Recent OHDSI Publications

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

Upcoming Community Calls

Date	Topic
Dec. 5	Recent Publications
Dec. 12	Happy Birthday OHDSI! Where Have We Come In 10 Years, and in 12 Months?
Dec. 19	Holiday-Themed Goodbye to 2023!
Dec. 26	No Call
Jan. 2	No Call
Jan. 9	Welcome Back! What Can OHDSI Accomplish in 2024?

Happy 10th Birthday to OHDSI

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

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*.

JAMIA Open, 2023, **6(4)**, ooad096
https://doi.org/10.1093/jamiaopen/ooad096
Research and Applications

Research and Applications

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

Rupa Makadia, PhD[1,2,*], Azza Shoaibi, PhD[1,2], Gowtham A Rao, MD[1,2], Anna Ostropolets, MD[1,3], Peter R. Rijnbeek, PhD[1,4], Erica A. Voss, MPH[1,2], Talita Duarte-Salles, PhD[1,5], Juan Manuel Ramírez-Anguita, PhD[6], Miguel A. Mayer, MD[7], Filip Maljković, MS[8], Spiros Denaxas, PhD[9,10], Fredrik Nyberg, PhD[11], Vaclav Papez, PhD[9], Anthony G. Sena, BA[1,2,4], Thamir M. Alshammari, PhD[1,12], Lana Y.H. Lai, PhD[1,13], Kevin Haynes, PharmD[2], Marc A. Suchard, MD[1,14], George Hripcsak, MD[1,3], Patrick B. Ryan, PhD[1,2,3]

[1]OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY 10027, United States, [2]Global Epidemiology, Janssen Pharmaceutical Research and Development, LLC, Titusville, NJ 08560, United States, [3]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY 10027, United States, [4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, 3000 CA, The Netherlands, [5]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, 08007, Spain, [6]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Barcelona, 08003, Spain, [7]Management Control Department, Parc de Salut Mar (PSMAR), Barcelona, 08007, Spain, [8]Research and Development, Heliant d.o.o, Belgrade, 11000, Serbia, [9]Institute of Health Informatics, University College London, London, NW1 2DA, United Kingdom, [10]British Heart Foundation Data Science Centre, HDR, London, NW1 2DA, United Kingdom, [11]School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, 40530, Sweden, [12]College of Pharmacy, Prince Sattam Bin Abdulaziz University, Riyadh, 11942, Saudi Arabia, [13]Division of Informatics, Imaging and Data Sciences, University of Manchester, Manchester, M13 9PL, United Kingdom, [14]Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90025, United States

*Corresponding author: Rupa Makadia, PhD, Janssen Pharmaceutical Research and Development, LLC, 1125 Trenton Harbourton Rd, Titusville, NJ 08560 (rmakadia@its.jnj.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 modifications (inpatient setting, standardization of code set, and code set changes) to the computable phenotype on background incidence rates.

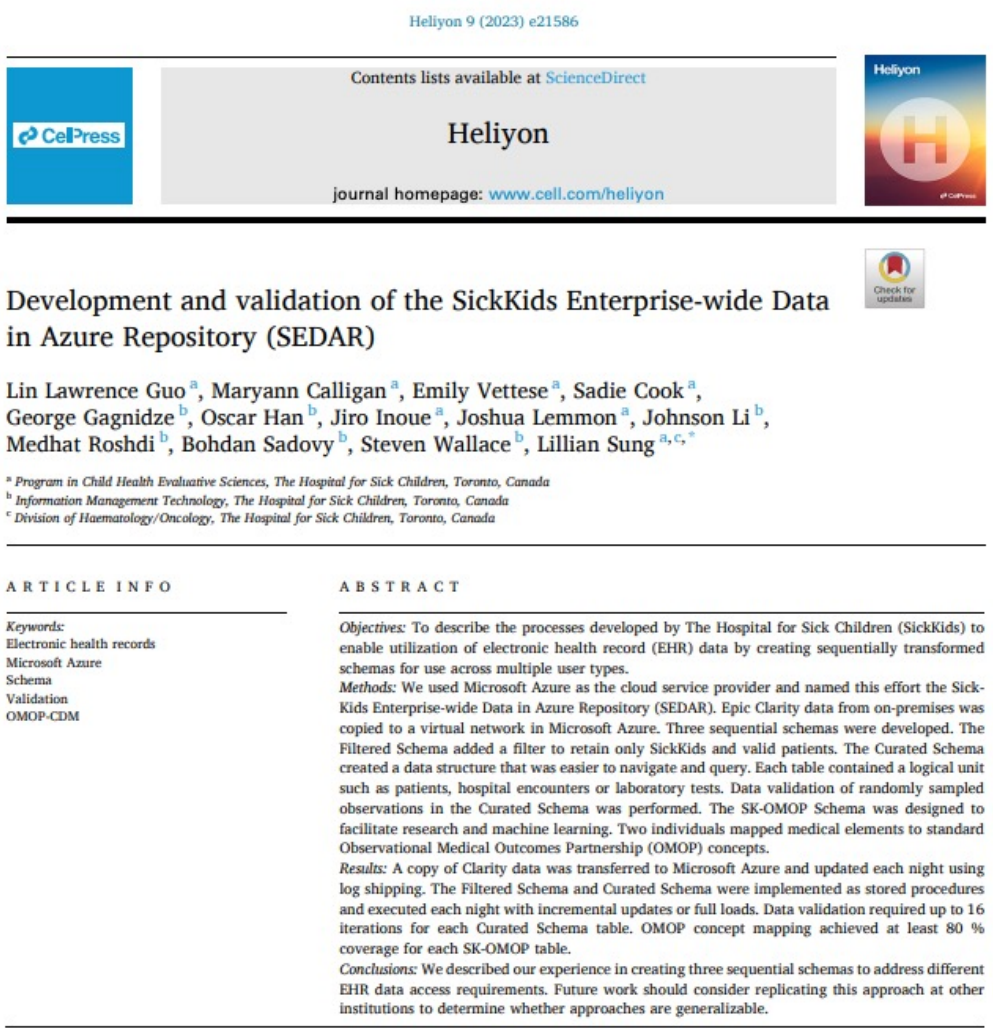
Results: Rate ratios (RRs) of the incidence rates from each computable phenotype definition varied across outcomes, with inpatient restriction showing the highest variation from 1 to 11.93. Standardization of code set RRs ranges from 1 to 1.64, and code set changes range from 1 to 2.52.

Discussion: The modification that has the highest impact is requiring inpatient place of service, leading to at least a 2-fold higher incidence rate in the base definition. Standardization showed almost no change when using source code variations. The strength of the effect in the inpatient restriction is highly dependent on the outcome. Changing definitions from broad to narrow showed the most variability by age/gender/database across phenotypes and less than a 2-fold increase in rate compared to the base definition.

Conclusion: Characterization of outcomes across a network of databases yields insights into sensitivity and specificity trade-offs when definitions are altered. Outcomes should be thoroughly evaluated prior to use for background rates for their plausibility for use across a global network.

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*.



Heliyon 9 (2023) e21586

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Development and validation of the SickKids Enterprise-wide Data in Azure Repository (SEDAR)

Lin Lawrence Guo [a], Maryann Calligan [a], Emily Vettese [a], Sadie Cook [a], George Gagnidze [b], Oscar Han [b], Jiro Inoue [a], Joshua Lemmon [a], Johnson Li [b], Medhat Roshdi [b], Bohdan Sadovy [b], Steven Wallace [b], Lillian Sung [a,c,*]

[a] Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Canada
[b] Information Management Technology, The Hospital for Sick Children, Toronto, Canada
[c] Division of Haematology/Oncology, The Hospital for Sick Children, Toronto, Canada

ARTICLE INFO

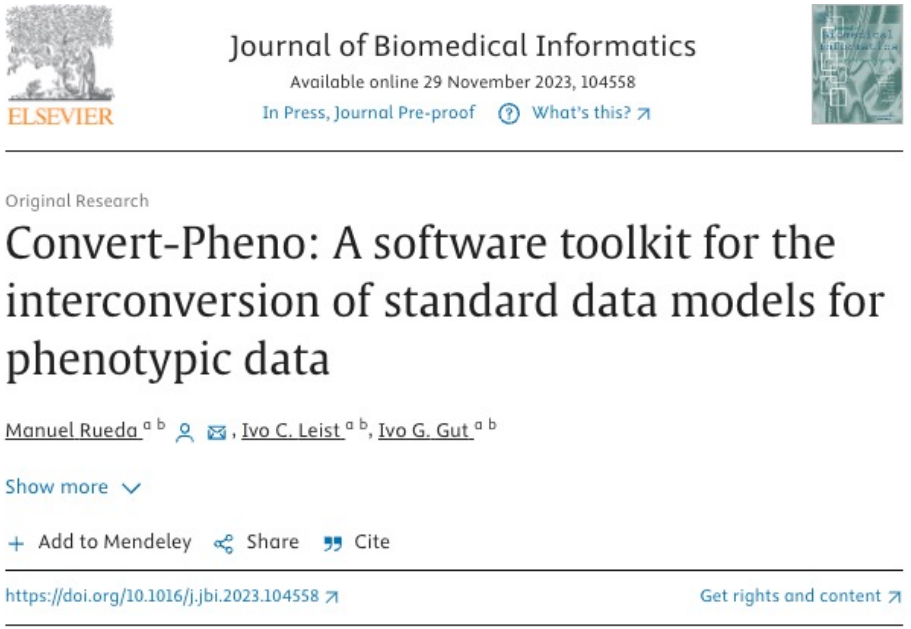
Keywords:
Electronic health records
Microsoft Azure
Schema
Validation
OMOP-CDM

ABSTRACT

Objectives: To describe the processes developed by The Hospital for Sick Children (SickKids) to enable utilization of electronic health record (EHR) data by creating sequentially transformed schemas for use across multiple user types.
Methods: We used Microsoft Azure as the cloud service provider and named this effort the SickKids Enterprise-wide Data in Azure Repository (SEDAR). Epic Clarity data from on-premises was copied to a virtual network in Microsoft Azure. Three sequential schemas were developed. The Filtered Schema added a filter to retain only SickKids and valid patients. The Curated Schema created a data structure that was easier to navigate and query. Each table contained a logical unit such as patients, hospital encounters or laboratory tests. Data validation of randomly sampled observations in the Curated Schema was performed. The SK-OMOP Schema was designed to facilitate research and machine learning. Two individuals mapped medical elements to standard Observational Medical Outcomes Partnership (OMOP) concepts.
Results: A copy of Clarity data was transferred to Microsoft Azure and updated each night using log shipping. The Filtered Schema and Curated Schema were implemented as stored procedures and executed each night with incremental updates or full loads. Data validation required up to 16 iterations for each Curated Schema table. OMOP concept mapping achieved at least 80 % coverage for each SK-OMOP table.
Conclusions: We described our experience in creating three sequential schemas to address different EHR data access requirements. Future work should consider replicating this approach at other institutions to determine whether approaches are generalizable.

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*.



ELSEVIER

Journal of Biomedical Informatics

Available online 29 November 2023, 104558

In Press, Journal Pre-proof What's this?

Original Research

Convert-Pheno: A software toolkit for the interconversion of standard data models for phenotypic data

Manuel Rueda [a b], Ivo C. Leist [a b], Ivo G. Gut [a b]

Show more

Add to Mendeley Share Cite

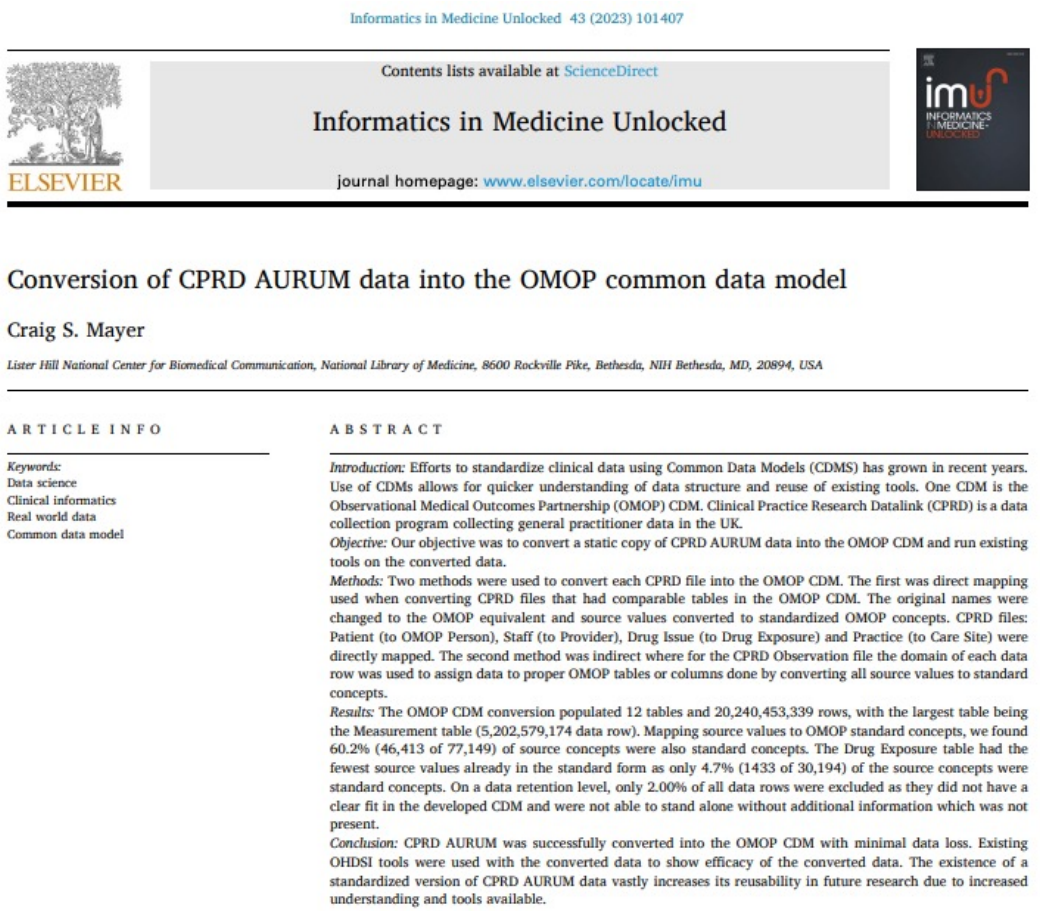
https://doi.org/10.1016/j.jbi.2023.104558 Get rights and content

Abstract

Efficient sharing and integration of phenotypic data is crucial for advancing biomedical research and enhancing patient outcomes in precision medicine and public health. To achieve this, the health data community has developed standards to promote the harmonization of variable names and values. However, the use of diverse standards across different research centers can hinder progress. Here we present Convert-Pheno, an open-source software toolkit that enables the interconversion of common data models for phenotypic data such as Beacon v2 Models, CDISC-ODM, OMOP-CDM, Phenopackets v2, and REDCap. Along with the software, we have created a detailed documentation that includes information on deployment and installation.

OHDSI Shoutouts!

Congratulations to **Craig Mayer** on the publication of **of Conversion of CPRD AURUM Data into the OMOP Common Data Model** in *Informatics in Medicine Unlocked*.



Informatics in Medicine Unlocked 43 (2023) 101407

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Conversion of CPRD AURUM data into the OMOP common data model

Craig S. Mayer

Lister Hill National Center for Biomedical Communication, National Library of Medicine, 8600 Rockville Pike, Bethesda, NIH Bethesda, MD, 20894, USA

ARTICLE INFO

Keywords:
Data science
Clinical informatics
Real world data
Common data model

ABSTRACT

Introduction: Efforts to standardize clinical data using Common Data Models (CDMS) has grown in recent years. Use of CDMs allows for quicker understanding of data structure and reuse of existing tools. One CDM is the Observational Medical Outcomes Partnership (OMOP) CDM. Clinical Practice Research Datalink (CPRD) is a data collection program collecting general practitioner data in the UK.

Objective: Our objective was to convert a static copy of CPRD AURUM data into the OMOP CDM and run existing tools on the converted data.

Methods: Two methods were used to convert each CPRD file into the OMOP CDM. The first was direct mapping used when converting CPRD files that had comparable tables in the OMOP CDM. The original names were changed to the OMOP equivalent and source values converted to standardized OMOP concepts. CPRD files: Patient (to OMOP Person), Staff (to Provider), Drug Issue (to Drug Exposure) and Practice (to Care Site) were directly mapped. The second method was indirect where for the CPRD Observation file the domain of each data row was used to assign data to proper OMOP tables or columns done by converting all source values to standard concepts.

Results: The OMOP CDM conversion populated 12 tables and 20,240,453,339 rows, with the largest table being the Measurement table (5,202,579,174 data row). Mapping source values to OMOP standard concepts, we found 60.2% (46,413 of 77,149) of source concepts were also standard concepts. The Drug Exposure table had the fewest source values already in the standard form as only 4.7% (1433 of 30,194) of the source concepts were standard concepts. On a data retention level, only 2.00% of all data rows were excluded as they did not have a clear fit in the developed CDM and were not able to stand alone without additional information which was not present.

Conclusion: CPRD AURUM was successfully converted into the OMOP CDM with minimal data loss. Existing OHDSI tools were used with the converted data to show efficacy of the converted data. The existence of a standardized version of CPRD AURUM data vastly increases its reusability in future research due to increased understanding and tools available.

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

Date	Time (ET)	Meeting
Wednesday	8 am	Psychiatry
Wednesday	4 pm	Vulcan/OHDSI Meeting
Thursday	9:30 am	Themis
Thursday	12 pm	Methods Research
Thursday	1 pm	OMOP CDM Oncology – Vocabulary/Development Subgroup
Thursday	7 pm	Dentistry
Friday	9 am	GIS – Geographic Information System Development
Friday	9 am	Phenotype Development & Evaluation
Friday	1 pm	Clinical Trials
Friday	10 pm	China Chapter
Monday	9 am	Vaccine Vocabulary
Monnday	10 am	Africa Chapter
Monday	11 am	Early-Stage Researchers
Monday	4 pm	Eyecare & Vision Research
Tuesday	9 am	OMOP CDM Oncology – Genomic Subgroup

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.

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

December Newsletter is Available

OHDSI

The Journey Newsletter (December 2023)

OHDSI officially formed on Dec. 16, 2013, and over the next 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 OHDSI during our Dec. 12 community call, and we provide a brief look back in our newsletter video podcast below. We also highlight papers, presentations and other updates from the last month in this OHDSI newsletter! #JoinTheJourney

Video Podcast: 10 Years of OHDSI

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

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 month. 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 collaborator showcase page to find all the research from the global symposium.

• Members of the OHDSI community provided a full-day workshop entitled "OHDSI RWE Revolution: Igniting Data Modernization with Harmonized Standards for Cutting-Edge Health Research" during the 2023 AMIA Symposium. The full slidedeck 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 OHDSI in action. Please join our Dec. 12 community call to reflect on 10 years of OHDSI — how the community formed and grew, and what we achieved through collaboration and open science along the way.

• **Andrew Williams** posted a document that describes how OHDSI workgroups form, the kinds of things they do, and the categories of workgroups that the OHDSI community is organized into. It is meant to help community members understand how to use OHDSI 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.

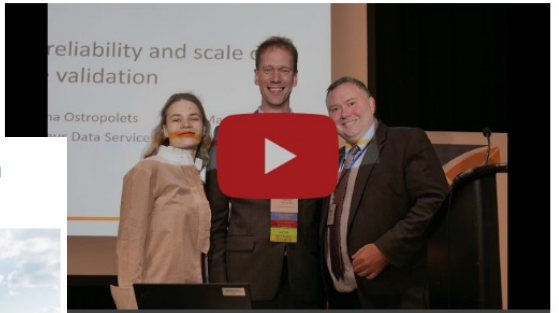
10th Anniversary of OHDSI Celebration Set For Dec. 12 Community Call

On Dec. 16, 2013, George Hripcsak 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 49 counties to the OMOP Common Data Model to cover more than 956 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*

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



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

idation is regarded as a necessary element of regulatory-grade
, but conducting case validation th
s time- and resource-intensive, has
y conducted in such a way that do
proper quantitative bias analysis.

November Publications

Henke E, Zoch M, Kallfelz M, Ruhnke T, Leutner LA, Spoden M, Günster C, Sedlmayr M, Bathelt F. Assessing the Use of German Claims Data Vocabularies for Research in the Observational Medical Outcomes Partnership Common Data Model: Development and Evaluation Study. JMIR Med Inform. 2023 Nov 7;11:e47959. doi: 10.2196/47959. PMID: 37942786; PMCID: PMC10653283.

Voss EA, Blacketer C, van Sandijk S, Moinat M, Kallfelz M, van Speybroeck M, Prieto-Alhambra D, Schuemie MJ, Rijnbeek PR. European Health Data & Evidence Network-learnings from building out a standardized international health data network. J Am Med Inform Assoc. 2023 Nov 10:ocad214. doi: 10.1093/jamia/ocad214. Epub ahead of print. PMID: 37952118.

Choi JY, Yoo S, Song W, Kim S, Baek H, Lee JS, Yoon YS, Yoon S, Lee HY, Kim KI. Development and Validation of a Prognostic Classification Model Predicting Postoperative Adverse Outcomes in Older Surgical Patients Using a Machine Learning Algorithm: Retrospective Observational Network Study. J Med Internet Res. 2023 Nov 13;25:e42259. doi: 10.2196/42259. PMID: 37955965; PMCID: PMC10682929.

Bu F, Schuemie MJ, Nishimura A, Smith LH, Kostka K, Falconer T, McLeggon JA, Ryan PB, Hripcsak G, Suchard MA. Bayesian safety surveillance with adaptive bias correction. Stat Med. 2023 Nov 27. doi: 10.1002/sim.9968. Epub ahead of print. PMID: 38010062.

Cai CX, Halfpenny W, Boland MV, Lehmann HP, Hribar M, Goetz KE, Baxter SL. Advancing Toward a Common Data Model in Ophthalmology: Gap Analysis of General Eye Examination Concepts to Standard Observational Medical Outcomes Partnership (OMOP) Concepts. Ophthalmol Sci. 2023 Aug 25;3(4):100391. doi: 10.1016/j.xops.2023.100391. PMID: 38025162; PMCID: PMC10630664.

mailchi.mp/ohdsi/december2023

December Newsletter is Available

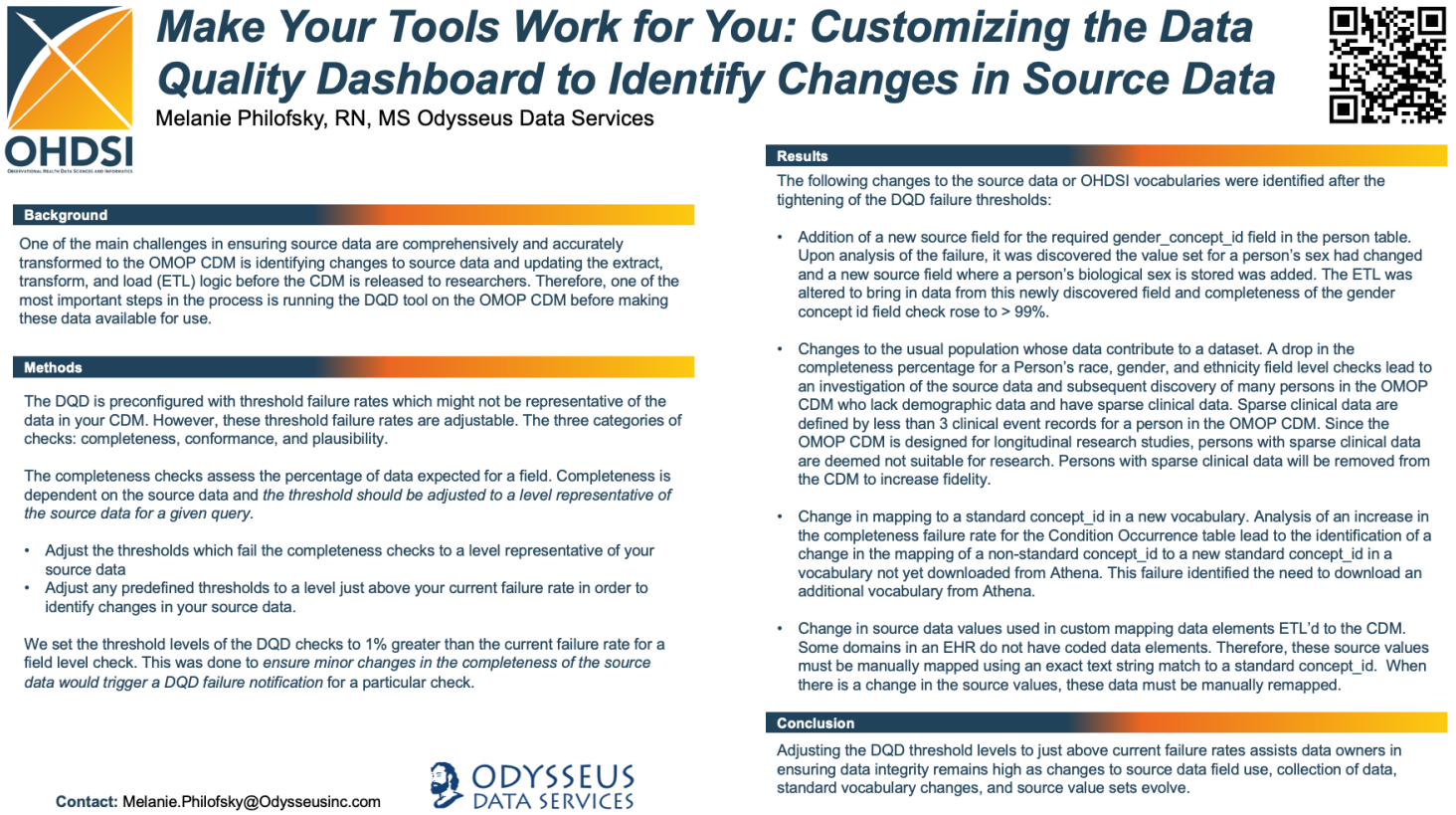
mailchi.mp/ohdsi/december2023

#OHDSISocialShowcase This Week

MONDAY

Make Your Tools Work for You: Customizing the Data Quality Dashboard to Identify Changes in Source Data

(Melanie Philofsky)



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 preconfigured 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.

Contact: Melanie.Philofsky@Odysseusinc.com

ODYSSEUS DATA SERVICES

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.
- Change 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'd 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.

#OHDSISocialShowcase This Week

TUESDAY

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)

Assessment of Pre-trained Observational Large Longitudinal models in OHDSI (APOLLO)

Martijn Schuemie[1,2], Yong Chen[3], Egill Fridgeirsson[4], Chungsoo Kim[5], Jenna Reps[1,4], Marc Suchard[2], Xiaoyu Wang[1,6], Chao Pang[7]

[1] Johnson & Johnson, [2] UCLA, [3] University of Pennsylvania, [4] Erasmus University Medical Center of Rotterdam, [5] Ajou University Graduate School of Medicine, [6] Duke University, [7] Columbia University

Read the abstract

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 pretrained 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** refines the pre-trained model for a specific task.

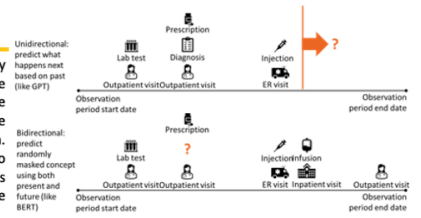


Figure 1. Types of pre-training

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.
- **phenotyping**, 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**: given 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.

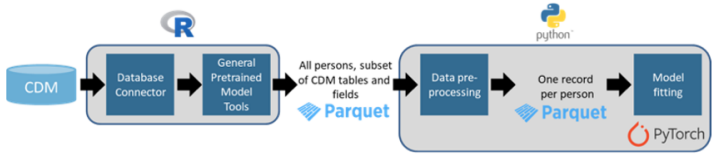


Figure 2. Overall architecture for pre-training

Contact: schuemie@ohdsi.org

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 person IDs, visit occurrence IDs, 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:

- Type of pre-training task: unidirectional or bidirectional? Predicting the next /masked 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, season, 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 regularization, 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 Merative Marketscan CCAE database. Download took 1.8 hours, and Parquet files take 3.1GB of disc space. Pre-processing took 10 minutes, resulting in Parquet files totaling 2.3GB. A single epoch of pre-training took 20 hours on an NVIDIA A10G for a 121-million-parameter model.

Conclusions

Despite existing uncertainties surrounding the applications of large pre-trained models to healthcare data at scale, the potential for transformative impacts is promising.

#OHDSISocialShowcase This Week

WEDNESDAY

A use case of OHDSI ATLAS in a high-throughput genome wide association study 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)

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

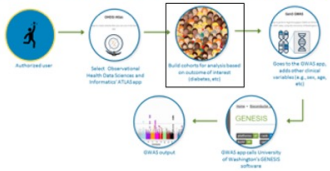
Craig C. Teerlink,[1,2] Hamid Saoudian,[1,2] Richard D. Boyce,[3] Philip S. Tsao,[4,5] Kyle M. Hernandez,[6] Victoria Zaksas,[6] Pieter Lukasse,[6] Andrew Prokhorenkov,[6] Noah Metoki-Shlubsky,[6] Robert L. Grossman,[6] Scott L DuVall[1,2]

[1]VA Informatics and Computing Infrastructure, [2]University of Utah School of Medicine, [3]University of Pittsburgh Department of Biomedical Informatics, [4]VA Palo Alto Health Care System, [5]Stanford University, [6]Center for Translational Data Science, University of Chicago

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 VADC contains data on ~650K subjects, including:
  - Frequency of the number of times a Veteran was diagnosed within PheCode 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



We use ATLAS to define cohorts for the GWAS

FIRST: Define an overall cohort to study (example: all MVP subjects)

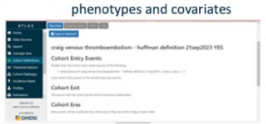
SECOND: Specify the logical criteria for dichotomous phenotypes

THIRD: Specify dichotomous covariates

1. Create an overall study cohort

2. Define concept sets to specify inclusion and exclusion criteria for dichotomous phenotype and covariate cohorts

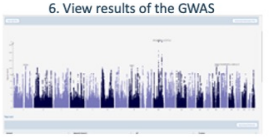
3. Use cohort definitions based on concept sets to define phenotypes and covariates



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



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

#OHDSISocialShowcase This Week

THURSDAY

Data-driven assessment of mental health among children and adolescents with food allergy

(Natalie Flaks-Manov, Inbal Goldshtein, Chen Yanover)

Data-driven assessment of mental health among children and adolescents with food allergy

PRESENTER: **Natalie Flaks-Manov**

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.

Natalie Flaks-Manov[1], Inbal Goldshtein[1], Chen Yanover[1]

[1] KI Research Institute

CONTACT: natalie@kinstitute.org.il

OHDSI

The Israeli Institute for Applied Research in Computational Health https://kinstitute.org.il/

Food allergies in children and adolescents are associated with an increased risk of mental health disorders, specifically eating disorders

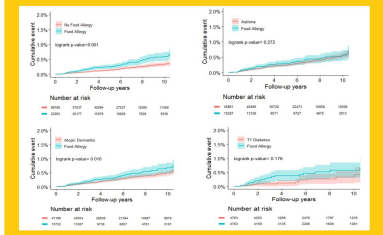


Figure 1: Kaplan-Meier cumulative probability of eating disorders among FA cohort vs. other chronic conditions

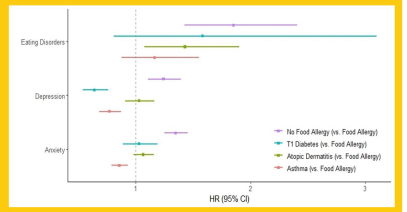


Figure 2: Hazard ratios from Cox models of mental outcomes in a FA cohort vs. no FA and vs. other chronic conditions (adjusted by age and sex)

Take a picture to **download** the **full paper**

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.

POTENTIAL IMPACT:

- ✓ There are no guidelines for assessing eating disorders in children with food allergies.
- ✓ This study should alert healthcare providers to the connection between food allergies and the development of eating disorders in later life.
- ✓ The awareness should drive the development of prevention strategies for binge eating disorders and, when necessary, prompt early referrals to multidisciplinary teams specialized in eating disorder treatment.

RESULTS

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

Cohorts[1]	Pre matching			Post matching		
	No Food Allergy	Food Allergy		No Food Allergy	Food Allergy	
Characteristic	N = 1,110,721	N = 23,263	p-value[2]	N = 69,789	N = 23,263	p-value[2]
Age, Median (IQR)	8.0 (4.0, 13.0)	5.0 (2.0, 10.0)	<0.001	5.0 (2.0, 10.0)	5.0 (2.0, 10.0)	>0.9
Female	538,701 (48%)	10,081 (43%)	<0.001	30,243 (43%)	10,081 (43%)	>0.9
Eating disorder outcome	2,663 (0.2%)	95 (0.4%)	<0.001	132 (0.2%)	95 (0.4%)	<0.001
Anxiety outcome	42,533 (3.8%)	1,075 (4.6%)	<0.001	1,998 (2.9%)	1,075 (4.6%)	<0.001
Depression outcome	25,249 (2.2%)	689 (2.1%)	0.2	946 (1.4%)	689 (2.1%)	<0.001
Cohorts[1]	**Asthma**	**Food Allergy**		**Asthma**	**Food Allergy**	
	N = 136,453	N = 16,992		N = 45,861	N = 15,287	
Age, Median (IQR)	8.0 (4.0, 12.0)	5.0 (2.0, 10.0)	<0.001	5.0 (3.0, 10.0)	5.0 (3.0, 10.0)	<0.001
Female	58,549 (43%)	7,683 (45%)	<0.001	20,547 (45%)	6,869 (45%)	>0.9
Eating disorder outcome	652 (0.5%)	65 (0.4%)	0.086	177 (0.4%)	64 (0.4%)	0.6
Anxiety outcome	10,767 (7.9%)	721 (4.2%)	<0.001	2,695 (5.9%)	713 (4.7%)	<0.001
Depression outcome	6,885 (5.0%)	333 (2.0%)	<0.001	1,372 (3.0%)	333 (2.2%)	<0.001
Cohorts[1]	**Atopic Dermatitis**	**Food Allergy**		**Atopic Dermatitis**	**Food Allergy**	
	N = 207,575	N = 15,746		N = 47,196	N = 15,732	
Age, Median (IQR)	5.0 (2.0, 10.0)	5.0 (2.0, 10.0)	0.001	5.0 (2.0, 10.0)	5.0 (2.0, 10.0)	0.93
Female	105,987 (51%)	6,799 (43%)	<0.001	20,361 (43%)	6,787 (43%)	>0.99
Eating disorder outcome	820 (0.4%)	69 (0.4%)	0.41	150 (0.3%)	69 (0.4%)	0.026
Anxiety outcome	11,705 (5.6%)	733 (4.7%)	<0.001	2,178 (4.6%)	733 (4.7%)	0.82
Depression outcome	6,595 (3.2%)	359 (2.3%)	<0.001	1,105 (2.3%)	359 (2.3%)	0.67
Cohorts[1]	**Type 1 Diabetes**	**Food Allergy**		**Type 1 Diabetes**	**Food Allergy**	
	N = 4,815	N = 23,215		N = 4,763	N = 4,763	
Age, Median (IQR)	11.0 (7.0, 14.0)	5.0 (2.0, 10.0)	<0.001	11.0 (7.0, 14.0)	11.0 (7.0, 14.0)	0.2
Female	2,169 (45%)	10,063 (43%)	0.054	2,162 (45%)	2,162 (45%)	>0.9
Eating disorder outcome	14 (0.3%)	95 (0.4%)	0.2	14 (0.3%)	21 (0.4%)	0.2
Anxiety outcome	378 (7.8%)	1,074 (4.6%)	<0.001	370 (7.8%)	348 (7.3%)	0.4
Depression outcome	340 (7.0%)	488 (2.1%)	<0.001	328 (6.9%)	200 (4.2%)	<0.001

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.

#OHDSISocialShowcase This Week

FRIDAY

Quantification of Symptom Documentation on Disease Diagnosis Date in Structured Claims Data: An Application of the OHDSI Phenotype Library

(Gowtham Rao, Azza Shoaibi)

Quantification of Symptom Documentation on Disease Diagnosis Date in Structured Claims Data: An Application of the OHDSI Phenotype Library

PRESENTER : **Gowtham A Rao.**

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 dyspnea, on the same day as a definitive related disease diagnosis In structured health insurance us claims data .
- We demonstrate how demonstrate how the Observational Health Data Sciences and Informatics (OHDSI) PhenotypeLibrary (PL) can be used to carry out a study within the OHDSI network.

METHODS:

- We utilized phenotype definitions from the OHDSI PL version 3.15.0[2].
- We selected 14 acute severe clinical conditions that are expected to have a sudden onset and short latent or indolent 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 R package using the function PhenotypeLibrary::getPlCohortDefinitionSet
- The obtained CohortDefinitionSet object was used in CohortGenerator to instantiate the initial cohorts..
- The study was run on five US claims data sources: The IQVIA® Adjudicated Health Plan Claims Data (PharMetrics Plus), Optum's Clinformatics® Data Mart (OptumDOD), The MerativeTM MarketScan® Commercial Database (CCAE), The MerativeTM MarketScan® Multi-State Medicaid Database (MDCD), and The MerativeTM MarketScan® Medicare Supplemental Database (MDCR).

Table 1

27 Asthma without COPD

-1:Composite [48.1% - 57.8%]
3:Cough or Sputum [10.1% - 14.4%]
5:Dyspnea [5.7% - 16.7%]
6:Fever [5.7% - 5.7%]
9:Sore throat [5.1% - 7.7%]
12:Rhinitis or common cold or Sinusitis [8.4% - 20.0%]
20:Bronchitis or Bronchiolitis [8.8% - 12.9%]
278:Pain or ache that is Non Chronic [10.2% - 17.4%]
327:Pharyngitis including tonsillitis [5.5% - 5.5%]
328:Wheezing [7.1% - 7.1%]
354:Nasal Congestion or Rhinitis or Common Cold [5.1% - 14.3%]
368:Sinusitis [5.6% - 6.8%]
392:Otalgia or Otitis [8.9% - 8.9%]

74 Hemorrhagic stroke with inpatient admission

-1:Composite [78.1% - 91.4%]
5:Dyspnea [5.7% - 20.0%]
7:Headache or Headache disorder [14.4% - 32.6%]
10:Nausea or Vomiting [5.9% - 11.4%]
11:Malaise and or fatigue [11.3% - 23.0%]
62:Generalized Seizure [8.3% - 16.6%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [10.1% - 20.4%]
194:Encephalopathy or its presentations [20.1% - 32.9%]
244:Dizziness or giddiness including motion sickness and vertigo [6.1% - 8.3%]
278:Pain or ache that is Non Chronic [25.8% - 52.8%]
331:Encephalopathy [6.0% - 6.8%]
332:Pain or ache that is Chronic [5.3% - 15.1%]
339:Hypotension [5.2% - 6.7%]
341:Loss of mentation including comma, syncope, altered consciousness [18.1% - 33.7%]
358:Acute Respiratory Failure among persons with no chronic respiratory failure [15.1% - 22.1%]
359:Acute Respiratory Failure [16.3% - 24.1%]

251 Acute pancreatitis

-1:Composite [76.3% - 92.4%]
4:Diarrhea [5.2% - 7.6%]
5:Dyspnea [9.3% - 16.3%]
10:Nausea or Vomiting [22.1% - 34.0%]
11:Malaise and or fatigue [7.3% - 9.9%]
57:Bleeding [5.9% - 6.1%]
189:Right Upper Quadrant Pain [8.4% - 15.1%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [6.5% - 8.9%]
194:Encephalopathy or its presentations [5.1% - 7.2%]
278:Pain or ache that is Non Chronic [54.2% - 79.2%]
280:Abdominal Pain or acute abdomen [56.3% - 78.1%]
281:Epigastric Pain [17.8% - 25.7%]
299:Acute gastrointestinal bleeding or perforation events [5.8% - 6.1%]
359:Acute Respiratory Failure [5.6% - 6.4%]

68 Heart failure with inpatient admission

-1:Composite [83.1% - 92.6%]
3:Cough or Sputum [6.1% - 7.9%]
5:Dyspnea [29.8% - 46.4%]
10:Nausea or Vomiting [6.0% - 7.2%]
11:Malaise and or fatigue [7.3% - 16.7%]
57:Bleeding [6.3% - 9.0%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [6.6% - 14.7%]
194:Encephalopathy or its presentations [5.4% - 11.9%]
244:Dizziness or giddiness including motion sickness and vertigo [5.1% - 5.1%]
278:Pain or ache that is Non Chronic [28.8% - 52.0%]
280:Abdominal Pain or acute abdomen [8.4% - 17.1%]
282:Joint Pain [12.6% - 24.0%]
332:Pain or ache that is Chronic [5.5% - 15.8%]
339:Hypotension [5.8% - 8.5%]
341:Loss of mentation including comma, syncope, altered consciousness [6.1% - 8.3%]
358:Acute Respiratory Failure among persons with no chronic respiratory failure [12.0% - 18.6%]
359:Acute Respiratory Failure [14.7% - 22.7%]
364:Sleep disorder [7.7% - 11.5%]

234 Appendicitis

-1:Composite [70.7% - 89.0%]
4:Diarrhea [5.3% - 6.8%]
5:Dyspnea [12.4% - 12.4%]
6:Fever [5.5% - 12.8%]
10:Nausea or Vomiting [12.2% - 33.8%]
11:Malaise and or fatigue [5.4% - 5.4%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [5.5% - 5.5%]
278:Pain or ache that is Non Chronic [61.1% - 81.3%]
280:Abdominal Pain or acute abdomen [65.0% - 84.5%]

258 Anaphylaxis or Anaphylactic shock events

-1:Composite [54.6% - 65.5%]
5:Dyspnea [10.8% - 23.1%]
10:Nausea or Vomiting [5.0% - 6.8%]
11:Malaise and or fatigue [3.2% - 3.2%]
12:Rhinitis or common cold or Sinusitis [8.8% - 26.4%]
241:Urticaria [8.0% - 14.9%]
278:Pain or ache that is Non Chronic [11.3% - 21.5%]
340:Hives, Erythema, Eruption, Urticaria [13.3% - 25.2%]
354:Nasal Congestion or Rhinitis or Common Cold [6.9% - 23.2%]
381:Skin Itching [5.7% - 6.5%]

70 Stroke with inpatient admission

-1:Composite [71.2% - 87.5%]
5:Dyspnea [6.2% - 20.3%]
7:Headache or Headache disorder [9.9% - 26.5%]
10:Nausea or Vomiting [7.1% - 7.7%]
11:Malaise and or fatigue [19.4% - 32.0%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [18.1% - 29.9%]
194:Encephalopathy or its presentations [19.3% - 32.7%]
232:Paresthesia [5.2% - 16.3%]
244:Dizziness or giddiness including motion sickness and vertigo [9.0% - 15.7%]
278:Pain or ache that is Non Chronic [28.2% - 50.5%]
331:Encephalopathy [6.5% - 8.0%]
339:Hypotension [5.9% - 6.9%]
341:Loss of mentation including comma, syncope, altered consciousness [5.4% - 20.4%]
359:Acute Respiratory Failure [8.0% - 12.5%]
364:Sleep disorder [5.3% - 8.1%]

249 Ischemic Stroke

-1:Composite [74.1% - 90.0%]
5:Dyspnea [6.4% - 22.6%]
7:Headache or Headache disorder [9.4% - 24.5%]
10:Nausea or Vomiting [7.0% - 7.3%]
11:Malaise and or fatigue [20.0% - 34.7%]
57:Bleeding [11.4% - 14.4%]
62:Generalized Seizure [5.8% - 13.6%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [18.6% - 32.4%]
194:Encephalopathy or its presentations [21.9% - 35.9%]
232:Paresthesia [5.7% - 17.3%]
244:Dizziness or giddiness including motion sickness and vertigo [8.8% - 16.1%]
278:Pain or ache that is Non Chronic [26.8% - 49.2%]
331:Encephalopathy [5.3% - 9.8%]
341:Loss of mentation including comma, syncope, altered consciousness [5.0% - 19.3%]
359:Acute Respiratory Failure [8.4% - 15.1%]

284 Myocarditis or Pericarditis

-1:Composite [59.6% - 68.3%]
3:Cough or Sputum [5.2% - 6.3%]
5:Dyspnea [21.9% - 35.9%]
6:Fever [5.1% - 6.5%]
11:Malaise and or fatigue [5.1% - 9.2%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [6.4% - 7.5%]
278:Pain or ache that is Non Chronic [39.7% - 51.9%]

71 Acute myocardial infarction with inpatient admission

-1:Composite [84.8% - 94.8%]
5:Dyspnea [18.8% - 40.7%]
10:Nausea or Vomiting [5.8% - 9.5%]
11:Malaise and or fatigue [5.1% - 14.8%]
57:Bleeding [5.8% - 9.3%]
62:Generalized Seizure [5.3% - 5.3%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [10.5% - 13.2%]
219:Sudden Cardiac arrest or cardiac death [5.3% - 6.2%]
244:Dizziness or giddiness including motion sickness and vertigo [5.0% - 5.0%]
273:Pathological Ventricular Tachycardia [5.1% - 5.6%]
274:Cardiac arrest or Pathological Ventricular tachycardia [7.3% - 9.9%]
278:Pain or ache that is Non Chronic [49.7% - 79.1%]
335:Anxiety or Fear [5.7% - 10.7%]
339:Hypotension [5.2% - 10.8%]
341:Loss of mentation including comma, syncope, altered consciousness [5.7% - 10.2%]
359:Acute Respiratory Failure [10.8% - 24.5%]
402:Ventilatory assist for respiratory findings with Acute Respiratory Failure [5.5% - 8.9%]
404:Ventricular Tachycardia, in an inpatient or Emergency room setting [5.4% - 6.0%]
405:Atrial Fibrillation or Flutter [7.8% - 24.6%]

248 Disseminated intravascular coagulation DIC

-1:Composite [89.6% - 94.6%]
3:Cough or Sputum [5.2% - 6.1%]
5:Dyspnea [16.1% - 41.1%]
6:Fever [7.5% - 15.2%]
11:Malaise and or fatigue [9.3% - 20.9%]
57:Bleeding [25.3% - 32.6%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [7.6% - 18.0%]
194:Encephalopathy or its presentations [18.5% - 31.6%]
278:Pain or ache that is Non Chronic [22.4% - 46.8%]
280:Abdominal Pain or acute abdomen [17.0% - 32.9%]
339:Hypotension [13.6% - 20.4%]
341:Loss of mentation including comma, syncope, altered consciousness [6.0% - 11.9%]
359:Acute Respiratory Failure [38.8% - 53.0%]

292 Hepatic Failure

-1:Composite [79.0% - 90.0%]
5:Dyspnea [14.0% - 29.4%]
6:Fever [5.1% - 5.1%]
10:Nausea or Vomiting [8.2% - 12.2%]
11:Malaise and or fatigue [9.3% - 20.3%]
57:Bleeding [11.3% - 16.0%]
61:Bradycardia or heart block with inpatient admission [5.2% - 8.9%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [8.1% - 18.3%]
193:Jaundice or Itching [5.1% - 8.0%]
219:Sudden Cardiac arrest or cardiac death [5.6% - 8.5%]
240:Muscle weakness or myonlegia [6.0% - 8.6%]
274:Cardiac arrest or Pathological Ventricular tachycardia [5.9% - 10.0%]
278:Pain or ache that is Non Chronic [20.3% - 35.8%]
280:Abdominal Pain or acute abdomen [20.2% - 34.2%]
289:Presence Of Cardiac Arrhythmia [13.8% - 37.0%]
332:Pain or ache that is Chronic [7.0% - 19.5%]
335:Anxiety or Fear [5.2% - 10.0%]
339:Hypotension [6.7% - 14.3%]
341:Loss of mentation including comma, syncope, altered consciousness [9.3% - 34.9%]
364:Sleep disorder [8.9% - 8.9%]
380:Jaundice [5.0% - 7.7%]

362 Acute Kidney Injury AKI

-1:Composite [80.6% - 89.5%]
3:Cough or Sputum [5.5% - 6.5%]
4:Diarrhea [5.1% - 6.3%]
5:Dyspnea [13.4% - 32.2%]
6:Fever [5.2% - 8.6%]
7:Headache or Headache disorder [5.1% - 5.4%]
10:Nausea or Vomiting [8.6% - 14.1%]
11:Malaise and or fatigue [9.1% - 20.4%]
57:Bleeding [7.7% - 10.2%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [8.4% - 19.0%]
194:Encephalopathy or its presentations [9.5% - 20.4%]
244:Dizziness or giddiness including motion sickness and vertigo [5.9% - 5.9%]
278:Pain or ache that is Non Chronic [26.5% - 45.3%]
280:Abdominal Pain or acute abdomen [16.2% - 27.9%]
282:Joint Pain [11.5% - 23.0%]
299:Acute gastrointestinal bleeding or perforation events [5.6% - 7.6%]
331:Encephalopathy [5.1% - 8.5%]
332:Pain or ache that is Chronic [6.9% - 16.5%]
339:Hypotension [9.0% - 13.5%]
341:Loss of mentation including comma, syncope, altered consciousness [5.7% - 11.3%]
358:Acute Respiratory Failure among persons with no chronic respiratory failure [12.5% - 16.1%]
359:Acute Respiratory Failure [14.3% - 19.9%]
364:Sleep disorder [5.5% - 9.5%]

329 Pneumonitis and lung infections

-1:Composite [50.8% - 65.3%]
3:Cough or Sputum [15.4% - 25.0%]
5:Dyspnea [13.4% - 32.3%]
6:Fever [5.8% - 17.1%]
9:Sore throat [5.1% - 6.9%]
11:Malaise and or fatigue [6.4% - 10.7%]
12:Rhinitis or common cold or Sinusitis [6.1% - 7.4%]
191:Fatigue, Asthenia, Malaise, Lethargy, Anorexia [5.6% - 9.7%]
278:Pain or ache that is Non Chronic [13.1% - 27.9%]

**All IDs in Table 1 are from OHDSI PL version 3.15.0*

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, cont'd.

- For calculating co-occurrence of symptoms on the same day disease, we utilized the unpublished CohortCovariates branch of FeatureExtraction.
- We created a composite cohort representing the occurrence of any expected symptoms for each disease. We achieved this by utilizing the CohortAlgebra R package
- We calculated the prevalence of each symptom in each condition across the data sources

RESULTS

- Table 1 reports on the prevalence (range) of the of related symptoms in 14 disease cohorts with their corresponding IDs.
- All IDs match those of the cohorts in the OHDSI PL.
- The proportion of persons with symptoms for each disease appeared consistent across data sources.
- Asthma had the lowest rate of symptom capture followed by conditions like pneumonitis, anaphylaxis, and myocarditis.
- Heart failure, myocardial infarction, stroke, appendicitis, and disseminated intravascular coagulation had higher rates of symptoms.
- Only one disease in one data source had a composite symptom capture of less than 50% (asthma)

DISCUSSION

- Our results indicate that a large proportion of individuals with acute illnesses have at least one symptom code recorded simultaneously in claims data.
- Except for one acute disease (27: Asthma without COPD) in one data source (MDCR), symptom occurrence was observed in at least 50% of individuals.
- researchers may consider using symptoms as input covariates in future research.
- Our work sets the groundwork for further work into 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 in an OHDSI network study and its synergy with standard OHDSI software.

Gowtham Rao, Azza Shoaibi

Johnson & Johnson Epidemiology
OFFICE OF THE CHIEF MEDICAL OFFICER

OHDSI

OHDSI HADES releases: DatabaseConnector 6.3.1

DatabaseConnector 6.3.1 | Reference | Articles | Changelog | HADES

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

Contents

Strategus Development Update

Strategus sub-team formation

Developers hades

anthonysena 3h

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)

Reply

Opening: Limerick Digital Cancer Research Centre

University of Limerick local time: 21-November-2023 14:39

Navigation Section

- Search
  Click here to search for vacancies. Please email erecruitment@ul.ie if you experience difficulties.
- Login
  Click to Login. UL Staff, you must be logged in to see Internal Vacancies
- Register
  Click to Register
- Vacancy Detail
  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 AND COMMITTED TO SELECTION ON MERIT"

corehr

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 2,000 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's 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 - €48,427 p.a. pro rata

PD2 €49,790 – €54,153 p.a. pro rata

Informal enquires regarding the post may be directed to:
Professor Aedin Culhane
School of Medicine
University of Limerick
Email: aedin.culhane@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

Apply

Deputy Director, Quantitative Sciences

Apply

Job Opening: Stanford University

Stanford | Office of Postdoctoral Affairs
All Postdocs. All the Time.

Home About Prospective Postdocs Current Postdocs Faculty Mentors Postdoc Admins

Prospective Postdocs

- How To Apply
- Open Postdoctoral Positions
- Finding a Faculty Mentor
- Cost of Living
- Housing
- Fellowships at Stanford
- Fellowships outside Stanford

Open Postdoctoral position, faculty mentor Brian Bateman

Our research team is looking for a postdoctoral scholar in perinatal pharmacoepidemiology. 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.

Our research group prioritizes a collaborative and inclusive team environment. The principal investigators are experienced mentors who are highly committed to supporting the postdoctoral scholar in advancing their career as a future independent investigator. The

Important Info

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

Other Mentor(s) if Applicable:
Stephanie Leonard

Stanford Departments and Centers:
Anesthes, 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