



Evidence Network in Action: The Semaglutide Study

OHDSI Community Call
Nov. 19, 2024 • 11 am ET



Upcoming Community Calls

Date	Topic
Nov. 19	Evidence Network in Action: Semaglutide Study
Nov. 26	Collaborator Showcase Honorees
Dec. 3	Recent OHDSI Publications
Dec. 10	How Did We Do In 2024?
Dec. 17	Holiday-Themed Final Call of 2024



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!



Congratulations to the team of **Francesco Dernie, George Corby, Abigail Robinson, James Bezer, Nuria Mercade-Besora, Romain Griffier, Guillaume Verdy, Angela Leis, Juan Manuel Ramirez-Anguita, Miguel A. Mayer, James T. Brash, Sarah Seager, Rowan Parry, Annika Jodicke, Talita Duarte-Salles, Peter R. Rijnbeek, Katia Verhamme, Alexandra Pacurariu, Daniel Morales, Luis Pinheiro, Daniel Prieto-Alhambra, and Albert Prats-Urbe** on the publication of **Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU)** in *Pharmacoepidemiology and Drug Safety*.

Pharmacoepidemiology and Drug Safety

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU)

Francesco Dernie^{1,2} | George Corby^{1,2} | Abigail Robinson^{1,2} | James Bezer^{1,2} | Nuria Mercade-Besora² | Romain Griffier³ | Guillaume Verdy³ | Angela Leis⁴ | Juan Manuel Ramirez-Anguita⁵ | Miguel A. Mayer^{4,5} | James T. Brash⁶ | Sarah Seager⁶ | Rowan Parry⁷ | Annika Jodicke² | Talita Duarte-Salles^{7,8} | Peter R. Rijnbeek⁷ | Katia Verhamme⁷ | Alexandra Pacurariu⁹ | Daniel Morales^{9,10} | Luis Pinheiro⁹ | Daniel Prieto-Alhambra^{2,7} | Albert Prats-Urbe²

¹Medical Sciences Division, University of Oxford, Oxford, UK | ²Pharmaco- and Device Epidemiology, Centre for Statistics in Medicines, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK | ³Public Health Department, Medical Information Service, Medical Informatics and Archiving Unit (IAM), University Hospital of Bordeaux, Bordeaux, France | ⁴Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute, Barcelona, Spain | ⁵Management and Control Department, Consorci Mar Parc de Salut de Barcelona, Barcelona, Spain | ⁶Real World Solutions, IQVIA, Brighton, UK | ⁷Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, The Netherlands | ⁸Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Universitat Autònoma de Barcelona, Barcelona, Spain | ⁹Real World Evidence Workstream, European Medicines Agency, Amsterdam, The Netherlands | ¹⁰Division of Population Health and Genomics, University of Dundee, Dundee, UK

Correspondence: Daniel Prieto-Alhambra (d.prietoalhambra@darwin-eu.org)

Received: 10 January 2024 | **Revised:** 2 October 2024 | **Accepted:** 6 October 2024

Funding: This work is part of the DARWIN EU initiative, funded by the European Medicines Agency. Francesco Dernie, Annika Jodicke, Daniel Prieto-Alhambra and Albert Prats-Urbe receive partial support from the National Institute for Health and Care Research (NIHR) in the form of the Oxford NIHR Biomedical Research Centre.

Keywords: pancreatic cancer | phenotyping | systemic lupus erythematosus

ABSTRACT

Purpose: The generation of representative disease phenotypes is important for ensuring the reliability of the findings of observational studies. The aim of this manuscript is to outline a reproducible framework for reliable and traceable phenotype generation based on real world data for use in the Data Analysis and Real-World Interrogation Network (DARWIN EU). We illustrate the use of this framework by generating phenotypes for two diseases: pancreatic cancer and systemic lupus erythematosus (SLE).

Methods: The phenotyping process involves a 14-steps process based on a standard operating procedure co-created by the DARWIN EU Coordination Centre in collaboration with the European Medicines Agency. A number of bespoke R packages were utilised to generate and review codelists for two phenotypes based on real world data mapped to the OMOP Common Data Model.

Results: Codelists were generated for both pancreatic cancer and SLE, and cohorts were generated in six OMOP-mapped databases. Diagnostic checks were performed, which showed these cohorts had broadly similar incidence and prevalence figures to previously published literature, despite significant inter-database variability. Co-occurrent symptoms, conditions, and medication use were in keeping with pre-specified clinical descriptions based on previous knowledge.

Conclusions: Our detailed phenotyping process makes use of bespoke tools and allows for comprehensive codelist generation and review, as well as large-scale exploration of the characteristics of the resulting cohorts. Wider use of structured and reproducible phenotyping methods will be important in ensuring the reliability of observational studies for regulatory purposes.



OHDSI Shoutouts!



Congratulations to the team of
**Aaron S Eisman, Elizabeth S Chen,
Wen-Chih Wu, Karen M Crowley,
Dilum P Aluthge, Katherine Brown,
and Indra Neil Sarkar** on the
publication of **Learning health
system linchpins: information
exchange and a common data
model** in *JAMIA*.

Journal of the American Medical Informatics Association, 2024, 1–11
<https://doi.org/10.1093/jamia/ocae277>
Research and Applications



Research and Applications

Learning health system linchpins: information exchange and a common data model

Aaron S. Eisman , MD, PhD^{*1,2,3}, **Elizabeth S. Chen**, PhD^{1,2,4}, **Wen-Chih Wu** , MD, MPH^{2,4,5}, **Karen M. Crowley**, MS, PhD¹, **Dilum P. Aluthge**, PhD^{1,2}, **Katherine Brown**, PhD, MSN, RN^{1,2}, **Indra Neil Sarkar**, PhD, MLIS^{1,2,4,6}

¹Center for Biomedical Informatics, Brown University, Providence, RI 02912, United States, ²The Warren Alpert Medical School, Brown University, Providence, RI 02912, United States, ³Yale School of Medicine, New Haven, CT 06510, United States, ⁴School of Public Health, Brown University, Providence, RI 02912, United States, ⁵Division of Cardiology, VA Providence Health Care, Providence, RI 02912, United States, ⁶Rhode Island Quality Institute, Providence, RI 02912, United States

*Corresponding author: Aaron S. Eisman, MD, PhD, Center for Biomedical Informatics, Brown University, Providence, RI 02912, United States (aaron.eisman@yale.edu)

Abstract

Objective: To demonstrate the potential for a centrally managed health information exchange standardized to a common data model (HIE-CDM) to facilitate semantic data flow needed to support a learning health system (LHS).

Materials and Methods: The Rhode Island Quality Institute operates the Rhode Island (RI) statewide HIE, which aggregates RI health data for more than half of the state's population from 47 data partners. We standardized HIE data to the Observational Medical Outcomes Partnership (OMOP) CDM. Atherosclerotic cardiovascular disease (ASCVD) risk and primary prevention practices were selected to demonstrate LHS semantic data flow from 2013 to 2023.

Results: We calculated longitudinal 10-year ASCVD risk on 62,999 individuals. Nearly two-thirds had ASCVD risk factors from more than one data partner. This enabled granular tracking of individual ASCVD risk, primary prevention (ie, statin therapy), and incident disease. The population was on statins for fewer than half of the guideline-recommended days. We also found that individuals receiving care at Federally Qualified Health Centers were more likely to have unfavorable ASCVD risk profiles and more likely to be on statins. CDM transformation reduced data heterogeneity through a unified health record that adheres to defined terminologies per OMOP domain.

Discussion: We demonstrated the potential for an HIE-CDM to enable observational population health research. We also showed how to leverage existing health information technology infrastructure and health data best practices to break down LHS barriers.

Conclusion: HIE-CDM facilitates knowledge curation and health system intervention development at the individual, health system, and population levels.

Key words: learning health system; health information exchange; common data model; atherosclerotic cardiovascular disease.



OHDSI Shoutouts!



Congratulations to the team of **Junqing Xie, Kim López-Güell, Daniel Dedman, Talita Duarte-Salles, Raivo Kolde, Raúl López-Blasco, Álvaro Martínez, Gregoire Mercier, Alicia Abellan, Johnmary T. Arinze, Zara Cuccu, Antonella Delmestri, Dominique Delseny, Sara Khalid, Chungsoo Kim, Ji-woo Kim, Kristin Kostka, Cora Loste, Lourdes Mateu, Miguel A. Mayer, Jaime Meléndez-Cardiel, Núria Mercadé-Besora, Mees Mosseveld, Akihito Nishimura, Hedvig M.E. Nordeng, Jessie O. Oyinlola, Laura Pérez-Crespo, Marta Pineda-Moncusí, Juan Manuel Ramírez-Anguita, Nhung T.H. Trinh, Anneli Uusküla, Bernardo Valdivieso, Theresa Burkard, Edward Burn, Martí Català, Daniel Prieto-Alhambra, Roger Paredes, and Annika M. Jödicke** on the publication of **Incidence of post-acute COVID-19 symptoms across healthcare settings in seven countries: an international retrospective cohort study using routinely-collected data** in *eClinical Medicine*.

Incidence of post-acute COVID-19 symptoms across healthcare settings in seven countries: an international retrospective cohort study using routinely-collected data



Junqing Xie,^{a,aa} Kim López-Güell,^{a,aa} Daniel Dedman,^b Talita Duarte-Salles,^{c,d} Raivo Kolde,^e Raúl López-Blasco,^f Álvaro Martínez,^g Gregoire Mercier,^h Alicia Abellan,ⁱ Johnmary T. Arinze,^j Zara Cuccu,^k Antonella Delmestri,^l Dominique Delseny,^m Sara Khalid,ⁿ Chungsoo Kim,^o Ji-woo Kim,^p Kristin Kostka,^{q,m} Cora Loste,^{r,w,x,y} Lourdes Mateu,^{r,w,x,y} Miguel A. Mayer,^r Jaime Meléndez-Cardiel,^l Núria Mercadé-Besora,^{a,c} Mees Mosseveld,^d Akihito Nishimura,^z Hedvig M. E. Nordeng,^{a,p} Jessie O. Oyinlola,^b Laura Pérez-Crespo,^z Marta Pineda-Moncusí,^z Juan Manuel Ramírez-Anguita,ⁿ Nhung T. H. Trinh,^o Anneli Uusküla,^z Bernardo Valdivieso,^{g,t} Theresa Burkard,^o Edward Burn,^o Martí Català,^o Daniel Prieto-Alhambra,^{a,d,*} Roger Paredes,^{h,q,u,v,x,y,z,ab} and Annika M. Jödicke^{a,ab}



^aNuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

^bCPRD, Medicines and Healthcare Products Regulatory Agency, London, UK

^cFundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

^dDepartment of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands

^eInstitute of Computer Science, University of Tartu, Tartu, Estonia

^fBiocomputing Unit, Aragon Health Sciences Institute (IACS), Zaragoza, Spain

^gThe Health Research Institute Hospital La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7a Planta, 46026, Valencia, Spain

^hDepartment of Infectious Diseases, Hospital Germans Trias i Pujol, Badalona, Catalonia, Spain

ⁱPublic Health Department, University Hospital of Montpellier, 34295 Montpellier, France

^jIDESP, Université de Montpellier, INSERM, 34000, Montpellier, France

^kDepartment of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Republic of Korea

^lBig Data Department, Health Insurance Review and Assessment Service, Wonju, Republic of Korea

^mThe OHDSI Center at the Roux Institute, Northeastern University, Portland, ME, USA

ⁿParc de Salut Mar, Hospital del Mar Medical Research Institute, Barcelona, Spain

^oPharmacoepidemiology and Drug Safety Research Group, Department of Pharmacy, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway

^pDepartment of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway

^qCenter for Global Health and Diseases, Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH, USA

^rInstitute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia

^sDepartment of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

^tThe University and Polytechnic La Fe Hospital of Valencia, Avenida Fernando Abril Martorell, 106 Torre H 1a Planta, 46026, Valencia, Spain

^uCIBER Infectious Diseases (CIBERINFEC), Institute of Health Carlos III (ISCIII), Madrid, Spain

^vUniversitat Autònoma de Barcelona, Catalonia, Spain

^wREICOP (Red de Investigación Covid Persistente), Madrid, Spain

^xFundació Lluita Contra les Infeccions, Badalona, Catalonia, Spain

^yUniversitat de Vic – UCC, Vic, Catalonia, Spain

^zIrsiCaba AIDS Research Institute, Germans Trias i Pujol Research Institute (IGTP), Can Ruti Campus, Badalona, Catalonia, Spain

Summary

Background The World Health Organisation (WHO) has identified a range of symptomatic manifestations to aid in the clinical diagnosis of post-COVID conditions, herein referred to as post-acute COVID-19 symptoms. We conducted an international network cohort study to estimate the burden of these symptoms in North American, European, and Asian populations.

eClinicalMedicine
2024;77: 102903
Published Online xxx
<https://doi.org/10.1016/j.eclim.2024.102903>



OHDSI Shoutouts!



Congratulations to the team of **Eizen Kimura, Yukinobu Kawakami, Shingo Inoue, and Ai Okajima** on the publication of **Mapping Drug Terms via Integration of a Retrieval-Augmented Generation Algorithm with a Large Language Model** in *Healthcare Informatics Research*.

Original Article

Healthc Inform Res. 2024 October;30(4):355-363.
<https://doi.org/10.4258/hir.2024.30.4.355>
pISSN 2093-3681 • eISSN 2093-369X



HIR

Healthcare Informatics Research

Mapping Drug Terms via Integration of a Retrieval-Augmented Generation Algorithm with a Large Language Model

Eizen Kimura¹, Yukinobu Kawakami¹, Shingo Inoue², Ai Okajima²

¹Department of Medical Informatics, Medical School of Ehime University, Toon, Ehime, Japan

²Yuimedi Inc., Tokyo, Japan

Objectives: This study evaluated the efficacy of integrating a retrieval-augmented generation (RAG) model and a large language model (LLM) to improve the accuracy of drug name mapping across international vocabularies. **Methods:** Drug ingredient names were translated into English using the Japanese Accepted Names for Pharmaceuticals. Drug concepts were extracted from the standard vocabulary of OHDSI, and the accuracy of mappings between translated terms and RxNorm was assessed by vector similarity, using the BioBERT-generated embedded vectors as the baseline. Subsequently, we developed LLMs with RAG that distinguished the final candidates from the baseline. We assessed the efficacy of the LLM with RAG in candidate selection by comparing it with conventional methods based on vector similarity. **Results:** The evaluation metrics demonstrated the superior performance of the combined LLM + RAG over traditional vector similarity methods. Notably, the hit rates of the Mixtral 8x7b and GPT-3.5 models exceeded 90%, significantly outperforming the baseline rate of 64% across stratified groups of PO drugs, injections, and all interventions. Furthermore, the r-precision metric, which measures the alignment between model judgment and human evaluation, revealed a notable improvement in LLM performance, ranging from 41% to 50% compared to the baseline of 23%. **Conclusions:** Integrating an RAG and an LLM outperformed conventional string comparison and embedding vector similarity techniques, offering a more refined approach to global drug information mapping.

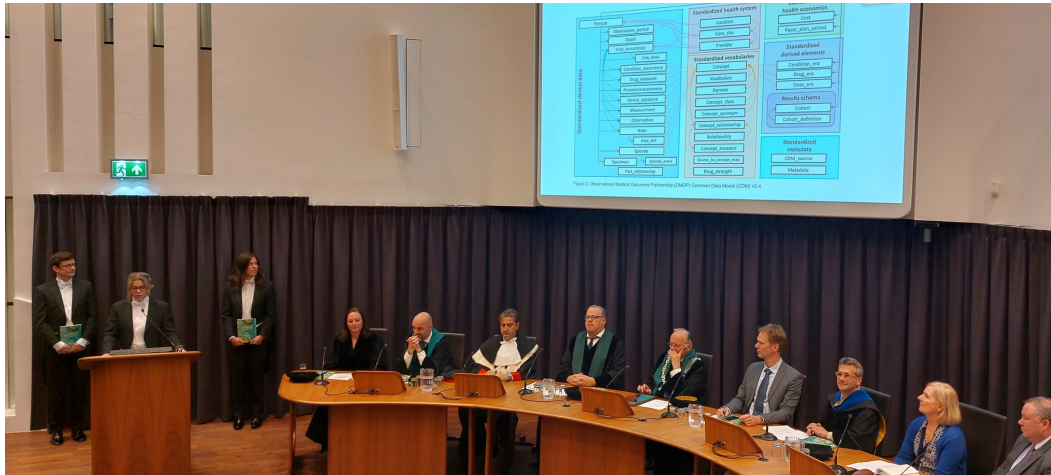
Keywords: RxNorm, Machine Learning, Computer Neural Network, Terminology, Controlled Vocabulary



OHDSI Shoutouts!



Congratulations to **Erica Voss** for successfully defending her dissertation at Erasmus University earlier today!





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	1 pm	Common Data Model
Wednesday	1 pm	Perinatal & Reproductive Health
Thursday	7 am	India Community Call
Thursday	9 am	OMOP CDM Oncology – Vocabulary/Development Subgroup
Thursday	9:30 am	Themis
Thursday	12 pm	Medical Devices
Thursday	12 pm	HADES
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development & Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Clinical Trials
Friday	11:30 am	Steering
Monday	10 am	CDM Survey Subgroup
Tuesday	9 am	OMOP CDM Oncology – Genomic Subgroup



Next CBER Best Seminar: Tomorrow!

Topic: Statistical methods for improving post-licensure vaccine safety surveillance

Presenter: Jennifer Clark Nelson, PhD,
Director of Biostatistics & Senior Investigator,
Biostatistics Division, Kaiser Permanente
Washington Health Research Institute.

Date/Time: Nov. 20, 11 am ET



ohdsi.org/cber-best-seminar-series



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



Agnes Kiragga, PhD

Lead - Data Science Program, African Population and Health Research Center (APHRC)

‘Promoting Data Science and Data Harmonization in Africa ’

November 21, 2024, 11am-12pm EST

Virtually via [Zoom](#)

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

TuftsMedicine
Tufts Medical Center



2024 APAC Symposium

Dec. 4-8 • Marina Bay Sands & National University of Singapore (NUS)

Dec. 4: Tutorial at NUS

Dec. 5-6: Main Conference at Marina Bay Sands

Dec. 7-8: Datathon at NUS



ohdsi.org/APAC2024



2024 Global Symposium



2024 OHDSI Global Symposium

Oct. 22-24 • New Brunswick, N.J. • Hyatt Regency Hotel

The 10th annual OHDSI Global Symposium brought together more than 470 global collaborators for three days of sharing research, building new connections and pushing forward our mission of improving health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

This page will host all materials from OHDSI2024, including video presentations (when available) from the main conference and tutorials, slide decks, posters, demos and more.

State of the Community

Where Have We Gone and Where Are We Going?
(George Hripsak, Columbia University)

Expand OHDSI Initiative for Eye Care and Ocular Imaging Challenge
(Amberlynn Reed, National Eye Institute)

Titan Awards
(George Hripsak, Columbia University & Marc Suchard, UCLA)



State of the Community Slides

Plenary: Value Proposition for Participating in OHDSI Network Studies like LEGEND-T2DM

Introduction to OHDSI Evidence Network / Marketplace
(Moderator: Clair Blacketer, Johnson & Johnson)

Reflections from US Department of Veterans Affairs
(Scott Duvall, VA)

Reflections from SIDIAP (Spain)
(Talita Duarte-Salles, IDIAP)

Reflections from a Global Commercial Data Provider
(Atif Adam, IQVIA)



Plenary: Value Proposition for Participating in OHDSI Network Studies like LEGEND-T2DM Slides

Plenary Q&A: Lessons Learned on LEGEND-T2DM Journey

Moderator: Fan Bu, University of Michigan

Panelists: LEGEND-T2DM co-authors



Plenary Q&A: Lessons Learned on LEGEND-T2DM Journey Slides

Plenary Panel: JACC-OHDSI Partnership

Moderators:
Nicole Pratt, University of South Australia
Marc Suchard, UCLA

Panelists:
Harlan Krumholz, Yale University
Seng Chan You, Yonsei University
Yuan Lu, Yale University



Plenary Panel: JACC-OHDSI Partnership Slides

2024 Global Collaborator Showcase

Observational Data Standards & Management

- 1 – [Application of OMOP Common Data Model to Disease Registry Data](#) (Vojtech Huser, Maria Rogozhnikina, Vlad Korsik, Teresa A. Simon, Peter Moorthamer, Dan Kiselev, Teresa A. Simon, Anastasia Vakhmistrova, Eugene Paulenkovich, Alexander Davydov, Michel Van Speybroeck)
- 2 – [Best Practices for Developing Disease-Specific Federated Networks: Insights from a Systemic Lupus Erythematosus Study](#) (Clair Blacketer, Frank DeFalco, Gowtham A Rao, Anna Sheahan, Michel Van Speybroeck, Martine Lewi, Federico Zazzetti)
- 3 – [Standardizing Rare Disease Patient Registry data to the OMOP-CDM](#) (Parag Shiralkar, Radhika Lakireddy, Sushma Ghanta, Sanket Kalyankar, Monika Brand)
- 4 – [Phederization – the federated network of Pulmonary Hypertension registries](#) (Eva-Maria Didden, Valerie van Baalen, Michel van Speybroeck, Monika Brand)
- 5 – [Lessons from mapping cancer information from European hospitals to ICD-O-3 conditions in OMOP](#) (Lars Halvorsen, Olivier Bouissou, Elisabeth Ross, Stelios Theophanous, Joëlle Thonnard, Piers Mahon)
- 6 – [SMEs optimization with high precision data ingestion of CAPRICORN onto OMOP at AllianceChicago](#) (Andrew Hamilton, Amro Hassan, Davera Gabriel, Guy Tsafnat)
- 7 – [Process of Conversion of Ukrainian Medical Data to OMOP CDM Format](#) (Bohdan Khilchevskiy, Denys Kaduk, Maksym Trofymenko, Polina Talapova, Tetiana Nesmilan, Max Ved, Inna Ageeva, Pavlova Olga, Holovko Tetiana, Shevchenko Natalia)
- 8 – [An evaluation of the transformation of large German EHR database to OMOP CDM](#) (Andreas Ochs, Milou Brand, Jack Brewster, Methosdios Typou, Meda Sandu, Joe Maskell, Meghan Pettine, Atif Adam, George Kafatos)
- 9 – [Adopting the OMOP Oncology CDM at the Helsinki University Hospital](#) (Valtteri Nieminen, Alexey Ryzhenkov, Johanna Sanoja, Salma Rachidi, Juho Laitinenmaa, Joonas Laitinen, Samu Eränen, Tomi Mäkelä, Eric Fey, Kimmo Porkka)
- 10 – [Going global, redeeming the local: an innovative approach to implement the OMOP CDM in two countries of the Global South](#) (Valentina Martufi, Emma Kalk, Enny S. Cruz, Juliana Araújo Prata de Faria, Adalton do Anjos Fonseca, Mauricio L. Barreto, Maria Yury Travassos Ichihara, Jessica Gammon, Nicki Tiffin, Chris Fourie, Danilo Luis Cerqueira Dias, Denise Moraes Pimenta, Tsaone Tamuhla, Andrew Boule, Themba Mutemaringa, Juan-Paul Hynek, Muzzamil Ismail, Julio Barbour Oliveira, Ricardo Felix Monteiro Neto, Júlia Pescarini, Fernanda Revoredo de Sousa, Marianne Costa e Silva Lage, Adam Loff, Melvin Moodley, Elzo Pereira Pinto Junior)
- 11 – [Transforming Clinical Trial Data to the OMOP CDM](#) (Cynthia Sung, Mike Hamdi, Zhen Lin, Tom Walpole, Rebecca Baker, Melissa Cook, Shital Desai, Priya Gopal, Dan Hartley, Vojtech Huser, Priya Meghrajani, Tra Nguyen, Paul Orna12, Katy Sadowski, Sebastiaan van Sandijk, Philip Solovoyev, Ramona Walls, Kenneth J. Wilkins, Qi Yang)
- 12 – [Streamlining Research Data Standardization: AI-READI Survey Instrument Data Elements and MoCA Measurement Data Elements are curated and mapped utilizing a Standardized Value Set Mapping Table for transformation into the OMOP Common Data Model](#) (Stephanie S. Hong, James Cavallon, Yvette Chen, Monique Bangudi, Jessica Mitchell, Dawn Matthies, Steven Chamberlin, Aaron Cohen, Julie Owens, Abigail Lucero, Sally Baxter, Christopher G Chute, Cecilia S. Lee, Aaron Lee, AI-READI consortium)
- 13 – [Institutionalizing data interoperability and the application of common data models in a health data and research center: CIDACS' experience in Brazil](#) (Valentina Martufi, Juliana Araújo Prata de Faria, Danilo Luis Cerqueira Dias, Elzo Pereira Pinto Junior, Roberto Carreiro, Pablo Ivan Ramos, Mauricio L. Barreto)
- 14 – [OMOP GIS Vocabulary Package for Observational Studies in Health Care and Public Health](#) (Maksym Trofymenko, Polina Talapova, Andrew Williams)
- 15 – [Enhancing Infectious Disease Data Integration and management through OMOP-CDM in South Korea](#) (Min Ho An, Seok Kim, ByungJin Choi, Sooyoung Yoo, Rae Woong Park, Ji Seon Oh)
- 16 – [FHIR to OMOP Cookbook – Mapping mCODE FHIR Resources for Observational Research](#) (Qi Yang, Guy Livne, Sebastian van Sandijk, May Terry)
- 17 – [Towards Reproducible Imaging Research: Implementation of DICOM to OMOP CDM](#) (Woo Yeon Park, Ben Martin, Gabriel Salvador, Blake Dewey, Teri Sippel Schmidt, Paul Nagy)
- 18 – [Leveraging UDI for Advanced Medical Device Tracking in OMOP-CDM](#) (Seojeong Shin, Yiju Park, Sejoeng Eom, Kyulee Jeon, Seng Chan You)
- 19 – [Inclusion of Intraocular pressure data into the University of California Health Data Warehouse](#) (William Halliwell, Shahin Hallaj, Ayan Patel, Catherine Q. Sun, Kerry Goetz, Michelle Hribar, Sally L. Baxter, on behalf of the OMOP Eye Care & Vision Research Workgroup)
- 20 – [A Collaborative Analytic Enclave for the Metabolic Dysregulation and Obesity Cancer Risk Program \(MedOC\) Consortium: Extensions of the OMOP Common Data Model for Translational Research](#) (Madhan Subramanian, Nisha Grover, Maddie Wheeler, Marinella Temprosa)
- 21 – [Expanding the OMOP Common Data Model to support Extracorporeal Life Support research](#) (Clemens Rieder, Oleg Zhuk, Ahmed Said, Peta M.A. Alexander, Dominik J. Hoechter)
- 22 – [ETing from your OMOP CDM to your OMOP CDM? An efficient solution to vocabulary migration](#) (Clair Blacketer, Anton Ivanov, Evanette Burrows, Dmitry Dymshyts, Frank DeFalco)
- 23 – [Evaluating the impact of different vocabulary versions on cohort definitions and CDM](#) (Dmitry Dymshyts, Frank DeFalco, Anna Ostropolets, Gowtham Rao, Azza Shoabi, Clair Blacketer)

ohdsi.org/OHDSI2024

2024 Our Journey is Online

OHDSI COLLABORATORS

Map of Collaborators

The OHDSI community brings together volunteers from around the world to establish open community data standards, develop open-source software, conduct methodological research, and apply scientific best practices to answer public health questions by generating reliable clinical evidence.

OHDSI COLLABORATORS

Our community is ALWAYS seeking new collaborators. Do you want to focus on data standards or methodological research? Are you passionate about open-source development or clinical applications? Do you have data that you want to be part of global network studies? Do you want to join a global community that truly values the benefits of open science? Add a dot to the map below and JOIN THE JOURNEY!

DATA STANDARDS

OMOP CDM Data Sources

The OMOP Common Data Model is an open community data standard, freely available to anyone who would like to standardize their patient-level data into a format that makes it easier to perform analyses and generate reliable evidence. OHDSI prides itself on stewarding the OMOP Common Data Model as a community resource, and actively encourages its adoption through various workshops, open-source tool development, and educational sessions, and collaborative support.

There are currently 544 data sources that come from 54 different countries which have been standardized to the OMOP Common Data Model. These data sources contain a range of patient-level observations from various data capture processes within routine clinical care, including electronic health records, administrative claims, registries, hospital systems, genomics and biobanks. Together, these data sources conservatively cover more than 874 million unique patient records, representing approximately 12% to the world's population.

DATA STANDARDS

OMOP CDM Adoption Map

OHDSI By The Numbers

- 4,294 collaborators
- 83 countries
- 21 time zones
- 6 continents
- 1 community

OHDSI PUBLICATIONS

Through disease-specific Federated Data Networks, OHDSI has enabled a range of research in various domains. This includes research on the impact of COVID-19, the effectiveness of various treatments, and the impact of various risk factors on health outcomes.

OHDSI PUBLICATIONS

Through disease-specific Federated Data Networks, OHDSI has enabled a range of research in various domains. This includes research on the impact of COVID-19, the effectiveness of various treatments, and the impact of various risk factors on health outcomes.

OHDSI PUBLICATIONS

Through disease-specific Federated Data Networks, OHDSI has enabled a range of research in various domains. This includes research on the impact of COVID-19, the effectiveness of various treatments, and the impact of various risk factors on health outcomes.

Our Journey

Where The OHDSI Community Has Been
And Where We Are Going
2024 edition



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



@OHDSI

www.ohdsi.org

#JoinTheJourney



ohdsi



#OHDSISocialShowcase This Week

Monday

FHIR to OMOP Cookbook - Mapping mCODE FHIR Resources for Observational Research

(**Qi Yang**, Guy Livne, Sebastian van Sandijik, May Terry)

FHIR-to-OMOP Cookbook - a cohesive approach to mapping FHIR Resources for Observational Research

PRESENTERS: Guy Livne, May Terry, Qi Yang

BACKGROUND

The conversion of Electronic Health Record (EHR) data from Health Level 7 (HL7®) Fast Healthcare Interoperability Resources (FHIR®) to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is becoming increasingly standardized for research, though inconsistencies exist among implementers. To address this, HL7 and the Observational Health Data Sciences and Informatics (OHDSI) formed a partnership in 2021, focusing on better alignment between FHIR and the OMOP CDM, including mapping minimum Common Oncology Data Elements (mCODE™) v2.0 FHIR resources to OMOP CDM v5.4. These efforts led to the creation of the FHIR-to-OMOP Cookbook, a guide for providing best practices for seamless integration of FHIR resources with the OMOP CDM.

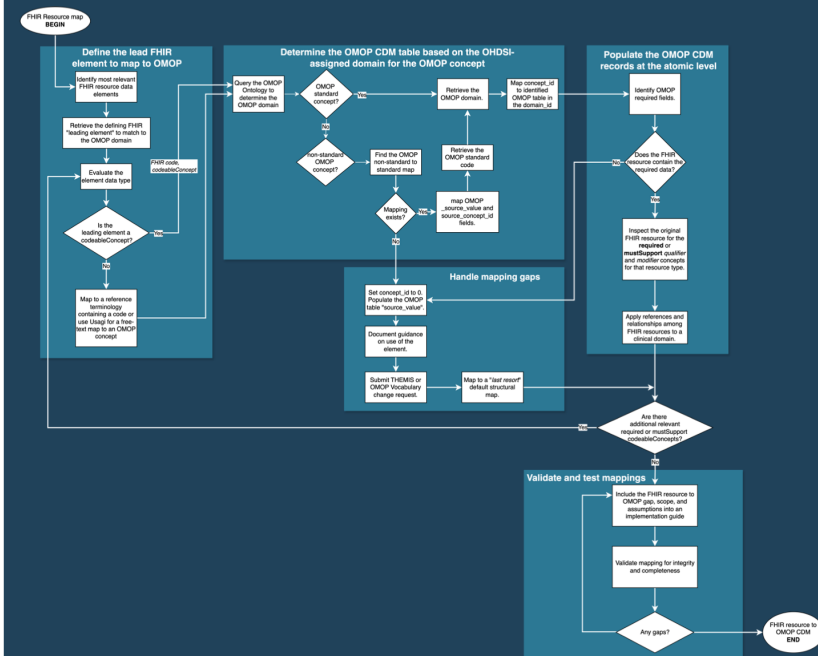
PRINCIPAL ARTIFACTS

1. Holistic mapping approach
2. Starter guide for implementers
3. Step-by-step practical guidelines

METHOD

1. Define the FHIR elements that are relevant to represent in OMOP.
2. Identify the OMOP concept that best matches each of the FHIR element.
3. Determine the OMOP CDM table based on the OHDSI-assigned domain for the OMOP concept.
4. Map related FHIR resources to the OMOP CDM table required fields.
5. Populate the OMOP CDM records at the atomic/record level.
6. Apply references and relationships among FHIR resources to a clinical domain from the original resource where possible.
7. Test OMOP CDM integrity and completeness.
8. Compare FHIR and OMOP representations for a known set of patient information for accuracy and identify gaps.

A cohesive approach for mapping FHIR to OMOP



Scan to download the FHIR-to-OMOP Cookbook



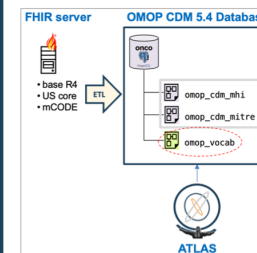
COOKBOOK VISION

Starter guide for implementers seeking to convert HL7 FHIR resources to the OMOP CDM.

The Cookbook compares the purpose, design principles of both FHIR and OMOP CDM. It proposes a methodology for mapping with example use cases and patterns in mapping FHIR resources to equivalent OMOP CDM elements.

TEST ENVIRONMENT

The FHIR-to-OMOP Cookbook methodology was tested in the May 2022 HL7 FHIR Connection with mappings from mCODE v2.0 and OMOP CDM v5.4.



RESULTS

Over 80% of the mCODE FHIR-based elements were mapped. Gaps and learnings from the mapping and testing process were documented in the FHIR-to-OMOP Cookbook.

CONCLUSION

Although FHIR and OMOP are widely adopted standards, their purpose and data modeling differences require further guidance, including a methodology on how to integrate both process and technical transformations. The FHIR-to-OMOP Cookbook is an initial step in providing this guidance, and can be used to inform other interoperability efforts such as the HL7 Vulcan FHIR to OMOP initiative.

Livne, G.¹, Terry, M.², Yang, Q.³

¹ Kineret, Israel, ² MITRE, Bedford, MA, ³ IQVIA, Durham, NC



© 2024 The MITRE Corporation. All rights reserved. Approved for Public Release. 23-03025-10. Distribution unlimited.



#OHDSISocialShowcase This Week

Tuesday

Enabling Clinical Trial Feasibility and Patient Finding Through the Use of the OMOP CDM

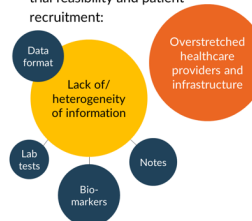
(Mui Van Zandt, Jason Hsu, Alex Nguyen, Gyeol Song, Sabrina Cheema, Michael Krupnick)

Enabling Clinical Trial Feasibility and Patient Finding Through the Use of the OMOP CDM

PRESENTER: Mui Van Zandt

INTRO

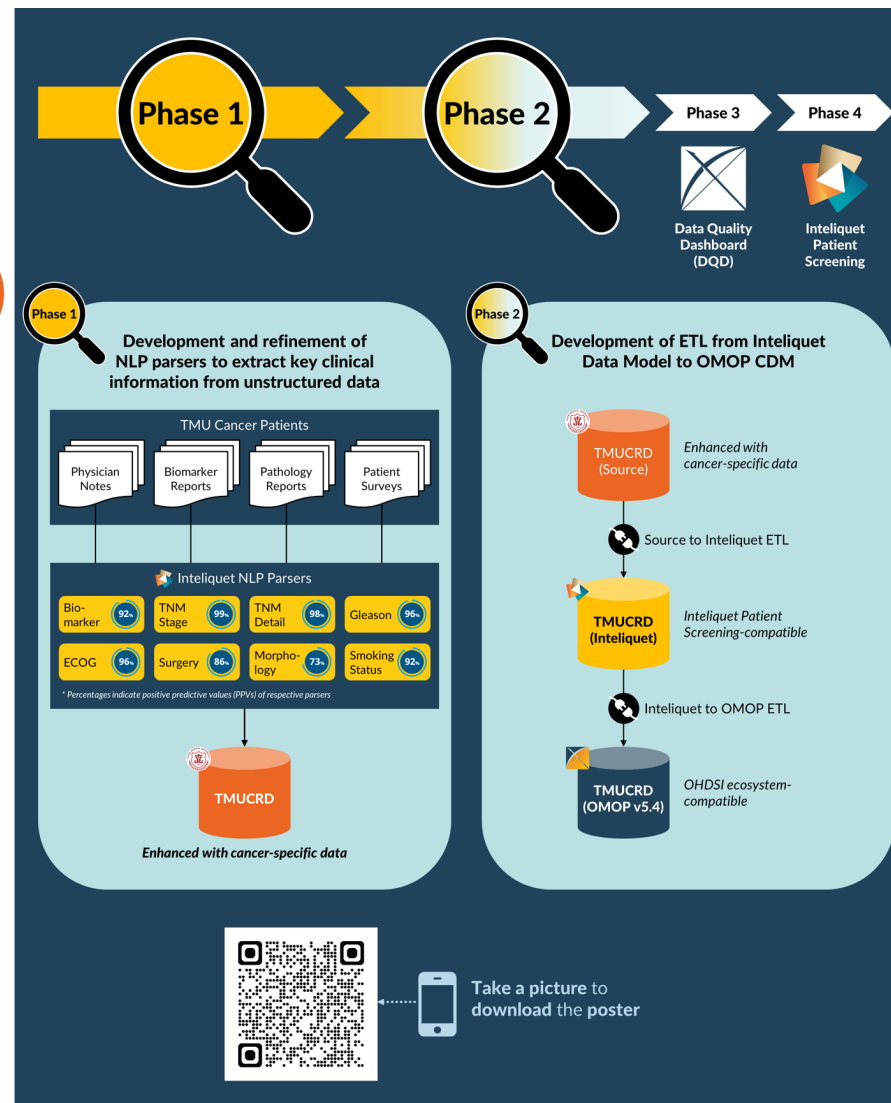
- Various challenges exist in clinical trial feasibility and patient recruitment:



- Recruitment for clinical trials in cancer pose further difficulties due to the complexity and specificity of eligibility criteria
- Data-centric and technology-enabled approaches can tackle these challenges

METHODS

- Phase 1:** IQVIA and Taipei Medical University (TMU) teams are currently collaborating to develop and refine natural language processing (NLP) parsers to extract key clinical information from unstructured data of cancer patients. These parsers enable extraction of cancer-specific data, such as biomarker and TNM information, from various sources. Cancer-specific data extracted will be used to enhance TMUCRD¹.
- Phase 2:** Using the enhanced TMUCRD, IQVIA and TMU are also in parallel developing 2 ETLs: (1) TMUCRD source to Inteliquet for compatibility with the Inteliquet Patient Screening² tool and; (2) Inteliquet to OMOP CDM v5.4 for compatibility with the OHDSI ecosystem and its tools.



METHODS (Continued)

- Phase 3:** Post the development of the Inteliquet to OMOP CDM v5.4. ETL conversion will be validated using OHDSI's Data Quality Dashboard (DQD) package.
- Phase 4:** Once all ETLs are validated, the Inteliquet Patient Screening tool will be installed and implemented into the clinical trial workflows of TMU.

Supplementary Information

- Taipei Medical University Clinical Research Database (TMUCRD) is an integrated database consisting of electronic medical records (EMRs) from three TMU affiliated hospitals - Taipei Medical University Hospital (TMUH), Wan-Fang Hospital (WFH) and Shuang-Ho Hospital (SHH). As of today, the database covers a vast range of medical records for approximately 4.15 million patients, spanning from the year 2004 to 2021.
- Inteliquet Patient Screening is an application that sits behind clinical trial sites' firewalls and is used to assist site staff with querying for eligible patients and matching patients to clinical trials. The tool has a graphical user interface (GUI) that allows users to easily and quickly conduct prescreening for clinical trials and identify potentially eligible patients. Users may also digitize protocols to automatically and continuously match patients to inclusion/exclusion criteria in the protocol against new data being ingested by the tool.

Mui Van Zandt, Jason Hsu, Alex Nguyen, Gyeol Song, Sabrina Cheema, Michael Krupnick





#OHDSISocialShowcase This Week

Wednesday

Utility of Large Language Models for Concept Set Curation

(Adit Anand, Anna Ostropolets, Patrick Ryan, George Hripcsak)



Utility of Large Language Models for Concept Set Curation

Adit Anand¹, Anna Ostropolets^{1,2}, Patrick Ryan^{1,2}, George Hripcsak¹

1. Department of Biomedical Informatics, Columbia University, New York, NY, USA

2. Janssen Research & Development, Raritan, NJ, USA

Key Points

- We evaluate large-language models' (LLMs) ability to curate concepts for EHR phenotyping from PHOEBE recommendations.
- ChatGPT-3.5-Turbo performs better than ChatGPT-4o, with varying performance across conditions, and ChatGPT-3.5-Turbo has better sensitivity than specificity when refining concept sets.
- Further work is needed to identify generalizable prompt engineering strategies across LLMs and determine the optimal role(s) LLMs should play in streamlining EHR phenotyping.

Background

- Constructing rule-based EHR phenotypes requires substantial informatics and clinical expertise. As a result, phenotyping algorithms are currently manually validated in an iterative manner to refine a phenotype's criteria.
- PHOEBE is a concept recommender system that facilitates developing clinical concepts in ATLAS.
- Efforts to generate rule-based phenotyping algorithms with LLMs have recognized the need for a "human-in-the-loop" approach to construct concept sets.
- Using 2 different versions of ChatGPT, we explore the utility of LLMs to curate clinical concepts for EHR phenotyping from PHOEBE recommendations.

Methods

- For each of 4 phenotypes, PHOEBE recommends an initial set of OMOP concepts belonging to the Condition domain.
- For each recommended concept, we prompt zero-shot ChatGPT-3.5 and ChatGPT-4o to generate a True/False answer for whether a recommended concept is specific to the phenotype.
- The concepts that ChatGPT approves are inputted into PHOEBE to produce a second set of recommended Condition concepts.
- This second list is validated by ChatGPT with the same prompt engineering strategy, and the approved concepts from each pass comprise the final concept set.
- We calculate the sensitivity, specificity, and AUROC for each final concept set using a manually-curated, gold-standard concept set.

Contact: adit.anand@columbia.edu

Results

	ChatGPT-3.5-Turbo			ChatGPT-4o		
	Sensitivity	Specificity	AUROC	Sensitivity	Specificity	AUROC
Acute Myocardial Infarction	0.879	0.479	0.679	0.094	0.985	0.539
Pulmonary Hypertension	0.000	0.693	0.347	0.000	1.000	0.500
Rheumatoid Arthritis	0.941	0.430	0.686	0.185	1.000	0.593
Type 1 Diabetes Mellitus	0.952	0.684	0.819	0.000	1.000	0.500

Table 1. Sensitivity, specificity, and AUROC of the concept set generated by PHOEBE recommendations and zero-shot ChatGPT prompting for each phenotype

Phenotype	Concept Name	ChatGPT-4o Explanation
Acute Myocardial Infarction	First myocardial infarction	A "first myocardial infarction" refers to the initial occurrence of a myocardial infarction (heart attack) in an individual. While having a first myocardial infarction means that the person has experienced an acute myocardial infarction (AMI), it is not specific to AMI in the sense that it does not exclusively indicate the presence of AMI. The term "first myocardial infarction" simply denotes the initial event of a heart attack and does not imply that every instance of AMI is a first myocardial infarction. Therefore, the condition "first myocardial infarction" is not specific to the malady "acute myocardial infarction" because it does not meet the criterion that if you have the condition, you definitely have the malady.
Rheumatoid Arthritis	Bilateral deformity of hands due to rheumatoid arthritis	While bilateral deformity of the hands can be a consequence of rheumatoid arthritis (RA), it is not specific to RA. This condition can also be seen in other forms of arthritis and joint diseases, such as osteoarthritis or psoriatic arthritis. Therefore, having bilateral deformity of the hands does not definitively mean that a person has RA, even though it is a common manifestation in advanced cases of RA.
Type 1 Diabetes Mellitus	Type 1 diabetes mellitus uncontrolled	The condition "Type 1 diabetes mellitus uncontrolled" is not specific to Type 1 Diabetes Mellitus (T1DM) in the sense that having uncontrolled T1DM does not definitively mean you have T1DM. The term "uncontrolled" simply indicates that the diabetes, which is already diagnosed as T1DM, is not well-managed or regulated, leading to poor blood glucose control. However, the presence of "uncontrolled" diabetes does not provide new diagnostic information about the type of diabetes; it only describes the state of management of an already diagnosed condition. Therefore, it does not meet the criteria of being specific to T1DM because it does not independently confirm the diagnosis of T1DM. It merely describes the status of a pre-existing diagnosis.

Table 2. The concept names and explanations for three phenotypically-relevant clinical concepts assessed by ChatGPT-4o to be not specific to their corresponding phenotypes

Results

- Table 1 demonstrates how the concept sets derived from using ChatGPT-4o have negligible sensitivity and thus have a statistically insignificant AUROC.
- These findings highlight that ChatGPT-4o performs worse relative to ChatGPT-3.5-Turbo for concept set validation.
- ChatGPT-3.5 Turbo distinguishes symptoms indicating a condition's onset and risk factors from the disease itself.
- ChatGPT-3.5-Turbo overestimates the specificity of symptoms which arise from the phenotype of interest.
- To identify what causes ChatGPT-4o to underperform, Table 2 highlights 3 PHOEBE-recommended concepts that ChatGPT-4o falsely identifies as not specific to their corresponding phenotype.
- ChatGPT-4o fails to account for concepts' semantic similarity and overvalues the distinctness of new information or modifiers.

Conclusions

- Filtering PHOEBE concept recommendations with ChatGPT-3.5-Turbo or ChatGPT-4o is a promising method to curate concept sets for phenotyping algorithms.
- Ongoing work is needed to develop generalizable, robust prompt engineering or few-shot prompting methods.
- These LLM-curated concept sets should be used in phenotyping algorithms and evaluated with existing OHDSI tools.
- Reproducing this study across multiple clinical domains within the OMOP CDM (e.g., Procedure, Drug, Measurement, etc.) would provide a more holistic assessment of our methodology.

References

1. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc*. 2013 Jan 1;20(1):117-21.
2. Hripcsak G, Duke JB, Shah NH, Reich CG, Huser V, Schumm M, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform*. 2015;216:574-8.
3. Banda JM, Seneviratne M, Hernandez-Boussard T, Shah NH. Advances in Electronic Phenotyping: From Rule-Based Definitions to Machine Learning Models. *Annu Rev Biomed Data Sci*. 2018 Jul 20;1(Volume 1):53-68.
4. Zheng NS, Feng Q, Kerckhoffer VE, Zhao J, Edwards TL, Cox NJ, et al. PheMap: a multi-resource knowledge base for high-throughput phenotyping within electronic health records. *J Am Med Inform Assoc*. 2020 Nov 1;27(11):1675-87.
5. Wu P, Gifford A, Meng X, U X, Campbell H, Varley T, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform*. 2019 Nov 29;7(4):e14325.
6. Ostropolets A, Hripcsak G, Knoll C, Ryan P. PHOEBE 2.0: selecting the right concept sets for the right patients using lexical, semantic, and data-driven recommendations. [cited 2024 Jun 7]; Available from: https://www.ohdsi.org/wp-content/uploads/2022/10/6-Ostropolets_Phebe2-0-abstract.pdf
7. Yan C, Ong HH, Grabowska ME, Krantz MS, Su WC, Dickson AL, et al. Large language models facilitate the generation of electronic health record phenotyping algorithms. *J Am Med Inform Assoc*. 2024 Apr 13;30(4):e72.
8. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*. 2012 Jan 1;19(1):54-60.
9. ChatGPT [Internet]. [cited 2023 Dec 13]. Available from: <https://chat.openai.com>



@OHDSI

www.ohdsi.org

#JoinTheJourney



ohdsi



#OHDSISocialShowcase This Week

Thursday

Electronic Frailty index and hazard of with MACE event in patients with Type 2 diabetes mellitus

(Daeun Hyeon, Sujin Gan, Rae Woong Park)



Electronic Frailty index and hazard of with MACE event in patients with Type 2 diabetes mellitus

Daeun Hyeon¹, Sujin Gan¹, Rae Woong Park^{1,2}

¹ Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea

² Department of Medical Sciences, Graduate School of Ajou University, Suwon, Republic of Korea



Background

- Type 2 diabetes mellitus (T2DM) significantly increases the risk of cardiovascular disease (CVD), particularly in older individuals with frailty and multimorbidity.
- Traditional frailty assessment tools are often complex and time-consuming, highlighting the need for EMR-based frailty indices to identify at-risk patients more efficiently.
- This study aims to investigate the association between frailty, as measured through electronic medical records (EMRs), and major adverse cardiovascular events (MACE) in T2DM patients.

Methods

1. Data preparation

- Observational medical outcomes partnership common data model (OMOP-CDM) database at Ajou University School of Medicine (AUSOM)
- Inclusion criteria
 - 1) 40 years and older
 - 2) Diagnosed with type 2 diabetes mellitus (T2DM)
 - 3) No history of major adverse cardiovascular events (MACE) ; myocardial infarction, cardiovascular disorders, acute ischemic heart disease, chronic ischemic heart disease and acute myocardial infarction.

2. Outcome

- Occurrence of MACE

3. Sensitivity analysis

- Dividing the participants into two age groups : 65 years and younger and 66 years and older.

3. Frailty index calculation

- Electronic medical record (EMR) data was used to calculate the Electronic Frailty Index (eFI).
 - The eFI was calculated by summing binarized variables, resulting in a score ranging from 0 to 1.
 - This score was divided by the number of variables per patient, excluding missing values.
 - The maximum value of eFI was divided into thirds, stratifying patients into three groups based on each interval.
 - Patients were categorized as normal, pre-frailty, or frailty based on their FI score.
- ##### 4. Statistical Analysis
- Cox proportional hazards regression model
 - Kaplan-Meier survival curves
 - Log-rank test

Conclusions

- This study shows an association between increasing eFI and the occurrence of MACE in patients with T2DM aged 40 years or older.
- The eFI used in this study has the advantage of not requiring separate frailty testing, and it showed the feasibility of using eFI in OMOP-CDM to screen for CVD risk groups in patients with T2DM.

Acknowledgement

- 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).
- This research was supported by a Government-wide R&D Fund project for infectious disease research (GFID), Republic of Korea (grant number: HG22C0024).

Results

- The risk of MACE was significantly higher in the frailty group compared to the normal group (Hazard Ratio [HR]: 1.68, 95% Confidence Interval [CI]: 1.38-2.04; $P < 0.05$) and in the pre-frailty group compared to the normal group (HR 1.44 (1.35-1.55); $P < 0.05$).

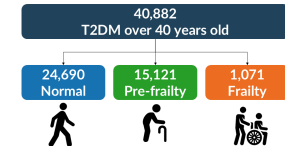


Figure 1. Classification of T2DM patients based on electronic Frailty Index

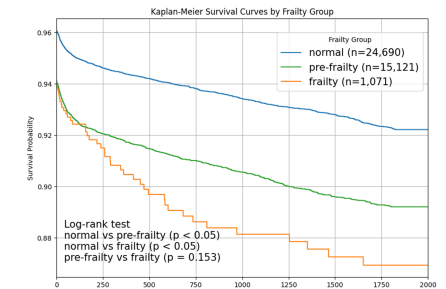


Figure 2. Survival probability curve for normal, pre-frailty, and frailty groups

Age	Frailty category	N	Hazard ratio (95% CI)
Below 65	Normal	15,714	1.0 (ref)
	Pre-Frailty	8,028	1.26 (1.13-1.40)
	Frailty	505	1.28 (0.90-1.83)
Over 65	Normal	8,976	1.0 (ref)
	Pre-Frailty	7,093	1.46 (1.33-1.61)
	Frailty	566	1.70 (1.34-2.16)

Table 1. Hazard ratio for MACE in Subgroup



#OHDSISocialShowcase This Week

Friday

Comparing IRB Review of OHDSI Network Studies: Sharing Experience and Guidance

(Ben Martin, Will Kelly, Christopher Mecoli, Khyzer Aziz, Haeun Lee, Star Liu, Paul Nagy)

Title: *Communicating OHDSI Network Studies to Local IRBs*

PRESENTER: Ben Martin

INTRO:

Progress in data standardization and the adoption of CDMs like OMOP have enabled observational study execution times to go from months to a matter of days. With these robust studies becoming the norm rather than remarkable achievements, it is imperative that informatics faculty and observational health researchers work to optimize their processes for working with the IRB.

METHODS:

- This project involved the collation experience and identification of important considerations for IRB review of OHDSI network studies.
- Individual experiences with the submission of OHDSI network study protocols for IRB review were shared amongst clinical research teams at Johns Hopkins.
- Feedback and guidance from data governance board members and IRB reviewers were examined across different projects and common issues and unique considerations were summarized.
- Common information required for IRB application was compared, and key considerations were compiled to develop a general-purpose IRB template (available below).



A generalized OHDSI IRB Protocol Template developed at Johns Hopkins from this compiled experience is publicly available for data partner support.

Accessible on the OHDSI Evidence Network GitHub repository

Comparing IRB Review of OHDSI Network Studies: Sharing Experience & Guidance



Take a picture to download the full paper

RESULTS:

As was expected with IRB review of observational research, the most common request from IRB reviewers was around the degree of data deidentification.

At JHU, the OMOP data resources are available as separate projections with three distinct levels of patient health information (PHI) sensitivity:

- 1) Fully identifiable projection of the OMOP CDM. This CDM does not contain any direct identifiers such as MRN, or patient names, but does contain fields such as date of birth and encounter dates.
- 2) OMOP CDM Limited Data Set; which has been determined by the Core for Clinical Research Data Acquisition to meet criteria for a HIPAA limited dataset.
- 3) Aggressively deidentified OMOP CDM; which has been determined by the Core for Clinical Research Data Acquisition to meet criteria for a limited dataset and has had sensitive data removed to further prevent reidentification.

ID	OHDSI Network Study	Key IRB Framing Summary	Specific Considerations	Review Type
A	Dermatomyositis Phenotype Development and Evaluation	<ul style="list-style-type: none">Fully identifiable OMOP projection;Lead site for a network study;Characterization via executing CohortDiagnostics and Phevaluator	Use of the patient-level prediction algorithm in the Phevaluator package raised concerns about FDA regulations around software as a medical device	Full Review
B	Impact of concomitant use of proton pump inhibitors and antiplatelet agents on cardiovascular adverse outcomes – A multicenter, multinational study using common data model	<ul style="list-style-type: none">Aggressively deidentified OMOP projection;Lead site for a network study;Patient-level prediction study of adverse drug events	Use of the patient-level prediction algorithm prompted requests for more information around FDA software as a medical device regulations Evaluation of adverse drug events for a broad category of drugs	Expedited Review
C	Estimating the risk of adverse drug events associated with broad-spectrum antibiotics in pediatric patients with sepsis	<ul style="list-style-type: none">Aggressively deidentified OMOP;Data partner for an external network study;Population-level effect estimation	Evaluation of a drug efficacy prompted full IRB review; the specificity of drug evaluation was also a difference between Study B	Full Review
D	Characterization of Maternal Health Outcomes	<ul style="list-style-type: none">Aggressively deidentified OMOP projection;Characterization of comorbidities	Single site study; descriptive analysis only	Exempt Status
E	Toxicity of time for Planned Tracheostomy	<ul style="list-style-type: none">Aggressively deidentified OMOP projection;Characterization of outcomes and healthcare service utilization, phenotype evaluation	Single site study; descriptive analysis and phenotype evaluation using Phevaluator	Exempt Status
F	DbDiagnostics for the OHDSI Data Partner Network	<ul style="list-style-type: none">Aggressively deidentified OMOP projection;Data partner for an external network study;Submission of dbDiagnostics profile	Multi-site study; submission of concept-level counts only (dbDiagnostics profile)	Exempt Status
G	Semaglutide and NAION & DR progression	<ul style="list-style-type: none">Aggressively deidentified OMOP projection;Data partner for an external network study;Population-level effect estimation	Multi-site study; execution of CohortMethod and SelfControlledCaseSeries modules	Change in Research to existing approved IRB

It is vital that the OHDSI community continues to support each other in optimizing our collaborations with local IRB review processes by continuing to share knowledge and experience in this domain.

¹Ben Martin; ²Will Kelly; ³Christopher Mecoli; ⁴Khyzer Aziz; ⁵Haeun Lee; ⁶Star Liu; ⁷Paul Nagy

1. Johns Hopkins School of Medicine, Baltimore, MD, USA.





Job Opening

Senior Program Officer, Clinical AI Innovation, Gates Foundation

Senior Program Officer, Public Health Surveillance, Data Integration, and Artificial Intelligence

Apply

What You'll Do

Are you passionate about using the power of information technology to support better surveillance and public health decision making and thereby improve health and reduce inequality in low- and middle-income countries? Do you have experience working in low- and middle-income countries on AI or digital public health initiatives? If so, we want you to join our team at the largest nonprofit fighting poverty, disease, and inequity around the world.

The Senior Program Officer, Public Health Surveillance, Data Integration, and Artificial Intelligence is a key member of the foundation's Surveillance unit in the Pneumonia and Pandemic Preparedness Team. This role will focus on public health informatics and AI for the Foundation's cross-cutting public health surveillance initiative and will also support other groups in PPP that are considering investments in public health informatics and AI. Beyond the PPP portfolio, this role requires cross-foundation collaboration with surveillance activities with other teams, including Africa Regional Office; Genomics, Epidemiology, & Modeling; Global Policy & Advocacy; and Global Health Agencies & Funds. This role also collaborates with the AI Core Team and Program Strategy Teams on other infectious disease surveillance programs to ensure coordination and integration of surveillance activities. Taking advantage of the promise of AI and preventing an AI digital divide in Africa and other areas of the global south is a high priority area for the Foundation. As such, this role will be responsible for developing our overarching strategy for public health surveillance applications in AI, as well as other applications of information technology to public health surveillance and epidemic readiness. This person will also provide technical assistance to other PSTs and advocate for the safe, responsible use of AI as a force multiplier to improve public health surveillance in LMICs.

- Create and implement a strategy for digital innovation, including AI, to improve actionable, integrated, and timely information for public health decision-makers in the global south
 - Develop a strategy for the Foundation's work on public health informatics and AI for public health surveillance, consistent with our theory of change, with quantifiable impact goals
 - Develop a clear understanding of specific ecosystem constraints and opportunities related to AI for public health surveillance
 - Map the pathways by which promising public health surveillance technologies transition from small, defined proof-of-concepts to minimal viable products and ultimately scaled products/applications that are sustainable across multiple countries
 - Identify a key set of partners and stakeholders for success in this focus area across the technical, advocacy, government, academic and funding spheres
- Use your scientific and technical expertise to identify, develop, and manage a portfolio of strategic investments related to our strategy
 - Review submitted concept notes and grant proposals to support projects and programs on public health informatics for surveillance, including global and national data architecture, norms, and standards; unique health ID and privacy-preserving record linkage; early warning systems for outbreak detection and notification
 - Shape and manage investments on AI tool effectiveness, safety, and regulatory strategies for AI-informed public health surveillance in LMICs
 - Develop an overarching evaluation plan to understand the viability of tools deployed in LMIC settings and why they succeed or fail
 - Ensure that funding is appropriately managed and where appropriate, done in collaboration with other foundation teams and external funders



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?

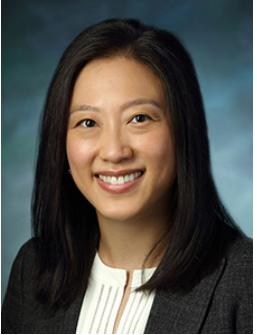
Where Are We Going?





Nov. 19: Evidence Network in Action

The Semaglutide Study



Cindy Cai

Assistant Professor of Ophthalmology
Wilmer Eye Institute at Johns Hopkins Hospital

Topic: Semaglutide and NAION: An OHDSI Network Study



Anthony Sena

Director, Observational Health Data Analytics
Johnson & Johnson

Topic: Strategus



Paul Nagy

Program Director for Graduate Training
in Biomedical Informatics and Data Science
Johns Hopkins University

Topic: Evidence Network



Ben Martin

Postdoctoral Fellow
Johns Hopkins University

Topic: Using the Results Schema



Linying Zhang

Assistant Professor of Biostatistics
Washington University

Topic: Methods



Erik Westlund

Assistant Scientist
Johns Hopkins University

Topic: Using the Results Schema



**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**