Collaborator Showcase Brainstorm

OHDSI Community Call

May 12, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

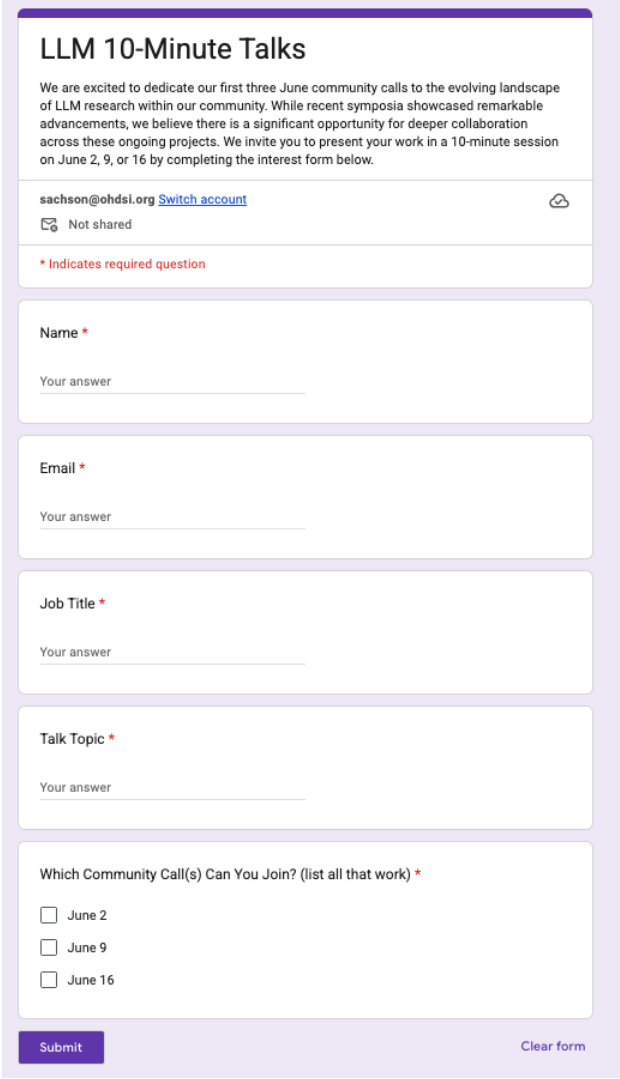
Upcoming Community Calls

Date	Topic
May 12	Collaborator Showcase Brainstorm (Submission Deadline is June 5)
May 19	MEDS (Medical Event Data Standard) & Potential Collaborations with OHDSI
May 26	Workgroup Spotlight: Vocabulary and Evidence Network
June 2	LLM Research Around The World, Session 1
June 9	LLM Research Around The World, Session 2
June 16	LLM Research Around The World, Session 3
June 23	CANCELLED: OHDSI Summer School at Columbia University
June 30	OMOP & OHDSI Research Spotlight

June: LLM Research Presentations

We are excited to dedicate our first three June community calls to the evolving landscape of LLM research within our community. While recent symposia showcased remarkable advancements, we believe there is a significant opportunity for deeper collaboration across these ongoing projects.

We invite you to present your work in a 10-minute session on June 2, 9, or 16 by completing the interest form or emailing sachson@OHDSI.org.



LLM 10-Minute Talks

We are excited to dedicate our first three June community calls to the evolving landscape of LLM research within our community. While recent symposia showcased remarkable advancements, we believe there is a significant opportunity for deeper collaboration across these ongoing projects. We invite you to present your work in a 10-minute session on June 2, 9, or 16 by completing the interest form below.

sachson@ohdsi.org Switch account

Not shared

* Indicates required question

Name *

Your answer

Email *

Your answer

Job Title *

Your answer

Talk Topic *

Your answer

Which Community Call(s) Can You Join? (list all that work) *

- [ ] June 2
- [ ] June 9
- [ ] June 16

Submit

Clear form

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Woo-Young Shin, Ha Young Jang, and Kiyoung Lee** on the recent publication of **Real-world risk assessment of combined cilostazol-rosuvastatin: a retrospective cohort study using Korean electronic health records** in *Frontiers in Pharmacology*.

frontiers in Pharmacology

TYPE Brief Research Report
PUBLISHED 17 April 2026
DOI 10.3389/fphar.2026.1752835

Check for updates

OPEN ACCESS

EDITED BY
Serafeim - Chrysovalantis Kotoulas,
Hippokration General Hospital, Greece

REVIEWED BY
Segundo Mariz,
Retired, Madrid, Spain
Weichieh Lee,
Chi Mei Medical Center, Taiwan

*CORRESPONDENCE
Woo-young Shin,
swy84@cau.ac.kr

RECEIVED 24 November 2025
REVISED 28 February 2026
ACCEPTED 25 March 2026
PUBLISHED 17 April 2026

CITATION
Shin W-y, Jang HY and Lee K (2026) Real-world risk assessment of combined cilostazol-rosuvastatin: a retrospective cohort study using Korean electronic health records.
Front. Pharmacol. 17:1752835.
doi: 10.3389/fphar.2026.1752835

COPYRIGHT

Real-world risk assessment of combined cilostazol-rosuvastatin: a retrospective cohort study using Korean electronic health records

Woo-young Shin[1]*, Ha Young Jang[2] and Kiyoung Lee[3]

[1]Department of Family Medicine, Chung-ang University Gwangmyeong Hospital, Chung-Ang University College of Medicine, Seoul, Republic of Korea, [2]College of Pharmacy, Gachon University, Incheon, Republic of Korea, [3]Department of Internal Medicine, Gachon University Gil Medical Center, Gachon University School of Medicine, Incheon, Republic of Korea

Background: Cilostazol and rosuvastatin are frequently co-prescribed for cardiovascular disease management, particularly in Asian populations. Despite widespread clinical use, comprehensive real-world safety data for this drug combination remain limited.
Objective: To evaluate the 1-year safety profile of cilostazol-rosuvastatin combination therapy compared with that of cilostazol monotherapy in a real-world Korean population using electronic health records.
Methods: This retrospective cohort study utilised the Observational Medical Outcomes Partnership (OMOP) Common Data Model from Gachon University Gil Hospital (2004–2025). Patients on stable rosuvastatin therapy (≥90 days) who had cilostazol added to their treatment (n = 262) were compared to cilostazol monotherapy patients (n = 6,323). Primary outcomes were incidence of atherosclerotic cardiovascular disease (ASCVD), bleeding, and myopathy at 90 and 365 days. Secondary outcomes included changes in liver enzyme levels and platelet count. Fisher's exact and Welch's t-tests were used for statistical analyses.
Results: The combination group had higher baseline comorbidities compared with the monotherapy group, including chronic kidney disease (45.8% vs. 32.1%), diabetes mellitus (55.0% vs. 29.2%), and concurrent antithrombotic therapy (61.1% vs. 27.9%). At the 365-day follow-up, no primary safety events occurred in the combination group (0/113) versus ASCVD 2/2,307 (0.1%), bleeding 3/2,307 (0.1%), and myopathy 1/2,307 (0.0%) in the monotherapy group (p = 1.00). Laboratory

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

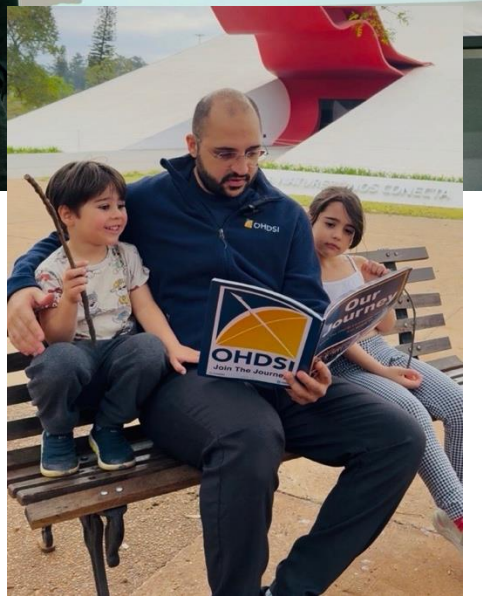
Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Generative AI and Analytics
Wednesday	7 am	Medical Imaging
Wednesday	8 am	Psychiatry
Wednesday	9 am	Patient-Level Prediction
Wednesday	2 pm	Natural Language Processing
Wednesday	7 pm	Eyecare and Vision Research
Thursday	10 am	Rare Diseases
Thursday	10 am	GIS - Geographic Information System
Thursday	10:30 am	Evidence Network
Thursday	7 pm	Dentistry
Friday	9 am	Waveform
Friday	10 am	Transplant
Friday	11:30 am	Steering Group
Monday	11 pm	ETL (formerly Databricks)
Monday	2 pm	Electronic Animal Health Records
Tuesday	9 am	Data2Evidence

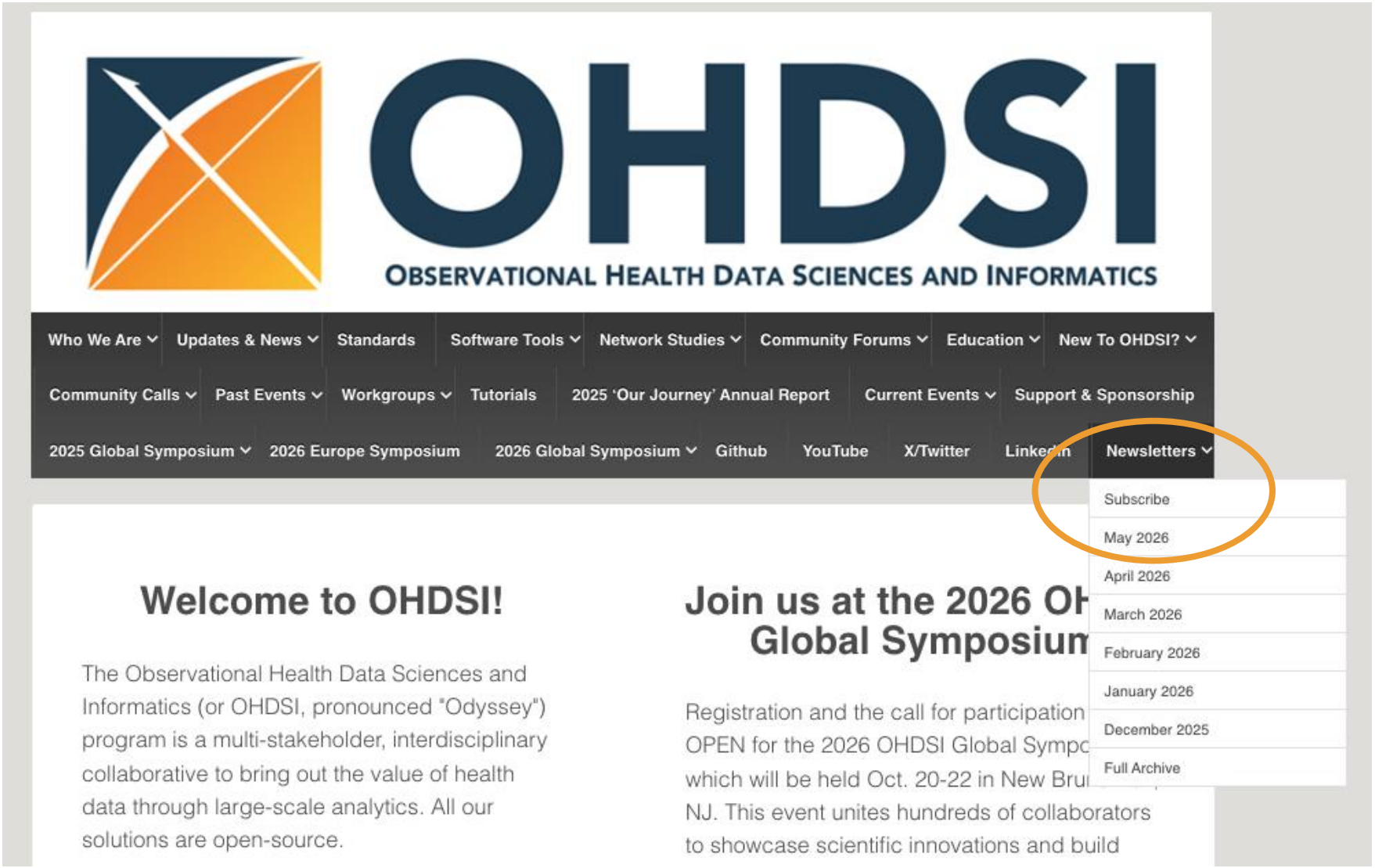
Spotlight: Julio Oliveira

In the latest collaborator spotlight, **Julio Oliveira** reflects on his path to OHDSI, its growth in Latin America, why now is the right time for the first LATAM Symposium, and more.



ohdsi.org/spotlight-julio-oliveira

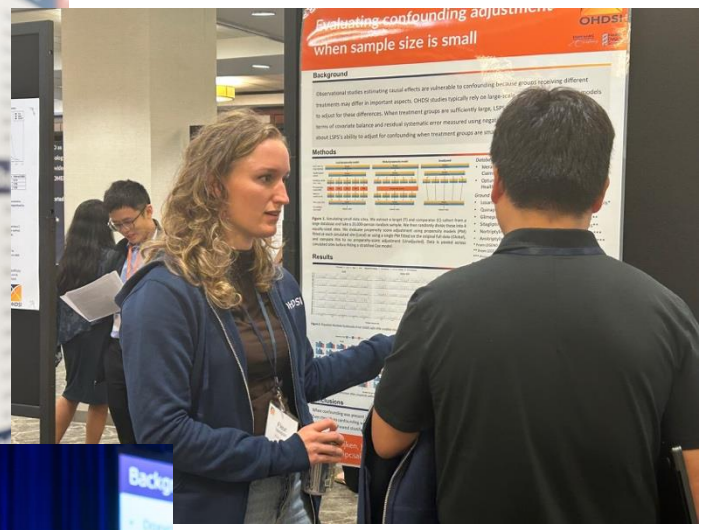
May Newsletter Is Available

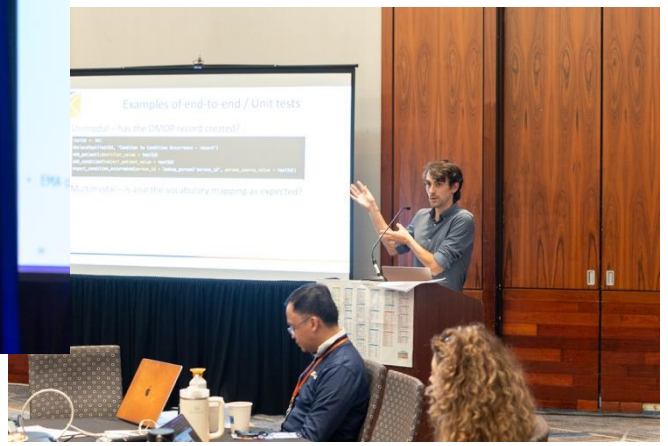


2026 OHDSI Global Symposium

The **call for participation** is open for the 2026 Global Symposium.

The submission deadline is June 5 at 8 pm ET.





24 Days Remaining!

ohdsi.org/OHDSI2026

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

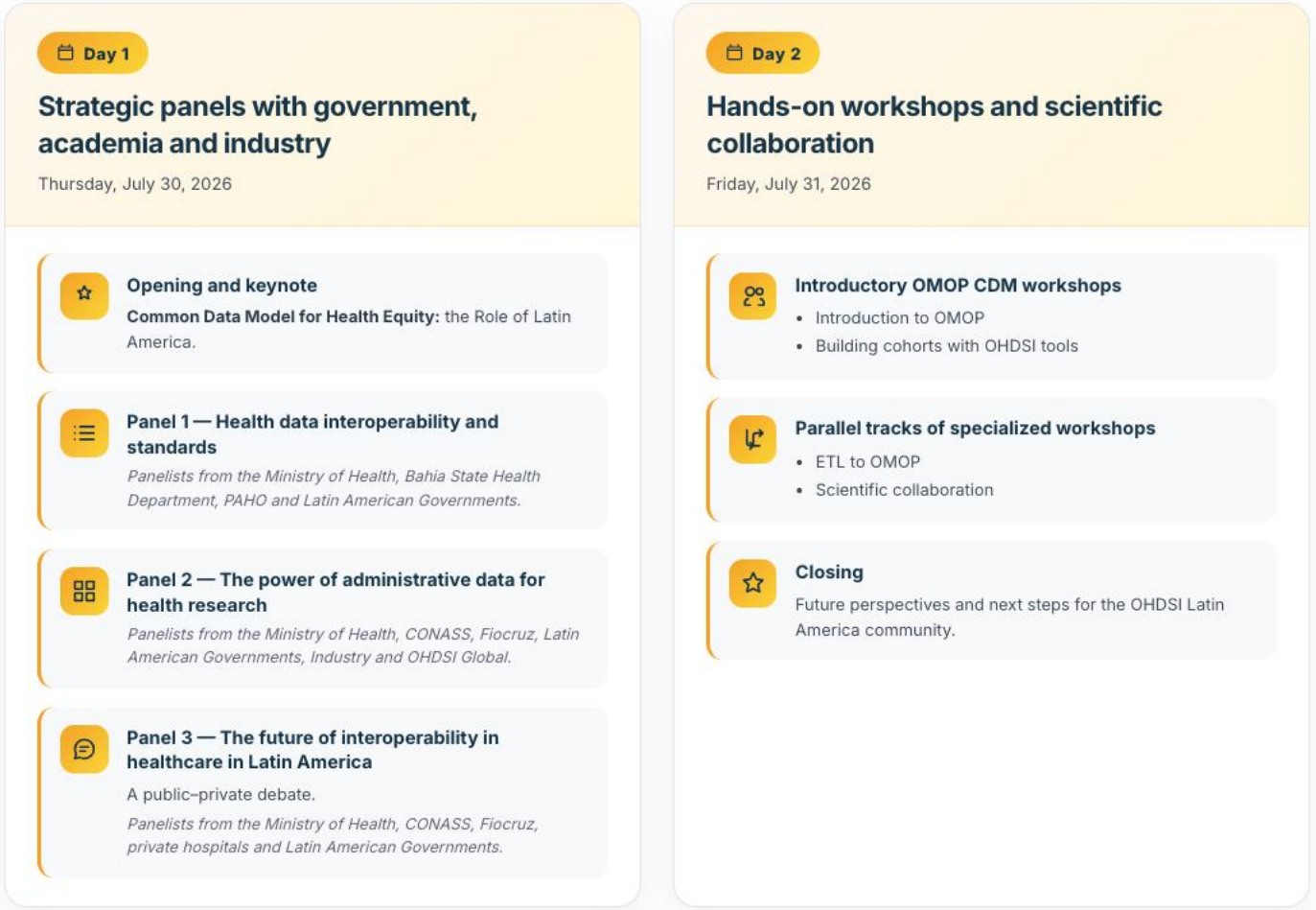
ohdsi.org/OHDSI2026

Maternal Fellowship Deadline: May 30

First Latin America Symposium – July 30-31

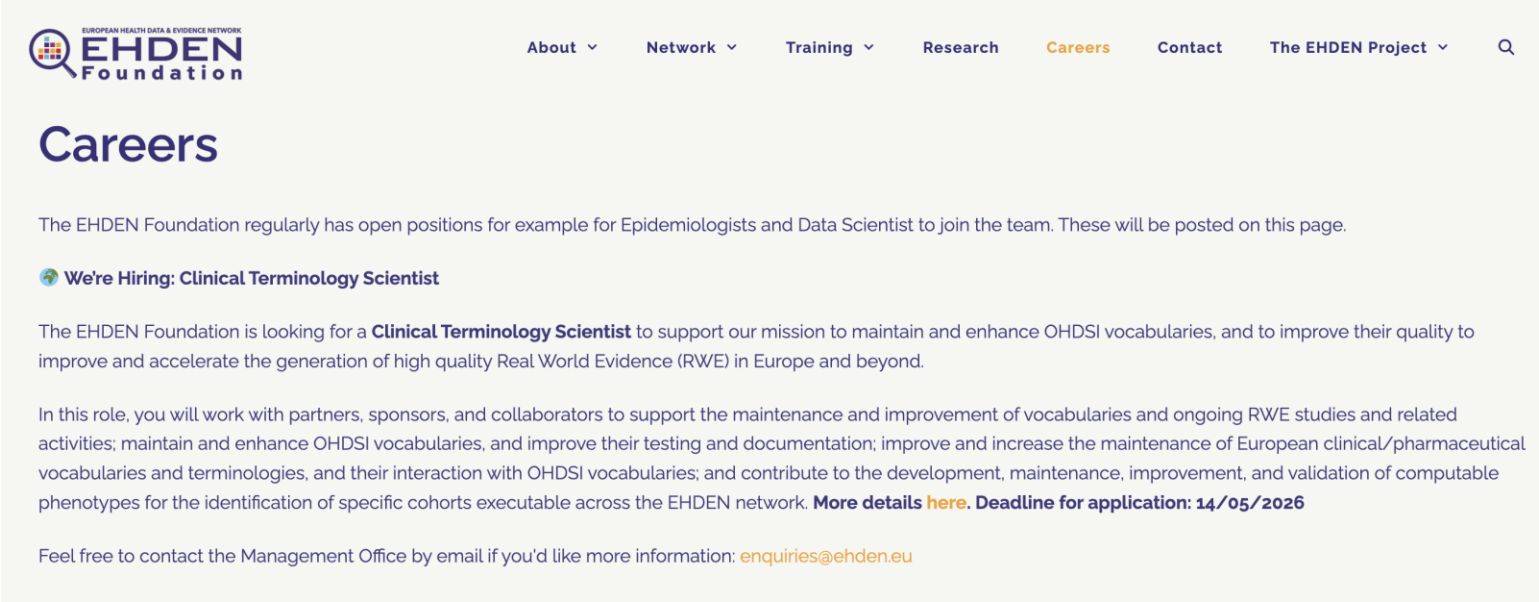
Registration is open for the first OHDSI Latin America Symposium, taking place July 30-31 in Salvador, Brazil.

The submission deadline for the showcase is May 17.



ohdsilatam.org

Opening: Clinical Terminology Scientist



EHDEN Foundation

About · Network · Training · Research · Careers · Contact · The EHDEN Project

Careers

The EHDEN Foundation regularly has open positions for example for Epidemiologists and Data Scientist to join the team. These will be posted on this page.

We're Hiring: Clinical Terminology Scientist

The EHDEN Foundation is looking for a **Clinical Terminology Scientist** to support our mission to maintain and enhance OHDSI vocabularies, and to improve their quality to improve and accelerate the generation of high quality Real World Evidence (RWE) in Europe and beyond.

In this role, you will work with partners, sponsors, and collaborators to support the maintenance and improvement of vocabularies and ongoing RWE studies and related activities; maintain and enhance OHDSI vocabularies, and improve their testing and documentation; improve and increase the maintenance of European clinical/pharmaceutical vocabularies and terminologies, and their interaction with OHDSI vocabularies; and contribute to the development, maintenance, improvement, and validation of computable phenotypes for the identification of specific cohorts executable across the EHDEN network. **More details here. Deadline for application: 14/05/2026**

Feel free to contact the Management Office by email if you'd like more information: enquiries@ehden.eu

EUROPEAN HEALTH DATA & EVIDENCE NETWORK
EHDEN Foundation

JOB DESCRIPTION AND SELECTION CRITERIA

Job title	**Clinical terminology scientist**
Location	Online work from Europe
Annual salary	€45,000 to €75,000 per annum, depending on experience and qualifications
Hours	Full time or Part-time
Contract type	Fixed-term for 1 year
Reporting to	EHDEN Foundation Board
Vacancy reference	26/001

Research topic	Clinical terminology, vocabularies
EHDEN Foundation web site	See website www.ehden.eu
Technical skills	Medical, pharmaceutical, or health sciences Healthcare data standards (HL7 FHIR, OMOP CDM) SQL programming Clinical terminologies (e.g. SNOMED CT, ICD-10) Relational database expertise

Application deadline for this position is May 14, 2026.

ehden.eu/careers

Columbia DBMI Summer School

The 2026 Summer School in Observational Health Data Science & Informatics, AI, and Real World Evidence

COLUMBIA | COLUMBIA UNIVERSITY DEPARTMENT OF BIOMEDICAL INFORMATICS

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

June 22–26, 2026, Columbia Biomedical Informatics

The Columbia OHDSI Summer School provides health professionals, researchers, and industry practitioners with an immersive, hands-on training to working with real-world health data and generating real-world evidence (RWE). Participants will explore the types of healthcare data captured during routine clinical care—such as electronic health records and administrative claims—and learn how to standardize these data using the OMOP Common Data Model to support collaborative, distributed research as part of a data network.

Over the course of the week, participants will engage with three real-world analytic use cases:

- **Clinical characterization** – using descriptive epidemiology to study disease natural history and treatment patterns
- **Population-level estimation** – applying causal inference to assess drug safety and comparative effectiveness
- **Patient-level prediction** – leveraging machine learning for early disease detection and precision medicine

Participants will be guided through the full RWE study lifecycle: from designing observational studies tailored to each use case, to applying open-source tools form the OHDSI community, and executing analyses across real-world data sources.

The curriculum combines foundational lectures on analytical methods with hands-on, interactive, faculty-led group exercises. In addition, participants will have dedicated time to develop and advance their own study concepts with personalized feedback and mentoring.

EUROPE

You've built your OMOP CDM. Now let's make it shine!

Take your OMOP implementation to the next level

- 2.5-day hands-on expert training (max. 30 participants)
- Work with Data Quality Dashboard, ATLAS, CohortDiagnostics
- Learn study design, validation, execution
- Optional: bring your own ETL + OMOP CDM

What to expect

- Practical OMOP expertise
- Concrete CDM improvements
- Skills for evidence generation

Practical

🗓️ 7–9 September 2026
📍 Rotterdam

#OHDSISocialShowcase This Week

Monday

Cohort-Pilot: Collaborative AI for Translating Natural Language into Actionable Cohort Analytics

(Alvaro A. Alvarez, Farnoosh Sheikhi, Priya Desai)

Cohort-Pilot:

Collaborative AI that Translates Natural Language into useable / ready-to-use SQL queries

Background

- Existing tools are powerful but still require manual concept curation, local SQL adjustments, and fragmented workflows with ad-hoc search.
- Cohort-Pilot is our current approach to address these limitations by translating Natural Language (NL) into validated queries for OMOP CDM on BigQuery.
- It combines human expertise with AI-driven automation to help create more accurate, custom cohorts. The key motivation is accessibility and reducing the need for SQL expertise so that cohorts can be defined directly in plain English.

Methods

- Cohort-pilot features a multi-agent architecture where specialized AI agents manage tasks such as natural language interpretation, queries generation, syntax correction, and heuristic validation.
- It supports comprehensive OMOP domain coverage, delivers data completeness insights, and offers flexible AI model integration, with ongoing development focused on enhancing reliability and refining functionality through community feedback.
- Vocabulary integration uses OMOP hierarchies and athena-client (supporting tool) for precise concept exploration *(see companion poster for details)*.
- Workflow employs an iterative Generate → Correct → Review loop to ensure high accuracy and reliability, integrates automated validation checks to verify SQL safety, and uses a modular design separating vocabulary lookups,

Cohort-Pilot turns plain English cohort definitions into tested BigQuery ready SQL for OMOP

Figure A — Cohort-Pilot Workflow

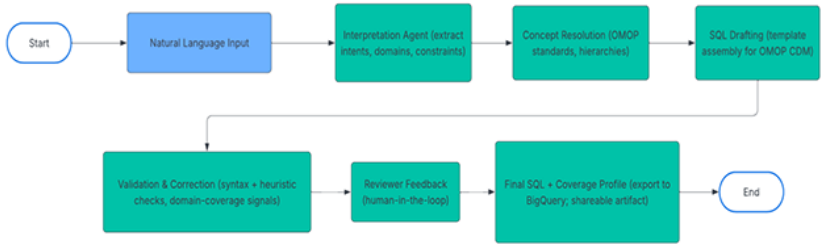
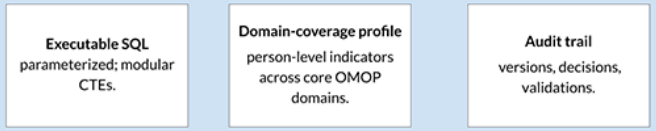


Figure B — Outputs Snapshot



Source Code

Alvaro A Alvarez,
Farnoosh Sheikhi,
Priya Desai

Technology & Digital Solution
Stanford Health Care
School of Medicine

Results

- Task-specialized agents improved reliability and consistency of SQL outputs.
- Explicit feedback loops markedly increased precision and recall of concept sets.
- Structured state management simplified debugging and reproducibility.
- Aggregated outputs with context reduced user errors and accelerated review.
- Next step: head-to-head comparisons vs ATLAS for accuracy, time-to-SQL, and completeness.

Conclusion

- Cohort-Pilot modernizes cohort construction by fusing human expertise with specialized AI agents.
- Delivers precision, transparency, and speed for OMOP workflows on BigQuery.
- Ongoing community validation to refine heuristics, templates, and review UX.

References

1. Hripcsak G. et al. *PNAS* 2016 — OHDSI network characterization.
2. Voss E.A. et al. *JAMIA* 2017 — OMOP CDM applications.
3. OHDSI Athena Vocabulary docs.
4. athena-client (support tool; details in companion poster)
5. GitHub: https://github.com/aandresalvarez/cohortgen .
6. Bridging the Language Gap: Generative Models for Efficient Medical Concept Discovery (OHDSI Global Symposium 2024, Collaborator Showcase honorees).

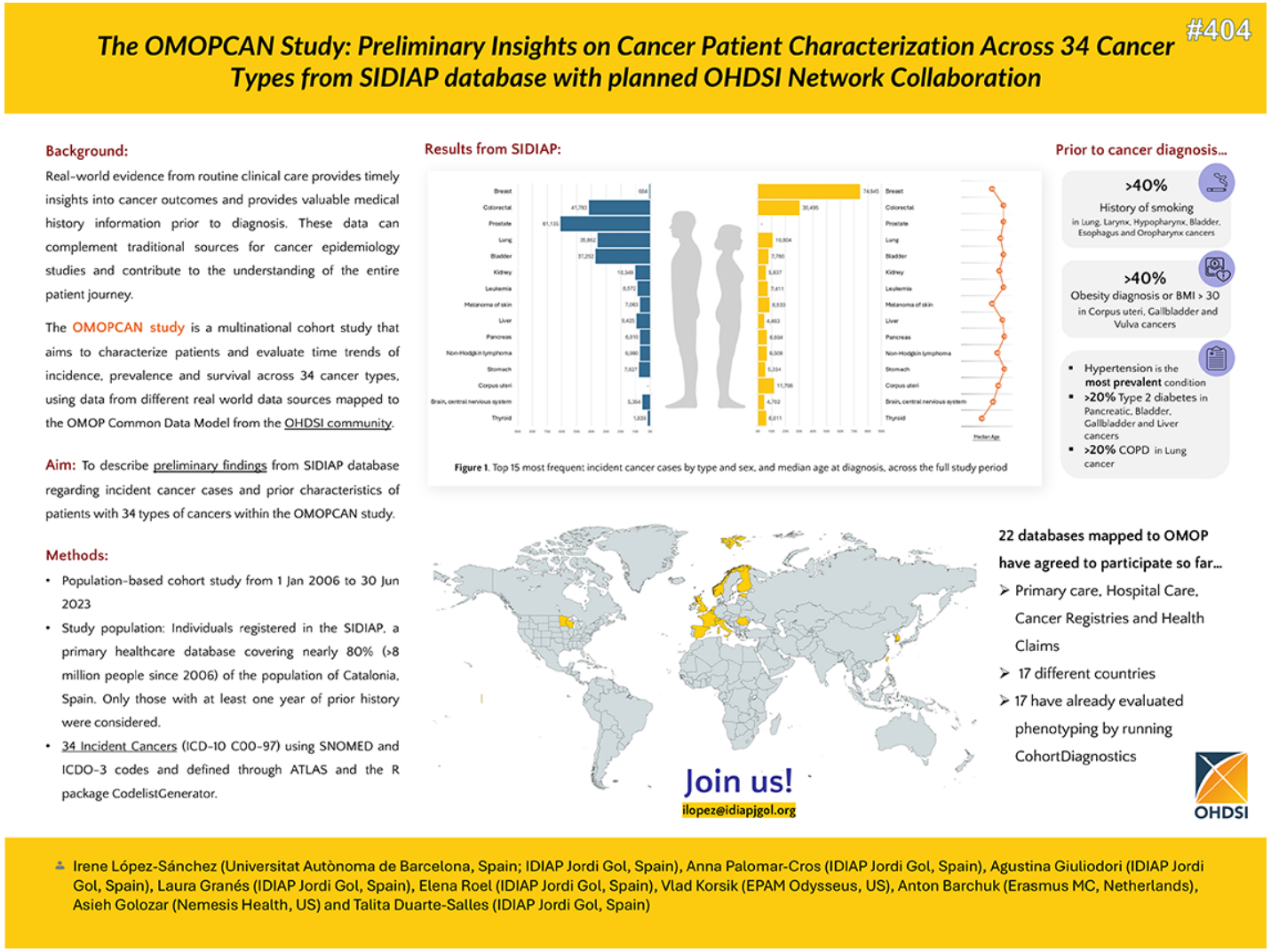
OHDSI

#OHDSISocialShowcase This Week

Tuesday

The OMOPCAN Study: Preliminary Insights on Cancer Patient Characterization Across 34 Cancer Types from SIDIAP database with planned OHDSI Network Collaboration

(Irene López Sánchez, Anna Palomar-Cros, Agustina Giulidori, Laura Granés, Elena Roel, Vlad Korsik, Anton Barchuk, Asieh Golozar, Talita Duarte-Salles)



#OHDSISocialShowcase This Week

Wednesday

Compositional Public Health Approaches to Observational Health Research

(Jacob S. Zelko, Nathaniel Osgood)

Compositional Public Health Approaches to Observational Health Research

PRESENTER: Jacob S. Zelko

INTRODUCTION

- As public health threats emerge, adaptive and robust analysis pipelines are crucial
- Managing multiple data sources in these scenarios can be labor-intensive and error-prone
- Compositional public health **empowers researchers** to **reclaim agency** through **abstraction** in **public health**.

BACKGROUND

Definition:
Compositional Public Health

*Compositional public health is an emerging field at the **intersection** of **category theory**, **epidemiology**, **systems science**, and **engineering**, and utilizes tools from applied category theory **for public health applications**.* [1, 2]

- Arose from public health work in:
  - Stock and Flow Modeling
  - Epidemic Modeling
  - Agent-Based Modeling

METHODS

Category Theory:
How Things Relate To Things

"Category theory takes a bird's eye view of mathematics. From high in the sky, details become invisible, but we can spot patterns that were impossible to detect from ground level." - Tom Leinster [3]

- **Category:** fundamental structure
  - Collection of objects
  - Arrows ($f : A \to B$)
- **Functor:** Can go between categories
  - Objects to objects, arrows to arrows
  - Notation: $F : C \to D$
- Domain applications include:
  - Higher Order Logic
  - Relational Algebra (i.e. SQL)
  - Programming Language Design

Compositional public health **empowers** researchers to **reclaim agency** through **abstraction** in **public health.**

AlgebraicJulia: Compositionality for Technical Computing

Overview

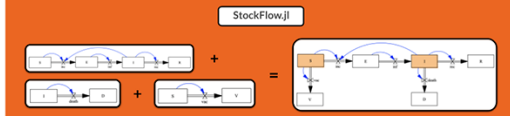
- Framework for applied [...] category theory
- Open source research software stack written in Julia
- Programming [...] interface for applications of category theory [4, 5]

ACSets.jl

- Built on top of ACSets (Attributed Copresheaves) [6]
- In-memory columnar database
- Can represent **Schema** instances:
  - Objects: Database tables
  - Arrows: Columns

Applied Category Theory in Epidemiology

StockFlow.jl



- Exploring how to compose stock and flow diagrams for population dynamics [7]
- Separates the syntax from the semantics of the problem
- Supports algebraic rewriting to prevent unnecessary computation

AlgebraicPetri.jl

- Design to build and compose Petri net agent based models [8]
- Simulate several different intervention policy scenarios
- Combine or stratify multiple ODE systems together

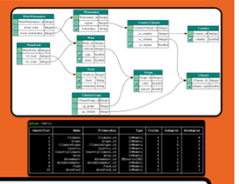
Data Science with Category Theoretic Machinery

Master's Project

- ACSet-ification of the OMOP CDM (via AlgebraicJulia)
- Harmonization with census, climate, and patient data (via JuliaHealth)
- Category theoretic analysis with conjunctive queries

Data Fabric

- A unified access protocol for heterogeneous data [9]
- Consists of an ACSet which records relations between data sources
- Supports federation across heterogeneous data sources



Interdisciplinary Collaborative Modeling

CatColab

- CatColab is a structure editor for categorical structures [10]
- Edited content is a structured object
- Provides mathematical syntax by construction
- Provides generality across structures

ModelCollab

- Application for system dynamics modelling [11]
- Collaborative user interface
- Backed by AlgebraicJulia tools
- Generates model scripts for analysis and development

APPLICATIONS

- Reason about data at a relationship level
- Compose, adjust, and reuse processes
- Expand more robustly to encompass other data resources

DISCUSSION

- Mixes public health domain needs with ACT approaches
- Bridges mathematicians and public health practitioners

FUTURE DIRECTIONS

- Formal concept analysis of phenotype definitions [12]
- How best to represent an OMOP CDM-based study categorically
- Preserving relationships in datasets with missing or partially available features
- Improving methodology to capture metadata and paradata of data sources.

CONCLUSION

Compositional public health provides a framing to couch state of the art observational health methods while adapting to growing needs in public health scenarios.

ACKNOWLEDGEMENTS

The Topos Institute, GATAS Lab, ACT community & AlgebraicJulia communities, Dr. Sean L. Wu, and Matt Cuffaro

REFERENCES

[1] Osgood, Nathaniel, Jacob S. Zelko, and Nastaran Jamali. "Applied Category Theory for Public Health." Transactions in Category Theory, June 2025. https://ologermaster.xyz/TAC133.html.
[2] Zelko, Jacob S. "An Introduction to Compositional Public Health." New York City Category Theory Seminar, Feb. 2025. Available at: https://www.youtube.com/watch?v=4QKHtmFaE
[3] Leinster, Tom. Basic category theory. Vol. 143. Cambridge University Press, 2014.
[4] J. P. Fairbanks, "Scientific computing with categories," in IWD4: Applied category theory for compositional decision making, american control conference 2025, 2025.
[5] J. P. Fairbanks and D. Lynch, "Computational category theory in applied mathematics," in 2023 joint mathematics meetings (JMM 2023), AMS.
[6] E. Patterson, D. Lynch, and J. Fairbanks, "Categorical data structures for technical computing," Compositionality, vol. 4, 2022.
[7] J. Baez, X. Li, S. Libkind, N. D. Osgood, and E. Patterson, "Compositional modeling with stock and flow diagrams," arXiv preprint arXiv:2209.08373, 2022.
[8] S. Libkind, A. Baas, M. Halter, E. Patterson, and J. P. Fairbanks, "An algebraic framework for structured epidemic modeling," Philosophical Transactions of the Royal Society A, vol. 380, no. 2233, p. 20210309, 2022.
[9] Zelko, Jacob S., Matt Cuffaro, and Wu L. Sean. "ACT-Informed Data Science: A Case Study in Public Health Research." 8th International Conference on Applied Category Theory (ACT), June 2025. https://gatalab.org/act2025/act2025.
[10] E. Patterson, "CatColab: Toward collaborative modeling with categorical logics," Topos Institute Berkeley Seminar, 2024.
[11] J. Baez, John C., Xiaoyan Li, and Nathaniel D. Osgood. "ModelCollab: Software for 'Compositional Modeling.'"
[12] Black, Adam, et al. "Constructing Vaccine Vocabulary Hierarchy Using Formal Concept Analysis." 2022 OHDSI Symposium, Oct. 2022.

Jacob S. Zelko, Nathaniel Osgood

UNIVERSITY OF SASKATCHEWAN

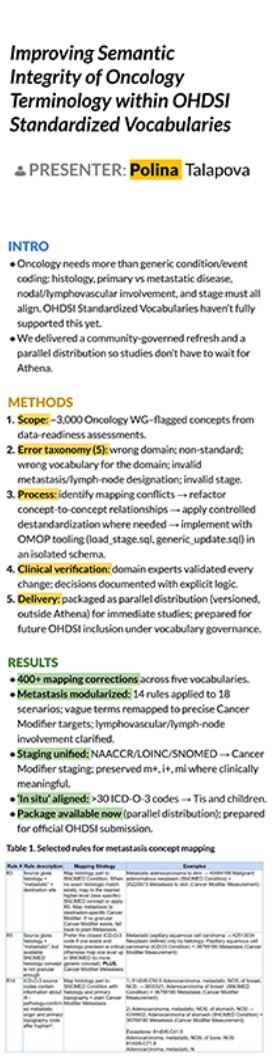
OHDSI

#OHDSISocialShowcase This Week

Thursday

Improving Semantic Integrity of Oncology Terminology within OHDSI Standardized Vocabularies

(Bohdan Khilchevskyi, Maksym Trofymenko, Polina Talapova, Denys Kaduk, Asieh Golozar, Christian Reich)



Improving Semantic Integrity of Oncology Terminology within OHDSI Standardized Vocabularies

PRESENTER: Polina Talapova

INTRO

- Oncology needs more than generic condition/event coding: histology, primary vs metastatic disease, nodal/lymphovascular involvement, and stage must all align. OHDSI Standardized Vocabularies haven't fully supported this yet.
- We delivered a community-governed refresh and a parallel distribution so studies don't have to wait for Athena.

METHODS

1. **Scope:** ~3,000 Oncology WG–flagged concepts from data-readiness assessments.
2. **Error taxonomy (5):** wrong domain; non-standard; wrong vocabulary for the domain; invalid metastasis/lymph-node designation; invalid stage.
3. **Process:** identify mapping conflicts → refactor concept-to-concept relationships → apply controlled destandardization where needed → implement with OMOP tooling (load_stage.sql, generic_update.sql) in an isolated schema.
4. **Clinical verification:** domain experts validated every change; decisions documented with explicit logic.
5. **Delivery:** packaged as parallel distribution (versioned, outside Athena) for immediate studies; prepared for future OHDSI inclusion under vocabulary governance.

RESULTS

- **400+ mapping corrections** across five vocabularies.
- **Metastasis modularized:** 14 rules applied to 18 scenarios; vague terms remapped to precise Cancer Modifier targets; lymphovascular/lymph-node involvement clarified.
- **Staging unified:** NAACCR/LOINC/SNOMED → Cancer Modifier staging; preserved m+, i+, mi where clinically meaningful.
- **'In situ' aligned:** >30 ICD-O-3 codes → Tis and children.
- **Package available now** (parallel distribution); prepared for official OHDSI submission.

Table 1. Selected rules for metastasis concept mapping

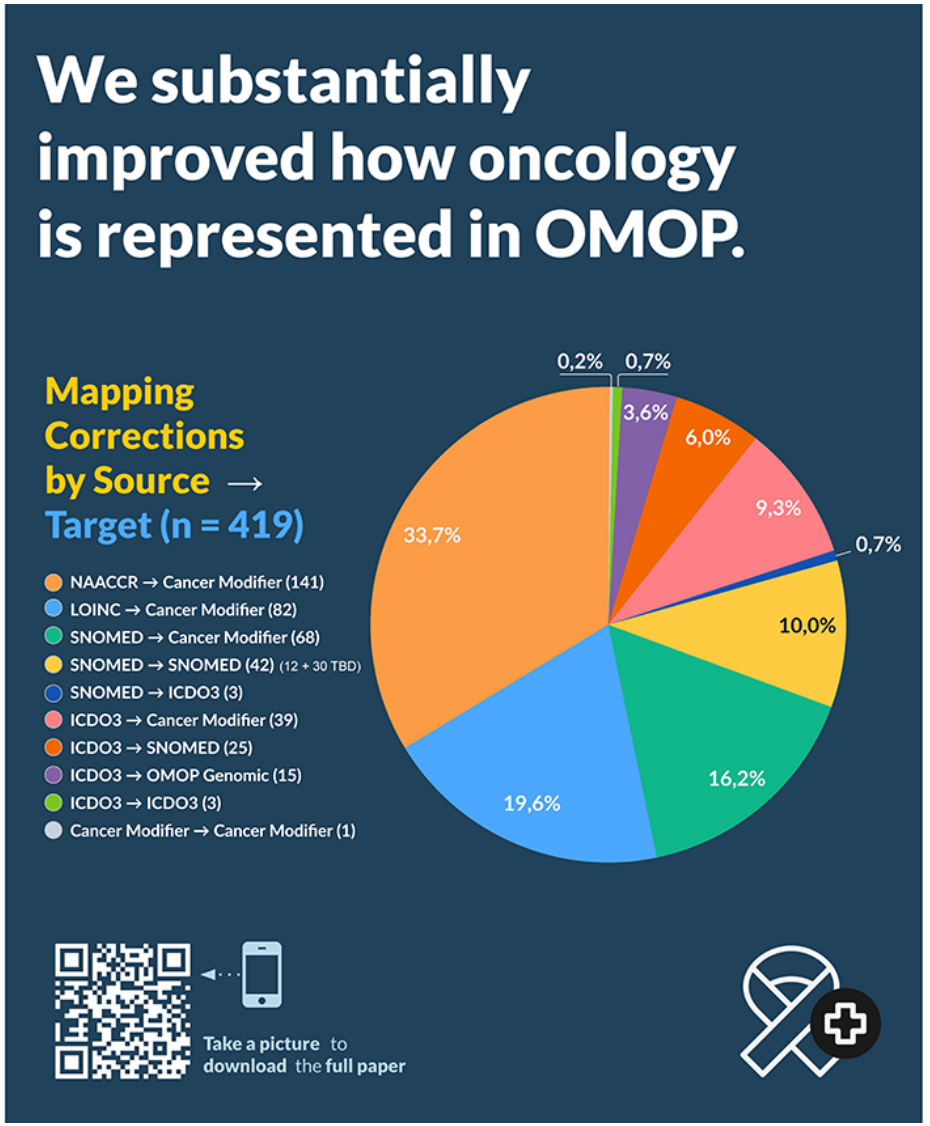


Table 2. Selected examples of metastasis concept mapping

Table 3. Mappings for lymph node and lymphovascular involvement concepts

Table 4. Mapping of oncology staging concepts to Cancer Modifier

LIMITATIONS

- Parallel package ≠ Athena (yet); version-pin to avoid drift.
- Initial focus on ~3,000 concepts: long-tail rarities still pending.
- Staging conventions evolve; monitor AJCC/UICC updates and rerun deltas each cycle.

CONCLUSIONS

- Significantly improved the conceptual integrity and interoperability of oncology-related OHDSI Standardized Vocabularies.
- Directly supports more precise cohort creation, staging alignment, and advanced oncology analytics across observational datasets.
- Demonstrates the value of community contribution - leveraging shared expertise, open governance, and iterative validation - to address complex vocabulary challenges.
- Provides strategies and mappings as a reproducible model for future vocabulary-refinement efforts.
- Reinforces the role of the OHDSI community in advancing high-quality, semantically consistent research infrastructure.
- Calls for continued participation to sustain and expand these gains.

THE TEAM:
Bohdan Khilchevskyi,
Maksym Trofymenko,
Polina Talapova,
Denys Kaduk,
Asieh Golozar,
Christian Reich

sciforce

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

#OHDSISocialShowcase This Week

Friday

Repurposing Drugs in the ICU through Real-world Data Analysis using OMOP CDM (OHDSI)

(Daniella Garofalo, Xiaohan Tanner Zhang, Amy Chuang, Dario Kuzmanovic, Reham Khan , Nicholas Mohr, Smith Heavner)

REPURPOSING DRUGS IN THE ICU THROUGH REAL-WORLD DATA ANALYSIS USING OMOP CDM (OHDSI)

Garofalo D[1], Zhang XT[2], Chuang A[1], Kuzmanovic D[1], Khan R[3], Mohr N[4], Heavner SF[5,6]
[1]Health Data Innovation Program (HDIP), Keck Medicine of USC, [2]Johns Hopkins Medicine Department of General Internal Medicine, [3]Society for Critical Care Medicine, [4]University of Iowa Medicine, Department of Emergency Medicine, [5]Critical Path Institute, Data Collaboration Center, [6]Clemson University, Department of Public Health Sciences

Keck Medicine of USC
SCCTSI Southern California Clinical and Translational Science Institute

BACKGROUND

The need for real-world data (RWD) and real-world evidence (RWE) in the clinical research industry is growing. While an established precedent in observational studies, a rapidly increasing body of guidance and methodological publications now outline use cases for RWD has expanded to interventional studies as well. These include:

- **Feasibility assessments** and **site selections** for randomized controlled trials (RCTs)
- **Exploratory analyses** to aid in the design of pragmatic trials
- **External controls** for non-randomized interventional studies

One area of interest is off-label drug use and drug repurposing.[1,2,3,4,5] Silico studies also heavily rely on RWD, potentially leading to hypotheses to be tested in further studies.[6] However, multiple proprietary data models across healthcare institutions make these data challenging to obtain.[7,8,9]

Repurposing Drugs in Intensive Care Units through RWD Analysis (REDISCOVER-ICU) is a multi-institutional effort led by the Critical Path Institute (C-Path) and the Society of Critical Care Medicine (SCCM). It aims to establish a versatile repository for RWD to support drug repurposing and clinical research in critical care. This effort seeks to harness the Observational Health Data Sciences and Informatics (OHDSI) data harmonization tools while curating large, high-quality, de-identified datasets for diseases with significant unmet clinical needs, including sepsis.

This poster highlights Keck Medicine of USC's approach to cohort definition, Extract, Transform, and Load (ETL) processes, and preliminary data quality findings through its participation in the REDISCOVER-ICU collaborative.

- Participating sites **contributed OMOP Common Data Model data**, enabling standardized and scalable analytics across disparate institutions and EHRs
- The initiative focused on generating **curated, de-identified cohorts** of hospitalized and ICU patients with sepsis
- At Keck Medicine of USC, we implemented a **site-specific pipeline to extract, validate, and prepare OMOP data** to include in the registry
- **Additional enhancements** included manual concept mapping, vocabulary reconciliation, and data quality checks to ensure accurate clinical event representation
- The curated dataset **supports future collaborative research** efforts in phenotyping, predictive modeling, and benchmarking care practices in the ICU setting

METHODS

We generated a de-identified cohort dataset of septic hospitalized patients from a local instance of the OMOP CDM., We also developed an ARES (Atlas-based Evidence Synthesis) to ensure comprehensive data quality and validate the integrity of the exported cohort. Outputs from the ARES included detailed data quality reports and visualizations crucial for identifying discrepancies, discovering clinical implausibility, guiding iterative feedback, and informing timely data refinements and fixes.[10] Key steps for data extraction and preparation included:

- **Identify the Cohort:** A sepsis cohort was identified from the local OMOP instance using SNOMED concept ID 132797 (Sepsis) and concept 43021283 (infection associated with vascular device), restricted to inpatient and ED-inpatient visits [visit_concept_id IN (9201, 262)] occurring after January 1, 2020
- **Construct the Domain Table:** Key OMOP domain tables were systematically filtered to include only records related to the identified cohort which formed the foundational datasets of the REDISCOVER-ICU registry
- **Enhance Concept Mapping:** A manual review of source code descriptions and corresponding source data were mapped to the REDISCOVER ICU concepts of interest. Data elements that matched the conceptual definitions but were missing from the initial extract (whether due to unmapped concept IDs or alternative vocabularies such as LOINC instead of SNOMED) were then added to ensure completeness.
- **Roll-up Concept for Rare Conditions:** Condition concepts with ≤10 occurrences were rolled up to their nearest parent concept with ≥10 occurrences to support meaningful aggregate analyses and address data sparsity
- **De-identify and Prepare Data Sharing:** Identifiable fields were removed to produce a de-identified version, fully compliant with data sharing policies governing the REDISCOVER-ICU registry
- **Check Data Quality:** Data quality checks from the ARES instance were extended.[11] Summary-level audits were performed on all domain tables to confirm record counts, validate date ranges, and ensure internal consistency across the dataset.
- **Profile Prevalence:** Prevalence profiles were generated across multiple domains to characterize the cohort and support future federated research. The domains included were:
  - **Conditions:** Aggregated by parent concept, calculated per patient
  - **Measurements:** Aggregated by parent concept, calculated per patient
  - **Drugs:** Aggregated by the drug ingredient
  - **Unmapped Drugs:** Drug source values with drug_concept_id = 0 and ≥20 instances flagged for manual review.
  - **Devices:** Aggregated by parent concept, calculated per patient

RESULTS

We identified 8,825 patients who met sepsis cohort criteria, accounting for 11,870 inpatient or emergency inpatient visits. Subsequent analyses from this defined cohort reflect OMOP CDM data extracts from Keck Medicine. Clinical characteristic and care pattern insights for this population during these visits include most prevalent conditions, clinical measures, administered drugs, and utilized devices observed.

Table 1. Top Prevalent Conditions, Drugs, and Devices/Procedures

Domain Table	Concept Name	Concept ID	Persons Count	% Persons	Avg Entries Per Person	Max Entries Per Person
Condition	Hypertensive Disorder	316,866	6,763	76	N/A	N/A
Condition	Chronic disease of the cardiovascular system	4,828,244	5,597	62	N/A	N/A
Condition	Chronic disease of the respiratory system	4,063,381	5,283	59	N/A	N/A
Condition	Diabetes Mellitus	201,820	4,786	54	N/A	N/A
Condition	Chronic Obstructive Lung Disease	255,573	4,604	52	N/A	N/A
Drug	Alka-Seltzer Plus cold and cough cool menthol, dextromethorphan hydrobromide, chlorpheniramine maleate, acetaminophen, phenylephrine hydrochloride capsule, liquid filled	44,261,362	8,307	96	21	806
Drug	Acetaminophen	1,125,315	8,249	95	21	835
Drug	Ondansetron	1,000,560	7,939	92	5	298
Drug	Sodium	19,136,648	7,806	90	8	745
Drug	Potassium Chloride	19,049,105	7,365	85	17	774
Device/ Procedure	Ventilator	45,761,108	8,583	97	276	8,480
Device/ Procedure	Oxygen ventilator	4,138,916	8,572	97	410	16,869

Table 2. Top 5 Most Prevalent Measurements

Concept Name	Concept ID	%tile 25	Median	%tile 75	%tile 95	Interpretation
Diastolic blood pressure	3012888	108	277	595	1514	For half of the patients, there were at least 277 records. Most patients (25th-75th percentile) had 108-595 records. 5% of patients had 1,514 or more records.
Heart rate	3027018	34	138	476	1794	For half of the patients, there were at least 138 records. Most patients (25th-75th percentile) had 34-476 records. 5% of patients had 1,794 or more records.
Glasgow Coma Scale/GCS	3032652	23	55	130	348	For half of the patients, there were at least 55 records. Most patients (25th-75th percentile) had 23-130 records. 5% of patients had 348 or more records.
Inhaled oxygen flow rate	3005629	12	50	146	488	For half of the patients, there were at least 50 records. Most patients (25th-75th percentile) had 12-146 records. 5% of patients had 488 or more records.
Creatinine and Glomerular filtration rate predicted panel - Serum, Plasma, or Blood	35615262	12	30	76	257	For half of the patients, there were at least 30 records. Most patients (25th-75th percentile) had 12-76 records. 5% of patients had 257 or more records.

CONCLUSIONS

The creation of this high-quality, de-identified sepsis cohort in the OMOP CDM provides a robust foundation for a range of observational studies, including clinical characterization, population-level effect estimation, and patient-level prediction within the REDISCOVER-ICU initiative. Building this cohort also allowed us to evaluate the scalability of shared ETL strategies across highly heterogeneous ICU data, including titrated medications, oxygen delivery devices, and other complex clinical variables, while leveraging OHDSI tools, such as Athena, to refine concept mappings and improve vocabulary harmonization.

It also revealed several practical lessons that improved our local OMOP implementation and deepened our understanding of how to prepare high-quality ICU data for collaborative research.

Participation in REDISCOVER ICU Benefits and Lessons Learned

- **Participation provided the local team a starting point** to build from shared cohort scripts rather than beginning from scratch, reducing setup time and complexity
- **Data Validation provided the opportunity to highlight the local use of non-standard concepts** by identifying missing or unmapped vocabularies (concept_id = 0) and site-specific codes that impact data retrieval
- **Process helped prioritize the standardization of custom concepts** by focusing on high-impact mappings that enabled cohort generation without requiring full-scale remediation upfront
- **Exposure to approaches and expertise from other sites** provided insights into handling ICU-specific data, such as ventilator usage stored in different OMOP domains
- **Learning from other sites** created professional development opportunities and identified knowledge gaps, expanding team expertise in OMOP, ETL processes, and vocabulary mapping through collaborative problem-solving

Benefits to the C-PATH Collaborative:

- **Testing Code Across Environments:** Running shared queries at multiple institutions revealed logic gaps and performance issues that would not appear in isolated test settings
- **Identifying Vocabulary Gaps:** Feedback from participating sites highlighted missing or mismatched concepts across vocabularies (e.g., cases where equivalent LOINC and SNOMED concepts needed inclusion in queries)
- **Sharing Data Appropriately:** Strengthened processes for privacy-conscious data sharing that support public health research, drug development, and critical care studies

REFERENCES

Where Are We Going?

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

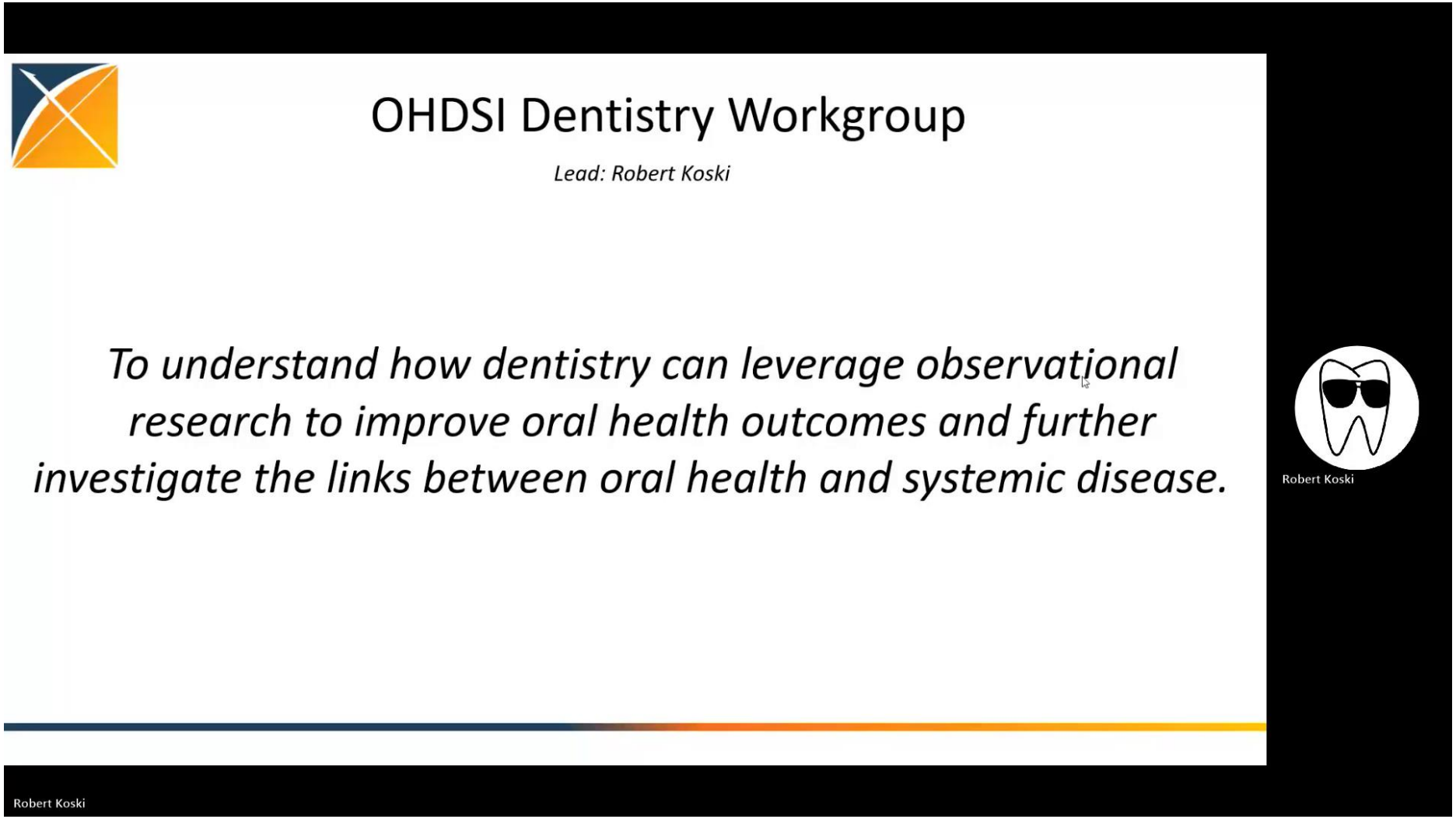
Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Dentistry OKRs



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