Updates in Methods Research

OHDSI Community Call
Oct. 8, 2024 • 11 am ET

October Community Calls

Date	Topic
Oct. 8	Recent Advances in Methodological Research
Oct. 15	Global Symposium Mad Minutes/Final Logistics
Oct. 22	No Meeting due to Global Symposium
Oct. 29	Welcome to OHDSI

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	Generative AI and Analytics
Tuesday	12 pm	CDM Vocabulary Subgroup
Tuesday	3 pm	OMOP CDM Oncology Outreach and Research Subgroup
Wednesday	9 am	Patient-Level Prediction
Wednesday	12 pm	Health Equity
Wednesday	2 pm	Natural Language Processing
Thursday	9:30 am	Network Data Quality
Thursday	10:30 am	Evidence Network
Thursday	12 pm	Medical Devices
Thursday	12 pm	Strategus
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Steering
Friday	11:30 am	Clinical Trials
Friday	11 pm	China Chapter
Monday	10 am	Africa Chapter
Monday	10 am	CDM Survey Workgroup
Monday	11 am	Early-Stage Researchers

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. 7-24: Collaborator Showcase Submission Review

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 Global Symposium

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

Registration is OPEN for the 2024 OHDSI Global Symposium. Collaborator Showcase notifications are taking place this week. Agendas and tutorial/workgroup schedules are posted.

Tuesday: Tutorials
Wednesday: Plenary/Showcase
Thursday: Workgroup Activities

ohdsi.org/OHDSI2024

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

Vipina Keloth, PhD

Associate Research Scientist in Biomedical Informatics and Data Science at Yale University School of Medicine

‘Exploring the realm of large language models for information retrieval in the biomedical domain’

October 31, 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

Improving completeness in mapping population based cancer registry data to OMOP CDM by using alternative source fields

(Tapio Niemi, Eloïse Martin, Vincent Faivre, Patrick Arveux, Simon Germann, Valérie Pittet, Jean Luc Bulliard)

Mapping cancer registry data to OMOP CDM: Using alternative source fields reduced unmapped cases by 95%

Improving completeness in mapping population-based cancer registry data to OMOP CDM by using alternative source fields

Background: OMOP CDM mappings do not exist for all source codes used in cancer registries. Ignoring unmapped cases reduce completeness and may bias surveillance indicators based on these mappings.
The work was done within the European Health Data and Evidence Network (EHDEN) project, and was based on the Vaud Cancer Registry's database, 1974-2018 (Switzerland).

Results

- The OHDSI data quality tools indicated 100% correctness. Comparison against a manual extraction showed over 99% correspondence.
- Additional use of ICD-10 reduced unmapped cases by 95% compared to ICD-O alone (Figure 1).
- General OMOP codes did not increase unmapped TNM stage codes (breast cancer cases, Figure 2).
- The ICD-8 death cause codes were challenging and only 73% of them were mapped (vs 96% ICD-9 and ICD-10).

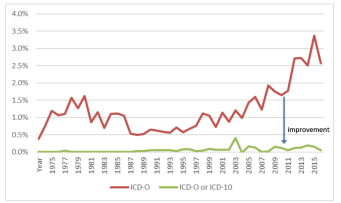


Figure 1. Annual differences (in %) between ICD-O and ICD-O and ICD-10 methods compared to the manual extraction of the source data

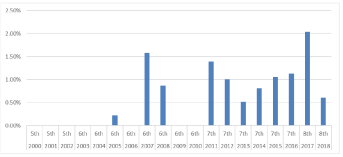


Figure 2. Percentage of unmapped TNM stage values for breast cancer cases diagnosed between 2000 and 2018 with used TNM versions

Methods

We improved cancer registry mapping to CDM focusing on completeness by:

1) Using ICD-10 codes when the ICD-O code was not mapped.
2) Using general OMOP TNM staging codes for unmapped TNM versions.
3) Mapping old ICD-8 death cause codes to ICD-9 codes first.

Our validation process included two steps:

1) Using OHDSI data quality tools, and
2) Comparing CDM data to a manually extracted source data.

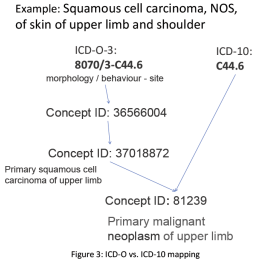


Figure 3: ICD-O vs. ICD-10 mapping

Discussion and Conclusions

- Common cancer types and recent versions of coding systems are straightforward to map to OMOP CDM, but rare types or older versions can need more work.
- A challenge is that ICD-O or ICD-10 codes can map to different OMOP concepts. This may require applying concept hierarchies in OMOP queries (Figure 3).

www.ehden.eu

Acknowledgements: EHDEN cancer registry teams Geneva, Luxemburg, Netherlands, and Norway

Contact:
tapio.niemi@unisante.ch

Tapio Niemi, Eloïse Martin, Vincent Faivre, Patrick Arveux, Simon Germann, Valérie Pittet, Jean-Luc Bulliard

unisanté
Centre universitaire de médecine générale et santé publique • Lausanne

#OHDSISocialShowcase

TUESDAY

Exploring Embedding Representations for Structured Data in the OMOP CDM

(Jorge Cerejo, Bernardo Neves, Simão Gonçalves, José Maria Moreira, Nuno A. da Silva, Francisca Leite)

Exploring Embedding Representations for Structured Data in the OMOP CDM

Use-Case for Long Hospitalization Prediction

Jorge Cerejo[1], Bernardo Neves[1,2], Simão Gonçalves[1], José Maria Moreira[1], Nuno A. da Silva[1], Francisca Leite[1]

[1]Hospital da Luz Learning Health, Luz Saúde, Lisboa
[2]Instituto Superior Técnico, Universidade de Lisboa

Introduction

Word embedding techniques have gained influence in the field of medical data analysis, as these techniques enable the representation of medical concepts [1,2].

This study explores the utilization of word embeddings to represent medical codes extracted from a Portuguese tertiary hospital's OMOP CDM database. By representing medical codes into a continuous vector space, we aim to capture inherent relationships between healthcare concepts, potentially enhancing predictions of long hospitalizations (>7days). Predicting the length of stay early can help in better resource allocation and patient care.

Methods

The real-world dataset used encompasses approximately 52M events across 6.5M visits and 500K patients, with an observation period spanning between 2007 and 2023.

We employed the word2vec algorithm to generate medical code embeddings, exploring diverse embedding spaces by varying parameters such as vector and window sizes [3]. Additionally, we expanded our vocabulary using ontology expansion [4].

We use the embeddings as features of machine learning (ML) models for predicting long hospitalization, comparing their performance with count-based features.

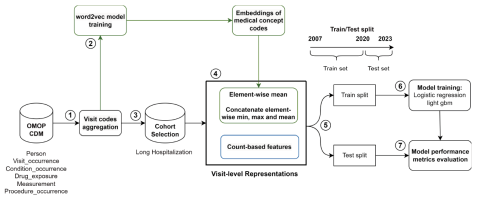
Results

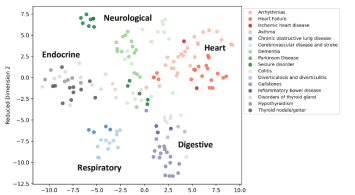
- Ontology expansion significantly increases the event counts of all patients; however, it does not improve predictive performance.
- Smaller embedding dimensions yield suboptimal performance, while dimensions exceeding 100 do not lead to significant improvements.
- Increasing window sizes fails to enhance model performance.
- Concatenation of element-wise min, max and mean representations was show to be the best approach for visit embedding representations.

Conclusions

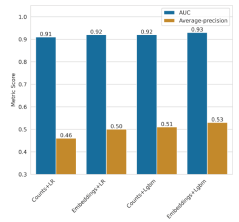
- Our study demonstrates the feasibility of creating embedding representations from structured data in the OMOP CDM.
- Ontology expansion resulted in inferior performance, introducing noise rather than enhancing contextual understanding.
- Explore advanced models with attention mechanisms for ontology expansion [5].
- Incorporate medical code descriptions in embedding creation to improve contextual understanding and code relationships [6].

Embedding representations demonstrate slightly **superior performance** over **count-based features**, while **capturing** data **relationships** and **dependencies.**





Ontology Expansion?	No					No				
Visit representation	Element-wise mean					Concatenation of Element-wise min, max and mean				
Window_size	5	10	20	30	50	5	10	20	30	50
Vector_size 10	0,911	0,912	0,909	0,913	0,911	0,918	0,921	0,918	0,922	0,918
Vector_size 50	0,918	0,917	0,918	0,918	0,917	0,925	0,925	0,926	0,924	0,925
Vector_size 100	0,921	0,919	0,917	0,919	0,917	0,926	0,926	0,924	0,926	0,924
Vector_size 300	0,920	0,921	0,920	0,921	0,920	0,926	0,926	0,925	0,926	0,925
Vector_size 500	0,921	0,921	0,920	0,921	0,920	0,926	0,926	0,925	0,925	0,926
Vector_size 1000	0,922	0,921	0,920	0,921	0,920	0,925	0,926	0,926	0,926	0,925
Ontology Expansion?	Yes					Yes				
Visit representation	Element-wise mean					Concatenation of Element-wise min, max and mean				
Window_size	5	10	20	30	50	5	10	20	30	50
Vector_size 10	0,871	0,873	0,875	0,878	0,876	0,896	0,897	0,904	0,907	0,907
Vector_size 50	0,901	0,900	0,902	0,903	0,902	0,911	0,910	0,912	0,915	0,916
Vector_size 100	0,905	0,906	0,906	0,908	0,907	0,911	0,914	0,918	0,918	0,918
Vector_size 300	0,911	0,910	0,911	0,911	0,911	0,921	0,920	0,912	0,919	0,920
Vector_size 500	0,912	0,913	0,913	0,912	0,912	0,923	0,922	0,922	0,922	0,920
Vector_size 1000	0,914	0,913	0,913	0,913	0,913	0,922	0,923	0,922	0,923	0,921



References

[1] Bajor et al. Embedding Complexity In the Data Representation Instead of In the Model: A Case Study Using Heterogeneous Medical Data. 2018.
[2] Beaney et al. Comparing natural language processing representations of disease sequences for prediction in the electronic healthcare record. Health Informatics; 2023
[3] Mikolov et al. Efficient Estimation of Word Representations in Vector Space. 2013.
[4] Choi et al. GRAM: Graph-based Attention Model for Healthcare Representation Learning. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax NS Canada: ACM; 2017
[5] Hur et al. Unifying Heterogeneous Electronic Health Records Systems via Text-Based Code Embedding. 2021.
[6] Rasmy et al. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. Npj Digit Med. 2021

HOSPITAL DA LUZ LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

OHDSI

#OHDSISocialShowcase

WEDNESDAY

Harmonization to the OMOIP CDM of cancer patient data at the Modena Oncology Center

(Enrico Calanchi, Luca Moscetti, Mirko Orsini, Laura Delsante, Enrica Martinelli, Andrea Spallanzani, Federica Bertolini, Elisa Pelorelli, Massimo Dominici)

Harmonization to the OMOIP CDM of cancer patient data at the Modena Oncology Center

Enrico Calanchi[1], Luca Moscetti[2], Mirko Orsini[1], Laura Delsante[1], Enrica Martinelli[3], Andrea Spallanzani[2], Federica Bertolini[2], Elisa Pelorelli[3],Massimo Dominici[3]
[1] DataRiver Srl, [2] Oncology and Hemathology Department, Azienda Ospedaliero-Universitaria di Modena – Policlinico, Modena , Italy, [3] Division of Medical Oncology, Department of Medical and Surgical Sciences for Children and Adults, University Hospital of Modena, Modena, Italy

Background

The evolution of treatments and in particular the development of target therapies and the advent of immunotherapy has revolutionized the therapeutic strategy in the treatment of early and metastatic cancers in the last 10 years. The availability of new treatments has outlined a scenario in which the clinical information relating to the individual patient, is insufficient to define the best treatment but requires constant integration of anatomo-pathological and molecular parameters, obtained through advanced technologies such as gene sequencing. The effectiveness of new drugs, both as monotherapy and in association with chemotherapy or hormone therapy, was correlated with various predictive factors deriving from immunohistochemical or molecular biology analyses. The constant changes in this treatment scenario, determined by registration clinical trials with careful selection of the patients included, determine the need to collect real-practice data to evaluate the feasibility of the aforementioned therapies also in daily outpatient clinical practice. Due to this urgency to provide evidence-based information, the EHDEN (European Health Data & Evidence Network) consortium has launched an extraordinary appeal to invite institutions in Europe to systemically share with all members of the network the data relating to patients suffering from oncological pathologies and treated according to clinical practice. The final aim tends to harmonize the data collected in order to create a common database that allows the rapid extrapolation of high quality evidence. The analysis of the database may allow the identification of any parameters to be able to propose Real Word Evidence studies in the future aimed at creating a platform with data, standardized and anonymized according to the computer language used by the platform EHDEN. This evidence will help the scientific community to better understand the impact of treatments on the population, bringing improvements to care in the oncology field also through the conduct of real world studies.

Methods

In this retrospective cohort study, all patients treated with solid tumors starting from the year 2001 at the COM are included. The objective of the study is the creation of a database through the collection of demographic and clinical parameters in order to evaluate the actual outcome of patients treated in normal clinical practice and who do not identify with the subjects enrolled in the clinical studies. This is a single-center, retrospective observational study. For the patients included in the study, clinical data, therapies and procedures performed has been collected retrospectively. In particular, the data of interest are: demographic information (gender, age, place of birth, ethnicity), dlinical data (only and exclusively the dates of the start and end of therapies and the dates and results of instrumental re-evaluations will be collected). Histopathological and molecular diagnosis, therapies performed: systemic therapies and locoregional treatments. All patients diagnosed with solid tumors, both prevalent and incident, in the period starting from the year 2001 were enrolled. Eligibility criteria were as follow Inclusion criteria, Age > 18 years, patients diagnosed with solid neoplasia since 2001 under the care of the Modena Oncology Center both in terms of new cases and in terms of patients already under treatment. All disease stages relating to solid tumors that have been the responsibility of oncology for surgery alone, for subsequent therapies or for follow-up. Different databases sources will be used: company, radiological examination, laboratory examination. The systems are integrated and individual patients has been identified by a unique ID code. The IT and Telematics Technologies Service executed patient extraction queries from the various company applications and was responsible for their anonymization. The data has been standardized by the DataRiver using an application provided by the EHDEN consortium, to be made accessible in aggregate form to the members of the network itself. The protocol has been approved by the local ethic committee in January 2024. The analysis, mapping and ETL of the COMNet data has been performed on docker platform hosted on a shared virtual server accessible via VPN to ensure maximum data confidentiality. The environment was based on the Linux operating system Ubuntu 22.04 to hosts the installation of the tools provided by EHDEN. The ETL has been implemented on a PostgreSQL instance to upload the data extracted from the SQL Server database of the information systems in csv format via Python script. Python script has been used also to create a support table for the additional transformations while the most of ETL consists in SQL scripts.

Results

The results highlight the increase in clinical data entry and digitalisation over the last 2 decades. An increase in the use of cancer drugs is observed after 2010, due to the availability of new therapies, especially after the availability of targeted therapies and the development of precision medicine. In addition, a parallel increase in clinical instrumental controls is observed, visible in the increase of measurement and observation information. The lack of data on visits prior to 2010, highlights the change in the method of clinical data collection due to the evolution of the medical record and the progressive digitalisation of the paper data collected in the decades-long history of the oncology center in Modena. (Fig.1 and 2)

Figure 1

Figure 2

Conclusion

The harmonization process highlighted numerous inconsistencies related to missing dates on information entry or diagnosis of conditions related to patients clinical data stored in the COMNet platform. The analysis of the harmonised data revealed the criticality of data fragmentation due to the incompleteness of the digitalisation process and the presence of different and unconnected software platforms. The objective of achieving greater interoperability and having an integrated view of the information will be one of the guidelines of the re-engineering process of the COMNet platform

OHDSI

Contact
Luca Moscetti
Department of Oncology-Hematology,
University Hospital of Modena,
Gruppo Oncologico Italiano Ricerca Clinica
Email: moscetti.luca@aou.mo.it
Website: www.aou.mo.it, www.goirc.org
Phone:+39 059 4223244

#OHDSISocialShowcase

THURSDAY

Developing a Clinical Decision Support System Software Prototype that Assists in the Management of Patients with Self-Harm based on the Use of National Registries Harmonized in OMOP CDM: the PERMANENS Project

(Angela Leis, Philippe Mortier, Franco Amigo, Madhav Bhargav, Susana Conde, Montserrat Ferrer, Oskar Flygare, Busenur Kizilaslan, Laura Latorre, Miguel Angel Mayer, Víctor Pérez Sola, Ana Portillo van Diest, Juan Manuel Ramírez-Anguita, Ferran Sanz, Gemma Vilagut, Jordi Alonso, Lars Mehlum, Ella Arensman, Johan Bjureberg, Manuel Pastor, Ping Qin, Sofia Bettella, Gabriel de Maeztu)

Developing a Clinical Decision Support System Software Prototype that Assists in the Management of Patients with Self-Harm based on the Use of National Registries Harmonized in OMOP CDM: the PERMANENS Project

European OHDSI Symposium, Rotterdam, 1-3 June 2024

Background

The damaging impact of self-harm and suicide is profound in society, with over 700,000 global suicides annually, signifying the loss of millions of life years and affecting numerous lives. Europe struggles with notable suicide rates but faces challenges in surveillance due to inconsistent definitions and methods. Emergency departments (EDs) serve as essential points of intervention for those who self-harm, yet there is variability in the quality of care provided, often influenced by stigma and a lack of specialist training. This inconsistency can affect treatment uptake and patient engagement.

In this study, we describe the process, usefulness, and challenges of harmonizing national registries focused on suicide attempts using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) applied in the Towards Personalized Clinical Management of Suicide Risk through Data Driven Clinical Decision Support using Translational Electronic Registry Data (PERMANENS) project.

Methods

The PERMANENS project employs a multidisciplinary approach, combining clinical mental health research, public health, biostatistics, and informatics to construct a Clinical Decision Support System (CDSS). This system is trained on diverse European registry data harmonized using the OMOP Common Data Model (CDM) and is designed to offer personalized risk assessments and management strategies for emergency department (ED) patients presenting with self-harm. The OMOP CDM, developed by the Observational Health Data Sciences and Informatics (OHDSI), is used to transform and harmonize all registry data from the four sites into a common data format using standard terminologies, vocabularies, and coding schemes.

The core of the CDSS is machine learning-based prediction models, designed to be clinically interpretable. These models predict key adverse outcomes, such as repeat self-harm or suicide, leveraging large datasets to understand complex interactions. Additionally, a systematic review creates a clinical knowledge base for effective interventions.

Table 1. The sources include the National Self-Harm Registry and the Hospital In-Patient Enquiry (Ireland); the Norwegian Patient Registry, the Norwegian Cause-of-Death Register, Statistics Norway's Events Database, and the Norwegian Central Population Registry (Norway); the Swedish National Patient Registry, the Swedish Cause-of-Death Register, the Swedish Total Population Register, and electronic healthcare records retrieved from Stockholm County Council (Sweden); and the Data Analytics Program for Health Research and Innovation (Catalonia, Spain).

	Ireland	Norway	Sweden		Catalonia (Spain)
Target population	All patients with hospital-presenting self-harm in the Republic of Ireland	All Norwegian residents	All psychiatric emergencies in Stockholm aged ≥ 18	All Swedish residents aged 18-39	All Catalan residents
N available for analysis	± 163,000 patients	± 5.4 mil	± 75,000 patients	± 5 mil	811,394 cases and controls
Time span	2002-2020	2008-2019	2016-2022	1993-2016	2014-2019

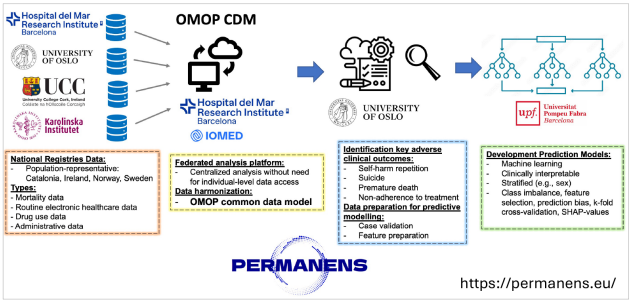


Figure 1. Pipeline utilizing OMOP CDM and Machine Learning technologies in the PERMANENS project, involving data harmonization, federated analysis, and development of predictive models for key clinical outcomes.

Conclusions

The PERMANENS project encapsulates the potential of personalized medicine in mental health care, particularly for those at risk of self-harm and suicide. The project stands out for its inclusive machine learning models, implementation research with stakeholders, and focus on personalized care using the OMOP Common Data Model. It emphasizes the indispensability of involving patients and healthcare professionals in the developmental process to ensure the system is aligned with real-world clinical needs. This initiative holds promise for revolutionizing suicide risk assessment and intervention in EDs, setting a new standard for personalized, data-informed healthcare solutions.

References

1. Global Burden of Disease Collaborative Network. Global burden of disease study 2019 (GBD 2019) results. *Seattle U S*. Published Online First: 2019.
2. Ribeiro JD, Franklin JC, Fox KR, *et al*. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychol Med*. 2016;46:225–36.
3.Stang PE, Ryan PB, Racoosin JA, *et al*. Advancing the Science for Active Surveillance: Rationale and Design for the Observational Medical Outcomes Partnership. *Ann Intern Med*. 2010;153:600–6.
4. Informatics OHDS and. *The Book of OHDSI*. https://ohdsi.github.io/TheBookOfOhdsi/ (accessed 13 December 2023).

Angela Leis, Philippe Mortier, Franco Amigo, Madhav Bhargav, Susana Conde, Montserrat Ferrer, Oskar Flygare, Busenur Kizilaslan, Laura Latorre, Miguel Angel Mayer, Víctor Pérez Sola, Ana Portillo van Diest, Juan Manuel Ramírez-Anguita, Ferran Sanz, Gemma Vilagut, Jordi Alonso, Lars Mehlum, Ella Arensman, Johan Bjureberg, Manuel Pastor, Ping Qin, Sofia Bettella, Gabriel de Maeztu

The PERMANENS project is supported by Instituto de Salud Carlos III (ISCIII) and by the European Union NextGenerationEU, Mecanismo para la Recuperación y la Resiliencia (AC22/00006; AC22/00045), the Swedish Innovation Agency (no. 2022-00549), the Research Council of Norway (project no. 342386) and the Health Research Board Ireland (EIA/PERMEID/2022) under the frame of ERA PerMed

#OHDSISocialShowcase

FRIDAY

Automation of STCM Vocabularies Review for OMOP CDM

(Wai Yi Man, Antonella Delmestri)

Automation of STCM Review saves from HOURS to DAYS of human time

Presenter: Wai Yi Man

OHDSI

Automation of STCM Vocabularies Review for OMOP CDM

Background: The SOURCE TO CONCEPT MAP (STCM) approach maps local source codes not present in the Athena vocabularies to standard concepts. The vocabularies keep evolving to allow updated and accurate data mapping, and this requires regular manual reviews: each concept ID in an STCM-tailored CSV-formatted vocabulary, which must be checked against Athena, and then updated if necessary. This process must be repeated at each new data release. We have implemented an algorithm to automate this process.

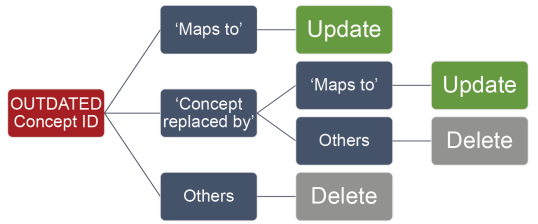
Results:

In the last release of CPRD GOLD* (July 2023 data release) there were over 5,000 concept IDs. Reviewing one concept typically takes between 30 seconds and 2 minutes for one person. By multiplying this by 5,000 concept IDs, the range of saved human hours per CPRD GOLD release extends between 41.7-166.7 hours. Our algorithm is fully automated, and it takes from seconds to minutes to review around ~5,000 concept IDs. This automation reduces the overall processing time, and removes any potential human errors, producing higher quality data and a more robust data mapping cycle.

*Clinical Practice Research Datalink (CPRD) GOLD is a well-established structured UK primary care data source with over twenty million patients and more than ten billion events in total. In 2022, CPRD GOLD, mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) was onboarded in DARWIN EU®. The University of Oxford is committed to onboarding in DARWIN EU® a new OMOP release of CPRD GOLD every six months.

Methods

Before starting the Extract, Transform, and Load (ETL) process, our algorithm, developed in a Python program, searches all outdated (non-standard or/and invalid) target concept IDs in the STCM files, and identifies the new associated standard concept IDs. For each outdated STCM target concept ID, there are three possible scenarios:



1. **The outdated STCM target concept ID is mapped to a new Athena standard concept ID (RELATIONSHIP_ID= 'Maps to') and the STCM target concept ID is updated.**
2. **The outdated STCM target concept ID is replaced by another Athena concept ID (RELATIONSHIP_ID= 'Concept replaced by'), which has then to be mapped. The mapping goes under an extra check as the replaced concept may be invalid or not standard. Likewise, the mapped concept is updated if a 'Maps to' concept exists or deleted if no 'Maps to' concept is found.** Duplication results from scenarios 1 and 2 are removed as 'Maps to' and 'Concept replaced by' relationships are not mutually exclusive for a concept.
3. **Others (neither 1 nor 2. A non-standard or/and invalid concept can exist without any relationships with other concepts). The concept is deleted.**

Wai Yi Man, Antonella Delmestri

NDORMS NUFFIELD DEPARTMENT OF ORTHOPAEDICS, RHEUMATOLOGY AND MUSCULOSKELETAL SCIENCES

UNIVERSITY OF OXFORD

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?

Oct 8: Advances in Methods Research

Yong Chen

Professor of Biostatistics, University of Pennsylvania

Topic: negative controls in vaccine research

George Hripcsak

Vivian Beaumont Allen Professor of Biomedical Informatics, Columbia University

Topic: new diagnostics for covariate balance in small samples

Shounak Chattopadhyay

Postdoctoral Scholar, UCLA

Topic: performance of the concurrent comparator design

Also: Update on the 2024-25 Vocabulary Roadmap

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