What Can OHDSI Achieve Together in 2024?

OHDSI Community Call
Jan. 9, 2024 • 11 am ET
# Upcoming Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 9</td>
<td>Welcome Back! What Can OHDSI Accomplish in 2024?</td>
</tr>
<tr>
<td>Jan. 16</td>
<td>Connections For Collaborations</td>
</tr>
<tr>
<td>Jan. 23</td>
<td>2023 UK Study-A-Thon Lessons Learned</td>
</tr>
<tr>
<td>Jan. 30</td>
<td>Phenotype Phebruary Introduction</td>
</tr>
<tr>
<td>Feb. 6</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 1</td>
</tr>
<tr>
<td>Feb. 13</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 2</td>
</tr>
<tr>
<td>Feb. 20</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 3</td>
</tr>
<tr>
<td>Feb. 27</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 4</td>
</tr>
</tbody>
</table>

@OHDSI  
www.ohdsi.org  
#JoinTheJourney
Three Stages of The Journey

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

Congratulations to the team of Pierre Heudel, Hugo Crochet, Thierry Durand, Philippe Zrounba, and Jean-Yves Blay on the publication of From data strategy to implementation to advance cancer research and cancer care: A French comprehensive cancer center experience in *PLOS Digital Health*.
OHDSI Shoutouts!

Congratulations to the team of Cynthia Yang, Egill Fridgeirsson, Jan Kors, Jenna Reps and Peter Rijnbeek on the publication of Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data in the Journal of Big Data.
OHDSI Shoutouts!

Congratulations to the team of Christian Reich, Anna Ostropolets, Patrick Ryan, Peter Rijnbeek, Martijn Schuemie, Alexander Davydov, Dmitry Dymshyts, and George Hripcsak on the publication of OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization in JAMIA.

Research and Applications

OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization

Christian Reich, MD,1,2,3,†, Anna Ostropolets, PhD,1,4,5, Patrick Ryan, PhD,1,4,6, Peter Rijnbeek, PhD,1,2,†, Martijn Schuemie, PhD,1,6, Alexander Davydov, MD,1,5, Dmitry Dymshyts, MD,1,6, George Hripcsak, MD,1,4

1Coordinating Center, Observational Health Data Sciences and Informatics, New York City NY 10032, United States, 2OHDSI Center at the Roux Institute, Northeastern University, Boston MA 02111, United States, 3Department of Medical Informatics, Erasmus University Medical Center, 3015 GD Rotterdam, The Netherlands, 4Department of Biomedical Informatics, Columbia University Medical Center, New York City NY 10032, United States, 5Oxford Data Services, Cambridge MA 02140, United States, 6Observational Health Data Analytics, Janssen Research & Development, Titusville NJ 08560, United States

*Corresponding author: Christian Reich, MD, OHDSI Center at the Roux Institute, Northeastern University, 100 Foro St, Portland ME 04101 (reich@ohdsi.org)

Abstract

Importance: The Observational Health Data Sciences and Informatics (OHDSI) is the largest distributed data network in the world encompassing more than 321 data sources with 2.1 billion patient records across 34 countries. It enables large-scale observational research through standardizing the data into a common data model (CDM). Observational Medical Outcomes Partnership (OMOP) CDM requires a comprehensive, efficient, and reliable ontology to support data harmonization.

Materials and methods: We created the OHDSI Standardized Vocabularies—a common reference ontology mandatory to all data sites in the network. It comprises imported and de-novenormalized ontologies containing concepts and relationships between them, and the praxis of converting the source data to the OMOP CDM based on these. It enables harmonization through assigned domains according to clinical categories, comprehensive coverage of entities within each domain, support for common-used international coding schemes, and standardization of semantically equivalent concepts.

Results: The OHDSI Standardized Vocabularies comprise over 1.5 million concepts from 136 vocabularies. They are used by hundreds of groups and several large data networks. More than 88,000 users have performed 50,000 downloads of the system. This open-source resource has proven to address an impediment of large-scale observational research—the dependence on the content of source data representation. With that, it has enabled efficient phenotyping, covariate construction, patient-level prediction, population-level estimation, and standard reporting.

Discussion and conclusion: OHDSI has made available a comprehensive, open vocabulary system that is unmatched in its ability to support global observational research. We encourage researchers to exploit it and contribute their use cases to this dynamic resource.

Key words: OHDSI; controlled vocabulary; common data model; observational data.
OHDSI Shoutouts!

Congratulations to the team of Qiong Wu, Jiayi Tong, Bingyu Zhang, Dazheng Zhang, Jiajie Chen, Yuqing Lei, Yiwen Lu, Yudong Wang, Lu Li, Yishan Shen, Jie Xu, L. Charles Bailey, Jiang Bian, Dimitri A. Christakis, Megan L. Fitzgerald, Kathryn Hirabayashi, Ravi Jhaveri, Alka Khaitan, Tianchen Lyu, Suchitra Rao, Hanieh Razzaghi, Hayden T. Schwenk, Fei Wang, Margot I. Gage Witvliet, Eric J. Tchetgen Tchetgen, Jeffrey S. Morris, Christopher B. Forrest, and Yong Chen on the publication of Real-World Effectiveness of BNT162b2 Against Infection and Severe Diseases in Children and Adolescents in Annals of Internal Medicine.
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>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>2 pm</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>Wednesday</td>
<td>4 pm</td>
<td>Vulcan/OHDSI Meeting</td>
</tr>
<tr>
<td>Wednesday</td>
<td>7 pm</td>
<td>Medical Imaging</td>
</tr>
<tr>
<td>Thursday</td>
<td>8 am</td>
<td>India Chapter</td>
</tr>
<tr>
<td>Thursday</td>
<td>9 am</td>
<td>Medical Devices</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>Strategus HADES 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>9 am</td>
<td>Phenotype Development and Evaluation</td>
</tr>
<tr>
<td>Friday</td>
<td>11 am</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Friday</td>
<td>11:30 am</td>
<td>Steering Group</td>
</tr>
<tr>
<td>Friday</td>
<td>10 pm</td>
<td>China Chapter</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Healthcare Systems Interest Group</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Africa Chapter</td>
</tr>
<tr>
<td>Monday</td>
<td>11 am</td>
<td>Data Bricks User Group</td>
</tr>
<tr>
<td>Tuesday</td>
<td>10 am</td>
<td>Registry</td>
</tr>
</tbody>
</table>
Collaborator Spotlight: Chungsoo Kim

Chungsoo Kim is a PhD candidate in the Department of Biomedical Informatics at Ajou University College of Medicine. He earned his Doctor of Pharmacy degree from the College of Pharmacy of the same university in 2019. His research interests include reliable real-world evidence for medication and prediction of individual drug effects/adverse events based on the OMOP common data models. He is also interested in data/analytics infrastructure for conducting data-driven research.

Since joining OHDSI in 2019, he has participated in and led several research projects at OHDSI. He currently participates in OHDSI working groups, including PatientLevelPrediction and the APAC group. He also served as a tutorial instructor for the 2019 OHDSI Korea International Symposium.

Chungsoo discusses his research focuses, his involvement in the OHDSI community, the growth of OHDSI around the Asia-Pacific region, and plenty more in the latest Collaborator Spotlight.

Can you discuss your research focuses at Ajou University?

The goal of my journey is to achieve patients’ better health through data-driven research. My research interest is broadly focusing on generating reliable real-world evidence, especially on medication. I’m interested in utilizing as much data and as many various methodologies as possible to produce results that ultimately benefit patients. All research I conducted is done using the OMOP CDM.
January 2024 OHDSI Newsletter

The Journey Newsletter (January 2024)

Happy New Year! OHDSI made exciting progress in 2023 around the areas of standardized data, vocabularies and open-source tools, as well as in building collaborations around the world. We do a bit of reflecting on what made 2023 special, and we start to imagine the possibilities for 2024, in the latest edition of The Journey newsletter! #JoinTheJourney

Video Podcast: A Look Back, A Look Ahead

In the latest On The Journey video, Patrick Ryan and Craig Sechion take a final look back at some of the highlights of 2023, and take a quick peak at some goals of 2024. Please join our first community call of the new year on Tuesday, Jan. 9 (11 am ET) for a full look at community goals for 2024.

Video Reflection: 10 Years of OHDSI

On Dec. 16, 2013, George Hripcevich led the official formation of the OHDSI community. Within a month, the first face-to-face meeting was held within the Department of Biomedical Informatics at Columbia University. How did we get from there to a global community of more than 3,800 collaborators? The Dec. 12 community call reflected on 10 years of OHDSI, with a video presentation led by Patrick Ryan.

The presentation highlights several of the firsts in the community, including its first publication (which now has more than 1,300 citations), first symposium in the United States, Europe and the Asia-Pacific region, first open-source tools, and plenty more. It also reflects on some of the clinical impacts made by the OHDSI community.

The video presentation is available here, while the slideshow (which includes the 2023 Year In Review slides) can be found here.

December Publications


CohortMethod

Introduction

CohortMethod is an R package for performing new-user cohort studies in an observational database in the OMOP Common Data Model.

Features

- Extracts the necessary data from a database in OMOP Common Data Model format.
- Uses a large set of covariates for both the propensity and outcome model, including for example all drugs, diagnoses, procedures, as well as age, comorbidity indexes, etc.
- Large scale regularized regression to fit the propensity and outcome models.
- Includes function for trimming, stratifying, matching, and weighting on propensity scores.
- Includes diagnostic functions, including propensity score distribution plots and plots showing covariate balance before and after matching and/or trimming.
- Supported outcome models are (conditional) logistic regression, (conditional) Poisson regression, and (conditional) Cox regression.
Empirical Calibration

Introduction

This R package contains routines for performing empirical calibration of observational study estimates. By using a set of negative control hypotheses, we can estimate the empirical null distribution of a particular observational study setup. This empirical null distribution can be used to compute a calibrated p-value, which reflects the probability of observing an estimated effect size when the null hypothesis is true, taking both random and systematic error into account, as described in the paper Interpreting observational studies: why empirical calibration is needed to correct p-values.

Also supported is empirical calibration of confidence intervals, based on the results for a set of negative and positive controls, as described in the paper Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data.

Features

- Estimate the empirical null distribution given the effect estimates of a set of negative controls.
- Estimate the calibrated p-value of a given hypothesis given the estimated empirical null distribution.
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
HADES Development Updates: PheValuator 2.2.11

PheValuator

PheValuator is part of HADES.

Introduction

The goal of PheValuator is to produce a large cohort of subjects each with a predicted probability for a specified health outcome of interest (HOI). This is achieved by developing a diagnostic predictive model for the HOI using the PatientLevelPrediction (PLP) R package and applying the model to a large, randomly selected population. These subjects can be used to test one or more phenotype algorithms.

Process Steps

The first step in the process, developing the evaluation cohort, is shown below:

**Step 1: Develop Evaluation Cohort from Diagnostic Predictive Model**
A Toxin Vocabulary for the OMOP CDM

**MONDAY**

**A Toxin Vocabulary for the OMOP CDM**

(Maksym Trofymenko, Polina Talapova, Tetiana Nesmiian, Andrew Williams, Denys Kaduk, Max Ved, Inna Ageeva)

**UNLOCKING Environmental Health Research with OMOP**

The **FIRST HIERARCHY** for toxins in OMOP:
- **ONT vocabulary**
- **3000+ exposures**
- **41,000+ synonyms**
- **77,000+ internal associations**
- **1800 mappings to OMOP Vocabularies**

**OMOP Toxin Vocabulary**

**External Relationships**
- Maps to: SNOMED, RxNorm, and ReForm Extension
- Has molecular location
- Has tissue location
- Has influence on process

**OMOP Curation**
- Curation by Curation Committee
- Curation by Curation Committee
- Curation by Curation Committee

**Internal Relationships**
- Has measurement
- Has measurement
- Has measurement

**THE TEAM**
- Maksym Trofymenko, Polina Talapova, Tetiana Nesmiian, Andrew Williams, Max Ved, Inna Ageeva
Developing a perinatal expansion table for the OMOP common data model

External validation using clinical domain knowledge from the SNOMED medical terms hierarchy

(LH John, EA Fridgeirsson, JA Kors, JM Reps, PR Rijnbeek)

**External validation using clinical domain knowledge from the SNOMED medical terms hierarchy**

**Background**

External validation is crucial for ensuring the reliability of prediction models on new data. However, performance often declines during external validation due to database heterogeneity caused by variations in record collection, regulatory guidelines, and database purposes. [3]

**Use Case**

Figure 1 depicts a hypothetical model developed on the Integrated Primary Care and Information, a Dutch GP dataset, with predictors Heart Failure, Depression, and COPD, which cannot be applied to a patient from an external database who has slightly different diagnoses. However, considering the contextual similarity, a medical expert may have been able to apply the model based on clinical domain knowledge.

![Incompatible model and patient record due to database heterogeneity.](image)

This work aims to utilize embeddings to approximate clinical concepts, specifically in the context of predicting dementia in persons aged 50-65 in the next five years. This approach may enable external validation of a model even when an exact match for predictors is not found in a patient's record.

**Methods**

Clinical domain knowledge is encoded in our vocabulary hierarchies. For example, SNOMED provides over one million ancestor-descendant relationships. Figure 2 shows a subset of 177 SNOMED relationships with the ancestor concept Clinical Finding as tree root. In this work, we embed the SNOMED hierarchy to obtain a latent space in which terms that resemble one another are positioned closer to each other, which will allow us to approximate missing concepts.

Nickel & Kiela introduced an efficient method to embed hierarchical data, such as the SNOMED hierarchy, into a lower-dimensional manifold (2) hierarchical data follows a tree structure. The number of descendants exponentially increases with distance from the root. To address the limitation of growing hierarchical data, which can exceed the available Euclidean space in Euclidean embeddings and cause overfitting if we attempt to solve it by adding more dimensions, Nickel & Kiela proposed using hyperbolic space instead. Hyperbolic space is characterized by constant negative curvature and is described by hyperbolic geometry. For this study, we will use the hyperbolic Poincaré disk model to embed our hierarchical data.

![Subset of SNOMED medical terms hierarchy with the concept Clinical Finding as the root.](image)

**Results**

![Discrimination of logistic regression using traditional concepts (left) and using the embeddings (right).](image)

We develop and externally validate logistic regression and gradient boosting models across five databases: Integrated Primary Care and Information, IBM MarketScan®, Medicare Supplemental, Icos Disease Analyzer Germany, Optum® de-identified Clininformatics® Data Mart, and Optum® de-identified Electronic Health Record. For development, the hyperbolic embeddings are mean aggregated to be passed into the models as input. We use conditions as sole predictors, which may result in relatively low discrimination performance as compared to models using also demographic information such as age.

![Discrimination of gradient boosting using traditional concepts (left) and using the embeddings (right).](image)

**Contact**

ljohn@erasmusmc.nl

---

We appreciate the feedback provided by @OHDSI, www.ohdsi.org, @JoinTheJourney, @OHDSI, www.ohdsi.org, and #JoinTheJourney. The content is based on our research and experiences, and any similarity with other works is coincidental.
Estimating the comparative risk of kidney failure associated with intravitreal anti-vascular endothelial growth factor (anti-VEGF) exposure in patients with blinding diseases

(Cindy X. Cai, Mary Grace Bowring Diep Tran, Paul Nagy, Michael Cook, Akihiko Nishimura, Jia Ng, Marc A. Suchard, Scott L. DuVall, Michael Matheny, Asieh Golozar, Anna Ostropolets, Evan Minty, Fan Bu, Brian Toy, Will Halfpenny, Michelle Hribar, Jody-Ann McLeggon, Thomas Falconer, Linying Zhang, Laurence Lawrence-Archer, George Hripcsak)
FRIDAY

Characteristics and outcomes of over a million inflammatory bowel disease subjects in seven countries: a multinational cohort study

(Chen Yanover, Ramit Magen-Rimon, Erica Voss, Joel Swerdel, Anna Sheahan, Nathan Hall, Jimyung Park, Rae Jae Lee, Sung Jae Shin, Seung In Lee, Kyung-Joo Lee, Thomas Falconer, Leonard Haas, Paul Nagy, Mary Bowring, Michael Cook, Steven Miller, Tal El-Hay, Maytal Bivas-Benita, Pinchas Akiva, Yehuda Chowers, Roni Weisshof)

INTRO
Crohn’s disease (CD) and ulcerative colitis (UC) are inflammatory bowel diseases (IBD) with consistently increasing incidence rates. These conditions significantly impact the quality of life of patients and families.

METHODS
• Study design: A multinational cohort study using routinely collected healthcare data from 16 OMOP-compliant databases (DBs).
• Study population: IBD cohorts include individuals with ≥2 IBD Dx or with IBD Dx + IBD medication Rx; CD and UC cohorts also require at least one diagnosis of the corresponding disease and none of the other.
• Characteristics, outcomes: Predefined features (demographics, condition groups, drug era groups) + 100 IBD-specific features during subjects’ entire history, 1Y before index date; 1M, 1.5, 2, 3, 5, 10Y and all-time following index date.

RESULTS

Characteristics and Outcomes of >1M Inflammatory Bowel Disease Patients

Disease Trajectory of Crohn’s Disease and Ulcerative Colitis Patients from Australia, Korea, Japan, the UK, Germany, France, and the USA

LIMITATIONS
• Potential differences in coding, reporting across DBs
• Concept sets outdated (vocabulary updates)
• HUGE amounts of data (>2G), challenging to view, handle
• Only binary attributes; no cross-strata info

CONTRIBUTORS

Chen Yanover, Tal El-Hay, Maytal Bivas-Benita, Pinchas Akiva, Yehuda Chowers, Roni Weisshof, Mary Bowring, Michael Cook, Steven Miller, Tal El-Hay, Maytal Bivas-Benita, Pinchas Akiva, Yehuda Chowers, Roni Weisshof

@OHDSI
www.ohdsi.org
#OHDSISocialShowcase
This Week

#JoinTheJourney

OHDSI
Opening: Data Steward at EBMD

Description

Are you looking for a job where you can make a difference and work in a non-profit?
Would you like to be a part of an ambitious and international organisation on the cutting edge of science?
Then this position might be right up your alley.

The EBMT is a non-profit medical and scientific organisation which hosts a unique patient registry providing a pool of data to perform studies and assess new trends.

OUR MISSION
Save and improve the lives of patients with blood-related disorders.

The Registry
Holding the data of over half a million patients, the EBMT registry is the starting point for all studies carried out through the EBMT working parties. The department focuses on data collection processes, data quality monitoring, and maintenance of the database.

YOUR MISSION
Responsible for collecting, collating, and evaluating issues and problems with data and enforcing data usage policies.

RESPONSIBILITIES AND TASKS

Data Stewardship:
- Design, implementation and testing of new data collection processes including data collection forms (DCFs) development.
- Take care of the mapping of new items from DCFs to the OMOP CDM
- Providing input on data quality reports
- Check and clean data on request and ad hoc.
- Data retrieval including designing data reports and data report running.
- Carry out computerized system validation activities.
- Supporting consolidation/harmonization of data
- Creating standard data definitions, and maintain a consistent use of data assets across the organization
- Documenting data policies and data standards
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?