Europe Symposium Review & Reflections

OHDSI Community Call

May 5, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

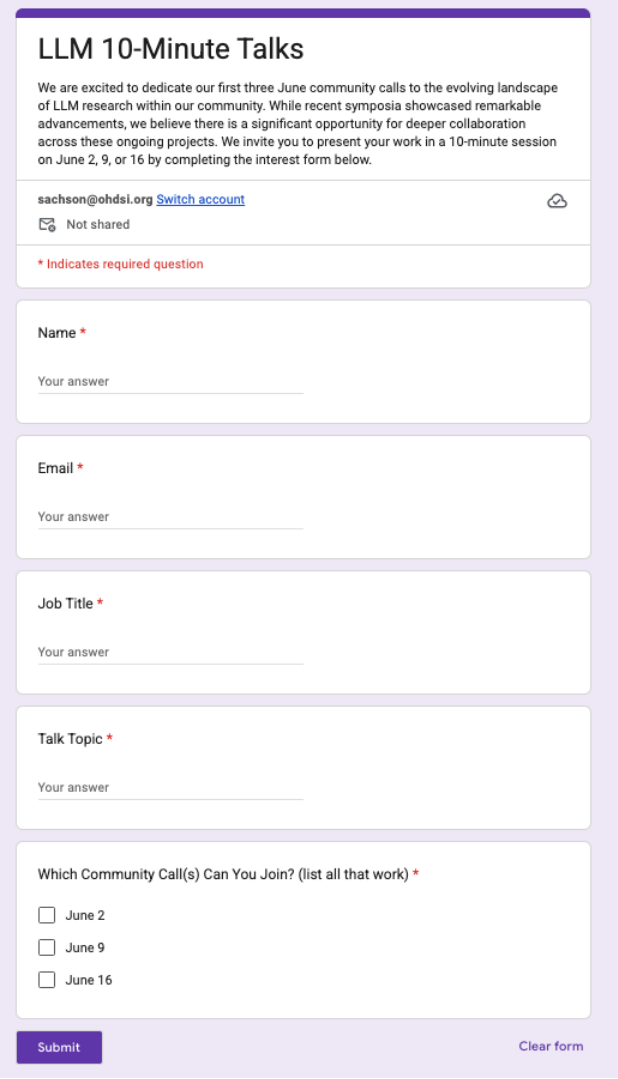
Upcoming Community Calls

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

June: LLM Research Presentations

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

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



LLM 10-Minute Talks

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

sachson@ohdsi.org Switch account

Not shared

* Indicates required question

Name *

Your answer

Email *

Your answer

Job Title *

Your answer

Talk Topic *

Your answer

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

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

Submit

Clear form

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Corina Konstantinou, Georgia Soursou, Samuel Abimbola, Pantelis Charisiadis, Angelos Kyriacou, Theofano Modestou, Michalis Tornaritis, Charalambos Hadjigeorgiou, Agapios Agapiou, Efstathios A. Elia, George Milis, Alexis Kyriacou, Lygia Eleftheriou, Zoi Tsimtsiou, Pantelis Natsiavas, Or Duek, Idan Menashe, Nathalia Bilenko, Itamar Grotto, Enkeleint A. Mechili, Mònica Guxens, Costas A. Christophi, Constantinos Deltas, and Konstantinos C. Makris** on the recent publication of **Designing a children's health exposomics study protocol: The CHILDREN_FIRST multi-country prospective cohort using multi-omics and personalized prevention approaches** in *PLOS One*.

PLOS One

STUDY PROTOCOL

Designing a children's health exposomics study protocol: The CHILDREN_FIRST multi-country prospective cohort using multi-omics and personalized prevention approaches

Corina Konstantinou[1], Georgia Soursou[1], Samuel Abimbola[1], Pantelis Charisiadis[1], Angelos Kyriacou[1], Theofano Modestou[2], Michalis Tornaritis[3], Charalambos Hadjigeorgiou[3], Agapios Agapiou[4], Efstathios A. Elia[4], George Milis[5], Alexis Kyriacou[5], Lygia Eleftheriou[6], Zoi Tsimtsiou[7], Pantelis Natsiavas[8], Or Duek[9], Idan Menashe[9], Nathalia Bilenko[9], Itamar Grotto[9], Enkeleint A. Mechili[10], Mònica Guxens[11,12,13,14,15], Costas A. Christophi[1,16], Constantinos Deltas[2,17¤], Konstantinos C. Makris[1,16*]

1 Cyprus International Institute for Environmental and Public Health, School of Health Sciences, Cyprus University of Technology, Limassol, Cyprus, **2** biobank.cy, Centre of Excellence in Biobanking and Biomedical Research, University of Cyprus, Nicosia, Cyprus, **3** Research and Education Institute of Child Health, Nicosia, Cyprus, **4** Department of Chemistry, University of Cyprus, Nicosia, Cyprus, **5** PHOEBE Research and Innovation LTD, Nicosia, Cyprus, **6** CP. Food Lab, Nicosia, Cyprus, **7** Department of Hygiene, Social-Preventive Medicine and Medical Statistics, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece, **8** Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, **9** Ben-Gurion University of the Negev, Beersheba, Israel, **10** Department of Healthcare, Faculty of Health Sciences, University of Vlora, Vlora, Albania, **11** ISGlobal, Barcelona, Spain, **12** ICREA, Barcelona, Spain, **13** Universitat Pompeu Fabra, Barcelona, Spain, **14** Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain, **15** Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC, University Medical Centre, Rotterdam, the Netherlands, **16** Department of Rehabilitation Sciences, School of Health Sciences, Cyprus University of Technology, Limassol, Cyprus, **17** School of Medicine, University of Cyprus, Nicosia, Cyprus

¤ Current address: Department of Basic and Clinical Sciences, University of Nicosia Medical School, Nicosia, Cyprus
* konstantinos.makris@cut.ac.cy

Check for updates

OPEN ACCESS

Citation: Konstantinou C, Soursou G, Abimbola S, Charisiadis P, Kyriacou A, Modestou T, et al. (2026) Designing a children's health exposomics study protocol: The CHILDREN_FIRST multi-country prospective cohort using multi-omics and personalized prevention approaches. PLoS One 21(4): e0326641. https://doi.org/10.1371/journal.pone.0326641

Editor: Elma Izze Da Silva Magalhães, Federal University of Santa Maria: Universidade Federal de Santa Maria, BRAZIL

Received: June 10, 2025

Accepted: April 1, 2026

Published: April 27, 2026

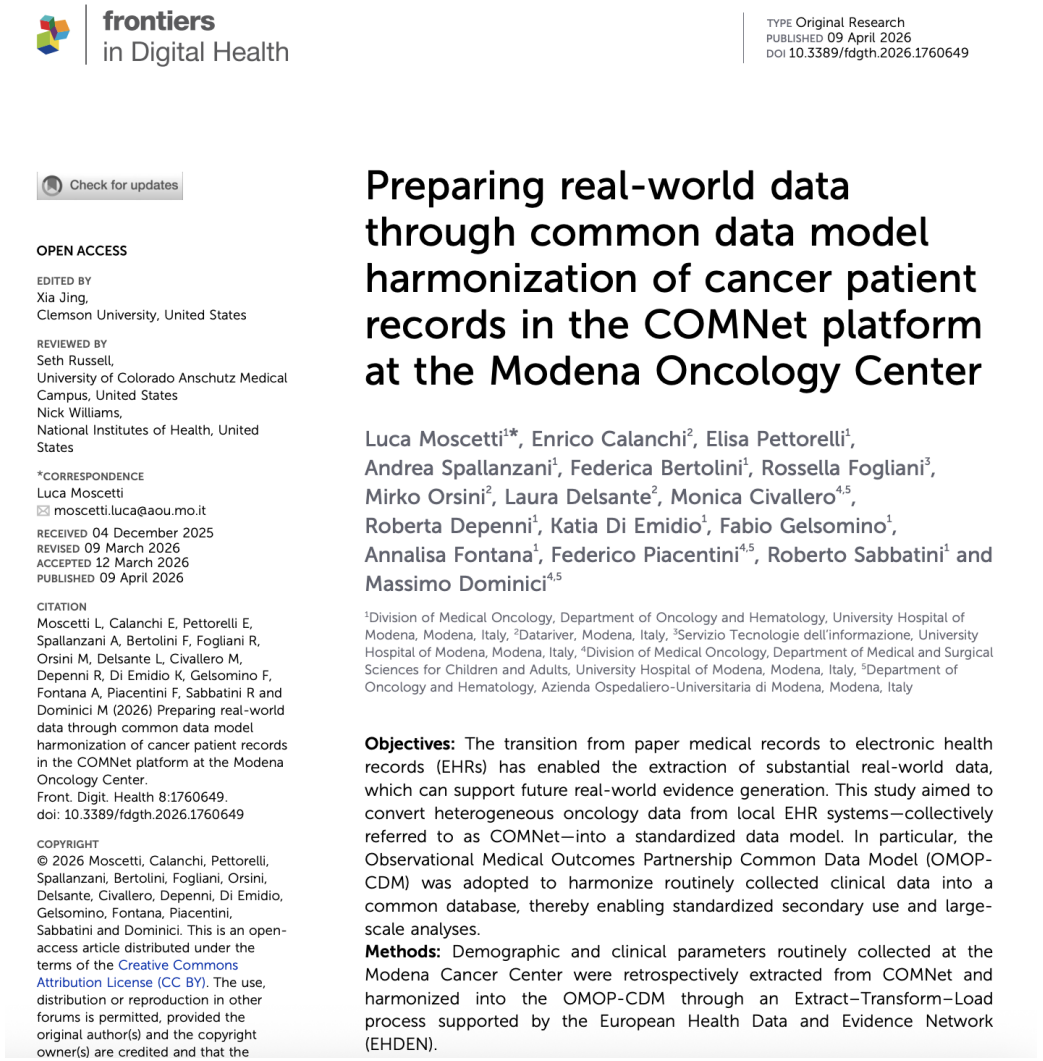
Data availability statement: No datasets were generated or analysed during the current study.

Abstract

Non-communicable diseases (NCDs) account for ~71% of all deaths globally, including 15 million premature deaths each year (deaths between 30–69 years of age). Instead of waiting until disease manifestation, focusing on the origins of NCDs during childhood offers a critical window of disease prevention and control. The CHILDREN_FIRST international cohort observatory study aims to investigate how the spatio-temporal evolution of the children's exposome profiles in the Mediterranean region influences early-life programming of chronic disease risk during the critical window of susceptibility in primary school years (6–11 years of age). The study protocol adopts the human exposome framework integrated with a personalized prevention approach, using multi-omics platforms and advanced machine learning algorithms implemented across Mediterranean countries, namely Cyprus, Greece, and Albania. The cohort will consist of children enrolled in the first grade of primary

OHDSI Shoutouts!

Congratulations to the team of **Luca Moscetti, Enrico Calanchi, Elisa Pettorelli, Andrea Spallanzani, Federica Bertolini, Rossella Fogliani, Mirko Orsini, Laura Delsante, Monica Civallero, Roberta Depenni, Katia Di Emidio, Fabio Gelsomino, Annalisa Fontana, Federico Piacentini, Roberto Sabbatini, and Massimo Dominici** on the recent publication of **Preparing real-world data through common data model harmonization of cancer patient records in the COMNet platform at the Modena Oncology Center** in *Frontiers in Digital Health*.



frontiers in Digital Health

TYPE Original Research
PUBLISHED 09 April 2026
DOI 10.3389/fdgth.2026.1760649

Check for updates

OPEN ACCESS

EDITED BY
Xia Jing,
Clemson University, United States

REVIEWED BY
Seth Russell,
University of Colorado Anschutz Medical Campus, United States
Nick Williams,
National Institutes of Health, United States

*CORRESPONDENCE
Luca Moscetti
moscetti.luca@aou.mo.it

RECEIVED 04 December 2025
REVISED 09 March 2026
ACCEPTED 12 March 2026
PUBLISHED 09 April 2026

CITATION
Moscetti L, Calanchi E, Pettorelli E, Spallanzani A, Bertolini F, Fogliani R, Orsini M, Delsante L, Civallero M, Depenni R, Di Emidio K, Gelsomino F, Fontana A, Piacentini F, Sabbatini R and Dominici M (2026) Preparing real-world data through common data model harmonization of cancer patient records in the COMNet platform at the Modena Oncology Center.
Front. Digit. Health 8:1760649.
doi: 10.3389/fdgth.2026.1760649

Preparing real-world data through common data model harmonization of cancer patient records in the COMNet platform at the Modena Oncology Center

Luca Moscetti[1]*, Enrico Calanchi[2], Elisa Pettorelli[1], Andrea Spallanzani[1], Federica Bertolini[1], Rossella Fogliani[3], Mirko Orsini[2], Laura Delsante[2], Monica Civallero[4,5], Roberta Depenni[1], Katia Di Emidio[1], Fabio Gelsomino[1], Annalisa Fontana[1], Federico Piacentini[4,5], Roberto Sabbatini[1] and Massimo Dominici[4,5]

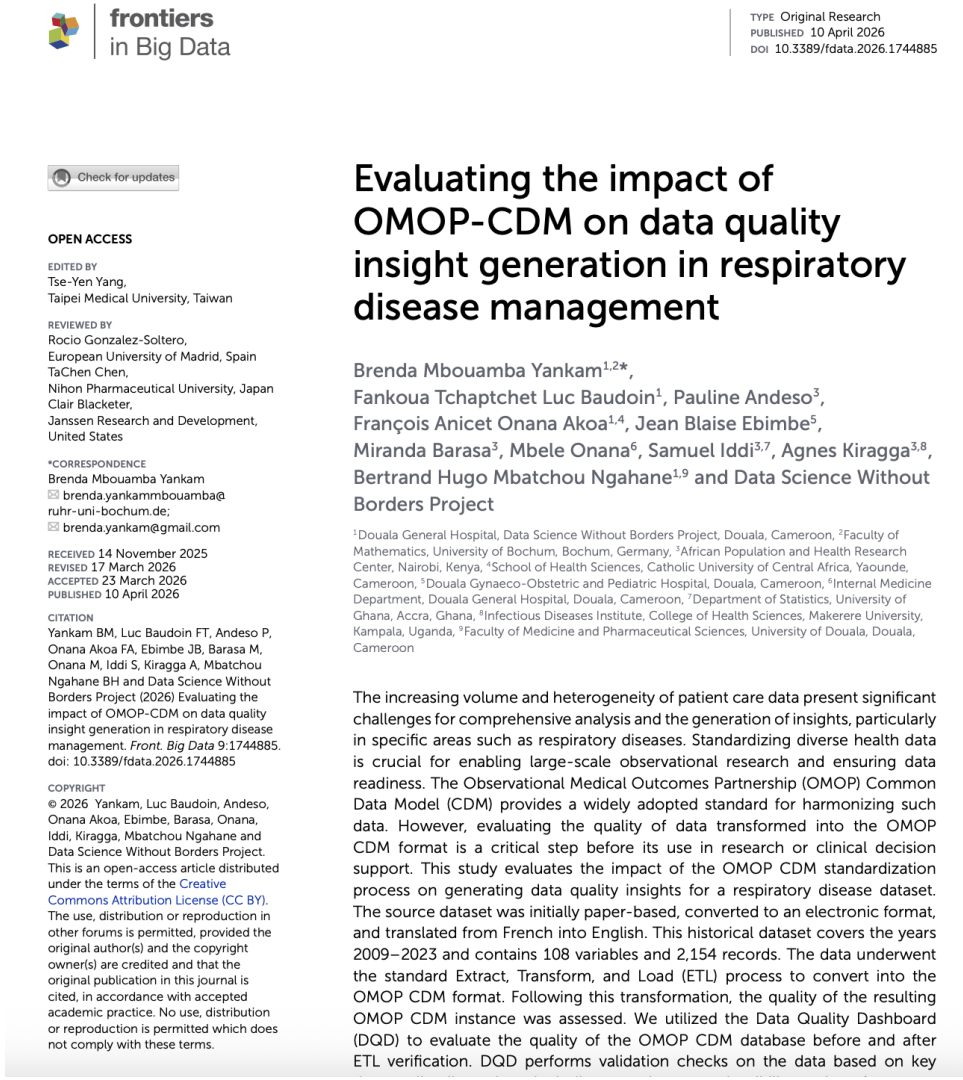
[1]Division of Medical Oncology, Department of Oncology and Hematology, University Hospital of Modena, Modena, Italy, [2]Datariver, Modena, Italy, [3]Servizio Tecnologie dell'informazione, University Hospital of Modena, Modena, Italy, [4]Division of Medical Oncology, Department of Medical and Surgical Sciences for Children and Adults, University Hospital of Modena, Modena, Italy, [5]Department of Oncology and Hematology, Azienda Ospedaliero-Universitaria di Modena, Modena, Italy

Objectives: The transition from paper medical records to electronic health records (EHRs) has enabled the extraction of substantial real-world data, which can support future real-world evidence generation. This study aimed to convert heterogeneous oncology data from local EHR systems—collectively referred to as COMNet—into a standardized data model. In particular, the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) was adopted to harmonize routinely collected clinical data into a common database, thereby enabling standardized secondary use and large-scale analyses.

Methods: Demographic and clinical parameters routinely collected at the Modena Cancer Center were retrospectively extracted from COMNet and harmonized into the OMOP-CDM through an Extract–Transform–Load process supported by the European Health Data and Evidence Network (EHDEN).

OHDSI Shoutouts!

Congratulations to the team of **Brenda Mbouamba Yankam, Fankoua Tchaptchet Luc Baudoin, Pauline Andeso, François Anicet Onana Akoa, Jean Blaise Ebimbe, Miranda Barasa, Mbele Onana, Samuel Iddi, Agnes Kiragga, Bertrand Hugo Mbatchou Ngahane and the Data Science Without Borders Project** on the recent publication of **Evaluating the impact of OMOP-CDM on data quality insight generation in respiratory disease management** in *Frontiers in Big Data*.



frontiers in Big Data

TYPE Original Research
PUBLISHED 10 April 2026
DOI 10.3389/fdata.2026.1744885

Check for updates

OPEN ACCESS

EDITED BY
Tse-Yen Yang,
Taipei Medical University, Taiwan

REVIEWED BY
Rocio Gonzalez-Soltero,
European University of Madrid, Spain
TaChen Chen,
Nihon Pharmaceutical University, Japan
Clair Blacketer,
Janssen Research and Development, United States

*CORRESPONDENCE
Brenda Mbouamba Yankam
brenda.yankammbouamba@ruhr-uni-bochum.de;
brenda.yankam@gmail.com

RECEIVED 14 November 2025
REVISED 17 March 2026
ACCEPTED 23 March 2026
PUBLISHED 10 April 2026

CITATION
Yankam BM, Luc Baudoin FT, Andeso P, Onana Akoa FA, Ebimbe JB, Barasa M, Onana M, Iddi S, Kiragga A, Mbatchou Ngahane BH and Data Science Without Borders Project (2026) Evaluating the impact of OMOP-CDM on data quality insight generation in respiratory disease management. *Front. Big Data* 9:1744885. doi: 10.3389/fdata.2026.1744885

COPYRIGHT
© 2026 Yankam, Luc Baudoin, Andeso, Onana Akoa, Ebimbe, Barasa, Onana, Iddi, Kiragga, Mbatchou Ngahane and Data Science Without Borders Project. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Evaluating the impact of OMOP-CDM on data quality insight generation in respiratory disease management

Brenda Mbouamba Yankam[1,2]*, Fankoua Tchaptchet Luc Baudoin[1], Pauline Andeso[3], François Anicet Onana Akoa[1,4], Jean Blaise Ebimbe[5], Miranda Barasa[3], Mbele Onana[6], Samuel Iddi[3,7], Agnes Kiragga[3,8], Bertrand Hugo Mbatchou Ngahane[1,9] and Data Science Without Borders Project

[1]Douala General Hospital, Data Science Without Borders Project, Douala, Cameroon, [2]Faculty of Mathematics, University of Bochum, Bochum, Germany, [3]African Population and Health Research Center, Nairobi, Kenya, [4]School of Health Sciences, Catholic University of Central Africa, Yaounde, Cameroon, [5]Douala Gynaeco-Obstetric and Pediatric Hospital, Douala, Cameroon, [6]Internal Medicine Department, Douala General Hospital, Douala, Cameroon, [7]Department of Statistics, University of Ghana, Accra, Ghana, [8]Infectious Diseases Institute, College of Health Sciences, Makerere University, Kampala, Uganda, [9]Faculty of Medicine and Pharmaceutical Sciences, University of Douala, Douala, Cameroon

The increasing volume and heterogeneity of patient care data present significant challenges for comprehensive analysis and the generation of insights, particularly in specific areas such as respiratory diseases. Standardizing diverse health data is crucial for enabling large-scale observational research and ensuring data readiness. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) provides a widely adopted standard for harmonizing such data. However, evaluating the quality of data transformed into the OMOP CDM format is a critical step before its use in research or clinical decision support. This study evaluates the impact of the OMOP CDM standardization process on generating data quality insights for a respiratory disease dataset. The source dataset was initially paper-based, converted to an electronic format, and translated from French into English. This historical dataset covers the years 2009–2023 and contains 108 variables and 2,154 records. The data underwent the standard Extract, Transform, and Load (ETL) process to convert into the OMOP CDM format. Following this transformation, the quality of the resulting OMOP CDM instance was assessed. We utilized the Data Quality Dashboard (DQD) to evaluate the quality of the OMOP CDM database before and after ETL verification. DQD performs validation checks on the data based on key

OHDSI Shoutouts!

Congratulations to the team of **Joseph Cronin, Olivia Wiper, Anthony Poncet, Keiran Tait, Bernard Cooke, Andrew Fry, Janie Baxter, and Robert Dürichen** on the recent publication of **ArcMAP – ML assisted medical concept mapping to accelerate NHS data standardization** in *Frontiers in Digital Health*.

frontiers in Digital Health

TYPE Original Research
PUBLISHED 14 April 2026
DOI 10.3389/fdgth.2026.1770903

OPEN ACCESS

EDITED BY
Angus Roberts,
King's College London, United Kingdom

REVIEWED BY
Markus Knott,
Klinikum Stuttgart, Germany
Komal Gilani,
Maastricht University, Netherlands

*CORRESPONDENCE
Joseph Cronin
joseph.cronin@arcturisdata.com

RECEIVED 18 December 2025
REVISED 04 March 2026
ACCEPTED 11 March 2026
PUBLISHED 14 April 2026

CITATION
Cronin J, Wiper O, Poncet A, Tait K, Cooke B, Fry A, Baxter J and Dürichen R (2026) ArcMAP – ML assisted medical concept mapping to accelerate NHS data standardization.
Front. Digit. Health 8:1770903.
doi: 10.3389/fdgth.2026.1770903

ArcMAP – ML assisted medical concept mapping to accelerate NHS data standardization

Joseph Cronin*, Olivia Wiper, Anthony Poncet, Keiran Tait, Bernard Cooke, Andrew Fry, Janie Baxter and Robert Dürichen

Arcturis Data, Kidlington, United Kingdom

The increasing use of electronic health records (EHRs) for real-world evidence (RWE) studies is hindered by substantial heterogeneity in data collection practices and local coding schemes across healthcare providers. Data standardization—particularly the mapping of locally defined medical concepts to standardized vocabularies—is therefore a critical but labour-intensive step, traditionally relying on extensive manual review by clinical experts. While a range of machine-learning (ML) approaches have been proposed to support medical concept mapping, their integration into practical, end-to-end workflows and their performance under real-world conditions remain insufficiently studied. In this work, we present ArcMAP, an end-to-end application that integrates a state-of-the-art biomedical representation model (BioLORD) into a human-in-the-loop workflow designed to streamline and accelerate medical concept mapping. ArcMAP provides a graphical user interface that enables clinical experts to efficiently review, validate, and correct automated mapping suggestions. A core component of the system is a continuous learning pipeline, in which expert feedback is systematically captured and used to update the underlying model, allowing ArcMAP to adapt to evolving coding practices and newly onboarded data sources. We conduct a comprehensive evaluation of ArcMAP across multiple deployment scenarios, including the impact of continuous fine-tuning, the onboarding of a new hospital, and a longitudinal real-world evaluation conducted over a two-month period using medication and laboratory test data from five UK-based NHS hospitals. Our results demonstrate the importance of domain-specific

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	9 am	Tidy R Programming with OMOP
Wednesday	10 am	Common Data Model
Thursday	7 am	Europe Community Call
Thursday	10 am	ATLAS/WebAPI
Thursday	10 am	Africa Chapter (ZOOM)
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	11 am	Clinical Trials
Friday	11:30 am	Steering Group
Friday	11 pm	China Chapter
Monday	9 am	Vaccine Vocabulary
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	10 am	CDM Survey

May Newsletter Is Available

OHDSI

The Journey Newsletter (May 2026)

Welcome to the May newsletter, which is filled with insights and outputs from our recent Europe Symposium as we carry that momentum toward a busy summer of collaboration. We are officially calling for submissions for the 2026 Global Symposium, so please ensure you share your work by the June 5 deadline to be part of an agenda that is now officially set. This issue also spotlights two outstanding community members and provides essential updates on our upcoming events, including the first-ever Latin America Symposium this July. #JoinTheJourney

Podcast: Europe Symposium, Global Research Opportunities at OHDSI 2026

In the May 2026 On The Journey podcast, Patrick Ryan and Craig Sachson discuss the 2026 Europe Symposium, including the weekend workshops and tutorials, as well as the wide-ranging research presented at the main conference. With the Global Symposium showcase submission date coming June 5, Patrick discusses the types of research he is hoping to see presented this October in New Jersey. *(If video does not appear, please click 'view this email in your browser.)*

Community Updates

Where Have We Been?

- **Europe Symposium:** The 2026 Europe Symposium recently concluded in Rotterdam, centering on the theme of "Continuous Collaboration for Living Evidence Generation." This year's event saw a record-breaking number of collaborator showcase submissions and a sold-out crowd on the SS Rotterdam; our thanks go out to the Erasmus MC team for hosting such a successful gathering.
- **Phenotype April:** Our community made significant strides in Phenotype April, where leaders from the Phenotype WG hosted focused sessions on live cohort building and evaluation using KEEPER. You can access all presentations further down in this newsletter or on our community calls page.
- **Global Showcase Agenda:** The 2026 Global Symposium agenda was announced, and it features a trio of community-developed plenaries, as well as our planned tutorials and workshop activities. Visit our event page to learn more about the sessions and activities planned for New Brunswick.

Where Are We Now?

- **Global Symposium Showcase Deadline:** 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. **The deadline to submit your brief report(s) for the Collaborator Showcase is June 5 at 8 pm ET.** More information about the collaborator showcase is available in this newsletter.
- **Maternal Health Fellowship Deadline:** The deadline to apply for the second OHDSI Maternal Health Fellowship is May 15. More details are available in this newsletter; you can apply here.
- **Latin America Symposium:** The first-ever OHDSI Latin America Symposium is coming to Salvador, Brazil, July 30-31, and there is still time to be part of the program. Abstracts for the collaborator showcase **are being accepted through May 17.** The agenda features workshops on OMOP essentials and network study opportunities. The organizing committee is seeking regional stakeholders; please contact Valentina Martufi or the Latin America team to assist.

Europe Symposium Focuses On Collaboration Opportunities For Evidence Generation; New OHDSI Europe Website Unveiled

The 2026 OHDSI Europe Symposium welcomed a record number of both participants and collaborator showcase submissions to Rotterdam, where the weekend theme was 'Continuous Collaboration for Living Evidence Generation'. Following two days of tutorials and workshops, the community joined together April 20 on the SS Rotterdam for a day of networking, sharing research, and looking ahead to the potential of a global healthcare surveillance system developed by the OHDSI community.

My Journey: Liesbet Peeters

In the latest installment of our "My Journey" series, Liesbet Peeters (Associate Professor at Hasselt University) shares how OHDSI provided the "vocabulary" she needed to tackle the complexities of global health data scale-up. From her deep dive into the community in 2018 to hosting the OHDSI Europe Symposium and launching OHDSI Belgium, Liesbet discusses why decentralized leadership and a "network of thinkers" are the keys to solving the wicked problems in healthcare. *(If video does not appear, please click 'view this email in your browser.)*

April Publications

Cheng W, Yu Z. Toward semantic interoperability of imaging and clinical data: reflections on the DICOM-OMOP integration framework. J Am Med Inform Assoc. 2026 Apr 1;33(4):939-940. doi: 10.1093/jamia/ocaf215. PMID: 41453141; PMCID: PMC13089545.

Kim C, Bu F, Blacketer C, Ostropolets A, Duarte-Salles T, Viernes B, Falconer T, Pistillo A, Li J, Yin C, Van Zandt M, Nagy P, Nishimura A, Minty E, You SC, Sawano M, Sawano S, Jeon JY, Aminorroaya A, Dhingra LS, Pedroso AF, Thangaraj P, Dorr DA, Pratt N, Man KKC, Lau WCY, Morales DR, Khera R, Schuemie MJ, Ryan PB, Hripcsak G, Krumholz HM, Suchard MA, Lu Y. Real-world evidence for comparative safety of second-line antihyperglycemic agents in older adults with type 2 diabetes. Nat Commun. 2026 Apr 4. doi: 10.1038/s41467-026-71307-0. Epub ahead of print. PMID: 41935054.

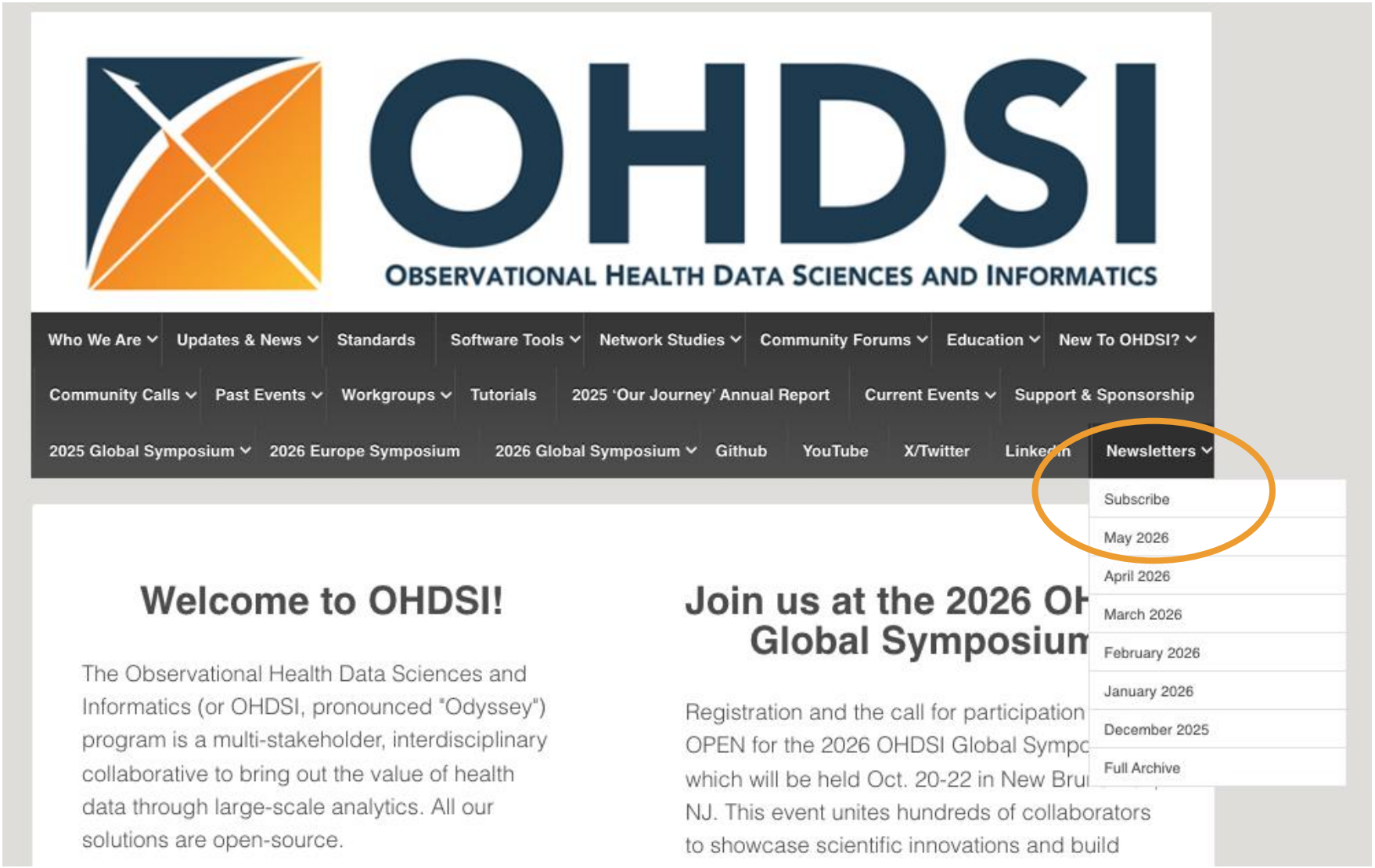
Pignatti F, El-Galaly TC, Kaiser M, Porkka K, Doeswijk R, Mol P, Rivera DR, Lerro CC, Rohr UP, Verpillat P, Valachis A, Trapani D, Di Maio M, Latino N, Cordoba R, Cherny N, Koopman M, Martins-Branco D, Pentheroudakis G, Postmus D. Assessing Overall Survival Benefits in Advanced Cancers: The Role of External Comparator Cohort Studies with Real-World Data. Clin Pharmacol Ther. 2026 Apr;119(4):1080-1087. doi: 10.1002/cpt.70213. Epub 2026 Feb 11. PMID: 41669939; PMCID: PMC12997495.

Spotnitz M, Giannini J, Clark E, Ostchega Y, Litwin TR, Berman L. Assessing data quality of rheumatoid and psoriatic arthritis patients in the *All of Us* Research Program. JAMIA Open. 2026 Mar 7;9(2):ooag028. doi: 10.1093/jamiaopen/ooag028. PMID: 41822200; PMCID: PMC12978248.

Du M, Prats-Uribe A, Mercadé-Besora N, Lopez-Guell K, Guo Y, Alcalde-Herraiz M, Chen X, Delmestri A, Man WY, Duarte-Salles T, Palomar A, Giuliodori A, Brađašević E, Jezidžić A, Bräuner E, Bruun S, Verhamme K, Mosseveld M, Brash JT, Vojinovic D, Kaczmarczyk I, Mendez A, Rijnbeek P, Prieto-Alhambra D, Burn E, Català M. CohortCharacteristics: an R package for population characterisation in observational studies using the OMOP common data model. Eur J Epidemiol. 2026 Apr 3. doi: 10.1007/s10654-025-01352-4. Epub ahead of print. PMID: 41933108.

Moscetti L, Calanchi E, Pettorelli E, Spallanzani A, Bertolini F, Fogliani R, Orsini M, Delsante L, Civallero M, Depenni R, Di Emidio K, Gelsomino F, Fontana A, Piacentini F, Sabbatini R, Dominici M. Preparing real-world data through common data model harmonization of cancer patient records in the COMNet platform at the Modena Oncology Center. Front Digit Health. 2026 Apr

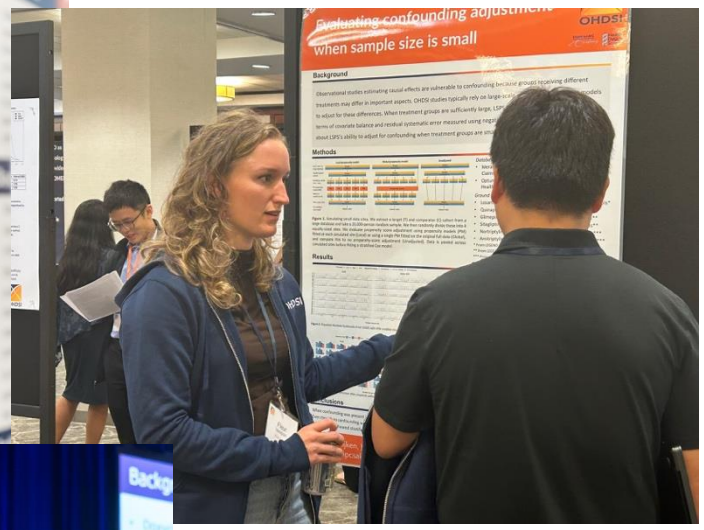
May Newsletter Is Available

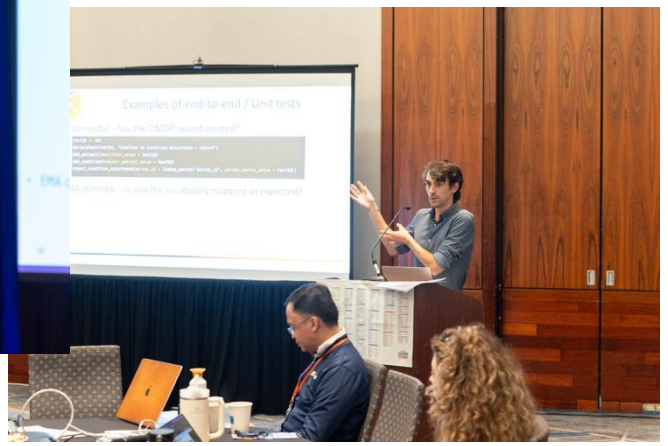


2026 OHDSI Global Symposium

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

The submission deadline is June 5 at 8 pm ET.





ONE MONTH AWAY

ohdsi.org/OHDSI2026

2026 OHDSI Global Symposium

Registration is OPEN for the **2026 OHDSI Global Symposium**, which will be held Oct. 20-22 in New Brunswick, N.J., USA.

Oct. 20: Tutorials
Oct. 21: Plenaries, Showcase
Oct. 22: Workgroup Activities

ohdsi.org/OHDSI2026

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

2026 OHDSI Global Symposium

There are opportunities to be both a **sponsor** and an **exhibitor** at the Global Symposium.

Please reach out to **symposium@OHDSI.org** for more information.

ohdsi.org/OHDSI2026

Maternal Fellowship Deadline: May 15

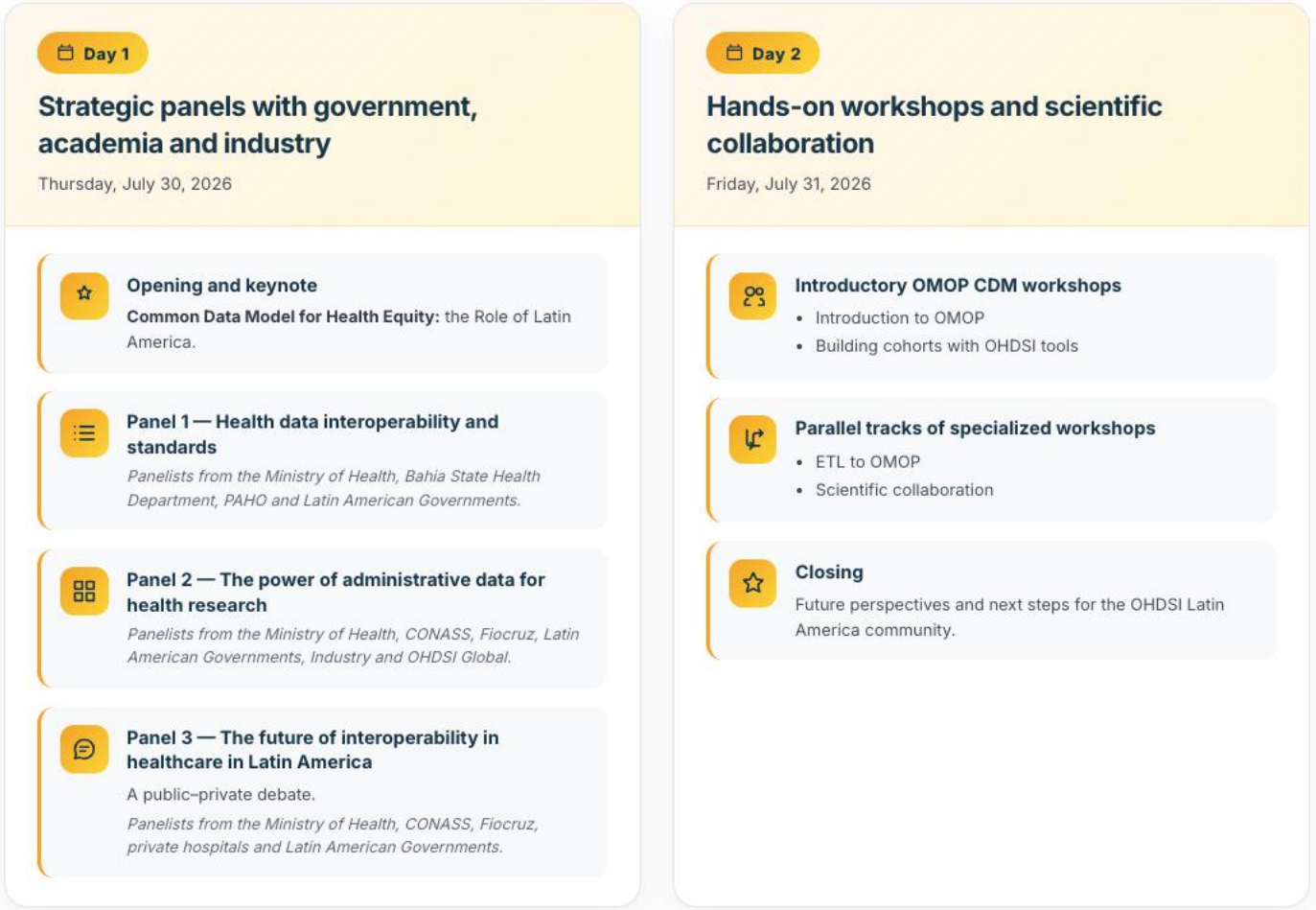
The second **OHDSI Maternal Health Fellowship** is designed to train clinical investigators for improved maternal and neonatal care. This fellowship offers three key components: **Career Development, Practice, and Networking**.

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.

First Latin America Symposium – July 30-31

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

The submission deadline for the showcase is May 17.



ohdsilatam.org

EUROPE

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

Take your OMOP implementation to the next level

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

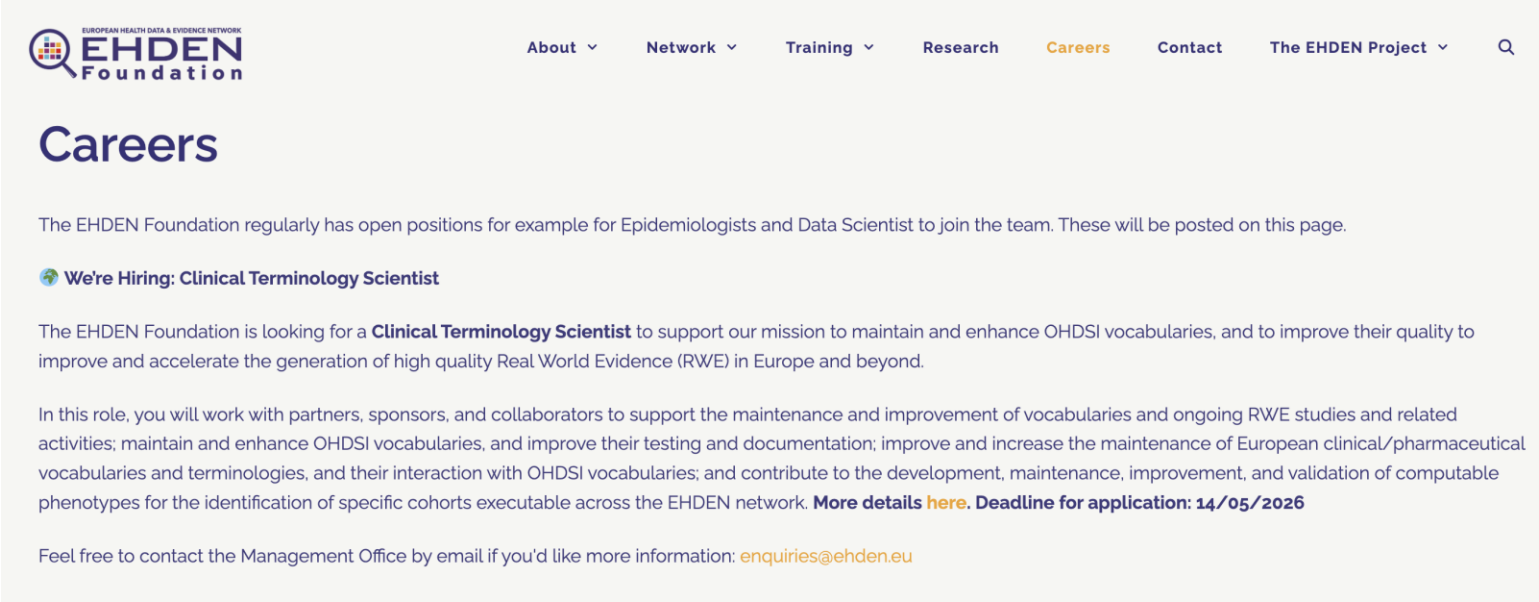
What to expect

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

Practical

🗓️ 7–9 September 2026
📍 Rotterdam

Opening: Clinical Terminology Scientist



EHDEN Foundation

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

Careers

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

We're Hiring: Clinical Terminology Scientist

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

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

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

EUROPEAN HEALTH DATA & EVIDENCE NETWORK
EHDEN Foundation

JOB DESCRIPTION AND SELECTION CRITERIA

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

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

Application deadline for this position is May 14, 2026.

ehden.eu/careers

#OHDSISocialShowcase This Week

Monday

Anatomical Location Auto-check and Standardization for Medical Images

(Qingrui Wang, Teri Sippel Schmidt, Paul Nagy, Blake E. Dewey)

Automated Anatomical Identification and Standardization for Medical Images

PRESENTER: Qingrui (Carrie) Wang

INTRO:

- Imaging technologists record the scanned body area as DICOM *(0018,0015) Body Part Examined* tag.
- For secondary use, this label is often missing/inaccurate, so anatomy-based cohort queries miss studies or include the wrong ones.
- The OMOP Imaging Extension enables series-level cohorting, which depends on reliable anatomy labels.
- Goal: **Validate and standardize anatomical labels automatically** and store the result for OMOP-based analytics.

METHODS

Pipeline

1. **Segment** organs/structures in CT/MR with TotalSegmentator.
2. **Compare** segmentation-derived anatomy to the DICOM tag (0018,0015) Body Part Examined.
3. **Validate**: mark match as correct; flag mismatch/absent for review.
4. **Standardize**: map validated anatomy to SNOMED CT concept IDs/terms.
5. **Augment**: keep the additional structures found by the model for future, finer-grained cohorting.

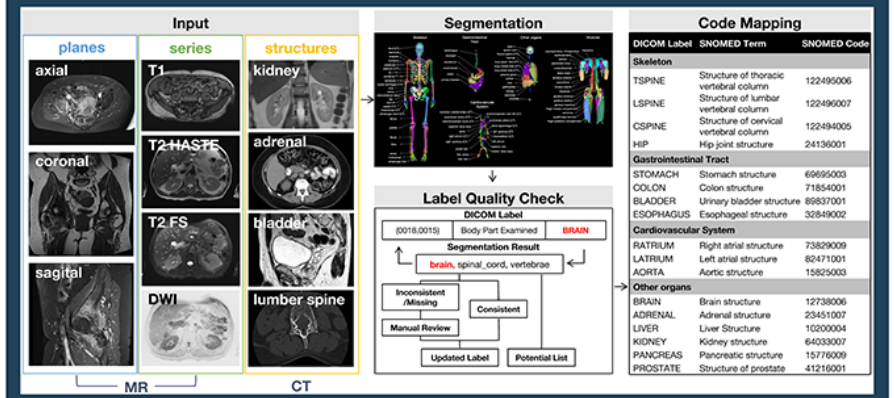
Data

- Public TCIA sample: 53 series (13 CT, 40 MR) from multiple institutions/scanners across ~16 anatomic regions and diverse sequences/planes/positions.

Outputs

- Validated primary anatomy per series.
- SNOMED concept code & term.
- List of additional detected anatomies (by modality).
- QC flags for failure/mismatch.

Automatically validate *Body Part Examined* in CT/MR with TotalSegmentator → SNOMED mapping



Take a picture to download the full abstract

Qingrui (Carrie) Wang, Teri Sippel Schmidt, Paul Nagy, Blake E. Dewey

RESULTS:

- Consistency: 50/53 series matched between DICOM anatomy and segmentation.
- Auto-labeling: 2 unlabeled series received reasonable primary anatomies from the segmenter.
- Flagged: 1 series was flagged (model miss due to poor image quality).
- Richer context: additional structures per series: CT ~70.0 (avg), MR ~15.4 (avg).

Table 1. Automated anatomical identification and standardization examples (red indicates a match with Body Part Examined)

Case	Modality	Series Description	Study Description	Status	SNOMED concept	SNOMED value
1	MR	COR T2 MRCP NAV	MR ABDOMEN NO CONTRAST	Needs Review	Pancreatic structure	15776009
2	CT	BLAD DELAY	CT CHEST ABDOMEN PELVIS W	match	Urinary bladder structure	89837001
3	MR	COR T1 TFE POST MPR	MR BRAIN WOW CONTRAST	match	Brain structure	12738006
4	MR	T2 SAG SM FOV	Prostate Fused-MRI-Pathology	match	Structure of prostate	41216001

Case	Body Part Examined	Detected Structures by TotalSegmentator
1	PANCREAS	colon, duodenum, femur_left, femur_right, gallbladder, gluteus_medius_left, gluteus_medius_right, gluteus_minimus_left, gluteus_minimus_right, hip_left, hip_right, iliac_artery_left, iliac_artery_right, iliac_vena_left, iliac_vena_right, iliopsoas_left, iliopsoas_right, intervertebral_discs, liver, lung_left, lung_right, sacrum, small_bowel, spleen, stomach, urinary_bladder, vertebrae
2	BLADDER	colon, femur_left, femur_right, gluteus_maximus_left, gluteus_maximus_right, gluteus_medius_left, gluteus_medius_right, gluteus_minimus_left, gluteus_minimus_right, hip_left, hip_right, iliac_artery_left, iliac_artery_right, iliac_vena_left, iliac_vena_right, iliopsoas_left, iliopsoas_right, inferior_vena_cava, prostate, sacrum, urinary_bladder
3	BRAIN	brain, intervertebral_discs, spinal_cord, vertebrae
4	PROSTATE	colon, femur_right, hip_left, hip_right, humerus_right, iliac_vena_right, prostate, urinary_bladder

CONCLUSIONS:

- A simple, reproducible workflow cleans and standardizes anatomy metadata in routine imaging.
- Improves cohort precision in OMOP-based studies and enables feature-level imaging analytics.

FUTURE STEPS:

- Scale to larger, multi-site datasets.
- Extend available labels: store multiple standardized labels per series with finer granularity.
- Integrate with OMOP Imaging Extension for broader, high-recall cohort queries.

JOHNS HOPKINS MEDICINE

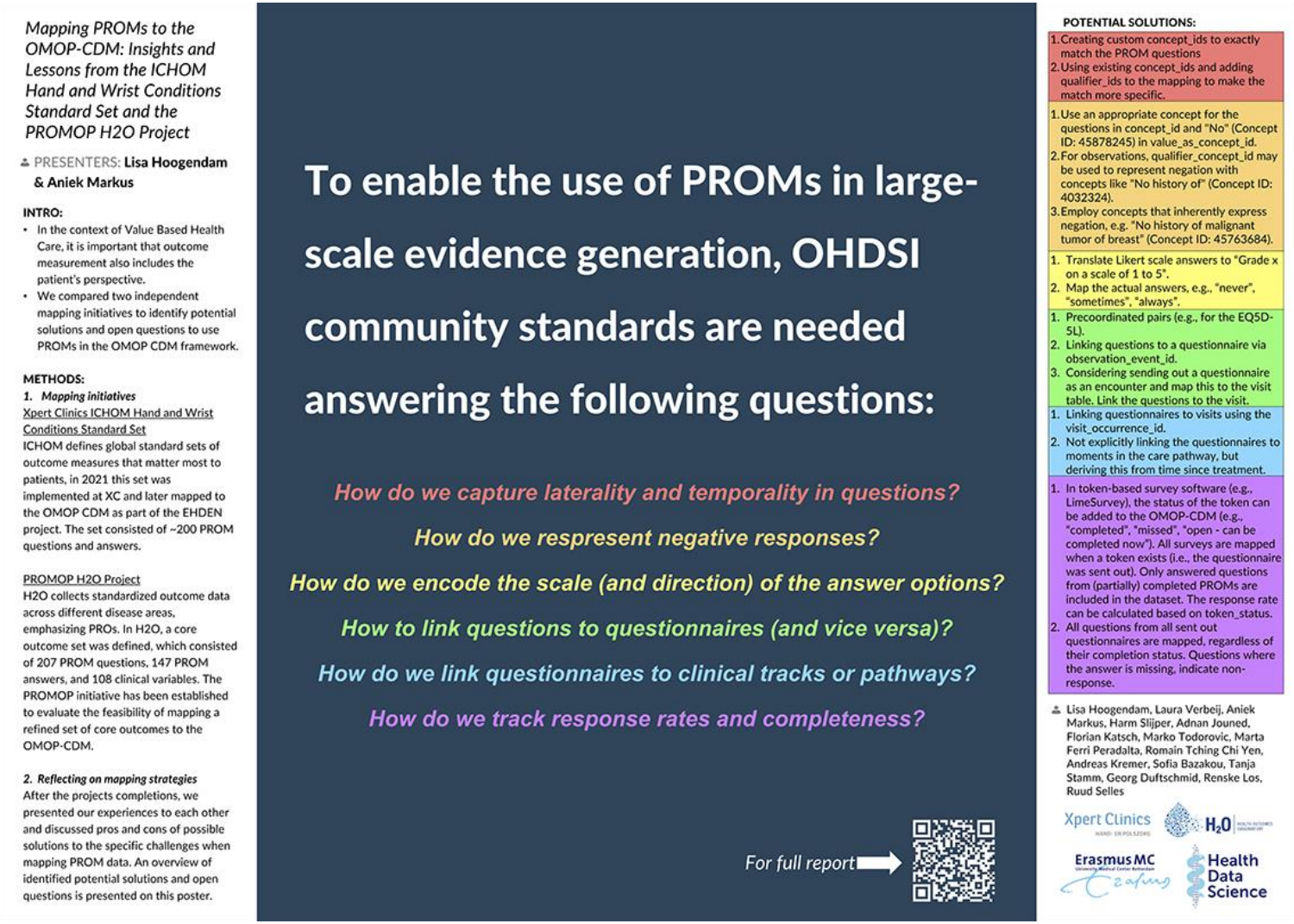
OHDSI

#OHDSISocialShowcase This Week

Tuesday

Mapping PROMs to the OMOP-CDM: Insights and Lessons from the ICHOM Hand and Wrist Conditions Standard Set and the PROMOP H2O Project

(Lisa Hoogendam, Laura Verbeij, Aniek Markus, Harm Slijper, Adnan Jouned, Florian Katsch, Marko Todorovic, Marta Ferri Peradalta, Romain Tching Chi Yen, Andreas Kremer, Sofia Bazakou,, Tanja Stamm, Georg Duftschmid, Renske Los, Ruud Selles)



Mapping PROMs to the OMOP-CDM: Insights and Lessons from the ICHOM Hand and Wrist Conditions Standard Set and the PROMOP H2O Project

PRESENTERS: **Lisa Hoogendam & Aniek Markus**

INTRO:

- In the context of Value Based Health Care, it is important that outcome measurement also includes the patient's perspective.
- We compared two independent mapping initiatives to identify potential solutions and open questions to use PROMs in the OMOP CDM framework.

METHODS:

1. Mapping initiatives

Xpert Clinics ICHOM Hand and Wrist Conditions Standard Set

ICHOM defines global standard sets of outcome measures that matter most to patients, in 2021 this set was implemented at XC and later mapped to the OMOP CDM as part of the EHDEN project. The set consisted of ~200 PROM questions and answers.

PROMOP H2O Project

H2O collects standardized outcome data across different disease areas, emphasizing PROs. In H2O, a core outcome set was defined, which consisted of 207 PROM questions, 147 PROM answers, and 108 clinical variables. The PROMOP initiative has been established to evaluate the feasibility of mapping a refined set of core outcomes to the OMOP-CDM.

2. Reflecting on mapping strategies

After the projects completions, we presented our experiences to each other and discussed pros and cons of possible solutions to the specific challenges when mapping PROM data. An overview of identified potential solutions and open questions is presented on this poster.

To enable the use of PROMs in large-scale evidence generation, OHDSI community standards are needed answering the following questions:

How do we capture laterality and temporality in questions?

How do we respresent negative responses?

How do we encode the scale (and direction) of the answer options?

How to link questions to questionnaires (and vice versa)?

How do we link questionnaires to clinical tracks or pathways?

How do we track response rates and completeness?

For full report

POTENTIAL SOLUTIONS:

1. Creating custom concept_ids to exactly match the PROM questions
2. Using existing concept_ids and adding qualifier_ids to the mapping to make the match more specific.

1. Use an appropriate concept for the questions in concept_id and "No" (Concept ID: 45878245) in value_as_concept_id.
2. For observations, qualifier_concept_id may be used to represent negation with concepts like "No history of" (Concept ID: 4032324).
3. Employ concepts that inherently express negation, e.g. "No history of malignant tumor of breast" (Concept ID: 45763684).

1. Translate Likert scale answers to "Grade x on a scale of 1 to 5".
2. Map the actual answers, e.g., "never", "sometimes", "always".

1. Precoordinated pairs (e.g., for the EQ5D-5L).
2. Linking questions to a questionnaire via observation_event_id.
3. Considering sending out a questionnaire as an encounter and map this to the visit table. Link the questions to the visit.

1. Linking questionnaires to visits using the visit_occurrence_id.
2. Not explicitly linking the questionnaires to moments in the care pathway, but deriving this from time since treatment.

1. In token-based survey software (e.g., LimeSurvey), the status of the token can be added to the OMOP-CDM (e.g., "completed", "missed", "open - can be completed now"). All surveys are mapped when a token exists (i.e., the questionnaire was sent out). Only answered questions from (partially) completed PROMs are included in the dataset. The response rate can be calculated based on token_status.
2. All questions from all sent out questionnaires are mapped, regardless of their completion status. Questions where the answer is missing, indicate non-response.

Lisa Hoogendam, Laura Verbeij, Aniek Markus, Harm Slijper, Adnan Jouned, Florian Katsch, Marko Todorovic, Marta Ferri Peradalta, Romain Tching Chi Yen, Andreas Kremer, Sofia Bazakou, Tanja Stamm, Georg Duftschmid, Renske Los, Ruud Selles

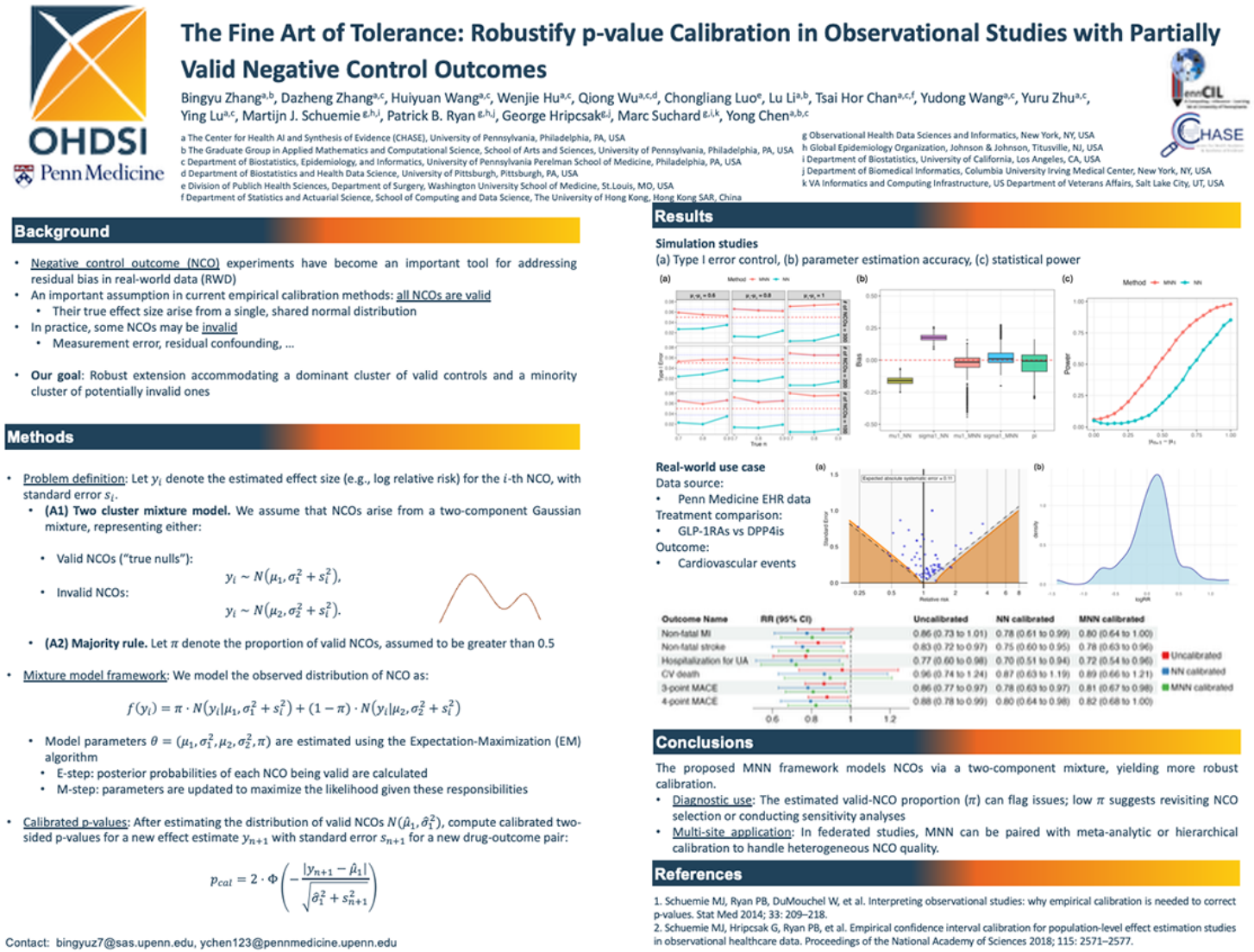
Xpert Clinics H2O Erasmus MC Health Data Science

#OHDSISocialShowcase This Week

Wednesday

The Fine Art of Tolerance: Robustify p-value Calibration in Observational Studies with Partially Valid Negative Control Outcomes

(Bingyu Zhang, Dazheng Zhang, Huiyuan Wang, Wenjie Hu, Qiong Wu, Chongliang Luo, Lu Li, Tsai Hor Chan, Yudong Wang, Martijn Schuemie, Patrick Ryan, George Hripcsak, Marc Suchard, Yong Chen)



The Fine Art of Tolerance: Robustify p-value Calibration in Observational Studies with Partially Valid Negative Control Outcomes

Bingyu Zhang[a,b], Dazheng Zhang[a,c], Huiyuan Wang[a,c], Wenjie Hu[a,c], Qiong Wu[a,c,d], Chongliang Luo[e], Lu Li[a,b], Tsai Hor Chan[a,c,f], Yudong Wang[a,c], Yuru Zhu[a,c], Ying Lu[a,c], Martijn J. Schuemie[g,h,i], Patrick B. Ryan[g,h,i], George Hripcsak[g,i], Marc Suchard[g,i,k], Yong Chen[a,b,c]

a The Center for Health AI and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA
b The Graduate Group in Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA
c Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
d Department of Biostatistics and Health Data Science, University of Pittsburgh, Pittsburgh, PA, USA
e Division of Publich Health Sciences, Department of Surgery, Washington University School of Medicine, St.Louis, MO, USA
f Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, Hong Kong SAR, China
g Observational Health Data Sciences and Informatics, New York, NY, USA
h Global Epidemiology Organization, Johnson & Johnson, Titusville, NJ, USA
i Department of Biostatistics, University of California, Los Angeles, CA, USA
j Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA
k VA Informatics and Computing Infrastructure, US Department of Veterans Affairs, Salt Lake City, UT, USA

Background

- Negative control outcome (NCO) experiments have become an important tool for addressing residual bias in real-world data (RWD)
- An important assumption in current empirical calibration methods: all NCOs are valid
  - Their true effect size arise from a single, shared normal distribution
- In practice, some NCOs may be invalid
  - Measurement error, residual confounding, ...
- **Our goal**: Robust extension accommodating a dominant cluster of valid controls and a minority cluster of potentially invalid ones

Methods

- Problem definition: Let y_i denote the estimated effect size (e.g., log relative risk) for the i-th NCO, with standard error s_i.
  - **(A1) Two cluster mixture model.** We assume that NCOs arise from a two-component Gaussian mixture, representing either:
    - Valid NCOs ("true nulls"): $y_i \sim N(\mu_1, \sigma_1^2 + s_i^2)$,
    - Invalid NCOs: $y_i \sim N(\mu_2, \sigma_2^2 + s_i^2)$.
  - **(A2) Majority rule.** Let π denote the proportion of valid NCOs, assumed to be greater than 0.5
- Mixture model framework: We model the observed distribution of NCO as:

$$f(y_i) = \pi \cdot N(y_i|\mu_1, \sigma_1^2 + s_i^2) + (1-\pi) \cdot N(y_i|\mu_2, \sigma_2^2 + s_i^2)$$

  - Model parameters $\theta = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \pi)$ are estimated using the Expectation-Maximization (EM) algorithm
    - E-step: posterior probabilities of each NCO being valid are calculated
    - M-step: parameters are updated to maximize the likelihood given these responsibilities
- Calibrated p-values: After estimating the distribution of valid NCOs $N(\hat{\mu}_1, \hat{\sigma}_1^2)$, compute calibrated two-sided p-values for a new effect estimate y_{n+1} with standard error s_{n+1} for a new drug-outcome pair:

$$p_{cal} = 2 \cdot \Phi\left(-\frac{|y_{n+1} - \hat{\mu}_1|}{\sqrt{\hat{\sigma}_1^2 + s_{n+1}^2}}\right)$$

Contact: bingyuz7@sas.upenn.edu, ychen123@pennmedicine.upenn.edu

Results

Simulation studies

(a) Type I error control, (b) parameter estimation accuracy, (c) statistical power

Real-world use case

Data source:
- Penn Medicine EHR data

Treatment comparison:
- GLP-1RAs vs DPP4is

Outcome:
- Cardiovascular events

Outcome Name	RR (95% CI)	Uncalibrated	NN calibrated	MNN calibrated
Non-fatal MI		0.86 (0.73 to 1.01)	0.78 (0.61 to 0.99)	0.80 (0.64 to 1.00)
Non-fatal stroke		0.83 (0.72 to 0.97)	0.75 (0.60 to 0.95)	0.78 (0.63 to 0.96)
Hospitalization for UA		0.77 (0.60 to 0.98)	0.70 (0.51 to 0.94)	0.72 (0.54 to 0.96)
CV death		0.96 (0.74 to 1.24)	0.87 (0.63 to 1.19)	0.89 (0.66 to 1.21)
3-point MACE		0.86 (0.77 to 0.97)	0.78 (0.63 to 0.97)	0.81 (0.67 to 0.98)
4-point MACE		0.88 (0.78 to 0.99)	0.80 (0.64 to 0.98)	0.82 (0.68 to 1.00)

Uncalibrated
NN calibrated
MNN calibrated

Conclusions

The proposed MNN framework models NCOs via a two-component mixture, yielding more robust calibration.

- Diagnostic use: The estimated valid-NCO proportion (π) can flag issues; low π suggests revisiting NCO selection or conducting sensitivity analyses
- Multi-site application: In federated studies, MNN can be paired with meta-analytic or hierarchical calibration to handle heterogeneous NCO quality.

References

1. Schuemie MJ, Ryan PB, DuMouchel W, et al. Interpreting observational studies: why empirical calibration is needed to correct p-values. Stat Med 2014; 33: 209–218.
2. Schuemie MJ, Hripcsak G, Ryan PB, et al. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. Proceedings of the National Academy of Sciences 2018; 115: 2571–2577.

#OHDSISocialShowcase This Week

Thursday

Understanding Community Needs for ATLAS: Results of the June 2025 OHDSI Feature Use and Prioritization Survey

(Christopher Knoll)

Understanding Community Needs for ATLAS
Results of the June 2025 OHDSI Feature Use and Prioritization Survey

PRESENTER: **Christopher Knoll**

INTRODUCTION:

- **ATLAS has evolved over 10 years, growing in features and complexity. For the 3.0 version, we wanted to determine features to trim and areas for improvement. A survey was created to discover community sentiment and areas of focus.**

METHODS

1. We constructed a 5-part survey that was presented at OHDSI community calls and implemented with Microsoft forms.
2. We collected frequency and importance responses from OHDSI community members.
3. We evaluated graphs to understand frequency and importance of ATLAS features. Through manual and GenAI processing, we summarized user feedback to determine the perceived strengths and weaknesses of the application.

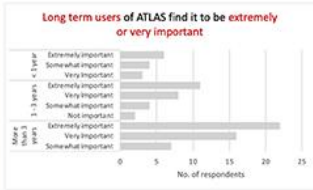
RESULTS

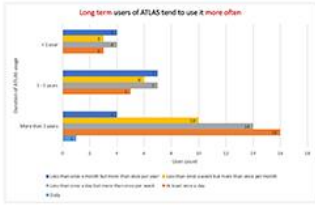
The survey reached almost 100 people. Overall, this graph shows higher importance and frequency of database characterization, vocabulary search, concept set and cohort definitions. Less used functions were estimation, prediction, reusables and feedback. User feedback cited strengths in core OMOP activities, but challenges in performance and usability (UI/UX).

Although ATLAS is regarded as a **widely valued tool** with **broad accessibility** that is **essential for cohort work,** with **global relevance** and **broad community support,** challenges include **usability for new users, performance, installation challenges, documentation** and **feature gaps.**

Take a picture to **download** the **full paper**

- **Top 5 priorities –**
  - **User experience and guidance** - UI/UX, easier onboarding, more guides
  - **Performance** – Improvements to speed, responsiveness, stability
  - **Concept set and cohort creation** – improvement in fuzzy search, syncing, versioning, visuals
  - **Interoperability and external connections** – REST API, HADES, Strategus connections
  - **Expanded reporting capabilities** – explanation for less technical users, cloudburst chart, cohort pathways and incidence rates





Christopher Knoll, Karthik Seetharaman

#OHDSISocialShowcase This Week

Friday

Data-Driven Identification of Comorbidities and Pharmacological Patterns in Patients with Sleep Disorders

(Praveen Kumar, Kristan A Schneider, Fariha Moomtaheen, Rajesh Upadhayaya, Scott A. Malec, Jeremy J. Yang, Cristian G. Bologa, Yiliang Zhu, Mauricio Tohen, Gerardo Villarreal, Douglas J. Perkins, Elliot M. Fielstein, Sharon E. Davis, Michael E. Matheny, Christophe G. Lambert)

Data-Driven Identification of Comorbidities and Pharmacological Patterns in Patients with Sleep Disorders

Praveen Kumar[1], Kristan A. Schneider[1], Fariha Moomtaheen[1], Rajesh Upadhayaya[1], Scott A. Malec[1,3], Jeremy J. Yang[1,3], Cristian G. Bologa[1,3], Yiliang Zhu[1], Mauricio Tohen[2], Gerardo Villarreal[2,3], Douglas J. Perkins[1], Elliot M. Fielstein[4,5], Sharon E. Davis[4,5], Michael E. Matheny[4,5], Christophe G. Lambert[1,3]

[1]University of New Mexico, Department of Internal Medicine, Albuquerque, NM, USA, [2]University of New Mexico, Department of Psychiatry & Behavioral Sciences, Albuquerque, NM, USA, [3]VA New Mexico Healthcare System, Albuquerque, NM, USA, [4]Vanderbilt University Medical Center, Department of Biomedical Informatics, Nashville, TN, USA, [5]Tennessee Valley Healthcare System VA, Nashville, TN, USA

Abstract

Sleep disorders are common yet underdiagnosed conditions that substantially contribute to poor health outcomes and rising healthcare costs. In the US, 50-70 million individuals are affected by chronic sleep or wakefulness disorders, with undiagnosed sleep apnea alone estimated to cost over $150 billion annually. Identifying at-risk individuals is critical to reducing the health and economic burden of sleep disorders. However, most prior research has relied on supervised learning approaches that treat patients without diagnosis codes as negative cases. This assumption overlooks the realities of underdiagnosis and undercoding in electronic health records (EHRs), thereby limiting predictive accuracy of supervised models. In this study, we aimed to characterize the clinical profiles of patients diagnosed with sleep disorders to support the identification of undiagnosed individuals. Using the *All of Us* EHR dataset mapped to the OMOP Common Data Model, we identified 83,610 patients diagnosed with sleep disorders. From their records, we extracted 41,068 unique conditions and medications documented prior to diagnosis. We then applied non-negative matrix factorization (NMF) to the patient-covariate matrix to find latent comorbidity and drug exposure patterns. The analysis revealed four clinically meaningful components: (1) chronic pain, anxiety, depression, and gastrointestinal symptoms; (2–3) perioperative and pain-management drug exposures, including anesthetics, opioids, sedatives, and antiemetics; and (4) cardiometabolic conditions such as hypertension, diabetes, obesity, and hyperlipidemia. These findings demonstrate that NMF can uncover pre-diagnosis profiles among patients with sleep disorders, offering a scalable framework for early detection of at-risk individuals and enabling timely intervention to reduce the long-term burden of sleep disorders.

Background

Sleep is a fundamental biological activity essential for overall health and well-being. Sufficient, uninterrupted sleep is critical for maintaining both physical and mental health. Insufficient or disrupted sleep can have detrimental effects, leading to a range of physical and mental health conditions[1,2]. Sleep disorders (such as insomnia, sleep apnea, hypersomnia, sleepwalking, and nightmares) are conditions that affect the timing, quantity, or quality of sleep and are linked to an increased risk of numerous health problems, including depression, obesity, type 2 diabetes, heart disease, dementia, and certain cancers[1,3,4]. Sleep disorders are among the most prevalent yet often overlooked health concerns. It is estimated that 50-70 million Americans suffer from chronic sleep or wakefulness disorders[1,4], and over 100 million Americans of all ages report not getting sufficient sleep[5]. Each year, these sleep-related conditions contribute significantly to national healthcare costs. For instance, in the US, the annual cost related to undiagnosed sleep apnea alone is estimated to amount to $150 billion; this figure does not include additional costs associated with health problems, productivity loss, and accidents, highlighting sleep disorders as a significant public health concern.

Given the substantial health and economic burden of sleep disorders, identifying at-risk individuals is essential to mitigate serious health consequences. Existing research has primarily relied on supervised classification methods to predict sleep disorders[3,6,7]. These methods typically treat uncoded patients as negative examples in machine learning (ML) models. However, the absence of a diagnosis code in a health record does not necessarily imply that the person is free of the condition as underdiagnosis, undercoding, incomplete data capture, and variations in clinical documentation practices are prevalent across healthcare data.

In this study, our objective is to characterize the clinical patterns of patients diagnosed with sleep disorders, enabling us to identify individuals at risk who may be undiagnosed and share similar profiles. Specifically, we aim to identify common comorbidities and medications among individuals diagnosed with any type of sleep disorder by employing the non-negative matrix factorization (NMF) algorithm[8]. This approach could ultimately facilitate timely interventions thereby preventing severe outcomes associated with sleep disorders.

Materials and Methods

- **Data Source:** *All of Us* electronic health records (EHRs) data mapped to the OMOP Common Data Model[9].
- **Sleep disorder phenotype:** G47* (physiological or medical origin) and F51* (psychological or behavioral origin).
- **Covariates:** Conditions (ICD-10-CM codes only) and Drugs (RxNorm codes).

Figure 1: Steps to generate Compressed Sparse Row (CSR) matrix for the unsupervised method, NMF.

- **ML method:** Unsupervised algorithm, NMF. NMF factorizes a non-negative $n \times m$-matrix X into two lower-rank non-negative matrices: a basis matrix W ($n \times k$) and a coefficient matrix H ($k \times m$), such that $X \approx WH$, where $k \ll min(n, m)$ is the number of latent components.
- NMF reconstruction error was used for the selection of number of components, k.

if $X \approx WH$ with $W \in \mathbb{R}^{n \times k}$ and $H \in \mathbb{R}^{k \times m}$, the reconstruction error is:

$$\|X - WH\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - (WH)_{ij})^2}$$

Results

Component 1	Component 2	Component 3	Component 4
1. Other Chronic Pain	1. Sodium chloride 9 MG	1. Sodium chloride 9 MG	1. Essential (primary) hypertension
2. Anxiety Disorder, Unspecified	2. Fentanyl 0.05 MG;	2. Fentanyl	2. Hyperlipidemia, unspecified
3. Gastro-Esophageal Reflux Disease Without Esophagitis	3. Midazolam 1 MG	3. Ondansetron 2 MG	3. Type 2 diabetes mellitus without complications
4. Low Back Pain	4. Propofol 10 MG	4. Midazolam	4. Atherosclerotic heart disease of native coronary artery without angina pectoris
5. Major Depressive Disorder, Single Episode, Unspecified	5. Naloxone hydrochloride 0.4 MG	5. Lidocaine	5. Obesity, unspecified
6. 2 Ml Ondansetron 2 Mg	6. Lidocaine hydrochloride 20 MG	6. Calcium chloride 0.0014 MEQ	6. Mixed hyperlipidemia
7. Unspecified Abdominal Pain	7. Potassium chloride 20 MEQ Extended Release Oral Tablet	7. Ondansetron	7. Gastro-esophageal reflux disease without esophagitis
8. Cervicalgia	8. Diphenhydramine Hydrochloride 50 MG	8. Propofol	8. Shortness of breath
9. Acetaminophen 325 Mg	9. Calcium chloride 0.0014 MEQ	9. 10 ML sodium chloride 9 MG	9. Chest pain, unspecified
10. Cough	10. 1 ML promethazine hydrochloride 25 MG	10. Midazolam 1 MG	10. Type 2 diabetes mellitus with hyperglycemia
11. Fluticasone Propionate 0.05 Mg	11. Dexamethasone phosphate 4 MG	11. Acetaminophen 325 MG Oral Tablet	11. Atorvastatin 40 MG Oral Tablet
12. 2 Ml Fentanyl 0.05 Mg	12. Ketorolac tromethamine 30 MG	12. Dexamethasone	12. Pure hypercholesterolemia, unspecified
13. Other Fatigue	13. Morphine sulfate 4 MG	13. Oxycodone hydrochloride 5 MG Oral Tablet	13. Aspirin 81 MG Delayed Release Oral Tablet
14. Other Specified Postprocedural States	14. Rocuronium bromide 10 MG	14. Oxycodone	14. Vitamin D deficiency, unspecified
15. Acute Upper Respiratory Infection, Unspecified	15. 50 ML magnesium sulfate 40 MG	15. Hydromorphone	15. Hypothyroidism, unspecified

Table 1. Top 15 covariates with the highest coefficients in each latent component of the NMF coefficient matrix (H), representing conditions and medications prior to the first sleep disorder diagnosis.

Figure 2: Set of top 10 features from each component along with their coefficient values within the respective components.

Figure 3: Number of patients assigned to each component based on their largest membership weight.

Discussion and Conclusions

- NMF, an unsupervised approach, is well-suited to health data without labeled negatives, enabling discovery of latent patterns of multimorbidity and polypharmacy for a given condition.
- All 4 NMF components revealed clinically interpretable patterns of pre-diagnosis conditions and medication exposures in patients later diagnosed with sleep disorders.
- Studies have found that chronic pain and mood disorders (anxiety, depression)[10], cardiometabolic conditions (hypertension, hyperlipidemia, diabetes, obesity, cardiovascular disease)[11], and exposure to opioids (fentanyl, morphine), sedatives (midazolam, propofol), and opioid antagonists (naloxone) are associated with poor sleep quality and increased risk of sleep disorders[12].
- Identifying common conditions and medications in diagnosed patients can help detect undiagnosed individuals at high risk for sleep disorders, enabling earlier intervention to prevent severe sleep-related problems.
- While we applied NMF only to conditions and medications in this study, the methodology can be extended to include other OMOP semantic domains, such as procedures, observations, and measurements.

References

1. Sleep Science and Sleep Disorders. https://www.ninds.nih.gov/science/sleep-science-and-sleep-disorders Accessed on: Sep 7, 2025.
2. Phan, D. V., Yang, N. P., Kuo, C. Y., & Chen, C. L. (2020). Sleep learning approaches for sleep disorder prediction in an asthma cohort. Journal of Asthma, 58(7), 903–911. https://doi.org/10.1080/02770903.2020.1742352
3. Ay S, Fayet O, Senin S, Chakraborty S, Barua PD, Loh HW, Elphick H, Molinari F, Acharya UR. A review of automated sleep disorder detection. Computers in biology and medicine. 2022 Nov 1;150:106100.
4. Institute of Medicine (US) Committee on Sleep Medicine and Research; Colten HR, Altevogt BM, editors. Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem. Washington (DC): National Academies Press (US); 2006. 3, Extent and Health Consequences of Chronic Sleep Loss and Sleep Disorders. Available from: https://www.ncbi.nlm.nih.gov/books/NBK19961/ Accessed on: Sep 7, 2025.
5. Sleep Disorders. https://my.clevelandclinic.org/health/diseases/21429-sleep-disorders Accessed on: Sep 7, 2025.
6. Kim WP, Kim HJ, Park SP, Lim JH, Cho EH, Lee H. Machine learning-based prediction of attention-deficit/hyperactivity disorder and sleep problems with wearable data in children. JAMA network open. 2023 Mar 1;6(3):e233923.
7. Ma EY, Kim JW, Lee S, Cho SN, Kim H, Kim JK. Combined unsupervised-supervised machine learning for phenotyping complex diseases with its application to obstructive sleep apnea. Scientific Reports. 2021 Feb 24;11(1):4457.
8. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. nature. 1999 Oct 21;401(6755):788-91.
9. Voss EA, Makadia R, Matcho A, et al. : Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc 2015; 22 (3): 553–64.
10. Morin, Charles M. Ph.D.; Gibson, Douglas Psy.D.; Wade, James Ph.D. Self-Reported Sleep and Mood Disturbance in Chronic Pain Patients. The Clinical Journal of Pain 14(4):p 311-314, December 1998.
11. Xu, Zhiwei BMa; Qian, Xuanfang MDb,c; Xu, Junyao PhDb,*. The cardiometabolic index as a predictor of sleep disorders and mortality: A cross-sectional study. Medicine 104(14):p e42029, April 04, 2025. | DOI: 10.1097/MD.0000000000042029
12. Angarita, G.A., Emadi, N., Hodges, S. et al. Sleep abnormalities associated with alcohol, cannabis, cocaine, and opiate use: a comprehensive review. Addict Sci Clin Pract 11, 9 (2016). https://doi.org/10.1186/s13722-016-0056-7

Acknowledgment: This research was supported by funding from the US National Institutes of Health, specifically, the National Institute of Mental Health grant R03MH129764, the National Library of Medicine grant R00LM013367, and infrastructure support from the National Center for Advancing Translational Sciences grant UL1TR001449. We gratefully acknowledge All of Us participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's All of Us Research Program for making available the participant data examined in this study.

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?

PRHeG Work Group
Perinatal & Reproductive Health Group
OKR’s 2026

PRHeG OKR #1 2026

Use the OHDSI Evidence Network to summarize pregnancy-related data assets, studies, and participation across the network.

- Key Results
  - Identify participating data partners contributing pregnancy-related data by drafting a proposal to the Evidence Network
  - Produce standardized summary tables and visualizations describing the pregnancy network and potential data sources for pregnancy research

Leads: Rupa Makadia
Supporting contributors: Allison Callahan, Stephanie Leonard, Cynthia Sung
Timelines: Q2
Progress Notes: Shared the framework/idea with the Evidence Network; need to draft proposal and bring back to the group

PRHeG OKR #2 2026

Objective Transform U.S. Birth Certificate data into a high-quality, research-ready OMOP CDM asset that accelerates maternal health insights and enables immediate use by investigators.

- Key Results
  - Complete end-to-end ETL of 2021 U.S. Birth Certificate data into the current OMOP CDM.
  - Publish documentation and data dictionary and enable the Maternal Health Research Fellowship (JHU)
  - Produce a standardized characterization report (counts, completeness, and distributions for core maternal and neonatal domains such as maternal age, parity, gestational age, birth outcomes) through the Maternal Health Research Fellowship (JHU)
  - Generalize and re-run the ETL for at least one additional year (e.g., 2020 or 2022) using the same pipeline.

Leads: Cynthia Sung
Supporting contributors: Sean O'Rilley, Paul Nagy, Hayden Spence, Stephanie Leonard
Timelines: Q2/Q3
Progress Notes: ETL has started, data transfer to JHU servers in progress

PRHeG OKR #3 2026

Support, mentor, and integrate Maternal Health Data Science Fellows into PRHeG and OHDSI research activities.

- Key Results
  - Establish structured mentorship, office hours, or working sessions
  - Engage fellows in active PRHeG studies, protocols, or phenotype development
  - Co-develop fellow-led deliverables (presentations, study packages, manuscripts, or tutorials)

Leads: Sean O'Reilly
Supporting contributors: MHF Faculty
Timelines: Q3/Q4
Progress Notes: Fellowship applications are in progress, fellows start in the Fall, starting monthly meetings with faculty

PRHeG OKR #4 2026

Deploy the submitted pregnancy algorithm in claims data through the OHDSI Evidence Network and evaluate its integrity, usability, and performance across diverse data sources, establishing confidence in its applicability for network-wide pregnancy research.

- Key Results
  - Write a protocol to deploy the pregnancy algorithm within the OHDSI Evidence Network using claims-based OMOP CDM data across multiple participating sites.
  - Evaluate algorithm integrity by assessing completeness, internal consistency, and face validity of identified pregnancy episodes across data sources.
  - Summarize cross-site findings and lessons learned (e.g., summary tables, diagnostics, and documentation) and draft manuscript for publication.

Leads: JHU Maternal Health Fellows
Supporting contributors: PRHeG working group/ MHF faculty
Timelines: Q4/Q1 2027
Progress Notes:

Phenotypes for Pregnancy Cohorts

Live birth OHDSI PhenotypeLibrary GitHub repository. https://github.com/OHDSI/PhenotypeLibrary/blob/main/inst/cohorts/1433.json

Stillbirth OHDSI PhenotypeLibrary GitHub repository. https://github.com/OHDSI/PhenotypeLibrary/blob/main/inst/cohorts/1432.json

Ectopic Pregnancy OHDSI PhenotypeLibrary GitHub repository. https://github.com/OHDSI/PhenotypeLibrary/blob/main/inst/cohorts/1431.json

Miscarriage or abortion OHDSI PhenotypeLibrary GitHub repository. https://github.com/OHDSI/PhenotypeLibrary/blobl/main/inst/cohorts/1434.json

Identification and dating of pregnancies in administrative healthcare claims databases: An updated, transparent, and transportable algorithm

Rupa Makadia[1], Erica A. Voss[1], Joel Swerdel[1], Jill Hardin[1], Amir Sarayani[1], Alexis A. Krumme[1], Kourtney J. Davis[1], Patrick B. Ryan[1], Melanie H. Jacobson[1]*

[1]Johnson & Johnson, Raritan, NJ

Running Title: Identifying and dating pregnancies in administrative claims

Keywords: Pregnancy, administrative claims, ICD, electronic healthcare data, real world data

Manuscript Submitted and In Review!!

OKR Progress/ Meetings

If anything sounds exciting, we would love to have you join! Each OKR will have its own workstream and then monthly calls are dedicated to topics to share and progress on our OKR's.

Please reach out to makadia@ohdsi.org OR join our working group!

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