ATLAS Deepdive: Data Sources and Vocabularies

OHDSI Community Call
June 10, 2025 • 11 am ET

Upcoming Community Calls

Date	Topic
June 10	ATLAS Deepdive: Data Sources and Vocabularies
June 17	ATLAS Deepdive: Cohorts and Conceptsets
June 24	ATLAS Deepdive: Characterization, Cohort Pathways, Incidence
July 1	ATLAS Deepdive: Technical and Administrative Capabilities
July 8	No Meeting – Europe Symposium
July 15	Europe Symposium Review
July 22	OMOP/OHDSI Research Spotlight

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Justin Bohn, James Gilbert, Christopher Knoll, David Kern, and Patrick Ryan** on the publication of **Large-scale Empirical Identification of Candidate Comparators for Pharmacoepidemiological Studies** in *Drug Safety.*

Drug Safety
https://doi.org/10.1007/s40264-025-01569-y

ORIGINAL RESEARCH ARTICLE

Large-scale Empirical Identification of Candidate Comparators for Pharmacoepidemiological Studies

Justin Bohn[1] · James P. Gilbert[1] · Christopher Knoll[1] · David M. Kern[1] · Patrick B. Ryan[1]

Accepted: 23 May 2025

Abstract

Background and Objective The new user cohort design has emerged as a best practice for the estimation of drug effects from observational data. However, despite its advantages, this design requires the selection and evaluation of comparators for appropriateness, a process that can be challenging. The objective of this work was to introduce an empirical approach to rank candidate comparators in terms of their similarity to a target drug in high-dimensional covariate space.

Methods We generated new user cohorts for each RxNorm ingredient and Anatomic Therapeutic Chemical level 4 class in five administrative claims databases then extracted aggregated pre-treatment covariate data for each cohort across five clinically oriented domains. We formed all pairs of cohorts with ≥ 1000 patients and computed a scalar similarity score, defined as the average of cosine similarities computed within each domain, for each pair. We then generated ranked lists of candidate comparators for each cohort.

Results Across up to 1350 cohorts forming 922,761 comparisons, drugs that were more similar in the Anatomic Therapeutic Chemical hierarchy had higher cohort similarity scores. The most similar candidate comparators for each of six example drugs corresponded to alternative treatments used in the target drug's indication(s), and choosing the top-ranked comparator for randomly selected drugs tended to produce balance on most covariates. This approach also ranked highly those comparators chosen in high-quality published new user cohort design studies.

Conclusion Empirical comparator recommendations may serve as a useful aid to investigators and could ultimately enable the automated generation of new user cohort design-derived evidence, a process that has previously been limited to self-controlled designs.

OHDSI Shoutouts!

Thank you **Lee Evans** for coming to the rescue of OHDSI services over the last week.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

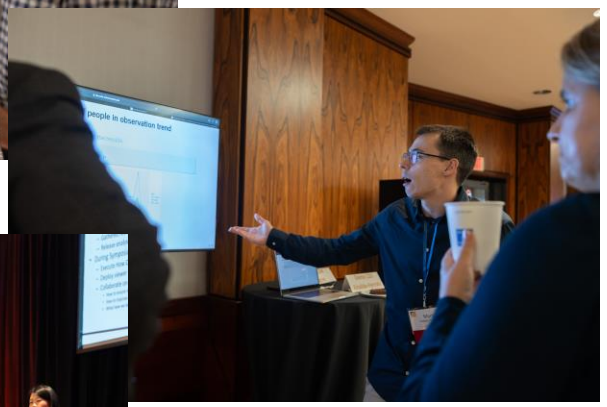
Where Are We Going?

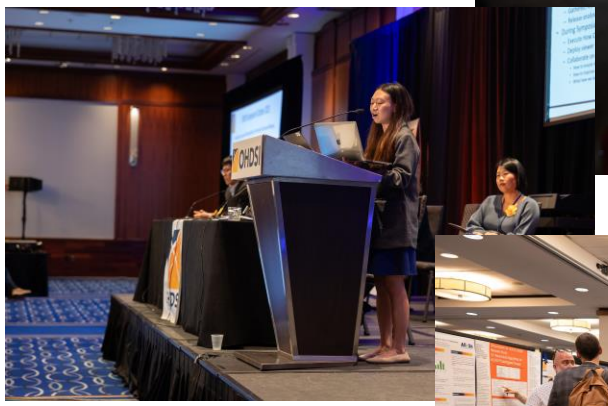
Upcoming Workgroup Calls

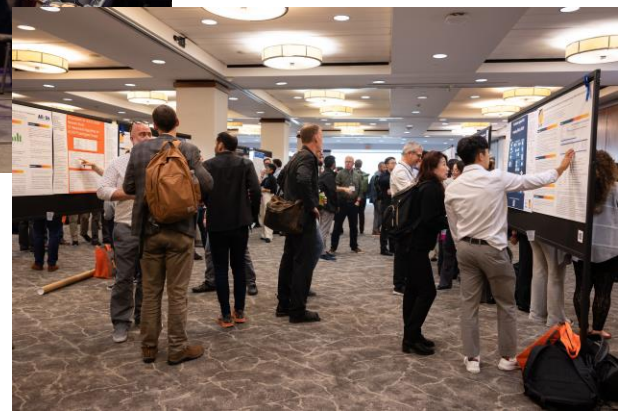
Date	Time (ET)	Meeting
Tuesday	12 pm	Generative AI and Analytics
Tuesday	3 pm	Oncology Outreach/Research Subgroup
Wednesday	7 am	Medical Imaging
Wednesday	9 am	Patient-Level Prediction
Wednesday	11 am	Common Data Model
Wednesday	7 pm	Eyecare and Vision Research
Thursday	9:30 am	Network Data Quality
Thursday	10 am	Rare Diseases
Thursday	10 am	Rehabilitation
Thursday	10:30 am	Evidence Network
Thursday	12 pm	Medical Devices
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	Transplant
Friday	10 am	GIS-Geographic Information System
Friday	11 am	Clinical Trials
Friday	11 pm	China Chapter
Monday	10 am	Healthcare Systems Interest Group
Monday	11 am	Data Bricks Interest Group
Monday	2 pm	Electronic Animal Health Records

THREE Weeks Remaining

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







More information about the collaborator showcase, including links to the submission form and poster templates, can be found on the #OHDSI2025 homepage.

Announcements

NCI's Office of Data Sharing (ODS) Childhood Cancer Data Jamboree

Day 1: Sept. 29, Building 10, NIH Bethesda Campus.

Day 2: Sept. 30, NCI Shady Grove Campus

Both Registration and Project Submissions are required

Virtual options will be provided.

NCI's Office of Data Sharing (ODS) 3rd Annual Symposium

Day 1 & Day 2: Sept. 30 – Oct. 1, NCI Shady Grove Campus or Virtual

Register now

Poster Abstract Submission

The NCI Office of Data Sharing presents the ODS Webinar Series:

Reviewing the 2024 NIH Public Access Policy

June 18, 2025. 2-3 PM EST

- Explain what is changing from the 2008 Public Access Policy
- Cover special guide notices relating to publishing rights and costs
- Question-and-Answer session
- We aim to prepare stakeholders for the new policy when it goes into effect on July 1, 2025

Registration is open to the public

NIH NATIONAL CANCER INSTITUTE

OHDSI on Bluesky

OHDSI is now on Bluesky!

You can now get updates on all community activities and see all global research through the #OHDSISocialShowcase on Bluesky.

bsky.app/profile/ohdsi.bsky.social

Europe Symposium Agenda

Symposium Agenda – July 7, 2025

Time	Topic	
8:00 - 9:00	*Registration & Coffee*	
9:00 - 9:10	Welcome to the European OHDSI Journey *(Speakers: Liesbet M. Peeters & Peter Rijnbeek)*	
9:10 - 9:30	Journey of OHDSI: Where have we been and where can we go together? *(Speaker: Patrick Ryan)*	
9:30 - 11:00	Impact of Leveraging OMOP CDM for Scalable and Reliable Evidence Generation Showcased by the National Nodes *(Moderators: Renske Los & Annelies Verbiest)*	
11:00 - 11:30	*Coffee Break*	
11:30 - 12:45	Collaborator Showcase: Rapid Fire Presentations *(Moderator: TBC)*	
12:45 - 13:45	*Lunch*	
13:45 - 16:00	OHDSI Collaborator Showcase	Early Investigator Mentor Meeting (14:00 - 15:00)
16:00 - 17:10	Bridging Policy and Practice: OHDSI's Role in Implementing the European Health Data Space *(Panel debate)* *(Confirmed speakers/moderators: Enrique Bernal-Delgado, Nick Marly, Talita Duarte-Salles, Patrick Ryan, Dipak Kalra)*	
17:10 - 17:30	Closing remarks *(Speakers: Liesbet M. Peeters & Peter Rijnbeek)*	

Agenda Saturday July 5, 2025

Time	Activity	Track 1A - Newcomers	Track 1B - Newcomers	Track 2 - Advanced	Track 3 - NN/WG
09:30 - 10:00	*Registration + coffee*				
10:00 - 12:30	Morning Session	**Introduction to OHDSI – Tutorial** Lead: Renske Los, Aniek Markus & Laura Verbeij (Erasmus MC) *Overview of OHDSI, key concepts, and an introduction to the OMOP Common Data Model*			**HADES hack-a-thon** Lead: Martijn Schuemie (J&J), Adam Black (Erasmus MC), Anthony Sena (Janssen R&D) *Hands-on coding and tool development in HADES*
12:30 - 13:30	*Lunch break*				
13:30 - 15:00	Afternoon Session I	**OMOP CDM & ETL Conventions** Lead: Maxim Moinat (Erasmus MC), Sofia Bazakou & Anne van Winzum (The Hyve)	**OHDSI Standardized Vocabularies for Research – Part 1.1** Lead: Anna Ostropolets (Janssen R&D), Polina Talapova (Sciforce), Vlad Korsik & Oleg Zhuk (Odysseus) *Concept sets & patient identification techniques.*		
15:00 - 15:30	*Coffee Break*				
15:30 - 17:00	Afternoon Session II		**OHDSI Standardized Vocabularies for Research – Part 1.2** Lead: Anna Ostropolets (Janssen R&D), Polina Talapova (Sciforce), Vlad Korsik & Oleg Zhuk (Odysseus) *Concept sets & patient identification techniques.*		
17:15 - 18:45*	**Optional - guided city tour Hasselt (with local specialities)*				

Agenda Sunday July 6, 2025

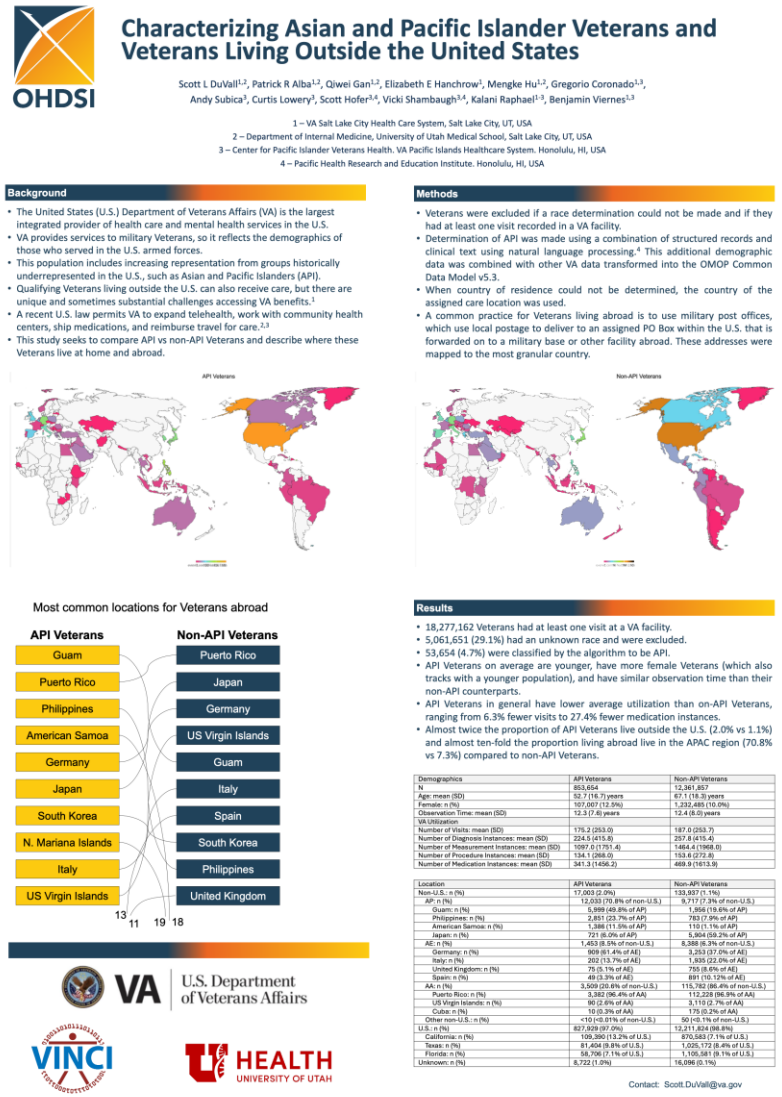
Time	Activity	Track 1A - Newcomers	Track 1B - Newcomers	Track 2 - Advanced	Track 3 - NN/WG
09:30 - 10:00	*Registration + coffee*				
10:00 - 12:30	Morning Session		**OHDSI Standardized Vocabularies for Research – Part 2** Lead: Anna Ostropolets (Janssen R&D), Polina Talapova (Sciforce), Vlad Korsik & Oleg Zhuk (Odysseus) *Final discussion & application of concept sets.*		**NN All Actors Meet Parallel NN meetings**
12:30 - 13:30	***Data Partners Lunch Break***				
13:30 - 15:00	Afternoon Session I	**Whirlwind Introduction to Open-Source Analytic Tools - Part 1** Lead: Martijn Schuemie (J&J), Adam Black (Erasmus MC), Anthony Sena (Janssen R&D) *Overview of HADES and other key OHDSI tools for analysis.*		**Running characterisation studies from beginning to end: a tutorial using DARWIN EU standardised analytics - Part 1** Lead: Daniel Prieto-Alhambra (Oxford University)	**NN All Actors Meet Parallel NN meetings**
15:00 - 15:30	*Coffee Break*				
15:30 - 17:00	Afternoon Session II	**Whirlwind Introduction to Open-Source Analytic Tools - Part 2** Lead: Martijn Schuemie (J&J), Adam Black (Erasmus MC), Anthony Sena (Janssen R&D) *Overview of HADES and other key OHDSI tools for analysis.*		**Running characterisation studies from beginning to end: a tutorial using DARWIN EU standardised analytics - Part 2** Lead: Daniel Prieto-Alhambra (Oxford University)	**OHDSI Europe NN leads meet** Lead: Renske Los (only NN leads/managers)
17:00 - 18:00*	**Optional - networking drink*				

#OHDSISocialShowcase This Week

Monday

Characterizing Asian and Pacific Islander Veterans and Veterans Living Outside the United States

(Scott L DuVall, Patrick R Alba, Qiwei Gan, Elizabeth E Hanchrow, Mengke Hu, Gregorio Coronado, Kalani Raphael, Andy Subica, Curtis Lowery, Scott Hofer, Vicki Shambaugh, Benjamin Viernes)

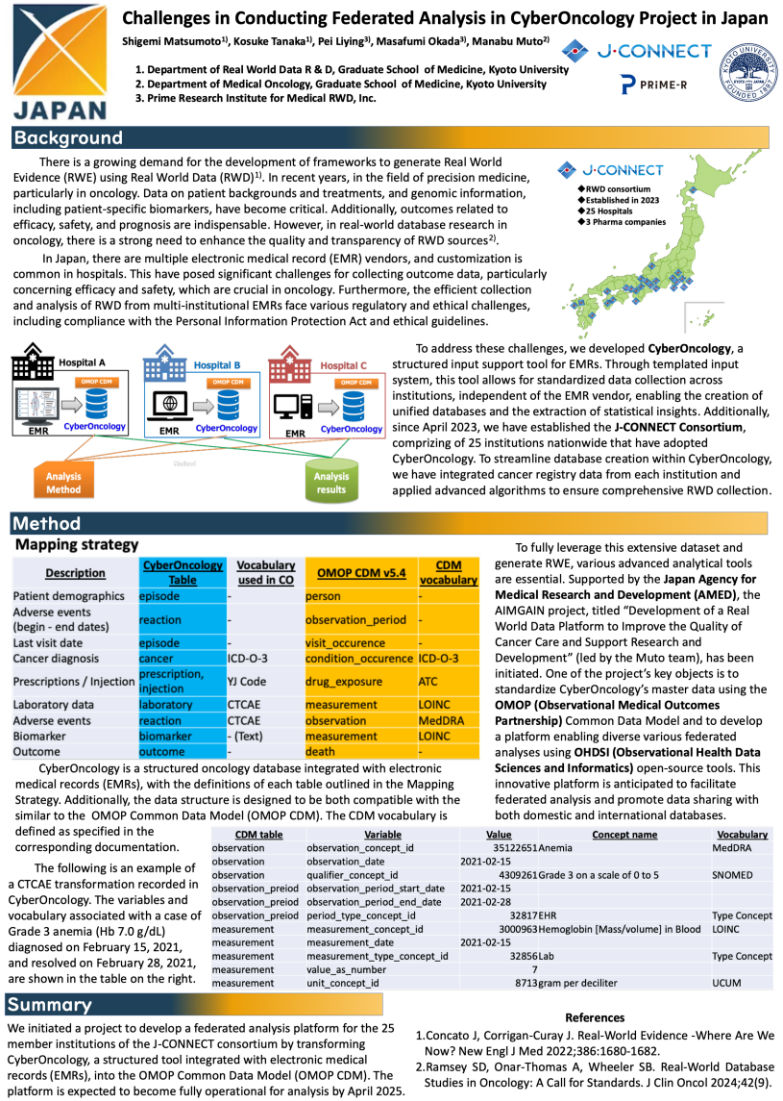


#OHDSISocialShowcase This Week

Tuesday

Challenges in Conducting Federated Analysis in CyberOncology Project in Japan

(Shigemi Matsumoto, Kosuke Tanaka, Liying Pei, Masafumi Okada, Manabu Muto)



Challenges in Conducting Federated Analysis in CyberOncology Project in Japan

Shigemi Matsumoto[1], Kosuke Tanaka[1], Pei Liying[1], Masafumi Okada[3], Manabu Muto[2]

1. Department of Real World Data R & D, Graduate School of Medicine, Kyoto University
2. Department of Medical Oncology, Graduate School of Medicine, Kyoto University
3. Prime Research Institute for Medical RWD, Inc.

JAPAN

J-CONNECT

PRIME-R

Background

There is a growing demand for the development of frameworks to generate Real World Evidence (RWE) using Real World Data (RWD)[1]. In recent years, in the field of precision medicine, particularly in oncology. Data on patient backgrounds and treatments, and genomic information, including patient-specific biomarkers, have become critical. Additionally, outcomes related to efficacy, safety, and prognosis are indispensable. However, in real-world database research in oncology, there is a strong need to enhance the quality and transparency of RWD sources[2].

In Japan, there are multiple electronic medical record (EMR) vendors, and customization is common in hospitals. This have posed significant challenges for collecting outcome data, particularly concerning efficacy and safety, which are crucial in oncology. Furthermore, the efficient collection and analysis of RWD from multi-institutional EMRs face various regulatory and ethical challenges, including compliance with the Personal Information Protection Act and ethical guidelines.

J-CONNECT
- RWD consortium
- Established in 2023
- 25 Hospitals
- 3 Pharma companies

Hospital A / Hospital B / Hospital C
EMR → CyberOncology
Analysis Method
Analysis results

To address these challenges, we developed **CyberOncology**, a structured input support tool for EMRs. Through templated input system, this tool allows for standardized data collection across institutions, independent of the EMR vendor, enabling the creation of unified databases and the extraction of statistical insights. Additionally, since April 2023, we have established the **J-CONNECT Consortium**, comprizing of 25 institutions nationwide that have adopted CyberOncology. To streamline database creation within CyberOncology, we have integrated cancer registry data from each institution and applied advanced algorithms to ensure comprehensive RWD collection.

Method

Mapping strategy

Description	CyberOncology Table	Vocabulary used in CO	OMOP CDM v5.4	CDM vocabulary
Patient demographics	episode	-	person	-
Adverse events (begin - end dates)	reaction	-	observation_period	-
Last visit date	episode	-	visit_occurence	-
Cancer diagnosis	cancer	ICD-O-3	condition_occurence	ICD-O-3
Prescriptions / Injection	prescription, injection	YJ Code	drug_exposure	ATC
Laboratory data	laboratory	CTCAE	measurement	LOINC
Adverse events	reaction	CTCAE	observation	MedDRA
Biomarker	biomarker	- (Text)	measurement	LOINC
Outcome	outcome	-	death	-

CyberOncology is a structured oncology database integrated with electronic medical records (EMRs), with the definitions of each table outlined in the Mapping Strategy. Additionally, the data structure is designed to be both compatible with the similar to the OMOP Common Data Model (OMOP CDM). The CDM vocabulary is defined as specified in the corresponding documentation.

The following is an example of a CTCAE transformation recorded in CyberOncology. The variables and vocabulary associated with a case of Grade 3 anemia (Hb 7.0 g/dL) diagnosed on February 15, 2021, and resolved on February 28, 2021, are shown in the table on the right.

CDM table	Variable	Value	Concept name	Vocabulary
observation	observation_concept_id	35122651	Anemia	MedDRA
observation	observation_date	2021-02-15		
observation	qualifier_concept_id	4309261	Grade 3 on a scale of 0 to 5	SNOMED
observation_preiod	observation_period_start_date	2021-02-15		
observation_preiod	observation_period_end_date	2021-02-28		
observation_preiod	period_type_concept_id	32817	EHR	Type Concept
measurement	measurement_concept_id	3000963	Hemoglobin [Mass/volume] in Blood	LOINC
measurement	measurement_date	2021-02-15		
measurement	measurement_type_concept_id	32856	Lab	Type Concept
measurement	value_as_number	7		
measurement	unit_concept_id	8713	gram per deciliter	UCUM

To fully leverage this extensive dataset and generate RWE, various advanced analytical tools are essential. Supported by the **Japan Agency for Medical Research and Development (AMED)**, the AIMGAIN project, titled "Development of a Real World Data Platform to Improve the Quality of Cancer Care and Support Research and Development" (led by the Muto team), has been initiated. One of the project's key objects is to standardize CyberOncology's master data using the **OMOP (Observational Medical Outcomes Partnership)** Common Data Model and to develop a platform enabling diverse various federated analyses using **OHDSI (Observational Health Data Sciences and Informatics)** open-source tools. This innovative platform is anticipated to facilitate federated analysis and promote data sharing with both domestic and international databases.

Summary

We initiated a project to develop a federated analysis platform for the 25 member institutions of the J-CONNECT consortium by transforming CyberOncology, a structured tool integrated with electronic medical records (EMRs), into the OMOP Common Data Model (OMOP CDM). The platform is expected to become fully operational for analysis by April 2025.

References

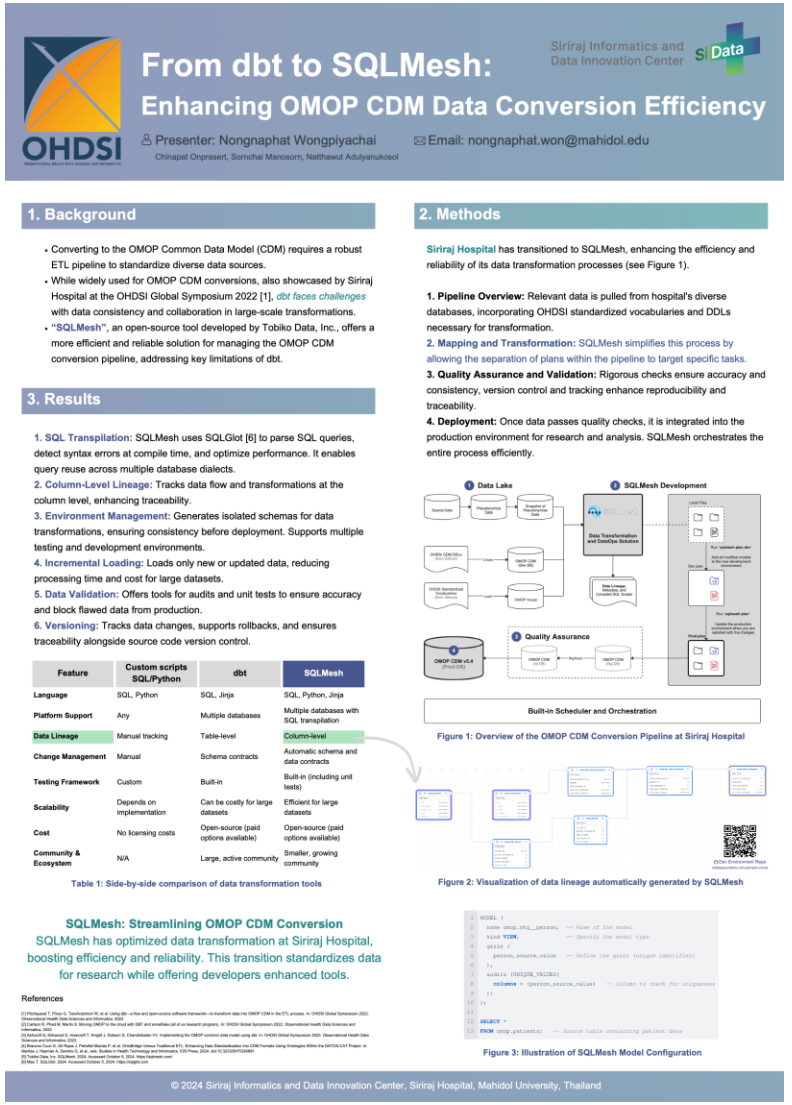
1. Concato J, Corrigan-Curay J. Real-World Evidence -Where Are We Now? New Engl J Med 2022;386:1680-1682.
2. Ramsey SD, Onar-Thomas A, Wheeler SB. Real-World Database Studies in Oncology: A Call for Standards. J Clin Oncol 2024;42(9).

#OHDSISocialShowcase This Week

Wednesday

From dbt to SQLMesh: Enhancing OMOP CDM Data Conversion Efficiency

(Nongnaphat Wongpiyachai, Chinapat Onprasert, Sornchai Manosorn, Natthawut Adulyanukosol)



From dbt to SQLMesh:
Enhancing OMOP CDM Data Conversion Efficiency

Siriraj Informatics and Data Innovation Center

Presenter: Nongnaphat Wongpiyachai
Email: nongnaphat.won@mahidol.edu
Chinapat Onprasert, Sornchai Manosorn, Natthawut Adulyanukosol

1. Background

- Converting to the OMOP Common Data Model (CDM) requires a robust ETL pipeline to standardize diverse data sources.
- While widely used for OMOP CDM conversions, also showcased by Siriraj Hospital at the OHDSI Global Symposium 2022 [1], *dbt faces challenges* with data consistency and collaboration in large-scale transformations.
- **"SQLMesh"**, an open-source tool developed by Tobiko Data, Inc., offers a more efficient and reliable solution for managing the OMOP CDM conversion pipeline, addressing key limitations of dbt.

2. Methods

Siriraj Hospital has transitioned to SQLMesh, enhancing the efficiency and reliability of its data transformation processes (see Figure 1).

1. **Pipeline Overview:** Relevant data is pulled from hospital's diverse databases, incorporating OHDSI standardized vocabularies and DDLs necessary for transformation.
2. **Mapping and Transformation:** SQLMesh simplifies this process by allowing the separation of plans within the pipeline to target specific tasks.
3. **Quality Assurance and Validation:** Rigorous checks ensure accuracy and consistency, version control and tracking enhance reproducibility and traceability.
4. **Deployment:** Once data passes quality checks, it is integrated into the production environment for research and analysis. SQLMesh orchestrates the entire process efficiently.

3. Results

1. **SQL Transpilation:** SQLMesh uses SQLGlot [6] to parse SQL queries, detect syntax errors at compile time, and optimize performance. It enables query reuse across multiple database dialects.
2. **Column-Level Lineage:** Tracks data flow and transformations at the column level, enhancing traceability.
3. **Environment Management:** Generates isolated schemas for data transformations, ensuring consistency before deployment. Supports multiple testing and development environments.
4. **Incremental Loading:** Loads only new or updated data, reducing processing time and cost for large datasets.
5. **Data Validation:** Offers tools for audits and unit tests to ensure accuracy and block flawed data from production.
6. **Versioning:** Tracks data changes, supports rollbacks, and ensures traceability alongside source code version control.

Feature	Custom scripts SQL/Python	dbt	SQLMesh
Language	SQL, Python	SQL, Jinja	SQL, Python, Jinja
Platform Support	Any	Multiple databases	Multiple databases with SQL transpilation
Data Lineage	Manual tracking	Table-level	Column-level
Change Management	Manual	Schema contracts	Automatic schema and data contracts
Testing Framework	Custom	Built-in	Built-in (including unit tests)
Scalability	Depends on implementation	Can be costly for large datasets	Efficient for large datasets
Cost	No licensing costs	Open-source (paid options available)	Open-source (paid options available)
Community & Ecosystem	N/A	Large, active community	Smaller, growing community

Table 1: Side-by-side comparison of data transformation tools

Figure 1: Overview of the OMOP CDM Conversion Pipeline at Siriraj Hospital

Figure 2: Visualization of data lineage automatically generated by SQLMesh

SQLMesh: Streamlining OMOP CDM Conversion

SQLMesh has optimized data transformation at Siriraj Hospital, boosting efficiency and reliability. This transition standardizes data for research while offering developers enhanced tools.

Figure 3: Illustration of SQLMesh Model Configuration

References

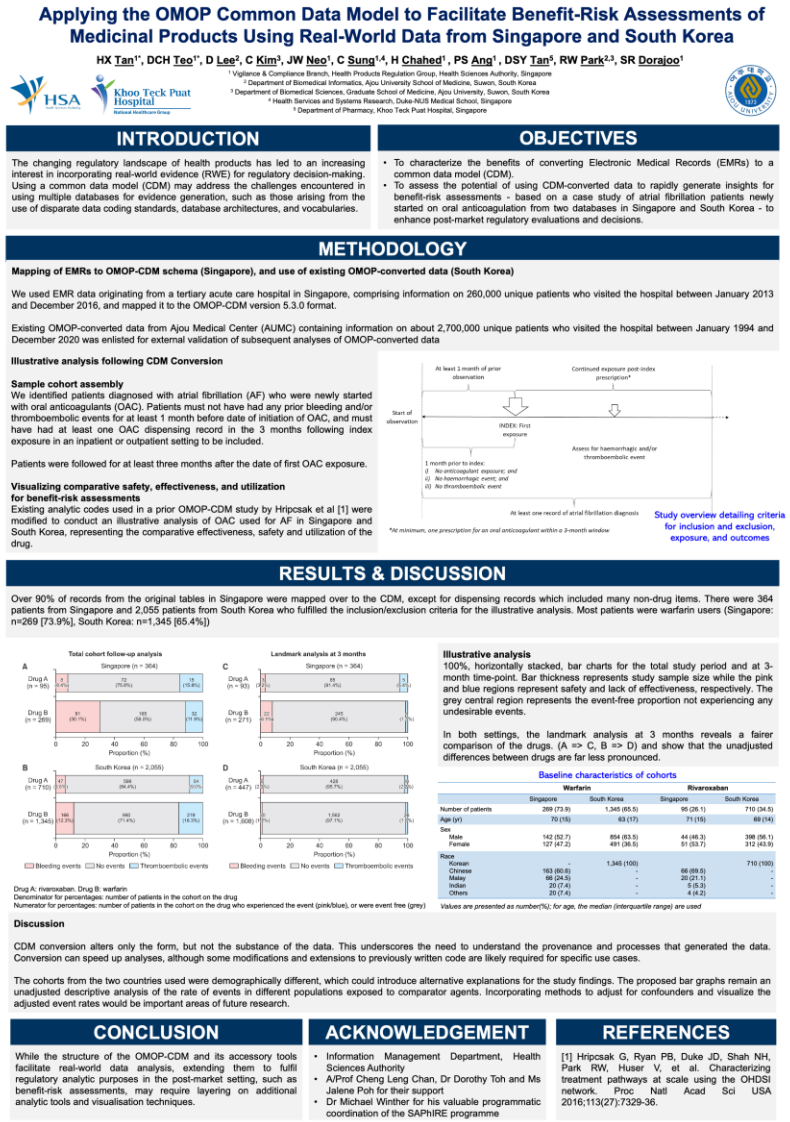
© 2024 Siriraj Informatics and Data Innovation Center, Siriraj Hospital, Mahidol University, Thailand

#OHDSISocialShowcase This Week

Thursday

Applying the OMOP Common Data Model to Facilitate Benefit-Risk Assessments of Medicinal Products Using Real-World Data from Singapore and South Korea

(Hui Xing Tan, Desmond Chun Hwee Teo, Dongyun Lee, Chungsoo Kim, Jing Wei Neo, Cynthia Sung, Haroun Chahed, Pei San Ang, Doreen Su Yin Ta5, Rae Woong Park, Sreemanee Raaj Dorajoo)



Applying the OMOP Common Data Model to Facilitate Benefit-Risk Assessments of Medicinal Products Using Real-World Data from Singapore and South Korea

HX Tan[1*], DCH Teo[1*], D Lee[2], C Kim[3], JW Neo[1], C Sung[1,4], H Chahed[1], PS Ang[1], DSY Tan[5], RW Park[2,3], SR Dorajoo[1]

[1] Vigilance & Compliance Branch, Health Products Regulation Group, Health Sciences Authority, Singapore
[2] Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea
[3] Department of Biomedical Sciences, Graduate School of Medicine, Ajou University, Suwon, South Korea
[4] Health Services and Systems Research, Duke-NUS Medical School, Singapore
[5] Department of Pharmacy, Khoo Teck Puat Hospital, Singapore

INTRODUCTION

The changing regulatory landscape of health products has led to an increasing interest in incorporating real-world evidence (RWE) for regulatory decision-making. Using a common data model (CDM) may address the challenges encountered in using multiple databases for evidence generation, such as those arising from the use of disparate data coding standards, database architectures, and vocabularies.

OBJECTIVES

- To characterize the benefits of converting Electronic Medical Records (EMRs) to a common data model (CDM).
- To assess the potential of using CDM-converted data to rapidly generate insights for benefit-risk assessments - based on a case study of atrial fibrillation patients newly started on oral anticoagulation from two databases in Singapore and South Korea - to enhance post-market regulatory evaluations and decisions.

METHODOLOGY

Mapping of EMRs to OMOP-CDM schema (Singapore), and use of existing OMOP-converted data (South Korea)

We used EMR data originating from a tertiary acute care hospital in Singapore, comprising information on 260,000 unique patients who visited the hospital between January 2013 and December 2016, and mapped it to the OMOP-CDM version 5.3.0 format.

Existing OMOP-converted data from Ajou Medical Center (AUMC) containing information on about 2,700,000 unique patients who visited the hospital between January 1994 and December 2020 was enlisted for external validation of subsequent analyses of OMOP-converted data

Illustrative analysis following CDM Conversion

Sample cohort assembly
We identified patients diagnosed with atrial fibrillation (AF) who were newly started with oral anticoagulants (OAC). Patients must not have had any prior bleeding and/or thromboembolic events for at least 1 month before date of initiation of OAC, and must have had at least one OAC dispensing record in the 3 months following index exposure in an inpatient or outpatient setting to be included.

Patients were followed for at least three months after the date of first OAC exposure.

Visualizing comparative safety, effectiveness, and utilization for benefit-risk assessments
Existing analytic codes used in a prior OMOP-CDM study by Hripcsak et al [1] were modified to conduct an illustrative analysis of OAC used for AF in Singapore and South Korea, representing the comparative effectiveness, safety and utilization of the drug.

Study overview detailing criteria for inclusion and exclusion, exposure, and outcomes

RESULTS & DISCUSSION

Over 90% of records from the original tables in Singapore were mapped over to the CDM, except for dispensing records which included many non-drug items. There were 364 patients from Singapore and 2,055 patients from South Korea who fulfilled the inclusion/exclusion criteria for the illustrative analysis. Most patients were warfarin users (Singapore: n=269 [73.9%], South Korea: n=1,345 [65.4%])

Drug A: rivaroxaban. Drug B: warfarin
Denominator for percentages: number of patients in the cohort on the drug
Numerator for percentages: number of patients in the cohort on the drug who experienced the event (pink/blue), or were event free (grey)

Illustrative analysis
100%, horizontally stacked, bar charts for the total study period and at 3-month time-point. Bar thickness represents study sample size while the pink and blue regions represent safety and lack of effectiveness, respectively. The grey central region represents the event-free proportion not experiencing any undesirable events.

In both settings, the landmark analysis at 3 months reveals a fairer comparison of the drugs. (A => C, B => D) and show that the unadjusted differences between drugs are far less pronounced.

Baseline characteristics of cohorts

	Warfarin		Rivaroxaban	
	Singapore	South Korea	Singapore	South Korea
Number of patients	269 (73.9)	1,345 (65.5)	95 (26.1)	710 (34.5)
Age (yr)	70 (15)	63 (17)	71 (15)	69 (14)
Sex				
Male	142 (52.7)	854 (63.5)	44 (46.3)	398 (56.1)
Female	127 (47.2)	491 (36.5)	51 (53.7)	312 (43.9)
Race				
Korean		1,345 (100)		710 (100)
Chinese	163 (60.6)	-	66 (69.5)	-
Malay	66 (24.5)	-	20 (21.1)	-
Indian	20 (7.4)	-	5 (5.3)	-
Others	20 (7.4)	-	4 (4.2)	-

Values are presented as number(%); for age, the median (interquartile range) are used

Discussion

CDM conversion alters only the form, but not the substance of the data. This underscores the need to understand the provenance and processes that generated the data. Conversion can speed up analyses, although some modifications and extensions to previously written code are likely required for specific use cases.

The cohorts from the two countries used were demographically different, which could introduce alternative explanations for the study findings. The proposed bar graphs remain an unadjusted descriptive analysis of the rate of events in different populations exposed to comparator agents. Incorporating methods to adjust for confounders and visualize the adjusted event rates would be important areas of future research.

CONCLUSION

While the structure of the OMOP-CDM and its accessory tools facilitate real-world data analysis, extending them to fulfil regulatory analytic purposes in the post-market setting, such as benefit-risk assessments, may require layering on additional analytic tools and visualisation techniques.

ACKNOWLEDGEMENT

- Information Management Department, Health Sciences Authority
- A/Prof Cheng Leng Chan, Dr Dorothy Toh and Ms Jalene Poh for their support
- Dr Michael Winther for his valuable programmatic coordination of the SAPhIRE programme

REFERENCES

[1] Hripcsak G, Ryan PB, Duke JD, Shah NH, Park RW, Huser V, et al. Characterizing treatment pathways at scale using the OHDSI network. Proc Natl Acad Sci USA 2016;113(27):7329-36.

#OHDSISocialShowcase This Week

Friday

Protective Effects of SGLT2 Inhibitors on Cardiovascular-Kidney-Metabolic (CKM) Syndrome Progression in Type 2 Diabetes with Chronic Kidney Disease: A Multi-Center Data Analysis Using OMOP-CDM

(Nguyen Phung-Anh, Christianus Heru Setiawan, Ching-Wen Chiu, Phan Thanh-Phuc, Jason C. Hsu)

OHDSI TAIWAN

TAIPEI MEDICAL UNIVERSITY

SGLT2 Inhibitors Mitigate Renal Decline and Cardiovascular Events in T2D Patients with CKD, Slowing CKM Syndrome Progression

Protective Effects of SGLT2 Inhibitors on Cardiovascular-Kidney-Metabolic (CKM) Syndrome Progression in Type 2 Diabetes with Chronic Kidney Disease: A Multi-Center Data Analysis Using OMOP-CDM

Background: This study aims to evaluate the protective effects of sodium-glucose co-transporter 2 inhibitors (SGLT2is) on renal and cardiovascular events in type 2 diabetes (T2D) patients with chronic kidney disease (CKD) within the context of Cardiovascular-Kidney-Metabolic (CKM) Syndrome. Specifically, we assess whether SGLT2is can reduce the progression of renal dysfunction and cardiovascular risks, providing real-world insights using multi-center clinical data.

Methods: We conducted a retrospective cohort study using the OMOP-common data model (CDM) with data sourced from the Taipei Medical University Clinical Research Database (TMUCRD) across three hospitals: Taipei Medical University Hospital, Wanfang Hospital, and Shuang Ho Hospital. The study included patients with T2D and CKD who received antidiabetic medications between 2008 and 2020. Propensity score matching was used to balance patient characteristics across SGLT2is, dipeptidyl peptidase-4 inhibitors (DPP4is), and thiazolidinediones (TZD) groups. A total of 5,005 patients were included, with 524 in the SGLT2is group. Primary outcomes were renal function markers (e.g., sustained ≥50% eGFR reduction, eGFR ≤15 mL/min/1.73 m², and initiation of kidney replacement therapy [KRT]), and the incidence of 4-point major adverse cardiovascular events (4P-MACE).

Figure 1: Cohort Selection Process

Results : The SGLT2is group exhibited significantly lower rates of renal function decline, with an adjusted hazard ratio (HR) of 0.49 (95% CI: 0.29-0.82) for a ≥50% reduction in eGFR, and an HR of 0.41 (95% CI: 0.22-0.77) for eGFR ≤15 mL/min/1.73 m², compared to the DPP4is and TZD groups. Additionally, the incidence of 4P-MACE was significantly reduced in the SGLT2is group (HR: 0.65, 95% CI: 0.47-0.90), including a notable reduction in cardiovascular deaths (HR: 0.37, 95% CI: 0.21-0.65) compared to both DPP4is and TZDs. Subgroup analyses indicated that male patients with pre-existing heart disease particularly benefited from SGLT2is (HR: 0.38, 95% CI: 0.15-0.97).

Conclusion : The use of SGLT2is in T2D patients with CKD significantly mitigates both renal function decline and cardiovascular events, supporting their efficacy in slowing the progression of CKM Syndrome. These findings, based on real-world clinical data from multiple centers, highlight SGLT2is as a valuable therapeutic option for reducing the burden of renal and cardiovascular complications in this high-risk population. This study is expected to be developed into a multinational cooperative research using OHDSI tools and OMOP CDM in the future.

Figure 2: 4P-MACE Outcomes in Patients with SGLT2i or Other Hypoglycemic Agents in Propensity-matched Cohort (1:4)

OHDSI

	Events, No.	Participant years of follow up	Incidence rate (Events/100 participants)	Crude model Hazard ratio (95% CI)	*P* value	Adjusted model* Hazard ratio (95% CI)	*P* value
4P-MACE							
SGLT2i	43	7867	8.0				
vs. non-SGLT2i	254	11433	12.1	0.65 (0.47, 0.90)	0.010	0.68 (0.49, 0.95)	0.024
SGLT2i	42	2842	8.0				
vs. DPP4i	253	11430	12.1	0.65 (0.47, 0.91)	0.012	0.72 (0.52, 1.00)	0.053
SGLT2i	36	2813	7.1				
vs. TZD	40	3336	7.9	0.91 (0.59, 1.43)	0.688	0.76 (0.49, 1.18)	0.217

†Abbreviations: DPP4i, dipeptidyl peptidase-4 inhibitor; MACE, major adverse cardiovascular event; SGLT2i, sodium–glucose cotransporter-2 inhibitors; TZD, thiazolidinedione.

*Adjusted for age, duration of type 2 DM, Charlson Comorbidity Index (CCI), and eGFR.

Nguyen Phung-Anh[1,2,3,4]; Christianus Heru Setiawan[5,6]; Ching-Wen Chiu[7]; Phan Thanh-Phuc[8]; Jason C. Hsu[2,3,4,8*]

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?

June 10: ATLAS Deepdive

Data Sources and Vocabularies

Christopher Knoll

Director, Observational Health Data Analytics
Janssen Research and Development

Join us throughout June to help create the roadmap for ATLAS!

Week 2 ATLAS Survey

These weekly surveys will help us build future versions of ATLAS!

We want to hear from you!

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