Network Studies
(and possible a software demo!)

OHDSI Community Call
Nov. 15, 2022 • 11 am ET
# Upcoming OHDSI Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov. 22</td>
<td>10-Minute Tutorials</td>
</tr>
<tr>
<td>Nov. 29</td>
<td>Workgroup Updates</td>
</tr>
<tr>
<td>Dec. 6</td>
<td>Fall Publications</td>
</tr>
<tr>
<td>Dec. 13</td>
<td>How Did We Do In 2022?</td>
</tr>
<tr>
<td>Dec. 20</td>
<td>Holiday-Themed Final Call of 2022</td>
</tr>
</tbody>
</table>
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
2022 Asia-Pacific (APAC) Symposium
2022 Asia-Pacific (APAC) Symposium
2022 Asia-Pacific (APAC) Symposium
2022 Asia-Pacific (APAC) Symposium
OHDSI Shoutouts!

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

Have a study published? Please send to sachson@ohdsi.org so we can share 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>1 pm</td>
<td>Common Data Model</td>
</tr>
<tr>
<td>Wednesday</td>
<td>9 am</td>
<td>FHIR and OMOP Data Model Harmonization Subgroup (ZOOM)</td>
</tr>
<tr>
<td>Wednesday</td>
<td>12 pm</td>
<td>Health Equity Journal Club</td>
</tr>
<tr>
<td>Wednesday</td>
<td>2 pm</td>
<td>FHIR and OMOP Terminologies Subgroup (ZOOM)</td>
</tr>
<tr>
<td>Wednesday</td>
<td>7 pm</td>
<td>Medical Imaging</td>
</tr>
<tr>
<td>Thursday</td>
<td>10 am</td>
<td>Data Quality Dashboard Development</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>HADES</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>Geographic Information System (GIS)</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Healthcare Special Interest Group</td>
</tr>
<tr>
<td>Tuesday</td>
<td>9 am</td>
<td>OMOP CDM Oncology</td>
</tr>
</tbody>
</table>

[ohdsi.org/upcoming-working-group-calls/](https://www.ohdsi.org/upcoming-working-group-calls/)
Open-Source Community WG Meeting

Please join the Nov. 23 Open-Source Community WG meeting, which will include a presentation from 2018 Titan Award winner Lee Evans focused on OHDSI open source software: continuous integration, automated testing, & test database infrastructure.

Wednesday, Nov. 23, 11 am ET
Join Anna Ostropolets’ Dissertation Defense

OHDSI veteran and 2018 Titan Award winner Anna Ostropolets will defend her Columbia University dissertation Wed., Nov. 30, on Generating Reliable and Responsive Observational Evidence: Reducing Pre-analysis Bias. The open session will be at 10 am ET on Zoom.

Wednesday, Nov. 30, 10 am ET
Deployment of an OMOP CDM-compatible NLP system for Rapid Development and Dissemination of a Long-COVID Extraction NLP task

(Andrew Wen, Liwei Wang, Huan He, Sunyang Fu, Sijia Liu, Hongfang Liu)

MONDAY
Impact of random oversampling and random undersampling on the development and validation of prediction models using observational health data (Cynthia Yang, Egill A. Fridgeirsson, Jan A. Kors, Jenna M. Reps, Peter R. Rijnbeek)

RESULTS
1. Figure 1 shows the AUROC differences in both AUCROC with stratified and AUCROC without stratifying across all prediction problems and databases for each sampling strategy and classifier on external validation. On average, random oversampling and random undersampling did not improve the AUCROC. For random oversampling with random forest, we observed a larger effect on model performance.
2. The calibration plots (Figure 2) and calibration diagnostics followed the original calibration curve on external validation. The calibration plot for random oversampling with random forest was improved compared to the calibration plot for random undersampling with random forest.
3. The area under the receiver operating characteristic curve (AUCROC) and stratified calibration plots were not improved by either sampling strategy.

TUESDAY

#OHDSISocialShowcase This Week

Don't over- or undersample when developing prediction models using large observational health data.
KorNER: Building Korean NER models for a manually annotated corpus from clinical notes using cross-lingual transfer learning

A. Park

BACKGROUND:
A list of studies have been explored on NER to extract important clinical concepts like Problem, Tests, and Treatments in English clinical notes. In the Korean medical field, research has been carried out on clinical notes written in Korean much less in a standardized way in OMOP-CDM.

OBJECTIVE:
To develop an efficient and reusable Korean NER framework to extract important clinical concepts based on OMOP-CDM

METHODS:
1. Data Source
- Korean medical discharge summary from the Korea Institute of Health and Social Affairs.
2. Core Features
- Important clinical concepts: HPI (Problem), Tests, Treatments
- Annotation was performed using Clinical vocabulary
3. Model Development
- Train, test, and evaluate were set as
- The model was fine-tuned on pre-trained language models
- BERT and XLM-Roberta were employed
- The model was fine-tuned on the same datasets
- The model was evaluated using standard metrics

RESULTS:
- XLM-ROBERTA-base model achieved the best performance
- BERT-ROBERTA-base and BERT-MULTILINGUAL-base models were compared on the baseline BERT-CDM
- XLM-ROBERTA-base model outperformed BERT-CDM

Table 1. Overall model performance

<table>
<thead>
<tr>
<th>Model</th>
<th>F1</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT-ROBERTA</td>
<td>0.741</td>
<td>0.779</td>
<td>0.768</td>
</tr>
<tr>
<td>BERT-MULTILINGUAL</td>
<td>0.769</td>
<td>0.812</td>
<td>0.801</td>
</tr>
<tr>
<td>XLM-ROBERTA</td>
<td>0.803</td>
<td>0.826</td>
<td>0.812</td>
</tr>
<tr>
<td>XLM-MULTILINGUAL</td>
<td>0.807</td>
<td>0.840</td>
<td>0.825</td>
</tr>
</tbody>
</table>

KorNER – An efficient and reusable Korean NER framework based on common data model

B. Test
- Softmax
- Linear layer

Pre-trained Language Models
- BERT-ROBERTA

Building Korean NER models for a manually annotated corpus from clinical notes using cross-lingual transfer learning (Jianfu Li, Jimyung Park, Xinyue Hu, Jingqi Wang, Rae Woong Park, Hua Xu)
Protocol for finding supplemental oxygen data in electronic health record (EHR) flowsheets for inclusion in the OMOP ETL

TANNER ZHANG, STEVEN MILLER, MICHAEL COOK, ALAN COLTRI, ZACHARY WANG, PAUL NAGY, JUSTIN RUCCI, GALINA LOZINSKI, MATTHEW ROBINSON

THURSDAY
COVID-19 Vaccine Administration Pathways in US Administrative Claims

(Kevin Haynes, Christopher Knoll, Rupa Makadia, Patrick Ryan)

ATLAS Cohort Pathways can be used to evaluate data quality by identifying potentially implausible pathways.
Northeastern University invites applications for multiple tenured/tenure-track faculty positions in support of an Impact Engine centered on large-scale observational health data science and informatics to start in the fall of 2023. These faculty will be core members of our Real-World Healthcare Navigator (RWHN) Impact Engine which aims to change how research is translated into clinical practice by establishing a sustainable service that leads the way in fully reproducing health studies.
The OHDSI Center at the Roux Institute seeks a postdoctoral fellow to join their team focused on developing statistical methods and applying them to observational data from large-scale federated datasets (e.g. electronic health records and administrative claims data), with specific applications to the safety of biologics. This research will directly improve our ability to use real world data to characterize patient populations, construct population level estimates relating exposures to health outcomes, and to enhance clinical decision making through improved patient-level predictions.
Opening: FDA/CDER

FDA/CDER’s Division of Hepatology and Nutrition is seeking a clinician with bioinformatics or biostatistics training to work with the Drug-Induced Liver Injury (DILI) Team to evaluate large datasets of liver-related data, collaborate on the Team’s review of drugs with hepatotoxicity signals, and help develop informatics-based processes in DILI evaluation across the Agency.

Contact Judy Racoosin at judith.racoosin@fda.hhs.gov for information about the application process (that will be through USAJOBS).
Opening: Tufts Medicine

Andrew Williams recently announced two exciting new openings at Tufts Medicine.

1) Senior Project Manager for a multisite multiyear grant standardizing critical care EHR and waveform data. (CHoRUS Bridge2AI)

2) Lead software developer and research data warehouse manager for Tufts Medicine’s OMOP instance and related services.

Remote work is possible for both positions.

1. Link for Senior Project Manager position: https://smrtr.io/bBVzh
2. Link for Lead Software Developer and Research Data Warehouse Manager position: https://jobs.smartrecruiters.com/TuftsMedicalCenter1/743999857980631-software-development-lead-res-g-c-ctsi

Andrew’s email: awilliams15@tuftsmedicalcenter.org
Research Associate (Data Scientist/Statistical Engineer), Johns Hopkins inHealth and Biostatistics Center

• Execute OHDSI studies (e.g. for cohort characterizations and comparative effectiveness) on Johns Hopkins’s EHR data to support clinicians;

• Collaborate with statisticians and clinicians to continuously integrate state-of-the-art statistical tools to the inHealth/OHDSI tool stack for deployment;

• Mentor trainees on data science and software development skills;

• Co-teach courses on observational health data analytics and data science skills at School of Medicine and Public Health;

• Facilitate adoption of the inHealth tools among the broader OHDSI community by contributing to OHDSI’s Health Analytics Data-to-Evidence Suite.

• https://apply.interfolio.com/114436
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?
Nov. 15: Open Network Studies

Expanding maternal and infant data from EHRs for pregnancy research
- Safety and Effectiveness of Anti-Hypertensive Medications in Pregnancy
- Project to Characterize Anti-Hypertensive, Anti-Coagulant, Anti-Diabetic and Antibiotic Medication Usage During Pregnancy and Postpartum

Alison Callahan (Instructor, Medicine • Stanford University)
Stephanie Leonard (Instructor, Obstetrics & Gynecology • Stanford University)
Louisa Smith (Assistant Professor, Health Sciences • Northeastern University)

Relative Risk of Cervical Neoplasms Associated with Copper and Levonorgestrel Secreting Intrauterine Devices: Real World Evidence from the OHDSI Network

Matthew Spotnitz (Postdoctoral Research Fellow • Columbia University)