Happy 10th Birthday OHDSI!

OHDSI Community Call
Dec. 12, 2023 • 11 am ET

Upcoming Community Calls

Date	Topic
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?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Marek Oja, Sirli Tamm, Kerli Mooses, Maarja Pajusalu, Harry-Anton Talvik, Anne Ott, Marianna Laht, Maria Malk, Marcus Lõo, Johannes Holm, Markus Haug, Hendrik Šuvalov, Dage Särg, Jaak Vilo, Sven Laur, Raivo Kolde, and Sulev Reisberg** on the publication of **Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned** in *JAMIA Open*.

JAMIA Open, 2023, **6(4)**, ooad100
https://doi.org/10.1093/jamiaopen/ooad100
Research and Applications

Research and Applications

Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned

Marek Oja, PhD[1,*], Sirli Tamm, MSc[1], Kerli Mooses, PhD[1], Maarja Pajusalu, MSc[1], Harry-Anton Talvik, MSc[1,2], Anne Ott, MSc[1], Marianna Laht, MD[1], Maria Malk, MSc[1], Marcus Lõo, MSc[1], Johannes Holm, MSc[1], Markus Haug, MSc[1], Hendrik Šuvalov, MSc[1], Dage Särg, MSc[1,2], Jaak Vilo, PhD[1,2], Sven Laur, PhD[1], Raivo Kolde, PhD[1], Sulev Reisberg, PhD[1,2]

[1]Institute of Computer Science, University of Tartu, 51009 Tartu, Estonia, [2]STACC, 51009 Tartu, Estonia

*Corresponding author: Marek Oja, PhD, Institute of Computer Science, University of Tartu, Narva mnt 18, 51009 Tartu, Estonia (marek.oja@ut.ee)

Author Contributions: Dr M. Oja and S. Tamm are considered co-first authors and Dr R Kolde and Dr S Reisberg are considered co-last authors of this work. In addition, they had full access to all the data in the study and take responsibility for the integrity of data and accuracy of the data analysis.

Abstract

Objective: To describe the reusable transformation process of electronic health records (EHR), claims, and prescriptions data into Observational Medical Outcome Partnership (OMOP) Common Data Model (CDM), together with challenges faced and solutions implemented.

Materials and Methods: We used Estonian national health databases that store almost all residents' claims, prescriptions, and EHR records. To develop and demonstrate the transformation process of Estonian health data to OMOP CDM, we used a 10% random sample of the Estonian population ($n = 150\,824$ patients) from 2012 to 2019 (MAITT dataset). For the sample, complete information from all 3 databases was converted to OMOP CDM version 5.3. The validation was performed using open-source tools.

Results: In total, we transformed over 100 million entries to standard concepts using standard OMOP vocabularies with the average mapping rate 95%. For conditions, observations, drugs, and measurements, the mapping rate was over 90%. In most cases, SNOMED Clinical Terms were used as the target vocabulary.

Discussion: During the transformation process, we encountered several challenges, which are described in detail with concrete examples and solutions.

Conclusion: For a representative 10% random sample, we successfully transferred complete records from 3 national health databases to OMOP CDM and created a reusable transformation process. Our work helps future researchers to transform linked databases into OMOP CDM more efficiently, ultimately leading to better real-world evidence.

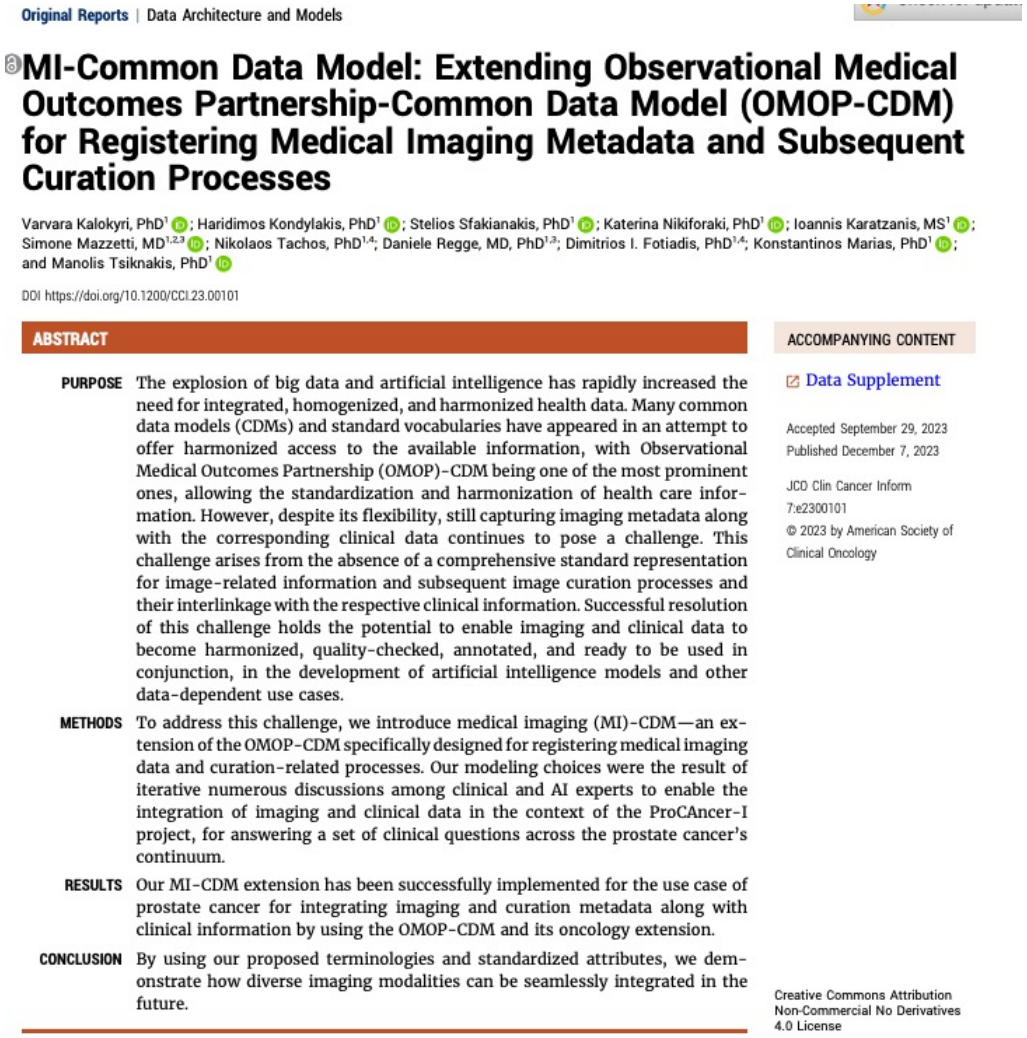
Lay Summary

Health data can be found in various sources and formats, making it challenging for researchers. To address this issue, one possible approach is to transform the data into a standardized common data model (CDM). In this study, we describe the process of converting electronic health records (EHR), claims, and prescriptions data into the Observational Medical Outcome Partnership (OMOP) CDM, along with the challenges faced and solutions implemented. We used Estonian national health databases containing information on claims, prescriptions, and EHR records of 10% of Estonian residents (MAITT dataset). The study describes how data were mapped to standardized vocabulary and successfully converted to the OMOP CDM. We discuss the encountered difficulties and problems and propose solutions to help future researchers transform linked databases into OMOP CDM more efficiently, leading to better real-world evidence.

Key words: OMOP; electronic health record; EHR; ETL; mapping.

OHDSI Shoutouts!

Congratulations to the team of **Varvara Kalokyri, Haridimos Kondylakis, Stelios Sfakianakis, Katerina Nikiforaki, Ioannis Karatzanis, Simone Mazzetti, Nikolaos Tachos, Daniele Regge, Dimitrios Fotiadis, Konstantinos Marias, and Manolis Tsiknakis** on the publication of **MI-Common Data Model: Extending Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM) for Registering Medical Imaging Metadata and Subsequent Curation Processes** in *JCO Clinical Cancer Informatics*.



Original Reports | Data Architecture and Models

MI-Common Data Model: Extending Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM) for Registering Medical Imaging Metadata and Subsequent Curation Processes

Varvara Kalokyri, PhD[1]; Haridimos Kondylakis, PhD[1]; Stelios Sfakianakis, PhD[1]; Katerina Nikiforaki, PhD[1]; Ioannis Karatzanis, MS[1]; Simone Mazzetti, MD[1,2,3]; Nikolaos Tachos, PhD[1,4]; Daniele Regge, MD, PhD[1,2]; Dimitrios I. Fotiadis, PhD[1,4]; Konstantinos Marias, PhD[1]; and Manolis Tsiknakis, PhD[1]

DOI https://doi.org/10.1200/CCI.23.00101

ABSTRACT

PURPOSE The explosion of big data and artificial intelligence has rapidly increased the need for integrated, homogenized, and harmonized health data. Many common data models (CDMs) and standard vocabularies have appeared in an attempt to offer harmonized access to the available information, with Observational Medical Outcomes Partnership (OMOP)-CDM being one of the most prominent ones, allowing the standardization and harmonization of health care information. However, despite its flexibility, still capturing imaging metadata along with the corresponding clinical data continues to pose a challenge. This challenge arises from the absence of a comprehensive standard representation for image-related information and subsequent image curation processes and their interlinkage with the respective clinical information. Successful resolution of this challenge holds the potential to enable imaging and clinical data to become harmonized, quality-checked, annotated, and ready to be used in conjunction, in the development of artificial intelligence models and other data-dependent use cases.

METHODS To address this challenge, we introduce medical imaging (MI)-CDM—an extension of the OMOP-CDM specifically designed for registering medical imaging data and curation-related processes. Our modeling choices were the result of iterative numerous discussions among clinical and AI experts to enable the integration of imaging and clinical data in the context of the ProCAncer-I project, for answering a set of clinical questions across the prostate cancer's continuum.

RESULTS Our MI-CDM extension has been successfully implemented for the use case of prostate cancer for integrating imaging and curation metadata along with clinical information by using the OMOP-CDM and its oncology extension.

CONCLUSION By using our proposed terminologies and standardized attributes, we demonstrate how diverse imaging modalities can be seamlessly integrated in the future.

ACCOMPANYING CONTENT

Data Supplement

Accepted September 29, 2023
Published December 7, 2023

JCO Clin Cancer Inform
7:e2300101

OHDSI Shoutouts!

Congratulations to the team of **Kyungseon Choi, Sang Jun Park, Sola Han, Yongseok Mun, Da Yun Lee, Dong-Jin Chang, Seok Kim, Sooyoung Yoo, Se Joon Woo, Kyu Hyung Park, and Hae Sun Suh** on the publication of **Patient-Centered Economic Burden of Exudative Age-Related Macular Degeneration: Retrospective Cohort Study** in *JMIR Public Health and Surveillance*.

Original Paper

Patient-Centered Economic Burden of Exudative Age-Related Macular Degeneration: Retrospective Cohort Study

Kyungseon Choi[1,2*], MSc, PharmD; Sang Jun Park[3*], MSc, MD, PhD; Sola Han[4], MSc, PharmD, PhD; Yongseok Mun[5], MD; Da Yun Lee[3], MS; Dong-Jin Chang[6], MSc, MD; Seok Kim[7], MPH; Sooyoung Yoo[7], PhD; Se Joon Woo[3], MD, PhD; Kyu Hyung Park[8], MD, PhD; Hae Sun Suh[1,2,9], MA, MPharm, PhD

[1]Department of Regulatory Science, Graduate School, Kyung Hee University, Seoul, Republic of Korea
[2]Institute of Regulatory Innovation Through Science, Kyung Hee University, Seoul, Republic of Korea
[3]Department of Ophthalmology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[4]Health Outcomes Division, College of Pharmacy, University of Texas at Austin, Austin, TX, United States
[5]Department of Ophthalmology, Kangnam Sacred Heart Hospital, Hallym University College of Medicine, Seoul, Republic of Korea
[6]Department of Ophthalmology and Visual Science, Yeouido St Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[7]Healthcare ICT Research Center, Office of eHealth Research and Businesses, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[8]Department of Ophthalmology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[9]College of Pharmacy, Kyung Hee University, Seoul, Republic of Korea
*these authors contributed equally

Corresponding Author:
Hae Sun Suh, MA, MPharm, PhD
College of Pharmacy
Kyung Hee University
26, Kyunghee-daero, Dongdaemun-gu
Seoul, 02447
Republic of Korea
Phone: 82 29619492
Email: haesun.suh@khu.ac.kr

Abstract

Background: Exudative age-related macular degeneration (AMD), one of the leading causes of blindness, requires expensive drugs such as anti–vascular endothelial growth factor (VEGF) agents. The long-term regular use of effective but expensive drugs causes an economic burden for patients with exudative AMD. However, there are no studies on the long-term patient-centered economic burden of exudative AMD after reimbursement of anti-VEGFs.

Objective: This study aimed to evaluate the patient-centered economic burden of exudative AMD for 2 years, including nonreimbursement and out-of-pocket costs, compared with nonexudative AMD using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

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	3 pm	OMOP CDM Oncology – Outreach/Research Subgroup
Wednesday	9 am	Patient-Level Prediction
Wednesday	1 pm	Perinatal & Reproductive Health
Wednesday	2 pm	Natural Language Processing
Wednesday	4 pm	Vulcan/OHDSI Meeting
Wednesday	7 pm	Medical Imaging
Thursday	8 am	India Chapter
Thursday	9:30 am	Data Network Quality
Thursday	7 pm	Dentistry
Friday	9 am	GIS – Geographic Information System General
Friday	10:30 am	Open-Source Community
Friday	11 am	Clinical Trials
Friday	11:30	Steering Group
Monday	10 am	Healthcare Systems Interest Group
Monday	11 am	Data Bricks User Group
Tuesday	10 am	Registry

OHDSI HADES releases: DatabaseConnector 6.3.2

DatabaseConnector 6.3.2 | Reference | Articles | Changelog | HADES

DatabaseConnector

R-CMD-check failing | codecov 60% | CRAN 6.3.2 | downloads 3293/month

DatabaseConnector is part of HADES.

Introduction

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

Features

- Create connections to the various database platforms:
  - MicrosoftSQL Server
  - Oracle
  - PostgresSql
  - Microsoft Parallel Data Warehouse (a.k.a. Analytics Platform System)
  - Amazon Redshift
  - Apache Impala
  - Google BigQuery

Links

View on CRAN

Browse source code

Report a bug

Ask a question

License

Apache License

Citation

Citing DatabaseConnect

Developers

Martijn Schuemie
Author, maintainer

Marc Suchard
Author

More about authors...

OHDSI HADES releases: SelfControlledCaseSeries 5.1.0

SelfControlledCaseSeries 5.1.0 | Reference | Articles | Changelog | HADES

SelfControlledCaseSeries

R-CMD-check passing | codecov 87%

SelfControlledCaseSeries is part of HADES.

Introduction

SelfControlledCaseSeries is an R package for performing Self-Controlled Case Series (SCCS) analyses in an observational database in the OMOP Common Data Model.

Features

- Extracts the necessary data from a database in OMOP Common Data Model format.
- Optionally add seasonality using a spline function.
- Optionally add age using a spline function.
- Optionally add calendar time using a spline function.
- Optionally correct for event-dependent censoring of the observation period.
- Optionally add many covariates in one analysis (e.g. all drugs).
- Options for constructing different types of covariates and risk windows, including pre-exposure windows (to capture contra-indications).

Links

Browse source code

Report a bug

Ask a question

License

Apache License 2.0

Citation

Citing SelfControlledCaseSeries

Developers

Martijn Schuemie
Author, maintainer

Patrick Ryan
Author

Trevor Shaddox
Author

Marc Suchard
Author

OHDSI HADES releases: DeepPatientLevelPrediction 2.0.2

SelfControlledCaseSeries 5.1.0 Reference Articles Changelog HADES

SelfControlledCaseSeries

R-CMD-check passing codecov 87%

SelfControlledCaseSeries is part of HADES.

Introduction

SelfControlledCaseSeries is an R package for performing Self-Controlled Case Series (SCCS) analyses in an observational database in the OMOP Common Data Model.

Features

- Extracts the necessary data from a database in OMOP Common Data Model format.
- Optionally add seasonality using a spline function.
- Optionally add age using a spline function.
- Optionally add calendar time using a spline function.
- Optionally correct for event-dependent censoring of the observation period.
- Optionally add many covariates in one analysis (e.g. all drugs).
- Options for constructing different types of covariates and risk windows, including pre-exposure windows (to capture contra-indications).

Links

Browse source code

Report a bug

Ask a question

License

Apache License 2.0

Citation

Citing SelfControlledCaseSeries

Developers

Martijn Schuemie
Author, maintainer

Patrick Ryan
Author

Trevor Shaddox
Author

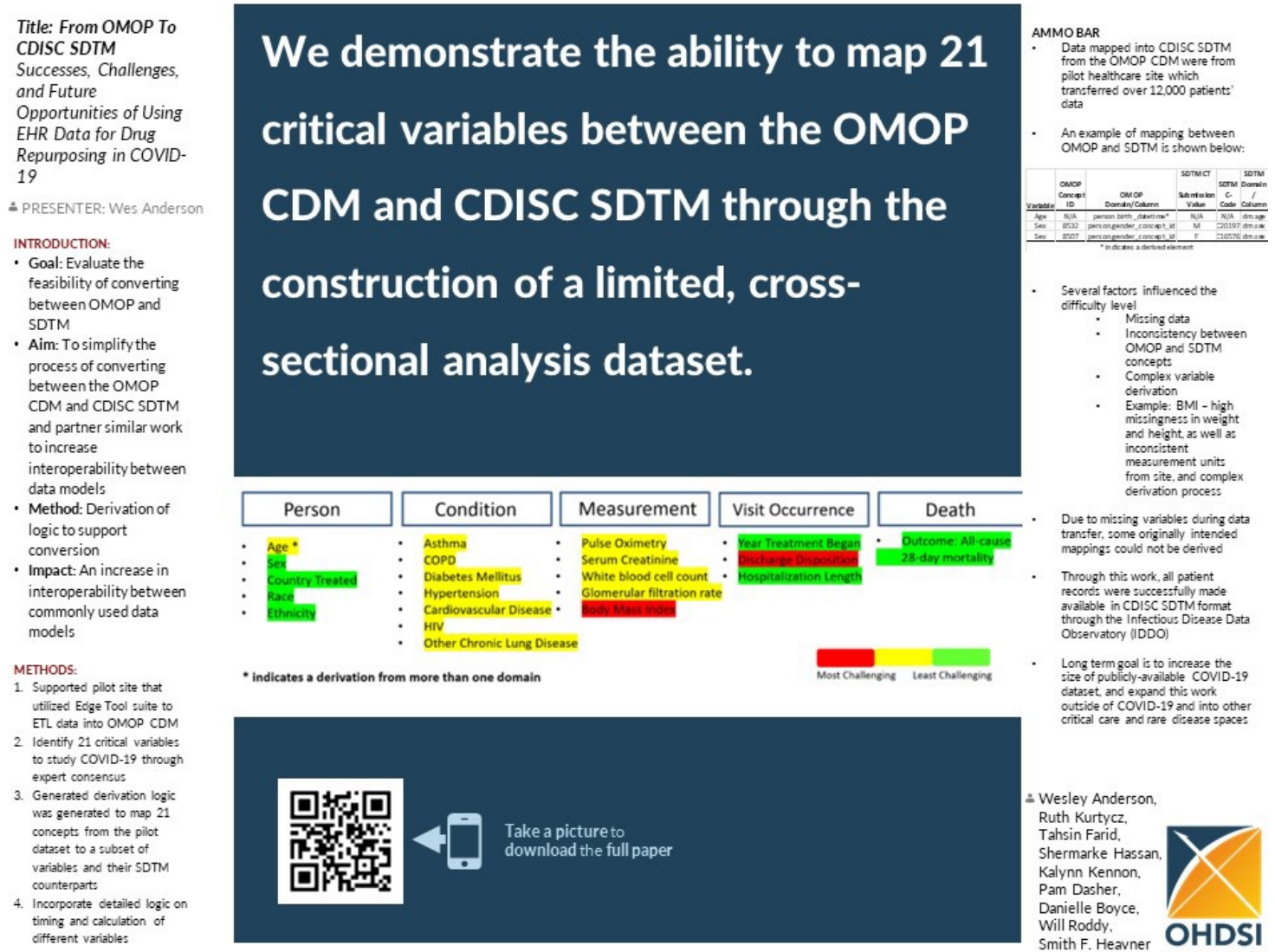
Marc Suchard
Author

#OHDSISocialShowcase This Week

MONDAY

From OMOP to CDISC SDTM: Successes, Challenges, and Future Opportunities of using EHR Data for Drug Repurposing in COVID-19

(Wesley Anderson, Ruth Kurtycz, Tahsin Farid, Shermarke Hassan, Kalynn Kennon, Pam Dasher, Danielle Boyce, Will Roddy, Smith F. Heavner)



Title: From OMOP To CDISC SDTM *Successes, Challenges, and Future Opportunities of Using EHR Data for Drug Repurposing in COVID-19*

PRESENTER: Wes Anderson

INTRODUCTION:

- **Goal:** Evaluate the feasibility of converting between OMOP and SDTM
- **Aim:** To simplify the process of converting between the OMOP CDM and CDISC SDTM and partner similar work to increase interoperability between data models
- **Method:** Derivation of logic to support conversion
- **Impact:** An increase in interoperability between commonly used data models

METHODS:

1. Supported pilot site that utilized Edge Tool suite to ETL data into OMOP CDM
2. Identify 21 critical variables to study COVID-19 through expert consensus
3. Generated derivation logic was generated to map 21 concepts from the pilot dataset to a subset of variables and their SDTM counterparts
4. Incorporate detailed logic on timing and calculation of different variables

We demonstrate the ability to map 21 critical variables between the OMOP CDM and CDISC SDTM through the construction of a limited, cross-sectional analysis dataset.

Person	Condition	Measurement	Visit Occurrence	Death
Age *	Asthma	Pulse Oximetry	Year Treatment Began	Outcome: All-cause 28-day mortality
Sex	COPD	Serum Creatinine	Discharge Disposition	
Country Treated	Diabetes Mellitus	White blood cell count	Hospitalization Length	
Race	Hypertension	Glomerular filtration rate		
Ethnicity	Cardiovascular Disease	Body Mass Index		
	HIV			
	Other Chronic Lung Disease			

* indicates a derivation from more than one domain

Most Challenging — Least Challenging

Take a picture to download the full paper

AMMO BAR

- Data mapped into CDISC SDTM from the OMOP CDM were from pilot healthcare site which transferred over 12,000 patients' data
- An example of mapping between OMOP and SDTM is shown below:

Variable	OMOP Concept ID	OMOP Domain/Column	SDTM CT Submission Value	SDTM C-Code	SDTM Domain / Column
Age	N/A	person.birth_datetime*	N/A	N/A	dm.age
Sex	8532	person.gender_concept_id	M	C20197	dm.sex
Sex	8507	person.gender_concept_id	F	C16576	dm.sex

* indicates a derived element

- Several factors influenced the difficulty level
  - Missing data
  - Inconsistency between OMOP and SDTM concepts
  - Complex variable derivation
  - Example: BMI – high missingness in weight and height, as well as inconsistent measurement units from site, and complex derivation process
- Due to missing variables during data transfer, some originally intended mappings could not be derived
- Through this work, all patient records were successfully made available in CDISC SDTM format through the Infectious Disease Data Observatory (IDDO)
- Long term goal is to increase the size of publicly-available COVID-19 dataset, and expand this work outside of COVID-19 and into other critical care and rare disease spaces

Wesley Anderson, Ruth Kurtycz, Tahsin Farid, Shermarke Hassan, Kalynn Kennon, Pam Dasher, Danielle Boyce, Will Roddy, Smith F. Heavner

OHDSI

#OHDSISocialShowcase This Week

TUESDAY

Mapping gravity value sets to OMOP CDM: The case of the food insecurity screening

(Adam Bouras, Davera Gabriel)

Mapping Gravity Value Sets to OMOP CDM

A food insecurity screening case study

PRESENTER: **Dr. Adam Bouras**

INTRODUCTION

The Office of the National Coordinator for Health IT (ONC) launched the USCDI and USCDI+ initiatives to establish domain, and program-specific datasets. As part of these initiatives, the ONC has developed data element lists for the following subdomains supporting Public Health programs: Case-Based Surveillance, Laboratory Data Exchange, Multi-Directional Exchange with Healthcare and Other Partners, Resource Reporting and Situational Awareness, and Risk Behaviors and Drivers of Inequity Data elements. Supporting the ONC aims, the HL7 Gravity Accelerator developed Social Determinants of Health (SDOH) Value Sets in response. However, only a handful of studies have been published to-date on the standards-based representation of SDOH screening tools. Further, these have yet to demonstrate that the SDOH standards could be used with real-world data or for data exchange between sites using different Common Data Models, with a few exceptions.

METHODS

1. Identified the assessment tools related to the SDOH developed by the Gravity project
2. Manually mapped food insecurity tool using Athena to OMOP CDM
3. Conducted a text similarity analysis to assess the semantic substitutability between screen tools mapped with non-mapped tools

Dr. Bouras would like to acknowledge the support of Dr. Davera Gabriel and the OHDSI OMOP + FHIR WG in developing this submission.

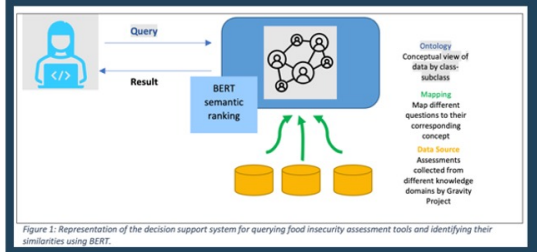
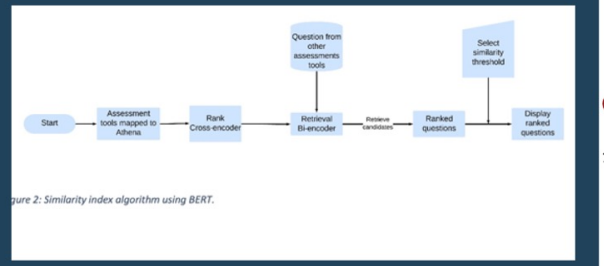


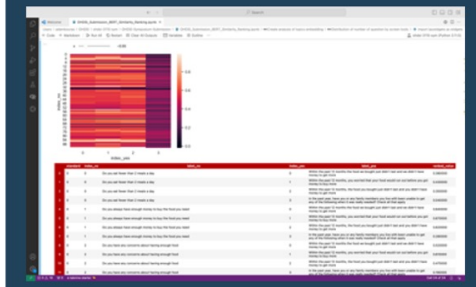
Figure 1: Representation of the decision support system for querying food insecurity assessment tools and identifying their similarities using BERT.

The decision support system for queries food insecurity assessment tools and identifying their similarities using BERT

Similarity index algorithm using BERT



gure 2: Similarity index algorithm using BERT.



Python Jupyter Notebook app displaying the results of the BERT

Take a picture to download the full paper

RESULTS

23 programs identified that were aimed at reducing food insecurity

24 screening tools

98 related questions to assess food insecurity

CONCLUSION

1. Only a few screening tools have been mapped to both FHIR and the OMOP CDM. Further work is needed to identify which questions from these tools should be standardized and mapped and to create corresponding resources within FHIR and OMOP CDM alike
2. Text similarity can be used to streamline standardization and consolidation across different screen tools and questions
3. Large Language Models (LLMs) such as BERT can compute semantic similarity in various Social Determinants of Health (SDOH) constructs. Doing so may encourage more effective utilization of SDOH constructs by Electronic Health Record (EHR) vendors and healthcare providers. LLMs may also provide the OHDSI community with the means to identify missing SDOH questions and answers. In turn, this could support prioritized SDOH content development in the OMOP CDM

JOHNS HOPKINS SCHOOL of MEDICINE · Tritonis inc · OHDSI

#OHDSISocialShowcase This Week

WEDNESDAY

Enhancing Precision and Validity: Leveraging Multiple Error-Prone Phenotypes in EHR-Based Association Studies

(Yiwen Lu, Jiayi Tong, Rebecca A Hubbard, Yong Chen)

Enhancing Precision and Validity: Leveraging Multiple Error-Prone Phenotypes in EHR-Based Association Studies

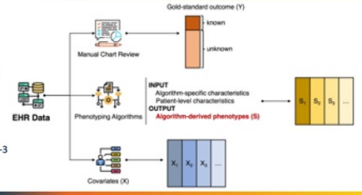
Yiwen Lu[1,2,*], Jiayi Tong[1,*], Rebecca A Hubbard[3], Yong Chen[1,3,4,5]

[1]Center for Health Analytics and Synthesis of Evidence (CHASE), Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[2]The Graduate Group in Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA
[3]Leonard Davis Institute of Health Economics, Philadelphia, PA, USA
[4]Penn Medicine Center for Evidence-based Practice (CEP), Philadelphia, PA, USA
[5]Penn Institute for Biomedical Informatics (IBI), Philadelphia, PA, USA
*: equally contributed

Code available here

Background

- **Objective**
  Present a novel bias-correction approach that harnesses the collective power of all available phenotypes derived from multiple algorithms.
- **Why do we study it?**
  - Manual chart review are often constrained by time and cost limitations[1-3]
  - Cannot get true estimation directly



Methods

- **Comparison between existing methods**

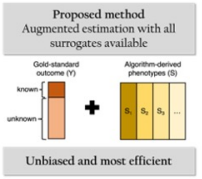
	Description	Pros and cons
Method 1	use validation set only	Unbiased but inefficient
Method 2	use full surrogate only	Efficient but biased
Method 3 (Tong2019)	Augmented estimation with single surrogate[4]	Unbiased, efficient but we want more efficiency

- **Proposed method**

Algorithm

1. Obtain
   (a) $\widehat{\beta}_V$ using validation set
   (b) $\widehat{\gamma}_F^a, \widehat{\gamma}_V^a$ using full set and validation part of k-th surrogate respectively
2. Compute covariance matrices Ω, Σ, Σ^* of $\widehat{\beta}_V - \beta_1$ joint with $\widehat{\gamma}_V - \widehat{\gamma}_F$.
3. Obtain the proposed augmented estimator $\widehat{\beta}_{AM}$ by

$$\widehat{\beta}_{AM} = \widehat{\beta}_V - \widehat{\Omega}^T \widehat{\Sigma}^{*-1} (\widehat{\gamma}_V - \widehat{\gamma}_F)$$



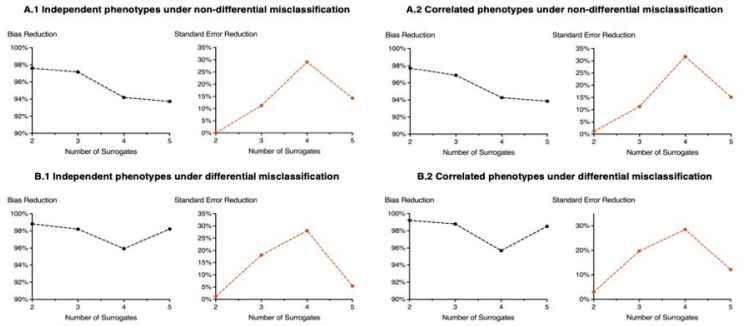
Simulation Settings

- **Two major settings:** nondifferential (Setting A) and differential (Setting B) misclassifications.
- **Two cases:** independent surrogate (Case 1) and correlated surrogates (Case 2)
- Simulated estimation using **different number of surrogates** (two to five)
- Complete data set of **3 sizes** (3000, 5000, and 10000) with the validation set of **4 sizes** (400, 600, 1000, and 2000).
- The complete set of **validation ratios** in our study ranged from 0.04 to 0.4.

References
[1] Williamson T, Green ME, Birtwhistle R, Khan S, Garies S, Wong ST, et al. Validating the 8 CPCSSN Case Definitions for Chronic Disease Surveillance in a Primary Care Database of Electronic Health Records. The Annals of Family Medicine. 2014 Jul 1;12(4):367–72.
[2] Inacio MCS, Paxton EW, Chen Y, Harris J, Eck E, Barnes S, et al. Leveraging Electronic Medical Records for Surveillance of Surgical Site Infection in a Total Joint Replacement Population. Infect Control Hosp Epidemiol. 2011 Apr;32(4):351–9.
[3] Tian TY, Zlateva I, Anderson DR. Using electronic health records data to identify patients with chronic pain in a primary care setting. J Am Med Inform Assoc. 2013 Dec;20(e2): e275–80.
[4] Tong J, Huang J, Chubak J, Wang X, Moore JH, Hubbard RA, et al. An augmented estimation procedure for EHR-based association studies accounting for differential misclassification. Journal of the American Medical Informatics Association. 2020 Feb 1;27(2):244–53.

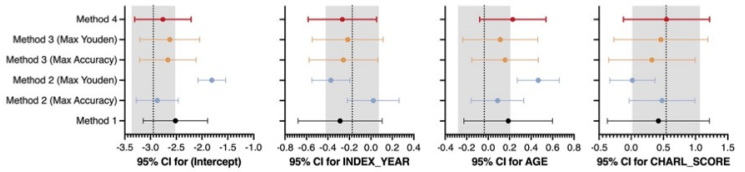
Contact: ychen123@upenn.edu yiwenlu@sas.upenn.edu

Simulation Result



Real Data Evaluation

- **Data source:** Kaiser Permanente Washington (KPW) healthcare system
- **Study cohorts:** 1063 patients aged 18 or older at the time of diagnosis of stage I-IIIA colon cancer between 1995 and 2014.
- **Result:** We included two algorithm-derived phenotypes and compared the point estimates and 95% confidence intervals (CI) for the association (in log odds ratio scale) between covariates and the occurrence of colon cancer events using different methods, where the validation ratio is approximately 0.3.



Conclusion

- We have presented an augmented estimation procedure that effectively addresses both **non-differential and differential classifications** by leveraging the full potential of **multiple independent or correlated** algorithm-derived phenotypes.
- Our method is applicable to OHDSI data.
- Our novel bias-correction approach enhances the efficiency of estimation, contributes to the **reliability and reproducibility** of findings, and ultimately fosters progress in evidence-based healthcare and population health research.

#OHDSISocialShowcase This Week

THURSDAY

Enabling Innovation at the Bedside using STARR-OMOP

(Priya Desai, Alison Callahan, Juan M. Banda, Nikesh Kotecha, Shreya Shah, Somalee Datta)

Enabling Innovation at the Bedside using STARR-OMOP

Priya Desai[1,2], Alison Callahan[1,2], Juan M. Banda[2], Nikesh Kotecha[2], Shreya Shah[1], Somalee Datta[1,2]

[1]Stanford School of Medicine, [2]Stanford Health Care

Stanford MEDICINE
Technology & Digital Solutions

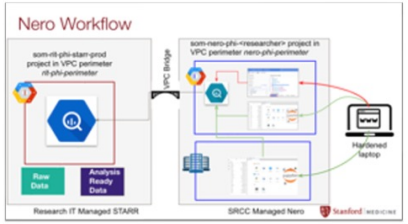
Background

Data coupled with artificial intelligence(AI) and machine learning(ML) can advance both the science and practice of medicine. Large amounts of patient data are now available due to the widespread adoption of electronic health records (EHRs), including imaging data from radiology and pathology etc. Academic Medical Centers (AMCs) are increasingly focused on creating data repositories to harness these data for research that translates to the bedside.

The STAnford Medicine Research data Repository (STARR) is managed by the Stanford Medicine Research Technology team, a unit that supports research at Stanford School of Medicine (SoM), Stanford Health Care(SHC) and Stanford Children's Health(SCH). It hosts multiple linked clinical data warehouses with a range of data types, all in one place on the cloud. STARR contains structured and unstructured, raw and "analysis-ready" data as well as tools to analyze these data. STARR datasets, the ETLs that convert the raw data to analysis ready forms, and Nero, our secure research computing platform are all hosted on a public cloud, specifically, the Google Cloud Platform. The vision of this platform is to provide researchers with the access and tools to explore clinical data seamlessly and accelerate the pace of bench to bedside research.

Methods

STARR-OMOP is a clinical data warehouse supported by STARR containing EHR data from the two hospitals standardized to OHDSI OMOP Common Data Model. This data warehouse contains data for ~3.7 million patients and is refreshed weekly. A PHI scrubbed version of STARR-OMOP is made available for AI and population health research. The data can be accessed programmatically (i.e. SQL queries) from Stanford's secure Big Data Computing platforms, Nero Cloud and its on-premise counterpart Carina, or analyzed using cohort tools (e.g., OHDSI ATLAS, ACE).



The STARR-OMOP datasets resides on BigQuery, Google's fully managed cloud data warehouse, in Research Technology owned Google Cloud Platform(GCP) projects, and users have only **read access** to the PHI scrubbed STARR-OMOP data via a Virtual Private Cloud(VPC) bridge from their own Nero GCP project. This setup allows our data and compute resources to remain secure, while providing researchers with a high-performance computing platform that can scale.

The typical workflow for a SoM researcher is to request a Nero GCP project with access to STARR data (Radiology Images, EHR data in OMOP etc), and once that has been made available, they can start exploring the data and testing their models and hypotheses. Hospital data scientists access the relevant STARR-OMOP datasets (PHI or PHI scrubbed) from secure GCP projects that are managed by the Research Technology team.

Contact: prd@stanford.edu

Impacting patient care using real world data

STARR-OMOP has enabled cutting edge research that is of high practical value to Stanford Medicine. Two examples:

1. Providing real world evidence at the bedside through a consultation service

The Green Button[2] project at Stanford Medicine leveraged observational patient data (in the OMOP CDM) to pilot an on-demand consultation service for providers and researchers to answer questions about patient care and outcomes. For each consultation request, custom patient cohorts were created, appropriate statistical analyses designed, summarized, and shared back with the requestor. The service used the Advanced Cohort Engine (ACE)[3], a search engine that operates over STARR-OMOP, to define and retrieve patient cohorts. An IRB approved pilot study of the service[4,5] found that evidence derived from observational data can fill important gaps in clinical knowledge which directly impacted patient care. The service was commercially launched by Atropos Health in 2020 and now serves a number of health systems.

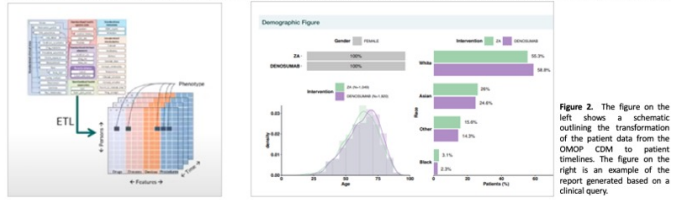


Figure 2. The figure on the left shows a schematic outlining the transformation of the patient data from the OMOP CDM to patient timelines. The figure on the right is an example of the report generated based on a clinical query.

2. Developing a Clinical Decision Support Tool to identify Risks and Care Gaps in Primary Care

The SHC AI Applied Research Team (HEART) & Codex Health have developed an AI risk-prediction tool trained on a STARR-OMOP dataset of over 70,000 primary care patients spanning seven years (2015-2022), to identify patients at high risk for ED visits and hospitalizations[6]. The tool ingests OMOP data to calculate risk metrics and displays patient risk scores with temporal evolution as well as clinical recommendations based on identified care gaps. A pilot is underway to better understand the feasibility and acceptability of the tool among clinical teams in primary care.

Assessing ML guided workflows at Stanford Health Care

The primary goal of the SHC Data Science team is to design & build infrastructure to support rapid deployment of ML systems which aid health care delivery and operations and develop processes to evaluate the value of ML-guided workflows being considered for deployment in SHC. STARR-OMOP provides crucial data about patients & their outcomes to assess the ML model's potential fairness, reliability[8,9] and usefulness[10] for patients and providers. These assessments inform deployment strategies and resource allocation for ML projects at SHC, and we are actively evaluating if models built using retrospective STARR-OMOP data can be used in real time at the point of care using EPIC-FHIR integration API's.

References

1. Datta S, et al. A new paradigm for accelerating clinical data science at Stanford Medicine, arXiv:2003.10534, Mar 2020, https://arxiv.org/abs/2003.10534
2. Longhurst C, Harrington R, Shah N. A. Green Button for Using Aggregate Patient Data at the Point of Care, Health Affairs, Jul 2014
3. Callahan A et al ACE: The Advanced Cohort Engine for searching longitudinal patients like mine. Journal of the American Medical Informatics Association (March 2021)
4. Callahan A, et al. Using Aggregate Patient at the Bedside via an On-Demand Consultation Service, NEJM Catalyst (Oct 2021)
Gombar S et al. Its time to learn from patients like mine. npj Digit. Med. 2,16 (2019)
6. Lin S, Shah S, Sattler A, Smith M. Predicting Avoidable Health Care Utilization: Practical Considerations for Artificial Intelligence/Machine Learning Models in Population Health. Mayo Clin Proc. 2022 Apr;97(4):653-657. doi: 10.1016/j.mayocp.2021.11.019. PMID: 35379419.
7. Stanford Health Care appoints inaugural chief data scientist (https://med.stanford.edu/news/all-news/2021/03/stanford-shah-inaugural-chief-data-scientist.html?utm_source=tldrai)
8. Liu S, et al. Considerations in the reliability and fairness audits of predictive models for advance care planning, Front Digit Health, Sept 2022
9. Caglero D, et alA framework toidentify ethical concerns with ML guided workflowsa case study of mortality prediction to guide advance care planning, JAM Inform Assoc, Apr 2023
10. Wornow M, et al. APLUS: A Python library for usefulness simulations of machine learning models in healthcare, J Biomed Inform, Mar 2023

#OHDSISocialShowcase This Week

FRIDAY

Guidance for Communication of the OHDSI Network Study Approach with Institutional Review Boards

(Ben Martin, Mary Grace Bowring, Paul Nagy)

Guidance for Communication of the OHDSI Network Study Approach with Institutional Review Boards

PRESENTER: **Ben Martin**

INTRO

Due to layers of data security and regulatory requirements around protected health information (PHI), institutional review boards (IRB) have a routine obligation to ensure proper handling of PHI involved in studies using real-world data (RWD).

The revised regulations combined with inter-institutional coordination amongst researchers using secondary RWD present an opportunity to minimize the time between study conception and data analysis for network studies.

OBJECTIVE

To provide guidance for clear communication of federated methods and data governance considerations of OHDSI network studies to local Institutional Review Boards.

METHODS

A review of OHDSI network study documentation, educational media, and researcher experience was performed.

Results were refined to provide common language and key communication points for use with IRB personnel and when submitting study protocols to data partners' local Institutional Review Boards.

Ben Martin, PhD[1]; Mary Grace Bowring, MPH[1]; Paul Nagy, PhD[1]; [1]Johns Hopkins School of Medicine, Baltimore, MD, USA.

Standardization of language when communicating network study methods **facilitates coordination with IRBs across federated governance structures.**

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

RESULTS

Key language and communication points for participating data partners to share with local IRBs:

- The OMOP CDM has mapped patient identifiers and should be considered a limited data set.
- Shift and truncate (SANT) methods are available to conduct randomizing date shifts for each patient while preserving temporal relationship.
- Each participating network site must register the study with their local IRB, if required by institutional governance.
- The study protocol document with rationale, background, objectives, methods, data analysis plan, and measures for protection of human subjects are published publicly on the OHDSI Studies GitHub under the corresponding study folder at https://github.com/ohdsi-studies.
- Row level patient data is never shared across separate institutions participating in a network study. Results are calculated behind the firewall of each site and aggregated before sharing.
- At each data evaluation stage prior to generation of results (i.e., data diagnostics, phenotype evaluation, and study diagnostics) only aggregate evaluations are observed by the external network study lead. Patient-level data is never shared between data partners at any point during the network study process.
- Each network site study team will have to work with their local IT team to install any dependent R packages needed to execute the OHDSI network study packages. Ensure this is permitted by under local data security and governance rules.
- Check with your local governance team to ensure that no agreements need to be executed for sharing the aggregated results.

JOHNS HOPKINS SCHOOL of MEDICINE

OHDSI

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?

Dec. 12: Happy 10th Birthday to OHDSI