Introduction to Phenotype Phebruary III

OHDSI Community Call
Jan. 30, 2024 • 11 am ET

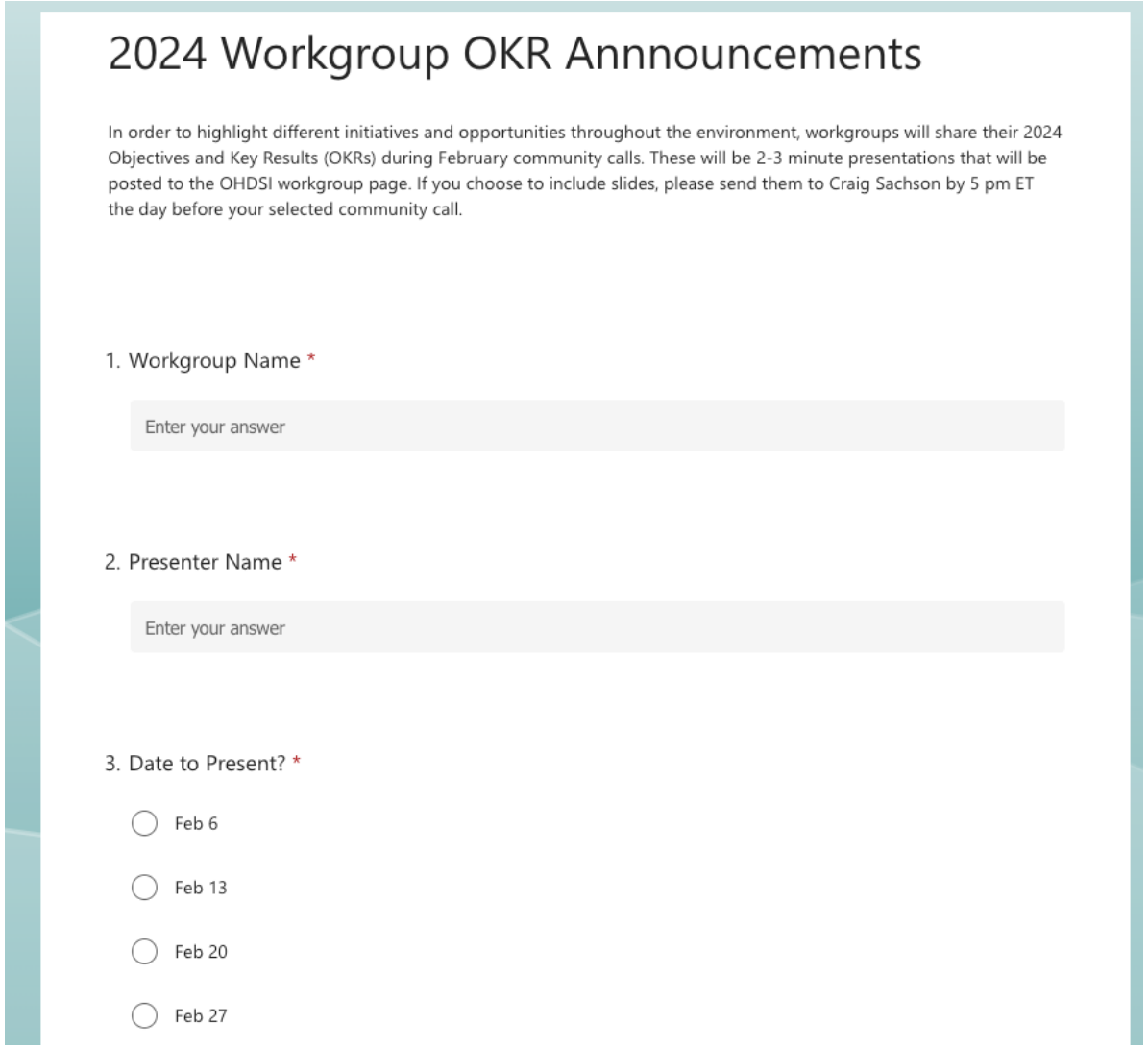
Upcoming Community Calls

Date	Topic
Jan. 30	Phenotype Phebruary Introduction
Feb. 6	Workgroup OKRs / Phenotype Phebruary Update 1
Feb. 13	Workgroup OKRs / Phenotype Phebruary Update 2
Feb. 20	Workgroup OKRs / Phenotype Phebruary Update 3
Feb. 27	Workgroup OKRs / Phenotype Phebruary Update 4

WG Leads: Please Sign Up For OKR Announcements

Currently Signed Up:

- Africa Chapter
- APAC
- CDM
- CDM Vocabulary Subgroup
- Dentistry
- Electronic Animal H.R.
- Eyecare & Vision Research
- FHIR + OMOP
- Generative AI and Analytics in Healthcare (GAIA)
- HADES
- Health Equity
- Healthcare Systems
- Medical Devices
- Methods Research
- NLP
- Network Data Quality
- Oncology
- Patient-Level Prediction
- Perinatal and Reproductive Health
- Phenotype Development & Evaluation
- Registry
- Steering Group
- Themis
- Vaccine Vocabulary



2024 Workgroup OKR Annnouncements

In order to highlight different initiatives and opportunities throughout the environment, workgroups will share their 2024 Objectives and Key Results (OKRs) during February community calls. These will be 2-3 minute presentations that will be posted to the OHDSI workgroup page. If you choose to include slides, please send them to Craig Sachson by 5 pm ET the day before your selected community call.

1. Workgroup Name *

Enter your answer

2. Presenter Name *

Enter your answer

3. Date to Present? *

- Feb 6
- Feb 13
- Feb 20
- Feb 27

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Soobeen Seol, Jimyung Park, Chungsoo Kim, Dong Yun Lee, and Rae Woong Park** on the publication of **RHEA: Real-World Observational Health Data Exploration Application** in Volume 310 of *Studies in Health Technology and Informatics.*

1474

MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231251

RHEA: Real-World Observational Health Data Exploration Application

Soobeen SEOL[a], Jimyung PARK[a], Chungsoo KIM[a], Dong Yun LEE[b] and Rae Woong PARK[a,b,1]

[a]*Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Republic of Korea;* [b]*Department of Biomedical Informatics, Ajou University School of Medicine, Republic of Korea*

ORCiD ID: Soobeen Seol https://orcid.org/0000-0002-7238-8986, Jimyung Park https://orcid.org/0000-0002-6998-2546, Chungsoo Kim https://orcid.org/0000-0003-1802-1777, Dong Yun Lee https://orcid.org/0000-0002-3678-9862 and Rae Woong Park https://orcid.org/0000-0003-4989-3287

Abstract. We developed a standardized framework named RHEA to represent longitudinal status of patient with cancer. RHEA generates a dashboard to visualize patients' data in the Observational Medical Outcomes Partnership-Common Data Model format. The generated dashboard consists of three main parts for providing the macroscopic characteristics of the patient: 1) cohort-level visualization, 2) individual-level visualization and 3) cohort generation.

Keywords. Common data model, data visualization, electronic health records

1. Introduction

For clinicians' appropriate clinical decision-making, patients' clinical data should be readily available. However, in general, patient data in electronic health records (EHRs) are fragmented and passive in nature, which makes it difficult to understand. We aim to develop a framework named REHA, a real-world observational health data exploration application, to reduce clinicians' cognitive efforts to understand patients' data and to provide insight into patients.

OHDSI Shoutouts!

Congratulations to the team of **Sujin Gan, Chungsoo Kim, Dong Yun Lee, and Rae Woong Park** on the publication of **Prediction Models for Readmission Using Home Healthcare Notes and OMOP-CDM** in Volume 310 of *Studies in Health Technology and Informatics.*

1438

MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231233

Prediction Models for Readmission Using Home Healthcare Notes and OMOP-CDM

Sujin Gan[a], Chungsoo Kim[a], Dong Yun Lee[b] and Rae Woong Park[a,b,1]
[a] *Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea*
[b] *Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea*
ORCiD ID: Sujin Gan https://orcid.org/0000-0003-3326-8428, Chungsoo Kim https://orcid.org/0000-0003-1802-1777, Dong Yun Lee https://orcid.org/0000-0002-3678-9862 Rae, Woong Park https://orcid.org/0000-0003-4989-3287

Abstract. This study developed readmission prediction models using Home Healthcare (HHC) documents via natural language processing (NLP). An electronic health record of Ajou University Hospital was used to develop prediction models (A reference model using only structured data, and an NLP-enriched model with structured and unstructured data). Among 573 patients, 63 were readmitted to the hospital. Five topics were extracted from HHC documents and improved the model performance (AUROC 0.740).

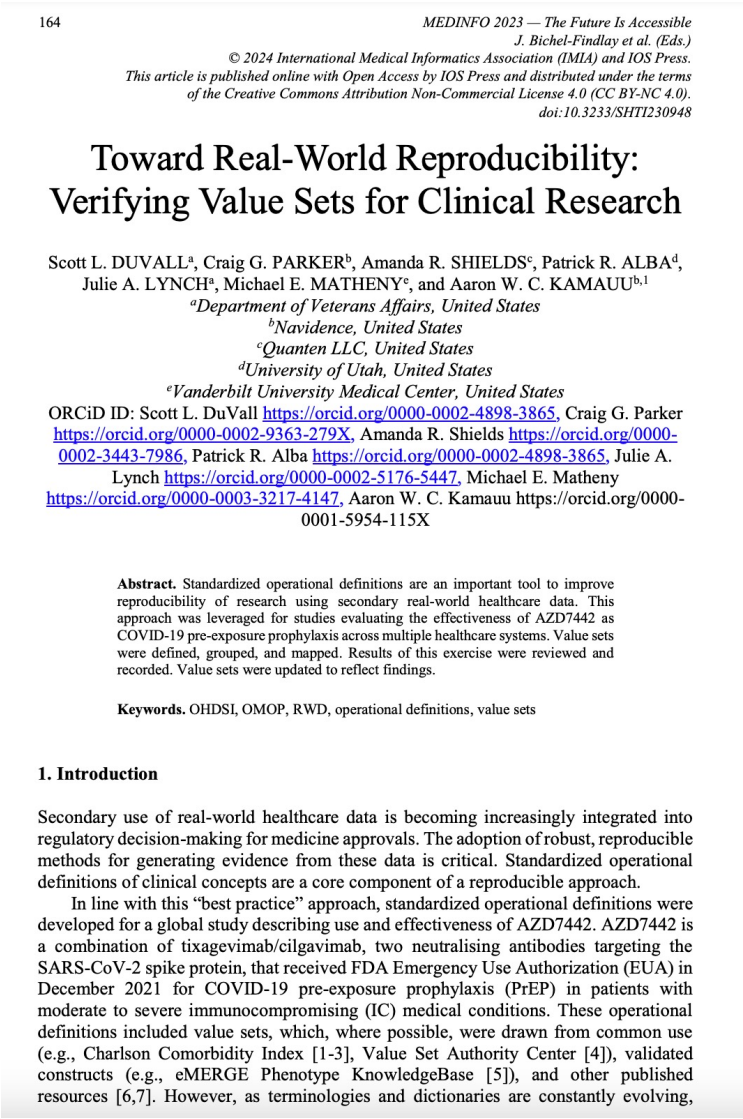
Keywords. Readmission, home healthcare, machine learning, prediction

1. Introduction

Readmission is an indicator of inpatient care quality and a major contributor to growing healthcare costs [1]. Therefore, identifying patients at high risk for readmission is crucial to reduce the likelihood of readmission. Home healthcare (HHC) is provided to discharged patients. Thus, HHC documents, which include post-discharge information from medical procedures to patient complaints, may contain hidden risk factors. Machine learning-based prediction models have been developed to assist in the identification of readmission risk factors, but they have not been applied to HHC documents.

OHDSI Shoutouts!

Congratulations to the team of **Scott DuVall, Craig Parker, Amanda Shields, Patrick Alba, Julie Lynch, Michael Matheny, and Aaron Kamauu** on the publication of **Toward Real-World Reproducibility: Verifying Value Sets for Clinical Research** in Volume 310 of *Studies in Health Technology and Informatics.*



MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230948

Toward Real-World Reproducibility: Verifying Value Sets for Clinical Research

Scott L. DUVALL[a], Craig G. PARKER[b], Amanda R. SHIELDS[c], Patrick R. ALBA[d], Julie A. LYNCH[a], Michael E. MATHENY[e], and Aaron W. C. KAMAUU[b,1]
[a]*Department of Veterans Affairs, United States*
[b]*Navidence, United States*
[c]*Quanten LLC, United States*
[d]*University of Utah, United States*
[e]*Vanderbilt University Medical Center, United States*
ORCiD ID: Scott L. DuVall https://orcid.org/0000-0002-4898-3865, Craig G. Parker https://orcid.org/0000-0002-9363-279X, Amanda R. Shields https://orcid.org/0000-0002-3443-7986, Patrick R. Alba https://orcid.org/0000-0002-4898-3865, Julie A. Lynch https://orcid.org/0000-0002-5176-5447, Michael E. Matheny https://orcid.org/0000-0003-3217-4147, Aaron W. C. Kamauu https://orcid.org/0000-0001-5954-115X

Abstract. Standardized operational definitions are an important tool to improve reproducibility of research using secondary real-world healthcare data. This approach was leveraged for studies evaluating the effectiveness of AZD7442 as COVID-19 pre-exposure prophylaxis across multiple healthcare systems. Value sets were defined, grouped, and mapped. Results of this exercise were reviewed and recorded. Value sets were updated to reflect findings.

Keywords. OHDSI, OMOP, RWD, operational definitions, value sets

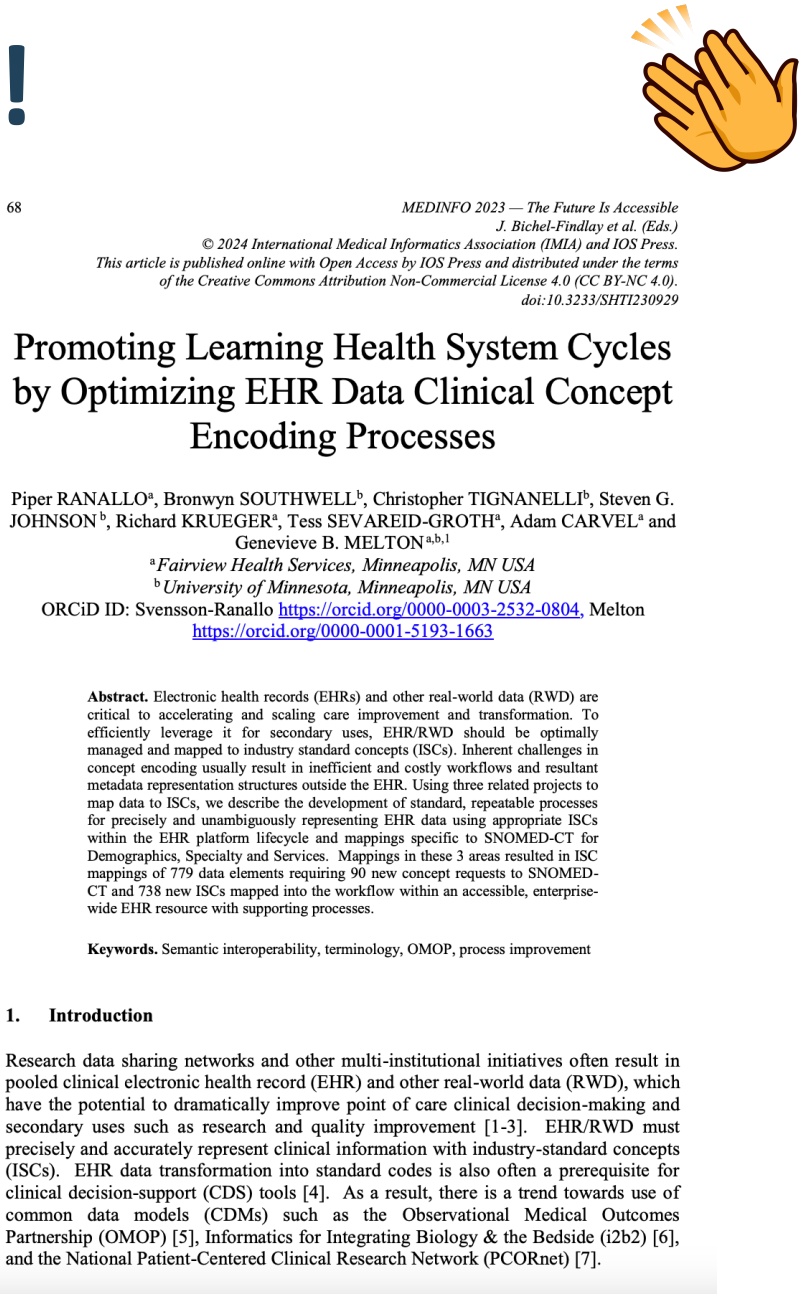
1. Introduction

Secondary use of real-world healthcare data is becoming increasingly integrated into regulatory decision-making for medicine approvals. The adoption of robust, reproducible methods for generating evidence from these data is critical. Standardized operational definitions of clinical concepts are a core component of a reproducible approach.

In line with this "best practice" approach, standardized operational definitions were developed for a global study describing use and effectiveness of AZD7442. AZD7442 is a combination of tixagevimab/cilgavimab, two neutralising antibodies targeting the SARS-CoV-2 spike protein, that received FDA Emergency Use Authorization (EUA) in December 2021 for COVID-19 pre-exposure prophylaxis (PrEP) in patients with moderate to severe immunocompromising (IC) medical conditions. These operational definitions included value sets, which, where possible, were drawn from common use (e.g., Charlson Comorbidity Index [1-3], Value Set Authority Center [4]), validated constructs (e.g., eMERGE Phenotype KnowledgeBase [5]), and other published resources [6,7]. However, as terminologies and dictionaries are constantly evolving,

OHDSI Shoutouts!

Congratulations to the team of **Piper Ranallo, Bronwyn Southwell, Christopher Tignanelli, Steven G. Johnson, Richard Krueger, Tess Sevareid-Groth, Adam Carvel, and Genevieve B. Melton** on the publication of **Promoting Learning Health System Cycles by Optimizing EHR Data Clinical Concept Encoding Processes** in Volume 310 of *Studies in Health Technology and Informatics.*



MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230929

Promoting Learning Health System Cycles by Optimizing EHR Data Clinical Concept Encoding Processes

Piper RANALLO[a], Bronwyn SOUTHWELL[b], Christopher TIGNANELLI[b], Steven G. JOHNSON[b], Richard KRUEGER[a], Tess SEVAREID-GROTH[a], Adam CARVEL[a] and Genevieve B. MELTON[a,b,1]
[a] *Fairview Health Services, Minneapolis, MN USA*
[b] *University of Minnesota, Minneapolis, MN USA*
ORCiD ID: Svensson-Ranallo https://orcid.org/0000-0003-2532-0804, Melton https://orcid.org/0000-0001-5193-1663

Abstract. Electronic health records (EHRs) and other real-world data (RWD) are critical to accelerating and scaling care improvement and transformation. To efficiently leverage it for secondary uses, EHR/RWD should be optimally managed and mapped to industry standard concepts (ISCs). Inherent challenges in concept encoding usually result in inefficient and costly workflows and resultant metadata representation structures outside the EHR. Using three related projects to map data to ISCs, we describe the development of standard, repeatable processes for precisely and unambiguously representing EHR data using appropriate ISCs within the EHR platform lifecycle and mappings specific to SNOMED-CT for Demographics, Specialty and Services. Mappings in these 3 areas resulted in ISC mappings of 779 data elements requiring 90 new concept requests to SNOMED-CT and 738 new ISCs mapped into the workflow within an accessible, enterprise-wide EHR resource with supporting processes.

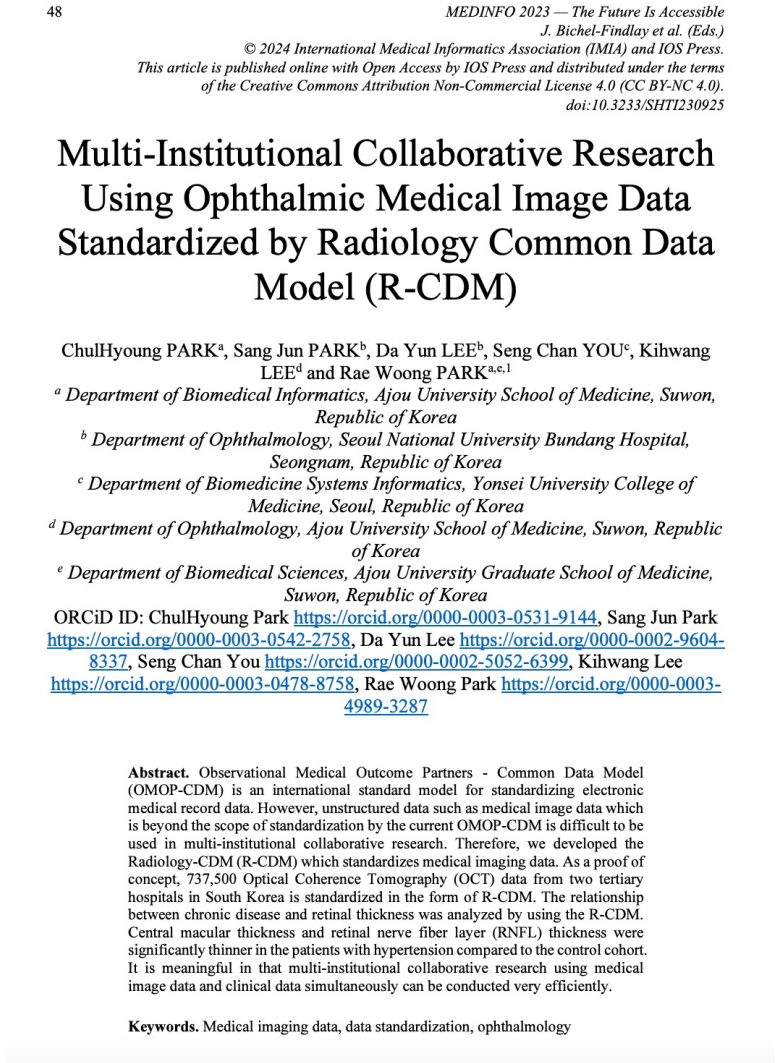
Keywords. Semantic interoperability, terminology, OMOP, process improvement

1. Introduction

Research data sharing networks and other multi-institutional initiatives often result in pooled clinical electronic health record (EHR) and other real-world data (RWD), which have the potential to dramatically improve point of care clinical decision-making and secondary uses such as research and quality improvement [1-3]. EHR/RWD must precisely and accurately represent clinical information with industry-standard concepts (ISCs). EHR data transformation into standard codes is also often a prerequisite for clinical decision-support (CDS) tools [4]. As a result, there is a trend towards use of common data models (CDMs) such as the Observational Medical Outcomes Partnership (OMOP) [5], Informatics for Integrating Biology & the Bedside (i2b2) [6], and the National Patient-Centered Clinical Research Network (PCORnet) [7].

OHDSI Shoutouts!

Congratulations to the team of **ChulHyoung Park, Sang Jun Park, Da Yun Lee, Seng Chan You, Kihwang Lee, and Rae Woong Park** on the publication of **Multi-Institutional Collaborative Research Using Ophthalmic Medical Image Data Standardized by Radiology Common Data Model (R-CDM)** in Volume 310 of *Studies in Health Technology and Informatics.*



MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230925

Multi-Institutional Collaborative Research Using Ophthalmic Medical Image Data Standardized by Radiology Common Data Model (R-CDM)

ChulHyoung PARK[a], Sang Jun PARK[b], Da Yun LEE[b], Seng Chan YOU[c], Kihwang LEE[d] and Rae Woong PARK[a,e,1]
[a] *Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea*
[b] *Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea*
[c] *Department of Biomedicine Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea*
[d] *Department of Ophthalmology, Ajou University School of Medicine, Suwon, Republic of Korea*
[e] *Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Republic of Korea*
ORCiD ID: ChulHyoung Park https://orcid.org/0000-0003-0531-9144, Sang Jun Park https://orcid.org/0000-0003-0542-2758, Da Yun Lee https://orcid.org/0000-0002-9604-8337, Seng Chan You https://orcid.org/0000-0002-5052-6399, Kihwang Lee https://orcid.org/0000-0003-0478-8758, Rae Woong Park https://orcid.org/0000-0003-4989-3287

Abstract. Observational Medical Outcome Partners - Common Data Model (OMOP-CDM) is an international standard model for standardizing electronic medical record data. However, unstructured data such as medical image data which is beyond the scope of standardization by the current OMOP-CDM is difficult to be used in multi-institutional collaborative research. Therefore, we developed the Radiology-CDM (R-CDM) which standardizes medical imaging data. As a proof of concept, 737,500 Optical Coherence Tomography (OCT) data from two tertiary hospitals in South Korea is standardized in the form of R-CDM. The relationship between chronic disease and retinal thickness was analyzed by using the R-CDM. Central macular thickness and retinal nerve fiber layer (RNFL) thickness were significantly thinner in the patients with hypertension compared to the control cohort. It is meaningful in that multi-institutional collaborative research using medical image data and clinical data simultaneously can be conducted very efficiently.

Keywords. Medical imaging data, data standardization, ophthalmology

OHDSI Shoutouts!

Congratulations to the team of **Seol Whan Oh, Soo Jeong Ko, Yun Seon Im, Surin Jung, Bo Yeon Choi, Jae Yoon Kim, Sunghyeon Park, Wona Choi, and In Young Choi** on the publication of **Development of Integrated Data Quality Management System for Observational Medical Outcomes Partnership Common Data Model** in Volume 310 of *Studies in Health Technology and Informatics.*

MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230985

Development of Integrated Data Quality Management System for Observational Medical Outcomes Partnership Common Data Model

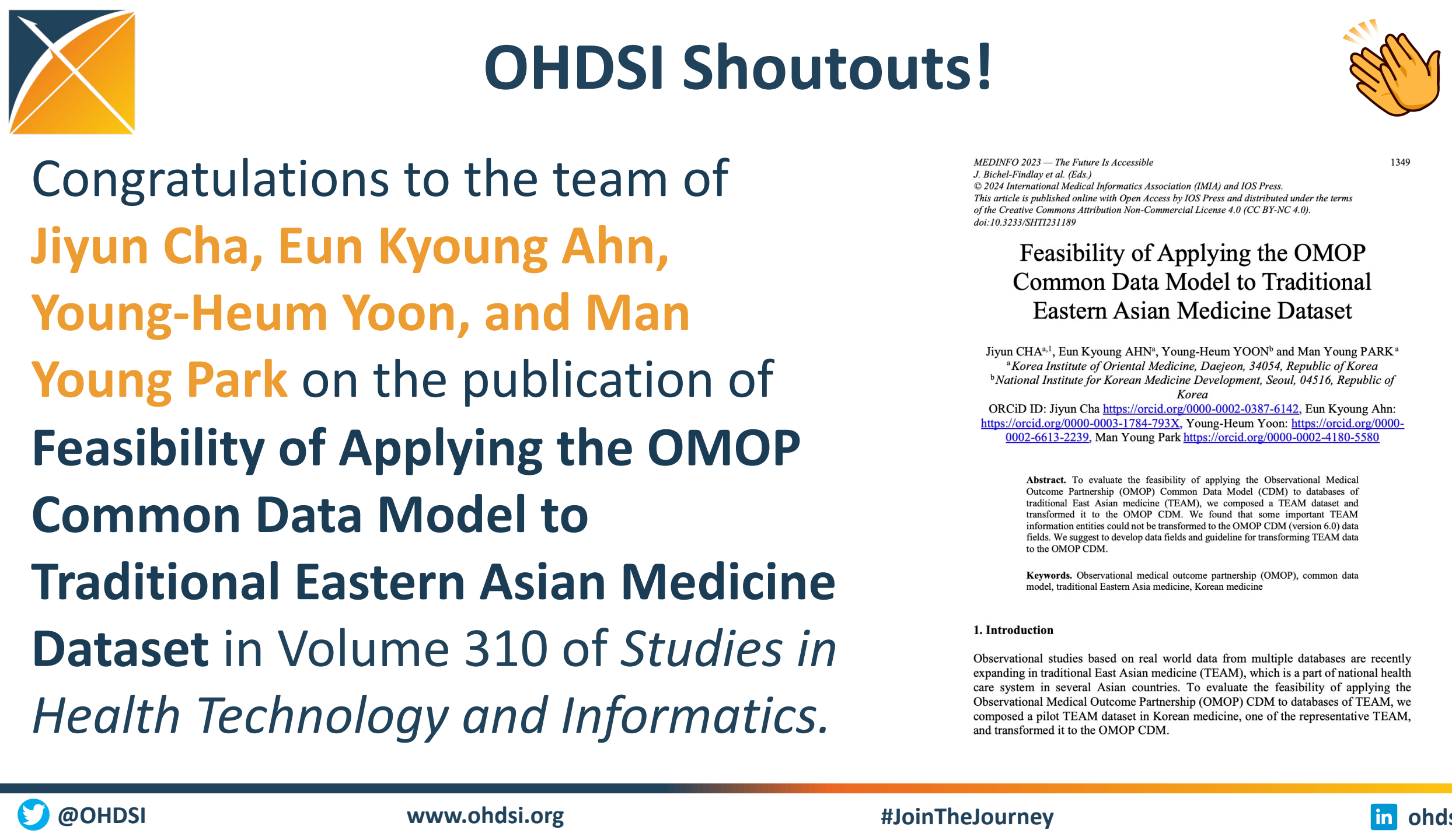
Seol Whan OH[a,b], Soo Jeong KO[a,b], Yun Seon IM[a,b], Surin JUNG[a], Bo Yeon CHOI[a,b], Jae Yoon KIM[a,b], Sunghyeon PARK[a,b], Wona CHOI[a,1] and In Young CHOI[a]
[a] *Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea*
[b] *Department of Biomedicine &Health Sciences, The Catholic University of Korea, Seoul, Republic of Korea*
ORCiD ID: Seol Whan OH https://orcid.org/0000-0002-0328-9634,Soo Jeong KO https://orcid.org/0000-0002-6550-9188,Yun Seon IM https://orcid.org/0000-0002-2510-1380, Surin JUNG https://orcid.org/0000-0002-3314-3185, Bo Yeon CHOI https://orcid.org/0000-0003-1311-2645, Jae Yoon KIM https://orcid.org/0000-0001-8847-9586, Sunghyeon PARK https://orcid.org/0000-0002-2235-4358, Wona CHOI https://orcid.org/0000-0003-0269-6374, In Young CHOI https://orcid.org/0000-0002-2860-9411

Abstract. The amount of research on the gathering and handling of healthcare data keeps growing. To support multi-center research, numerous institutions have sought to create a common data model (CDM). However, data quality issues continue to be a major obstacle in the development of CDM. To address these limitations, a data quality assessment system was created based on the representative data model OMOP CDM v5.3.1. Additionally, 2,433 advanced evaluation rules were created and incorporated into the system by mapping the rules of existing OMOP CDM quality assessment systems. The data quality of six hospitals was verified using the developed system and an overall error rate of 0.197% was confirmed. Finally, we proposed a plan for high-quality data generation and the evaluation of multi-center CDM quality.

Keywords. Data quality, common data model, data quality management system

OHDSI Shoutouts!

Congratulations to the team of **Jiyun Cha, Eun Kyoung Ahn, Young-Heum Yoon, and Man Young Park** on the publication of **Feasibility of Applying the OMOP Common Data Model to Traditional Eastern Asian Medicine Dataset** in Volume 310 of *Studies in Health Technology and Informatics.*



MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231189

Feasibility of Applying the OMOP Common Data Model to Traditional Eastern Asian Medicine Dataset

Jiyun CHA[a,1], Eun Kyoung AHN[a], Young-Heum YOON[b] and Man Young PARK[a]
[a]*Korea Institute of Oriental Medicine, Daejeon, 34054, Republic of Korea*
[b]*National Institute for Korean Medicine Development, Seoul, 04516, Republic of Korea*
ORCiD ID: Jiyun Cha https://orcid.org/0000-0002-0387-6142, Eun Kyoung Ahn: https://orcid.org/0000-0003-1784-793X, Young-Heum Yoon: https://orcid.org/0000-0002-6613-2239, Man Young Park https://orcid.org/0000-0002-4180-5580

Abstract. To evaluate the feasibility of applying the Observational Medical Outcome Partnership (OMOP) Common Data Model (CDM) to databases of traditional East Asian medicine (TEAM), we composed a TEAM dataset and transformed it to the OMOP CDM. We found that some important TEAM information entities could not be transformed to the OMOP CDM (version 6.0) data fields. We suggest to develop data fields and guideline for transforming TEAM data to the OMOP CDM.

Keywords. Observational medical outcome partnership (OMOP), common data model, traditional Eastern Asia medicine, Korean medicine

1. Introduction

Observational studies based on real world data from multiple databases are recently expanding in traditional East Asian medicine (TEAM), which is a part of national health care system in several Asian countries. To evaluate the feasibility of applying the Observational Medical Outcome Partnership (OMOP) CDM to databases of TEAM, we composed a pilot TEAM dataset in Korean medicine, one of the representative TEAM, and transformed it to the OMOP CDM.

OHDSI Shoutouts!

Congratulations to the team of **Martijn Schuemie, Jenna Reps, Adam Black, Frank Defalco, Lee Evans, Egill Fridgeirsson, James P. Gilbert, Chris Knoll, Martin Lavallee, Gowtham A. Rao, Peter Rijnbeek, Katy Sadowski, Anthony Sena, Joel Swerdel, Ross D. Williams, and Marc Suchard** on the publication of **Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research** in Volume 310 of *Studies in Health Technology and Informatics.*

966

MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231108

Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research

Martijn SCHUEMIE[a,b,c,1], Jenna REPS[a,b,d], Adam BLACK[a,e], Frank DeFALCO[a,b], Lee EVANS[a,f], Egill FRIDGEIRSSON[a,d], James P. GILBERT[a,b], Chris KNOLL[a,b], Martin LAVALLEE[a,g], Gowtham A. RAO[a,b], Peter RIJNBEEK[a,d], Katy SADOWSKI[a,h], Anthony SENA[a,b,d], Joel SWERDEL[a,b], Ross D. WILLIAMS[a,d] and Marc SUCHARD[a,c,i]

[a] *Observational Health Data Science and Informatics, New York, NY, USA*
[b] *Observational Health Data Analytics, Johnson & Johnson, Titusville, NJ, USA*
[c] *Department of Biostatistics, UCLA, Los Angeles, CA, USA*
[d] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[e] *Odysseus Data Services Inc., Cambridge, MA, USA*
[f] *LTS Computing LLC, West Chester, PA, USA*
[g] *Virginia Commonwealth University, Richmond, VA, USA*
[h] *TrialSpark Inc., New York, NY, USA*
[i] *VA Informatics and Computing Infrastructure, Department of Veterans Affairs, Salt Lake City, UT, USA*

Abstract. The Health-Analytics Data to Evidence Suite (HADES) is an open-source software collection developed by Observational Health Data Sciences and Informatics (OHDSI). It executes directly against healthcare data such as electronic health records and administrative claims, that have been converted to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. Using advanced analytics, HADES performs characterization, population-level causal effect estimation, and patient-level prediction, potentially across a federated data network, allowing patient-level data to remain locally while only aggregated statistics are shared. Designed to run across a wide array of technical environments, including different operating systems and database platforms, HADES uses continuous integration with a large set of unit tests to maintain reliability. HADES implements OHDSI best practices, and is used in almost all published OHDSI studies, including some that have directly informed regulatory decisions.

Keywords. Observational research, software, open-source, machine learning, epidemiology

OHDSI Shoutouts!

Congratulations to the team of **Najia Ahmadi, Quang Vu Nguyen, Martin Sedlmayr, and Markus Wolfien** on the publication of **A comparative patient-level prediction study in OMOP CDM: applicative potential and insights from synthetic data** in *Scientific Reports.*

scientific reports

Check for updates

OPEN

A comparative patient-level prediction study in OMOP CDM: applicative potential and insights from synthetic data

Najia Ahmadi[1,3]✉, Quang Vu Nguyen[1,3], Martin Sedlmayr[1] & Markus Wolfien[1,2]

The emergence of collaborations, which standardize and combine multiple clinical databases across different regions, provide a wealthy source of data, which is fundamental for clinical prediction models, such as patient-level predictions. With the aid of such large data pools, researchers are able to develop clinical prediction models for improved disease classification, risk assessment, and beyond. To fully utilize this potential, Machine Learning (ML) methods are commonly required to process these large amounts of data on disease-specific patient cohorts. As a consequence, the Observational Health Data Sciences and Informatics (OHDSI) collaborative develops a framework to facilitate the application of ML models for these standardized patient datasets by using the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). In this study, we compare the feasibility of current web-based OHDSI approaches, namely ATLAS and "Patient-level Prediction" (PLP), against a native solution (R based) to conduct such ML-based patient-level prediction analyses in OMOP. This will enable potential users to select the most suitable approach for their investigation. Each of the applied ML solutions was individually utilized to solve the same patient-level prediction task. Both approaches went through an exemplary benchmarking analysis to assess the weaknesses and strengths of the PLP R-Package. In this work, the performance of this package was subsequently compared versus the commonly used native R-package called *Machine Learning in R 3* (mlr3), and its sub-packages. The approaches were evaluated on performance, execution time, and ease of model implementation. The results show that the PLP package has shorter execution times, which indicates great scalability, as well as intuitive code implementation, and numerous possibilities for visualization. However, limitations in comparison to native packages were depicted in the implementation of specific ML classifiers (e.g., Lasso), which may result in a decreased performance for real-world prediction problems. The findings here contribute to the overall effort of developing ML-based prediction models on a clinical scale and provide a snapshot for future studies that explicitly aim to develop patient-level prediction models in OMOP CDM.

OHDSI Shoutouts!

Congratulations to **Ines Reinecke** on successfully defending her doctoral thesis at Technische Universitat Dresden.

Dr. Reinecke's dissertation was titled "Secondary use of German inpatient medication data in international studies while preserving semantics." Congrats!

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 and Perioperative Medicine
Wednesday	3 pm	Vulcan/OHDSI Meeting (ZOOM)
Thursday	9:30 am	Themis
Thursday	11 am	Industry
Thursday	12 pm	Medical Devices
Thursday	12 pm	Methods Research
Thursday	1 pm	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	7 pm	Dentistry
Friday	10 am	GIS – Geographic Information System
Friday	11:30 am	Clinical Trials
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Africa Chapter
Tuesday	9 am	ATLAS/WebAPI
Tuesday	10 am	Common Data Model

OHDSI Evidence Network

OHDSI Evidence Network

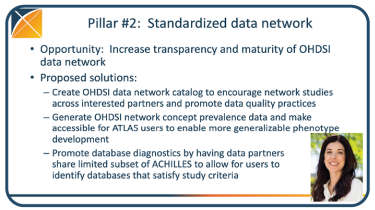
OHDSI is proud to have a global community dedicated to generating real-world evidence and which recognizes the opportunity to collaborate together as part of a distributed network based on standardized data and standardized analytics.

The OHDSI Evidence Network consists of organizations equipped with access to one or more databases standardized to the OMOP CDM who express a keen interest in participating in OHDSI network studies. Collaboratively, OHDSI Evidence Network partners share aggregate summary statistics about their databases, which are used to support Database Diagnostics, helping identify databases within the network that are fit-for-use for particular research questions. Additionally, partners have the opportunity to opt in and contribute to network studies proposed by the OHDSI community.

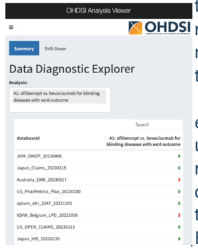
The recent SOS challenge serves as a compelling demonstration of the OHDSI Evidence Network's current capabilities and its promising future potential. We wholeheartedly encourage all organizations that are adopting the OMOP CDM and aspire to apply standardized analytics for the reliable generation of real-world evidence to become part of the OHDSI Evidence Network.

A message from Common Data Model workgroup lead Clair Blacketer …

During the first community call of 2023, Patrick Ryan unveiled the strategic priorities for the OHDSI Community for the year. Among these, a key focus is on enhancing the transparency and maturity of the OHDSI network.



To address this objective, we began by considering how network studies are currently conducted, recognizing the challenges and complexities faced by collaborating organizations when contributing to the body of evidence. This investigation led to the creation of Database Diagnostics, a tool designed to answer a critical question: when tackling a specific research inquiry, which data sources within the OHDSI Evidence Network are the most relevant and suitable for generating robust evidence?



This innovative approach leverages aggregated summary statistics from each data source, obtained through the open-source tool dbProfile. It evaluates data fitness-for-use across various dimensions, including patient demographics, domain coverage, longitudinal data availability, and the capture of target, comparator, and outcome variables. The overarching vision was to establish these database profiles as the foundation to enable the OHDSI Evidence Network.

Organizations and Data Sources in the OHDSI Evidence Network

Ajou University • Ajou University
Casa di Cura Igea • Casa di Cura Igea
Clinical Center of Montenegro • Clinical Center of Montenegro
Columbia University Medical Center • Columbia University Medical Center
Hong Kong University • UK THIN
IQVIA • Australia EMR
IQVIA • Disease Analyzer France
IQVIA • Disease Analyzer Germany
IQVIA • Japan Claims
IQVIA • Japan HIS
IQVIA • Longitudinal Patient Database (LPD) in Belgium
IQVIA • Longitudinal Patient Database (LPD) in France
IQVIA • Longitudinal Patient Database (LPD) in Italy
IQVIA • Longitudinal Patient Database (LPD) in Spain
IQVIA • OMOP US Hospital Data Master
IQVIA • Pharmetrics Plus
IQVIA • UK Medical Research Data EMIS
IQVIA • UK Medical Research Data THIN
IQVIA • US Open Claims
Janssen Research & Development • JMDC
Janssen Research & Development • Merative® Marketscan® Commercial Claims and Encounters
Janssen Research & Development • Merative® Marketscan® Medicare Supplemental
Janssen Research & Development • Merative® Marketscan® Multi-State Medicaid
Janssen Research & Development • Optum's Clinformatics® Data Mart - Date of Death
Janssen Research & Development • Optum's Clinformatics® Data Mart - Socio-Economic Status
Janssen Research & Development • Optum's Longitudinal EHR Repository
Janssen Research & Development • Premier Healthcare Database
Johns Hopkins University • Johns Hopkins University
National University of Singapore • National University of Singapore
Northeastern • IQVIA Pharmetrics Plus
Organization Name • Data Source Name
Taipei Medical University • Taipei Medical University
Tufts University Medical Center • Tufts University Medical Center
University of Nebraska Medical Center • University of Nebraska Medical Center
University of Southern California • Keck Medical Center
US Department of Veteran's Affairs • US Department of Veteran's Affairs
Yinzhou Bigdata Platform • Yinzhou Bigdata Platform

On March 28, 2023, the OHDSI Global Community initiated the Save Our Sisyphus (SOS) Challenge, a groundbreaking opportunity for collaborative research involving simultaneous participation in four different network studies. What made it truly remarkable was that any organization interested in joining the OHDSI Evidence Network could contribute to these studies by sharing their database profiles for the data sources they had access to. These profiles were centrally aggregated at the OHDSI Central Coordinating Center, enabling us to empirically determine which of the four study questions each data source was best suited to address. This inaugural OHDSI Evidence Network endeavor encompassed 36 diverse adata sources from 16 different organizations. Not only did this foster rapid evidence generation and collaboration during the SOS Challenge, but it also positioned us for future collaborations on additional network studies as part of the OHDSI Evidence Network.

Join The OHDSI Evidence Network

If you are interested in becoming a part of the OHDSI Evidence Network and contributing to advancing evidence-based healthcare, please use the provided QR code to complete a brief form about your organization and your data source. A member of the OHDSI Network Data Quality Working Group will reach out to you to explore this exciting opportunity further!

HADES Development Updates: SelfControlledCaseSeries 5.1.1

SelfControlledCaseSeries 5.1.1 | 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).
- Optionally use regularization on all covariates except the outcome of interest.

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

Transforming the Optum® Enriched Oncology module to OMOP CDM

(Dmitry Dymshyts, Clair Blacketer)

Title: Transforming the Optum® Enriched Oncology module to OMOP CDM

PRESENTER: **Dmitry Dymshyts**

INTRO

- The Optum® Enriched Oncology Data set is a group of tables that can supplement the Optum® de-identified Electronic Health Record dataset.
- It includes specific oncology concepts important for understanding the progression of the disease, which often not available in structured formats, particularly the tumor, node, and metastasis (TNM) values, stage information and biomarkers.
- As of 2022, there are approximately 1.9 million patients with at least one solid tumor ICD-9 or ICD-10 diagnosis included in the data set.

METHODS

1. The data mapping was done with help of the OHDSI tools White Rabbit and Rabbit-in-a-hat*.
2. The concept mapping was done by a semantic analysis of source concepts and with help of OHDSI Usagi tool**. Data and concepts mapping is based on the OHDSI Oncology Working group guidelines***.
3. The data cleansing was done by removing clinically impossible events, for example, the same cancer being "in situ" and "invasive" at the same day.

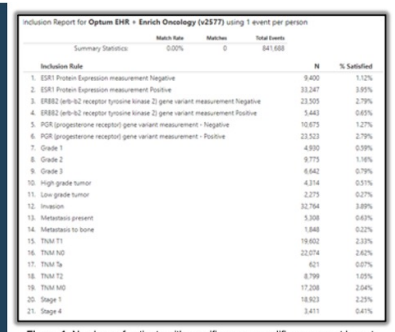
RESULTS

- The Optum EHR data now contains important characteristics of neoplastic disorders: staging, grading, histology, behavior, TNM stage, genetic markers, tumor size, etc. See example in **Figure 1**.
- This enables observational research on a very specific groups of patients or detect a specific outcome. This is important in development of novel targeted therapies and better understanding of how existing drugs affect patients.

* https://github.com/OHDSI/WhiteRabbit/
** https://github.com/OHDSI/usagi/
*** https://ohdsi.github.io/OncologyWG/

The OHDSI Standardized Vocabularies are robust enough to support transformation of oncology data, including staging, grading, histology and genetic markers of neoplastic disorders.

More information available by the QR code



Inclusion Report for **Optum EHR + Enrich Oncology (v2577)** using 1 event per person

	Match Rate	Matches	Total Events
Summary Statistics	0.00%	0	841,688

	Inclusion Rule	N	% Satisfied
1.	ESR1 Protein Expression measurement Negative	9,400	1.12%
2.	ESR1 Protein Expression measurement Positive	33,247	3.95%
3.	ERBB2 (erb-b2 receptor tyrosine kinase 2) gene variant measurement Negative	23,505	2.79%
4.	ERBB2 (erb-b2 receptor tyrosine kinase 2) gene variant measurement Positive	5,443	0.65%
5.	PGR (progesterone receptor) gene variant measurement - Negative	10,673	1.27%
6.	PGR (progesterone receptor) gene variant measurement - Positive	23,523	2.79%
7.	Grade 1	4,930	0.59%
8.	Grade 2	9,775	1.16%
9.	Grade 3	6,642	0.79%
10.	High grade tumor	4,314	0.51%
11.	Low grade tumor	2,275	0.27%
12.	Invasion	32,764	3.89%
13.	Metastasis present	5,308	0.63%
14.	Metastasis to bone	1,848	0.22%
15.	TNM T1	19,602	2.33%
16.	TNM N0	22,074	2.62%
17.	TNM Ta	621	0.07%
18.	TNM T2	8,799	1.05%
19.	TNM M0	17,208	2.04%
20.	Stage 1	18,923	2.25%
21.	Stage 4	3,411	0.41%

Figure 1. Numbers of patients with specific cancer modifiers amongst breast cancer patients

Highlights

- **Data mapping** is straightforward and source data fits well into the OMOP CDM in general.

- **Concept mapping**:
The following categories of concepts were mapped
  - Conditions: combination of histology, topography and behavior (adenocarcinoma of breast in situ)
  - Biomarkers (erb-b2 receptor tyrosine kinase 2 (ERBB2 or HER2/neu))
  - TNM stage
  - Other cancer modifiers: grade, metastases location, summary stage, advanced/localized, evaluation systems (Binet Stage, Durie/Salmon Stage, ECOG performance status, FIGO Stage, Gleason, Gleason score), tumor size, etc.

- **Relatively small number of patients have cancer modifiers documented** (see Figure 1):
only about 4% of patients have Estrogen receptor status with result documented.
But the proportion between Positive and Negative statuses is the same as reported in the literature; same applicable for the ERBB2 and Progesterone receptor measurements.

Limitations/needs for future development:

- **Treatment response:**
Good / excellent / complete / partial / complete pathologic / etc. – No corresponding concepts in the Episode domain.
- **Tumor progression** will be mapped to the Episode table in the next iteration
- **Line of therapy** to be mapped to the HemOnc vocabulary
- **Data cleansing algorithms** to be developed

Dmitry Dymshyts, Janssen
Clair Blacketer
Erasmus MC
OHDSI

#OHDSISocialShowcase This Week

TUESDAY

Community Contribution to the OHDSI Vocabularies: moving towards collaborative shared resource

(Oleg Zhuk, Anna Ostropolets, Alexander Davydov, Christian Reich)

Community Contribution to the OHDSI Vocabularies: moving towards collaborative shared resource

PRESENTER: **Oleg Zhuk**

BACKGROUND

OHDSI Vocabularies as of the beginning of 2023 encompass more than 130 vocabularies that are imported, manipulated and released by the OHDSI Vocabulary Team, which is a part of the OHDSI CDM Working Group.

METHODS

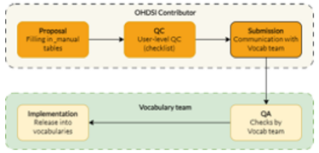
We split the use cases of community contribution into two groups: relatively simple, such as the addition or modification of relationships, and more complex, such as the modification of standard vocabularies and hierarchies, and then developed system to support them. The first group of use cases is now fully supported. We are currently working with contributors to support second group.

RESULTS

We incorporated **four contributions in August 2023 Vocabularies release**: two new non-standard vocabularies, new RxNorm Extension concepts to support international drug schemas and changes to ICD10CM-SNOMED mapping. We have **12** more contributions on the way.

The vocabulary team continues with additional QA and incorporates proposed changes into the vocabulary ecosystem (Figure 1).

Figure 1. Community contribution dataflow



Scan QR to go to Community Contribution Guidelines

Fix your own Vocabulary:

Templates and guidelines let you

- **add** your source vocabulary
- **add** missing concepts
- **fix** domains, concept names and mappings

QC checklists help you don't do anything wrong

Talk to us if you have more complex use cases!

Published guidelines cover the following use cases (Table 1).

Table 1. Supported contribution use cases

#	Type
T1	Adding new non-standard concept(s) to an existing vocabulary
T2	Adding new synonym(s) to an existing concept(s)
T3	Adding a mapping to an existing concept
T4	Adding a new vocabulary as non-standard with mappings (full or partial) to a standard vocabulary
T5	Modifying attributes of an existing concept(s)
T6	Modifying mapping for an existing concept
T7	Promoting non-standard concepts to standard

Depending on the use case, the contributor needs to submit the following tables: *concept_manual, concept_relationship_manual, concept_synonym_manual, metadata, checklist*

The Vocabulary team stores metadata about the contribution. Their format is compatible with SSSOM. **Date of submission, license status and contact details (email, name, organisation)** are collected per contribution and other parameters are collected per relationship (Table 2).

Table 2. Metadata collected per relationship

confidence	confidence in the new mapping (0-1)
predicate_id	type of matching (exactMatch, narrowMatch, broadMatch)
mapping_source	if applicable
mapping_justification	how the mapping was arrived at: ManualMappingCuration, LexicalMatching, etc.
mapping_tool	if applicable

We are collecting more complex use cases to develop infrastructure enabling *shared vocabulary development* by other community members (***stewards***). Talk to us if you want to support a vocabulary!

Oleg Zhuk,
Anna Ostropolets,
Alexander Davydov,
Christian Reich

ODYSSEUS

OHDSI

#OHDSISocialShowcase This Week

WEDNESDAY

Evaluating confounding adjustment when sample size is small

(Martijn Schuemie, Marc A. Suchard, Akihiko Nishimura, Linying Zhang, George Hripcsak)

Evaluating confounding adjustment when sample size is small

Martijn Schuemie[1,2], Marc A. Suchard[2], Akihiko Nishimura[3], Linying Zhang[4], George Hripcsak[4]

[1] Observational Health Data Analytics, Johnson & Johnson, [2] Department of Biostatistics, University of California, Los Angeles, [3] Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, [4] Department of Biomedical Informatics, Columbia University Medical Center

Read the abstract

Background

Observational studies estimating causal effects are vulnerable to confounding because groups receiving different treatments may differ in important aspects. OHDSI studies typically rely on large-scale propensity score (LSPS) models to adjust for these differences. When treatment groups are sufficiently large, LSPS has proven to work well, both in terms of covariate balance and residual systematic error measured using negative controls. However, little is known about LSPS's ability to adjust for confounding when treatment groups are small. To complicate matters, prior research shows that our ability to measure covariate balance — using the standardized difference of means (SDM) — degrades when sample size is limited.

Methods

To measure performance of LSPS under small sample sizes, we take a large study population and randomly divide it into smaller partitions to simulate different data sites, as shown in **Figure 1**.

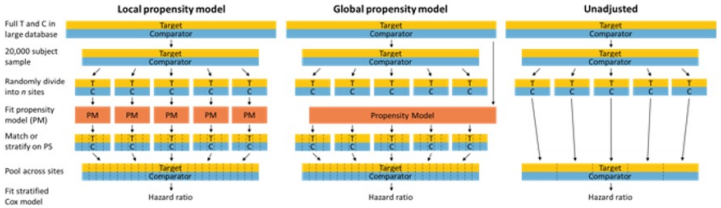


Figure 1. Simulating small data sites. We extract a target (T) and comparator (C) cohort from a large database and take a 20,000-person random sample. We then randomly divide these into *n* equally-sized sites. We evaluate propensity score adjustment using propensity models (PM) fitted at each simulated site (Local) or using a single PM fitted on the original full data (Global), and compare this to no propensity-score adjustment (Unadjusted). Data is pooled across simulated sites before fitting a stratified Cox model.

Ground truth

- Lisinopril vs hydrochlorothiazide (HCTZ), with 76 negative controls*
- Lisinopril vs metoprolol, with 76 negative controls*
- Sitagliptin vs glimepiride, with 94 negative controls**
- Sitagliptin vs liraglutide, with 94 negative controls**

* From LEGEND-HTN
** From LEGEND-T2DM

Databases

- Merative MarketScan MDCD
- Metarive MarketScan MDCR
- Optum® de-identified Electronic Health Record dataset (Optum EHR).

Metrics

- Expected Absolute Systematic Error (EASE) is computed by first fitting a Gaussian distribution to the estimated negative control hazard ratios, and then taking the expected absolute value of that distribution.
- Maximum standardized difference of mean (SDM) is computed by dividing the difference between the mean in T and C by the standard deviation for each covariate and taking the maximum of the absolute value.

Sampled sites

- 5 sites of 4,000 persons
- 10 sites of 2,000 persons
- 20 sites of 1,000 persons
- 40 sites of 500 persons
- 80 sites of 250 persons
- 160 sites of 125 persons

Contact: schuemie@ohdsi.org

Results

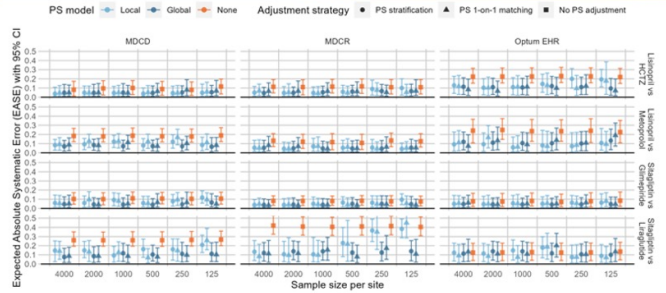


Figure 2. Expected Absolute Systematic Error (EASE) with 95% credible intervals per sample size.

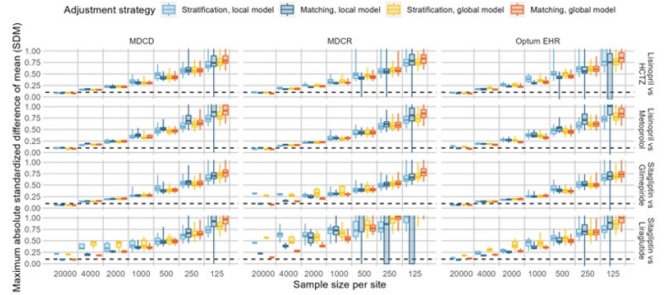


Figure 3. Maximum absolute SDM per sample size. Max SDM is computed at each site, resulting in a distribution characterized by box plots. A max SDM below 0.1 (dashed line) is considered to indicate balance.

Conclusions

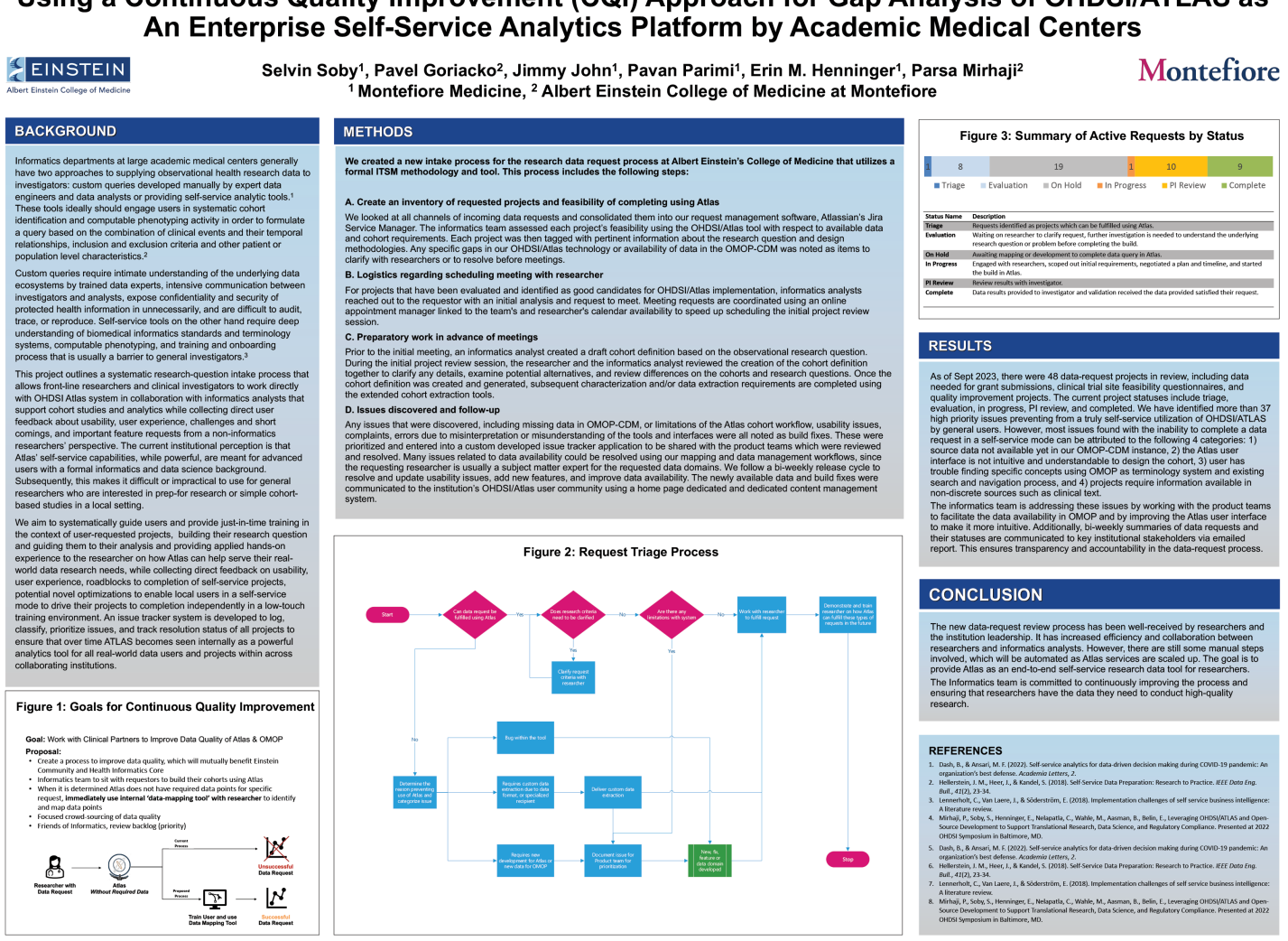
- **Several target-comparator-database combinations already show little confounding in the unadjusted analyses** as measured by EASE. Here, locally-fitted propensity models did not make systematic error worse but also had no opportunity to improve.
- When confounding was detected in the unadjusted analysis, **LSPS was able to adjust for confounding at all but the smallest sample sizes**. No breakdown in performance as measured by EASE was observed when sample size >= 1,000. In many cases, sample size >= 250 was sufficient.
- Even though no confounding was observed (after adjustment) in most situations, max SDM always suggested large imbalance, meaning **our balance metric does not function when sample size is small** (*n* <= 4,000).

#OHDSISocialShowcase This Week

THURSDAY

Using a Continuous Quality Improvement (CQI) Approach for Gap Analysis of OHDSI/ATLAS as An Enterprise Self-Service Analytics Platform by Academic Medical Centers

(Selvin Soby, Pavel Goriacko, Jimmy John, Pavan Parimi, Erin M. Henninger, Parsa Mirhaji)



Using a Continuous Quality Improvement (CQI) Approach for Gap Analysis of OHDSI/ATLAS as An Enterprise Self-Service Analytics Platform by Academic Medical Centers

EINSTEIN Albert Einstein College of Medicine

Selvin Soby[1], Pavel Goriacko[2], Jimmy John[1], Pavan Parimi[1], Erin M. Henninger[1], Parsa Mirhaji[2]
[1] Montefiore Medicine, [2] Albert Einstein College of Medicine at Montefiore

Montefiore

BACKGROUND

Informatics departments at large academic medical centers generally have two approaches to supplying observational health research data to investigators: custom queries developed manually by expert data engineers and data analysts or providing self-service analytic tools.[1] These tools ideally should engage users in systematic cohort identification and computable phenotyping activity in order to formulate a query based on the combination of clinical events and their temporal relationships, inclusion and exclusion criteria and other patient or population level characteristics.[2]

Custom queries require intimate understanding of the underlying data ecosystems by trained data experts, intensive communication between investigators and analysts, expose confidentiality and security of protected health information in unnecessarily, and are difficult to audit, trace, or reproduce. Self-service tools on the other hand require deep understanding of biomedical informatics standards and terminology systems, computable phenotyping, and training and onboarding process that is usually a barrier to general investigators.[3]

This project outlines a systematic research-question intake process that allows front-line researchers and clinical investigators to work directly with OHDSI Atlas system in collaboration with informatics analysts that support cohort studies and analytics while collecting direct user feedback about usability, user experience, challenges and short comings, and important feature requests from a non-informatics researchers' perspective. The current institutional perception is that Atlas' self-service capabilities, while powerful, are meant for advanced users with a formal informatics and data science background. Subsequently, this makes it difficult or impractical to use for general researchers who are interested in prep-for research or simple cohort-based studies in a local setting.

We aim to systematically guide users and provide just-in-time training in the context of user-requested projects, building their research question and guiding them to their analysis and providing applied hands-on experience to the researcher on how Atlas can help serve their real-world data research needs, while collecting direct feedback on usability, user experience, roadblocks to completion of self-service projects, potential novel optimizations to enable local users in a self-service mode to drive their projects to completion independently in a low-touch training environment. An issue tracker system is developed to log, classify, prioritize issues, and track resolution status of all projects to ensure that over time ATLAS becomes seen internally as a powerful analytics tool for all real-world data users and projects within across collaborating institutions.

Figure 1: Goals for Continuous Quality Improvement

Goal: Work with Clinical Partners to Improve Data Quality of Atlas & OMOP

Proposal:

- Create a process to improve data quality, which will mutually benefit Einstein Community and Health Informatics Core
- Informatics team to sit with requestors to build their cohorts using Atlas
- When it is determined Atlas does not have required data points for specific request, **immediately use internal 'data-mapping tool' with researcher** to identify and map data points
- Focused crowd-sourcing of data quality
- Friends of Informatics, review backlog (priority)

METHODS

We created a new intake process for the research data request process at Albert Einstein's College of Medicine that utilizes a formal ITSM methodology and tool. This process includes the following steps:

A. Create an inventory of requested projects and feasibility of completing using Atlas

We looked at all channels of incoming data requests and consolidated them into our request management software, Atlassian's Jira Service Manager. The informatics team assessed each project's feasibility using the OHDSI/Atlas tool with respect to available data and cohort requirements. Each project was then tagged with pertinent information about the research question and design methodologies. Any specific gaps in our OHDSI/Atlas technology or availability of data in the OMOP-CDM was noted as items to clarify with researchers or to resolve before meetings.

B. Logistics regarding scheduling meeting with researcher

For projects that have been evaluated and identified as good candidates for OHDSI/Atlas implementation, informatics analysts reached out to the requestor with an initial analysis and request to meet. Meeting requests are coordinated using an online appointment manager linked to the team's and researcher's calendar availability to speed up scheduling the initial project review session.

C. Preparatory work in advance of meetings

Prior to the initial meeting, an informatics analyst created a draft cohort definition based on the observational research question. During the initial project review session, the researcher and the informatics analyst reviewed the creation of the cohort definition together to clarify any details, examine potential alternatives, and review differences on the cohorts and research questions. Once the cohort definition was created and generated, subsequent characterization and/or data extraction requirements are completed using the extended cohort extraction tools.

D. Issues discovered and follow-up

Any issues that were discovered, including missing data in OMOP-CDM, or limitations of the Atlas cohort workflow, usability issues, complaints, errors due to misinterpretation or misunderstanding of the tools and interfaces were all noted as build fixes. These were prioritized and entered into a custom developed issue tracker application to be shared with the product teams which were reviewed and resolved. Many issues related to data availability could be resolved using our mapping and data management workflows, since the requesting researcher is usually a subject matter expert for the requested data domains. We follow a bi-weekly release cycle to resolve and update usability issues, add new features, and improve data availability. The newly available data and build fixes were communicated to the institution's OHDSI/Atlas user community using a home page dedicated and dedicated content management system.

Figure 2: Request Triage Process

Figure 3: Summary of Active Requests by Status

Status Name	Description
Triage	Requests identified as projects which can be fulfilled using Atlas.
Evaluation	Waiting on researcher to clarify request, further investigation is needed to understand the underlying research question or problem before completing the build.
On Hold	Awaiting mapping or development to complete data query in Atlas.
In Progress	Engaged with researchers, scoped out initial requirements, negotiated a plan and timeline, and started the build in Atlas.
PI Review	Review results with investigator.
Complete	Data results provided to investigator and validation received the data provided satisfied their request.

RESULTS

As of Sept 2023, there were 48 data-request projects in review, including data needed for grant submissions, clinical trial site feasibility questionnaires, and quality improvement projects. The current project statuses include triage, evaluation, in progress, PI review, and completed. We have identified more than 37 high priority issues preventing from a truly self-service utilization of OHDSI/ATLAS by general users. However, most issues found with the inability to complete a data request in a self-service mode can be attributed to the following 4 categories: 1) source data not available yet in our OMOP-CDM instance, 2) the Atlas user interface is not intuitive and understandable to design the cohort, 3) user has trouble finding specific concepts using OMOP as terminology system and existing search and navigation process, and 4) projects require information available in non-discrete sources such as clinical text.

The informatics team is addressing these issues by working with the product teams to facilitate the data availability in OMOP and by improving the Atlas user interface to make it more intuitive. Additionally, bi-weekly summaries of data requests and their statuses are communicated to key institutional stakeholders via emailed report. This ensures transparency and accountability in the data-request process.

CONCLUSION

The new data-request review process has been well-received by researchers and the institution leadership. It has increased efficiency and collaboration between researchers and informatics analysts. However, there are still some manual steps involved, which will be automated as Atlas services are scaled up. The goal is to provide Atlas as an end-to-end self-service research data tool for researchers.

The Informatics team is committed to continuously improving the process and ensuring that researchers have the data they need to conduct high-quality research.

REFERENCES

1. Dash, B., & Ansari, M. F. (2022). Self-service analytics for data-driven decision making during COVID-19 pandemic: An organization's best defense. *Academia Letters*, 2.
2. Hellerstein, J. M., Heer, J., & Kandel, S. (2018). Self-Service Data Preparation: Research to Practice. *IEEE Data Eng. Bull.*, *41*(2), 23-34.
3. Lennerholt, C., Van Laere, J., & Söderström, E. (2018). Implementation challenges of self service business intelligence: A literature review.
4. Mirhaji, P., Soby, S., Henninger, E., Nelapatla, C., Wahle, M., Aasman, B., Belin, E., Leveraging OHDSI/ATLAS and Open-Source Development to Support Translational Research, Data Science, and Regulatory Compliance. Presented at 2022 OHDSI Symposium in Baltimore, MD.
5. Dash, B., & Ansari, M. F. (2022). Self-service analytics for data-driven decision making during COVID-19 pandemic: An organization's best defense. *Academia Letters*, 2.
6. Hellerstein, J. M., Heer, J., & Kandel, S. (2018). Self-Service Data Preparation: Research to Practice. *IEEE Data Eng. Bull.*, *41*(2), 23-34.
7. Lennerholt, C., Van Laere, J., & Söderström, E. (2018). Implementation challenges of self service business intelligence: A literature review.
8. Mirhaji, P., Soby, S., Henninger, E., Nelapatla, C., Wahle, M., Aasman, B., Belin, E., Leveraging OHDSI/ATLAS and Open-Source Development to Support Translational Research, Data Science, and Regulatory Compliance. Presented at 2022 OHDSI Symposium in Baltimore, MD.

#OHDSISocialShowcase This Week

FRIDAY

Risk of aortic aneurysm or dissection following exposure to fluoroquinolone antibiotics

(Jack L. Janetzki, Jung Ho Kim, Jung Ah Lee, Seng Chan You, Nicole L. Pratt)

Risk of aortic aneurysm or dissection following exposure to fluoroquinolone antibiotics

PRESENTER: **Seng Chan You**

INTRODUCTION

- Fluoroquinolone (FQ) antibiotics are a broad-spectrum class of antibiotics. Whilst FQ antibiotics are generally well tolerated, a meta-analysis of four observational studies found that FQ antibiotics were associated with an increased risk of aortic diseases compared to other antibiotics (adjusted odds ratio 2.10; 95% CI 1.65-2.68).
- However, the quality of this study was rated as moderate and the results from later studies were conflicted.
- We aimed to determine whether exposure to FQ are associated with aortic aneurysm or dissections after initiating treatment

METHODS

- This study is being conducted through international collaboration through OHDSI network in Save Our Sisyphus (SOS) challenge.
- The study utilised data from January 2010 to December 2019.
- We used a new-user comparative cohort design including all patients diagnosed with urinary tract infection (UTI) and initiating either treatment with systemic FQ or an active comparator of either trimethoprim (+/- sulfamethoxazole) (TMP) cephalosporin (CPH) in outpatient setting.
- Primary outcome is a composite outcomes of hospitalization due to aortic aneurysm or aortic dissection within 60 days.
- The suite of diagnostics is employed to evaluate balance, empirical equipoise, and systematic bias. Only results passing the predefined suite of diagnostics are used for the meta-analysis.

Studies identifying **comparative risk of aortic diseases following fluoroquinolone** exposure from a **single data source or single health system** may be **vulnerable to bias** as there is **substantial heterogeneity** in preference of treatment

Comparative risk of aortic dissection or aneurysm in fluoroquinolone users versus trimethoprim or cephalosporin users. Shades are proportional to overlap of preference score distributions.

Source	Matched(n)	FQ	TMP	HR (95% CI)	Max SDM	PS overlap	EASE	Diagnostics
CUMC(US)	3996	<7.69	7.69	0.45(0.08-2.58)	0.12	0.92	0.15	PASS
IBM CCAE(US)	505839	0.58	0.53	1.07(0.70-1.63)	0.03	0.94	0.04	PASS
IBM MDCD(US)	88791	2.87	2.09	1.36(0.84-2.19)	0.03	0.88	0.03	PASS
Optum DOD(US)	410842	3.75	4.53	0.83(0.68-1.01)	0.05	0.93	0.04	PASS
Optum EHR(US)	478507	2.23	2.34	0.92(0.73-1.15)	0.05	0.91	0.06	PASS
PharMetrics(US)	358621	1.12	0.96	1.15(0.80-1.66)	0.03	0.93	0.03	PASS
VA(US)	108202	4.46	6.40	0.67(0.50-0.91)	0.04	0.86	0.05	PASS
TMUDB(TW)	2328	<13.60	0.00	NA(NA- NA)	0.14	0.50	0.50	FAIL
AUSOM(KR)	61	0.00	0.00	NA(NA- NA)	0.57	0.62	NA	FAIL
NHIS-NSC(KR)	442	0.00	0.00	NA(NA- NA)	0.30	0.77	NA	FAIL
Yonsei(KR)	251	0.00	0.00	NA(NA- NA)	0.33	0.40	NA	FAIL
JMDC(JP)	722	0.00	0.00	NA(NA- NA)	0.21	0.48	NA	FAIL
Japan Claims(JP)	1106	0.00	0.00	NA(NA- NA)	0.20	0.36	NA	FAIL
LPD Australia(AU)	947	0.00	0.00	NA(NA- NA)	0.16	0.55	NA	FAIL

Source	Matched(n)	FQ	CPH	HR (95% CI)	Max SDM	PS overlap	EASE	Diagnostics
CUMC(US)	4867	6.31	8.84	0.73(0.22- 2.42)	0.09	0.58	0.14	PASS
IBM CCAE(US)	222632	1.01	0.85	1.16(0.71- 1.91)	0.08	0.43	0.05	PASS
IBM MDCD(US)	67180	5.19	3.52	1.45(0.96- 2.21)	0.03	0.43	0.04	PASS
Optum DOD(US)	274869	6.29	7.03	0.90(0.72- 1.13)	0.06	0.47	0.07	PASS
Optum EHR(US)	338470	4.10	4.50	0.89(0.73- 1.09)	0.09	0.52	0.04	PASS
PharMetrics(US)	211877	2.41	1.84	1.29(0.92- 1.81)	0.06	0.44	0.03	PASS
VA(US)	68155	9.25	10.46	0.86(0.65- 1.13)	0.06	0.47	0.05	PASS
TMUDB(TW)	7912	<4.08	4.05	0.77(0.19- 3.08)	0.10	0.69	0.20	PASS
AUSOM(KR)	528	0.00	0.00	NA(NA- NA)	0.30	0.99	NA	FAIL
NHIS-NSC(KR)	5073	<6.18	<6.17	1.60(0.11-24.44)	0.13	0.94	0.35	FAIL
Yonsei(KR)	3581	<8.90	<8.97	0.80(0.03-20.72)	0.12	0.91	0.47	FAIL
JMDC(JP)	9569	<3.26	<3.26	0.48(0.03- 7.48)	0.07	0.89	0.27	FAIL
Japan Claims(JP)	15060	<2.06	0.00	NA(NA- NA)	0.07	0.56	0.13	PASS
LPD Australia(AU)	889	0.00	0.00	NA(NA- NA)	0.18	0.58	NA	FAIL

RESULTS

- Among 14 databases, results from 7 and 9 databases passed diagnostics for comparison with TMP and CPH, respectively.
- Overlap of preference score (PS) distributions across databases is heterogeneous across data sources. In results form US, more PS overlap is identified in comparison of FQ with TMP than comparison with CPH, while vice versa in results from non-US data sources.
- In meta-analysis using only results passing diagnostics, the hazard ratio of primary outcome was 0.91 (95%CI: 0.73-1.15) in comparison with TMP and 1.01 (95% CI: 0.82-1.25), indicating no different risk of primary outcome between groups.

CONCLUSION

- We found that use of fluoroquinolone in patients with UTI did not result in different risk of following aortic dissection or aneurysm within 60 days compared with use of TMP or CPH.
- Studies identifying comparative risk of aortic diseases following fluoroquinolone exposure from a single data source or single health system may be vulnerable to bias as there is substantial heterogeneity in preference of treatment.

Jack L. Janetzki[1], Jung Ho Kim[2], Jung Ah Lee[2], Seng Chan You[2], Nicole L. Pratt[1]

[1] University of South Australia
[2] Yonsei University College of Medicine

University of South Australia OHDSI

(I²DB)

Postdoc/Senior Data Analyst Opening at WashU

The Zhang Lab at Washington University School of Medicine in St. Louis has **one postdoct/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

Washington University School of Medicine in St. Louis

Opening: Epidemiology UX/Web Design Intern at J&J

Career Programs

Epidemiology UX/Web Design Intern

JOB TITLE	Epidemiology UX/Web Design Intern
FUNCTION	Career Programs
SUB FUNCTION	Non-LDP Intern/Co-Op
LOCATION	Raritan, New Jersey, United States
DATE POSTED	Jan 19 2024
REQUISITION NUMBER	2406163977W

DESCRIPTION

Janssen Research & Development, L.L.C., a division of Johnson & Johnson's Family of Companies is recruiting for Epidemiology UX/Web Design Intern. This position is a member of the Observational Health Data Analytics (OHDA) team. OHDA's mission is to improve the lives individuals and quality of healthcare by efficiently generating real-world evidence from the world's observational health data, transparently disseminating evidence-based insights to real-world decision-makers, and objectively advancing the science and technology behind reliab

Apply Now

Opening: Research Information Specialist at UNC

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

- Home
- Search Jobs
- Careers At Carolina Notifications
- Log In /Create Account
- Help
- Working at Carolina

Research Informatics Specialist

Bookmark this Posting | Print Preview | Apply for this Job

Please see Special Instructions for more details.

Working hours are Monday-Friday, 8:00 am – 6:00 pm EST with flexibility available within that window.

Posting Information

Posting Information

Department	TraCS Institute-429801
Career Area	Information Technology
Posting Open Date	12/13/2023
Application Deadline	01/30/2024
Open Until Filled	No
Position Type	Permanent Staff (EHRA NF)
Working Title	Research Informatics Specialist
Appointment Type	EHRA Non-Faculty
Position Number	20060002
Vacancy ID	NF0007640
Full Time/Part Time	Full-Time Permanent
FTE	1

Position Summary

Responsibilities include:
* Perform SQL-based programming against UNC's clinical data warehouse to identify patient cohorts and develop patient datasets.
* Consult with and collaborate with researchers to ensure programming work aligns with project needs.
* Develop ETL (extract, transform, and load) and data integration processes to support common data models (OMOP, PCORnet) using appropriate technologies (SQL, Python, or R).
* Carefully following UNC's regulatory and governance policy to ensure data integrity and security.
* In collaboration with IDSci team, identify potential enhancements in current workflows and data architecture.
* Implement quality assurance strategies, such as data validation and peer code review.
* Write and maintain up-to-date supporting documentation. Ensure code is well-commented and use GitLab/GitHub to manage code changes and track data lineage.
* Provide technical leadership and direction for assigned projects and/or data requests.

Minimum Education and Experience Requirements

Master's and 1-2 years' experience; or Bachelors and 2-4 years' experience; or will accept a combination of related education and experience in substitution.

Required Qualifications, Competencies, and Experience

This position requires two or more years of relevant work experience and:

* Expert-level knowledge of SQL programming, data modeling, and relational database systems such as Oracle, Microsoft SQL Server, MySQL, etc.
* Past experience working with health care data in an analytic capacity, particularly electronic health record and/or claims data.
* Demonstrable past experience in scoping technical projects in terms of length of time, competencies and cost. Individual will be expected to manage multiple projects at once while delivering high-quality work on time.
* Excellent written and oral business communication skills. Public speaking at meetings and conferences may be required. The ability to clearly convey technical concepts to non-technical clients is a must.

Opening: Data Steward at EBMD

Description

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

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

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

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

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

RESPONSIBILITIES AND TASKS

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

Where Are We Going?

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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Jan 30: Introduction to Phenotype Phebruary

Azza Shoaibi

Director, Observational Health Data Analytics
Janssen Research and Development

Anna Ostropolets

Associate Director, Observational Health Data Analytics
Janssen Research and Development

Jamie Weaver

Associate Director, Observational Health Data Analytics
Janssen Research and Development

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