Welcome to OHDSI, pt 2

OHDSI Community Call
Oct. 31, 2023 • 11 am ET
# Upcoming Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct. 31</td>
<td>Welcome to OHDSI, part 2</td>
</tr>
<tr>
<td>Nov. 7</td>
<td>Meet The Titans</td>
</tr>
<tr>
<td>Nov. 14</td>
<td>Collaborator Showcase Honorees</td>
</tr>
<tr>
<td>Nov. 21</td>
<td>Showcase Software Demos</td>
</tr>
<tr>
<td>Nov. 28</td>
<td>TBA</td>
</tr>
<tr>
<td>Dec. 5</td>
<td>Recent Publications</td>
</tr>
<tr>
<td>Dec. 12</td>
<td>How Did OHDSI Do This Year?</td>
</tr>
<tr>
<td>Dec. 19</td>
<td>Holiday-Themed Goodbye to 2023!</td>
</tr>
</tbody>
</table>
Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?
Congratulations to the team of Berta Raventós, Martí Català, Mike Du, Yuchen Guo, Adam Black, Ger Inberg, Xintong Li, Kim López-Güell, Danielle Newby, Maria de Ridder, Cesar Barboza, Talita Duarte-Salles, Katia Verhamme, Peter Rijnbeek, Daniel Prieto Alhambra, and Edward Burn on the publication of IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model in *Pharmacoepidemiology & Drug Safety*. 
Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and Deidentified Clinical Dataset

**Best Community Contribution Honorees!**

**Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and Deidentified Clinical Dataset**

**Presenter:** Stephanie Hong

**INTRO:**

The National COVID Cohort Collaborative (N3C) dataset is a platform that provides researchers access to COVID-related patient EHR data in OMOP COM format. It is the largest longitudinal registry of COVID-related patient EHR data in the US. CMS claims data is also transformed into OMOP COM format using code map service. N3C COVID patient cohort is now linked to CMS claims data via Privacy-Preserving Linkage (PRL). As a result, N3C EHR datasets in OMOP COM format are enriched with the following additional CMS claims data:

- Visit/Visit level (inpatient, outpatient, emergency, or other)
- Drug prescription
- Part B
- Long-term care
- Durable medical equipment
- Other services

**METHODS:**

1. CMS claims files in wide format are parsed and pivoted into long format. The clinical concept codes are organized into a condensed format per patient per visit for efficient data transformation.
2. The condensed dataset is then used by the Code Map service to generate the clinical concept translation table. The unified version of the OMOP vocabulary tables is used to perform the translation from the source code to OMOP concept ID.
3. The generated code map service table is used as input in the data pipeline to transform the CMS claims datasets into OMOP COM format.
4. The data pipeline is built to generate CMS dataset in OMOP COM format with N3C PRL linkage.
5. N3C data is enriched with CMS data per PRL linked N3C patient. In cases where N3C person ID is duplicated, a Global ID is provided for each.

**RESULT:**

- All sites are participating in N3C.
- All sites are participating in N3C CMS-PRL linkage
- N3C CMS-patient = 1,938,841
- N3C-PRL linked CMS-patients = 839,999
- Total rows of data in N3C CMS-PRL dataset
- CMS-patients = 5,962,207
- N3C dataset matched by CMS
- Among the PRL-linked patient, in average, additional data is available from CMS.

**Data Standards**

A timeline with no gaps, either overlapping or contiguous, is used to construct the visit approach.

**Contact:** shang17@jh.edu

**Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and Deidentified Clinical Dataset**

(Stephanie Hong, Thomas Richards, Benjamin Amor, Tim Schwab, Philip Sparks, Maya Choudhury, Saad Ljazouli, Peter Leese, Amin Manna, Christophe Roeder, Tanner Zhang, Lisa Eskenazi, Bryan Laraway, James Cavallon, Eric Kim, Shijia Zhang, Emir Amaro Syallendra, Shawn O’Neil, Davera Gabriel, Sigfried Gold, Tricia Francis, Andrew Girvin, Emily Pfaff, Anita Walden, Harold Lehmann, Melissa Haendel, Ken Gersing, Christopher G Chute)
Generating Synthetic Electronic Health Records in OMOP using GPT

(Chao Pang, Xinzhuo Jiang, Nishanth Parameshwar Pavinkurve, Krishna S. Kalluri, Elise L. Minto, Karthik Natarajan)

Methods

Research

Best Community Contribution Honorees!

Generating Synthetic Electronic Health Records in OMOP using GPT

Chao Pang, Xinzhuo Jiang, Nishanth Parameshwar Pavinkurve, Krishna S. Kalluri, Elise L. Minto, Karthik Natarajan

Columbia University Irving Medical Center, Department of Biomedical Informatics

Background

This work focuses on synthetic data generation and demonstrates the capability of training a GPT model using a patient representation derived from CDHR-BERT, enabling the generation of patient sequences that can be seamlessly converted to the OMOP data format in a bidirectional manner.

Current approach (Bag of Words + GAM Model)

Use cases of synthetic EHR data:
- Personalized epidemic simulation
- Test development
- Clinical validation
- Training and education
- Comprehensive evaluation of new care pathways

Methods – Framework

OMOP

Methods – Patient Representation and OMOP Converter

Patient Representation and OMOP Converter

Context and patient representation of contexts inside OMOP tables (convertable)

OMOP Converter

OMOPs

Result

Concept Prevalence

Co-occurrence Metric

Machine Learning Performance Metrics

Conclusions

- First framework generated longitudinal synthetic EHR data using OMOP CDM
- Designed an innovative patient representation by incorporating temporal information which allowed for an accurate reconstruction of patient medical timelines as compared to state-of-the-art methods
- Comprehensive evaluation procedures showed that the synthetic data preserved the fundamental patient characteristics of the real population

Contact: CDHR-BERT@data.ccm.columbia.edu
Best Community Contribution Honorees!

GUSTO Data Vault: Laying the foundations for an open science system with OMOP Data Catalogue

(Cindy Ho, Li Ting Ang, Maisie Ng, Hang Png, Shuen Lin Tan, Estella Ye, Sunil Kumar Raja, Mengling Feng, Johan G Eriksson, Mukkesh Kumar)

GUSTO OMOP Data Catalogue lays the foundations for developing cross-study OMOP Data Catalogues expanded across APAC and global OHDSI data partners, enabling database level characterizations.

Open-Source Development
Patient’s outcomes after endoscopic retrograde cholangiopancreatography (ERCP) using reprocessed duodenoscope accessories: a descriptive study using real-world data


Background
- ERCP: Significant impact on management and prognosis of biliary and pancreatic diseases
- Concerns related to duodenoscope-related infections due to material reprocessing

Study objectives using an OMOP CDH harmonized dataset from Brazil:
- To compare the % of readmissions post-ERCP between single-use (NSUG) and non-single-use (NSUG) institutions

Methods
- Data source: Brazilian national administrative database (DHFASUS), including the Hospital and Ambulatory Information Systems. A deterministic linkage algorithm was developed to connect both datasets.
- Inclusion and exclusion criteria:
  - Patients with a history of cancer
  - ERCP procedures, excluding due to nes, acute pancreatitis, or cholangitis
  - Readmission within 30 days
  - Causes for readmission: nes, acute pancreatitis, or cholangitis

Results

| NSUG | NSUG
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total readmitted patients</td>
<td>660</td>
</tr>
<tr>
<td>NSUG readmitted patients</td>
<td>877</td>
</tr>
</tbody>
</table>

Mean age (y): NSUG: 55.0 (19.0); NSUG: 55.0 (17.9). NSUG: single-use group; NSUG: non-single-use group. NSUG: non-single-use group. Readmission NSUG: 4.8% (4.8) NSUG: 2.9% NSUG: 65% higher.

Conclusions
- We found a greater % of readmission of patients following ERCP procedures in the NSUG institutions compared to those observed in the NSUG institutions.
- Limitations: unbalanced number and geographical distribution of NSUG and NSUG institutions, descriptive analysis, and no adjustment for potential confounders.
- Next steps: estimation study, controlling for potential confounders and dealing with unbalanced data.
- Clinical importance: advance the understanding of material reprocessing implications and to inform clinical decision-making and optimal practices for ERCP management.

Identification of NSUG and NSUG hospitals:
- 3 NSUG institutions: one institute from the Northeast and two from the Midwest of Brazil
- 15 NSUG institutions: twelve institutions from the Northeast, two from the North, and one from the Southeast of Brazil

Statistical analysis: NSUG

Clinical Applications
OHDSI Shoutouts!

Any shoutouts from the community? Please share and help promote and celebrate OHDSI work!

Do you have anything you want to share? Please send to sachson@ohdsi.org so we can highlight during this call and on our social channels. Let’s work together to promote the collaborative work happening in OHDSI!
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
## Upcoming Workgroup Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Time (ET)</th>
<th>Meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday</td>
<td>12 pm</td>
<td>Common Data Model Vocabulary Subgroup</td>
</tr>
<tr>
<td>Wednesday</td>
<td>2 am</td>
<td>Methods Research</td>
</tr>
<tr>
<td>Wednesday</td>
<td>7 am</td>
<td>Medical Imaging</td>
</tr>
<tr>
<td>Wednesday</td>
<td>8 am</td>
<td>Psychiatry</td>
</tr>
<tr>
<td>Thursday</td>
<td>1 pm</td>
<td>OMOP CDM Oncology Vocabulary/Development Subgroup</td>
</tr>
<tr>
<td>Thursday</td>
<td>7 pm</td>
<td>Dentistry</td>
</tr>
<tr>
<td>Friday</td>
<td>9 am</td>
<td>GIS – Geographic Information System General</td>
</tr>
<tr>
<td>Friday</td>
<td>11 am</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Healthcare Systems Interest Group</td>
</tr>
<tr>
<td>Monday</td>
<td>6 pm</td>
<td>OMOP &amp; FHIR</td>
</tr>
<tr>
<td>Tuesday</td>
<td>9 am</td>
<td>ATLAS</td>
</tr>
<tr>
<td>Tuesday</td>
<td>10 am</td>
<td>Common Data Model</td>
</tr>
</tbody>
</table>
Automated Concept Mapping System for OMOP using Vector Representations and Cross-hospital Mapping

(Martina Carres, Gabriel Maeztu, Mónica Arrúe)
Mapping Dental Use Cases to the OMOP-CDM: Vocabulary and Common Data Model Evaluation

(Tuesday)

**Mapping Dental Use Cases to the OMOP-CDM: Vocabulary and Common Data Model Evaluation**

**Presenters:** Robert Koski, Gopikrishnan Chandrasekharan, William D. Duncan

The OMOP-CDM has the potential to elevate observational research in dentistry.

The Dentistry Workgroup is leading the effort.

**Methods**

1. Develop a use case
2. Create synthetic data with Synthea
3. Map the data to the OMOP-CDM
4. Explore the mapping and gain key insights

Amongst patients that received a posterior composite restoration, how many patients experienced restoration failure within five years?

**Additional concepts for the use case are currently being mapped:**

- Billing codes for dental procedures
- Caries code
- Dental devices
- Tooth-centric (anatomical site) mapping
- Survival of tooth

**Future Use Cases**

- Characterization of Obstructive Sleep Apnea treatments
- Care pathways for special needs dental patients
- Periodontal disease management

Robert Koski, Gopikrishnan Chandrasekharan, William D. Duncan
Bayesian sparse logistic regression models exhibit smaller bias and sparser models than L1-regularized models by incorporating prior information on our parameters of interest.
Comparing Penalization Methods for Linear Models on Large Observational Health Data

(Egill A. Fridgeirsson, Ross D. Williams, Peter Rijnbeek, Marc Suchard, Jenna Reps)

Title: Comparing Penalization Methods for Linear Models on Large Observational Health Data

Presented by: Egill A. Fridgeirsson

Introduction:
- We explore the impact of various penalty techniques on predictive model performance aiming to create more robust and generalizable models for better healthcare outcomes.

Methods:
1. We study L1 (LASSO), L2 (Ridge), L1L2 (ElasticNet) and L0 (iterative hard thresholding and broken adaptive ridge) penalized logistic regression models.
2. We study 21 outcomes occurring in the year after patients start pharmacovigilance treatment for major depressive disorder.
3. We use five databases (IBM CCAE, IBM NDCI, IBM NDCD, Optum EHR and Optum Claims).
4. We evaluate discrimination (AUC) and calibration (C-index) and use critical difference diagrams to investigate significant differences in ratios between methods.

Results:

Critical difference diagram (AUC)

LASSO is the best penalization method for discrimination. For parsimony and calibration L0 penalties are best.
Prediction of End Stage Renal Disease in Patients with Type 2 Diabetes Mellitus

(Methods)

1. Data source: Electronic Health Record (EHR) data from five secondary or tertiary hospitals standardized to Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) version 5.1.2.
2. Included Patients: Newly diagnosed with T2DM at a single hospital and had no history of ESRD.
3. Outcome: Occurrence of ESRD defined as eGFR<15 ml/min or transplantation or dialysis.
4. Model development:
   1. Generate several models using several machine learning algorithms.
   2. Select 3 algorithms which showed best performance.
   3. Hyperparameter tuning of each of these algorithms.
5. Model Validation:
   - Internal validation: using test set of the data that was used to develop model (25% training set / 25% test set).
   - External validation: using external data set.
   - Model Performance Evaluation:
     - Area Under the Receiver Operating Characteristic Curve (AUROC).
     - Area Under the Precision-Recall Curve (AUPRC).
     - Positive Predictive Value and Accuracy at different threshold level.
     - Calibration Curve.

(out of a total of 4,240 variables, 270 were included as predictors)

- Measurement above creatinine and urea and nitrogen in serum plasma / Condition of benign hypertension, renal impairment history were considered to be important.)
Collaborator Spotlight: Atif Adam

Dr. Atif Adam is a systems scientist and researcher boasting over a decade of diversified experience spanning academia, industry, and entrepreneurial ventures. He attained his doctorate in Health Systems Science and Computational Epidemiology. In addition, Dr. Adam completed his clinical training in Internal Medicine and secured master’s degrees in Health Policy and Spatial Epidemiology.

His research probes the nuanced relationships between chronic cardiometabolic diseases, mental health, cognitive aging, and health disparities. During his academic appointments at institutions such as Johns Hopkins and Harvard, Dr. Adam pioneered innovative simulation frameworks for cardiometabolic diseases and rigorously evaluated care pathways for the most vulnerable populations. To this end, he employs advanced statistical, geospatial, and systems modeling methodologies. Transitioning into the digital health space, Dr. Adam co-founded and assumed the role of Chief R&D Officer for the digital mental health startup, Rose Health. In this capacity, he harnessed large-scale data and sensor-based models to curate evidence-based digital solutions, poised for proactive patient monitoring.

In his present role as the Associate Director of Epidemiology at IQVIA, Dr. Adam channels his expertise to spearhead transformative real-world evidence (RWE) initiatives. Within the OMOP team at IQVIA, he merges his deep understanding of health systems, an unwavering commitment to health equity, and knowledge in data science to develop and deliver robust RWE studies at scale. Beyond mere discovery, Dr. Adam is ardently devoted to mentorship, nurturing, and guiding the forthcoming generation of health scientists towards a more informed and equitable healthcare horizon. He discusses his career journey, challenges in health equity and how OHDSI is dealing with them, advice for newcomers in OHDSI, and plenty more in the latest Collaborator Spotlight.

You joined the scientific review committee for the global symposium this year, so what stood out about the variety of research you reviewed for the collaborator showcase?

Joining the scientific review committee for this year's global symposium was a profound journey through cutting-edge developments in health research. As I sifted through the collaborator showcase submissions, it wasn't just the breadth of topics that stood out but also the depth and diversity of approaches employed.

Beyond addressing intriguing questions and hypotheses, there was a noticeable emphasis on methodological innovations. Abstracts showcased a wide range of methods-oriented work, shedding light on enhancements to traditional models and introducing avant-garde techniques. This focus on refining and redefining methods signifies a maturing field, one that's continuously introspecting and evolving.

Several submissions highlighted new collaborations, bringing together multifaceted teams with varied expertise. These collaborations spanned regions and bridged disciplines, underscoring the interdisciplinary nature of modern health research (including Generative AI). Adding new data assets enriched the research landscape, allowing for multifaceted analyses and richer insights.

For me, the collaborator showcase was a microcosm of the future of OHDSI and RWE research. From innovative hypotheses to pioneering methods and new collaborations to the application of advanced models, it was a vivid testament to the dynamism and promise of the global health OHDSI community.

Atif Adam (right) was a member of the 2023 Scientific Review Committee, and he moderated a series of lightning talks at the Global Symposium.
Global Symposium Homepage

2023 OHDSI Symposium
Oct. 20-22 • East Brunswick, New Jersey

The 2023 OHDSI Global Symposium welcomed more than 440 of our global collaborators together for three days of sharing research, forging new connections and pushing forward together the OHDSI mission of improving health by empowering a community of collaborators to generate the evidence that promotes better health decisions and better care.

This page will be home to all materials from the global symposium. Check back for all video presentations from the event.

#JoinTheJourney #OHDSI2023

Video coming soon

State of the Community

Various leaders within OHDSI shared a presentation on the state of the community, with specific focuses on data standards, vocabulary enhancements and open-source development. Speakers included:

George Hripcsak, Columbia University
Clair Blacketer, Johnson & Johnson
Alexander Davydov, Odysseyus Data Services
Katy Sadowski, Boehringer Ingelheim
Peter Rijnbeek, Erasmus MC
Mengling Momin' Feng, National University of Singapore

State of the Community Slides

2023 Global Collaborator Showcase
Observational Data Standards & Management

1. FindOMOP – a population-based data network (Javier Grando-Teboneno, Punto Kositweng, Pia Tjäder, Sampo Kuikurinen, Guenter Klingstedt, Anna Hammers, Peresepheh Doshi, Oscar Brück, Leena Hakonarson, Anna Kallia, Marco Hautarthen, Toni Mikkola, Marianna Niami, Pasi Rikala, Simo Rytänen, Anna Manuennern, Antti Vuori, Joonee Demmert, Eric Fey, Tenho Kiiro, Aimo Vihma, Tarja Laitinen, Antti Perkkö)


3. Augmenting the National COVID Cohort Collaborative (N3): Dataset with Medscape and Medscape Meds and Clinical Dataset (Stephanie Hong, Thomas Richards, Benjamin Amor, Tim Schwab, Phil Sparks, Maya Chouhtouy, Said Lazouz, Peter Leen, Amin Mamm, Christophe Roeder, Tanner Zhang, Lisa Eskinian, Bryan Laceray, James Cavallion, Eric Kim, Shija Zhang, Emil Amaro Sambelina, Shawn O'Neil, Daisuke Kubota, Rydghold Gudin, Tricia Francis, Andrew Grin, Emily Pfaff, Antoni Walden, Marcel Lehmann, Melissa Haemel, Ken Gersing, Christopher G. Chute)

4. Integrating Clinical and Laboratory research data using the OMOP CDQI. (Edward A. Frankenberg, Chun Yang, Vassilis Madeya Mela Harkness, Ayssal Goodyear)

5. Development of Medical Imaging Data Standardization for Imaging-Based Observational Research. OMOP Common Data Model Extension (Wale Yeon Park, Kyuey Jeon, Tari Sipilä Schmidt, Haraldins Kondrashok, Song Chan You, Paul Nagy)


7. Conversion of a Krooitra Precision Medicine Center into a Common Data Model: A Case Study (Dzachary Wang, Will Kelly, Paul Nagy, Christopher A. Musick)

8. Implementing a common data model in endometriology: Congruence of general gynec examination measures to standard OMOP concepts across two major EHR systems (Justin C. Quinn, William Hartmann, Cindy X. Cai, Sally L. Bader, Brian C. Toy)

9. End-to-End Data Quality Management. Integrating Capture and Cleaning Modes to the Data Quality Dashboard (Frank DeFalco, Clair Blacketer)


11. A Trial Vocabulary by the OMOP CDQI (Makoto Tsuboiyama, Polina Tseltsa, Tetsuya Nishimura, Andrew Williams, Cenys Rudak, Max Ved, Inês Aguiar)

12. Challenges and opportunities in adapting OMOP CDQI in Brazilian healthcare: a report from Hospital Israelita Albert Einstein (Maria Alahani, Uly Alencar, Patricia Ranieri, Maíra Terra de Lima Freitas, Diego Pastro, Amanda Gomes Ribeiro, Cecília Augusto Mostizzoli, Gabriela Chuffo Tunes, Elizette Dau, Gabriela Mesquita de Souza, Soraya Yuki Astolfi, Adriana José Pereira, Edison Amaral)

13. Transforming the OhclayrPharmacology. ontology module to OMOP CDQI (Christy Cyn, Clair Blacketer)

14. Mapping Multi-Layered Omics Data to OMOP (John Mehta, Sherry Lee)

15. Development of extravascular common data model (P-CDA) - Leveraging extravascular scales (Don Yon Lee, Chiorga Kim, Rae Wong Park)

16. Brazilian administrative data for real-world research: a comparative lineage procedure and OMOP CDQI harmonization (Jessica Mayumi Masuyama, Juliana Cesar de Oliveira)

17. Integration of Clinical and Genomic Data Using the OMOP Common Data Model in a Focused Data Network in Benin (Tahiana Jatsimoro, Murat Akand, Joris Robert Vermeech, Dries Rombaut, Michel Van Speybroeck, Martine Lewi, Valerie Vandeweerd)

ohdsi.org/OHDSI2023

@OHDSI

www.ohdsi.org

#JoinTheJourney
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?
What Questions Did The Community Ask?

- What is the sustainability model?
- How do I learn more about the open problems, especially around methods development?
- Is there a dictionary of Greek named tools and what they are used for?
- Can you clarify the roles of the coordinating center(s) and whether and how members are to engage with them? For example writing grants with them?
- How to troubleshoot the install and understand the OHDSI tools?
- Would be awesome if there was a Roadmap/Guide for CDM implementation with tools associated with each step along the way. This may be a lot to ask.
- Wondering if there is collaboration that exists between OHDSI and the military (i.e., DHA - Defense Health Agency)?
- Does OHDSI have a workgroup 'need' level? Like one workgroup may have a tons of contribution and other not so much, where could one help most.
- Any logistic recommendations for members from Institutions that don’t have Microsoft accounts? It was a challenge to get connected to teams with my institutional email without that.
- Is the list of members of the OHDSI network posted anywhere?
- How can someone have a presentation for more than 8 minutes in the OHDSI US conference like Martijn and Patrick have every year?