



# Collaborator Showcase Brainstorm

OHDSI Community Call  
May 12, 2026 • 11 am ET



# 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	<b>CANCELLED: OHDSI Summer School at Columbia University</b>
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](mailto: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*.**



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



#### OPEN ACCESS

EDITED BY  
Serafeim - Chrysovalantis Kotoulas,  
Hippokraton 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  
© 2026 Shin, Jang and Lee. This is an  
open-access article distributed under the  
terms of the Creative Commons  
Attribution License (CC BY). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

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

Woo-young Shin<sup>1\*</sup>, Ha Young Jang<sup>2</sup> and Kiyoung Lee<sup>3</sup>

<sup>1</sup>Department of Family Medicine, Chung-ang University Gwangmyeong Hospital, Chung-Ang University College of Medicine, Seoul, Republic of Korea, <sup>2</sup>College of Pharmacy, Gachon University, Incheon, Republic of Korea, <sup>3</sup>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

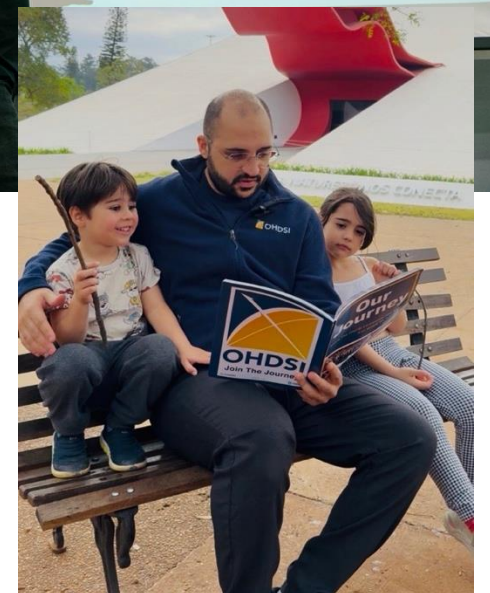
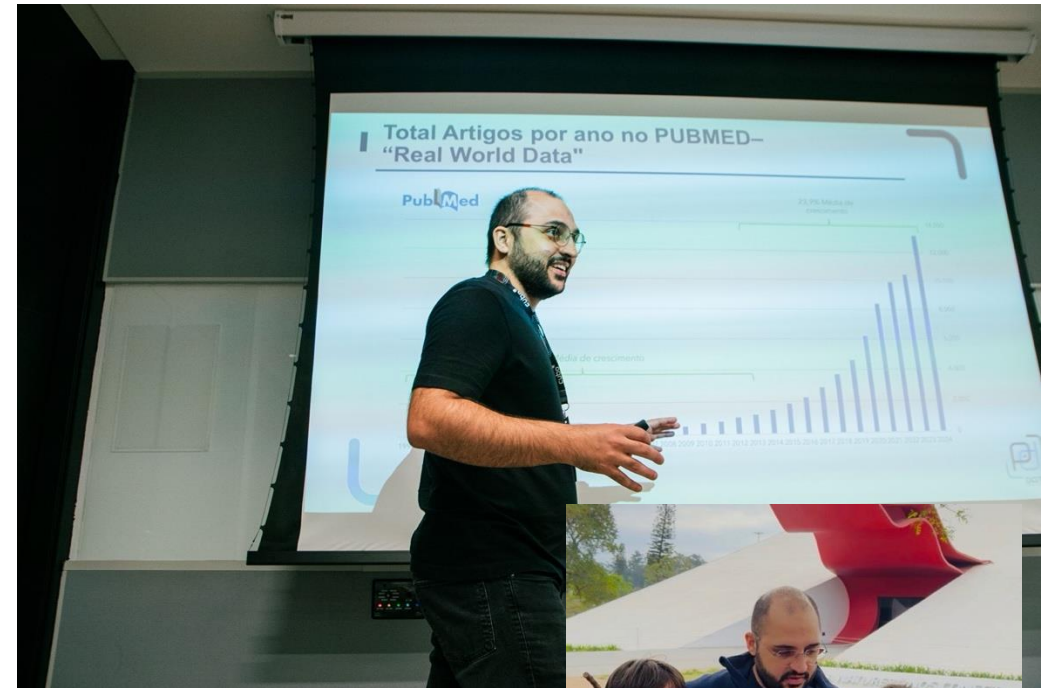


Date	Time (ET)	Meeting
Tuesday	12 pm	Generative AI and Analytics
Wednesday	7 am	Medical Imaging
Wednesday	8 am	Psychiatry
Wednesday	9 am	Patient-Level Prediction
Wednesday	2 pm	Natural Language Processing
Wednesday	7 pm	Eyecare and Vision Research
Thursday	10 am	Rare Diseases
Thursday	10 am	GIS - Geographic 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](https://ohdsi.org/spotlight-julio-oliveira)



# May Newsletter Is Available

The screenshot shows the OHDSI website header and navigation menu. The OHDSI logo is prominently displayed at the top left. Below it, the text "OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS" is visible. The navigation menu is a dark grey bar with white text and dropdown arrows. The "Newsletters" dropdown menu is open, showing a list of newsletter issues from "Subscribe" to "Full Archive". The "May 2026" option is highlighted with an orange circle. Below the navigation menu, the main content area features two columns of text. The left column is titled "Welcome to OHDSI!" and the right column is titled "Join us at the 2026 OHDSI Global Symposium".

**OHDSI**  
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Who We Are ▾ Updates & News ▾ Standards Software Tools ▾ Network Studies ▾ Community Forums ▾ Education ▾ New To OHDSI? ▾

Community Calls ▾ Past Events ▾ Workgroups ▾ Tutorials 2025 'Our Journey' Annual Report Current Events ▾ Support & Sponsorship

2025 Global Symposium ▾ 2026 Europe Symposium 2026 Global Symposium ▾ Github YouTube X/Twitter LinkedIn **Newsletters ▾**

- Subscribe
- May 2026**
- April 2026
- March 2026
- February 2026
- January 2026
- December 2025
- Full Archive

## Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

## Join us at the 2026 OHDSI Global Symposium

Registration and the call for participation OPEN for the 2026 OHDSI Global Symposium which will be held Oct. 20-22 in New Brunswick, NJ. This event unites hundreds of collaborators to showcase scientific innovations and build

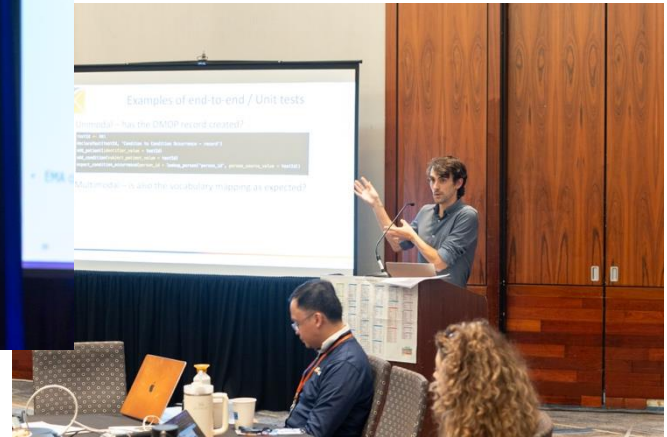


# 2026 OHDSI Global Symposium

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](https://ohdsi.org/OHDSI2026)

[www.ohdsi.org](https://www.ohdsi.org)

#JoinTheJourney





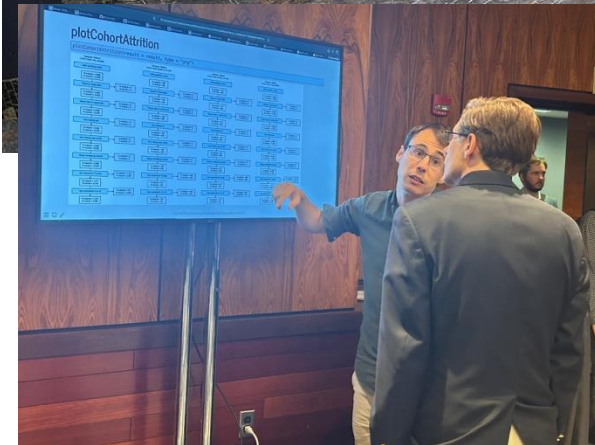
# 2026 OHDSI Global Symposium

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](https://ohdsi.org/OHDSI2026)



# Maternal Fellowship Deadline: May 30



## 2026 Maternal Health Fellowship

### Career Development



- Create evidence from real-world data
- Leverage standard data models for reproducible research
- Build skills on effective network studies

### Practice



- Design effective observational research protocols
- Master OHDSI tools
- Write papers & grants

### Networking



- Build relationships with mentors & fellow learners
- Coordinate with colleagues in the OHDSI data network, spanning 450 sites worldwide & 960 million unique patients

Want to build your career?

Generate reproducible evidence by leading multi-institutional studies!



Find out more and apply via [this link](#)

by **May 31<sup>st</sup>, 2026 !**



# 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.

## Day 1

### Strategic panels with government, academia and industry

Thursday, July 30, 2026



#### Opening and keynote

**Common Data Model for Health Equity: the Role of Latin America.**



#### Panel 1 — Health data interoperability and standards

*Panelists from the Ministry of Health, Bahia State Health Department, PAHO and Latin American Governments.*



#### Panel 2 — The power of administrative data for health research

*Panelists from the Ministry of Health, CONASS, Fiocruz, Latin American Governments, Industry and OHDSI Global.*



#### Panel 3 — The future of interoperability in healthcare in Latin America

A public-private debate.  
*Panelists from the Ministry of Health, CONASS, Fiocruz, private hospitals and Latin American Governments.*

## Day 2

### Hands-on workshops and scientific collaboration

Friday, July 31, 2026



#### Introductory OMOP CDM workshops

- Introduction to OMOP
- Building cohorts with OHDSI tools



#### Parallel tracks of specialized workshops

- ETL to OMOP
- Scientific collaboration



#### Closing

Future perspectives and next steps for the OHDSI Latin America community.

[ohdsilatam.org](https://ohdsilatam.org)

[www.ohdsi.org](https://www.ohdsi.org)

#JoinTheJourney





# Opening: Clinical Terminology Scientist



## 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 <a href="http://www.ehden.eu">www.ehden.eu</a>
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



- 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](mailto:enquiries@ehden.eu)

# Application deadline for this position is May 14, 2026.

[ehden.eu/careers](http://ehden.eu/careers)



# Columbia DBMI Summer School

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

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 from 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.





# The OMOP Practitioner

Data quality . Cohort design . Real-world evidence

Expert training . Rotterdam . September 7-9, 2026

OMOP CDM SCHEME



NEW

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

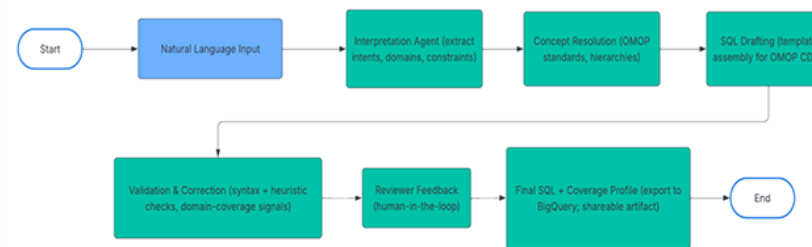
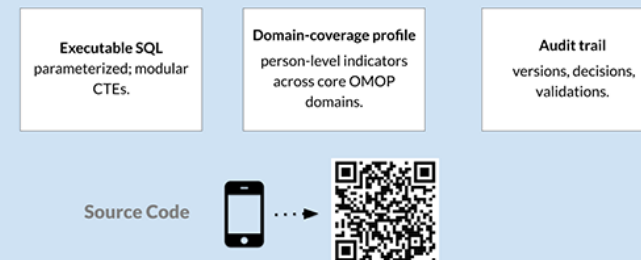


Figure B – Outputs Snapshot



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

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





# #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)

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

#404

##### Background:

Real-world evidence from routine clinical care provides timely insights into cancer outcomes and provides valuable medical history information prior to diagnosis. These data can complement traditional sources for cancer epidemiology studies and contribute to the understanding of the entire patient journey.

The **OMOPCAN study** is a multinational cohort study that aims to characterize patients and evaluate time trends of incidence, prevalence and survival across 34 cancer types, using data from different real world data sources mapped to the OMOP Common Data Model from the **OHDSI community**.

**Aim:** To describe preliminary findings from SIDIAP database regarding incident cancer cases and prior characteristics of patients with 34 types of cancers within the OMOPCAN study.

##### Methods:

- Population-based cohort study from 1 Jan 2006 to 30 Jun 2023
- Study population: Individuals registered in the SIDIAP, a primary healthcare database covering nearly 80% (>8 million people since 2006) of the population of Catalonia, Spain. Only those with at least one year of prior history were considered.
- **34 Incident Cancers** (ICD-10 C00-97) using SNOMED and ICDO-3 codes and defined through ATLAS and the R package CodelistGenerator.

##### Results from SIDIAP:

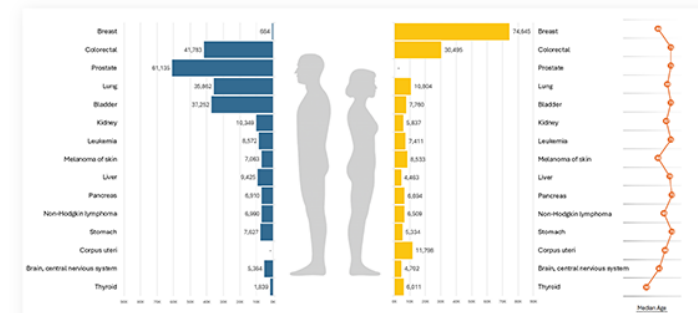


Figure 1. Top 15 most frequent incident cancer cases by type and sex, and median age at diagnosis, across the full study period

##### Prior to cancer diagnosis...

- **>40%** History of smoking in Lung, Larynx, Hypopharynx, Bladder, Esophagus and Oropharynx cancers
- **>40%** Obesity diagnosis or BMI > 30 in Corpus uteri, Gallbladder and Vulva cancers
- Hypertension is the **most prevalent** condition
- **>20%** Type 2 diabetes in Pancreatic, Bladder, Gallbladder and Liver cancers
- **>20%** COPD in Lung cancer



- 22 databases mapped to OMOP have agreed to participate so far...
- Primary care, Hospital Care, Cancer Registries and Health Claims
  - 17 different countries
  - 17 have already evaluated phenotyping by running CohortDiagnostics



▲ Irene López-Sánchez (Universitat Autònoma de Barcelona, Spain; IDIAP Jordi Gol, Spain), Anna Palomar-Cros (IDIAP Jordi Gol, Spain), Agustina Giulidori (IDIAP Jordi Gol, Spain), Laura Granés (IDIAP Jordi Gol, Spain), Elena Roel (IDIAP Jordi Gol, Spain), Vlad Korsik (EPAM Odysseus, US), Anton Barchuk (Erasmus MC, Netherlands), Asieh Golozar (Nemesis Health, US) and Talita Duarte-Salles (IDIAP Jordi Gol, Spain)



# #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]

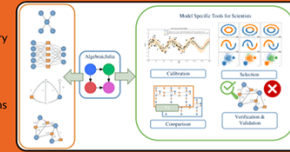
- Category: fundamental structure
  - Collection of objects
  - Arrows ( $f: A \rightarrow B$ )
- Functor: Can go between categories
  - Objects to objects, arrows to arrows
  - Notation:  $F: C \rightarrow D$
- Domain applications include:
  - Higher Order Logic
  - Relational Algebra (i.e. SQL)
  -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

```
Employee = ACSet{Employee}
Employee : Employee
Employee : Employee
Employee : Employee
Employee : Employee
Employee : Employee
Employee : Employee
Employee : Employee
Employee : Employee
Employee : Employee
Employee : Employee
```

#### 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

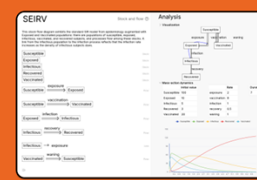
- A unified access protocol for heterogeneous data [9]
- Consists of an ACSet which records relations between data sources
- Supports federation across heterogeneous data sources



#### Interdisciplinary Collaborative Modeling

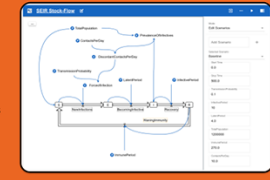
##### CatColab

- 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] Digne, Nathalie, David S. Zaka, and Nathaniel Osgood. "Applied Category Theory for Public Health." *Transactions in Category Theory* June 2022. <https://doi.org/10.1017/S2474785822000001>

[2] Zelko, Jacob S. "An Introduction to Compositional Public Health." *New York City Category Theory Seminar*, Feb. 2023. Available at <https://www.youtube.com/watch?v=102111111111>

[3] Leinster, Tom. *Basic category theory*. Vol. 143. Cambridge University Press, 2014.

[4] F. Farfaglia. "Science computing with categories." *Workshop on Applied Category Theory for Computational Modelling*, American Institute of Mathematics, 2022.

[5] F. Farfaglia and G. Leung. "Compositional category theory in applied mathematics." *2022 Joint Mathematics Meetings (JMM 2022)*, AMS.

[6] Patterson, E. and F. Farfaglia. "Compositional category theory for technical computing." *Compositionality*, vol. 4, 2022.

[7] Bran, K. L., S. Liband, N. D. Osgood, and E. Patterson. "Compositional modeling with stock and flow diagrams." *arXiv preprint arXiv:2201.02011*, 2022.

[8] Liband, A. Bran, M. Hahn, E. Patterson, and J. F. Farfaglia. "An algebraic framework for structural epidemiology." *Preprint*, *Transactions in Category Theory*, vol. 2022, 2022.

[9] Zelko, Jacob S., Matt Cuffaro, and Nathaniel Osgood. "ACE: An Open-Source Algebraic Category Theory Framework for Public Health Research." *8th International Conference on Applied Category Theory (ACT)*, June 2023. [https://doi.org/10.1007/978-3-031-38202-0\\_10](https://doi.org/10.1007/978-3-031-38202-0_10)

[10] E. Patterson. "CatColab: toward collaborative modeling with category theory." *Topos Institute Berkeley Seminar*, 2023.

[11] Bran, Jacob S., Nathaniel Osgood, and Nathaniel Osgood. "ModelCollab: Software for Compositional Modeling." *12th Black Alliance for Community Organizing and Leadership Development (BALD) Conference*, 2022.

Jacob S. Zelko, Nathaniel Osgood





# #OHDSISocialShowcase This Week

## Thursday

# Improving Semantic Integrity of Oncology Terminology within OHDSI Standardized Vocabularies

(**Bohdan Khilchevskiy**, 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 Athens.

### 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+, ix, 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.

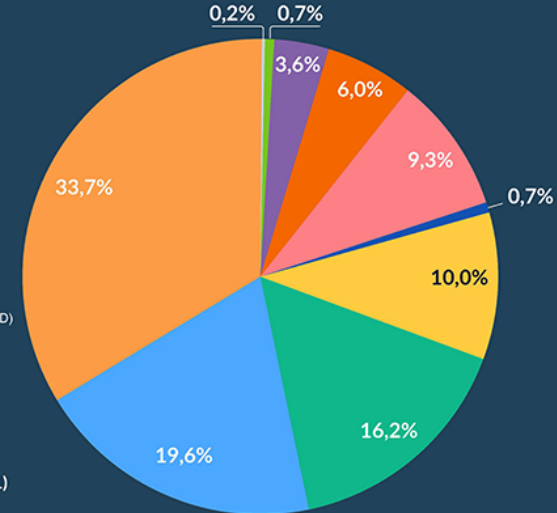
Table 1. Selected rules for metastasis concept mapping

NAACCR	Mapping	SNOMED
801	Metastatic disease	Metastatic neoplasm (C40.00)
802	Primary site	Primary site of neoplasm (C40.01)
803	Primary site	Primary site of neoplasm (C40.02)
804	Primary site	Primary site of neoplasm (C40.03)
805	Primary site	Primary site of neoplasm (C40.04)
806	Primary site	Primary site of neoplasm (C40.05)
807	Primary site	Primary site of neoplasm (C40.06)
808	Primary site	Primary site of neoplasm (C40.07)
809	Primary site	Primary site of neoplasm (C40.08)
810	Primary site	Primary site of neoplasm (C40.09)
811	Primary site	Primary site of neoplasm (C40.10)
812	Primary site	Primary site of neoplasm (C40.11)
813	Primary site	Primary site of neoplasm (C40.12)
814	Primary site	Primary site of neoplasm (C40.13)
815	Primary site	Primary site of neoplasm (C40.14)
816	Primary site	Primary site of neoplasm (C40.15)
817	Primary site	Primary site of neoplasm (C40.16)
818	Primary site	Primary site of neoplasm (C40.17)
819	Primary site	Primary site of neoplasm (C40.18)
820	Primary site	Primary site of neoplasm (C40.19)
821	Primary site	Primary site of neoplasm (C40.20)
822	Primary site	Primary site of neoplasm (C40.21)
823	Primary site	Primary site of neoplasm (C40.22)
824	Primary site	Primary site of neoplasm (C40.23)
825	Primary site	Primary site of neoplasm (C40.24)
826	Primary site	Primary site of neoplasm (C40.25)
827	Primary site	Primary site of neoplasm (C40.26)
828	Primary site	Primary site of neoplasm (C40.27)
829	Primary site	Primary site of neoplasm (C40.28)
830	Primary site	Primary site of neoplasm (C40.29)
831	Primary site	Primary site of neoplasm (C40.30)
832	Primary site	Primary site of neoplasm (C40.31)
833	Primary site	Primary site of neoplasm (C40.32)
834	Primary site	Primary site of neoplasm (C40.33)
835	Primary site	Primary site of neoplasm (C40.34)
836	Primary site	Primary site of neoplasm (C40.35)
837	Primary site	Primary site of neoplasm (C40.36)
838	Primary site	Primary site of neoplasm (C40.37)
839	Primary site	Primary site of neoplasm (C40.38)
840	Primary site	Primary site of neoplasm (C40.39)
841	Primary site	Primary site of neoplasm (C40.40)
842	Primary site	Primary site of neoplasm (C40.41)
843	Primary site	Primary site of neoplasm (C40.42)
844	Primary site	Primary site of neoplasm (C40.43)
845	Primary site	Primary site of neoplasm (C40.44)
846	Primary site	Primary site of neoplasm (C40.45)
847	Primary site	Primary site of neoplasm (C40.46)
848	Primary site	Primary site of neoplasm (C40.47)
849	Primary site	Primary site of neoplasm (C40.48)
850	Primary site	Primary site of neoplasm (C40.49)
851	Primary site	Primary site of neoplasm (C40.50)
852	Primary site	Primary site of neoplasm (C40.51)
853	Primary site	Primary site of neoplasm (C40.52)
854	Primary site	Primary site of neoplasm (C40.53)
855	Primary site	Primary site of neoplasm (C40.54)
856	Primary site	Primary site of neoplasm (C40.55)
857	Primary site	Primary site of neoplasm (C40.56)
858	Primary site	Primary site of neoplasm (C40.57)
859	Primary site	Primary site of neoplasm (C40.58)
860	Primary site	Primary site of neoplasm (C40.59)
861	Primary site	Primary site of neoplasm (C40.60)
862	Primary site	Primary site of neoplasm (C40.61)
863	Primary site	Primary site of neoplasm (C40.62)
864	Primary site	Primary site of neoplasm (C40.63)
865	Primary site	Primary site of neoplasm (C40.64)
866	Primary site	Primary site of neoplasm (C40.65)
867	Primary site	Primary site of neoplasm (C40.66)
868	Primary site	Primary site of neoplasm (C40.67)
869	Primary site	Primary site of neoplasm (C40.68)
870	Primary site	Primary site of neoplasm (C40.69)
871	Primary site	Primary site of neoplasm (C40.70)
872	Primary site	Primary site of neoplasm (C40.71)
873	Primary site	Primary site of neoplasm (C40.72)
874	Primary site	Primary site of neoplasm (C40.73)
875	Primary site	Primary site of neoplasm (C40.74)
876	Primary site	Primary site of neoplasm (C40.75)
877	Primary site	Primary site of neoplasm (C40.76)
878	Primary site	Primary site of neoplasm (C40.77)
879	Primary site	Primary site of neoplasm (C40.78)
880	Primary site	Primary site of neoplasm (C40.79)
881	Primary site	Primary site of neoplasm (C40.80)
882	Primary site	Primary site of neoplasm (C40.81)
883	Primary site	Primary site of neoplasm (C40.82)
884	Primary site	Primary site of neoplasm (C40.83)
885	Primary site	Primary site of neoplasm (C40.84)
886	Primary site	Primary site of neoplasm (C40.85)
887	Primary site	Primary site of neoplasm (C40.86)
888	Primary site	Primary site of neoplasm (C40.87)
889	Primary site	Primary site of neoplasm (C40.88)
890	Primary site	Primary site of neoplasm (C40.89)
891	Primary site	Primary site of neoplasm (C40.90)
892	Primary site	Primary site of neoplasm (C40.91)
893	Primary site	Primary site of neoplasm (C40.92)
894	Primary site	Primary site of neoplasm (C40.93)
895	Primary site	Primary site of neoplasm (C40.94)
896	Primary site	Primary site of neoplasm (C40.95)
897	Primary site	Primary site of neoplasm (C40.96)
898	Primary site	Primary site of neoplasm (C40.97)
899	Primary site	Primary site of neoplasm (C40.98)
900	Primary site	Primary site of neoplasm (C40.99)
901	Primary site	Primary site of neoplasm (C40.00)
902	Primary site	Primary site of neoplasm (C40.01)
903	Primary site	Primary site of neoplasm (C40.02)
904	Primary site	Primary site of neoplasm (C40.03)
905	Primary site	Primary site of neoplasm (C40.04)
906	Primary site	Primary site of neoplasm (C40.05)
907	Primary site	Primary site of neoplasm (C40.06)
908	Primary site	Primary site of neoplasm (C40.07)
909	Primary site	Primary site of neoplasm (C40.08)
910	Primary site	Primary site of neoplasm (C40.09)
911	Primary site	Primary site of neoplasm (C40.10)
912	Primary site	Primary site of neoplasm (C40.11)
913	Primary site	Primary site of neoplasm (C40.12)
914	Primary site	Primary site of neoplasm (C40.13)
915	Primary site	Primary site of neoplasm (C40.14)
916	Primary site	Primary site of neoplasm (C40.15)
917	Primary site	Primary site of neoplasm (C40.16)
918	Primary site	Primary site of neoplasm (C40.17)
919	Primary site	Primary site of neoplasm (C40.18)
920	Primary site	Primary site of neoplasm (C40.19)
921	Primary site	Primary site of neoplasm (C40.20)
922	Primary site	Primary site of neoplasm (C40.21)
923	Primary site	Primary site of neoplasm (C40.22)
924	Primary site	Primary site of neoplasm (C40.23)
925	Primary site	Primary site of neoplasm (C40.24)
926	Primary site	Primary site of neoplasm (C40.25)
927	Primary site	Primary site of neoplasm (C40.26)
928	Primary site	Primary site of neoplasm (C40.27)
929	Primary site	Primary site of neoplasm (C40.28)
930	Primary site	Primary site of neoplasm (C40.29)
931	Primary site	Primary site of neoplasm (C40.30)
932	Primary site	Primary site of neoplasm (C40.31)
933	Primary site	Primary site of neoplasm (C40.32)
934	Primary site	Primary site of neoplasm (C40.33)
935	Primary site	Primary site of neoplasm (C40.34)
936	Primary site	Primary site of neoplasm (C40.35)
937	Primary site	Primary site of neoplasm (C40.36)
938	Primary site	Primary site of neoplasm (C40.37)
939	Primary site	Primary site of neoplasm (C40.38)
940	Primary site	Primary site of neoplasm (C40.39)
941	Primary site	Primary site of neoplasm (C40.40)
942	Primary site	Primary site of neoplasm (C40.41)
943	Primary site	Primary site of neoplasm (C40.42)
944	Primary site	Primary site of neoplasm (C40.43)
945	Primary site	Primary site of neoplasm (C40.44)
946	Primary site	Primary site of neoplasm (C40.45)
947	Primary site	Primary site of neoplasm (C40.46)
948	Primary site	Primary site of neoplasm (C40.47)
949	Primary site	Primary site of neoplasm (C40.48)
950	Primary site	Primary site of neoplasm (C40.49)
951	Primary site	Primary site of neoplasm (C40.50)
952	Primary site	Primary site of neoplasm (C40.51)
953	Primary site	Primary site of neoplasm (C40.52)
954	Primary site	Primary site of neoplasm (C40.53)
955	Primary site	Primary site of neoplasm (C40.54)
956	Primary site	Primary site of neoplasm (C40.55)
957	Primary site	Primary site of neoplasm (C40.56)
958	Primary site	Primary site of neoplasm (C40.57)
959	Primary site	Primary site of neoplasm (C40.58)
960	Primary site	Primary site of neoplasm (C40.59)
961	Primary site	Primary site of neoplasm (C40.60)
962	Primary site	Primary site of neoplasm (C40.61)
963	Primary site	Primary site of neoplasm (C40.62)
964	Primary site	Primary site of neoplasm (C40.63)
965	Primary site	Primary site of neoplasm (C40.64)
966	Primary site	Primary site of neoplasm (C40.65)
967	Primary site	Primary site of neoplasm (C40.66)
968	Primary site	Primary site of neoplasm (C40.67)
969	Primary site	Primary site of neoplasm (C40.68)
970	Primary site	Primary site of neoplasm (C40.69)
971	Primary site	Primary site of neoplasm (C40.70)
972	Primary site	Primary site of neoplasm (C40.71)
973	Primary site	Primary site of neoplasm (C40.72)
974	Primary site	Primary site of neoplasm (C40.73)
975	Primary site	Primary site of neoplasm (C40.74)
976	Primary site	Primary site of neoplasm (C40.75)
977	Primary site	Primary site of neoplasm (C40.76)
978	Primary site	Primary site of neoplasm (C40.77)
979	Primary site	Primary site of neoplasm (C40.78)
980	Primary site	Primary site of neoplasm (C40.79)
981	Primary site	Primary site of neoplasm (C40.80)
982	Primary site	Primary site of neoplasm (C40.81)
983	Primary site	Primary site of neoplasm (C40.82)
984	Primary site	Primary site of neoplasm (C40.83)
985	Primary site	Primary site of neoplasm (C40.84)
986	Primary site	Primary site of neoplasm (C40.85)
987	Primary site	Primary site of neoplasm (C40.86)
988	Primary site	Primary site of neoplasm (C40.87)
989	Primary site	Primary site of neoplasm (C40.88)
990	Primary site	Primary site of neoplasm (C40.89)
991	Primary site	Primary site of neoplasm (C40.90)
992	Primary site	Primary site of neoplasm (C40.91)
993	Primary site	Primary site of neoplasm (C40.92)
994	Primary site	Primary site of neoplasm (C40.93)
995	Primary site	Primary site of neoplasm (C40.94)
996	Primary site	Primary site of neoplasm (C40.95)
997	Primary site	Primary site of neoplasm (C40.96)
998	Primary site	Primary site of neoplasm (C40.97)
999	Primary site	Primary site of neoplasm (C40.98)
1000	Primary site	Primary site of neoplasm (C40.99)

## We substantially improved how oncology is represented in OMOP.

### Mapping Corrections by Source → Target (n = 419)

- NAACCR → Cancer Modifier (141)
- LOINC → Cancer Modifier (82)
- SNOMED → Cancer Modifier (68)
- SNOMED → SNOMED (42) (12 + 30 TBD)
- SNOMED → ICDO3 (3)
- ICDO3 → Cancer Modifier (39)
- ICDO3 → SNOMED (25)
- ICDO3 → OMOP Genomic (15)
- ICDO3 → ICDO3 (3)
- Cancer Modifier → Cancer Modifier (1)



Take a picture to download the full paper



Table 2. Selected examples of metastasis concept mapping

Source	Target	Mapping
801	Metastatic disease	Metastatic neoplasm (C40.00)
802	Primary site	Primary site of neoplasm (C40.01)
803	Primary site	Primary site of neoplasm (C40.02)
804	Primary site	Primary site of neoplasm (C40.03)
805	Primary site	Primary site of neoplasm (C40.04)
806	Primary site	Primary site of neoplasm (C40.05)
807	Primary site	Primary site of neoplasm (C40.06)
808	Primary site	Primary site of neoplasm (C40.07)
809	Primary site	Primary site of neoplasm (C40.08)
810	Primary site	Primary site of neoplasm (C40.09)
811	Primary site	Primary site of neoplasm (C40.10)
812	Primary site	Primary site of neoplasm (C40.11)
813	Primary site	Primary site of neoplasm (C40.12)
814	Primary site	Primary site of neoplasm (C40.13)
815	Primary site	Primary site of neoplasm (C40.14)
816	Primary site	Primary site of neoplasm (C40.15)
817	Primary site	Primary site of neoplasm (C40.16)
818	Primary site	Primary site of neoplasm (C40.17)
819	Primary site	Primary site of neoplasm (C40.18)
820	Primary site	Primary site of neoplasm (C40.19)
821	Primary site	Primary site of neoplasm (C40.20)
822	Primary site	Primary site of neoplasm (C40.21)
823	Primary site	Primary site of neoplasm (C40.22)
824	Primary site	Primary site of neoplasm (C40.23)
825	Primary site	Primary site of neoplasm (C40.24)
826	Primary site	Primary site of neoplasm (C40.25)
827	Primary site	Primary site of neoplasm (C40.26)
828	Primary site	Primary site of neoplasm (C40.27)
829	Primary site	Primary site of neoplasm (C40.28)
830	Primary site	Primary site of neoplasm (C40.29)
831	Primary site	Primary site of neoplasm (C40.30)
832	Primary site	Primary site of neoplasm (C40.31)
833	Primary site	Primary site of neoplasm (C40.32)
834	Primary site	Primary site of neoplasm (C40.33)
835	Primary site	Primary site of neoplasm (C40.34)
836	Primary site	Primary site



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

### 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.<sup>1,2,3,4,5</sup> Silico studies also heavily rely on RWD, potentially leading to hypotheses to be tested in further studies.<sup>6</sup> However, multiple proprietary data models across healthcare institutions make these data challenging to obtain.<sup>7,8,9</sup>

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.<sup>10</sup> 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  $\leq 10$  occurrences were rolled up to their nearest parent concept with  $\geq 10$  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.<sup>11</sup> 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  $\geq 20$  instances flagged for manual review.
  - **Devices:** Aggregated by parent concept, calculated per patient

### REFERENCES

1. Coates AJ, Gargan-Craig J. Real-world evidence - where are we now? *N Engl J Med*. 2022 May 18;386(20):1965-6.
2. Food and Drug Administration. Framework for FDA's Real-World Evidence Program [Internet]. Silver Spring (MD): US Food and Drug Administration; 2022 [cited 2023 Jun 26]. Available from: <https://www.fda.gov/oc/real-world-evidence-program>
3. Kucharski AJ, Simonsen L, Funk DJ. Real-world data and evidence in the development and validation of health care decision-making. *PLoS Med*. 2018;15(12):e1002604.
4. Ousey B, Patel A, Heaton A, Heaton A, Heaton A. Use of real-world evidence to drive drug development strategy and inform clinical trial design. *Clin Pharmacol Ther*. 2022;113(1):17-26.
5. Chen A, Kirk A. *Atlas-based Evidence Synthesis: Identifying Discrepancies and Assessing Data Quality for Evidence Synthesis*. *Clin Pharmacol Ther*. 2023;113(1):17-26.
6. Hinkle N, Choi BA, Chinnam K, Heavner SF, Ousey B. *Atlas-based Evidence Synthesis: Identifying Discrepancies and Assessing Data Quality for Evidence Synthesis*. *Clin Pharmacol Ther*. 2023;113(1):17-26.

### 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	Prevalence Count	% Patients	Avg Entries Per Patient	Max Entries Per Patient
Condition	Hypertension Disorder	376,866	6,763	76	NA	NA
Condition	Chronic disease of the cardiovascular system	4,828,244	5,887	67	NA	NA
Condition	Chronic disease of the respiratory system	4,263,261	5,283	59	NA	NA
Condition	Diabetes Mellitus	201,829	4,768	54	NA	NA
Condition	Chronic Obstructive Lung Disease	255,373	4,654	52	NA	NA
Drug	Atorvastatin	1,125,375	8,240	93	21	535
Drug	Ondansetron	1,000,960	7,939	90	3	289
Drug	Sodium	18,116,648	7,658	86	4	743
Drug	Propranolol	18,048,105	7,260	82	17	174
Device/Procedure	Ventilator	40,761,109	4,860	55	276	8,480
Device/Procedure	Oxygenator	4,108,316	4,872	55	410	13,863

Table 2. Top 5 Most Prevalent Measurements

Concept Name	Concept ID	Units	Median	Units %	Units %	Interpretation
Diastolic blood pressure	3013886	mm	277	985	14.14	For half of the patients, there were at least 277 mmHg. Most patients (20%-70% percentile) had 100-200 mmHg. 5% of patients had $\geq 150$ at more records.
Heart rate	3027010	beats	130	476	17.04	For half of the patients, there were at least 130 beats. Most patients (20%-70% percentile) had 50-100 beats. 5% of patients had $\geq 170$ at more records.
Change Core Temperature	3050952	deg	0.6	130	348	For half of the patients, there were at least 0.6 degrees. Most patients (20%-70% percentile) had 0.2-0.6 degrees. 5% of patients had $\geq 1.0$ at more records.
Inhaled oxygen flow rate	3050929	l	50	146	188	For half of the patients, there were at least 50 l. Most patients (20%-70% percentile) had 10-30 l. 5% of patients had $\geq 100$ at more records.
Oxidation and Corrosion Potential	3042020	mmHg	10	30	76	For half of the patients, there were at least 10 mmHg. Most patients (20%-70% percentile) had 0-10 mmHg. 5% of patients had $\geq 20$ at 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 stratified 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

10. Minkov A, Shallicka M, Kharitonov S, Kharitonov S, Kharitonov S, et al. Guiding principles for the conduct of observational critical care research for continuous disease 2019 guidelines and beyond. The Society of Critical Care Medicine Discovery Unit Structure and Regulatory Issues Unmet Study. *Open Access J Clin Res*. 2022;6(1):1-10.
11. Heavner SF, Ousey B, Simonsen L, Funk DJ, et al. Atlas-based Evidence Synthesis: Identifying Discrepancies and Assessing Data Quality for Evidence Synthesis. *Clin Pharmacol Ther*. 2023;113(1):17-26.
12. Williams BA. Connecting epidemiologic cohorts from electronic health record data. *J Am Stat Assoc*. 2021;116(542):1515-1525.
13. Observational Health Data Science and Informatics (OHDSI). Atlas-based Evidence Synthesis: Identifying Discrepancies and Assessing Data Quality for Evidence Synthesis [Internet]. Available from: <https://www.ohdsi.org/atlases/evidence-synthesis/> [cited 2023 Jun 26].
14. Heavner SF, Ousey B, Simonsen L, Funk DJ, et al. Atlas-based Evidence Synthesis: Identifying Discrepancies and Assessing Data Quality for Evidence Synthesis. *Clin Pharmacol Ther*. 2023;113(1):17-26.





# 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



## OHDSI Dentistry Workgroup

*Lead: Robert Koski*

*To understand how dentistry can leverage observational research to improve oral health outcomes and further investigate the links between oral health and systemic disease.*



Robert Koski

Robert Koski



**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](https://ohdsi.org/community-calls-2026)**



**Find your workgroup.**

**Fuel our mission.**

**[ohdsi.org/workgroups](https://ohdsi.org/workgroups)**