Meet The Titans

OHDSI Community Call
Nov. 7, 2023 • 11 am ET
# Upcoming Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<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>Happy Birthday OHDSI! Where Have We Come In 10 Years, and in 12 Months?</td>
</tr>
<tr>
<td>Dec. 19</td>
<td>Holiday-Themed Goodbye to 2023!</td>
</tr>
</tbody>
</table>
Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and Deidentified Clinical Dataset

**INTRO:** The National COVID Cohort Collaborative (N3C) data hub is a platform that provides researchers access to COVID-related patient data in the OMOP (Open Biomedical Open Data Standards) format. It is the largest centralized repository of COVID-related patient EHR data in the U.S. CMS claims data is also transferred to OMOP-OMOP format using open map services. N3C COVID patient cohort is now linked to CMS claims data via Privacy Preserving Record Linkage (PPRL). As a result, N3C EHR datasets in OMOP-OMOP format are enriched with the following additional CMS claims data:

- Inpatient: Drug prescription
- Part B: Long-term care
- Durable medical equipment

**METHODS:**

1. CMS claim files in wide format are paired 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 are used to perform the translation from the source code to OMOP concept IDs.

3. The generated code map service table is used to input in the data pipeline to translate the CMS claims datasets into OMOP-OMOP format.

4. The data pipeline is built to generate CMS dataset in OMOP-OMOP format with PPRL linkage.

5. N3C data is enriched with CMS data per PPRL-linked N3C patient. In cases where N3C person IDs are duplicated, a Global ID is provided for each.

**RESULT:**

- All sites are participating in N3C.
- All sites are participating in N3C-OHR-PRL linkage.
- N3C: 81,841.79
- N3C-PRL linked CMS patients: 81,841.79
- Total rows of data in CMS: 81,841.79
- N3C dataset enriched by CMS.

**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 Syailandra, Shawn O’Neil, Davera Gabriel, Sigfried Gold, Tricia Francis, Andrew Girvin, Emily Pfaff, Anita Walden, Harold Lehmann, Melissa Haendel, Ken Gersing, Christopher G Chute

**Contact:** shpeng@ohsu.edu
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

Methods

Research

Generating Synthetic Electronic Health Records in OMOP using GPT

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

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.

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

GUSTO OMOP Data Catalogue

Methods:
- OMOP Data Catalogue makes GUSTO cohort-specific CDW fields to be discovered across the Person, Condition, Observation and Measurement selections in the global research environment.
- Metadata is described with reference to the CDW Field Concept, Codebook, Name, Subject Year, Visit Timepoint, Description and Details.

Poster Title: GUSTO 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

Our future work includes the exploration of GUSTO OMOP Data Catalogue partnerships with similar cross-study initiatives such as the OXERA OMOP Consortium. Future release of OMOP Data Catalogue Luminare Part II.

Scan to visit GUSTO Data Vault

Scan to download the abstract

Global Impact of GUSTO Data Vault

www.ohdsi.org

#JoinTheJourney
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

Objectives:
To compare the % of readmission post-ERCP between single-use (NSUG) and Non-single-use (SUG) institutions

Methods:
Data source: Brazilian national administrative database (DHFASIS), 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, extending due to severe acute pancreatitis, or cholangitis
- Readmission within 30 days
- Causes for readmission: severe acute pancreatitis, or cholangitis

Identification of SUG and NSUG hospitals:
- 3 SUG 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: NIsa

Results:
Table 1: Descriptive information of total and readmitted patients in SUG and NSUG

<table>
<thead>
<tr>
<th></th>
<th>SUG</th>
<th>NSUG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Readmitted patients</td>
<td>660</td>
<td>20</td>
</tr>
<tr>
<td>NSUG</td>
<td></td>
<td>887</td>
</tr>
<tr>
<td>%</td>
<td>30.9</td>
<td>50.0</td>
</tr>
<tr>
<td>Mean age (y)</td>
<td>55.0 (19.0)</td>
<td>55.0 (17.9)</td>
</tr>
</tbody>
</table>

Conclusion:
We found a greater % of readmission of patients following ERCP procedures in the SUG institutions compared to those observed in the NSUG institutions.

Clinical importance: advance the understanding of material reprocessing implications and to inform clinical decision-making and optimal practices for ERCP management.
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>Wednesday</td>
<td>9 am</td>
<td>Patient-Level Prediction</td>
</tr>
<tr>
<td>Wednesday</td>
<td>12 pm</td>
<td>Health Equity</td>
</tr>
<tr>
<td>Wednesday</td>
<td>2 pm</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>Thursday</td>
<td>8 am</td>
<td>India Chapter</td>
</tr>
<tr>
<td>Thursday</td>
<td>9:30 am</td>
<td>Data Network Quality</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>Medical Devices</td>
</tr>
<tr>
<td>Thursday</td>
<td>7 pm</td>
<td>Dentistry</td>
</tr>
<tr>
<td>Friday</td>
<td>9 am</td>
<td>Phenotype Development &amp; Evaluation</td>
</tr>
<tr>
<td>Friday</td>
<td>9 am</td>
<td>GIS – Geographic Information System Development</td>
</tr>
<tr>
<td>Friday</td>
<td>1 pm</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Friday</td>
<td>11 pm</td>
<td>China Chapter</td>
</tr>
<tr>
<td>Monday</td>
<td>9 am</td>
<td>Vaccine Vocabulary</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Africa Chapter</td>
</tr>
<tr>
<td>Monday</td>
<td>11 am</td>
<td>Early-Stage Researchers</td>
</tr>
<tr>
<td>Tuesday</td>
<td>9 am</td>
<td>OMOP CDM Oncology Genomic Subgroup</td>
</tr>
</tbody>
</table>
Global Symposium Homepage

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

The 2023 OHDSI Global Symposium welcomed more than 400 of our global collaborators together for three days of sharing research, forging new connections and pushing forward the OHDSI mission of improving health by empowering a community to collaborate and generalize evidence to promote better health decisions and better care.

This page will be home to all materials from the global symposium. Check back in the coming days for all videos presentations from the event.

2023 Global Collaborator Showcase
Observational Data Standards & Management

2 - FINDCMOP: a customizable data model layer Javier Garcia-Tahonera, Porto Kostokosta, Pia Tapanainen, Zampee Kulikarni, Gustav Ringqvist, Arina Hanea, Penelope Diao, Oscar Elgr, Leena Hakinlampi, Anna Kala, Marco Pelizzari, Toru Mikabuchi, Blerim Varol, Pasi Ruokala, Omar Poyhonen, Arina Velikanova, Mira Vahvanen, Joonas Dittmar, Esa Pyy, Teemu Kiila, Tapio Leppa, Kristo Pirttila
3 - OHDSI in EHR-PRIM: Successes, Challenges, and Future Directions of Joint EHR-PRIM and OHDSI Data Science and Demographic Task Forces Srinivasan Iyengar, Thomas Richards, Benjamin Amor, Tim Schwalb, Philip Suzuki, Maya Clouthier, Daniel Lupis, Peter Dacalos, Yvonne Maro, Christopher Roberge, Tanner Zhang, Lisa Estes, Brian Lamay, James Cavannon, Eric Heden, Shoa Zhang, Emil Ahsanfarzadabadi, Shahab Chelu, Denise Gabel, Sigithe Gold, Tirol France, Andrew Qimin, Emily Piask, Anna Wink, Harold Lehmann, Melissa Harris, Ken Canning, Christopher G Chute
4 - Innovations in clinical and research data sharing using the CMOP-CDM: Evidence A. Freisleraberg, Chuan Yang, Victor-von Restorfi Idies Verklaar, Alyssa Goodrich
5 - Development of Medical Imaging Data Standards for Imaging-Based Observational Research: CMOP CDM Model Extension Wha Yoon Park, Koojae Lee, Tim Sopel Schmitt, Hanminh Konopka, Sang Cheol Yu, Paul Rago
7 - Overview of a Newly Founded Medical Center in a Complex Data Model: A Case Study (Fakery Yang, William Li, Paula Rago, Christopher A McCall)

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

George Hripcsak, Columbia University
Ciaran Blackmore, Johnson & Johnson
Alexandre Ouyers, Opistos Data Services
Kay Sadowiski, Baxter International
Peter Ripjebek, Einstein MC
Mingling ‘Wanxi’ Peng, National University of Singapore

2023 Collaborator Showcase Posters & Software Demos

State of the Community Slides

Survey data from evidence can be challenging alone but is greatly rewarded through community collaboration. In this talk, we will introduce newcomers to OHDSI. Specifically, about the tools, practices, and open-source approach to evidence generation that makes the OHDSI community has developed and evolved over the past decade.

Faculty will highlight the ways community members can participate as well as receive value from the community's outputs. The course will include topics such as open community data standards – including the CMOP Common Data Model and OHDSI Standardized Vocabulary, open-source analytic tools

Tutorial: Introduction to OHDSI

ohdsi.org/OHDSI2023

www.ohdsi.org

#JoinTheJourney
November Newsletter

Community Updates

Where Have We Been?
- The 2023 OHDSI Global Symposium was held Oct. 20-22 in East Brunswick, N.J., and it included a main conference and a full weekend of activities, including the "Welcome to OHDSI" tutorial. All presentations, slides and showcase submissions from OHDSI2023 are now available on the event homepage.
- The 2023 Tian Awards were presented at the OHDSI Global Symposium. Congratulations to our 2023 honorees, who were both nominated and selected by fellow community members!
  - Data Standards: Gautham Rani and Azaa Shubshri
  - Methodological Research: Jieyi (Jesus) Tang
  - Open-Source Development: Kathy Sedivy
  - Clinical Applications: Center for Surgical Science
  - Community Leadership: Nicole Pratt
  - Community Collaboration: Cynthia Sung
  - Community Support: Gyalen Song

Where Are We Now?
- Thirty-five organizations and data sources were introduced as the original members of the OHDSI Evidence Network during the State of the Community presentation. These are among the 594 data sources mapped to the OMOP CD, and the data source workflow map on the OHDSI website.
- The latest edition of the Our Journey annual report was shared at the Global Symposium. You can now make sure you are following OHDSI on Twitter, Facebook, and Instagram, and follow the latest OHDSI news.

OHDSI Videoicast: 2023 Symposium Review

In the latest On The Journey video, Patrick Ryan and Craig Taylor reflect on the OHDSI 2023 Global Symposium, including notes from the State of the Community talk, and both the plenary (improving the reliability and scale of case validation) and panel (Lessons learned from OHDSI network studies) sessions. They also discuss the largest collaborator showcase in OHDSI history, the weekend events, and more. If video does not appear, click View this email in your browser!

mailchi.mp/ohdsi/november2023

OHDSI 2023 Focuses On Large-Scale Evidence Generation & Community Collaboration

The 2023 OHDSI Global Symposium brought together more than 430 community members from around the world for this three-day event filled with opportunities to learn, connect and form new relationships.

The main conference was held during Day 1, and featured a plenary on improving the reliability and scale of case validation, a State of the Community presentation by several leaders in the community, and a panel on lessons learned from OHDSI network studies. The collaborative showcase included a record number of posters, software demos and lightning talks, and the closing session was an interactive session on large-scale collaboration, racecape-style.

There were also two days of workshops, workshop meetings and an introduction to OHDSI tutorial. Videos and slides from all presentations and the tutorial are available on the symposium homepage. Thank you to those who both volunteered their time to make the event a success or joined us to help push forward OHDSI's mission of improving health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

OHDSI Foundation

The OHDSI Foundation is a non-profit organization that supports the mission of improving health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

Focus Areas

- Large-Scale Evidence Generation
- Community Collaboration

OHDSI Ecosystem

The OHDSI Ecosystem is a community-driven initiative to improve the reliability and scale of case validation.

OHDSI Network

The OHDSI Network is a community of researchers, practitioners and patients working together to improve health through data-driven evidence.
OHDSI at AMIA

Saturday, Nov 11, 2023
8:30 AM - 4:30 PM CST

OHDSI RWE Revolution: Igniting Data Modernization with Harmonized Standards for Cutting-Edge Health Research

In response to the 21st Century Cures Act and the growing importance of Real-world evidence (RWE), the FDA has released guidance on RWE data for regulatory decision-making. Despite RWE’s potential to improve clinical studies, challenges remain in rapidly utilizing RWE for decision-making due to the volume and diversity of real-world data, further emphasized by the COVID-19 pandemic. The Observational Health Data Science and Informatics (OHDSI) developed the Observational Medical Outcomes Partnership (OMOP) to address these challenges and ensure the quality of RWE. OMOP focuses on the development of a common data model and standardized analytics to facilitate meaningful comparisons across different RWE data sources and research studies. Building on the need to better understand RWE and OMOP, the workshop gathers leading experts from various fields to discuss three major themes: (1) understanding the origin and barriers of real-world data for healthcare research and the role that OHDSI/OMOP has played in improving the use of RWE for healthcare research; (2) showcasing the potential of RWE analysis across multiple data types with OMOP CDM; and (3) discussing the challenges and opportunities to adopt RWE for secondary use in research and development. Participants will engage with in-depth topics such as data transformation, cohort definitions, diagnostic methods, visualization techniques, and practical applications of cohort diagnostics in real-world scenarios. This event aligns closely with the broader informatics interests of the attendees. It aims to enhance their understanding of the synergies and opportunities at the intersection of OHDSI, OMOP CDM, and healthcare.

Speaker(s):
Speaker: Alf Adam, PhD, MD, MPH, Johns Hopkins Bloomberg School of Public Health
Speaker: Mul Van Zandt, MS, IQVIA
Speaker: Paul Negy
Speaker: Mengjing Feng, PhD, National University of Singapore
Speaker: Christian Reich, PhD, MD, OHDSI
Speaker: Zhen Lin, PhD, Memorial Herman Texas Medical Center

Location: Churchill A
Session Code: W62
Session Credits: 6.00
OHDSI HADES releases: SelfControlledCaseSeries 5.0.0

SelfControlledCaseSeries

SelfControlledCaseSeries is part of HADES.

Introduction

SelfControlledCaseSeries is an R package for performing Self-Controlled Case Series (SCCS) analyses in an observational database in the OMOP Common Data Model.

Features

- Extracts the necessary data from a database in OMOP Common Data Model format.
- Optionally add seasonality using a spline function.
- Optionally add age using a spline function.
- Optionally add calendar time using a spline function.
- Optionally correct for event-dependent censoring of the observation period.
- Optionally add many covariates in one analysis (e.g., all drugs).
- Options for constructing different types of covariates and risk windows, including pre-exposure windows (to capture contraindications).
- Optionally use regularization on all covariates except the outcome of interest.
OHDSI HADES releases: DeepPatientLevelPrediction 2.0.1

DeepPatientLevelPrediction

Introduction

DeepPatientLevelPrediction is an R package for building and validating deep learning patient-level predictive models using data in the OMOP Common Data Model format and OHDSI PatientLevelPrediction framework.


Features

- Adds deep learning models to use in the OHDSI PatientLevelPrediction framework.
- Allows to add custom deep learning models.
- Includes an MLP, ResNet and a Transformer
- Allows to use all the features of PatientLevelPrediction to validate and explore your model performance.

Technology
OHDSI HADES releases: DataQualityDashboard 2.5.0

DataQualityDashboard

DataQualityDashboard is part of HADES.

The goal of the Data Quality Dashboard (DQD) project is to design and develop an open-source tool to expose and evaluate observational data quality.

Introduction

This package will run a series of data quality checks against an OMOP CDM instance (currently supports v5.4, v5.3 and v5.2). It systematically runs the checks, evaluates the checks against some pre-specified threshold, and then communicates what was done in a transparent and easily understandable way.

Overview

The quality checks were organized according to the Kahn Framework which uses a system of categories and contexts that represent strategies for assessing data quality. For an introduction to the kahn framework please click here.

Using this framework, the Data Quality Dashboard takes a systematic-based approach to running data quality checks. Instead of writing thousands of individual checks, we use “data quality check types”. These “check types” are more general, parameterized data quality checks into which OMOP tables, fields, and concepts can be substituted to represent a singular data quality idea. For example, one check type might be written as
OHDSI HADES releases: CohortExplorer 0.1.0

CohortExplorer

Introduction

This software tool is designed to extract data from a randomized subset of individuals within a cohort and make it available for exploration in a 'Shiny' application environment. It retrieves date-stamped, event-level records from one or more data sources that represent patient data in the Observational Medical Outcomes Partnership (OMOP) data model format. This tool features a user-friendly interface that enables users to efficiently explore the extracted profiles, thereby facilitating applications, such as reviewing structured profiles. The output of this R-package is a self-contained R shiny that contains person-level data for review.

Warning

- Contains person level data. This package is not to be considered de-identified.
- Please do not share the output with others as it may violate protected health information.
- .ROS file in output contains PHI.
Conversion of a Myositis Precision Medicine Center into a Common Data Model: A Case Study

(Zachary Wang, Will Kelly, Paul Nagy, Christopher A Mecoli)

“OMOP-ifying” data from inside an academic institution is hard... Here’s how we did it!
Making NLP-derived data actionable within the OHDSI ecosystem

(Michael Gurley, Kyle Zollo-Venecek, Andrew Williams, Daniel Smith, Robert Miller, Vipina Kuttichi Keloth, Hua Xu)

Deposing NLP-derived data into an EHR-based OMOP instance in accordance with newly proposed conventions enables the creation of patient cohorts not possible based solely on EHR-derived data.

Michael Gurley
Kyle Zollo-Venecek
Andrew Williams
Daniel Smith
Vipina Kuttichi Keloth
Hua Xu

#OHDSISocialShowcase
This Week
TUESDAY
Title: Demonstrating Utility of the Edge Tool Suite through Clinical Trial Emulation

PRESENTER: Ruth Kurtycz

INTRO:
- Real-world data (RWD) can support repurposing approved medications for new uses. OMOP standardizes RWD, aiding research and sharing, but implementing OMOP can be resource-intensive. The Edge Tool suite streamlines data conversion. Here, we assess the utility of data converted with the Edge Tool suite for potential research using an emulation of the RECOVERY COVID-19 clinical trial.

METHODS:
1. The Edge Tool Suite enables ETL into the OMOP CDM. Propensity score matching at a 1:3 ratio was used to match patients with and without dexamethasone administration on 11 key variables (e.g., patient age).
2. A pilot healthcare site provided over 10,000 COVID-19 patient records from March 2020 to March 2022, with exclusion criteria applied to align with the RECOVERY trial.
3. Logistic regression and survival analyses were performed on the matched data to assess the impact of Dexamethasone on 28-day mortality.

RESULTS:
- The analysis found that dexamethasone reduced 28-day mortality in COVID-19 patients receiving oxygen alone or mechanical ventilation, aligning with the RECOVERY trial. However, without oxygen support stratification, this result was not confirmed.

We demonstrate the utility of RWD from the Edge Tool Suite to replicate findings of a clinical trial for COVID-19. Results indicate approach has potential to be used to assess the efficacy of treatments for emerging diseases.

WEDNESDAY

Demonstrating Utility of the Edge Tool Suite through Clinical Trial Emulation

(Ruth Kurtycz, Wesley Anderson, Allan J. Walkey, Kerry A. Howard, Smith F. Heavner)
THURSDAY

OHDSI Network Study Execution Framework and Templating

(Ben Martin, Cindy Cai, Asieh Golozar, Paul Nagy)

INTRO:
- To enhance robustness and reliability of OHDSI network studies, it is essential to define sideline key steps and status indicators of study development and execution.

OBJECTIVES:
- Define the key process steps common amongst network studies, to improve reproducibility and reproducibility and help set expectations for researchers looking to engage in network studies.
- Develop standardized human and computer readable indicators of network study status, progress, and needs.

METHODS:
- Consensus from individuals with experience in leading OHDSI network studies demarcated nine fundamental stages that all network studies must progress through towards completion.
- A standard set of human readable and computable data artifacts attributed to each network study are derived from the delimitation of these stages and from existing documentation of completed network studies.

RESULTS:
- Figure 1: nine stages of a network study were identified as, in order: protocol development, data diagnostics, phenotype development, phenotype evaluation, analysis specifications, network execution, study diagnostics, evidence synthesis, and results evaluation.
- Table 1: a proposed set of data artifacts for study progress monitoring and facilitation.

CONCLUSION:
- OHDSI network studies share a great degree of common methodology and challenges. Laying out the key steps with a common framework, providing clarity and direction through each of these common stages, and identifying key information for monitoring progress amongst the community will facilitate progression and use shared experience to overcome repetitive challenges. The framework of network study stages and set of study progress artifacts proposed here needs to be refined and internalized by the OHDSI community at large.

Table 1. Data Artifacts for Network Study Monitoring

<table>
<thead>
<tr>
<th>Study Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRB materials are sufficient for review</td>
<td>[No, Yes]</td>
</tr>
<tr>
<td>Cohort definition available</td>
<td>[No, Yes]</td>
</tr>
<tr>
<td>Data partner recruitment status</td>
<td>[Not Ready, Open, Closed]</td>
</tr>
<tr>
<td>Deadline for adding new data partners</td>
<td>[MM/DD/YYYY]</td>
</tr>
<tr>
<td>Statistical partner recruitment status</td>
<td>[Not Ready, Open, Closed]</td>
</tr>
<tr>
<td>Deadline for adding new statistician partners</td>
<td>[MM/DD/YYYY]</td>
</tr>
<tr>
<td>Clinical domain partner recruitment status</td>
<td>[Not Ready, Open, Closed]</td>
</tr>
<tr>
<td>Deadline for adding new clinical domain partners</td>
<td>[MM/DD/YYYY]</td>
</tr>
</tbody>
</table>

@OHDSI
www.ohdsi.org
#OHDSISocialShowcase
This Week

#JoinTheJourney
Using Contrastive Principal Component Analysis to Identify Post-acute Sequelae of SARS-CoV-2 Infection Subphenotypes: an EHR-Based Cohort from the RECOVER Program

(Xiaokang Liu, Yishan Shen, Naimin Jing, Christopher B. Forrest, Yong Chen)

FRIDAY

Using Contrastive Principal Component Analysis to Identify Post-acute Sequelae of SARS-CoV-2 Infection Subphenotypes: an EHR-Based Cohort from the RECOVER Program

**Background**

- The post-acute sequelae of SARS-CoV-2 infection (PASC), known as "long COVID", refers to a range of persistent or new symptoms that emerge after the acute phase of COVID-19 infection. These symptoms can endure for weeks or months following the initial infection and can affect various body systems. PASC symptoms can vary widely between individuals, which brings challenges to the diagnosis and treatment of PASC patients.
- To gain more understanding of dominant symptom co-occurrence patterns of PASC and develop effective treatments, identifying subtypes (also known as subphenotypes) of PASC is of great interest to both health care providers and patients.

- Since the highly heterogeneous spectrum of PASC clinical features can overlap with features of other diseases, the subphenotypes identified by traditional clustering methods may not be specific to PASC.
- With electronic health records (EHR) for both COVID-19 test-positive and test-negative patients extracted from the PEDSnet COVID-19 Database, we applied a contrastive principal component analysis method (cPCA) to help derive PASC subphenotypes for children. This study aims to provide more insights into PASC and facilitate tailored interventions for affected children.

**Method: Contrastive Principal Component Analysis**

- **Input:** target dataset \(X_t\), background dataset \(Y_b\).
- **Target:** identify prominent trends that are specific to a target dataset, which is of the main interest to the researchers, relative to a comparison background dataset.
- **Method:** calculate variance-covariance matrices \(\Sigma_t\) and \(\Sigma_b\). Then, the contrastive projection directions are the vectors \(v = \Sigma_t^{-\frac{1}{2}}(\Sigma_t - \Sigma_b)v\) where \(\lambda\) determines the desired contrast level.
- **Output:** subspaces that capture a significant amount of variation within the target data, while exhibiting minimal variation in the background. The features within this subspace encapsulate structures specific to the target data.
- **Clustering:** project the target data onto this subspace and use k-means to discover the clustering patterns unique to the target data relative to the background.

**Application**

- **PASC subphenotyping analysis:**
  - Target dataset: EHR of COVID-19 test-positive patients;
  - Background dataset: EHR of COVID-19 test-negative patients, contains information regarding general disease patterns not specific to PASC.

**References**

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?
Meet The Titans