Bridging Evidence Gaps: OHDSI's Guideline-Driven Network Studies

OHDSI Community Call
May 20, 2025 • 11 am ET

Upcoming Community Calls

Date	Topic
May 20	Bridging Evidence Gaps: OHDSI's Guideline-Driven Network Studies
May 27	Collaborator Showcase Brainstorm (Deadline is July 1)
June 3	The Journey of ATLAS
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

Three Stages of The Journey

Where Have We Been?

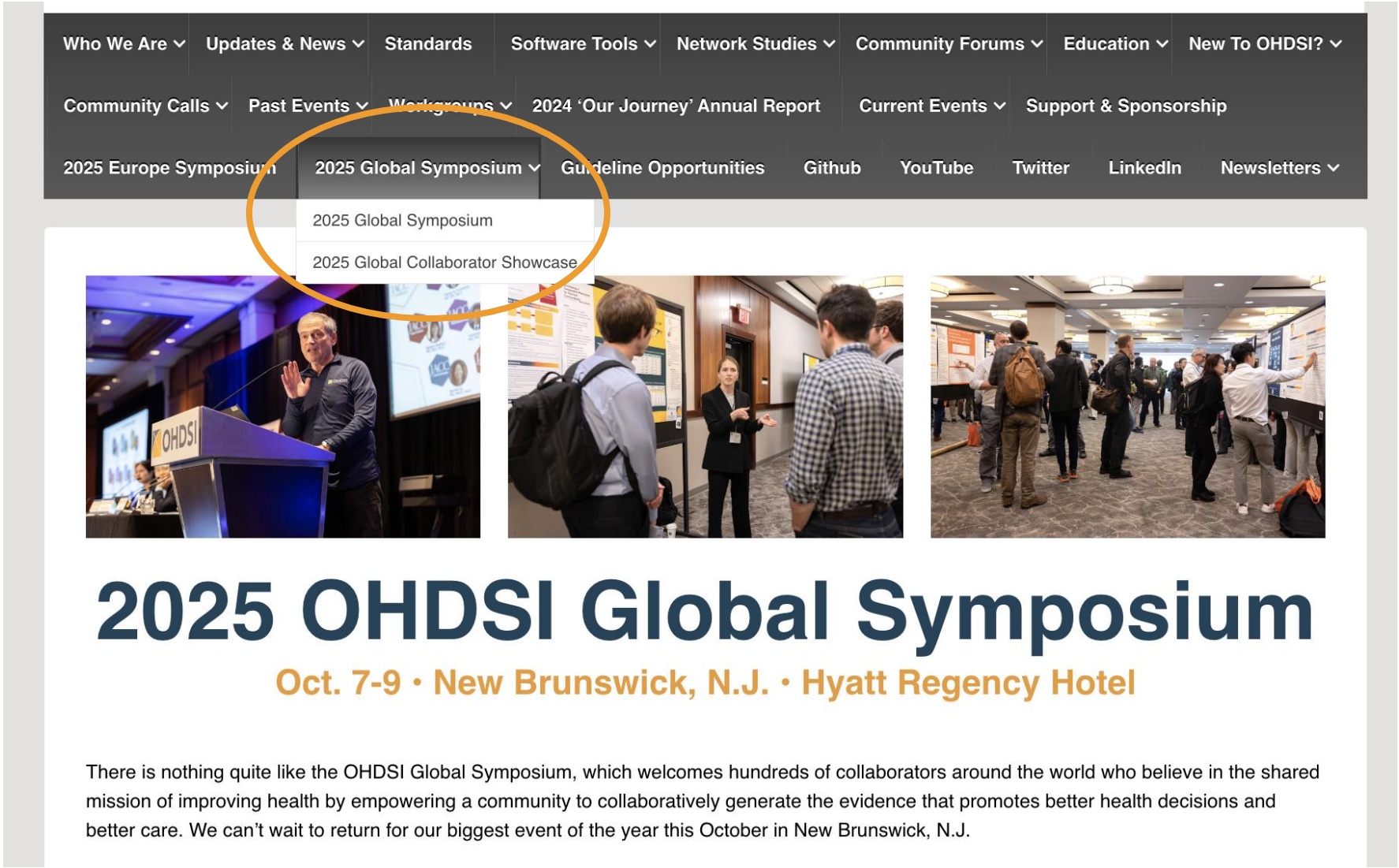
Where Are We Now?

Where Are We Going?

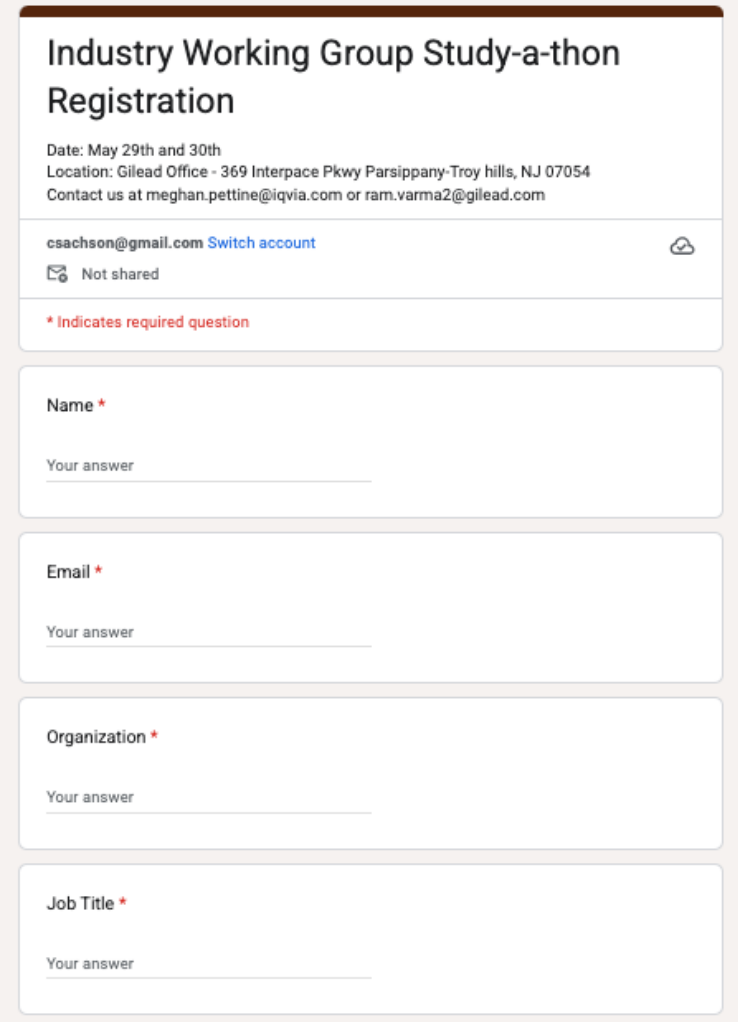
Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	CDM Vocabulary Subgroup
Tuesday	12 pm	ATLAS
Thursday	9:30 am	Network Data Quality
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Africa Chapter
Monday	10 am	Getting Started
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	9:30 am	CDM Survey

Symposium Event Page is Live



Industry Workgroup Studyathon



Industry Working Group Study-a-thon Registration

Date: May 29th and 30th
Location: Gilead Office - 369 Interpace Pkwy Parsippany-Troy hills, NJ 07054
Contact us at meghan.pettine@iqvia.com or ram.varma2@gilead.com

csachson@gmail.com Switch account

Not shared

* Indicates required question

Name *

Your answer

Email *

Your answer

Organization *

Your answer

Job Title *

Your answer

Where: Parsippany, New Jersey

Hosted By: Gilead

The OHDSI Industry Working Group is pleased to announce a two day in-person study-a-thon. The primary objective is to provide stakeholders with a tangible example of the advantages of joining the OHDSI network. These benefits include:

- The OMOP Common Data Model (CDM) aligns with institutional goals and highlights the advantages of a standardized data model.
- Accelerated timelines for conducting network studies.
- A high-quality and user-friendly technology stack.
- Established trust in OMOP vocabularies.
- A comprehensive solution for conducting efficient studies.

This demonstration aims to underscore the value and effectiveness of the OHDSI network, encouraging stakeholders to recognize its potential and benefits.

Interested in learning more? Please join the OHDSI Industry Working Group during our monthly meeting. Past meeting recordings and notes are available on Teams.

ATLAS Usage & Feedback Survey

Atlas Survey

General

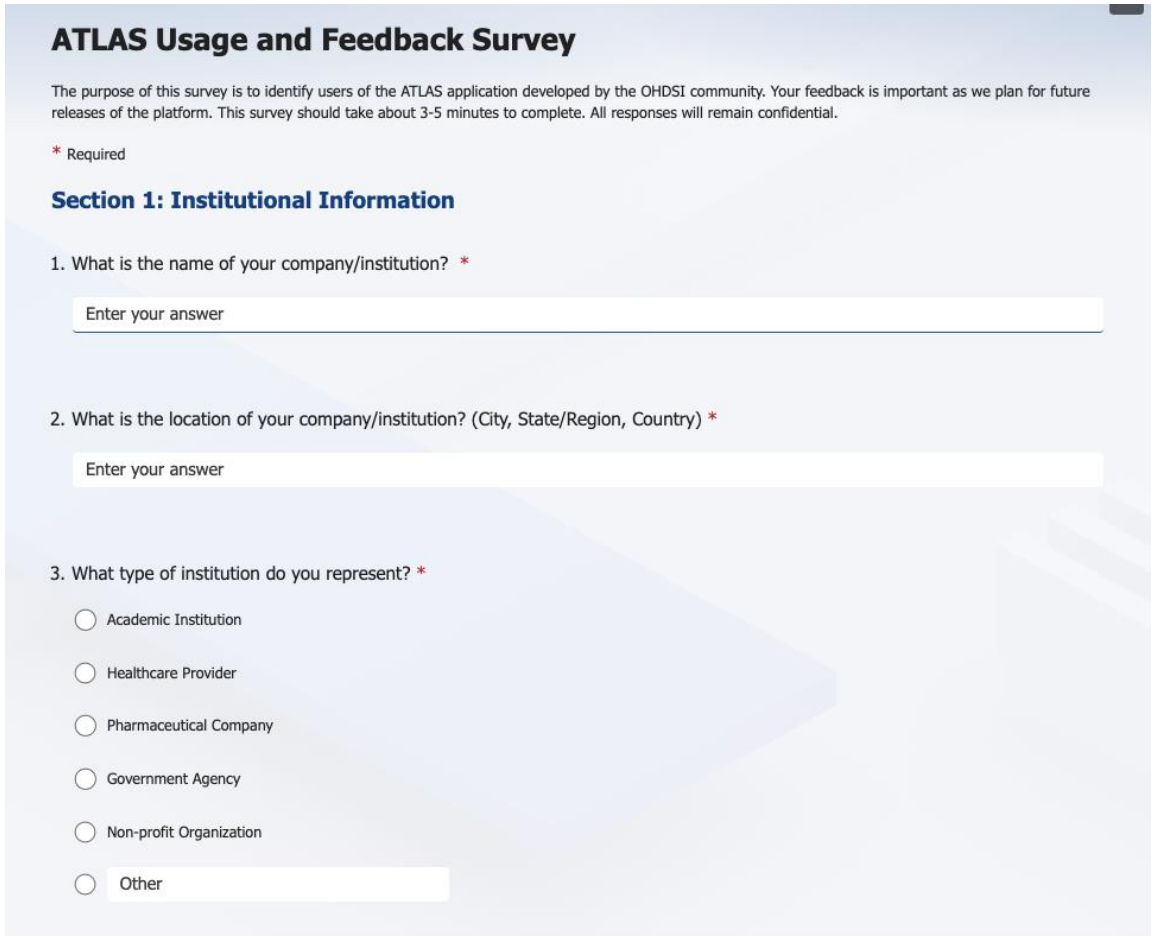
Chris_Knoll 5d

The ATLAS working group has put together a short survey (Microsoft Forms 6) to help us identify who is using ATLAS in our community. If you are not using ATLAS, we'd still ask you to fill in this survey so you can help us to identify any barriers for adoption in your company/institution.

Additionally, this survey will ask if you'd like to be interviewed for feedback on your usage of ATLAS. Data4Life (https://www.data4life.care/ 1) is working closely with our working group to conduct interviews (~1hr) that will help inform the future direction of the application.

We'd appreciate if you could fill out this survey and consider speaking with Data4Life regarding your experiences with ATLAS.

Tagging @anthonysena



ATLAS Usage and Feedback Survey

The purpose of this survey is to identify users of the ATLAS application developed by the OHDSI community. Your feedback is important as we plan for future releases of the platform. This survey should take about 3-5 minutes to complete. All responses will remain confidential.

* Required

Section 1: Institutional Information

1. What is the name of your company/institution? *

Enter your answer

2. What is the location of your company/institution? (City, State/Region, Country) *

Enter your answer

3. What type of institution do you represent? *

- Academic Institution
- Healthcare Provider
- Pharmaceutical Company
- Government Agency
- Non-profit Organization
- Other

Link on community calls page

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

Georgie Kennedy, PhD

Senior Research Fellow, University of New South Wales and the Ingham Institute

'Learning from Real-World Cancer Data: Maturing data pipelines to support research that impacts clinical care'

May 29, 2025, 9-10am EDT

Virtually via Zoom

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

TuftsMedicine
Tufts Medical Center

Job Opening

Manager, Observational Health Data Analytics, Johnson & Johnson

Johnson&Johnson

Data Analytics & Computational Sciences | Raritan / United States of America

Manager, Observational Health Data Analytics - Raritan, NJ

Primary responsibilities:

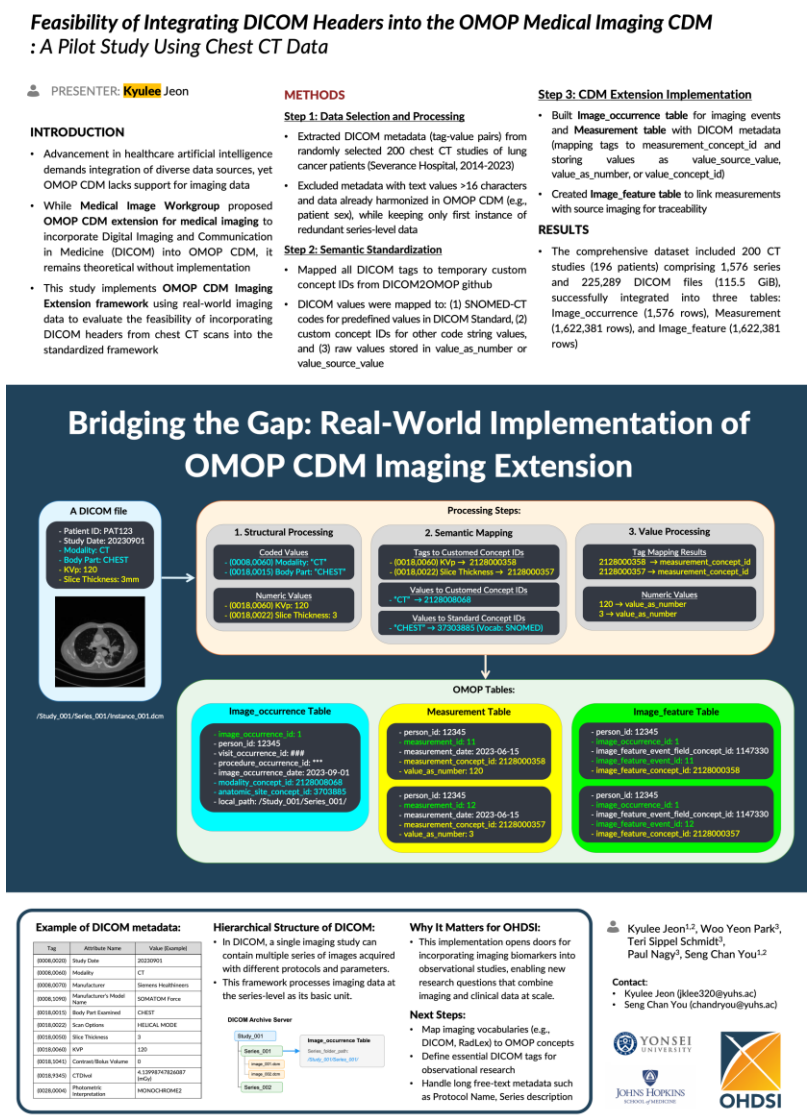
- Work closely with colleagues within GEO.
- Contribute to the successful delivery of observational analyses for clinical characterization, population-level effect estimation, and patient-level prediction to meet the needs of Johnon & Johnson's scientific and business functions.
- Contribute to the design of observational database analysis, including authoring protocol and analysis plans.
- Contribute to the execution of observational database analyses by using standardized analytical tools and writing statistical programs against internal and external observational data resources.
- Contribute to innovating, evaluating, and establishing scientific best practices around the design and conduct of observational analysis and accompanying processes to ensure the reliability of real-world evidence.
- Contribute to the design and development of software and analytical tools that encode scientific best practices into solutions that enable real-world evidence generation and dissemination.
- Contribute to the development and evolution of scientific and industry standards for observational data harmonization, ensuring their appropriate application across the Johnson & Johnson real-world data ecosystem, and leading the evaluation and characterization of observational data for their fitness-for-use to address clinical questions from across the organization.

#OHDSISocialShowcase This Week

Monday

Feasibility of Integrating DICOM Headers into the OMOP Medical Imaging Common Data Model (MI-CDM): A Pilot Study Using Chest CT Data

(Kyulee Jeon, Woo Yeon Park, Teri Sippel Schmidt, Paul Nagy, Seng Chan You)



Feasibility of Integrating DICOM Headers into the OMOP Medical Imaging CDM : A Pilot Study Using Chest CT Data

PRESENTER: Kyulee Jeon

INTRODUCTION

- Advancement in healthcare artificial intelligence demands integration of diverse data sources, yet OMOP CDM lacks support for imaging data
- While **Medical Image Workgroup** proposed **OMOP CDM extension for medical imaging** to incorporate Digital Imaging and Communication in Medicine (DICOM) into OMOP CDM, it remains theoretical without implementation
- This study implements **OMOP CDM Imaging Extension framework** using real-world imaging data to evaluate the feasibility of incorporating DICOM headers from chest CT scans into the standardized framework

METHODS

Step 1: Data Selection and Processing

- Extracted DICOM metadata (tag-value pairs) from randomly selected 200 chest CT studies of lung cancer patients (Severance Hospital, 2014-2023)
- Excluded metadata with text values >16 characters and data already harmonized in OMOP CDM (e.g., patient sex), while keeping only first instance of redundant series-level data

Step 2: Semantic Standardization

- Mapped all DICOM tags to temporary custom concept IDs from DICOM2OMOP github
- DICOM values were mapped to: (1) SNOMED-CT codes for predefined values in DICOM Standard, (2) custom concept IDs for other code string values, and (3) raw values stored in value_as_number or value_source_value

Step 3: CDM Extension Implementation

- Built **Image_occurrence table** for imaging events and **Measurement table** with DICOM metadata (mapping tags to measurement_concept_id and storing values as value_source_value, value_as_number, or value_concept_id)
- Created **Image_feature table** to link measurements with source imaging for traceability

RESULTS

- The comprehensive dataset included 200 CT studies (196 patients) comprising 1,576 series and 225,289 DICOM files (115.5 GiB), successfully integrated into three tables: Image_occurrence (1,576 rows), Measurement (1,622,381 rows), and Image_feature (1,622,381 rows)

Example of DICOM metadata:

Tag	Attribute Name	Value (Example)
(0008,0020)	Study Date	20230901
(0008,0060)	Modality	CT
(0008,0070)	Manufacturer	Siemens Healthineers
(0008,1090)	Manufacturer's Model Name	SOMATOM Force
(0018,0015)	Body Part Examined	CHEST
(0018,0022)	Scan Options	HELICAL MODE
(0018,0050)	Slice Thickness	3
(0018,0060)	KVP	120
(0018,1041)	Contrast/Bolus Volume	0
(0018,9345)	CTDIvol	4.139978747826087 [mGy]
(0028,0004)	Photometric Interpretation	MONOCHROME2

Hierarchical Structure of DICOM:

- In DICOM, a single imaging study can contain multiple series of images acquired with different protocols and parameters.
- This framework processes imaging data at the series-level as its basic unit.

DICOM Archive Server

Why It Matters for OHDSI:

- This implementation opens doors for incorporating imaging biomarkers into observational studies, enabling new research questions that combine imaging and clinical data at scale.

Next Steps:

- Map imaging vocabularies (e.g., DICOM, RadLex) to OMOP concepts
- Define essential DICOM tags for observational research
- Handle long free-text metadata such as Protocol Name, Series description

Kyulee Jeon[1,2], Woo Yeon Park[3], Teri Sippel Schmidt[3], Paul Nagy[3], Seng Chan You[1,2]

Contact:
- Kyulee Jeon (jklee320@yuhs.ac)
- Seng Chan You (chandryou@yuhs.ac)

YONSEI UNIVERSITY
JOHNS HOPKINS SCHOOL of MEDICINE
OHDSI

#OHDSISocialShowcase This Week

Tuesday

Risk of aortic aneurysm or dissection following use of fluoroquinolones: a multinational network cohort study

(Jack L Janetzki, Jung Ho Kim, Nicole Pratt, Seng Chan You)

Risk of aortic aneurysm or dissection following use of fluoroquinolones: a multinational network cohort study

PRESENTER: Jack Janetzki

INTRODUCTION:

- International regulators warned about a rare increased risk of potentially fatal aortic aneurysm (AA) and aortic dissection (AD) with use of fluoroquinolone (FQ) antibiotics.
- Prior studies were poorly designed which generated moderate evidence and conflicting findings.

METHODS

1. We conducted a retrospective cohort study via large scale distributed network analysis on 14 databases from 5 countries.
2. We included patients aged ≥35 years of age who were taking FQ or a comparator antibiotic (trimethoprim +/- sulfamethoxazole(TMP) or cephalosporins (CPHs) for treatment of urinary tract infection (UTI) between January 2010 and December 2019. The primary outcome was occurrence of AA/AD within 60 days of exposure.
3. Cox proportional hazards models were used to estimate risk of outcome after 1:1 PS matching.
4. Results were calibrated based on results of negative control outcomes.
5. We employed objective diagnostics to evaluate the analytic method performance.
6. Results were pooled across databases using Bayesian random effects meta-analysis.

In a study both large and exact,

Fluoroquinolones held to the fact.

No aneurysm woe

For the data did show—

They're a choice we need not retract.

Take a picture to view the results OHDSI analysis viewer

RESULTS

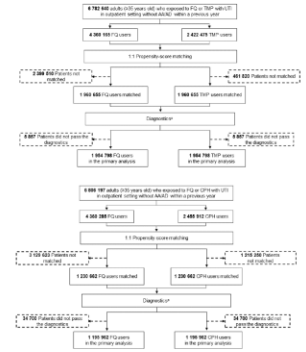


Figure 1: Study Flowchart of Patients Initiating Fluoroquinolones, Trimethoprim with or without Sulfamethoxazole, or Cephalosporins for Urinary Tract Infection

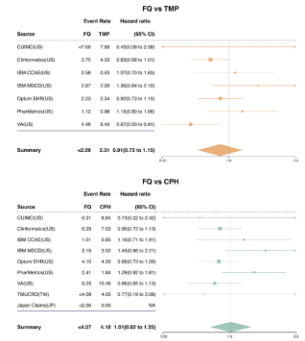


Figure 2: Meta-analytic Comparative Risk of Aortic Aneurysm or Dissection Within 60 Days After Treatment Initiation for Urinary Tract Infection

Jack Janetzki, Jung Ho Kim, Nicole Pratt, and Seng Chan You, on behalf of the 2023 OHDSI SOS Challenge Fluoroquinolone team

University of South Australia

OHDSI

#OHDSISocialShowcase This Week

Wednesday

Mapping Thai Medicine Terminology to RxNorm: Lessons Learned in Standard Vocabulary Integration

(Krittaphas Chaisutyakorn, Peamboon Thomchotpong, Natpatchara Pongjirapat, Jirapat Aiamsopon, Natthawut Adulyanukosol)

OHDSI

Siriraj Informatics and Data Innovation Center SiData

Mapping Thai Medicine Terminology to RxNorm: Lesson Learned in Standard Vocabulary Integration

Presenter: Krittaphas 'Krit' Chaisutyakorn Email: krittaphas.cha@mahidol.edu

Background

Siriraj Hospital is currently transforming 20+ years of EHR data into the OMOP CDM [1] to support observational research. Vocabulary mapping is a crucial process, as having standardized medical concepts is essential for multicenter international collaboration. In this report, we present our methods, mapping processes, and obstacles related to mapping Thai Medicine Terminology (TMT) [2] to RxNorm medication codes.

Method

First, we extracted the source data, which included medication source codes, medication descriptions, and medication usage frequencies. The medication codes included were retrieved from Siriraj Hospital's EHR data, representing a subset of the TMT codes. The mapping process was performed at the code level, encompassing a triad of ingredient, strength, and form. The medication codes were then sorted by usage frequency. Next, the source data was input into the **Usagi** vocabulary mapping tool [3], which returned the automatic matching results.

Figure 1: Diagram of the mapping process

Next, we split the mapping process into 2 steps. In the first step, the automated matched codes were quickly reviewed. We approved the automated matches only when they were unquestionably correct (e.g., same medication name, dose, and form). This step was completed by one person. No manual edits were performed in this step. This approach helped accelerate the process, allowing even individuals with minimal medical knowledge to complete this step.

In the second step, we carried out the manual mapping process. This involved correcting incorrect results from the automated step. Four physicians were assigned medication codes, sorted by their usage frequency.

TLDR

- Breaking down the mapping process into steps make the task much more achievable
- Only 60% of medication codes can cover over 95% of the total medication usage count in our Hospital
- In the total of 15 hours, we can cover more than 95% of the medication usage frequency by using proper tools and methods.

Result

Figure 2: Number and percentage of mapped codes of each step

Our hospital has a total of 2,749 medication codes in use. After the automated matching and first approval processes 1,213 codes were correctly mapped and approved, covering 80% of the medication usage frequency.

In the manual mapping process, we extended the vocabulary to cover 1,778 codes, resulting in 96% coverage of the total usage frequency.

1st-round: 1 person, Total time 5 hours

Manual Review: 4 people, Average time 2.5 hours/person, Total time 10 hours

Figure 3: Time and Person investment in the project

The total time investment for this project was 3 weeks. The first round approval process was completed by single person in a total of 5 hours. The manual mapping process was completed by 4 physicians, each spending an average of 2.5 hours.

Even though the automatic matching result from Usagi tool can help quickening the mapping process, there are some types of medications that the Usagi tool tends to fail. In our observation, these medications are as follow:

- Intravenous medication with varied concentration
- Combination of medications
- Vaccines

References

1. Data Standardization - OHDSI;. Available from: https://www.ohdsi.org/data-standardization/
2. Thai Medicines Terminology (TMT) - Thai Health Information Standards Development Center (THIS);. Available from: https://www.this.or.th/service/tmt/
3. USAGI for vocabulary mapping - OHDSI;. Available from: https://www.ohdsi.org/analytic-tools/usagi/

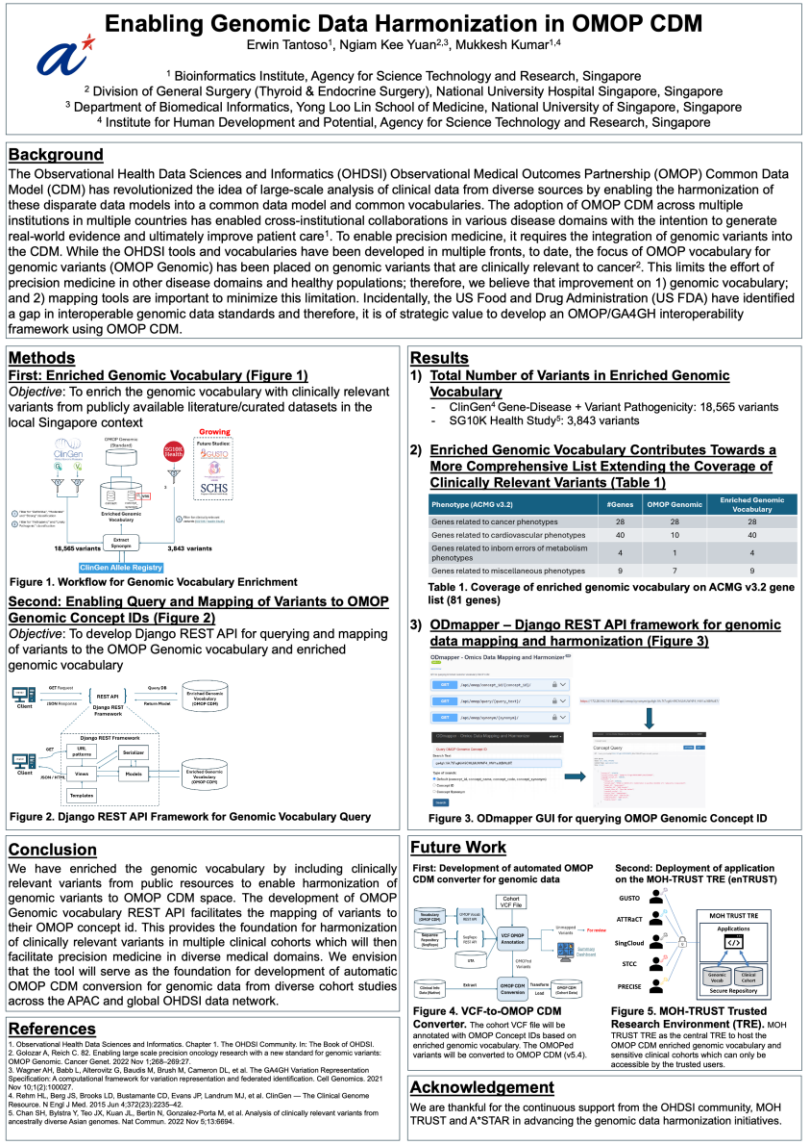
Krittaphas Chaisutyakorn, Peamboon Thomchotpong, Natpatchara Pongjirapat, Jirapat Aiamsopon, and Natthawut Adulyanukosol

#OHDSISocialShowcase This Week

Thursday

Enabling Genomic Data Harmonization in OMOP CDM

(Erwin Tantoso, Ngiam Kee Yuan, Mukkesh Kumar)



Enabling Genomic Data Harmonization in OMOP CDM

Erwin Tantoso[1], Ngiam Kee Yuan[2,3], Mukkesh Kumar[1,4]

[1] Bioinformatics Institute, Agency for Science Technology and Research, Singapore
[2] Division of General Surgery (Thyroid & Endocrine Surgery), National University Hospital Singapore, Singapore
[3] Department of Biomedical Informatics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[4] Institute for Human Development and Potential, Agency for Science Technology and Research, Singapore

Background

The Observational Health Data Sciences and Informatics (OHDSI) Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) has revolutionized the idea of large-scale analysis of clinical data from diverse sources by enabling the harmonization of these disparate data models into a common data model and common vocabularies. The adoption of OMOP CDM across multiple institutions in multiple countries has enabled cross-institutional collaborations in various disease domains with the intention to generate real-world evidence and ultimately improve patient care[1]. To enable precision medicine, it requires the integration of genomic variants into the CDM. While the OHDSI tools and vocabularies have been developed in multiple fronts, to date, the focus of OMOP vocabulary for genomic variants (OMOP Genomic) has been placed on genomic variants that are clinically relevant to cancer[2]. This limits the effort of precision medicine in other disease domains and healthy populations; therefore, we believe that improvement on 1) genomic vocabulary; and 2) mapping tools are important to minimize this limitation. Incidentally, the US Food and Drug Administration (US FDA) have identified a gap in interoperable genomic data standards and therefore, it is of strategic value to develop an OMOP/GA4GH interoperability framework using OMOP CDM.

Methods

First: Enriched Genomic Vocabulary (Figure 1)

Objective: To enrich the genomic vocabulary with clinically relevant variants from publicly available literature/curated datasets in the local Singapore context

Figure 1. Workflow for Genomic Vocabulary Enrichment

Second: Enabling Query and Mapping of Variants to OMOP Genomic Concept IDs (Figure 2)

Objective: To develop Django REST API for querying and mapping of variants to the OMOP Genomic vocabulary and enriched genomic vocabulary

Figure 2. Django REST API Framework for Genomic Vocabulary Query

Results

1) **Total Number of Variants in Enriched Genomic Vocabulary**
   - ClinGen[4] Gene-Disease + Variant Pathogenicity: 18,565 variants
   - SG10K Health Study[5]: 3,843 variants

2) **Enriched Genomic Vocabulary Contributes Towards a More Comprehensive List Extending the Coverage of Clinically Relevant Variants (Table 1)**

Phenotype (ACMG v3.2)	#Genes	OMOP Genomic	Enriched Genomic Vocabulary
Genes related to cancer phenotypes	28	28	28
Genes related to cardiovascular phenotypes	40	10	40
Genes related to inborn errors of metabolism phenotypes	4	1	4
Genes related to miscellaneous phenotypes	9	7	9

Table 1. Coverage of enriched genomic vocabulary on ACMG v3.2 gene list (81 genes)

3) **ODmapper – Django REST API framework for genomic data mapping and harmonization (Figure 3)**

Figure 3. ODmapper GUI for querying OMOP Genomic Concept ID

Conclusion

We have enriched the genomic vocabulary by including clinically relevant variants from public resources to enable harmonization of genomic variants to OMOP CDM space. The development of OMOP Genomic vocabulary REST API facilitates the mapping of variants to their OMOP concept id. This provides the foundation for harmonization of clinically relevant variants in multiple clinical cohorts which will then facilitate precision medicine in diverse medical domains. We envision that the tool will serve as the foundation for development of automatic OMOP CDM conversion for genomic data from diverse cohort studies across the APAC and global OHDSI data network.

Future Work

First: Development of automated OMOP CDM converter for genomic data

Figure 4. VCF-to-OMOP CDM Converter. The cohort VCF file will be annotated with OMOP Concept IDs based on enriched genomic vocabulary. The OMOPed variants will be converted to OMOP CDM (v5.4).

Second: Deployment of application on the MOH-TRUST TRE (enTRUST)

Figure 5. MOH-TRUST Trusted Research Environment (TRE). MOH TRUST TRE as the central TRE to host the OMOP CDM enriched genomic vocabulary and sensitive clinical cohorts which can only be accessible by the trusted users.

References

1. Observational Health Data Sciences and Informatics. Chapter 1. The OHDSI Community. In: The Book of OHDSI.
2. Golozar A, Reich C. 82. Enabling large scale precision oncology research with a new standard for genomic variants: OMOP Genomic. Cancer Genet. 2022 Nov 1;268–269:27.
3. Wagner AH, Babb L, Alterovitz G, Baudis M, Brush M, Cameron DL, et al. The GA4GH Variation Representation Specification: A computational framework for variation representation and federated identification. Cell Genomics. 2021 Nov 10;1(2):100027.
4. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen — The Clinical Genome Resource. N Engl J Med. 2015 Jun 4;372(23):2235–42.
5. Chan SH, Bylstra Y, Tao JX, Kuan JL, Bertin N, Gonzalez-Porta M, et al. Analysis of clinically relevant variants from ancestrally diverse Asian genomes. Nat Commun. 2022 Nov 5;13:6694.

Acknowledgement

We are thankful for the continuous support from the OHDSI community, MOH TRUST and A*STAR in advancing the genomic data harmonization initiatives.

#OHDSISocialShowcase This Week

Friday

Personalised prediction of chronic kidney disease progression in patients with chronic kidney disease stages 3-5: a multicentre study using the machine learning approach

(Trung Toan Duong, Minh Tri Nguyen, Chia-Te Liao, Ngoc Hoang Le, Thanh Phuc Phan, Chih-Wei Huang, Jason C. Hsu, Alex P.A Nguyen)

OHDSI TAIWAN

Personalised prediction of chronic kidney disease progression in patients with chronic kidney disease stages 3–5: a multicentre study using the machine learning approach

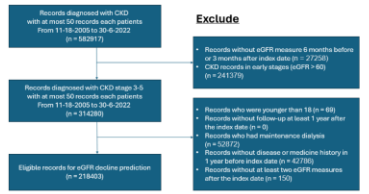
Authors: Minh Tri Nguyen 1, Trung Toan Duong 2 3, Jason C. Hsu 4 5 6 7, Alex PA. Nguyen 5 6 7 8 *
Affiliation: 1. Institute of Data Science, College of Management, Taipei Medical University, New Taipei City, Taiwan; 2. International PhD program of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; 3. Cho Ray Hospital, Ministry of Health, Ho Chi Minh City, Vietnam; 4. International Ph.D. Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan; 5. Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei Medical University, Taipei, Taiwan; 6. Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan; 7. Research Center of Data Science on Healthcare Industry, College of Management, Taipei Medical University, Taipei, Taiwan; 8. Graduate Institute of Data Science, College of management, Taipei Medical University, Taipei, Taiwan; * Corresponding author.

Background

Chronic kidney disease (CKD) progression, calculated through estimated glomerular filtration rate (eGFR), is an important measurement to maintain patients' health and prevent other complications such as hypertension. This study aims to train and evaluate machine learning (ML) models for CKD progression prediction within 1-year timeframe among patients with CKD stages 3–5.

Methods

Figure 1: Enrollment process



Electronic health record data from the Taipei Medical University clinical research database (TMUCRD) were used for the retrospective dataset. The TMUCRD database has been converted into the OMOP Common Data Model. Our cohort includes patients with CKD stages 3–5 between 2005 and 2021, with a maximum follow-up of 1 year. Patient demographics, comorbidities, medications, and laboratory data were used to develop models. We divide the dataset into training and testing sets and evaluate the model with 5-fold cross-validation to guarantee robust performance. Area under the curve (AUC), sensitivity, specificity, and accuracy were employed as evaluation metrics.

Results

After the enrollment process, 11488 patients were included in model training. The Light Gradient-Boosting Machine model achieved the best results in predicting 5% and 25% eGFR decline, with AUC values of 0.76 and 0.82, respectively. Based on SHAPLEY value calculations, important features that contributed to the prediction's results included baseline eGFR, eGFR slope, and blood urea nitrogen (BUN).

Table 1: Features importances in predicting 5% eGFR decline (left figure) and 25% eGFR decline (right figure)

Conclusions

This study demonstrates the effectiveness of applying an ML approach for predicting CKD progression for patients with CKD stages 3–5. These findings can be used for personalized prevention and treatment strategies and discovering patients at risk for CKD decline. We plan to expand this study into a multicenter study in the future.

Contact: tri.m.nguyen.1999@gmail.com

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