OHDSI/OMOP Research Spotlight

OHDSI Community Call
April 1, 2025 • 11 am ET

Upcoming Community Calls

Date	Topic
Apr. 1	Recent OHDSI Publications
Apr. 8	Strategus Update & Review
Apr. 15	Treatment Patterns
Apr. 22	Current Practices in Estimation and Prediction
Apr. 29	DevCon 2025 Review
May 6	Evidence Synthesis
May 13	Maternal Health Fellowship Review
May 20	Guideline-Driven Evidence Study Review
May 27	Collaborator Showcase Brainstorm (Deadline is July 1)

Three Stages of The Journey

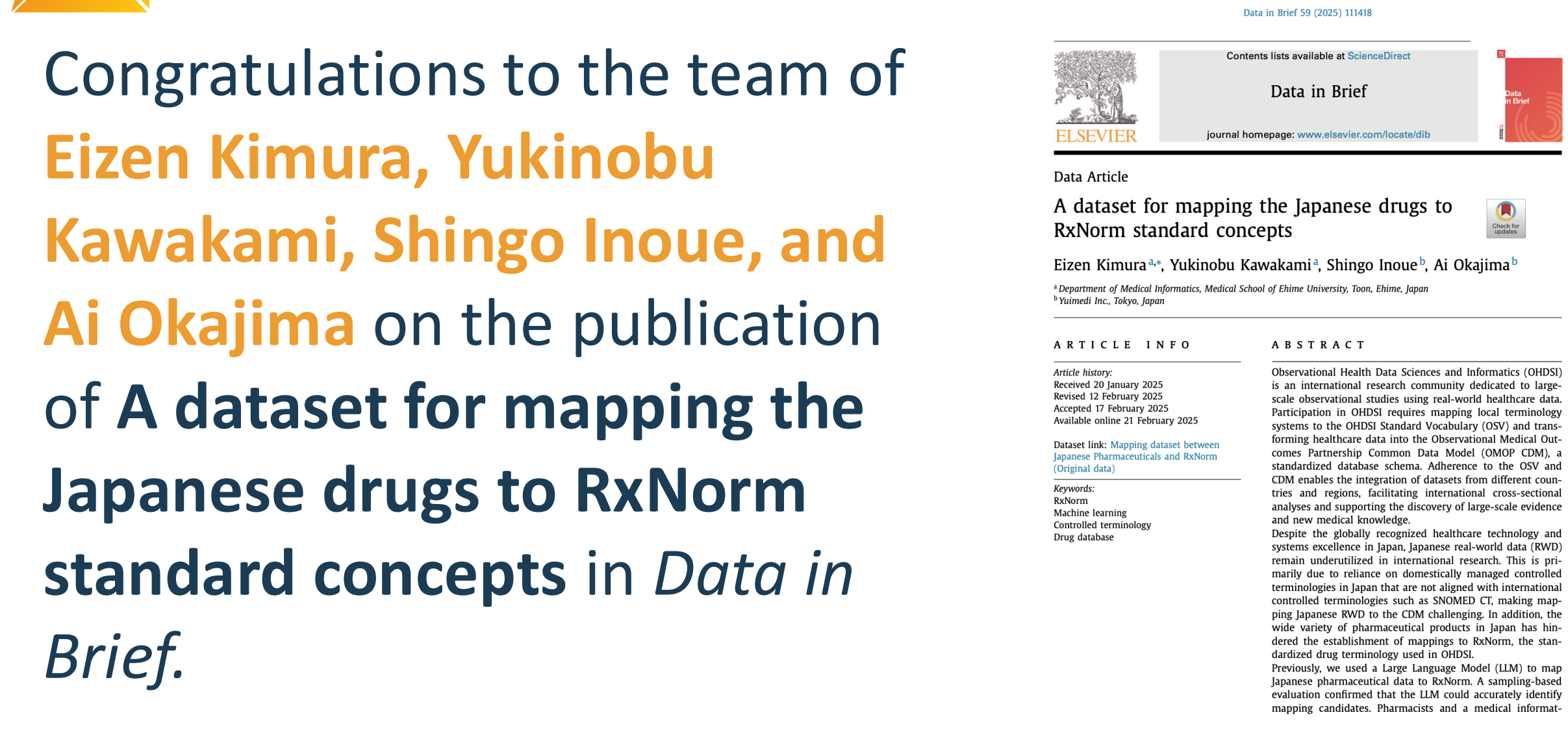
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Eizen Kimura, Yukinobu Kawakami, Shingo Inoue, and Ai Okajima** on the publication of **A dataset for mapping the Japanese drugs to RxNorm standard concepts** in *Data in Brief*.



Data in Brief 59 (2025) 111418

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A dataset for mapping the Japanese drugs to RxNorm standard concepts

Eizen Kimura[a,*], Yukinobu Kawakami[a], Shingo Inoue[b], Ai Okajima[b]

[a] Department of Medical Informatics, Medical School of Ehime University, Toon, Ehime, Japan
[b] Yuimedi Inc., Tokyo, Japan

ARTICLE INFO

Article history:
Received 20 January 2025
Revised 12 February 2025
Accepted 17 February 2025
Available online 21 February 2025

Dataset link: Mapping dataset between Japanese Pharmaceuticals and RxNorm (Original data)

Keywords:
RxNorm
Machine learning
Controlled terminology
Drug database

ABSTRACT

Observational Health Data Sciences and Informatics (OHDSI) is an international research community dedicated to large-scale observational studies using real-world healthcare data. Participation in OHDSI requires mapping local terminology systems to the OHDSI Standard Vocabulary (OSV) and transforming healthcare data into the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), a standardized database schema. Adherence to the OSV and CDM enables the integration of datasets from different countries and regions, facilitating international cross-sectional analyses and supporting the discovery of large-scale evidence and new medical knowledge.

Despite the globally recognized healthcare technology and systems excellence in Japan, Japanese real-world data (RWD) remain underutilized in international research. This is primarily due to reliance on domestically managed controlled terminologies in Japan that are not aligned with international controlled terminologies such as SNOMED CT, making mapping Japanese RWD to the CDM challenging. In addition, the wide variety of pharmaceutical products in Japan has hindered the establishment of mappings to RxNorm, the standardized drug terminology used in OHDSI.

Previously, we used a Large Language Model (LLM) to map Japanese pharmaceutical data to RxNorm. A sampling-based evaluation confirmed that the LLM could accurately identify mapping candidates. Pharmacists and a medical informat-

OHDSI Shoutouts!

Congratulations to the team of **Florian Katsch, Rada Hussein, Tanja Stamm, and Georg Duftschmid** on the publication of **Converting Health Level 7 Clinical Document Architecture (CDA) documents to Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) by leveraging CDA Template definitions** in *JAMIA Open*.

JAMIA Open, 2025, **8(2)**, ooaf022
https://doi.org/10.1093/jamiaopen/ooaf022
Research and Applications

Research and Applications

Converting Health Level 7 Clinical Document Architecture (CDA) documents to Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) by leveraging CDA Template definitions

Florian Katsch, MSc*[1,2], Rada Hussein, PhD[2], Tanja Stamm, PhD[1], Georg Duftschmid, PhD[1]

[1]Center for Medical Data Science, Medical University of Vienna, 1090 Vienna, Austria, [2]Ludwig Boltzmann Institute for Digital Health and Prevention, 5020 Salzburg, Austria

*Corresponding author: Florian Katsch, MSc, Ludwig Boltzmann Institute for Digital Health and Prevention, Lindhofstr. 22, 5020 Salzburg, Austria (florian.katsch@dhp.lbg.ac.at)

Abstract

Objectives: This work aims to develop a methodology for transforming Health Level 7 (HL7) Clinical Document Architecture (CDA) documents into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The described method seeks to improve the Extract, Transform, Load (ETL) design process by using HL7 CDA Template definitions and the CDA Refined Message Information Model (CDA R-MIM).

Material and Methods: Our approach utilizes HL7 CDA Templates to define structural and semantic mappings. Supported by the CDA R-MIM for semantic alignment with the OMOP CDM, we developed a tool named CDA Rabbit that enables the generation of Rabbit-In-a-Hat project files from HL7 CDA Template definitions and could be successfully integrated into the existing toolchain around OMOP.

Results: We tested our approach using 13 CDA Templates from the Austrian national EHR System (ELGA) and 430 anonymized CDA test documents that were mapped to 10 OMOP CDM tables. The data quality assessment, using OMOP's DataQualityDashboard, showed a 99% pass rate, indicating a robust and accurate data transformation.

Conclusion: This study presents a novel framework for transforming HL7 CDA documents into OMOP CDM using template definitions and CDA R-MIM. The methodology improves semantic interoperability, mapping reusability, and ETL design efficiency. Future work should focus on automating code generation, improving data profiling, and addressing cyclic dependencies within CDA templates. The presented approach supports improved secondary use of health data and research while adhering to standardized data models and semantics.

Discussion: Using CDA Templates for ETL design addresses common ETL challenges, such as data accessibility during ETL design, by decoupling the process from the actual CDA instances. Future work could focus on extending this approach to automatically generate boilerplate code, address cyclic dependencies within CDA Templates, and adapt the method for the use with FHIR profiles.

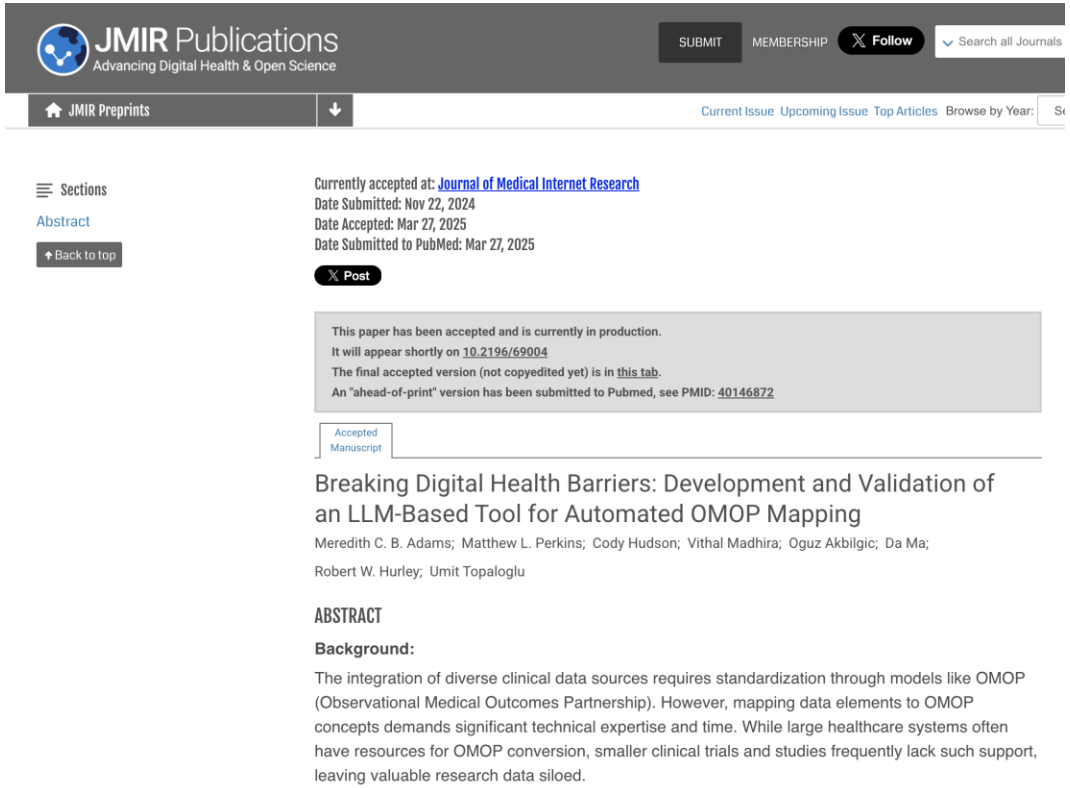
Lay Summary

This study describes a new method for converting electronic health records stored in Health Level 7 Clinical Document Architecture (CDA) format into the Observational Medical Outcomes Partnership Common Data Model, a widely used standard for health-care data analysis. The process uses CDA Templates, which define the structure of CDAs, to streamline data mapping and improve the accuracy of data transformation. A software tool has been developed to support this process. The proposed new method improves the reuse of health-care data for research. Future improvements will focus on automating the software generation process, refining data quality checks, and overcoming challenges related to the complex nature of CDA Templates.

Key words: OMOP; HL7 CDA; HL7 Templates; Health Information Interoperability.

OHDSI Shoutouts!

Congratulations to the team of **Meredith C B Adams, Matthew L Perkins, Cody Hudson, Vithal Madhira, Oguz Akbilgic, Da Ma, Robert W Hurley, and Umit Topaloglu** on the publication of **Breaking Digital Health Barriers: Development and Validation of an LLM-Based Tool for Automated OMOP Mapping** in the *Journal of Medical Internet Research.*



JMIR Publications
Advancing Digital Health & Open Science

SUBMIT MEMBERSHIP Follow Search all Journals

JMIR Preprints

Current Issue Upcoming Issue Top Articles Browse by Year:

Sections
Abstract
Back to top

Currently accepted at: Journal of Medical Internet Research
Date Submitted: Nov 22, 2024
Date Accepted: Mar 27, 2025
Date Submitted to PubMed: Mar 27, 2025

Post

This paper has been accepted and is currently in production.
It will appear shortly on 10.2196/69004
The final accepted version (not copyedited yet) is in this tab.
An "ahead-of-print" version has been submitted to Pubmed, see PMID: 40146872

Accepted Manuscript

Breaking Digital Health Barriers: Development and Validation of an LLM-Based Tool for Automated OMOP Mapping

Meredith C. B. Adams; Matthew L. Perkins; Cody Hudson; Vithal Madhira; Oguz Akbilgic; Da Ma; Robert W. Hurley; Umit Topaloglu

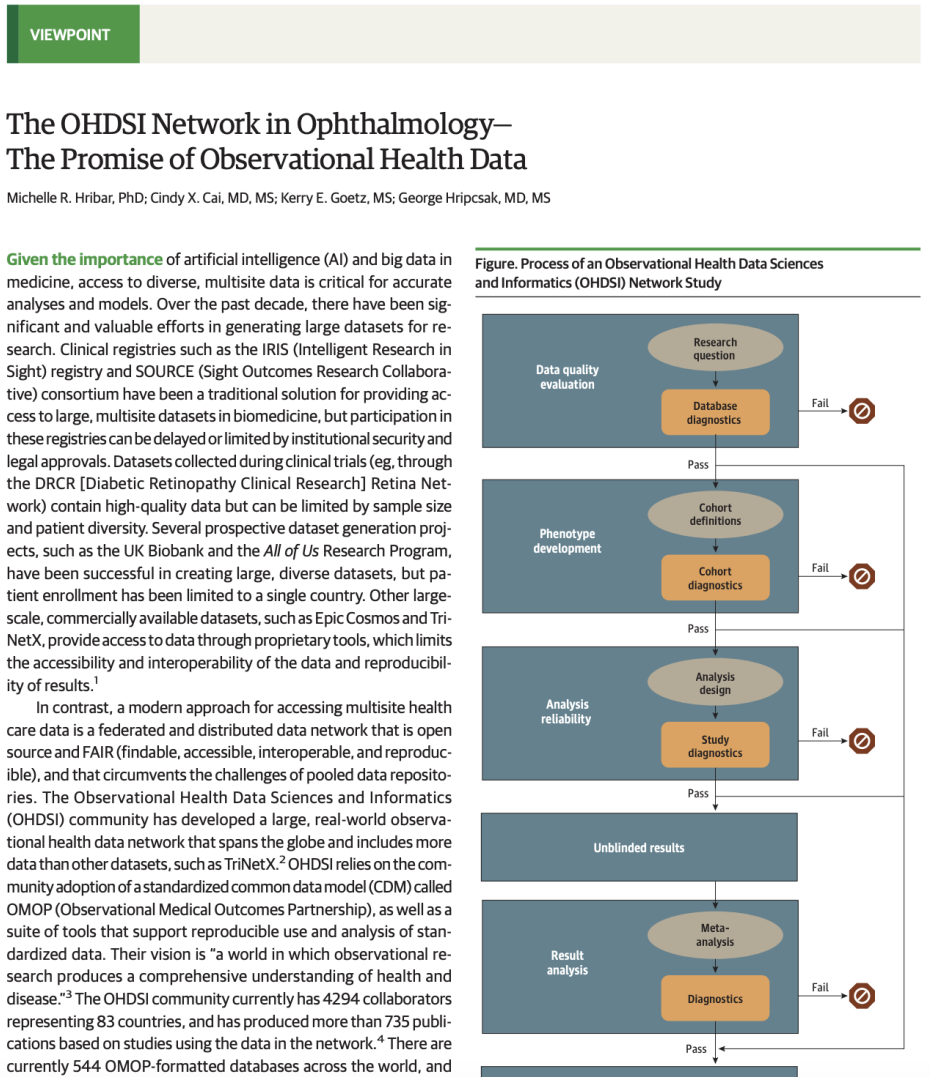
ABSTRACT

Background:

The integration of diverse clinical data sources requires standardization through models like OMOP (Observational Medical Outcomes Partnership). However, mapping data elements to OMOP concepts demands significant technical expertise and time. While large healthcare systems often have resources for OMOP conversion, smaller clinical trials and studies frequently lack such support, leaving valuable research data siloed.

OHDSI Shoutouts!

Congratulations to the team of **Michelle Hribar, Cindy X Cai, Kerry E. Goetz, and George Hripcsak** on the publication of **The OHDSI Network in Ophthalmology-The Promise of Observational Health Data** in *JAMA Ophthalmology.*



VIEWPOINT

The OHDSI Network in Ophthalmology— The Promise of Observational Health Data

Michelle R. Hribar, PhD; Cindy X. Cai, MD, MS; Kerry E. Goetz, MS; George Hripcsak, MD, MS

Given the importance of artificial intelligence (AI) and big data in medicine, access to diverse, multisite data is critical for accurate analyses and models. Over the past decade, there have been significant and valuable efforts in generating large datasets for research. Clinical registries such as the IRIS (Intelligent Research in Sight) registry and SOURCE (Sight Outcomes Research Collaborative) consortium have been a traditional solution for providing access to large, multisite datasets in biomedicine, but participation in these registries can be delayed or limited by institutional security and legal approvals. Datasets collected during clinical trials (eg, through the DRCR [Diabetic Retinopathy Clinical Research] Retina Network) contain high-quality data but can be limited by sample size and patient diversity. Several prospective dataset generation projects, such as the UK Biobank and the *All of Us* Research Program, have been successful in creating large, diverse datasets, but patient enrollment has been limited to a single country. Other large-scale, commercially available datasets, such as Epic Cosmos and TriNetX, provide access to data through proprietary tools, which limits the accessibility and interoperability of the data and reproducibility of results.[1]

In contrast, a modern approach for accessing multisite health care data is a federated and distributed data network that is open source and FAIR (findable, accessible, interoperable, and reproducible), and that circumvents the challenges of pooled data repositories. The Observational Health Data Sciences and Informatics (OHDSI) community has developed a large, real-world observational health data network that spans the globe and includes more data than other datasets, such as TriNetX.[2] OHDSI relies on the community adoption of a standardized common data model (CDM) called OMOP (Observational Medical Outcomes Partnership), as well as a suite of tools that support reproducible use and analysis of standardized data. Their vision is "a world in which observational research produces a comprehensive understanding of health and disease."[3] The OHDSI community currently has 4294 collaborators representing 83 countries, and has produced more than 735 publications based on studies using the data in the network.[4] There are currently 544 OMOP-formatted databases across the world, and

Figure. Process of an Observational Health Data Sciences and Informatics (OHDSI) Network Study

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Atlas
Wednesday	8 am	Psychiatry
Wednesday	7 pm	Medical Imaging
Thursday	11 am	Themis
Thursday	11 am	Industry
Thursday	12 pm	Methods Research
Thursday	1 pm	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	2 pm	Early-Stage Researchers
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS - Geographic Information System
Friday	10 am	Transplant
Friday	11:30 am	Steering
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Healthcare Systems Getting Started Subgroup
Tuesday	9 am	OMOP CDM Oncology Genomic Subgroup
Tuesday	9:30 am	Common Data Model

DevCon 2025: April 25

Agenda

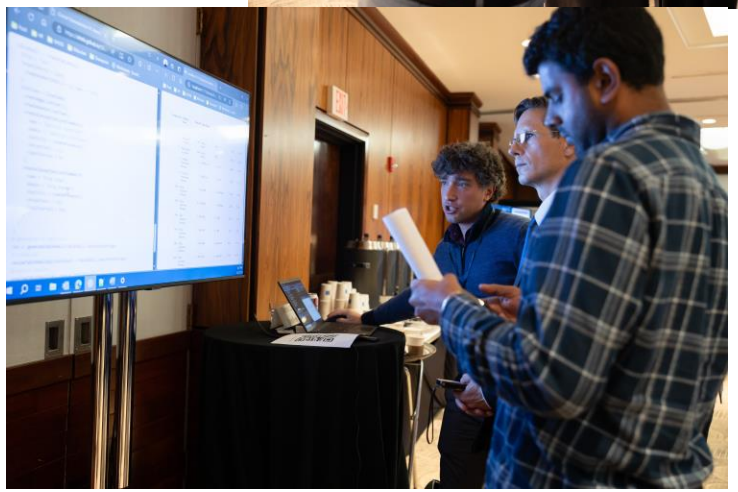
9:00 – 9:15am ET · Welcome & Introduction
- Paul Nagy, Johns Hopkins University

9:15 – 11:30am ET · OHDSI Projects Lightning Talks
- Stabilizing Gaia Core – Robert Miller, Miller Data Solutions
- CustomVocabularyBuilder – Jared Houghtaling, Tufts University
- CohortConstructor – Núria Mercadé-Besora, University of Oxford
- Updates on Strategus – Anthony Sena, Johnson & Johnson
- Experiences with SQLMesh/CICD integration with Databricks – Vishnu Chandrabalan, Lancashire Teaching Hospitals NHS Foundation Trust
- Updates from the Technical Advisory Board – Frank Defalco, Johnson & Johnson

11:30 – 12:30pm ET · Developer dialogue: Dev ops, DBT and, of course, LLMs
Moderator: Katy Sadowski, Boehringer Ingelheim
- Eduard Korchmar, EPAM Systems
- Egill Fridgeirsson, Erasmus MC
- Martin Lavallee, Boehringer Ingelheim
- Lawrence Adams, Artificial Intelligence Centre for Value Based Healthcare

12:30 – 1:00pm ET · Break

1:00 – 2:00pm ET · Sustainable Open-Source Ecosystems Panel
Moderator: Paul Nagy, Sean O'Reilly
- Data4Life – Peter Hoffmann
- The Hyve – Jan Blom/Wouter Franke
- Cognome – James Green

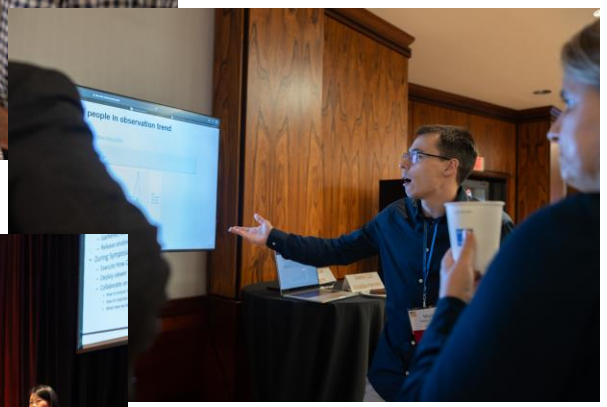


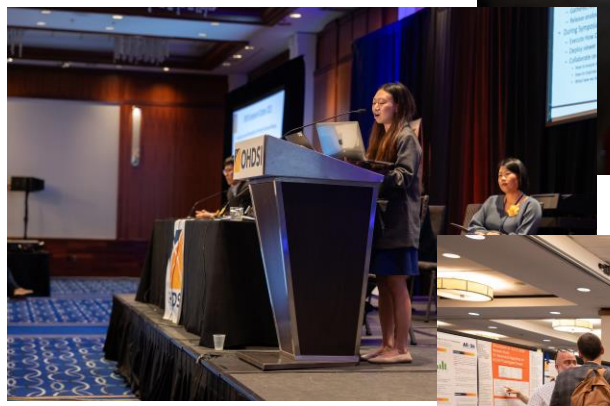
ohdsi.org/DevCon2025

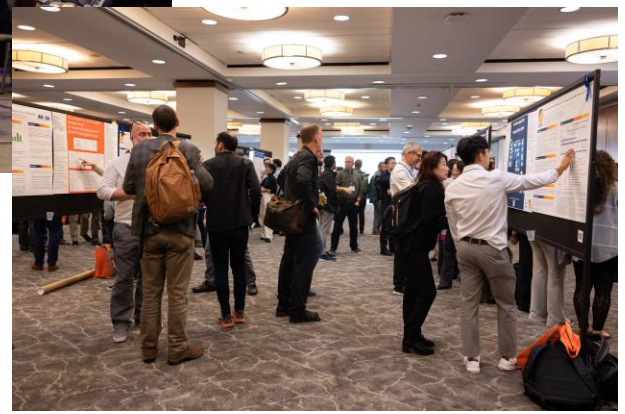
Save The Date!

The submission deadline for the 2025 Global Symposium Collaborator Showcase is July 1.

The showcase will be accepting both posters and software demos, as well as interest in hosting lightning talks. More information on the symposium, including abstract submission and registration links, will be available soon.







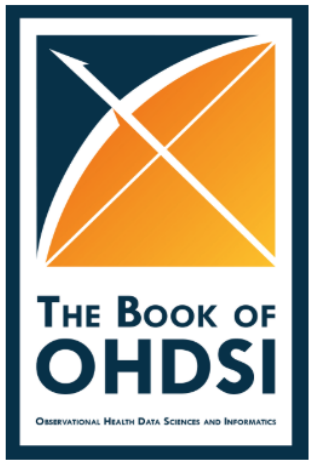
Japanese Book of OHDSI is Available

OHDSIの本

Observational Health Data Sciences and Informatics (OHDSI)

2025-03-15

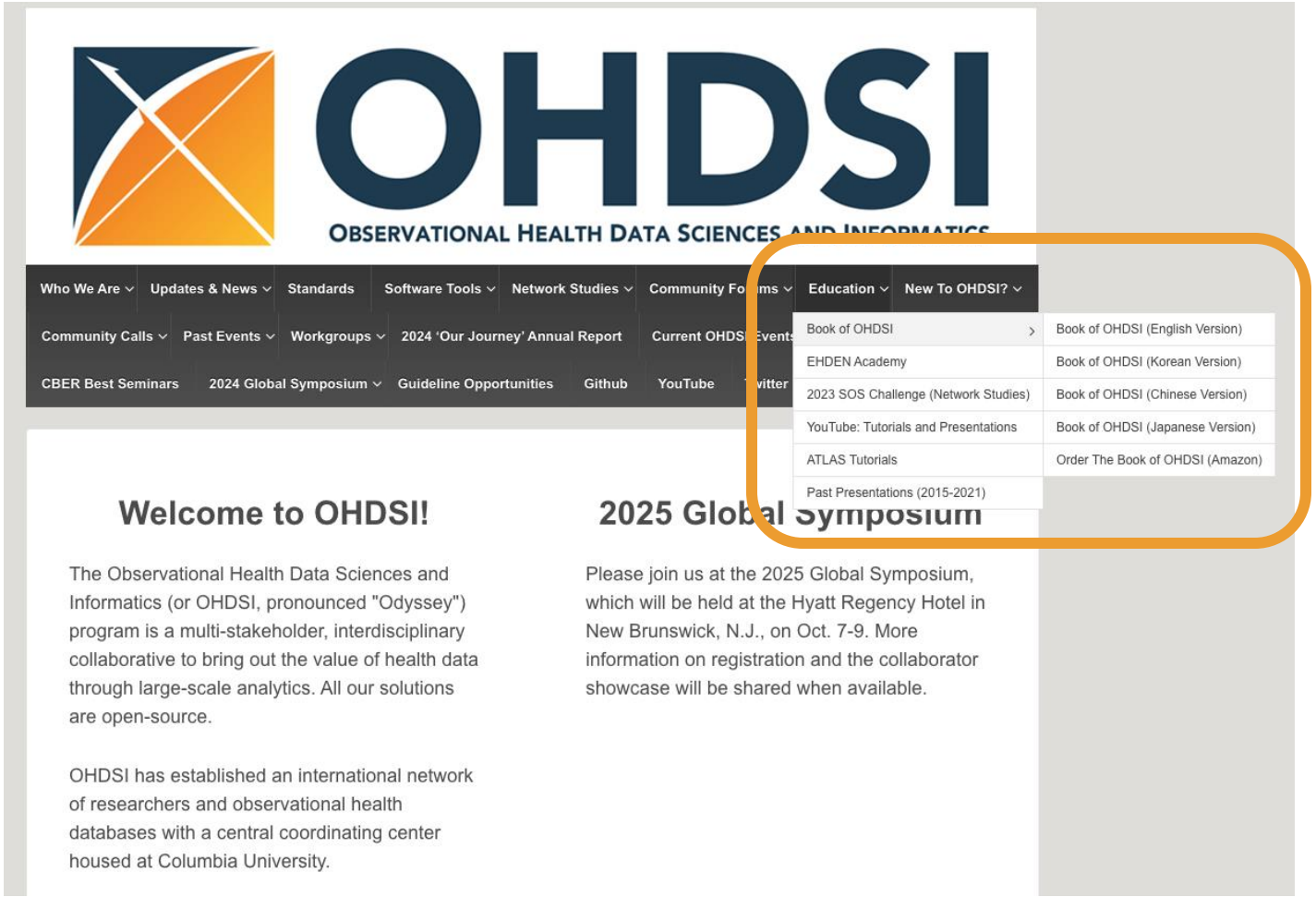
序章



これは、OHDSI コラボレーションについての本です。この本は、OHDSIコミュニティにより作成され、OHDSIに関するすべての知識の中心的なリポジトリとして役立つことを目指しています。この本はオープンソース開発ツールを通じてコミュニティにより維持される生きた文書であり、絶えず進化しています。オンライン版は無料で http://book.ohdsi.org から利用でき、常に最新バージョンを表示します。物理的なコピー（訳者注：英語版）はAmazon で原価価格で入手可能です。

この本の目標

この本は、OHDSIの中心的な知識リポジトリとなることを目的としており、OHDSIコミュニティ、OHDSI データ標準、およびOHDSI ツールについて説明します。本書は、OHDSIの初心者とベテランの両方を対象としており、必要な理論とそれに続く手順を提供する実用的な内容を目指しています。本書を読んだ後には、OHDSIが何であり、どのようにしてその旅に参加できるかを理解できます。共通データモデルおよび標準ボキャブラリとは何か、また、それらが観察医療データベースを標準化のためにどのように使用されるかを学びます。これらのデータの主な使用例である特性評価、集団レベルの推定、および患者レベルの予測について学びます。これらの3つすべての活動をサポートするOHDSIのオープンソースツールとその使用方法についても読みます。データ品質、臨床的妥当性、ソフトウエアの妥当性、方法の妥当性に関する章は、生成されたエビ



April OHDSI Newsletter

OHDSI

The Journey Newsletter (April 2025)

Standardized vocabularies and data fitness were core themes for OHDSI in March. Our vocabulary team provided an update on a major refresh, and we hosted multiple sessions on ensuring data quality for observational research. Learn more about both, as well as the growing Evidence Network, the developing Book of OHDSI 2.0, and plenty more in our latest newsletter. #JoinTheJourney

Podcast: Vocabularies and Data Fitness

Vocabularies & data fitness have been hot topics in the community over the last month. In the latest On The Journey podcast, Patrick Ryan and Craig Sachson discuss the recent refresh of the OHDSI Standardized Vocabularies, and they highlight the importance of downloading the updated version. They also discuss recent updates on the Evidence Network and how the Data Diagnostics tool impacts community research. *(If image does not appear, please click view this email in your browser)*

Community Updates

Where Have We Been?

- The Winter 2025 vocabulary refresh was recently released and includes several domain changes, newly added concepts, concept changes and more. Learn more about this and check out the presentation later in this newsletter.
- Our Phenotype Phebruary team, which included over 40 collaborators, built 165 cohort definitions and wrote the clinical descriptions for 13 studies. The team produced three Atlas and CohortDiagnostics demos and nearly completed the phenotyping for 85% of our clinical guideline studies. Thank you to everybody involved; check out the full final update.

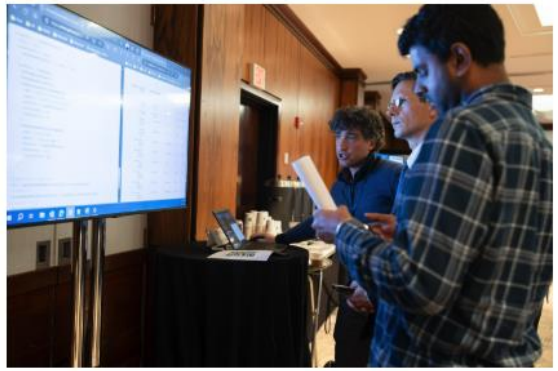
Where Are We Now?

- **Christian Reich** and **Sarah Seager** are leading an effort to publish a second edition of the Book of OHDSI, which will include updates to previous text and new paragraphs/chapters. This work is taking place within the Education workgroup; if you would like to join this effort, please sign up here.
- April community calls will feature several useful tutorials for our ongoing guideline-driven network studies, starting with updates on Strategus, an R package for coordinating and executing analytics using HADES modules, on April 8. Please visit our community calls page for a schedule and all past recordings.
- Our fourth annual DevCon will be held Friday, April 25, from 9 am - 2 pm ET. This virtual session continues to connect our global open-source community so that we can learn about recent updates and discuss ways to continue enhancing the future of OHDSI open-source software. More agenda information and the meeting link is available later in this newsletter.

Where Are We Going?

- The 2025 Global Symposium will be held Oct. 7-9 at the Hyatt Regency Hotel in New Brunswick, NJ, USA. Agenda and registration information will be shared when available. **The submission deadline to take part in the Global Symposium Collaborator Showcase is July 1.**
- Registration is open for the Europe Symposium, which will be held July 5-7 in the "Old Prison" building of Hasselt University in Hasselt, Belgium. More information is available later in this newsletter.
- The #OHDSISocialShowcase continues this month on our LinkedIn, Twitter/X and Instagram feeds. Please make sure you are following OHDSI on our social channels to receive daily updates on the research presented by our community.

DevCon to Spotlight Open-Source Innovation, Sustainability, and the Future



The OHDSI community will host DevCon 2025, the fourth annual gathering dedicated to advancing open-source development and collaboration, on April 25 within our Teams environment (9 am – 2 pm ET). This event will bring together developers and innovators to explore the latest tools, technologies, and strategies shaping the future of open-source software in healthcare and data science. The event kicks off with an exciting series of talks showcasing cutting-edge OHDSI projects, including updates on core infrastructure, cohort construction, and novel integrations with modern data platforms.

Engage in a dynamic developer dialogue on key topics such as DevOps, DBT, and the growing role of large language models in open-source development. This interactive session will provide insights from industry leaders on emerging trends, challenges, and opportunities in the evolving open-source landscape.

The day concludes with a panel on building sustainable open-source ecosystems, where experts will share their experiences in fostering long-term collaboration, innovation, and community-driven development. As the open-source movement continues to grow, understanding sustainable models is more critical than ever. Don't miss this opportunity to connect with fellow developers, exchange ideas, and contribute to the future of open-source technology at OHDSI DevCon 2025. Please use the links below to learn more and join the session.

DevCon 2025 Agenda

Join DevCon 2025 (April 25, 9 am - 2 pm ET)

March Publications

Cohen I, Diao Z, Goyal P, Gupta A, Hawk K, Malcom B, Malicki C, Sharma D, Sweeney B, Weiner SG, Venkatesh A, Taylor RA. Mapping Emergency Medicine Data to the Observational Medical Outcomes Partnership Common Data Model: A Gap Analysis of the American College of Emergency Physicians Clinical Emergency Data Registry. J Am Coll Emerg Physicians Open. 2025 Jan 10;6(1):100016. doi: 10.1016/j.acepjo.2024.100016. PMID: 40012646; PMCID: PMC11853007.

Chai Y, Lam ICH, Man KKC, Hayes JF, Wan EYF, Li X, Chui CSL, Lau WCY, Lin X, Yin C, Fan M, Chan EW, Wong ICK, Luo H. Psychiatric and neuropsychiatric sequelae of COVID-19 within 2 years: a multinational cohort study. BMC Med. 2025 Mar 7;23(1):144. doi: 10.1186/s12916-025-03952-z. PMID: 40055683; PMCID: PMC11887073.

Spotnitz M, Giannini J, Ostchega Y, Goff SL, Anandan LP, Clark E, Litwin TR, Berman L. Assessing the Data Quality Dimensions of Partial and Complete Mastectomy Cohorts in the *All of Us* Research Program: Cross-Sectional Study. JMIR Cancer. 2025 Mar 11;11:e59298. doi: 10.2196/59298. PMID: 40068169; PMCID: PMC11918980.

Dhaenens BAE, Moinat M, Didden EM, Ammour N, Oostenbrink R, Rijnbeek P. Identifying patients with neurofibromatosis type 1 related optic pathway glioma using the OMOP CDM. Eur J Med Genet. 2025 Mar 17;75:105011. doi: 10.1016/j.ejmg.2025.105011. Epub ahead of print. PMID: 40107446.

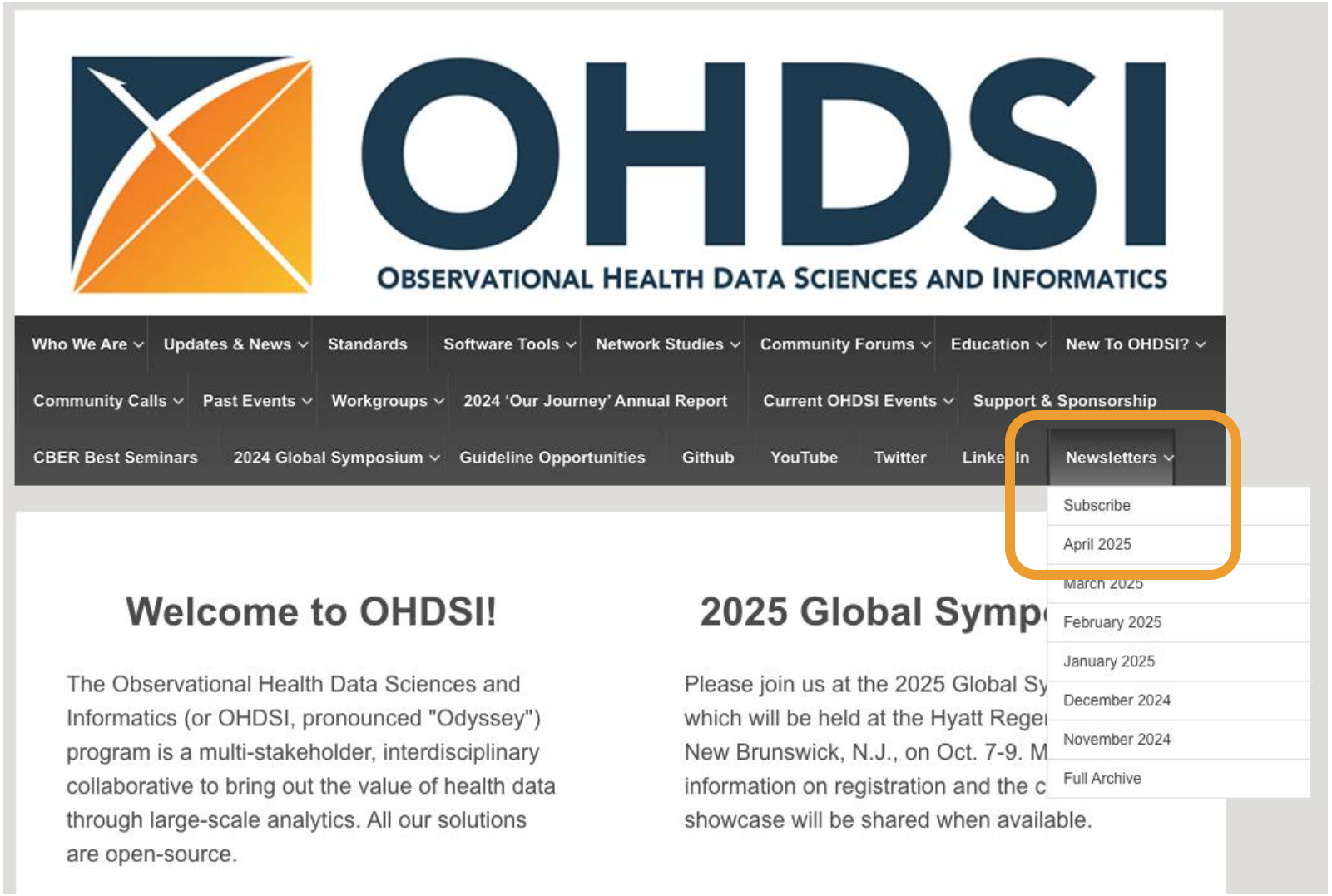
Lai LY, Sakalis V, Chatzichristos C, de la Parra I, Steinbeisser C, Golozar A, de Meulder B, Hijazy A, Snijder R, Feng Q, Falconer T, Cornford P, Bjartell A, Evans-Axelsson S, N'Dow J, Gandaglia G, Rivas JG. Baseline characteristics and clinical outcomes of prostate cancer patients on delayed palliative management: a PIONEER analysis based on big data. Cent European J Urol. 2024;77(3):403-410. doi: 10.5173/ceju.2024.83. Epub 2024 Aug 18. PMID: 40115485; PMCID: PMC11921962.

Delange B, Popoff B, Séité T, Lamer A, Parrot A. LinkR: An open source, low-code and collaborative data science platform for healthcare data analysis and visualization. Int J Med Inform. 2025 Mar 10;199:105876. doi: 10.1016/j.ijmedinf.2025.105876. Epub ahead of print. PMID: 40121766.

Kimura E, Kawakami Y, Inoue S, Okajima A. A dataset for mapping the Japanese drugs to RxNorm standard concepts. Data Brief. 2025 Feb 21;59:111418. doi: 10.1016/j.dib.2025.111418. PMID: 40124300; PMCID: PMC11926709.

Katsch F, Hussein R, Stamm T, Duftschmid G. Converting Health Level 7 Clinical Document Architecture (CDA) documents to Observational Medical

April OHDSI Newsletter



Analysis April Begins

2025 OHDSI Focus Areas

Guideline-driven Evidence Generation	Evidence-driven Data standardization	Evidence-driven Open Source Development	Evidence-driven Collaborative Education
Dry January: Guideline review to determine evidence needs where RWE could potentially contribute	**Phenotype Phebruary:** Develop/evaluate cohorts needed to support filling the evidence gaps	**March to Data Fitness:** Evidence network to determine which partners are appropriate to generate which evidence	
Analysis April: Prepare protocol and analysis specification to initiate network execution	**Meta-analysis May:** Collaborative interpretation of results from across network	**Journey to June:** Mid-year reflection on evidence generation process and progress	
Spread-the-Word Second Half: Focus on Evidence Dissemination			
July: OHDSI EU	August:	September:	
October: OHDSI Global (tbd)	November:	December: OHDSI APAC	

#OHDSISocialShowcase This Week

Monday

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)

Enhancing Infectious Disease Data Integration and management through OMOP-CDM in South Korea

Min Ho An, MD[1,2*], Seok Kim, M.P.H[3*], ByungJin Choi, MD[1,2], Sooyoung Yoo, Ph.D[3], Rae Woong Park, MD, Ph.D[1] , Ji Seon Oh, MD, Ph.D[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]Office of eHealth Research and Business, Seoul National University Bundang Hospital, Republic of Korea
[4]Information Medicine, Big Data Research Center, Asan Medical Center, Seoul, Republic of Korea
*These authors are equally contributed to this work

Background

- The Platform for Harmonizing and Accessing Data in Real-time on Infectious Disease Surveillance (PHAROS) was initiated to address challenges in data integration and management.
- PHAROS focuses on developing an integrated infectious disease data management system based on the OMOP-CDM in Korea, with the goal of enhancing real-time clinical information collection and improving treatment and disease management strategies.
- To support this, data encompassing microbial test results, infectious disease consultation notes, vaccination-related information, emergency room data, and legal infectious disease reports, were utilized, aimed at improving accessibility and ensuring clear representation of information.
- The codes within infectious disease consultation notes, vaccination-related information, emergency room data, and legal infectious disease reports are newly mapped and integrated as CDM records.
- Moreover, to address challenge of identifying detailed culture information, we developed new Extract Transform Load (ETL) method that suits to specifically store data drawn from specimen culture.
- While this model maintains the relationship between microbial tests and drug resistance, it captures various aspects of culture information without requiring additional data tables, thus improving the comprehensiveness and utility of information from specimen culture.

Methods

- In this study, OMOP-CDM was utilized to include infection-related clinical data. We used CDM version 5.4 without any additional columns. Infectious disease department consultation notes are integrated into the CDM's Note domain using specific concept ids, with consultation request recorded in the observation table.
- Additionally, vaccination-related reports are thoroughly documented in the drug domain, with dose information recorded in the observation table for detailed tracking.
- Primary symptom information from the National Emergency Department Information System (NEDIS) system is integrated by mapping chief complaints to SNOMED-CT and inserting them into the condition table or the observation table if no suitable mapping exists.
- We also utilized patient travel history from legal communicable disease reports. Particularly, Microbial test results were stored across three tables: specimens were stored in the specimen table, cultured microorganisms and antibiotic susceptibility results were stored in the measurement table, and the type of microorganism identified were stored in the observation table.
- These tables were designed to be linked using connection keys, facilitating the proper extraction of necessary data for various purposes.

Conclusions

- This study addresses infectious disease data integration challenges using the OMOP-CDM framework, standardizing clinical data for better accessibility and comprehensiveness. The new ETL method stores detailed culture information without extra tables, preserving key relationships between microbial tests and drug resistance. This approach may enhance research, supports rapid outbreak response, and improves disease management

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)
- 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: minho.an23@gmail.com

Results

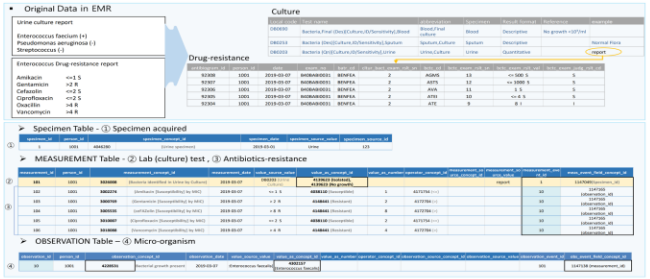


Figure 1. Culture Modeling Table Schema

Type	Number of Patients	Number of Data	Detailed Data Items
Person	239,310	239,310	De-identified ID, gender, birthdate, etc.
Visit Occurrence	239,108	5,928,625	Visit start/end time, visit type (outpatient, inpatient, etc.)
Condition Occurrence	238,859	84,261,109	Diagnosis code, diagnosis date
Drug Exposure	238,769	9,565,547	Drug code, prescription date, drug quantity, etc.
Procedure	238,707	159,246,096	Procedure code, procedure date
Measurement	227,579	58,070,197	Measurement code, result, unit (continuous, categorical, text, etc.)
Device	224,278	9,867,621	Medical device code, order date, amount
Death	222,246	14,820,106	Death date, cause of death
Observation	201,959	78,984,878	Other clinical information, observation date
Specimen	6,211	6,211	Specimen code, collection date, quantity, unit

Table 1. Converted Data Summery in Ajou University Hospital

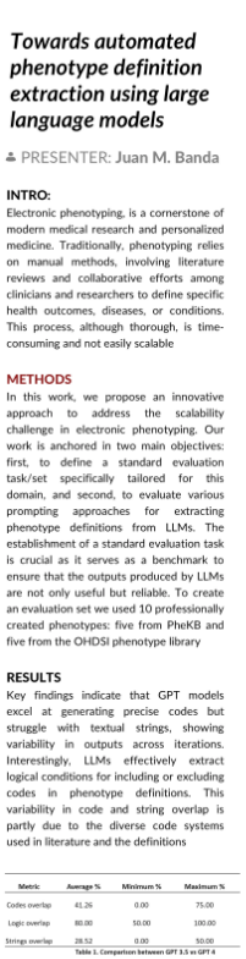
- A total of 560 codes for infection types, testing procedures, antimicrobial sensitivity, and travel history were mapped. Additionally, the National Emergency Department Information System (NEDIS) was mapped to include 1,114 codes for major symptoms and issues.
- A total of 2,226 codes were mapped for legal infectious diseases. Furthermore, how infection-specific data such as microbial tests and antibiotic susceptibility results are stored in the CDM is illustrated in Figure 1.
- The information for specimen acquisition is recorded in the specimen table with the corresponding specimen concept ID (①).
- The results of laboratory (culture) tests are documented in the measurement table in "value_as_concept_id," indicating the existence of microorganisms by "isolated" or "no growth" (4139623) and linked to the specimen table through the "measurement_event_id" and "meas_event_field_concept_id" to trace the source (②).
- Additionally, antibiotic susceptibility data (③) is loaded into the measurement table. The differentiation from laboratory (culture) tests is achieved by using "meas_event_field_concept_id" with the related field as "observation_id".
- Lastly, the type of identified microorganism is recorded in the observation table (④), with the presence identified by observation_concept_id, and the name of the microorganism designated in "value_as_concept_id". This data is linked through the field "observation_id" matched with "measurement_event_id" in the measurement table.

#OHDSISocialShowcase This Week

Tuesday

Towards automated phenotype definition extraction using large language models

(Ramya Tekumalla, Juan M. Banda)



Towards automated phenotype definition extraction using large language models

PRESENTER: Juan M. Banda

INTRO:
Electronic phenotyping, is a cornerstone of modern medical research and personalized medicine. Traditionally, phenotyping relies on manual methods, involving literature reviews and collaborative efforts among clinicians and researchers to define specific health outcomes, diseases, or conditions. This process, although thorough, is time-consuming and not easily scalable

METHODS
In this work, we propose an innovative approach to address the scalability challenge in electronic phenotyping. Our work is anchored in two main objectives: first, to define a standard evaluation task/set specifically tailored for this domain, and second, to evaluate various prompting approaches for extracting phenotype definitions from LLMs. The establishment of a standard evaluation task is crucial as it serves as a benchmark to ensure that the outputs produced by LLMs are not only useful but reliable. To create an evaluation set we used 10 professionally created phenotypes: five from PheKB and five from the OHDSI phenotype library

RESULTS
Key findings indicate that GPT models excel at generating precise codes but struggle with textual strings, showing variability in outputs across iterations. Interestingly, LLMs effectively extract logical conditions for including or excluding codes in phenotype definitions. This variability in code and string overlap is partly due to the diverse code systems used in literature and the definitions

Metric	Average %	Minimum %	Maximum %
Codes overlap	41.26	0.00	75.00
Logic overlap	80.00	50.00	100.00
Strings overlap	28.52	0.00	50.00

Table 1. Comparison between GPT 3.5 vs GPT 4

While LLMs, currently, produce seemingly convincing definitions, they are highly inconsistent and inaccurate compared to human created definitions

However, there is promise in terms of augmenting the human-guide process, and with the creation of smaller domain specific models

Take a picture to download the full paper

Using Biomedical Content Explorer linked with PubDictionaries, ICD10, and ICD10-CM dictionaries, we compared GPT-3.5 and GPT-4 in matching phenotype codes. The results highlight the models' weaknesses, particularly their inaccuracies and hallucinations. These issues were more pronounced for less-documented phenotypes, underscoring the need for cautious use and meticulous verification of LLM-generated data.

Model	Metric	Average %	Minimum %	Maximum %
GPT 4	Codes overlap	50.94	20.00	88.89
	Logic overlap	90.00	50.00	100.00
	Strings overlap	48.59	0.00	100.00
GPT 3.5	Codes overlap	27.51	10.00	85.20
	Logic overlap	70.20	0.00	90.00
	Strings overlap	41.28	0.00	75.12

Table 2. Comparison between human definition vs GPT models

Figure 1. Comparisons of GPT hallucinations when producing codes

Conclusions:
Our exploration of LLMs for automating phenotype definition extraction highlights their potential to enhance scalability and efficiency in digital healthcare

While GPT-3.5 and GPT-4 show promise in generating medically relevant codes, challenges remain in achieving consistent textual output and avoiding inaccuracies

The study underscores the need for robust evaluation and validation frameworks to ensure LLM reliability

Despite hallucinations and inconsistencies, GPT models can serve as valuable initial steps or augmentation tools, significantly streamlining and improving electronic phenotyping methodologies

Ramya Tekumalla and Juan M. Banda

Stanford HEALTH CARE

OHDSI

#OHDSISocialShowcase This Week

Wednesday

Can we combine propensity score modeling and patient level prediction to make counterfactual predictions?

(Jenna Reps, Chris Knoll)

Can we combine propensity score modelling and patient-level prediction to make counter-factual predictions?

Presenter: Jenna Reps

BACKGROUND

- The OHDSI community have developed advanced frameworks for population level causal inference and patient-level prediction [1].
- However, due to confounding or other forms of bias commonly seen in observational data it is currently not possible to use the current OHDSI tools and large observational healthcare data for counterfactual prediction.
- If patients were randomly assigned treatment, it would be possible to developed prediction models that can personalize treatment
- Propensity score models are often used in population level causal inference to mimic randomization [2]. In this feasibility study we investigate using propensity score modelling to create a data set that is like a randomized clinical trial and then apply patient-level prediction .

METHODS

Data:
Optum's de-identified Clinformatics® Data Mart Database (Optum Clinformatics®) mapped to the OMOP CDM. Optum Clinformatics® is a large US insurance claims database.

Task:
We identified new users of lisinopril or amlodipine with >365 days prior observation, aged 18 or older with an exposure on or after 2020. We restricted to patients with a history of hypertension. The CohortMethod R package was used to create a propensity model to predict the probability of being given lisinopril. The prediction target population was created by trimming to a preference score between 0.3-0.7 and then applying 1-1 matching.

We then developed patient-level prediction models, using the R PatientLevelPrediction package [3], for the task: *which patients in the target population will have liver injury during treatment exposure*. Liver injury is listed as a known side effect of taking lisinopril.

Finally, we apply the model to compare the mean predicted risk of liver injury if 1) all patients had lisinopril (and no amlodipine) and 2) all patients had amlodipine (and no lisinopril). The ratio of this is the estimated population level effect from counterfactual prediction.

Analysis:
A gradient boosting machine model with treatment as a predictor as well as age/sex, conditions grouped using the hierarchy and drug ingredients was developed. The model development used 75% of the data for model training and 25% of the data for model testing. 3-fold cross validation was applied in the training data to learn the optimal hyper-parameters. A final model was fit using the optimal hyper-parameters and all the training data.

Metrics:
The model performance was accessed internally by comparing the predicted risk and the true risk in the testing data. Model performance was evaluated using the area under the receiver operating characteristic curve (AUROC), calibration-in-the-large and brier score.

We then calculated the predicted risk for all test patients if they had been given lisinopril and the predicted risk for all test patients if they had been given amlodipine. We divided the mean lisinopril risk by the mean amlodipine risk to estimate the population risk ratio. This was compared with the relative risk obtained from a propensity matched cohort method analysis.

We propose a simple approach to counterfactual prediction using observational data in certain situations

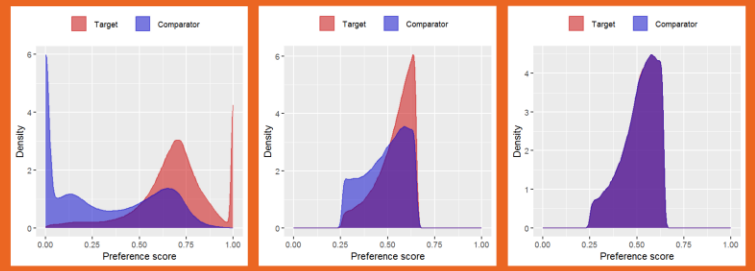


Figure 1 - Propensity score plots for the full population, trimmed population and trimmed then matched populations.

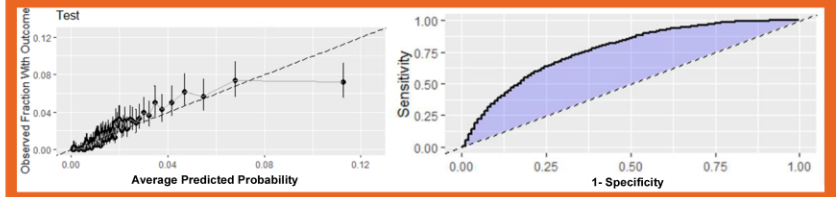


Figure 2 - Calibration and ROC plots for the test data.

Preliminary results show promise at being able to personalize treatment effects…

RESULTS

Data Size
We identified 978,893 new users of lisinopril or amlodipine with inclusion criteria. After trimming 398,943 remained (195,009 lisinopril and 203,934 amlodipine). After 1:1 matching 320,932 remained in our final target population (160,466 each). The propensity plots for the full, trimmed and trimmed than matched as shown in Figure 1.

10,272 patients were excluded due to having prior liver injury, leaving 310,660 patients. 4380 patients (1.4%) experienced liver injury while exposed to the drugs.

Prediction Model
The outcome model obtained an AUROC of 0.77 (0.76-0.78) in the test data and a calibration-in-the-large predicted risk of 0.01402 and observed risk of 0.01410. Figure 2 shows the discrimination and calibration plots.

Population-level effect estimates
Patient-level prediction: The mean predicted risk when all patients (test patients) were set to have lisinopril and not amlodipine was 0.01548191 (0.01540209) and the mean predicted risk when all patients (test patients) were set to have amlodipine and not lisinopril was 0.01267826 (0.0126168). This gives a population risk ratio of 1.22 (1.22 in test set patients only).

Cohort Method: The odds ratio estimate from cohort method after trimming and matching was 1.39 (1.31 - 1.48). When restricting to the same patients used in the prediction test set, the estimate was 1.16 (0.91 - 1.48).

CONCLUSIONS

- We were able to develop a good prediction model for liver injury in the new drug users.
- The estimates were similar: population effect estimate from the patient-level prediction model was 1.22 and the cohort method population level effect estimate was 1.16 (0.91-1.48).
- We proposed a simple methodology for causal prediction, but more tasks need to be evaluated to see whether the population level performance consistently appears to be close to the cohort method.
- The methodology is only applicable when a patient could have been given an alternative treatment. If a patient has a high propensity for one treatment, then there is likely a clinical reason for them to have that treatment and therefore personalizing their treatment is not suitable.
- If future work shows the counterfactual prediction works, then it can be fed into risk-benefit analyses to help patients pick their optimal treatment.

REFERENCES

1. Schuemie M., Reps J., Black A, DeSilvio F., Evans L., Fridgeirsson E., Gilbert I.P., Knoll C., Lavallee M., Rao GA., Rijnbeek P., Sadecwicz K., Sena A., Swerdel J., Williams RD., Suchard M. *Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research.* Stud Health Technol Inform. 2024 Jan 25;310:966-970. doi: 10.3233/SHTI231106.
2. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011 May;46(3):399-424. doi: 10.1080/00273171.2011.568786. Epub 2011 Jun 8. PMID: 21818162; PMCID: PMC3144483.
3. Jenna M Reps, Martijn J Schuemie, Marc A Suchard, Patrick B Ryan, Peter R Rijnbeek, Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data, *Journal of the American Medical Informatics Association*, Volume 25, Issue 8, August 2018, Pages 969–975, https://doi.org/10.1093/jamia/ocy032

Jenna Reps and Chris Knoll
Janssen Research and Development, Titusville, NJ, United States

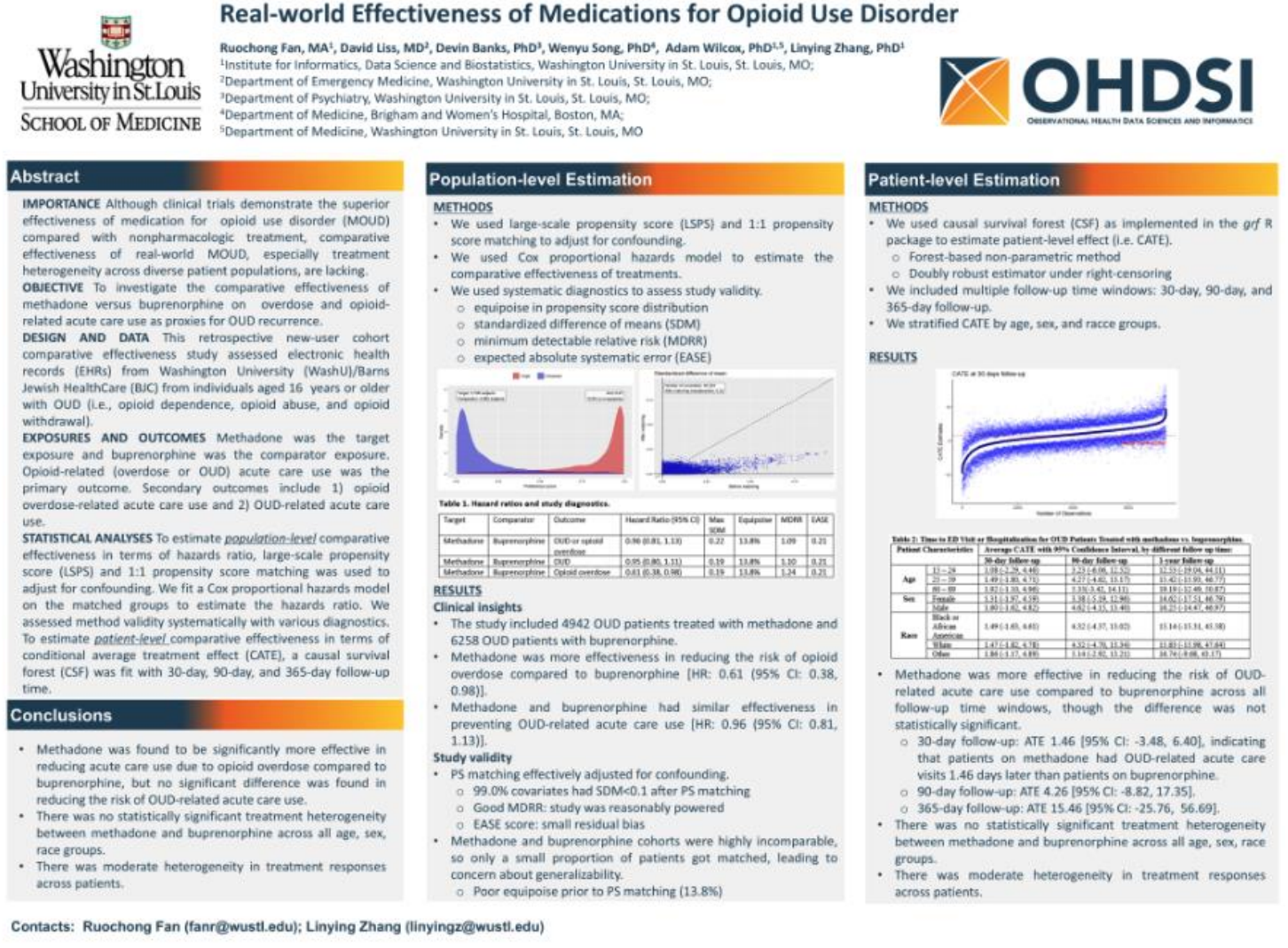
Johnson&Johnson
Global Epidemiology Organization

#OHDSISocialShowcase This Week

Thursday

Real-world Effectiveness of Medications for Opioid Use Disorder (RWE-MOUD)

(Ruochong Fan, David Liss, Devin Banks, Wenyu Song, Adam Wilcox, Linying Zhang)



Real-world Effectiveness of Medications for Opioid Use Disorder

Ruochong Fan, MA[1], David Liss, MD[2], Devin Banks, PhD[3], Wenyu Song, PhD[4], Adam Wilcox, PhD[1,5], Linying Zhang, PhD[1]
[1]Institute for Informatics, Data Science and Biostatistics, Washington University in St. Louis, St. Louis, MO;
[2]Department of Emergency Medicine, Washington University in St. Louis, St. Louis, MO;
[3]Department of Psychiatry, Washington University in St. Louis, St. Louis, MO;
[4]Department of Medicine, Brigham and Women's Hospital, Boston, MA;
[5]Department of Medicine, Washington University in St. Louis, St. Louis, MO

Washington University in St. Louis School of Medicine

OHDSI Observational Health Data Sciences and Informatics

Abstract

IMPORTANCE Although clinical trials demonstrate the superior effectiveness of medication for opioid use disorder (MOUD) compared with nonpharmacologic treatment, comparative effectiveness of real-world MOUD, especially treatment heterogeneity across diverse patient populations, are lacking.
OBJECTIVE To investigate the comparative effectiveness of methadone versus buprenorphine on overdose and opioid-related acute care use as proxies for OUD recurrence.
DESIGN AND DATA This retrospective new-user cohort comparative effectiveness study assessed electronic health records (EHRs) from Washington University (WashU)/Barns Jewish HealthCare (BJC) from individuals aged 16 years or older with OUD (i.e., opioid dependence, opioid abuse, and opioid withdrawal).
EXPOSURES AND OUTCOMES Methadone was the target exposure and buprenorphine was the comparator exposure. Opioid-related (overdose or OUD) acute care use was the primary outcome. Secondary outcomes include 1) opioid overdose-related acute care use and 2) OUD-related acute care use.
STATISTICAL ANALYSES To estimate *population-level* comparative effectiveness in terms of hazards ratio, large-scale propensity score (LSPS) and 1:1 propensity score matching was used to adjust for confounding. We fit a Cox proportional hazards model on the matched groups to estimate the hazards ratio. We assessed method validity systematically with various diagnostics. To estimate *patient-level* comparative effectiveness in terms of conditional average treatment effect (CATE), a causal survival forest (CSF) was fit with 30-day, 90-day, and 365-day follow-up time.

Conclusions

- Methadone was found to be significantly more effective in reducing acute care use due to opioid overdose compared to buprenorphine, but no significant difference was found in reducing the risk of OUD-related acute care use.
- There was no statistically significant treatment heterogeneity between methadone and buprenorphine across all age, sex, race groups.
- There was moderate heterogeneity in treatment responses across patients.

Population-level Estimation

METHODS

- We used large-scale propensity score (LSPS) and 1:1 propensity score matching to adjust for confounding.
- We used Cox proportional hazards model to estimate the comparative effectiveness of treatments.
- We used systematic diagnostics to assess study validity.
  - equipoise in propensity score distribution
  - standardized difference of means (SDM)
  - minimum detectable relative risk (MDRR)
  - expected absolute systematic error (EASE)

Table 1. Hazard ratios and study diagnostics.

Target	Comparator	Outcome	Hazard Ratio (95% CI)	Max SDM	Equipoise	MDRR	EASE
Methadone	Buprenorphine	OUD or opioid overdose	0.90 (0.81, 1.13)	0.22	13.8%	1.09	0.21
Methadone	Buprenorphine	OUD	0.95 (0.80, 1.31)	0.19	13.8%	1.10	0.21
Methadone	Buprenorphine	Opioid overdose	0.61 (0.38, 0.98)	0.19	13.8%	1.24	0.21

RESULTS

Clinical insights

- The study included 4942 OUD patients treated with methadone and 6258 OUD patients with buprenorphine.
- Methadone was more effectiveness in reducing the risk of opioid overdose compared to buprenorphine [HR: 0.61 (95% CI: 0.38, 0.98)].
- Methadone and buprenorphine had similar effectiveness in preventing OUD-related acute care use [HR: 0.96 (95% CI: 0.81, 1.13)].

Study validity

- PS matching effectively adjusted for confounding.
  - 99.0% covariates had SDM<0.1 after PS matching
  - Good MDRR: study was reasonably powered
  - EASE score: small residual bias
- Methadone and buprenorphine cohorts were highly incomparable, so only a small proportion of patients got matched, leading to concern about generalizability.
  - Poor equipoise prior to PS matching (13.8%)

Patient-level Estimation

METHODS

- We used causal survival forest (CSF) as implemented in the *grf* R package to estimate patient-level effect (i.e. CATE).
  - Forest-based non-parametric method
  - Doubly robust estimator under right-censoring
- We included multiple follow-up time windows: 30-day, 90-day, and 365-day follow-up.
- We stratified CATE by age, sex, and racce groups.

RESULTS

Patient Characteristics		Average CATE with 95% Confidence Interval, by different follow-up time:		
		30-day follow-up	90-day follow-up	1-year follow-up
Age	12 – 24	1.08 (-2.79, 4.44)	3.73 (-6.08, 12.52)	12.53 (-19.04, 44.11)
	25 – 39	1.49 (-1.80, 4.71)	4.27 (-4.82, 13.17)	13.40 (-17.93, 46.77)
	40 – 89	1.82 (-1.33, 4.98)	5.16 (-3.42, 14.11)	15.19 (-17.49, 50.87)
Sex	Female	1.31 (-1.97, 4.59)	3.58 (-5.59, 12.96)	14.62 (-17.51, 46.76)
	Male	1.80 (-1.62, 4.82)	4.62 (-4.15, 13.40)	16.21 (-14.47, 46.97)
Race	Black or African American	1.49 (-1.65, 4.61)	4.32 (-4.37, 13.02)	13.14 (-15.31, 41.58)
	White	1.47 (-1.82, 4.78)	4.52 (-4.76, 13.34)	13.83 (-13.96, 47.64)
	Other	1.84 (-1.17, 4.89)	5.14 (-2.82, 13.21)	16.74 (-9.68, 43.17)

- Methadone was more effective in reducing the risk of OUD-related acute care use compared to buprenorphine across all follow-up time windows, though the difference was not statistically significant.
  - 30-day follow-up: ATE 1.46 [95% CI: -3.48, 6.40], indicating that patients on methadone had OUD-related acute care visits 1.46 days later than patients on buprenorphine.
  - 90-day follow-up: ATE 4.26 [95% CI: -8.82, 17.35].
  - 365-day follow-up: ATE 15.46 [95% CI: -25.76, 56.69].
- There was no statistically significant treatment heterogeneity between methadone and buprenorphine across all age, sex, race groups.
- There was moderate heterogeneity in treatment responses across patients.

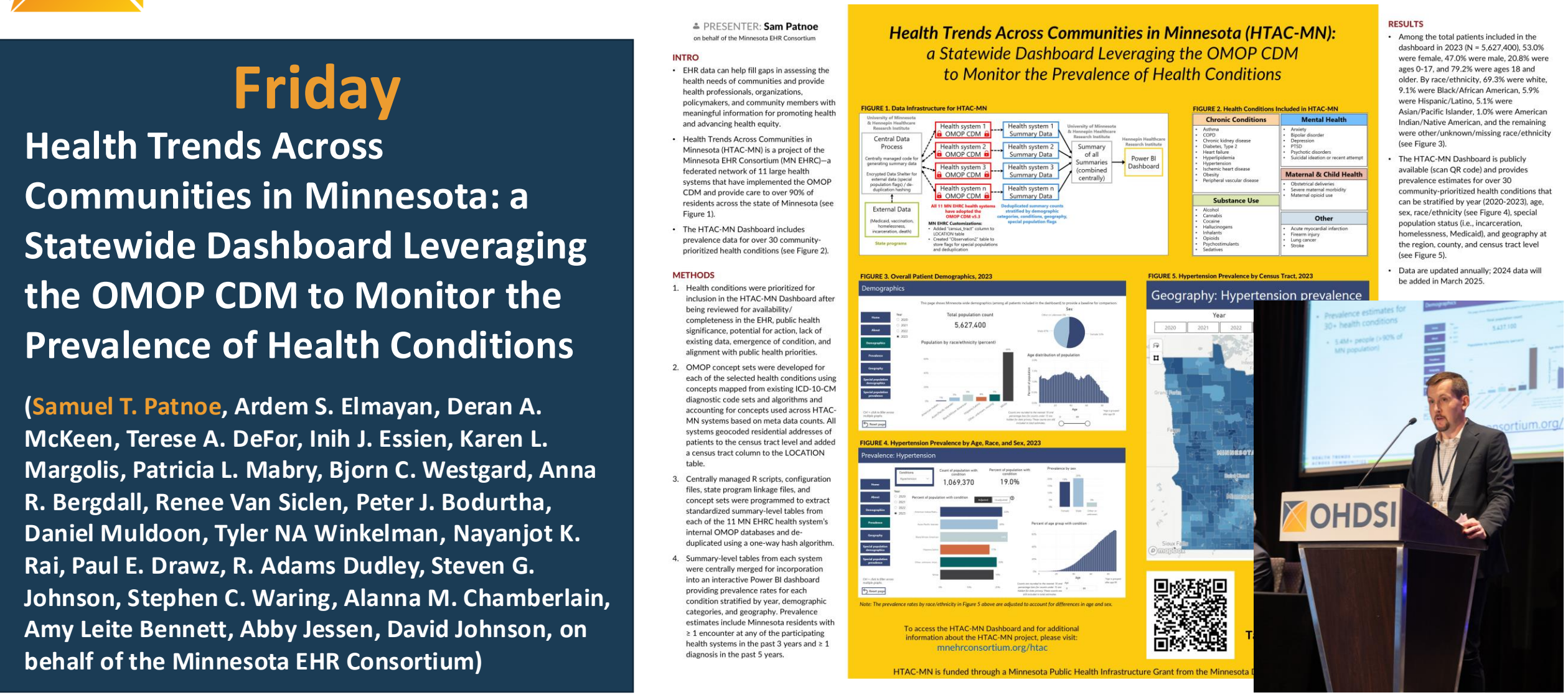
Contacts: Ruochong Fan (fanr@wustl.edu); Linying Zhang (linyingz@wustl.edu)

#OHDSISocialShowcase This Week

Friday

Health Trends Across Communities in Minnesota: a Statewide Dashboard Leveraging the OMOP CDM to Monitor the Prevalence of Health Conditions

(Samuel T. Patnoe, Ardem S. Elmayan, Deran A. McKeen, Terese A. DeFor, Inih J. Essien, Karen L. Margolis, Patricia L. Mabry, Bjorn C. Westgard, Anna R. Bergdall, Renee Van Siclen, Peter J. Bodurtha, Daniel Muldoon, Tyler NA Winkelman, Nayanjot K. Rai, Paul E. Drawz, R. Adams Dudley, Steven G. Johnson, Stephen C. Waring, Alanna M. Chamberlain, Amy Leite Bennett, Abby Jessen, David Johnson, on behalf of the Minnesota EHR Consortium)



PRESENTER: **Sam Patnoe**
on behalf of the Minnesota EHR Consortium

INTRO

- EHR data can help fill gaps in assessing the health needs of communities and provide health professionals, organizations, policymakers, and community members with meaningful information for promoting health and advancing health equity.
- Health Trends Across Communities in Minnesota (HTAC-MN) is a project of the Minnesota EHR Consortium (MN EHRC)—a federated network of 11 large health systems that have implemented the OMOP CDM and provide care to over 90% of residents across the state of Minnesota (see Figure 1).
- The HTAC-MN Dashboard includes prevalence data for over 30 community-prioritized health conditions (see Figure 2).

METHODS

1. Health conditions were prioritized for inclusion in the HTAC-MN Dashboard after being reviewed for availability/completeness in the EHR, public health significance, potential for action, lack of existing data, emergence of condition, and alignment with public health priorities.
2. OMOP concept sets were developed for each of the selected health conditions using concepts mapped from existing ICD-10-CM diagnostic code sets and algorithms and accounting for concepts used across HTAC-MN systems based on meta data counts. All systems geocoded residential addresses of patients to the census tract level and added a census tract column to the LOCATION table.
3. Centrally managed R scripts, configuration files, state program linkage files, and concept sets were programmed to extract standardized summary-level tables from each of the 11 MN EHRC health system's internal OMOP databases and de-duplicated using a one-way hash algorithm.
4. Summary-level tables from each system were centrally merged for incorporation into an interactive Power BI dashboard providing prevalence rates for each condition stratified by year, demographic categories, and geography. Prevalence estimates include Minnesota residents with ≥ 1 encounter at any of the participating health systems in the past 3 years and ≥ 1 diagnosis in the past 5 years.

Health Trends Across Communities in Minnesota (HTAC-MN): a Statewide Dashboard Leveraging the OMOP CDM to Monitor the Prevalence of Health Conditions

FIGURE 1. Data Infrastructure for HTAC-MN

FIGURE 2. Health Conditions Included in HTAC-MN

FIGURE 3. Overall Patient Demographics, 2023

FIGURE 4. Hypertension Prevalence by Age, Race, and Sex, 2023

Note: The prevalence rates by race/ethnicity in Figure 5 above are adjusted to account for differences in age and sex.

FIGURE 5. Hypertension Prevalence by Census Tract, 2023

To access the HTAC-MN Dashboard and for additional information about the HTAC-MN project, please visit:
mnehrconsortium.org/htac

HTAC-MN is funded through a Minnesota Public Health Infrastructure Grant from the Minnesota D

RESULTS

- Among the total patients included in the dashboard in 2023 (N = 5,627,400), 53.0% were female, 47.0% were male, 20.8% were ages 0-17, and 79.2% were ages 18 and older. By race/ethnicity, 69.3% were white, 9.1% were Black/African American, 5.9% were Hispanic/Latino, 5.1% were Asian/Pacific Islander, 1.0% were American Indian/Native American, and the remaining were other/unknown/missing race/ethnicity (see Figure 3).
- The HTAC-MN Dashboard is publicly available (scan QR code) and provides prevalence estimates for over 30 community-prioritized health conditions that can be stratified by year (2020-2023), age, sex, race/ethnicity (see Figure 4), special population status (i.e., incarceration, homelessness, Medicaid), and geography at the region, county, and census tract level (see Figure 5).
- Data are updated annually; 2024 data will be added in March 2025.

Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

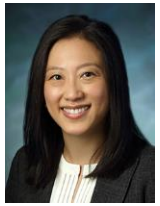
Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

April 1: OHDSI/OMOP Research Spotlight



Cindy Cai • Johns Hopkins University

Semaglutide and Nonarteritic Anterior Ischemic Optic Neuropathy • *JAMA Ophthalmology*

Chen Yanover • KI Research Institute

Characteristics and Outcomes of Over a Million Patients with Inflammatory Bowel Disease in Seven Countries: Multinational Cohort Study and Open Data Resource • *Digestive Diseases and Sciences*

Mitchell Conover • Janssen Research and Development

Objective study validity diagnostics: a framework requiring pre-specified, empirical verification to increase trust in the reliability of real-world evidence • *JAMIA*

Jiayi Tong • Johns Hopkins University

DisC2o-HD: Distributed causal inference with covariates shift for analyzing real-world high-dimensional data • *Journal of Machine Learning Research*

Naimin Jing, Yiwen Lu • University of Pennsylvania

Evaluating the Bias, type I error and statistical power of the prior Knowledge-Guided integrated likelihood estimation (PIE) for bias reduction in EHR based association studies • *Journal of Biomedical Informatics*

The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

Links are sent out weekly and available at:
ohdsi.org/community-calls-2025