



Workgroup Spotlight: Health Systems Interest Group and Surgery & Perioperative Medicine

OHDSI Community Call
July 21, 2026 • 11 am ET



Upcoming Community Calls

Date	Topic
July 21	Workgroup Spotlight: Health Systems, Surgery and Perioperative Medicine
July 28	OHDSI Newcomer Introductions
Aug. 4	Clinical Definitions
Aug. 11	Oncology Research in OHDSI
Aug. 18	Workgroup Spotlight: Eyecare and Vision Research, Rehabilitation
Aug. 25	Asia-Pacific Regional Updates
Sept. 1	Vocabulary Refresh, Summer 2026 Version
Sept. 8	Workgroup Spotlight: Medical Imaging, TBA
Sept. 15	Global Symposium Preview



July 28: Newcomer Introductions

The annual **Welcome OHDSI Newcomers** community call will be held July 28. If you are relatively new to the community or hope to have a bigger role, this is a great opportunity to introduce yourself.

Please tell us who you are and where you work, what are your research interests, how you hope to help the community, and how OHDSI can help your own research journey.





Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?



OHDSI Shoutouts!



Congratulations to the team of
**Mikhail Shubov, Mareile Beernink,
Jasmin Carus, Alexander Johannes
Wiederhold, the AI-CARE Consortium
& Christopher Gundler** on the recent
publication of **Enhancing
stratification for survival analyses
across standardized data sources** in
*BMC Medical Informatics and
Decision Making*.

Shubov et al. *BMC Medical Informatics and Decision Making* (2026) 26:261
<https://doi.org/10.1186/s12911-026-03666-z>

BMC Medical Informatics
and Decision Making

RESEARCH

Open Access



Enhancing stratification for survival analyses across standardized data sources

Mikhail Shubov^{1†}, Mareile Beernink^{1†}, Jasmin Carus², Alexander Johannes Wiederhold¹,
the AI-CARE Consortium and Christopher Gundler^{1*}

Abstract

Background Patient stratification is crucial for advancing personalized medicine yet is complicated due to the fragmented and variable nature of healthcare data. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) addresses the challenges regarding the data by offering a standardized framework for data integration across diverse sources and configurations. Recent advancements in machine learning, particularly transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT), have demonstrated significant potential in extracting deep patient representations from electronic health records.

Methods This study assesses the efficacy of BERT-based patient representation learning using OMOP CDM data for patient stratification. We harmonize originally incompatible datasets, including the established MIMIC-IV-2.2 and lung cancer data from the German cancer registry Schleswig-Holstein, within the OMOP CDM framework. BERT is pre-trained on the MIMIC-IV-2.2 dataset, and the derived representations are utilized to generate patient embeddings from the cancer registry (test) data. We employ k-means clustering on the embeddings to stratify patient subgroups. To evaluate whether the embeddings are useful for clustering, we divide the original cancer registry data into the corresponding groups and conduct survival analyses on selected columns for each cluster. The clustering method's effectiveness is assessed by comparing survival models trained on these clusters with naïve clusters derived only from the original dataset. In addition, we included a clinical expert review, in which a physician assessed the resulting cluster assignments for clinical plausibility and interpretability.

Results Our approach effectively identifies patient similarities across datasets and allowed for efficient patient stratification. Survival analyses show varied performance depending on the model and cluster characteristics, with up to a 11% improvement over naïve k-means clusters, demonstrating the benefits of transfer learning. For some groups of patients, the corresponding accuracy of the survival analysis increased by up to 7%, emphasizing the value of stratifying homogeneous subgroups.



OHDSI Shoutouts!



Congratulations to the team of **Yuchen Guo, Edward Burn, Annika M. Jödicke, Núria Mercadé-Besora, Kim Lopez-Güell, Xihang Chen, Mike Du, Xintong Li, Raivo Kolde, Marek Oja, Mees Mosseveld, Katia Verhamme, Julieta Politi, Antonella Delmestri, Wai Yi Man, Peter Rijnbeek, Daniel Prieto-Alhambra, and Martí Català** on the recent publication of **DrugUtilisation: An R Package to Support Drug Utilisation Research Using the OMOP Common Data Model in Pharmacoepidemiology & Drug Safety.**

Pharmacoepidemiology and Drug Safety

WILEY

ORIGINAL ARTICLE OPEN ACCESS

DrugUtilisation: An R Package to Support Drug Utilisation Research Using the OMOP Common Data Model

Yuchen Guo¹ | Edward Burn¹ | Annika M. Jödicke¹ | Núria Mercadé-Besora¹ | Kim Lopez-Güell¹ | Xihang Chen¹ | Mike Du¹ | Xintong Li¹ | Raivo Kolde² | Marek Oja² | Mees Mosseveld³ | Katia Verhamme³ | Julieta Politi³ | Antonella Delmestri¹ | Wai Yi Man¹ | Peter Rijnbeek³ | Daniel Prieto-Alhambra^{1,3} | Martí Català¹

¹Pharmaco- and Device Epidemiology Group, Health Data Sciences, Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK | ²Institute of Computer Science, University of Tartu, Tartu, Estonia | ³Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands

Correspondence: Martí Català (m.catalasabate@darwin-eu.org)

Received: 30 October 2025 | Revised: 21 May 2026 | Accepted: 7 July 2026

Keywords: drug dose | drug exposure | observational studies | OMOP CDM | treatment adherence | treatment discontinuation | treatment pattern

ABSTRACT

Purpose: To develop and describe *DrugUtilisation*, an open-source R package that facilitates drug utilisation studies using data mapped to the OMOP Common Data Model (CDM).

Methods: Core functionalities include creating drug user cohorts, identifying and summarising indications, describing the duration and dose of medication/s and assessing treatment adherence. The package works with packages developed within the DARWIN EU initiative to support study-specific workflows. We show the package's workflow by analysing the use of simvastatin in three European real-world databases.

Results: This paper outlines the *DrugUtilisation* package's functions and demonstrates their application with a clinical example of simvastatin use in databases from the United Kingdom, Estonia and the Netherlands. We generated results including cohort counts, indication summaries, measures of dose and duration, as well as publication-ready tables and figures. We implemented comprehensive unit tests and standardised output format, which ensured consistency across databases and minimised coding errors.

Conclusion: The development of this software allows for researchers to quickly perform common drug utilisation analyses, while also providing the foundation for additional, bespoke study-specific analyses.



OHDSI Shoutouts!



Congratulations to the team of **Haeun Lee, Benjamin Martin, Andreea Creanga, Evan Minty, Asieh Golozar, Cindy Cai, Cynthia Sung, Sean O'Reilly, Khyzer Aziz, and Paul Nagy** on the recent publication of **Assessment of real-world evidence research competencies in federated research networks: a maternal health fellowship program evaluation** in *JAMIA Open*.

JAMIA Open, 2026, 9(4), ooag128
<https://doi.org/10.1093/jamiaopen/ooag128>
Research and Applications

AMIA OXFORD
ADVANCING MEDICAL RESEARCH. LEADING THE WAY.

Assessment of real-world evidence research competencies in federated research networks: a maternal health fellowship program evaluation

Haeun Lee, MS^{*1}, Benjamin Martin, PhD¹, Andreea A. Creanga, MD, PhD², Evan Minty, MD, MSc³, Asieh Golozar, MD, PhD^{1,4}, Cindy X. Cai, MD^{1,5}, Cynthia Sung, PhD⁶, Sean O'Riley, BS¹, Khyzer Aziz, MD^{1,7}, Paul Nagy, PhD¹

¹Biomedical Informatics and Data Science, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, MD, 21205, United States

²Department of Gynecology and Obstetrics, The Johns Hopkins University School of Medicine, Baltimore, MD, 21287, United States

³Faculty of Medicine, O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, 2500, Canada

⁴Nemesis Health, New York, NY, 10009, United States

⁵Wilmer Eye Institute, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, MD, United States

⁶Center of Regulatory Excellence, Duke-NUS Medical School, Singapore

⁷Department of Neonatology, The Johns Hopkins University School of Medicine, Baltimore, MD, United States

*Corresponding author: Haeun Lee, MS, Biomedical Informatics and Data Science, Johns Hopkins School of Medicine, Johns Hopkins University, 2024 E. Monument Street, Baltimore, MD, 21205, United States (hlee292@jh.edu)

Abstract

Background and Significance: Generating real-world evidence (RWE) approaches across federated evidence networks based on real-world data requires learning new methodological skills. However, standardized frameworks for defining and assessing competencies for large-scale RWE research remain limited. This study aims to develop an assessment framework for RWE studies and evaluate the competencies of RWE research professionals.

Methods: We adapted the Joint Task Force (JTF) Core Competency Framework for RWE research using the Observational Medical Outcomes Partnership (OMOP) Common Data Model. Through iterative expert review, 4 new domains were created, including Ethics and Governance, Protocol Development, Study Operations, and Data Analysis and Informatics, and 4 original JTF domains were adapted. Participants completed pre- and post-training self-assessments on a 10-point scale. Paired t-tests with effect sizes were used to evaluate changes in competency scores.

Results: Seventeen participants from academic, data, and clinical roles completed the assessment. Mean self-reported competency scores increased by 2.5 points across the 8 competency domains, with statistically significant changes observed in all domains. The largest gain was observed in Leadership and Professionalism (mean increase 3.66; 95% CI, 2.69-4.62).

Discussion: The RWE research competency assessment tool captured changes in self-reported skill proficiency, perceived readiness to conduct RWE studies, and methodological knowledge gaps. These findings suggest the potential to identify competency gaps, tailor educational interventions, and inform quality benchmarks for RWE research training programs.



OHDSI Shoutouts!



Congratulations to the team of
Sulev Reisberg, Marek Oja, Kerli Mooses, Sirli Tamm, Ami Sild, Harry-Anton Talvik, Sven Laur, Raivo Kolde, and Jaak Vilo on the recent publication of **Data Resource Profile: EST-Health-30** in the *International Journal of Epidemiology*.

International Journal of Epidemiology, 2026, 55(4), dyag113
<https://doi.org/10.1093/ije/dyag113>
Data Resource Profile



Data Resource Profile: EST-Health-30

Sulev Reisberg^{1,2,*}, Marek Oja^{1,2}, Kerli Mooses¹, Sirli Tamm¹, Ami Sild¹, Harry-Anton Talvik^{1,2}, Sven Laur¹, Raivo Kolde¹, Jaak Vilo^{1,2}

¹Institute of Computer Science, University of Tartu, Tartu, 51009, Estonia

²STACC, Tartu, 51009, Estonia

*Corresponding author. Institute of Computer Science, University of Tartu, Narva mnt 18, 51009, Tartu, Estonia. E-mail: sulev.reisberg@ut.ee

Keywords Estonia, health dataset, EHR, pharmacy, linked data, OMOP

Key Features

- EST-Health-30 is a population-representative dataset of complete health records for a random 30% sample of the Estonian population (~500 000 individuals) spanning 2012 to the present, enabling population-level epidemiological analyses with annual updates.
- The dataset is constructed by using a random sampling approach based on hashed password-protected personal identifiers, ensuring consistent inclusion over time with unbiased population coverage.
- Individual-level data are linked across multiple nationwide databases, including electronic health records, claims, prescriptions, and cancer and cause-of-death registry data, enabling multimodal analyses of health trajectories.
- All data are standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) version 5.4 using international vocabularies (e.g. SNOMED CT, RxNorm, LOINC), supporting reproducibility and participation in federated research networks.
- The dataset is accessible through a secure processing environment compliant with the European Health Data Space framework.

Data resource basics

EST-Health-30 is a population-based health dataset that includes complete health records for a randomly sampled 30% of the Estonian population, linked across five national databases: electronic health records from primary and secondary care, claims, laboratory test results, pharmacy data, and cancer and cause-of-death registries, from 2012 onwards. The dataset currently contains data through to 2024 and includes 509 858 individuals.

It is updated annually through to 2026, with each release typically including data up to the end of the previous calendar year. The expected lag between data generation and availability is ~6–24 months, depending on the source (e.g. manually curated cancer

Despite negative natural growth since the 1990s, the population has remained relatively stable since 2000 due to positive net migration, with a spike in 2022 driven by Ukrainian refugees (2.6% increase). The life expectancy for those born in 2024 is 79.5 years (75.0 years for men and 83.4 years for women), which is slightly below the European Union (EU) average of 81.7 years [1, 2].

Estonian healthcare system

A comprehensive overview of the Estonian healthcare system is provided by Kasekamp *et al.* [3]. Around 94% of the population is insured through the state-funded system managed by the Estonian Health Insurance Fund (EHIF). Roughly half of those in-



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	10 am	Common Data Model
Wednesday	7 pm	Medical Imaging
Thursday	10 am	GIS – Geographic Information System
Friday	9 am	Waveform
Friday	10 am	Transplant
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	9 am	Open EHR and OMOP
Tuesday	9 am	Data2Evidence



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.

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



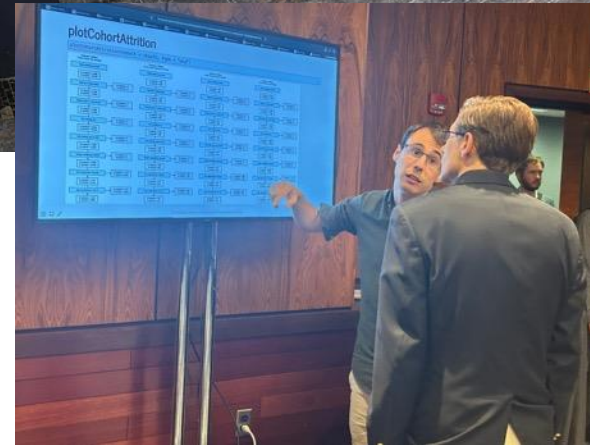
2026 OHDSI Global Symposium

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



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**



2026 Global Symposium Agenda

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



2026 Symposium Workgroup Activities

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

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

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

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



Atlas 3 Demo Site

- Presented by Peter Hoffmann at 7/7 OHDSI Community Call
- We are asking for community feedback
- Published to <https://atlas-preview.ohdsi.org/atlas/>
 - Db Authentication enabled: use ohdsi/ohdsi
 - Will enable anonymous access in a future update
- Discussions at <https://github.com/OHDSI/Atlas3/discussions>



Wanted: HADES Developers

Help wanted - developers for HADES shiny viewer

■ Developers



jreps

5d

We have a set of R packages for developing modular shiny apps to explore the standardized results returned via HADES packages.

- OhdsiShinyAppBuilder - this builds the shiny app and defines what modules to include (e.g., cohort explorer, characterization, estimation, ...)
- OhdsiShinyModules - the main shiny code for the key HADES packages are here
- OhdsiReportGenerator - this is a set of SQL that pulls out friendly data.frames with the results

However, there is a lot of potential for improvements in these packages including (but not limited to):

- adding bookmarking to the URL for easy sharing ([Chapter 11 Bookmarking | Mastering Shiny](#))
- improving display time (improving data extraction speed)
- improving the UI

If you are a developer with knowledge or interest in R or shiny or web development/design and UI and want to help out with HADES development please post below or message me.

I'm planning to set up regular meetings for this work group.



#OHDSISocialShowcase This Week

Monday

Adoption of the OHDSI Geographic Information Systems (GIS) Workgroup for Schizophrenia Research

(**Manos Hatzakis**, Evangelos Georgaras, Iskanter Bensenousi, Despoina Ntenekou, Athanasios Kakasis, Claudia García-Barragán, Lourdes Martínez-Martínez, Miguel Giráldez-Álvarez, Jesús Moreno-Conde, Alberto Moreno-Conde, Marta Quaresma Giraldez, Kiko Núñez, Christos Chatzichristos)

Adoption of the OHDSI *GIS Workgroup* for Schizophrenia Research

Manos Hatzakis, Evangelos Georgaras, Iskanter Bensenousi, Despoina Ntenekou, Athanasios Kakasis, Claudia García-Barragán, Lourdes Martínez-Martínez, Miguel Giráldez-Álvarez, Jesús Moreno-Conde, Alberto Moreno-Conde, Marta Quaresma Giraldez, Kiko Núñez, Christos Chatzichristos



Background

Schizophrenia affects ~1% of the global population and remains a major cause of disability. Early symptoms often overlap with depression and bipolar disorder, delaying diagnosis and treatment. Reliable biomarkers and scalable approaches for early intervention are still lacking. VOLABIOS addresses this challenge by integrating multi-omics, AI/ML, and real-world data from six European countries within the OMOP Common Data Model. In addition to clinical and biological information, environmental and geographic factors are increasingly recognized as important drivers of schizophrenia risk and progression. To support this effort, we adopted the OHDSI GIS Workgroup framework to harmonize spatially resolved environmental exposures within OMOP CDM, enabling standardized and reproducible schizophrenia research.

OMOP CDM — GIS Schema Extensions

Entity-relationship diagram showing the Extension, Exposure and Location, History extension tables and their links to the main OMOP CDM



The ETL pipeline followed a three-step approach aligned with OHDSI best practices:

1. Mapping environmental exposure variables to standard OMOP vocabularies, extending conventions using GIS Workgroup guidance where needed.
2. Representing spatial attributes through location, exposure, and observation period tables aligned with OMOP CDM conventions.
3. Linking environmental exposures geographically and temporally to individual records via standardized location identifiers.

Two retrospective datasets were converted to the OMOP CDM following GIS Workgroup recommendations:

- A. **Geohex dataset:** clinical data linked to environmental exposures (air quality, land use, noise, deprivation, light pollution) from European registries.
- B. **Synthetic dataset:** patient-level clinical + demographic data with geo-referenced census tract IDs from Virgen Macarena University Hospital, Seville.

Conclusion & Next Steps

This work successfully adopts the OHDSI GIS Workgroup conventions to harmonize retrospective environmental exposure data into the OMOP CDM for schizophrenia research. This method facilitates reproducible, interoperable analyses with current OHDSI tools by allowing standardized, spatially and temporally resolved representation of environmental determinants. This foundation offers a scalable path for upcoming multi-site and federated studies within the OHDSI network and reinforces the incorporation of environmental context into schizophrenia research. This approach aligns with VOLABIOS goals by enabling interoperable, privacy-aware, and clinically trusted data pipelines for early detection. Future work will expand GIS concept sets, refine exposure modeling, and evaluate model performance in prospective clinical settings. Future work will investigate adding volatiles data to the OMOP CDM, inspired by the OHDSI Genomics Workgroup. Because volatile organic compounds may capture early biological alterations linked to schizophrenia, volatiles could support earlier diagnosis and better patient stratification when integrated with clinical and environmental data.

References

1. Hery M, Rivai A. Schizophrenia. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. [updated 2024 Feb 23; cited 2026 Feb 6]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK533664/>
2. European Commission. VOLABIOS - Volatile organic compounds and multi-omics biomarkers for early detection of schizophrenia [Internet]. CORDIS; 2024 [cited 2026 Feb 6]. Available from: <https://cordis.europa.eu/project/id/10105862>
3. Observational Health Data Sciences and Informatics (OHDSI). Geographic Information Systems (GIS) Workgroup [Internet]. OHDSI; [cited 2026 Feb 6]. Available from: <https://ohdsi.github.io/GIS/>
4. Moreno A, Moreno J, González V, Salas S, Segura C, de Luque V, et al. GeoHealth: Geographic Information System for Health Management and Clinical, Epidemiological and Translational Research. Stud Health Technol Inform. 2022;290:1084-1088. doi:10.3233/SH1202079.



"This project has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 10105862. © 2024 VOLABIOS. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [name of the granting authority]. Neither the European Union nor the granting authority can be held responsible for them."





#OHDSISocialShowcase This Week

Tuesday

From Lived Experience to Longitudinal Evidence: Designing an OMOP-Compatible Export Layer for Prospective Cancer Patient-Reported Outcomes in Sweden

(**Máté Bendegúz Horváth**, Carlos Manriquez, Lars Halvorsen, Christian Högberg)

41-74% of cancer treatment toxicities go undetected by clinicians. Oncoly captures what happens between clinic visits and standardizes it to OMOP CDM v5.4, making the patient's lived experience a research asset.

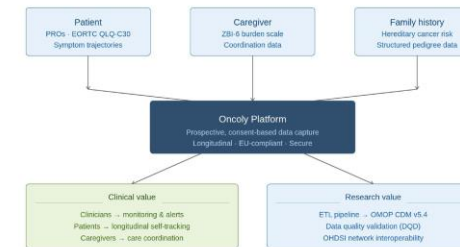
From Lived Experience to Longitudinal Evidence: Designing an OMOP-Compatible Export Layer for Prospective Cancer PRO, Caregiver Burden, and Hereditary Risk Data in Sweden



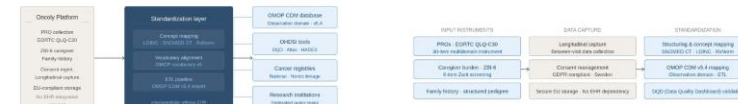
Background



Results



Methods



Key Deliverables



Limitation: Mapping completeness (EORTC QLQ-C30 → OMOP Observation domain) under evaluation via DQD checks

Conclusions:
1. Addresses three major gaps in observational oncology data.
2. Demonstrates feasibility of integrating prospective PROs into OMOP.
3. Enables longitudinal, patient-centered research within OHDSI.
4. Provides a replicable framework for digital health platforms



Lars Halvorsen
edenceHealth

Máté Bendegúz Horváth, Carlos Manriquez
Oncoly AB

Christian Högberg
Passion 2 Improve Sverige AB



(**Petteri Hovi**, Eero Poukka, Ida-Liisa Kolari, Javier Gracia Tabuenca, FinnGen, Gustav Klingstedt)

Petteri Hovi, Eero Poukka, Ida-Liisa Kolari, Javier Garcia Tabuenca, FinnGen, Gustav Klingstedt. *Institute for Health and Welfare, Helsinki, Finland*



#OHDSISocialShowcase This Week

Thursday

Bridging OMOP patient records and patient supplement files through object storage solutions

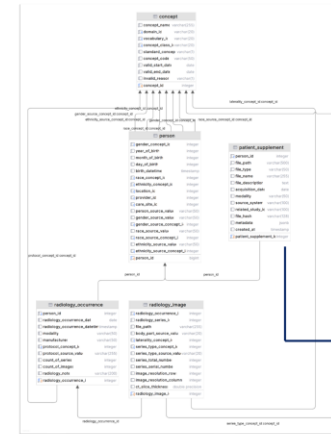
(Frederic Jung, Stef Rommes, Gökhan Ertaylan)

Extending OMOP with Object Storage for Cohort-Driven Access to Unstructured Data

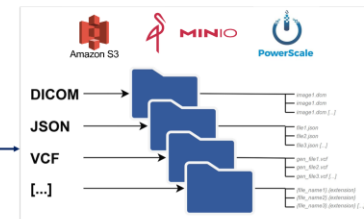
Title: *Bridging OMOP patient records and patient supplement files through object storage solutions*



The OHDSI OMOP Common Data Model (CDM) is widely used for harmonizing clinical data from Electronic Health Records (EHR) but provides limited native support for linking large, unstructured patient-level assets such as medical imaging, genomics, and documents. These data types are increasingly essential, particularly for the evaluation of AI as a medical device, where reproducible and cohort-driven access to supplementary files is critical. By design, the OMOP CDM cannot capture the full complexity or volume of such source data, making external storage and linkage unavoidable. Moreover, many analytical pipelines and medical software tools require direct access to the original source files rather than derived or summarized representations, reinforcing the need for robust mechanisms to associate OMOP records with their corresponding raw data assets. Several approaches have been proposed to address this limitation, such as domain-specific extensions including the Radiology Common Data Model (R-CDM) and more recently the Medical Imaging CDM (MI-CDM). However, these solutions are typically focused on a single data modality and lack the generality required to support heterogeneous unstructured data types. Any solution for integrating such data with OMOP must therefore be highly flexible, and deployable across diverse technical environments, while preserving the integrity and interoperability of the OMOP CDM.



This work was heavily inspired by the R-CDM, which links OMOP records to DICOM objects via external file references and structured metadata. We generalize this concept by introducing a lightweight extension table (`patient_supplement`) that links OMOP `person_ids` to object-stored files across multiple modalities. The approach enables nearly fully automated, scalable retrieval of patient supplement files for cohort-based analyses. It supports multiple storage backends, and large file volumes without impacting OMOP performance. The proposed pattern has been validated using local file systems as well as S3-compatible object storage solutions (MinIO, PowerScale).



This work demonstrates a pragmatic and OMOP-compatible extension pattern for integrating object storage with cohort-driven workflows without altering or constraining the core CDM. By layering a single, lightweight extension table on top of an existing OMOP instance, the approach preserves the OMOP CDM spirit while enabling access to supplementary patient-level data. The required setup is intentionally simple, relying on minimal schema additions and straightforward ETL logic for integrating object storage with OMOP without modifying the core CDM.



Frédéric Jung¹, Stef Rommes¹, Gökhan Ertaylan¹
¹. VITO, Vlaamse Instelling voor Technologisch Onderzoek, Mol, Belgium





#OHDSISocialShowcase This Week

Friday

Design and Implementation of an Automated OMOP Database Migration Tool from PostgreSQL to SQL Server

(Mandickel Donnex Kamtengeni,
Antonella Delmestri)

Design and Implementation of an Automated OMOP Database Migration Tool from PostgreSQL to SQL Server

Background

The Observational Medical Outcomes Partnership (OMOP) common data model (CDM) standardises observational health data across institutions. However, its use differs across Relational Database Management System (RDBMS) platforms such as PostgreSQL and Microsoft SQL Server, making database migration complex and inconsistent. A reliable, reproducible framework is therefore needed to accurately migrate schemas and data while preserving OMOP CDM integrity.

Objectives

To design and implement a modular, transparent, and maintainable database migration tool for transferring OMOP CDM databases across different RDBMS environments, using PostgreSQL to Microsoft SQL Server as a case study.

Methods

The proposed modular framework for transferring OMOP CDM databases across RDBMS platforms as shown in Figure 1.

1. **Extraction:** Batch scripts export tables to CSV files.
2. **Loading:** T-SQL bulk insert exported data into staging tables.
3. **Schema Creation:** OMOP CDM DDL scripts from Observational Health Data Sciences and Informatics (OHDSI) create tables and objects.
4. **Constraints:** Keys and indexes are applied post-load.
5. **Automation:** Execution is fully automated in Batch.
6. **Configuration:** Parameters are ready for multi-environment use.
7. **Logging:** Execution status and errors are recorded.

Results

Implementation of the proposed migration approach demonstrated the successful recreation of the OMOP CDM databases from PostgreSQL to Microsoft SQL Server through extraction, transfer and reconstruction. Schema objects, primary keys, indexes and foreign keys were recreated on the target system with no structural inconsistencies observed during the database reconstruction. Execution logs confirmed the completion of each migration stage without errors, demonstrating suitability for repeatable and maintainable OMOP CDM migrations across heterogeneous RDBMS environments.

Conclusion

The proposed migration framework addresses a key interoperability challenge faced by institutions deploying OMOP CDM databases across RDBMS. The staged bulk-loading strategy effectively eliminates possible issues related to constraint enforcement and data type incompatibilities during migration by separating data ingestion from schema enforcement. This approach supports systematic recreation of schemas, tables, primary keys, indexes and foreign keys in the target SQL Server environment.

This design choice improves transparency by making structural transformations explicit and reviewable, while also supporting version control and long-term maintainability. Limitations of the current framework include the small number of RDBMS used. Future enhancements will focus on extending the framework across other RDBMS, optimising performance, and automating the execution and integration of RDBMS-specific DDL artefacts provided by OHDSI.

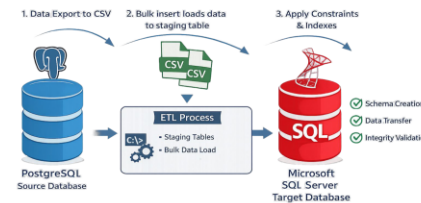


Figure 1. OMOP CDM migration workflow from PostgreSQL to Microsoft SQL Server



GitHub



Mandickel Kamtengeni¹, Antonella Delmestri¹
¹Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
University of Oxford, Oxford, United Kingdom
Email: mandickel.kamtengeni@ndorms.ox.ac.uk





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?



**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://www.ohdsi.org/community-calls-2026)



Find your workgroup.

Fuel our mission.

ohdsi.org/workgroups