Education in OHDSI: Lessons Learned

OHDSI Community Call

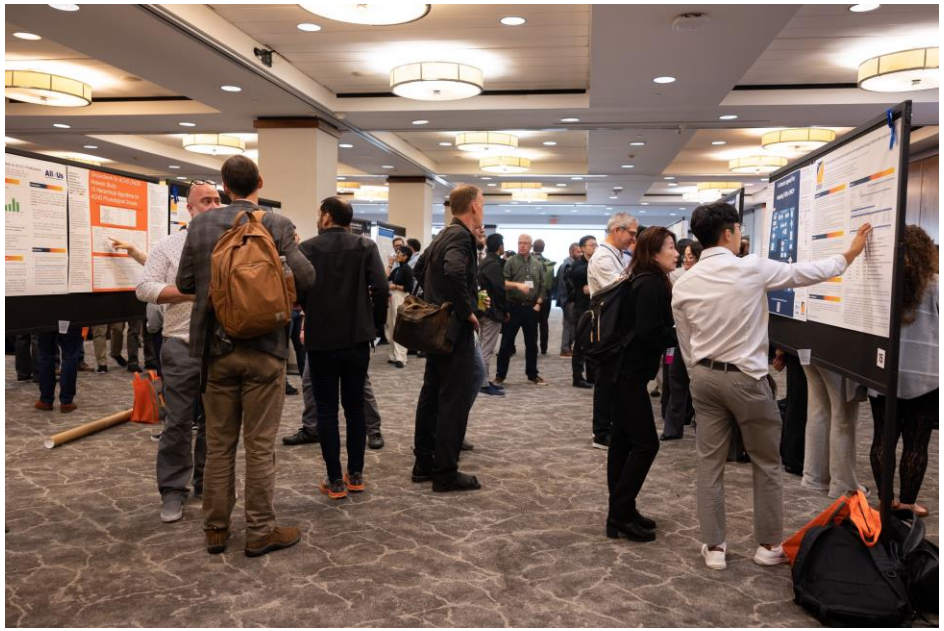
Sept. 23, 2025 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Upcoming Community Calls

Date	Topic
Sept. 23	Educating on OHDSI: Lessons Learned
Sept. 30	OHDSI 2025 Poster Preview Mad Minutes / Symposium Logistics
Oct. 7	No Call – OHDSI Symposium
Oct. 14	Welcome to OHDSI
Oct. 21	Meet the Titans

Sept. 30: OHDSI2025 Mad Minutes



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Kim López-Güell, Martí Català, Daniel Dedman, Talita Duarte-Salles, Raivo Kolde, Raúl López-Blasco, Álvaro Martínez, Gregoire Mercier, Alicia Abellan, Johnmary T Arinze, Theresa Burkard, Edward Burn, Zara Cuccu, Antonella Delmestri, Dominique Delseny, Sara Khalid, Chungsoo Kim, Ji-Woo Kim, Kristin Kostka, Cora Loste, Miguel A Mayer, Jaime Meléndez-Cardiel, Núria Mercadé-Besora, Mees Mosseveld, Akihito Nishimura, Hedvig Me Nordeng, Jessie O Oyinlola, Roger Paredes, Laura Pérez-Crespo, Marta Pineda-Moncusí, Juan Manuel Ramírez-Anguita, Nhung T H Trinh, Anneli Uusküla, Bernardo Valdivieso, Daniel Prieto-Alhambra, Junqing Xie, Lourdes Mateu, and Annika M Jödicke** on the publication of **Clusters of post-acute COVID-19 symptoms: a latent class analysis across 9 databases and 7 countries** in the *Journal of Clinical Epidemiology.*

ELSEVIER

Check for updates

Journal of Clinical Epidemiology 185 (2025) 111867

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

Clusters of post-acute COVID-19 symptoms: a latent class analysis across 9 databases and 7 countries

Kim López-Güell[a,#], Martí Català[a,#], Daniel Dedman[b], Talita Duarte-Salles[c,d], Raivo Kolde[e], Raúl López-Blasco[f], Álvaro Martínez[g], Gregoire Mercier[h,i], Alicia Abellan[c], Johnmary T. Arinze[d], Theresa Burkard[a], Edward Burn[a], Zara Cuccu[b], Antonella Delmestri[a], Dominique Delseny[h], Sara Khalid[a], Chungsoo Kim[j], Ji-woo Kim[k], Kristin Kostka[a,l], Cora Loste[m,n,o,p], Miguel A. Mayer[q], Jaime Meléndez-Cardiel[f], Núria Mercadé-Besora[a,c], Mees Mosseveld[d], Akihito Nishimura[r], Hedvig ME. Nordeng[s,t], Jessie O. Oyinlola[b], Roger Paredes[m,n,p,u,v,w,x], Laura Pérez-Crespo[c], Marta Pineda-Moncusí[a], Juan Manuel Ramírez-Anguita[q], Nhung T.H. Trinh[s], Anneli Uusküla[y], Bernardo Valdivieso[g,z], Daniel Prieto-Alhambra[a,d,*], Junqing Xie[a], Lourdes Mateu[m,n,o,p,v,†], Annika M. Jödicke[a,†]

[a]*Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK*
[b]*CPRD, Medicines and Healthcare products Regulatory Agency, London, UK*
[c]*Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain*
[d]*Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[e]*Institute of Computer Science, University of Tartu, Tartu, Estonia*
[f]*Biocomputing Unit, Aragon Health Sciences Institute (IACS), Zaragoza, Spain*
[g]*The Health Research Institute Hospital La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7a planta, Valencia 46026, Spain*
[h]*Public Health Department, University Hospital of Montpellier, Montpellier 34295, France*
[i]*IDESP, Université de Montpellier, INSERM, Montpellier 34000, France*
[j]*Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Republic of Korea*
[k]*Big Data Department, Health Insurance Review and Assessment Service, Wonju, Republic of Korea*
[l]*The OHDSI Center at the Roux Institute, Northeastern University, Portland, ME, USA*
[m]*Department of Infectious Diseases & irsiCaixa AIDS Research Institute, Hospital Universitari Germans Trias i Pujol, Badalona, Catalonia, Spain*
[n]*University of Vic- Central University of Catalonia (UVic-UCC), Vic, Spain*
[o]*REICOP (Red de Investigación Covid Persistente), Madrid, Spain*
[p]*Fundació Lluita Contra les Infeccions, Badalona, Catalonia, Spain*
[q]*Parc de Salut Mar, Hospital del Mar Medical Research Institute, Barcelona, Spain*
[r]*Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA*
[s]*Department of Pharmacy, Pharmacoepidemiology and Drug Safety Research Group, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway*
[t]*Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway*
[u]*Department of Pathology, Center for Global Health and Diseases, Case Western Reserve University School of Medicine, Cleveland, OH*
[v]*Universitat Autònoma de Barcelona, Catalonia, Spain*

OHDSI Shoutouts!

Congratulations to the team of **George Kafatos, Joe Maskell, Olia Archangelidi, and David Neasham** on the publication of **Validation of the transformed clinical practice research datalink (CPRD) GOLD and aurum data into the OMOP common data model** in the *Health Informatics Journal.*

Health Informatics Journal
Volume 31, Issue 3, July 2025
© The Author(s) 2025, Article Reuse Guidelines
https://doi.org/10.1177/14604582251381270

Sage Journals

Review Article

Validation of the transformed clinical practice research datalink (CPRD) GOLD and aurum data into the OMOP common data model

George Kafatos, Joe Maskell, Olia Archangelidi, and David Neasham

Abstract

Objective: To assesses the transformation of UK Clinical Practice Research Datalink (CPRD) databases into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) version 5.3.1. **Methods:** A systematic approach was used to generate medical code lists and compare prevalent and incident counts between the source and OMOP CDM versions. **Results:** The results showed, for CPRD General Practitioner Online Database (GOLD) database, 89.5% of clinical events had no or very small differences in prevalent and incident event counts between the two versions of the database. The differences for CPRD Aurum were even smaller, with 97.4% of events showing no or very small differences in counts between the source and OMOP versions. Some observed discrepancies were due to codes being mapped into different tables. **Conclusion:** The study findings confirm the consistency of the OMOP transformation and provide confidence in analyses that query CPRD OMOP-transformed data.

Keywords

CPRD Aurum, CPRD GOLD, OMOP, medical codes

OHDSI Shoutouts!

Congratulations to the team of **Aparajita Kashyap, Maryam Aziz, Tony Y Sun, Sharon Lipsky Gorman, Jessica Opoku-Anane, and Noémie Elhadad** on the publication of **Investigating racial disparities in drug prescriptions for patients with endometriosis** in *NPJ Women's Health*.

npj | women's health

Article

https://doi.org/10.1038/s44294-025-00053-3

Investigating racial disparities in drug prescriptions for patients with endometriosis

Check for updates

Aparajita Kashyap[1] ✉, Maryam Aziz[2], Tony Y Sun[1], Sharon Lipsky Gorman[1], Jessica Opoku-Anane[3] & Noémie Elhadad[1,2]

We assess racial disparities in medication prescription patterns for endometriosis patients across Medicaid administrative claims data. We use ATC 3rd level drug codes to identify drug classes prescribed significantly more frequently for endometriosis patients than a comparison cohort of non-endometriosis patients. Temporal prevalence differences of prescriptions (pre- vs. post-diagnosis) were also examined. The endometriosis cohort comprised 16,372 endometriosis patients (23.3% Black, 66.0% White). Of the 28 drug classes examined, 17 were prescribed significantly less in Black patients and 4 were prescribed significantly more in Black patients. Of the 17 drugs prescribed more often in White patients, 13 have larger disparities pre-diagnosis than post-diagnosis. In the non-endometriosis cohort (n = 3,663,904), 21 drug classes were prescribed significantly more in White patients and 6 were prescribed significantly more in Black patients. Our analysis identifies disparities in prescriptions practices between White and Black endometriosis patients, notably in pain management and comorbidity treatment.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	ATLAS/WebAPI
Wednesday	9 am	Oncology Outreach/Research Subgroup
Wednesday	9 am	Health Economics and Value Assessment (HEVA)
Wednesday	10 am	Surgery and Perioperative Medicine
Wednesday	10 am	Women of OHDSI
Wednesday	12 pm	Latin America
Thursday	9:30 am	Network Data Quality
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Friday	2 pm	Vaccine Vocabulary
Monday	10 am	Getting Started Subgroup
Tuesday	10 am	CDM Survey Subgroup

Congratulations, 2025 Titan Award nominees!

Agnes Kiragga • Akihiko Nishimura • Alexey Manoylenko • ALS TDI's Real World Evidence Team • Andrew Williams • Andrew Kanter • Aniek Markus • Anna Ostropolets • Anthony Sena • Asieh Golozar • ATLAS Development Team • Ben Martin • Bill O'Brien • Bingyu Zhang • Carlos Diaz • Chungsoo Kim • Christopher Knoll • Clair Blacketer • Craig Sachson • Critical Path Institute's Data Science and Data Engineering team • Cynthia Sung • Daniel Prieto-Alhambra • DARWIN-EU Team • Data4Life Team • Dave Kern • Davera Gabriel • Department of Biomedical Systems Informatics, Yonsei University College of Medicine • Deran Mckeen • Diane Corey • Egill Fridgeirsson • Eric Fey • Evanette Burrows • Eye Care and Vision Research WG • FHIR to OMOP WG • Freija Descamps • German Soto • Greg Klebanov • Hannah Lee • Harry Reyes Nieva • HealthPartners Institute • Henrik John • Ian Braun • Ilse Vermeulen • IQVIA OMOP DARWIN Team • IQVIA OMOP Productized Analytics Team • James Gilbert • Jamie Weaver • Jared Houghtaling • Jason Hsu • Jenna Reps • Jiwon Um • Joel Swerdel • John Gresh • Justin Bohn • Katia Verhamme • Lars Halvorsen • Liesbet Peeters • Lotte Geys • Maarten van Kessel • Marc Suchard • Marti Catala Sabate • Martijn Schuemie • Marty Alvarez • Maxim Moinat • Michael Matheny • Michel Walravens • Mike Pauley • Milou Brand • Mitchell Conover • Mukkesh Kumar • OHDSI Belgium Team • Patricia Mabry • Patrick Ryan • Pavan Sudhakar • Peter Hoffmann • Peter Rijnbeek • Polina Talapova • Renske Los • REWARD Team • Richard Boyce • Roger Carlson • Sam Patnoe • SciForce Team • Treatment Patterns Team • Vaccine Vocabulary Team • Will Roddy

Jamie Weaver Scholarship at University of Oxford

About the Project

This scholarship and work has been proposed to continue and expand work started by the late James (Jamie) Weaver. Jamie was a talented and bright data scientist and DPhil student working with us on the use of methods to minimise the impact of outcome misclassification in real world evidence (RWE). Funding has been secured, from the Medical Sciences Division, Brasenose College, and NDORMS, for this project to continue his important work on this extremely relevant topic; the successful candidate will be assigned to Brasenose College.

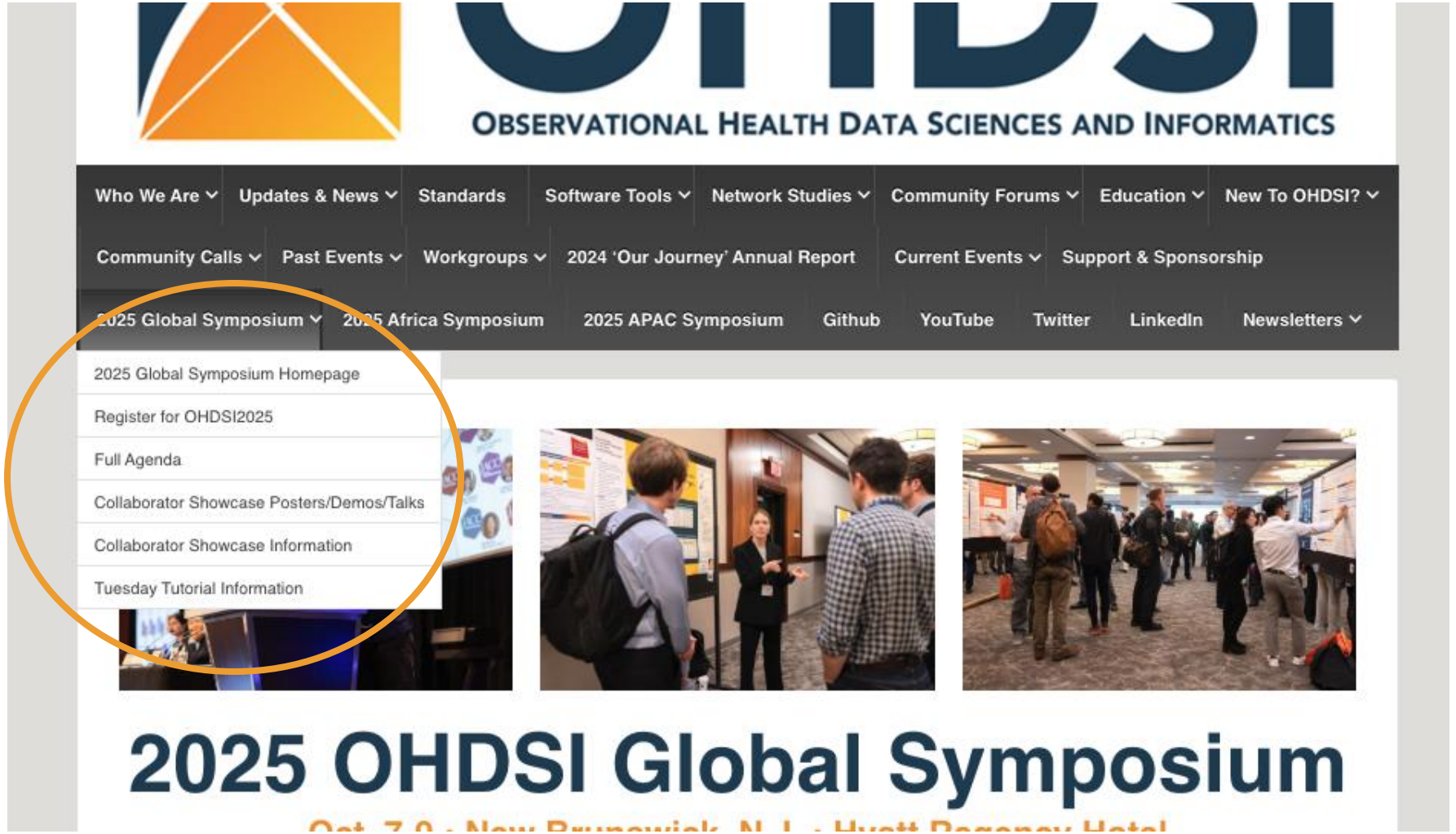
Real world evidence (RWE) is generated by leveraging and processing large routinely collected health data. Despite difficulties in the analysis of such information for causal inference purposes, RWE has recently been shown as a reliable source of data when used using adequate methods for trial emulation [1, 2]. We participate in multiple European and international networks to generate reliable information to inform, amongst others, regulatory decision making and health technology assessments.

Through ongoing collaborations, we leverage multiple international datasets mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model in a federated manner. Previous work led by our student Jamie Weaver uncovered the impact of outcome misclassification on the estimation of background rates of adverse events, and proposed new methods to account for this in future studies [3].

Through this 3-year PhD funded studentship, we aim to investigate how novel methods can be applied to measure and account for outcome misclassification in RWE studies, by researching:

1. The use and application of artificial intelligence (and specifically large language models) for the generation and validation of computable phenotypes
2. The impact of outcome misclassification in different data assets
3. The performance of existing and novel methods to account for outcome misclassification in international RWE studies

Global Symposium: Oct. 7-9



ohdsi.org/ohdsi2025

Global Symposium: Oct. 7-9

Agenda ·

Time (ET)	Session/Topic
7:00 am - 8:00 am	Lite Breakfast ar
8:00 am - 12:00 pm	*Introductory Tut* An Introduction t Faculty: Erica Vo Pennsylvania; K of South Australi Vocabulathon 20 Lead: Alexander
12:00 pm - 1:00 pm	Buffet Lunch for
1:00 pm - 5:00 pm	*Advanced Tutori* Developing and to the OMOP Co Faculty: Clair Bl University; Evan Mahidol Universi Using the OHDSI Faculty: Anna O Data Services; P Clinical Characte Evidence Faculty: Patrick Hsin Yi "Cindy" Population-Level Real-World Evid Faculty: George Johnson; Linyin Columbia Univer Patient-Level Pre Evidence Faculty: Jenna R Ross Williams, E
5:00 pm - 6:00 pm	Collaborator Sho
6:00 pm - 8:00 pm	Networking Rece

Agenda · Wednesday, Oct. 8

Time (ET)	Topic
7:00 am - 8:00 am	Lite Breakfast and Registration, Exhibits
7:15 am - 7:45 am	Newcomer Orientation Paul Nagy, Johns Hopkins University
8:00 am - 9:00 am	State of the Community: Welcome to OHDSI George Hripcsak, Columbia University
9:00 am - 9:30 am	Group Networking Activity
9:30 am - 10:15 am	Collaborator Showcase Poster/Software Demo Session #1
10:15 am - 12:00 pm	Plenary: Why network studies are necessary to improve trust in evidence Martijn Schuemie, Johnson & Johnson; Asieh Golozar, Nemesis Health; Cindy Cai, Johns Hopkins University; Patrick Ryan, Johnson & Johnson, Columbia University
12:00 pm - 1:00 pm	Buffet Lunch, Exhibits
1:00 pm - 2:00 pm	Plenary: Reflections on the evolution of pre- and postmarket safety review in CDER over 3 decades Judy Racoosin, US Food and Drug Administration (retired)
2:00 pm - 2:45 pm	Collaborator Showcase Lightning Talk Session #1 Moderator: Harry Reyes Nieva, Columbia University Bridging Standards: Creating OMOP data via Fast Healthcare Interoperability Resources (FHIR) and Health Information Networks Stephanie Hong, Johns Hopkins University OMOP Waveform Extension: A Schema for Integrating Physiological Signals and Derived Features into the OMOP CDM Jared Houghtaling, Tufts University Improving VSAC to OMOP Mapping Using LLM Assisted Curation Robert Barrett, Johns Hopkins University Evaluating the effectiveness of using Large Language Models for the development of concept sets Joel Swerdel, Johnson & Johnson Validating a Scalable Approach to Data Fitness-for-Use: Database Diagnostics Applied to LEGEND-T2DM Clair Blacketer, Johnson & Johnson

* agenda is subject to change

Agenda · Wednesd

Time (ET)	Topic
7:00 am - 8:00 am	Lite Breakfast and Registration, Exhibits
7:15 am - 7:45 am	Newcomer Orientation Paul Nagy, Johns Hopkins University
8:00 am - 9:00 am	State of the Community: Welcome to OH George Hripcsak, Columbia University
9:00 am - 9:30 am	Group Networking Activity
9:30 am - 10:15 am	Collaborator Showcase Poster/Software
10:15 am - 12:00 pm	Plenary: Why network studies are neces Martijn Schuemie, Johnson & Johnson; Cindy Cai, Johns Hopkins University; Pa Columbia University
12:00 pm - 1:00 pm	Buffet Lunch, Exhibits
1:00 pm - 2:00 pm	Plenary: Reflections on the evolution of CDER over 3 decades Judy Racoosin, US Food and Drug Admini
2:00 pm - 2:45 pm	Collaborator Showcase Lightning Talk S Moderator: Harry Reyes Nieva, Columbia Bridging Standards: Creating OMOP data Resources (FHIR) and Health Information Stephanie Hong, Johns Hopkins Univers OMOP Waveform Extension: A Schema f and Derived Features into the OMOP CDM Jared Houghtaling, Tufts University Improving VSAC to OMOP Mapping Usin Robert Barrett, Johns Hopkins University Evaluating the effectiveness of using La development of concept sets Joel Swerdel, Johnson & Johnson Validating a Scalable Approach to Data F Diagnostics Applied to LEGEND-T2DM Clair Blacketer, Johnson & Johnson

Agenda · Thursday, Oct. 9

Time (ET)	Meetings
7:00 am - 8:00 am	Lite Breakfast, Exhibits
8:00 am - 10:00 am	Session 1 of Workgroup Activities Featuring: Africa Chapter, APAC Chapter, Medical Imaging, GIS - Geographic Information System, HADES Hackathon, Oncology, Common Data Model, ATLAS/WebAPI, Phenotype Development and Evaluation, Dentistry, and Latin America
10:00 am - 10:30 am	Break, Exhibits
10:30 am - 12:30 pm	Session 2 of Workgroup Activities Featuring: Perinatal and Reproductive Health, Industry, Natural Language Processing, GIS - Geographic Information System, HADES Hackathon, Oncology, Common Data Model, ATLAS/WebAPI, Phenotype Development and Evaluation, Early-Stage Researchers, and Vocabularies
12:30 pm - 1:30 pm	Buffet Lunch and Exhibits
1:30 pm - 3:30 pm	Session 3 of Workgroup Activities Featuring: Surgery and Perioperative Medicine, Rare Diseases, Medical Devices, Psychiatry, HADES Hackathon, Health Equity, Evidence Network Data Partners, Eyecare and Vision Research, Women of OHDSI, CDM Survey
3:45 pm - 5:00 pm	Workgroup Summary

* agenda is subject to change

ohdsi.org/ohdsi2025

Global Symposium: Oct. 7-9

OHDSI

2025 Collaborator Showcase Presenters

October 7 – Pre Showcase - 6:00pm-8:00pm
October 8 – Collaborator Showcase
9:30am-10:15am, 2:45pm-3:30pm, 3:30pm-4:15pm

Community Building (#s 1-8)

1	Building the OHDSI Evidence Network – A Global, Open, Federated Collaboration	Clair Blacketer, Haeun Lee, Benjamin Martijn, Burrows, Ben Gerber, Pantelis Natsiavas, Aar Vadsariya, Hanieh Razzaghi, Paul Nagy
2	Characterizing the OHDSI Evidence Network – A Global Snapshot of Real-World Data Partners	Clair Blacketer, Evanette Burrows, Ben Gerbe Huser, Paul Nagy
3	Australian Health Data Evidence Network (AHDEN): Building a National Data Infrastructure for Standardised, Federated Health Data Research	Roger Ward, Nicole Pratt, Graeme Hart, Ilan Clair Sullivan, Blanca Gallego Luxan, Georgin
4	Progress and Challenges of the OHDSI Africa Chapter	Cynthia Sung, Agnes Kiragga, David Amadi, Yohannes Amare, Onana Akoa Anciet, Paulin Daniel Ankrah, Alex Asiimwe, Chidi Asuzu, To Bhattacharjee, Adam Bouras, Geert Byttebier Coorevits, Kluivert B. Duah, Luc Baudoin Fank Fourie Yacob Gebretensae, Jay Greenfield, La Halvorsen, Jared Houghtaling, Katherine Joh Andrew S. Kanter, Johnblack Kabukye, Mack Charlie Maere Maureen Ng'etich, Michael Oc Ogoe, Bolu Oluwalade, James Orwa, Nahend Garbya, Amelia Taylor, Marleen Temmermar Marc Twagirumukiza, Mirjam van Reisen, Ils Michel Walravens, Andrew Williams
5	From Fragmentation to Federation: A Multi-Partner OMOP Implementation in Uganda Enabling Global Real-World Evidence Generation	Francis Kanyike, Annet Nanungi, Harriet Dick Adam, James Brash, Thu Do, Caroline Otike, Bogart, Alex Asiimwe, Mui Van Zandt, Cissy Mutuluuza
6	OHDSI India Digital Health CoE and National Registry Pilots	Swetha, Parthi, Louis, Vikram, Anurag, Rintu
7	Data Coordinating Center for the OHDSI Ophthalmic Network: A Proposal for the NEI OHDSI Challenge	Michelle R. Hribar, Mohammad Adibuzzaman Brinks, Aiyin Chen, David Huang, Hiroshi Ishikawa, Yali Jia, Elizabeth Silbermann, Xubo Song, Ou Tan

Software Demonstrations (#s 501-516)

501	dqdbt: Continuous Data Quality Testing for OMOP ETL with dbt	Katy Sadowski, Lawrence Adams, Thomas Wylie
502	Summarizing FHIR® to OMOP Transformation Exceptions using Generative AI	Ron Sweeney, Hannah Kimura, Qi Li
503	Usagi-on-the-Web: A Cloud-Based Collaborative Platform for Vocabulary Mapping	Natthawut Adulyanukosol
504	Advancing Electronic Clinical Quality Measure (eCQM) Interoperability: Model Context Protocol (MCP)-Orchestrated CQL-to-OMOP Translation	Star Liu, Robert B Barrett, Kyle Zollo-Venecek, Benjamin Riesser, Benjamin Martin
505	Federated Platform for Clinical Data Mediation: Enhancing Interoperability with OMOP and NLP	Mónica Arrúe, María Quijada, Paula Chocrón, Josep Cordón, Gabriel de Maeztu
506	Enhancing OMOP Concept Mapping in Data2Evidence: A Comparative Study of Full-Text and Semantic Search	Zhi Min, Peter Hoffmann
507	The OMOP Annotator: A Database Agnostic Tool for Reviewing and Augmenting the Patient Record	Amy Yates, Erik Benton, Izabelle Humes, Matthew Lawhead, Heath Harrelson, Imogen Bentley, Rumel Mahmood, William Hersh, Steven Bedrick
508	Automated OMOP Concept Mapping Using Multi-Agent Large Language Models and Graph-Enhanced Semantic Retrieval	Adil Ahmed, Selvin Soby, Boudewijn Aasman, Parsa Mirhaji
509	EHR Browser: A Web Tool to Explore OMOP-CDM Health Records by Concept Hierarchy, Mappings, and Temporal Trends	Veronica Lorenzini, Javier Gracia-Tabuenca, Nicola Cerioli, FinnGen, Mary Pat Reeve
510	Advances in ARES: Evolving Observational Data Management and Systematic Review Capabilities	Frank DeFalco, Evanette Burrows, Clair Blacketer, Mikhail Iontsev
511	DarwinBenchmark: Evaluating cohort generation and analytics in OMOP CDM databases	Ioanna Nika, Maxim Moniat, Guido van Leeuwen, Ross Williams

Lightning Talks and Lightning Talk Posters (#s 601-610)

601	Bridging Standards: Creating OMOP data via Fast Healthcare Interoperability Resources (FHIR) and Health Information Networks	Stephanie Hong, Thanaphop Na Nakhonphanom, Andrew Laitman, Matthew Owens, Anne Bailey, Bryan Laraway, Tanner Zhang, Yvette Chen, Richard Moffitt, Rob Schuff, Tursynay Issabekova, Christopher Chute, Josh Lemieux, Melissa Haendel, William Hogan, Emily Pfaff, Shahim Essaid
602	OMOP Waveform Extension: A Schema for Integrating Physiological Signals and Derived Features into the OMOP CDM	Jared Houghtaling, Polina Talapova, Brian Gow, Manlik Kwong, Andrew J King, Benjamin Moody, Mike Kriley, Tom Pollard, Andrew E. Williams
603	Improving VSAC to OMOP Mapping Using LLM Assisted Curation	Robert Barrett, Star Liu, Kyle Zollo-Venecek, Benjamin Riesser, Benjamin Martin
604	Evaluating the effectiveness of using Large Language Models for the development of concept sets	Joel Swerdel, Dmytro Dymshyts, Anna Ostropolets, Azza Shoaibi, Patrick Ryan, Martijn Schuemie
605	Validating a Scalable Approach to Data Fitness-for-Use: Database Diagnostics Applied to LEGEND-T2DM	Clair Blacketer, Patrick B. Ryan, George Hripcsak, Marc Suchard, Fan Bu, Can Yin, Martijn J. Schuemie, Peter R. Rijnbeek
606	Causal Inference with Multi-Modal Foundation Models: A Case Study of Anti-VEGF Injections in Diabetic Macular Edema	Siqi Sun, Cindy X. Cai, Ruochong Fan, Saiyu You, Diep Tran, P. Kumar Rao, Marc A. Suchard, Yixin Wang, Linying Zhang
607	LATTE: A One-shot Lossless Algorithm for Federated Target Trial Emulation with Application to Alzheimer's Disease and Related Dementia Drug Repurposing Using Decentralized Data	Lu Li, Qiong Wu, Yiwen Lu, Kyra S. O'Brien, Bingyu Zhang, Ting Zhou, Jiayi Tong, Dazheng Zhang, Yuqing Lei, Huilin Tang, Yun Lu, David Asch, Yong Chen
608	From Data Quality to Clinical Quality – Episodes as Enablers for Next Generation Dashboarding	Georgina Kennedy, Shalini Vinod, Gui Mei Xiong, Nasreen Kaadan, Merran Findlay, April Matt, Marnie Harris, Arya Shinde, Shuang Liang, Carolyn Mazariego, Tim Churches, Louisa Jorm, Victoria Bray, Angela Berthelsen, Phan Sayaloune, Geoff Delaney
609	Heterogeneity of Treatment Effects Across Nine Glucose-Lowering Drug Classes in Type 2 Diabetes: Extension of the LEGEND-T2DM Network Study	Hsin Yi Chen, Thomas Falconer, Anna Ostropolets, Tara V. Anand, Xinzhuo Jiang, David Dávila-García, Linying Zhang, Ruochong Fan, George Hripcsak
610	DARWIN EU® - A multi-national network cohort and self-controlled case series study of the effect of doxycycline versus active comparators on the risk of suicidality in individuals with acne	Nicholas B. Hunt, Guido J. van Leeuwen, Maarten van Kessel, Anna Palomar-Cros, Antonella Delmestri, Agustina Giuliodori, Talita Duarte Salles, Mandickel Kamtengeni, Ross D. Williams, Daniel Prieto Alhambra, Katia Verhamme

ohdsi.org/ohdsi2025

Africa Symposium: Nov. 10-12

The first-ever OHDSI Africa Symposium will be held Nov. 10-12 in Kampala, Uganda, at the Joint Clinical Research Centre (JCRC) and Mestil Hotel. The event will begin with a dedicated one-day training course at JCRC, followed by a two-day main conference at the Mestil Hotel. Here are some important dates for you to save to your calendar:

Collaborator Showcase

- Submissions deadline: passed
- Submissions review: September 10 – 30
- Notification of acceptance: October 5

ohdsi.org/africa2025

APAC Symposium: Dec. 6-7

The 2025 OHDSI APAC Symposium will be held Dec. 6-7 in Shanghai, China at the Shanghai Jiao Tong University. It will feature a 1-day tutorial and a 1-day main conference. Here are some important dates for you to save to your calendar:

Collaborator Showcase

- Submissions deadline: passed
- Submissions review: September 8 – October 9
- Notification of acceptance: October 17

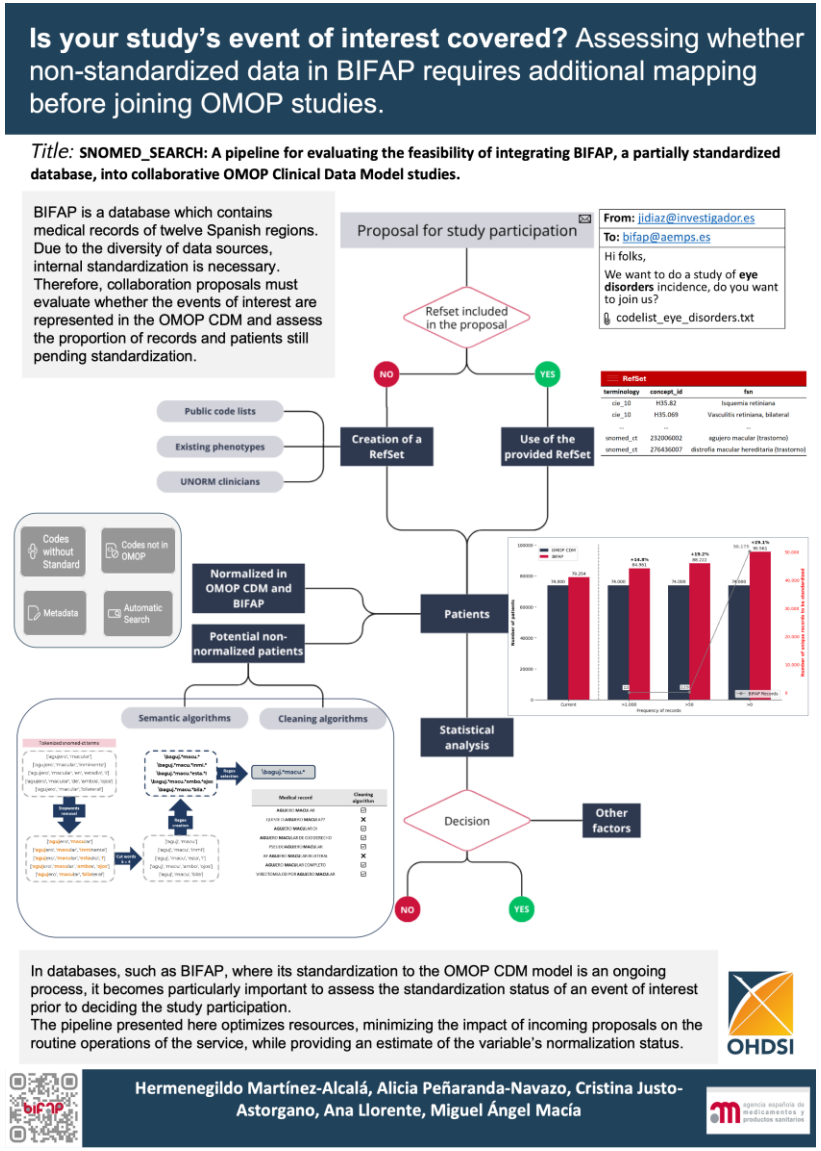
ohdsi.org/apac2025

#OHDSISocialShowcase This Week

Monday

SNOMED_SEARCH: A pipeline for evaluating the feasibility of integrating BIFAP, a partially standardized database, into collaborative OMOP Clinical Data Model studies

(Hermenegildo Martinez-Alcala, Cristina Justo-Astorgano, Alicia Peñaranda-Navazo, Ana Llorente, Miguel Ángel Macía)

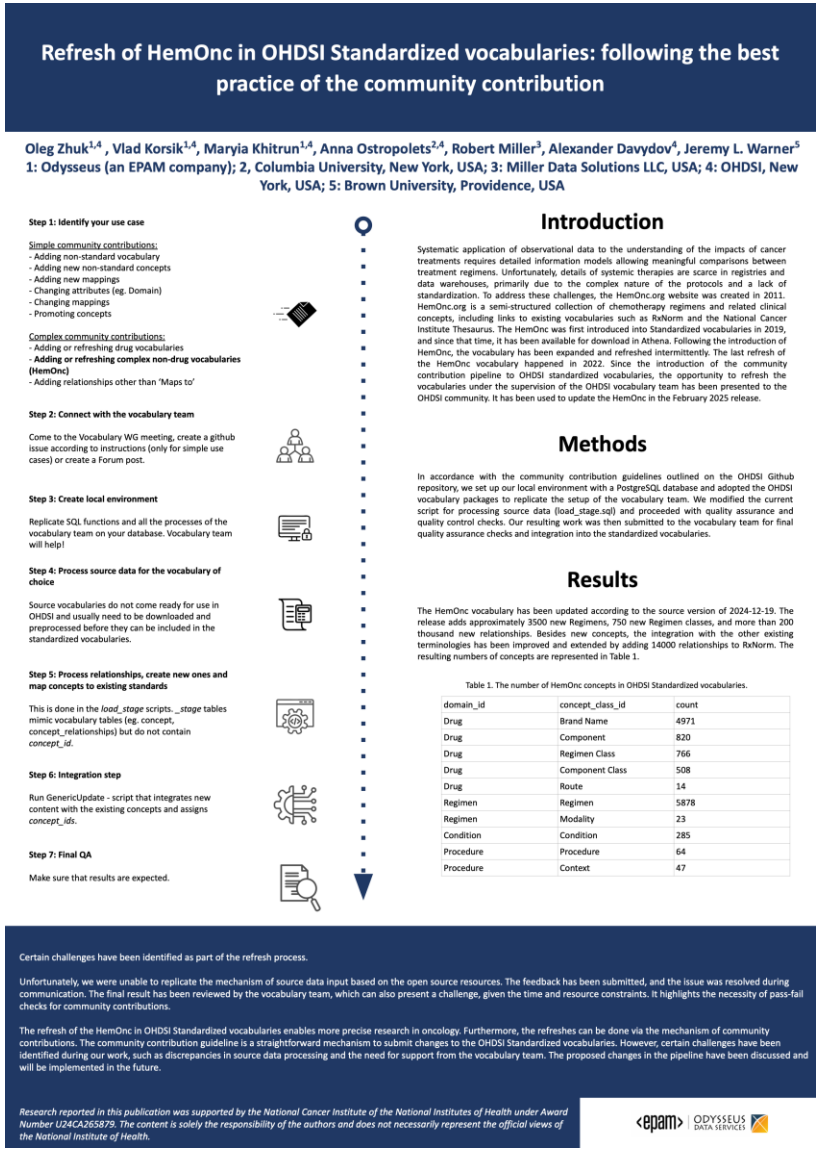


#OHDSISocialShowcase This Week

Tuesday

Refresh of the HemOnc in OHDSI Standardized vocabularies: following the best practice of the community contribution

(Oleg Zhuk, Vlad Korsik, Maryia Khitrun, Anna Ostropolets, Robert Miller, Alexander Davydov, Jeremy Warner)



Refresh of HemOnc in OHDSI Standardized vocabularies: following the best practice of the community contribution

Oleg Zhuk[1,4] , Vlad Korsik[1,4], Maryia Khitrun[1,4], Anna Ostropolets[2,4], Robert Miller[3], Alexander Davydov[4], Jeremy L. Warner[5]
1: Odysseus (an EPAM company); 2, Columbia University, New York, USA; 3: Miller Data Solutions LLC, USA; 4: OHDSI, New York, USA; 5: Brown University, Providence, USA

Step 1: Identify your use case

Simple community contributions:
- Adding non-standard vocabulary
- Adding new non-standard concepts
- Adding new mappings
- Changing attributes (eg. Domain)
- Changing mappings
- Promoting concepts

Complex community contributions:
- Adding or refreshing drug vocabularies
- **Adding or refreshing complex non-drug vocabularies (HemOnc)**
- Adding relationships other than 'Maps to'

Step 2: Connect with the vocabulary team

Come to the Vocabulary WG meeting, create a github issue according to instructions (only for simple use cases) or create a Forum post.

Step 3: Create local environment

Replicate SQL functions and all the processes of the vocabulary team on your database. Vocabulary team will help!

Step 4: Process source data for the vocabulary of choice

Source vocabularies do not come ready for use in OHDSI and usually need to be downloaded and preprocessed before they can be included in the standardized vocabularies.

Step 5: Process relationships, create new ones and map concepts to existing standards

This is done in the *load_stage* scripts. *_stage* tables mimic vocabulary tables (eg. concept, concept_relationships) but do not contain *concept_id*.

Step 6: Integration step

Run GenericUpdate - script that integrates new content with the existing concepts and assigns *concept_ids*.

Step 7: Final QA

Make sure that results are expected.

Introduction

Systematic application of observational data to the understanding of the impacts of cancer treatments requires detailed information models allowing meaningful comparisons between treatment regimens. Unfortunately, details of systemic therapies are scarce in registries and data warehouses, primarily due to the complex nature of the protocols and a lack of standardization. To address these challenges, the HemOnc.org website was created in 2011. HemOnc.org is a semi-structured collection of chemotherapy regimens and related clinical concepts, including links to existing vocabularies such as RxNorm and the National Cancer Institute Thesaurus. The HemOnc was first introduced into Standardized vocabularies in 2019, and since that time, it has been available for download in Athena. Following the introduction of HemOnc, the vocabulary has been expanded and refreshed intermittently. The last refresh of the HemOnc vocabulary happened in 2022. Since the introduction of the community contribution pipeline to OHDSI standardized vocabularies, the opportunity to refresh the vocabularies under the supervision of the OHDSI vocabulary team has been presented to the OHDSI community. It has been used to update the HemOnc in the February 2025 release.

Methods

In accordance with the community contribution guidelines outlined on the OHDSI Github repository, we set up our local environment with a PostgreSQL database and adopted the OHDSI vocabulary packages to replicate the setup of the vocabulary team. We modified the current script for processing source data (load_stage.sql) and proceeded with quality assurance and quality control checks. Our resulting work was then submitted to the vocabulary team for final quality assurance checks and integration into the standardized vocabularies.

Results

The HemOnc vocabulary has been updated according to the source version of 2024-12-19. The release adds approximately 3500 new Regimens, 750 new Regimen classes, and more than 200 thousand new relationships. Besides new concepts, the integration with the other existing terminologies has been improved and extended by adding 14000 relationships to RxNorm. The resulting numbers of concepts are represented in Table 1.

Table 1. The number of HemOnc concepts in OHDSI Standardized vocabularies.

domain_id	concept_class_id	count
Drug	Brand Name	4971
Drug	Component	820
Drug	Regimen Class	766
Drug	Component Class	508
Drug	Route	14
Regimen	Regimen	5878
Regimen	Modality	23
Condition	Condition	285
Procedure	Procedure	64
Procedure	Context	47

Certain challenges have been identified as part of the refresh process.

Unfortunately, we were unable to replicate the mechanism of source data input based on the open source resources. The feedback has been submitted, and the issue was resolved during communication. The final result has been reviewed by the vocabulary team, which can also present a challenge, given the time and resource constraints. It highlights the necessity of pass-fail checks for community contributions.

The refresh of the HemOnc in OHDSI Standardized vocabularies enables more precise research in oncology. Furthermore, the refreshes can be done via the mechanism of community contributions. The community contribution guideline is a straightforward mechanism to submit changes to the OHDSI Standardized vocabularies. However, certain challenges have been identified during our work, such as discrepancies in source data processing and the need for support from the vocabulary team. The proposed changes in the pipeline have been discussed and will be implemented in the future.

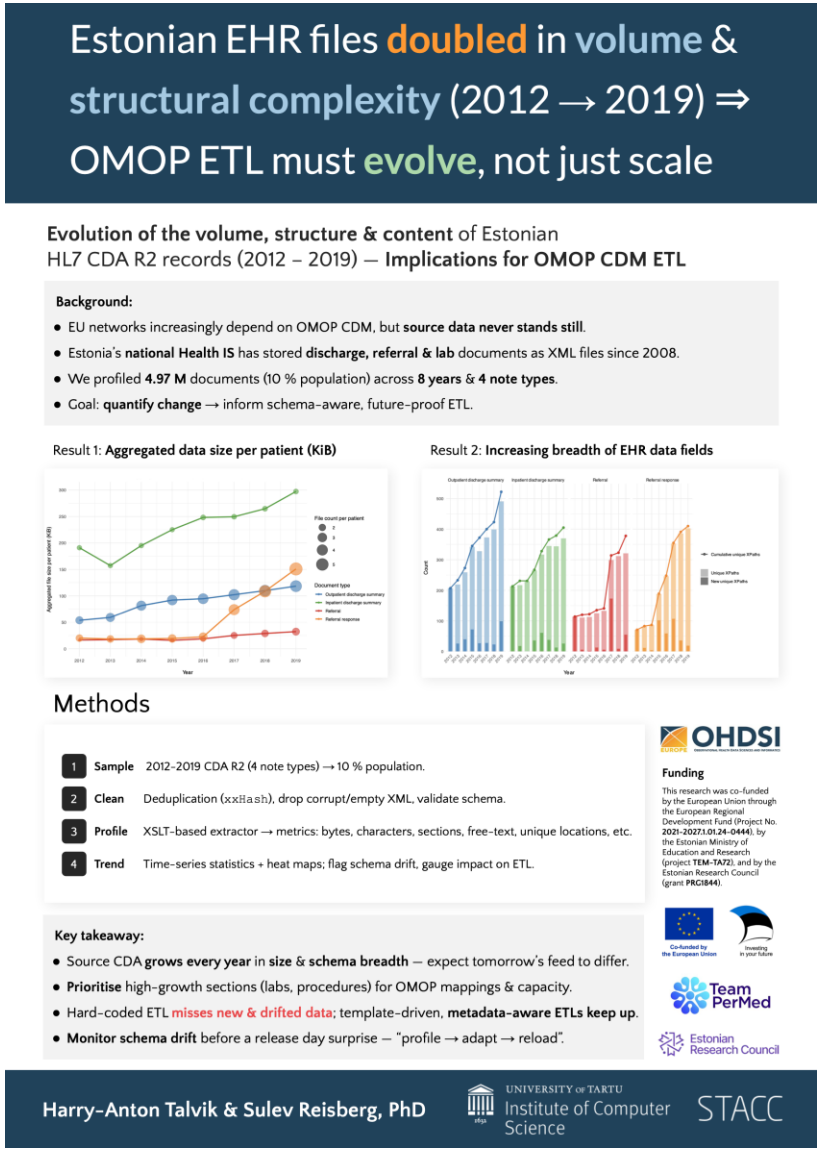
Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number U24CA265879. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health.

#OHDSISocialShowcase This Week

Wednesday

Standardized Terminology Gap Analysis In OHDSI Vocabularies For The Critical Care Domain And Modular Extension Strategy: Progress From The Bridge2AI For Clinical Care CHoRUS Project

(Polina Talapova, Jared Houghtaling, Andrew Williams, Manlik Kwong, Gilles Clermont, Soojin Park, Brian Gow, Tom Pollard)



#OHDSISocialShowcase This Week

Thursday

Moving towards standard observation period settings for OMOP-based EHR data: results from the HERON-UK network

(Andy South, Hiba Junaid, Jennifer Lane, Usama Rahman, Ben Eaton, Xavier Griffin, Ana Cavalcante, Steve Harris, Ofran Almossawi, Lydia Briggs, Daniel Key, Timothy Howcroft, Vishnu Chandrabalan, Peter S Hall, Mahéva Vallet, Marta Pasikowska, Nicola Symmers, Colin McLean, Spyro Nita, Daniel Dedman, Zara Cuccu, Stelios Theophanous, Geoff Hall, Edward Bolton, Cecilia Campanile, Elin Rowlands, Danielle Newby, Daniel Prieto Alhambra, Edward Burn, Marti Catala)

Incidence estimates can vary 10-fold depending on observation period definition

Moving towards standard observation period settings for OMOP-based EHR data: results from the HERON-UK network

Background: Observation period in the OMOP CDM is used to define when a patient can be included in an analysis. Defining the start, and particularly the stop, of an observation period from EHR is not straightforward, and decisions when setting these will affect future study results. We characterised 12 different definitions of observation period and evaluated their impact in incidence rates.

Methods

Observation period definitions: (1) First to data extraction; (2) First to last record; (3) Inpatient hospitalization; (4) Active record; (5) Active + 180 days persistence; (6) Active + 180 days surveillance; (7) 365 days persistence; (8) 365 days surveillance; (9) 545 days persistence; (10) 545 days surveillance; (11) 730 days persistence; (12) 730 days surveillance. Persistence: concatenate records within window. Surveillance: add time post-record, merge overlaps.

Characterisation: number of total observation periods, number of records per person, length of observation periods and time between observation periods.

Incidence: ciprofloxacin incidence rate was calculated during 2023 with a washout of 30 days.

Data Sources: Barts Health and UCLH are London hospitals. GOSH is a pediatric hospital, IDRIL Lancashire hospital and CPRD AURUM is primary care.

Results

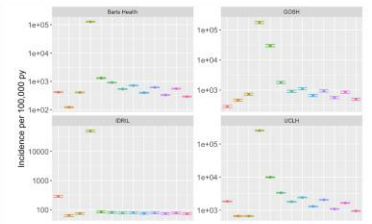


Figure 1. Incidence per 100,000 person-years of ciprofloxacin in 2023.

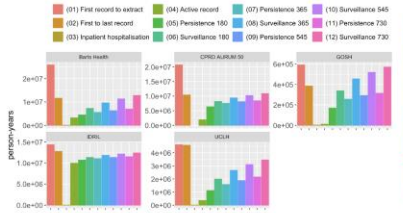


Figure 2. Number of person-years included in each one of the observation periods definition.

Table 1. Characterisation of observation periods. *median and IQR of length in days. †median and IQR of time between observation periods. ‡median and IQR of number of records per person.

Conclusions: The Observation period definitions can greatly impact study outcomes, such as incidence rates. To ensure consistency, the HERON-UK network recommends using '**first record to extraction**' as default. This captures the full data history and leverages national mortality data for reliable end dates. It also allows **flexible sensitivity analyses** by shortening periods within study code.

This work is supported by HDRUK: https://www.hdruk.ac.uk/about-us/funders/.

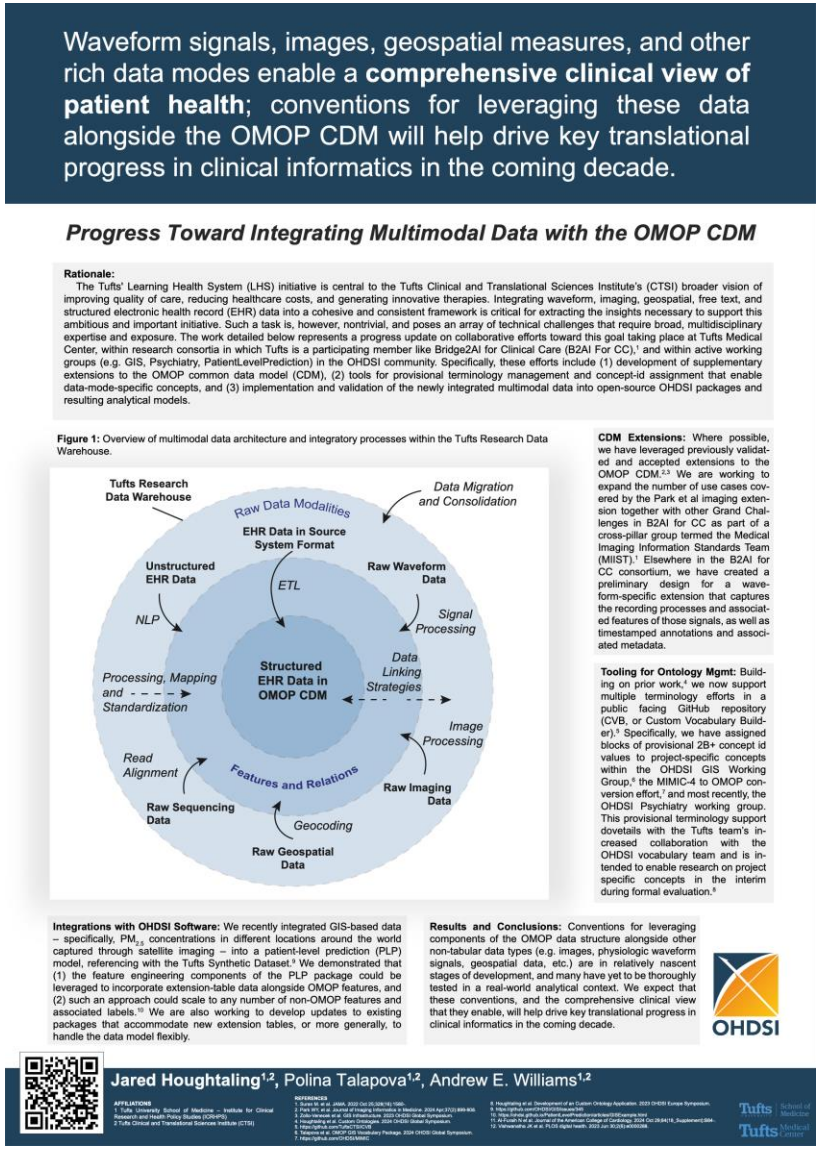
Andy South[1], Hiba Junaid[2,3,4], Jennifer Lane[2,3], Usama Rahman[2,3], Ben Eaton[2,3], Xavier Griffin[2,3], Ana Cavalcante[5], Steve Harris[5,11], Ofran Almossawi[6], Lydia Briggs[6], Daniel Key[6], Timothy Howcroft[7], Vishnu Chandrabalan7, Peter S Hall[8], Mahéva Vallet[8], Marta Pasikowska[8], Nicola Symmers[8], Colin McLean[8], Spyro Nita[8], Daniel Dedman[9], Zara Cuccu[9], Stelios Theophanous[10], Geoff Hall[10], Edward Bolton[10], Cecilia Campanile[1], Elin Rowlands[1], Danielle Newby[1], Daniel Prieto-Alhambra[1,12], Edward Burn[1], Marti Catala[1]
[1]University of Oxford, Oxford [2] Queen Mary University [3]Barts Life Sciences [5] UCLH [6]GOSH [7]Lancashire Teaching Hospitals [8] DataLoch [9] CPRD, MHRA [10] Leeds Teaching Hospitals [11] Institute of Health Informatics [12] Erasmus MC

#OHDSISocialShowcase This Week

Friday

Progress Toward Integrating Multimodal Data with the OMOP CDM

(Jared Houghtaling, Polina Talapova, Andrew E. Williams)



Waveform signals, images, geospatial measures, and other rich data modes enable a **comprehensive clinical view of patient health**; conventions for leveraging these data alongside the OMOP CDM will help drive key translational progress in clinical informatics in the coming decade.

Progress Toward Integrating Multimodal Data with the OMOP CDM

Rationale:
The Tufts' Learning Health System (LHS) initiative is central to the Tufts Clinical and Translational Sciences Institute's (CTSI) broader vision of improving quality of care, reducing healthcare costs, and generating innovative therapies. Integrating waveform, imaging, geospatial, free text, and structured electronic health record (EHR) data into a cohesive and consistent framework is critical for extracting the insights necessary to support this ambitious and important initiative. Such a task is, however, nontrivial, and poses an array of technical challenges that require broad, multidisciplinary expertise and exposure. The work detailed below represents a progress update on collaborative efforts toward this goal taking place at Tufts Medical Center, within research consortia in which Tufts is a participating member like Bridge2AI for Clinical Care (B2AI For CC),[1] and within active working groups (e.g. GIS, Psychiatry, PatientLevelPrediction) in the OHDSI community. Specifically, these efforts include (1) development of supplementary extensions to the OMOP common data model (CDM), (2) tools for provisional terminology management and concept-id assignment that enable data-mode-specific concepts, and (3) implementation and validation of the newly integrated multimodal data into open-source OHDSI packages and resulting analytical models.

Figure 1: Overview of multimodal data architecture and integratory processes within the Tufts Research Data Warehouse.

CDM Extensions: Where possible, we have leveraged previously validated and accepted extensions to the OMOP CDM.[2,3] We are working to expand the number of use cases covered by the Park et al imaging extension together with other Grand Challenges in B2AI for CC as part of a cross-pillar group termed the Medical Imaging Information Standards Team (MIIST).[1] Elsewhere in the B2AI for CC consortium, we have created a preliminary design for a waveform-specific extension that captures the recording processes and associated features of those signals, as well as timestamped annotations and associated metadata.

Tooling for Ontology Mgmt: Building on prior work,[4] we now support multiple terminology efforts in a public facing GitHub repository (CVB, or Custom Vocabulary Builder).[5] Specifically, we have assigned blocks of provisional 2B+ concept id values to project-specific concepts within the OHDSI GIS Working Group,[6] the MIMIC-4 to OMOP conversion effort,[7] and most recently, the OHDSI Psychiatry working group. This provisional terminology support dovetails with the Tufts team's increased collaboration with the OHDSI vocabulary team and is intended to enable research on project specific concepts in the interim during formal evaluation.[8]

Integrations with OHDSI Software: We recently integrated GIS-based data – specifically, $PM_{2.5}$ concentrations in different locations around the world captured through satellite imaging – into a patient-level prediction (PLP) model, referencing with the Tufts Synthetic Dataset.[9] We demonstrated that (1) the feature engineering components of the PLP package could be leveraged to incorporate extension-table data alongside OMOP features, and (2) such an approach could scale to any number of non-OMOP features and associated labels.[10] We are also working to develop updates to existing packages that accommodate new extension tables, or more generally, to handle the data model flexibly.

Results and Conclusions: Conventions for leveraging components of the OMOP data structure alongside other non-tabular data types (e.g. images, physiologic waveform signals, geospatial data, etc.) are in relatively nascent stages of development, and many have yet to be thoroughly tested in a real-world analytical context. We expect that these conventions, and the comprehensive clinical view that they enable, will help drive key translational progress in clinical informatics in the coming decade.

OHDSI

Jared Houghtaling[1,2], Polina Talapova[1,2], Andrew E. Williams[1,2]

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?

Sept. 23: Education in OHDSI - Lessons Learned

George Hripcsak

Vivian Beaumont Allen Professor of Biomedical Informatics, Columbia University

Topic: OHDSI Summer School at Columbia DBMI

Dani Prieto-Alhambra

Section Head and Professor in Health Data Sciences, University of Oxford
Deputy Director of DARWIN EU Coordination Centre and Professor, Erasmus MC

Topic: Real World Evidence Summer School at Oxford

Paul Nagy

Head of Biomedical Informatics and Associate Professor, Johns Hopkins University

Topic: OHDSI in Johns Hopkins Postgraduate Education / OHDSI Maternal Health Fellowship

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-2025