OHDSI/OMOP Research Spotlight

OHDSI Community Call
July 22, 2025 • 11 am ET

Upcoming Community Calls

Date	Topic
July 22	OMOP/OHDSI Research Spotlight
July 29	Asia-Pacific Regional Updates
Aug. 5	No Meeting
Aug. 12	Newcomer Introductions
Aug. 19	TBA
Aug. 26	Large-Language Model Innovations in OHDSI
Sept. 2	Standardized Vocabulary Summer Refresh Update
Sept. 9	Global Symposium Preview

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Woo Yeon Park, Teri Sippel Schmidt, Gabriel Salvador, Kevin O'Donnell, Brad Genereaux, Kyulee Jeon, Seng Chan You, Blake E Dewey, Paul Nagy, and the Alzheimer's Disease Neuroimaging Initiative** on the publication of **Breaking data silos: incorporating the DICOM imaging standard into the OMOP CDM to enable multimodal research** in *JAMIA*.

J Am Med Inform Assoc. 2025 Jul 18:ocaf091. doi: 10.1093/jamia/ocaf091. Online ahead of print.

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

Woo Yeon Park [1], Teri Sippel Schmidt [1], Gabriel Salvador [1], Kevin O'Donnell [2], Brad Genereaux [1 3], Kyulee Jeon [4 5], Seng Chan You [4 5], Blake E Dewey [1 6], Paul Nagy [1]; Alzheimer's Disease Neuroimaging Initiative

Collaborators, Affiliations + expand

PMID: 40680297 DOI: 10.1093/jamia/ocaf091

Abstract

Objective: This work incorporates the Digital Imaging Communications in Medicine (DICOM) Standard into the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) to standardize and accurately represent imaging studies, such as acquisition parameters, in multimodal research studies.

Materials and methods: DICOM is the internationally adopted standard that defines entities and relationships for biomedical imaging data used for clinical imaging studies. Most of the complexity in the DICOM data structure centers around the metadata. This metadata contains information about the patient and the modality acquisition parameters. We parsed the DICOM vocabularies in Parts 3, 6, and 16 to obtain structured metadata definitions and added these as custom concepts in the OMOP CDM vocabulary. To validate our pipeline, we harvested and transformed DICOM metadata from magnetic resonance images in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

OHDSI Shoutouts!

Congratulations to the team of **Jiayi Tong, Jenna M. Reps, Chongliang Luo, Yiwen Lu, Lu Li, Juan Manuel Ramirez-Anguita, Milou T. Brand, Scott L. DuVall, Thomas Falconer, Alex Mayer Fuentes, Xing He, Michael E. Matheny, Miguel A. Mayer, Bhavnisha K. Patel, Katherine R. Simon, Marc A. Suchard, Guojun Tang, Benjamin Viernes, Ross D. Williams, Mui van Zandt, Fei Wang, Jiang Bian, Jiayu Zhou, David A. Asch and Yong Chen** on the publication of **Unlocking efficiency in real-world collaborative studies: a multi-site international study with one-shot lossless GLMM algorithm** in *NPJ Digital Medicine.*

npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-01846-1

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

Check for updates

Jiayi Tong[1,2,3] ✉, Jenna M. Reps[4,5,6], Chongliang Luo[7], Yiwen Lu[1,2], Lu Li[1,2], Juan Manuel Ramirez-Anguita[8], Milou T. Brand[9], Scott L. DuVall[10,11], Thomas Falconer[12], Alex Mayer Fuentes[13], Xing He[14,15], Michael E. Matheny[16,17], Miguel A. Mayer[8], Bhavnisha K. Patel[16,17], Katherine R. Simon[16,17], Marc A. Suchard[11,18], Guojun Tang[19], Benjamin Viernes[11], Ross D. Williams[6], Mui van Zandt[9], Fei Wang[20], Jiang Bian[14,15], Jiayu Zhou[21], David A. Asch[22,23] & Yong Chen[1,2,23] ✉

The widespread adoption of real-world data has given rise to numerous healthcare-distributed research networks, but multi-site analyses still face administrative burdens and data privacy challenges. In response, we developed a Collaborative One-shot Lossless Algorithm for Generalized Linear Mixed Models (COLA-GLMM), the first-ever algorithm that achieves both *lossless* and *one-shot* properties. COLA-GLMM ensures accuracy against the gold standard of pooled data while requiring only summary statistics and completes within a single communication round, eliminating the usual back-and-forth overhead. We further introduced an enhanced version that employs homomorphic encryption to reduce the risks of summary statistics misuse at the coordinating center. The simulation studies showed near-exact agreement with the gold standard in parameter estimation, with relative differences of 7.8×10^{-6}%–3.0% under various cell suppression settings. We also validated COLA-GLMM on eight international decentralized databases to identify risk factors for COVID-19 mortality. Together, these results show that COLA-GLMM enables accurate, low-burden, and privacy-preserving multi-site research.

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	9 am	Oncology Outreach/Research Subgroup
Wednesday	11 am	Common Data Model
Wednesday	12 pm	Latin America
Thursday	9:30 am	Network Data Quality
Thursday	12 pm	Medical Devices
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	10 am	Transplant
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Tuesday	9 am	Data2Evidence

Africa Symposium: Nov. 10-12

Who We Are | Updates & News | Standards | Software Tools | Network Studies | Community Forums | Education | New To OHDSI?

Community Calls | Past Events | Workgroups | 2024 Our Journey' Annual Report | Current Events | Support & Sponsorship

2025 Europe Symposium | 2025 Global Symposium | 2025 Africa Symposium | GitHub | YouTube | Twitter | LinkedIn | Newsletters

Join Us At The Inaugural OHDSI Africa Symposium

Nov. 10-12, 2025 · Joint Clinical Research Centre (JCRC) & Mestil Hotel Kampala




The inaugural OHDSI Africa Symposium will be held in Kampala at the Joint Clinical Research Centre (JCRC) and Mestil Hotel. Our community is delighted to introduce a new face-to-face opportunity in Africa, where OHDSI is growing at an exciting pace. We hope you will join us for this historical moment.

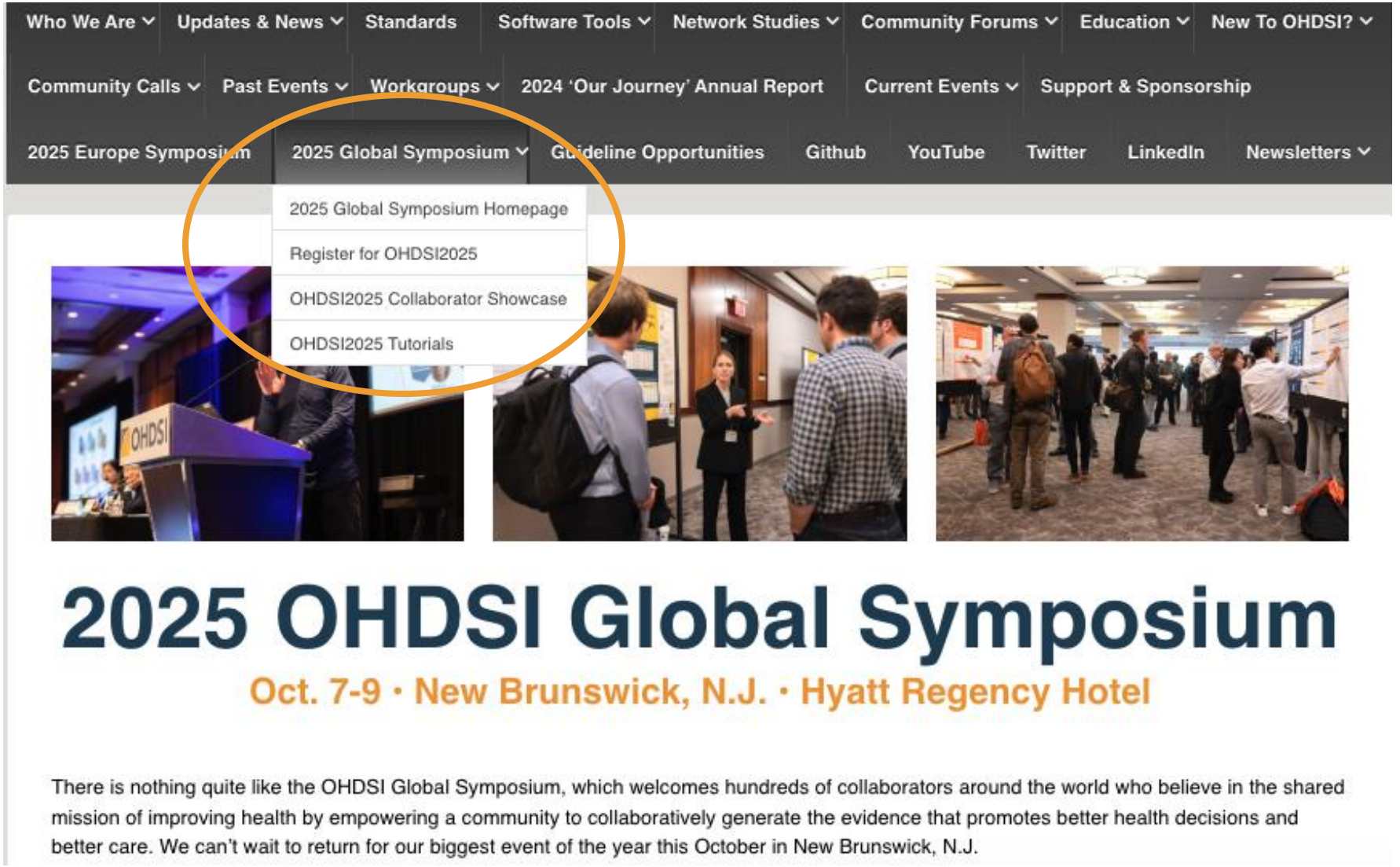
The first OHDSI Africa symposium will be hosted by JCRC and will begin with a dedicated one-day training course at JCRC, followed by a two-day main conference at Mestil hotel. Below are some important dates for you to save to your calendar:

Collaborator Showcase
- Submissions deadline: September 10
- Submissions review: September 11-30
- Notification of acceptance: October 5

Symposium
- Tutorial: November 10 at JCRC
- Main conference: November 11-12 at Mestil Hotel

Register Me for the 2025 OHDSI Africa Symposium!

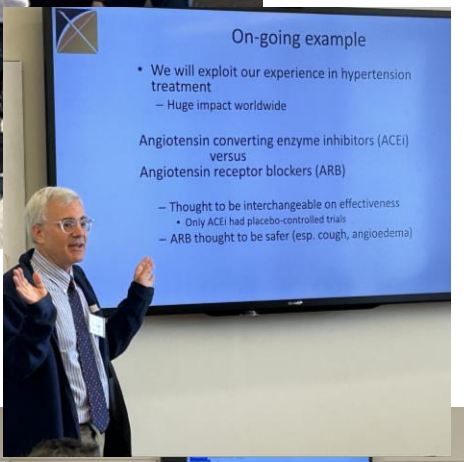
Global Symposium: Oct. 7-9

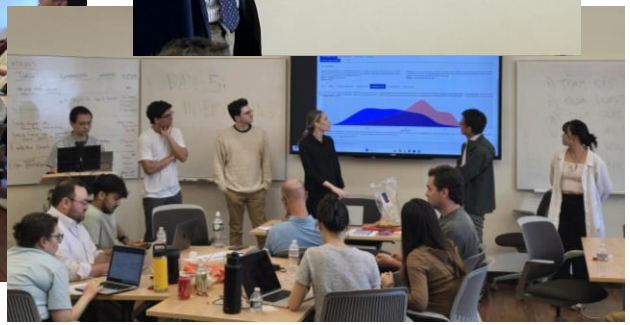


OHDSI Summer School at Columbia

The first-ever Summer School in Observational Health Data Science & Informatics, AI, and Real World Evidence at Columbia University was held last week.

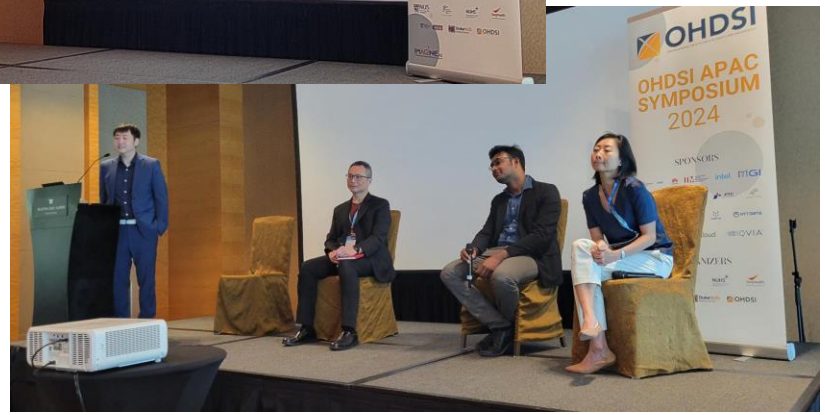
Patrick Ryan, George Hripcsak, Anna Ostropolets and Karthik Natarajan served as faculty for this event.





2025 APAC Symposium

The 2025 OHDSI APAC Symposium will be held Dec. 6-7 in Shanghai, China.



Information on registration and abstract deadline will be posted when available.

The Center for Advanced Healthcare Research Informatics (CAHRI) at Tufts Medicine welcomes:

Tiffany Callahan, PhD

Senior Machine Learning Research Scientist at SandboxAQ

‘Agentic Mixture-of-Workflows for Multi-Modal Chemical Search’

July 31, 2025, 11am-12pm EDT

Virtually via Zoom

Please contact Marty Alvarez at malvarez2@tuftsmedicalcenter.org for calendar invite or questions.

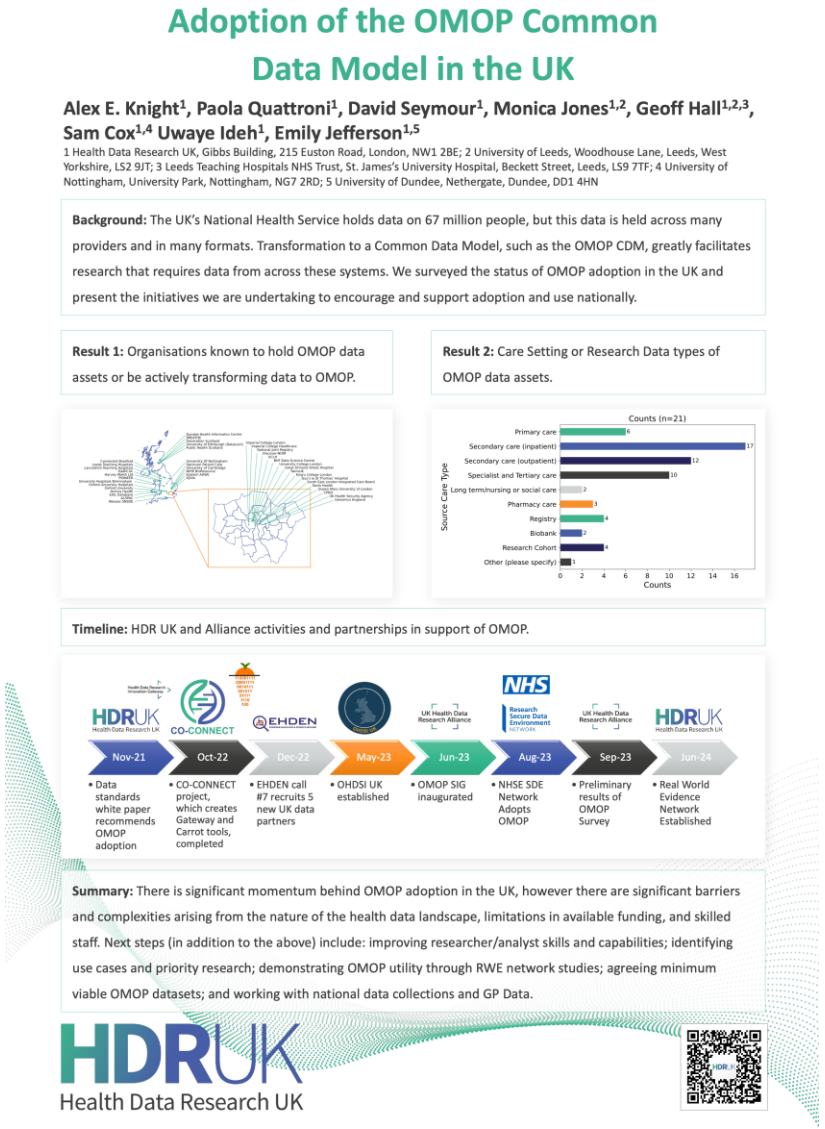
TuftsMedicine
Tufts Medical Center

#OHDSISocialShowcase This Week

Monday

Adoption of the OMOP Common Data Model in the UK

(Alex E. Knight, Paola Quattroni, David Seymour, Monica Jones, Geoff Hall, Sam Cox, Uwaye Ideh, Emily Jefferson)



#OHDSISocialShowcase This Week

Tuesday

OMOP model evaluation for representing GENIE oncology research project data

(Vojtech Huser, Sebastian van Sandijk)

OMOP model evaluation for representing GENIE oncology research project data

Vojtech Huser[1], Sebastian van Sandijk[1]
1: EPAM Odysseus

BACKGROUND

Confirming validity of research results of healthcare observational research on external datasets is good practice. Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) facilitates analysis portability and conducting external validations. Our objective was to evaluate whether a cancer research dataset can be represented using OMOP model and what lessons learned can be communicated to the OHDSI CDM workgroup.

Materials: American Association for Cancer Research established in 2015 a research project in oncology called GENIE (Genomics Evidence Neoplasia Information Exchange). This project made available several datasets. We used GENIE non-small cell lung cancer (NSCLC) dataset.[1] This dataset relies on some features of GENIE registry.[2]

METHODS

If GENIE data were to be represented in OMOP CDM, it would be possible to repeat analysis of GENIE data on other OMOP datasets (for example using data from the 'All of Us' research program). While full conversion of GENIE data into OMOP format was out of scope for our study, as a first step in this ambition, we evaluated whether OMOP model can fully capture GENIE lung cancer data. We used de-identified patient level GENIE NSCLC data and we also retrieved publicly posted corresponding data dictionary of GENIE data elements.[3]

RESULTS

Many publications demonstrate the ability of the OMOP model to capture Electronic Health Record (EHR) data and healthcare claims data. There is also growing experience with using OMOP model for registry data and clinical trials data. GENIE represents a combination of EHR data and Case Report Form (CRF) data captured via an Electronic Data Capture (EDC) system. Some data elements in GENIE contain results of manual chart review. For example, GENIE categorical data element of 'Medical Oncologist Assessment of Change in Cancer Status' (element_id: md_ca_status) has 5 permissible values of: Improving/Responding; Stable/No change; Mixed; Progressing/Worsening/Enlarging; Not stated/Indeterminate). In OMOP model, the OBSERVATION table can be used to capture this data. An identical representation approach was chosen by the All of Us NIH study.[5]

To represent dates, GENIE uses relative paradigm as 'days since birth' for all events.[6] Using year of birth, OMOP representation would introduce some assumptions (e.g., July 1st as default day and month element of birth date) and impute necessary absolute days. This is necessary because the OMOP model requires absolute days.

GENIE medication history data only capture active ingredients administered and not full details about brand name, strength, dose form, or quantity. For ingredients, GENIE uses aggregated ingredient drug terms. Two examples would be: example 1 'Fluorouracil (5FU, 5Fluorouracil, 5Fluracil, AccuSite, Adrucil, Carac, Fluouracil, Flurablastin, Fluracedyl, Fluracil, Fluril, Fluroblastin, Ribofluor, Ro29757); example 2 'Oxaliplatin(1OHP, Al Heng, DACPLAT, Dacotin, ELOXATIN, Eloxatine, JM83)'. If GENIE terms would be mapped and converted to OMOP drug ingredient concepts, it would be possible to formulate a medication analysis and such an analysis would be portable across OMOP datasets. As one special drug concept, GENIE allows capture of administration of an investigational drugs in the context of a clinical trial. The drug recorded would be 'investigational drug' and it would not specify which drug exactly, just the category of investigational). In OMOP, such data rows would have to be represented using 'concept 0 approach' or using a custom local concept for investigational drug ("two billion" concept approach). GENIE medication data also include an inferred construct of drug regimen. EPISODE table in OMOP oncology extension could capture these inferred regiments.[7]

GENIE diagnostic data are organized as cancer diagnosis history and as CRF data elements indicating comorbidities. GENIE registry uses OncoTree terminology for cancer diagnosis. For example, data would use OncoTree code of 'NSCLCPD' (in data element OCOTREE_CODE). Additional two data elements provide additional information: data element CANCER_TYPE has value 'Non-Small Cell Lung Cancer' and data element CANCER_TYPE_DETAILED has value 'Poorly Differentiated Non-Small Cell Lung Cancer'. To achieve comparable analyses on non-oncology specified OMOP data (e.g., All Of Us), diagnosis history data would be captured in the CONDITION_OCCURRENCE OMOP table relying on mapping of OncoTree concepts to standard diagnostic SNOMED CT concept. To describe the mapping using an example, concept with concept_id of 777920 (representing NSCLPDP code in OncoTree) is mapped to standard concept of 'Non-small cell lung cancer' (concept_id of 4115276; vocabulary: SNOMED CT).

Full overview by GENIE source data category and by OMOP target table will be included in the poster and at our project repository at https://github.com/informaticsrepo/omop-onc

CONCLUSIONS

GENIE data can be captured in OMOP model; however, to achieve comparable analyses, careful attention to conventions must be paid. We specifically considered the case of All of Us OMOP dataset to guide our evaluation and see if comparable data would be present and whether necessary mappings currently exist. Mappings of ICD10CM cancer diagnoses (in All Of Us) and OncoTree diagnoses (in GENIE) do not map to identical standard concepts and proper traversing the terminology hierarchy would be essential for any analysis.

For medication analyses, inferring drug regimen episodes from medication history data is an important unsolved challenge because differences in the inference methodology will lead to non-comparable OMOP cancer-related datasets. For some data rows, OMOP lacks suitable convention to capture investigational drugs that would allow across datasets analyses of what cancer types and what types of patients receive investigational drugs.

Discussion

Current oncology workgroup OMOP extension is not formally included in official OMOP v5.4 specification.[8] As a result, All of Us data may be captured on misaligned level of granularity and not capture detailed cancer data in comparable way. Our evaluation revealed some OMOP model limitations that we hope to submit to the OHDSI CDM workgroup, THEMIS workgroup and/or other workgroups. Current standard terminology for diagnoses, SNOMED CT, does not allow full capture of detailed cancer diagnosis and the oncology OMOP extension is the best approach for this data. However, other disease domains may encounter a similar challenge as oncology did. The best compromise may be development of expected expanded target concepts (in measurement and/or observation domain) for each such domain. OHDSI community and CDM workgroup may need to design a formal format specification that would work for number of disease domains. This specification would communicate what expected expanded concepts are preferred target concept to unite the community around. The challenge is analogous to the clinical research informatics challenge of formulating research common data elements [8] but in OMOP model context it can be firmly grounded to currently present concepts in the OMOP Athena terminology layer.

Limitations

We did not include GENIE genomic data in this phase 1 of our evaluation. As future work, we hope to assess to what extend OMOP can capture such data. This is especially important because GENIE project's strength lies in providing extensive genomic data in addition to clinical data for cancer patients. We expect further input to OHDSI workgroups stemming from this future phase 2.

REFERENCES

1. Non-small cell lung cancer GENIE dataset [Internet]. Available from: https://www.aacr.org/professionals/research/aacr-project-genie/bpc/nsclc
2. Acebedo A, Bedard PL, Brown S, Ceca E, Fiandalo M, Fuchs H, et al. Collaborating across sectors in service of open science, precision oncology, and patients: an overview of the AACR Project GENIE (Genomics Evidence Neoplasia Information Exchange) Biopharma Collaborative (BPC). ESMO Real World Data Digit Oncol. 2025 Mar 1;7:100097.
3. GENIE NSLC Dataset Documentation [Internet]. Available from: https://www.aacr.org/wp-content/uploads/2022/05/GENIE-BPC-NSCLC-v2.0-public-Analytic-Data-Guide-1.pdf
4. Mayer CS, Huser V. Learning important common data elements from shared study data: The All of Us program analysis. Pry JM, editor. PLOS ONE. 2023 Jul 7;18(7):e0283601.
5. Choudhury NJ, Lavery JA, Brown S, de Bruijn I, Jee J, Tran TN, et al. The GENIE BPC NSCLC Cohort: A Real-World Repository Integrating Standardized Clinical and Genomic Data for 1,846 Patients with Non–Small Cell Lung Cancer. Clin Cancer Res. 2023 Sep 1;29(17):3418–28.
6. Warner JL, Dymshyts D, Reich CG, Gurley MJ, Hochheiser H, Moldwin ZH, et al. HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. J Biomed Inform. 2019 Aug 1;96:103239.
7. OMOP extensions [Internet]. Available from: https://ohdsi.github.io/CommonDataModel/typesOfAdditions.html#ExpansionExtension
8. Kush RD, Warzel D, Kush MA, Sherman A, Navarro EA, Fitzmartin R, et al. FAIR data sharing: The roles of common data elements and harmonization. J Biomed Inform. 2020 Jul 1;107:103421.

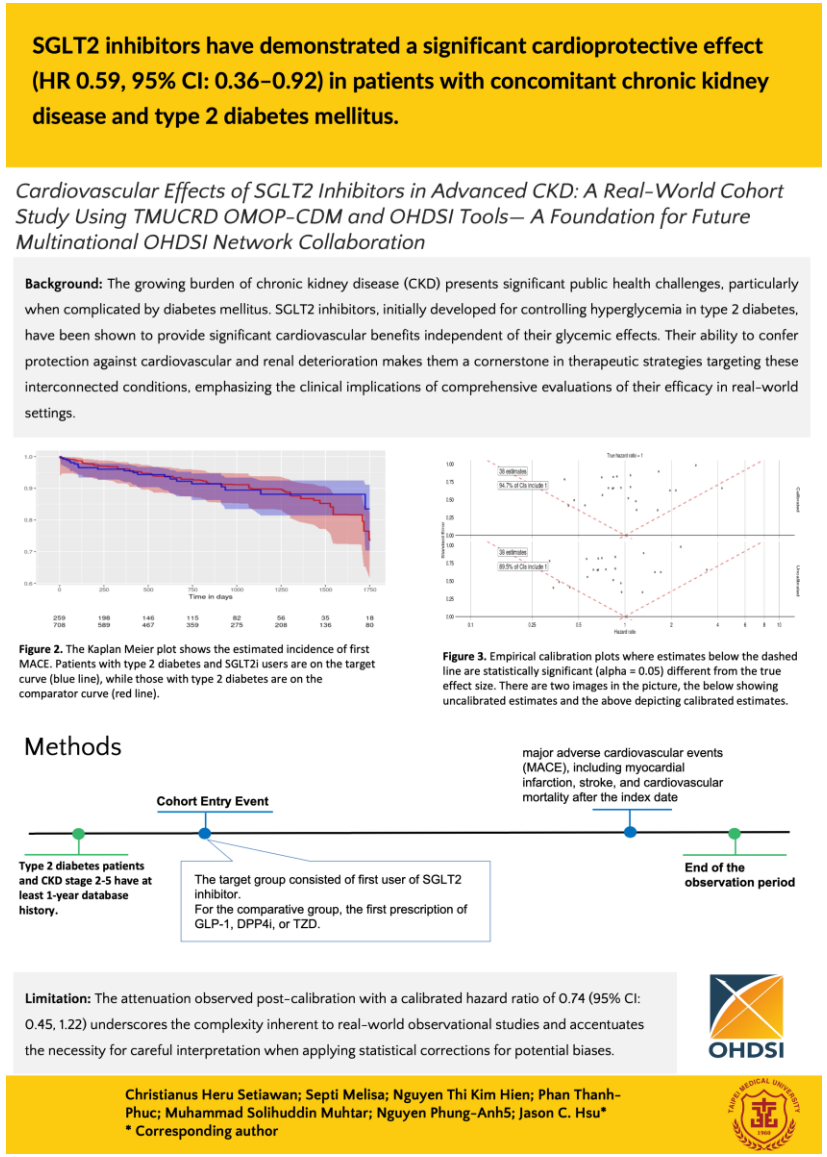
<epam> | ODYSSEUS DATA SERVICES

#OHDSISocialShowcase This Week

Wednesday

Cardiovascular Effects of SGLT2 Inhibitors in Advanced CKD: A Real-World Cohort Study Using TMUCRD OMOP-CDM and OHDSI Tools— A Foundation for Future Multinational OHDSI Network Collaboration

(Christianus Heru Setiawan, Septi Melisa, Nguyen Thi Kim Hien, Phan Thanh-Phuc, Muhammad Solihuddin Muhtar, Nguyen Phung-Anh, Jason C. Hsu)



SGLT2 inhibitors have demonstrated a significant cardioprotective effect (HR 0.59, 95% CI: 0.36–0.92) in patients with concomitant chronic kidney disease and type 2 diabetes mellitus.

Cardiovascular Effects of SGLT2 Inhibitors in Advanced CKD: A Real-World Cohort Study Using TMUCRD OMOP-CDM and OHDSI Tools— A Foundation for Future Multinational OHDSI Network Collaboration

Background: The growing burden of chronic kidney disease (CKD) presents significant public health challenges, particularly when complicated by diabetes mellitus. SGLT2 inhibitors, initially developed for controlling hyperglycemia in type 2 diabetes, have been shown to provide significant cardiovascular benefits independent of their glycemic effects. Their ability to confer protection against cardiovascular and renal deterioration makes them a cornerstone in therapeutic strategies targeting these interconnected conditions, emphasizing the clinical implications of comprehensive evaluations of their efficacy in real-world settings.

Figure 2. The Kaplan Meier plot shows the estimated incidence of first MACE. Patients with type 2 diabetes and SGLT2i users are on the target curve (blue line), while those with type 2 diabetes are on the comparator curve (red line).

Figure 3. Empirical calibration plots where estimates below the dashed line are statistically significant (alpha = 0.05) different from the true effect size. There are two images in the picture, the below showing uncalibrated estimates and the above depicting calibrated estimates.

Methods

Limitation: The attenuation observed post-calibration with a calibrated hazard ratio of 0.74 (95% CI: 0.45, 1.22) underscores the complexity inherent to real-world observational studies and accentuates the necessity for careful interpretation when applying statistical corrections for potential biases.

OHDSI

Christianus Heru Setiawan; Septi Melisa; Nguyen Thi Kim Hien; Phan Thanh-Phuc; Muhammad Solihuddin Muhtar; Nguyen Phung-Anh5; Jason C. Hsu*
* Corresponding author

#OHDSISocialShowcase This Week

Thursday

Accelerating OMOP Common Data Model Adoption for Sustainable Cross-National Atopic Dermatitis Research: Implementation Across Five European Registers

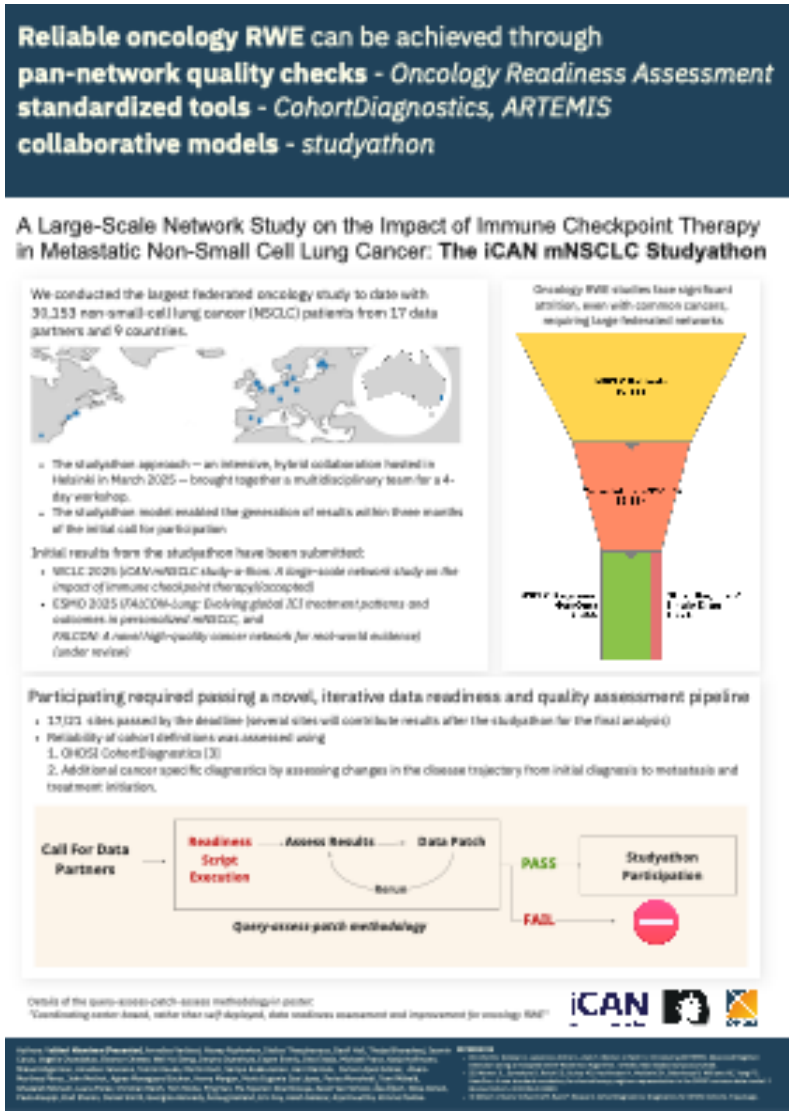
(Axel Hertzschuch, MD Nazrul Islam, Niels Steen Krogh, Man Fung Tsoi, Bolaji Coker, Merel C. Postema, Ahmet Akkoc, David Prieto Merino, Jochen Schmitt, Carsten Flohr, Elizaveta Gribaleva, Thomas Birkner)

#OHDSISocialShowcase This Week

Friday

Large-Scale Network Study on the Impact of Immune Checkpoint Therapy in Metastatic Non-Small Cell Lung Cancer: The iCAN mNSCLC Study-a-Thon

(Valtteri Nieminen, Annelies Verbiest, Alexey Ryzhenkov, Stelios Theophanous, Geoff Hall, Thejas Bharadwaj, Jasmin Carus, Vagelis Chandakas, Eleanor Cheese, Wei Hai Deng, Dmytro Dymshyts, Espen Enerly, Otto Ettala, Michael Franz, Katja Hoffmann, Mikael Högerman, Annelies Janssens, Tommi Kauko, Martin Koch, Sampo Kukkurainen, Harri Rantala, Carlos López Gómez, Álvaro Martínez Pérez, John Methot, Agnes Moesgaard Eschen, Henry Morgan, María Eugenia Gas López, Parisa Movahedi, Tomi Mäkelä, Ghazaleh Niknam, Laura Perez, Christian Reich, Tom Stone, Ping Sun, Pia Tajanen-Doumbouya, Zarah Van Schoor, Åsa Öjlert, Ilkka Ilonen, Paula Kauppi, Elad Sharon, Daniel Smith, Georgina Kennedy, Åslaug Helland, Eric Fey, Asieh Golozar, Aija Knuuttila, Kimmo Porkka)



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?

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