OHDSI
Observational Health Data Sciences and Informatics

Welcome to OHDSI

OHDSI Community Call
Oct. 29, 2024 • 11 am ET

Upcoming Community Calls

Date	Topic
Oct. 29	Welcome to OHDSI
Nov. 5	Meet The 2024 Titans
Nov. 12	Next Steps in Evidence Dissemination
Nov. 19	Evidence Network in Action: Semiglutide 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

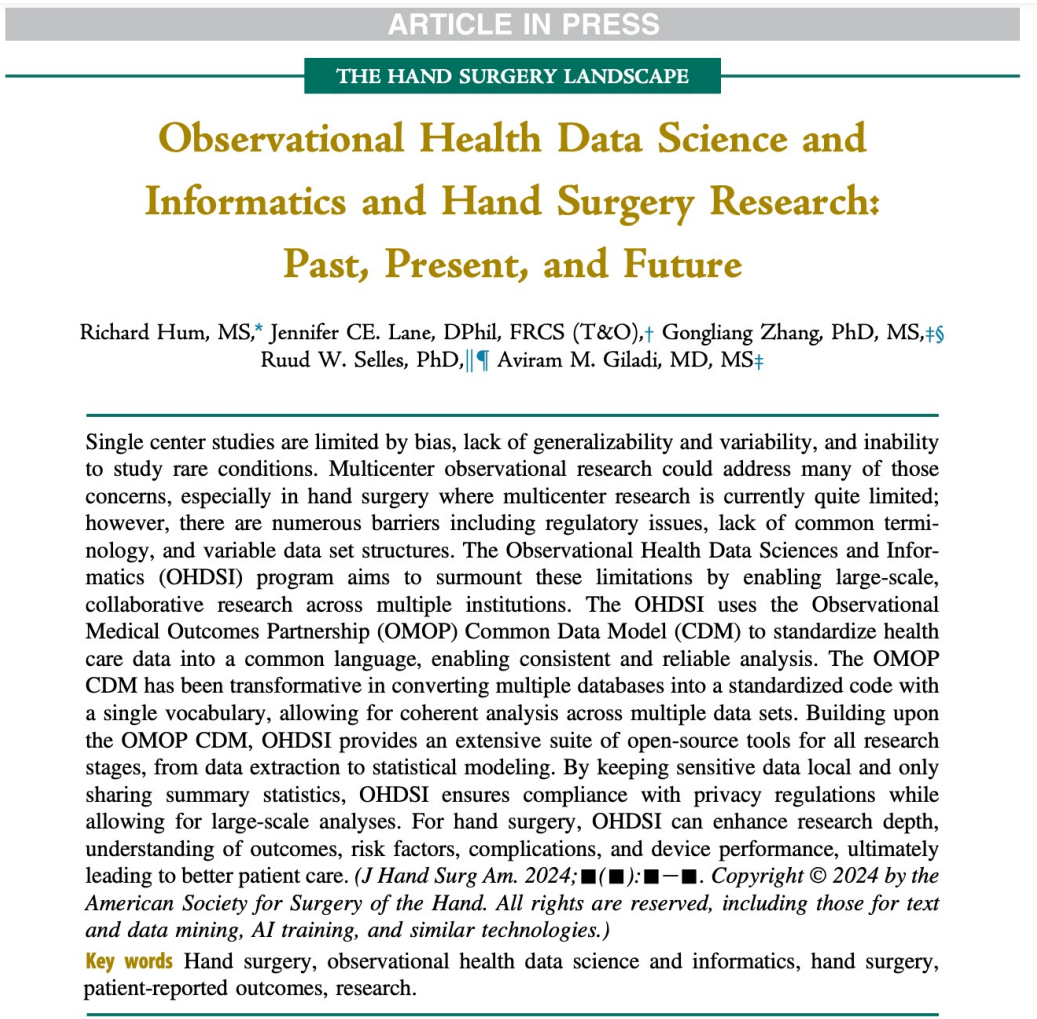
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Richard Hum, Jennifer CE. Lane, Gongliang Zhang, Ruud Selles, and Aviram M. Giladi** on the publication of **Observational Health Data Science and Informatics and Hand Surgery Research: Past, Present, and Future** in *The Journal of Hand Surgery*.



ARTICLE IN PRESS

THE HAND SURGERY LANDSCAPE

Observational Health Data Science and Informatics and Hand Surgery Research: Past, Present, and Future

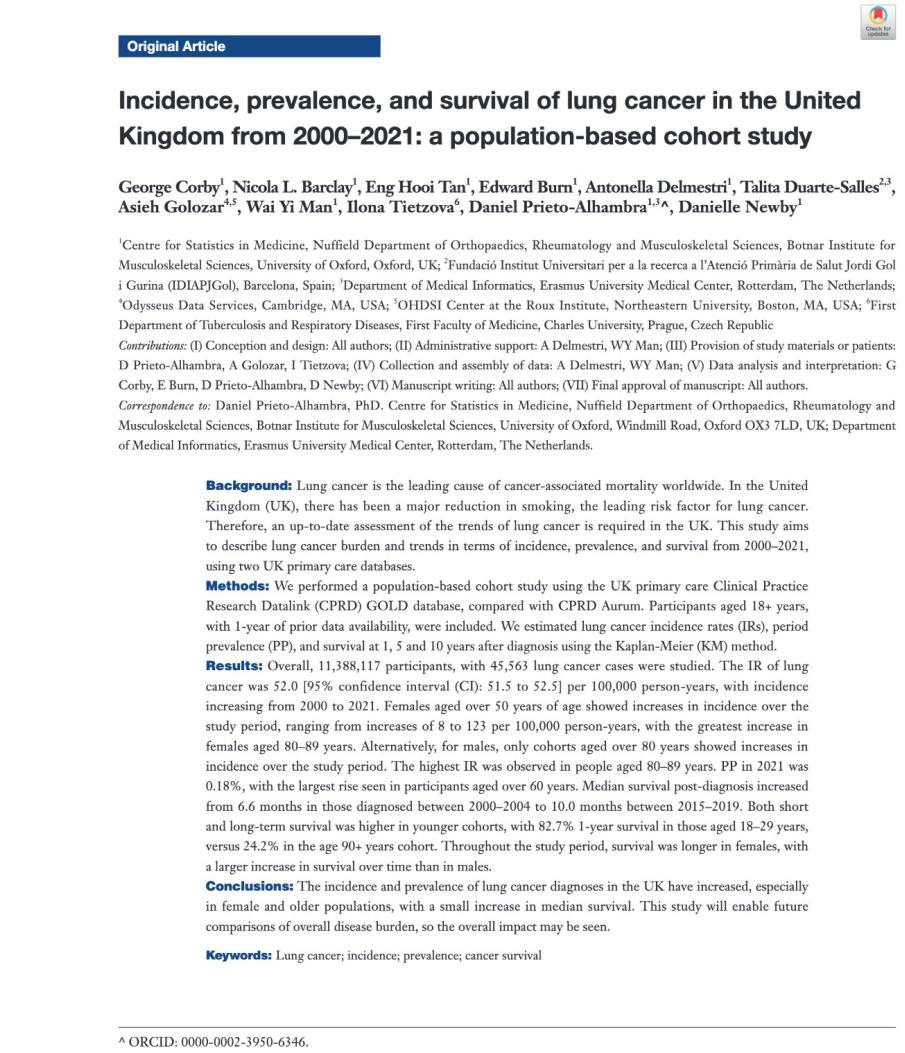
Richard Hum, MS,* Jennifer CE. Lane, DPhil, FRCS (T&O),† Gongliang Zhang, PhD, MS,‡§ Ruud W. Selles, PhD,‖¶ Aviram M. Giladi, MD, MS‡

Single center studies are limited by bias, lack of generalizability and variability, and inability to study rare conditions. Multicenter observational research could address many of those concerns, especially in hand surgery where multicenter research is currently quite limited; however, there are numerous barriers including regulatory issues, lack of common terminology, and variable data set structures. The Observational Health Data Sciences and Informatics (OHDSI) program aims to surmount these limitations by enabling large-scale, collaborative research across multiple institutions. The OHDSI uses the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to standardize health care data into a common language, enabling consistent and reliable analysis. The OMOP CDM has been transformative in converting multiple databases into a standardized code with a single vocabulary, allowing for coherent analysis across multiple data sets. Building upon the OMOP CDM, OHDSI provides an extensive suite of open-source tools for all research stages, from data extraction to statistical modeling. By keeping sensitive data local and only sharing summary statistics, OHDSI ensures compliance with privacy regulations while allowing for large-scale analyses. For hand surgery, OHDSI can enhance research depth, understanding of outcomes, risk factors, complications, and device performance, ultimately leading to better patient care. *(J Hand Surg Am. 2024;■(■):■−■. Copyright © 2024 by the American Society for Surgery of the Hand. All rights are reserved, including those for text and data mining, AI training, and similar technologies.)*

Key words Hand surgery, observational health data science and informatics, hand surgery, patient-reported outcomes, research.

OHDSI Shoutouts!

Congratulations to the team of George Corby, Nicola Barclay, Eng Hooi Tan, Edward Burn, Antonella Delmestri, Talita Duarte-Salles, Asieh Golozar, Wai Yi Man, Ilona Tietzova, Daniel Prieto-Alhambra, and Danielle Newby on the publication of **Incidence, prevalence, and survival of lung cancer in the United Kingdom from 2000–2021: a population-based cohort study** in *Translational Lung Cancer Research*.



Original Article

Incidence, prevalence, and survival of lung cancer in the United Kingdom from 2000–2021: a population-based cohort study

George Corby[1], Nicola L. Barclay[1], Eng Hooi Tan[1], Edward Burn[1], Antonella Delmestri[1], Talita Duarte-Salles[2,3], Asieh Golozar[4,5], Wai Yi Man[1], Ilona Tietzova[6], Daniel Prieto-Alhambra[1,3]^, Danielle Newby[1]

[1]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Institute for Musculoskeletal Sciences, University of Oxford, Oxford, UK; [2]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain; [3]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands; [4]Odysseus Data Services, Cambridge, MA, USA; [5]OHDSI Center at the Roux Institute, Northeastern University, Boston, MA, USA; [6]First Department of Tuberculosis and Respiratory Diseases, First Faculty of Medicine, Charles University, Prague, Czech Republic

Contributions: (I) Conception and design: All authors; (II) Administrative support: A Delmestri, WY Man; (III) Provision of study materials or patients: D Prieto-Alhambra, A Golozar, I Tietzova; (IV) Collection and assembly of data: A Delmestri, WY Man; (V) Data analysis and interpretation: G Corby, E Burn, D Prieto-Alhambra, D Newby; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Daniel Prieto-Alhambra, PhD. Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Institute for Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, UK; Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands.

Background: Lung cancer is the leading cause of cancer-associated mortality worldwide. In the United Kingdom (UK), there has been a major reduction in smoking, the leading risk factor for lung cancer. Therefore, an up-to-date assessment of the trends of lung cancer is required in the UK. This study aims to describe lung cancer burden and trends in terms of incidence, prevalence, and survival from 2000–2021, using two UK primary care databases.

Methods: We performed a population-based cohort study using the UK primary care Clinical Practice Research Datalink (CPRD) GOLD database, compared with CPRD Aurum. Participants aged 18+ years, with 1-year of prior data availability, were included. We estimated lung cancer incidence rates (IRs), period prevalence (PP), and survival at 1, 5 and 10 years after diagnosis using the Kaplan-Meier (KM) method.

Results: Overall, 11,388,117 participants, with 45,563 lung cancer cases were studied. The IR of lung cancer was 52.0 [95% confidence interval (CI): 51.5 to 52.5] per 100,000 person-years, with incidence increasing from 2000 to 2021. Females aged over 50 years of age showed increases in incidence over the study period, ranging from increases of 8 to 123 per 100,000 person-years, with the greatest increase in females aged 80–89 years. Alternatively, for males, only cohorts aged over 80 years showed increases in incidence over the study period. The highest IR was observed in people aged 80–89 years. PP in 2021 was 0.18%, with the largest rise seen in participants aged over 60 years. Median survival post-diagnosis increased from 6.6 months in those diagnosed between 2000–2004 to 10.0 months between 2015–2019. Both short and long-term survival was higher in younger cohorts, with 82.7% 1-year survival in those aged 18–29 years, versus 24.2% in the age 90+ years cohort. Throughout the study period, survival was longer in females, with a larger increase in survival over time than in males.

Conclusions: The incidence and prevalence of lung cancer diagnoses in the UK have increased, especially in female and older populations, with a small increase in median survival. This study will enable future comparisons of overall disease burden, so the overall impact may be seen.

Keywords: Lung cancer; incidence; prevalence; cancer survival

^ ORCID: 0000-0002-3950-6346.

OHDSI Shoutouts!

Congratulations to the team of **Sung Hwan Joo, Seungwon Yang, Suhyun Lee, Seok Jun Park, Taemin Park, Sang Youl Rhee, Jae Myung Cha, Sandy Jeong Rhie, Hyeon Seok Hwang, Yang Gyun Kim, and Eun Kyoung Chung** on the publication of **Trends in Antidiabetic Drug Use and Safety of Metformin in Diabetic Patients with Varying Degrees of Chronic Kidney Disease from 2010 to 2021 in Korea: Retrospective Cohort Study Using the Common Data Model** in *Pharmaceuticals*.

pharmaceuticals MDPI

Article

Trends in Antidiabetic Drug Use and Safety of Metformin in Diabetic Patients with Varying Degrees of Chronic Kidney Disease from 2010 to 2021 in Korea: Retrospective Cohort Study Using the Common Data Model

Sung Hwan Joo [1,2,†], Seungwon Yang [1,2,3,†], Suhyun Lee [3,†], Seok Jun Park [1,2], Taemin Park [1,2,3], Sang Youl Rhee [4,5], Jae Myung Cha [6], Sandy Jeong Rhie [7,8], Hyeon Seok Hwang [9,*], Yang Gyun Kim [10,*] and Eun Kyoung Chung [1,2,3,11,*]

1 Department of Regulatory Science, College of Pharmacy, Graduate School, Kyung Hee University, Seoul 02447, Republic of Korea; shsonic95@khu.ac.kr (S.H.J.); syang345@khu.ac.kr (S.Y.); psjqkr0824@khu.ac.kr (S.J.P.); ptyoon1@gmail.com (T.P.)
2 Institute of Regulatory Innovation through Science (IRIS), Kyung Hee University, Seoul 02447, Republic of Korea
3 Department of Pharmacy, College of Pharmacy, Kyung Hee University, Seoul 02447, Republic of Korea; sh198410@khu.ac.kr
4 Center for Digital Health, Medical Science Research Institute, College of Medicine, Kyung Hee University, Seoul 02447, Republic of Korea; rheesy@khu.ac.kr
5 Department of Endocrinology and Metabolism, School of Medicine, Kyung Hee University, Seoul 02447, Republic of Korea
6 Division of Gastroenterology, Department of Internal Medicine, Kyung Hee University Hospital at Gangdong, School of Medicine, Kyung Hee University, Seoul 05278, Republic of Korea; drcha@khu.ac.kr
7 College of Pharmacy, Ewha Womans University, Seoul 03760, Republic of Korea; sandy.rhie@ewha.ac.kr
8 Graduate School of Pharmaceutical Sciences, Ewha Womans University, Seoul 03760, Republic of Korea
9 Division of Nephrology, Department of Internal Medicine, Kyung Hee University Hospital, Kyung Hee University College of Medicine, Seoul 02447, Republic of Korea
10 Division of Nephrology, Department of Internal Medicine, College of Medicine, Kyung Hee University, Seoul 02447, Republic of Korea
11 Department of Pharmacy, Kyung Hee University Hospital at Gangdong, Seoul 05278, Republic of Korea
* Correspondence: hwanghsne@gmail.com (H.S.H.); apple8840@hanmail.net (Y.G.K.); cekchung@khu.ac.kr (E.K.C.)
† These authors contributed equally to this work.

Abstract: Background/Objectives: This study aimed to investigate trends in antidiabetic drug use and assess the risk of metformin-associated lactic acidosis (MALA) in patients with chronic kidney disease (CKD). Methods: A retrospective observational analysis based on the common data model was conducted using electronic medical records from 2010 to 2021. The patients included were aged ≥18, diagnosed with CKD and type 2 diabetes, and had received antidiabetic medications for ≥30 days. MALA was defined as pH ≤ 7.35 and arterial lactate ≥4 mmol/L. Results: A total of 8318 patients were included, with 6185 in CKD stages 1–2 and 2133 in stages 3a–5. Metformin monotherapy was the most prescribed regimen, except in stage 5 CKD. As CKD progressed, metformin use significantly declined; insulin and meglitinides were most frequently prescribed in end-stage renal disease. Over the study period, the use of SGLT2 inhibitors (13.3%) and DPP-4 inhibitors (24.5%) increased significantly, while sulfonylurea use decreased ($p < 0.05$). Metformin use remained stable in earlier CKD stages but significantly decreased in stage 3b or worse. The incidence rate (IR) of MALA was 1.22 per 1000 patient-years, with a significantly increased IR in stage 4 or worse CKD ($p < 0.001$). Conclusions: Metformin was the most prescribed antidiabetic drug in CKD patients in Korea with a low risk of MALA. Antidiabetic drug use patterns varied across CKD stages, with a notable decline in metformin use in advanced CKD and a rise in SGLT2 inhibitor prescriptions, underscoring the need for further optimized therapy.

check for updates

Citation: Joo, S.H.; Yang, S.; Lee, S.; Park, S.J.; Park, T.; Rhee, S.Y.; Cha, J.M.; Rhie, S.J.; Hwang, H.S.; Kim, Y.G.; et al. Trends in Antidiabetic Drug Use and Safety of Metformin in Diabetic Patients with Varying Degrees of Chronic Kidney Disease from 2010 to 2021 in Korea: Retrospective Cohort Study Using the Common Data Model. *Pharmaceuticals* **2024**, *17*, 1369. https://doi.org/10.3390/ph17101369

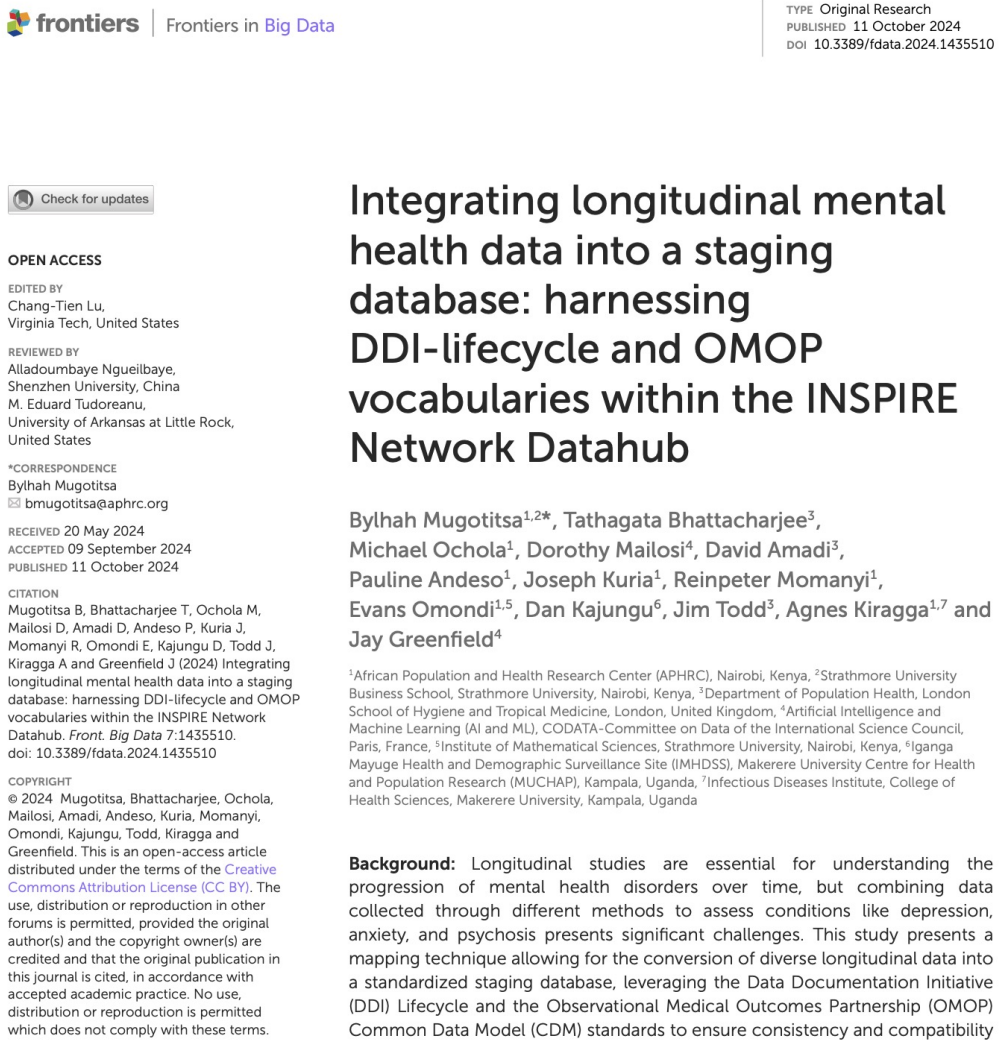
Academic Editor: Agnieszka Sliwinska

Received: 5 September 2024
Revised: 30 September 2024
Accepted: 4 October 2024
Published: 14 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

OHDSI Shoutouts!

Congratulations to the team of **Bylhah Mugotitsa, Tathagata Bhattacharjee, Michael Ochola, Dorothy Mailosi, David Amadi, Pauline Andeso, Joseph Kuria, Reinpeter Momanyi, Evans Omondi, Dan Kajungu, Jim Todd, Agnes Kiragga, and Jay Greenfield** on the publication of **Integrating longitudinal mental health data into a staging database: harnessing DDI-lifecycle and OMOP vocabularies within the INSPIRE Network Datahub** in *Frontiers in Big Data*.



frontiers | Frontiers in Big Data

TYPE Original Research
PUBLISHED 11 October 2024
DOI 10.3389/fdata.2024.1435510

Check for updates

OPEN ACCESS

EDITED BY
Chang-Tien Lu,
Virginia Tech, United States

REVIEWED BY
Alladoumbaye Ngueilbaye,
Shenzhen University, China
M. Eduard Tudoreanu,
University of Arkansas at Little Rock, United States

*CORRESPONDENCE
Bylhah Mugotitsa
bmugotitsa@aphrc.org

RECEIVED 20 May 2024
ACCEPTED 09 September 2024
PUBLISHED 11 October 2024

CITATION
Mugotitsa B, Bhattacharjee T, Ochola M, Mailosi D, Amadi D, Andeso P, Kuria J, Momanyi R, Omondi E, Kajungu D, Todd J, Kiragga A and Greenfield J (2024) Integrating longitudinal mental health data into a staging database: harnessing DDI-lifecycle and OMOP vocabularies within the INSPIRE Network Datahub. *Front. Big Data* 7:1435510. doi: 10.3389/fdata.2024.1435510

COPYRIGHT
© 2024 Mugotitsa, Bhattacharjee, Ochola, Mailosi, Amadi, Andeso, Kuria, Momanyi, Omondi, Kajungu, Todd, Kiragga and Greenfield. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Integrating longitudinal mental health data into a staging database: harnessing DDI-lifecycle and OMOP vocabularies within the INSPIRE Network Datahub

Bylhah Mugotitsa[1,2*], Tathagata Bhattacharjee[3], Michael Ochola[1], Dorothy Mailosi[4], David Amadi[3], Pauline Andeso[1], Joseph Kuria[1], Reinpeter Momanyi[1], Evans Omondi[1,5], Dan Kajungu[6], Jim Todd[3], Agnes Kiragga[1,7] and Jay Greenfield[4]

[1]African Population and Health Research Center (APHRC), Nairobi, Kenya, [2]Strathmore University Business School, Strathmore University, Nairobi, Kenya, [3]Department of Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom, [4]Artificial Intelligence and Machine Learning (AI and ML), CODATA-Committee on Data of the International Science Council, Paris, France, [5]Institute of Mathematical Sciences, Strathmore University, Nairobi, Kenya, [6]Iganga Mayuge Health and Demographic Surveillance Site (IMHDSS), Makerere University Centre for Health and Population Research (MUCHAP), Kampala, Uganda, [7]Infectious Diseases Institute, College of Health Sciences, Makerere University, Kampala, Uganda

Background: Longitudinal studies are essential for understanding the progression of mental health disorders over time, but combining data collected through different methods to assess conditions like depression, anxiety, and psychosis presents significant challenges. This study presents a mapping technique allowing for the conversion of diverse longitudinal data into a standardized staging database, leveraging the Data Documentation Initiative (DDI) Lifecycle and the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) standards to ensure consistency and compatibility across datasets.

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

OHDSI Shoutouts!

Three Stages of The Journey

Where Have We Been?

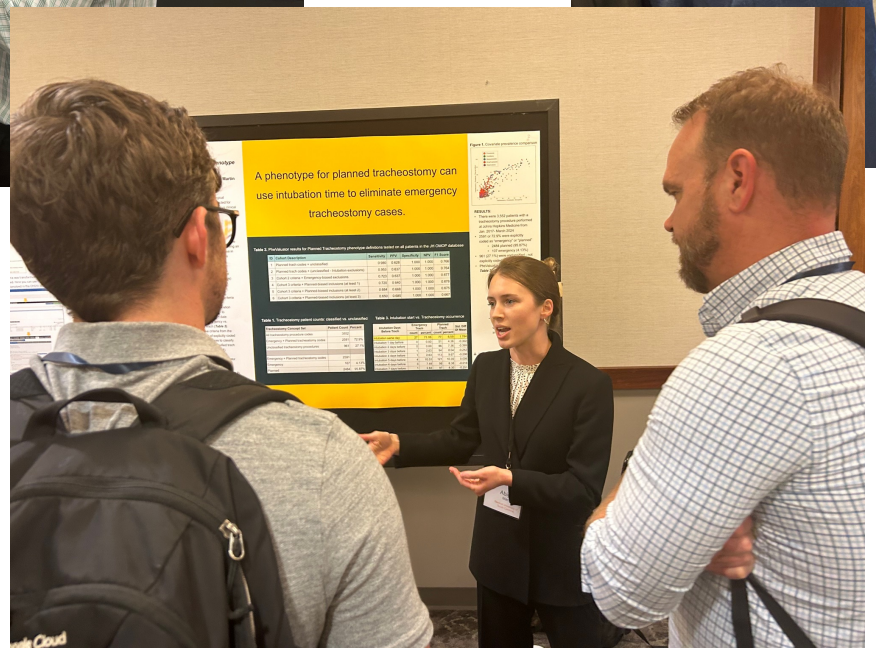
Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	CDM Vocabulary Subgroup
Wednesday	7 am	Medical Imaging
Wednesday	10 am	Surgery and Perioperative Medicine
Wednesday	4 pm	Joint Vulcan/OHDSI Meeting
Thursday	8 am	Medical Devices
Thursday	7 pm	Dentistry
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Steering
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Healthcare Systems Interest Group
Tuesday	9 am	Atlas
Tuesday	10 am	Common Data Model

Vote For #OHDSI2024 Showcase Honors



NEI/OHDSI Session: Oct. 30, 12 pm ET

National Eye Institute
Research Today...Vision Tomorrow

Search the site | Search

Learn About Eye Health | Grants and Training | Research at NEI | About NEI

Home » About NEI » News and Events » Events » OHDSI Research Opportunities: Harnessing Healthcare Databases for Improved Outcomes

OHDSI Research Opportunities: Harnessing Healthcare Databases for Improved Outcomes

October 30, 2024
12:00 PM to 1:00 PM ET

Scheduled Speakers
Michelle Hribar
Cindy Cai
Patrick Ryan

2024 APAC Symposium

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

Registration is OPEN!

Preliminary Dates To Know

Oct. 31: Notification of Acceptance

Symposium Agenda

Dec. 4: Tutorial at NUS

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

Dec. 7-8: Datathon at NUS

ohdsi.org/APAC2024

2024 APAC Symposium

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

Day 1 (December 4) – Tutorial at NUS

9:00 – 12:00 • Introduction of OHDSI/OMOP, ETL Process
12:00 – 13:00 • Lunch
13:00 – 17:00 • OHDSI Analytical Tools

Day 2-3 (December 5-6) – Main conference at Marina Bay Sands

Dec. 5

13:30 – 13:40 • Opening
13:40 – 14:10 • OHDSI for Real-World Evidence (RWE)
14:10 – 15:00 • OHDSI APAC Regional Chapter Updates
15:00 – 15:30 • Break
15:30 – 15:45 • OHDSI APAC Updates
15:45 – 16:45 • Community-Wide ETL Project: Recap and Lessons Learned
16:45 – 17:05 • Large Language Model and OHDSI
17:05 – 17:25 • HL7 Singapore Chapter and OHDSI Singapore Chapter Collaboration
17:25 – 17:30 • Closing

Dec. 6

9:00 – 9:40 • Opening
9:40 – 9:50 • Introduction of 2024 APAC Study
9:50 – 12:00 • 2024 APAC Study: Journey from Data to Evidence
12:00 – 13:30 • Lunch and Poster Presentations
13:30 – 14:30 • 2024 APAC Study: Panel Discussion
14:30 – 15:15 • Lightning Talks
15:15 – 15:30 • Closing

ohdsi.org/APAC2024

The Center for Advanced Healthcare Research Informatics (CAHRI) October talk has been RESCHEDULED.

Dr. Vipina Keloth October 31st talk:

'Exploring the realm of large language models for information extraction in the biomedical domain'

is rescheduled due to a conflict with another Tufts center symposium.

A new calendar invite has already been sent to those who previously reached out.

New date: January 30, 2025, 11am-12pm EST
Virtually via Zoom

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

TuftsMedicine
Tufts Medical Center

#OHDSISocialShowcase

MONDAY

Harmonization of Biobanco-iMM Rheumatology Collection data to OMOP CDM

(Catarina Tomé, Enrico Calanchi, Laura Delsante, Ângela Afonso, Daniel Silva, Ana Rita Lopes and João Eurico Fonseca)

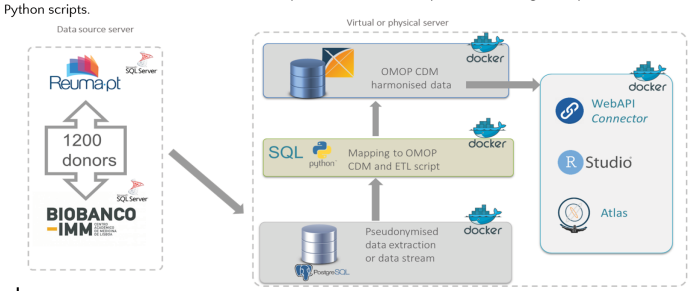
ADAPTION OF THE OMOP CDM FOR RHEUMATOLOGY:
A Portuguese experience.

Harmonization of Biobanco-iMM Rheumatology Collection data to OMOP CDM

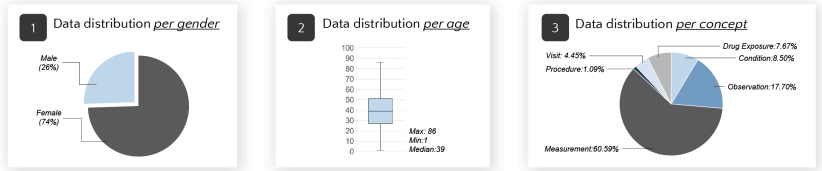
Background: The OMOP Common Data Model (OMOP CDM) is an option to store patient data and to use these in an international context.
Biobanco-iMM includes biological samples (from surgery, biopsies, blood samples) which are voluntarily donated with permission for preservation and future use in biomedical research. Biobanco-iMM has a collection for rheumatology that is connected with Reuma.pt - the Rheumatic Diseases Portuguese Register from Portuguese Society of Rheumatology (SPR).
Reuma.pt protocols include structured information about socio-demographic data and information about standard diagnostic criteria and clinical features of each disease.
This project aimed to convert the Biobanco-iMM Rheumatology Collection data to the OMOP Common Data Model (CDM) version 5.4, using data from Biobanco-iMM and Reuma.pt databases.

Methods

Source data was extracted from the Biobanco-iMM and Reuma.pt databases to be loaded onto a PostgreSQL DBMS instance (postgreSQL 10+190ubuntu0.1), where CDM was also implemented. ETL was implemented through SQL procedures and orchestrated through Python scripts.



Results



Conclusion: Data harmonization, besides to contribute to the data quality improvement, also facilitates the development of relevant clinical projects. In addition, the implementation of country nodes strongly improves its outcome. In the future, CDM materialization will be manually updated based on the needs expressed by Biobanco-iMM and Rheumatology Service.

Limitation: A few Biobanco-iMM Rheumatology Collection data was not standardized due to the lack of appropriate standard vocabulary.

OHDSI

Catarina Tomé, Enrico Calanchi, Laura Delsante, Ângela Afonso,
Daniel Silva, Ana Rita Lopes and João Eurico Fonseca

BIOBANCO -IMM

#OHDSISocialShowcase This Week

TUESDAY

Leveraging FHIR for a generic EHR-to-OMOP ETL

(Sebastiaan van Sandijk, Renske Los)

Leveraging FHIR for a generic EHR-to-OMOP ETL
Can we make an ETL process reusable?

PRESENTER: **Sebastiaan van Sandijk**

INTRO

- **Harmonization of EHR data can be cumbersome.** The speed, effort, and quality of a conversion of EHR data to OMOP heavily depend on factors that are not always transparent in the resulting CDM.
- One way to improve data harmonization and to reduce the effort involved in **OMOPing EHR data**, is to (re)use a **generic ETL**.
- There are currently **no detailed guidelines** for a generic process.
- We explore how EHR data sets extracted according to standard, **FHIR-based specifications** can provide a **standardized basis for a generic OMOP-ETL**.

METHODS

1. Analysis of the specifications of the Dutch BGZ '**Basisgegevensset Zorg**', **European Patient Summary** (EPS) and **International Patient Summary** (IPS).
2. Preliminary **comparison of test data** created according to these specifications. A comparison of extractions from three different EHR systems in underway.
3. **Discussions** with European standardization experts and projects about evolving developments (IPS in particular)

RESULTS

- Data elements included in **FHIR-based** data sets are **more clearly specified** and could provide a useful basis for a generic ETL.
- The BGZ is a structured data set extracted in separate queries. The IPS is **a document with structured data elements**. The latter will improve harmonization of extracted data from different EHR systems.

FHIR-based standards like the **International Patient Summary (IPS)** can be a useful basis for a **generic OMOP-ETL**. **Metadata extensions are needed** to accommodate research requirements.

- However, also the IPS is a 'snapshot' of what is in the EHR system, and the specifications do not require metadata about context, scope and completeness of the extracted data. This would result in a lack of transparency about the (observation)period covered and completeness of the data in the summary.
- Further exploration of data sets that are extracted from various EHR systems (conformant to BGZ and/or IPS specifications) is essential.
- So far, we anticipate different vendors to implement the standards in more or less slightly different manners.
- Because these are FHIR-based specifications, however, we expect that the extracted data sets still can be processed in generic data profiling, mapping, and transformation processes. This would include semantic mappings, making use of the specified (inter)national coding systems.

CONCLUSION

- Standardized FHIR-based data sets may well prove to form a useful basis for a reusable ETL process.
- The specifications currently show important gaps, especially regarding metadata that are important for research and researchers.
- Detailed comparison of practical extractions from different EHR systems is still needed.
- We work towards recommendations for (metadata) extensions in the IPS specification.

Sebastiaan van Sandijk[1,3]
Renske Los [2,3]
[1] Odysseus Data Services, [2] Erasmus Medical Center, [3] OHDSI NL
*Sebastiaan.van.Sandijk@odysseusinc.com

ODYSSEUS DATA SERVICES INC
Erasmus MC University Medical Center Rotterdam
OHDSI

#OHDSISocialShowcase

WEDNESDAY

dsOMOP: Federated Analysis of Harmonized Clinical Data Combining OMOP CDM and DataSHIELD in a DATOS-CAT Cohort Use Case

(David Sarrat-González, Judith Martinez-Gonzalez, Xavier Escribà-Montagut, Aikaterini Lymperidou, Ramón Mateo, Rafael de Cid, Juan R González and Alberto Labarga)

dsOMOP: Federated Analysis of Harmonized Clinical Data Combining OMOP CDM and DataSHIELD in a DATOS-CAT Cohort Use Case

David Sarrat-González[1], Judith Martinez-Gonzalez[2,3], Xavier Escribà-Montagut[1], Aikaterini Lymperidou[3,4], Ramón Mateo[1,3], Rafael de Cid[4], Juan R González[1] and Alberto Labarga[2]

[1]Barcelona Institute for Global Health (ISGlobal). [2]Barcelona Supercomputing Center (BSC). [3]Institute for Bioengineering of Catalonia (IBEC). [4]Genomes for Life - GCAT Lab - Germans Trias i Pujol Research Institute (IGTP).

Introduction

The **DATOS-CAT project** aims to enhance the visibility and scientific impact of population-based cohorts developed in Catalonia, such as GCAT | Genomes for Life and the COVICAT-CONTENT subcohort. These cohorts provide valuable data for research, but the variability and sensitive nature of the data pose significant challenges, especially in terms of data sharing between institutions. To address these issues, the project has adopted the **Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)** to enable **standardized analysis** and **DataSHIELD** technology to perform **federated non-disclosive statistical analyses**.

What is DataSHIELD?

DataSHIELD is a technology that enables the analysis of sensitive health data across different research sites **without physically sharing the data**. It works by allowing statistical analyses to be performed on individual-level data, but these data never leave their original location. Instead, DataSHIELD **transfers the commands to the data** and then **aggregates the results from each site**. This means researchers can collaborate and analyze combined datasets from various cohorts without ever accessing the raw, sensitive data directly. This allows DATOS-CAT to maintain **patient confidentiality** and adhering to **privacy regulations** (such as GDPR) while still enabling a thorough and collaborative analysis of clinical data across its multiple participant sources.

Figure 1: Multi-site DataSHIELD infrastructure architecture using Opal servers

The dsOMOP package

The **dsOMOP package** is designed to facilitate the interaction with remote databases formatted in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) within a DataSHIELD environment. It provides a suite of functions that allow users to **fetch** and **transform data** from these databases into a format that is **intelligible** and **usable** within the DataSHIELD analytical workflow. This direct integration ensures that data analysis complies with the DataSHIELD's **security model** and **disclosure checks** system, which are crucial for maintaining the privacy and security of the data.

Structure

The dsOMOP ecosystem comprises **two essential components** designed to work in tandem: the server-side package (**dsOMOP**) and the client-side package (**dsOMOPClient**).

- **dsOMOP** is installed on the DataSHIELD server by the institution's data manager and is responsible for direct interactions with the OMOP CDM databases. It retrieves, transforms, and integrates the data in a format compatible with DataSHIELD's analytical tools.
  Code available at: **https://github.com/isglobal-brge/dsOMOP**
- **dsOMOPClient** is used locally by researchers and data analysts, it orchestrates the communication between the dsOMOP package on the server and the target OMOP CDM database, allowing for the construction of use-case specific datasets.
  Code available at: **https://github.com/isglobal-brge/dsOMOPClient**

Collaborating centers

Dataset construction

The dsOMOP package is designed to enable users to orchestrate the **creation of datasets** that meet their **specific research needs**. The package allows users to **define their data queries** with precision, ensuring that only the relevant data is fetched. This process is facilitated by dsOMOP's robust support for database interaction, which includes methods for examining table contents, columns, and concept catalogs. Users can filter data based on their specific criteria, making the data selection process both flexible and efficient.

In addition to data selection, dsOMOP **automatically translates all concepts** present in the database into a format that is **easily understandable** for users. This is achieved by utilizing the concept names from the registered vocabularies within the OMOP CDM. As a result, the retrieved data is presented in a **user-friendly format**, allowing researchers to quickly grasp the content and context of their datasets.

The package also handles the **categorization of data** based on its nature. For instance, it can distinguish between **longitudinal data**, which involves multiple records over time for the same entity, and **relational data**, which links multiple instances across different tables. This automatic categorization ensures that data is always served in the **most appropriate format** according to its nature, making it **readily accessible** for subsequent analysis. All of this is performed without the need for supervision by the researcher.

Support for community extensions

While dsOMOP acts as an interface between DataSHIELD servers and OMOP CDM databases, the **potential for automation** or **streamlining** of processes through the creation of supplementary functions, scripts, and packages is vast.

An example of this approach is **dsOMOPHelper**, a complementary package we have developed alongside dsOMOP. This package **significantly reduces the complexity** of using dsOMOP for most simple use cases, where data from an OMOP CDM database may be used for epidemiological studies within the DataSHIELD environment.

Code available at: **https://github.com/isglobal-brge/dsOMOPHelper**

We **strongly encourage** the community to develop tools that build upon dsOMOP, tailoring them to specific use cases and research needs. Such community-driven development not only enhances the utility of dsOMOPClient but also fosters a collaborative ecosystem around the combined use of both DataSHIELD and OMOP CDM.

Results

Thanks to the development and implementation of the dsOMOP package, we have been enabled to effectively **combine and integrate** the use of various **core components** of the DATOS-CAT project. This integration has allowed us to successfully perform **comprehensive** and **federated analyses** on **harmonized clinical data**.

Specifically, we have been able to:

- **Integrate data** from the various participant cohorts into the DataSHIELD environment in a **standardized format** using the OMOP CDM format.
- **Conduct epidemiological studies** while maintaining strict adherence to **privacy regulations** and ensuring **data confidentiality**.
- Leverage the full suite of **DataSHIELD analytical tools** to perform **non-disclosive statistical analyses** on combined datasets from **multiple sources**.
- Facilitate **data sharing** and **collaboration** between different research centers, enabling a more integrated and **cooperative research environment**.
- Enhance the ability to perform **cross-cohort analyses**, **combining insights** from multiple datasets to generate more robust and comprehensive **research findings**.

The dsOMOP package has proven to be a crucial asset in achieving the project's objectives, enabling **efficient**, **secure**, and **collaborative** data analysis across diverse and sensitive clinical datasets.

Funding

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
Next Generation Catalunya
Generalitat de Catalunya
Funded by the "Complementary Plan for Biotechnology Applied to Health," coordinated by the Institute for Bioengineering of Catalonia (IBEC) within the framework of the Recovery, Transformation, and Resilience Plan (C17.I1) - Funded by the European Union - NextGenerationEU

#OHDSISocialShowcase

THURSDAY

Cloud-based, Automated Solution for Transforming Clinical Datasets to the OMOP CDM

(Simon Thompson, Abigail Carter)

Cloud-based, Automated Solution for Transforming Clinical Datasets to the OMOP CDM

Simon Thompson & Abigail Carter
Genomics England

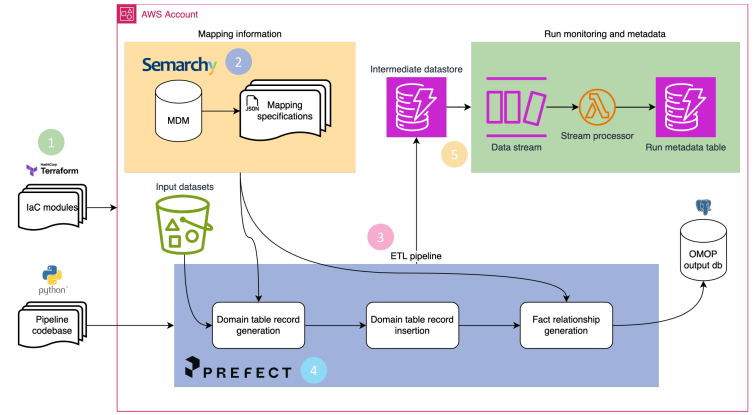
BACKGROUND

Genomics England is a public-funded body that partners with the NHS to provide whole genome sequencing diagnostics and equip researchers with a large genomic database, the National Genomic Research Library (NGRL). The NGRL holds clinical data on over 100,000 patients. The information is currently held in separate data models without standardised concepts or enumerations. There is a desire to transform as much of this diverse dataset as possible to the OMOP CDM.

PIPELINE REQUIREMENTS

The pipeline should:

- be cloud-based (in AWS) in its entirety and use serverless technology where appropriate,
- be high-throughput and process data in parallel,
- store mapping information in a format that is easily interrogated and updated,
- be secure, have a high level of testing and monitoring, and capture failing records at runtime.



SOLUTION

1. Infrastructure is deployed using Terraform (an infrastructure-as-code tool).
2. Mapping information for all datasets and concepts is stored in a Master Data Management (MDM) platform, Semarchy (see poster by Carter et al. for further information).
3. Individual pipeline components (written in Python) for generating OMOP domain table records, generating fact relationship records, and inserting records into a database, are generic and parametised by the MDM platform.
4. Prefect orchestrates the pipeline and manages the parallel execution of processes.
5. Dynamodb, a serverless NoSQL database, is used as an intermediate datastore and feeds processes that monitor each pipeline run. Data is encrypted at rest and in transit throughout.

Simon Thompson
Data Engineer – Research Data Products Squad
simon.thompson@genomicsengland.co.uk

Abigail Carter
Data Wrangler – Research Data Layer Squad
abigail.carter@genomicsengland.co.uk

#OHDSISocialShowcase

Job Opening

Senior Program Officer, Clinical AI Innovation, Gates Foundation

Senior Program Officer, Clinical AI Innovation

Apply

Seattle, WA

Full time

Posted 6 Days Ago

B020184

The Foundation

We are the largest nonprofit fighting poverty, disease, and inequity around the world. Founded on a simple premise: people everywhere, regardless of identity or circumstances, should have the chance to live healthy, productive lives. We believe our employees should reflect the rich diversity of the global populations we aim to serve. We provide an exceptional benefits package to employees and their families which include comprehensive medical, dental, and vision coverage with no premiums, generous paid time off, paid family leave, foundation-paid retirement contribution, regional holidays, and opportunities to engage in several employee communities. As a workplace, we're committed to creating an environment for you to thrive both personally and professionally.

Your Role

Are you passionate about using the power of AI to reduce inequality in low- and middle-income countries? Do you have experience working in developing countries on AI and digital 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, Clinical AI Innovation is a key member of the AI team. This role will support several teams at the Foundation who are considering and investing in multiple applications of AI in Health, which is a high priority area for the Foundation. As such, this individual will be responsible for developing our overarching strategy to healthcare applications in AI; conceptualising, investing and managing investments in health applications of AI; providing advice and technical assistance to other program teams considering investment in this area; advocate for the safe, responsible use of AI as force multiplier to reducing inequality in health in LMICs.

What You'll Do

Develop the foundations' approach to AI and health

- Ensure we have an approach to evaluation of clinical AI applications/ use cases
- This would include existing and planned investment in multiple applications of AI in health across diagnostics, end user engagement, decision support and decision sciences for health
- Develop a clear understanding of specific ecosystem constraints and opportunities related to AI in health
- Identify a key set of partners and stakeholders in order to be successful in this focus area across the technical, advocacy, government, academic and funding spheres

Where Are We Going?

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

Please feel free to promote your #OHDSI2024 workshop or workgroup activity!

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

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