmitro Dymshyts, Anna Saura Lazaro Description • Explore existing pipelines for phenotyping within OHDSI • Explore open-source tools that exists to help researchers with individual phenotyping tasks, including creating clinical description for the clinical idea, literature review and concept sets, phenotype evaluation and storage & phenotype maintenance across different OHDSI Standardized Vocabularies versions Target audience Researchers who design studies and build phenotypes and anyone who want to learn more about phenotyping

Quiz

- What is the current version of the OMOP-CDM?
- To participate in a network study, you need to share your source data. True or false?
- If you don't have data, you can't join OHDSI. True or false?
- Treatment patterns are awesome. True or false?
- OHDSI will add value to my organization. True or false?

WANT TO KNOW MORE ABOUT OHDSI?

Learn more:

- Book of OHDSI: https://ohdsi.github.io/TheBookOfOhdsi/
- Follow tutorials on the EHDEN academy: https://academy.ehden.eu
- Read publications about OHDSI tools and pipelines

➔ Start getting hands-on experience!

Stay updated: https://www.ohdsi.org/ohdsi-news-updates/

- Weekly community calls – Tuesday 17:00-18:00
- Plenty of workgroups (https://www.ohdsi.org/workgroups/)