Phenotype Aphril Kickoff

OHDSI Community Call
March 31, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Upcoming Community Calls

Date	Topic
Mar. 31	Kickoff to Phenotype April
Apr. 7	Phenotype April, Week 1: Live Build of Cohorts
Apr. 14	Phenotype April, Week 2: KEEPER Evaluation Session
Apr. 21	NO MEETING / EUROPE SYMPOSIUM
Apr. 28	Phenotype April, Week 4: Final Evaluation and Learnings
May 5	Europe Symposium Review/Phenotype April Finale

Three Stages of The Journey

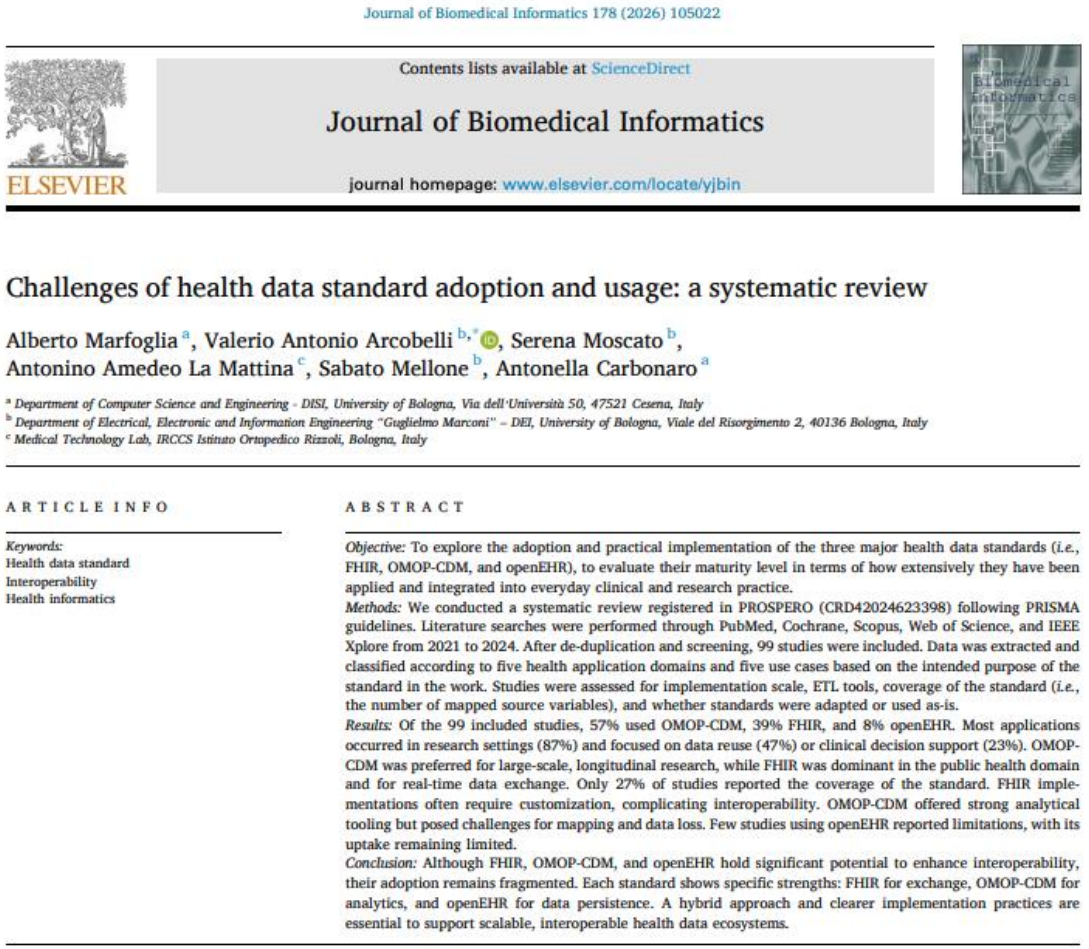
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Alberto Marfoglia, Valerio Antonio Arcobelli, Serena Moscato, Antonino Amedeo La Mattina, Sabato Mellone, and Antonella Carbonaro** on the recent publication of **Challenges of health data standard adoption and usage: a systematic review** in the *Journal of Biomedical Informatics*.



Journal of Biomedical Informatics 178 (2026) 105022

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Challenges of health data standard adoption and usage: a systematic review

Alberto Marfoglia [a], Valerio Antonio Arcobelli [b,*], Serena Moscato [b], Antonino Amedeo La Mattina [c], Sabato Mellone [b], Antonella Carbonaro [a]

[a] Department of Computer Science and Engineering - DISI, University of Bologna, Via dell'Università 50, 47521 Cesena, Italy
[b] Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" – DEI, University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy
[c] Medical Technology Lab, IRCCS Istituto Ortopedico Rizzoli, Bologna, Italy

ARTICLE INFO

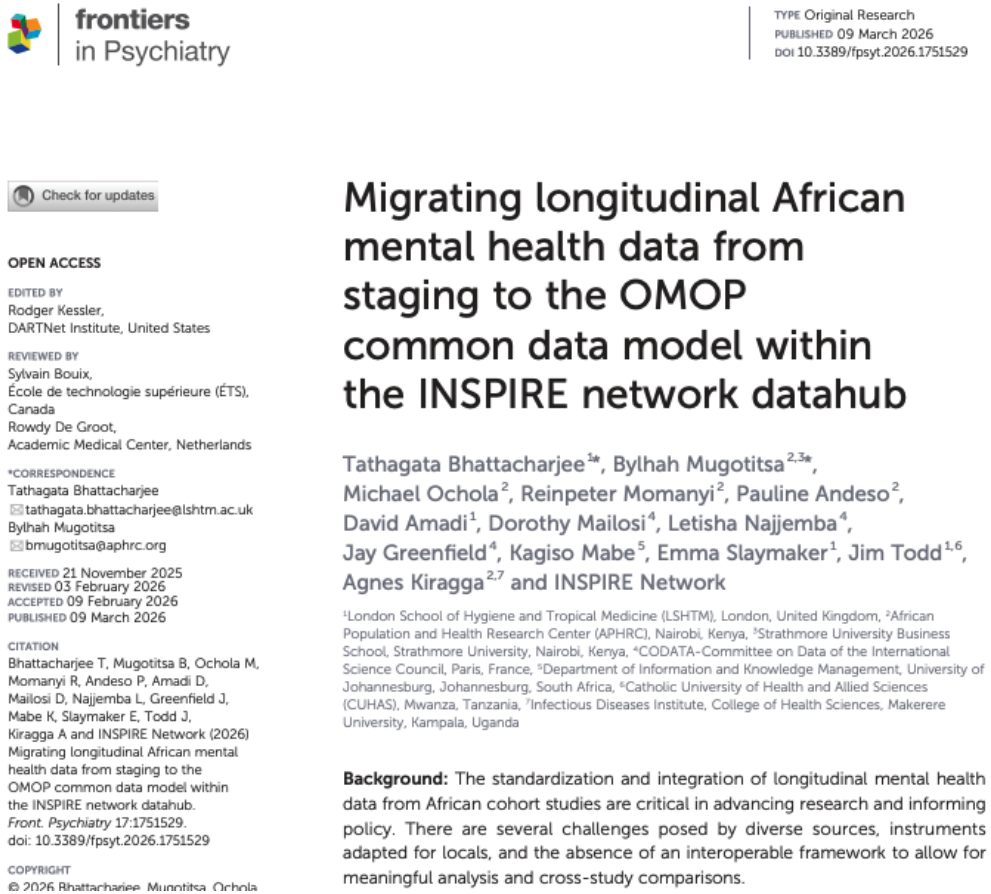
Keywords:
Health data standard
Interoperability
Health informatics

ABSTRACT

Objective: To explore the adoption and practical implementation of the three major health data standards (*i.e.*, FHIR, OMOP-CDM, and openEHR), to evaluate their maturity level in terms of how extensively they have been applied and integrated into everyday clinical and research practice.
Methods: We conducted a systematic review registered in PROSPERO (CRD42024623398) following PRISMA guidelines. Literature searches were performed through PubMed, Cochrane, Scopus, Web of Science, and IEEE Xplore from 2021 to 2024. After de-duplication and screening, 99 studies were included. Data was extracted and classified according to five health application domains and five use cases based on the intended purpose of the standard in the work. Studies were assessed for implementation scale, ETL tools, coverage of the standard (*i.e.*, the number of mapped source variables), and whether standards were adapted or used as-is.
Results: Of the 99 included studies, 57% used OMOP-CDM, 39% FHIR, and 8% openEHR. Most applications occurred in research settings (87%) and focused on data reuse (47%) or clinical decision support (23%). OMOP-CDM was preferred for large-scale, longitudinal research, while FHIR was dominant in the public health domain and for real-time data exchange. Only 27% of studies reported the coverage of the standard. FHIR implementations often require customization, complicating interoperability. OMOP-CDM offered strong analytical tooling but posed challenges for mapping and data loss. Few studies using openEHR reported limitations, with its uptake remaining limited.
Conclusion: Although FHIR, OMOP-CDM, and openEHR hold significant potential to enhance interoperability, their adoption remains fragmented. Each standard shows specific strengths: FHIR for exchange, OMOP-CDM for analytics, and openEHR for data persistence. A hybrid approach and clearer implementation practices are essential to support scalable, interoperable health data ecosystems.

OHDSI Shoutouts!

Congratulations to the team of **Tathagata Bhattacharjee, Bylhah Mugotitsa, Michael Ochola, Reinpeter Momanyi, Pauline Andeso, David Amadi, Dorothy Mailosi, Letisha Najjemba, Jay Greenfield, Kagiso Mabe, Emma Slaymaker, Jim Todd, Agnes Kiragga, and the INSPIRE Network** on the recent publication of **Migrating longitudinal African mental health data from staging to the OMOP common data model within the INSPIRE network datahub** in *Frontiers in Psychiatry*.



frontiers in Psychiatry

TYPE Original Research
PUBLISHED 09 March 2026
DOI 10.3389/fpsyt.2026.1751529

Check for updates

OPEN ACCESS

EDITED BY
Rodger Kessler,
DARTNet Institute, United States

REVIEWED BY
Sylvain Bouix,
École de technologie supérieure (ÉTS), Canada
Rowdy De Groot,
Academic Medical Center, Netherlands

*CORRESPONDENCE
Tathagata Bhattacharjee
tathagata.bhattacharjee@lshtm.ac.uk
Bylhah Mugotitsa
bmugotitsa@aphrc.org

RECEIVED 21 November 2025
REVISED 03 February 2026
ACCEPTED 09 February 2026
PUBLISHED 09 March 2026

CITATION
Bhattacharjee T, Mugotitsa B, Ochola M, Momanyi R, Andeso P, Amadi D, Mailosi D, Najjemba L, Greenfield J, Mabe K, Slaymaker E, Todd J, Kiragga A and INSPIRE Network (2026) Migrating longitudinal African mental health data from staging to the OMOP common data model within the INSPIRE network datahub. *Front. Psychiatry* 17:1751529. doi: 10.3389/fpsyt.2026.1751529

Migrating longitudinal African mental health data from staging to the OMOP common data model within the INSPIRE network datahub

Tathagata Bhattacharjee[1]*, Bylhah Mugotitsa[2,3]*, Michael Ochola[2], Reinpeter Momanyi[2], Pauline Andeso[2], David Amadi[1], Dorothy Mailosi[4], Letisha Najjemba[4], Jay Greenfield[4], Kagiso Mabe[5], Emma Slaymaker[1], Jim Todd[1,6], Agnes Kiragga[2,7] and INSPIRE Network

[1]London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom, [2]African Population and Health Research Center (APHRC), Nairobi, Kenya, [3]Strathmore University Business School, Strathmore University, Nairobi, Kenya, [4]CODATA-Committee on Data of the International Science Council, Paris, France, [5]Department of Information and Knowledge Management, University of Johannesburg, Johannesburg, South Africa, [6]Catholic University of Health and Allied Sciences (CUHAS), Mwanza, Tanzania, [7]Infectious Diseases Institute, College of Health Sciences, Makerere University, Kampala, Uganda

Background: The standardization and integration of longitudinal mental health data from African cohort studies are critical in advancing research and informing policy. There are several challenges posed by diverse sources, instruments adapted for locals, and the absence of an interoperable framework to allow for meaningful analysis and cross-study comparisons.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Wednesday	8 am	Psychiatry
Wednesday	7 pm	Medical Imaging
Thursday	10 am	ATLAS/WebAPI
Thursday	10 am	GIS – Geographic Information System
Thursday	11 am	Industry
Thursday	11 am	Themis
Thursday	12 pm	Methods Research
Thursday	1 pm	Oncology Vocabulary/Development Subgroup
Thursday	2 pm	Early-Stage Researchers
Thursday	7 pm	Dentistry
Friday	9 am	Waveform
Friday	10 am	Transplant
Friday	11:30 am	Steering Group
Monday	10 am	Healthcare Systems Interest Group

OHDSI2026 Registration Opens

Registration and the **call for participation** are now open for the **2026 OHDSI Global Symposium**, held Oct. 20-22 at the Hyatt Regency in New Brunswick, N.J.

ohdsi.org/OHDSI2026

2026 Symposium Tutorials – Session 1

- An Introduction to the Journey from Data to Evidence Using OHDSI
- An Introduction to ATLAS
- Bringing FAIR to Imaging Research with the Medical Imaging OMOP Extension
- Complex Phenotyping at Scale with and without LLMs Using PhenotypeR
- OHDSI Leadership Storytelling Workshop
- Mastering OMOP: Transforming EHR Data with Practical Strategies, Best Practices, and OHDSI Integration

2026 Symposium Tutorials – Session 2

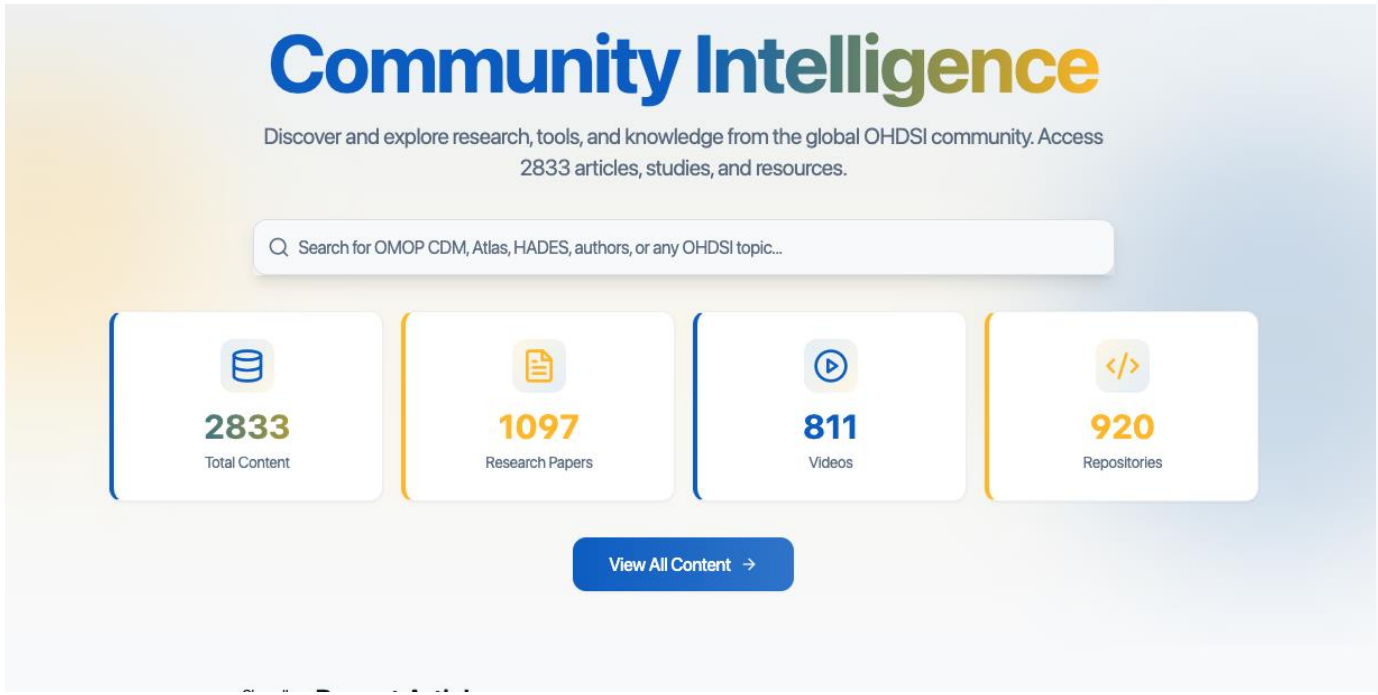
- Building and Using the OHDSI Evidence Network: From Data Partner to Federated Study Execution
- From Multi-Modal Data to Real-World Evidence: Hands-on with the Data2Evidence Platform for OMOP Data Curation and Analytics
- Integrating Geospatial Data Into OMOP CDM
- Introduction to OHDSI Phenotype Development & Evaluation
- OHDSI Standardized Vocabularies on FHIR: A Deep Dive Using the Echidna Terminology Server
- Using OMOP Model in Registry Context & Clinical Trials Standardization Context: Conventions, Past Use Cases, SDTM & Regulatory Consideration, Challenges

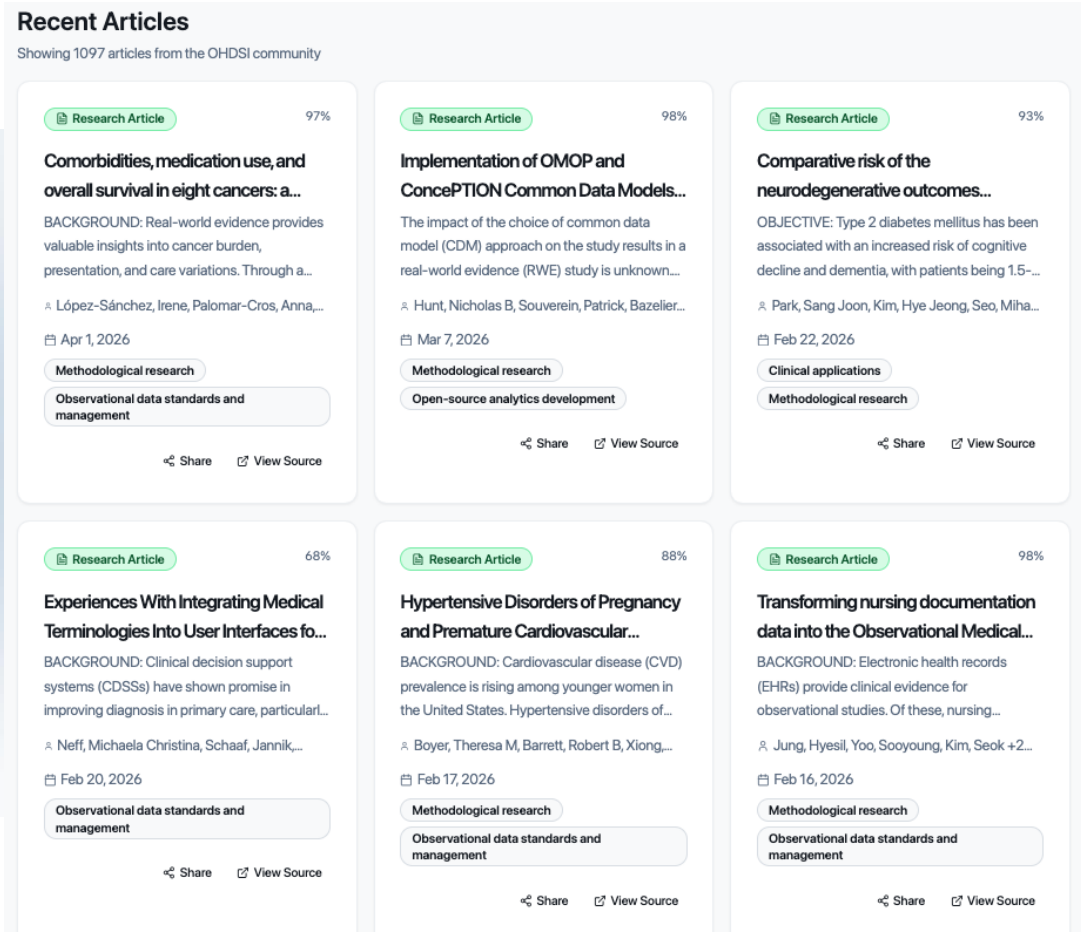
Maternal Fellowship Opens

The second **OHDSI Maternal Health Fellowship** is designed to train clinical investigators for improved maternal and neonatal care. This fellowship offers three key components: **Career Development, Practice, and Networking**.

Supported by both the OHDSI community and the NIH IMPROVE initiative, the program focuses on training clinical investigators in observational research methods to enable them to conduct reproducible research and generate real-world evidence.

Community Dashboard





dash.ohdsi.org

OMOP School in Stockholm, Sweden

passion2improve edence Health

The OMOP School

3+1-day OMOP CDM Bootcamp

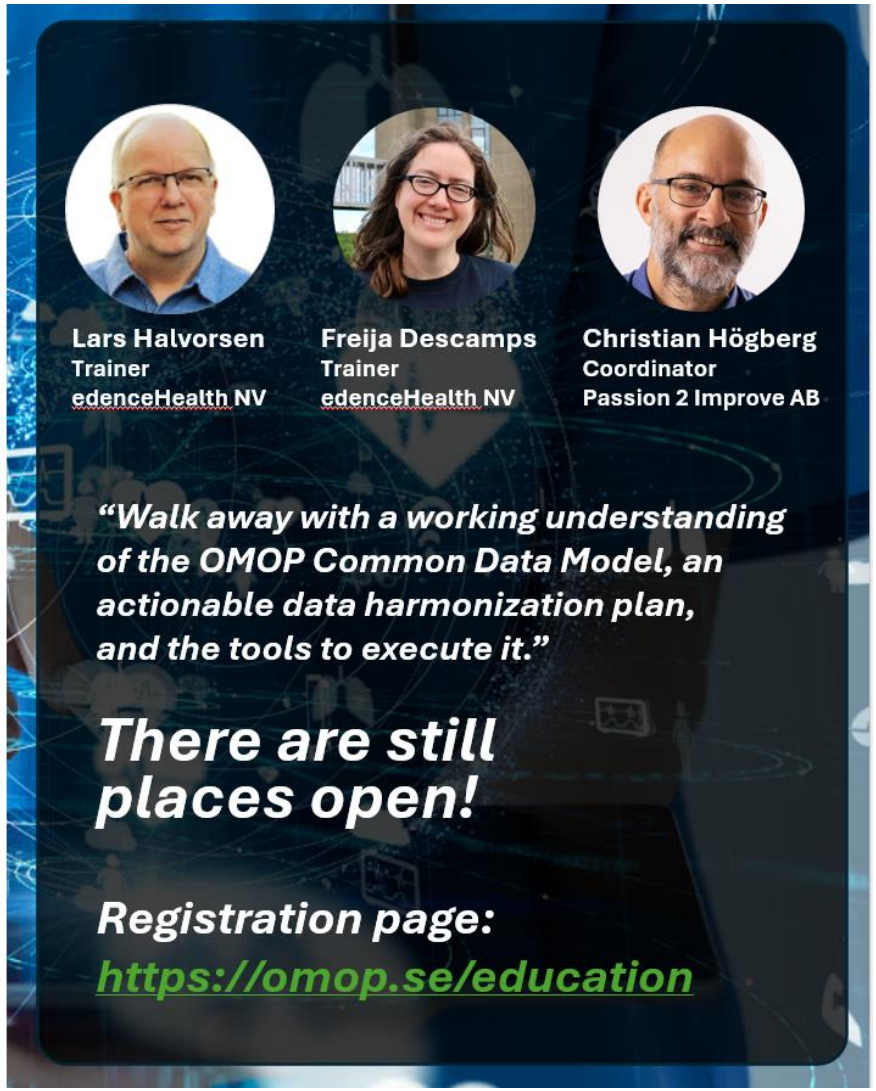
A hands-on training and workshop that turns your data harmonization vision into reality.

May 26th – 28th
+ May 29th (optional extra day for Use Cases Deep Dive)
09:00-17:00 Tue-Thu, 09:00-16:00 Fri
Stockholm, Sweden (venue TBD)

Learning objectives:
- Explain the role of standardization in federated research
- Understand the OMOP Common Data Model and how it can be applied
- Perform semantic mapping using OMOP vocabularies
- Design and implement an OMOP ETL process
- Evaluate and improve data quality
- Conduct standardized observational analyses
- Execute an end-to-end mini-harmonization project
- Use selective OHDSI methods and tools

Who to attend?
- Health Data Owners/Data Custodians
- Data Scientists, Clinical Researchers & Epidemiologists
- IT/Data Architects/ETL Developers/Health Informatics Specialists
- Digital Transformation Leads, Registry Directors, Healthcare Strategists
- Healthcare Policy Stakeholders

OMOP4SWEDEN!
EVIDENCE GENERATION FOR BETTER HEALTHCARE



UK Symposium Call for Abstracts Opens

HDR UK Event

OHDSI UK 2026

We're delighted to announce that OHDSI UK 2026 will be held on the 18th of September at the University of Nottingham. For the first time, there will also be an OMOP training day on the 17th of September.

Share this page

OHDSI (Observational Health Data Sciences and Informatics, pronounced "Odyssey") is an international community of stakeholders dedicated to unlocking the value of health data through large-scale analytics. OHDSI promotes open science and collaboration in health data research with a key focus on adoption of the OMOP Common Data Model, a global standard for harmonising data and facilitating federated analytics across institutions. **Find out more about OHDSI.**

Call for Abstracts

We invite you to submit an abstract for consideration at OHDSI UK 2026. Whether you wish to present a poster, software demo, or lightning talk, we welcome contributions from across the community. Abstract submission is available via this form, and the deadline is 1st May 2026. Please use this template to prepare your abstract and save it as a PDF, and start your file name with the surname (family name) of the presenting author.

Key dates:

Registration Opens: 20th April 2026

Registration Closes: 4th September 2026

Abstract Submission Opens: 20th March 2026

Abstract Submission Deadline: 1st May 2026

Training day: 17th September 2026

Symposium: 18th September 2026

First Latin America Symposium – July 30-31

2026 Europe Symposium

The 2026 OHDSI Europe Symposium returns to Rotterdam next year and will be held **April 18-20.**

Registration is open on the **OHDSI & OHDSI Europe** web sites.

Time	Symposium Agenda - Monday April 20, 2026	Location
8:00	***Registration and Coffee***	Queen's Lounge
9:00	**Welcome to OHDSI Europe** Dr. Renske Los, Department of Medical Informatics, *Erasmus MC* Dr. Aniek Markus, Department of Medical Informatics, *Erasmus MC*	Theatre
9:05	**Journey of OHDSI** Prof. Peter Rijnbeek, Chair Department of Medical Informatics, *Erasmus MC*	Theatre
9:30	**Collaborator Showcase - part 1** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC*	Theatre
10:00	**Speed networking**	Theatre
10:15	***Coffee Break & posters National Nodes***	Queen's Lounge
11:15	**Collaborator Showcase - part 2** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC*	Theatre
11:45	**Dreaming about the OHDSI journey ahead** Dr. Patrick Ryan, Vice President, Observational Health Data Analytics, *Johnson & Johnson* Dr. Renske Los, Department of Medical Informatics, *Erasmus MC*	Theatre
12:15	***Lunch break & networking & posters/demo's*** *(Early investigator meeting - 13:00-13:45 Queen's Lounge)*	La Fontaine & Odyssee Room
13:45	**From dreams to reality** OHDSI Titan Award winners	Theatre
14:30	**Propositions for collaboration from the National Nodes** National Node leads	Theatre
14:45	***Coffee break & posters/demo's***	La Fontaine & Odyssee Room
16:15	**The OH Factor** To be announced	Theatre
17:00	**Closing**	Theatre
17:15	***Networking reception***	Queen's Lounge

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.

#OHDSISocialShowcase This Week

Monday

Powering a Personal Health Record Analytic Environment using the OHDSI CDM and Google Colab

(Janos Hajagos)

Powering a Personal Health Record Analytic Environment using the OHDSI CDM and Google Colab

Motivation:

Provide analytic access to your own medical record

Very few people have direct access to their complete medical data and if they do it is in a range of different formats, such as, PDFs and CSV files. Individuals move and see different providers so a single analytic record is likely to be incomplete.

Combine emerging sensor data with EHR data

Mobile health data devices have matured significantly over the last decade from initially counting steps to now providing detailed biometric data including: heart rate, breathing rate, blood oxygen level, and continuous glucose monitoring. This data is not normally used clinically as the volume of data is large and measurements are not clinically validated. By combining EHR data with mobile health data into a standardize form we can begin the process of scientifically evaluating sensor data.

Standardizing to a CDM (n > 1)

Individuals with complex medical histories have disjointed care records. By building their own PHR an individual can document their care and understand how their physiology is changing during their life path.

By standardizing to a single data model this would allow the aggregation of multiple PHRs. Individuals could elect to donate a de-identified version of their PHR record to support research.

Learning the OHDSI CDM

The OHDSI CDM is the de facto standard for research data warehouses based on EHR (Electronic Health Record) data. Understanding how your own data is transformed and mapped to OHDSI CDM allows learners to understand the data transformation process. By writing queries against your own data you can understand the complexities of working with EHR data.

Solution:

We map commonly downloadable XML files from patient portals and mobile health devices to the OHDSI CDM (v.5.4). The tables are stored as Apache Parquet files. We use PySpark in the Google Colab environment to support rich visualization and data analyses.

What is an XML C-CDA Document?

C-CDAs XML are a mandated format that all major EHR vendors in the United States support for Meaningful Use. They are widely generated to support health data exchanges. C-CDA documents can be downloaded by patients through a patient portal. CDAs are in an XML format and contain standardized data elements which can easily be mapped to the OHDSI CDM.

What about FHIR?

FHIR offers a REST API for improved access to EHR data. The same approach here could be done with FHIR resources. Currently patient portals and C-CDAs are the most universal way to access and download your own medical record in a structured format.

Building on existing work

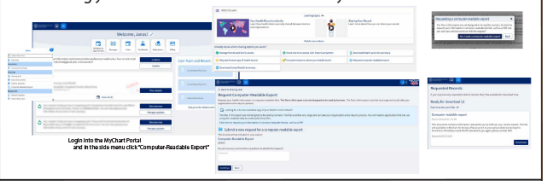
We have developed a data model called Prepared Source Format (PSF) which is easier to target for ETL writers. We have previously implemented a mapping script from PSF to the OHDSI CDM. In addition, we utilize the Spark computation environment which provides powerful tools to work with small and very large data sets. The same PSF to OHDSI mapping script which does the conversion is currently being used to convert an entire Health System's EHR data to the OHDSI CDM. Google Colab supports PySpark allowing for little or no cost to run Spark data pipelines. This work extends this pipeline by writing a program to extract clinical mappable elements from C-CDA XML documents.

Alternative to Google Colab

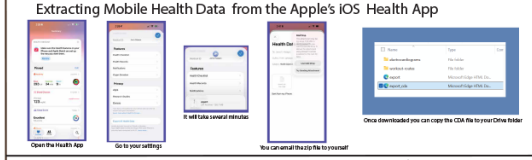
The pipeline used to extract elements from C-CDAs can be run on any recent Python (>3.8). The mapping script to OHDSI can be run in a Docker container (source code available on GitHub) and on a Spark Cluster.

DIY (Do It Yourself) Guide

Obtaining your electronic medical record from MyChart

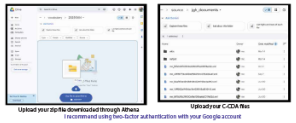


Extracting Mobile Health Data from the Apple's iOS Health App



Getting Concept Files

Staging in Google Drive

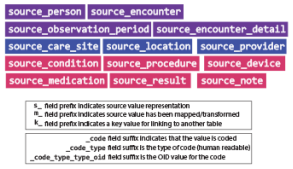


Configuring Colab

Executing the Pipeline

Querying with PySpark

PSF Format (Intermediary Format)



SQL Template

We map data elements extracted from C-CDAs to PSF

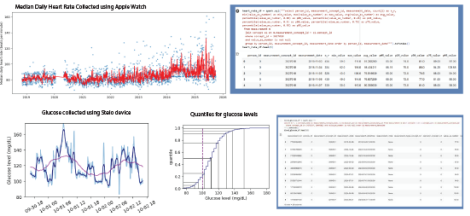
Mapping results:

Source table and OHDSI CDM table row counts

CDA File Name	Vendor	Measurements #	Measurements # concepts	Observations #	Observations # concepts	Drug Exposures #	Drug Exposures # concepts	Condition Occurrences #	Condition Occurrences # concepts	Notes #	Notes # concepts	Start	End
CCDA_2_2_20241016.xml	Allscripts					4	3			5	1	11/16/2017	9/24/2024
export_cda.xml	Apple	823,158	7									3/28/2016	6/30/2025
DOC0002.XML	Epic	3	3	3	2	1	1					9/8/2020	9/24/2024
ccda.xml	Flatiron	441	139			8	5	12	4			9/15/2022	9/18/2024
summary-32282035.xml	Oracle/Cerner					1	1					2/28/2015	3/1/2015
summary-03092035.xml	Oracle/Cerner	25	24									3/9/2015	3/9/2015
summary-04262017.xml	Oracle/Cerner									1	1	4/24/2017	4/24/2017
summary-05052017.xml	Oracle/Cerner									1	1	5/5/2017	5/5/2017

At the time of abstract submission (June 2025) a total of 8 CDA XMLs files from 5 different vendors were converted to the OHDSI CDM covering over 10 years of personal health data.

Analytics in Colab (Jupyter Notebook Environment)



Analyze extracted text using LLMs

	n_r	n	note_class_concept_id	concept_name
0	21	1	3008735	Progress note
1	17	1	3085855	Laboratory report
2	15	1	3085854	Diagnostic imaging study
3	5	1	40771137	Hematology+Medical oncology Progress note
4	3	1	3085122	Pathology study
5	2	1	40771156	Ophthalmology Progress note
6	1	1	4481709	Urology Progress note

C-CDAs can contain either full text notes and embedded documents such as PDFs. We use PyPDF library to extract text from the PDFs. This texts is then mapped to the note table in the OHDS CDM. LLM (Large Language Models) like Gemini can be used summarize the information in the notes.

Conclusions

This work demonstrates that a single person's health data obtained manually from multiple sources can be converted to the OHDSI CDM and analyzed in a single database. This preliminary conversion covered multiple EHR sources and patient captured data devices. The current OHDSI vocabulary does not include an OHDSI type for data collected by the patient. It is important to distinguish data collected in routine patient care and those collected using personal health devices as they have different sampling frequencies and quality. If the OHDSI CDM will be used for self-collected patient device data it will require significant database scaling due to the large volume of data a single person can generate.

Future directions

There are subject areas in the C-CDA documents that are not currently being extracted. These elements can be extracted to provide a richer analytic data set and include: immunizations, visits, and the social history section.

Janos G. Hajagos, Ph.D.
Department of Biomedical Informatics
Stony Brook University

https://github.com/jhajagos/PHR2OHDSI/

Presented at the 2025 OHDSI Symposium
October 8, 2025

#OHDSISocialShowcase This Week

Tuesday

Considerations for De-identification of the OMOP Common Data Model

(Jose Posada, Natasha Flowers, Priya Desai)

Considerations for De-identification of the OMOP-CDM

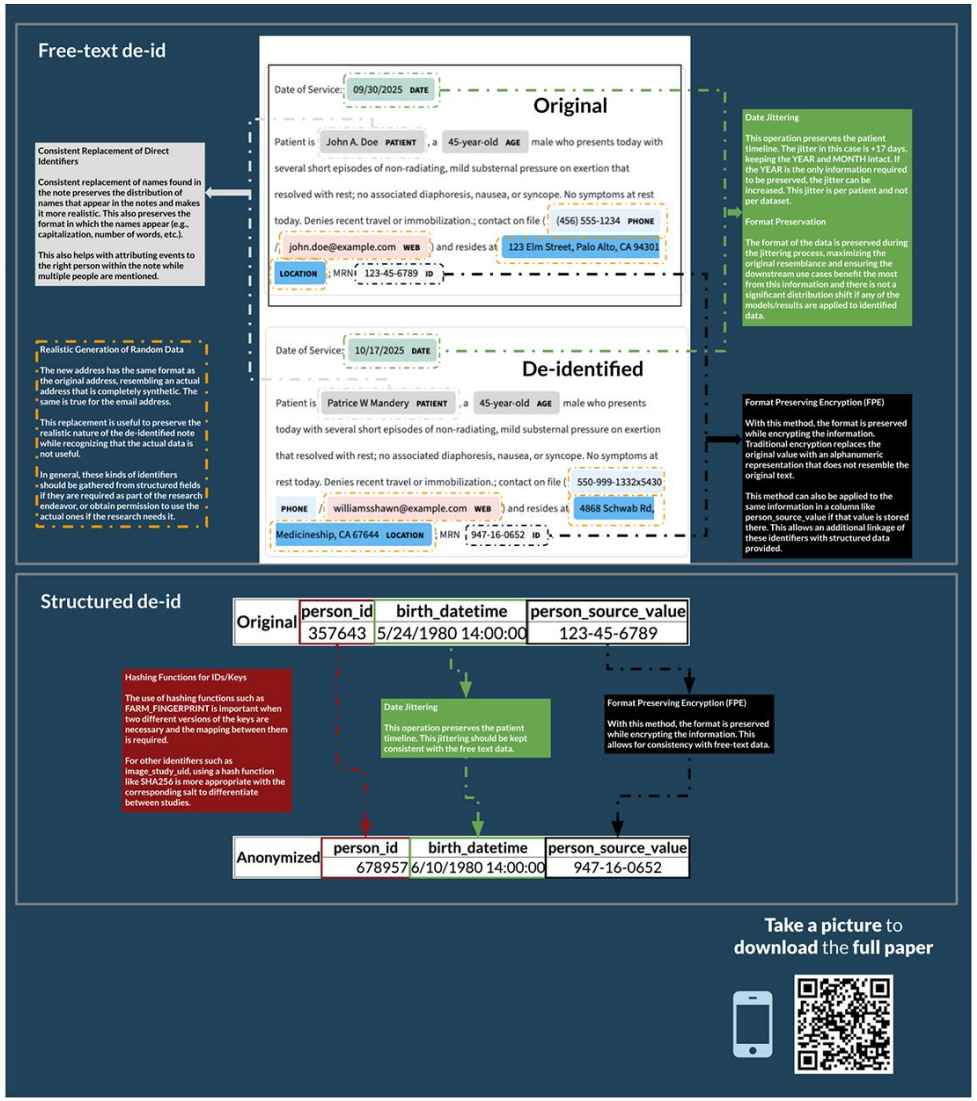
Jose D. Posada, PhD
Natasha Flowers, MSc
Priya Desai, MSc

Technology & Digital Solutions
Stanford Health Care and School of Medicine

The OMOP-CDM contains several types of Personal Identifiable Information (PII) that must be addressed during de-identification.

Direct identifiers like **person_source_value** require random identifier generation or **deterministic encryption** with secure key management, while dates typically need consolidation or patient-level shifting to preserve event timelines.

Source_value fields throughout the OMOP-CDM tables present particular challenges as they may contain free-form text with embedded PII, especially in fully populated implementations where provenance data is preserved for **concept_id** mapping and measurement details.



Recommendations

De-identification
- ***Structured data***
  - Date shifting and grouping
  - Hashing
  - Format Preserving Encryption (FPE)
  - Standard Encryption
- ***Free-text:*** Mostly source values and *note_text*
  - Extensive Allow lists
  - Standard NLP
  - LLM
  - Hiding in Plain sight (HIPS)
    - Format preservation
    - Consistent replacement
    - Concordance with structured data (e.g. date shifting)

Risk Assessment and expert determination
- DUAs and access/use infrastructure
- Enclaves

Stanford MEDICINE

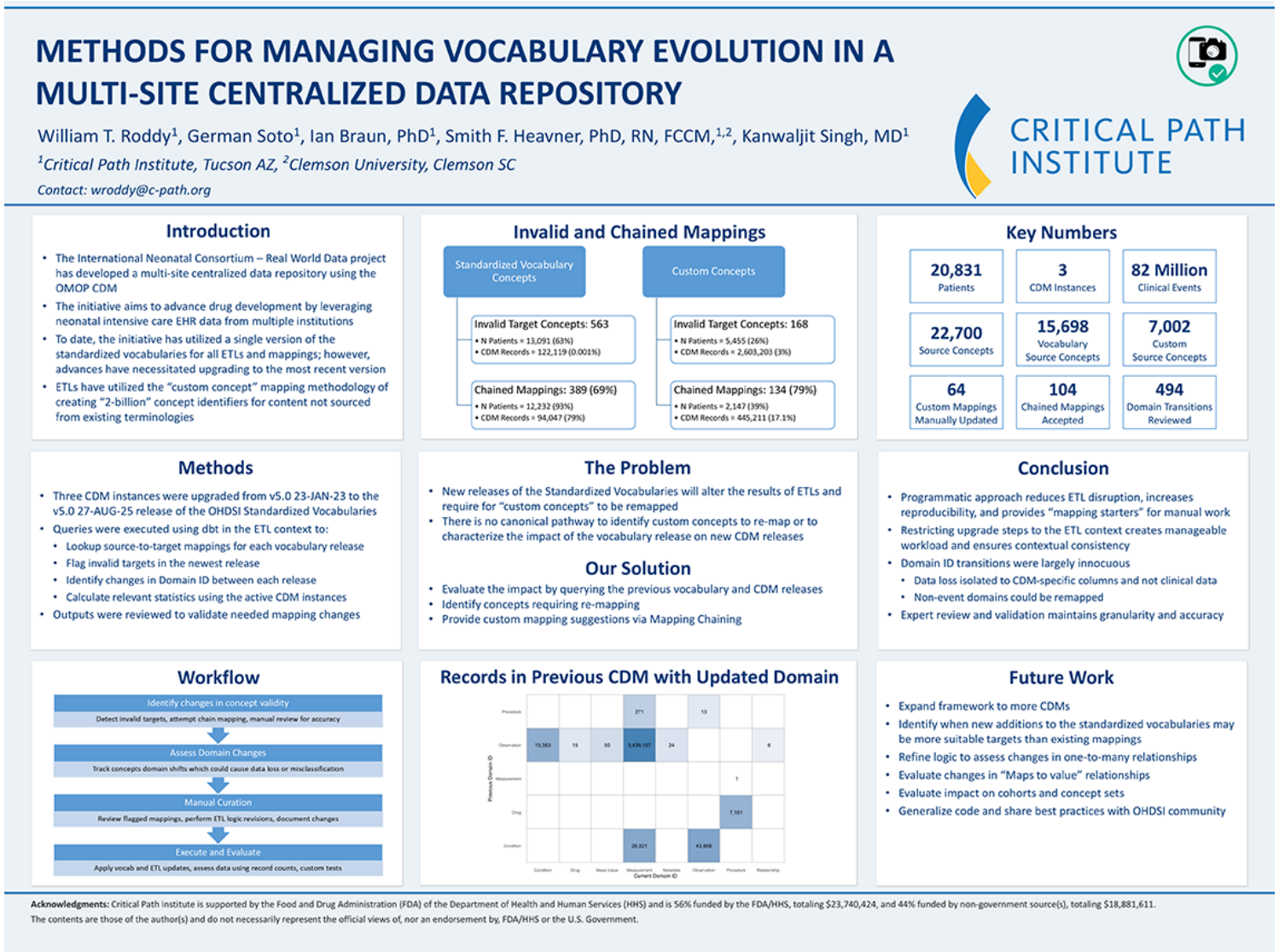
OHDSI

#OHDSISocialShowcase This Week

Wednesday

Methods for Managing Vocabulary Evolution in a Multi-Site Centralized Data Repository

(William T. Roddy, German Soto, Ian Braun, Smith F. Heavner, Kanwaljit Singh)



METHODS FOR MANAGING VOCABULARY EVOLUTION IN A MULTI-SITE CENTRALIZED DATA REPOSITORY

William T. Roddy[1], German Soto[1], Ian Braun, PhD[1], Smith F. Heavner, PhD, RN, FCCM,[1,2], Kanwaljit Singh, MD[1]

[1]Critical Path Institute, Tucson AZ, [2]Clemson University, Clemson SC

Contact: wroddy@c-path.org

CRITICAL PATH INSTITUTE

Introduction

- The International Neonatal Consortium – Real World Data project has developed a multi-site centralized data repository using the OMOP CDM
- The initiative aims to advance drug development by leveraging neonatal intensive care EHR data from multiple institutions
- To date, the initiative has utilized a single version of the standardized vocabularies for all ETLs and mappings; however, advances have necessitated upgrading to the most recent version
- ETLs have utilized the "custom concept" mapping methodology of creating "2-billion" concept identifiers for content not sourced from existing terminologies

Invalid and Chained Mappings

Standardized Vocabulary Concepts	Custom Concepts
Invalid Target Concepts: 563 • N Patients = 13,091 (63%) • CDM Records = 122,119 (0.001%)	Invalid Target Concepts: 168 • N Patients = 5,455 (26%) • CDM Records = 2,603,203 (3%)
Chained Mappings: 389 (69%) • N Patients = 12,232 (93%) • CDM Records = 94,047 (79%)	Chained Mappings: 134 (79%) • N Patients = 2,147 (39%) • CDM Records = 445,211 (17.1%)

Key Numbers

20,831 Patients	3 CDM Instances	82 Million Clinical Events
22,700 Source Concepts	15,698 Vocabulary Source Concepts	7,002 Custom Source Concepts
64 Custom Mappings Manually Updated	104 Chained Mappings Accepted	494 Domain Transitions Reviewed

Methods

- Three CDM instances were upgraded from v5.0 23-JAN-23 to the v5.0 27-AUG-25 release of the OHDSI Standardized Vocabularies
- Queries were executed using dbt in the ETL context to:
  - Lookup source-to-target mappings for each vocabulary release
  - Flag invalid targets in the newest release
  - Identify changes in Domain ID between each release
  - Calculate relevant statistics using the active CDM instances
- Outputs were reviewed to validate needed mapping changes

The Problem

- New releases of the Standardized Vocabularies will alter the results of ETLs and require for "custom concepts" to be remapped
- There is no canonical pathway to identify custom concepts to re-map or to characterize the impact of the vocabulary release on new CDM releases

Our Solution

- Evaluate the impact by querying the previous vocabulary and CDM releases
- Identify concepts requiring re-mapping
- Provide custom mapping suggestions via Mapping Chaining

Conclusion

- Programmatic approach reduces ETL disruption, increases reproducibility, and provides "mapping starters" for manual work
- Restricting upgrade steps to the ETL context creates manageable workload and ensures contextual consistency
- Domain ID transitions were largely innocuous
  - Data loss isolated to CDM-specific columns and not clinical data
  - Non-event domains could be remapped
- Expert review and validation maintains granularity and accuracy

Workflow

Records in Previous CDM with Updated Domain

Future Work

- Expand framework to more CDMs
- Identify when new additions to the standardized vocabularies may be more suitable targets than existing mappings
- Refine logic to assess changes in one-to-many relationships
- Evaluate changes in "Maps to value" relationships
- Evaluate impact on cohorts and concept sets
- Generalize code and share best practices with OHDSI community

Acknowledgments: Critical Path Institute is supported by the Food and Drug Administration (FDA) of the Department of Health and Human Services (HHS) and is 56% funded by the FDA/HHS, totaling $23,740,424, and 44% funded by non-government source(s), totaling $18,881,611.
The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the U.S. Government.

#OHDSISocialShowcase This Week

Thursday

A Framework for Understanding Bias and Real-World Clinical Electronic Health Record Data

(Haeun Lee, Harold Lehmann, Benjamin Martin, Paul Nagy)

A Framework for Understanding Bias and Real-World Clinical Electronic Health Record Data

Haeun Lee

INTRO:

Real-world evidence (RWE) from electronic health records (EHRs) plays a crucial role in healthcare decision-making, offering advantages in speed, cost, and population diversity compared to traditional randomized clinical trials. However, the transformation of raw clinical data into reliable evidence faces numerous biases that remain inadequately addressed. This study aims to identify and categorize bias sources throughout the evidence-generation pipeline.

METHODS

We developed a bias framework through synthesis of established RWE generation pathways and iterative review by a multidisciplinary expert panel of clinicians, informaticians, and biostatisticians. To examine bias in RWD, we conducted a scoping review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) framework, searching PubMed and Web of Science for English publications (2016-2025) using "electronic health records" AND "bias" combined with related terms including "patient", "healthcare", "diagnostic", "coding", "data", "extraction", "standards", "phenotype", "CDM", "ETL", "transformation", and "mapping".

Results

- Analysis of 215 papers identified 209 distinct bias sources across three levels (Healthcare, Data, Management, Research) and seven domains, showing how biases cascade through the clinical-to-research pipeline
- Approximately 20% of all bias sources originates at the access to medical care level (Level1), followed by data acquisition and measurement (Level 3, 18.7%), and data analytics (Level 7, 17.2%)
- Healthcare-level biases (N = 108) were most frequently mentioned in the literature

Conclusion

- This study developed a multi-level framework across healthcare, data management, and research domains to identify and categorize 209 sources of bias in clinical evidence generation
- Biases showed interconnected properties that span multiple levels and amplify throughout the data lifecycle, indicating the need for comprehensive bias evaluation across the entire data journey
- This framework can serve as a trigger to help methodologists make sure they have considered important sources of bias

Next Steps

- Develop a hierarchical clustering taxonomy of bias, stratifying across a 7-level framework to elucidate patterns of systematic error propagation throughout the evidence generation pipeline
- Quantify bias magnitude through emulated trial frameworks, measuring the differential between actual data and estimates in selected clinical scenarios

Haeun Lee, MS, Harold Lehmann, MD, PhD, Benjamin Martin, PhD, Paul Nagy, PhD

JOHNS HOPKINS SCHOOL of MEDICINE

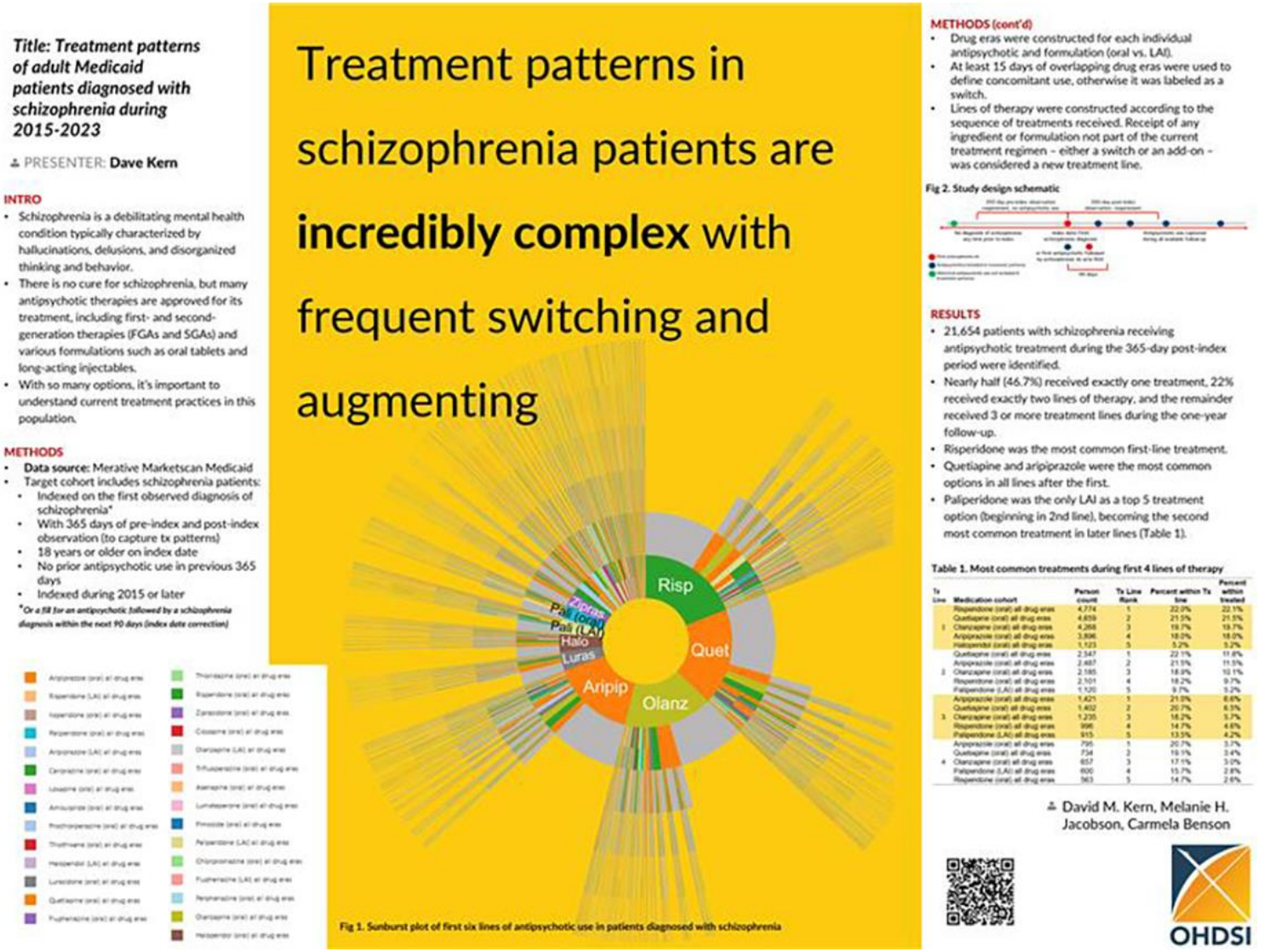
OHDSI

#OHDSISocialShowcase This Week

Friday

Treatment patterns of adult Medicaid patients diagnosed with schizophrenia during 2015-2023

(David M. Kern, Melanie H. Jacobson, Carmela Benson)



Title: Treatment patterns of adult Medicaid patients diagnosed with schizophrenia during 2015-2023

PRESENTER: **Dave Kern**

INTRO

- Schizophrenia is a debilitating mental health condition typically characterized by hallucinations, delusions, and disorganized thinking and behavior.
- There is no cure for schizophrenia, but many antipsychotic therapies are approved for its treatment, including first- and second-generation therapies (FGAs and SGAs) and various formulations such as oral tablets and long-acting injectables.
- With so many options, it's important to understand current treatment practices in this population.

METHODS

- **Data source:** Merative Marketscan Medicaid
- Target cohort includes schizophrenia patients:
  - Indexed on the first observed diagnosis of schizophrenia*
  - With 365 days of pre-index and post-index observation (to capture tx patterns)
  - 18 years or older on index date
  - No prior antipsychotic use in previous 365 days
  - Indexed during 2015 or later

*Or a fill for an antipsychotic followed by a schizophrenia diagnosis within the next 90 days (index date correction)

Treatment patterns in schizophrenia patients are **incredibly complex** with frequent switching and augmenting

Fig 1. Sunburst plot of first six lines of antipsychotic use in patients diagnosed with schizophrenia

METHODS (cont'd)

- Drug eras were constructed for each individual antipsychotic and formulation (oral vs. LAI).
- At least 15 days of overlapping drug eras were used to define concomitant use, otherwise it was labeled as a switch.
- Lines of therapy were constructed according to the sequence of treatments received. Receipt of any ingredient or formulation not part of the current treatment regimen – either a switch or an add-on – was considered a new treatment line.

Fig 2. Study design schematic

RESULTS

- 21,654 patients with schizophrenia receiving antipsychotic treatment during the 365-day post-index period were identified.
- Nearly half (46.7%) received exactly one treatment, 22% received exactly two lines of therapy, and the remainder received 3 or more treatment lines during the one-year follow-up.
- Risperidone was the most common first-line treatment.
- Quetiapine and aripiprazole were the most common options in all lines after the first.
- Paliperidone was the only LAI as a top 5 treatment option (beginning in 2nd line), becoming the second most common treatment in later lines (Table 1).

Table 1. Most common treatments during first 4 lines of therapy

Tx Line	Medication cohort	Person count	Tx Line Rank	Percent within Tx line	Percent within treated
1	Risperidone (oral) all drug eras	4,774	1	22.0%	22.1%
	Quetiapine (oral) all drug eras	4,659	2	21.5%	21.5%
	Olanzapine (oral) all drug eras	4,268	3	19.7%	19.7%
	Aripiprazole (oral) all drug eras	3,896	4	18.0%	18.0%
	Haloperidol (oral) all drug eras	1,123	5	5.2%	5.2%
2	Quetiapine (oral) all drug eras	2,347	1	22.1%	11.8%
	Aripiprazole (oral) all drug eras	2,487	2	21.5%	11.5%
	Olanzapine (oral) all drug eras	2,185	3	18.6%	10.1%
	Risperidone (oral) all drug eras	2,101	4	18.2%	9.7%
	Paliperidone (LAI) all drug eras	1,120	5	9.7%	5.2%
3	Aripiprazole (oral) all drug eras	1,421	1	21.5%	6.6%
	Quetiapine (oral) all drug eras	1,402	2	20.7%	6.5%
	Olanzapine (oral) all drug eras	1,235	3	18.2%	5.7%
	Risperidone (oral) all drug eras	996	4	14.7%	4.6%
	Paliperidone (LAI) all drug eras	915	5	13.5%	4.2%
4	Aripiprazole (oral) all drug eras	795	1	20.7%	3.7%
	Quetiapine (oral) all drug eras	734	2	19.1%	3.4%
	Olanzapine (oral) all drug eras	657	3	17.1%	3.0%
	Paliperidone (LAI) all drug eras	600	4	15.7%	2.8%
	Risperidone (oral) all drug eras	563	5	14.7%	2.6%

David M. Kern, Melanie H. Jacobson, Carmela Benson

OHDSI

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?

OHDSI Canada

Purpose

OHDSI Canada builds a national community to advance the adoption and use of the OMOP Common Data Model and OHDSI tools across Canadian health research institutions. Through knowledge sharing, aligning standards, developing resources, and collaboration, we aim to enable multi-site observational studies that generate reliable evidence to improve health.

Vision

OHDSI Canada envisions a future where Canadian health data are widely harmonized using OMOP, enabling multi-centre federated research across Canada and broad participation in global research networks. Supported by a strong and collaborative community, OHDSI Canada will be a leader in open health data science and contribute to the global OHDSI mission of generating evidence to improve health and care.

Governance and Membership

Steering Committee

- Provides strategic guidance on priorities, activities, overall direction and leadership for the chapter; liaises and ensures alignment with OHDSI

Name	Institution	Region
Alison Park	ICES / HDRN Canada	ON
Anne Martel	Sunnybrook Research Institute	ON
Darya Zhukova	EPAM Systems	ON
Evan Minty	University of Calgary	AB
Lillian Sung	SickKids	ON
Lisa Lix	University of Manitoba / HDRN Canada	MB
Lisiane Leal	CHU Sainte-Justine Research Centre	QC
Mahmoud Azimaee	Centre for Recovery Excellence	AB
Sayem Borhan	St. Joseph's Hamilton / McMaster University	ON

Governance and Membership

Membership

Name	Institution	Region
Ali Mussavi	Unity Health (GEMINI / VITAL)	ON
Anis Sharafoddini	Unity Health (GEMINI)	ON
Anya Okhmatovskaia	McGill University	QC
Ashley Girgis	Terry Fox Research Institute (DHDP)	CAN
Augustine Amakiri	University of Calgary	AB
Brad Millson	IQVIA	CAN
Colin Josephson	University of Calgary	AB
Danielle Boyce	ALS Therapy Development Institute / Johns Hopkins	US
David Buckeridge	McGill University / MUHC	QC
Elaheh Najafi	University of Toronto	ON
Ella Young	BC PHSA	BC
Elodie Portales-Casamar	CHU Sainte-Justine	QC
Georgina Archbold	HDRN Canada	ON
Gord Mawdsley	Sunnybrook	ON
James Gilbert	Johnson & Johnson	CAN
James White	University of Calgary / Libin / PULSE	AB

Governance and Membership

Membership

Name	Institution	Region
Janine Kaye	CIHI	CAN
Jeff Bakal	Alberta Health Shared Services	AB
Karen Cranston	UHN	ON
Kieran Shah	Unity Health (GEMINI)	ON
Loren Li	BCPHSA	BC
Max Wang	SickKids	ON
Michael Hardisty	Sunnybrook Research Institute	ON
Odile Sheehy	CHU Sainte-Justine	QC
Patrick Ryan	OHDSI	US
Peter Gill	SickKids	ON
Robert Platt	McGill University	QC
Sarah Culgin	St. Joseph's Hamilton / The Research Institute of St. Joe's	ON
Son Chau	UHN	ON
Theo Chan	Student	ON

2026 Objectives and Key Results

Objective 1: Establish OHDSI Canada as a Formal Regional Chapter

- Submit registration and obtain approval from the OHDSI Steering Committee
- Establish Terms of Reference including governance structure, participation commitments, meeting frequency
- Set up communication infrastructure
- Hold regular chapter meetings
- Establish 2 subgroups within the Chapter: 'Canadian Vocabularies and Terminologies WG' and 'Evidence Generation Network WG' to pursue specific goals

2026 Objectives and Key Results

Objective 2: Build and Engage the Canadian OHDSI Community

- Create and maintain an OHDSI Canada member directory
- Host an OHDSI Canada symposium

Objective 3: Develop OHDSI Canada OMOP CDM Resources

- Launch a Canadian OMOP CDM repository/database containing vocabulary mappings and ETL documentation
- Complete mapping of at least 1 high-priority Canadian vocabulary

2026 Objectives and Key Results

Objective 4: Strengthen OMOP CDM Capacity Within the Canadian Research Community

- Deliver educational sessions for OHDSI Canada members by invited international OHDSI experts

Objective 5: Demonstrate Value Through Evidence Generation

- Design and initiate a multi-site observational study involving 3+ institutions
- Submit abstract on OHDSI Canada activities or study findings to at least one national conference

2026 Objectives and Key Results

Objective 6: Identify Funding Opportunities for Sustainability

- Conduct review of funding landscape

Objective 7: Develop OHDSI Canada Advocacy Strategy

- Develop an OHDSI Canada advocacy strategy outlining key messages, priority audiences, and engagement approaches

Objective 8: Report Progress and Maintain Alignment with the Global OHDSI Community

- Participate in at least 2 OHDSI global community calls
- Submit annual activity and achievement report to OHDSI Coordinating Centre by fiscal year end

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