March Madness & April Olympians

OHDSI Community Call
March 12, 2024 • 11 am ET
## Upcoming Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar. 12</td>
<td>March Madness &amp; April Olympians</td>
</tr>
<tr>
<td>Mar. 19</td>
<td>NO MEETING</td>
</tr>
<tr>
<td>Mar. 26</td>
<td>Recent OHDSI Publications</td>
</tr>
<tr>
<td><em>coming in April</em></td>
<td><em>CDM Month (hear more about this during today’s call)</em></td>
</tr>
</tbody>
</table>
March 26: Recent OHDSI Publications

**Tathagata Bhattacharjee • University of London**
INSPIRE datahub: a pan-African integrated suite of services for harmonising longitudinal population health data using OHDSI tools • *Frontiers in Digital Health*

**Sulev Resiberg • University of Tartu**
Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned • *JAMIA Open*

**Fan Bu • University of Michigan**
Bayesian safety surveillance with adaptive bias correction • *Statistics in Medicine*

**Jen Wooyeon Park • Johns Hopkins University**
Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension • *Journal of Imaging Informatics in Medicine*

**Christian Reich • Odysseus**
OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization • *JAMIA*
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>Registry</td>
</tr>
<tr>
<td>Tuesday</td>
<td>6 pm</td>
<td>Eyecare &amp; Vision Research</td>
</tr>
<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>3 pm</td>
<td>Vulcan/OHDSI</td>
</tr>
<tr>
<td>Wednesday</td>
<td>2 pm</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>Thursday</td>
<td>9:30 am</td>
<td>Network Data Quality</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>Medical Devices</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>Strategus HADES Subgroup</td>
</tr>
<tr>
<td>Thursday</td>
<td>7 pm</td>
<td>Dentistry</td>
</tr>
<tr>
<td>Friday</td>
<td>10 am</td>
<td>GIS – Geographic Information System</td>
</tr>
<tr>
<td>Friday</td>
<td>10:30 am</td>
<td>Open-Source Community</td>
</tr>
<tr>
<td>Friday</td>
<td>11:30 am</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Friday</td>
<td>11:30 am</td>
<td>Steering Group</td>
</tr>
<tr>
<td>Monday</td>
<td>10 pm</td>
<td>Africa Chapter</td>
</tr>
<tr>
<td>Monday</td>
<td>11 am</td>
<td>Data Bricks User Group</td>
</tr>
<tr>
<td>Monday</td>
<td>2 pm</td>
<td>Electronic Animal Health Records</td>
</tr>
</tbody>
</table>
Evidence Network WG Meeting
14 March 2024 10:30am EST

Methods for Mapping GIS Data

Kyle Zollo-Venecek
DevCon 2024: April 26, 9 am-3 pm ET

The third annual OHDSI DevCon will be held virtually on Friday, April 26, from 9 am-3 pm ET.

Join leaders from our Open-Source Community for a day to both welcome and inform both new and veteran developers within the OHDSI Community.
The **2024 OHDSI Global Symposium** will be held Oct. 22-24 at the Hyatt Regency Hotel in New Brunswick, NJ.

Tentative symposium format:
- **Oct. 22** – tutorials
- **Oct. 23** – plenaries, collaborator showcase
- **Oct. 24** – workgroup activities
Registration is now OPEN for the 2024 OHDSI Europe Symposium, which will be held June 1-3 in Rotterdam, Netherlands.

June 1 – tutorial/workshop
June 2 – tutorial/workshop
June 3 – main conference
ETL-Synthea Package v2.0.0 Release

This update includes the following:

- Update of the synthea package to support synthea 3.1.0 and 3.20
- Location and caresite ETL support
- Update to the package to split intermediate steps and event table loading @katy-sadowski @NACHC-CAD
- Updates to export to SQLite to support OMOP v5.3 and v5.4 @mccullen
- Update to github.io documentation
- Improved version support for CreateCDMIndexAndConstraintScripts, backupCDM, createPrunedTables, getEventConceptId, pruneCDM, restoreCDMTable scripts
- Resolution of formatting and lint issues
- Removal of redundant joins from allvisitable query
- Addition of the synthea version to the source description for debugging

Contributors

mccullen, katy-sadowski, and NACHC-CAD
Acute ST-Elevation Myocardial Infarction (STEMI) Network Study

Join Our CVD OHDSI Network Study on Acute STEMI!

Hello OHDSI Community,

We (@mirza_khan @mbrand @atifadam ) are thrilled to announce an exciting opportunity for collaboration in new network study focusing on Acute ST-Elevation Myocardial Infarction (STEMI). This study promises to deepen our understanding of STEMI patients' characteristics and identify incidence rates across multiple real-world data datasets.

Why This Study Matters:

- Acute myocardial infarction is a leading cause of hospital admission in the U.S., with STEMI being a critical subtype requiring immediate attention.
- Accurate, scalable, and generalizable identification and characterization of STEMI in multi-country real-world data has numerous benefits, including informing resource allocation, promoting use of effective therapies and interventions, and improving cardiovascular health.

Study Highlights:

- A cohort study using administrative claims or EHR data mapped to the OMOP CDM across the OHDSI network.
- Aims to understand patient characteristics and incidence rates of acute STEMI.
- Utilizes various standardized analytics in the OHDSI community, including the Strategus pipeline and HADES library.

We Need Your Participation!

- Your expertise and data can significantly contribute to this large-scale study.
- Collaboration is key to achieving a comprehensive and diverse understanding of STEMI across different populations and healthcare systems.
SciForce invite you to join a free webinar

**JACKALOPE PLUS:**
The Power of ML for Healthcare Data Mapping & Management

**SPEAKER:**
Denys Kaduk (MD, Data Scientist at Medical Team, SciForce)

With its advanced ML algorithms, **Jackalope aims to revolutionize the field of data management and mapping.** This tool facilitates a seamless transition to OHDSI’s OMOP CDM, ensuring greater efficiency.

**JOIN ONLINE:**
Save the date: **14th of March, 5 pm, GMT +2**

To find out more about the event and reserve your spot you can here [https://www.sciforce.tech/](https://www.sciforce.tech/)
March 14: Current Approaches for Distributed Analysis

Federated Analysis
State of the Science Collective Learning Series

Panel Discussion:
Current Approaches for Distributed Analysis

Thursday, March 14
10:00 a.m. PT | 1:00 p.m. ET
Bladder cancer - a quality benchmark utilizing FHIR and OMOP

(Andries Clinckaert, Valerie Vandeweerd, Murat Akand, Charlotte De Vlieghere, Bart Vannieuwenhuyse, Michel Van Speybroek, Frank Van der Aa, Martine Lewi, Christos Chatzichristos)

Background

- The field of oncology relies heavily on high quality and granular patient-level data.
- One of the main challenges for multi-center benchmarks and potentially analysis across different centers is the quality of the data.
- Transurethral resection of bladder tumor (TURBT) is a treatment for non-muscle invasive bladder cancer (NMIBC).

Methods

Quality Indicators (QIs) - FHIR:
- QIs can serve as a valuable mitigation strategies.
- Focus specifically on the Tre-TURBT for patients who had no detrusor muscle present.

\[ \% \text{ofpatients who had no detrusor muscle present} = \frac{\text{All TURBTs for no DM - AND subsequent TURBT within 6 weeks}}{\text{All TURBTs for no DM}} \]

- Only patients where no detrusor muscle was found in the resection specimen are included.
- The tumor cannot be Ta or T1-grade (known at time of TURBT), after the initial TURBT.
- Each sequence of a TURBT procedure followed by an MDT is counted.

Longitudinal Patient Trajectory (LPT) - OMOP:
- By leveraging the temporal and relational aspects of the OMOP data, we examined the sequence of events within the disease trajectory.

Results

Quality Indicators:
- 9% of patients who should undergo no-TURBT had a structured report.
- Delayed interval (84 days) - 35%.
- Treated interval (123 days) - 43%.
- Small final cohort of 31 patients.
- Too strict inclusion or inaccuracy / partially used reporting forms.

Conclusions

- Incomplete documentation and inconsistent data export can significantly impact data quality.
- Implement mitigation strategies. Data quality dashboard extended with disease specific rules:
  - A warning is generated when a T-stage is recorded without accompanying a TURBT procedure.
- By leveraging FHIR and OMOP in BE EHR data, we can ensure comprehensive data inclusion and improve data registration and quality.
- By leveraging standardized data protocols healthcare organizations can optimize data quality and enhance decision-making.

Acknowledgements

The current research was implemented within Athena project, which is funded by Flanders Innovation & Entrepreneurship (VLAIO). Project number: HIC.2019.2528.
Using MONAI Pre-Trained Models for Colorectal Tissue Type Phenotyping: A Feasibility Study to Integrate Deep Learning Model Results using the Medical Extension OMOP CDM

(Shijia Zhang, Woo Yeon Park, Blake Dewey, Paul Nagy)

**Introduction**

The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) has introduced new imaging extensions aiming to standardize medical imaging data. The OMOP CDM developed by Observational Health Data Sciences and Informatics (OHDSI) community is in charge of conducting consistent, real-world analysis of observational health data through open-source software. The extension to this model is significant as it enables the medical images to the Electronic Health Record data.

We extracted meaningful features from medical images for the continuous phenotyping of imaging extension of the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) review challenges.

Our study aims to recognize tissue types from pre-trained machine learning models on pathology images directly into the structured tables of the OMOP CDM using a transfer learning approach. It facilitates the evaluation of deep phenotyping from imaging biomarkers with health outcomes recorded in electronic medical records.

**Materials and Methods**

We employed transfer learning to utilize a pre-trained tumor nucleus model from the Medical Open Network for AI (MONAI) model. Based on the ResNet 50 architecture, it was initially trained on pathology images to recognize specific tumor features.

Our primary step was to further specialize the model for colorectal pathology. To achieve this, we fine-tuned using the PathMNIST Colon Pathology training dataset. The dataset contains colorectal tissue images categorized into polyps, including adenomas, adenocarcinomas, normal mucosa, and mucinous cancers. These images were then used for developing a deep learning model for colorectal adenocarcinomas epithelium. The aim of this fine-tuning was to enhance the model’s capability in identifying features unique to colorectal pathology.

After the fine-tuning, we assessed the model’s performance using a separate testing dataset from the PathMNIST Colon Pathology dataset. To fine-tune the training, especially the tissue types adenoma, were then integrated into the MONAI framework and made available in the OMOP CDM system.

Finally, to benchmark our progress, we compared the performance of our fine-tuned MONAI model to a separate fine-tuned Standard ResNet 18 model.

**Results**

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>True Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>0.89</td>
<td>0.90</td>
<td>0.91</td>
<td>125</td>
</tr>
<tr>
<td>Adenoma</td>
<td>0.90</td>
<td>0.85</td>
<td>0.88</td>
<td>125</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>0.94</td>
<td>0.86</td>
<td>0.91</td>
<td>125</td>
</tr>
<tr>
<td>Normal Mucosa</td>
<td>0.92</td>
<td>0.88</td>
<td>0.90</td>
<td>125</td>
</tr>
<tr>
<td>Mucinous Cancers</td>
<td>0.90</td>
<td>0.85</td>
<td>0.87</td>
<td>125</td>
</tr>
<tr>
<td>Colorectal Adenocarcinoma Epithelium</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
<td>125</td>
</tr>
</tbody>
</table>

**Conclusion**

Our study illustrates the potential of using fine-tuned pre-trained machine learning models to generate the OMOP CDM with phenotyping evidence derived from pathology images. This provides a more structured way to incorporate medical imaging findings into the OMOP CDM, without the process of sorting through unstructured reports. Moreover, this work shows the potential avenues to pursue imaging-based federated learning against multiple medical institutions using the OMOP CDM imaging extension, paving the way for widespread collaborative research and diagnosis.
WEDNESDAY

A tool for empirically identifying and reviewing candidate comparators for Pharmacoepidemiological studies

(Justin Bohn, Jamie P. Gilbert, Christopher Knoll, David M. Kern, Patrick B. Ryan)
The necessity of validity diagnostics when drawing causal inferences from observational data

(James Weaver, Erica A Voss, Guy Cafri, Kathleen Beyrau, Michelle Nashleanas, Robert Suruki)
FRIDAY
Identification of HIV positive individuals across multiple datasets

(Craig Mayer)

Identification of HIV positive individuals across multiple datasets

PRESENTER: Craig Mayer

INTRO:
- Condition attribution is key for appropriate patient capture
- Attribution may depend on amount and type of data available

METHODS
1. Analyzed 3 datasets
   1. All of Us (EHR and Research)
   2. UK Biobank (Research)
   3. Clinical Practice Research Datalink (EHR)
2. Analyzed 3 domains
   1. Condition
   2. Observation
      1. Self-reported
      2. Measurement
         1. Confirmatory tests
         2. No screening tests
3. Found patients from each domain and measured patient capture and crossover between domains

RESULTS
For All of Us, each HIV attribution method produced additional cases.
For UK Biobank, all HIV positive individuals were captured through self-reporting. The limited measurement and condition data captured no additional HIV positive individuals.
For CPDRD, all HIV positive individuals were captured via condition. No additional HIV positive individuals were found via the measurement data.

Ideal HIV attribution method may vary by type of dataset

For robust case capture, HIV attribution may require attribution methods from multiple domains.

HIV Counts by Domain

<table>
<thead>
<tr>
<th>Domain</th>
<th>All of Us</th>
<th>UK Biobank</th>
<th>CPDRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>5,185</td>
<td>214</td>
<td>93,374</td>
</tr>
<tr>
<td>Observation/Self-reported</td>
<td>1,046</td>
<td>484</td>
<td>x</td>
</tr>
<tr>
<td>Measurement</td>
<td>3,025</td>
<td>18</td>
<td>2,600</td>
</tr>
</tbody>
</table>

Case crossover by domain

<table>
<thead>
<tr>
<th>Domains</th>
<th>All of Us</th>
<th>UK Biobank</th>
<th>CPDRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition and Observation</td>
<td>1,031</td>
<td>194</td>
<td>x</td>
</tr>
<tr>
<td>Condition and Measurement</td>
<td>2,403</td>
<td>16</td>
<td>2,806</td>
</tr>
<tr>
<td>Measurement and Observation</td>
<td>575</td>
<td>16</td>
<td>x</td>
</tr>
<tr>
<td>Condition, Measurement and Observation</td>
<td>990</td>
<td>18</td>
<td>x</td>
</tr>
</tbody>
</table>
Opening: Biomedical Informatics Data Scientist at Stanford

Biomedical Informatics Data Scientist
1.0 FTE • Full time • Day - 08 Hour • R2335119 • Hybrid • 84866 IT RESEARCH • Technology & Digital Solutions • 455 Broadway, REDWOOD CITY, California

If you’re ready to be part of our legacy of hope and innovation, we encourage you to take the first step and explore our current job openings. Your best is waiting to be discovered.

Day - 08 Hour (United States of America)
This is a Stanford Health Care job.

A Brief Overview
The Biomedical Informatics Data Scientist will partner with researchers and clinicians to enable effective and efficient use of data and resources available via Stanford’s research clinical data repository (STARR) including the Electronic Health Records in the OMOP Common Data Model, radiology and cardiology imaging data and associated metadata, and new data types as they get integrated along with their databases and respective cohort query tools and interfaces e.g., OHDSI ATLAS. This individual will enable researchers to maximize their understanding, interpretation and use of these clinical and research tools for more informed and productive research, clinical trials, patient care and quality outcome projects.

Clean, extract, transform and analyze various kinds of clinical data to create analysis-ready datasets that follow the FAIR (Findable, Accessible, Interoperable and Re-usable) principles. Partner with researchers and clinicians to enable effective and efficient use of Stanford Clinical data and resources for the advancement of research and the educational mission.
The Zhang Lab at Washington University School of Medicine in St. Louis has **one postdoc/senior data analyst position** to work on **causal machine learning** and **responsible AI** for reliable real-world evidence generation.

- More details at [https://linyingzhang.com](https://linyingzhang.com)
  - Postdoc: [https://linyingzhang.com/files/Postdoc.pdf](https://linyingzhang.com/files/Postdoc.pdf)
  - Data analyst: [https://linyingzhang.com/files/Analyst.pdf](https://linyingzhang.com/files/Analyst.pdf)
- If interested, please send CV and cover letter to linyinzing@wustl.edu
### Epidemiology UX/Web Design Intern

<table>
<thead>
<tr>
<th>JOB TITLE</th>
<th>Epidemiology UX/Web Design Intern</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUNCTION</td>
<td>Career Programs</td>
</tr>
<tr>
<td>SUB FUNCTION</td>
<td>Non-LDP Intern/Co-Op</td>
</tr>
<tr>
<td>LOCATION</td>
<td>Raritan, New Jersey, United States</td>
</tr>
<tr>
<td>DATE POSTED</td>
<td>Jan 19 2024</td>
</tr>
<tr>
<td>REQUISITION NUMBER</td>
<td>2406163977W</td>
</tr>
</tbody>
</table>

**DESCRIPTION**

Janssen Research & Development, LLC., a division of Johnson & Johnson's Family of Companies is recruiting for Epidemiology UX/Web Design Intern. This position is a member of the Observational Health Data Analytics (OHDA) team. OHDA's mission is to improve the lives of patients and quality of healthcare by efficiently generating real-world evidence from the world's observational health data, transparently disseminating evidence-based insights to real-world decision-makers, and objectively advancing the science and technology behind reliable evidence.
Director, RWE at Gilead

Director, RWE - Data Science - OHDSI

Responsibilities:
Collaborate with researchers and data scientists to understand project requirements and translate them into OHDSI-compatible solutions. Work with databases, ensuring data integrity and optimization for OHDSI-related queries and analyses. Perform data analyses in OHDSI-related tools like ATLAS. Customize and extend OHDSI tools and applications to meet specific project needs. Collaborate with cross-functional teams to troubleshoot and resolve technical issues related to OHDSI implementations. Stay informed about OHDSI community updates, best practices, and emerging trends in observational health data research. Contribute to the development and documentation of data standards and conventions within the OHDSI community.
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: ohdssi.org/community-calls