



HowOften: Findings, Current Projects & Next Steps

OHDSI Community Call
July 16, 2024 • 11 am ET



Upcoming Community Calls

Date	Topic
July 16	HowOften Initiative & Early Results
July 23	Building The OHDSI Evidence Network Sprint
July 30	Advances in Patient-Level Prediction
Aug. 6	Building The OHDSI Evidence Network Sprint
Aug. 13	Global Symposium Tutorials
Aug. 20	Building The OHDSI Evidence Network Sprint
Aug. 27	canceled due to ISPE 2024
Sept. 3	New Standardized Vocabularies Release



July 23: Building The Evidence Network, Session II



Clair Blacketer

Director, Epidemiology Analytics, Janssen Research & Development, Inc.



Paul Nagy

Deputy Director, Johns Hopkins Medicine Technology Innovation Center
Director of Education, Biomedical Informatics and Data Science Graduate
Training Programs



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!



Congratulations to the team of
**Kyulee Jeon, Woo Yeon Park,
Charles E Kahn Jr, Paul Nagy, Seng
Chan You, and Soon Ho Yoon** on the
publication of **Advancing Medical
Imaging Research Through
Standardization: The Path to Rapid
Development, Rigorous Validation,
and Robust Reproducibility in
Investigative Radiology.**

REVIEW ARTICLE

OPEN

Advancing Medical Imaging Research Through Standardization *The Path to Rapid Development, Rigorous Validation, and Robust Reproducibility*

Kyulee Jeon, BS, Woo Yeon Park, MS, Charles E. Kahn, Jr, MD, MS, FACR, Paul Nagy, PhD,
Seng Chan You, MD, PhD, and Soon Ho Yoon, MD, PhD^{ORCID}

Abstract: Artificial intelligence (AI) has made significant advances in radiology. Nonetheless, challenges in AI development, validation, and reproducibility persist, primarily due to the lack of high-quality, large-scale, standardized data across the world. Addressing these challenges requires comprehensive standardization of medical imaging data and seamless integration with structured medical data.

Developed by the Observational Health Data Sciences and Informatics community, the OMOP Common Data Model enables large-scale international collaborations with structured medical data. It ensures syntactic and semantic interoperability, while supporting the privacy-protected distribution of research across borders. The recently proposed Medical Imaging Common Data Model is designed to encompass all DICOM-formatted medical imaging data and integrate imaging-derived features with clinical data, ensuring their provenance.

The harmonization of medical imaging data and its seamless integration with structured clinical data at a global scale will pave the way for advanced AI research in radiology. This standardization will enable federated learning, ensuring privacy-preserving collaboration across institutions and promoting equitable AI through the inclusion of diverse patient populations. Moreover, it will facilitate the development of foundation models trained on large-scale, multimodal datasets, serving as powerful starting points for specialized AI applications. Objective and transparent algorithm validation on a standardized data infrastructure will enhance reproducibility and interoperability of AI systems, driving innovation and reliability in clinical applications.

Key Words: radiology, diagnostic imaging, data standardization, observational study, artificial intelligence, reproducibility of results, multimodal data analysis, federated analysis

(Invest Radiol 2025;00: 00-00)

Since 2010, there has been a remarkable increase in the number of published papers utilizing artificial intelligence (AI) in medical research.¹ Notably, one fifth of these publications dealt with medical imaging, which emerged as the most significant area in the paradigm shift of medical research toward AI.² This trend reflects the fact that the field of radiology has been at the forefront of AI research within the medical domain.

The predominance of radiology in medical AI research stems from multiple factors. The advancements in deep learning for computer vision, especially since the development of AlexNet in 2012,³ have significantly enhanced the field of medical imaging.⁴ These technological breakthroughs have achieved unprecedented precision in tasks essential to radiological analysis, such as image classification, object detection, and segmentation.^{3,5,6} Meanwhile, the progress in computer vision has been facilitated by the assembly of extensive datasets such as ImageNet, which is openly accessible and comprises over 14 million annotated images.⁷ However, constructing comparable datasets in the medical field remains largely impractical. Medical data are not primarily gathered for research purposes but are recorded during the delivery of patient care, which vary widely according to the practices of each healthcare institution. Consequently, the data exhibit significant variations in format and content both across and within institutions, making it exceptionally challenging to standardize, manage, or amalgamate effectively.

Unlike in other healthcare fields, the widespread adoption of the Digital Imaging and Communications in Medicine (DICOM) standard has been pivotal in advancing radiological studies. As DICOM has been implemented across almost every device, it allows for the integration of medical images from various sources within Picture Archiving and Communication Systems (PACS).⁸⁻¹¹ This integration has been further

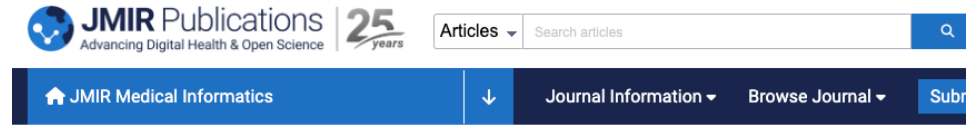
Downloaded from https://www.ahajournals.org/ by guest on 07/19/2024



OHDSI Shoutouts!



Congratulations to the team of **Hyerim Ji, Seok Kim, Leonard Sunwoo, Sowon Jang, Ho-Young Lee, and Sooyoung Yoo** on the publication of **Integrating Clinical Data and Medical Imaging in Lung Cancer: Feasibility Study Using the Observational Medical Outcomes Partnership Common Data Model Extension** in *JMIR Medical Informatics*.



Published on 12.7.2024 in Vol 12 (2024)
Preprints (earlier versions) of this paper are available at <https://preprints.jmir.org/preprint/59187>, first published April 04, 2024.



Integrating Clinical Data and Medical Imaging in Lung Cancer: Feasibility Study Using the Observational Medical Outcomes Partnership Common Data Model Extension

Hyerim Ji^{1,2}; Seok Kim¹; Leonard Sunwoo³; Sowon Jang³; Ho-Young Lee^{1,4}; Sooyoung Yoo¹

Article Authors Cited by Tweetations Metrics

- Abstract
- Introduction
- Methods
- Results
- Discussion
- References
- Abbreviations
- Copyright

Abstract

Background:

Digital transformation, particularly the integration of medical imaging with clinical data, is vital in personalized medicine. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) standardizes health data. However, integrating medical imaging remains a challenge.

Objective:

This study proposes a method for combining medical imaging data with the OMOP CDM to improve multimodal research.

Methods:

Our approach included the analysis and selection of digital imaging and communications in medicine header tags, validation of data formats, and alignment according to the OMOP CDM framework. The Fast Healthcare Interoperability Resources ImagingStudy profile guided our consistency in column naming and definitions. Imaging Common Data Model (I-CDM), constructed using the entity-attribute-value model, facilitates scalable and efficient medical imaging data management. For patients with lung cancer diagnosed between 2010 and 2017, we introduced 4 new tables—IMAGING_STUDY, IMAGING_SERIES, IMAGING_ANNOTATION, and FILEPATH—to standardize various imaging-related data and link to clinical data.



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
Wednesday	4 pm	Vulcan/OHDSI
Thursday	9 am	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	9:30 am	Themis
Thursday	12 pm	Medical Devices
Friday	10 am	GIS-Geographic Information System
Friday	10:30 am	Open-Source Community
Friday	11:30 am	Steering Group
Friday	11:30 am	Clinical Trials
Monday	10 am	Africa Chapter
Monday	10 am	CDM Survey Subgroup
Tuesday	9 am	OMOP CDM Oncology Vocabulary/Development Subgroup

Is Semaglutide Associated with Blinding Eye Diseases??

JAMA Ophthalmology | **Original Investigation**

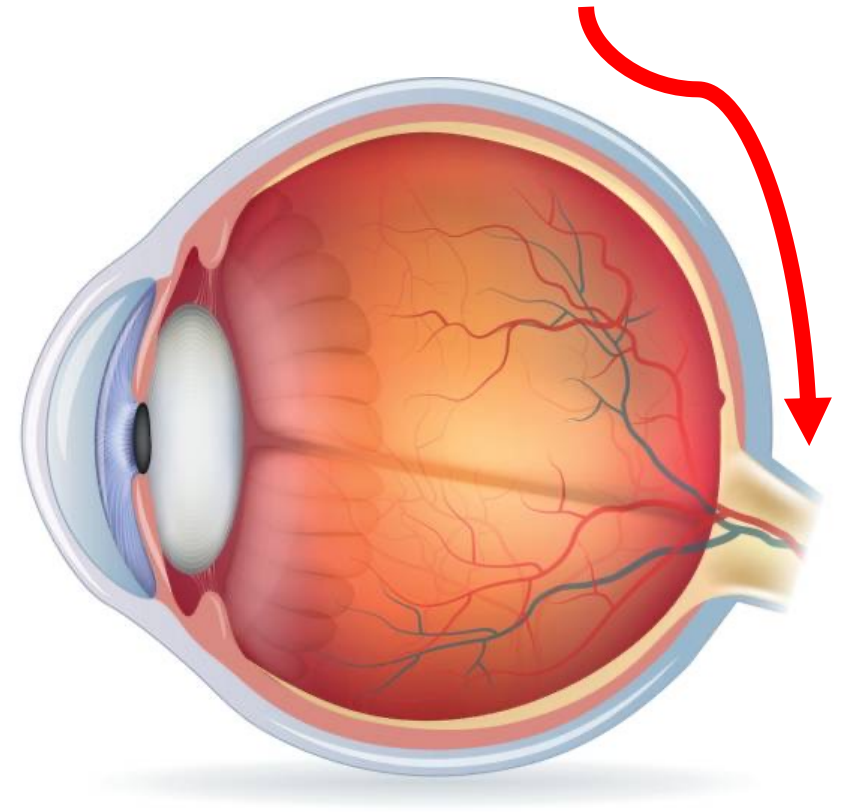
Risk of Nonarteritic Anterior Ischemic Optic Neuropathy in Patients Prescribed Semaglutide

Jimena Tatiana Hathaway, MD, MPH; Madhura P. Shah, BS; David B. Hathaway, MD; Seyedeh Maryam Zekavat, MD, PhD; Drenushe Krasniqi, BA; John W. Gittinger Jr, MD; Dean Cestari, MD; Robert Mallery, MD; Bardia Abbasi, MD; Marc Bouffard, MD; Bart K. Chwalisz, MD; Tais Estrela, MD; Joseph F. Rizzo III, MD

Hazard Ratio of NAION 4.28 (95% CI: 1.62 – 11.29, $P < .001$)

“The best approaches to confirm, refute, or re- fine our findings would be to conduct a much larger, retrospective, multicenter population-based cohort study; a prospective, randomized clinical study; or a postmarket analysis of all GLP-1 RA drugs.”

NAION = stroke of the optic nerve





Next CBER Best Seminar: July 17

Speaker: Yonas Ghebremichael-Weldeselassie, Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK

Topic: A modified self-controlled case series method for event-dependent exposures and high event-related mortality, with application to COVID-19 vaccine safety

Date/Time: Wednesday, July 17, 11 am ET

ohdsi.org/cber-best-seminar-series

Upcoming Seminars

— July 17, 2024 (11 am) - Yonas Ghebremichael-Weldeselassie, Warwick Medical School

Topic: A modified self-controlled case series method for event-dependent exposures and high event-related mortality, with application to COVID-19 vaccine safety

Presenter: Yonas Ghebremichael-Weldeselassie, Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK

[Watch This Seminar](#)

Abstract:

We propose a modified self-controlled case series (SCCS) method to handle both event-dependent exposures and high event-related mortality. This development is motivated by an epidemiological study undertaken in France to quantify potential risks of cardiovascular events associated with COVID-19 vaccines. Event-dependence of vaccinations, and high event-related mortality, are likely to arise in other SCCS studies of COVID-19 vaccine safety. Using this case study and simulations to broaden its scope, we explore these features and the biases they may generate, implement the modified SCCS model, illustrate some of the properties of this model, and develop a new test for presence of a dose effect. The model we propose has wider application, notably when the event of interest is death.

Bio: Yonas Weldeselassie is a Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK. He graduated in statistics and demography from University of Asmara, Eritrea and went on to become an assistant lecturer in Mekelle University, Ethiopia, and then a Senior Research Fellow in Medical Statistics at Warwick Medical School, division of Population Evidence and Technologies. He earned a Msc in Biostatistics from Hasselt University, Belgium and PhD in statistics from the Open University, UK. After working as a research associate, on MRC project 'Software tools and online resources for the self-controlled case series method and its extensions', at the department of mathematics and statistics, the Open University since 2014, he joined Warwick Medical School in June 2017. His main research interest is in medical statistics specially in the methodological development and application of the self-controlled case series (SCCS) method. He published a book on SCCS with Paddy Farrington and Heather Whitaker, and he is currently working on early prediction of gestational diabetes mellitus.





#OHDSI2024 Registration Is Open!

Registration is OPEN for the 2024 OHDSI Global Symposium, which will be held **Oct. 22-24** at the **Hyatt Regency Hotel in New Brunswick, N.J., USA.**

Tuesday: Tutorials

Wednesday: Plenary/Showcase

Thursday: Workgroup Activities

ohdsi.org/OHDSI2024





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



Melissa Haendel, PhD

Director of Precision Health & Translational Informatics and the Sarah Graham Kenan Distinguished Professor in the Department of Genetics at The University of North Carolina at Chapel Hill and co-founder of the Monarch Initiative and the National Covid Cohort Collaborative

‘Journeys across the translational divide: making healthcare and basic research data interoperable’

July 25, 2024, 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

Advancing Certification and Evaluation of Medical Device Software in the EU using OMOP

(Frédéric Jung, Chang Sun, Mahmoud Ibrahim, Gökhan Ertaylan)

REALM – innovative solutions for the creation and evaluation of AI medical device software.

Title: *Advancing Certification and Evaluation of Medical Device Software in the EU using OMOP.*



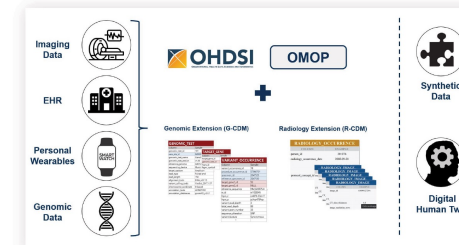
Background: In the landscape of healthcare, the increasing complexity and autonomy of **Medical Device Software (MDS)** present significant challenges in their certification and post-market monitoring, particularly in adapting to **real-world** healthcare settings. To address these challenges, the REALM project, also known as Real-world-data enabled assessment for health regulatory decision-making, aims to provide a robust testing infrastructure for the **evaluation** and certification of MDS in the European healthcare industry. By emphasizing transparency through the use of **OMOP** (Observational Medical Outcomes Partnership) databases and extensions, REALM seeks to offer stakeholders and regulatory bodies detailed and transparent insights into the performance of **AI** models embedded in MDS, crucial for ensuring trust and reliability in medical software solutions.

Figure1: REALM Partners across Europe.



REALM is a significant collaborative effort involving 15 partners across Europe (Figure 1). Within this consortium, five partners act as **demonstrators**, providing **AI** models for evaluation of the REALM capabilities to generate relevant testing dataset and evaluate how **AI** models performed. Given the **diversity** of data types required by the demonstrators (and future **AI** models) REALM heavily rely on the **OMOP CDM** and Extensions (**R-CDM** – Park et al., 2022 & **G-CDM** – Shin et al., 2019) to harmonize source data (Figure 2) from the **REALM data catalogue** and provide a unique way to query and generate test datasets. Following this approach, REALM also plans to incorporate data from the embedded **synthetic data generator** and **digital human twin** into its evaluation framework. By integrating these additional resources, REALM aims to further enhance the diversity and richness of its testing datasets, providing a more comprehensive **assessment** of **AI** model performance across various healthcare scenarios.

Figure2: REALM interoperability vision & harmonization plan.



Conclusion: REALM's integration of diverse data sources using OMOP CDM together with the REALM's rigorous framework provide **regulatory bodies** with a powerful **sandbox environment** for a transparent and precise assessment of medical AIs.



Frédéric Jung¹, Chang Sun², Mahmoud Ibrahim², Gökhan Ertaylan¹

1) VITO, Vlaamse Instelling voor Technologisch Onderzoek, Mol, Belgium

2) Institute of Data Science, Maastricht University, Maastricht, the Netherlands



#OHDSISocialShowcase

TUESDAY

SNOMED overhaul and its impact on ETL and phenotyping

(Masha Khitrun, Alexander Davydov, Oleg Zhuk)

SNOMED overhaul and its impact on ETL and phenotyping

Masha Khitrun¹, Alexander Davydov¹, Oleg Zhuk¹

¹Odysseus Data Services Inc., Cambridge, MA



Background: Over the years, the Vocabulary team has been working on integrating SNOMED CT into the ecosystem of OHDSI Standardized vocabularies. However, due to its comprehensive structure, multiple adjustments to SNOMED vocabulary ETL logic and interventions on the content level have been necessary, leading to the accumulation of bugs and discrepancies over the years. The SNOMED load_stage script that integrates the SNOMED into the OMOP vocabularies, has grown larger and more complex than anticipated, resulting in significant delays of OHDSI releases and a time lag between the OMOP version of SNOMED and SNOMED sources.

We present the results of a comprehensive overhaul of SNOMED in OHDSI vocabularies. This overhaul included both technical changes to the load_stage, aimed at simplifying future releases, and content changes designed to optimize cohort creation and ETL process.

Methods: The vocabulary development follows the guiding principles outlined in the Book of OHDSI¹, ensuring adherence to established standards and practices within the OHDSI framework. Both developer² and end-user³ documentation is maintained and made publicly accessible on GitHub, allowing for transparency and collaboration within the community. To assess the impact of vocabulary changes on ETL processes, we conducted a comprehensive analysis leveraging completed and ongoing ETL projects. This analysis provided insights into the challenges posed by vocabulary modifications.

Recognizing the significant impact of vocabulary changes on ETL⁴, we employ analytical methods tailored to mitigate these challenges. These methods⁵, accompanied by our internal quality control approach⁶ which includes the collection of vocabulary statistics, and a bunch of specific vocabulary checks⁷ enable us to address the implications of vocabulary updates on ETL workflows and phenotyping proactively.

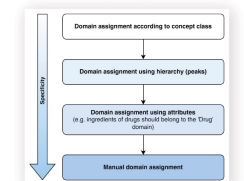
Table 1. Main changes implemented during the SNOMED overhaul and their estimated impact on ETL and Phenotyping

Changes	No. of concepts involved ^a	Impact on ETL	Impact on Phenotypes
Domain changes	24021	Low	High
De-standardization of concepts	3533	High	Low**
Split of pre-coordinated measurements and allergies	5357	Medium	High
Mapping of secondary neoplasms to Cancer Modifier	1042	High	High
Mapping following replacement links	9393	Low	Low
Retirement of the UK Drug Extension Module	464010	Possible	Low

^a Overall number of SNOMED concepts: 1 084 286

^{**} De-standardized concepts represent non-clinical events

Figure 1. Domain assignment algorithm in SNOMED



Results:

- **Domain assignment was improved** in its stability and consistency (Figure 1). As a result of this change, you may need to change the tables of interest (eg. querying **Condition_occurrence** instead of **Observation**) in the process of cohort creation. However, **semantic "grey zones"** still exist, where the domain assignment is a matter of debate due to the **ambiguity of concept interpretation**. Domain flows (Table 2) in these grey zones were discovered using analytical methods⁵, and domain improvement here is a **constant iterative process**.
- **Pre-coordinated SNOMED measurements** were moved from the **Condition** to the **Measurement** domain and splitted in a **post-coordinated way**. Thus, if you have previously used pre-coordinated SNOMED measurements, you should now start looking into the **Measurement** table where these concepts live as the **Measurement / Value** pairs.
- **Secondary neoplasms** were mapped to Standard concepts in **Cancer Modifier** vocabulary that belong to the **Measurement** domain. Thus, now you should look into the **Measurement** table to find the Secondary neoplasm concepts.
- In the course of the overhaul **110 SNOMED concepts** in the **Measurement** domain, mainly representing vital sign measurements, were mapped to the **Standard LOINC** concepts. These mappings are **erroneous** as SNOMED has a higher position in the hierarchy than LOINC, and they may affect the **hierarchy of measurements**. These issue is supposed to be solved in the course of the next release.
- We improved the creation of **"Maps to"** relationships following the SNOMED sources' **replacement links**. As a result of this change, more concepts are now mapped to Standard (Figure 3), and the **number of events in cohorts may increase**.
- Concepts that belong to **Attribute**, **Location** (except countries), **Social Context** (except concepts that carry the semantics of relatives, religion, occupation), **Physical Force**, and **Physical Object** (except concepts in the Device domain) concept classes have been de-Standardized in the course of the overhaul.
- We have performed the **retirement of the UK Drug Extension** module aimed to declutter the vocabulary, with concepts deprecated and linked to their equivalents in the dm-d vocabulary⁸.

Figure 2. Number of concepts with mapping compared with the upgraded concepts

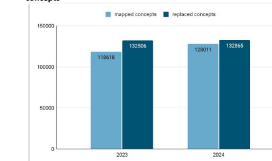


Table 2. Changes of domains for Standard SNOMED concepts over 2020-2024 years

2020	2021	2022	2023	2024	count
Condition	Condition	Condition	Condition	Observation	14216
Observation	Observation	Observation	Observation	Meas Value	2796
Observation	Observation	Measurement	Measurement	Measurement	1978
Observation	Observation	Observation	Measurement	Measurement	1589
Procedure	Procedure	Procedure	Procedure	Observation	1308
Condition	Condition	Condition	Condition	Measurement	847
Observation	Observation	Observation	Language	Language	834
Observation	Observation	Observation	Procedure	Procedure	614
Observation	Observation	Drug	Drug	Observation	500
Procedure	Procedure	Procedure	Measurement	Measurement	410
Condition	Condition	Observation	Observation	Observation	259

Conclusion:

The overhaul of the SNOMED vocabulary in OMOP has yielded significant improvements in ontology structure, cohort creation, and mapping efficiency. These enhancements contribute to more accurate data analysis, better research outcomes, and increased interoperability within the healthcare ecosystem. However, it is essential to consider their impact on ETL processes and phenotyping algorithms. The adjustments made to domain assignments and concept mappings may require updates to existing ETL workflows, and researchers should carefully review their phenotyping algorithms to ensure compatibility with the updated vocabulary structure and content.

References:

1. Observational Health Data Sciences and Informatics. The Book of OHDSI.
2. <https://github.com/OHDSI/Vocabulary-v5.0/tree/master/SNOMED>
3. <https://github.com/OHDSI/Vocabulary-v5.0/wiki/Vocab-SNOMED>
4. <https://forums.ohdsi.org/t/cpt-hierarchy-errors-lost-children-in-2023-and-changed-domains/18383>
5. Dmitry Dynmhyts, Frank DeFalco, Anthony Molinaro, Clair Blacketer. An Evaluation and maintenance of cohorts and concept sets in the OMOP Vocabulary Evolution. July 2023, Conference: OHDSI European Symposium 2023.
6. https://github.com/OHDSI/Vocabulary-v5.0/tree/master/working/packages/QA_TESTS
7. https://github.com/OHDSI/Vocabulary-v5.0/blob/master/working/manual_checks_after_generic_update.sql
8. https://github.com/OHDSI/Vocabulary-v5.0/releases/tag/v20240229_170921714.000000



#OHDSISocialShowcase

WEDNESDAY

OMOPification of real world cancer data to enable privacy-preserving analytics for cancer research

(Prabash Galgane Banduge, Anne-Lore Bynens, Cedric Gillissen, Andre Dekker, Petros Kalendralis, Pascal Suppers, Alberto Traverso, Lizza Hendriks, Aiara Lobo Gomes)

Challenges in harmonising data across multiple biobanks

Karyn Mégy¹, Rebecca Akhanemhe¹, Ben Hollis¹, Ali Abbasi¹, Amanda O'Neill¹, Shikta Das¹, Stewart MacArthur¹, Sean O'Dell¹, Sebastian Wasilewski¹, Quanli Wang², Slavé Petrovski¹, Jen Harrow¹.

¹. Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK. ². Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Waltham, MA, USA

1. Background

Early-stage incorporation of human genomic data into the assessment of drug targets has been shown to significantly increase drug pipeline success rates. Large biobanks such as UK Biobank, combining genetic and clinical data on 0.5 million individuals, offer an unprecedented opportunity to evaluate effects of genetic variants on a broad collection of traits. Statistical power, however, comes from the size as well as the ethnic diversity of those biobanks. AstraZeneca's Centre for Genomics Research is establishing one of the world most comprehensive and diverse genetic resource, combining genetic and phenotypic data for multiple biobanks. This work describes the challenges faced when harmonizing such datasets, enabling their cross analysis.

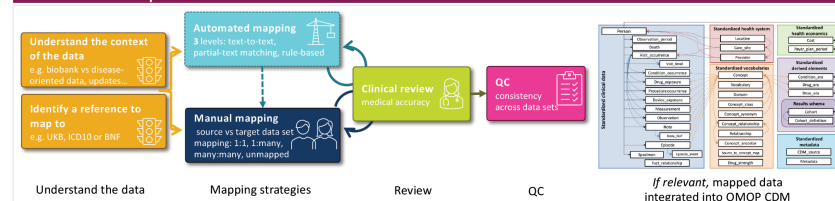
Different data sets, several data types, multiple standards

UK Biobank (UKB) is one of the golden standard, in terms of data diversity, sources but also coding systems. However, it is very reflective of the UK population and health care system. Biobanks from different countries will be using different coding system, different units (e.g. *HbA1c*: mmol/mol vs. %), medication names (e.g. *metformin* vs. *metformina*), and the local language, making comparison of those data sets challenging.

Source of health data available in our cohorts						
	Hospital data	Primary care	Cancer data	Questions	Free text	And also...
UK Biobank	WHO ICD9 & 10	Read2 & 3	yes	formatted	-	Lab. proc.
US cohort #1	CM ICD9 & 10	CM ICD9 & 10	CM ICD9 & 10	-	-	Lab. proc.
UK cohort #1	WHO ICD9 & 10	Read2 & 3	yes	-	-	Lab. proc.
FinnGen	WHO ICD8, 9 & 10					
MCPS	-	-	WHO ICD10	WHO ICD10	yes	Lab.

ICD: International Classification of Diseases, versions 8, 9 and 10. In the WHO or ICDM system.
Read2: International Classification of Diseases Clinical Terms; Lab & proc: Laboratory results and procedures

2. Harmonisation process



3. Results

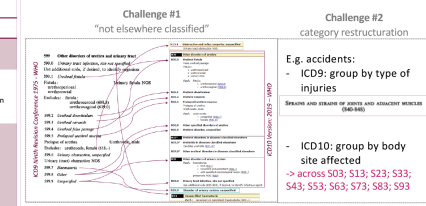
E.g.1: mapping a small disease-oriented data set to a large biobank

We mapped one of our data sets, a small disease-oriented resource, to UK Biobank. An initial mapping was done manually, following the process described above, high-quality but timely. As a test, we then performed NLP (Natural Language Processing) on that same data set, reducing the mapping time from months to a week, however <50% of terms could be mapped. => NLP followed by manual mapping would be most efficient strategy in term of time and accuracy. The final mapping will be transformed into the OMOP common data model

Approach	Description	Example
Text-to-text match	Exact Text match	"Age" = "Age at recruitment"
Partial text match	Step 1: Sub-word matching, part of substring matches Field ID Step 2: Sub-word matching and matching all words present in column	"Supplement" = "Supplements" "weight loss" = "loss in weight"
Rule-based match	Identify key terms and match to UKB Field ID Layer 1: Age or family history-based fields can be identified using "age", "th", "brother/mother" terms Layer 2: Diseases and symptoms can be identified using ontologies and classified as cancer vs non-cancer Layer 3: Category and Question information to identify terms like self-reported, subsection etc.	20002 (self-reported non-cancer)

E.g.2: mapping across ontologies, from UKB ICD9 to ICD10

UK Biobank diagnoses are encoded both in the ICD9 and in ICD10 WHO classifications. Following our harmonisation process, we have mapped the ICD9 terms present in the UKB data set to ICD10 and, according to the FAIR principles, are returning the results to UK Biobank so that they can be shared with the community (manuscript in preparation). => In total, we mapped 751 ICD9 codes to 573 unique ICD10 codes, with 85% having a 1:1 mapping.



4. Take home messages

- Harmonising data sets require to understand the data and adapt the strategy when needed.
- NLP followed by manual mapping is the most efficient strategy for mapping across cohorts
- Importance of return of data for use by the community (FAIR principle)
- Gap in OMOP: mapping of images & medications

Acknowledgements - We would like to thank the participants and investigators in the UK Biobank study who made this work possible. We also acknowledge contribution from members of the AstraZeneca Genomics Initiative and the AZ IT Knowledge Engineering Team.





#OHDSISocialShowcase

THURSDAY

The association between comorbid depression and insulin initiation in type 2 diabetes A cohort OHDSI study

(**Christianus Heru Setiawan**, Daniel C.A. Nugroho , Phan Thanh Phuc , Septi Melisa, Muhammad Solihuddin Muhtar, Nguyen Phung Anh, Jason C. Hsu)

The study found a significant correlation between depression and a higher probability of initiating insulin treatment, with an initial hazard ratio of 1.38.

The association between comorbid depression and insulin initiation in type 2 diabetes: A cohort OHDSI study

Background: Individuals diagnosed with type 2 diabetes have a higher risk of experiencing depression compared to those without the condition. Hyperglycemia-induced neurochemical dysregulation promotes the progression of type 2 diabetes. Furthermore, depression can lead to poor outcomes and may cause insulin resistance. This comorbidity may fail diabetes oral medications, and insulin therapy may be required.

Result: We analyzed data from 35,589 patients, and after PS matching (1:4), we obtained 1903 patients for the target group and 5857 patients for the comparator group. We examined the association between depression comorbid with the outcome of insulin initiation. Depression was found to be significantly associated with insulin initiation, with a hazard ratio of 1.38 (95% CI: 1.11, 1.71).

Figure 1. The Kaplan Meier plot shows the estimated incidence of first initiation of insulin use. Patients with type 2 diabetes and depression are on the target curve (blue line), while those with type 2 diabetes are on the comparator curve (red line).

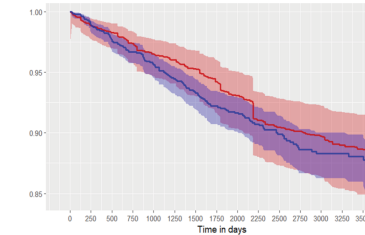
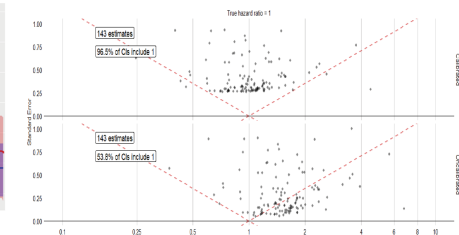
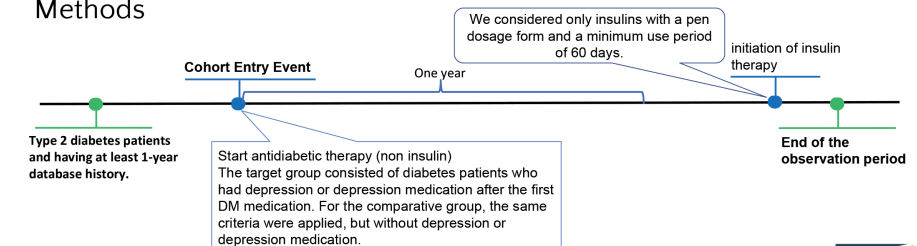


Figure 2. Empirical calibration plots where estimates below the dashed line are statistically significant ($\alpha = 0.05$) different from the true effect size. There are two images in the picture, the below showing uncalibrated estimates and the above depicting calibrated estimates.



T2DM and depression patients 1,903 1,795 1,661 1,531 1,402 1,258 1,141 1,042 934 829 704 578 476 391 298
T2DM patients 5,857 5,288 4,844 4,389 3,983 3,445 3,055 2,743 2,431 2,135 1,836 1,491 1,183 925 719

Methods



Limitation: Furthermore, after refining the findings using negative controls, the effect size estimates were recalibrated, revealing no significant difference in the hazard of insulin initiation after calibration. This recalibrated outcome indicates a hazard ratio of 0.92, suggesting that the observed connection between depression and insulin initiation may be more complex than initially thought, possibly influenced by unmeasured confounding factors.



Christianus Heru Setiawan, Daniel C.A. Nugroho, Phan Thanh-Phuc, Septi Melisa, Muhammad Solihuddin Muhtar, Nguyen Phung-Anh, Jason C. Hsu





#OHDSISocialShowcase

FRIDAY

Empowering research with seamless data flow and research-ready, anonymised data in OMOP CDM: Learnings from the design of WAYFIND-R, a global precision oncology registry and research platform

(**Tom Stone**, Yuri Pyatkin, Ana Ferro, Dimitar Toshev)

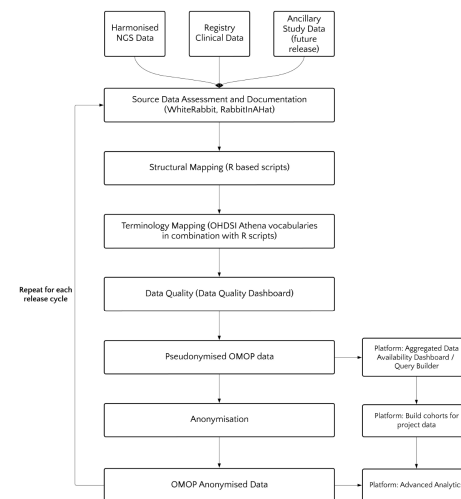


Empowering research requires **seamless** delivery of **high quality data**. The WAYFIND-R® platform enables automation and **accelerates insights generation** from primary data collection to research-ready data.

Title: Empowering research with seamless data flow and research-ready, anonymised data in OMOP CDM: Learnings from the design of WAYFIND-R, a global precision oncology registry and research platform

Background: WAYFIND-R is a global precision oncology registry (NCT04529122) and has the aim to advance science and provide the scientific community worldwide with access to real-world data, enabling epidemiological and clinical research, and collaborations across research groups. The WAYFIND-R® Data Sharing and Collaboration Platform enables researchers to access anonymised clinico-genomic data from the registry transformed to the OMOP CDM within a secure research environment.

Methods



Outcomes

- Researchers able to access quality OMOP anonymised aggregate and individual-level data from registry
- Automation of concept mapping simplifies review process on an ongoing basis
- Defined data quality processes allow for routine checking of data and actions for each data release cycle. For example queries raised within clinical database
- ETL automation ensures consistency and reproducible output
- Changes to the structure registry clinical database require ongoing assessments to determine impact to data and platform
- Implementation of cohort definition within platform

Opportunities for Collaboration

- Gene and biomarker ontologies
- Oncology extension
- Registry studies - our learnings

Acknowledgements:

We thank the patients and their families who take part in WAYFIND-R, as well as the staff, research coordinators, and investigators at each participating institution.



Tom Stone¹, Yuri Pyatkin², Ana Ferro¹, Dimitar Toshev²

¹ Roche Products Limited, Welwyn Garden City, UK; ² F. Hoffmann-La Roche Ltd, Basel, Switzerland





Opening: Health Data Scientist, Erasmus MC

[Jobs](#)[About Erasmus MC](#) ▾[Research facilities](#) ▾[Career](#) ▾[NL](#) / [EN](#)

Health Data Scientist

Would you like to be responsible for the creation and execution of study software to support our researchers? Join our team!

🕒 Closing date 18-07-2024 # 27.16.24.TG

🕒 36 hours

🎓 University of Applied Sciences

🛠️ IT

Apply ▶

Job description

The [Health Data Science group](#) at the department of Medical Informatics is looking for an experienced R programmer to develop open-source analytics.

You will be responsible for the creation and execution of study software to support the work of epidemiologists. This will include designing, developing, documenting, and maintaining R code that will be executed against health data that is standardized to the OMOP Common Data Model (OMOP-CDM). This data model is maintained by the Observational Health Data Sciences and Informatics (OHDSI, www.ohdsi.org) initiative. The Department is leading the Data Analysis and Real World Interrogation Network coordination centre ([DARWIN EU®](#)) in which a large data network is created with data sources that are mapped to the OMOP-CDM. The aim of this network is to provide fast and reliable evidence to the European Medicines Agency. Your work will enable the use of health data at an unprecedented scale in Europe and will facilitate the execution of impactful studies to improve patient care. We recommend applicants to review the [DARWIN EU ® software packages](#) to get an understanding of the growing set of R packages.



Opening: Sr AD, Real World Evidence & Analytics Boehringer Ingelheim

SR AD, Real World Evidence & Analytics

[Apply Now](#)

[< Back](#)

JOB ID - 13278

Description

The purpose of this job is to:

- Generate real world evidence (RWE) to support in-line and pipeline products.
- Provide statistical advice on the analysis of real world data (RWD) to various internal and external stakeholders.
- Contribute to the RWD acquisition strategy and tool evaluation.



Opening: Lead Director, RWE Distributed Research CVS Health

Lead Director, RWE Distributed Research

Apply



Hybrid



PA - Blue Bell

IL - Northbrook

CT - Hartford

RI - Woonsocket

AZ - Scottsdale

[View All 7 Locations](#) ▾



Full time



Posted 6 Days Ago



R0268183

Bring your heart to CVS Health. Every one of us at CVS Health shares a single, clear purpose: Bringing our heart to every moment of your health. This purpose guides our commitment to deliver enhanced human-centric health care for a rapidly changing world. Anchored in our brand — with heart at its center — our purpose sends a personal message that how we deliver our services is just as important as what we deliver.

Our Heart At Work Behaviors™ support this purpose. We want everyone who works at CVS Health to feel empowered by the role they play in transforming our culture and accelerating our ability to innovate and deliver solutions to make health care more personal, convenient and affordable.

About us



Our Work Experience is the combination of everything that's unique about us: our culture, our core values, our company meetings, our commitment to sustainability, our recognition programs, but most importantly, it's our people. Our

[Read More](#) ▾



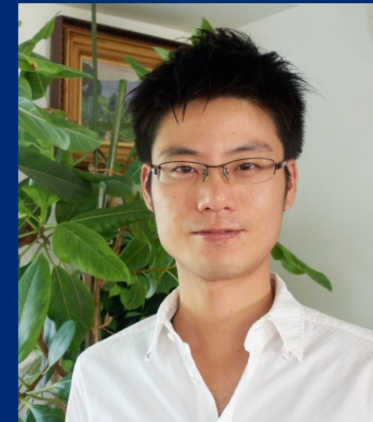
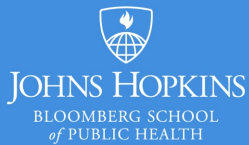
Openings: Postdoctoral Fellow, Johns Hopkins Univ.

PHARMACOEPIDEMIOLOGY POST-DOCTORAL TRAINING PROGRAM

Co-Directors: Caleb Alexander, MD, MS and Jodi Segal, MD, MPH

The **Pharmacoepidemiology Training Program** at the Johns Hopkins Bloomberg School of Public Health (BSPH) is currently **seeking to support postdoctoral fellows**. All supported trainees work with core faculty on existing or newly developed research projects on pharmacoepidemiology, so as to optimize the safe and effective use of medicines to treat heart, lung and blood diseases in the United States. |

Deadline for applications: rolling





Where Are We Going?

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





Three Stages of The Journey

Where Have We Been?

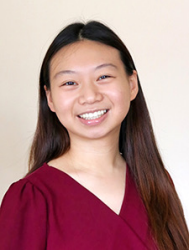
Where Are We Now?

Where Are We Going?





July 16: HowOften Initiative & Early Results



Hsin Yi "Cindy" Chen

MD-PhD Student

Columbia University Department of Biomedical Informatics



Azza Shoaibi

Director, Observational Health Data Analytics

Janssen Research and Development



Elise Ruan

Clinical Informatics Fellow

NewYork-Presbyterian Hospital/Columbia University



George Hripcsak

Professor of Biomedical Informatics

Columbia University



**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**