



April Olympians #4 / CDM & Themis Process Overview

OHDSI Community Call
April 23, 2024 • 11 am ET



Upcoming Community Calls

Date	Topic
April 23	April Olympians Update Presentation: CDM & Themis Process Overview
April 30	April Olympians Update Presentation: What We Achieved & How You Can Use It
May 7	DevCon 2024 Review
May 14	10-Minute Tutorials
May 21	Open Studies in the OHDSI Community
May 28	Collaborator Showcase Brainstorm
June 4	NO CALL – EUROPEAN SYMPOSIUM
June 11	European Symposium Review
June 18	Application of LLMs In Evidence Generation Process
June 25	Recent OHDSI Publications



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!



Congratulations to the team of **Roger Ward, Christine Mary Hallinan, David Ormiston-Smith, Christine Chidgey, and Dougie Boyle** on the publication of **The OMOP common data model in Australian primary care data: Building a quality research ready harmonised dataset** in *PLOS One*.

PLOS ONE

RESEARCH ARTICLE

The OMOP common data model in Australian primary care data: Building a quality research ready harmonised dataset

Roger Ward, Christine Mary Hallinan^{*}, David Ormiston-Smith, Christine Chidgey, Dougie Boyle

Health & Biomedical Research Information Technology Unit (HaBIC R2), Department of General Practice and Primary Care, Faculty of Medicine, Dentistry & Health Sciences, The University of Melbourne, Parkville, Victoria, Australia

^{*} hallinan@unimelb.edu.au



Abstract

Background

The use of routinely collected health data for secondary research purposes is increasingly recognised as a methodology that advances medical research, improves patient outcomes, and guides policy. This secondary data, as found in electronic medical records (EMRs), can be optimised through conversion into a uniform data structure to enable analysis alongside other comparable health metric datasets. This can be achieved with the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), which employs a standardised vocabulary to facilitate systematic analysis across various observational databases. The concept behind the OMOP-CDM is the conversion of data into a common format through the harmonisation of terminologies, vocabularies, and coding schemes within a unique repository. The OMOP model enhances research capacity through the development of shared analytic and prediction techniques; pharmacovigilance for the active surveillance of drug safety; and 'validation' analyses across multiple institutions across Australia, the United States, Europe, and the Asia Pacific. In this research, we aim to investigate the use of the open-source OMOP-CDM in the PATRON primary care data repository.

OPEN ACCESS

Citation: Ward R, Hallinan CM, Ormiston-Smith D, Chidgey C, Boyle D (2024) The OMOP common data model in Australian primary care data: Building a quality research ready harmonised dataset. *PLoS ONE* 19(4): e0301557. <https://doi.org/10.1371/journal.pone.0301557>

Editor: Dong Keon Yon, Kyung Hee University School of Medicine, REPUBLIC OF KOREA

Received: December 18, 2023

Accepted: March 15, 2024

Published: April 18, 2024

Copyright: © 2024 Ward et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and



OHDSI Shoutouts!



Congratulations to the team of **Christian Gulden, Philipp Macho, Ines Reinecke, Cosima Strantz, Hans-Ulrich Prokosch, and Romina Blasini** on the publication of **recruIT: A cloud-native clinical trial recruitment support system based on Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) and the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)** in *Computers in Biology and Medicine*.

Computers in Biology and Medicine 174 (2024) 108411



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)
Computers in Biology and Medicine
journal homepage: www.elsevier.com/locate/compbimed



recruIT: A cloud-native clinical trial recruitment support system based on Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) and the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)

Christian Gulden^{a,*}, Philipp Macho^b, Ines Reinecke^c, Cosima Strantz^a, Hans-Ulrich Prokosch^a, Romina Blasini^d

^a Friedrich-Alexander-Universität Erlangen-Nürnberg, Department of Medical Informatics, Biometrics and Epidemiology, Medical Informatics, Erlangen, Germany
^b Medical Informatics, Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany
^c Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany
^d Institute of Medical Informatics, Justus Liebig University, Giessen, Germany

ABSTRACT

Background: Clinical trials (CTs) are foundational to the advancement of evidence-based medicine and recruiting a sufficient number of participants is one of the crucial steps to their successful conduct. Yet, poor recruitment remains the most frequent reason for premature discontinuation or costly extension of clinical trials. **Methods:** We designed and implemented a novel, open-source software system to support the recruitment process in clinical trials by generating automatic recruitment recommendations. The development is guided by modern, cloud-native design principles and based on Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) as an interoperability standard with the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) being used as a source of patient data. We evaluated the usability using the system usability scale (SUS) after deploying the application for use by study personnel. **Results:** The implementation is based on the OMOP CDM as a repository of patient data that is continuously queried for possible trial candidates based on given clinical trial eligibility criteria. A web-based screening list can be used to display the candidates and email notifications about possible new trial participants can be sent automatically. All interactions between services use HL7 FHIR as the communication standard. The system can be installed using standard container technology and supports more sophisticated deployments on Kubernetes clusters. End-users (n = 19) rated the system with a SUS score of 79.9/100. **Conclusion:** We contribute a novel, open-source implementation to support the patient recruitment process in clinical trials that can be deployed using state-of-the-art technologies. According to the SUS score, the system provides good usability.



OHDSI Shoutouts!



Congratulations to the team of **Giorgio Gandaglia, Francesco Pellegrino, Asieh Golozar, Bertrand De Meulder, Thomas Abbott, Ariel Achtman, Muhammad Imran Omar, Thamir Alshammari, Carlos Areia, Alex Asimwe, Katharina Beyer, Anders Bjartell, Riccardo Campi, Philip Cornford, Thomas Falconer, Qi Feng, Mengchun Gong, Ronald Herrera, Nigel Hughes, Tim Hulsen, Adam Kinnaird, Lana Y.H. Lai, Gianluca Maresca, Nicolas Mottet, Marek Oja, Peter Prinsen, Christian Reich, Sebastiaan Remmers, Monique J. Roobol, Vasileios Sakalis, Sarah Seager, Emma J. Smith, Robert Snijder, Carl Steinbeisser, Nicolas H. Thurin, Ayman Hijazy, Kees van Bochove, Roderick C.N. Van den Bergh, Mieke Van Hemelrijck, Peter-Paul Willemse, Andrew E. Williams, Nazanin Zounemat Kermani, Susan Evans-Axelsson, Alberto Briganti, James N'Dow, on behalf of the PIONEER Consortium** on the publication of **Clinical Characterization of Patients Diagnosed with Prostate Cancer and Undergoing Conservative Management: A PIONEER Analysis Based on Big Data** in *European Urology*.

EUROPEAN UROLOGY 85 (2024) 457–465

available at www.sciencedirect.com
journal homepage: www.europeanurology.com



Clinical Characterization of Patients Diagnosed with Prostate Cancer and Undergoing Conservative Management: A PIONEER Analysis Based on Big Data

Giorgio Gandaglia^{a,b,*}, Francesco Pellegrino^b, Asieh Golozar^{c,d}, Bertrand De Meulder^e, Thomas Abbott^f, Ariel Achtman^g, Muhammad Imran Omar^{a,h}, Thamir Alshammariⁱ, Carlos Areia^j, Alex Asimwe^k, Katharina Beyer^l, Anders Bjartell^m, Riccardo Campi^{a,n,o}, Philip Cornford^p, Thomas Falconer^q, Qi Feng^f, Mengchun Gong^{r,s}, Ronald Herrera^k, Nigel Hughes^t, Tim Hulsen^u, Adam Kinnaird^v, Lana Y.H. Lai^w, Gianluca Maresca^x, Nicolas Mottet^a, Marek Oja^{y,z}, Peter Prinsen^{aa}, Christian Reich^{bb}, Sebastiaan Remmers^{cc}, Monique J. Roobol^{cc}, Vasileios Sakalis^{dd}, Sarah Seager^{ee}, Emma J. Smith^a, Robert Snijder^f, Carl Steinbeisser^k, Nicolas H. Thurin^{ff}, Ayman Hijazy^e, Kees van Bochove^{gg}, Roderick C.N. Van den Bergh^{hh}, Mieke Van Hemelrijckⁱ, Peter-Paul Willemse^{a,ii}, Andrew E. Williams^{jj}, Nazanin Zounemat Kermani^{kk}, Susan Evans-Axelsson^{k,‡}, Alberto Briganti^{a,b,‡}, James N'Dow^{a,h,‡}, on behalf of the PIONEER Consortium



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





Upcoming Workgroup Calls



Date	Time (ET)	Meeting
Wednesday	9 am	OMOP CDM Oncology Outreach/Research Subgroup
Wednesday	10 am	Surgery and Perioperative Medicine
Wednesday	12 pm	Latin America
Wednesday	1 pm	Perinatal and Reproductive Health
Wednesday	3 pm	Joint Vulcan/OHDSI Meeting
Thursday	9:30 am	Network Data Quality
Thursday	12 pm	Medical Devices
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Clinical Trials
Monday	10 am	Africa Chapter



DevCon 2024: April 26, 9 am-3 pm ET

Morning Agenda

9:00 am – Introduction (Adam Black, Paul Nagy)

9:15 am – Developers Panel and Lightning Talks (Katy Sadowski)

- *OHDSI/OMOP – The hard way is the easy way!* (Vishnu V Chandrabalan)
- *Moving OMOP to the Cloud With DBT and Snowflake* (Roger Carlson)
- *Use cases for ORMs in OMOP* (Georgina Kennedy)
- *Carrot: code-free OMOP ETL without full data access* (Sam Cox)
- *Rabbit-in-a-blender - an ETL pipeline to transform your EMR data into OMOP* (Pieter-jan Lammertyn)

10:45 am – Darwin EU[®] Developers Update (Adam Black)

12:00 pm – Break

Afternoon Agenda

12:30 pm – OHDSI Ecosystem Updates

- TAB Update (Frank DeFalco)
- Strategus Update (Anthony Sena)
- Broadsea Update (Lee Evans)
- Kheiron Updates (Paul Nagy)

1:15 pm – JACKALOPE PLUS The Power of ML for Healthcare Data Mapping & Management (Denys Kaduk)

2:00 pm - An Introduction to Knowledge Graphs using PheKnowLator and OMOP2OBO with Example Applications in Drug Surveillance and Computational Phenotyping (Tiffany Callahan)



#OHDSI2024 Registration Is Open!

Registration is now OPEN for the 2024 OHDSI Global Symposium, which will be held Oct. 22-24 at the Hyatt Regency Hotel in New Brunswick, N.J., USA.

Tuesday: Tutorials

Wednesday: Plenary/Showcase

Thursday: Workgroup Activities

ohdsi.org/OHDSI2024





#OHDSI2024 Collaborator Showcase

Submissions are now being accepted for the 2024 Global Symposium Collaborator Showcase.

All submissions are due by 8 pm ET on Friday, June 21.

Notification of acceptance will be made by Tuesday, Aug. 20.



ohdsi.org/OHDSI2024



Maternal Health Data Science Fellowship

This program is designed to empower clinical investigators to leverage emerging technologies for improved maternal and neonatal care while reducing morbidity and mortality.

Three main components of this program:

1) Career Development (create evidence, leverage data models, build skills on network studies)

2) Practice (design effective observational research protocols, master tools, write papers/grants)

3) Networking (build relationships with mentors, learners, coordinate with global OHDSI collaborators)

Application deadline: May 15

Want to build your career?

Generate reproducible evidence by leading multi-institutional studies!

Learn more & apply!





RWE Workshop at AIME24: Call for Submissions!

Workshop: AI for Reliable and Equitable Real-World Evidence Generation in Medicine

<https://medicine.utah.edu/dbmi/aime/ai-reliable>

Organizing Committee

Linying Zhang
Adam Wilcox
Yves Lussier

Scientific Program Committee

Peter Rijnbeek Mattia Prosperi
Larry Han Xia Ning
Xiaoqian Jiang Yifan Peng

Opening Keynote

George Hripcsak

IMPORTANT DATES

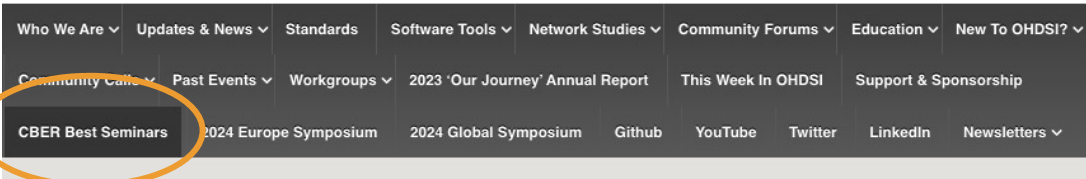
May 31, 2024 | Submission Deadline

June 14, 2024 | Notice of Acceptance

July 12, 2024 | Workshop

AIME 2024
22nd International Conference on Artificial Intelligence in Medicine
Salt Lake City, Utah, USA, July 9-12
Hosted by the University of Utah

CBER BEST Seminar Recording Is Posted



CBER BEST Seminar Series

The [CBER BEST Initiative](#) Seminar Series is designed to share and discuss recent research of relevance to ongoing and future surveillance activities of CBER regulated products, namely biologics. The series focuses on safety and effectiveness of biologics including vaccines, blood components, blood-derived products, tissues and advanced therapies. The seminars will provide information on characteristics of biologics, required infrastructure, study designs, and analytic methods utilized for pharmacovigilance and pharmacoepidemiologic studies of biologics. They will also cover information regarding potential data sources, informatics challenges and requirements, utilization of real-world data and evidence, and risk-benefit analysis for biologic products. The length of each session may vary, and the presenters will be invited from outside FDA.

Below you will find details of upcoming CBER BEST seminars, including virtual links that will be open to anybody who wishes to attend. Speakers who give their consent to be recorded will also have their presentations included on this page; you can find those sessions below the list of upcoming speakers.





— April 17: Yong Chen, University of Pennsylvania

Topic: Real-World Effectiveness of BNT162b2 Against Infection and Severe Diseases in Children and Adolescents: causal inference under misclassification in treatment status

Presenter: Dr. Yong Chen, Professor & Director of the Center for Health AI and Synthesis of Evidence (CHASE) at the University of Pennsylvania

Description: The current understanding of long-term effectiveness of the BNT162b2 vaccine across diverse U.S. pediatric populations is limited. We assessed the effectiveness of BNT162b2 against various strains of the SARS-CoV-2 virus using data from a national collaboration of pediatric health systems (PDSnet). We emulated three target trials to assess the real-world effectiveness of BNT162b2 during the Delta and Omicron variant periods. In the U.S., immunization records are often captured and stored across multiple disconnected sources, resulting in incomplete vaccination records in patients' electronic health records (EHR). We implemented a novel trial emulation pipeline accounting for possible misclassification bias in vaccine documentation in EHRs. The effectiveness of the BNT162b2 vaccine was estimated from the Poisson regression model with confounders balanced via propensity score stratification. This study suggests BNT162b2 was effective among children and adolescents in Delta and Omicron periods for a range of COVID-19-related outcomes and is associated with a lower risk for cardiac complications.

Bio: Dr. Yong Chen is tenured Professor of Biostatistics and the Founding Director of the Center for Health AI and Synthesis of Evidence (CHASE) at the University of Pennsylvania. He is an elected fellow of American Statistical Association, International Statistical Institute, Society for Research Synthesis Methodology, American College of Medical Informatics, and American Medical Informatics Association. He founded the Penn Computing, Inference and Learning (PennCIL) lab at the University of Pennsylvania, focusing on clinical evidence generation and evidence synthesis using clinical and real-world data. During pandemic, Dr. Chen is serving as biostatistics core director for a national multi-center study on Post-Acute Sequelae of SARS CoV-2 infection (PASC), involving more than 9 million pediatric patients across 40 health systems.



ohdsi.org/cber-best-seminar-series



#OHDSISocialShowcase This Week

MONDAY

Enhancing Data Quality Management: Introducing Capture and Cleanse Modes to the Data Quality Dashboard

(Frank DeFalco, Clair Blacketer)

Enhancing Data Quality Management

Introducing Capture and Cleanse Modes to the Data Quality Dashboard

PRESENTER: Frank J DeFalco

INTRO

- The DataQualityDashboard package was updated with new features to support capture and cleanse. The implementation of the capture and cleanse feature has demonstrated promising results in improving data quality management.

METHODS

- New run modes named 'capture' and 'cleanse' were added to the Data Quality Dashboard package
- New elements were added to the check description data to codify logic for these new run modes
- All data quality checks are identified as eligible for capture and cleanse through the existing check description functionality.

ENHANCED VISIBILITY

- By capturing all failing data quality data owners can review the data quality failures in a dedicated schema.

PROACTIVE DATA CLEANSING

- Eliminating records that failed data quality checks allows organizations to improve the reliability of downstream analytics

Capture data quality issues and Cleanse them from your data to ensure your organization uses Research Quality Data.

```
DataQualityDashboard::executeDqChecks (
    runMode = "capture"
)
```



Take a picture to download the full paper

CAPTURE MODE

- Provides the ability to identify data records that fail specific data quality checks and captures copies of the affected records to a user-specified schema.
- With capture mode, organizations can preserve and characterize the failing records, gaining valuable insights for further analysis and investigation of their data quality issues

CLEANSE MODE

- This mode provides the ability to automatically remove failing records from a data source.
- By leveraging the cleanse mode, organizations can maintain a cleaner and more reliable dataset by eliminating records that fail data quality checks, ensuring data integrity and accuracy.
- A systematic approach to data cleansing provides a reproducible way to eliminate failing records as part of a data operations pipeline.

Frank DeFalco, Clair Blacketer





#OHDSISocialShowcase This Week

TUESDAY

Making OMOP Happen: An Implementation Science Approach

(Maya Younoszai, Pam Dasher, Danielle Boyce, Smith Heavner)

Making OMOP Happen An Implementation Science Approach

PRESENTER: Maya Younoszai

INTRO:

- Multicenter projects bring together sites at all stages of their OHDSI journey
- OHDSI projects are complicated and require the coordination of resources internal and external to the participating sites.
- The EPIS implementation science framework allows us to understand the successes and challenges of the CURE ID project

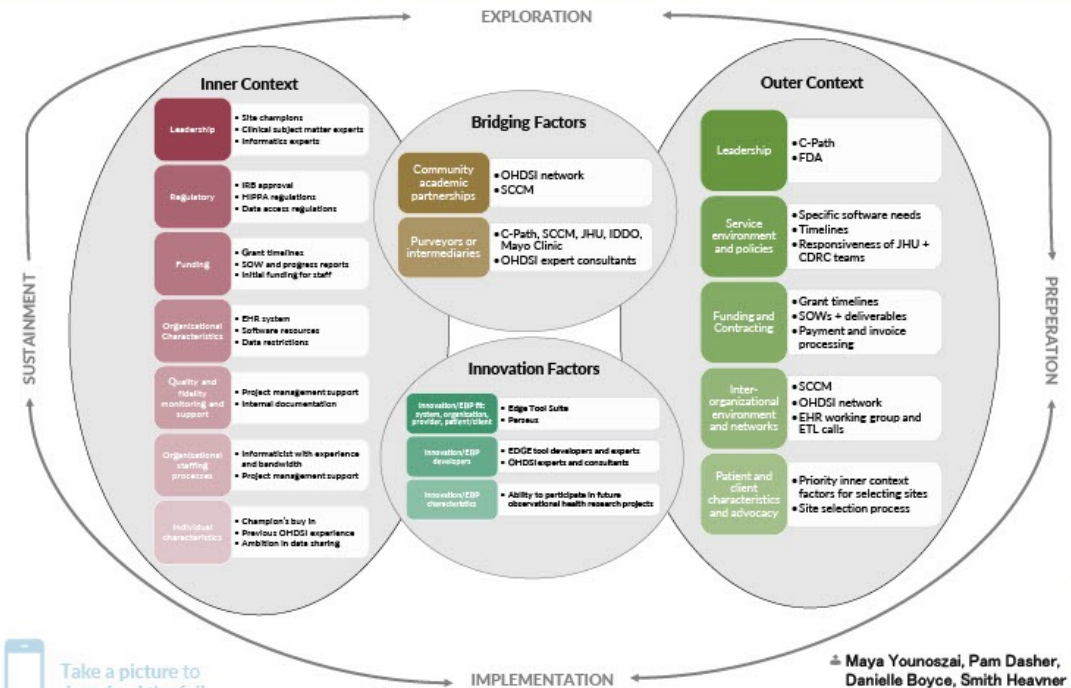
METHODS

1. Reviewed calls and emails from hospital sites to understand major successes and roadblocks
2. Identified recurring themes with multiple
3. Performed member checking to confirm results with a variety of stakeholders involved in many steps of the project

RESULTS

- Inner context (specifically technological and compliance factors) is a major predictor of site success
- Strong bridging factors, including OHDSI community involvement, helped support sites through setbacks
- Highlighting and emphasizing innovation factors led to faster connections and interest to higher-level leadership

Having a strategic plan for implementation and utilizing tools like the EPIS framework from the outset of projects can improve efficiency, reduce redundancy, and expedite problem solving.



Take a picture to download the full paper

Maya Younoszai, Pam Dasher, Danielle Boyce, Smith Heavner





#OHDSISocialShowcase This Week

WEDNESDAY

Evaluation of Study Execution using Large-Scale Analytics: A Machine Learning Approach to Assess Pre-Exposure Prophylaxis (PrEP) Utilization in the Real-World

(Nag Mani, Xiwen Huang, Li Tao, Hu Li)

Evaluation of Study Execution using Large-Scale Analytics: A Machine Learning Approach to Assess Pre-Exposure Prophylaxis (PrEP) Utilization in the Real-World

Nag Mani, Xiwen Huang, Li Tao, Hu Li
Gilead Sciences, Inc., Foster City, California, US

Key Findings

- This study demonstrates the value of using a large-scale analytics (LSA) platform with Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) data to efficiently identify similarities and differences across multiple real-world data sources.
- Our analysis of cohort characteristics, model performance, and top predictors for pre-exposure prophylaxis (PrEP) use emphasizes the need to study diverse data sources to understand data strengths and limitations.
- This approach generates reliable real-world evidence (RWE), which will ultimately drive improvements in PrEP use in the real world.

Conclusions

- This analysis identified similarities and differences of demographics in target and outcome cohorts across data sources, namely variations in geographic location, age, and sex at birth.
- Machine learning methodology applied in the LSA showed that PrEP prevalence in both target cohorts across the three data sources ranged from 0.15% to 1.6%, highlighting the limited adoption of PrEP among populations vulnerable to HIV infection.
- Sex at birth and exposure to HIV were two of the top predictors for PrEP use observed across data sources.
- This multi-dataset analysis generates reliable RWE, furthermore, standardization of the OMOP CDM ensures findings are scalable, aiding application of effective PrEP implementation strategies in key populations.

Future Work

- To use RWE and machine learning models to identify potential individuals who could benefit from PrEP via HIV access point needs.
- To improve model performance through application of ensemble techniques such as stacking and boosting, e.g., XGBoost.
- To leverage the LSA to expand these findings to new data sources from the US and Europe.

Introduction

- RWE studies aim to generate reliable evidence using individual-level observational data¹.
- LSA platforms that use machine learning methodology are increasingly being used to analyze the similarities and commonalities across different sources of real world data. This can help address challenges with multi-dataset analysis and development of accurate and reliable evidence.
- This approach has been applied to identify populations who could benefit from PrEP in real-world settings in the US.
- Despite the proven effectiveness of PrEP, its adoption remains limited among key populations vulnerable to HIV². Identifying these key populations using real-world data continues to pose challenges.

Objective

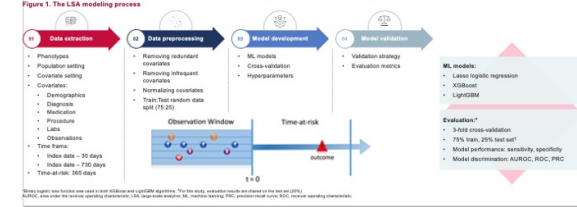
- To provide comprehensive characterization of RWE on the prevalence of HIV PrEP use and the predictors of PrEP use via an LSA platform and machine learning approach, using three data sources.

Methods

- Data sources/cohort descriptions**
 - Individual-level data were obtained from three data sources in OMOP CDM version 5.3 between 1 January 2018 and 30 June 2022:
 - IQVIA Pharmetrics Plus Claims (Claims)
 - IQVIA Ambulatory EMR (EMR)
 - HealthVerity (HV)
 - The target cohorts were:
 - 1. Sexually transmitted disease (STD)³
 - 2. Other key population⁴
 - The outcome cohort was defined as PrEP users (identified for each target cohort for each data source).
 - All individuals aged ≥18 and older on the last occurrence (index date) of the condition outlined for the target cohorts were included, and those with history of HIV, immunosuppression, pregnancy, hepatitis B infection, chronic hepatitis B, or prior diagnosis of HIV were excluded.
 - The modeling process (Figure 1) was developed using the Patient Level Prediction package⁵.
 - The top predictors of PrEP use were analyzed (as identified by the best predictive model for each data source). The sensitivity, specificity, and discrimination of the model were assessed.

¹Observational research: methods, practice, and reporting. *BMJ*. 2015;350:g759. doi:10.1136/bmj.g759. ²World Health Organization. *Global Health Observatory (GHO) Data*. Available from: <https://data.who.int/dashboards/gov>. ³Centers for Disease Control and Prevention. *Sexually Transmitted Infections Treatment Guidelines - 2021*. Atlanta, GA: US Department of Health and Human Services; 2021. ⁴Centers for Disease Control and Prevention. *Sexually Transmitted Infections Treatment Guidelines - 2021*. Atlanta, GA: US Department of Health and Human Services; 2021. ⁵Centers for Disease Control and Prevention. *Sexually Transmitted Infections Treatment Guidelines - 2021*. Atlanta, GA: US Department of Health and Human Services; 2021.

Methods (continued)



Results

- Demographics**
 - Characteristics reporting of cohort characteristics showed similarities and differences of population distribution in target and outcome cohorts across data sources, for example:
 - Geographic variations: most of the IQVIA Claims dataset population were concentrated in Florida and Texas, while in the IQVIA EMR and the HV datasets most individuals were from California.
 - Age variations:
 - In the other key population target cohort and in the outcome cohort, age distribution was similar across datasets, with most of the population aged <40 years, and with a declining trend in age.
 - In the STD target cohort, most individuals in the IQVIA Claims and IQVIA EMR datasets were aged <40 years but fewer were aged >60 years compared with the HV dataset.
 - In the other key population target cohort, more individuals in the IQVIA EMR dataset were aged <18 years (15%) compared with the IQVIA Claims and HV datasets (1%).
 - Sex at birth variations:
 - In the other key population cohort, sex at birth was evenly distributed in the IQVIA Claims dataset, however, the IQVIA EMR and HV datasets reported more females than males.
 - In the STD target cohort, across datasets most individuals were reported female (70%).
 - In the outcome cohort, across datasets most individuals were reported male (90-95%).
- Prevalence of PrEP**
 - Lasso logistic regression, XGBoost, and LightGBM models used to predict PrEP use demonstrated that the prevalence of PrEP in both cohorts across the three data sources ranged from 0.15% to 1.6% (Table 1).
 - In the other key population cohort:
 - XGBoost performed the best in the IQVIA Claims and HV datasets.
 - The Lasso logistic regression model outperformed more complex boosting methods in the IQVIA EMR dataset.
 - In the STD cohort:
 - XGBoost demonstrated the best performance across data sources.

Table 1. Model results for each target cohort for each data source: Claims, EMR, and HV

Cohort	Individuals	Prevalence of PrEP	Model	AUROC (95% CI)		AUPRC	Sensitivity (%)		Specificity (%)	
				Claims	EMR		Claims	EMR	Claims	EMR
Other key population	11707	1.66	Claims	Lasso Logistic Regression	0.95 (0.94, 0.96)	0.36	97.6	79.2	97.7	71.2
			EMR	Lasso Logistic Regression	0.96 (0.95, 0.97)	0.42	97.7	71.2	97.7	71.2
			HV	XGBoost	0.97 (0.96, 0.97)	0.44	98.0	79.2	98.0	79.2
	43559	0.65	Claims	Lasso Logistic Regression	0.89 (0.88, 0.90)	0.11	92.4	66.5	92.4	66.5
			EMR	Lasso Logistic Regression	0.82 (0.79, 0.85)	0.02	92.4	47.6	92.4	47.6
			HV	XGBoost	0.87 (0.86, 0.90)	0.07	95.5	61.4	95.5	61.4
72803	1.64	Claims	Lasso Logistic Regression	0.89 (0.87, 0.90)	0.11	93.6	65.0	93.6	65.0	
		EMR	Lasso Logistic Regression	0.88 (0.88, 0.91)	0.10	93.0	62.9	93.0	62.9	
		HV	XGBoost	0.89 (0.88, 0.91)	0.13	93.3	79.9	93.3	79.9	
92303	0.16	Claims	Lasso Logistic Regression	0.92 (0.91, 0.94)	0.47	96.1	66.6	96.1	66.6	
		EMR	Lasso Logistic Regression	0.92 (0.91, 0.94)	0.43	96.7	67.7	96.7	67.7	
		HV	XGBoost	0.89 (0.82, 0.94)	0.5	96.1	79.9	96.1	79.9	
35448	1.20	Claims	Lasso Logistic Regression	0.92 (0.91, 0.94)	0.23	94.9	72.0	94.9	72.0	
		EMR	Lasso Logistic Regression	0.86 (0.85, 0.87)	0.08	93.5	62.7	93.5	62.7	
		HV	XGBoost	0.89 (0.88, 0.90)	0.23	94.5	69.4	94.5	69.4	
173659	0.51	Claims	Lasso Logistic Regression	0.91 (0.89, 0.93)	0.09	92.8	73.8	92.8	73.8	
		EMR	Lasso Logistic Regression	0.90 (0.88, 0.92)	0.07	92.3	70.4	92.3	70.4	
		HV	XGBoost	0.82 (0.81, 0.84)	0.18	94.4	71.4	94.4	71.4	

Best results are bolded on the best performing model for the data source. The model with the highest AUROC is highlighted in green. The model with the highest AUPRC is highlighted in blue. The model with the highest sensitivity is highlighted in red. The model with the highest specificity is highlighted in purple. The model with the highest AUC is highlighted in yellow. The model with the highest AUC is highlighted in yellow. The model with the highest AUC is highlighted in yellow.

References: 1. *Stat Sci*. 2021;36(1):1-16. 2. *Mayo Clin Proc*. 2022;97(1):1-11. 3. *Centers for Disease Control and Prevention*. PrEP for HIV Prevention in the US. <https://www.cdc.gov/hiv/prep/>. 4. *Journal of the American Medical Association*. 2018;320(19):1969-1976. 5. *Chen T, Guehri C*. (2018). In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 788-794). New York, NY, USA: ACM. 6. *Kie G, et al*. *AIDS Research and Human Immunology*. 2017;30:446-454.

Acknowledgments: Medical writing support was provided by Anfield MedComms (Macclesfield, UK), an Inizio company, and funded by Gilead Sciences, Inc. Disclosures: NM, XH, LT and HL are all employees and shareholders of Gilead Sciences, Inc. Correspondence: nagmani@gilead.com



Results (continued)

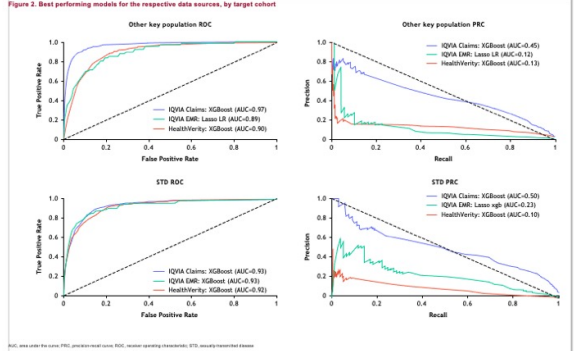


Table 2. Top predictors across data sources for the other key population cohort

Claims (XGBoost)	EMR (Logistic)	Marketplace (XGBoost)
1. Exposure to HIV	1. Exposure to HIV	1. Sex at birth
2. Pharmacy visit	2. Male heterosexual behavior	2. Exposure to HIV
3. Sex at birth	3. SARS-CoV-2 (COVID-19) vaccine	3. Male heterosexual behavior
4. Hepatitis B surface antigen test	4. Antiretroviral therapy screening	4. Exposure to potentially hazardous substance
5. Exposure to STD	5. COVID test	5. Outpatient visit count in past 2 years
6. Number of conditions in past 2 years	6. Exposure to viral hepatitis	6. Exposure to viral hepatitis
7. Observation time (days)	7. Creatinine blood test	7. Finding missing to missing and sexual activity
8. Creatinine blood test	8. Blood pressure vital	8. Other viral vaccines
9. Viral vaccines	9. Hepatitis B surface antigen test	9. Hepatitis B surface antigen test
10. Outpatient visit count in last 30 days	10. Basic metabolic panel test	10. Number of conditions in past 2 years

Table 3. Top predictors across data sources for the STD cohort

IQVIA Claims (XGBoost)	IQVIA EMR (XGBoost)	Marketplace (XGBoost)
1. Sex at birth	1. Sex at birth	1. Sex at birth
2. Pharmacy visit	2. High-risk sexual behavior	2. Male heterosexual behavior
3. Bacterial infectious disease	3. Rapid plasma reagin test	3. Exposure to HIV
4. Antibody test for Hepatitis B surface antigen	4. Number of observations in past 30 days	4. Outpatient visit count in past 2 years
5. Observation time (days)	5. Observation time (days)	5. Observation time (days)
6. Hepatitis B surface antigen test	6. HIV seronegativity, combination assay, screening test	6. Chlamydia infection, Pharyngeal ulceration
7. Hepatitis C antibody test	7. Outpatient visit count in past 30 days	7. Outpatient visit count in past 30 days
8. Exposure to HIV	8. Viral disease of HIV	8. Infectious disease of liver
9. Number of drugs in past 2 years	9. Number of drugs in past 2 years	9. Number of drugs in past 2 years
10. Number of visits in past 2 years	10. Hepatitis B surface antigen measurement test	10. Number of observations in past 2 years



#OHDSISocialShowcase This Week

THURSDAY

Validation and Comparison of Frailty Indexes: An OHDSI Network Study

(Chen Yanover, Louisa Smith, Tal El-Hay, Brianne Olivieri-Mui, Maytal Bivas-Benita, Robert Cavanaugh, Pinchas Akiva, Chelsea N. Wong, Ariela Orkaby)

Title: Validation and Comparison of Frailty Indexes: An OHDSI Network Study

INTRO

A frailty index (FI) is a marker of overall health status and vulnerability, used to identify those at increased risk for adverse health outcomes; typically, a sum of health indicators ("deficits") across diverse health domains. We aimed to validate and compare electronic health record (EHR)-based FIs across multiple health care settings and geographies.

METHODS

- **Study design:** A multinational cohort study using routinely collected healthcare data from 5 OMOPed DBs
- **Study population:** Individuals ≥ 40 years old, with ≥ 1 year of observation prior to an index date – a random visit for UK data sources and PharMetrics+; and 1 year following recruitment date in the AoU data.
- **EHR-based FIs:** UK electronic Frailty Index (eFI) and US Veterans Affairs Frailty Index (VA-FI), computed on 1Y lookback period.



Frailty Status Shows Similar Trends across Healthcare Systems, but Different Prevalence

• Data sources

IQVIA™ Adjudicated Health Plan Claims	PharMetrics+	<input checked="" type="checkbox"/>	
All of Us Research Program	AoU	<input checked="" type="checkbox"/>	
IQVIA™ Medical Research Data – EMIS	IMRD-EMIS	<input type="checkbox"/>	
IQVIA™ Medical Research Data – UK	IMRD-UK	<input type="checkbox"/>	
The UK Biobank	UKBB	<input checked="" type="checkbox"/>	

Geography USA; UK
 Data type Admin claims; EHRs; EHRs + Questionnaires
 Included visits Outpatient; Inpatient



Poster, abstract



Code

CONCLUSIONS

- ✓ Expected FI and deficit trends (e.g., 1age, osteoporosis F > M)
- ✓ Substantial differences in frailty prevalence between USA, UK

LIMITATIONS

- FI code lists (originally, Read, ICD) may be incomplete
- Potential differences in coding, reporting within the various healthcare systems

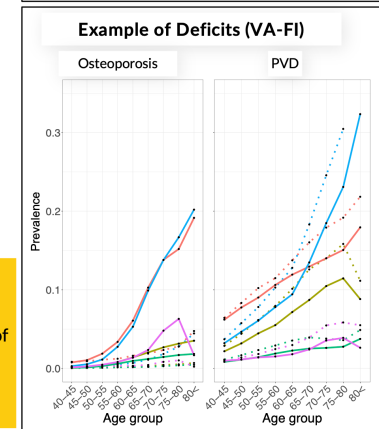
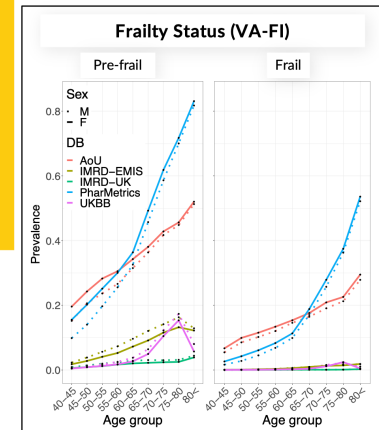
Chen Yanover¹, Louisa Smith^{2,3}, Tal El-Hay¹, Brianne Olivieri-Mui^{2,3,4}, Maytal Bivas-Benita¹, Robert Cavanaugh³, Pinchas Akiva¹, Chelsea N. Wong^{3,4}, Ariela Orkaby^{5,6,7}

¹KI Research Institute, Kfar Malal, Israel; ²The Roux Institute, Northeastern University; ³Department of Health Sciences, Northeastern University; ⁴The Marcus Institute for Aging Research, Hebrew SeniorLife, Harvard Medical School; ⁵Veteran Affairs Boston Healthcare System, Boston, Mass; ⁶Department of Medicine/Division of Aging, Brigham and Women's Hospital, Boston, Mass; ⁷Harvard Medical School, Boston, Mass.

RESULTS

Characteristics of the study populations

	N	% Female	40-75y	>75y
PharMetrics+	5,292,854	53.6%	64.7%	35.3%
AoU	189,746	60.6%	87.5%	12.2%
IMRD-EMIS	1,103,278	50%	73.3%	26.7%
IMRD-UK	3,051,179	50.7%	75.9%	24.1%
UKBB	470,226	53.8%	98.0%	2.0%





#OHDSISocialShowcase This Week

FRIDAY

Broadsea 3.0: "BROADening the ohdSEA"

(**Ajit Londhe**, Lee Evans, Sanjay Udoshi)



OHDSI Broadsea Evolution

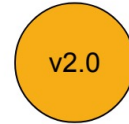
"Broadsea is the easiest way to install (& upgrade) the OHDSI tools"



v1.0

Atlas/WebAPI & RStudio Docker images on Mac/Linux/Windows

10k+ Downloads



v2.0

Pre-populated demo postgres database image & Traefik reverse proxy



v3.0

Docker profiles for a-la-carte services, more Traefik networking, environment variable driven deployment, new OHDSI apps, build from Git

```
docker-compose --profile default up -d
```

<https://github.com/OHDSI/Broadsea>



Opening: Research Assistant, University of Oxford



UK date and time: 23-April-2024 15:25

Applicant Options

- › [New Search](#)
- › [Login](#)
- › [Job Details](#)
- › [Help](#)
- › [Terms of Use & Privacy Policy](#)

Job Details

Research Assistant in Health Data Sciences

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

We have an exciting opportunity for a Research Assistant in Health Data Sciences to join the Pharmaco- and Device epidemiology research group led by Professor Daniel Prieto-Alhambra at the Botnar Research Centre, NDORMS, University of Oxford. The NDORMS Pharmaco- and Device epidemiology research group is involved in a number of national and international studies exploring the conditions of use (adherence, compliance, off and on-label use) of a number of licensed drugs, devices, and vaccines for the prevention and treatment of human disease in 'real world' (routine practice) conditions.

As a Research Assistant in Health Data Sciences you will contribute to the programming of analytical pipelines for the analysis of routinely collected data mapped to the OMOP Common Data Model. You will analyse real world data to address regulatory questions related to the prevalence/incidence of disease, use of medicines/vaccines, and the risks or benefits of medicines/vaccines or devices. You will prepare analytical packages to run a number of pre-specified analyses, contribute to wider project planning, including ideas for new research projects and gather, analyse, and present scientific data from a variety of sources.

You will hold a relevant BA or MSc degree in Mathematics, Engineering, or a related field. Knowledge of medical statistics and experience analysing large datasets, experience in biostatistics and/or health data sciences and experience in the programming of R packages are essential. Experience in propensity scores, overlap weighting, inverse probability weighting and/or similar methods, expertise in pharmaco or vaccine epidemiology and experience of working with electronic medical records/routinely collected data are desirable.

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

The closing date for this position is 12 noon on 10 May 2024. You will be required to upload a CV and supporting statement as part of your online application.

Contact Person : HR Team, NDORMS
Contact Phone :
Pay Scale : STANDARD GRADE 6
Salary (£) : £32,332 - £38,205 p.a

Vacancy ID : 172348
Closing Date & Time : 10-May-2024 12:00
Contact Email : hr@ndorms.ox.ac.uk



Opening: Biomedical Informatics Data Scientist at Stanford



[Who We Are](#) ▾

[What We Offer](#) ▾

[How We Hire](#) ▾

[Career Areas](#) ▾

[Search](#) .

Biomedical Informatics Data Scientist

1.0 FTE • Full time • Day - 08 Hour • R2335119 • Hybrid • 84866 IT RESEARCH • Technology & Digital Solutions • 455 Broadway, REDWOOD CITY, California •

1.0 FTE Full time Day - 08 Hour R2335119 Hybrid 84866 IT RESEARCH Technology & Digital Solutions 455 Broadway, REDWOOD CITY, California

If you're ready to be part of our legacy of hope and innovation, we encourage you to take the first step and explore our current job openings. Your best is waiting to be discovered.

Day - 08 Hour (United States of America)

This is a [Stanford Health Care](#) job.

A Brief Overview

The Biomedical Informatics Data Scientist will partner with researchers and clinicians to enable effective and efficient use of data and resources available via Stanford's research clinical data repository (STARR) including the Electronic Health Records in the OMOP Common Data Model, radiology and cardiology imaging data and associated metadata, and new data types as they get integrated along with their databases and respective cohort query tools and interfaces e.g., OHDSI ATLAS. This individual will enable researchers to maximize their understanding, interpretation and use of these clinical and research tools for more informed and productive research, clinical trials, patient care and quality outcome projects.

Clean, extract, transform and analyze various kinds of clinical data to create analysis-ready datasets that follow the FAIR (Findable, Accessible, Interoperable and Re-usable) principles. Partner with researchers and clinicians to enable effective and efficient use of Stanford Clinical data and resources for the advancement of research and the educational mission.

Postdoc/Senior Data Analyst Opening at WashU

The Zhang Lab at Washington University School of Medicine in St. Louis has **one postdoc/senior data analyst position** to work on **causal machine learning** and **responsible AI** for reliable real-world evidence generation.



PI: Linying Zhang, PhD

- More details at <https://linyingzhang.com>
 - Postdoc:
<https://linyingzhang.com/files/Postdoc.pdf>
 - Data analyst:
<https://linyingzhang.com/files/Analyst.pdf>
- If interested, please send CV and cover letter to linyingz@wustl.edu





Director, RWE at Gilead

About Us



Director, RWE - Data Science - OHDSI

Apply

Gilead Sciences, Inc. is a biopharmaceutical company that has pursued and achieved breakthroughs in medicine for more than three decades, with the goal of creating a healthier world for all people. The company is committed to advancing innovative medicines to prevent and treat life-threatening diseases, including HIV, viral hepatitis and cancer. Gilead operates in more than 35 countries worldwide, with headquarters in Foster City, California.

Responsibilities:

Collaborate with researchers and data scientists to understand project requirements and translate them into OHDSI-compatible solutions. Work with databases, ensuring data integrity and optimization for OHDSI-related queries and analyses. Perform data analyses in OHDSI-related tools like ATLAS. Customize and extend OHDSI tools and applications to meet specific project needs. Collaborate with cross-functional teams to troubleshoot and resolve technical issues related to OHDSI implementations. Stay informed about OHDSI community updates, best practices, and emerging trends in observational health data research. Contribute to the development and documentation of data standards and conventions within the OHDSI community.



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?





April 23: CDM & Themis Process Overview



Clair Blacketer

Director, Observational Health Data Analytics
Janssen Research & Development



Melanie Philofsky

Senior Business Analyst and Project Manager
Odysseus Data Services, Inc.



The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

Links are sent out weekly and available at:
ohdsi.org/community-calls