OHDSI/OMOP Research Spotlight

OHDSI Community Call

Sept. 16, 2025 • 11 am ET

Upcoming Community Calls

Date	Topic
Sept. 16	OHDSI/OMOP Research Spotlight
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. 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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Clair Blacketer, Frank DeFalco, Mitchell Conover, Patrick Ryan, Martijn Schuemie, and Peter Rijnbeek** on the publication of **Evaluation of the impact of defining observable time in real-world data on outcome incidence** in *JAMIA*.

Journal of the American Medical Informatics Association, 2025, **32(9)**, 1434–1444
https://doi.org/10.1093/jamia/ocaf119
Advance access publication 22 July 2025
Research and Applications

Research and Applications

Evaluation of the impact of defining observable time in real-world data on outcome incidence

Clair Blacketer, MPH*[1,2,3], Frank J. DeFalco, BA[1,3], Mitchell M. Conover, PhD[1,3], Patrick B. Ryan, PhD[1,3,4], Martijn J. Schuemie, PhD[1,5], Peter R. Rijnbeek, PhD[1,2]

[1]Coordinating Center, Observational Health Data Sciences and Informatics (OHDSI), New York, NY, 10032, United States, [2]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, NL, 3015 GD, United States, [3]Johnson & Johnson, Raritan, NJ, 08869, United States, [4]Department of Biomedical Informatics, Columbia University, New York, NY, 10032, United States, [5]Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, 90095, United States

*Corresponding author: Clair Blacketer, MPH, Observational Health Data Analytics, Johnson & Johnson, 920 US Route 202, Raritan, NJ 08869, United States (mblacke@its.jnj.com)

Abstract

Objective: In real-world data (RWD), defining the observation period—the time during which a patient is considered observable—is critical for estimating incidence rates (IRs) and other outcomes. Yet, in the absence of explicit enrollment information, this period must often be inferred, introducing potential bias.

Materials and Methods: This study evaluates methods for defining observation periods and their impact on IR estimates across multiple database types. We applied 3 methods for defining observation periods: (1) a persistence + surveillance window approach, (2) an age- and gender-adjusted method based on time between healthcare events, and (3) the min/max method. These were tested across 11 RWD databases, including both enrollment-based and encounter-based sources. Enrollment time was used as the reference standard in eligible databases. To assess the impact on epidemiologic results, we replicated a prior study of adverse event incidence, comparing IRs and calculating mean squared error between methods.

Results: Incidence rates decreased as observation periods lengthened, driven by increases in the person-time denominator. The persistence + surveillance method produced estimates closest to enrollment-based rates when appropriately balanced. The min/max approach yielded inconsistent results, particularly in encounter-based databases, with greater error observed in databases with longer time spans.

Discussion: These findings suggest that assumptions about data completeness and population observability significantly affect incidence estimates. Observation period definitions substantially influence outcome measurement in RWD studies.

Conclusion: Standardized, transparent approaches are necessary to ensure valid, reproducible results—especially in databases lacking defined enrollment.

Key words: observation period; real-world data; incidence rates; data standardization; data quality.

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
Tuesday	12 pm	CDM Vocabulary Subgroup
Wednesday	8 am	Psychiatry
Wednesday	9 am	Health Economics and Value Assessment (HEVA)
Wednesday	11 am	Common Data Model
Wednesday	1 pm	Perinatal and Reproductive Health
Wednesday	7 pm	Medical Imaging
Thursday	8 am	India Community Call
Thursday	9 am	Oncology Vocabulary/Development Subgroup
Thursday	11 am	Themis
Thursday	12 pm	HADES
Thursday	7 pm	Dentistry
Friday	10 am	GIS-Geographic Information System
Friday	10 am	Transplant
Friday	10:30 am	Open-Source Community
Friday	11:30 am	Steering
Friday	2 pm	Vaccine Vocabulary
Monday	10 am	Healthcare Systems Interest Group
Tuesday	9 am	Data2Evidence
Tuesday	9 am	Oncology Genomic 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

Science for a Sustainable Future: Showcasing Science Collaboration

The role and contribution of **science in attaining the United Nations Sustainable Development Goals (SDGs)** will be the central theme of the Science Summit. The objective is to enable science collaborations to demonstrate how science supports the attainment of the UN SDGs and Agenda 2030.

The Summit will examine what **enabling policy, regulatory and financial environments** are needed to implement and sustain the science mechanisms required to support genuinely global scientific collaborations across continents, nations and themes.

Scientific discovery through the analysis of massive data sets is at hand. This data-enabled approach to science, research and development will be necessary if the SDGs are to be achieved.

SCIENCE FOR GLOBAL CHALLENGES →

Full programme is here

https://sciencesummitnyc.org/

Sep, Thu 18 | 08:30 AM - 10:30 AM | Virtual | Public

Standardizing Health Data and Analytics to Accelerate Clinical Impact and Global Reach: Part 1

Theme: Digital / AI

Observational Health Data Science and Informatics (OHDSI) is a global community that uses data harmonized to the OMOP Common Data Model, standardized vocabulary, data quality checks and validated analytics to produce rigorous evaluation of big data from existing health databases. Through sharing of computer codes and summary statistics instead of patient-level data, OHDSI preserves privacy while enabling collaboration across institutions, countries, and continents. Large-scale, real-world studies through OHDSI network collaborations have revealed valuable insights into clinical care and public health.

Speakers:

Agnes Kiragga
Global Health Leader,...

Organization: African Population Health and Research Centre

Chan Seng You
Assistant Professor

Organization: Yonsei University College of Medicine

NicolePratt
Professor, Biostatistic...

Organization: University of South Australia

George Hripcsak
Professor, Biomedical...

Organization: Columbia University

Register | Session details

Sep, Thu 18 | 11:00 AM - 12:45 PM | Virtual | Public

Standardizing Health Data and Analytics to Accelerate Clinical Impact and Global Reach: Part 2

Theme: Digital / AI

Observational Health Data Science and Informatics (OHDSI) is a global community that uses data harmonized to the OMOP Common Data Model, standardized vocabulary, data quality checks and validated analytics to produce large-scale evaluation of real world data. Through sharing of computer codes and summary statistics instead of patient-level data, OHDSI preserves privacy while enabling collaboration across institutions, countries, and continents. Large-scale, real-world studies by OHDSI members have revealed valuable insights into clinical care and public health.

Speakers:

Cynthia Sung
Adjunct Associate...

Organization: Duke-NUS Medical School Centre of Regulatory Excellence

Patrick Ryan
VP Janssen...

Organization: OHDSI Observational Health Data Science and Informatics

Katia Verhamme
Associate Professor of...

Organization: Erasmus University Medical Center

Peter Rijnbeek
Professor, Medical...

Organization: Erasmus University Medical Center

Julio Oliveira
CEO

Organization: Precision Data

Registration links

Part 1 Sep 18, 8:30-10:30 EDT: https://event.sciencesummitnyc.org/list-of-sessions/detail/131

Part 2: Sep 18, 11:00-13:00 EDT https://event.sciencesummitnyc.org/list-of-sessions/detail/130

Full programme here: https://event.sciencesummitnyc.org/list-of-sessions

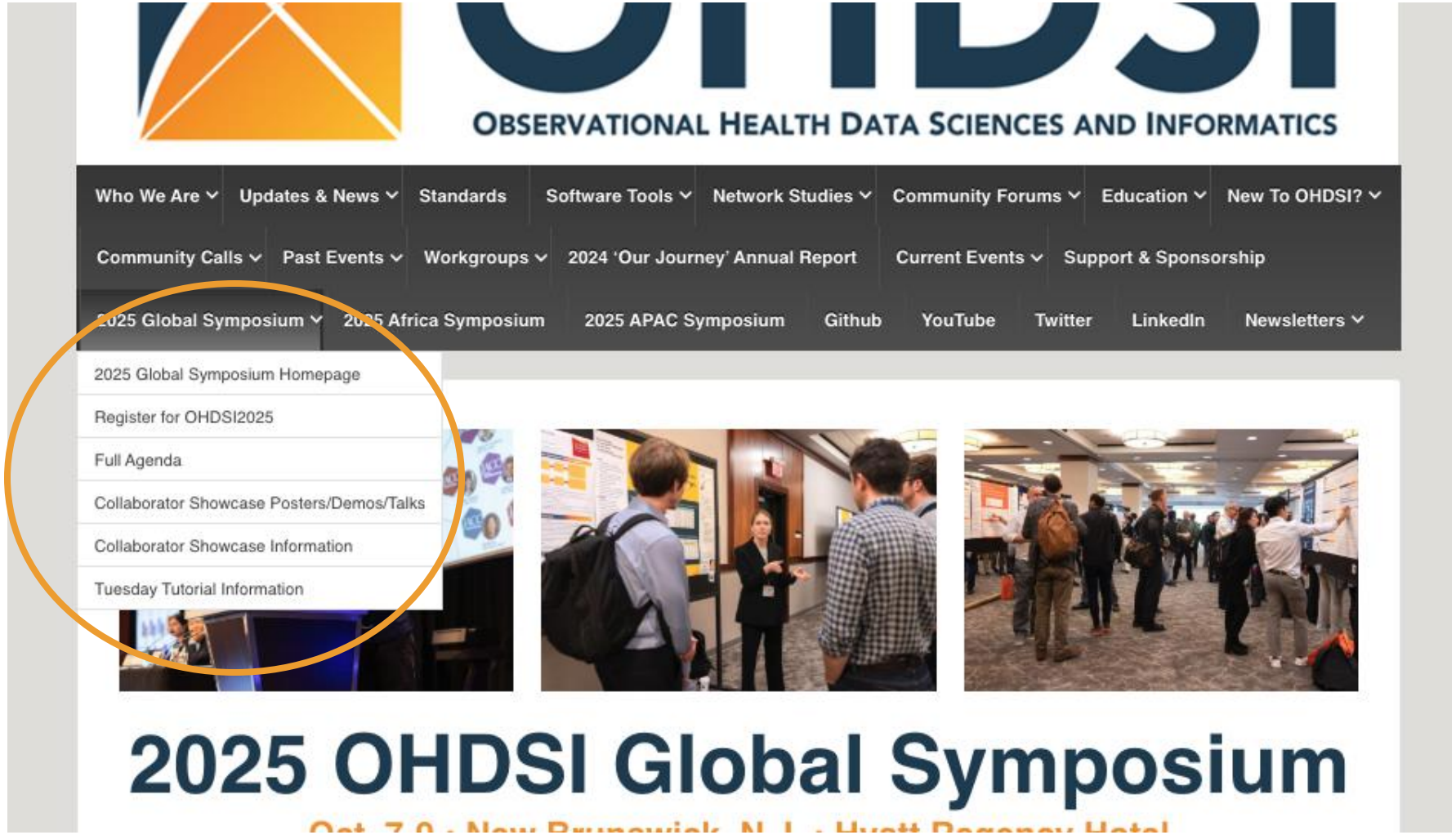
Part 1 (8:30 am ET)

1. *Observational Health Data Science and Informatics (OHDSI): Inclusive and Collaborative Science.* George Hripcsak
2. *Promoting Data Harmonization and Data Science in Africa*. Agnes Kiragga
3. *Rapid Response to the Covid-19 Pandemic Using a National Scale Database.* Chan Seng You
4. *OHDSI in Asia and the Pacific Rim.* Nicole Pratt
5. *Q&A Session*

Part 2 (11 am ET)

1. *Enabling Reliable Evidence Generation from Real-word Data in Europe.* Peter Rijnbeek
2. *DARWIN-EU® – Delivering Real World Evidence to Support Regulatory Decision-making by the European Medicines Agency.* Katia Verhamme
3. *OHDSI Adoption and Current Implementation Landscape in Latin America.* Julio Cesar Barbour Oliveira
4. *Learning Opportunities for OHDSI Skills Development.* Cynthia Sung
5. *Clinical and Public Health Impact of OHDSI.* Patrick Ryan
6. *Q&A Session*

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

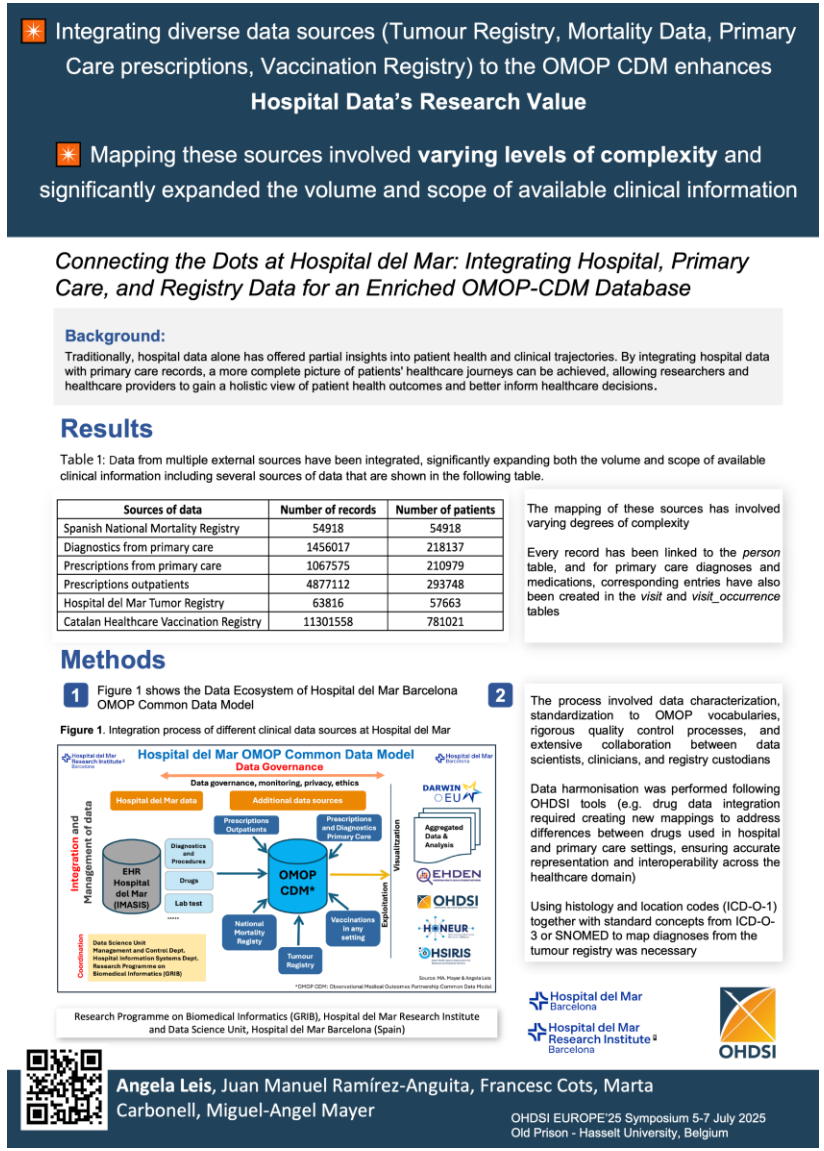
ohdsi.org/apac2025

#OHDSISocialShowcase This Week

Monday

Connecting the dots at Hospital del Mar: integrating hospital, primary care and registry data for an enriched OMOP-CDM database

(Angela Leis, Juan Manuel Ramírez-Anguita, Miguel Angel Mayer)

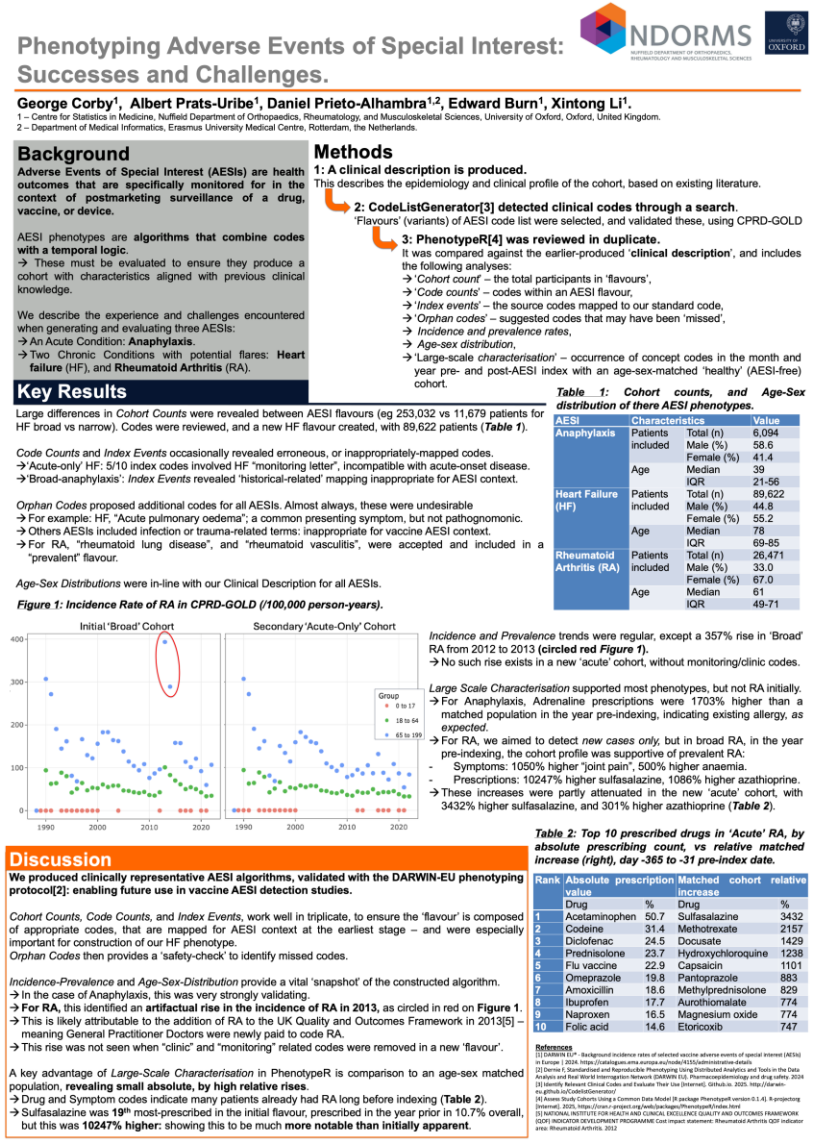


#OHDSISocialShowcase This Week

Tuesday

Phenotyping Adverse Events of Special Interest: Successes and Challenges

(George Corby, Albert Prats-Uribe, Daniel Prieto-Alhambra, Edward Burn, Xintong Li)



Phenotyping Adverse Events of Special Interest: Successes and Challenges.

George Corby[1], Albert Prats-Uribe[1], Daniel Prieto-Alhambra[1,2], Edward Burn[1], Xintong Li[1].

1 – Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom.
2 – Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, the Netherlands.

Background

Adverse Events of Special Interest (AESIs) are health outcomes that are specifically monitored for in the context of postmarketing surveillance of a drug, vaccine, or device.

AESI phenotypes are **algorithms that combine codes with a temporal logic**.
→ These must be evaluated to ensure they produce a cohort with characteristics aligned with previous clinical knowledge.

We describe the experience and challenges encountered when generating and evaluating three AESIs:
→ An Acute Condition: **Anaphylaxis**.
→ Two Chronic Conditions with potential flares: **Heart failure** (HF), and **Rheumatoid Arthritis** (RA).

Methods

1: A clinical description is produced.
This describes the epidemiology and clinical profile of the cohort, based on existing literature.

2: CodeListGenerator[3] detected clinical codes through a search.
'Flavours' (variants) of AESI code list were selected, and validated these, using CPRD-GOLD

3: PhenotypeR[4] was reviewed in duplicate.
It was compared against the earlier-produced '**clinical description**', and includes the following analyses:
→ '*Cohort count*' – the total participants in 'flavours',
→ '*Code counts*' – codes within an AESI flavour,
→ '*Index events*' – the source codes mapped to our standard code,
→ '*Orphan codes*' – suggested codes that may have been 'missed',
→ *Incidence and prevalence rates*,
→ *Age-sex distribution*,
→ 'Large-scale *characterisation*' – occurrence of concept codes in the month and year pre- and post-AESI index with an age-sex-matched 'healthy' (AESI-free) cohort.

Key Results

Large differences in *Cohort Counts* were revealed between AESI flavours (eg 253,032 vs 11,679 patients for HF broad vs narrow). Codes were reviewed, and a new HF flavour created, with 89,622 patients (***Table 1***).

Code Counts and *Index Events* occasionally revealed erroneous, or inappropriately-mapped codes.
→'Acute-only' HF: 5/10 index codes involved HF "monitoring letter", incompatible with acute-onset disease.
→'Broad-anaphylaxis': *Index Events* revealed 'historical-related' mapping inappropriate for AESI context.

Orphan Codes proposed additional codes for all AESIs. Almost always, these were undesirable
→ For example: HF, "Acute pulmonary oedema"; a common presenting symptom, but not pathognomonic.
→ Others AESIs included infection or trauma-related terms: inappropriate for vaccine AESI context.
→ For RA, "rheumatoid lung disease", and "rheumatoid vasculitis", were accepted and included in a "prevalent" flavour.

Age-Sex Distributions were in-line with our Clinical Description for all AESIs.

Table 1: Cohort counts, and Age-Sex distribution of there AESI phenotypes.

AESI	Characteristics		Value
Anaphylaxis	Patients included	Total (n)	6,094
		Male (%)	58.6
		Female (%)	41.4
	Age	Median	39
		IQR	21-56
Heart Failure (HF)	Patients included	Total (n)	89,622
		Male (%)	44.8
		Female (%)	55.2
	Age	Median	78
		IQR	69-85
Rheumatoid Arthritis (RA)	Patients included	Total (n)	26,471
		Male (%)	33.0
		Female (%)	67.0
	Age	Median	61
		IQR	49-71

Figure 1: Incidence Rate of RA in CPRD-GOLD (/100,000 person-years).

Incidence and Prevalence trends were regular, except a 357% rise in 'Broad' RA from 2012 to 2013 (**circled red *Figure 1***).
→ No such rise exists in a new 'acute' cohort, without monitoring/clinic codes.

Large Scale Characterisation supported most phenotypes, but not RA initially.
→ For Anaphylaxis, Adrenaline prescriptions were 1703% higher than a matched population in the year pre-indexing, indicating existing allergy, *as expected*.
→ For RA, we aimed to detect *new cases only*, but in broad RA, in the year pre-indexing, the cohort profile was supportive of prevalent RA:
- Symptoms: 1050% higher "joint pain", 500% higher anaemia.
- Prescriptions: 10247% higher sulfasalazine, 1086% higher azathioprine.
→ These increases were partly attenuated in the new 'acute' cohort, with 3432% higher sulfasalazine, and 301% higher azathioprine (***Table 2***).

Table 2: Top 10 prescribed drugs in 'Acute' RA, by absolute prescribing count, vs relative matched increase (right), day -365 to -31 pre-index date.

Rank	Absolute prescription value		Matched cohort relative increase	
	Drug	%	Drug	%
1	Acetaminophen	50.7	Sulfasalazine	3432
2	Codeine	31.4	Methotrexate	2157
3	Diclofenac	24.5	Docusate	1429
4	Prednisolone	23.7	Hydroxychloroquine	1238
5	Flu vaccine	22.9	Capsaicin	1101
6	Omeprazole	19.8	Pantoprazole	883
7	Amoxicillin	18.6	Methylprednisolone	829
8	Ibuprofen	17.7	Aurothiomalate	774
9	Naproxen	16.5	Magnesium oxide	774
10	Folic acid	14.6	Etoricoxib	747

Discussion

We produced clinically representative AESI algorithms, validated with the DARWIN-EU phenotyping protocol[2]: enabling future use in vaccine AESI detection studies.

Cohort Counts, *Code Counts*, and *Index Events*, work well in triplicate, to ensure the 'flavour' is composed of appropriate codes, that are mapped for AESI context at the earliest stage – and were especially important for construction of our HF phenotype.
Orphan Codes then provides a 'safety-check' to identify missed codes.

Incidence-Prevalence and *Age-Sex-Distribution* provide a vital 'snapshot' of the constructed algorithm.
→ In the case of Anaphylaxis, this was very strongly validating.
→ **For RA**, this identified an **artifactual rise in the incidence of RA in 2013**, as circled in red on **Figure 1**.
→ This is likely attributable to the addition of RA to the UK Quality and Outcomes Framework in 2013[5] – meaning General Practitioner Doctors were newly paid to code RA.
→ This rise was not seen when "clinic" and "monitoring" related codes were removed in a new 'flavour'.

A key advantage of *Large-Scale Characterisation* in PhenotypeR is comparison to an age-sex matched population, **revealing small absolute, by high relative rises**.
→ Drug and Symptom codes indicate many patients already had RA long before indexing (**Table 2**).
→ Sulfasalazine was **19th** most-prescribed in the initial flavour, prescribed in the year prior in 10.7% overall, but this was **10247% higher:** showing this to be much **more notable than initially apparent**.

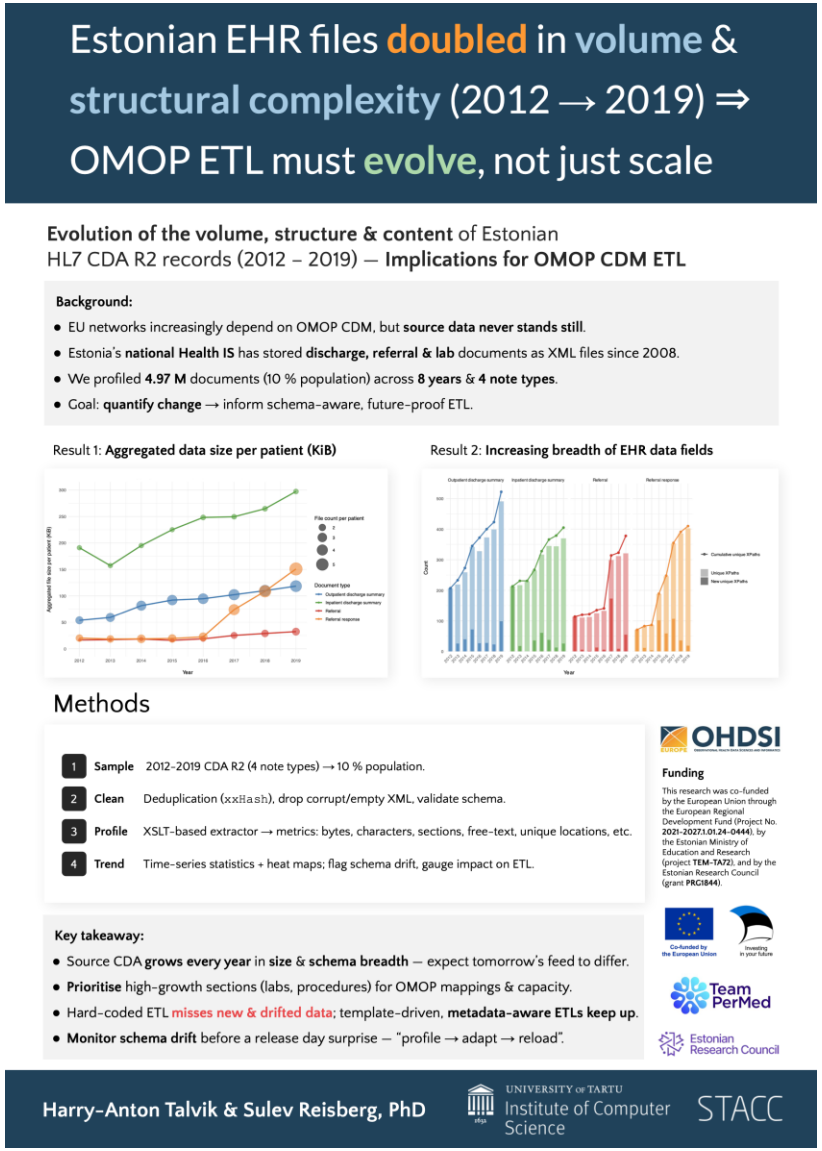
References
[1] DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe | 2024. https://catalogues.ema.europa.eu/node/4153/administrative-details
[2] Dernie F, Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN) EU®. Pharmacoepidemiology and drug safety. 2024
[3] Identify Relevant Clinical Codes and Evaluate Their Use [Internet]. GitHub.io. 2025. http://darwin-eu.github.io/CodelistGenerator/
[4] Assess Study Cohorts Using a Common Data Model [R package PhenotypeR version 0.1.4]. R-projectorg [Internet]. 2025; https://cran.r-project.org/web/packages/PhenotypeR/index.html
[5] NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE QUALITY AND OUTCOMES FRAMEWORK (QOF) INDICATOR DEVELOPMENT PROGRAMME Cost impact statement: Rheumatoid Arthritis QOF indicator area: Rheumatoid Arthritis. 2012

#OHDSISocialShowcase This Week

Wednesday

Evolution of the volume, structure & content of Estonian HL7 CDA R2 records (2012 – 2019) — Implications for OMOP CDM ETL

(Harry-Anton Talvik, Sulev Reisberg)



#OHDSISocialShowcase This Week

Thursday

Experiences with the development of an EORTC source vocabulary to represent QoL questionnaire data in OMOP

(Vlad Korsik, Maryia Rahozhkina, Sebastiaan van Sandijk, Peter Prinsen, Mieke Van Hemelrijck, Charlotte Moss)

Experiences with the development of an EORTC source vocabulary to represent QoL questionnaire data in OMOP

Vlad Korsik,MD[1] , Maryia Rahozhkina[1], Sebastiaan van Sandijk[1], Peter Prinsen[2], Mieke Van Hemelrijck[3], Charlotte Moss[3]
1: Odysseus (an EPAM company); 2, IKNL; 3: Transforming cancer OUtcomes through Research, King's College London

Introduction

The EORTC (European Organization for Research and Treatment of Cancer) Quality of Life group (QLG) develops and validates health-related quality of life (HRQoL) questionnaires for use in clinical trials, clinical practice, and academic research. Through collaborations with other EORTC research groups, and by engaging multidisciplinary experts, the QLG aims to understand how cancer and its treatments impact HRQoL across diverse populations and contexts. Despite their role in capturing patient-centered outcomes, HRQoL data is often underrepresented and remains largely absent from big data analyses. This data is typically poorly converted to OMOP as standard terminology due to its entity-attribute-value/post-coordination nature. Here we describe our experiences with developing an OMOP (source) vocabulary for HRQoL data using a prototype ontology model based on terminology content in the EORTC Item Library.

Methods

Human-in-the-loop approaches for formal logic development and individual-to-class mappings were utilized to minimize risks of incorrect terminological ETL. Analysis of the EORTC Quality of Life Questionnaire (QLQ) Item Library and real-world source (RWS) data evaluation was conducted to identify key concepts/classes and relationships associated with the various types of questionnaire modules in the EORTC Item Library. The resulting prototype ontology model was iteratively evaluated at collaborators' sessions and subsequently ratified. Upon ratification, a data (JSON formatted) scraping algorithm was introduced and terminology source storing was enabled at OMOP Vocabulary server side. Filtering rules were applied to exclude draft survey content. Concept classes and relationships that overlap with preexisting OMOP elements were mapped accordingly, to secure absence of redundancy of OMOP metadata elements. Concept codes were defined to permit intrinsic uniqueness first and were subsequently enriched with semantic context attributes to ease data manipulation during ETL, based on the RWS patterns that were identified.

Image 1. Schematic overview implementation-process

Results

Image 2. logic description of building a concept_code (left) and example of formal representation of entities within-EORTC QLQ vocabulary

The semantic corpus of OMOP, which adopted the EORTC QLQ vocabulary, comprises 4,094 concepts and 15,732 relationships. To enhance the existing scope of OMOP relationships with domain-specific metadata, 3 new relationships (along with their reciprocal counterparts) were added. All processed items belong to the Event Domains within the OMOP CDM schema.

Given that EORTC QLQ instruments are standardized, most concepts were mapped to the Measurement domain, facilitating the integration and use of relevant data elements and enabling data representation approaches (e.g., post-coordination) specific to the table. The only new type of entities created de novo were Answer concepts. Detailed visualizations of the results are presented in Images 2 and 3.

The terminological ETL code and related documentation have been published on the OHDSI Vocabulary 5.0 GitHub, and the actual terminological content is available for browsing and download at athena.ohdsi.org.

Image 3. OMOP CDM Domain and class distribution

Discussion

The development and integration of vocabularies, such as the EORTC QLQ vocabulary, into established healthcare data standards like OMOP (Observational Medical Outcomes Partnership) is both essential and highly sought after. This need is driven by the growing importance of patient-reported outcome measures (PROMs) in healthcare analytics.

Our work demonstrates the feasibility of adopting complex survey instruments as source terminologies for OHDSI use cases. Moving forward, our efforts will focus on implementing this vocabulary into real-world data conversion and analytics pipelines. This will align with ongoing OMOP expansions into survey and PROMs data, such as the value_as_concept_id initiative

We also believe that further semantic augmentation can offer significant benefits for ETL (Extract, Transform, Load) developers and researchers.

Potential enhancements in terminology include:

- Creating precoordinated pairs for OMOP-specific semantic mappings (e.g., event mappings).
- Establishing relationships with clinical conditions (e.g., SNOMED Neoplastic axis, HemOnc condition classes).
- Enriching synonyms to support multilingual data operations.
- Adapting non-formalized attributes and incorporating new attributes related to questionnaires.

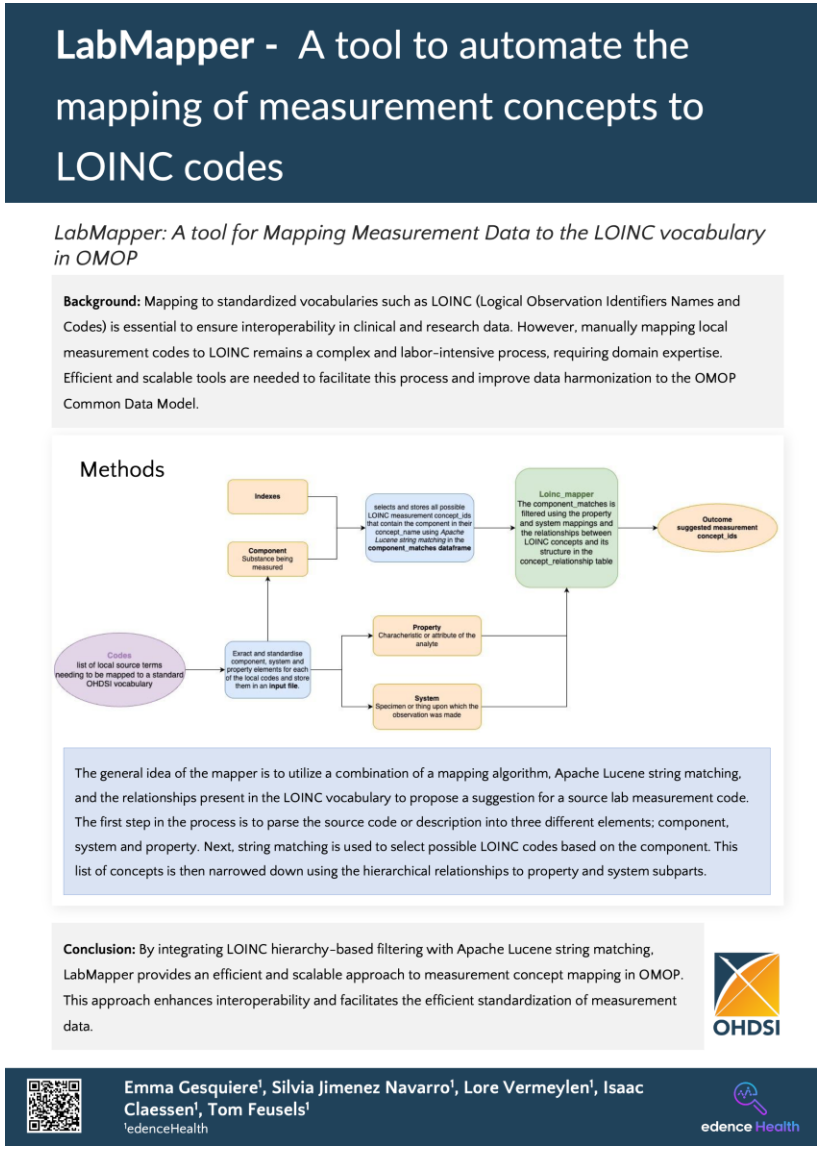
The proposed EORTC vocabulary is poised to strengthen the use of EORTC questionnaire data in network research, particularly in oncology. Future steps will focus on improving the adoption and implementation of this vocabulary, ensuring its utility in validation studies and across relevant data and research networks.

#OHDSISocialShowcase This Week

Friday

LabMapper: A tool for Mapping Measurement Data to the LOINC vocabulary in OMOP

(Emma Gesquiere, Silvia Jimenez Navarro, Lore Vermeylen, Isaac Claessen, Tom Feusels)



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 16: OHDSI/OMOP Research Spotlight

Jessie Tong

Assistant Professor, Johns Hopkins University

Unlocking efficiency in real-world collaborative studies: a multi-site international study with one-shot lossless GLMM algorithm • *NPJ Digital Medicine*

Kim López Güell

Dphil Candidate, University of Oxford

Clusters of post-acute COVID-19 symptoms: a latent class analysis across 9 databases and 7 countries • *Journal of Clinical Epidemiology*

Jen Wooyeon Park

PhD Student, Johns Hopkins University

Breaking data silos: incorporating the DICOM imaging standard into the OMOP CDM to enable multimodal research • *JAMIA*

Abigail Newbury

PhD Student, Columbia University

Multi-domain rule-based phenotyping algorithms enable improved GWAS signal • *NPJ Digital Medicine*

Benjamin Martin

Postdoctoral Fellow, Johns Hopkins University

Identification of Adult Dermatomyositis Patients Using Real-World Data Sources • *Arthritis Care and Research*

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