Will AI Help or Hurt OHDSI in its Mission to Generate RWE?

OHDSI Community Call

Sept. 22, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Upcoming Community Calls

Date	Topic
Sept. 22	Will AI Help or Hurt OHDSI in its Mission to Generate RWE
Sept. 29	OHDSI/OMOP Research Spotlight
Oct. 6	DARWIN EU Progress
Oct. 13	Global Symposium Poster Preview / Final Logistics
Oct. 20	NO CALL – Global Symposium
Oct. 27	Welcome to OHDSI: What We Do, How Can You Make an Impact
Nov. 3	Meet the Titans
Nov. 10	Collaborator Showcase Honoree Presentations
Nov. 17	Early-Stage Researcher Presentations

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Eun Hee Yu, Hyun Joo Lee, Young Mi Han, and Jong Kil Joo** on the recent publication of **Association Between Endometriosis and Autoimmune Thyroid Disease Using a Multicenter Observational Medical Outcomes Partnership (OMOP) Common Data Model** in the *Journal of Clinical Medicine*.

Journal of Clinical Medicine

MDPI

Article

Association Between Endometriosis and Autoimmune Thyroid Disease Using a Multicenter Observational Medical Outcomes Partnership (OMOP) Common Data Model

Eun Hee Yu [1,2], Hyun Joo Lee [1,2], Young Mi Han [3] and Jong Kil Joo [1,2,*]

[1] Department of Obstetrics and Gynecology and Biomedical Research Institute, Pusan National University Hospital, Busan 49241, Republic of Korea; hoyyeh@naver.com (E.H.Y.); atouchof@pusan.ac.kr (H.J.L.)
[2] Department of Obstetrics and Gynecology, School of Medicine, Pusan National University, Busan 49241, Republic of Korea
[3] R&D Team, Visual Terminology Co., Ltd., Busan 49241, Republic of Korea; maryhan01@gmail.com
* Correspondence: jongkilj@hanmail.net

Abstract

Background: Endometriosis affects approximately 6–10% of women of reproductive age—an estimate that varies substantially with the diagnostic standard applied—and is increasingly recognized as a systemic inflammatory disease. Growing evidence implicates shared immunological pathways between endometriosis and autoimmune thyroid disease, yet large-scale real-world evidence from standardized multi-site data remains limited. We evaluated the association between endometriosis and newly recorded autoimmune thyroid disease using federated Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) data from 12 Korean hospitals. **Methods**: We conducted a propensity score (PS)-matched cohort study using OMOP-CDM version 5.3 data from 12 Korean tertiary academic medical centers. Women aged 18–60 years with a first recorded endometriosis diagnosis were matched 1:1 to controls without endometriosis on age, calendar year, hypertension, and selected Charlson comorbidity index components. Site-specific Cox proportional hazards models, fitted to patient-level records locally at each institution, estimated hazard ratios (HRs) for newly recorded autoimmune thyroid disease, defined as Hashimoto's thyroiditis or Graves' disease. Pooled estimates were derived using DerSimonian–Laird random-effects meta-analysis; leave-one-out meta-analysis, meta-regression on site-specific follow-up ratio, and an E-value were used to assess robustness. **Results**: After 1:1 PS matching, 71,619 women with endometriosis and 71,619 matched controls were included. Mean follow-up was 2693 days in the endometriosis group and 1677 days in the control group. A directionally positive but statistically inconclusive associ-

OHDSI Shoutouts!

Congratulations to the team of **Anna Ostropolets, Vlad Korsik, Tatsiana Skuhareuskaya, Aleh Zhuk, Maryia Khitrun, Alexander Davydov, Dmitry Dymshyts, Christian Reich, George Hripcsak, and Patrick Ryan** on the recent publication of **Real-World Use of Controlled Terminologies, Ontologies, and Vocabularies for Evidence Generation Across a Large International Observational Network: Challenges and Lessons Learned From a Mixed Method Study** in *JMIR Medical Informatics.*

JMIR MEDICAL INFORMATICS Ostropolets et al

Original Paper

Real-World Use of Controlled Terminologies, Ontologies, and Vocabularies for Evidence Generation Across a Large International Observational Network: Challenges and Lessons Learned From a Mixed Method Study

Anna Ostropolets[1,2], MD, PhD; Vlad Korsik[3], MD; Tatsiana Skuhareuskaya[3], MD; Aleh Zhuk[3], MD; Maryia Khitrun[3], MD; Alexander Davydov[3], MD; Dmitry Dymshyts[1], MD; Christian Reich[4*], MD, PhD; George Hripcsak[2*], MD, MS; Patrick Ryan[2,5*], PhD

[1]Observational Health Data Analytics, Johnson & Johnson, Raritan, NJ, United States
[2]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, United States
[3]Odysseus, an EPAM Company, Cambridge, MA, United States
[4]Nemesis Health, New York, NY, United States
[5]Observational Health Data Analytics, Johnson&Johnson (United States), Titusville, NJ, United States
*these authors contributed equally

Corresponding Author:

Anna Ostropolets, MD, PhD
Observational Health Data Analytics
Johnson & Johnson
Raritan, NJ 08869
United States
Email: ao2671@cumc.columbia.edu

Abstract

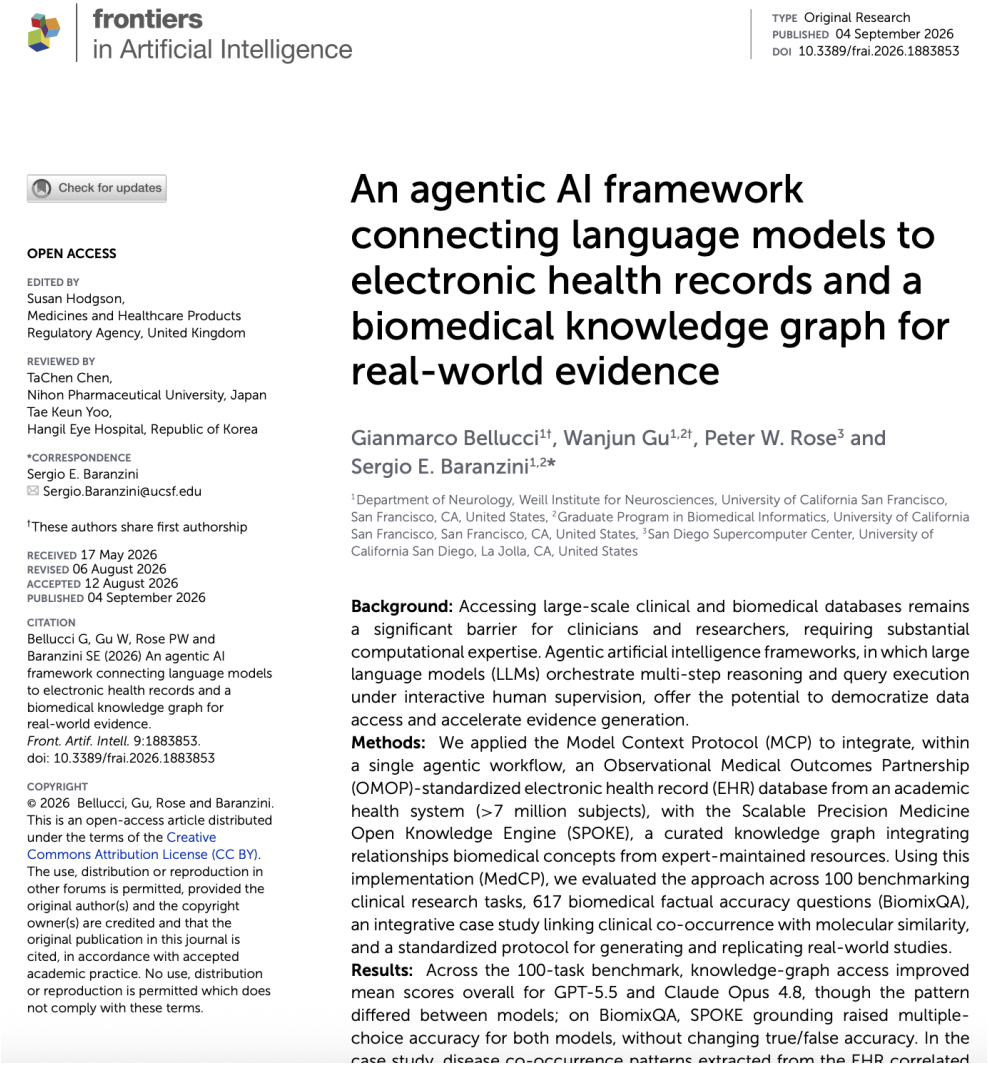
Background: Large-scale international real-world evidence generation benefits from terminology harmonization. Despite widespread adoption of standardized vocabularies, their effective use and long-term sustainability at scale remain poorly understood.

Objective: This study aimed to examine real-world terminology and code use in data across a federated network of observational data sources.

Methods: We conducted a 2-part survey of researchers and data owners within the Observational Health Data Sciences and Informatics community on their terminology use, challenges, and needs, accompanied by the analysis of code use across a subset of real-world data sources.

OHDSI Shoutouts!

Congratulations to the team of **Gianmarco Bellucci, Wanjun Gu, Peter Rose, and Sergio Baranzini** on the recent publication of **An agentic AI framework connecting language models to electronic health records and a biomedical knowledge graph for real-world evidence** in *Frontiers of Artificial Intelligence*.



frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 04 September 2026
DOI 10.3389/frai.2026.1883853

Check for updates

OPEN ACCESS

EDITED BY
Susan Hodgson,
Medicines and Healthcare Products Regulatory Agency, United Kingdom

REVIEWED BY
TaChen Chen,
Nihon Pharmaceutical University, Japan
Tae Keun Yoo,
Hangil Eye Hospital, Republic of Korea

*CORRESPONDENCE
Sergio E. Baranzini
Sergio.Baranzini@ucsf.edu

†These authors share first authorship

RECEIVED 17 May 2026
REVISED 06 August 2026
ACCEPTED 12 August 2026
PUBLISHED 04 September 2026

CITATION
Bellucci G, Gu W, Rose PW and Baranzini SE (2026) An agentic AI framework connecting language models to electronic health records and a biomedical knowledge graph for real-world evidence.
Front. Artif. Intell. 9:1883853.
doi: 10.3389/frai.2026.1883853

An agentic AI framework connecting language models to electronic health records and a biomedical knowledge graph for real-world evidence

Gianmarco Bellucci[1†], Wanjun Gu[1,2†], Peter W. Rose[3] and Sergio E. Baranzini[1,2*]

[1]Department of Neurology, Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, United States, [2]Graduate Program in Biomedical Informatics, University of California San Francisco, San Francisco, CA, United States, [3]San Diego Supercomputer Center, University of California San Diego, La Jolla, CA, United States

Background: Accessing large-scale clinical and biomedical databases remains a significant barrier for clinicians and researchers, requiring substantial computational expertise. Agentic artificial intelligence frameworks, in which large language models (LLMs) orchestrate multi-step reasoning and query execution under interactive human supervision, offer the potential to democratize data access and accelerate evidence generation.
Methods: We applied the Model Context Protocol (MCP) to integrate, within a single agentic workflow, an Observational Medical Outcomes Partnership (OMOP)-standardized electronic health record (EHR) database from an academic health system (>7 million subjects), with the Scalable Precision Medicine Open Knowledge Engine (SPOKE), a curated knowledge graph integrating relationships biomedical concepts from expert-maintained resources. Using this implementation (MedCP), we evaluated the approach across 100 benchmarking clinical research tasks, 617 biomedical factual accuracy questions (BiomixQA), an integrative case study linking clinical co-occurrence with molecular similarity, and a standardized protocol for generating and replicating real-world studies.
Results: Across the 100-task benchmark, knowledge-graph access improved mean scores overall for GPT-5.5 and Claude Opus 4.8, though the pattern differed between models; on BiomixQA, SPOKE grounding raised multiple-choice accuracy for both models, without changing true/false accuracy. In the case study, disease co-occurrence patterns extracted from the EHR correlated

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Vocabulary Subgroup
Thursday	7:30 am	India Community Call
Thursday	9 am	Phenotype Development and Evaluation
Thursday	10 am	GIS – Geographic Information System
Thursday	10 am	Africa Chapter
Thursday	11 am	Perinatal & Reproductive Health
Thursday	12 pm	Medical Devices
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Monday	9 am	Vaccine Vocabulary
Tuesday	9 am	OpenEHR & OMOP
Tuesday	10 am	CDM Survey

Titan Nominations Close Today!

The OHDSI Titan Awards recognize those who have gone above and beyond to foster community engagement, lead research and development efforts, and make significant contributions towards OHDSI's mission.

Individuals, teams or collaborating institutions can be nominated for an award, and you can make as many nominations as you would like.

Nominations will be accepted until 8pm ET TONIGHT.

ohdsi.org/titan-awards

4th Annual EU Symposium

2026 OHDSI Global Symposium

Registration is OPEN for the **2026 OHDSI Global Symposium**, which will be held Oct. 20-22 in New Brunswick, N.J., USA.

Oct. 20: Tutorials
Oct. 21: Plenaries, Showcase
Oct. 22: Workgroup Activities

ohdsi.org/OHDSI2026

Registration Opens For Canada Symposium

Registration is now open for the 2026 OHDSI Canada Symposium, held Nov. 17-18 in Toronto, Ont.

Nov. 17: Workshops

Track A: From Source Data to OMOP

Track B: From OMOP to Evidence

Nov. 18: Symposium

The main symposium day will bring the Canadian OHDSI community together to share experiences, showcase emerging work, and explore opportunities for collaboration.

APAC Symposium: Nov. 13-15

The 2026 APAC Symposium will be held Nov. 13-15 at Yonsei University in Seoul, South Korea.

Nov. 13: Main Conference
Nov. 14: Tutorials, Datathon
Nov. 15: Datathon Team Activities

Registration is open! All details can be found on the event homepage.

ohdsi.org/APAC2026

In the era of foundation models, does data standardization still matter?

> *LLMs read raw clinical text directly — no ETL, no mapping.*
> *So is the OMOP CDM a trust mechanism for AI evidence, or a legacy cost we keep paying out of habit?*

WHY NOW · THE GAP

- ✓ Clinicians were asked — AMA, 2026
- ✓ Global experts in medical informatics, AI, bioethics, and etc were asked — npj Digital Medicine, 2026

It's not a clean split — the same person can hold both views, shifting by domain, data type and task.

IN SOME DOMAINS — A LEGACY COST

If natural language becomes the universal interface, do rigid vocabularies still earn their keep?

IN OTHERS — THE GUARDRAIL YOU CAN'T DROP

Ontologies keep LLM agents accurate, auditable and interoperable. Lexical tools alone break on messy terms.

TAKE THE SURVEY

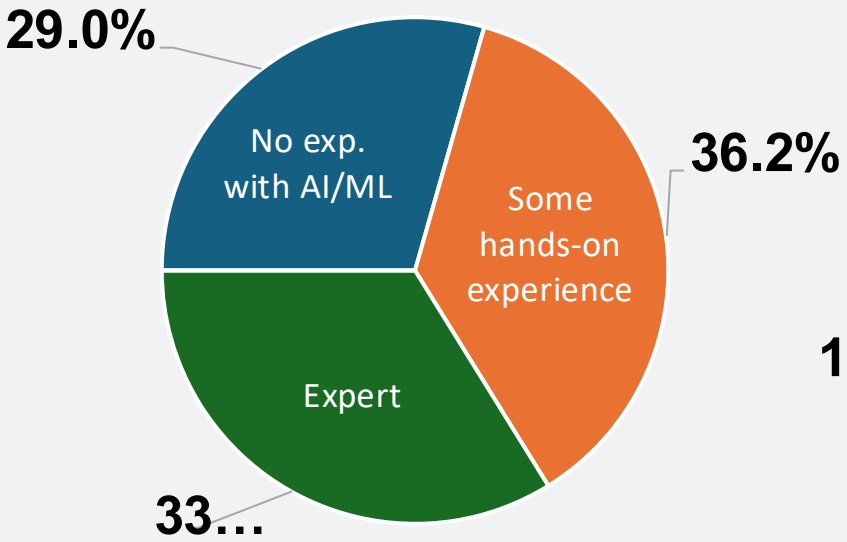
≈ 7 minutes — do it now, it counts!

Final week to participate (closing Oct 4!)

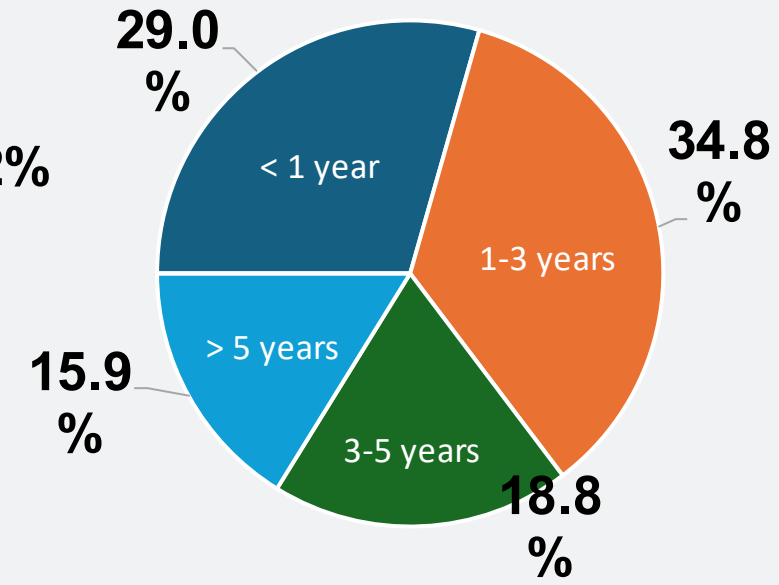
68 responses so far – who has answered?

Respondents include Working Group contributors (CDM, Vocabulary, HADES, THEMIS, Generative AI) across academic, industry (~26%), hospital, and regulatory settings.

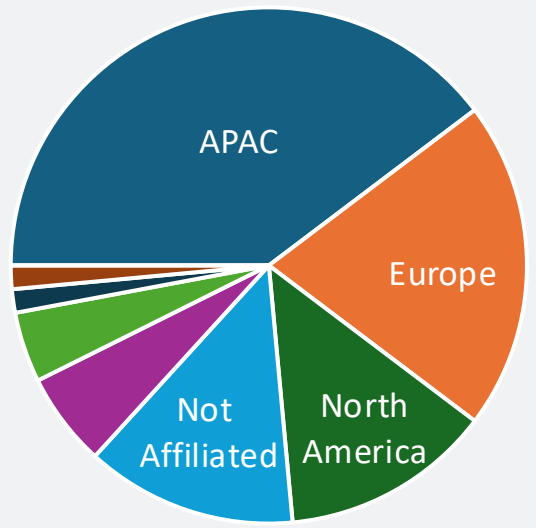
AI/ML experience



Years in OHDSI



OHDSI Region



- APAC 39.1%
- Europe 20.3%
- North America 13.0%
- Not affiliated 13.0%

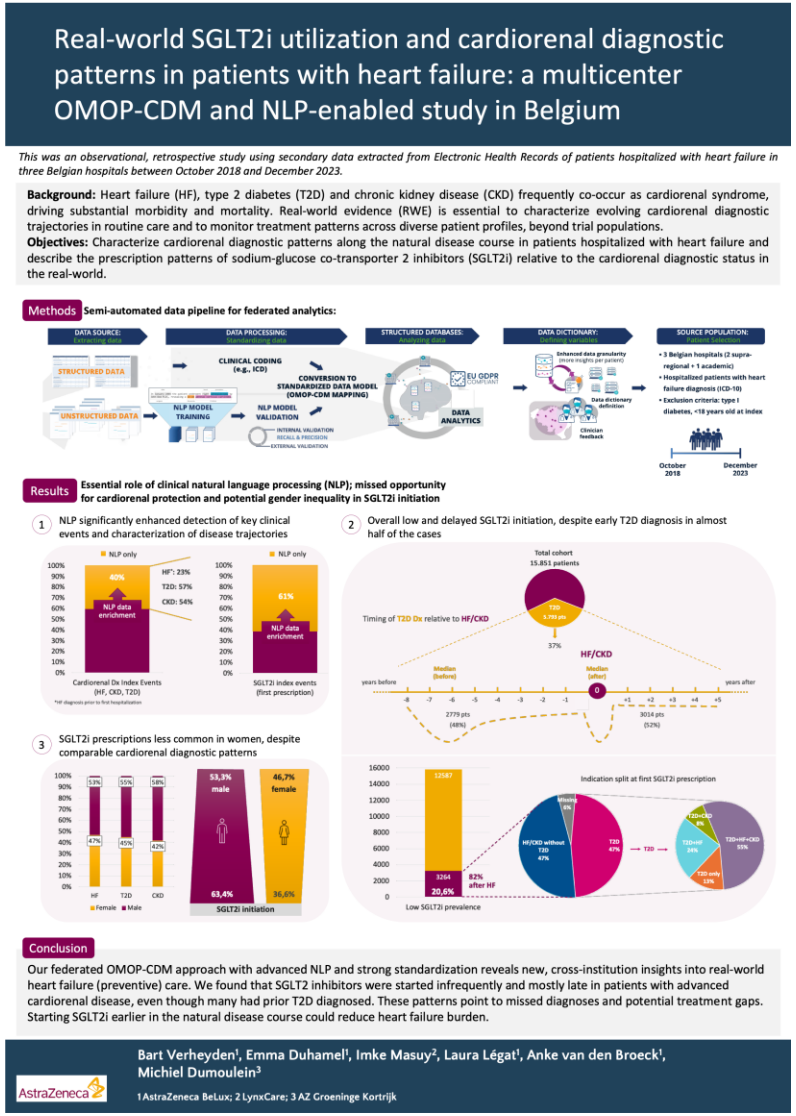
Results will be presented @ 2026 APAC Symposium discussion session
Any questions?, contact: jiwonum@ohdsi.org / chandryou@yuhs.ac

#OHDSISocialShowcase This Week

Monday

Real-World Evidence on SGLT2 inhibitor utilization across cardiorenal phenotypes in Belgium: a federated OMOP-CDM and NLP-enabled hospital network study

(Emma Duhamel, Bart Verheyden, Imke Masuy, Laura Légat, Anke Van den Broeck, Michiel Dumoulein)



Real-world SGLT2i utilization and cardiorenal diagnostic patterns in patients with heart failure: a multicenter OMOP-CDM and NLP-enabled study in Belgium

This was an observational, retrospective study using secondary data extracted from Electronic Health Records of patients hospitalized with heart failure in three Belgian hospitals between October 2018 and December 2023.

Background: Heart failure (HF), type 2 diabetes (T2D) and chronic kidney disease (CKD) frequently co-occur as cardiorenal syndrome, driving substantial morbidity and mortality. Real-world evidence (RWE) is essential to characterize evolving cardiorenal diagnostic trajectories in routine care and to monitor treatment patterns across diverse patient profiles, beyond trial populations.
Objectives: Characterize cardiorenal diagnostic patterns along the natural disease course in patients hospitalized with heart failure and describe the prescription patterns of sodium-glucose co-transporter 2 inhibitors (SGLT2i) relative to the cardiorenal diagnostic status in the real-world.

Methods **Semi-automated data pipeline for federated analytics:**

Results **Essential role of clinical natural language processing (NLP); missed opportunity for cardiorenal protection and potential gender inequality in SGLT2i initiation**

1 NLP significantly enhanced detection of key clinical events and characterization of disease trajectories

2 Overall low and delayed SGLT2i initiation, despite early T2D diagnosis in almost half of the cases

3 SGLT2i prescriptions less common in women, despite comparable cardiorenal diagnostic patterns

Conclusion

Our federated OMOP-CDM approach with advanced NLP and strong standardization reveals new, cross-institution insights into real-world heart failure (preventive) care. We found that SGLT2 inhibitors were started infrequently and mostly late in patients with advanced cardiorenal disease, even though many had prior T2D diagnosed. These patterns point to missed diagnoses and potential treatment gaps. Starting SGLT2i earlier in the natural disease course could reduce heart failure burden.

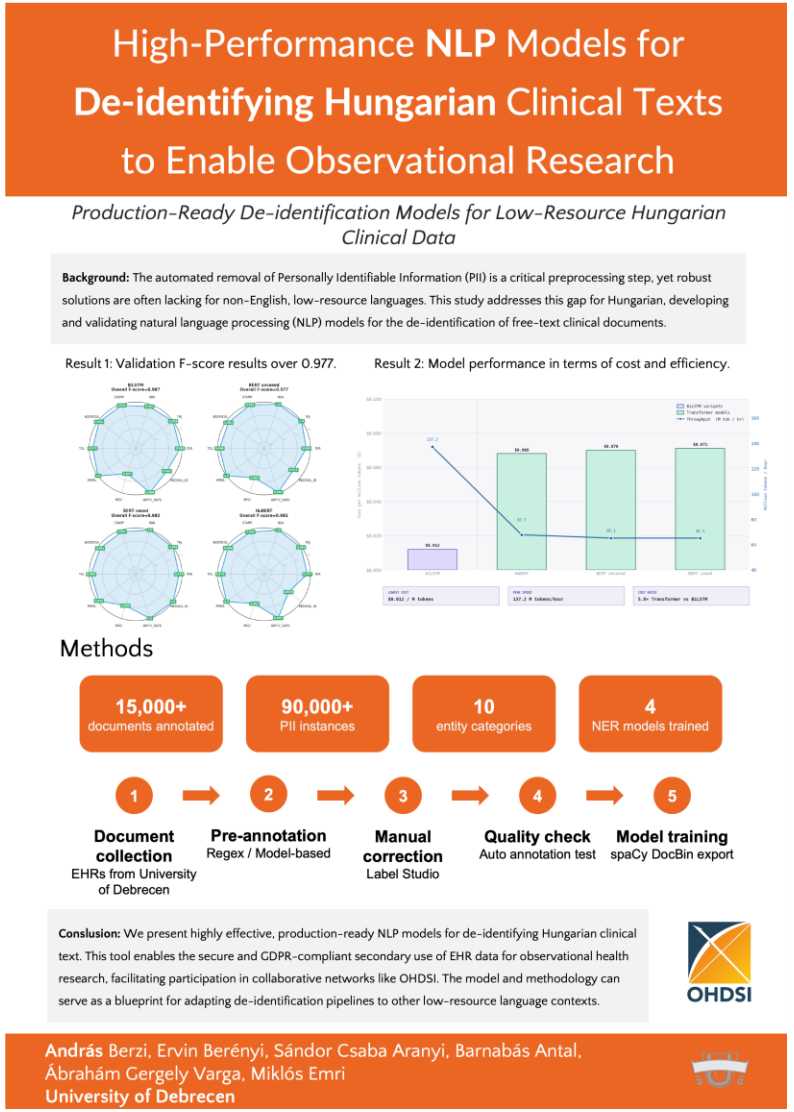
Bart Verheyden[1], Emma Duhamel[1], Imke Masuy[2], Laura Légat[1], Anke van den Broeck[1], Michiel Dumoulein[3]
1 AstraZeneca BeLux; 2 LynxCare; 3 AZ Groeninge Kortrijk

#OHDSISocialShowcase This Week

Tuesday

High-Performance NLP Models for De-identifying Hungarian Clinical Texts to Enable Observational Research

(András Berzi, Ervin Berényi, Zita Képes, Sándor Csaba Aranyi, Barnabás Antal, Ábrahám Gergely Varga, Miklós Emri)

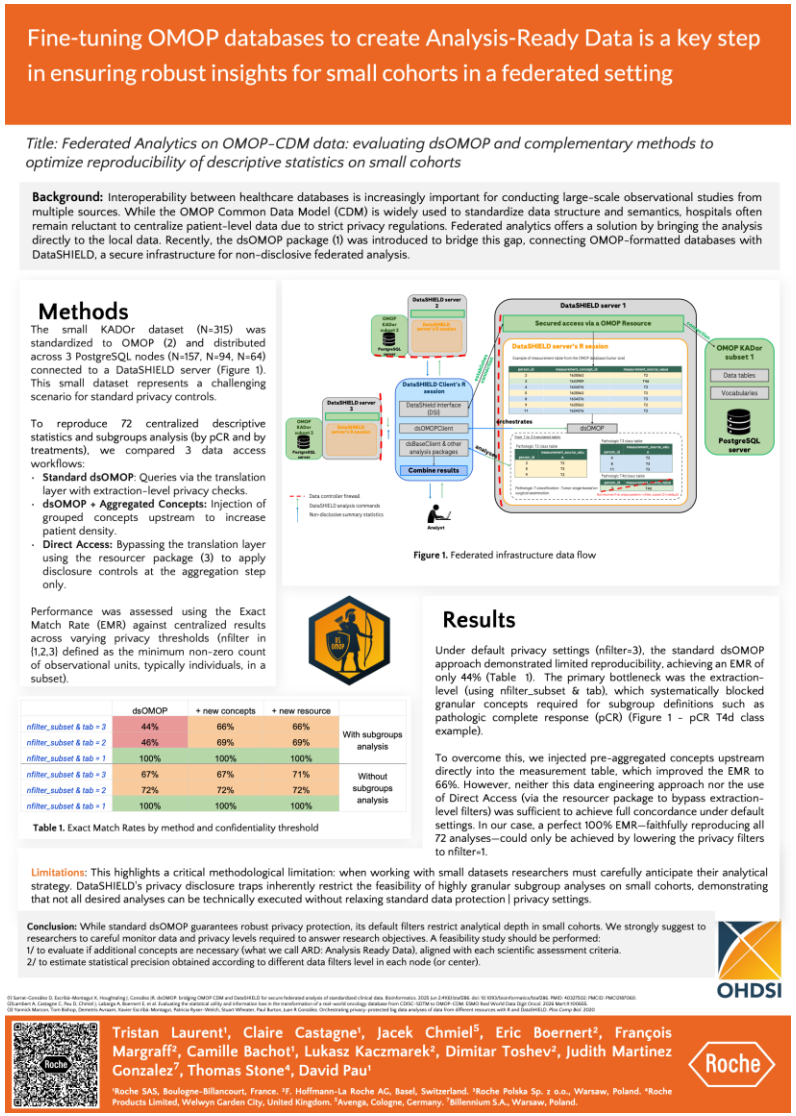


#OHDSISocialShowcase This Week

Wednesday

Fine-tuning OMOP databases to create Analysis-Ready Data is a key step in ensuring robust insights for small cohorts in a federated setting

(Tristan Laurent, Jacek Chmiel, Camille Bachot, Dimitar Toshev, François Margraff, Lukasz Kaczmarek, Eric Boernert, Judith Martinez Gonzalez, Thomas Stone, David Pau, Claire Castagne)



Fine-tuning OMOP databases to create Analysis-Ready Data is a key step in ensuring robust insights for small cohorts in a federated setting

Title: Federated Analytics on OMOP-CDM data: evaluating dsOMOP and complementary methods to optimize reproducibility of descriptive statistics on small cohorts

Background: Interoperability between healthcare databases is increasingly important for conducting large-scale observational studies from multiple sources. While the OMOP Common Data Model (CDM) is widely used to standardize data structure and semantics, hospitals often remain reluctant to centralize patient-level data due to strict privacy regulations. Federated analytics offers a solution by bringing the analysis directly to the local data. Recently, the dsOMOP package (1) was introduced to bridge this gap, connecting OMOP-formatted databases with DataSHIELD, a secure infrastructure for non-disclosive federated analysis.

Methods

The small KADOr dataset (N=315) was standardized to OMOP (2) and distributed across 3 PostgreSQL nodes (N=157, N=94, N=64) connected to a DataSHIELD server (Figure 1). This small dataset represents a challenging scenario for standard privacy controls.

To reproduce 72 centralized descriptive statistics and subgroups analysis (by pCR and by treatments), we compared 3 data access workflows:

- **Standard dsOMOP**: Queries via the translation layer with extraction-level privacy checks.
- **dsOMOP + Aggregated Concepts:** Injection of grouped concepts upstream to increase patient density.
- **Direct Access:** Bypassing the translation layer using the resourcer package (3) to apply disclosure controls at the aggregation step only.

Performance was assessed using the Exact Match Rate (EMR) against centralized results across varying privacy thresholds (nfilter in {1,2,3} defined as the minimum non-zero count of observational units, typically individuals, in a subset).

Figure 1. Federated infrastructure data flow

Results

Under default privacy settings (nfilter=3), the standard dsOMOP approach demonstrated limited reproducibility, achieving an EMR of only 44% (Table 1). The primary bottleneck was the extraction-level (using nfilter_subset & tab), which systematically blocked granular concepts required for subgroup definitions such as pathologic complete response (pCR) (Figure 1 – pCR T4d class example).

To overcome this, we injected pre-aggregated concepts upstream directly into the measurement table, which improved the EMR to 66%. However, neither this data engineering approach nor the use of Direct Access (via the resourcer package to bypass extraction-level filters) was sufficient to achieve full concordance under default settings. In our case, a perfect 100% EMR—faithfully reproducing all 72 analyses—could only be achieved by lowering the privacy filters to nfilter=1.

	dsOMOP	+ new concepts	+ new resource	
nfilter_subset & tab = 3	44%	66%	66%	With subgroups analysis
nfilter_subset & tab = 2	46%	69%	69%	
nfilter_subset & tab = 1	100%	100%	100%	
nfilter_subset & tab = 3	67%	67%	71%	Without subgroups analysis
nfilter_subset & tab = 2	72%	72%	72%	
nfilter_subset & tab = 1	100%	100%	100%	

Table 1. Exact Match Rates by method and confidentiality threshold

Limitations: This highlights a critical methodological limitation: when working with small datasets researchers must carefully anticipate their analytical strategy. DataSHIELD's privacy disclosure traps inherently restrict the feasibility of highly granular subgroup analyses on small cohorts, demonstrating that not all desired analyses can be technically executed without relaxing standard data protection | privacy settings.

Conclusion: While standard dsOMOP guarantees robust privacy protection, its default filters restrict analytical depth in small cohorts. We strongly suggest to researchers to careful monitor data and privacy levels required to answer research objectives. A feasibility study should be performed:
1/ to evaluate if additional concepts are necessary (what we call ARD: Analysis Ready Data), aligned with each scientific assessment criteria.
2/ to estimate statistical precision obtained according to different data filters level in each node (or center).

OHDSI

(1) Sarrat-Considine D, Escriba-Montagut X, Houghtaling J, Considine JR. dsOMOP: bridging OMOP CDM and DataSHIELD for secure federated analysis of standardized clinical data. Bioinformatics. 2025 Jun 2:41(6):btaf286. doi: 10.1093/bioinformatics/btaf286. PMID: 40337502; PMCID: PMC12187060.
(2) Laurent G, Castagne C, Pau D, Chmiel J, Lebeaja A, Boernert E, et al. Evaluating the statistical utility and information loss in the transformation of a real-world oncology database from CDISC SDTM to OMOP-CDM. ESMO Real World Data Digit Oncol. 2026 Mar;9:100666.
(3) Yannick Marcon, Tom Bishop, Demetris Avraam, Xavier Escribà Montagut, Patricia Ryser-Welch, Stuart Wheater, Paul Burton, Juan R González. Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD. Plos Comp Biol. 2020

Tristan Laurent[1], Claire Castagne[1], Jacek Chmiel[5], Eric Boernert[2], François Margraff[2], Camille Bachot[1], Lukasz Kaczmarek[2], Dimitar Toshev[2], Judith Martinez Gonzalez[7], Thomas Stone[4], David Pau[1]

[1]Roche SAS, Boulogne-Billancourt, France. [2]F. Hoffmann-La Roche AG, Basel, Switzerland. [3]Roche Polska Sp. z o.o., Warsaw, Poland. [4]Roche Products Limited, Welwyn Garden City, United Kingdom. [5]Avenga, Cologne, Germany. [7]Billennium S.A., Warsaw, Poland.

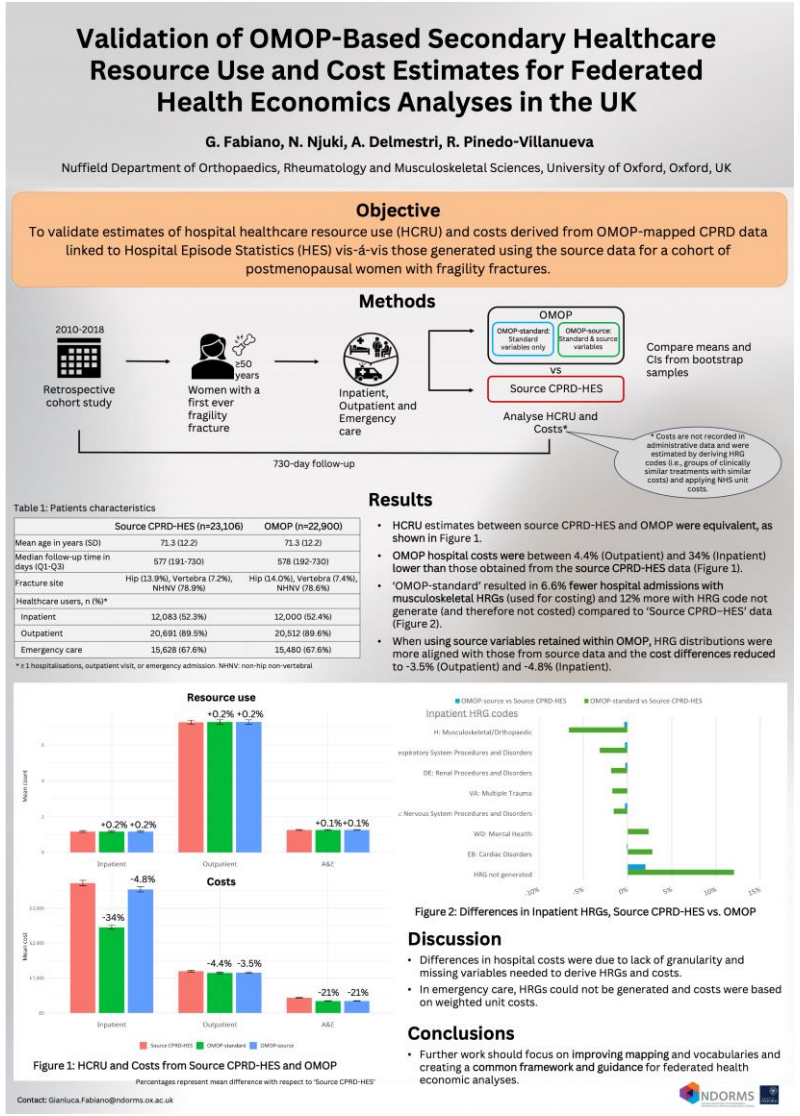
Roche

#OHDSISocialShowcase This Week

Thursday

Validation of OMOP-Based Secondary Healthcare Resource Use and Cost Estimates for Federated Health Economics Analyses in the UK

(Gianluca Fabiano, Njoki Njuki, Antonella Delmestri, Rafael Pinedo-Villanueva)

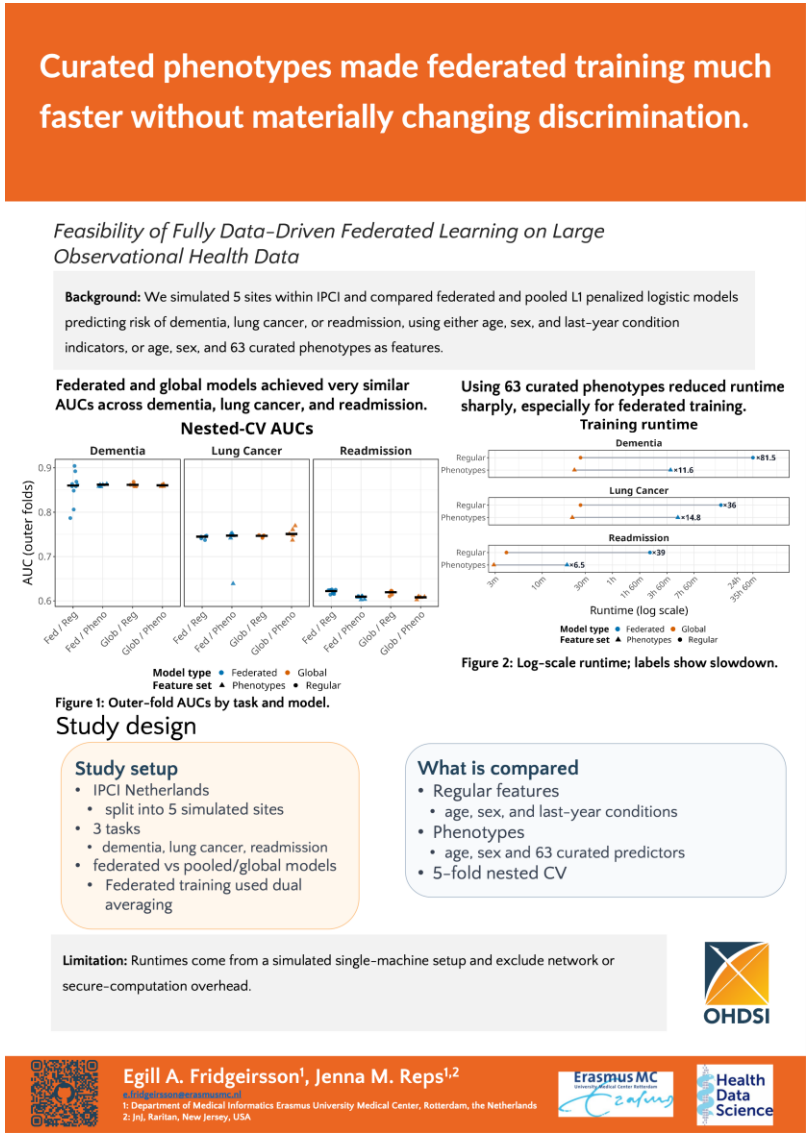


#OHDSISocialShowcase This Week

Friday

Feasibility of Fully Data-Driven Federated Learning on Large Observational Health Data

(Egill Fridgeirsson, Jenna Reps)



Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

2026 OHDSI Global Symposium

Registration is OPEN for the **2026 OHDSI Global Symposium**, which will be held Oct. 20-22 in New Brunswick, N.J., USA.

Oct. 20: Tutorials
Oct. 21: Plenaries, Showcase
Oct. 22: Workgroup Activities

ohdsi.org/OHDSI2026

TUTORIAL

An Introduction to the Journey from Data to Evidence Using OHDSI

Tue Oct. 20
8:00 - 12:00

Meet the people behind this journey

Five perspectives. One OHDSI journey: from data to evidence.

Erica Voss
Johnson & Johnson

The transformer

Data standards & transformation

Ross Williams
Erasmus MC

The analyst

Network research & prediction

Katy Sadowski
Boehringer Ingelheim

The builder

Data quality & open-source tools

Mitchell Conover
Johnson & Johnson

The evidence maker

Clinical applications & causal inference

Ilse Vermeulen
Erasmus MC / OHDSI BE

The connector

Community & implementation

OHDSI
Observational Health Data Sciences and Informatics

#JoinTheJourney

An Intoduction To ATLAS

Preliminary Agenda

8:00am – 8:15am	Introduction • Who should be using Atlas	Chris
8:15am – 9:00am	The CDM & Data Sources • The CDM Data model • Database Characterization in Atlas	Chris
9:00am – 9:30am	Vocabulary, Concepts and Concept Sets • Vocabulary Overview (Domains, Codes) • Concepts: Standard vs. Non-Standard • Searching and Filtering • Record Counts (Record and Descendants) • Building Concept Sets	Jack
9:30am – 11am	Cohort Definitions • What is a Phenotype? • Cohort Definitions: Logic to build a Phenotype • Generation and Reports	Peter
11:00am – 11:45am	Analytics • Characterization • Incidence Rates • Pathways	Jack
11:45am – 12:00pm	Question & Answer	All

Bringing FAIR to Imaging Research with the Medical Imaging OMOP Extension

Participants will learn how to:

- Understand the design principles of the Medical Imaging OMOP extension (MI-CDM)
- Index and characterize DICOM archives from clinical or research repositories
- Standardize DICOM metadata using OMOP-compatible terminology
- Load imaging metadata into searchable OMOP tables
- Represent imaging-derived features while preserving provenance
- Perform multimodal analyses by combining imaging, clinical, and demographic data using standard OHDSI tools such as ATLAS
- Apply FAIR principles to imaging data within the OMOP ecosystem

Who should attend?

- OHDSI community members interested in medical imaging
- Researchers conducting observational or multimodal studies
- Data engineers and informaticians responsible for imaging ETL pipelines
- Institutions planning to integrate imaging into their OMOP CDM

By the end of the tutorial, attendees will be able to:

- Build an imaging-enhanced OMOP dataset from DICOM archives
- Integrate imaging features with clinical data for downstream analyses
- Apply reproducible, FAIR-aligned workflows for imaging research within OHDSI

OHDSI Change Leaders & Leadership

TUTORIAL Tuesday 20th (4h)

Interactive Hands-on

Not recorded • Needs physical participation

Amplifying our impact, through your ability to lead and create change.

Strengthen your ability to **drive change**, **engage stakeholders**, and **grow adoption** of OHDSI and OMOP in your environment.

You'll work on **YOUR priorities** and leave with practical tools, methods, insights, and a **concrete action plan** to can execute.

Who Should Attend

- National Node Leaders (Both current & Newly Becoming)
- Aspiring Change Leaders (Researchers, Data Scientists & Engineers)
- Anyone Supporting OHDSI Growth & Adoption (Industry, Innovators…etc)

Your Faculty

- Liesbet Peeters - NN Lead Belgium & Professor University of Hasselt
- Christian Hogberg - NN Lead Sweden & Change Leader Expert
- Chris Baldwin - Commercial Director Unison & Communication Expert

Become a stronger advocate, connector, facilitator & change leader.

What You'll Do — 4 Interactive Parts

Part I

Start with the End in Mind

Define your personal goal & pick your priority use case to work on.

Part II

Becoming a Change Leader

Apply change management & leadership best practices.

Part III

Building the Bridge from You to Them

Craft a narrative for your stakeholders that communicates value for them.

Part IV

Bringing It All Together

Refine your action plan & get peer + facilitator feedback.

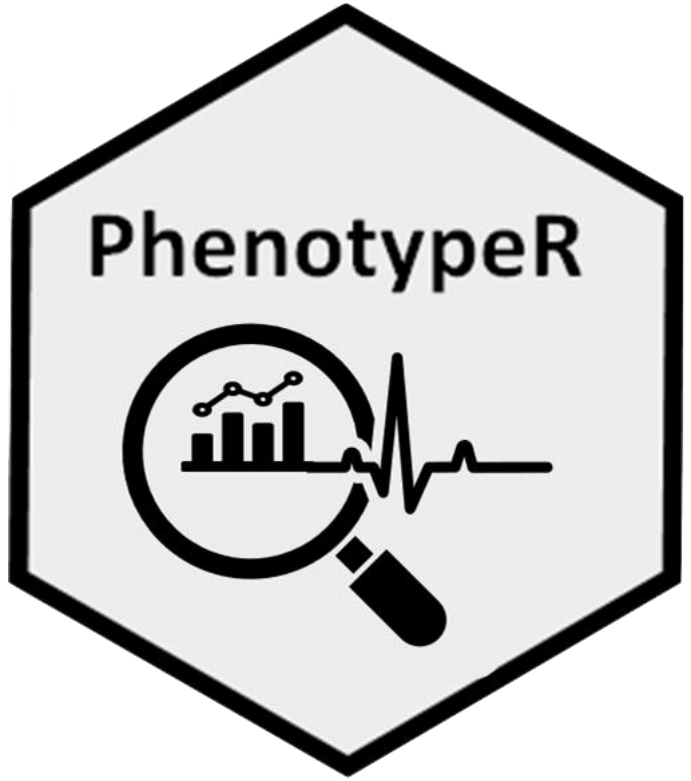
GO! Your stakeholders are ready & waiting — act now.

#JoinTheJourney

Register: www.ohdsi.org/ohdsi2026/

Complex Phenotyping at Scale with and without LLMs Using PhenotypeR

- Understand best practices for designing and evaluating clinical phenotypes
- Learn how to use the PhenotypeR R package to generate phenotype diagnostics
- Discover approaches to generating clinical expectations for diagnostics using human experts, LLMs, or both
- Gain experience in reviewing whether diagnostics match clinical expectations



https://ohdsi.github.io/PhenotypeR

Morning Session (8 am - 12 pm ET)

Mastering OMOP

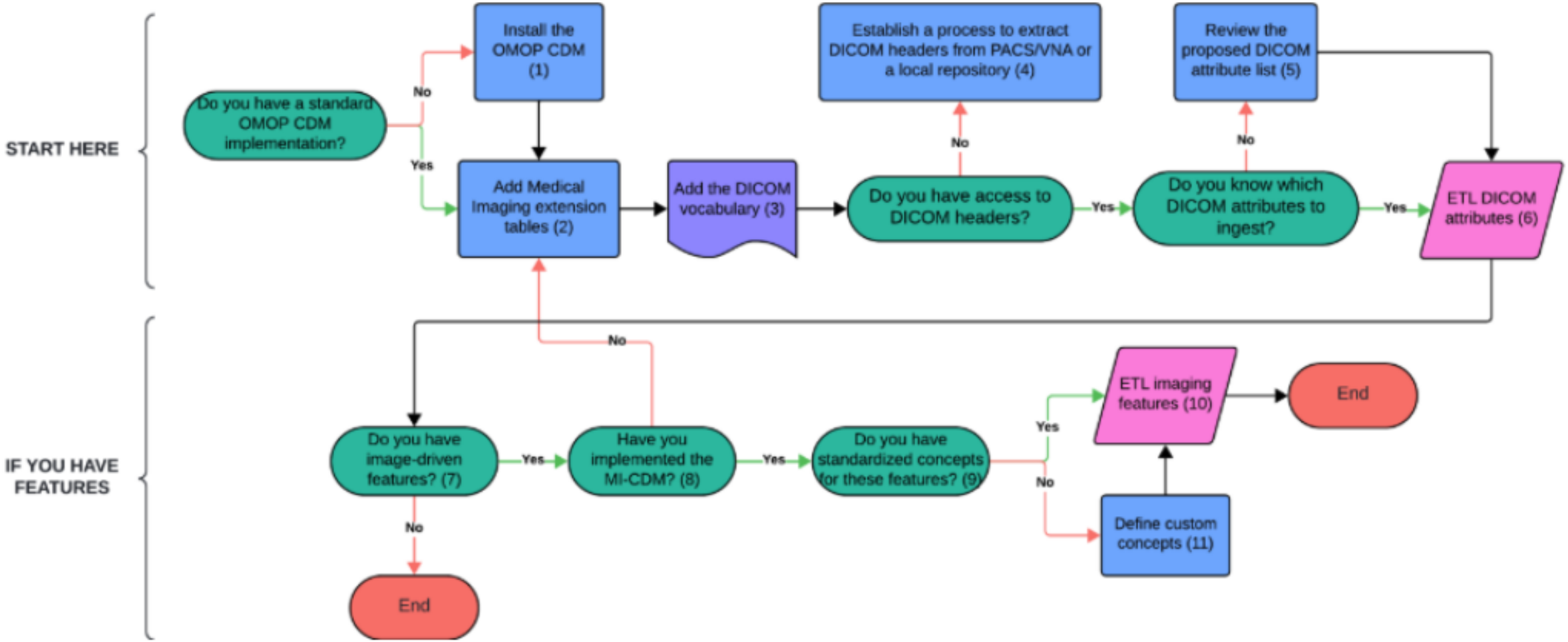
Transforming EHR Data with Practical Strategies, Best Practices, and OHDSI Integration

8am – 12pm on October 20th

- Strategic Institutional Value (ROI & Impact)
- OMOP vs. EHR Fundamentals
- Theory to Practice
- Empowering Modern, Standardized Research

Imaging Evidence Network - getting started

- https://github.com/OHDSI/ImageWG/wiki/



From Multi-Modal Data to Real-World Evidence:
Hands-on with the Data2Evidence Platform for OMOP

Bioinformatics Institute
A*STAR BII

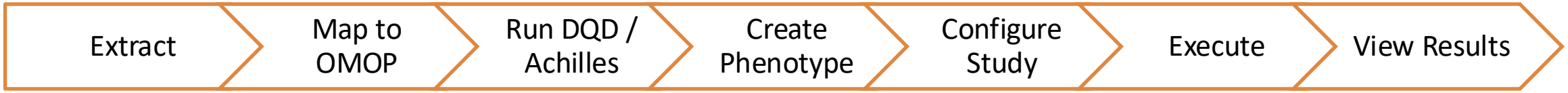
data4life

The dataset

GUSTO - Growing Up in Singapore Towards healthy Outcomes - follows more than 1,100 children from birth since 2009, pairing clinical records, questionnaires and biosamples with longitudinal imaging.

What sets this session apart

Most sessions go deep on one stage of the journey. This one runs its full length: a single dataset travelling from raw files to a shareable evidence artifact in one sitting, with the developers in the room.



Notes: No prior OMOP CDM experience required · laptop and setup instructions provided in advance

Introduction to OHDSI
Phenotype Development & Evaluation

Accurate phenotyping is a cornerstone of reliable observational research, yet it remains one of the greatest methodological challenges in real-world data analytics. Observational data are inherently prone to misclassification and these errors can meaningfully influence study validity. This mid-level OHDSI tutorial is designed for participants who are familiar with OHDSI standardized vocabularies and already comfortable developing cohorts in ATLAS, and who now seek to deepen their scientific understanding and strengthen their applied phenotype development skills. The tutorial will begin with an overview of the science of misclassification error, covering key concepts such as sensitivity, specificity, predictive value, and index event misclassification. Participants will learn how these errors arise within observational data sources and how they directly affect effect estimation, transportability, and downstream decision-making. Building on this foundation, the session will transition to OHDSI's established best practices for phenotype development and evaluation. Using real examples, we will walk through tools and methods that support transparent, reproducible, and scalable phenotype definitions—including cohort diagnostics, standard vocabulary exploration, benchmarking against population characteristics. The latter portion of the tutorial introduces an emerging opportunity for improving phenotype development: an AI-aided iterative workflow. We will demonstrate how AI/LLM can support iterative refinement as part of phenotype development.

By the end of this tutorial, participants will have a richer understanding of the scientific foundations of phenotype misclassification, practical experience applying OHDSI best practices, and early exposure to how AI-enabled workflows can enhance rigor and reproducibility. This session is ideal for researchers ready to advance from simply using ATLAS to mastering phenotype design as a scientific discipline.

2026 OHDSI Global Symposium

Oct 20-22 • New Brunswick, NJ • Hyatt Regency Hotel

Integrating Geospatial Data into OMOP CDM: A Hands-On Tutorial with the OHDSI GIS and GAIA Toolchain

Join the GIS Workgroup Team on October 20, 2026 | 1-5pm EST

Register now at ohdsi.org/ohdsi2026

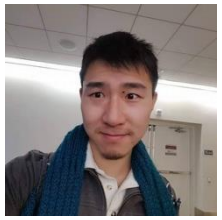
Jared Houghtaling
Johnson & Johnson

Robert Miller
Miller Data Solutions, LLC

Polina Talapova
SciForce

Tim Norris
University of Miami

Jay Greenfield
CODATA



Jim Phuong
University of Washington

Tim Rossi
Architectural Medicine, LLC

Jacob Zelko
Northeastern University

David Amadi
CODATA

Jiawen Liao
Cleveland Clinic Research

Using OMOP Model in Registry Context & Clinical Trials Standardization Context: Conventions, Past Use Cases, SDTM & Regulatory Consideration, Challenges

Registry and clinical trial (CT) data often need to fill the gap not covered by traditional RWD. If RWD follows OMOP, using the same model for harmonization of registry and CT data has advantages.

Outline:

1. Special considerations found in registry and CT data (e.g., converting relative dates to OMOP compliant dates, CRF data, adverse drug event (seriousness, severity, and relatedness between drug and AE))
2. Use of 2B concepts (custom vocabularies) vs OMOP vocabularies
3. Use cases published in literature (e.g., UK Biobank, OHDSI 2024 poster: Application of OMOP Common Data Model to Disease Registry Data, AllOfUs)
4. Other use cases (AACR GENIE)
5. OMOP and SDTM comparison (+ FDA regulatory considerations)
6. Combining trial and RWD data (external control arm) considerations (data granularity mismatch, RWD limited to routine healthcare, lack of advanced COAs, data collection not regulatory grade)

2026 Global Symposium Agenda

Start	End	Topic	Presenter/Lead
7:15 am	8:00 am	Newcomer Welcome	Paul Nagy/Marc Suchard
8:00 am	8:30 am	State of the Community	George Hripcsak
8:30 am	9:15 am	OHDSI Year In Review	Early-Stage Researcher WG
9:15 am	10:00 am	Collaborator Showcase: Posters and Demos (Session 1)	
10:00 am	11:00 am	Plenary 1: Federated Learning Meets Negative Control Calibration: Toward Reliable Multi-Site Evidence Generation	Yong Chen
11:00 am	12:00 pm	Plenary 2: Beyond the Defaults: How the OHDSI Community is Adapting, Extending, and Reimagining Its Tools	Scott Duvall
12:00 pm	1:00 pm	Network & Lunch	
1:00 pm	2:00 pm	Plenary 3: The role of national initiatives in supporting sustainability, collaboration, and growth of OHDSI	Ed Burn
2:00 pm	2:45 pm	Collaborator Showcase: Lightning Talks (Session 1)	5 presenters
2:45 pm	3:30 pm	Collaborator Showcase: Posters and Demos (Session 2)	
3:30 pm	4:15 pm	Collaborator Showcase: Posters and Demos (Session 3)	
4:15 pm	5:00 pm	Collaborator Showcase: Lightning Talks (Session 2)	5 presenters
5:00 pm	6:00 pm	Titan Awards, Closing	Patrick Ryan
6:00 pm	8:30 pm	Dinner on your own	
8:30 pm	11:30 pm	OHDSI Jam Session	Martijn Schuemie

2026 Symposium Workgroup Activities

Session 1 (8 am – 10 am): Eyecare and Vision Research, GIS – Geographic Information Systems, Early-Stage Researchers, Industry, Tidy R Programming with OMOP, HADES Hackathon, Generative AI and Foundation Models, Phenotype Development and Evaluation, Oncology, Health Equity, Vocabularies, APAC

Session 2 (10:30 am – 12:30 pm): Perinatal and Reproductive Health, Waveform, Medical Imaging, Industry, Tidy R Programming with OMOP, HADES Hackathon, Generative AI and Foundation Models, Phenotype Development and Evaluation, Oncology, Health Equity, Vocabularies, Rare Disease

Session 3 (1:30 pm – 3:30 pm): Dentistry, GIS & Waveform Cross-Pollination Meeting, Evidence Network, Women of OHDSI, Psychiatry, HADES Hackathon, Natural Language Processing, Health Economics & Value Assessment, ATLAS/WebAPI, CDM Survey, Surgery & Perioperative Medicine

Session 4 (3:30 pm – 5 pm): Workgroup Summary Session

The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

Links are sent out weekly and available at:

ohdsi.org/community-calls-2026

Find your workgroup.

Fuel our mission.

ohdsi.org/workgroups