Welcome To Phenotype Phebruary II

OHDSI Community Call
Jan. 31, 2023 • 11 am ET
# Upcoming OHDSI Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<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?
Congratulations to the team of Ines Reinecke, Joscha Siebel, Saskia Fuhrmann, Andreas Fischer, Martin Sedlmayr, Jens Weidner, and Franziska Bathelt on the publication of Assessment and Improvement of Drug Data Structuredness From Electronic Health Records: Algorithm Development and Validation in JMIR Medical Informatics.

**Abstract**

**Background:** Digitalization offers a multitude of opportunities to gain insights into current diagnostics and therapies from retrospective data. In this context, real-world data and their accessibility are of increasing importance to support unbiased and reliable research on big data. However, routinely collected data are not readily usable for research owing to the unstructured nature of health care systems and a lack of interoperability between these systems. This challenge is evident in drug data.

**Objective:** This study aimed to present an approach that identifies and increases the structuredness of drug data while ensuring standardization according to Anatomical Therapeutic Chemical (ATC) classification.

**Methods:** Our approach was based on available drug prescriptions and a drug catalog and consisted of 4 steps. First, we performed an initial analysis of the structuredness of local drug data to define a point of comparison for the effectiveness of the overall approach. Second, we applied 3 algorithms to unstructured data that translated text into ATC codes based on string comparisons in terms of ingredients and product names and performed similarity comparisons based on Levenshtein distance. Third, we validated the results of the 3 algorithms with expert knowledge based on the 1000 most frequently used prescription terms. Fourth, we performed a final validation to determine the increased degree of structuredness.

**Results:** Initially, 47.73% (n=943,980) of 1,768,153 drug prescriptions were classified as unstructured. With the application of the 3 algorithms, the percentage of drug prescriptions was increased to 85.18% (n=1,509,899) based on the 1000 most frequent medication prescriptions. In this regard, the combination of algorithms 1, 2, and 3 resulted in a correctness level of 100% (with 57,264 ATC codes identified), algorithms 1 and 3 resulted in 99.6% (with 152,404 codes identified), and algorithms 1 and 2 resulted in 95.9% (with 39,672 codes identified).

**Conclusions:** As shown in the first analysis step of our approach, the availability of a product catalog to select during the documentation process is not sufficient to generate structured data. Our 4-step approach reduces the problems and reliably increases the structuredness automatically. Similarity matching shows promising results, particularly for entities with no connection to a product catalog. However, further enhancement of the correctness of such a similarity matching algorithm needs to be investigated in future work.
Congratulations to the team of Junqing Xie, James T. Brash, Cigdem Turkmen, Stefan Driessen, Giustino Varrassi, George Argyriou, Sarah Seager, Christian Reich and Daniel Prieto-Alhambra on the publication of Risk of COVID-19 Diagnosis and Hospitalisation in Patients with Osteoarthritis or Back Pain Treated with Ibuprofen Compared to Other NSAIDs or Paracetamol: A Network Cohort Study in Drugs.
Congratulations to the team of Hao Luo, Wallis C. Y. Lau, Yi Chai, Carmen Olga Torre, Robert Howard, Kathy Y. Liu, Xiaoyu Lin, Can Yin, Stephen Fortin, David M. Kern, Dong Yun Lee, Rae Woong Park, Jae-Won Jang, Celine S. L. Chui, Jing Li, Christian Reich, Kenneth K. C. Man, and Ian C. K. Wong on the publication of Rates of Antipsychotic Drug Prescribing Among People Living With Dementia During the COVID-19 Pandemic in JAMA Psychiatry.
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>3 pm</td>
<td>OMOP CDM Oncology Outreach/Research Subgroup</td>
</tr>
<tr>
<td>Wednesday</td>
<td>2 am</td>
<td>Population-Level Estimation (Eastern Hemisphere)</td>
</tr>
<tr>
<td>Wednesday</td>
<td>8 am</td>
<td>Psychiatry</td>
</tr>
<tr>
<td>Wednesday</td>
<td>11 am</td>
<td>Open-Source Community</td>
</tr>
<tr>
<td>Wednesday</td>
<td>12 pm</td>
<td>Health Equity</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>Population-Level Estimation (Western Hemisphere)</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>11 am</td>
<td>Clinical Trials</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>
</tbody>
</table>
OHDSI HADES releases: SqlRender 1.12.0

SqlRender

SqlRender is part of HADES.

Introduction

This is an R package for rendering parameterized SQL, and translating it to different SQL dialects. SqlRender can also be used as a stand-alone Java library and a command-line executable.

Features

- Supports a simple markup syntax for making SQL parameterized, and renders parameterized SQL (containing the markup syntax) to executable SQL.
- The syntax supports defining default parameter values.
- The syntax supports if-then-else structures.
- Has functions for translating SQL from one dialect (Microsoft SQL Server) to other dialects (Oracle, PostgreSQL, Amazon RedShift, Impala, IBM Netezza, Google BigQuery, Microsoft PDW, Snowflake, Azure Synapse, Apache Spark and SQLite).
- Can be used as R package, Java library, or as stand-alone executable through a command-line interface.

Links

- View on CRAN
- Browse source code
- Report a bug
- Ask a question

License

Apache License 2.0

Citation

Citing SqlRender

Developers

Martijn Schuemie
Author, maintainer

Marc Suchard
Author
DatabaseConnector

Introduction

This R package provides function for connecting to various DBMSs. Together with the SqlRender package, the main goal of DatabaseConnector is to provide a uniform interface across database platforms: the same code should run and produce equivalent results, regardless of the database back end.

Features

- Create connections to the various database platforms:
  - Microsoft SQL Server
  - Oracle
  - PostgreSQL
CirceR

CirceR is part of HADES.

Introduction

A R-wrapper for Circe, a library for creating queries for the OMOP Common Data Model. These queries are used in cohort definitions (CohortExpression) as well as custom features (CriteriaFeature). This package provides convenient wrappers for Circe functions, and includes the necessary Java dependencies.

Features

- Convert a JSON cohort expression into a markdown print-friendly presentation.
- Convert a JSON cohort expression into SQL.

Examples

Links

Browse source code at https://github.com/OHDSI/CirceR/
Report a bug at https://github.com/OHDSI/CirceR/issues
Ask a question at http://forums.ohdsi.org

License

Apache License 2.0

Developers

Chris Knoll
Author, maintainer
Martijn Schuemie
Author

Dev status
Anna Ostropolets introduced a vocabulary landscape assessment survey to directly inform which vocabularies and activities the vocabulary team prioritizes in 2023.

The deadline is Feb. 23.
Save Our Sisyphus Challenge

The task of taking a research study from idea through design through execution through publication can seem a daunting challenge, much like rolling a boulder up a hill. That task is all the more challenging when researchers try to go it alone, as each step requires a distinct set of skills. Observational study design requires epidemiologic understanding and statistical methodological expertise. Implementing a study design requires statistical programming ability. Interpreting and reporting results requires domain knowledge of the clinical problem.

But when you are part of the OHDSI community, you never have to go it alone. And as a team effort, what seems an arduous task can become an efficient and effective process.

We are seeking important research questions that you want to contribute and participate in to take from idea to publication. The OHDSI community will provide support through every step of the process, working with you to design an appropriate protocol, implement a network analysis package, execute across OHDSI data partners, and prepare a manuscript for publication. Our goal is to collaboratively complete this network study over the course of 8 weeks across April and May, using the open-source tools and process that OHDSI has.

https://forms.gle/DySfETJPtmwgquKv9
Save the Date! April 21: DevCon

OHDSI DevCon 2022 Welcomes & Mentors New Contributors To Our Open-Source Environment

Watch All Eight Workshops, Talks & The Panel From DevCon Below

The Open-Source Community hosted the first DevCon on Friday, April 22 as a way of accepting and mentoring new contributors to our environment. Organized by Paul Nagy and Adam Black, the event included eight workshops, talks and a panel discussion to both welcome and engage both current and future developers within OHDSI.

All videos from this session have or will be uploaded to this page. A big announcement from DevCon was the formation of the Khiron Contributor Cohort, which will help onboard and mentor open-source developers in the community. If you are interested in joining the effort, please fill out the application.

To learn more about the Khiron Contributor Cohort, please check out the State of the Open Source Community presentation below.

Teams invite will go out at a later date.
Save the Date!
October 20-22, OHDSI Global Symposium

Location and more details coming soon
Next CBER Best Seminar

The CBER BEST Seminar Series returns Wed., Feb. 8, at 11 am ET, as 2022 Titan Award recipient Fan Bu will provide a presentation on Bayesian Safety Surveillance with Adaptive Bias Correction.

Speaker: Dr. Fan Bu (UCLA)

Description: In this presentation, we will discuss a collaborative project with the FDA CBER BEST Initiative to improve on post-market vaccine safety surveillance procedures through Bayesian sequential analysis. Post-market surveillance on approved vaccine products is essential for addressing safety concerns. The goal is to detect rare or high-risk adverse events that often go undetected in clinical trials due to limited sample sizes. Collaborating with FDA CBER, we have developed a Bayesian alternative surveillance procedure that tackles these challenges in sequential analysis of observational data. The standard statistical approach for surveillance is Maximum Sequential Probability Ratio Test (MaxSPRT). Through comprehensive empirical evaluations on large-scale observational healthcare databases, we show that, compared to MaxSPRT, our Bayesian method offers more flexibility on the surveillance schedule, more transparency and interpretability in decision-making, and better error control through statistical correction of bias in observational data.
Job Opening

Job Details

Database Programmer

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD

We are seeking to appoint a highly qualified and dedicated Database Programmer to join the Health Data Sciences research group led by Professor Daniel Pinto-Alhambra at the Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), Oxford.

You will join an outstanding, multi-disciplinary and friendly Group of motivated and cutting-edge researchers and to contribute to clinical research by providing technical knowledge, software engineering expertise and data insight.

As a Database Programmer you will Develop new database applications for big clinical data to meet project requirements and deadlines, provide software feedback and carry out software improvement, extension, integration and further development on existing code. You will contribute to the harmonisation, curation, and processing of large clinical datasets and develop code to validate, test, document and maintain database applications. You will also represent the project, team, and the University in collaboration meetings, conferences and at external meetings.

You will have a Degree in computer science, software engineering, health informatics or an equivalent combination of training and professional experience. Proven understanding and experience in one or more RDBMSs and SQL dialects (e.g. PostgreSQL), excellent skills in at least one high level programming language (e.g. Python, C#, C++) and excellent analytical and problem-solving skills with great attention to detail are essential. Experience in common data models (CDMs) and in the extract, transform, and load (ETL) process, knowledge of R and/or RStudio and working experience in a research environment are desirable.

This is a full-time fixed-term appointment for 2 years.

The closing date for this position is 12 noon on Monday 27 February 2023. You will be required to upload a CV and supporting statement as part of your online application.

Contact Person: HR Team, NDRMS
Contact Phone: 
Vacancy ID: 163066
Closing Date & Time: 27-Feb-2023 12:00
ICPE 2023 Abstract Deadline: Feb. 13

ICPE 2023 Call for Abstracts
Submission Deadline: February 13, 2023

Abstract submissions for the 39th International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE 2023) are now being accepted online.

Call for Abstracts
ICPE 2023 will be a live event held at the Halifax Convention Centre, Halifax, Nova Scotia, Canada, August 23-27, 2023. Virtual presentations are not permitted for the event; all presentations must be delivered in person. If you submit an abstract, it is with the intention that you will physically attend the conference to present it.

The ICPE 2023 is a unique forum for the exchange of scientific information from the fields of pharmacoepidemiology and therapeutic risk management among those in the pharmaceutical industry, government, academia, service
Constructing a vaccine vocabulary hierarchy using formal concept analysis

**INTRO:**
Vaccine concepts in the OMOP CDM Vocabulary lack comprehensiveness and consistencies. A manually curated hierarchy is difficult to maintain, and care providers and analysts cannot consistently access the larger number of concepts in existing vaccine vocabularies.

**METHOD:**
1. Decompose each vaccine code into its attributes (e.g., indication, mechanism of action). (This step requires manual work from vaccine experts.)
2. Convert decomposition to a "formal context." (This step requires manual work from vaccine experts.)
3. Build Formal Concept Analysis Application to create hierarchical relationships and "Webs of Yale"

**RESULTS:**
- Complete formal concept analysis application
- Fully created vaccine vocabulary hierarchy

**FCA is a rigorous method for building ontologies from a set of items and their attributes. In this work, we demonstrate the utility of the algorithm applied to OMOP vaccine vocabularies.**

**Comparison with manual hierarchy**
- The hierarchy produced by FCA is guaranteed to be a lattice: every pair of nodes in the resulting graph will have a unique least upper bound and a unique greatest lower bound.

**MONDAY**

Constructing vaccine vocabulary hierarchy using formal concept analysis

(Adam Black, Yupeng Li, Denys Kaduk, Licong Cui, Rashmie Abeysinghe, Lixia Yao)
#OHDSISocialShowcase This Week

**OHDSI Phenotype Phebruary: lessons learned**

Azza Shoaibi, Joel Swerdel, Allan Wu, Gowtham Rao, Adam Black, Evan Minty, Asieh Golozar, Rupa Makadia, Jill Hardin, Erica Voss, Tiffany J. Callahan, Juan Banda, Anna Ostropolets, Claudia Pulgarin, Marcela Rivera, David Vizcaya, Patrick Ryan

BACKGROUND:
- Phenotypes are the foundational elements in almost every real-world analysis. Yet, the science of phenotype development and evaluation is relatively immature.
- In February of 2022, OHDSI initiated “Phenotype Phebruary: 28 days, 28 phenotypes.”
- Each day, an OHDSI collaborator followed a process to phenotype one of the 28 clinical ideas.

METHODS:
- The community forum served as a platform for all members to collaborate by exploring the posted definitions, reviewing the results provided, replying with reflections or by executing cohort definitions and CohortDiagnostics in additional databases, and sharing consequent learnings.

**RESULTS**:
- Table 1 summarizes the 28 phenotypes that were discussed by the community. Table 2 summarizes lessons learned with the following structure: clinical description, phenotype development, and phenotype evaluation. The themes identified belonged to 3 different types of lessons: tips, strategies, challenges, and opportunities.

**CONCLUSIONS**:
- Phenotyping is complex, multidimensional and requires exchange of knowledge, learnings, and insights across collaboration from different background and expertise.
- Large-scale characterization (e.g., CDC), Diagnostic predictive models (e.g., PhenoPower), and structured review of patient profiles are potentially effective and novel strategies for phenotype evaluation.
- We are getting closer to a standardized process. But further collaboration is needed to formalize a scalable and reproducible process and establish empirically-driven objective diagnostics.

**TUESDAY**

OHDSI Phenotype Phebruary: lessons learned (Azza Shoaibi, Joel Swerdel, Allan Wu, Gowtham Rao, Adam Black, Evan Minty, Asieh Golozar, Rupa Makadia, Jill Hardin, Erica Voss, Tiffany J. Callahan, Juan Banda, Anna Ostropolets, Claudia Pulgarin, Marcela Rivera, David Vizcaya, Patrick Ryan)
Adjusting for Healthcare Utilization Improves the Performance of Self-Controlled Case Series Studies using Electronic Health Records

Undina Gisladottir, Nicholas Tatonetti

Department of Biostatistics, Columbia University

Introduction

- Adjusting for healthcare utilization aims to account for the potential confounding effect of healthcare utilization on the outcome of interest in self-controlled case series studies.
- Electronic health records (EHRs) provide a rich source of data for adjusting healthcare utilization.
- Prior studies have shown that adjusting for healthcare utilization can improve the performance of self-controlled case series studies.

Methods

Data and Cohort Selection

- Patients with at least one hospitalization in the study period were included.
- The cohort was restricted to patients with at least one prescription refill.
- The study period was defined as the period from 2005 to 2014.

Cohort Description

- Patients were stratified by their healthcare utilization.
- The healthcare utilization was categorized into low, medium, and high.

Results

- Adjusting for healthcare utilization improved the performance of self-controlled case series studies.
- The adjusted odds ratios were closer to the null value compared to the unadjusted analysis.

Conclusion

- Adjusting for healthcare utilization improves the performance of self-controlled case series studies.
- The findings suggest that healthcare utilization is an important confounding factor in these studies.
- Future research should consider adjusting for healthcare utilization in self-controlled case series studies using EHRs.
Delirium prediction in patients with trauma and comparison of predictors across trauma center and non-trauma center (Su Jin Gan, Dong Yun Lee, Jimyung Park, Rae Woong Park)
The OHDSI Community Dashboard: Tracking the Health and Impact of the Open Science Observational Health Data Sciences and Informatics Community (Star Liu, Asieh Golozar, Jody-Ann McLeggon, Adam Black, Paul Nagy)

Welcome to the OHDSI Community Dashboard

<table>
<thead>
<tr>
<th>PubMed OHDSI Manuscripts</th>
<th>YouTube Videos</th>
<th>Ehden Courses</th>
<th>Working Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>515 (2,092 authors)</td>
<td>820 (211K+ hours watched)</td>
<td>19 (3,276+ course completions)</td>
<td>28 (3000+ members)</td>
</tr>
</tbody>
</table>

Data as of: 01-30-2023

Observational Health Data Sciences and Informatics (OHDSI) is an open science community. OHDSI’s mission is to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care. The OHDSI Community Dashboard is a tool to highlight the progress we are making toward this mission and the collective accomplishments and impact of our community. A goal of the dashboard is help our community identify how members can see the OHDSI eco-system as an interconnected system to make a larger impact. We hope you find these tools useful staying up to date with all the activities in OHDSI as well as finding new colleagues in our community to collaborate with. Dashboards are developed to represent various aspects of the OHDSI community activities.
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?
Jan. 31: Introduction to Phenotype Phebruary

Patrick Ryan
Vice President, Observational Health Data Analytics, Janssen Research and Development, Inc.; Adjunct Assistant Professor, Columbia University

Gowtham Rao
Senior Director, Observational Health Data Analytics, Janssen Research and Development, Inc.; Phenotype Development & Evaluation Workgroup Lead

Azza Shoaibi
Associate Director, Observational Health Data Analytics, Janssen Research and Development, Inc.; OHDSI2022 presenter on “OHDSI Phenotype Phebruary: lessons learned”