Recent OHDSI/OMOP Publications

OHDSI Community Call
Dec. 3, 2024 • 11 am ET

Upcoming Community Calls

Date	Topic
Dec. 3	Recent OHDSI Publications
Dec. 10	How Did We Do In 2024?
Dec. 17	Holiday-Themed Final Call of 2024
Dec. 24	No Call
Dec. 31	No Call
Jan. 7	What Can OHDSI Go In 2025?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Seok Woo Hong and Jeong-Hyun Kang** on the publication of **Antinuclear Positivity and Malignant Transformation Potential of Oral Potentially Malignant Disorder** in *Oral Diseases*.

ORAL DISEASES
Leading in Oral, Maxillofacial, Head & Neck Medicine

ORIGINAL ARTICLE

Antinuclear Positivity and Malignant Transformation Potential of Oral Potentially Malignant Disorder

Seok Woo Hong, Jeong-Hyun Kang

First published: 25 November 2024 | https://doi.org/10.1111/odi.15208

Funding: This research was supported by funds from the Korean Academy of Laboratory & Diagnostic Dentistry, intramural research funds from Ajou University Medical Center, and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2022-00165960).

Read the full text > PDF TOOLS SHARE

ABSTRACT

Objective

This multicenter cohort study aimed to investigate the impact of antinuclear antibody (ANA) positivity on the malignant transformation potential of oral potentially malignant disorders (OPMDs) using the Observational Health Sciences and Informatics network tools for the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Materials and Methods

Retrospective cohort data from five medical centers were analyzed. Cox regression and 1:4 propensity score matching followed by aggregated meta-analysis were used to evaluate association between ANA positivity and the risk of oral squamous cell carcinoma (OSCC). The index date was defined as the first day of diagnosis of OPMD. The target and comparator cohorts comprised patients who tested positive and negative for ANA within 180 days before or after the index date, respectively.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Wednesday	8 am	Psychiatry
Thursday	11 am	Industry
Thursday	1 pm	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	7 pm	Dentistry
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Clinical Trials
Friday	11:30 am	Steering Group
Tuesday	9 am	OMOP CDM Oncology Genomic Subgroup

2024 APAC Symposium

Dec. 4-8 • Marina Bay Sands & National University of Singapore (NUS)

Dec. 4: Tutorial at NUS

Dec. 5-6: Main Conference at Marina Bay Sands

Dec. 7-8: Datathon at NUS

ohdsi.org/APAC2024

2024 APAC Symposium

Dec. 4-8 • Marina Bay Sands & National University of Singapore (NUS)

Day 1 (December 4) – Tutorial at NUS

8:30 – 9:00 • Registration/Tea
9:00 – 9:20 • OHDSI/OMOP Introduction (Mui Van Zandt)
9:20 – 10:00 • OMOP CDM and Vocabulary (Mukkesh Kumar, Cindy Ho)
10:00 – 10:30 • OMOP Conversion Process (Erica Voss)
10:30 – 10:40 • Break
10:40 – 12:00 • ETL Exercises (Gyeol Song)
12:00 – 13:30 • Lunch
13:30 – 14:10 • OHDSI Analyses: From Questions to TCO (Martijn Schuemie)
14:10 – 14:50 • OHDSI Analyses: Building Cohorts (Patrick Ryan)
14:50 – 15:10 • Tea Break
15:10 – 15:50 • OHDSI Analyses: Building Cohorts Hands-On (Patrick Ryan)

Dec. 5

13:30 – 13:40 • Day 1 Opening (Mengling 'Mornin' Feng)
13:40 – 14:10 • OHDSI for Real-World Evidence (RWE) (Patrick Ryan)
14:10 – 14:25 • Charting our APAC Journey: Lessons from the Past, Visions for the Future (Mui Van Zandt)
14:25 – 15:15 • OHDSI APAC Regional Chapter Updates (Rae Woong Park, Hua Xu, Nicole Pratt, Tatsuo Hiramatsu, Jason Hsu, Mengling 'Mornin' Feng, Parthiban Sulur)
15:15 – 15:45 • Tea Break
15:45 – 16:15 • 2024 APAC ETL Project (Mui Van Zandt, Gyeol Song, Steven Yong, Satish Kumar Anbazhagan, Kosuke Tanaka, Santan Maddi)
16:15 – 16:45 • 2024 APAC ETL Project: Panel Discussion (Gyeol Song, Steven Yong, Satish Kumar Anbazhagan, Kosuke Tanaka, Santan Maddi)
16:45 – 17:05 • OHDSI Evidence Network (Erica Voss)
17:05 – 17:25 • Large Language Model and OHDSI: Part 1 (Hua Xu)
17:25 – 17:35 • Large Language Model and OHDSI: Part 2 (Hyeonsik Kim)
17:35 – 17:55 • HL7 Singapore and OHDSI Singapore Collaboration (Adam Chee, Mengling 'Mornin' Feng)
17:55 – 18:00 • Day 1 Closing

Dec. 6

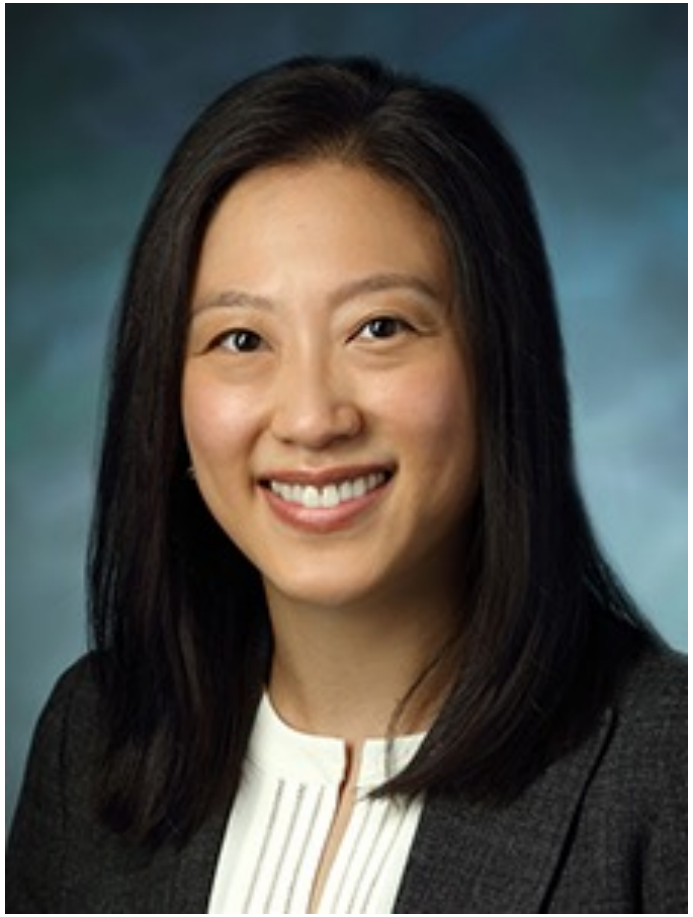
9:00 – 9:10 • Overview of the International and Singapore Standards Ecosystem (Aik Lam Khor)
9:10 – 9:20 • TRUST: Enabling Safe Data Exchange and Our OMOP Journey (Mingshi Koh)
9:20 – 9:30 • OMOP Common Data Model: Journey Towards Singapore's National Data Standardization for Real-World Evidence Generation (Mukkesh Kumar)
9:30 – 10:00 • Panel Discussion (Aik Lam Khor, Mingshi Koh, Mukkesh Kumar)
10:00 – 10:20 • Tea Break
10:20 – 10:30 • Use of OHDSI to Evaluate Safety Signals (Mengling 'Mornin' Feng)
10:30 – 10:50 • LEGEND-T2DM Study Introduction (Marc Suchard)
10:50 – 11:00 • 2024 APAC Study Introduction (Sreemanee Dorajoo)
11:00 – 12:00 • 2024 APAC Study: Journey from Data to Evidence (Evelyn Goh, Nicole Pratt)
12:00 – 13:30 • Lunch and Posters/Exhibitors
13:30 – 14:30 • 2024 APAC Study: Panel Discussion (Sreemanee Dorajoo, Yuhe Ke, Keiko Asao, Seng Chan You)
14:30 – 15:30 • Lightning Talks
15:30 – 15:40 • Day 2 Closing (Mengling 'Mornin' Feng)

ohdsi.org/APAC2024

Collaborator Spotlight: Cindy Cai

Dr. Cindy Cai is the Jonathan and Marcia Javitt Rising Professor of Ophthalmology at Johns Hopkins University and a retina specialist seeing patients at the Wilmer Eye Institute's locations in the Baltimore, Maryland area. Her primary focuses are in medical and surgical retina treatments, including: diabetic retinopathy, diabetic macular edema, and age-related macular degeneration.

A co-lead of the Eyecare and Vision Research Workgroup, Cindy is currently leading another OHDSI network study focused on Semaglutide and Nonarteritic Anterior Ischemic Optic Neuropathy. The 2024 Titan Award for Clinical Applications honoree, she discusses her career journey, her experience running her first community network study, opportunities in vision research using real-world data, and plenty more in the latest collaborator spotlight.



ohdsi.org/spotlight-Cindy-Cai

December Newsletter

The Journey Newsletter (December 2024)

Real-world evidence is at the center of OHDSI research, and it was a key topic in the community throughout November. One community call was focused on generating it through the OHDSI Evidence Network, while another looked at potential ways to improve the dissemination process. Evidence will be shared in

November Publications

Tekumalla R, Banda JM. Towards automated phenotype definition extraction using large language models. Genomics Inform. 2024 Oct 31;22(1):21. doi: 10.1186/s44342-024-00023-2. PMID: 39482749; PMCID: PMC11529293.

Ferrandiz-Espadin R, Rabasa G, Gasman S, McGinley B, Stovall R, Jafarzadeh SR, Liew JW, Dubreuil M. Disparities in time to diagnosis of Radiographic Axial Spondyloarthritis. J Rheumatol. 2024 Nov 4:jrheum.2024-0574. doi: 10.3899/jrheum.2024-0574. Epub ahead of print. PMID: 39486857.

Tong J, Li L, Reps JM, Lorman V, Jing N, Edmondson M, Lou X, Jhaveri R, Kelleher KJ, Pajor NM, Forrest CB, Bian J, Chu H, Chen Y. Advancing Interpretable Regression Analysis for Binary Data: A Novel Distributed Algorithm Approach. Stat Med. 2024 Nov 3. doi: 10.1002/sim.10250. Epub ahead of print. PMID: 39489875.

Pyun JM, Lee I, Lee K, Kim MH, Park C, Yang HJ. Effect of Choline Alfoscerate on the Progression From Mild Cognitive Impairment to Dementia: Distributed Network Analysis of a Multicenter Korean Database Using a Common Data Model. Dement Neurocogn Disord. 2024 Oct;23(4):202-211. doi: 10.12779/dnd.2024.23.4.202. Epub 2024 Oct 7. PMID: 39512703; PMCID: PMC11538851.

Community Updates

Where Have We Been?

• Five posters/demos were honored as Best Community Contribution Award winners following the 2024 Global Symposium. You can see the posters here, and presentations about each are available at the bottom of this newsletter.
  Observational Data Standards and Management: Gap Analysis of Static Automated Perimetry Concept Representation in OMOP CDM (**Shahin Hallaj**)
  Methodological Research: Towards automated phenotype definition extraction using large language models (**Ramya Tekumalla**)
  Open-Source Analytics Development: Bridging the Language Gap: Generative Models for Efficient Medical Concept Discovery (**Alvaro Alvarez**)
  Clinical Applications: Health Trends Across Communities in Minnesota: a Statewide Dashboard Leveraging the OMOP CDM to Monitor the Prevalence of Health Conditions (**Samuel Patnoe**)
  Community: Improving Team Science Through "Thons" Reflections on the April Olympians Community Event (**Clair Blacketer**)
• **Jennifer Clark Nelson**, Director of Biostatistics & Senior Investigator at Kaiser Permanente Washington Health Research Institute, presented a talk on Statistical methods for improving post-licensure vaccine safety surveillance during a Nov. 20 edition of the CBER BEST Seminar Series.
• The latest edition of "Our Journey: Where the OHDSI Community Has Been, and Where We Are Going" was distributed at the Global Symposium, and the

2024 Asia-Pacific Symposium Focuses On "When OHDSI Meets With AI"

The 2024 OHDSI Asia-Pacific (APAC) Symposium will be held December 4-8 in Singapore. Regional co-chairs Mengling 'Mornin' Feng and Ngial Kee Yuan will lead this event, which has a theme of "When OHDSI Meets with AI."

The five-day event begins with a tutorial on Dec. 4 at the National University of Singapore (NUS), and it will include sessions on OMOP, the ETL process, and OHDSI analytics tools. The main conference will be held Dec. 5-6 at the Marina Bay Sands; the full agenda for both days is available on the event homepage below. There will be a two-day datathon on Dec. 7-8 at NUS.

All details about the event, including a link to register, is available on the APAC 2024 homepage.

APAC 2024 Information & Registration

Podcast: Evidence Dissemination, Semaglutide Network Study, APAC Symposium

In the latest On The Journey videocast, Patrick Ryan and Craig Sachson discuss a recent community call focused on evidence dissemination within the community, including the new Evidence Translation workgroup. They also discuss the recent Evidence Network study on semaglutide, and they look ahead to the 2024 APAC

November Presentations

Nov. 5: Meet the Titans
Full Presentation (Janetzki, Cai, Zhuk, Zhang, Adulyannukosol, Blacketer, Camprubi, Katzman, Lavallee)

Nov. 12: Next Steps in Evidence Dissemination
Full Presentation (Ryan, Schuemie, Pratt)

Nov. 19 Evidence Network in Action: The Semiglutide Study
Full Presentation (Cai, Zhang, Nagy, Sena, Westlund, Martin)

Nov. 26: OHDSI 2024 Best Community Contribution Honorees
Gap Analysis of Static Automated Perimetry Concept Representation in OMOP CDM (Hallaj)
Towards automated phenotype definition extraction using large language models (Tekumalla)
Bridging the Language Gap: Generative Models for Efficient Medical Concept Discovery (Alvarez)
Health Trends Across Communities in Minnesota: a Statewide Dashboard Leveraging the OMOP CDM to Monitor the Prevalence of Health Conditions (Patnoe)
Improving Team Science Through "Thons" Reflections on the April Olympians Community Event (Blacketer)

December Newsletter

2024 Global Symposium

2024 OHDSI Global Symposium

Oct. 22-24 · New Brunswick, N.J. · Hyatt Regency Hotel

The 10th annual OHDSI Global Symposium brought together more than 470 global collaborators for three days of sharing research, building new connections and pushing forward our mission of improving health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

This page will host all materials from OHDSI2024, including video presentations (when available) from the main conference and tutorials, slide decks, posters, demos and more.

State of the Community

Where Have We Gone and Where Are We Going?
(George Hripcsak, Columbia University)

Expand OHDSI Initiative for Eye Care and Ocular Imaging Challenge
(Amberlynn Reed, Natiional Eye Institute)

Titan Awards
(George Hripcsak, Columbia University & Marc Suchard, UCLA)

State of the Community Slides

Plenary: Value Proposition for Participating in OHDSI Network Studies like LEGEND-T2DM

Introduction to OHDSI Evidence Network / Marketplace
(Moderator: Clair Blacketer, Johnson & Johnson)

Reflections from US Department of Veterans Affairs
(Scott Duvall, VA)

Reflections from SIDIAP (Spain)
(Talita Duarte-Salles, IDIAP)

Reflections from a Global Commercial Data Provider
(Atif Adam, IQVIA)

Plenary: Value Proposition for Participating in OHDSI Network Studies like LEGEND-T2DM Slides

Plenary Q&A: Lessons Learned on LEGEND-T2DM Journey

Moderator: Fan Bu, University of Michigan

Panelists: LEGEND-T2DM co-authors

Plenary Q&A: Lessons Learned on LEGEND-T2DM Journey Slides

Plenary Panel: JACC-OHDSI Partnership

Moderators:
Nicole Pratt, University of South Australia
Marc Suchard, UCLA

Panelists:
Harlan Krumholz, Yale University
Seng Chan You, Yonsei University
Yuan Lu, Yale University

Plenary Panel: JACC-OHDSI Partnership Slides

ohdsi.org/OHDSI2024

2024 Global Collaborator Showcase

Observational Data Standards & Management

1 – Application of OMOP Common Data Model to Disease Registry Data (Vojtech Huser, Maria Rogozhkina, Vlad Korsik, Teresa A. Simon, Peter Moorthamer, Dan Kiselev, Teresa A. Simon, Anastasia Vakhmistrova, Eugene Paulenkovich, Alexander Davydov, Michel Van Speybroeck)
2 – Best Practices for Developing Disease-Specific Federated Networks: Insights from a Systemic Lupus Erythematosus Study (Clair Blacketer, Frank DeFalco, Gowtham A Rao, Anna Sheahan, Michel Van Speybroeck, Martine Lewi, Federico Zazzetti)
3 – Standardizing Rare Disease Patient Registry data to the OMOP-CDM (Parag Shiralkar, Radhika Lakireddy, Sushma Ghanta, Sanket Kalyankar)
4 – PHederation – the federated network of Pulmonary Hypertension registries (Eva-Maria Didden, Valerie van Baalen, Michel van Speybroeck, Monika Brand)
5 – Lessons from mapping cancer information from European hospitals to ICD-O-3 conditions in OMOP (Lars Halvorsen, Olivier Bouissou, Elisabeth Ross, Stelios Theophanous, Joëlle Thonnard, Piers Mahon)
6 – SMEs optimization with high precision data ingestion of CAPriCORN CDM onto OMOP at AllianceChicago (Andrew Hamilton, Amro Hassan, Davera Gabriel, Guy Tsafnat)
7 – Process of Conversion of Ukrainian Medical Data to OMOP CDM Format (Bohdan Khilchevskyi, Denys Kaduk, Maksym Trofymenko, Polina Talapova, Tetiana Nesmiian, Max Ved, Inna Ageeva, Pavlova Olga, Holovko Tetiana, Shevchenko Natalia)
8 – An evaluation of the transformation of large German EHR database to OMOP CDM (Andreas Ochs, Milou Brand, Jack Brewster, Methosdios Typou, Meda Sandu, Joe Maskell, Meghan Pettine, Atif Adam, George Kafatos)
9 – Adopting the OMOP Oncology CDM at the Helsinki University Hospital (Valtteri Nieminen, Alexey Ryzhenkov, Johanna Sanoja, Salma Rachidi, Juho Lähteenmaa, Joonas Laitinen, Samu Eränen, Tomi Mäkelä, Eric Fey, Kimmo Porkka)
10 – Going global, redeeming the local: an innovative approach to implement the OMOP CDM in two countries of the Global South (Valentina Martufi, Emma Kalk, Enny S. Cruz, Juliana Araújo Prata de Faria, Adalton do Anjos Fonseca, Maurício L. Barreto, Maria Yury Travassos Ichihara, Jessica Gammon, Nicki Tiffin, Chris Fourie, Danilo Luis Cerqueira Dias, Denise Moraes Pimenta, Tsaone Tamuhla, Andrew Boulle, Themba Mutemaringa, Juan-Paul Hynek, Muzzammil Ismail, Julio Barbour Oliveira, Ricardo Felix Monteiro Neto, Júlia Pescarini, Fernanda Revoredo de Sousa, Marianne Costa e Silva Lage, Adam Loff, Melvin Moodley, Elzo Pereira Pinto Junior)
11 – Transforming Clinical Trial Data to the OMOP CDM (Cynthia Sung, Mike Hamidi, Zhen Lin, Tom Walpole, Rebecca Baker, Melissa Cook, Shital Desai, Priya Gopal, Dan Hartley, Vojtech Huser, Priya Meghrajani, Tra Nguyen, Paul Orona12, Katy Sadowski, Sebastiaan van Sandijk, Philip Solovyev, Ramona Walls, Kenneth J. Wilkins, Qi Yang)
12 – Streamlining Research Data Standardization: AI-READI Survey Instrument Data Elements and MoCA Measurement Data Elements are curated and mapped utilizing a Standardized Value Set Mapping Table for transformation into the OMOP Common Data Model (Stephanie S. Hong, James Cavallon, Yvette Chen, Monique Bangudi, Jessica Mitchell, Dawn Matthies, Steven Chamberlin, Aaron Cohen, Julie Owens, Abigail Lucero, Sally Baxter, Christopher G Chute, Cecilia S. Lee, Aaron Lee, AI-READI consortium)
13 – Institutionalizing data interoperability and the application of common data models in a health data and research center: CIDACS' experience in Brazil (Valentina Martufi, Juliana Araújo Prata de Faria, Danilo Luis Cerqueira Dias, Elzo Pereira Pinto Junior, Roberto Carreiro, Pablo Ivan Ramos, Maurício L. Barreto)
14 – OMOP GIS Vocabulary Package for Observational Studies in Health Care and Public Health (Maksym Trofymenko, Polina Talapova, Andrew Williams)
15 – Enhancing Infectious Disease Data Integration and management through OMOP-CDM in South Korea (Min Ho An, Seok Kim, ByungJin Choi,Sooyoung Yoo,Rae Woong Park,Ji Seon Oh)
16 – FHIR to OMOP Cookbook – Mapping mCODE FHIR Resources for Observational Research (Qi Yang, Guy Livne, Sebastian van Sandijk, May Terry)
17 – Towards Reproducible Imaging Research: Implementation of DICOM to OMOP CDM (Woo Yeon Park, Ben Martin, Gabriel Salvador, Blake Dewey, Teri Sippel Schmidt, Paul Nagy)
18 – Leveraging UDI for Advanced Medical Device Tracking in OMOP-CDM (Seojeong Shin, Yiju Park, Sujeong Eom, Kyulee Jeon, Seng Chan You)
19 – Inclusion of intraocular pressure data into the University of California Health Data Warehouse (William Halfpenny*, Shahin Halla*, Ayan Patel, Catherine Q. Sun, Kerry Goetz, Michelle Hribar, Sally L. Baxter, on behalf of the OMOP Eye Care & Vision Research Workgroup)
20 – A Collaborative Analytic Enclave for the Metabolic Dysregulation and Obesity Cancer Risk Program (MeDOC) Consortium: Extensions of the OMOP Common Data Model for Translational Research (Madhan Subramanian, Nisha Grover, Maddie Wheeler, Marinella Temprosa)
21 – Expanding the OMOP Common Data Model to support Extracorporeal Life Support research (Clemens Rieder, Oleg Zhuk, Ahmed Said, Peta M.A. Alexander, Dominik J. Hoechter)
22 – ETIng from your OMOP CDM to your OMOP CDM? An efficient solution to vocabulary migration (Clair Blacketer, Anton Ivanov, Evanette Burrows, Dmitry Dymshyts, Frank DeFalco)
23 – Evaluating the impact of different vocabulary versions on cohort definitions and CDM (Dmitry Dymshyts, Frank DeFalco, Anna Ostropolets, Gowtham Rao, Azza Shoaibi, Clair Blacketer)

2024 Global Symposium

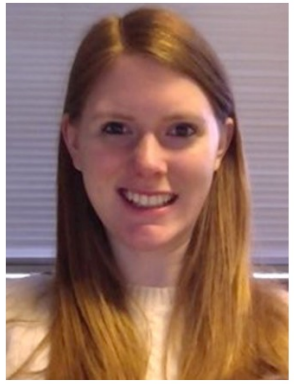
An Introduction to the Journey from Data to Evidence Using OHDSI

The journey from data to evidence can be challenging alone but is greatly enabled through community collaboration. In this half-day tutorial, we will introduce newcomers to OHDSI. Specifically, about the tools, practices, and open-science approach to evidence generation that the OHDSI community has developed and evolved over the past decade.

Faculty will highlight the ways community individuals can participate as well as receive value from the community's outputs. The course will include topics such as open community data standards – including the OMOP Common Data Model and OHDSI Standardized Vocabularies, open-source analytic tools – including HADES and ATLAS, and the conduct of open network studies for methodological research & clinical applications. **Location: Regency Ballroom DE, Main Level**

Faculty

Daniel Prieto-Alhambra
University of Oxford; Erasmus M.C.



Jenna Reps
Janssen Research & Development

Mui Van Zandt
IQVIA

Erica Voss
Janssen Research & Development

Linying Zhang
Washington University in St. Louis

ohdsi.org/OHDSI2024

2024 Global Symposium

Developing and Evaluating Your Extract, Transform, Load (ETL) Process to the OMOP Common Data Model

The OMOP Common Data Model has become one of the most widely used international health data standards. Standardizing data to the OMOP CDM requires development of an extract, transform, load (ETL) procedure that converts source data into the CDM structure while observing the appropriate conventions and adhering to the OHDSI standardized vocabularies. The OHDSI community maintains and provides resources for the OMOP CDM standard, Standardized Vocabularies, and THEMIS ETL conventions, and has developed a series of open-source analytic tools to support both ETL development and evaluation (including WhiteRabbit, CDMInspection, and DataQualityDashboard).

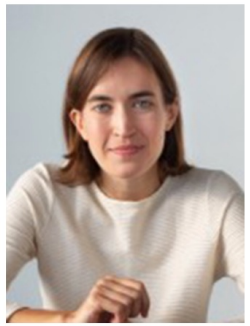
In this tutorial, students will learn about the tools and practices developed by the OHDSI community to support the journey to establish and maintain an ETL to standardize your data to OMOP CDM and enable standardized evidence generation across a data network. **Location: Regency Ballroom E, Main Level**

Faculty

Clair Blacketer
Janssen Research & Development

Evannette Burrows
Janssen Research & Development

Melanie Philofsky
Odysseus Data Services, Inc.



Katy Sadowski
Boehringer Ingelheim

ohdsi.org/OHDSI2024

2024 Global Symposium

Using the OHDSI Standardized Vocabularies for Research

The OHDSI Standardized Vocabularies serves as a foundation to data standardization process within the OMOP CDM. It also can be tremendously useful tool for enabling the appropriate design of analyses that can be executed across a network of databases. A core component within essentially all analysis is the specification of phenotypes and associated code lists to represent exposures, outcomes, and other features.

In this tutorial, students will learn how to take advantage of the OHDSI standardized vocabularies as an analytic tool to support your research, including searching for relevant clinical concepts, navigating concept relationships, creating Conceptsets and understanding source codes that map within these expressions. Students will also learn where the OHDSI standardized vocabularies is used throughout OHDSI's standardized analytic tools. **Location: Regency Ballroom F, Main Level**

Faculty

Anna Ostropolets
Janssen Research & Development

Vlad Korsik
Odysseus Data Services, Inc.

Azza Shoaibi
Janssen Research & Development

Polina Talapova
SciForce

Oleg Zhuk
Odysseus Data Services, Inc.

ohdsi.org/OHDSI2024

2024 Global Symposium

So, You Think You Want To Run an OHDSI Network Study?

Reliable real-world evidence generation requires appropriate analyses applied to data sources fit-for-purpose for the research question of interest. The OHDSI community has developed open-source standardized analytics tools that can be executed across a network of OMOP CDM databases and processes to facilitate collaborations between researchers throughout the evidence generation process from design through implementation and dissemination.

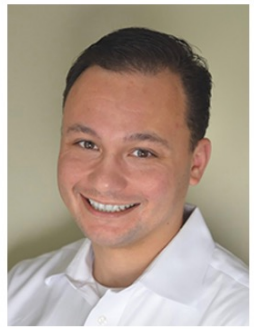
In this tutorial, students will learn about the steps along the journey to turn your research question into reliable evidence and how to lead an OHDSI network study. **Location: Regency Ballroom C, Main Level**

Faculty

Yong Chen
University of Pennsylvania

Ben Martin
Johns Hopkins University

Nicole Pratt
University of South Australia



Anthony Sena
Janssen Research & Development

Andrew Williams
Tufts University

Seng Chan You
Yonsei University Health System

ohdsi.org/OHDSI2024

2024 Global Symposium

Conducting 'Off-The-Shelf' Characterization Studies Using DARWIN EU® Tools and the OMOP CDM

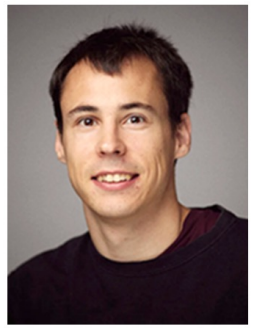
The European Medicines Agency (EMA) and the European Medicines Regulatory Network established the Data Analysis and Real-World Interrogation Network (DARWIN EU®) coordination center to provide timely and reliable evidence on the use, safety and effectiveness of medicines for human use, including vaccines, from real world healthcare databases across the European Union (EU). The DARWIN EU team has established a data network standardized to the OMOP CDM and has developed a series of open-source analytics tools that run atop the OMOP CDM to conduct characterization studies for disease natural history, drug utilization, and treatment patterns.

In this tutorial, students will learn from leaders in the DARWIN EU team about how to execute characterization analyses against their OMOP CDM instance using DARWIN EU packages, including how to define inputs to the standardized analytics and how to interpret standardized results. Students will also learn how DARWIN EU tools relate to and connect with OHDSI's broader open-source analytics ecosystem. **Location: Regency Ballroom D, Main Level**

Faculty

Edward Burn
University of Oxford

Daniel Prieto-Alhambra
University of Oxford;
Erasmus M.C.



Martí Català Sabaté
University of Oxford

Maarten van Kessel
Erasmus M.C.

ohdsi.org/OHDSI2024

CDM Survey Subgroup Landscape Assessment

The CDM Survey Subgroup invites colleagues who have or are going to design, develop, and/or implement research surveys and use them with the OMOP CDM to share information about those efforts by completing this survey. Your completion of this 10-15 minute survey will provide information to the CDM workgroup about OMOP utilization among survey research teams. The CDM Survey subgroup is a collaborative effort, led by a team at the National Cancer Institute, to develop standardized approaches and best practices for helping research teams better integrate survey data elements into the OMOP common data model.

The survey deadline is Dec. 31, 2024.

LANDSCAPE ASSESSMENT

- **Activities**
  - Invite representatives from cohorts with experience using the CDM for survey data to share their knowledge and challenges.
  - Conduct a community survey to gather information on experiences and needs related to survey data in the CDM.
  - Review the most used Common Data Elements (CDMs) as a foundation for developing standards, tools, and best practices.
- **Key Result**
  - A comprehensive report summarizing survey CDM mapping resources, challenges, and identified development priorities (vocabulary, standards, tools, best practices) to be shared with the OHDSI community.

WHO SHOULD PARTICIPATE

- You have survey data and you've mapped it to the OMOP CDM
- You have survey data and you would like to map it to the OMOP CDM
- You are in the process of developing a survey(s) and plan to map to the OMOP CDM
- Multiple perspectives from the same team
- Multiple surveys from the same person

#OHDSISocialShowcase This Week

Monday

OMOP GIS Vocabulary Package for Observational Studies in Health Care and Public Health

(Maksym Trofymenko, Polina Talapova, Andrew Williams)

OMOP GIS Vocabulary Package for Observational Studies in Health Care and Public Health

PRESENTER:
Polina Talapova

polina.talapova
@sciforce.tech

INTRO

Why should you care? Health outcomes are influenced by a complex interplay of geography, environment, and social determinants. However, traditional health records don't capture this. Our study matters because it bridges that gap: giving researchers the tools they need to explore how where you live, the air you breathe, and your social context affect your health.

METHODS

We built the OMOP GIS Vocabulary Package by:

1. Collected Data: Extracting geographic, environmental, and social data (SDOH) from reliable sources (T3DB, AHRQ, and others).

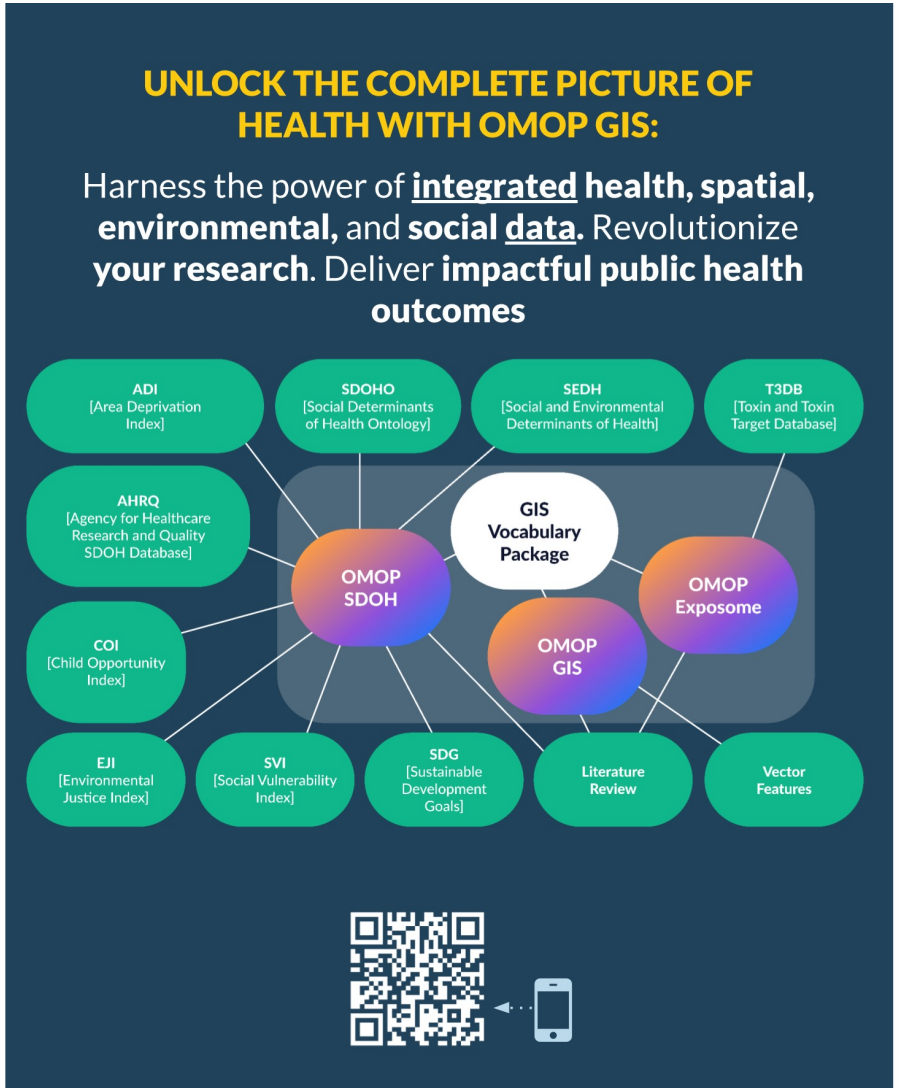
2. Tested Integration: We linked this data with health records in OMOP-compliant databases, using standardized vocabularies for spatial and environmental data, and extended existing OMOP vocabularies to capture new elements like social vulnerability and toxin exposure.

RESULTS

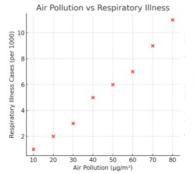
We successfully developed and validated a vocabulary package that integrates **9,918 terms** (Table 1) with **9,605 associations to OMOP**, and **89,311 internal relationships.**

Table 1. Term Distribution Across Vocabularies within the GIS Vocabulary Package

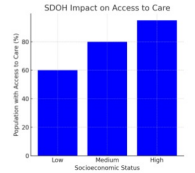
vocabulary_id	number of terms
OMOP SDOH	2,892
OMOP Exposome	6,867
OMOP GIS	159
Total	9,918



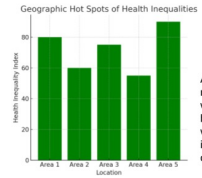
The Vocabulary Package enables a wide range of analyses, some options are demonstrated in the examples below.



A scatter plot shows the relationship between air pollution levels (in µg/m³) and respiratory illness cases (per 1,000 population).



A bar chart showing how access to care increases with socioeconomic status, highlighting disparities in healthcare access.



A bar chart representing areas with significant health inequalities, with higher values indicating greater disparities

CONCLUSION

The OMOP GIS Vocabulary Package advances health research by enabling the integration of spatial, environmental, and social data with clinical records. Its standardized terminologies enhance the exploration of the combined influences on health outcomes. Future developments will focus on refining the package and testing it with diverse use cases to ensure adaptability.

Maksym Trofymenko,
Polina Talapova,
Andrew Williams

sciforce

Tufts | CTSI Tufts Clinical and Translational Science Institute

OHDSI

#OHDSISocialShowcase This Week

Tuesday

Medical Device Standard Terminology Overview, Comparison and Analysis

(Asiyah Yu Lin, Michael Matheny, Andrew Williams, Seng Chan You)

Medical Device Standard Terminology Overview, Comparison, and Analysis

PRESENTER: **Asiyah Yu Lin**

INTRO:

- **A bottleneck to generating Real World Evidence (RWE) for medical device regulatory decision making is the lack of associated meta-data with the devices in existing terminologies** that limit the ability to evaluate common materials, manufacturing techniques or changes in component versioning across time, and other associated elements. It is **challenging to crosswalk between terminologies**, as the coverage of the interoperability between them is variable and heterogenous.

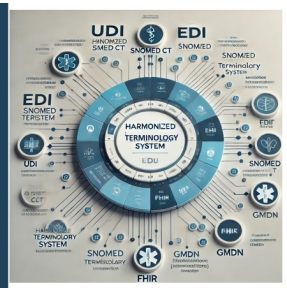
METHODS

WG conducted various comparisons: UDI vs. EDI vs. SNOMED CT, UDI vs. OMOP, UDI vs. FHIR resources, or SNOMED CT vs. GMDN. The workgroup members also explored the OHDSI Vocabulary's Medical Device component with a focus on RWE use.

RESULTS

- Critical gaps in meta-data characteristics of medical devices needed for categories of use cases such as analysis of component composition, version changes over time, among others. This degrades the capacity to automatically execute studies by requiring manual data collection on an ad-hoc basis. .

A harmonized and interoperable medical device terminologies, serving as the single source of truth, is needed for medical device RWE studies.



- UDI is more granular than SNOMED CT and supports more granular medical device RWE study designs.
- UDI and FHIR's medical device resources have similar meta-data to describe medical devices, however, these data are not included in the OMOP meta-data tables.
- Limitations remain in transforming observational data for medical devices into the OMOP CDM.
- Although able to use UDI in organizational supply chain data, the inventory systems will need to be connected to patient data from the EHR and available to support utilization.

Asiyah Yu Lin, Michael E. Matheny, Andrew Williams, Seng Chan You

OHDSI

#OHDSISocialShowcase This Week

Wednesday

An Explorative Study about the Latent Space of Clinical Foundation Models Based on a Common Data Model Database

(Min-Gyu Kim, Jin Yang Kim, Dong Yun Lee, Rae Woong Park, Joon-Kyung Seong)

An Explorative Study about the Latent Space of Clinical Foundation Models Based on a Common Data Model Database

Min-Gyu Kim[1, 2], Dong Yun Lee[1, 2], Jin Yang Kim[3], Rae Woong Park[1, 2], Joon-Kyung Seong[3, 4]
[1]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea
[2]Department of Medical Sciences, Graduate School of Ajou University, Suwon, Republic of Korea
[3]Department of Artificial Intelligence, Korea University, Seoul, South Korea
[4]School of Biomedical Engineering, Korea University, Seoul, South Korea

Background

Recently, there have been researches about clinical foundation models (FMs), which have shown advantages over traditional prediction model. While metrics like F1 score can explain the performance of a model objectively, they are usually inadequate for understanding the internal structure of the model. Also, methods to train such models are still limited to analogies from the language domain.
There are many methods available that enable model understanding, such as visualizing self-attention of each layer or dimension reduction in the latent space. In this study, we aim to understand how we should train clinical foundation models by first training a model using our own data based on OMOP-CDM and visualizing the latent space of the trained model.

Methods

We trained a transformer model based on the bidirectional transformer (BERT) architecture, using data from Ajou university hospital standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Patient records were first translated into a time series format. Additional information such as patient age and gender were prepended to the input series as separate tokens. To provide a better understanding about the domains defined by OMOP-CDM, each token was added to the embedding about its domain, i.e. condition, drug, measurement.
The model was trained using masked language modeling. 15% of the tokens were randomly masked and the model predicted the original tokens. 1% of the total training data was randomly selected, and the CLS tokens of the sample were calculated. The tokens were then reduced to seven dimensions using Uniform Manifold Approximation (UMAP) and clustered with Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). The result was visualized using t-distributed stochastic neighbor embedding (t-SNE) by reducing to a 2-dimensional plane. The resulting visualization was inspected, and cluster formation was manually evaluated using Term Frequency-Inverse Document Frequency (TF-IDF).

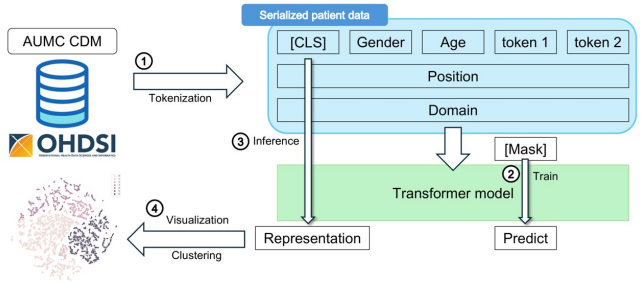


Figure 1. Study flow. First, the OMOP-CDM in Ajou university medical center was transformed into serial data according to each patient, including basic patient information such as gender and age. The data was then fed through a transformer model,

Results

Training loss converged and the model with the least validation error was selected. The clusters were not immediately recognizable with the IDs only, but some was specific enough to make weak assumptions about the cluster. For example, cluster 5 had measurements related to health screening.
The visualization of clusters using representative tokens showed better results in cluster membership. While some tokens representing a cluster was not present for most of the patient data within that cluster, certain tokens clearly showed patterns of grouping (Figure 2), closely resembling the distribution of the cluster.

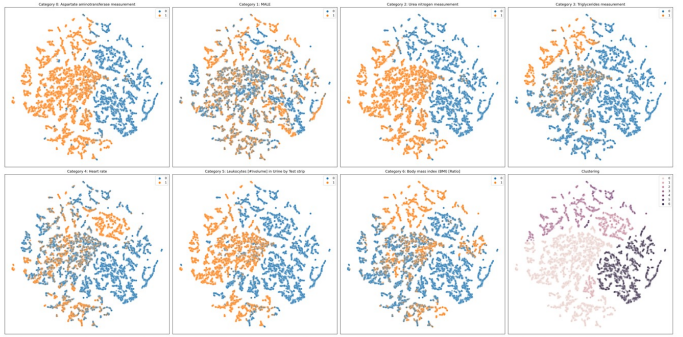


Figure 2. Top 1 representative token of each cluster visualized. (Blue) Patients without representative concept ID of cluster 0 to 6. (Orange) Patients with representative concept ID of cluster 0 to 6.

Conclusions

In this study, we trained a BERT-based clinical foundation model using data from electronic health record converted to OMOP-CDM. The latent space was visualized using dimension reduction techniques and clusters with explainable characteristics were found in some cases. A better optimized approach with different architectures or training method may lead to a better intuitive understanding about the data contained using OMOP-CDM.

Acknowledgement

This research was funded a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR16C0001) and this research was supported by a Government-wide R&D Fund project for infectious disease research (GFID), Republic of Korea (grant number: HG22C0024, KH124685).

Contact: manjmin6@gmail.com / Min-Gyu Kim

#OHDSISocialShowcase This Week

Thursday

Harmonization of routine care data from hospitals in the Digital Oncology Network for Europe (DigiONE) into Observational Medical Outcomes Partnership (OMOP) databases reveals changes in the number of new primary cancers diagnosed and 12-month survival during COVID-19 lockdowns

(Stelios Theophanous, Hayley Fenton, Aiara Lobo Gomes, Elisabeth Ross, Joëlle Thonnard, Andrea Wolf, Christian Brandts, Anne-Lore Bynens, Geoff Hall, Sara Bachir, Edward Bolton, Olivier Bouissou, Daniel Brucker, Susan Lara Cheeseman, Alix Collard, Andre Dekker, Prabash Galgane Banduge, Lars Halvorsen, Dennis Kadioglu, Petros Kalendralis, Junaid Khan, Piers Mahon, Timo Schneider, Linnea Lilja Ilona Schumann, Sarah Seager, Alberto Traverso, Aline van Maanen, Cédric van Marcke, Abishaa Vengadeswaran, Janina Wörmann, Tasmeia Yousaf, Rosie McDonald, Elin Hallan Naderi)

Harmonization of routine care data from hospitals in the Digital Oncology Network for Europe (DigiONE) into Observational Medical Outcomes Partnership (OMOP) databases reveals changes in the number of new primary cancers diagnosed and 12-month survival during COVID-19 lockdowns

S Theophanous*[1], H Fenton*[2], A Lobo Gomes[3], E Ross[4], J Thonnard[5], A Wolf[6], C Brandts[6], A-L Bynens[7], G Hall[1], S Bachir[8], E Bolton[1], O Bouissou[4], D Brucker[6], S Cheeseman[1], A Collard[5], A Dekker[3], P Galgane Banduge[3], L Halvorsen[9], I Kaczmarczyk[10], D Kadioglu[8], P Kalendralis[3], J Khan[11], P Mahon[2], T Schneider[6], L Schumann[6], A Traverso[12], A van Maanen[5], C van Marcke[5], A Vengadeswaran[8], J Wörmann[6], J Yeap[2], T Yousaf[4], R McDonald**[2], E Hallan Naderi***+4

*Joint first authors; **Joint last authors; +corresponding author: elinad@ous-hf.no. [1]Leeds Teaching Hospitals NHS Trust, United Kingdom. [2]IQVIA Ltd, United Kingdom. [3]Department of Radiation Oncology (Maastro), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, The Netherlands. [4]Oslo University Hospital, Norway. [5]Cliniques Universitaires Saint-Luc, Brussels, Belgium. [6]University Cancer Center, University Hospital Frankfurt, Germany. [7]Maastricht University Medical Center+, The Netherlands. [8]Goethe University Frankfurt, University Medicine, Institute of Medical Informatics (IMI), Frankfurt am Main, Germany. [9]edenceHealth NV, Belgium. [10]OMOP & PPG Solutions, IQVIA Ltd, London, United Kingdom. [11]IQVIA Ltd, India. [12]IRCCS San Raffaele Hospital, Milan, Italy.

INTRODUCTION

- The curation, standardization, and harmonization of routine care data across hospitals is complex due to its often unstructured, incomplete, and dissimilar nature.
- This study standardized and harmonized data in OMOP across 5 hospitals in Europe to investigate the impact of COVID-19 lockdowns on new cancer diagnoses and 12-month survival covering the period before the pandemic, during peak lockdowns, and the subsequent recovery phases.

METHODS

- Five academic, public, and general oncology treating DigiONE hospitals transformed a core dataset from their electronic medical records, into local OMOP databases to establish the first pan-cancer European hospital OMOP network for federated analysis[1] (Figure 1).
- Patients with a diagnosis of one of the 11 cancer groups (breast, prostate, lung, colorectal, upper gastrointestinal, hematological, gynecological, head and neck, non-prostate urological, melanoma, brain) between 01-Jul-2018 and 30-Sep-2023 were included.
- Interrupted time-series analysis using generalized least squares was applied to assess changes in the volume of cancer diagnoses at each hospital, as well as overall as a network for each cancer group. Data between Jul-2018 to Jun-2019 were used as the baseline.
- For patients with a 12-month follow up, 12-month overall survival was described by three-month (quarterly) cohorts based on index date.

Figure 1. Approach to harmonizing data across five hospitals

Study design committee aligned on objectives and variable requirements
Study documentation development – protocol & statistical analysis plan
Ethics approval from participating hospitals
Data source identification with each hospital data lake
Data standardisation • Mapping of source diagnosis codes into standard OMOP concepts[2]
Extract, transform, load (ETL) by hospitals following DigiONE network ETL handbook
OMOP database quality check (QC) following DigiONE network QC points
Medical code list development and validation by clinicians
Patient identification from OMOP databases using SQL • Output a flat file with patient minimal dataset for analysis, which sits within respective hospital
Analysis scripts sent to hospital databases for on-premise analysis against the minimal dataset • Output hospital-level aggregated results to be shared externally
Meta-analysis of hospital-level results for network level results

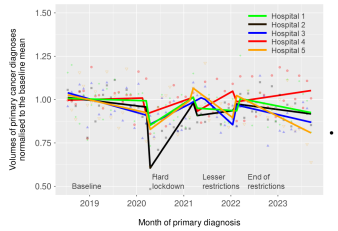
RESULTS

- The study included 124,682 patients across 5 hospitals. There was a four-fold difference in the average number of monthly primary cancer diagnoses between the smallest and the largest hospital (Table 1).
- The mean age at index was 63.9 years and 50.13% were male. There was no statistical significance change in age and sex groups over the study period.

Table 1: Volume of monthly diagnoses of all eligible cancer groups over the study period

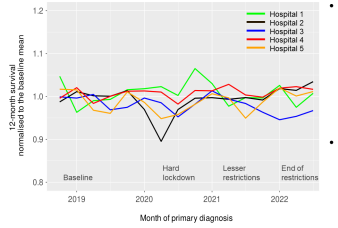
	Hosp 1	Hosp 2	Hosp 3	Hosp 4	Hosp 5
Average	197	356	240	763	172
Minimum	149	186	187	592	111
Maximum	255	428	290	907	220

Figure 2. Change in monthly primary cancers diagnoses in each hospital, across phase transitions using country-specific dates



- A statistically significant change in monthly volumes of primary cancers from baseline to hard lockdowns (all in March 2020) was observed in Hospitals 1 and 3 (p<0.05) and in Hospitals 2 and 4 (p<0.001).
- Hospitals 2 and 4 also showed a notable change in monthly volumes of primary cancers from hard lockdowns to lesser restrictions (end March to mid April 2021, p<0.005).

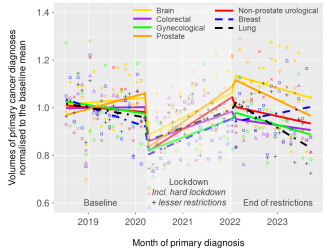
Figure 3. Change in proportion of patients surviving to 12 months from index in each hospital, across phase transitions using country-specific dates



- Proportion of newly diagnosed cancer patients surviving 12 months mirrored the drop in volume of primary cancer diagnoses from baseline to hard lockdown, particularly in Hospitals 2, 3 and 5.
- The trend of poorer survival during hard lockdowns which returns to pre-pandemic levels by the end of the study period was observed in four hospitals.
- Nevertheless, the lack of data collected on disease stage and cause of death, including COVID-19 related deaths, limits the interpretation of the current survival data and its association with the volume of primary cancer diagnoses.

Figure 4. Change in new diagnosis volumes in each cancer group as a network

Only shows cancer groups with a significant step change



- The overall number of primary cancer diagnoses significantly decreased from baseline to lockdown across the network (p<0.005).
- Statistically significant declines were observed in prostate (p<0.005), brain, colorectal, non-prostate urological, and gynaecological (p<0.05) cancers.
- Lung and breast cancer showed near-significant declines (p=0.057 and p=0.078, in dashed regression lines).

CONCLUSION

- This first pan-cancer study from the DigiONE network showcases the ability of 5 European hospitals to standardize and harmonize data in OMOP for over 120,000 patients collectively, collaborating on **data analysis in a privacy-preserving manner without transfer of patient level data**.
- The findings presented here showcase the use of the harmonized data for **comparative analyses of care practices and outcomes across different countries**, as well as **trends at the European network level**.
- The robust approach to OMOP data integration and harmonization at DigiONE hospitals not only **enhances the overall data reliability**, but also **expands the analyzable patient population**, facilitating faster collaborations for generating reliable real-world evidence in precision oncology research.

References

1. Mahon et al. Nat. Med., 30, 334–337 (2024)
2. Fenton et al. OHDSI Europe 2024 Poster

Acknowledgement

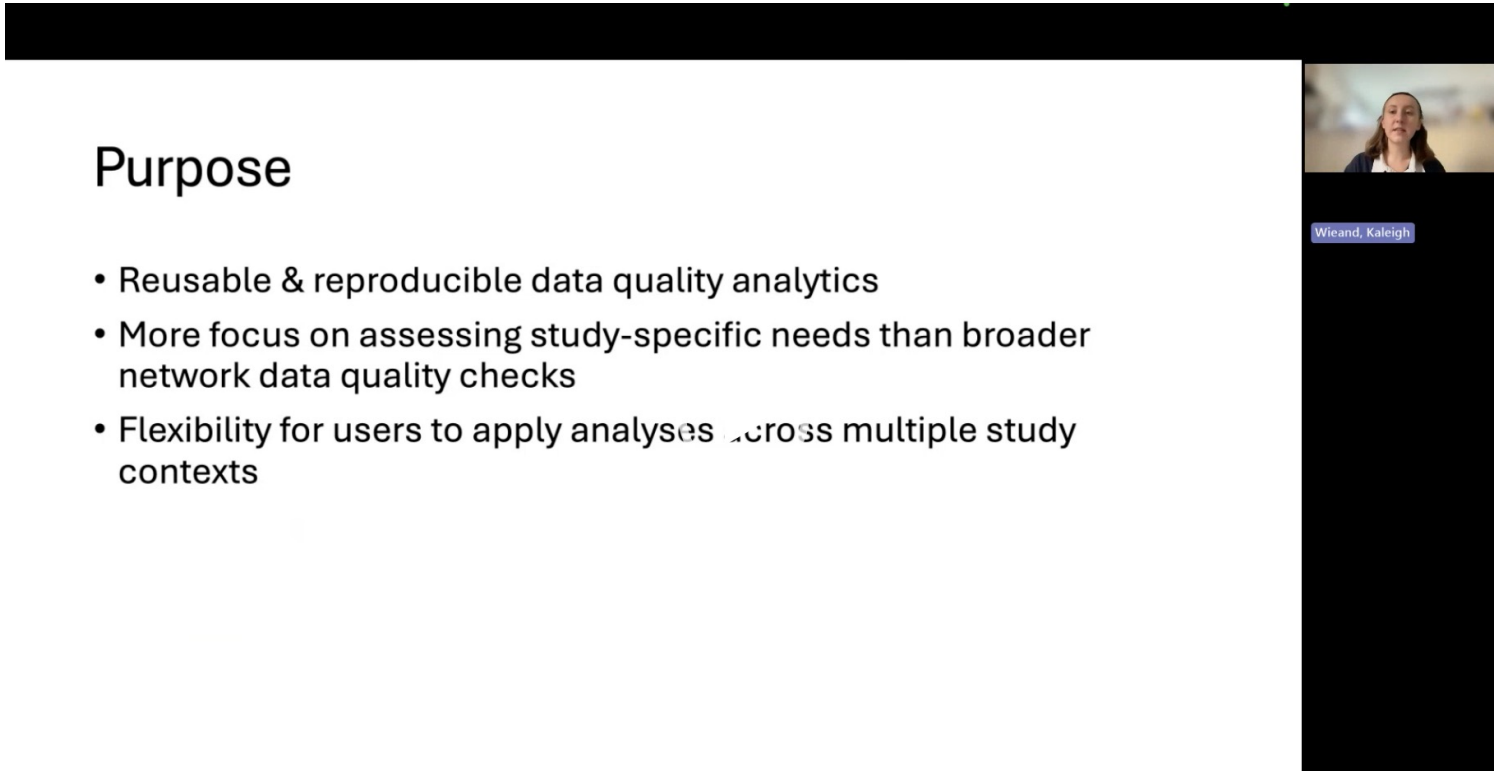
This work is part of the DigiCORE pan-European research network comprised of academic cancer centers and two industrial members (IQVIA and Illumina). The DigiONE project was funded by IQVIA and Illumina.

#OHDSISocialShowcase This Week

Friday

Executing a Reusable Framework for Study-Specific Data Quality Analysis (Demo)

(Kaleigh Wieand, Hanieh Razzaghi, Kim Dickinson, Michael Kahn, Jason Roy)



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?

Dec. 3: Recent OHDSI/OMOP Publications

Mathilde Fruchart

Data Scientist
CHU de Lille

Transforming Primary Care Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study

Aaron Eisman

PGY-1 Internal Medicine Resident
Yale School of Medicine

Learning health system linchpins: information exchange and a common data model

Albert Prats-Uribe

Senior Clinical Research Fellow in Public Health, Health Data Sciences
Oxford University

Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU)

Jiayi (Jessie) Tong

Assistant Professor, Department of Biostatistics
Johns Hopkins University

Advancing Interpretable Regression Analysis for Binary Data: A Novel Distributed Algorithm Approach

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