Welcome to OHDSI

OHDSI Community Call
Oct. 24, 2023 • 11 am ET

Upcoming Community Calls

Date	Topic
Oct. 24	Welcome to OHDSI
Oct. 31	Welcome to OHDSI, part 2
Nov. 7	Meet The Titans
Nov. 14	Collaborator Showcase Honorees
Nov. 21	Showcase Software Demos
Nov. 28	TBA
Dec. 5	Recent Publications
Dec. 12	How Did OHDSI Do This Year?
Dec. 19	Holiday-Themed Goodbye to 2023!

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Anna Ostropolets, George Hripcsak, Syed A Husain, Lauren R Richter, Matthew Spotnitz, Ahmed Elhussein, and Patrick Ryan** on the publication of **Scalable and interpretable alternative to chart review for phenotype evaluation using standardized structured data from electronic health records** in *JAMIA*.

Journal of the American Medical Informatics Association, 2023, 1–11
https://doi.org/10.1093/jamia/ocad202
Research and Applications

Research and Applications

Scalable and interpretable alternative to chart review for phenotype evaluation using standardized structured data from electronic health records

Anna Ostropolets, MD, PhD*[,1], George Hripcsak, MD, MS[1,2], Syed A. Husain, MD, MPH[3], Lauren R. Richter, MD, MS[1], Matthew Spotnitz, MD, MPH[1], Ahmed Elhussein, MD, MS[1], Patrick B. Ryan, PhD[1,4]

[1]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY 10032, United States, [2]Medical Informatics Services, New York-Presbyterian Hospital, New York, NY 10032, United States, [3]Division of Nephrology, Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY 10032, United States, [4]Observational Health Data Analytics, Janssen Research and Development, Titusville, NJ 08560, United States

*Corresponding author: Anna Ostropolets, MD, PhD, Columbia University Medical Center, 622 West 168th Street, PH-20, New York, NY 10032, USA (ao2671@cumc.columbia.edu)

Abstract

Objectives: Chart review as the current gold standard for phenotype evaluation cannot support observational research on electronic health records and claims data sources at scale. We aimed to evaluate the ability of structured data to support efficient and interpretable phenotype evaluation as an alternative to chart review.

Materials and Methods: We developed Knowledge-Enhanced Electronic Profile Review (KEEPER) as a phenotype evaluation tool that extracts patient's structured data elements relevant to a phenotype and presents them in a standardized fashion following clinical reasoning principles. We evaluated its performance (interrater agreement, intermethod agreement, accuracy, and review time) compared to manual chart review for 4 conditions using randomized 2-period, 2-sequence crossover design.

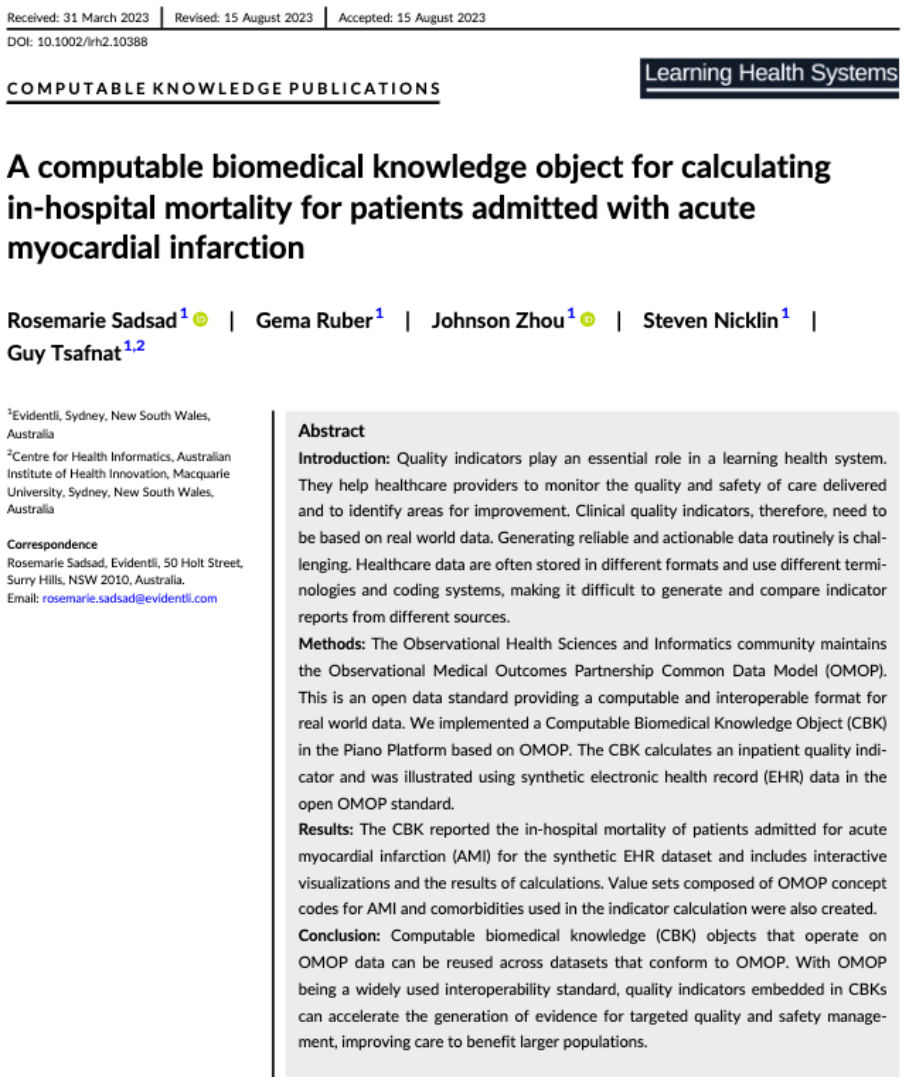
Results: Case ascertainment with KEEPER was twice as fast compared to manual chart review. 88.1% of the patients were classified concordantly using charts and KEEPER, but agreement varied depending on the condition. Missing data and differences in interpretation accounted for most of the discrepancies. Pairs of clinicians agreed in case ascertainment in 91.2% of the cases when using KEEPER compared to 76.3% when using charts. Patient classification aligned with the gold standard in 88.1% and 86.9% of the cases respectively.

Conclusion: Structured data can be used for efficient and interpretable phenotype evaluation if they are limited to relevant subset and organized according to the clinical reasoning principles. A system that implements these principles can achieve noninferior performance compared to chart review at a fraction of time.

Key words: chart review; phenotyping; observational studies; case adjudication; case ascertainment.

OHDSI Shoutouts!

Congratulations to the team of **Rosemarie Sadsad, Gema Ruber, Johnson Zhou, Steven Nicklin, and Guy Tsafnat** on the publication of **A computable biomedical knowledge object for calculating in-hospital mortality for patients admitted with acute myocardial infarction** in *Learning Health Systems.*



Received: 31 March 2023 | Revised: 15 August 2023 | Accepted: 15 August 2023
DOI: 10.1002/lrh2.10388

COMPUTABLE KNOWLEDGE PUBLICATIONS

Learning Health Systems

A computable biomedical knowledge object for calculating in-hospital mortality for patients admitted with acute myocardial infarction

Rosemarie Sadsad[1] | Gema Ruber[1] | Johnson Zhou[1] | Steven Nicklin[1] | Guy Tsafnat[1,2]

[1]Evidentli, Sydney, New South Wales, Australia
[2]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, New South Wales, Australia

Correspondence
Rosemarie Sadsad, Evidentli, 50 Holt Street, Surry Hills, NSW 2010, Australia.
Email: rosemarie.sadsad@evidentli.com

Abstract

Introduction: Quality indicators play an essential role in a learning health system. They help healthcare providers to monitor the quality and safety of care delivered and to identify areas for improvement. Clinical quality indicators, therefore, need to be based on real world data. Generating reliable and actionable data routinely is challenging. Healthcare data are often stored in different formats and use different terminologies and coding systems, making it difficult to generate and compare indicator reports from different sources.

Methods: The Observational Health Sciences and Informatics community maintains the Observational Medical Outcomes Partnership Common Data Model (OMOP). This is an open data standard providing a computable and interoperable format for real world data. We implemented a Computable Biomedical Knowledge Object (CBK) in the Piano Platform based on OMOP. The CBK calculates an inpatient quality indicator and was illustrated using synthetic electronic health record (EHR) data in the open OMOP standard.

Results: The CBK reported the in-hospital mortality of patients admitted for acute myocardial infarction (AMI) for the synthetic EHR dataset and includes interactive visualizations and the results of calculations. Value sets composed of OMOP concept codes for AMI and comorbidities used in the indicator calculation were also created.

Conclusion: Computable biomedical knowledge (CBK) objects that operate on OMOP data can be reused across datasets that conform to OMOP. With OMOP being a widely used interoperability standard, quality indicators embedded in CBKs can accelerate the generation of evidence for targeted quality and safety management, improving care to benefit larger populations.

OHDSI Shoutouts!

Congratulations to the team of **Evgeniy Krastev, Dimitar Tcharaktchiev, and Simeon Abanos** on the publication of **Application of OMOP Common Data Model for Data Integration: The Bulgarian Diabetes Register** in *Volume 309: Telehealth Ecosystems in Practice in Studies in Health Technology and Informatics.*

Application of OMOP Common Data Model for Data Integration: The Bulgarian Diabetes Register

Authors	Evgeniy Krastev, Dimitar Tcharaktchiev, Simeon Abanos
Pages	141 - 142
DOI	10.3233/SHTI230761
Category	Research Article
Series	Studies in Health Technology and Informatics
Ebook	Volume 309: Telehealth Ecosystems in Practice

Abstract

This paper considers mapping of the Bulgarian Diabetes Register(BDR) onto OMOP Common Data Model (CDM). Research results are referenced and plans for analysis of drug shortages from federated data sources are outlined.

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

Any shoutouts from the community? Please share and help promote and celebrate OHDSI work!

Do you have anything you want to share? Please send to sachson@ohdsi.org so we can highlight during this call and on our social channels.
Let's work together to promote the collaborative work happening in OHDSI!

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Wednesday	10 am	Surgery & Perioperative Medicine
Thursday	9:30 am	Network Data Quality
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development & Evaluation
Friday	9 am	GIS – Geographic Information System Development
Friday	1 pm	Clinical Trials
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Africa Chapter
Monday	6 pm	OMOP & FHIR
Tuesday	12 pm	Common Data Model Vocabulary Subgroup

#OHDSISocialShowcase

ohdsi.org/europe2023-showcase

#OHDSISocialShowcase

MONDAY

Building an OMOP CDM repository from OpenEHR: implementation experience of INFOBANCO

(Miguel Pedrera-Jiménez, Antonio Díaz Holgado, David Moner-Cano, Paula Rubio-Mayo, Noelia García-Barrio, Julián Jiménez-Carramiñana, Diego Boscá-Tomás, Juan Luis Cruz-Bermúdez, Javier de la Cruz-Bertolo, José Luis Bernal-Sobrino, Pablo Serrano-Balazote)

Implementation and usage of an **OMOP CDM** populated from **OpenEHR CDR**

Building an OMOP CDM repository from OpenEHR: implementation experience of INFOBANCO

Background: This study describes the process of building an OMOP CDM repository from an OpenEHR Clinical Data Repository (CDR) at Hospital Universitario 12 de Octubre (H12O) and Primary Care (PC) in Madrid Region, Spain, within the INFOBANCO platform. This OMOP CDM repository has supported the participation in several international data-driven consortiums, such as EHDEN, HONEUR and DARWIN.

Result 1: Mapping between Data Lake domains, OpenEHR archetypes and OMOP CDM.

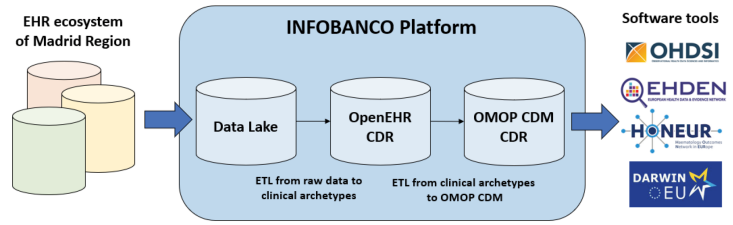
DL domain	OpenEHR archetypes	OMOP CDM tables
Patient	openEHR-EHR-CLUSTER.person.v1 openEHR-EHR-EVALUATION.gender.v1 openEHR-EHR-EVALUATION.ethnicity.v1 openEHR-EHR-EVALUATION.birth_summary.v0 openEHR-EHR-EVALUATION.death_summary.v0	Person Death
Encounter	openEHR-EHR-ADMIN_ENTRY.episode_institution.v0	Visit_ocurrence
Observation	openEHR-EHR-CLUSTER.laboratory_test_analyte.v1 openEHR-EHR-CLUSTER.genomic_variant_result.v1 *Specific archetypes for clinical observations semantics*	Observation Measurement
Diagnosis	openEHR-EHR-EVALUATION.problem_diagnosis.v1 openEHR-EHR-EVALUATION.family_history.v2	Condition_ocurrence
Medication	openEHR-EHR-CLUSTER.medication.v2 openEHR-EHR-ACTION.medication.v1 openEHR-EHR-INSTRUCTION.medication_order.v3	Drug_exposure
Procedure	openEHR-EHR-ACTION.procedure.v1	Procedure_ocurrence

Result 2: Data obtained into the OMOP CDM repository as of April 20, 2023.

OMOP CDM table	Records	Patients	Concepts
Person	292,305	292,305	-
Death	7634	7634	-
Visit_ocurrence	586,118	172,055	-
Observation	883,653	191,363	550
Measurement	361,756	102,651	15
Condition_ocurrece	1,760,706	201,839	7326
Procedure_ocurrence	502,778	58,871	2483
Drug_exposure	2,556,679	194,734	2576

Methods

INFOBANCO platform aims to create a **regional real-world data ecosystem** that provides services to clinicians, researchers and managers. It combines data from multiple heterogeneous sources in standards CDR.



Limitation: This study reports on INFOBANCO first stage. Further mapping between concept models, domains involved and data loading in the OMOP CDM repository are ongoing until complete coverage of the whole population attended in H12O and PC. In addition, a data quality protocol for continued analysis and validation is under development.

Miguel Pedrera-Jiménez, Antonio Díaz-Holgado, David Moner-Cano, Paula Rubio-Mayo, Noelia García-Barrio,Julián Jiménez-Carramiñana, Diego Boscá-Tomás, Juan Luis Cruz-Bermúdez, Javier de la Cruz-Bertolo, José Luis Bernal-Sobrino, Pablo Serrano-Balazote

FONDO EUROPEO DE DESARROLLO REGIONAL
Una manera de hacer Europa

#OHDSISocialShowcase

TUESDAY

Data OMOPisation of cancer data at Cliniques universitaires Saint-Luc

(Joëlle Thonnard, Frédéric Calay, Audrey Timmermans, Cédric Van Marcke)

An OMOP Cancer Data warehouse for data Quality Control, collaborative Real World Evidence Studies and data submission to the Belgian Cancer Registry

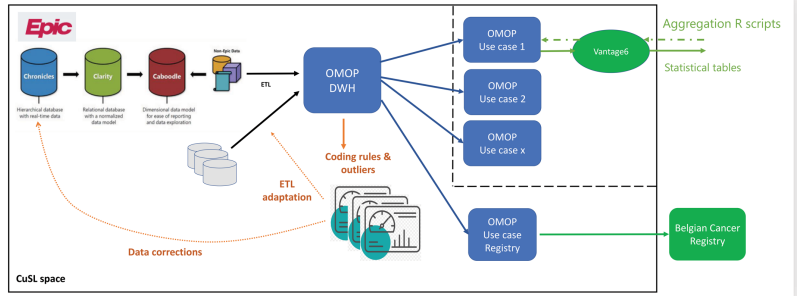
INSTITUT ROI ALBERT II CANCER ET HÉMATOLOGIE

Data OMOPisation of cancer data at Cliniques universitaires Saint-Luc

Background: Our hospital including a large cancer center increased significantly its digital maturity end 2020 following the implementation of an integrated Electronic Medical Record (EMR) based on EPIC®.

To make the multiple collected data Findable, Available, Interoperable and Reusable (FAIR), we decided to set-up an OMOP data warehouse (DWH) in collaboration with other European centres within the DigiONE project.

Results: Planned architecture



Methods

- 40 essential concepts identified for DigiONE: many already structured, evaluation of NLP needs.
- Extraction, transformation and loading (ETL) from EMR to OMOP DWH requires to agree on mapping between centers and according to international standards.
- The OMOP DWH allows to apply Quality Control rules and detection of outliers. Corrections will be implemented in the source data in the EMR.
- We need to set-up a federated research tool.
- We need to adapt our data governance and technologies according to the fast evolving legal framework.

Opportunities: Solid collaborations are needed between DigiONE members and beyond with the OMOP user community in order to reach CDMs and availability of FAIR data for further federated research and learning.

OHDSI

Joëlle Thonnard[1], Frédéric Calay[1], Audrey Timmermans[1], Cédric Van Marcke[1],
[1] Clinique universitaires Saint-Luc, Brussels, Belgium

#OHDSISocialShowcase

WEDNESDAY

Performance Improvement of Post-ETL in OMOP CDM

(Wai Yi Man, Antonella Delmestri)

Performance Improvement of Post-ETL in OMOP CDM

Presenter: Wai Yi Man

Background: Transformation of real-world data into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is not simply an Extract-Transform-Load (ETL) process. It also requires the building of primary keys, indexes and constraints using OHDSI provided sequential SQL code before the standardized data can be used. This implementation is a post-ETL operation, whose execution time depends closely on the data dimension, and can be very time consuming. The simpler approach for building primary keys, indexes, and constraints on OMOP CDM tables is running one after the other the OMOP CDM GitHub provided SQL scripts, one for PK, one for indexes and one for constraints. However, this operation could become significantly more efficient by splitting and merging these files in a way that allows concurrency to be used.

Result 1: The time complexity of execution of PK and indexes

(where R is the number of rows in the T_i table, and n is the number of tables)

$$O(R(T_1)) + O(R(T_2)) + \cdots + O(R(T_n)) = \sum_{i=1}^{n} O(R(T_i)) \Rightarrow \max_{1 \le i \le n} O(R(T_i))$$

Result 2: The time complexity execution of foreign keys when processed on target tables T_i with parent tables P_j

(where z is the number of shift)

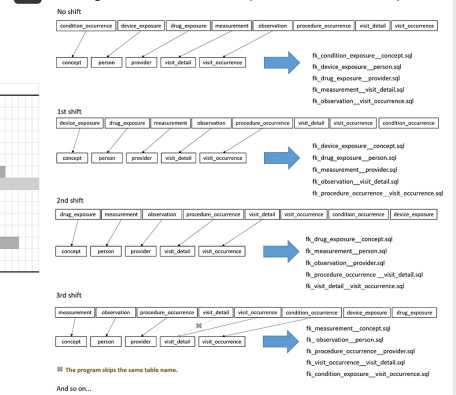
$$\sum_{i=1}^{n} \left(\sum_{\substack{j=1 \\ T_i \neq P_j}}^{m} O\left(R(T_i, P_j)\right) \right) \Rightarrow \sum_{z=0}^{n-1} \max_{\substack{1 \le j \le m \\ i = mod(j+z-1,n)+1}} O\left(R(T_i, P_j)\right)$$

Methods

1 Running PK and indexes .sql files in parallel

ID	File name	Time
1	pk_idxs_location.sql	
2	pk_idxs_care_site.sql	
3	pk_idxs_provider.sql	
4	pk_idxs_person.sql	
5	pk_idxs_observation_period.sql	
6	pk_idxs_visit_occurrence.sql	
7	pk_idxs_visit_detail.sql	
8	pk_idxs_observation.sql	
9	pk_idxs_measurement.sql	
10	pk_idxs_condition_occurrence.sql	
11	pk_idxs_procedure_occurrence.sql	
12	pk_idxs_device_exposure.sql	
13	pk_idxs_drug_exposure.sql	
14	pk_idxs_condition_era.sql	
15	pk_idxs_drug_era.sql	

2 Shift algorithm that identifies .sql files suitable to run in parallel



Wai Yi Man, Antonella Delmestri

NDORMS
NUFFIELD DEPARTMENT OF ORTHOPAEDICS, RHEUMATOLOGY AND MUSCULOSKELETAL SCIENCES
UNIVERSITY OF OXFORD

#OHDSISocialShowcase

THURSDAY

Changes in Incidence of Screening and Diagnostic Tests, Breast, Colorectal, Lung and Prostate Cancer, Before, During and After the UK National COVID-19 Lockdowns: A Cohort Study

(Nicola L. Barclay, Annika M. Jödicke, Xihang Chen, Antonella Delmestri, Berta Raventós, Wai Yi Man, Danielle Newby, Daniel Prieto-Alhambra, Marta Pineda-Moncusí, Martí Català)

COVID-19 lockdown reduced incidence of breast, colorectal, lung and prostate cancer, resulting in potentially **~62,000 missed cancer diagnoses**

Changes in Incidence of Screening and Diagnostic Tests, Breast, Colorectal, Lung and Prostate Cancer, Before, During and After the UK National COVID-19 Lockdowns: A Cohort Study

Background: Breast, colorectal, lung and prostate cancers are the most common causes of cancer death in the UK. Due to the COVID-19 pandemic many health systems postponed cancer screening and diagnostic tests, resulting in delays in diagnosis and treatment. We aimed is to understand whether cancer-related screening programmes, diagnostic tests and referrals, and incidence of four cancers, were affected by COVID-19 lockdown in the UK; and whether rates normalised to pre-pandemic levels by December 2021.

Figure 1. Incidence Rate Ratios of cancer diagnoses, in each lockdown period, with the pre-pandemic era (Jan 2017 to Feb 2020) as a reference

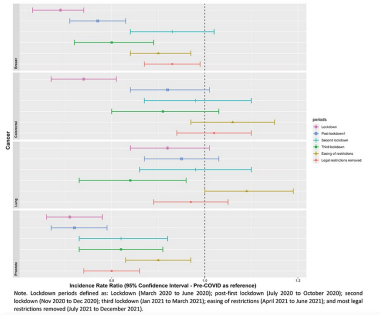
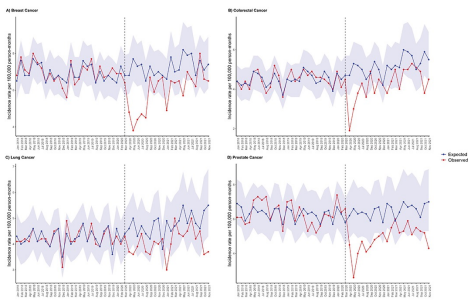


Figure 2. Expected and observed incidence rates per 100,000 person months for cancer diagnoses in CPRD GOLD UK from Jan 2018 to Dec 2021



Methods: A retrospective cohort study of electronic health records from UK primary care, using data from the Clinical Practice Research Datalink (CPRD) GOLD database, mapped to the Observational and Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Incidence rate ratios of first-ever diagnoses of breast, colorectal, lung and prostate cancer and their screening, diagnostic tests, and referrals were calculated (see Figure 1). Negative binomial regression models were run to forecast expected rates from March 2020 to December 2021 based two years data prior to the pandemic (see Figure 2).

Limitation: It is likely that the estimated shortfall in screening/diagnostic tests, and cancer diagnosis rates in the present study, are underestimates, given that many of these diagnoses are likely to be made in hospital settings, not captured in primary care data.

Nicola Barclay[1], Marta Pineda Moncusí[1], Annika Jödicke [1], Daniel Prieto-Alhambra [1], Berta Raventós[2], Danielle Newby [1], Antonella Delmestri [1], Wai Yi Man [1], Xihang Chen [1], Marti Català [1]

1 Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK.
2 Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

#OHDSISocialShowcase

FRIDAY

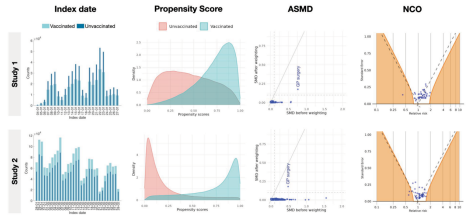
COVID-19 vaccines effectiveness against thromboembolic complications in the post-acute phase of the COVID-19 infection: a staggered cohort study using UK primary care electronic health records

(Núria Mercadé-Besora, Wai Man, Antonella Delmestri, Clara Prats, Daniel Prieto-Alhambra, Annika M Jödicke, Martí Català)

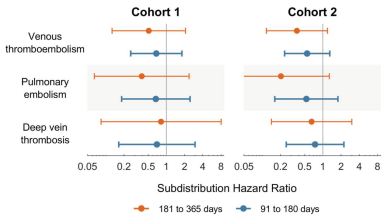
No effects were seen for COVID-19 vaccines and subacute post COVID-19 thromboembolic complications.

Title: ***COVID-19 vaccines effectiveness against thromboembolic complications in the post-acute phase of the COVID-19 infection: a staggered cohort study using UK primary care electronic health records***

Background: Research suggests that SARS-CoV-2 infection leads to a high risk of thromboembolic complications, both immediately and in the post-acute phase after infection. Effect of COVID-19 vaccination on these subacute outcomes remains unknown.



Result 1: Cohort balance.
(Index date) Distribution of index dates within enrolment periods. **(Propensity score)** Distribution of PS for vaccinated vs. unvaccinated cohorts. **(ASMD)** Absolute Standardised Mean Differences before and after overlap weighting. **(NCO)** Negative Control Outcomes subdistribution hazard ratios vs. standard deviation, and empirical calibration with 95% confidence intervals.



Result 2: Subdistribution Hazard Ratios (sHR)
sHR < 1 indicates protective vaccine effects against developing the post-acute COVID complication.

Methods

Data: UK primary care records from Clinical Practice Research Datalink (CPRD) AURUM, mapped to OMOP CDM.

Study desing and pulation:

- We conducted a staggered cohort study following the UK Goverment vaccine roll-out.
- First cohort enrolled people aged >= 75 years between 4th Jan 2021 and 27th Jan 2021. People receiving a first COVID-19 vaccine dose within this period constituted the vaccinated cohort (VC).
- Second cohort enrollment went from 28th Jan 2021 to 28th Feb 2021. Eligible individuals were >= 65 years, clinically extremely vulnerable adults or at-risk patients, and all unvaccinated persons from the first cohort.

Index date was the date of first dose for the VC. For the unvaccinated cohort (UV) index dates were randomly assigned within the enrolment period following the distribution of dates in the VC. Individuals with a history of COVID-19, or vaccination against COVID-19 before index date were excluded.

Follow-up end at first recording of outcome of interest, death, end of data availability, or first vaccine dose (UV).

Outcomes of interest: deep vein thrombosis (DVT), pulmonary embolism (PE), and venous thromboembolism (DVT+PE), in two-time windows: 91 to 180 days, and 181 to 365 days post-COVID.

Methods to account for confounding:

- **Observed confounding**: Propensity Score (PS) Overlap Weighting. Variables included in PS equation were age, GP surgery, region, prior observation years, number of previous outpatient visits, and number of previous PCR tests, in addition of covariates selected via LASSO regression.
- **Unobserved confounding:** 43 Negative Control Outcomes (NCO).

Metrics:

Fine-Gray model was used to estimate Subdistribution Hazard Ratios (sHR) for each outcome, and effect estimates were corrected using empirical calibration.

Limitations: Healthcare workers were eligible for vaccination in both cohorts, however we could not identify them and thus, they were excluded from the study. Inherent limitation of observational studies were accounted for using state-of-the-art methods like large scale propensity score overlap weighting and empirical calibration using negative control outcomes.

Núria Mercadé-Besora, Wai Yi Man, Antonella Delmestri, Clara Prats, Daniel Prieto-Alhambra, Annika M Jödicke, Martí Català

OHDSI HADES releases: SqlRender 1.16.1

DataQualityDashboard 2.4.1 | Get started | Reference | Articles | Changelog | HADES

DataQualityDashboard

DataQualityDashboard is part of HADES.

The goal of the Data Quality Dashboard (DQD) project is to design and develop an open-source tool to expose and evaluate observational data quality.

Introduction

This package will run a series of data quality checks against an OMOP CDM instance (currently supports v5.4, v5.3 and v5.2). It systematically runs the checks, evaluates the checks against some pre-specified threshold, and then communicates what was done in a transparent and easily understandable way.

Overview

The quality checks were organized according to the Kahn Framework[1] which uses a system of categories and contexts that represent strategies for assessing data quality. For an introduction to the kahn framework please click here.

Using this framework, the Data Quality Dashboard takes a systematic-based approach to running data quality checks. Instead of writing thousands of individual checks, we use "data quality check types". These "check types" are more general, parameterized data quality checks into which OMOP tables, fields, and concepts can be substituted to represent a singular data quality idea. For example, one check type might be written as

Links

Browse source code

Report a bug

Ask a question

DQD Example Output

License

Apache License (>= 2)

Citation

Citing DataQualityDash

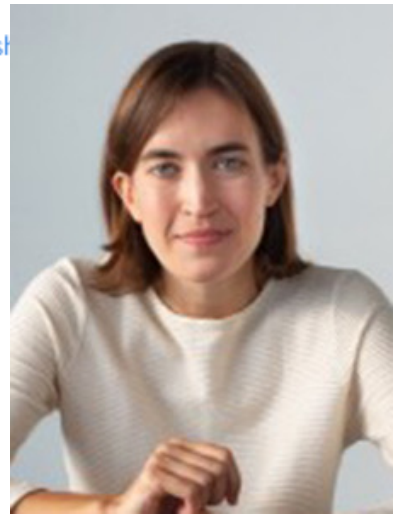
Developers

Katy Sadowski
Author, maintainer

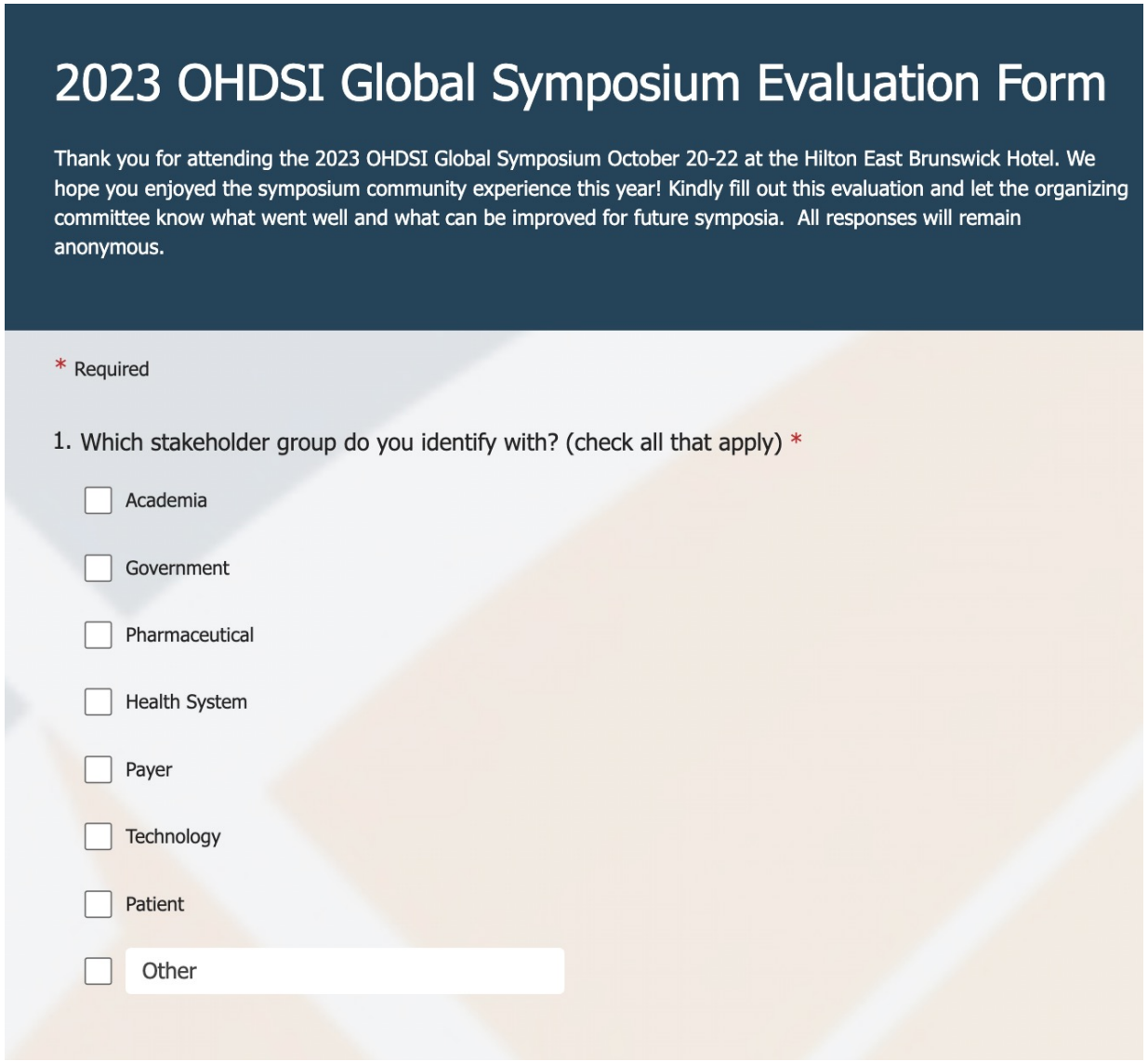
Clair Blacketer
Author

Ajit Londhe
Author

Anthony Sena
Author



Please Fill Out The Symposium Evaluation!



2023 OHDSI Global Symposium Evaluation Form

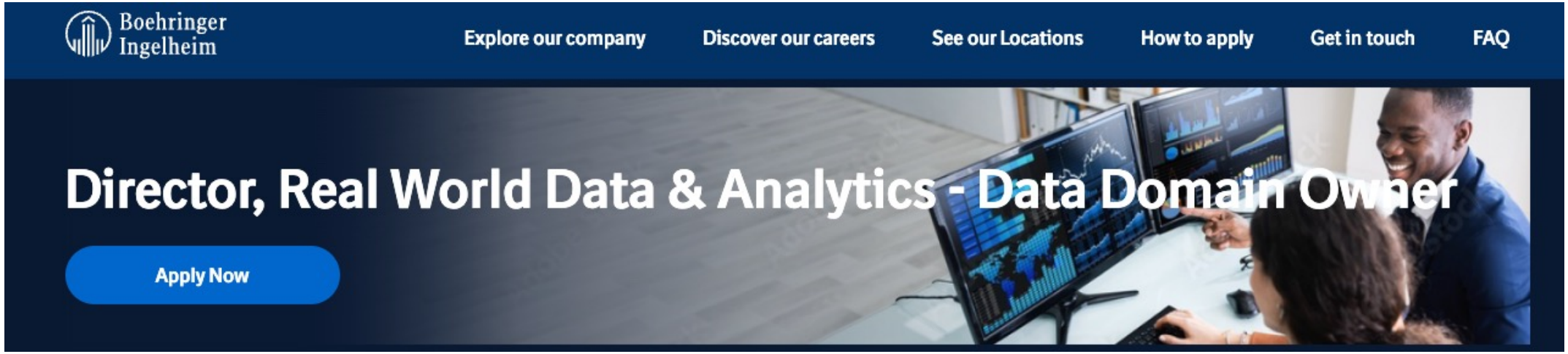
Thank you for attending the 2023 OHDSI Global Symposium October 20-22 at the Hilton East Brunswick Hotel. We hope you enjoyed the symposium community experience this year! Kindly fill out this evaluation and let the organizing committee know what went well and what can be improved for future symposia. All responses will remain anonymous.

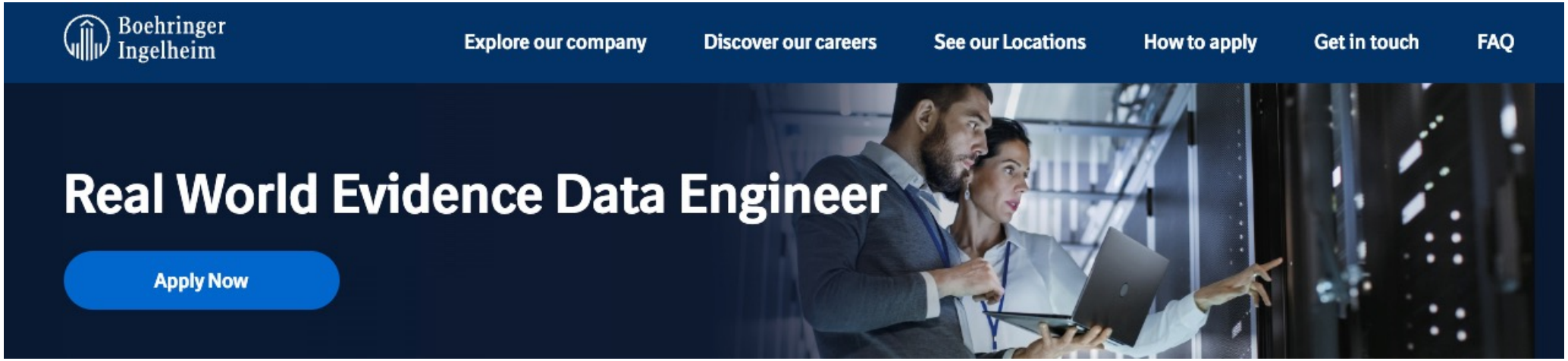
* Required

1. Which stakeholder group do you identify with? (check all that apply) *

- [ ] Academia
- [ ] Government
- [ ] Pharmaceutical
- [ ] Health System
- [ ] Payer
- [ ] Technology
- [ ] Patient
- [ ] Other

Openings at Boehringer Ingelheim





Opening: Postdoctoral Associate/Data Analyst

Job Announcement: Postdoctoral Associate/Data Analyst - LEGEND Hypertension Project

Position: Postdoctoral Associate/Data Analyst

Organization: Yale University, School of Medicine

Location: 195 Church Street, 5th floor, New Haven, CT, 06510

Application Deadline: Rolling basis

Job Description:

We are seeking a talented and dedicated Postdoctoral Associate/Data Analyst to join our dynamic team. In this role, you will play a pivotal part in advancing our mission of improving health outcomes through data-driven research. You will have the opportunity to work with diverse healthcare datasets, develop innovative analytical methods, and collaborate with experts in the field.

The Postdoctoral Associate/Data Analyst should possess significant experience in R and Rstudio, with specific expertise in database management using PostgreSQL—critical requirements within the OHDSI network. Your responsibilities will include assisting the Principal Investigator (Dr. Yuan Lu from Yale University) and Co-Investigator (Drs. Marc Suchard from UCLA) in creating the analytic tool stack and performing related analyses.

Key Responsibilities:

- Collaborate with multidisciplinary teams to design and execute data analysis projects.
- Develop and implement statistical and machine learning models for healthcare data.
- Perform data extraction and preprocessing tasks to prepare datasets for analysis.
- Conduct exploratory data analysis and visualization to extract insights from healthcare data.
- Assist in the development and maintenance of OHDSI's open-source tools and resources.
- Communicate findings and insights through reports, presentations, and publications.
- Stay up-to-date with the latest advancements in data science and healthcare informatics.

Email: y.lu@yale.edu

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?