Where Are We Going in 2023?

OHDSI Community Call
Jan. 10, 2022 • 11 am ET
## Upcoming OHDSI Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 10</td>
<td>Where Can OHDSI Go in 2023?</td>
</tr>
<tr>
<td>Jan. 17</td>
<td>OHDSI Speed Dating</td>
</tr>
<tr>
<td>Jan. 24</td>
<td>Collaborations For Strategic Priorities</td>
</tr>
<tr>
<td>Jan. 31</td>
<td>Introduction to Phenotype Phebruary</td>
</tr>
<tr>
<td>Feb. 7</td>
<td>Phenotype Phebruary Weekly Update + Workgroup Plans for 2023</td>
</tr>
<tr>
<td>Feb. 14</td>
<td>Phenotype Phebruary Weekly Update + Workgroup Plans for 2023</td>
</tr>
<tr>
<td>Feb. 21</td>
<td>Phenotype Phebruary Weekly Update + Workgroup Plans for 2023</td>
</tr>
<tr>
<td>Feb. 28</td>
<td>Phenotype Phebruary Weekly Update + Workgroup Plans for 2023</td>
</tr>
</tbody>
</table>
# Upcoming OHDSI Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 10</td>
<td>Where Can OHDSI Go in 2023?</td>
</tr>
<tr>
<td>Jan. 17</td>
<td>OHDSI Speed Dating</td>
</tr>
<tr>
<td>Jan. 24</td>
<td>Collaborations For Strategic Priorities</td>
</tr>
<tr>
<td>Jan. 31</td>
<td>Introduction to Phenotype Phebruary</td>
</tr>
<tr>
<td>Feb. 7</td>
<td>Phenotype Phebruary Weekly Update + Workgroup Plans for 2023</td>
</tr>
<tr>
<td>Feb. 14</td>
<td>Phenotype Phebruary Weekly Update + Workgroup Plans for 2023</td>
</tr>
<tr>
<td>Feb. 21</td>
<td>Phenotype Phebruary Weekly Update + Workgroup Plans for 2023</td>
</tr>
<tr>
<td>Feb. 28</td>
<td>Phenotype Phebruary Weekly Update + Workgroup Plans for 2023</td>
</tr>
</tbody>
</table>
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
OHDSI Shoutouts


Generalisable long COVID subtypes: Findings from the NIH N3C and RECOVER programmes


Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
The Jackson Laboratory for Genomic Medicine, 33 Discovery Drive, Farmington, CT, USA.
Anesthesia, Dipartimento di Informatica, Università degli Studi di Milano, Milan, Italy.
Sapienza University, Rome, Italy.

Congratulations to the team of Xiang Cheng, Meiling Cheng, Liyi Yu and Xuan Xiao on the publication of iADRGSE: A Graph-Embedding and Self-Attention Encoding for Identifying Adverse Drug Reaction in the Earlier Phase of Drug Development in the International Journal of Molecular Sciences.

Article

iADRGSE: A Graph-Embedding and Self-Attention Encoding for Identifying Adverse Drug Reaction in the Earlier Phase of Drug Development

Xiang Cheng, Meiling Cheng, Liyi Yu and Xuan Xiao

Department of Computer, Hangzhou Dianzi University, Hangzhou 310018, China
* Correspondence: xiangcheng@hdu.edu.cn; Tel: +86-571-8684-2866

Abstract: Adverse drug reactions (ADRs) are a major issue to be addressed by the pharmaceutical industry. Early and accurate detection of potential ADRs contributes to enhancing drug safety and reducing financial expenses. The majority of the approaches that have been employed to identify ADRs are limited to determining whether a drug exhibits an ADR, rather than identifying the exact type of ADR. By introducing the “multi-level fusion-fusion deep-learning model”, a new prediction, called iADRGSE, has been developed, which can be used to identify adverse drug reactions at the early stage of drug discovery. iADRGSE integrates a self-attentive module and a graph-network module that can extract one-dimensional sub-structure sequence information and two-dimensional chemical-structure graph information of drug molecules. As a demonstration, cross-validation and independent testing were performed with iADRGSE on a dataset of ADRs classified into 27 categories, based on the source of organ classification. In addition, experiments comparing iADRGSE with approaches such as NPF were conducted on the CMOP dataset, using the jackknife test method. Experiments show that iADRGSE was superior to existing state-of-the-art predictors.

Keywords: adverse drug reaction; graph isomorphism network; self-attention; multi-label learning

1. Introduction

Adverse drug reactions (ADRs) or side effects are substantially harmful or distressing reactions, and are described as adverse responses to drugs beyond their anticipated therapeutic effects [1]. In the United States, it is estimated that ADRs result in over 100,000 patient deaths per year [2] and the cost of ADR-related morbidity was USD 328.4 billion in 2016 [3]. The process of drug development involves a lot of monetary resources because it involves a lot of clinical trials and tests [1]. Many ADRs are not detected in the early stages of drug development, owing to restricted trial samples and time [1]. Thus, ADRs not only jeopardize patient health but also result in wasted healthcare costs, and are considered as a major global public health problem. Traditional laboratory experiments to identify potential ADRs are not only cumbersome and cost-effective, but also less effective in the earlier phase. In recent years, algorithms in silico have been employed to speed up the prediction process and reduce drug-development costs.

Among the existing studies, some utilize data mining to analyze potential ADRs from large amounts of data and various sources of information; others adopt machine learning methods to predict ADRs.

The available databases of ADRs have some limitations. Among the databases collected by the spontaneous reporting systems (SRS) and FDA, Adverse Event Reporting System (FAERS) are not comprehensive enough, and there are problems such as repeated declaration. Drugs in the Side Effect Resource (SIDER) are limited to FDA-approved drugs only. The content of the European Medicines Agency (EMA) and other large-scale databases is complicated, and has no special retrieval of ADRs, which cause a lot of inconvenience for the use of data. Considering the limitations of the existing database, some researchers
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?
<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>Tuesday</td>
<td>3 pm</td>
<td>OMOP CDM Oncology – Outreach/Research Subgroup</td>
</tr>
<tr>
<td>Tuesday</td>
<td>6 pm</td>
<td>Eyecare and Vision Research</td>
</tr>
<tr>
<td>Wednesday</td>
<td>9 am</td>
<td>Patient-Level Prediction</td>
</tr>
<tr>
<td>Wednesday</td>
<td>2 pm</td>
<td>Natural Language Processing</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</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 and Evaluation</td>
</tr>
<tr>
<td>Friday</td>
<td>9 am</td>
<td>GIS – Geographic Information System General</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>10 am</td>
<td>Healthcare Systems Interest Group</td>
</tr>
</tbody>
</table>
Upcoming Workgroup Calls

Welcome

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced “Odyssey”) program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University.

Building A Healthier World Together

The 2022 OHDSI Symposium focused on the theme of “Building A Healthier World Together” and it featured presentations and researchers from collaborators around the world. Please visit the symposium homepage to see the videos, slides and all other output from this three-day event.

2022 OHDSI Global Symposium

ohdsi.org/workgroups
Interested in presenting to the CDM WG?

Enhancing workgroup collaboration has been a focus over the last year.

Clair Blacketer has a few thoughts for any workgroup that is considering presenting to the CDM WG.
January Newsletter Is Available

Community Updates

Where Have We Been?
- 2022 saw a return to in-person events, several new workgroups and the Kherion Cohort, community activities like Phenotype Presentations, DevCon and the Early-Stage Researchers Career Speaker Series, and plenty more. Patrick Ryan provided a review of the year, which you can watch later in the newsletter.
- The OHDSI community produced a record-setting 111 publications in 2022, and have now seen 2,077 authors participate in OHDSI-related papers since our inception. You can read any of these studies via the new Publications Dashboard, developed by Paul Nagy, Stan Liu and their team this past year.
- The 2022 Asia-Pacific (APAC) Symposium took place in November, and all talks and slides from both the main conference and the full-day tutorial are now available at the APAC Symposium homepage.

Where Are We Now?
- The open-source tools that empower OHDSI's global research initiatives are not only available to the community, but they are also developed by the community. Leaders from around the world have developed tools that provide the foundation for OHDSI collaborators to engage in robust, reliable and reproducible observational health research. Many of these developers have provided "15-Minute Tutorials" that can help aid your understanding of how to use these tools in research.
- The EHDEN Consortium announced that 22 data partners from across 13 countries have been selected from the final open call to join the EHDEN data network. The data partners from this call represent almost 200 million patient records, originating from various care settings, adding to the approximately 630 million records with the 166 partners working with EHDEN from the prior six calls.

Where Are We Going?
- Community calls will resume Jan. 10 (11 am ET) and continue each Tuesday as a way to connect as a community, share updates, learn from each other, and continue to move forward together. A new call invite will go out the week of January 2, but you can also find the new link on our community calls page.
- The OHDSISocialShowcase will continue in 2023, as research from the OHDSI Symposium is featured each weekday on the OHDSI Twitter and LinkedIn accounts. Please follow us, learn about these new developments in our community, and share with your networks!

How Can You Join The Journey?
- The OHDSI research community strives to promote better health decisions and care through globally standardized health data, continuously developing large-scale analytics and a spirit of collaboration throughout open science. We are proud to have more than 3,200 collaborators across six continents, as well as health records for about 928 million unique patients from around the world. We are always looking for new collaborators, so if this sounds exciting to you, please read about how you can Join The Journey!
Oxford Real World Evidence Summer School

Oxford Summer School 2023: Real World Evidence using the OMOP Common Data Model

**Brief Description:**
Our Real World Evidence Summer School will provide participants with the tools and concepts necessary to plan and execute Real World Evidence studies, with a focus on the use of the OMOP common data model. The course will have morning lectures followed by afternoon practicals where concepts discussed in the morning will be put in practice with hands-on sessions. Practical sessions will have two tracks: a) for those interested in the design of studies and use of existing analytical and data curation tools; and b) for more advanced data scientists and programmers interested in the development or modification of analytical code using R.

**Registration:** It is now open

**Venue:** Lady Margaret Hall Talbot Hall Theatre, Norham Gardens, Oxford OX2 6QA

**Date:** 19th-23rd June 2023

For booking please use [Booking Information](#)

Please see the Preliminary Programme [here](#)

**AUDIENCE:**
Pharmacists, clinicians, academics (including statisticians, epidemiologists, and related MSc/PhD students); Industry (pharmacy or device) or Regulatory staff with an interest in the use of routinely collected data for research.

**LEARNING GOALS:**
Collaborator Spotlight: Thamir AlShammary

Thamir AlShammary, an advisor to the President of the Saudi Food and Drug Authority (SFDA), has been an active contributor to the OHDSI community for several years. He discusses his background, his journey into OHDSI and the impact he has seen, and why OHDSI can be a difference maker in generating trustworthy evidence, tools and best practices within the community, in the latest edition of the Collaborator Spotlight.

ohdsi.org/spotlight-thamir-alshammary
Analysis of Influencing Factors of Mortality in COVID-19 Patients: A Retrospective Cohort Study

Do Duy Khang, Phung-Anh Nguyen, Chang-I Chen, Chung-Chien Huang, Carlos Shu-Kei Lam, Noi Yar, Christine Y. Lu, Chi-Tsun Cheng, Jason C. Hsu

Taipei Medical University, Taiwan

Contact: jasonhua@tmu.edu.tw

TUESDAY

Analysis of Influencing Factors of Mortality in COVID-19 Patients: A Retrospective Cohort Study (Do Duy Khang, Phung-Anh Nguyen, Chang-I Chen, Chung-Chien Huang, Carlos Shu-Kei Lam, Noi Yar, Christine Y. Lu, Chi-Tsun Cheng, Jason C. Hsu)
The manifold presentations of PROMS and questionnaires: patient-reported outcomes in OMOP use cases

**PRESENTED BY:** Sebastiaan van Sandijk

**INTRODUCTION**

- PROMS and ODI questionnaires are essential in various use cases. An HTA (Health Technology Assessment), health appraisal and drug safety assessments, clinical guideline development, etc.
- Currently, however, OMOP-CDM and OHDSI tools do not fully support PROMS data or standardized analytics with the necessary metadata.
- Guidelines or conventions for editing PROMS data in OMOP-CDM do not exist. This is going to work aims to propose some recommendations.

**METHODS**

1. Use case driven approach working with data partners in the EHDEN network that are in the process of converting their data.
2. Inventory of common PROMS and questionnaires as well as their use cases: why and what in OMOP?
3. Inventory of existing issues as well as other useful solutions to address these issues.
4. Discussion of potential best practice recommendations are discussed for selected PROMS / questionnaires, focusing on the quick wins.
5. (Future step) Test the proposed recommendations in selected focused case studies, with live data partners, to determine value added and validity of the approach or recommendations.

**RESULTS AND RECOMMENDATIONS**

- Need to standardize survey questionnaires, registries, information or recommended data sets, cohort studies, screenings, intake forms, etc. These all exist but are not used consistently with their own implications.
- Common issues:
  - Lack of standardization in use;
  - Inconsistent / label differentiation;
  - “valued” / “non-valued”;
  - Disease-specific / general;
  - Time period;
  - Preprocess vs. “data type” sources; strings, BLOB.
- “Prevalence” and “confirm of use” should be better represented in OMOP - and used in analytics.

**DISCUSSION AND CONCLUSION**

- With our use case driven approach we expect to be able to formulate some recommendations for PROMS data in OMOP - for future use.
- Initial focus will be on validated questionnaires with refined business rules - like EORTC-30, PCRS-CTCAE, and some semantically overlapping EICHM-Standard Sets.

**REFERENCES**

- Sebastiaan van Sandijk*1, Peter Prinsen, Mieke van Hemelrijck, Michael Kallfelz, Dalia Dawoud*2

*1 Odyssey Data Services - NIN; *2 EORTC-Phoenix; *1OD2
1s.deves@ehden.org

**PROMS and questionnaires are not all the same; implications and requirements vary - for clinical practice and OMOP use cases.**

Start from what is in OMOP and examine how (validated) questionnaire scores relate to clinical standard concepts.

Conventions needed for Provenance and Context of Use!

**The manifold presentations of PROMS and questionnaires: patient-reported outcomes in OMOP use cases (Sebastiaan van Sandijk, Peter Prinsen, Mieke van Hemelrijck, Michael Kallfelz, Dalia Dawoud)**
#OHDSISocialShowcase This Week

Machine Learning for Predicting Patients at Risk of Prolonged Opioid Use Following Surgery (Behzad Naderalvojoud, Tina Hernandez-Boussard)

THURSDAY
Background
Cognitive impairment following stroke has wide prevalence ranging from 20% to 40%. Further, stroke and the subtypes, including ischemic stroke, transient ischemic attack and intracerebral hemorrhage, significantly increase the long-term risk of dementia after 5 to 10 years. The incidence rate of post-stroke dementia increases yearly, though the relative risk gradually decreases. The study aims to predict the dementia development one year after stroke diagnosis (false date).

Methods
The study conducted on TMCDC® from January 2018 to September 2017. The inclusion, exclusion and outcome criteria are selected based on ICD-9 and ICD-10 codes. We include all patients with history of stroke, insomnia, cognitive impairment and other codes related with the diseases (332.5, 433.1, 434.1, 436, 436.3, 438.2, 438.3, 438.4, 438.5, 438.81, 438.82, 438.83, 438.9). We exclude psychiatric disorder, sleep apnea, traumatic brain injury, cancer, Parkinson’s disease, and cognitive impairment from the outcome (331, 290.8, 780, 784.4, 786.0, 786.2, 786.3, 786.6, 787.1, 787.2, 787.3, 787.4, 787.5, 787.8, 788.0, 788.8, 788.9, 789.0, 789.1, 789.2, 789.8, 789.9). The outcome and risk factors used in the study are: Cognitive impairment, Stroke dementia, Uncontrolled, Stroke dementia with depression, Depression in conditions classified elsewhere, Alzheimer’s disease, Frontotemporal dementia, and Scleroderma degeneration of brain (331.8, 332.5, 430, 430.2, 430.3, 430.4, 430.5, 430.6, 430.7, 430.8, 430.9, 430.91, 430.92, 430.93, 430.94, 430.95, 430.96, 430.97, 430.98, 430.99, 430.999). The patient with outcome at least one year after stroke index data are labeled by Y, and the rest without outcome labeled by N. We use recent literature to compare the performance of many different machine learning algorithms.

Results
Preliminary result performed on the hospital data (59% out of 834 patients). LighGBM algorithm gives the best AUC metrics on the 30 hold-out validation. The scores are 0.82 for accuracy, 0.81 for AUC, 0.57 for precision, 0.52 for recall, and 0.21 for F1. In the current model, we only use ICD and gender as the features.

Conclusions
We can see the performance in Figure 1. The true positive is still less than the false negative. This is to be expected since we use common features between two labels. The Figure 2 shows the optimal threshold is pretty low, 0.27, much lower than default 0.5 for binary classification. Figure 3 shows the distribution of the data based on labels. We can see that some positive labels are overlapping with the negative ones (have the same features).

Acknowledgment
We appreciate the technical assistance provided by the staff of Taipei Medical University’s Office of Information Technology.

One-year Post-Stroke Prediction on Cognitive Impairment Development: A Machine Learning Approach (Muhammad Solihuuddin Muhtar, Faizul Hasan, Alex P.A. Nguyen, Jason C. Hsu, Hsiao-Yean Chiu)
Describing treatment with antidiabetics in patients with T2D and moderate to severe CKD across a network of OMOP databases (Martin Lavallee, David Vizcaya, Ron Herrera, Niki Oberprieler, Glen James, Darya Kosareva, Asieh Golozar)

Large-scale characterization of six anti-glycemic drug classes enables detection of treatment variation across geographies and settings
Examining differential measurement error in phenotype algorithms due to age, sex, and disease prevalence differences using PheValuator (Joel Swerdel, Jenna Reps)
Lowering the OMOP ETL Barrier for Clinical Registries

PRESENTER: Smith Heavner

RESULTS
- Possible to lower OMOP ETL barrier
- Default configuration transformations
- Support and feedback from experienced site
- Project continues with site implementing OMOP using Parsmus

Pilot site implemented OMOP in less than 200 hours

CONCLUSION
It is feasible to reduce the ETL implementation time by providing default configuration transformations along with assistance and feedback on the process. Further reduction in the person-hours required to perform an OMOP ETL will be evaluated with the Parsmus web-based OMOP ETL project and cloud provider deployments of Atlas and the Q2D. Our goal is to increase the adoption of OMOP in sites with fewer resources and enable wider participation in high-quality clinical registries with sufficient patient numbers and data variables to perform appropriate observational research techniques to control for potential confounders (e.g., propensity score matching).

Methods
With funding from HHS Assistant Secretary for Planning and Evaluation.
1. Developed tools and resources to disseminate harmonization methods developed by Johns Hopkins University
2. Expanding CME IQ from physician-entered reports to automated FAH data collection
3. Promote conversion of non-common data model (CDM) systems to OMOP standards
4. Developed minimal dataset for drug repurposing research in COVID-19 as a use case

INTRODUCTION
Goal: To systematically collect harmonized data from how clinicians use existing drugs in new ways to treat diseases with limited or no treatment options.

Aims: To simplify ETL process and create pathway for real-world data to be made available in CUME IQ.

Method: Data harmonization using OMOP CDM

Impact: Secure deidentified data elements for assessing the effectiveness of repurposed treatments for diseases of high unmet clinical need.

WEDNESDAY

Lowering the OMOP ETL Barrier for Clinical Registries (Smith Heavner, Trayson Llano, Zachary Wang, Marco Schito, Heather Stone, Pam Dasher, Tresha Russel, Vishakha Kumar, Ben Saeks, Michael Cooke, Rahul Kashyap, Matthew Robinson, Paul Nagy)
Einstein-ATLAS: Leveraging OHDSI/ATLAS and Open-Source Development to Support Translational Research, Data Science, and Regulatory Compliance (Parsa Mirhaji, Selvin Soby, Erin Henninger, Chandra Nelapatla, Manuel Wahle, Boudewijn Aasman, Eran Bellin)
Federated learning for quantifying racial disparities in kidney graft failure rates using US registry data from 29,468 patients across 149 transplant centers (Jiayi Tong, Yishan Shen, Alice Xu, Chongliang Luo, Mackenzie Edmondson, Ruowang Li, Di Wang, Kevin He, David A. Asch, Yong Chen)
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?