



Meet New Members of the OHDSI Community

OHDSI Community Call
July 2, 2024 • 11 am ET



Upcoming Community Calls

Date	Topic
June 25	Recent OHDSI Publications
July 2	Newcomer Introductions
July 9	Building The OHDSI Evidence Network Sprint
July 16	HowOften Initiative & Early Results
July 23	Building The OHDSI Evidence Network Sprint
July 30	Advances in Patient-Level Prediction
Aug. 6	Building The OHDSI Evidence Network Sprint
Aug. 13	Global Symposium Tutorials
Aug. 20	Building The OHDSI Evidence Network Sprint
Aug. 27	canceled due to ISPE 2024
Sept. 3	New Standardized Vocabularies Release



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!



Congratulations to the team of **Fateme Nateghi Haredasht, Sajjad Fouladvand, Steven Tate, Min Min Chan, Joannas Jie Lin Yeow, Kira Griffiths, Ivan Lopez, Jeremiah W. Bertz, Adam S. Miner, Tina Hernandez-Boussard, Chwen-Yuen Angie Chen, Huiqiong Deng, Keith Humphreys, Anna Lembke, L. Alexander Vance, and Jonathan H. Chen** on the publication of **Predictability of buprenorphine-naloxone treatment retention: A multi-site analysis combining electronic health records and machine learning in *Addiction*.**

Received: 7 February 2024 | Accepted: 19 May 2024
DOI: 10.1111/add.16587

RESEARCH REPORT

ADDICTION

SSA

Predictability of buprenorphine-naloxone treatment retention: A multi-site analysis combining electronic health records and machine learning

Fateme Nateghi Haredasht^{1,2,3} | Sajjad Fouladvand^{1,2,3} | Steven Tate⁴ | Min Min Chan^{5,6} | Joannas Jie Lin Yeow^{5,6} | Kira Griffiths^{5,6} | Ivan Lopez^{1,2,3} | Jeremiah W. Bertz⁷ | Adam S. Miner⁴ | Tina Hernandez-Boussard^{1,2,3} | Chwen-Yuen Angie Chen⁸ | Huiqiong Deng⁴ | Keith Humphreys⁴ | Anna Lembke⁴ | L. Alexander Vance^{5,6} | Jonathan H. Chen^{1,2,3}

¹Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, California, USA

²Division of Hospital Medicine, Stanford University, Stanford, California, USA

³Clinical Excellence Research Center, Stanford University, Stanford, California, USA

⁴Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California, USA

⁵Holmusk Technologies, Inc., Singapore, Singapore

⁶Holmusk Technologies, Inc., New York, New York, USA

⁷Center for the Clinical Trials Network, National Institute on Drug Abuse, North Bethesda, Maryland, USA

⁸Division of Primary Care and Population Health, Department of Medicine Stanford University School of Medicine, Stanford, California, USA

Correspondence

Fateme Nateghi Haredasht, Stanford Center for Biomedical Informatics Research, 3180 Porter Drive, Palo Alto, CA, 94304, USA.

Abstract

Background and aims: Opioid use disorder (OUD) and opioid dependence lead to significant morbidity and mortality, yet treatment retention, crucial for the effectiveness of medications like buprenorphine-naloxone, remains unpredictable. Our objective was to determine the predictability of 6-month retention in buprenorphine-naloxone treatment using electronic health record (EHR) data from diverse clinical settings and to identify key predictors.

Design: This retrospective observational study developed and validated machine learning-based clinical risk prediction models using EHR data.

Setting and cases: Data were sourced from Stanford University's healthcare system and Holmusk's NeuroBlu database, reflecting a wide range of healthcare settings. The study analyzed 1800 Stanford and 7957 NeuroBlu treatment encounters from 2008 to 2023 and from 2003 to 2023, respectively.

Measurements: Predict continuous prescription of buprenorphine-naloxone for at least 6 months, without a gap of more than 30 days. The performance of machine learning prediction models was assessed by area under receiver operating characteristic (ROC-AUC) analysis as well as precision, recall and calibration. To further validate our approach's clinical applicability, we conducted two secondary analyses: a time-to-event analysis on a single site to estimate the duration of buprenorphine-naloxone treatment



OHDSI Shoutouts!



Congratulations to the team of **Gyunam Park, Yaejin Lee, and Minsu Cho** on the publication of **Enhancing healthcare process analysis through object-centric process mining: Transforming OMOP common data models into object-centric event logs** in the *Journal of Biomedical Informatics*.



Journal of Biomedical Informatics

Available online 27 June 2024, 104682

In Press, Journal Pre-proof [What's this?](#)



Original Research

Enhancing healthcare process analysis through object-centric process mining: Transforming OMOP common data models into object-centric event logs

[Gyunam Park](#)^a , [Yaejin Lee](#)^b , [Minsu Cho](#)^b  

[Show more](#) 

[+](#) [Add to Mendeley](#) [🔗 Share](#) [📄 Cite](#)

<https://doi.org/10.1016/j.jbi.2024.104682> 

[Get rights and content](#) 

Abstract

Objectives:

This study aims to enhance the analysis of healthcare processes by introducing Object-Centric Process Mining (OCPM). By offering a holistic perspective that accounts for the interactions among various objects, OCPM transcends the constraints of conventional patient-centric process mining approaches, ensuring a more detailed and inclusive understanding of healthcare dynamics.



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





Upcoming Workgroup Calls



Date	Time (ET)	Meeting
Wednesday	9 am	Psychiatry
Wednesday	4 pm	Joint Vulcan/OHDSI meeting
Thursday	11 am	Industry
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Clinical Trials
Friday	11:30 am	Steering Group
Monday	10 am	CDM Survey Subgroup
Monday	10 am	Africa Chapter
Monday	11 am	Early-Stage Researchers
Tuesday	9 am	OMOP CDM Oncology Genomic Subgroup



Health Equity / OMOP + FHIR Cross-Working Group Collaboration

Kelly Davidson, MN, MSc, RN, CPMHN(C), CTSS
University of Victoria, British Columbia

HL7 Gender Harmony Project: Sex and Gender Representation

Wednesday, July 10 at 12pm ET ([Meeting Link](#))





OHDSI Evidence Network



5 Rationale and Background

The Observational Health Data Sciences and Informatics (OHDSI) federated network is a collaborative effort aimed at leveraging healthcare data from multiple institutions for large-scale federated observational research. In its current state there are over 500 data sources from over 49 countries mapped to the OMOP Common Data Model, the standard that enables such ambitious evidence generation. One major challenge of federated network studies is the assessment of network data quality, study feasibility and data fitness-for-use across these data sources in such a way that does not strain the time and resources of data holders while still supporting rigorous evidence generation that engenders trust and buy-in from the larger research community.

To facilitate collaborative research efforts and ensure the quality and integrity of the data across the OHDSI network, it is imperative to understand the characteristics and variability of the databases within the network. This study aims to collect summary statistics from participating sites to describe the databases and learn about the network as a whole. The output of the study will inform and enhance the research capabilities of the OHDSI community by enabling rapid data quality and fitness-for-use assessments.

5.1 Research Questions

The main research question of this study is:

What are the population-level characteristics of the databases within the OHDSI federated network?



Please fill out a short google form to indicate your intent to join the study



July Newsletter Is Available



The Journey Newsletter (July 2024)

Evidence has been a popular word in the OHDSI community recently. We focused a community call around the application of large language models in the evidence generation process, and we initiated an effort to build and develop the OHDSI Evidence Network. Learn more about these and everything else happening around the community in the latest OHDSI newsletter.

[#JoinTheJourney](#)

Video Podcast: Evidence Network, LLMs & More

OHDSI On The Journey #JoinTheJourney

In the latest On The Journey videocast, Patrick Ryan and Craig Sachson discuss the OHDSI Evidence Network, the emerging potential of large language models within the community, and reflections on the Europe Symposium. (If video does not appear, please click "View this email in your browser")

Community Updates

Where Have We Been?

- The 2024 OHDSI Europe Symposium was held June 1-3 in Rotterdam, Neth., and welcomed 350 collaborators for three days of sharing, learning and networking. It included two days of workshops/tutorials and a full-day conference that highlighted exciting progress ongoing in Europe, including the growth among National Nodes, progress by DARWIN EU®, and plenty more. Leaders from the event [shared a recap during a recent community call](#).
- OHDSI collaborators answered the call once again to share their research at the annual global symposium. We received more than 140 brief report submissions, including over 20 software demos, for the 2024 Global Symposium, which will be held Oct. 22-24 (more on that below). Congratulations to everybody in the community who shared an abstract, and thank you to the members of our Scientific Review Committee, who have begun the process of reviewing the submissions.
- Videos/slides from recent tutorials in both Japan ([Towards the Promotion of RWD Utilization Using OMOP CDM](#)) and Thailand ([Transforming Evidence Generation in Thailand with OHDSI/OMOP](#)) are now available on the OHDSI website.

Where Are We Now?

- The OHDSI Evidence Network, first introduced at the 2023 Global Symposium, consists of organizations equipped with access to one or more databases standardized to the OMOP CDM who express a keen interest in participating in OHDSI network studies. Building the OHDSI Evidence Network will be a major initiative this summer, and OMOP CDM users can share their interest in joining [by filling out this form](#).
- Yonas Ghebremichael-Weldeselassie, Senior Research Fellow in Medical Statistics at Warwick Medical School, will lead the next edition of the [CBER BEST Seminar Series](#) on Wednesday, July 17 (11 am ET). Please check out the event homepage for the meeting link, additional details and all past presentations, including a session on [Applying Machine Learning in Distributed Networks to Support Activities for Post-Market Surveillance of Medical Products: Opportunities, Challenges, and Considerations](#) by Harvard's Jenna Wong last month.

Large Language Models Can Enhance OHDSI Evidence Generation Process

June 18: Application of LLMs in RWE Generation



Yilu Fang
PhD Student, Columbia University



João Almeida
Assistant Professor, Chief Information Security Officer,
University of Aveiro



Martijn Schuemie
Research Fellow, Epidemiology Analytics, Janssen Research and Development

Large language models can analyze large datasets, extract insights, and generate evidence-based reports, aiding in real-world decision-making by providing accurate, comprehensive information efficiently. The [OHDSI Generative AI and Foundational Models workgroup](#) focuses on advancing healthcare research and improve patient outcomes through the innovative application of generative AI and foundational models.

Three members of the OHDSI global community joined the June 18 community call to present recent research in the area of large language models. You can find the video and slides for each presentation using the link below.

Knowledge-guided Generative AI For Automated Taxonomy Learning From Drug Labels – Yilu Fang • PhD Student, Columbia University
A Chatbot to Streamline Biomedical Data Discovery and Analysis – João Almeida • Chief Information Security Officer, University of Aveiro
Generative AI for real-world evidence – Martijn Schuemie • Research Fellow, Epidemiology Analytics, Janssen Research and Development

[LLMs Presentations & Slides](#)

OHDSI Evidence Network Introduces New Protocol & Network Study, Welcomes OMOP Users Into Crucial Community Research Asset

The OHDSI Evidence Network consists of organizations equipped with access to one or more databases standardized to the OMOP CDM who express a keen interest in participating in OHDSI network studies. Collaboratively, OHDSI Evidence Network partners share aggregate summary statistics about their databases, which are used to support Database Diagnostics, helping identify databases within the network that are fit-for-use for particular research questions. Additionally, partners have the opportunity to opt in and contribute to network studies proposed by the OHDSI community.

To carry out our mission, we need an active and willing global network of data partners, and we need the ability to quickly identify those that might be the right fit for a specific clinical research question. Last year we piloted this effort through the Save our Sisyphus challenge and are now ready to move forward based on our learnings. If you would like to learn more please check out [this update](#) from the June 11 community call.

The OHDSI Evidence Network workgroup is excited to initiate a network study that will describe the OHDSI Network in a publication, and will also create an open public resource designed to facilitate evidence generation faster and better than ever by building on methodologies developed by thought leaders around the world. You can access the protocol below. Come join us on this exciting journey!

[Access the OHDSI Evidence Network Protocol \(Brief Survey\)](#)



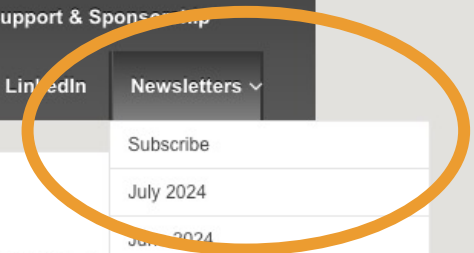
July Newsletter Is Available



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

- Who We Are ▾
- Updates & News ▾
- Standards
- Software Tools ▾
- Network Studies ▾
- Community Forums ▾
- Education ▾
- New To OHDSI? ▾
- Community Calls ▾
- Past Events ▾
- Workgroups ▾
- 2023 'Our Journey' Annual Report
- This Week In OHDSI
- Support & Sponsorship
- CBER Best Seminars
- 2024 Europe Symposium
- 2024 Global Symposium ▾
- Github
- YouTube
- Twitter
- LinkedIn
- Newsletters ▾



- Subscribe
- July 2024
- June 2024
- May 2024
- April 2024
- March 2024
- February 2024
- Full Archive

Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

Join Us At The 2024 Symposium

Registration is now open for the 2024 Symposium, which will be held October 1-3 at the Hyatt Regency Hotel in New Brunswick, NJ, USA. Check out the event page for details on the collaborator showcase, tutorial offerings, workshop activities, and more!



Next CBER Best Seminar: July 17

Speaker: Yonas Ghebremichael-Weldeselassie, Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK

Topic: Yonas Ghebremichael-Weldeselassie, Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK

Date/Time: Wednesday, July 17, 11 am ET

ohdsi.org/cber-best-seminar-series

Upcoming Seminars

— July 17, 2024 (11 am) - Yonas Ghebremichael-Weldeselassie, Warwick Medical School

Topic: A modified self-controlled case series method for event-dependent exposures and high event-related mortality, with application to COVID-19 vaccine safety

Presenter: Yonas Ghebremichael-Weldeselassie, Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK

[Watch This Seminar](#)

Abstract:

We propose a modified self-controlled case series (SCCS) method to handle both event-dependent exposures and high event-related mortality. This development is motivated by an epidemiological study undertaken in France to quantify potential risks of cardiovascular events associated with COVID-19 vaccines. Event-dependence of vaccinations, and high event-related mortality, are likely to arise in other SCCS studies of COVID-19 vaccine safety. Using this case study and simulations to broaden its scope, we explore these features and the biases they may generate, implement the modified SCCS model, illustrate some of the properties of this model, and develop a new test for presence of a dose effect. The model we propose has wider application, notably when the event of interest is death.

Bio: Yonas Weldeselassie is a Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK. He graduated in statistics and demography from University of Asmara, Eritrea and went on to become an assistant lecturer in Mekelle University, Ethiopia, and then a Senior Research Fellow in Medical Statistics at Warwick Medical School, division of Population Evidence and Technologies. He earned a Msc in Biostatistics from Hasselt University, Belgium and PhD in statistics from the Open University, UK. After working as a research associate, on MRC project 'Software tools and online resources for the self-controlled case series method and its extensions', at the department of mathematics and statistics, the Open University since 2014, he joined Warwick Medical School in June 2017. His main research interest is in medical statistics specially in the methodological development and application of the self-controlled case series (SCCS) method. He published a book on SCCS with Paddy Farrington and Heather Whitaker, and he is currently working on early prediction of gestational diabetes mellitus.





#OHDSI2024 Registration Is Open!

Registration is OPEN for the 2024 OHDSI Global Symposium, which will be held **Oct. 22-24** at the **Hyatt Regency Hotel in New Brunswick, N.J., USA.**

Tuesday: Tutorials

Wednesday: Plenary/Showcase

Thursday: Workgroup Activities

ohdsi.org/OHDSI2024





AI for Reliable and Equitable RWE Generation in Medicine Workshop: July 9

The workshop focuses on advancing the understanding and exploring the transformative role of artificial intelligence (AI) in analyzing real-world data (RWD) for real-world evidence (RWE) generation.



AIME Workshop on AI for Reliable and Equitable Real-World Evidence Generation in Medicine
July 9, 1:45pm - 6:00pm

Keynote Speakers



George Hripcsak, MD, MS,
*Professor, Biomedical Informatics
Columbia University*



Scott L. DuVall, PhD
*Director, VA Informatics and Computing Infrastructure
VA Salt Lake City Health Care System
Professor, Department of Internal Medicine
Division of Epidemiology, University of Utah*

Panelists



David K. Vawdrey, PhD
*Chief Data Informatics Officer
Steele Institute for Health Innovation
Geisinger*



Adam Wilcox, PhD
*Professor, Medicine
Director, Center for Applied Clinical Informatics
Washington University in St. Louis*



Linying Zhang, PhD
*Assistant Professor, Biostatistics
Washington University in St. Louis*

Organizers and Sponsors



Institute for Informatics,
Data Science and Biostatistics



Rising Stars



Michael Oberst, PhD
*Assistant Professor, Computer Science
Johns Hopkins University
"Auditing Fairness under Unobserved Confounding"*



Zhiyu Wan, PhD
*Postdoctoral fellow, Biomedical Informatics
Vanderbilt University
"Promoting Responsive Real-World Data Sharing in Medicine Using AI Agents"*



Victoria Tiase, PhD, RN-BC,
*Research Assistant Professor,
Biomedical Informatics
University of Utah
"Developing a Logical Data Model for Nursing Workload"*



Laura Wiley, PhD, MS,
*Associate Professor, Biomedical Informatics
Chief Data Scientist, Health Data Compass
University of Colorado Anschutz Medical Campus*

<https://medicine.utah.edu/dbmi/aime/ai-reliable>



#OHDSISocialShowcase

MONDAY

Transforming lung cancer EHR data into the OMOP CDM: A case study of Non Small Cell Lung Cancer

(Evangelos Chandakas (Handakas), Ping Sun)

High robustness of OHDSI tools and OMOP CDM to support transformation of lung oncology real world data

Title: Transforming lung cancer EHR data into the OMOP CDM: A case study of Non-Small Cell Lung Cancer

Background: Common Data Models (CDMs) are essential for data harmonization, leading to significant improvements in healthcare and research domains. CDMs enhance transparency, increase the reliability of medical research, and ensure efficient, timely evidence generation for decision-making¹. Despite the continuous progress and development of CDMs in oncology real-world data^{2,3}, challenges remain. Lung cancer is the leading cause of cancer-related mortality worldwide, with an estimated 2.2 million new diagnoses and 1.8 million deaths annually cases⁴. Non-Small Cell Lung Cancer (NSCLC) accounts for 80-85% of these cases⁴. This study evaluates the adoption of the OMOP CDM for lung cancer oncology real-world data, exploring the opportunities and challenges of implementing the OMOP CDM in lung oncology data.

Results:

- Applied to anonymized clinical and laboratory data revealed a high success rate, with over 99% of fields effectively transformed into OMOP CDM concepts, affirming the robustness of the data transformation process.
- American Joint Committee on Cancer (AJCC) cancer staging manual (eighth edition) able to accurately translated cancer stages while retaining essential clinical details.

Methods and Material:

- Database: Flatiron Enhanced Datamart (EDM), a subset of patients with Non-Small Cell Lung Cancer (NSCLC).
- Observational and retrospective data of over 90,000 anonymized patients.
- Data mapping using ATHENA, USAGI and in-house R and SQL pipelines.
- Quality assessment using in-house R and SQL pipelines.
- Large Language Model (ChatGPT 4).

Conclusions:

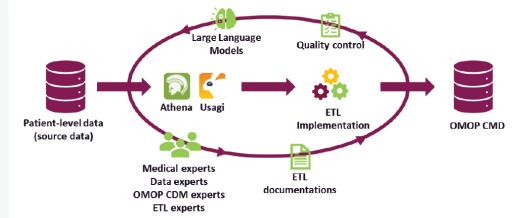
- This study highlighted the significant challenges in mapping NSCLC patient data to the OMOP CDM and presented a framework for addressing these challenges.
- We underscored the importance of collaboration and quality assurance measures in ensuring data accuracy and reliability in oncology.
- We demonstrated the potential of a common data model to support large-scale clinical and translational research initiatives.
- Large Language Model can lead to more efficient ETL workflows and improved decision-making capabilities.

References:

1. Kent, S. et al. Common problems, common data model solutions: evidence generation for health technology assessment. *Pharmacoeconomics* 39, 275-285 (2021).
2. Osterman, T. J., Terry, M. & Miller, R. S. Improving cancer data interoperability: the promise of the Minimal Common Oncology Data Elements (mCODE) initiative. *JCO Clinical Cancer Informatics* 4, 983-1001 (2020).
3. Carus, J. et al. Mapping the Oncological Basis Dataset to the Standardized Vocabularies of a Common Data Model: A Feasibility Study. *Cancers* 15, 4059 (2023).
4. International Agency for Research on Cancer, L. W. *Globocan 2022 Fact Sheet - Cancer today*, <https://gco.iarc.fr/Today/facts> (2022).
5. <https://www.ohdsi.org/software-tools/>

Highlights

- Tumour progression and metastasis were effectively integrated into the *Episode* and *Episode Events* tables, with additional mapping to the *Observation* and *Observation period* tables to ensure comprehensive capture of these events within the ATLAS⁵.
- Drug regimens were also mapped to *Drug Exposure* and *Drug Era* tables.
- Limitations encountered during the ETL process was the transformation of general data concept or outer subsets categories (e.g. "other types of mutations").
- Line of therapy were mapped with HemOnc vocabulary.
- ChatGPT provided significant robust solutions in ELT implantation and accelerated the preparation of ETL documentation.



Evangelos Chandakas (Handakas), Ping Sun





#OHDSISocialShowcase

TUESDAY

Four Complexities when mapping NCRAS to the OMOP CDM

(**Laura Kerr**, Abigail Carter)

Four Complexities when mapping NCRAS to the OMOP CDM



Laura Kerr, Abigail Carter
Genomics England

INTRODUCTION

Genomics England (GEL) is a global leader in enabling genomic medicine and research, focused on creating a world where everyone benefits from genomic healthcare.

- It enriches its primary clinical data (participant information) with secondary data (supporting healthcare records) which includes the National Cancer Registration and Analysis (NCRAS) data.
- The data is made available to researchers in an isolated Research Environment.
- There is a desire from researchers to run federated queries, however this is hampered by the disparate sources of data and their differing data models.
- GEL have therefore mapped NCRAS to the OMOP CDM to empower researchers and have made the mappings publicly available.

MAPPING APPROACH

1. Retrieve source attributes and enumerations from NCRAS data dictionary.
2. Identify OMOP domain best suited to source attribute.
3. Manually map source attributes to concepts by matching on descriptions.
4. Standard attributes are used wherever possible.

COMPLEXITIES

1. MAPPING GRANULARITY

- Opted to always map to the most granular concept that wouldn't lead to researchers inferring information that wasn't present in the source data.
- Utilised the CDM to provide context where desired levels of granularity have not been possible.
- For example, the most granular match found for the source enumeration shown below is 'Index of Multiple Deprivation (England)'. The value_as_string' field, has then been used to provide more detail.

Source Attribute	Source Enumeration	OMOP Domain Fields	Field Value
quintile2004	Deprivation Quintile 2004	observation_concept_id	S0812802
	1 - least deprived	value_as_string	Deprivation Quintile 2004, 1 - least deprived

- GEL is interested in defining a new vocabulary to support these niche cases.

3. CLINICAL CODES

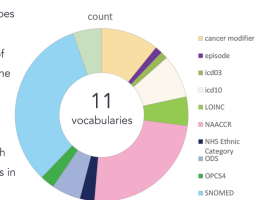
- The NCRAS data is a collection of secondary data sources.
- Clinical coded source attributes in cases therefore vary in format.
- Deviations in format need reformatting to maximise mapping quality. A simple example is shown below.

Source Attribute	Source Value	Target Value
site_inst19_02	C049	C049

- A benefit of mapping the data to OMOP CDM is the identification of coded entries that require such reformatting, and their target format.

2. VOCABULARIES

- The NCRAS dataset describes a broad scope of events.
- The use of a broad range of vocabularies to represent the data reflects this.
- GEL are interested in understanding how researchers will interact with the diversity of vocabularies in the future.



4. ONCOLOGY EXTENSION

- Used the Oncology extension on GEL mental health datasets as the structure is ideal for grouping many different episodes of care.
- More data is required to use the NCRAS dataset to populate episodes of care and patient pathway domains.
- GEL do not wish to infer information that we do not have so have not used the oncology extension here.

CONCLUSION

The OMOP mapping gave a high success rate, with almost all clinical information being mapping to the OMOP CDM. GEL expects the mapped NCRAS data to vastly improve the user experience in our research environment. It is recognised that there is scope for improvement in the mappings and feedback on the mappings is very welcome as GEL hope to iteratively improve their quality and depth.



https://gitlab.com/genomicsengland/genomics_england_publications/public-omop-mappings

Laura Kerr
Data Manager
laura.kerr@genomicsengland.co.uk

Abigail Carter
Data Wrangler
abigail.carter@genomicsengland.co.uk





#OHDSISocialShowcase

WEDNESDAY

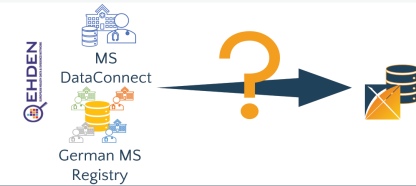
Piloting the Transformation of Multiple Sclerosis Real-World Data to the OMOP CDM: Lessons Learned

(**Tina Parciak**, Kirstin Tumler, Alexander Stahman, Emma Gesquiere, Freija Descamps, Liesbet Peeters)

OMOP CDM for data from multiple sclerosis registries and cohorts? Possible, but ...

Piloting the Transformation of Multiple Sclerosis Real-World Data to the OMOP CDM: Lessons Learned

Background: OMOP CDM is a promising option for data from MS registries and cohorts as it could enable analysis within and outside the MS community. Since OMOP was not originally designed for registry data, especially of a chronic, relapsing and progressive disease, a piloting transformation for two MS datasets was done.



Lessons Learned	
Datasets from MS registries and cohorts differ from EHR data. OMOP CDM lacks features or concepts for such data.	<ul style="list-style-type: none"> Differences in data collection methods, the lack of standards and free text fields use. Existence of relapses as a disease characteristic Existence of many different symptoms and comorbidities without granular information Importance of negative or "no evidence of" results in longitudinal follow-up
Standardisation to OMOP CDM can still result in heterogeneous outputs.	<ul style="list-style-type: none"> Create a mapping guideline with general transformation rules to promote consistency in OMOP MS databases. Additionally, encourage the use of the STEM table for wide formatted data, e.g. registry data, resulting in a simpler and more data-driven ETL implementation.
Exchange of experiences and alignment for registry-type data transformations is necessary.	<ul style="list-style-type: none"> Establishment of OHDSI Registry workgroup. Identified challenges e.g. survey data, non-valuable answers, linkage of events/fields, observation periods, missing concepts, and patient-reported outcomes.
Transforming data from MS registries and cohorts demands substantial time investment and interdisciplinary knowledge.	<ul style="list-style-type: none"> For the use case of MS, a tool will be developed that maps data in a specific format, the MSDA Core Dataset, to the OMOP CDM. Lowering the threshold for getting "OMOP-ready" and improving the speed and harmonisation in the mapping process.

Tina Parciak^{1,2,3}, Kirstin Tümler⁴, Alexander Stahmann⁵, Emma Gesquiere⁶, Freija Descamps⁶, Liesbet M. Peeters^{1,2,3}

¹University MS Center (UMSC), Hasselt - Pol, BELGIUM; ²U Hasselt, Biomedical Research Institute (BIOMED), Agoralaan, 3590 Diepenbeek, BELGIUM; ³U Hasselt, Data Science Institute (DSI), Agoralaan, 3590 Diepenbeek, BELGIUM; ⁴German Center for Diabetes Research (DZD), Munich, GERMANY; ⁵German MS Register by the German MS Society, MS Research and Project Development gGmbH (MSRP), Hannover, GERMANY; ⁶edincerehealth NV, Veiligheids 33 A, 2550 Kottich, BELGIUM





#OHDSISocialShowcase

THURSDAY

Towards all-Island sharing of Irish lymphoid blood cancers data: landscape and gap analysis

(Kluyvert Boakye Duah, Michael Quinn, Eva Szegezdi, Lisa Crawford, Aedin C. Culhane, Mark Lawler, Siobhan Glavey, Ruth Clifford, and Ian M. Overton)

Towards all-Island sharing of Irish lymphoid blood cancers data: landscape and gap analysis

Kluyvert Boakye Duah^{1,9}, Michael Quinn², Eva Szegezdi^{3,9}, Lisa Crawford¹, Aedin C. Culhane^{4,9}, Mark Lawler^{1,9}, Siobhan Glavey^{5,6,9}, Ruth Clifford^{4,7,9}, and Ian M. Overton^{1,8,9,*}

¹Patrick G. Johnson Centre for Cancer Research, Queen's University Belfast, UK, ²Belfast Trust, Health and Social Care Northern Ireland, UK, ³School of Biological and Chemical Sciences, University of Galway, Ireland, ⁴Limerick Digital Cancer Research Centre, Health Research Institute, School of Medicine, University of Limerick, Ireland, ⁵Department of Haematology, Beaumont Hospital, Ireland, ⁶Department of Pathology, Royal College of Surgeons in Ireland, Ireland, ⁷Department of Haematology, University Hospital Limerick, Ireland, ⁸Health Data Research Wales and Northern Ireland, Queen's University Belfast, UK, ⁹Health-Hub for Cancer.

*Correspondence address: E-mail: I.Overton@qub.ac.uk



Background: Sharing health data significantly improves public health, clinical care, personal care and associated research. However, privacy laws, limited data standardisation and interoperability, and insufficient data integration have made it difficult for health data to be shared across health institutions and borders. We examine the health data landscape in Northern Ireland (NI) and Ireland (IE). Also, a federated approach with multiparty homomorphic encryption is proposed to analyse and share Chronic Lymphocytic Leukaemia (CLL) and Multiple Myeloma (MM) data on the Island.

Figure 1: Cancer-related data ecosystem in NI.

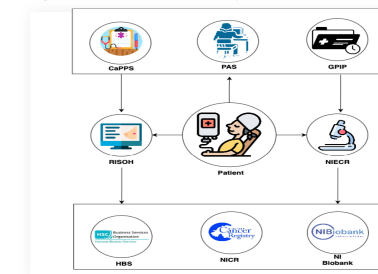
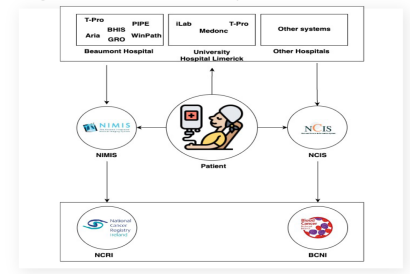
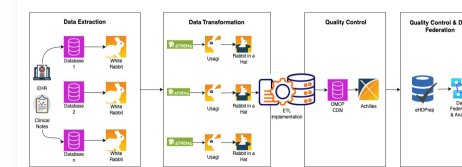


Figure 2: Cancer-related data ecosystem in IE.

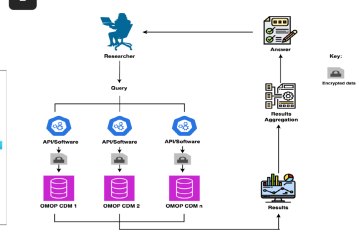


Future work

1 Mapping of Irish lymphoid blood cancers data workflow.



2 Proposed data federation and analysis workflow.



Method: Data managers and haematologists of selected health institutions were interviewed in a semi-structured manner. The interviews were done to assess the current state of Irish cancer data including CLL and MM. The nine key areas covered were a) Tools/systems/software for collecting data, b) Data sources, c) Data quality, d) Variables collected, e) FAIR principle, f) Data sharing, g) Conditions inhibiting data sharing, h) Data ownership, and i) Data governance.





#OHDSISocialShowcase

FRIDAY

Exploring Drug Utilization Patterns in Osteoporosis Therapy

(Balqis Istiqomah Gusbela, Septi Melisa, Ming-Hung Teng, Daniel C.A Nugroho, Jason C. Hsu)

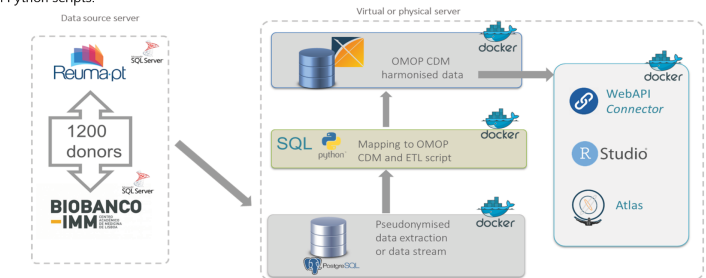
ADAPTION OF THE OMOP CDM FOR RHEUMATOLOGY: A Portuguese experience.

Harmonization of Biobanco-iMM Rheumatology Collection data to OMOP CDM

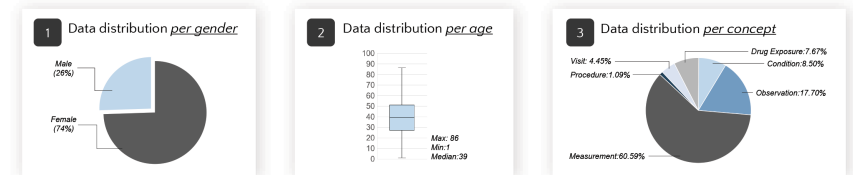
Background: The OMOP Common Data Model (OMOP CDM) is an option to store patient data and to use these in an international context. Biobanco-iMM includes biological samples (from surgery, biopsies, blood samples) which are voluntarily donated with permission for preservation and future use in biomedical research. Biobanco-iMM has a collection for rheumatology that is connected with Reuma.pt - the Rheumatic Diseases Portuguese Register from Portuguese Society of Rheumatology (SPR). Reuma.pt protocols include structured information about socio-demographic data and information about standard diagnostic criteria and clinical features of each disease. This project aimed to convert the Biobanco-iMM Rheumatology Collection data to the OMOP Common Data Model (CDM) version 5.4, using data from Biobanco-iMM and Reuma.pt databases.

Methods

Source data was extracted from the Biobanco-iMM and Reuma.pt databases to be loaded onto a PostgreSQL DBMS instance (postgresql 10+190ubuntu0.1), where CDM was also implemented. ETL was implemented through SQL procedures and orchestrated through Python scripts.



Results



Conclusion: Data harmonization, besides to contribute to the data quality improvement, also facilitates the development of relevant clinical projects. In addition, the implementation of country nodes strongly improves its outcome. In the future, CDM materialization will be manually updated based on the needs expressed by Biobanco-iMM and Rheumatology Service.

Limitation: A few Biobanco-iMM Rheumatology Collection data was not standardized due to the lack of appropriate standard vocabulary.



Catarina Tomé, Enrico Calanchi, Laura Delsante, Ângela Afonso, Daniel Silva, Ana Rita Lopes and João Eurico Fonseca





Opening: Sr AD, Real World Evidence & Analytics Boehringer Ingelheim

SR AD, Real World Evidence & Analytics

[Apply Now](#)

[< Back](#)

JOB ID - 13278

Description

The purpose of this job is to:

- Generate real world evidence (RWE) to support in-line and pipeline products.
- Provide statistical advice on the analysis of real world data (RWD) to various internal and external stakeholders.
- Contribute to the RWD acquisition strategy and tool evaluation.



Opening: Lead Director, RWE Distributed Research CVS Health

Lead Director, RWE Distributed Research

Apply

Hybrid

Full time

PA - Blue Bell

Posted 6 Days Ago

IL - Northbrook

R0268183

CT - Hartford

RI - Woonsocket

AZ - Scottsdale

[View All 7 Locations](#) ▾

Bring your heart to CVS Health. Every one of us at CVS Health shares a single, clear purpose: Bringing our heart to every moment of your health. This purpose guides our commitment to deliver enhanced human-centric health care for a rapidly changing world. Anchored in our brand – with heart at its center – our purpose sends a personal message that how we deliver our services is just as important as what we deliver.

Our Heart At Work Behaviors™ support this purpose. We want everyone who works at CVS Health to feel empowered by the role they play in transforming our culture and accelerating our ability to innovate and deliver solutions to make health care more personal, convenient and affordable.

About us



Our Work Experience is the combination of everything that's unique about us: our culture, our core values, our company meetings, our commitment to sustainability, our recognition programs, but most importantly, it's our people. Our

[Read More](#) ▾



Opening: Postdoctoral Researcher, University of Oxford



UK date and time: 18-June-2024 15:41

Applicant Options

- › [New Search](#)
- › [Login](#)
- › [Job Details](#)
- › [Help](#)
- › [Terms of Use & Privacy Policy](#)

Job Details

Postdoctoral Researcher in Real World Evidence

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD

We have an exciting opportunity for a Postdoctoral Researcher in Real World Evidence to join our Pharmaco- and Device epidemiology research group led by Professor Daniel Prieto-Alhambra at the Botnar Research Centre, NDORMS, University of Oxford. The NDORMS Pharmaco- and Device epidemiology research group is involved in a number of national and international studies exploring the conditions of use (adherence, compliance, off and on-label use) of a number of licensed drugs, devices, and vaccines for the prevention and treatment of human disease in 'real world' (routine practice) conditions.

As a Postdoctoral Researcher in Real World Evidence you will be leading or co-leading real world evidence studies, analysing real world health data mapped to the OMOP common data model and write study reports and scientific manuscripts. You will be responsible for development of analysis plans, protocols, ethical (and similar panel) submissions, governance and regulatory submissions as required for ongoing and future studies. You will carry out collaborative projects with colleagues in partner institutions, and research groups (both in public and private sector) and manage your own academic research.

You will hold a Doctoral (or be near completion) degree in epidemiology, biostatistics, real world evidence, health data sciences, or a related field. You will have experience in the use of R for statistical analysis together with experience of analysing real world data. Good track record of peer reviewed scientific publications, excellent team working and communication skills are also essential. Experience in the analysis and/or interpretation of OMOP-mapped data and experience designing and conducting cohort, self-controlled, and similar studies are desirable.

This is a full-time fixed-term appointment for 2 years.

The closing date for this position is 12 noon on 1 July 2024. You will be required to upload a CV and supporting statement as part of your online application.

Contact Person : HR Team, NDORMS
Contact Phone :
Pay Scale : STANDARD GRADE 7
Salary (£) : £36,024 - £44,263 p.a.

Vacancy ID : 173456
Closing Date & Time : 01-Jul-2024 12:00
Contact Email : hr@ndorms.ox.ac.uk



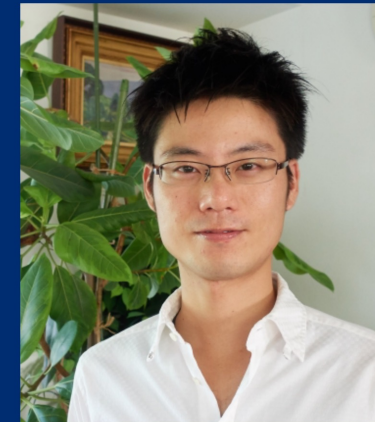
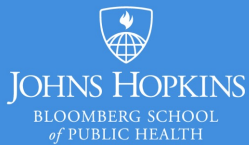
Openings: Postdoctoral Fellow, Johns Hopkins Univ.

PHARMACOEPIDEMIOLOGY POST-DOCTORAL TRAINING PROGRAM

Co-Directors: Caleb Alexander, MD, MS and Jodi Segal, MD, MPH

The **Pharmacoepidemiology Training Program** at the Johns Hopkins Bloomberg School of Public Health (BSPH) is currently **seeking to support postdoctoral fellows**. All supported trainees work with core faculty on existing or newly developed research projects on pharmacoepidemiology, so as to optimize the safe and effective use of medicines to treat heart, lung and blood diseases in the United States. |

Deadline for applications: rolling





Opening: Junior Research Software Engineer, Tufts



INFORMATICS

Research Services

COVID-19 Information and Resources

Data and Safety Monitoring Board (DSMB) Program

Center for Clinical Trials (CCT)

Program Evaluation

Qualitative and Mixed Methods Service

Clinical Trial Design Labs

Dissemination and Implementation (D&I) Core

Science Communications



“Our Informatics team can help you collect and manage research data, develop databases, and identify study participants. We’ll find the best data collection solution for your study. To get started, please submit a request below.”

William Harvey, MD, MSc, FACR
Co-Director, Informatics and Tufts Medical Center CMIO

Overview

We participate in development of a robust institutional informatics infrastructure, enabling research teams to maintain their focus on scientific discovery and analyses rather than on data wrangling. Our infrastructure and support systems are dynamic, to keep pace with the changing and interdependent fields of health informatics, bioinformatics, statistics, and data science; expandable, to accommodate new data types and analytic methods; and scalable, to support efficient and methodologically rigorous multisite/institution research. These defining traits allow us to elucidate novel methods and operational principles, harmonize datasets, and create pipelines for data sharing and analytics.



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?







Meet Our Newest Members

Welcome to OHDSI! - Please introduce yourself

General


 This is the first time assarabbiya has posted — let's welcome them to our community!

 **assarabbiya** 13d

Hi all,
I'm Jae-Won and I'm trying to learn more about OHDSI CDM so we can transform our EHR data onto it. Any guidance on where to start would be wonderful!
I'm currently watching the recording of OHDSI2023 Tutorial here: <https://www.youtube.com/watch?v=7-T9a811uN8>

1 Reply     Reply

 This is the first time Lui has posted — let's welcome them to our community!

 **Lui** Luisa Martinez 5d

Hello everyone! My name is Luisa Martinez and I'm a data scientist with clinical background. I have just started working as a Postdoctoral scientist at J&J Innovative Medicine department. I will be involved in OMOP related projects so I am righth now taking some of the OHDSI courses and reviewing the resources. I am looking forward to join the community and learn from one another!! Hope to be in touch!

    Reply

 **MPhilofsky** Melanie Philofsky   **assarabbiya** 20h

Hello [@assarabbiya](#) and welcome to OHDSI!

I highly suggest you start with [EHDEN Academy](#). The first course you should take is OMOP CDM and Standardized Vocabularies. Make sure you have a thorough understanding of the content before moving on to the Infrastructure; Extract, Transform and Load; Introduction to Usagi Code Mappings and for an ETL.

The Health Systems Interest Group on [MS Teams](#) assists health systems and you can also post specific questions to the [CDM Builders](#) group on the forums.




Meet Our Newest Members

Welcome to OHDSI! - Please introduce yourself


General

This is the first time assarabbiya has posted — let's welcome them to our community!


 **assarabbiya** 13d
Hi all,
I'm Jae-Won and I'm trying to learn more about OHDSI CDM so we can transform our EHR data onto it. Any guidance on where to start would be wonderful!
I'm currently watching the recording of OHDSI2023 Tutorial here: <https://www.youtube.com/watch?v=7-T9a811uN8>

1 Reply     Reply

This is the first time Lui has posted — let's welcome them to our community!

 **Lui** Luisa Martinez 5d
Hello everyone! My name is Luisa Martinez and I'm a data scientist with clinical background. I have just started working as a Postdoctoral scientist at J&J Innovative Medicine department. I will be involved in OMOP related projects so I am righth now taking some of the OHDSI courses and reviewing the resources. I am looking forward to join the community and learn from one another!! Hope to be in touch!

    Reply

 **MPhilofsky** Melanie Philofsky  **assarabbiya** 20h

Hello [@assarabbiya](#) and welcome to OHDSI!

I highly suggest you start with [EHDEN Academy](#). The first course you should take is OMOP CDM and Standardized Vocabularies. Make sure you have a thorough understanding of the content before moving on to the Infrastructure; Extract, Transform and Load; Introduction to Usagi Code Mappings and for an ETL.

The Health Systems Interest Group on [MS Teams](#) assists health systems and you can also post specific questions to the [CDM Builders](#) group on the forums.

OHDSI Workgroups

OHDSI's central mission is to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care. We work towards that goal in the areas of data standards, methodological research, open-source analytics development, and clinical applications.

Our workgroups present opportunities for all community members to find a home for their talents and passions, and make meaningful contributions. We are always looking for new collaborators. Learn more about these workgroups by checking out this page.

See an area where you want to contribute? Please Join The Journey!

Join A Workgroup

Workgroup Meeting Schedule

Get to Know the OHDSI Workgroups

[Africa Chapter](#)

[APAC](#)

[ATLAS/WebAPI](#)

[Clinical Trials](#)

[Common Data Model](#)

[CDM Survey Subgroup](#)

[CDM Vocabulary Subgroup](#)

[Dentistry](#)

[Early-Stage Researchers](#)

[Electronic Animal Health Records](#)

[Eye Care & Vision Research](#)

[FHIR and OMOP](#)

[Generative AI & Analytics in Healthcare \(GAIA\)](#)

[GIS – Geographic Information System](#)

[HADES](#)

[Health Equity](#)

[Healthcare Systems](#)

[Industry](#)

[Latin America](#)

[Medical Devices](#)

[Medical Imaging](#)

[Methods Research](#)

[Natural Language Processing](#)

[Network Data Quality](#)

[Oncology](#)

[Open-Source Community](#)

[Patient-Level Prediction](#)

[Perinatal and Reproductive Health](#)

[Phenotype Development & Evaluation](#)

[Psychiatry](#)

[Rehabilitation](#)

[Registry](#)

[Steering Group](#)

[Surgery and Perioperative Medicine](#)

[Themis](#)

[Vaccine Vocabulary](#)



Meet Our Newest Members

Thejas Bharadwaj

Ashlin Harris

Richard Paskach

Jo Yeleswarapu

Varsha Borhade

Mike Enger

Hozefa A. Divan

Shavawn Morgan

Mary Regan

Ahmed Anas Awad

Fares Alahdab

Esmond Urwin

Ivy Cerelia Valerie

Ann-Marie Jankowski

Mostafa Samy

Ondrej Klempir



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