Recent OHDSI Publications

OHDSI Community Call
Dec. 5, 2023 • 11 am ET
## Upcoming Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<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>
<tr>
<td>Dec. 26</td>
<td>No Call</td>
</tr>
<tr>
<td>Jan. 2</td>
<td>No Call</td>
</tr>
<tr>
<td>Jan. 9</td>
<td>Welcome Back! What Can OHDSI Accomplish in 2024?</td>
</tr>
</tbody>
</table>
Happy 10th Birthday to OHDSI
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
Congratulations to the team of Rupa Makadia, Azza Shoaibi, Gowtham Rao, Anna Ostropolets, Peter Rijnbeek, Erica Voss, Talita Duarte-Salles, Juan Manuel Ramírez-Anguita, Miguel Mayer, Filip Maljković, Spiros Denaxas, Fredrik Nyberg, Vaclav Papez, Anthony Sena, Thamir Alshammari, Lana Lai, Kevin Haynes, Marc Suchard, George Hripcsak, and Patrick Ryan on the publication of Evaluating the impact of alternative phenotype definitions on incidence rates across a global data network in JAMIA Open.

Research and Applications

Evaluating the impact of alternative phenotype definitions on incidence rates across a global data network

Rupa Makadia, PhD1,2,*, Azza Shoaibi, PhD1,3, Gowtham A Rao, MD1,2, Anna Ostropolets1, MD1,3, Peter R. Rijnbeek, PhD1,4, Erica A. Voss1, MPH1,5, Talita Duarte-Salles1, PhD1,5, Juan Manuel Ramírez-Anguita, PhD1, Miguel A. Mayer, MD1, Filip Maljkovic1, MS1, Spiros Denaxas1, PhD1,2, Fredrik Nyberg1, PhD1,2, Vaclav Papez1, PhD2, Anthony G. Sena1, BA1,2,4, Thamir M. Alshammari, PhD1,1,1, Lana Y. Lai1, PhD1,1,1, Kevin Haynes, PharmD1, Marc A. Suchard1, MD1,4, George Hripcsak1, PhD2, and Patrick B. Ryan1, PhD2,3,4

1OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY 10027, United States; 2Global Epidemiology, Janssen Pharmaceutical Research and Development, LLC, Titusville, NJ 08096, United States; 3Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY 10032, United States; 4Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, 3000 CA, The Netherlands; 5Fundació Universitariat per a la Recerca i l’Assistència Princep de Salut Jordi Gol i Buvat (IRAP-JGB), Barcelona, 08035, Spain; 6Research Programme on Biomedical Informatics (IRBIB), Hospital del Mar Medical Research Institute (IMMAR), Barcelona, 08035, Spain; 7Management Group Department, Parc de Salut Mar (IPSAM), Barcelona, 08035, Spain; 8Research and Development, Wellcentrics, Alexandria, VA 22314, United States; 9Institute of Health Informatics, University College London, London, NW1 2DT, United Kingdom; 10School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, 43190, Sweden; 11College of Pharmacy, Prince Sultan Biomedical University, Riyadh, 11432, Saudi Arabia; 12Division of Informatics, Imaging, and Data Sciences, University of Manchester, Manchester, M13 9PL, United Kingdom; 13Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90025, United States

*Corresponding author: Rupa Makadia, PhD, Janssen Pharmaceutical Research and Development, LLC, 1101-200 Twenty Harrington Rd, Titusville, NJ 08096 (makadia@srl.com)

Abstract

Objective: Developing accurate phenotype definitions is critical in obtaining reliable and reproducible background rates in safety research. This study aims to illustrate the differences in background incidence rates by comparing definitions for a given outcome.

Materials and Methods: We used 16 data sources to systematically generate and evaluate outcomes for 13 adverse events and their overall background rates. We examined the effect of different changes in the phenotype definition on background rates.

Results: The incidence rate range of five different phenotype definitions varied across outcomes, with an increase in incidence rate from 1 to 11.53. Standardization of code sets ranged from 1 to 1.54, and code set changes ranged from 1 to 2.52.

Discussion: The model that has the highest impact is required to implement a positive change in the outcome incidence rate. Standardization showed almost no change when using source code variations. The strength of this effect in the incremental reduction is highly dependent on the outcome. Changing definitions from broad to narrow showed the most variability by aggregating different source code sets.

Conclusion: The model that has the highest impact is required to implement a positive change in the outcome incidence rate. Standardization showed almost no change when using source code variations. The strength of this effect in the incremental reduction is highly dependent on the outcome. Changing definitions from broad to narrow showed the most variability by aggregating different source code sets.

@OHDSI www.ohdsi.org #JoinTheJourney
OHDSI Shoutouts!

Congratulations to the team of Lin Lawrence Guo, Maryann Calligan, Emily Vettese, Sadie Cook, George Gagnidze, Oscar Han, Jiro Inoue, Joshua Lemmon, Johnson Li, Medhat Roshdi, Bohdan Sadovy, Steven Wallace, and Lillian Sung on the publication of Development and validation of the SickKids Enterprise-wide Data in Azure Repository (SEDAR) in Heliyon.
OHDSI Shoutouts!

Congratulations to the team of Manuel Rueda, Ivo Leist, and Ivo Gut on the publication of Convert-Pheno: A software toolkit for the interconversion of standard data models for phenotypic data in the Journal of Biomedical Informatics.
Congratulations to Craig Mayer on the publication of Conversion of CPRD AURUM Data into the OMOP Common Data Model in Informatics in Medicine Unlocked.
OHDSI Shoutouts: 1st UK Study-A-Thon 💥
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>8 am</td>
<td>Psychiatry</td>
</tr>
<tr>
<td>Wednesday</td>
<td>4 pm</td>
<td>Vulcan/OHDSI Meeting</td>
</tr>
<tr>
<td>Thursday</td>
<td>9:30 am</td>
<td>Themis</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>Methods Research</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>GIS – Geographic Information System Development</td>
</tr>
<tr>
<td>Friday</td>
<td>9 am</td>
<td>Phenotype Development &amp; Evaluation</td>
</tr>
<tr>
<td>Friday</td>
<td>1 pm</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Friday</td>
<td>10 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>Monday</td>
<td>4 pm</td>
<td>Eyecare &amp; Vision Research</td>
</tr>
<tr>
<td>Tuesday</td>
<td>9 am</td>
<td>OMOP CDM Oncology – Genomic Subgroup</td>
</tr>
</tbody>
</table>
Dec. 10 Career Speaker Series: Fan Bu
organized by the Early-Stage Researchers WG

Fan Bu, the soon-to-be Assistant Professor in Biostatistics at the University of Michigan, will be the featured guest at the Dec. 10 (11 am) Career Speaker Series event.

Fan is a leading researcher in OHDSI’s vaccine safety surveillance collaboration with the FDA CBER Best Initiative and has collaborated on several OHDSI network studies.

Fan Bu
Postdoctoral Researcher, UCLA

MONDAY
DEC 11, 2023

TIME
11 AM - 12 PM EST

JOIN: MS TEAMS

bit.ly/OHDSILeaders
Collaborator Spotlight: Alison Callahan

Alison Callahan, an Instructor in the Center for Biomedical Informatics and Clinical Data Scientist in the Stanford Health Care Data Science Team, discusses her career journey, how OHDSI impacts her research at Stanford, critical knowledge gaps that can be addressed by the Perinatal and Reproductive Health workgroup, and more in the latest edition of the Collaborator Spotlight.

ohdsi.org/spotlight-alison-callahan
Community Updates

Where Have We Been?
- We honored both our Titan Award winners and our Best Community Contribution honorees from the 2023 Global Symposium during community calls this past week. Several of the open-source software presenters also provided live demos of their work. Check the November presentations below to find all of these talks, or check out the collaboration resources page to find all of the research from the global symposium.
- Members of the OHDSS community provided a full-day workshop entitled "OHDSS RAVE Revolution: Springing Data Modernization with Harmonized Standards for Cutting Edge Health Research" during the 2022 ANA Symposium. The full schedule is available here. Thank you to everybody involved in leading or taking part in this workshop.

Where Are We Now?
- We are approaching one full decade of OHDSS in action. Please join our Dec. 12 Community call to reflect on 10 years of OHDSS — how the community formed and grew, and what we achieved through collaboration and open science along the way.
- Andrew Williams posted a piece (a book that describes how OHDSS workflows form, the kinds of things they do, and the categories of workgroups that the OHDSS community organized into). It is meant to help community members understand how to use OHDSS workgroups to get things done. It also provides some tips on running workgroups. There is a request for comments until Jan. 1, 2024. To read more and access the document, please visit the forum post.

The Journey Newsletter (December 2023)

OHSI officially formed on Dec. 16, 2013, and over the past decade, more than 3,700 people joined the journey to collaboratively generate the real-world evidence that promotes better health decisions and better care. We will celebrate a decade of OHDSS during our Dec. 12 community call, and we provide a brief look back in our newsletter posted below. We also highlight papers, presentations, and other updates from the last month in this newsletter. Check the Journey.

Video Podcast: 10 Years of OHDSS

In the latest On The Journey video, Patrick Ryan and Craig Sachson reflect on 10 years of OHDSS. They discuss the origin and how collaboration has happened, and they play a game of 'there and now' around OHDSS's main collaborative successes. Please join the Dec. 12 community call for a full celebration of 10 years of OHDSS.

December Newsletter is Available

On Dec. 18, 2013, George Hipsas 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 (see photo above). There was a sense of hopefulness and belief that this initiative and its mission to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care could make an impact.

They will also be the first to admit they had no idea just how far they would come in one decade, including:
- welcoming 3,800 researchers from across 83 countries
- mapping 534 data sources from 48 countries to the OMOP Common Data Model to cover more than 86 million patient records
- developing and maintaining more than 40 open-source tools to enable and empower globally shared research
- impacting both clinical and regulatory decision-making around the world

November Publications


OHDSI 2023 Plenary Video: Improving the reliability and scale of case validation

OHDSS Symposium Plenary: Improving the reliability and scale of case validation

Validation is regarded as a necessary element of regulatory-grade but conducting case validation in a time- and resource-intensive, har
young conducted in such a way that does proper quantitative bias analyses.

mailchi.mp/ohdsi/november2023
Welcome to OHDSI!

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

2023 Global Symposium Posters Now Available

Thank you to everybody who joined the 2023 Global Symposium. All talks, demos and more are now available on the symposium homepage.

mailchi.mp/ohdsi/december2023
Make Your Tools Work for You: Customizing the Data Quality Dashboard to Identify Changes in Source Data

Melanie Philofsky, RN, MS Odysseus Data Services

Background

One of the main challenges in ensuring source data are comprehensively and accurately transformed to the OMOP CDM is identifying changes to source data and updating the extract, transform, and load (ETL) logic before the CDM is released to researchers. Therefore, one of the most important steps in the process is running the DQD tool on the OMOP CDM before making these data available for use.

Methods

The DQD is pre-configured with threshold failure rates which might not be representative of the data in your CDM. However, these threshold failure rates are adjustable. The three categories of checks: completeness, conformance, and plausibility. The completeness checks assess the percentage of data expected for a field. Completeness is dependent on the source data and the threshold should be adjusted to a level representative of the source data for a given query.

- Adjust the thresholds which fail the completeness checks to a level representative of your source data.
- Adjust any predefined thresholds to a level just above your current failure rate in order to identify changes in your source data.

We set the threshold levels of the DQD checks to 1% greater than the current failure rate for a field level check. This was done to ensure minor changes in the completeness of the source data would trigger a DQD failure notification for a particular check.

Results

The following changes to the source data or OHDSI vocabularies were identified after the tightening of the DQD failure thresholds:

- Addition of a new source field for the required gender_concept_id field in the person table. Upon analysis of the failure, it was discovered the value set for a person's sex had changed and a new source field where a person's biological sex is stored was added. The ETL was altered to bring in data from this newly discovered field and completeness of the gender concept id field check rose to >99%.

- Changes to the usual population whose data contribute to a dataset. A drop in the completeness percentage for a Person's race, gender, and ethnicity field level checks lead to an investigation of the source data and subsequent discovery of many persons in the OMOP CDM who lack demographic data and have sparse clinical data. Sparse clinical data are defined by less than 3 clinical event records for a person in the OMOP CDM. Since the OMOP CDM is designed for longitudinal research studies, persons with sparse clinical data are deemed not suitable for research. Persons with sparse clinical data will be removed from the CDM to increase fidelity.

- Changes in mapping to a standard concept_id in a new vocabulary. Analysis of an increase in the completeness failure rate for the Condition Occurrence table lead to the identification of a change in the mapping of a non-standard concept_id to a new standard concept_id in a vocabulary not yet downloaded from Athena. This failure identified the need to download an additional vocabulary from Athena.

- Change in source data values used in custom mapping data elements ETL to the CDM. Some domains in an EHR do not have coded data elements. Therefore, these source values must be manually mapped using an exact text string match to a standard concept_id. When there is a change in the source values, these data must be manually remapped.

Conclusion

Adjusting the DQD threshold levels to just above current failure rates assists data owners in ensuring data integrity remains high as changes to source data field use, collection of data, standard vocabulary changes, and source value sets evolve.
Assessment of Pre-trained Observational Large Longitudinal models in OHDSI (APOLLO)

(Martijn Schuemie, Yong Chen, Egill Fridgeirsson, Chungsoo Kim, Jenna Reps, Marc Suchard, Xiaoyu Wang, Chao Pang)

TUESDAY

Background
Large language models (LLMs) have recently received significant attention because of their ability to comprehend complex linguistic structures, enabling, among other things, ChatGPT to participate in human-like conversations. These models have been applied to various domains, extending beyond text to include images processing, as exemplified by projects like DALL-E and MidJourney.

The Assessment of Pre-trained Observational Large Longitudinal models in OHDSI (APOLLO) project aims to explore the feasibility of employing pre-trained models in the analysis of large healthcare databases, including electronic health records and administrative claims. The main form of these databases is time-stamped sets of codes, such as diagnosis codes, procedure codes, drug codes, and other time-stamped values, such as laboratory measurements.

Approach
Deep learning models like LLMs are commonly used in two stages: pre-training on a large dataset, often millions of persons’ data in the OMOP Common Data Model (CDM), where the model learns to predict withheld information. This can be done in a forward-only (similar to GPT) or bidirectional (like BERT) manner, as shown in Figure 1. Then, fine-tuning defines the pre-trained model for a specific task.

Potential fine-tunable tasks
After pre-training, the model may be fine-tuned for a wide range of tasks, which may include:
- Patient-level prediction, where a pre-trained model may prove more accurate with less training data than current non-pre-trained models.
- Missing value imputation, which is almost identical to the bidirectional pre-training task.
- Phnetaging, which can be thought of as a type of imputation.
- Patient clustering, where nodes in the hidden layers may represent subgroups of interest.
- Causal effect estimation, either by using the model for computing propensity scores, or directly eliciting effects learned by the pre-trained model.
- Counterfactual prediction, gives a choice between various treatment options, what is expected to happen to a patient in the future, for each treatment option?

We also suspect more potential applications will become apparent in the future.

Architecture
The current pre-training architecture, illustrated in Figure 2, utilizes the OHDSI DatabaseConnector R package to establish a connection with the CDM database and extract data for either a sample or the full set of persons. Subsequently, these data are stored locally in the efficient Apache Parquet format. The stored data comprises a subset of the CDM tables and columns, encompassing all clinical domain tables (except the notes tables), and includes persons 5%, visit occurrence 5%, concept IDs, numeric values, and several vocabulary tables.

Because research in LLMs is done almost exclusively in Python, the remainder of the pipeline is implemented in Python. A pre-processing script converts the CDM data to sequence-information per person, before fitting the model using the PyTorch library.

Planned evaluation
Initial evaluations will use simulated data only. We have developed a simple simulator that uses a hidden-state Markov model to generate data in CDM format, including data for fine-tuning prediction tasks.

Subsequent evaluations for patient-level prediction and causal effect estimation in real-world settings will rely on existing OHDSI benchmarks. For other tasks new benchmarks will be developed. Where possible, performance will be compared to the current state-of-the-art, such as the algorithms implemented in the OHDSI PatientLevelPrediction package, and the CohortMethod package using large-scale propensity scores.

There are many analysis choices when developing general pre-trained models, as well as when fine-tuning.

These include:
- Types of pre-training task: unifactorial or bidirectional? Predicting the next, framed event by choosing among all possible events, or by choosing among a limited set of candidates automatically selected for the training?
- Choice of input and output representation, including:
- How to represent elapsed time between events?
- Whether and how to encode and embed time, age, seasons, the day of the week, etc.
- Which features to include. Should only the most prevalent concepts be included? Should drugs be mapped to ingredients? Etc.
- Model architecture, such as number of layers and number of nodes per layer, but also choice of activation functions.
- Training parameters, such as regularisation, learning rate, and number of epochs.
- Data sources to use.

A set of combinations of these choices will need to be established and evaluated using the benchmarks.

Feasibility study
In a feasibility study, the GeneralPretrainedModelTools package was used to take a sample of two million persons from the Mative Metakarst CCAC database. Download took 1.8 hours, and Parquet files take 3.6GB of disk space. Pre-processing took 50 minutes, resulting in Parquet files totaling 3.5GB. A single epoch of pre-training took 20 hours on an NVIDIA A100 for a 122-million-parameter model.

Conclusions
Despite existing uncertainties surrounding the applications of large pre-trained models to healthcare data as a whole, the potential for transformative impacts is promising.
#OHDSISocialShowcase This Week

**A use case of OHDSI's ATLAS tool in a biobank-scale GWAS pipeline**

Craig C. Teerlink, Hamid Saoudian, Richard Boyce, Philip S. Tsao, Michael E. Matheny, Marc A. Suchard, Kyle M. Hernandez, Robert Grossman, Scott L. DuVall

## Background
- Infrastructure and data governance limitations have prevented widespread use of Million Veteran Program (MVP) data among the research community.
- The VA Data Commons was introduced as a solution for scaling up access to MVP data and significantly boost computational capabilities.
- The VA Data Commons is a cloud-based analytic environment that allows VA-credentialed research teams to securely access and perform genome-wide association studies (GWAS) using MVP data.
- The VA Data Commons contains data on ~500K subjects, including:
  - Frequency of the number of times a Veteran was diagnosed within PhenoType categories.
  - Prescriptions/medications for both inpatient and outpatient settings.
  - Frequency, min, max, mean, and median of Labs performed in VA.
  - Demographic information.
  - MVP survey data.
- Data was converted to the OMOP common data model.
- Our goal is to provide a "no-code" environment for users to create cohorts and run GWAS. We have incorporated Observational Health Data Sciences and Informatics (OHDSI)’s ATLAS tool for phenotype and covariate selection.
- OHDSI’s ATLAS tool is a Graphical User Interface (GUI) that removes the need to code.
- ATLAS can be used by researchers to construct complex phenotype definitions.
- ATLAS is widely used among clinical researchers (supports reproducibility).

## VA Data Commons GWAS pipeline workflow

1. **Create an overall study cohort**
2. **Define concept sets to specify inclusion and exclusion criteria for dichotomous phenotype and covariate cohorts**
3. **Use cohort definitions based on concept sets to define phenotypes and covariates**
4. **Once cohorts are generated, they are available for other programs to use**
5. **Once created, cohorts are sent to the GWAS app**
6. **View results of the GWAS**

## We use ATLAS to define cohorts for the GWAS

- **FIRST:** Define an overall cohort to study (e.g., all MVP subjects).
- **SECOND:** Specify the logical criteria for dichotomous phenotypes.
- **THIRD:** Specify dichotomous covariates.

## Conclusions

As the VA Data Commons is made available to VA and non-VA-credentialed users in the near future, we anticipate that users will have powerful phenotyping capability due to the incorporation of the ATLAS tool, which will optimize wide-spread utilization of the MVP data set.

Contact: craig.teerlink@va.gov
Data-driven assessment of mental health among children and adolescents with food allergy

INTRO:
Children with chronic diseases like asthma, diabetes, and obesity are at a higher risk of developing mental health disorders than their healthy peers. However, the mental consequences of food allergies (FA), which are on the rise but not classified as chronic diseases, remain insufficiently researched.

OBJECTIVE:
To examine the association between food allergies and mental health in children and adolescents. To compare the mental well-being of those with FA to those without, as well as to children with common chronic conditions.

CONCLUSIONS:
This large population-based study indicates an elevated risk of developing anxiety, depression, and eating disorders in children with FA compared to controls without FA. Eating disorders remain notably prevalent among food-allergic children compared to other disease groups, with statistical significance particularly evident compared to the atopic dermatitis group.

RESULTS
Table 1: Characteristics of FA cohorts vs. no FA and other chronic conditions

METHODS
Study population: Patients aged 0-18 years from 2000 to 2021.
Five cohorts were defined based on the occurrence of a specific condition: food allergy, asthma, atopic dermatitis, and type 1 diabetes (without a history of FA), and a cohort consisting of a population without FA was defined, using random physician visits as an index date.
Matching: FA patients were matched to no FA cohorts by age, sex, and index date (1:3 for all conditions, except from T1D, where it’s 1:1).

Three mental health outcomes: Diagnoses of anxiety, depression, and eating disorders after the index year.

Data source: IQVIA Medical Research Data, IMRD contains longitudinal non-identified patient electronic healthcare records collected from UK General Practitioner clinical systems incorporating data from THIN, a Cegedim database.

Statistical analysis:
Time to outcome was described with Kaplan-Meier (KM) curves and compared using a log-rank test with robust variance estimation to account for matching. The Cox regression model was used to estimate adjusted hazard ratios (HR) while controlling for age and sex.
Quantification of Symptom Documentation on Disease Diagnosis Date in Structured Claims Data: An Application of the OHDSI Phenotype Library

**INTRO:**
- Patient symptoms are important data elements that can be used in various clinical research applications. However, it is unclear if symptoms are documented in observational data sources.
- We quantified the occurrence of symptom codes, such as fever, cough, and fevers, on the same day as a defined related disease diagnosis in structured health insurance claims data.
- We demonstrated how to use the Observational Health Data Sciences and Informatics (OHDSI) Phenotype Engine (PE) to carry out a study within the OHDSI network.

**METHODS:**
- We utilized phenotype definitions from the OHDSI PE version 3.15.0.
- We selected 11 acute severe clinical conditions that are expected to have a sudden onset and short latent or incident period.
- Cohort IDs representing symptom phenotypes expected for each selected acute disease were identified in the OHDSI Phenotype Library based on clinical expertise.
- We retrieved disease and symptom cohorts from the OHDSI PhenotypeLibrary Pkg package using the function `PhenotypeLibrary?pckg/phenodefinition.txt`.
- The `PhenotypeLibrary&phenodefinition.txt` object was used in CohortGenerator to instantiate the initial cohort.
- The study was run on free US claims data sources: The CARDIAlink Adapted National Plan Claims Data (PhenX/Pro2), Optum’s Enrollight Environmental Health Outcomes (EHO), The MarketScan Commercial Database (CCN), The MarketScan MarketScan Healthcare Claims Database (HCDC), and The MarketScan MarketScan Medicare Supplemental Database (MSMCD).

**RESULTS:**
- 

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
<th>% of Total</th>
</tr>
</thead>
</table>
| Acute severe illness without COPD | 521 | 32%
| Acute severe illness with COPD | 291 | 18%
| Acute severe illness with COPD and symptom | 180 | 11%
| Acute severe illness without symptom | 140 | 9%
| Acute severe illness with symptom but not COPD | 130 | 8%

*Data in Table 1 are from OHDSI PE version 3.15.0*.

**DISCUSSION:**
- Our results indicate that a large proportion of individuals with acute illness have at least one symptom code recorded simultaneously in claims data.
- Except for one acute disease (27: Acute illness without COPD), in one data source (HCDC), syndrome occurrence was observed in at least 50% of individuals.
- Women were more likely to report symptoms as input variables in future research.
- Our work sets the groundwork for further work to improve patterns and differences in symptom documentation, thus opening avenues for a deeper understanding of healthcare utilization and patient’s journey and health outcomes.
- This study demonstrates an application of the OHDSI PhenotypeLibrary to OHDSI network studies and its syncing with standard OHDSI software.

**REFERENCES:**

- Gowtham Rao, Azza Shoaibi

---

**Large proportion of individuals with acute illnesses have at least one symptom code recorded simultaneously in claims data.**

*The OHDSI PhenotypeLibrary makes cohort definitions referenceable and reusable.*

---

**METHODS, useful:**
- For calculating or occurrence of symptoms on the same day disease, we utilized the unpublished Cohort/Outcome branch of FeatureDetection.
- We created a composite cohort representing the occurrence of any expected symptoms for each disease. We achieved this by utilizing the CohortDetection package.
- We calculated the prevalence of each variable in each condition across the data sources.

**FRIDAY**

#OHDSISocialShowcase This Week
OHDSI HADES releases: DatabaseConnector 6.3.1

DatabaseConnector 6.3.1

Bugfixes:
1. Fixed `dbFetch()` for DBI drivers, no longer ignoring `n` argument.
2. Fix bulk import for Postgres on MacOs.

DatabaseConnector 6.3.0 2023-11-08

Changes:
1. On Snowflake always using `QUOTED_IDENTIFIERS_IGNORE_CASE=TRUE` to avoid name mismatches when using quotes.
2. Updated Redshift drivers.
3. Added unit tests for all supported platforms.

Bugfixes:
1. Fix bug on BigQuery where wait time was too short to avoid rate limit error.

DatabaseConnector 6.2.4 2023-09-07
Strategus sub-team formation

In the HADES Working Group, we’ve discussed and decided to form a sub-team focused on the design of Strategus software for OHDSI network studies. There has been a lot of discussion of Strategus here on the forums link, in the HADES workgroup, the Save Our Sisyphus Challenge, the 2023 OHDSI Hack-a-thon and of course on the Strategus GitHub Issue Tracker.

Now we’d like to formalize the work around the Strategus project into a sub-team of the HADES Working Group and we want to open this up to developers in the OHDSI community that are interested in collaborating. I have opened a poll on the HADES Working Group OHDSI Teams Channel to see who is interested in meeting and some options for meeting days/times. Please feel use that link to vote and to join the sub-team! I’m aiming to start this sub-team in January 2024.

(If you don’t have access to the OHDSI Teams environment, please see: OHDSI Workgroups – OHDSI and click the “Join A Workgroup” link)
Opening: Limerick Digital Cancer Research Centre

Navigation Section

- Search
  Click here to search for vacancies. Please email us@ohdsi.org if you experience difficulties.
- Login
  Click to Login. Staff, you must be logged in to view Internal Vacancies
- Register
  Click to Register
- Vacancy Details
  Click to view the current job specification
- Terms & Conditions
  Agree to the terms and conditions
- Human Resources Website
  "The University of Limerick is an equal opportunities employer committed to education on merit only."

Job Spec

Advertisement/Information for Applicants

Please click on Information for Applicants/Job Description link below for full job

Post Doctoral Researcher (Level 1 or 2) in Cancer Digital Health Real World Evidence (2 Positions)

With over 18,000 students and 2,600 members of staff, the University of Limerick (UL) is an energetic, research-led and enterprising institution with a proud record in innovation and excellence in education, research and scholarship. The dynamic, entrepreneurial and pioneering values which drive UL's mission and strategy ensure that we capitalise on local, national and international engagement and connectivity. We are renowned for providing an outstanding student experience and conducting leading-edge research. Our commitment is to make a difference by shaping the future through educating and empowering our students.

With the River Shannon as a unifying focal point, UL is situated on a superb riverside campus of over 130 hectares. Outstanding recreational, cultural and sporting facilities further enhance the campus' exceptional learning and research environment.

Applications are invited for the following position:

Faculty of Education & Health Sciences

School of Medicine

Post Doctoral Researcher (Level 1 or 2) in Cancer Digital Health Real World Evidence (2 Positions) Specific Purpose Contract

Salary Scales: PD1 €42,033 - €49,427 p.a. pro rata
PD2 €49,799 - €54,153 p.a. pro rata

Informal enquires regarding the post may be directed to:
Professor Aidan Gough
School of Medicine
University of Limerick
Email: aidan.gough@ul.ie

"This is a professional training and development role and the training and development relevant to this position will be completed within the period of the contract. Postdoctoral Researchers appointed will be expected to complete the Researcher Career Development Programme."

The closing date for receipt of applications is Friday, 15th December 2023.
Applications must be completed online before 12 noon, Irish Standard Time on the closing date.

The University of Limerick supports blended working
Openings: Bill and Melinda Gates Foundation

Distinguished Scientist, Artificial Intelligence & Large Language Models

Deputy Director, Quantitative Sciences
Job Opening: Stanford University

Prospective Postdocs

Open Postdoctoral position, faculty mentor Brian Bateman

How To Apply

Open Postdoctoral Positions

Finding a Faculty Mentor

Cost of Living

Housing

Fellowships at Stanford

Fellowships outside Stanford

Our research team is looking for a postdoctoral scholar in perinatal pharmacoeconomics. The scholar will work closely with Drs. Brian Bateman and Stephanie Leonard on NIH-funded research projects on the comparative safety and effectiveness of medications in pregnancy and related research topics. Our projects employ advanced analytical methods in large databases, which include claims data and electronic health record data in conventional structures and in common data models. Current topical focus areas include mental health, behavioral health and cardiovascular health of people who are pregnant or postpartum.

Important Info

Faculty Sponsor (Last, First Name):
Bateman, Brian

Other Mentor(s) if Applicable:
Stephanie Leonard

Stanford Departments and Centers:
Anesthesiology, Periop & Pain Med

Postdoc Appointment Term:
Initial appointment is 1 year with renewal after the first year for an additional 1-2 years by mutual agreement

Appointment Start Date: Flexible start date

Group or Departmental Website:
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?
Recent OHDSI Publications

Multinational patterns of second line antihyperglycaemic drug initiation across cardiovascular risk groups: federated pharmacoepidemiological evaluation in LEGEND-T2DM

**Lovedeep Dhingra and Arya Aminorroaya** • Postdoctoral Associates, Yale School of Medicine

Scalable Infrastructure Supporting Reproducible Nationwide Healthcare Data Analysis toward FAIR Stewardship

**Chungsoo Kim** • PhD Candidate, Ajou University

Transforming the Information System for Research in Primary Care (SIDIAP) in Catalonia to the OMOP Common Data Model and Its Use for COVID-19 Research

**Berta Raventós** • Predoctoral Researcher, Institut Universitari d'Investigació en Atenció Primària (IDIAP) Jordi Gol

European Health Data & Evidence Network—learnings from building out a standardized international health data network

**Erica Voss** • Senior Director, Janssen Research & Development

Scalable and interpretable alternative to chart review for phenotype evaluation using standardized structured data from electronic health records

**Anna Ostropolets** • Director, Head of Innovation Lab, Odysseus Data Services, Inc