



Fun With Clinical Definitions

OHDSI Community Call
Aug. 4, 2026 • 11 am ET



Upcoming Community Calls

Date	Topic
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



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?



OHDSI Shoutouts!



Congratulations to the team of **Colin T Purmal, Michael E Matheny, Lucila Ohno-Machado, Gary Tarasovsky, Rick Larsen, and Mary A Whooley** on the recent publication of **Success and Challenges Querying OMOP-transformed EHR Data from Different Healthcare Organizations** in *ACI Open*.

Article published online: 2025-08-26

e42 Case Report



Success and Challenges Querying OMOP-transformed EHR Data from Different Healthcare Organizations

Colin T. Purmal^{1,2} Michael E. Matheny^{3,4} Lucila Ohno-Machado⁵ Gary Tarasovsky^{1,2}
Rick Larsen² Mary A. Whooley^{1,2}

¹ San Francisco VA Health Care System, San Francisco, California, United States

² University of California, San Francisco, California, United States

³ VA Tennessee Valley Health Care System, Nashville, Tennessee, United States

⁴ Vanderbilt University, Nashville, Tennessee, United States

⁵ Department of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, Connecticut, United States

Address for correspondence Colin Purmal, MD, San Francisco VA Medical Center, 4150 Clement Street, San Francisco, CA, 94121, United States (e-mail: Colin.purmal@va.gov).

ACI Open 2025;9:e42–e53.

Abstract

Background Adoption of a common data model (CDM), such as the Observational Medical Outcomes Partnership (OMOP) CDM, is a critical component of health information exchange, but the extent to which CDMs facilitate patient-level interoperability is unclear. We sought to determine the feasibility of using a CDM for health information exchange between two healthcare organizations.

Materials and Methods We executed a single Structured Query Language (SQL) query on OMOP-transformed data from the University of California, San Francisco (UCSF) and the San Francisco VA Healthcare System for five patients.

Results The SQL query was successfully executed, and complementary healthcare information was obtained for five of five patients, but interoperability was limited by (1) lack of uniform patient identifiers, (2) use of different coding vocabularies, and (3) variations in mapping of source data to the CDM.

Conclusion Although transformation of EHR data to a CDM can facilitate health information exchange, our study suggests that patient-level interoperability of a CDM may require further alignment of both semantics and syntax.

Keywords

- ▶ interoperability
- ▶ common data model
- ▶ health information exchange



OHDSI Shoutouts!



Congratulations to the team of **Gabriel L. O. Salvador, Jen Wooyeon Park, Teri Sippel Schmidt, Yiju Park, Kyulee Jeon, Seng Chan You, Paul Nagy, Blake Dewey & the Alzheimer's Disease Neuroimaging Initiative** on the recent publication of **Advancing the FAIRness of Multimodal Imaging Research Through the OMOP MI-CDM Framework: A Case Replication Study in Alzheimer's Disease** in the *Journal of Imaging Informatics in Medicine*.

Journal of Imaging Informatics in Medicine
<https://doi.org/10.1007/s10278-026-02146-0>



Advancing the FAIRness of Multimodal Imaging Research Through the OMOP MI-CDM Framework: A Case Replication Study in Alzheimer's Disease

Gabriel L. O. Salvador¹ · Jen Wooyeon Park¹ · Teri Sippel Schmidt¹ · Yiju Park^{2,3} · Kyulee Jeon^{2,3} · Seng Chan You^{2,3} · Paul Nagy¹ · Blake Dewey¹ · Alzheimer's Disease Neuroimaging Initiative

Received: 7 April 2026 / Revised: 23 June 2026 / Accepted: 11 July 2026
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2026

Abstract

The objective of this study is to demonstrate an end-to-end approach for operationalizing the Findable, Accessible, Interoperable, and Reusable (FAIR) principles in multimodal medical imaging research using standardized data models and reproducible computational workflows, illustrated by reproducing the design and directional findings of a published Alzheimer's disease (AD) imaging study. Clinical and imaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI-4) were harmonized within the Observational Medical Outcomes Partnership Common Data Model and its Medical Imaging extension (OMOP MI-CDM). MRI acquisition metadata (DICOM) were extracted and mapped to standardized concepts, while clinical variables were integrated via reproducible extract-transform-load processes. Interoperable phenotypes were defined using OHDSI tools. Hippocampal volumes were derived from T1-weighted MRI using a fully automated machine learning segmentation pipeline (OpenMap-T1). Imaging attributes and derived measurements were stored as structured, provenance-preserving records in OMOP MI-CDM. We replicated a reference study evaluating hippocampal volume differences across AD, mild cognitive impairment (MCI), and cognitively normal controls, stratified by age and sex. We included 289 participants and 545 MRI studies. Across age- and sex-stratified cohorts, mean hippocampal volumes showed consistent directional reductions in AD compared with controls, with intermediate values in MCI, matching trends reported in the replication study. FAIR principles can be operationalized across the full imaging research pipeline using OMOP MI-CDM and automated analysis workflows. This framework enables transparent cohort definition, reproducible image processing, and interoperable reuse of machine learning-derived imaging features to support scalable validation and reproducible multimodal imaging research.

Keywords Alzheimer's disease · OHDSI OMOP CDM · OpenMap-T1 · Machine learning · DICOM



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	7 am	Medical Imaging
Wednesday	9 am	Tidy R Programming with OMOP
Wednesday	10 am	Common Data Model
Thursday	6:30 am	India Community Call
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
Friday	9 am	Waveform
Friday	10 am	Respiratory Research
Friday	10 am	Transplant
Friday	11:30 am	Steering
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	9 am	Data2Evidence

August Newsletter is Available



The Journey Newsletter (August 2026)

Check out the enhancements coming in ATLAS 3.0, learn more about a record-breaking Global Symposium showcase, check out an impressive collection of updates and publications from July, as well as spotlights on Maxim Moinat and Misha Khitrin, and plenty more in the August 2026 newsletter.

[#JoinTheJourney](#)

Podcast: ATLAS 3.0, Showcase Themes



On The Journey

Craig Sachson Patrick Ryan

OHDSI

In the August 2026 On The Journey podcast, Patrick Ryan and Craig Sachson discuss ATLAS as a community tool, enhancements being made for ATLAS 3.0 and why community support is critical now. They also discussed the 2026 collaborator showcase, including how AI was used in the review process, and some themes that came out of a record-breaking submission total. (If video does not appear, please click 'View this email in your browser')

Community Updates

Where Have We Been?

- **Record Submission Total:** Acceptances went out to more than 160 submissions for [the 2026 Global Symposium](#), including more than 130 posters and 24 demos. There will be 10 lightning talks this year, and those are listed later in this newsletter. Thank you to the 41 volunteers in our Scientific Review Committee for their time and efforts in the peer-review process.
- **Global Growth:** The first LATAM Symposium was held July 30-31 in Salvador, Brazil, with the goal of bringing together the public sector, industry and academia to advance health data standardization and strengthen an ethical, open and inclusive regional community. The symposium worked to accelerate OMOP Common Data Model adoption in Brazil and across the region, fostering integration, scientific collaboration and social impact. Thank you to [Valentina Martufi](#) and [Julio Oliveira](#) for leading this initiative.
- **New Workgroup:** [Jenna Reys](#) recently posted a call for developers with knowledge or interest in R or shiny or web development/design and UI to join the REVEAL Workgroup (Results Enhancements & Visualization for Evidence, Analysis & Learning). If you are interested, [please respond to the forum post or sign up here](#).

Where Are We Now?

- **Clinical Definitions:** A critical aspect of the phenotype development and evaluation process is aligning on a clinical definition that provides a full specification for each of the indications, exposures, and outcomes of interest in our studies. But how good are we as a community at reaching that shared understanding? During the Aug. 4 [community call](#), [Patrick Ryan](#) will lead us in examining the topic of clinical definitions by highlighting work in our community and through a series of interactive exercises.
- **ATLAS 3.0 Feedback:** [Peter Hoffmann](#) provided a [preview of the ATLAS 3 demo site](#) during the July 7 community call. The ATLAS workgroup is looking for feedback about the new site. [The demo is available here](#) (Db Authentication enabled: use ohdsi/ohdsi, and anonymous access will be enabled in a future update). [Discussions will be held here](#). More details are included later in this newsletter. Thank you for collaborating on this critical open-source tool for our community.
- **Survey Requests:** [Jiwon Um](#), a PhD student at Yonsei University in Korea, [introduced a brief survey focusing on OHDSI in the Era of Medical AI](#). Medical AI is reshaping how observational research evidence is generated and used. The OHDSI community has spent more than a decade building infrastructure — OMOP CDM, standardized vocabularies, federated networks, open-source methodology — that may have a distinctive role in this transition. This survey asks how you, as a member of the community, see that role and the tensions surrounding it. Also, [Jiwoo Seo](#), a PhD student at Florida State University, announced a study on the motivations that drive experts to continuously contribute to biomedical ontology communities like OHDSI. [You can take part in this survey here](#).

Rebuilt for Performance: Explore the ATLAS 3.0 Preview and Share Your Feedback



What's in the ATLAS 3.0

Available	Removed — not carried forward
✓ Cohort definitions — fully redesigned builder ✓ Concept sets, vocabulary search & exploration ✓ Data sources: CDM dashboards & reports ✓ Characterisation & feature analysis ✓ Incidence rates & cohort pathways ✓ Patient profiles ✓ Roles, permissions & version history ✓ Auth: DB, LDAP/AD, OAuth, SAML, OpenID	✗ Population-level estimation ✗ Patient-level prediction ✗ Hierarchical reports & legacy cohort analysis <i>In line with the interview feedback: After focus on cohort design and characterisation, population-level analysis fits in the R / HANDE ecosystem.</i>

ATLAS is a free, open-source web platform that empowers researchers to explore and analyze standardized healthcare data. By integrating with the OMOP Common Data Model, ATLAS enables users to search medical vocabularies, define specific patient cohorts, and examine population data through an intuitive web interface—all without needing direct database access or writing complex code.

The upcoming ATLAS 3.0 release completely modernizes the platform behind the scenes to deliver a faster, smoother user experience. On the surface, users will benefit from a refreshed interface with updated search tools, sharper interactive charts, and instant patient counts as you refine your study criteria. Under the hood, a new modular design makes the tool significantly more responsive and gives technical teams the flexibility to build custom extensions tailored to their organization's unique research needs.

As we prepare for the official rollout, we invite community members and partner organizations to explore the preview release, test its new features, and help shape its final release. You can check out the source code and deployment guides on [GitHub](#) or review the [ATLAS 3.0 Preview slides](#) for a deeper look at what's new. Please share your feedback, bug reports, and suggestions with the OHDSI development team on [GitHub](#) to help ensure ATLAS 3.0 delivers the best possible experience for researchers worldwide.

[ATLAS 3.0 Preview](#)

[GitHub: ATLAS 3.0](#)

[Video: ATLAS 3.0 Presentation & Demo](#)

My Journey: Maxim Moinat



My Journey

OHDSI

In the latest installment of our "My Journey" series, Maxim Moinat (Researcher at the Medical Informatics Department of the Erasmus Medical Center) shares his love for turning "messy real-world data" into high-level evidence. Maxim discusses his pride in building open-source ETL tools used across the global OHDSI community, the warmth of the network's family-like culture, and the massive trajectory of the DARWIN EU initiative as it scales up parallel studies to impact international healthcare decision-making. (If video does not appear, please click 'View this email in your browser')

Global Showcase To Feature 160+ Submissions, Highlighting Wide Breadth of Research



The Collaborator Showcase remains one of the highlight events during any OHDSI symposium, and the one coming this October in New Jersey should be one of the most memorable. Following a record-breaking total of showcase submissions, our amazing Scientific Review committee of 41 volunteers (see graphic below) accepted more than 160 posters, software demos or lightning talks to share during the 2026 Global Symposium.

July Publications

Guo Y, Burn E, Jódicke AM, Mercadé-Besora N, Lopez-Güell K, Chen X, Du M, Li X, Kolde R, Oja M, Mosseveld M, Verhamme K, Politi J, Delmestri A, Man WY, Rijnbeek P, Prieto-Alhambra D, Català M. [DrugUtilisation: An R Package to Support Drug Utilisation Research Using the OMOP Common Data Model](#). Pharmacoepidemiol Drug Saf. 2026 Jul;35(7):e70433. doi: 10.1002/pds.70433. PMID: 42448616; PMCID: PMC13368618.

Park J, Kim JH, Kim YS, Lim G, Song HJ, Rhee SY, Hwang S, Cho D, Kim M, Kim BY, Jeong CW, Kim DY, Park HA, Kim MH, Kim SH, Seo WW, Kim WJ, Suh YS, Park RW, Shin JY. [GLP-1 Receptor Agonists for Weight Loss and Risk of Major Safety Outcomes: A Multicentre Cohort Study](#). Diabetes Obes Metab. 2026 Jul 6. doi: 10.1111/dom.71065. Epub ahead of print. PMID: 42410329.

Rueda M, Ramírez-Anguita JM, López-Sánchez V, Aguiló-Castillo S, López MEG, Labarga A, Mayer MÁ, Esteve JR, Gut IG. [Enhancing semantic interoperability in precision medicine: converting OMOP CDM to Beacon v2 in the Spanish IMPaCT-Data project](#). BMC Med Inform Decis Mak. 2026 Jul 8. doi: 10.1186/s12911-026-03649-0. Epub ahead of print. PMID: 42420997.

Shubov M, Beernink M, Carus J, Wiederhold AJ; AI-CARE Consortium; Gundler C. [Enhancing stratification for survival analyses across standardized data sources](#). BMC Med Inform Decis Mak. 2026 Jul 13;26(1):261. doi: 10.1186/s12911-026-03666-z. PMID: 42443871; PMCID: PMC13361719.

López-Sánchez I, Palomar-Cros A, Giuliodori A, Granés L, Pérez-Crespo L, Raventos B, Burn E, Barchuk A, Verbiest A, Eteve-Pitsaer C, Newby D, Rowlands EJ, Enerly E, Jadhav G, De Schutter H, Hsu JC, Evers J, Vrancic J, Peruchet-Noray L, Sung L, Kiss L, Van Swieten M, Šekerić M, Mosseveld M, Mabry PL, Kolde R, Patnoe S, You SC, Kim S, Reisberg S, Nguyen TKH, Lehtonen T, Myklebust TÅ, Bagyuza Z, Golozar A, Roel E, Duarte-Salles T. [Temporal trends in epidemiology and patient characteristics of 36 cancers: a protocol for a multinational population-based cohort study using OMOP-standardised databases to investigate CANcer \(OMOPCAN\)](#). BMJ Open. 2026 Jul 21;16(7):e119069. doi: 10.1136/bmjopen-2026-119069. PMID: 42481199; PMCID: PMC13410706.

Ozonoe O, Okonor O, Dike U, Adedeji T. [Enabling interoperability across disparate health data sources using Large Language Models](#). PLOS Digit Health. 2026 Jul 23;5(7):e0001511. doi: 10.1371/journal.pdig.0001511. PMID: 42490570; PMCID: PMC13395378.

Kim H, Kim M, Kim S, You SC. [From study design to executable code: automating target trial emulation with large language models](#). JAMIA Open. 2026 Jul 9;9(4):oag131. doi: 10.1093/jamiaopen/oag131. PMID: 42428529; PMCID: PMC13348706.

[#JoinTheJourney](#)

www.ohdsi.org





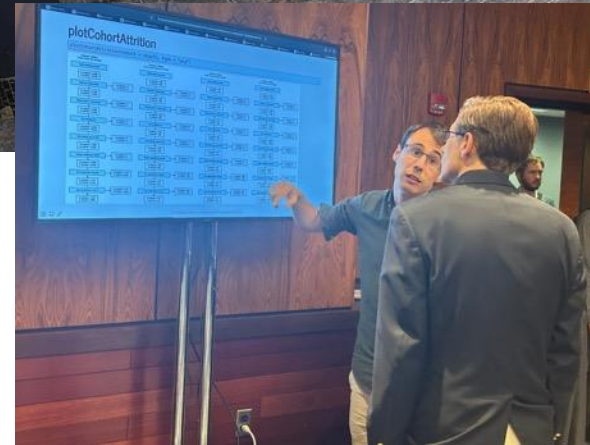
2026 OHDSI Global Symposium

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



APAC Symposium: Nov. 13-15

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

Nov. 13: Main Conference

Nov. 14: Tutorials, Datathon

Nov. 15: Datathon Team Activities

Registration will open soon, and the showcase submission deadline is August 17 at 8 pm ET.



ohdsi.org/APAC2026



#OHDSISocialShowcase This Week

Monday

Declarative Data Standardization to OMOP-CDM using LinkML copia

(Alberto Labarga)

Declarative Data Standardization to OMOP-CDM using LinkML

Alberto Labarga¹, Tom Stone²
¹MirrorHealth SL, Pamplona, Spain, ²Roche Products Limited, Welwyn Garden City, UK

The secondary use of clinical data for research is severely constrained by data heterogeneity across healthcare systems. Although the OMOP Common Data Model (OMOP-CDM)¹ provides a robust and widely adopted standard for observational research, the Extract-Transform-Load (ETL) process required to populate OMOP remains a major barrier. Traditional ETL pipelines rely on bespoke, imperative code (e.g., SQL or Python), making them costly to develop, difficult to maintain, and hard to reproduce. This poster presents a paradigm shift in OMOP standardization, reframing ETL as a declarative, model-driven, and AI-augmented process using the Linked Data Modeling Language (LinkML)³.

LinkML

Declarative data model

A LinkML schema defines the structure and semantics of a data model. It's the top-level specification that declares:

- What kinds of things exist (classes)
- What attributes or properties they have (slots)
- What controlled values those properties can take (enums)
- How these elements relate to each other



Figure 1: LinkML schema outline



Figure 2: Source data LinkML profile generated with schemata



Figure 3: Target data LinkML profile for OMOP-CDM v5.1

LinkML-map

Declarative transformation

We implemented a declarative clinical data to OMOP-CDM standardization framework built on the LinkML ecosystem. First, source database schemas were automatically generated using LinkML Schema Automator⁴, producing formal class-slot-enum definitions directly from relational databases. These schemas were semantically enriched using BioPortal-based annotators to suggest candidate mappings to OMOP standard vocabularies. Second, a canonical LinkML representation of the OMOP-CDM was used as the target schema. Third, a large language model (LLM) was employed to generate a LinkML-Map⁵ transformation specification aligning source and OMOP schemas. To ensure reliable structured output, we used BAML's schema-aligned parsing⁶, constraining the LLM to produce syntactically valid LinkML-Map artifacts. Finally, the LinkML-Map Transformer executes the declarative mappings to generate OMOP-CDM-conformant data, with validation performed against the OMOP LinkML schema.



Figure 4: LinkML-based standardization pipeline

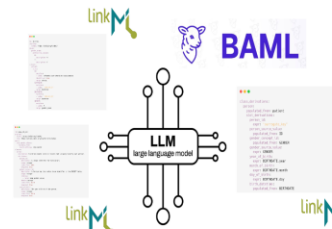


Figure 5: BAML's schema-aligned parsing is used to generate proper LinkML-map definition for the transformation between two LinkML data models using any LLM

Results

The proposed approach yielded qualitative improvements over traditional ETL pipelines. All transformation logic was externalized into human-readable, version-controllable YAML artifacts, improving transparency, auditability, and reproducibility. The use of structured schemas and schema-constrained LLM outputs substantially reduced malformed mappings and simplified expert review. The workflow promoted modularity by separating source modeling, vocabulary grounding, mapping generation, and execution, allowing independent evolution of each component. Rather than replacing human expertise, the framework functioned as a mapping co-pilot, automating repetitive tasks while enabling domain experts to review and refine mappings.

Conclusion

This work demonstrates that declarative, model-driven ETL using LinkML can significantly lower the technical barrier to OMOP-CDM adoption. By replacing brittle imperative scripts with explicit schemas and mapping specifications, the framework enhances maintainability, reproducibility, and trust. Structured AI integration improves efficiency while preserving human-in-the-loop governance, which remains essential given the semantic ambiguity of clinical data and the limitations of LLMs. The approach balances standardization and local customization, enabling institutions to adapt mappings to local data while ensuring strict OMOP compliance. Declarative data standardization provides a scalable foundation for federated observational research and aligns well with OHDSI's goals of reusable, transparent, and shareable analytics pipelines.

References

- [1] OHDSI. The Book of OHDSI. <https://ohdsi.github.io/TheBookOfOhdsi/>
- [2] Moxon, Sierra AT, et al. "LinkML: An Open Data Modeling Framework." GigaScience (2025)
- [3] Moxon, Schema Automator. <https://linkml.io/schema-automator/>
- [4] LinkML-Map Documentation. <https://linkml.io/linkml-map/>
- [5] Boundary ML (BAML). <https://github.com/baml-org>



mirrorhealth.eu



#OHDSISocialShowcase This Week

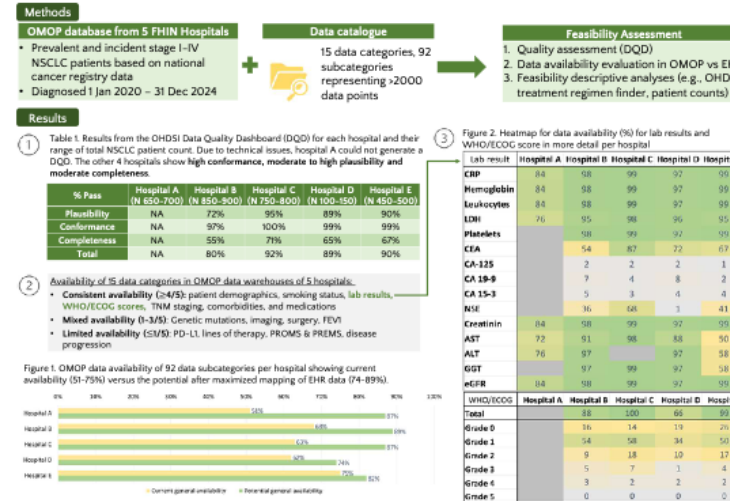
Tuesday

The Belgian FHIN-hospitals Industry Data Alliance (BELFHINDA) - Outcomes and learnings from a feasibility assessment in non-small cell lung cancer (Karen Crabbé, Peter De Jaeger, Siel Depestele, Kim Denturck, Aldo Elsen, Katoo Muylle)

The Belgian FHIN-hospitals Industry Data Alliance (BELFHINDA) : Outcomes and Learnings from a Feasibility Assessment in NSCLC

Feasibility analysis using datapoints available in the OMOP model from 5 FHIN hospitals, combined with data quality evaluation and basic analysis of the NSCLC cohort to inform readiness for RWD studies.

Background: The Federated Health Innovation Network (FHIN) VZW is a collaborative initiative uniting 11 (non-)academic hospitals in Belgium to facilitate data-driven care by leveraging data harmonization via OMOP CDM. The network focusses on building OMOP data warehouses based on mainly structured data in the participating hospitals, creating a federated platform to collaborate and started developing prediction models in lung and prostate cancer. FHIN partners via the Belgian pharmaceutical Industry Association, pharma.be, with pharmaceutical companies to conduct real world data (RWD) research in the lung cancer domain. This abstract presents the outcomes and learnings from both the FHIN network and industry perspectives of the feasibility study.



Conclusion: Results show the potential of the FHIN NSCLC database (n=2800-3050) for leveraging RWD research questions. Completeness should be improved through maximized mapping and by unlocking unstructured data across all hospitals.

Industry perspective: Federated, multi-hospital queries in Belgium are feasible with FHIN within short timelines (6 months), confirming their potential as a valuable RWD partner. Governance enables multi-stakeholder collaboration while preserving data sovereignty. Advanced RWD studies require increased data availability and analytics capabilities.

FHIN perspective: Confirmation of value in building a network around OMOP to enable high quality, scalable, and reusable RWD research. Alignment on standards, tools, governance and federated analytics allow hospitals with heterogeneous data to jointly perform research. FHIN is a mature partner and provides a strong foundation for new partners to co-create.

Karen Crabbé¹, Siel Depestele², Peter De Jaeger², Kim Denturck², Aldo Elsen³ and Katoo Muylle⁴ on behalf of BELFHINDA-Lung group

¹pharma.be, Belgium; ²BADAR, AZ Delta Roeselare, Belgium; ³MED Belgium; ⁴AstraZeneca Belux

References: 1. Ollermann C, Denford R, De Jaeger P, et al. Data-driven medical practices and enhanced patient care through OMOP CDM and federated learning in 9 Belgian hospitals. Presented at OHDSI Europe 2024.

Acknowledgments: The BELFHINDA-Lung group: for the hospital partners (Annelies Verbeke, Renier Weert, Stephanie Delellier, Karel Van Brantegem, Bram De Caluwé, Hilde Philips, Sofie Vandemorre, Jeroen Willems, Lieve Vermeulen, Alissa Ahmed, Zarah Van Schoor, Frenck Goring) and for the industry partners (David Serrano, Katrijke van der Hoeven, Jens Tytgat, Bieke van der Valken, Géraldine Deroose, Ann Curias, Catheline de Moyer, Bart Verheyden, Jo Jansen, Geert Van Casteren, Elise Daquennes, Jan Verbeke, Bart Vermeulen, Ludo De Rop, Hilde Van Camperhout, Sander Burgmans)



#OHDSISocialShowcase This Week

Wednesday

Enhancing Multinational Drug Safety Studies: Mapping the Nordic Drug Product Identifiers (Varenummer) to the OMOP Standard

(**Saeed Hayati**, Zeenat Fatima Chatha, Petrica Olteanu, Hedvig Nordeng)

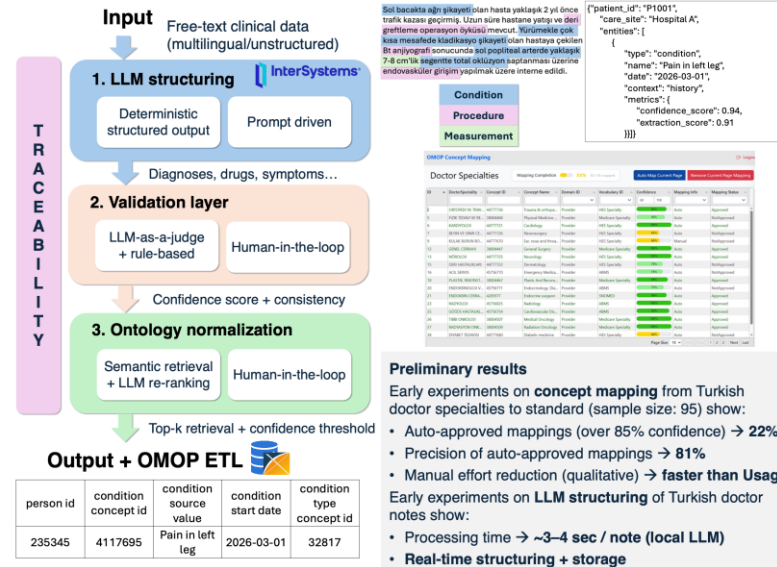
Explainable, Governance-Ready Clinical NLP Pipeline for OMOP-CDM Using Locally Deployed LLMs

Alisson Licona¹, Muhammad Waseem¹, Alex Loleit¹
¹Traverse Health

Clinical free-text notes contain **rich patient information** but remain difficult to **reuse for large-scale research, limiting integration** into standardized frameworks such as OMOP-CDM. While NLP methods have improved clinical information extraction, traditional rule-based systems lack flexibility, and modern machine learning approaches often face **challenges in interpretability, scalability, and cross-lingual adaptation**. Mapping local clinical concepts —especially from non-English data— to standardized vocabularies remains a critical bottleneck.

We propose an **explainable, end-to-end clinical NLP pipeline** that transforms free-text into standardized, research-ready data aligned with OMOP-CDM.

A preliminary implementation has been developed, focusing on the ontology normalization stage. In parallel, an LLM-based structuring module has been implemented to convert free-text into structured representations, currently undergoing validation.



Conclusion

We present an **explainable, end-to-end framework** for structuring clinical free text into OMOP-aligned data. The pipeline demonstrates the potential to enhance standardization, **transparency**, and scalable research. Local LLMs ensure **data governance and accessible deployment**.

Future work will focus on **full pipeline integration, large-scale validation, and benchmarking** against existing OMOP mapping approaches.





#OHDSISocialShowcase This Week

Thursday

Experiences with OMOPping Rehabilitation Data: A Unified ETL Strategy for Nine Distinct Clinical Datasets

(**Petros Patias**, Charalampos Georgiadis, Themistoklis Roustanis, George-Robert Patias, Alexandros Zacharegas, the PREPARE Project Group)

PREPARE project uses OHDSI tools for privacy-preserving clinical prediction
A Unified ETL Strategy for Nine Clinical Datasets

Experiences with OMOPping Rehabilitation Data A Unified ETL Strategy for Nine Distinct Clinical Datasets

Background: The Rehabilitation research relies on real-world observational data to model the patients' progression, to stratify populations, and to support personalized care pathways. Different rehabilitation organizations have heterogeneous data, ranging from clinical measurements and PROMs to functional assessments and free-text clinical information. The PREPARE Project addresses these challenges across nine rehabilitation clinical. PREPARE is implemented aiming at integration of heterogeneous rehabilitation datasets, using a unified analytical framework to support predictive modelling and personalized rehabilitation planning. This achieved using OMOP Common Data Model and standardizing rehabilitation data using a unified ETL strategy and tool-supported methodology, highlighting how OHDSI standards were used to support rehabilitation research needs rather than generic data integration

Results - OMOP Standardization Progress

Clinical cases Completed to-date: 6 / 9 (67%)

ETL Iterations: 3-5 cycles per partner

Key Challenges: High (heterogeneity, vocab gaps)

The unified ETL environment facilitated early identification of semantic inconsistencies and supported informed decisions on data inclusion, transformation, or exclusion, securing consistency on data quality principles.

Conclusion

The PREPARE project illustrates that the OMOP CDM can successfully facilitate rehabilitation research when implemented through a Unified ETL Framework. This approach incorporated the ETL tools to transform and unify rehabilitation data, enabling Clinicians-Centered Mapping, and embedded Data Quality principles, leading to a complete OHDSI Alignment in guiding targeted rehabilitation research.

Unified ETL framework and tools bridges data standards and real-world rehabilitation practice

Methods

Unified ETL strategy for rehabilitation data standardization

Data Context - Heterogeneous rehabilitation datasets

- Clinical data
- PROMs
- Functional assessments
- Free-text records

Unified ETL Strategy - Aligned with EU data quality principles:

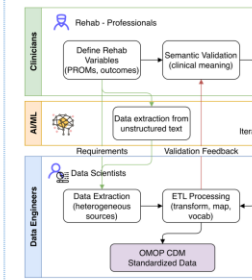
- Relevance
- Coherence
- Completeness
- Secondary-use utility

Minimum Viable OMOP Dataset - Defined per clinical case:

Clinicians + Data Scientists Focus on:

- Predictive variables
- Functional outcomes
- Temporal events

Workflow



Tool & Methodology

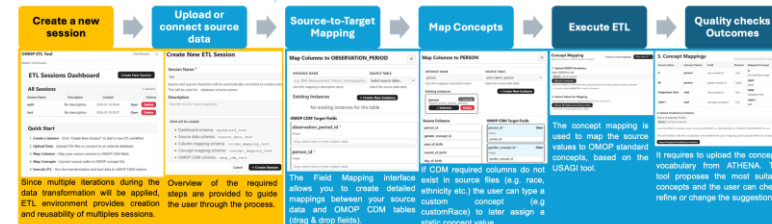
Unified ETL Tool

Supports:

- Schema mapping
- Data extraction & transformation
- Vocabulary mapping (ATHENA)
- Validation & iteration

Methodological Features

- Iterative process → session-based ETL
- Semantic mapping → ATHENA + clinician refinement - Custom concepts where needed
- Data quality feedback → OHDSI-aligned checks
- Traceability → Transparent & reusable mappings



Petros Patias, Charalampos Georgiadis, Themistoklis Roustanis, George-Robert Patias, Alexandros Zacharegas and the PREPARE Project Group

Funded by the European Union. UK participants in Horizon Europe Project PREPARE are supported by UKRI grant number 10086219 (Triateral Research). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HADEA) or UKRI. Neither the European Union nor the granting authority nor UKRI can be held responsible for them. Grant Agreement 101080288 PREPARE HORIZON-HLTH-2022-TOOL-12-01



#OHDSISocialShowcase This Week

Friday

Early adulthood adiposity trajectory and all-cause mortality among obesity-related cancer patients in Catalonia, Spain

(**Mariane Alves**, Agustina Giuliadori, Laura Pérez-Crespo, Elena Roel, Andrea Pistillo, Irene López-Sánchez, Anna Palomar-Cros, Talita Duarte-Salles)

Longer duration and greater cumulative exposure to overweight and obesity in early adulthood were associated with higher risk of all-cause mortality among individuals with obesity-related cancer

Early adulthood adiposity trajectory and all-cause mortality among obesity-related cancer patients in Catalonia, Spain

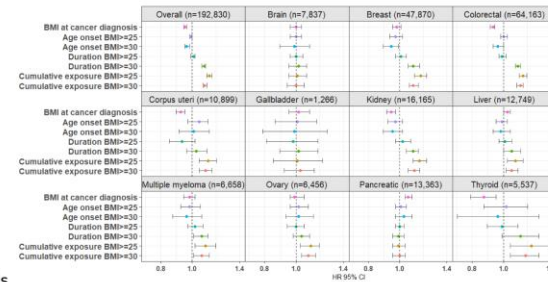
Background

Excessive adiposity and body weight are established risk factors for several types of cancer and have also been associated with increased overall and cancer-specific mortality. This study aims to investigate the association between early adulthood adiposity trajectory and risk of all-cause mortality in individuals diagnosed with 11 obesity-related cancers using data from SIDIAP database mapped to OMOP.

Result 1: Characteristics of the individuals with obesity-related cancer diagnosis



Result 2: Association between BMI-derived exposures and all-cause mortality per an increase of 1-SD in individuals diagnosed with obesity-related cancer overall and by each cancer type



Methods

Data source

SIDIAP (2010–2023) - Catalonia, Spain

Overweight/obesity exposures

Age at Onset
Duration (years)
Cumulative exposure
BMI at diagnosis

Population

18–40 years: BMI trajectories
41–90 years: Obesity-related cancer diagnosis

Outcome

All-cause mortality

Data processing

Multilevel time raster imputation (BMI trajectories)
Multiple imputation (smoking status, alcohol risk, MEDEA deprivation index)

Analysis

Cox models (HR, 95% CI)
Adjusted + Stratified
Pooled results - Rubin's rule

Limitation: Use of multiple imputed BMI measurements due to the extensive 13-year follow-up. Each participant's BMI values were imputed using their existing measurements and those of similar individuals.



Mariane Alves, Agustina Giuliadori, Laura Pérez-Crespo, Elena Roel, Andrea Pistillo, Irene López-Sánchez, Anna Palomar-Cros, Talita Duarte-Salles





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



Find your workgroup.

Fuel our mission.

ohdsi.org/workgroups