Research Breakouts for OHDSI2025 Showcase

OHDSI Community Call
May 27, 2025 • 11 am ET

Upcoming Community Calls

Date	Topic
May 27	Collaborator Showcase Brainstorm (Deadline is July 1)
June 3	The Journey of ATLAS
June 10	ATLAS Deepdive: Data Sources and Vocabularies
June 17	ATLAS Deepdive: Cohorts and Conceptsets
June 24	ATLAS Deepdive: Characterization, Cohort Pathways, Incidence
July 1	ATLAS Deepdive: Technical and Administrative Capabilities
July 8	No Meeting – Europe Symposium

Three Stages of The Journey

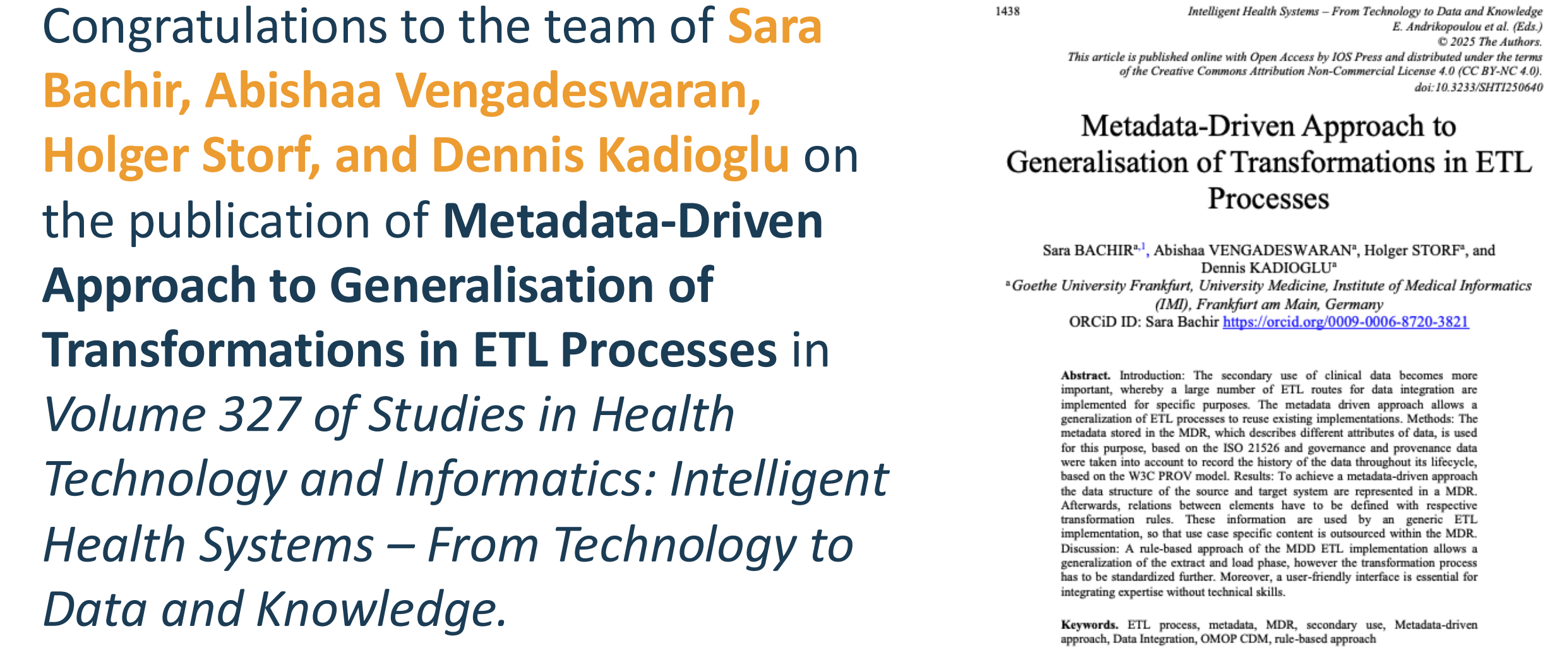
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Sara Bachir, Abishaa Vengadeswaran, Holger Storf, and Dennis Kadioglu** on the publication of **Metadata-Driven Approach to Generalisation of Transformations in ETL Processes** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*



Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250640

Metadata-Driven Approach to Generalisation of Transformations in ETL Processes

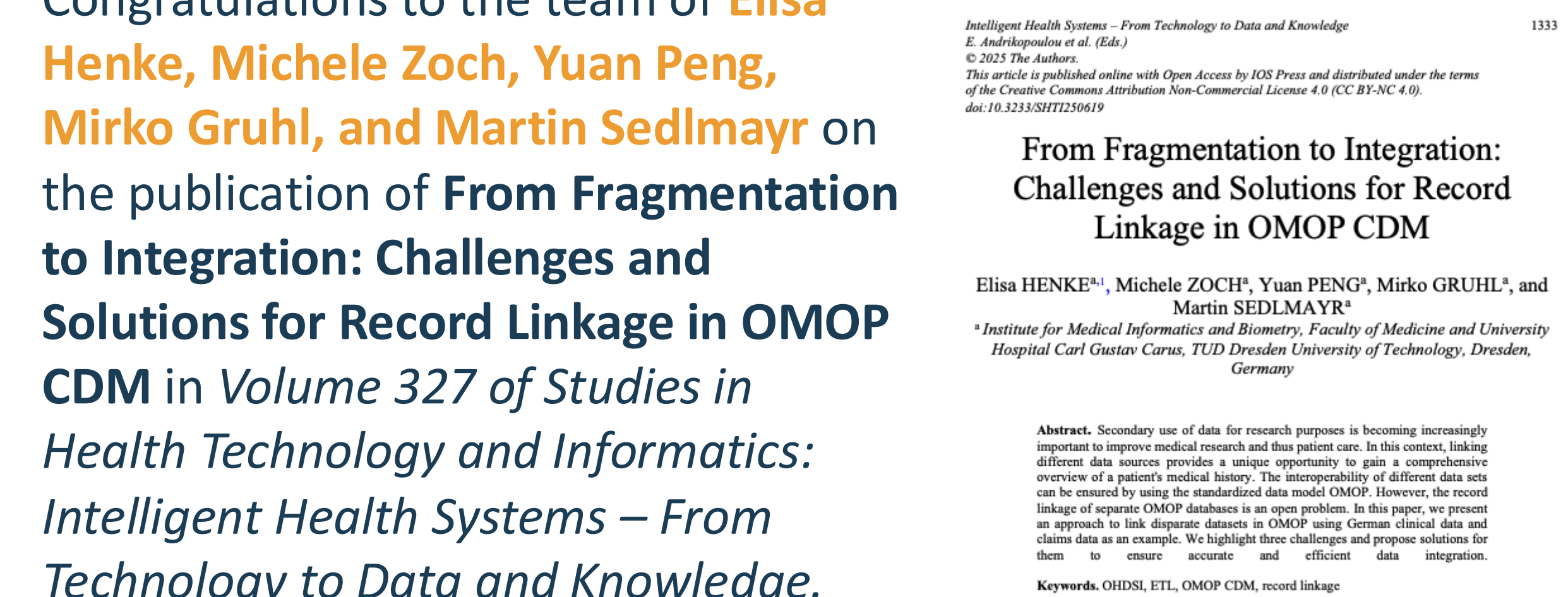
Sara BACHIR[a,1], Abishaa VENGADESWARAN[a], Holger STORF[a], and Dennis KADIOGLU[a]
[a]*Goethe University Frankfurt, University Medicine, Institute of Medical Informatics (IMI), Frankfurt am Main, Germany*
ORCiD ID: Sara Bachir https://orcid.org/0009-0006-8720-3821

Abstract. Introduction: The secondary use of clinical data becomes more important, whereby a large number of ETL routes for data integration are implemented for specific purposes. The metadata driven approach allows a generalization of ETL processes to reuse existing implementations. Methods: The metadata stored in the MDR, which describes different attributes of data, is used for this purpose, based on the ISO 21526 and governance and provenance data were taken into account to record the history of the data throughout its lifecycle, based on the W3C PROV model. Results: To achieve a metadata-driven approach the data structure of the source and target system are represented in a MDR. Afterwards, relations between elements have to be defined with respective transformation rules. These information are used by an generic ETL implementation, so that use case specific content is outsourced within the MDR. Discussion: A rule-based approach of the MDD ETL implementation allows a generalization of the extract and load phase, however the transformation process has to be standardized further. Moreover, a user-friendly interface is essential for integrating expertise without technical skills.

Keywords. ETL process, metadata, MDR, secondary use, Metadata-driven approach, Data Integration, OMOP CDM, rule-based approach

OHDSI Shoutouts!

Congratulations to the team of **Elisa Henke, Michele Zoch, Yuan Peng, Mirko Gruhl, and Martin Sedlmayr** on the publication of **From Fragmentation to Integration: Challenges and Solutions for Record Linkage in OMOP CDM** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*



Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250619

From Fragmentation to Integration: Challenges and Solutions for Record Linkage in OMOP CDM

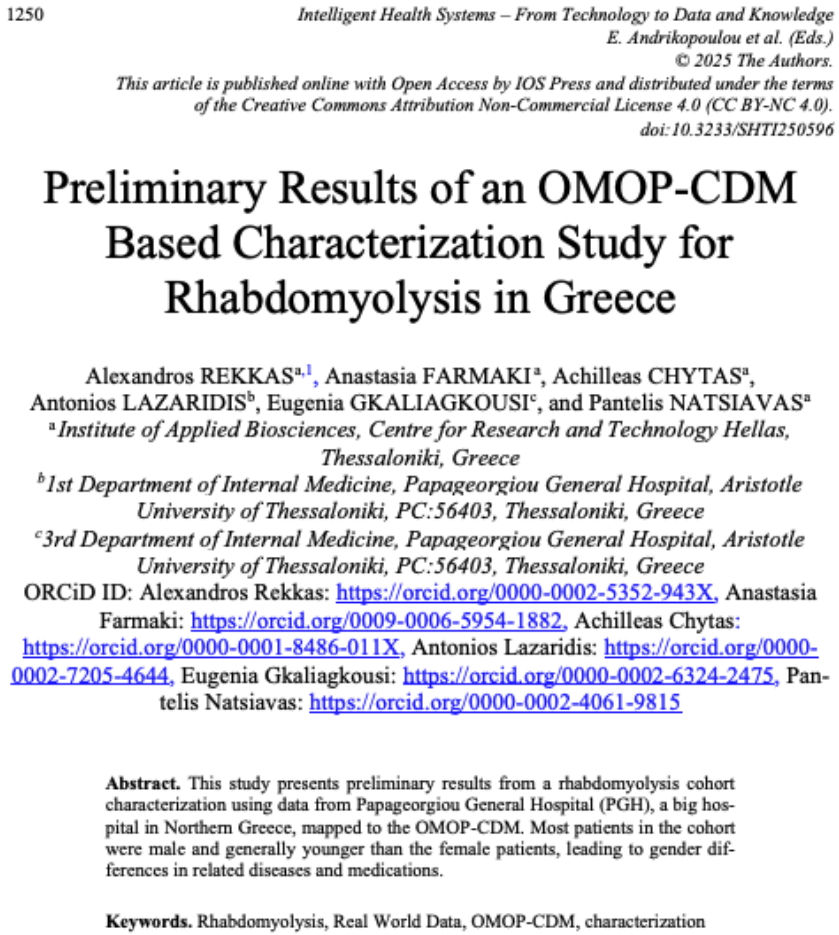
Elisa HENKE[a,1], Michele ZOCH[a], Yuan PENG[a], Mirko GRUHL[a], and Martin SEDLMAYR[a]
[a] *Institute for Medical Informatics and Biometry, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany*

Abstract. Secondary use of data for research purposes is becoming increasingly important to improve medical research and thus patient care. In this context, linking different data sources provides a unique opportunity to gain a comprehensive overview of a patient's medical history. The interoperability of different data sets can be ensured by using the standardized data model OMOP. However, the record linkage of separate OMOP databases is an open problem. In this paper, we present an approach to link disparate datasets in OMOP using German clinical data and claims data as an example. We highlight three challenges and propose solutions for them to ensure accurate and efficient data integration.

Keywords. OHDSI, ETL, OMOP CDM, record linkage

OHDSI Shoutouts!

Congratulations to the team of **Alexandros Rekkas, Anastasia Farmaki, Achilleas Chytas, Antonios Lazaridis, Eugenia Gkaliagkousi, Pantelis Natsiavas** on the publication of **Preliminary Results of an OMOP-CDM Based Characterization Study for Rhabdomyolysis in Greece** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*



doi:10.3233/SHTI250596

Preliminary Results of an OMOP-CDM Based Characterization Study for Rhabdomyolysis in Greece

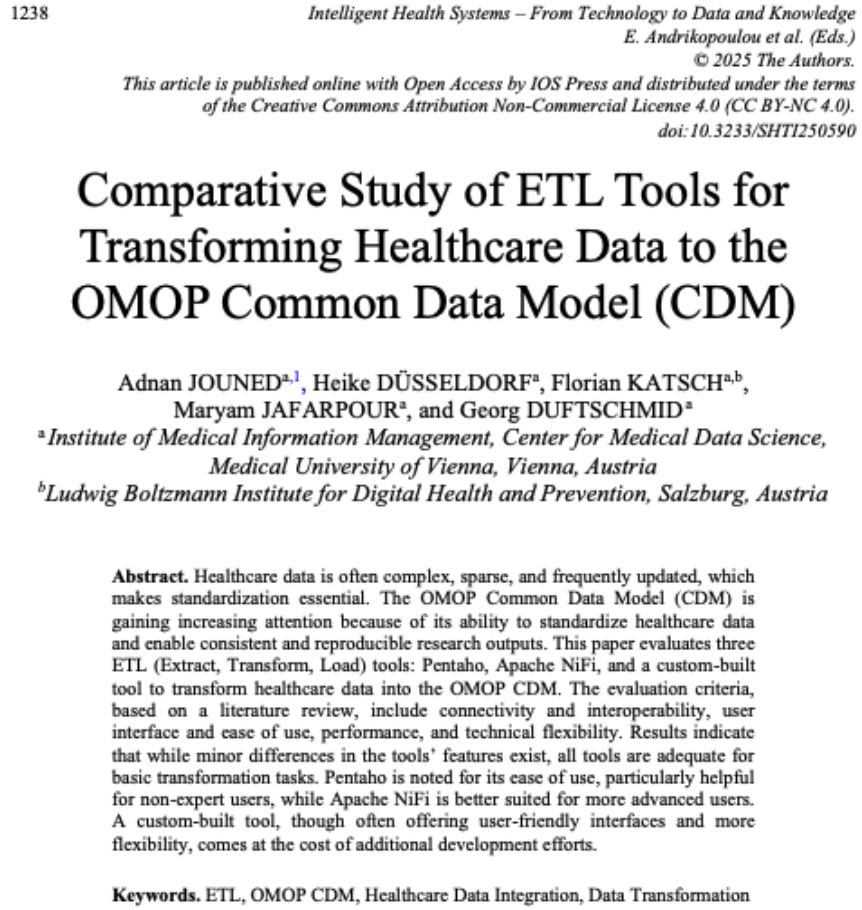
Alexandros REKKAS[a,1], Anastasia FARMAKI[a], Achilleas CHYTAS[a], Antonios LAZARIDIS[b], Eugenia GKALIAGKOUSI[c], and Pantelis NATSIAVAS[a]
[a]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece
[b]1st Department of Internal Medicine, Papageorgiou General Hospital, Aristotle University of Thessaloniki, PC:56403, Thessaloniki, Greece
[c]3rd Department of Internal Medicine, Papageorgiou General Hospital, Aristotle University of Thessaloniki, PC:56403, Thessaloniki, Greece
ORCiD ID: Alexandros Rekkas: https://orcid.org/0000-0002-5352-943X, Anastasia Farmaki: https://orcid.org/0009-0006-5954-1882, Achilleas Chytas: https://orcid.org/0000-0001-8486-011X, Antonios Lazaridis: https://orcid.org/0000-0002-7205-4644, Eugenia Gkaliagkousi: https://orcid.org/0000-0002-6324-2475, Pantelis Natsiavas: https://orcid.org/0000-0002-4061-9815

Abstract. This study presents preliminary results from a rhabdomyolysis cohort characterization using data from Papageorgiou General Hospital (PGH), a big hospital in Northern Greece, mapped to the OMOP-CDM. Most patients in the cohort were male and generally younger than the female patients, leading to gender differences in related diseases and medications.

Keywords. Rhabdomyolysis, Real World Data, OMOP-CDM, characterization

OHDSI Shoutouts!

Congratulations to the team of **Adnan Jouned, Heike Düsseldorf, Florian Katsch, Maryam Jafarpour, and Georg Duftschmid** on the publication of **Comparative Study of ETL Tools for Transforming Healthcare Data to the OMOP Common Data Model (CDM)** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*



doi:10.3233/SHTI250590

Comparative Study of ETL Tools for Transforming Healthcare Data to the OMOP Common Data Model (CDM)

Adnan JOUNED[a,1], Heike DÜSSELDORF[a], Florian KATSCH[a,b], Maryam JAFARPOUR[a], and Georg DUFTSCHMID[a]
[a] *Institute of Medical Information Management, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria*
[b] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*

Abstract. Healthcare data is often complex, sparse, and frequently updated, which makes standardization essential. The OMOP Common Data Model (CDM) is gaining increasing attention because of its ability to standardize healthcare data and enable consistent and reproducible research outputs. This paper evaluates three ETL (Extract, Transform, Load) tools: Pentaho, Apache NiFi, and a custom-built tool to transform healthcare data into the OMOP CDM. The evaluation criteria, based on a literature review, include connectivity and interoperability, user interface and ease of use, performance, and technical flexibility. Results indicate that while minor differences in the tools' features exist, all tools are adequate for basic transformation tasks. Pentaho is noted for its ease of use, particularly helpful for non-expert users, while Apache NiFi is better suited for more advanced users. A custom-built tool, though often offering user-friendly interfaces and more flexibility, comes at the cost of additional development efforts.

Keywords. ETL, OMOP CDM, Healthcare Data Integration, Data Transformation

OHDSI Shoutouts!

Congratulations to the team of **Achim Michel-Backofen, Romina Blasini, Jördis Beck, and Kurt Marquardt** on the publication of **Building a Research Infrastructure with REDCap and FHIR** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*

Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250457

Building a Research Infrastructure with REDCap and FHIR

Achim MICHEL-BACKOFEN [ab], Romina BLASINI [a], Jördis BECK [a], and Kurt MARQUARDT [b]
[a] *Data Integration Center, Institute of Medical Informatics, Justus Liebig University, Giessen, Germany*
[b] *University Hospital of Giessen, Giessen, Germany*

Abstract. At the university hospital of Giessen since 2018 a research infrastructure has been implemented as part of the German Medical Informatics Initiative. A central FHIR repository acts as the core part of this architecture where clinical data from various subsystems is consolidated. In the past, research projects were connected via FHIR or research databases like i2b2 or OMOP to this core dataset. REDCap was added as an electronic data capture platform especially for local research projects that need to combine data entry together with the available dataset on the central FHIR repository. Meanwhile this infrastructure has been successfully used within 3 research projects in internal medicine and psychosomatic medicine.

Keywords. FHIR, REDCap, ETL, electronic data capture

OHDSI Shoutouts!

Congratulations to the team of **Prabath Jayathissa, Lukas Rohatsch, Stefan Sauermann, and Rada Hussein** on the publication of **OMOP-on-FHIR: Integrating the Clinical Data Through FHIR Bundle to OMOP CDM** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*

Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250432

OMOP-on-FHIR: Integrating the Clinical Data Through FHIR Bundle to OMOP CDM

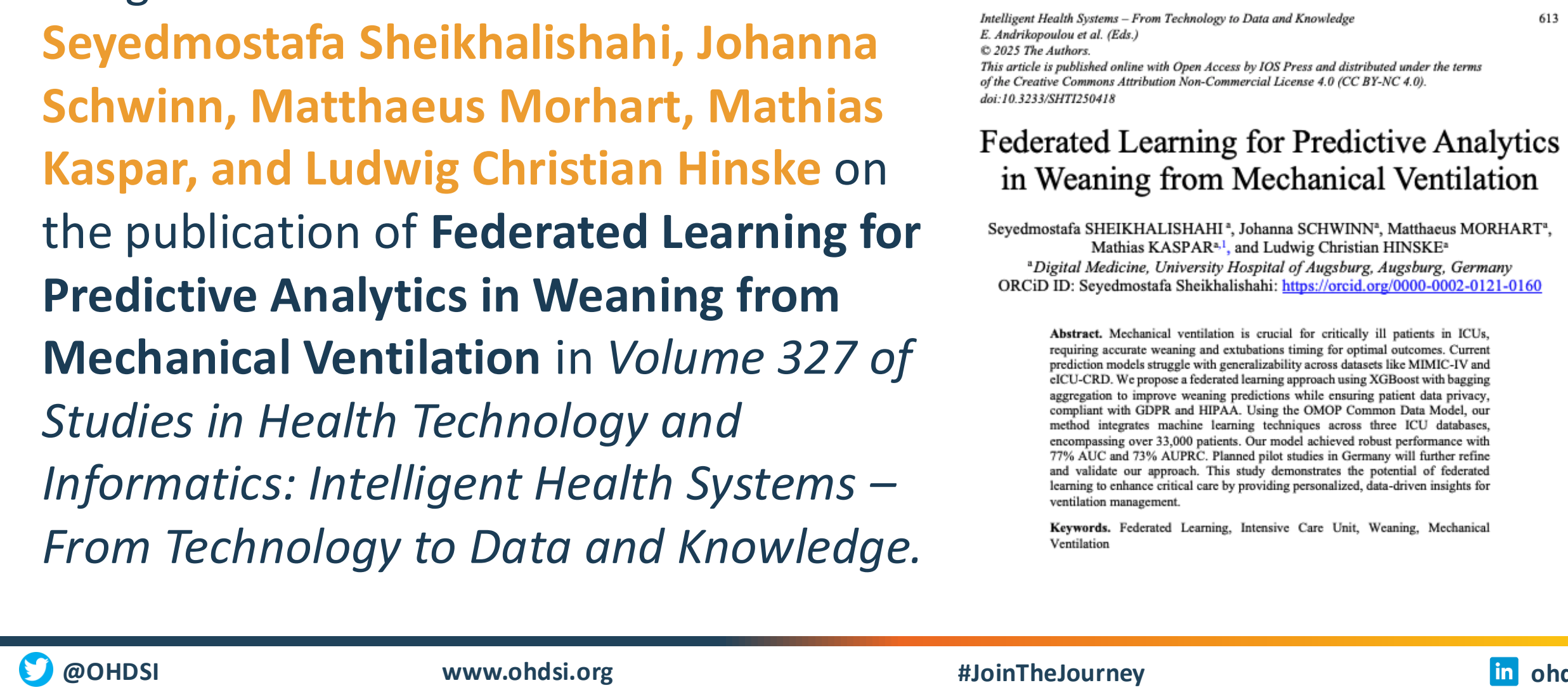
Prabath JAYATHISSA[a,1], Lukas ROHATSCH[b], Stefan SAUERMANN[b], and Rada HUSSEIN[a]
[a]*Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
[b]*University of Applied Sciences/FH-Technikum Wien, Vienna, Austria*
ORCiD ID: Prabath Jayathissa https://orcid.org/0000-0003-3056-503X, Lukas ROHATSCH https://orcid.org/0000-0002-6168-2717, Stefan SAUERMANN https://orcid.org/0000-0003-0824-9989,Rada Hussein https://orcid.org/0000-0003-1257-4848

Abstract. The harmonization of the OMOP Common Data Model (CDM) with HL7 FHIR aims to enhance interoperability in clinical research by harmonizing diverse healthcare datasets. This process, referred to as OMOP-on-FHIR, leverages FHIR Bundles for real-time clinical data exchange and transforms these resources into OMOP CDM format using an ETL process. The ETL pipeline, facilitated by tools like XSLT, enables the extraction, transformation, and loading of data while maintaining semantic consistency. By bridging these two standards, OMOP-on-FHIR promotes the seamless exchange of data across clinical systems and research-oriented databases, supporting global health studies, advanced analytics, and personalized medicine. This methodology advances cross-border research by providing a standardized approach to data management and analysis, thereby improving healthcare outcomes.

Keywords. OMOP, ETL Pipeline, FHIR, Clinical Research, Healthcare Interoperability

OHDSI Shoutouts!

Congratulations to the team of **Seyedmostafa Sheikhalishahi, Johanna Schwinn, Matthaeus Morhart, Mathias Kaspar, and Ludwig Christian Hinske** on the publication of **Federated Learning for Predictive Analytics in Weaning from Mechanical Ventilation** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*



Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250418

Federated Learning for Predictive Analytics in Weaning from Mechanical Ventilation

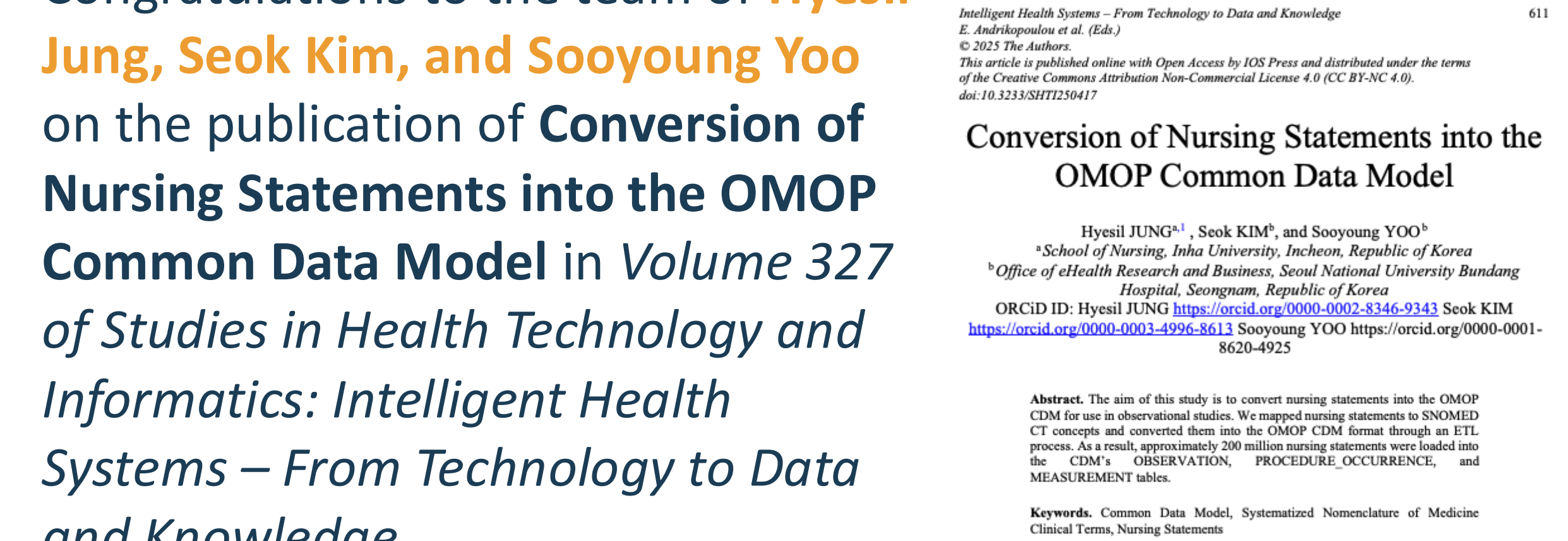
Seyedmostafa SHEIKHALISHAHI [a], Johanna SCHWINN[a], Matthaeus MORHART[a], Mathias KASPAR[a,1], and Ludwig Christian HINSKE[a]
[a] *Digital Medicine, University Hospital of Augsburg, Augsburg, Germany*
ORCiD ID: Seyedmostafa Sheikhalishahi: https://orcid.org/0000-0002-0121-0160

Abstract. Mechanical ventilation is crucial for critically ill patients in ICUs, requiring accurate weaning and extubations timing for optimal outcomes. Current prediction models struggle with generalizability across datasets like MIMIC-IV and eICU-CRD. We propose a federated learning approach using XGBoost with bagging aggregation to improve weaning predictions while ensuring patient data privacy, compliant with GDPR and HIPAA. Using the OMOP Common Data Model, our method integrates machine learning techniques across three ICU databases, encompassing over 33,000 patients. Our model achieved robust performance with 77% AUC and 73% AUPRC. Planned pilot studies in Germany will further refine and validate our approach. This study demonstrates the potential of federated learning to enhance critical care by providing personalized, data-driven insights for ventilation management.

Keywords. Federated Learning, Intensive Care Unit, Weaning, Mechanical Ventilation

OHDSI Shoutouts!

Congratulations to the team of **Hyesil Jung, Seok Kim, and Sooyoung Yoo** on the publication of **Conversion of Nursing Statements into the OMOP Common Data Model** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*



Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250417

Conversion of Nursing Statements into the OMOP Common Data Model

Hyesil JUNG[a,1], Seok KIM[b], and Sooyoung YOO[b]
[a]*School of Nursing, Inha University, Incheon, Republic of Korea*
[b]*Office of eHealth Research and Business, Seoul National University Bundang Hospital, Seongnam, Republic of Korea*
ORCiD ID: Hyesil JUNG https://orcid.org/0000-0002-8346-9343 Seok KIM https://orcid.org/0000-0003-4996-8613 Sooyoung YOO https://orcid.org/0000-0001-8620-4925

Abstract. The aim of this study is to convert nursing statements into the OMOP CDM for use in observational studies. We mapped nursing statements to SNOMED CT concepts and converted them into the OMOP CDM format through an ETL process. As a result, approximately 200 million nursing statements were loaded into the CDM's OBSERVATION, PROCEDURE_OCCURRENCE, and MEASUREMENT tables.

Keywords. Common Data Model, Systematized Nomenclature of Medicine Clinical Terms, Nursing Statements

OHDSI Shoutouts!

Congratulations to the team of **Johanna Schwinn, Seyedmostafa Sheikhalishahi, Matthaeus Morhart, Mathias Kaspar, and Ludwig Christian Hinske** on the publication of **A Federated Learning Model for the Prediction of Blood Transfusion in Intensive Care Units** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*

Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250311

A Federated Learning Model for the Prediction of Blood Transfusion in Intensive Care Units

Johanna SCHWINN[a,1], Seyedmostafa SHEIKHALISHAHI[a], Matthaeus MORHART[a], Mathias KASPAR[a], and Ludwig Christian HINSKE[a]
[a]*Digital Medicine, University Hospital of Augsburg, Augsburg, Germany*
ORCiD ID: Johanna Schwinn https://orcid.org/0009-0000-3107-1619

Abstract. Accurate prediction of blood transfusion requirements is crucial for patient outcomes and resource management in clinical settings. We developed a machine learning model using XGBoost to predict the need for a blood transfusion 2 hours in advance based on up to 7 hours of prior data from two large databases, MIMIC-IV and eICU-CRD. Our federated model showed promising results, with F1 scores of 0.72 and 0.66, respectively.

Keywords. Federated Learning, ICU, Flower framework, blood transfusion, OMOP CDM

OHDSI Shoutouts!

Congratulations to the team of **Angela Leis, Philippe Mortier, Franco Amigo, Madhav Bhargav, Susana Conde, Montserrat Ferrer, Oskar Flygare, Busenur Kizilaslan, Laura Latorre Moreno, Miguel-Angel Mayer, Víctor Pérez Sola, Ana Portillo van Diest, Juan Manuel Ramírez-Anguita, Ferran Sanz, Gemma Vilagut, Jordi Alonso, Lars Mehlum, Ella Arensman, Johan Bjureberg, Manuel Pastor, and Ping Qin** on the publication of **Machine Learning-Based Clinical Decision Support System for Suicide Risk Management: The PERMANENS Project** in *Volume 327 of Studies in Health Technology and Informatics: Intelligent Health Systems – From Technology to Data and Knowledge.*

Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250308

Machine Learning-Based Clinical Decision Support System for Suicide Risk Management: The PERMANENS Project

Angela LEIS[a,b,1], Philippe MORTIER[a,c], Franco AMIGO[a,c], Madhav BHARGAV[d], Susana CONDE[a], Montserrat FERRER[a,c], Oskar FLYGARE[e], Busenur KIZILASLAN[f], Laura LATORRE MORENO[a], Miguel-Angel MAYER[a,b], Víctor PÉREZ SOLA[g,h], Ana PORTILLO VAN DIEST[a,c], Juan Manuel RAMÍREZ-ANGUITA[a,b], Ferran SANZ [b], Gemma VILAGUT[a,c], Jordi ALONSO [a,b,c], Lars MEHLUM[f], Ella ARENSMAN[d], Johan BJUREBERG[e], Manuel PASTOR[b], and Ping QIN[f]
[a] *Hospital del Mar Research Institute, Barcelona, Spain*
[b] *Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona*
[c] *CIBER of Epidemiology and Public Health, Carlos III Health Institute Madrid, Spain*
[d] *School of Public Health & National Suicide Research Foundation, University College Cork, Ireland*
[e] *Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Stockholm, Sweden*
[f] *National Centre for Suicide Research and Prevention, Institute of Clinical Medicine, University of Oslo, Oslo, Norway*
[g] *Neuropsychiatry and Drug Addiction Institute, Barcelona MAR Health Park Consortium PSMAR, Barcelona, Spain*
[h] *CIBER of Mental Health and Carlos III Health Institute, Madrid, Spain*
ORCiD ID: Angela Leis https://orcid.org/0000-0003-4780-7111
Philippe Mortier https://orcid.org/0000-0003-2113-6241

Abstract. The PERMANENS European project addresses the global public health challenge of self-harm and suicide by developing a machine learning-based Clinical Decision Support System (CDSS) to assist emergency departments (EDs) in providing personalized care. With over 700,000 suicides annually, suicide prevention is critical, especially in Europe where consistent surveillance is lacking. The project harmonizes national suicide attempt registries from regions in Spain, Ireland, Norway, and Sweden using the OMOP Common Data Model (CDM) to create a comprehensive database for real-time analysis.

Keywords. Suicide, self-harm, machine learning, clinical support system

OHDSI Shoutouts!

Congratulations to the team of **Jongchan Park, Hye Jin Chang, Kyung Joo Hwang, Sun Hyung Yum, Chang Eun Park, Joo Hee Kim, and Miran Kim** on the publication of **Association of COX-2 Selectivity in Pain Medication Use with Endometriosis Incidence: Retrospective Cohort Study** in the *Yonsei Medical Journal*.

Original Article
Yonsei Med J 2025 Jun;66(6):374-382
https://doi.org/10.3349/ymj.2024.0255
Yonsei Medical Journal YMJ
pISSN: 0513-5796 · eISSN: 1976-2437

Association of COX-2 Selectivity in Pain Medication Use with Endometriosis Incidence: Retrospective Cohort Study

Jongchan Park[1]*, Hye Jin Chang[2]*, Kyung Joo Hwang[2], Sun Hyung Yum[2], Chang Eun Park[1], Joo Hee Kim[3], and Miran Kim[1,2]

[1]Department of Convergence Healthcare Medicine, Graduate School of Ajou University, Suwon;
[2]Department of Obstetrics & Gynecology, Ajou University School of Medicine, Suwon;
[3]Regulatory Strategy Center for Combination Products, Ajou University, Suwon, Korea.

Purpose: This retrospective cohort study aimed to investigate the association between the use of pain medications with varying cyclooxygenase-2 (COX-2) selectivity and the incidence of endometriosis (EMS) in women.
Materials and Methods: Medical records from January 1, 1994, to December 31, 2022, were retrospectively analyzed. The cohort included 33406 patients diagnosed with any pain-related condition who were prescribed either selective COX-2 inhibitors or non-steroidal anti-inflammatory drugs (NSAIDs). Patients were followed for up to 5 years from the cohort entry date. The incidence of EMS was compared between the two medication groups using Cox proportional hazards models, adjusting for confounding factors such as age, past drug use, and prior diagnosis.
Results: The incidence rates of EMS were 3.00 per 1000 person-years in the COX-2 inhibitor group and 3.97 per 1000 person-years in the NSAIDs group. After adjustment for confounders, the hazard ratio for EMS incidence in the COX-2 inhibitor group compared to the NSAIDs group was 0.77 [95% confidence interval (CI), 0.63 to 0.93; $p<0.01$], indicating a significantly lower risk in the COX-2 inhibitor group. Subgroup analysis revealed that this association was particularly significant in younger women aged 20–44 years, with a hazard ratio of 0.71 (95% CI, 0.54 to 0.95; $p<0.05$) in this age group.
Conclusion: The findings suggest that COX-2 inhibitors may reduce the incidence of EMS compared to traditional NSAIDs, highlighting their potential as a strategic option for managing EMS, particularly among younger women. Further prospective studies are needed to confirm these findings.

Key Words: Endometriosis, COX-2, OMOP-CDM, Celecoxib, Ibuprofen, Cohort Study

OHDSI Shoutouts!

Congratulations to the team of **Xinyu Zhou, Lovedeep Singh Dhingra, Arya Aminorroaya, Philip Adejumo, and Rohan Khera** on the publication of **A Novel Sentence Transformer-based Natural Language Processing Approach for Schema Mapping of Electronic Health Records to the OMOP Common Data Model** in the *AMIA Annual Symposium Proceedings Archive.*

A Novel Sentence Transformer-based Natural Language Processing Approach for Schema Mapping of Electronic Health Records to the OMOP Common Data Model

Xinyu Zhou BS[1], Lovedeep Singh Dhingra MBBS[2], Arya Aminorroaya MD, MPH[2], Philip Adejumo BS[2], Rohan Khera MD, MS[1,2,3,4]
[1]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA; [2]Yale School of Medicine, New Haven, CT, USA; [3]Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA; [4]Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA

Abstract
Mapping electronic health records (EHR) data to common data models (CDMs) enables the standardization of clinical records, enhancing interoperability and enabling large-scale, multi-centered clinical investigations. Using 2 large publicly available datasets, we developed transformer-based natural language processing models to map medication-related concepts from the EHR at a large and diverse healthcare system to standard concepts in OMOP CDM. We validated the model outputs against standard concepts manually mapped by clinicians. Our best model reached out-of-box accuracies of 96.5% in mapping the 200 most common drugs and 83.0% in mapping 200 random drugs in the EHR. For these tasks, this model outperformed a state-of-the-art large language model (SFR-Embedding-Mistral, 89.5% and 66.5% in accuracy for the two tasks), a widely used software for schema mapping (Usagi, 90.0% and 70.0% in accuracy), and direct string match (7.5% and 7.5% accuracy). Transformer-based deep learning models outperform existing approaches in the standardized mapping of EHR elements and can facilitate an end-to-end automated EHR transformation pipeline.

Introduction
Data standards, such as the Observational Medical Outcomes Partnership (OMOP) common data model (CDM), play a crucial role in enabling collaboration across diverse health systems by providing a uniform data standard for organizing the electronic health record (EHR) (1-5). However, transforming EHR data to the standardized CDMs remains challenging. For instance, a key challenge is the semantic mapping of the EHR elements to their equivalent standard concepts in the CDM. These free-form text elements are often represented in multiple ways in the EHR, limiting the possibility of a one-to-one string matching-based system, which is commonly used in mapping structured elements. Moreover, EHR elements such as drugs present with frequent variations in dosage and frequency, making mapping to the corresponding standardized concepts even more challenging.

OHDSI Shoutouts!

Congratulations to the team of **Yuanzhen Yue, Ashok Khanal, Tianchu Lyu, Sharon Weissman, and Chen Liang** on the publication of **EHR Phenotyping Methods for Measuring Treatment Adherence Among People Living With HIV in All of Us: Towards Disparities and Inequalities in HIV Care Continuum** in the *AMIA Annual Symposium Proceedings Archive.*

EHR Phenotyping Methods for Measuring Treatment Adherence Among People Living With HIV in All of Us: Towards Disparities and Inequalities in HIV Care Continuum

Yuanzhen Yue, MS[1#], Ashok Khanal, BPharm[1#], Tianchu Lyu, MPH[1*], Sharon Weissman, MD[2,3], Chen Liang, PhD[1,4,5]

[1]University of South Carolina Arnold School of Public Health, Columbia, South Carolina, USA; [2]University of South Carolina School of Medicine, Columbia, South Carolina, USA; [3]Prisma Health, Columbia, South Carolina, USA; [4]University of Washington, School of Medicine, Department of Biomedical Informatics and Medical Education, Seattle, Washington [5]National Institutes of Health, Bethesda, Washington DC

[#] Equally contributed.
[*] Corresponding author: TLYU@email.sc.edu

Abstract

HIV treatment adherence is among the most important determinants of HIV outcomes. However, only 50% of people living with HIV in the US were retained in care. Measuring HIV treatment adherence in the clinical settings is feasible but when it comes to the growing number of multi-site Electronic Health Records (EHR), there has been a dearth of research for adequate informatics methods to handle EHR. We sought to address this gap by developing a cluster of metrics for measuring HIV treatment adherence via EHR phenotyping methods. Our methods were developed and tested in the All of Us research program. We also performed preliminary analyses to explore disparities in HIV treatment adherence and demographic factors contributing to poor adherence. This study paves the way for systematic data mining and analyses for the HIV care continuum, disparities, and inequality research on All of Us and other EHR normalized with the OMOP Common Data Model.

Introduction

Advances in antiretroviral therapy (ART), including greater effectiveness and access, have significantly extended the lifespan and improved the quality of life for people living with HIV (PLWH) [1]. As a result, treatment adherence – encompassing linkage to care, retention in care, and adherence to ART – has emerged as the most crucial determinant of treatment outcomes. According to the Centers for Disease Control and Prevention (CDC), among 1.2 million PLWH in the United States at the end of 2019, an estimated 87% were diagnosed, 66% have been initiated with HIV medical care, yet only 50% were retained in care [2]. CDC defines retention rate as the percentage of PLWH with diagnosis who had two or more CD4 or viral load tests, performed at least three months apart. Disparities of the HIV medical care retention rate persist and are significantly associated with sex, race/ethnicity, HIV infection risk factors, and geographical locations [3].

OHDSI Shoutouts!

Congratulations to the team of **Xiaojin Li, Yan Huang, Licong Cui, Shiqiang Tao, and Guo-Qiang Zhang** on the publication of **Optimizing Medication Querying Using Ontology-Driven Approach with OMOP: with an application to a large-scale COVID-19 EHR dataset** in the *AMIA Annual Symposium Proceedings Archive.*

Optimizing Medication Querying Using Ontology-Driven Approach with OMOP: with an application to a large-scale COVID-19 EHR dataset

Xiaojin Li[1,3⋆], Yan Huang[1,3], Licong Cui[2,3], Shiqiang Tao[1,3], Guo-Qiang Zhang[1,2,3⋆]
[1]McGovern Medical School, [2]McWilliams School of Biomedical Informatics, [3]Texas Institute for Restorative Neurotechnologies
The University of Texas Health Science Center at Houston, Houston, Texas 77030, USA

Abstract

Efficient querying for medication information in Electronic Health Record (EHR) datasets is crucial for effective patient care and clinical research. To address the complexity and data volume challenges involved in efficient medication information retrieval, we propose an ontology-driven medication query (ODMQ) optimization approach, leveraging the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Integrating semantic ontology structures from the OMOP CDM can help enhance query accuracy and efficiency by broadening the scope of relevant medication terms like drug names, National Drug Codes, and generics, resulting in more comprehensive query outcomes than traditional methods. ODMQ significantly reduces manual search time and enhances query capabilities. We validate ODMQ's efficacy using real-world COVID-19 EHR data, demonstrating improved query performance. Through a comprehensive manual review, ODMQ ensures that expanded search terms are relevant to user inputs. It also includes an intuitive query interface and visualizes patient history for result validation and exploration.

1 Introduction

Electronic Health Record (EHR) serve as digital repositories containing comprehensive patient health information, encompassing medical histories, diagnoses, treatments, and medication records [1, 2]. These digitized records have revolutionized healthcare delivery by enabling efficient data storage, retrieval, and sharing among healthcare providers [3, 4]. However, the sheer volume and heterogeneity of EHR datasets present significant challenges in extracting actionable insights, particularly regarding medication-related information [5, 6].

To facilitate standardized data representation and analysis, several tools have been developed, such as Unified Medical Language System (UMLS) Terminology Services, and Systematized Medical Nomenclature for Medicine–Clinical Terminology (SNOMED CT), RxNorm and Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The UMLS Terminology Services, maintained by the United States National Library of Medicine, play a crucial role in facilitating interoperability and standardization in biomedical and healthcare informatics by providing access to a rich and diverse set of standardized terminologies and ontologies [7]. SNOMED CT offers a standardized and scientifically validated framework for representing clinical data collected by healthcare professionals. Its integration into EHR enhances the potential for efficient data utilization and ensures improved documentation quality [8]. The OMOP CDM, managed by the Observational Health Data Sciences and Informatics (OHDSI), provides a standardized framework for organizing and harmonizing disparate healthcare data sources, thereby facilitating interoperability and enabling large-scale analytics across diverse healthcare settings [9] .

Three Stages of The Journey

Where Have We Been?

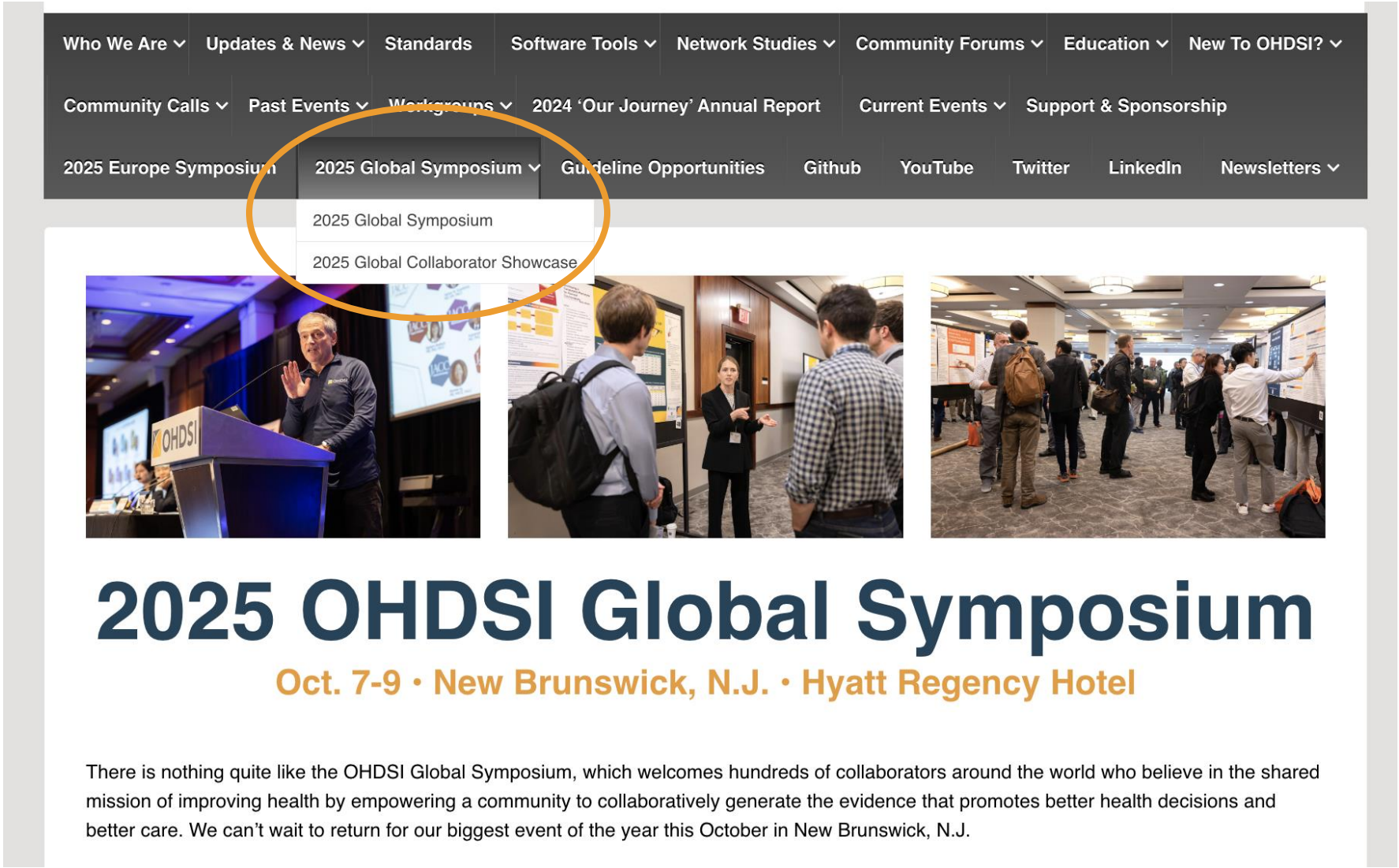
Where Are We Now?

Where Are We Going?

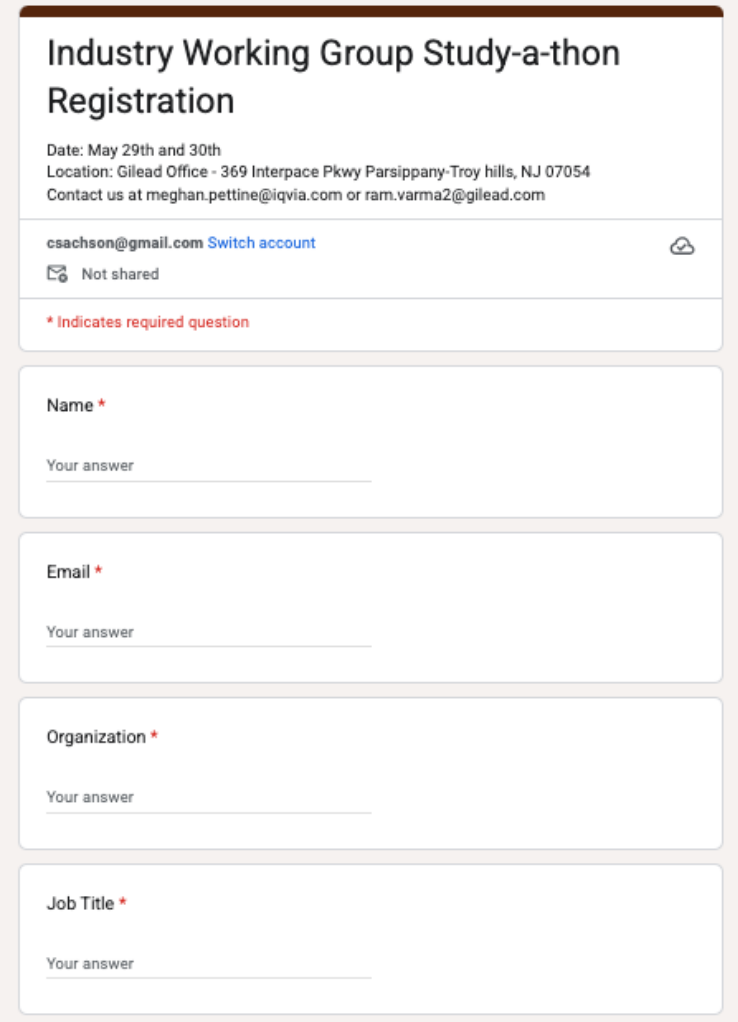
Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	ATLAS
Wednesday	9 am	Oncology Outreach/Research Subgroup
Wednesday	10 am	Surgery and Perioperative Medicine
Wednesday	10 am	Women of OHDSI
Wednesday	12 pm	Latin America
Wednesday	7 pm	Medical Imaging
Thursday	8 am	Medical Devices
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	10 am	Transplant
Friday	11:30 am	Steering
Monday	10 am	Healthcare Systems Interest Group
Tuesday	9:30 am	Common Data Model

Symposium Event Page is Live



Industry Workgroup Studyathon



Industry Working Group Study-a-thon Registration

Date: May 29th and 30th
Location: Gilead Office - 369 Interpace Pkwy Parsippany-Troy hills, NJ 07054
Contact us at meghan.pettine@iqvia.com or ram.varma2@gilead.com

csachson@gmail.com Switch account

Not shared

* Indicates required question

Name *

Your answer

Email *

Your answer

Organization *

Your answer

Job Title *

Your answer

Where: Parsippany, New Jersey

Hosted By: Gilead

The OHDSI Industry Working Group is pleased to announce a two day in-person study-a-thon. The primary objective is to provide stakeholders with a tangible example of the advantages of joining the OHDSI network. These benefits include:

- The OMOP Common Data Model (CDM) aligns with institutional goals and highlights the advantages of a standardized data model.
- Accelerated timelines for conducting network studies.
- A high-quality and user-friendly technology stack.
- Established trust in OMOP vocabularies.
- A comprehensive solution for conducting efficient studies.

This demonstration aims to underscore the value and effectiveness of the OHDSI network, encouraging stakeholders to recognize its potential and benefits.

Interested in learning more? Please join the OHDSI Industry Working Group during our monthly meeting. Past meeting recordings and notes are available on Teams.

ATLAS Usage & Feedback Survey

Atlas Survey

General

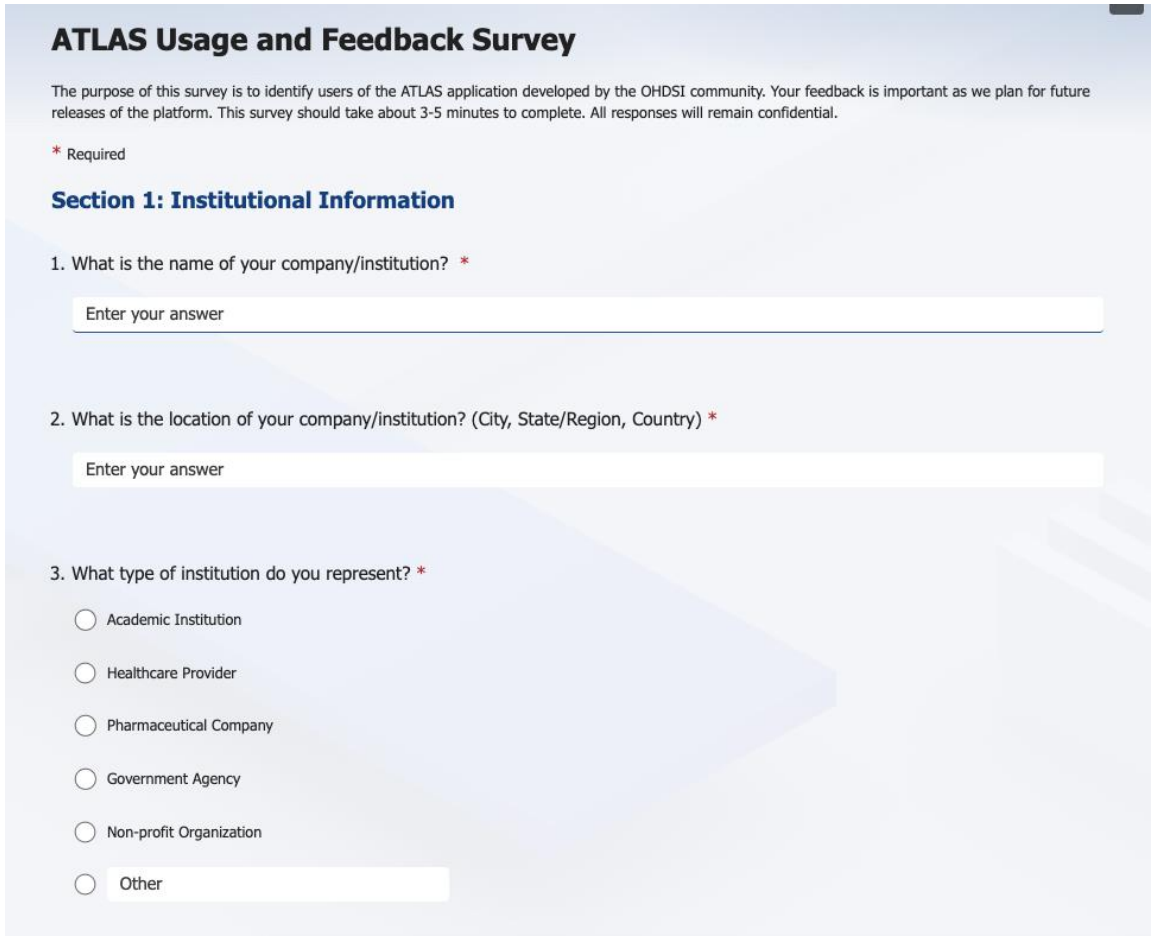
Chris_Knoll 5d

The ATLAS working group has put together a short survey (Microsoft Forms 6) to help us identify who is using ATLAS in our community. If you are not using ATLAS, we'd still ask you to fill in this survey so you can help us to identify any barriers for adoption in your company/institution.

Additionally, this survey will ask if you'd like to be interviewed for feedback on your usage of ATLAS. Data4Life (https://www.data4life.care/ 1) is working closely with our working group to conduct interviews (~1hr) that will help inform the future direction of the application.

We'd appreciate if you could fill out this survey and consider speaking with Data4Life regarding your experiences with ATLAS.

Tagging @anthonysena



ATLAS Usage and Feedback Survey

The purpose of this survey is to identify users of the ATLAS application developed by the OHDSI community. Your feedback is important as we plan for future releases of the platform. This survey should take about 3-5 minutes to complete. All responses will remain confidential.

* Required

Section 1: Institutional Information

1. What is the name of your company/institution? *

Enter your answer

2. What is the location of your company/institution? (City, State/Region, Country) *

Enter your answer

3. What type of institution do you represent? *

- Academic Institution
- Healthcare Provider
- Pharmaceutical Company
- Government Agency
- Non-profit Organization
- Other

Link on community calls page

The Center for Advanced Healthcare Research Informatics (CAHRI) at Tufts Medicine welcomes:

Georgie Kennedy, PhD

Senior Research Fellow, University of New South Wales and the Ingham Institute

'Learning from Real-World Cancer Data: Maturing data pipelines to support research that impacts clinical care'

May 29, 2025, 9-10am EDT

Virtually via Zoom

Please contact Marty Alvarez at malvarez2@tuftsmedicalcenter.org for calendar invite or questions.

TuftsMedicine
Tufts Medical Center

Job Opening

Manager, Observational Health Data Analytics, Johnson & Johnson

Johnson&Johnson

Data Analytics & Computational Sciences | Raritan / United States of America

Manager, Observational Health Data Analytics - Raritan, NJ

Primary responsibilities:

- Work closely with colleagues within GEO.
- Contribute to the successful delivery of observational analyses for clinical characterization, population-level effect estimation, and patient-level prediction to meet the needs of Johnon & Johnson's scientific and business functions.
- Contribute to the design of observational database analysis, including authoring protocol and analysis plans.
- Contribute to the execution of observational database analyses by using standardized analytical tools and writing statistical programs against internal and external observational data resources.
- Contribute to innovating, evaluating, and establishing scientific best practices around the design and conduct of observational analysis and accompanying processes to ensure the reliability of real-world evidence.
- Contribute to the design and development of software and analytical tools that encode scientific best practices into solutions that enable real-world evidence generation and dissemination.
- Contribute to the development and evolution of scientific and industry standards for observational data harmonization, ensuring their appropriate application across the Johnson & Johnson real-world data ecosystem, and leading the evaluation and characterization of observational data for their fitness-for-use to address clinical questions from across the organization.

#OHDSISocialShowcase This Week

Monday

Asian and/or Pacific Islander: Unmasking health disparity within commonly aggregated diverse populations in the US Department of Veterans Affairs

(Benjamin Viernes, Scott L DuVall, Patrick R Alba, Qiwei Gan, Elizabeth E Hanchrow, Mengke Hu, Gregorio Coronado, Andrew M Subica, Curtis Lowery, Scott Hofer, Vicki Shambaugh, Kalani Raphael)

OHDSI

Asian and/or Pacific Islander: Unmasking health disparity within commonly aggregated diverse populations in the US Department of Veterans Affairs

Benjamin Viernes[1,3], Scott L DuVall[1,2], Patrick R Alba[1,2], Qiwei Gan[1,2], Elizabeth E Hanchrow[1], Mengke Hu[1,2], Gregorio Coronado[1,3], Andy Subica, Curtis Lowery[4], Scott Hofer[1,4], Vicki Shambaugh[1,4], Kalani Raphael[1,3]

1 – VA Salt Lake City Health Care System, Salt Lake City, UT, USA
2 – Department of Internal Medicine, University of Utah Medical School, Salt Lake City, UT, USA
3 – Center for Pacific Islander Veterans Health. VA Pacific Islands Healthcare System. Honolulu, HI, USA
4 – Pacific Health Research and Education Institute. Honolulu, HI, USA

Background

- The United States' Government Office of Management and Budget (OMB) first defined the Asian/Pacific Islander (API) race/ethnicity category in 1977
- Despite recommending separating Asian and Native Hawaiian or Pacific Islander (NHPI) categories in 1997, API is still common in US federal data
- The API category masks differences between the diverse patient sub-populations in the US Department of Veterans Affairs (VA)
  - All patients with ancestry in South, Southeast, and East Asia
  - All patients with ancestry in the original peoples of the Pacific islands and Oceania.

Disaggregated sub-categories of Asian/Pacific Islander race/ethnicity

- This analysis sought to describe VA patients with API race/ethnicity and it's NHPI sub-categories.
- Given that NHPI are at higher risk of cardio-metabolic diseases, analyses sought to outline how different categorizations of race/ethnicity for these patients can mask these risks.

Methods

- Using the VA corporate data warehouse (CDW), supplemented with natural language processing of CDW clinical notes, described in a previous OHDSI abstract, API and its NHPI, and NHPI sub-group cohorts were derived.
- VA patients who are US Veterans, with at least 1 visit recorded in the VA OMOP Visit_Occurrence table that could be identified within the categories API, NHPI, Polynesian, Micronesian, or Melanesian were included in the analysis.
- The VA OMOP CDM was used for characterization of the cohorts identified.
  - Demographics, VA observations, measurements, diagnoses, and interactions
  - Specific cardio-metabolic conditions related for which NHPI populations are generally at higher risk.

Asian/Pacific Islander and sub-group identification

- Cardio-metabolic conditions associated with NHPI morbidity that were included are Chronic Kidney Disease (CKD), Cardiovascular Disease (CVD), and Diabetes Mellitus (DM)
  - Phenotype definitions from the Quan (2005) Charlson Comorbidity Index were used
  - Patients with at least 2 recorded diagnoses were determined to have the condition
- Some proportion of API could have been disaggregated to an Asian race/ethnicity category, but that was not the focus of this specific analysis

Results

- Of the 853,654 VA Patients that were identified as API, 21.2% (180,804) could be disaggregated to the level of NHPI.
- Among NHPI, 27% (48,890) could be identified as Polynesian, 9.9% (17,864) as Micronesian, and 1.3% (2,383) as Melanesian.
- A majority of API and NHPI Veterans could not be disaggregated to lower levels using current methods.
- The API population were younger with lower utilization than its NHPI sub-group.
- Rates of CKD, CVD, and DM were about twice as high among NHPI compared to API
  - Among NHPI the highest rates were found among Polynesian patients

	API	NHPI	Polynesian	Micronesian	Melanesian
N*	853,654	180,804	48,890	17,864	2,383
Deceased - N (%)	113,011	49,271	8,768	3,013	230
Age - Mean (SD)	52.74 (16.70)	60.91 (18.84)	59.07 (17.42)	55.71 (16.82)	47.23 (14.66)
Male - N (%)	746,646 (87.46)	158,663 (87.75)	41,515 (84.92)	15175 (84.95)	2,052 (86.11)
Years of VA Observation - Mean (SD)	12.28 (7.57)	13.63 (7.88)	15.43 (7.42)	13.63 (7.38)	11.90 (7.37)
Measurements - Mean (SD)	224.5 (415.8)	316.5 (501.7)	410.5 (590.4)	287.4 (481.6)	211.5 (397.9)
Diagnoses - Mean (SD)	1,097.0 (1751.4)	1,614.0 (2200.6)	1,944.0 (2,455.6)	1,266.0 (1,924.8)	1,060.0 (1,764.6)
VA Interactions - Mean (SD)	175.0 (253.0)	236.0 (300.9)	326.0 (352.8)	232.0 (283.5)	173.0 (257.4)
NHPI-related Conditions - N (%)					
CKD	106,797 (12.5)	43,186 (23.9)	14,647 (30.0)	4,277 (23.9)	263 (11.0)
CVD	118,816 (13.9)	51,887 (28.7)	14,635 (29.9)	4,139 (23.2)	269 (11.3)
DM	185,285 (21.7)	66,285 (36.7)	21,092 (43.1)	6,282 (35.2)	480 (18.5)

* - Patients in any of the sub-populations of API may exist within multiple categories.

Prevalence of Cardio-metabolic diseases among API and NHPI sub-populations

Conclusions

- **NHPI have higher rates of cardio-metabolic diseases.**
- This is not the first analysis to show that grouping NHPI together with API is problematic due to the masking of important disparities.
- Current data does not allow for a complete picture due to the inability to completely identify NHPI race/ethnicity.
- **Further disaggregation is necessary** for standardized data that can accurately report disparities.
- More work is required in both the standardization of granular race/ethnicity categories for API populations, and for novel methods to disaggregate current data sources.

Previous related work

Git Repo:

OHDSI Global Symposium Abstract:

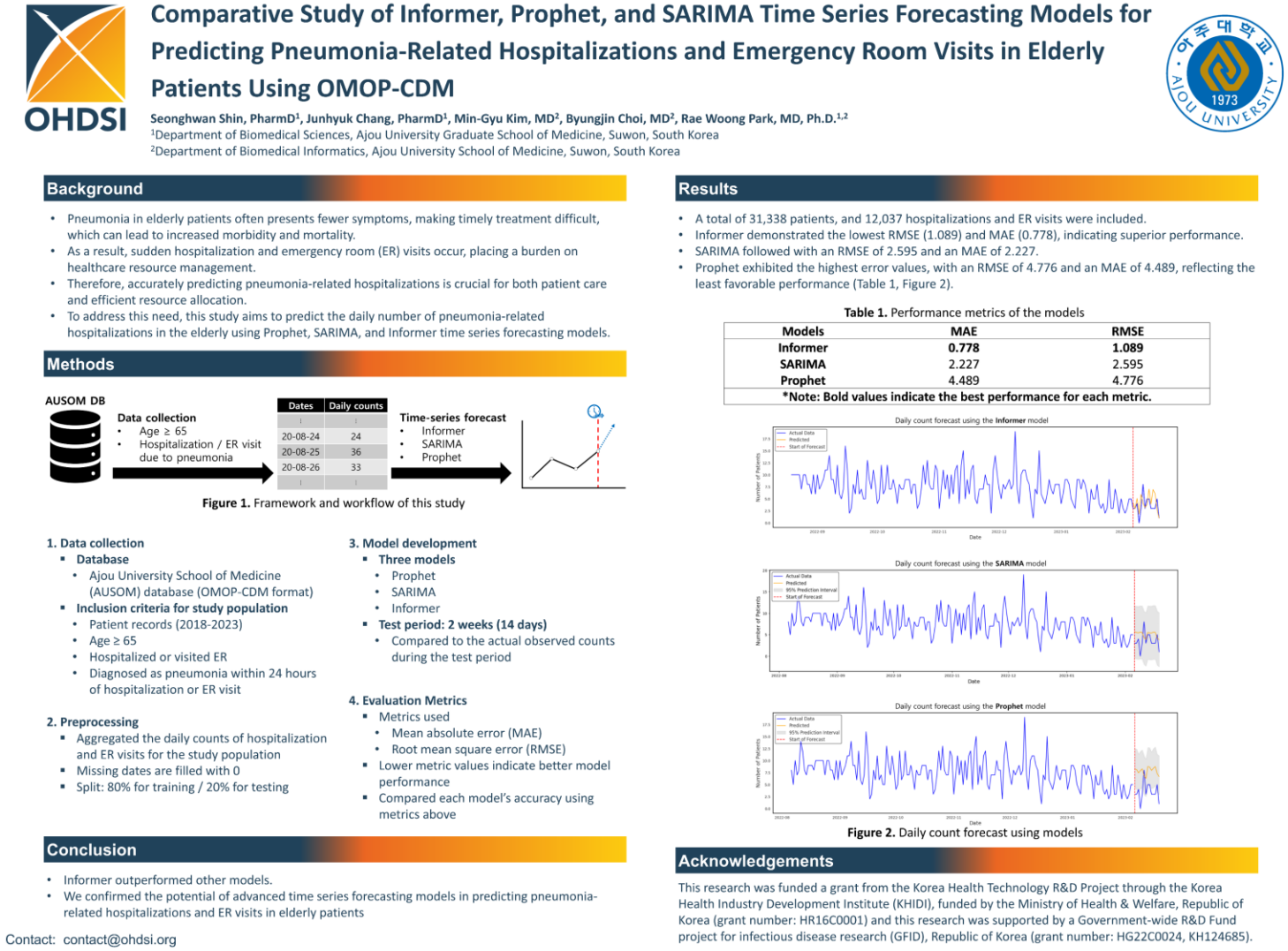
Contact: Benjamin.Viernes@va.gov

#OHDSISocialShowcase This Week

Tuesday

Comparative Study of Informer, Prophet, and SARIMA Time Series Forecasting Models for Predicting Pneumonia-Related Hospitalizations and Emergency Room Visits in Elderly Patients Using OMOP-CDM

(Seonghwan Shin, Junhyuk Chang, Min-Gyu Kim, Byungjin Choi, Rae Woong Park)



Comparative Study of Informer, Prophet, and SARIMA Time Series Forecasting Models for Predicting Pneumonia-Related Hospitalizations and Emergency Room Visits in Elderly Patients Using OMOP-CDM

Seonghwan Shin, PharmD[1], Junhyuk Chang, PharmD[1], Min-Gyu Kim, MD[2], Byungjin Choi, MD[2], Rae Woong Park, MD, Ph.D.[1,2]
[1]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, South Korea
[2]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea

Background

- Pneumonia in elderly patients often presents fewer symptoms, making timely treatment difficult, which can lead to increased morbidity and mortality.
- As a result, sudden hospitalization and emergency room (ER) visits occur, placing a burden on healthcare resource management.
- Therefore, accurately predicting pneumonia-related hospitalizations is crucial for both patient care and efficient resource allocation.
- To address this need, this study aims to predict the daily number of pneumonia-related hospitalizations in the elderly using Prophet, SARIMA, and Informer time series forecasting models.

Methods

AUSOM DB

Data collection
- Age ≥ 65
- Hospitalization / ER visit due to pneumonia

Dates	Daily counts
20-08-24	24
20-08-25	36
20-08-26	33

Time-series forecast
- Informer
- SARIMA
- Prophet

Figure 1. Framework and workflow of this study

1. Data collection
- **Database**
  - Ajou University School of Medicine (AUSOM) database (OMOP-CDM format)
- **Inclusion criteria for study population**
  - Patient records (2018-2023)
  - Age ≥ 65
  - Hospitalized or visited ER
  - Diagnosed as pneumonia within 24 hours of hospitalization or ER visit

2. Preprocessing
- Aggregated the daily counts of hospitalization and ER visits for the study population
- Missing dates are filled with 0
- Split: 80% for training / 20% for testing

3. Model development
- **Three models**
  - Prophet
  - SARIMA
  - Informer
- **Test period: 2 weeks (14 days)**
  - Compared to the actual observed counts during the test period

4. Evaluation Metrics
- Metrics used
  - Mean absolute error (MAE)
  - Root mean square error (RMSE)
- Lower metric values indicate better model performance
- Compared each model's accuracy using metrics above

Conclusion

- Informer outperformed other models.
- We confirmed the potential of advanced time series forecasting models in predicting pneumonia-related hospitalizations and ER visits in elderly patients

Contact: contact@ohdsi.org

Results

- A total of 31,338 patients, and 12,037 hospitalizations and ER visits were included.
- Informer demonstrated the lowest RMSE (1.089) and MAE (0.778), indicating superior performance.
- SARIMA followed with an RMSE of 2.595 and an MAE of 2.227.
- Prophet exhibited the highest error values, with an RMSE of 4.776 and an MAE of 4.489, reflecting the least favorable performance (Table 1, Figure 2).

Table 1. Performance metrics of the models

Models	MAE	RMSE
Informer	**0.778**	**1.089**
SARIMA	2.227	2.595
Prophet	4.489	4.776

*Note: Bold values indicate the best performance for each metric.

Daily count forecast using the **Informer** model

Daily count forecast using the **SARIMA** model

Daily count forecast using the **Prophet** model

Figure 2. Daily count forecast using models

Acknowledgements

This research was funded a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR16C0001) and this research was supported by a Government-wide R&D Fund project for infectious disease research (GFID), Republic of Korea (grant number: HG22C0024, KH124685).

#OHDSISocialShowcase This Week

Wednesday

Exploring the interplay between metabolic syndrome and brain volume in depression: Basis for Phenotype-Based Classification

(Sujin Gan, Narae Kim, Bumhee Park, Rae Woong Park)

Exploring the interplay between metabolic syndrome and brain volume in depression: Basis for Phenotype-Based Classification

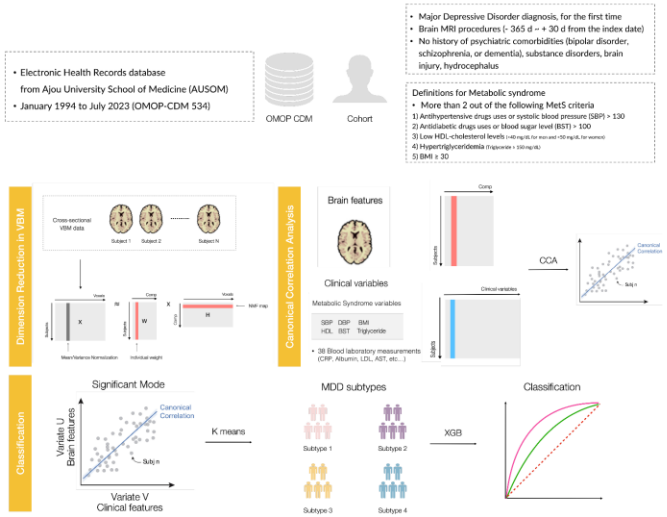
Sujin Gan[1], Narae Kim[1,3], Bumhee Park[2,3], and Rae Woong Park[1, 3]

[1] Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea
[2] Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea
[3] Office of Biostatistics, Medical Research Collaborating Center, Ajou Research Institute for Innovative Medicine, Ajou University Medical Center, Suwon, South Korea

Background

- The bidirectional relationship between major depressive disorder (MDD) and metabolic syndrome (MetS) suggests that each may exacerbate the other.
- While underlying mechanisms remain underexplored, brain structure and hematological markers are potential links.
- This study hypothesizes that integrating brain volume and clinical features may reveal distinct subgroups related to MetS in MDD patients.

Methods



Conclusions

- This study identified 4 brain components using non-negative matrix factorization (NMF), revealing significant correlations with metabolic features. Integrating NMF-derived brain features with clinical variables improved the classification performance of metabolic syndrome (MetS) in MDD patients.
- These findings suggest that subgroups, defined by brain morphology and clinical features, may play a key role in understanding and managing metabolic conditions in this population.

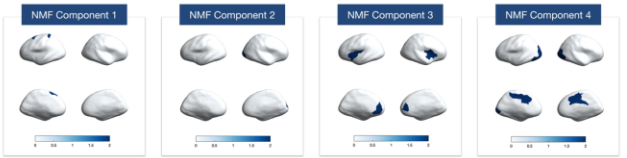
Results

1. Study population characteristics

- A total of 150 patients was selected based on the inclusion and exclusion criteria, 76 patients with MetS and 74 without MetS (with Mets: 52 females [68.4%]; age year, mean [SD] 61.5 ± 13.8; without Mets: 53 females [71.6%]; age year, mean [SD] 56.2 ± 1.66;).

2. NMF-derived brain features and clustering analysis

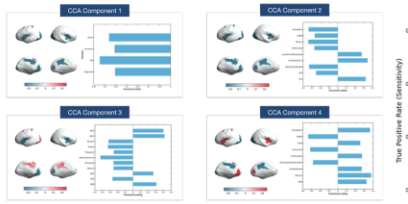
- Through NMF, 200 regions brain volume was reduced to 4 components, corresponding to structural brain networks



- These components were integrated with 44 clinical variables, and through canonical correlation analysis (CCA) followed by K-means clustering, two distinct subtypes were identified Mean silhouette (0.523).

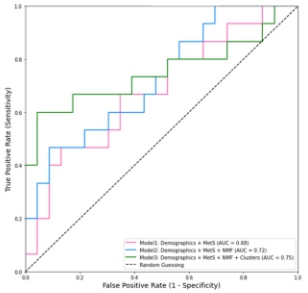
3. CCA Multivariate patterns of brain imaging and clinical variables

- Through CCA, 4 components were identified, with X loadings corresponding to clinical variables and Y loadings corresponding to structural brain networks, where 200 brain regions were reduced to 4 components



4. Classification model performance

- The classification model's performance improved by integrating NMF-derived brain components and cluster features, with the AUC increasing from 0.68 (baseline) to 0.75 in the final model.



Fundings

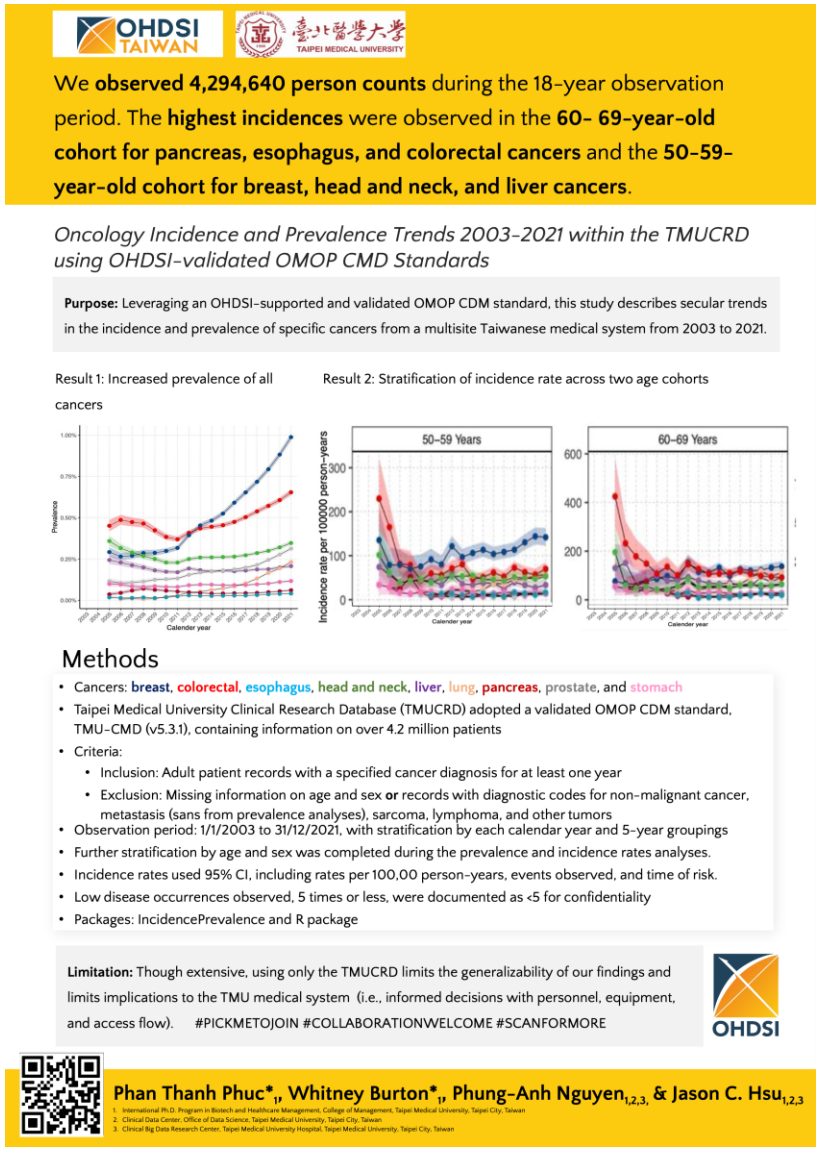
- This research was funded a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR16C0001) and was supported by a Government-wide R&D Fund project for infectious disease research (GFID), Republic of Korea (grant number: HG22C0024, KH124685).

#OHDSISocialShowcase This Week

Thursday

Oncology Incidence and Prevalence Trends 2005-2021 within the TMUCRD using OHDSI-validated OMOP CMD Standards

(Whitney Burton, Phan Thanh Phuc, Phung-Anh Nguyen, Jason C. Hsu)

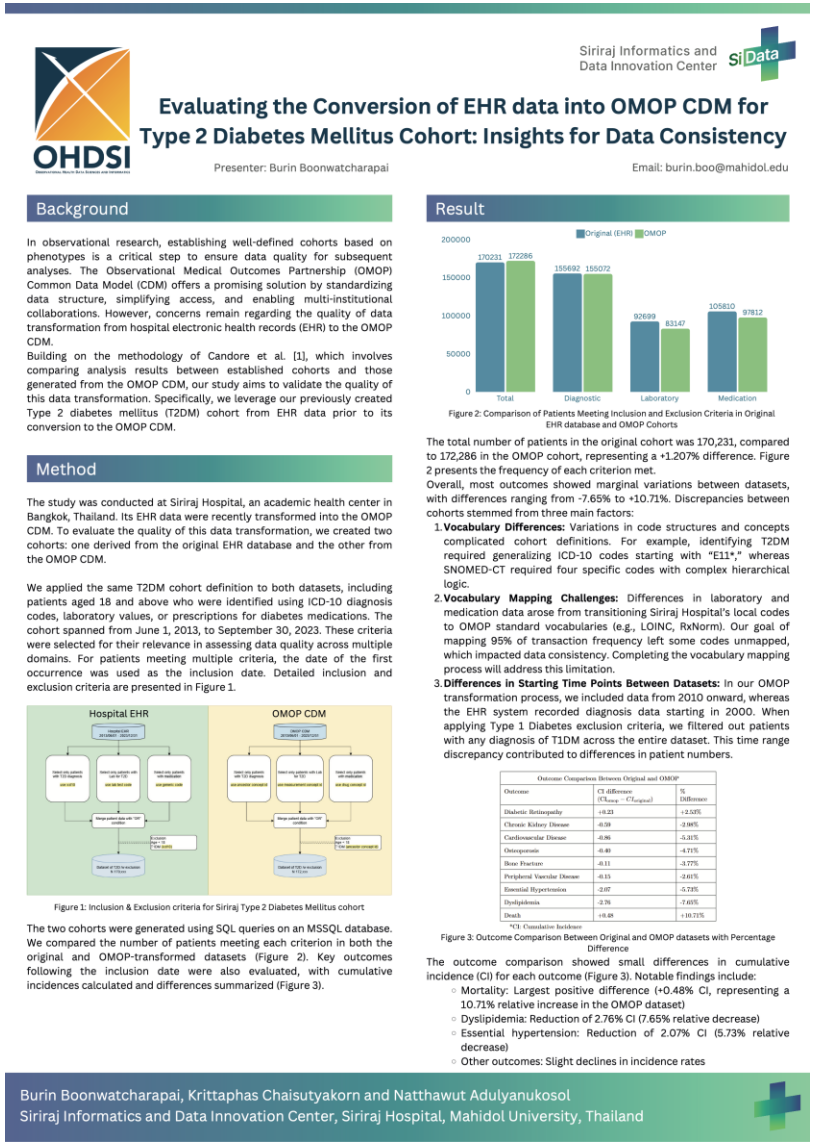


#OHDSISocialShowcase This Week

Friday

Evaluating the Conversion of EHR data into OMOP CDM for Type 2 Diabetes Mellitus Cohort: Insights for Data Consistency

(Burin Boonwatcharapai, Krittaphas Chaisutyakorn, Natthawut Adulyanukosol)



Siriraj Informatics and Data Innovation Center

Evaluating the Conversion of EHR data into OMOP CDM for Type 2 Diabetes Mellitus Cohort: Insights for Data Consistency

Presenter: Burin Boonwatcharapai
Email: burin.boo@mahidol.edu

Background

In observational research, establishing well-defined cohorts based on phenotypes is a critical step to ensure data quality for subsequent analyses. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) offers a promising solution by standardizing data structure, simplifying access, and enabling multi-institutional collaborations. However, concerns remain regarding the quality of data transformation from hospital electronic health records (EHR) to the OMOP CDM.

Building on the methodology of Candore et al. [1], which involves comparing analysis results between established cohorts and those generated from the OMOP CDM, our study aims to validate the quality of this data transformation. Specifically, we leverage our previously created Type 2 diabetes mellitus (T2DM) cohort from EHR data prior to its conversion to the OMOP CDM.

Method

The study was conducted at Siriraj Hospital, an academic health center in Bangkok, Thailand. Its EHR data were recently transformed into the OMOP CDM. To evaluate the quality of this data transformation, we created two cohorts: one derived from the original EHR database and the other from the OMOP CDM.

We applied the same T2DM cohort definition to both datasets, including patients aged 18 and above who were identified using ICD-10 diagnosis codes, laboratory values, or prescriptions for diabetes medications. The cohort spanned from June 1, 2013, to September 30, 2023. These criteria were selected for their relevance in assessing data quality across multiple domains. For patients meeting multiple criteria, the date of the first occurrence was used as the inclusion date. Detailed inclusion and exclusion criteria are presented in Figure 1.

Figure 1: Inclusion & Exclusion criteria for Siriraj Type 2 Diabetes Mellitus cohort

The two cohorts were generated using SQL queries on an MSSQL database. We compared the number of patients meeting each criterion in both the original and OMOP-transformed datasets (Figure 2). Key outcomes following the inclusion date were also evaluated, with cumulative incidences calculated and differences summarized (Figure 3).

Result

Figure 2: Comparison of Patients Meeting Inclusion and Exclusion Criteria in Original EHR database and OMOP Cohorts

The total number of patients in the original cohort was 170,231, compared to 172,286 in the OMOP cohort, representing a +1.207% difference. Figure 2 presents the frequency of each criterion met.

Overall, most outcomes showed marginal variations between datasets, with differences ranging from -7.65% to +10.71%. Discrepancies between cohorts stemmed from three main factors:

1. **Vocabulary Differences:** Variations in code structures and concepts complicated cohort definitions. For example, identifying T2DM required generalizing ICD-10 codes starting with "E11*," whereas SNOMED-CT required four specific codes with complex hierarchical logic.
2. **Vocabulary Mapping Challenges:** Differences in laboratory and medication data arose from transitioning Siriraj Hospital's local codes to OMOP standard vocabularies (e.g., LOINC, RxNorm). Our goal of mapping 95% of transaction frequency left some codes unmapped, which impacted data consistency. Completing the vocabulary mapping process will address this limitation.
3. **Differences in Starting Time Points Between Datasets:** In our OMOP transformation process, we included data from 2010 onward, whereas the EHR system recorded diagnosis data starting in 2000. When applying Type 1 Diabetes exclusion criteria, we filtered out patients with any diagnosis of T1DM across the entire dataset. This time range discrepancy contributed to differences in patient numbers.

Outcome Comparison Between Original and OMOP

Outcome	CI difference ($CI_{omop} - CI_{original}$)	% Difference
Diabetic Retinopathy	+0.23	+2.53%
Chronic Kidney Disease	-0.59	-2.98%
Cardiovascular Disease	-0.86	-5.31%
Osteoporosis	-0.40	-4.71%
Bone Fracture	-0.11	-3.77%
Peripheral Vascular Disease	-0.15	-2.61%
Essential Hypertension	-2.07	-5.73%
Dyslipidemia	-2.76	-7.65%
Death	+0.48	+10.71%

*CI: Cumulative Incidence

Figure 3: Outcome Comparison Between Original and OMOP datasets with Percentage Difference

The outcome comparison showed small differences in cumulative incidence (CI) for each outcome (Figure 3). Notable findings include:

- Mortality: Largest positive difference (+0.48% CI, representing a 10.71% relative increase in the OMOP dataset)
- Dyslipidemia: Reduction of 2.76% CI (7.65% relative decrease)
- Essential hypertension: Reduction of 2.07% CI (5.73% relative decrease)
- Other outcomes: Slight declines in incidence rates

Burin Boonwatcharapai, Krittaphas Chaisutyakorn and Natthawut Adulyanukosol
Siriraj Informatics and Data Innovation Center, Siriraj Hospital, Mahidol University, Thailand

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?

National Eye Institute
Research Today...Vision Tomorrow

National Eye Institute: Expand OHDSI Initiative for Eye Care and Ocular Imaging Challenge

Amberlynn Reed, Assistant Director
Office of Data Science and Health Informatics
National Eye Institute
National Institutes of Health

May 27, 2025

Winning Teams -- $250K each

- **Mass Eye and Ear (MEE)**
  Project: Expanding Ophthalmic Data in OMOP for Vision Research
  Lead: Michael Boland
  Enhancing OMOP CDM across U.S. and U.K. institutions with mapped eye care data and robust documentation.

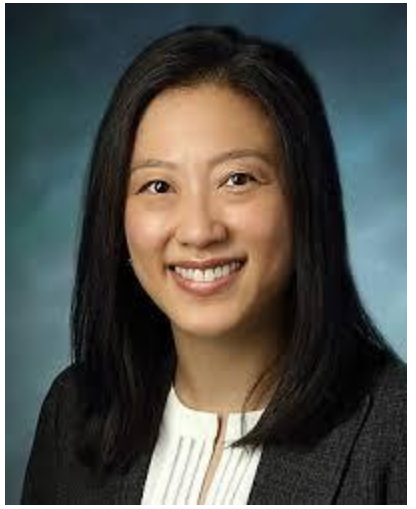
- **Oregon Health & Science University (OHSU)**
  Project: Data Coordinating Center for OHDSI Ophthalmic Network
  Lead: Michelle Hribar
  Creating tools to integrate imaging and exams into OMOP, enabling AI and federated learning.

- **Columbia University**
  Project: Eye Care and Vision Drug Characterization
  Lead: George Hripcsak
  Using OHDSI data to track ophthalmic medication use, adherence, and adverse events.

- **Stanford University**
  Project: Pediatric Eye Research Data Network
  Lead: Gayathri Srinivasan
  Building a longitudinal pediatric eye care network by mapping PEDSNet data into OMOP.

Honorable Mentions

Johns Hopkins University
Project: Equitable Access and Treatment for Diabetic Macular Edema
Lead: Cindy Cai
Using OHDSI data to study disparities in DME treatment across populations.

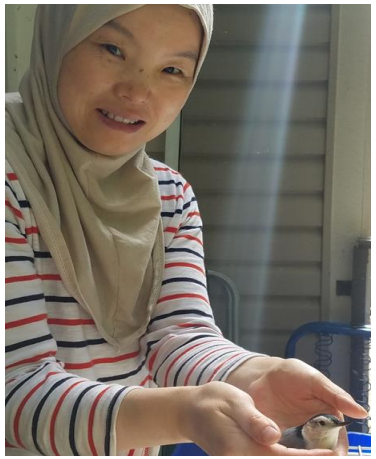


Columbia University
Project: AI-Federated Learning to Detect Thyroid Eye Disease
Lead: Kaveri Thakoor
Developing privacy-preserving tools to diagnose TED from facial imagery.

Expert Evaluation Panel

The expert evaluation panel consisted of interdisciplinary experts with specialized knowledge critical to assessing submissions:

- Evaluators assessed submissions based on innovation, feasibility, and potential impact on expanding the OHDSI network for vision research
- Their insights played a key role in selecting top submissions for the next phase of review



Asiyah Lin
Axil Informatics, former NIH Data Scholar

Patrick Ryan
Columbia University

Brian Stagg
University of Utah

Marc Suchard
UCLA

Federal Judging Panel

The Federal Judges bring diverse expertise relevant to the Challenge, including biomedical and clinical research, health informatics and data science, epidemiology and statistical analysis, vision science and ophthalmology, and translational research and policy implementation.

Role in the Challenge:

- Assessed feasibility, impact, and alignment with NEI's goals
- Selected submissions to recommend to NEI leadership based on rigorous scoring criteria

James Gao
NEI, NIH

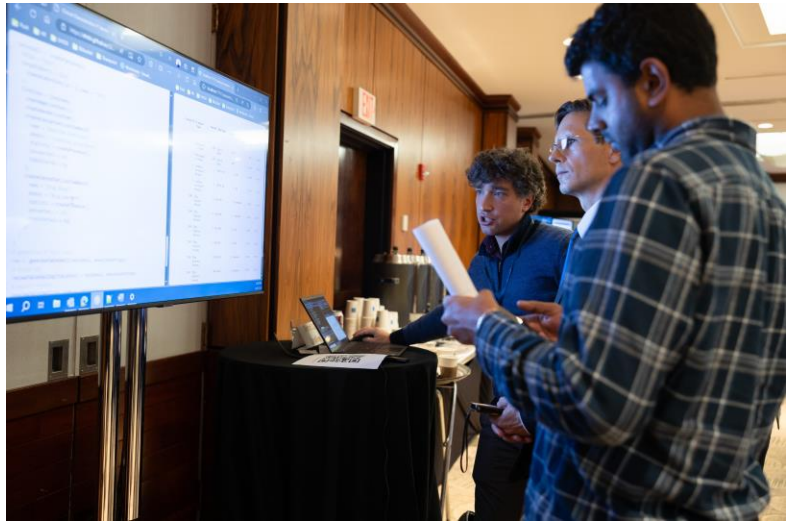
Azadeh Shaibi
NEI, NIH

Jonathan Pollock
NIDA, NIH

Ken Wilkins
NIDDK

May 27: #OHDSI2025 Collaborator Showcase

Brainstorm Breakouts



The collaborator showcase deadline is July 1

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-2025