What Can OHDSI Achieve Together in 2024?

OHDSI Community Call

Jan. 9, 2024 • 11 am ET

Upcoming Community Calls

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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Yoon Jin Choi, Tae Jun Kim, Chang Seok Bang, Yong Kang Lee, Moon Won Lee, Su Youn Nam, Woon Geon Shin, and Seung In Seo** on the publication of **Changing trends and characteristics of peptic ulcer disease: A multicenter study from 2010 to 2019 in Korea** in the *World Journal of Gastroenterology*.

World Journal of Gastroenterology

About the Journal | Submit a Manuscript | Current Issue | Search All Articles

Home / Archive / Volume 29, Issue 44

This Article | Retrospective Study | Open Access

Abstract
Core Tip
Article Highlights
Full Article (PDF)
Full Article with Cover (PDF)
Full Article (WORD)
Full Article (XML)
Full Article (HTML)
Audio
PubMed Central
PubMed
CrossRef
Google Scholar
Similar Articles (15)
Timeline of Article Publication (6)
Authors Evaluation (1)
Article Quality Tracking (1)
Reference Citation Analysis (0)
Academic Content and Language Evaluation of This Article
CrossCheck and Google

World J Gastroenterol. Nov 28, 2023; 29(44): 5882-5893
Published online Nov 28, 2023. doi: 10.3748/wjg.v29.i44.5882

Changing trends and characteristics of peptic ulcer disease: A multicenter study from 2010 to 2019 in Korea

Yoon Jin Choi, Tae Jun Kim, Chang Seok Bang, Yong Kang Lee, Moon Won Lee, Su Youn Nam, Woon Geon Shin, Seung In Seo

Yoon Jin Choi, Department of Internal Medicine, National Cancer Center, Goyang-si 13620, South Korea

Tae Jun Kim, Department of Internal Medicine, Samsung Medical Center, Seoul 06351, South Korea

Chang Seok Bang, Department of Internal Medicine, Chuncheon Sacred Heart Hospital, Chuncheon 24253, South Korea

Yong Kang Lee, Department of Internal Medicine, National Health Insurance Service Ilsan Hospital, Goyang-si 10444, South Korea

Moon Won Lee, Department of Internal Medicine, Pusan National University School of Medicine, Busan 50463, South Korea

Su Youn Nam, Division of Gastroenterology and Hepatology, Department of Internal Medicine, Kyungpook National University Chilgok Hospital, Daegu 41404, South Korea

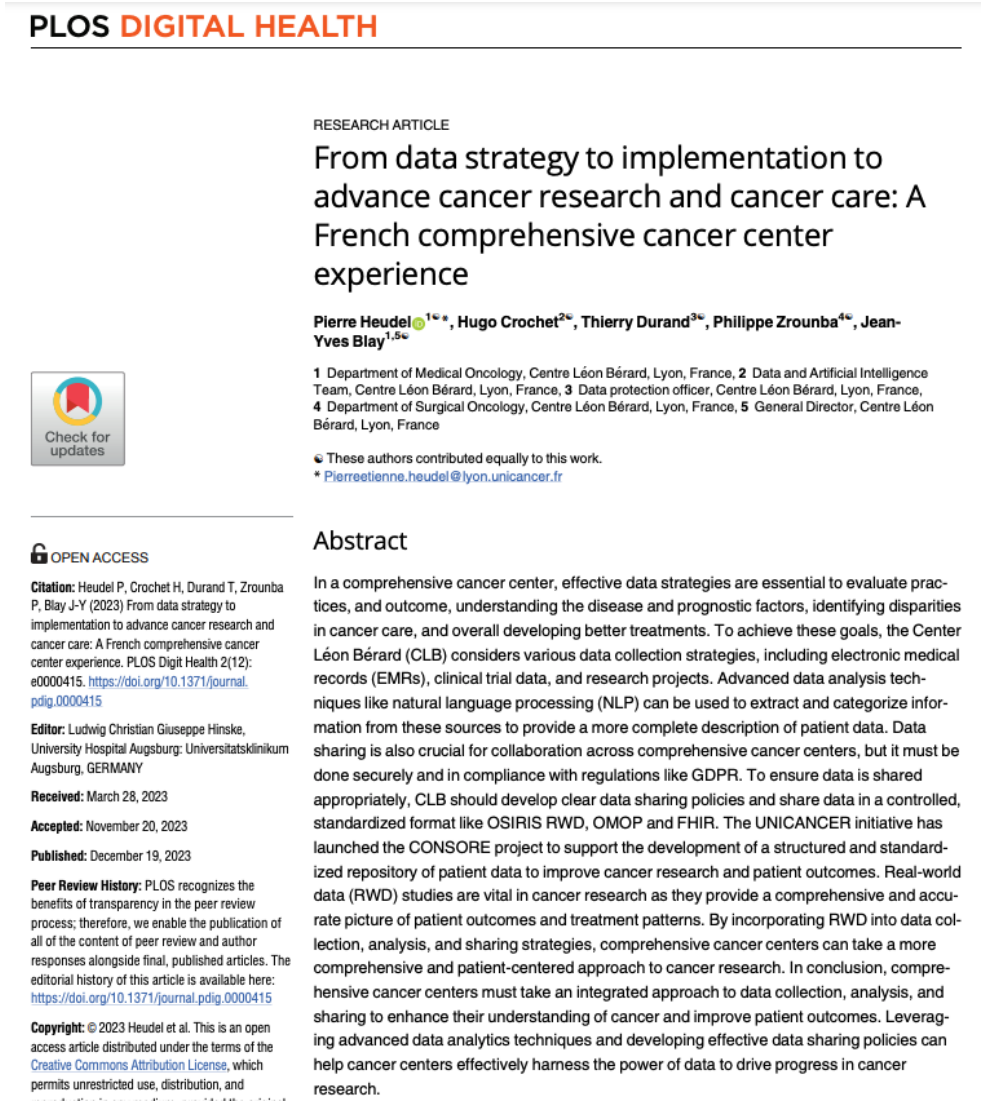
Woon Geon Shin, Department of Internal Medicine, Kangdong Sacred Heart Hospital, Hallym University College of Medicine, Seoul 05355, South Korea

Woon Geon Shin, Seung In Seo, Institute for Liver and Digestive Diseases, Hallym University, Chuncheon 24253, South Korea

Seung In Seo, Division of Gastroenterology, Department of Internal Medicine, Kangdong Sacred Heart Hospital, Seoul 05355, South Korea

OHDSI Shoutouts!

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



PLOS DIGITAL HEALTH

RESEARCH ARTICLE

From data strategy to implementation to advance cancer research and cancer care: A French comprehensive cancer center experience

Pierre Heudel[1☯*], **Hugo Crochet**[2☯], **Thierry Durand**[3☯], **Philippe Zrounba**[4☯], **Jean-Yves Blay**[1,5☯]

1 Department of Medical Oncology, Centre Léon Bérard, Lyon, France, **2** Data and Artificial Intelligence Team, Centre Léon Bérard, Lyon, France, **3** Data protection officer, Centre Léon Bérard, Lyon, France, **4** Department of Surgical Oncology, Centre Léon Bérard, Lyon, France, **5** General Director, Centre Léon Bérard, Lyon, France

☯ These authors contributed equally to this work.
* Pierreetienne.heudel@lyon.unicancer.fr

Check for updates

OPEN ACCESS

Citation: Heudel P, Crochet H, Durand T, Zrounba P, Blay J-Y (2023) From data strategy to implementation to advance cancer research and cancer care: A French comprehensive cancer center experience. PLOS Digit Health 2(12): e0000415. https://doi.org/10.1371/journal.pdig.0000415

Editor: Ludwig Christian Giuseppe Hinske, University Hospital Augsburg: Universitatsklinikum Augsburg, GERMANY

Received: March 28, 2023

Accepted: November 20, 2023

Published: December 19, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pdig.0000415

Abstract

In a comprehensive cancer center, effective data strategies are essential to evaluate practices, and outcome, understanding the disease and prognostic factors, identifying disparities in cancer care, and overall developing better treatments. To achieve these goals, the Center Léon Bérard (CLB) considers various data collection strategies, including electronic medical records (EMRs), clinical trial data, and research projects. Advanced data analysis techniques like natural language processing (NLP) can be used to extract and categorize information from these sources to provide a more complete description of patient data. Data sharing is also crucial for collaboration across comprehensive cancer centers, but it must be done securely and in compliance with regulations like GDPR. To ensure data is shared appropriately, CLB should develop clear data sharing policies and share data in a controlled, standardized format like OSIRIS RWD, OMOP and FHIR. The UNICANCER initiative has launched the CONSORE project to support the development of a structured and standardized repository of patient data to improve cancer research and patient outcomes. Real-world data (RWD) studies are vital in cancer research as they provide a comprehensive and accurate picture of patient outcomes and treatment patterns. By incorporating RWD into data collection, analysis, and sharing strategies, comprehensive cancer centers can take a more comprehensive and patient-centered approach to cancer research. In conclusion, comprehensive cancer centers must take an integrated approach to data collection, analysis, and sharing to enhance their understanding of cancer and improve patient outcomes. Leveraging advanced data analytics techniques and developing effective data sharing policies can help cancer centers effectively harness the power of data to drive progress in cancer research.

OHDSI Shoutouts!

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

Yang *et al. Journal of Big Data* (2024) 11:7
https://doi.org/10.1186/s40537-023-00857-7

Journal of Big Data

RESEARCH Open Access

Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data

Cynthia Yang[1*], Egill A. Fridgeirsson[1], Jan A. Kors[1], Jenna M. Reps[2] and Peter R. Rijnbeek[1]

*Correspondence:
c.yang@erasmusmc.nl

[1] Department of Medical Informatics, Erasmus University Medical Center, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands
[2] Observational Health Data Analytics, Janssen Research and Development, Titusville, NJ, USA

Abstract

Background: There is currently no consensus on the impact of class imbalance methods on the performance of clinical prediction models. We aimed to empirically investigate the impact of random oversampling and random undersampling, two commonly used class imbalance methods, on the internal and external validation performance of prediction models developed using observational health data.

Methods: We developed and externally validated prediction models for various outcomes of interest within a target population of people with pharmaceutically treated depression across four large observational health databases. We used three different classifiers (lasso logistic regression, random forest, XGBoost) and varied the target imbalance ratio. We evaluated the impact on model performance in terms of discrimination and calibration. Discrimination was assessed using the area under the receiver operating characteristic curve (AUROC) and calibration was assessed using calibration plots.

Results: We developed and externally validated a total of 1,566 prediction models. On internal and external validation, random oversampling and random undersampling generally did not result in higher AUROCs. Moreover, we found overestimated risks, although this miscalibration could largely be corrected by recalibrating the models towards the imbalance ratios in the original dataset.

Conclusions: Overall, we found that random oversampling or random undersampling generally does not improve the internal and external validation performance of prediction models developed in large observational health databases. Based on our findings, we do not recommend applying random oversampling or random undersampling when developing prediction models in large observational health databases.

Keywords: Patient-level prediction, Clinical prediction model, Class Imbalance Problem, Machine learning, External validation, Clinical decision support

OHDSI Shoutouts!

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

Journal of the American Medical Informatics Association, 2024, 1–8
https://doi.org/10.1093/jamia/ocad247
Research and Applications

Research and Applications

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

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

[1]Coordinating Center, Observational Health Data Sciences and Informatics, New York City NY 10032, United States, [2]OHDSI Center at the Roux Institute, Northeastern University, Portland ME 04101, United States, [3]Department of Medical Informatics, Erasmus University Medical Center, 3015 GD Rotterdam, The Netherlands, [4]Department of Biomedical Informatics, Columbia University Medical Center, New York City NY 10032, United States, [5]Odysseus Data Services, Cambridge MA 02142, United States, [6]Observational Health Data Analytics, Janssen Research & Development, Titusville NJ 08560, United States

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

Abstract

Importance: The Observational Health Data Sciences and Informatics (OHDSI) is the largest distributed data network in the world encompassing more than 331 data sources with 2.1 billion patient records across 34 countries. It enables large-scale observational research through standardizing the data into a common data model (CDM) (Observational Medical Outcomes Partnership [OMOP] CDM) and requires a comprehensive, efficient, and reliable ontology system to support data harmonization.

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

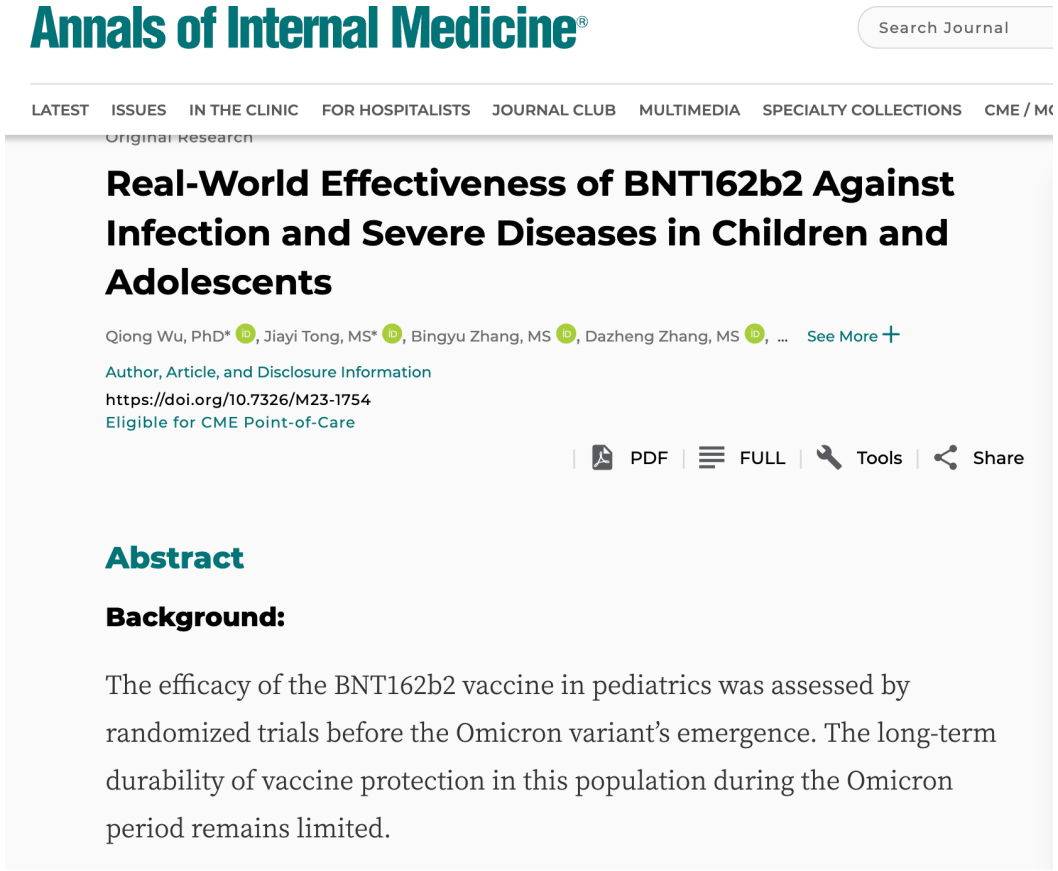
Results: The OHDSI Standardized Vocabularies comprise over 10 million concepts from 136 vocabularies. They are used by hundreds of groups and several large data networks. More than 8600 users have performed 50 000 downloads of the system. This open-source resource has proven to address an impediment of large-scale observational research—the dependence on the context of source data representation. With that, it has enabled efficient phenotyping, covariate construction, patient-level prediction, population-level estimation, and standard reporting.

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

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

OHDSI Shoutouts!

Congratulations to the team of **Qiong Wu, Jiayi Tong, Bingyu Zhang, Dazheng Zhang, Jiajie Chen, Yuqing Lei, Yiwen Lu, Yudong Wang, Lu Li, Yishan Shen, Jie Xu, L. Charles Bailey, Jiang Bian, Dimitri A. Christakis, Megan L. Fitzgerald, Kathryn Hirabayashi, Ravi Jhaveri, Alka Khaitan, Tianchen Lyu, Suchitra Rao, Hanieh Razzaghi, Hayden T. Schwenk, Fei Wang, Margot I. Gage Witvliet, Eric J. Tchetgen Tchetgen, Jeffrey S. Morris, Christopher B. Forrest, and Yong Chen** on the publication of **Real-World Effectiveness of BNT162b2 Against Infection and Severe Diseases in Children and Adolescents** in *Annals of Internal Medicine*.



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Common Data Model Vocabulary Subgroup
Tuesday	6 pm	Eyecare & Vision Research
Wednesday	9 am	Patient-Level Prediction
Wednesday	2 pm	Natural Language Processing
Wednesday	4 pm	Vulcan/OHDSI Meeting
Wednesday	7 pm	Medical Imaging
Thursday	8 am	India Chapter
Thursday	9 am	Medical Devices
Thursday	9:30 am	Data Network Quality
Thursday	12 pm	Strategus HADES Subgroup
Thursday	7 pm	Dentistry
Friday	9 am	GIS – Geographic Information System General
Friday	9 am	Phenotype Development and Evaluation
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering Group
Friday	10 pm	China Chapter
Monday	10 am	Healthcare Systems Interest Group
Monday	10 am	Africa Chapter
Monday	11 am	Data Bricks User Group
Tuesday	10 am	Registry

Collaborator Spotlight: Chungsoo Kim

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

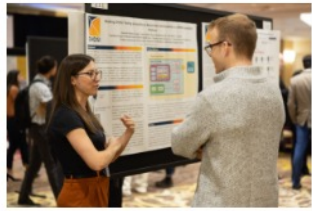
Since joining OHDSI in 2019, he has participated in and led several research projects at OHDSI. He currently participates in OHDSI working groups, including PatientLevelPrediction and the APAC group. He also served as a tutorial instructor for the 2019 OHDSI Korea International Symposium.

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

Can you discuss your research focuses at Ajou University?

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

January 2024 OHDSI Newsletter



The Journey Newsletter (January 2024)

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

Video Podcast: A Look Back, A Look Ahead

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

Video Reflection: 10 Years of OHDSI

On Dec. 16, 2013, George 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. How did we get from there to a global community of more than 3,800 collaborators? The Dec. 12 community call reflected on 10 years of OHDSI, with a video presentation led by Patrick Ryan.

The presentation highlights several of the firsts in the community, including its first publication (which now has more than 1,000 citations), first symposia in the United States, Europe and the Asia-Pacific region, first open-source tools, and plenty more. It also reflects on some of the clinical impacts made by the OHDSI community.

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

December Publications

Rueda M, Leist IC, Gut IG. Convert-Pheno: A software toolkit for the interconversion of standard data models for phenotypic data. J Biomed Inform. 2023 Nov 29;149:104558. doi: 10.1016/j.jbi.2023.104558. Epub ahead of print. PMID: 38035971.

Mayer CS. Conversion of CPRD AURUM Data into the OMOP Common Data Model. Inform Med Unlocked. 2023;43:101407. doi: 10.1016/j.imu.2023.101407. Epub 2023 Nov 10. PMID: 38046363; PMCID: PMC10688258.

Oja M, Tamm S, Mooses K, Pajusalu M, Talvik HA, Ott A, Laht M, Malk M, Lõo M, Holm J, Haug M, Šuvalov H, Särg D, Vilo J, Laur S, Kolde R, Reisberg S. Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned. JAMIA Open. 2023 Dec 5;6(4):ooad100. doi: 10.1093/jamiaopen/ooad100. PMID: 38058679; PMCID: PMC10697784.

Kalokyri V, Kondylakis H, Sfakianakis S, Nikiforaki K, Karatzanis I, Mazzetti S, Tachos N, Regge D, Fotiadis DI, Marias K, Tsiknakis M. MI-Common Data Model: Extending Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM) for Registering Medical Imaging Metadata and Subsequent Curation Processes. JCO Clin Cancer Inform. 2023 Sep;7:e2300101. doi: 10.1200/CCI.23.00101. PMID: 38061012; PMCID: PMC10715775.

Choi K, Park SJ, Han S, Mun Y, Lee DY, Chang DJ, Kim S, Yoo S, Woo SJ, Park KH, Suh HS. Patient-Centered Economic Burden of Exudative Age-Related Macular Degeneration: Retrospective Cohort Study. JMIR Public Health Surveill. 2023 Dec 8;9:e49852. doi: 10.2196/49852. PMID: 38064251; PMCID: PMC10746973.

Blasini R, Buchowicz KM, Schneider H, Samans B, Sohrabi K. Implementation of inclusion and exclusion criteria in clinical studies in OHDSI ATLAS software. Sci Rep. 2023 Dec 18;13(1):22457. doi: 10.1038/s41598-023-49560-w. PMID: 38105303; PMCID: PMC10725886.

HADES Development Updates: CohortMethod 5.2.0

CohortMethod 5.2.0 Reference Articles Changelog HADES

CohortMethod

R-CMD-check passing codecov 89%

CohortMethod is part of HADES.

Introduction

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

Features

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

Links

Browse source code

Report a bug

Ask a question

License

Apache License 2.0

Citation

Citing CohortMethod

Developers

Martijn Schuemie
Author, maintainer

Marc Suchard
Author

Patrick Ryan
Author

HADES Development Updates: EmpiricalCalibration 3.1.2

EmpiricalCalibration 3.1.2 | Reference | Articles | Changelog | HADES

EmpiricalCalibration

R-CMD-check passing | codecov 84% | CRAN 3.1.2 | downloads 648/month

EmpiricalCalibration is part of HADES.

Introduction

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

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

Features

- Estimate the empirical null distribution given the effect estimates of a set of negative controls.
- Estimate the calibrated p-value of a given hypothesis given the estimated empirical null distribution.

Links

View on CRAN

Browse source code

Report a bug

Ask a question

License

Apache License 2.0

Citation

Citing EmpiricalCalibration

Developers

Martijn Schuemie
Author, maintainer

Marc Suchard
Author

HADES Development Updates: DeepPatientLevelPrediction 2.0.3

DeepPatientLevelPrediction 2.0.3 | Get started | My first deep learning model | Reference | Articles | Changelog | HADES

DeepPatientLevelPrediction

R-CMD-check passing | codecov 99%

Introduction

DeepPatientLevelPrediction is an R package for building and validating deep learning patient-level predictive models using data in the OMOP Common Data Model format and OHDSI PatientLevelPrediction framework.

Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018;25(8):969-975.

Features

- Adds deep learning models to use in the OHDSI PatientLevelPrediction framework.
- Allows to add custom deep learning models.
- Includes an MLP, ResNet and a Transformer
- Allows to use all the features of PatientLevelPrediction to validate and explore your model performance.

Technology

Links

Browse source code

Report a bug

Ask a question

License

Apache License 2.0

Citation

Citing DeepPatientLevelPrediction

Developers

Egill Fridgeirsson
Author, maintainer

Jenna Reps
Author

Seng Chan You
Author

Chungsoo Kim
Author

Henrik John

HADES Development Updates: PheValuator 2.2.11

PheValuator 2.2.11 Reference Articles Changelog hadesLogo

PheValuator

PheValuator is part of HADES.

Introduction

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

Process Steps

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

Step 1: Develop Evaluation Cohort from Diagnostic Predictive Model

Create **Evaluation Cohort** of 1M randomly selected subjects from database

Links

Browse source code

Report a bug

Ask a question

License

Apache License 2.0

Citation

Citing PheValuator

Developers

Joel N. Swerdel

Maintainer

Dev status

R-CMD-check passing

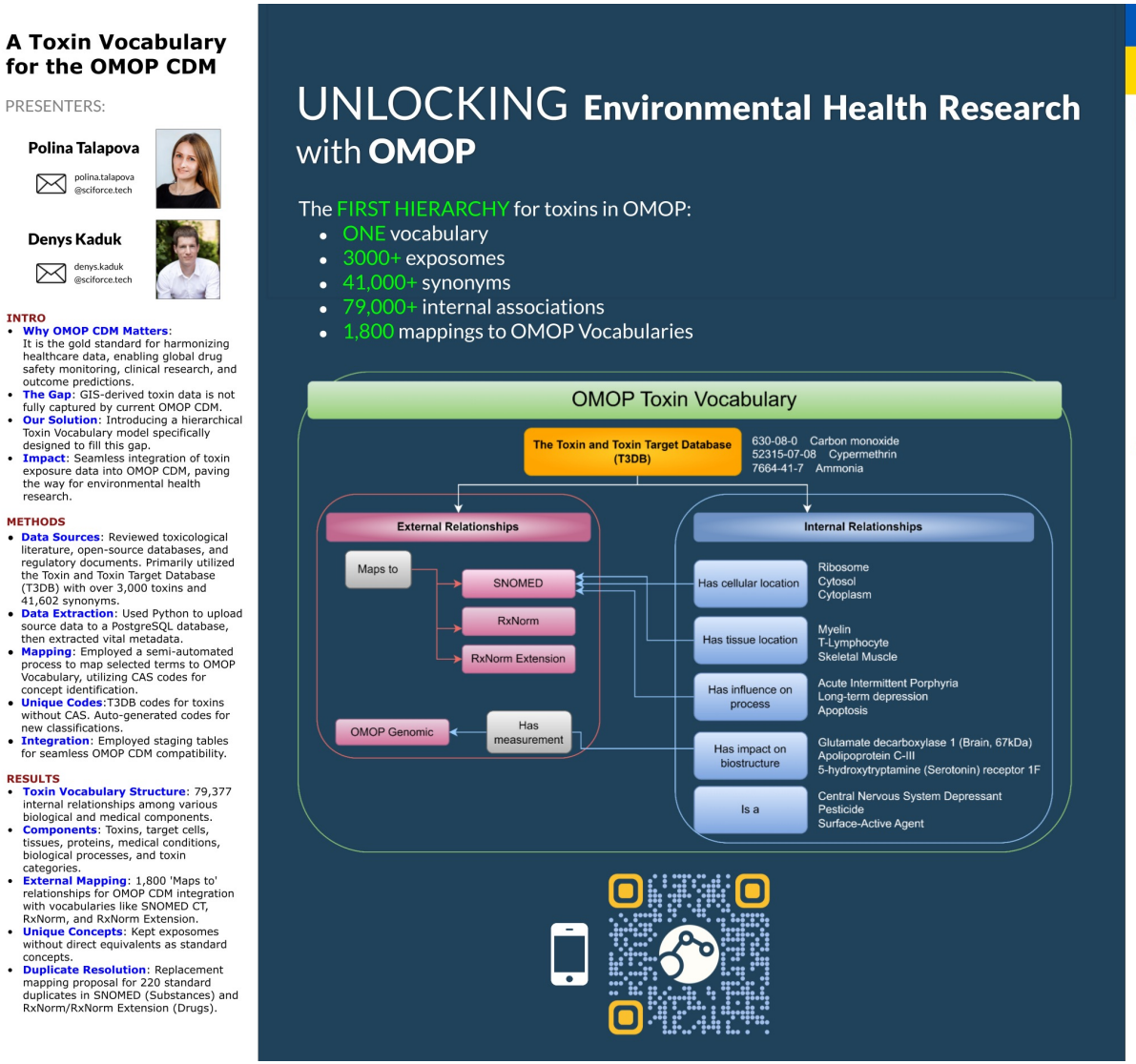
codecov 72%

#OHDSISocialShowcase This Week

MONDAY

A Toxin Vocabulary for the OMOP CDM

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



A Toxin Vocabulary for the OMOP CDM

PRESENTERS:

Polina Talapova
polina.talapova @sciforce.tech

Denys Kaduk
denys.kaduk @sciforce.tech

INTRO

- **Why OMOP CDM Matters**: It is the gold standard for harmonizing healthcare data, enabling global drug safety monitoring, clinical research, and outcome predictions.
- **The Gap**: GIS-derived toxin data is not fully captured by current OMOP CDM.
- **Our Solution**: Introducing a hierarchical Toxin Vocabulary model specifically designed to fill this gap.
- **Impact**: Seamless integration of toxin exposure data into OMOP CDM, paving the way for environmental health research.

METHODS

- **Data Sources**: Reviewed toxicological literature, open-source databases, and regulatory documents. Primarily utilized the Toxin and Toxin Target Database (T3DB) with over 3,000 toxins and 41,602 synonyms.
- **Data Extraction**: Used Python to upload source data to a PostgreSQL database, then extracted vital metadata.
- **Mapping**: Employed a semi-automated process to map selected terms to OMOP Vocabulary, utilizing CAS codes for concept identification.
- **Unique Codes**:T3DB codes for toxins without CAS. Auto-generated codes for new classifications.
- **Integration**: Employed staging tables for seamless OMOP CDM compatibility.

RESULTS

- **Toxin Vocabulary Structure**: 79,377 internal relationships among various biological and medical components.
- **Components**: Toxins, target cells, tissues, proteins, medical conditions, biological processes, and toxin categories.
- **External Mapping**: 1,800 'Maps to' relationships for OMOP CDM integration with vocabularies like SNOMED CT, RxNorm, and RxNorm Extension.
- **Unique Concepts**: Kept exposomes without direct equivalents as standard concepts.
- **Duplicate Resolution**: Replacement mapping proposal for 220 standard duplicates in SNOMED (Substances) and RxNorm/RxNorm Extension (Drugs).

UNLOCKING **Environmental Health Research** with **OMOP**

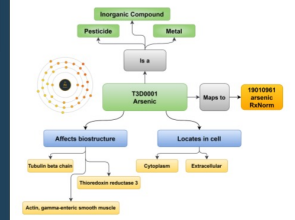
The FIRST HIERARCHY for toxins in OMOP:

- ONE vocabulary
- 3000+ exposomes
- 41,000+ synonyms
- 79,000+ internal associations
- 1,800 mappings to OMOP Vocabularies

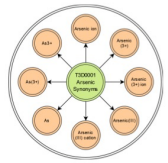
PRINCIPAL DEVELOPER:

Maksym Trofymenko
maksym.trofymenko @sciforce.tech

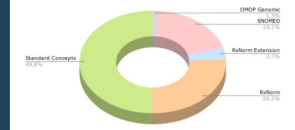
Example of Arsenic's hierarchy in the vocabulary:



Example of Arsenic's synonyms in the vocabulary:



Toxin's Mappings Distribution per Target Vocabulary:



THE TEAM:

Maksym Trofymenko, Polina Talapova, Tetiana Nesmiian, Andrew Williams, Max Ved, Inna Ageeva

sciforce

Tufts CTSI

OHDSI

#OHDSISocialShowcase This Week

TUESDAY

Developing a perinatal expansion table for the OMOP common data model

(Alicia Abellan, Edward Burn, Nhung Trinh, Theresa Burkard, Sergio Fernández-Bertolín, Eimir Hurley, Clara Rodriguez, Elena Segundo, Daniel R. Morales, Hedvig Nordeng, Talita Duarte-Salles)

What ?

- We propose a perinatal expansion with the aim to facilitate future network pregnancy studies using the OMOP Common Data Model (CDM).

Why ?

- The perinatal period is a critical stage of development especially vulnerable to many conditions and risk factors. And pregnant individuals are often excluded from clinical trials.
- Many observational healthcare databases contain detailed information related to the perinatal period.
- The OMOP CDM is a widely used data model for harmonizing observational healthcare data in a standardized format.
- The current OMOP-CDM lacks the structure to define a pregnancy episode and therefore lacks the representation of important features that are associated with the perinatal period.

How ?

- We developed a perinatal expansion involving input from domain experts and stakeholders.
- We defined new fields and content following the structure and vocabularies given OMOP-CDM ontological framework principles.
- We tested the expansion using SIDIAP (EHR from Primary Care, Spain) and Norwegian databases (national registries, Norway).
- We developed a diagnostics package for quality control assessment. Available at github.com/oxford-pharmacoepi/PETDiagnostics.
- We conducted a descriptive analysis on the captured perinatal data.

IDIAP JORDI GOL · EHDEN · Erasmus MC · NDORMS

This perinatal expansion for the OMOP CDM supports network studies in perinatal research

Table 1. Extract of the proposed perinatal expansion for the OMOP CDM

CDM Field	Description
Pregnancy table	
Required fields	
person_id	Unique identifier of the person for whom the pregnancy is recorded
pregnancy_id	Unique identifier of each pregnancy episode
pregnancy_start_date	Date when the pregnancy episode started (based on ultrasound estimations or calculated from last menstrual period if ultrasound is missing)
pregnancy_end_date	Date when the pregnancy episode ended
pregnancy_outcome_id	Outcome of the pregnancy (livebirth, miscarriage [<20 weeks], stillbirth [≥20 weeks], elective termination of pregnancy)
pregnancy_mode_delivery	How the delivery was initiated (vaginal, c-section)
gestational_length_in_day	Length of gestation in days
pregnancy_single	Single pregnancy (yes if single, no if multiple)
Infant table	
Required fields	
pregnancy_id	Unique identifier of each pregnancy episode
fetus_id	Unique identifier of each fetus

Figure 1. Number of starting pregnancies per year

*ETOP: Elective termination of pregnancy

Figure 2. Distribution of gestational length in weeks by pregnancy outcome

Results:

- The expansion consisted of two separate tables: a pregnancy table and an infant table (linked through the individual pregnancy), each with required and optional variables (Table 1).
- The perinatal expansion was successfully implemented in SIDIAP and Norwegian databases. Quality assessment demonstrated accurate capture of perinatal characteristics with minimal missing data.
- We captured **646,530** (2006-2020) pregnancies in **SIDIAP** and **746,671** (2008-2020) pregnancies in **Norwegian** databases (Figure 1).
- We described pregnancy outcomes (e.g., 0.5% stillbirths in SIDIAP and 0.4% in Norway), gestational length (median [IQR] in days, SIDIAP: 273 [56-280]; Norway: 280 [273-286]), number of fetus/infants (Norway: 758,806), and birth weight (median [IQR] in grams, Norway: 3,520 [3,175-3,860]), among other relevant perinatal variables
- The distribution of gestational weeks by pregnancy outcome is depicted in Figure 2.

Conclusions:

- We developed a perinatal expansion to the OMOP-CDM that accurately captures detailed information for perinatal research providing additional information to what is usually retrieved from algorithms
- The expansion can also be used to store algorithm derived perinatal information usually retrieving required fields of the expansion
- The diagnostics package will enable to identify successful implementation of the perinatal expansion in future database mappings with pregnancy and mother-child related information.
- This perinatal expansion is a big step forwards enabling and supporting future perinatal research network studies

Alicia Abellan, Edward Burn, Nhung TH Trinh, Theresa Burkard, Sergio Fernández-Bertolín, Eimir Hurley, Clara Rodriguez, Elena Segundo, Daniel R. Morales, Hedvig Nordeng, Talita Duarte-Salles

theresa.burkard@ndorms.ox.ac.uk

tduarte@idiapjgol.org

#OHDSISocialShowcase This Week

WEDNESDAY

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

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

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

LH John[1], EA Fridgeirsson[1], JM Reps[1,2], JA Kors[1], PR Rijnbeek[1]
[1]Erasmus University Medical Center; [2]Janssen Research and Development

Background

External validation is crucial for ensuring the reliability of prediction models on new data. However, performance often declines during external validation due to database heterogeneity caused by variations in record collection, regulatory guidelines, and database purposes. [1]

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

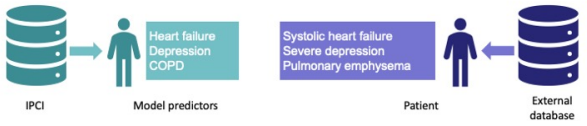


Figure 1. Incompatible model and patient record due to database heterogeneity.

This work aims to utilizes embeddings to approximate clinical concepts, specifically in the context of predicting dementia in persons aged 55-85 in the next five years. This approach may enable external validation of a model even when an exact match for predictors is not found in a patient's record.

Methods

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

Figure 2. Subset of SNOMED medical terms hierarchy with the concept *Clinical finding* as the root.

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

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

Results

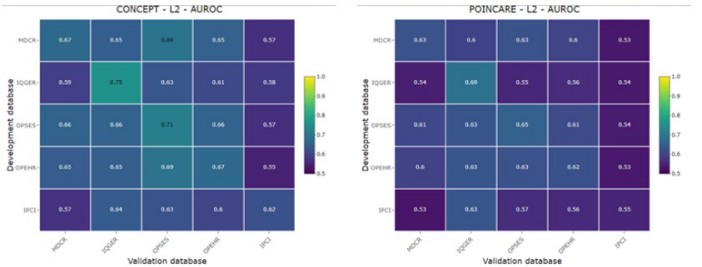


Figure 3. Discrimination of logistic regression using traditional concepts (left) and using the embeddings (right).

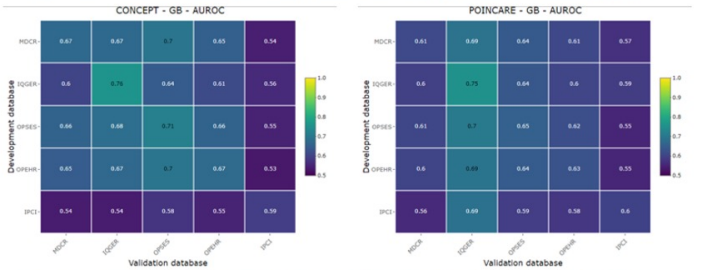


Figure 4. Discrimination of gradient boosting using traditional concepts (left) and using the embeddings (right).

Hyperbolic embeddings do not improve internal or external validation performance of logistic regression models (Figure 3). However, using gradient boosting we can observe that models trained on Integrated Primary Care and Information transport better to Iqvia Disease Analyzer Germany and vise versa. Therefore, we believe clinical domain knowledge from the SNOMED medical terms hierarchy can in some cases be used to improve external validation performance of a clinical prediction model. Future work will investigate under what exact circumstances this holds true and whether more complex models such as a Transformer will have improved validation performance, since training can be done directly on the embedding sequences. Transformers can take the embedding sequence as input directly without the mean aggregation step, which may further improve performance.

1. Chen, D., Liu, S., Kingsbury, P. *et al.* Deep learning and alternative learning strategies for retrospective real-world clinical data. *npj Digit. Med.* 2, 43 (2019).
2. Nickel, Maximillian, and Douwe Kiela. "Poincaré embeddings for learning hierarchical representations." *Advances in neural information processing systems* 30 (2017).

Contact: l.john@erasmusmc.nl

#OHDSISocialShowcase This Week

THURSDAY

Estimating the comparative risk of kidney failure associated with intravitreal anti-vascular endothelial growth factor (anti-VEGF) exposure in patients with blinding diseases

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

Title:
Estimating the comparative risk of kidney failure associated with intravitreal anti-vascular endothelial growth factor (anti-VEGF) exposure in patients with blinding diseases

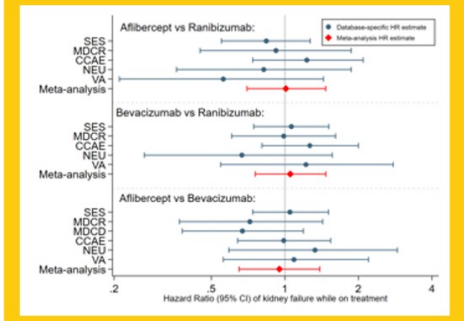
PRESENTER: **Cindy X. Cai**

INTRO:
- Anti-VEGF medications (when given systemically) have adverse kidney effects
- Intravitreal administration leads to systemic absorption (bevacizumab > aflibercept >> ranibizumab)
- Some recommend use of ranibizumab to lower risk of kidney failure (dialysis +/- kidney transplant, aka ESKD/ESRD)
- Unclear if there is evidence to support this practice

PURPOSE:
- A) characterize the incidence of kidney failure associated with intravitreal anti-VEGF exposure
- B) estimate the comparative risk of kidney failure associated with intravitreal anti-VEGF exposure between medications
- C) predict an individual's risk for kidney failure with intravitreal anti-VEGF use

METHODS:
- SOS Challenge: 12 databases (6 administrative claims, 6 electronic health record) in OHDSI Network (standardized to OMOP CDM) [1,2]
- Subjects: adults ≥18 years, new users of ≥3 monthly intravitreal anti-VEGF, for a blinding disease (DR/DME, AMD, RVO), with ≥365 days of prior observation
- Outcome: kidney failure while on-treatment (drug exposure <180 days between injections)
- Incidence rate standardized to 2015 U.S. population by age and sex
- Propensity score method to match patients using 1:1 propensity score matching [3]
- Cox proportional hazards models to estimate risk of kidney failure, meta-analysis to estimate single network-wide estimate [4]
- Machine learning models (L1-regularized logistic regression) to predict risk of kidney failure 6-24 months after cohort entry [5]

Hazard Ratio Estimates for Risk of Kidney Failure Comparing Ranibizumab, Aflibercept, and Bevacizumab



No statistically significant differences in the risk of kidney failure associated with intravitreal anti-VEGF exposure comparing ranibizumab, aflibercept, and bevacizumab

Incidence of Kidney Failure Associated with Intravitreal anti-VEGF exposure

	Patients at Risk	On Treatment Time (person-years)	Number of Outcomes	Incidence Proportion (per 100 persons)	Standardized Incidence Proportion (per 100 persons)§	Incidence Rate of Kidney Failure (per 100 person-years)
Ranibizumab*						
CCAE[1]	3799	3083.8	40	1.08	0.70	1.30
MDCR[2]	7604	8412.6	48	0.63	0.82	0.57
MDCD[3]	1265	1122.3	17	1.34	2.39	1.52
OptumEHR[4]	2520	2491.0	15	0.60	0.61	0.60
SES[5]	8048	9848.2	65	0.81	0.84	0.66
JMDC[6]	203	133.6	0	0.00	0.00	0.00
JHME[7]	19	12.9	0	0.00	0.00	0.00
NEU[8]	2084	2181.3	14	0.67	0.40	0.64
CUMC[9]	117	114.4	0	0.00	0.00	0.00
VA[10]	3943	2538.6	25	0.63	0.25	0.98
USC[11]	7	9.9	0	0.00	0.00	0.00
Aflibercept						
CCAE	3319	3251.7	56	1.69	1.67	1.72
MDCR	4644	5536.7	23	0.50	0.16	0.42
MDCD	1717	1796.1	32	1.86	1.45	1.78
OptumEHR	3282	4698.5	24	0.73	0.39	0.51
SES	8056	11011.5	72	0.89	1.41	0.65
JMDC	205	190.7	1	0.49	0.33	0.52
JHME	574	656.0	8	1.39	1.83	1.22
NEU	3696	3713.0	17	0.46	0.41	0.46
CUMC	335	466.3	0	0.00	0.00	0.00
VA	6266	6245.2	56	0.89	0.79	0.90
USC	59	89.0	1	1.69	0.32	1.12
Bevacizumab						
CCAE	10508	6777.1	104	0.99	0.68	1.54
MDCR	10625	9050.0	50	0.47	0.11	0.55
MDCD	3845	2632.7	70	1.82	2.20	2.66
OptumEHR	11933	12648.0	69	0.58	1.34	0.55
SES	52642	50615.9	317	0.60	1.07	0.63
JMDC	0	0.0	0	NA	0.93	NA
JHME	286	226.2	2	0.70	0.33	0.88
NEU	8331	6279.9	25	0.30	0.00	0.40
CUMC	74	41.1	0	0.00	0.00	0.00
VA	10037	5930.2	58	0.58	0.30	0.98
USC	27	20.1	0	0.00	0.00	0.00

	Standardized Incidence Proportion of Kidney Failure (per 100 persons)	Incidence Rate of Kidney Failure (per 100 person-years)
Intravitreal anti-VEGF	0.68	0.74
Blinding diseases	4.91	0.79
Literature (US Renal Data System)	0.0363 (in 2020)	

* Only patients who contribute at least 1 day to time-at-risk are included in this analysis.
§ Standardized to the 2015 U.S. Population by age and sex
1 CCAE = IBM Health MarketScan Commercial Claims and Encounters Database, USA commercially insured <65 years
2 MDCR = IBM Health MarketScan Medicare Supplemental and Coordination of Benefits Database, USA commercially insured 65+ years
3 MDCD = IBM Health MarketScan Multi-State Medicaid Database, USA Medicaid enrollees
4 OptumEHR = Optum(R) de-identified Electronic Health Record Dataset, USA combined insured claims and electronic health records
5 SES = Optum's Clinformatics Extended Data Mart - Socio-economic Status, USA commercially insured
6 JMDC = Japan Medical Data Center, Japan insurance claims <65 years
7 JHME = Johns Hopkins Medical Enterprise, USA non-profit academic medical center
8 NEU = PharMetrics Plus, USA commercially insured <65 years
9 CUMC = Columbia University Medical Center, USA non-profit academic medical center
10 VA = Department of Veterans Affairs, veterans in the USA
11 USC = University of Southern California, USA non-profit academic medical center

Area Under the Curve (AUROC) of Prediction Models

Exposure Group	Minimum AUROC	Mean AUROC	Maximum AUROC
Ranibizumab	0.813	0.877	0.915
Aflibercept	0.855	0.913	0.979
Bevacizumab	0.733	0.866	0.929

Database included: CCAE, MDCR, MDCD, OptumEHR, SES, NEU

OHDSI Tools Used
- ATLAS
- PheValuator
- Strategus execution pipeline to call Hades Packages (CohortGenerator, Characterization, Cohort Incidence, Cohort Method, PatientLevelPrediction)
- EvidenceSynthesis

References:
1. GitHub - ohdsi-studies/AntiVegfKidneyFailure [Internet]. GitHub. [cited 2023 Jun 9]. Available from: https://github.com/ohdsi-studies/AntiVegfKidneyFailure
2. OHDSI Analysis Viewer. Accessed August 27, 2023. https://data.ohdsi.org/AntiVegfKidneyFailure/
3. Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. Int J Epidemiol. 2018 Dec 1;47(6):2005–14.
4. Schuemie MJ, Chen Y, Madigan D, Suchard MA. Combining cox regressions across a heterogeneous distributed research network facing small and zero counts. Stat Methods Med Res. 2022 Mar;31(3):438–50.
5. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018 Aug 1;25(8):969–75.

Author List

Cindy X. Cai	Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, MD
Mary Grace Bowring	Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD
Diep Tran	Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, MD
Paul Nagy	Department of Biomedical Informatics and Data Science, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, MD
Michael Cook	Johns Hopkins University, Baltimore, MD
Akihiko Nishimura	Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
Jia Ng	Division of Kidney Diseases and Hypertension, Donald and Barbara School of Medicine at Hofstra/Northwell, NY
Marc A. Suchard	VA Informatics and Computing Infrastructure, US Department of Veterans Affairs, Salt Lake City, UT; and Department of Biostatistics, University of California Los Angeles, Los Angeles, CA
Scott L. DuVall	VA Informatics and Computing Infrastructure, US Department of Veterans Affairs, Salt Lake City, UT; and Department of Internal Medicine Division of Epidemiology, University of Utah School of Medicine, Salt Lake City, UT
Michael Matheny	VA Informatics and Computing Infrastructure, Tennessee Valley Healthcare System, Nashville, TN; and Department of Biomedical Informatics, Vanderbilt University, Nashville, TN
Asieh Golozar	Odysseus Data Services, Inc., Cambridge, MA, OHDSI Center at the Roux Institute, Northeastern University, Boston, MA
Anna Ostropolets	Odysseus Data Services, Inc., Cambridge, MA
Evan Minty	O'Brien Center for Public Health, Department of Medicine, University of Calgary, Canada
Fan Bu	Department of Biostatistics, University of California - Los Angeles, Los Angeles, CA
Brian Toy	Roski Eye Institute, Keck School of Medicine, University of Southern California; Los Angeles, CA
Will Halfpenny	University of Cambridge, Cambridge, UK
Michelle Hribar	National Eye Institute, National Institutes of Health, Bethesda, MD; and Casey Eye Institute, Oregon Health & Science University, Portland, OR
Jody-Ann McLeggon	Department of Biomedical Informatics, Columbia University
Thomas Falconer	Department of Biomedical Informatics, Columbia University
Linying Zhang	Department of Biomedical Informatics, Columbia University
George Hripcsak	Department of Biomedical Informatics, Columbia University
Laurence Lawrence-Archer	Odysseus Data Services, Inc., Cambridge, MA, OHDSI Center at the Roux Institute, Northeastern University, Boston, MA
Nathan Hall	Janssen Research and Development, Titusville, NJ
Azza Shoaibi	Janssen Research and Development, Titusville, NJ
Jenna Reps	Janssen Research and Development, Titusville, NJ
Anthony Sena	Janssen Research and Development, Titusville, NJ
Clair Blacketer	Janssen Research and Development, Titusville, NJ
Joel Swerdel	Janssen Research and Development, Titusville, NJ
Patrick Ryan	Janssen Research and Development, Titusville, NJ
	And many others in the OHDSI Community

JOHNS HOPKINS MEDICINE · WILMER EYE INSTITUTE · OHDSI

#OHDSISocialShowcase This Week

FRIDAY

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

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

PRESENTER: **Chen** Yanover

Characteristics and Outcomes of >1M Inflammatory Bowel Disease Patients

Disease Trajectory of Crohn's Disease and Ulcerative Colitis Patients from Australia 🇦🇺, Korea 🇰🇷, Japan 🇯🇵, the UK 🇬🇧, Germany 🇩🇪, France 🇫🇷, and the USA 🇺🇸

INTRO

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

METHODS

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

- **Data sources**

IBM® MarketScan® Commercial Claims DB	CCAE	■
IBM® MarketScan® Multi-State Medicaid DB	MDCD	■
IBM® MarketScan® Medicare Supplemental DB	MDCR	■
Optum's de-identified Clinformatics® Data Mart DB	CDM	■
IQVIA™ Adjudicated Health Plan Claims	PharMetrics+	■
Optum® de-identified Electronic Health Record DB	Optum EHR	□
Columbia University Irving Medical Center	CUIMC	□
Johns Hopkins Medicine	JHM	□
IQVIA™ Adjudicated Health Plan Claims Data	AMB-EHR	◥
IQVIA™ Disease Analyzer – France	France	□
IQVIA™ Disease Analyzer – Germany	Germany	□
IQVIA™ Medical Research Data – UK	IMRD-UK	□
Insurance claims from Japan	JMDC	■
Ajou University School of Medicine	AUSOM	□
Kangdong Sacred Heart Hospital	KDH	□
IQVIA Australian Longitudinal Patient Data	Australia	□

Geography: USA; Europe; Asia; Australia

Data type: ■ Admin claims; □ EHRs; ◥ Claims + EHRs

Included visits: Outpatient; Inpatient; ER

LIMITATIONS

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

Poster, abstract | Extended Table 1 | Full results

RESULTS

	Crohn's Disease		Ulcerative colitis	
	N	% Female	N	% Female
CCAE	84,959	55.6%	109,500	55.2%
MDCD	16,538	65.2%	16,431	65.9%
MDCR	12,809	58.5%	26,242	58.7%
CDM	61,376	56.1%	93,009	56.9%
PharMetrics+	93,217	53.8%	123,149	53.7%
Optum EHR	118,610	59.5%	119,188	58.2%
CUIMC	4,438	56.6%	3,922	59.0%
JHM	1128	62.5%	892	57.5%
AMB-EHR	51,328	61.6%	58,404	58.6%
France	320	62.5%	299	51.5%
Germany	8,321	59.3%	10,668	53.2%
IMRD-UK	6,936	54.4%	13,924	49.1%
JMDC	1,837	27.4%	12,527	35.6%
AUSOM	431	35.7%	660	36.4%
KDH	44	38.6%	93	39.8%
Australia	210	16.2%*	210	15.7%*

*In the Australian database, 59% of subjects have no designated sex; 13.3% of CD and 18.1 of UC subjects identified as males.

- ✓**% female CD and UC patients in Asia lower than in Europe and USA**, in full cohorts, adults and Asian strata
- ✓Noticeable **decrease, over time, in the average age** of CD diagnosis; weaker trend in UC
- ✓**Abdominal pain reported in ⅓ of IBD patients** in US claims DBs during the year preceding diagnosis; much lower rates in other DBs
- ✓**Anxiety and depression prior to CD and UC diagnosis increased**, in some DBs nearly doubled, since 2000
- ✓**Anxiety, depression, abdominal pain** prior to IBD diagnosis **significantly more common in females** than males
- ✓**No decrease in surgical procedures,** despite improved diagnosis, treatment options

CONTRIBUTORS

KI Research Institute	Chen Yanover, Tal El-Hay, Maytal Bivas-Benita, Pinchas Akiva	**Ajou University**	Rae Woong Park, Kwang Jae Lee, Sung Jae Shin
Janssen R&D, LLC	Erica A Voss, Joel Swerdel, Anna Sheahan, Nathan Hall	**Kandong Sacred Heart Hospital**	Seung In Seo, Kyung-Joo Lee
Columbia University	Jimyung Park, Thomas Falconer	**Johns Hopkins**	Leonard Haas, Paul Nagy, Mary Bowring, Michael Cook, Steven Miller
		Rambam Medical Center	Ramit Magen-Rimon, Yehuda Chowers, Roni Weisshof

OHDSI

Opening: Data Steward at EBMD

Description

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

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

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

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

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

RESPONSIBILITIES AND TASKS

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

Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?