Introduction to the HADES TreatmentPatterns Package

OHDSI Community Call
April 15, 2025 • 11 am ET

OHDSI
Observational Health Data Sciences and Informatics

Upcoming Community Calls

Date	Topic
Apr. 15	Treatment Patterns
Apr. 22	Current Practices in Estimation and Prediction
Apr. 29	DevCon 2025 Review
May 6	Evidence Synthesis
May 13	Maternal Health Fellowship Review
May 20	Guideline-Driven Evidence Study Review
May 27	Collaborator Showcase Brainstorm (Deadline is July 1)

Three Stages of The Journey

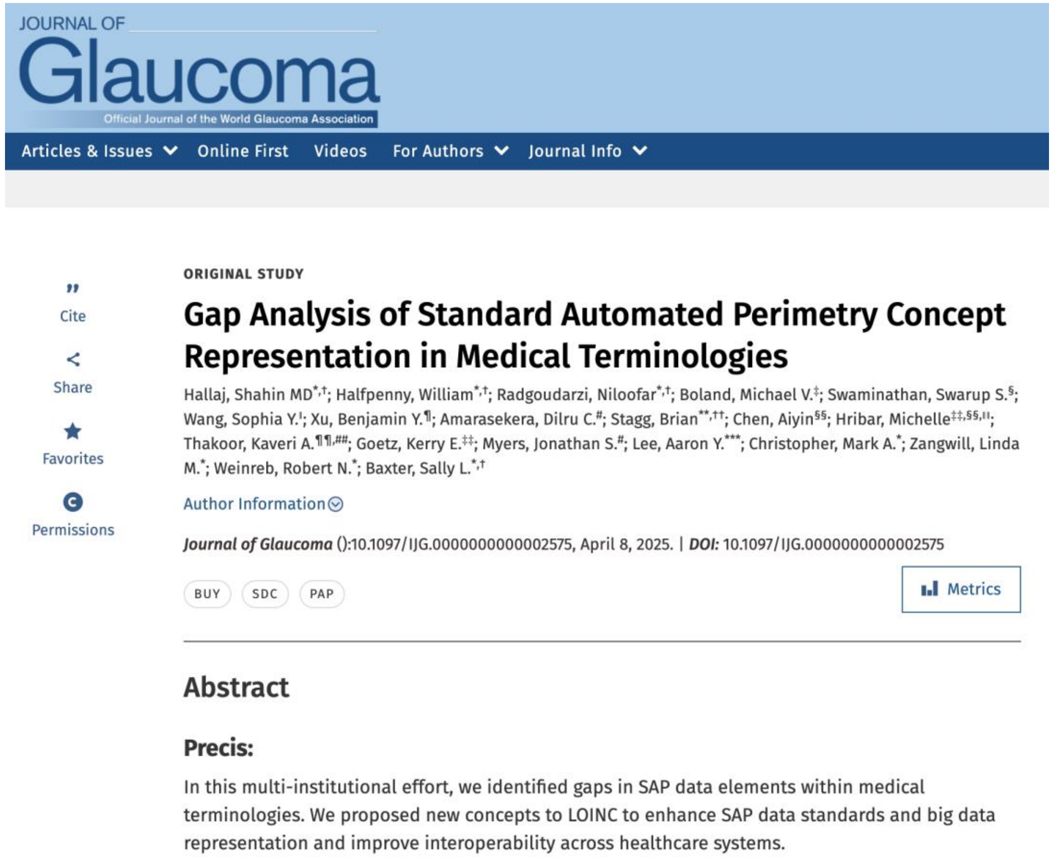
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Shahin Hallaj, William Halfpenny, Niloofar Radgoudarzi, Michael V Boland, Swarup S Swaminathan, Sophia Y Wang, Benjamin Y Xu, Dilru C Amarasekera, Brian Stagg, Aiyin Chen, Michelle Hribar, Kaveri A Thakoor, Kerry E Goetz, Jonathan S Myers, Aaron Y Lee, Mark A Christopher, Linda M Zangwill, Robert N Weinreb, and Sally L Baxter** on the publication of **Gap Analysis of Standard Automated Perimetry Concept Representation in Medical Terminologies** in the *Journal of Glaucoma.*



JOURNAL OF Glaucoma

Official Journal of the World Glaucoma Association

Articles & Issues | Online First | Videos | For Authors | Journal Info

Cite | Share | Favorites | Permissions

ORIGINAL STUDY

Gap Analysis of Standard Automated Perimetry Concept Representation in Medical Terminologies

Hallaj, Shahin MD[*,†]; Halfpenny, William[*,†]; Radgoudarzi, Niloofar[*,†]; Boland, Michael V.[‡]; Swaminathan, Swarup S.[§]; Wang, Sophia Y.[∥]; Xu, Benjamin Y.[¶]; Amarasekera, Dilru C.[#]; Stagg, Brian[**,††]; Chen, Aiyin[§§]; Hribar, Michelle[‡‡,§§,∥∥]; Thakoor, Kaveri A.[¶¶,##]; Goetz, Kerry E.[‡‡]; Myers, Jonathan S.[#]; Lee, Aaron Y.[***]; Christopher, Mark A.[*]; Zangwill, Linda M.[*]; Weinreb, Robert N.[*]; Baxter, Sally L.[*,†]

Author Information

Journal of Glaucoma ():10.1097/IJG.0000000000002575, April 8, 2025. | ***DOI:*** 10.1097/IJG.0000000000002575

BUY | SDC | PAP | Metrics

Abstract

Precis:

In this multi-institutional effort, we identified gaps in SAP data elements within medical terminologies. We proposed new concepts to LOINC to enhance SAP data standards and big data representation and improve interoperability across healthcare systems.

OHDSI Shoutouts!

Congratulations to the team of **Elisa Henke, Stephan Lorenz, Michele Zoch, Martin Sedlmayr, and Yuan Peng** on the publication of **Mapping National Vocabularies to International Standards Using OHDSI Standardized Vocabularies** in *Volume 323 of Studies in Health Technology and Informatics.*

Envisioning the Future of Health Informatics and Digital Health 349
J. Mantas et al. (Eds.)

doi:10.3233/SHTI250110

Mapping National Vocabularies to International Standards Using OHDSI Standardized Vocabularies

Elisa HENKE[a,1], Stephan LORENZ[a], Michele ZOCH[a], Martin SEDLMAYR[a] and Yuan PENG[a]

[a] *Institute for Medical Informatics and Biometry, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany*

ORCiD ID: Elisa Henke 0000-0002-5002-2676

Abstract. Ensuring semantic interoperability in international studies is crucial. In this context, the mapping of national to international vocabularies is necessary. The Standardized Vocabularies of OHDSI provide such a mapping, which forms the basis for semantic interoperability in the standardized data model OMOP CDM. The aim of this paper is to provide a guideline for vocabulary mapping that supports developers in efficiently implementing the technical application of mappings into the ETL process for transforming data to OMOP CDM. By implementing materialized views and creating a decision tree, we provide a solid foundation for efficient semantic mapping in OMOP CDM. With our work, we mark an important step in realizing international observational studies based on OMOP CDM.

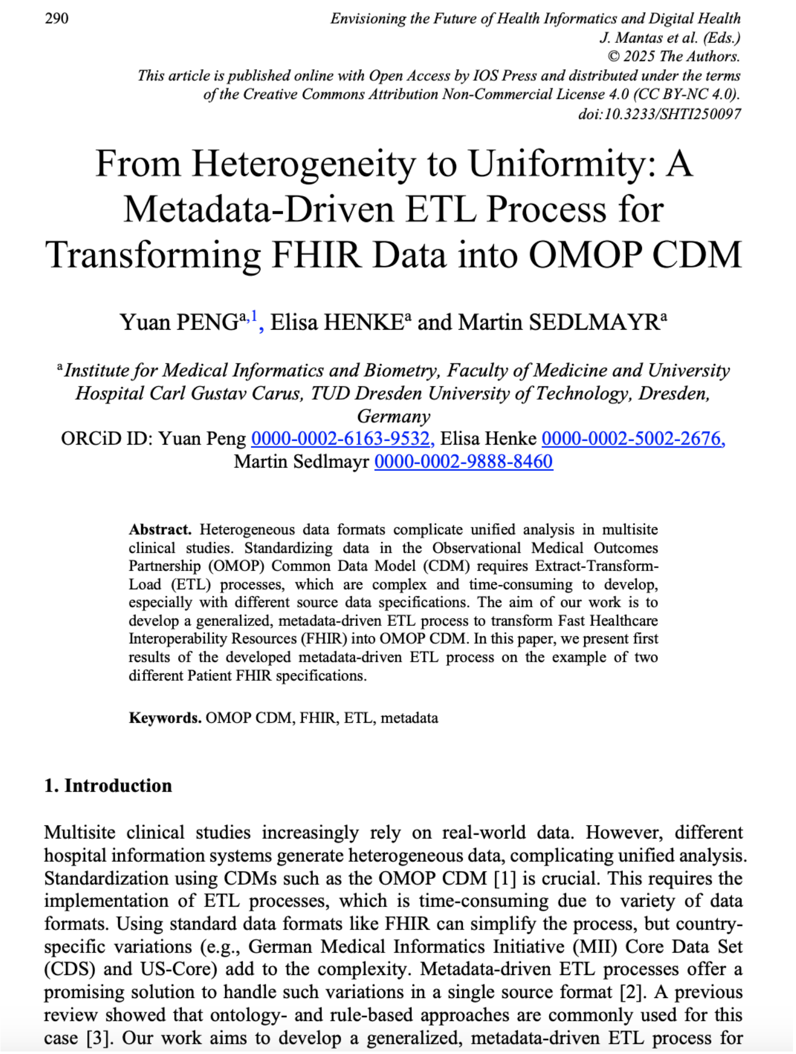
Keywords. OHDSI, OMOP CDM, vocabularies, semantic interoperability

1. Introduction

Research with secondary medical data across healthcare institutions requires the data to be interoperable. An important level of interoperability is semantic interoperability, which deals with the creation of a common understanding of message content. To ensure semantic interoperability, vocabularies with clearly defined codes have been introduced at national level, such as the International Classification of Diseases, Tenth Revision, German Edition (ICD-10-GM) for documenting diagnoses in Germany. However, these national, proprietary vocabularies cannot be used in full for international research as they may differ in structure, terminology or granularity from other standards.

OHDSI Shoutouts!

Congratulations to the team of **Yuan Peng, Elisa Henke, and Martin Sedlmayr** on the publication of **From Heterogeneity to Uniformity: A Metadata-Driven ETL Process for Transforming FHIR Data into OMOP CDM** in *Volume 323 of Studies in Health Technology and Informatics.*



Envisioning the Future of Health Informatics and Digital Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI250097

From Heterogeneity to Uniformity: A Metadata-Driven ETL Process for Transforming FHIR Data into OMOP CDM

Yuan PENG[a,1], Elisa HENKE[a] and Martin SEDLMAYR[a]

[a] *Institute for Medical Informatics and Biometry, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany*
ORCiD ID: Yuan Peng 0000-0002-6163-9532, Elisa Henke 0000-0002-5002-2676, Martin Sedlmayr 0000-0002-9888-8460

Abstract. Heterogeneous data formats complicate unified analysis in multisite clinical studies. Standardizing data in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) requires Extract-Transform-Load (ETL) processes, which are complex and time-consuming to develop, especially with different source data specifications. The aim of our work is to develop a generalized, metadata-driven ETL process to transform Fast Healthcare Interoperability Resources (FHIR) into OMOP CDM. In this paper, we present first results of the developed metadata-driven ETL process on the example of two different Patient FHIR specifications.

Keywords. OMOP CDM, FHIR, ETL, metadata

1. Introduction

Multisite clinical studies increasingly rely on real-world data. However, different hospital information systems generate heterogeneous data, complicating unified analysis. Standardization using CDMs such as the OMOP CDM [1] is crucial. This requires the implementation of ETL processes, which is time-consuming due to variety of data formats. Using standard data formats like FHIR can simplify the process, but country-specific variations (e.g., German Medical Informatics Initiative (MII) Core Data Set (CDS) and US-Core) add to the complexity. Metadata-driven ETL processes offer a promising solution to handle such variations in a single source format [2]. A previous review showed that ontology- and rule-based approaches are commonly used for this case [3]. Our work aims to develop a generalized, metadata-driven ETL process for

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
Tuesday	12 pm	ATLAS
Wednesday	7 am	Medical Imaging
Wednesday	1 pm	Perinatal & Reproductive Health
Thursday	8 am	Medical Devices
Thursday	11 am	Themis
Thursday	12 pm	HADES
Thursday	7 pm	Dentistry
Friday	10 am	GIS - Geographic Information System
Friday	10 am	Transplant
Friday	10:30 am	Open-Source Community
Friday	11:30 am	Steering
Monday	11 am	Data Bricks User Group
Monday	11 am	Book of OHDSI
Monday	2 pm	Electronic Animal Health Records
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	9:30 am	Common Data Model

DevCon 2025: April 25

Agenda

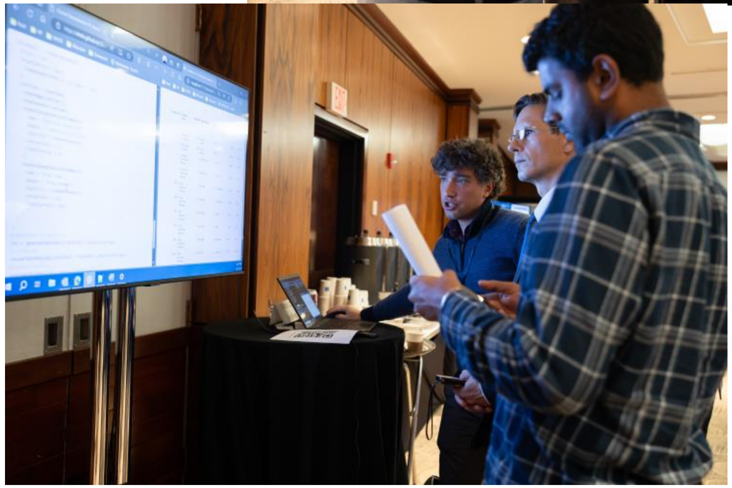
9:00 – 9:15am ET · Welcome & Introduction
- Paul Nagy, Johns Hopkins University

9:15 – 11:30am ET · OHDSI Projects Lightning Talks
- Stabilizing Gaia Core – Robert Miller, Miller Data Solutions
- CustomVocabularyBuilder – Jared Houghtaling, Tufts University
- CohortConstructor – Núria Mercadé-Besora, University of Oxford
- Updates on Strategus – Anthony Sena, Johnson & Johnson
- Experiences with SQLMesh/CICD integration with Databricks – Vishnu Chandrabalan, Lancashire Teaching Hospitals NHS Foundation Trust
- Updates from the Technical Advisory Board – Frank Defalco, Johnson & Johnson

11:30 – 12:30pm ET · Developer dialogue: Dev ops, DBT and, of course, LLMs
Moderator: Katy Sadowski, Boehringer Ingelheim
- Eduard Korchmar, EPAM Systems
- Egill Fridgeirsson, Erasmus MC
- Martin Lavallee, Boehringer Ingelheim
- Lawrence Adams, Artificial Intelligence Centre for Value Based Healthcare

12:30 – 1:00pm ET · Break

1:00 – 2:00pm ET · Sustainable Open-Source Ecosystems Panel
Moderator: Paul Nagy, Sean O'Reilly
- Data4Life – Peter Hoffmann
- The Hyve – Jan Blom/Wouter Franke
- Cognome – James Green



ohdsi.org/DevCon2025

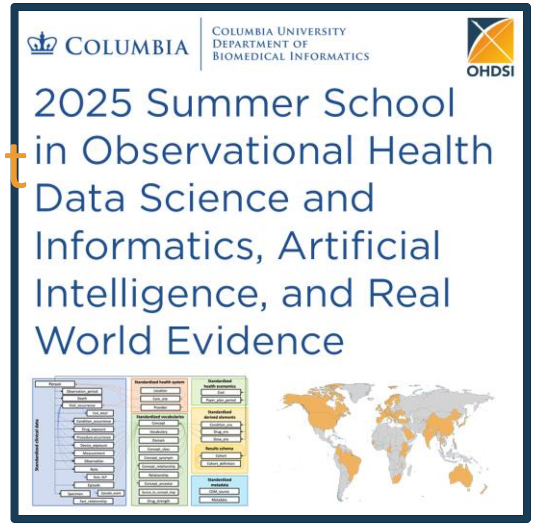
Save The Date!

The submission deadline for the 2025 Global Symposium Collaborator Showcase is July 1.

The showcase will be accepting both posters and software demos, as well as interest in hosting lightning talks. More information on the symposium, including abstract submission and registration links, will be available soon.

Join the OHDSI Summer School!

Registration is open for the first ever OHDSI Summer School, held July 14-18, 2025, at the Columbia University Department of Biomedical Informatics.



The Columbia Summer School in Observational Health Data Science and Informatics, Artificial Intelligence, and Real World Evidence (RWE) offers health professionals, researchers and industry practitioners the opportunity to gain familiarity and hands-on experience with real world data and generating real world evidence. Participants will learn about the different types of healthcare data captured during routine clinical care, including electronic health records and administrative records, and how these data can be standardized to the OMOP Common Data Model to enable distributed data network research.

Meet Our Faculty

George Hripcsak, MD MS
Vivian Beaumont Allen Professor of Biomedical Informatics

Patrick Ryan, PhD
Adjuct Assistant Professor of Biomedical Informatics

Anna Ostropolets, MD PhD
Adjuct Assistant Professor of Biomedical Informatics

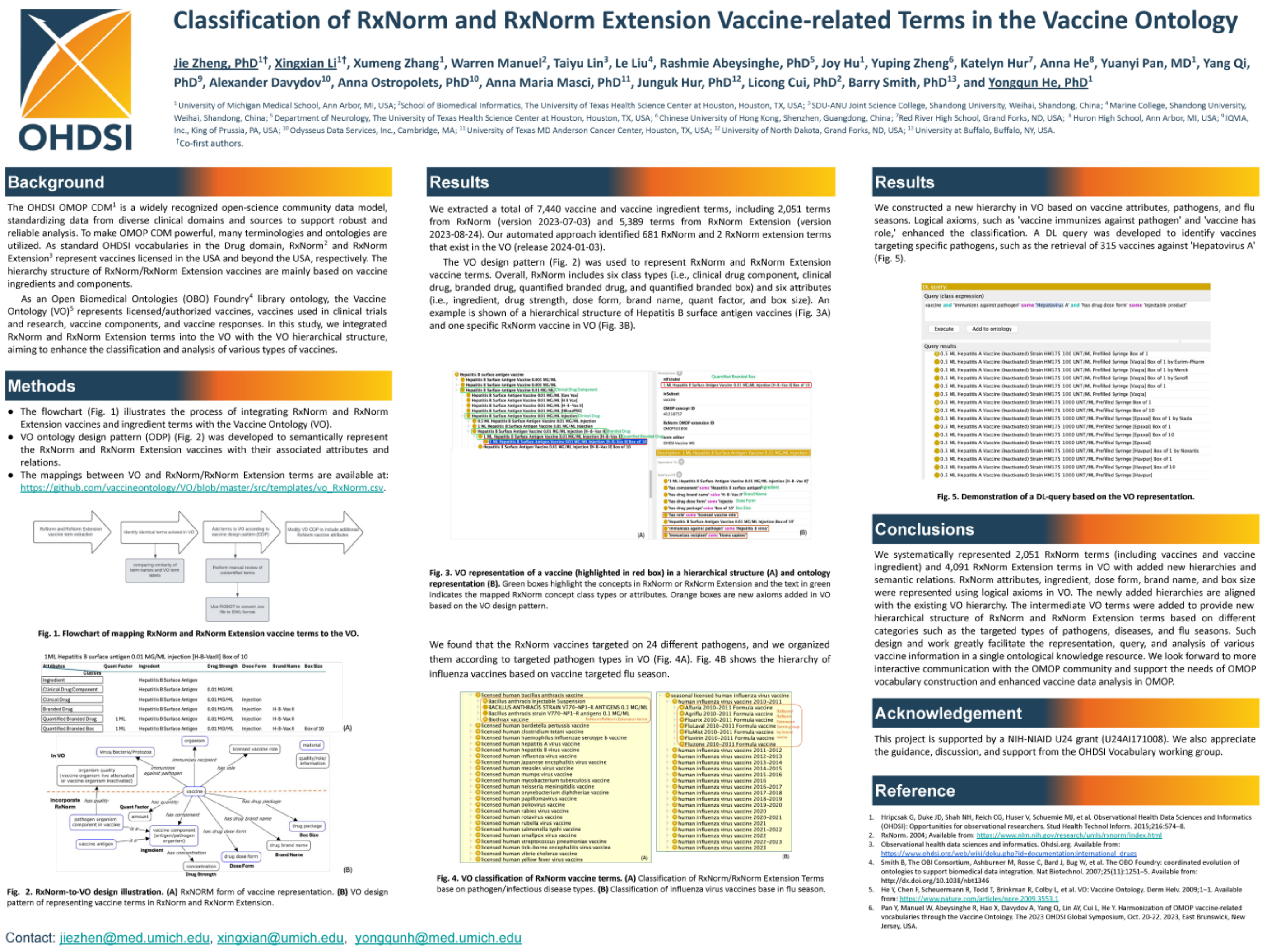
Karthik Natarajan, PhD
Assistant Professor of Biomedical Informatics

#OHDSISocialShowcase This Week

Monday

Classification of RxNorm and RxNorm Extension Vaccine-related Terms in the Vaccine Ontology

(Jie Zheng, Xingxian Li, Ellen Zhang, Warren Manuel, Rashmie Abeysinghe, Joy Hu, Yuping Zheng, Taiyu Lin, Katelyn Hur, Anna He, Yang Qi, Alexander Davydov, Anna Ostropolets, Anna Maria Masci, Junguk Hur, Licong Cui, Barry Smith, Yongqun He)



Classification of RxNorm and RxNorm Extension Vaccine-related Terms in the Vaccine Ontology

Jie Zheng, PhD[1†], Xingxian Li[1†], Xumeng Zhang[1], Warren Manuel[2], Taiyu Lin[3], Le Liu[4], Rashmie Abeysinghe, PhD[5], Joy Hu[1], Yuping Zheng[6], Katelyn Hur[7], Anna He[8], Yuanyi Pan, MD[1], Yang Qi, PhD[9], Alexander Davydov[10], Anna Ostropolets, PhD[10], Anna Maria Masci, PhD[11], Junguk Hur, PhD[12], Licong Cui, PhD[2], Barry Smith, PhD[13], and Yongqun He, PhD[1]

[1] University of Michigan Medical School, Ann Arbor, MI, USA; [2] School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA; [3] SDU-ANU Joint Science College, Shandong University, Weihai, Shandong, China; [4] Marine College, Shandong University, Weihai, Shandong, China; [5] Department of Neurology, The University of Texas Health Science Center at Houston, Houston, TX, USA; [6] Chinese University of Hong Kong, Shenzhen, Guangdong, China; [7] Red River High School, Grand Forks, ND, USA; [8] Huron High School, Ann Arbor, MI, USA; [9] IQVIA, Inc., King of Prussia, PA, USA; [10] Odysseus Data Services, Inc., Cambridge, MA; [11] University of Texas MD Anderson Cancer Center, Houston, TX, USA; [12] University of North Dakota, Grand Forks, ND, USA; [13] University at Buffalo, Buffalo, NY, USA.

†Co-first authors.

Background

The OHDSI OMOP CDM[1] is a widely recognized open-science community data model, standardizing data from diverse clinical domains and sources to support robust and reliable analysis. To make OMOP CDM powerful, many terminologies and ontologies are utilized. As standard OHDSI vocabularies in the Drug domain, RxNorm[2] and RxNorm Extension[3] represent vaccines licensed in the USA and beyond the USA, respectively. The hierarchy structure of RxNorm/RxNorm Extension vaccines are mainly based on vaccine ingredients and components.

As an Open Biomedical Ontologies (OBO) Foundry[4] library ontology, the Vaccine Ontology (VO)[5] represents licensed/authorized vaccines, vaccines used in clinical trials and research, vaccine components, and vaccine responses. In this study, we integrated RxNorm and RxNorm Extension terms into the VO with the VO hierarchical structure, aiming to enhance the classification and analysis of various types of vaccines.

Methods

- The flowchart (Fig. 1) illustrates the process of integrating RxNorm and RxNorm Extension vaccines and ingredient terms with the Vaccine Ontology (VO).
- VO ontology design pattern (ODP) (Fig. 2) was developed to semantically represent the RxNorm and RxNorm Extension vaccines with their associated attributes and relations.
- The mappings between VO and RxNorm/RxNorm Extension terms are available at: https://github.com/vaccineontology/VO/blob/master/src/templates/vo_RxNorm.csv.

Fig. 1. Flowchart of mapping RxNorm and RxNorm Extension vaccine terms to the VO.

1ML Hepatitis B surface antigen 0.01 MG/ML injection [H-B-VaxII] Box of 10

Classes / Attributes	Quant Factor	Ingredient	Drug Strength	Dose Form	Brand Name	Box Size
Ingredient		Hepatitis B Surface Antigen				
Clinical Drug Component		Hepatitis B Surface Antigen	0.01 MG/ML			
Clinical Drug		Hepatitis B Surface Antigen	0.01 MG/ML	Injection		
Branded Drug		Hepatitis B Surface Antigen	0.01 MG/ML	Injection	H-B-Vax II	
Quantified Branded Drug	1 ML	Hepatitis B Surface Antigen	0.01 MG/ML	Injection	H-B-Vax II	
Quantified Branded Box	1 ML	Hepatitis B Surface Antigen	0.01 MG/ML	Injection	H-B-Vax II	Box of 10

(A)

Fig. 2. RxNorm-to-VO design illustration. (A) RxNORM form of vaccine representation. **(B)** VO design pattern of representing vaccine terms in RxNorm and RxNorm Extension.

Contact: jiezhen@med.umich.edu, xingxian@umich.edu, yongqunh@med.umich.edu

Results

We extracted a total of 7,440 vaccine and vaccine ingredient terms, including 2,051 terms from RxNorm (version 2023-07-03) and 5,389 terms from RxNorm Extension (version 2023-08-24). Our automated approach identified 681 RxNorm and 2 RxNorm extension terms that exist in the VO (release 2024-01-03).

The VO design pattern (Fig. 2) was used to represent RxNorm and RxNorm Extension vaccine terms. Overall, RxNorm includes six class types (i.e., clinical drug component, clinical drug, branded drug, quantified branded drug, and quantified branded box) and six attributes (i.e., ingredient, drug strength, dose form, brand name, quant factor, and box size). An example is shown of a hierarchical structure of Hepatitis B surface antigen vaccines (Fig. 3A) and one specific RxNorm vaccine in VO (Fig. 3B).

Fig. 3. VO representation of a vaccine (highlighted in red box) in a hierarchical structure (A) and ontology representation (B). Green boxes highlight the concepts in RxNorm or RxNorm Extension and the text in green indicates the mapped RxNorm concept class types or attributes. Orange boxes are new axioms added in VO based on the VO design pattern.

We found that the RxNorm vaccines targeted on 24 different pathogens, and we organized them according to targeted pathogen types in VO (Fig. 4A). Fig. 4B shows the hierarchy of influenza vaccines based on vaccine targeted flu season.

Fig. 4. VO classification of RxNorm vaccine terms. (A) Classification of RxNorm/RxNorm Extension Terms base on pathogen/infectious disease types. **(B)** Classification of influenza virus vaccines base in flu season.

Results

We constructed a new hierarchy in VO based on vaccine attributes, pathogens, and flu seasons. Logical axioms, such as 'vaccine immunizes against pathogen' and 'vaccine has role,' enhanced the classification. A DL query was developed to identify vaccines targeting specific pathogens, such as the retrieval of 315 vaccines against 'Hepatovirus A' (Fig. 5).

Fig. 5. Demonstration of a DL-query based on the VO representation.

Conclusions

We systematically represented 2,051 RxNorm terms (including vaccines and vaccine ingredient) and 4,091 RxNorm Extension terms in VO with added new hierarchies and semantic relations. RxNorm attributes, ingredient, dose form, brand name, and box size were represented using logical axioms in VO. The newly added hierarchies are aligned with the existing VO hierarchy. The intermediate VO terms were added to provide new hierarchical structure of RxNorm and RxNorm Extension terms based on different categories such as the targeted types of pathogens, diseases, and flu seasons. Such design and work greatly facilitate the representation, query, and analysis of various vaccine information in a single ontological knowledge resource. We look forward to more interactive communication with the OMOP community and support the needs of OMOP vocabulary construction and enhanced vaccine data analysis in OMOP.

Acknowledgement

This project is supported by a NIH-NIAID U24 grant (U24AI171008). We also appreciate the guidance, discussion, and support from the OHDSI Vocabulary working group.

Reference

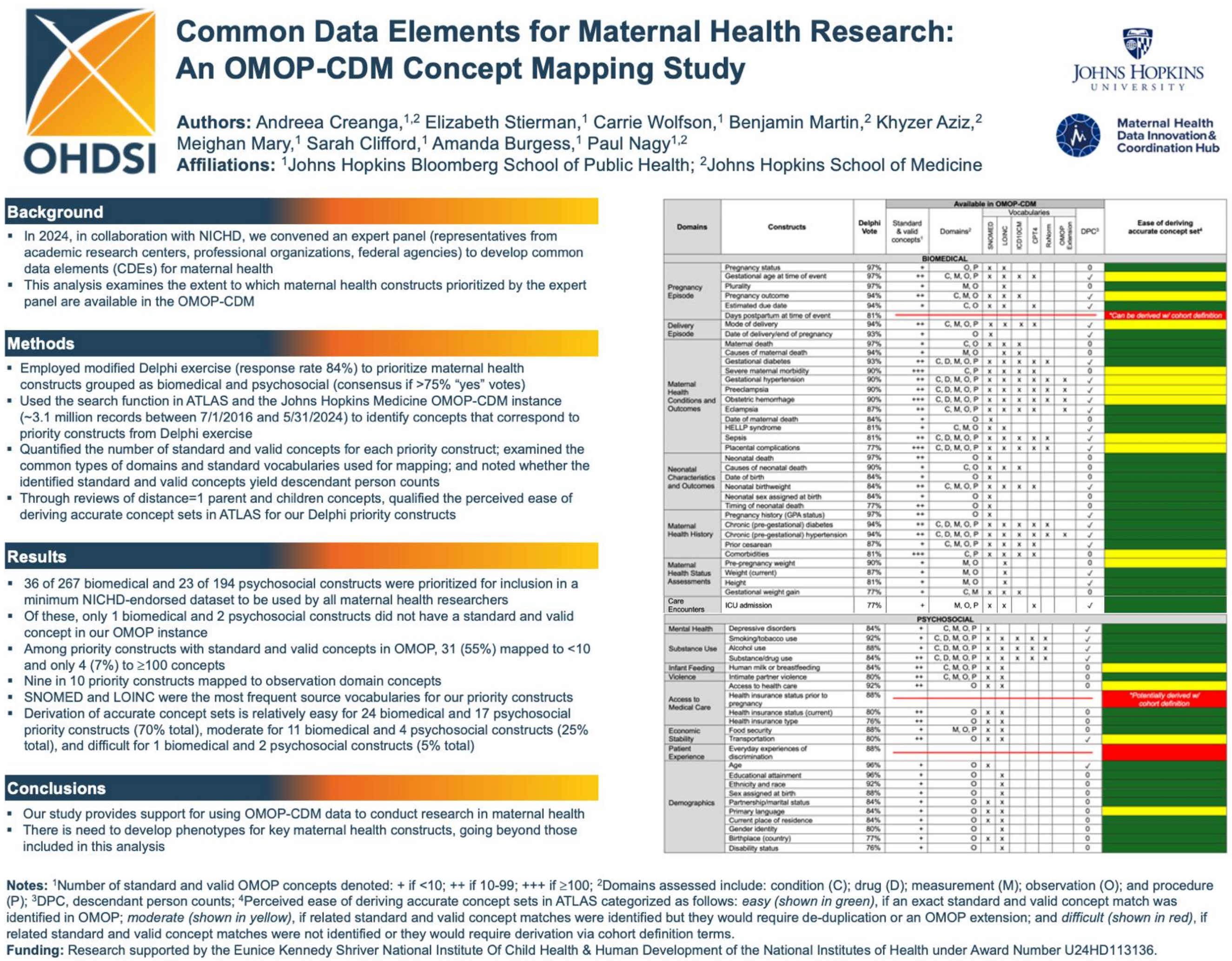
1. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. Stud Health Technol Inform. 2015;216:574–8.
2. RxNorm. 2004; Available from: https://www.nlm.nih.gov/research/umls/rxnorm/index.html
3. Observational health data sciences and informatics. Ohdsi.org. Available from: https://www.ohdsi.org/web/wiki/doku.php?id=documentation:international_drugs
4. Smith B, The OBI Consortium, Ashburner M, Rosse C, Bard J, Bug W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007;25(11):1251–5. Available from: http://dx.doi.org/10.1038/nbt1346
5. He Y, Chen F, Scheuermann R, Todd T, Brinkman R, Colby L, et al. VO: Vaccine Ontology. Derm Helv. 2009;1–1. Available from: https://www.nature.com/articles/npre.2009.3553.1
6. Pan Y, Manuel W, Abeysinghe R, Hao X, Davydov A, Yang Q, Lin AY, Cui L, He Y. Harmonization of OMOP vaccine-related vocabularies through the Vaccine Ontology. The 2023 OHDSI Global Symposium, Oct. 20-22, 2023, East Brunswick, New Jersey, USA.

#OHDSISocialShowcase This Week

Tuesday

Common Data Elements for Maternal Health Research: An OMOP-CDM Concept Mapping Study

(**Andreea Creanga**, Elizabeth Stierman, Carrie Wolfson, Benjamin Martin, Khyzer Aziz, Meighan Mary, Sarah Clifford, Amanda Burgess, Paul Nagy)



Common Data Elements for Maternal Health Research: An OMOP-CDM Concept Mapping Study

Authors: Andreea Creanga,[1,2] Elizabeth Stierman,[1] Carrie Wolfson,[1] Benjamin Martin,[2] Khyzer Aziz,[2] Meighan Mary,[1] Sarah Clifford,[1] Amanda Burgess,[1] Paul Nagy[1,2]
Affiliations: [1]Johns Hopkins Bloomberg School of Public Health; [2]Johns Hopkins School of Medicine

JOHNS HOPKINS UNIVERSITY

Maternal Health Data Innovation & Coordination Hub

Background

- In 2024, in collaboration with NICHD, we convened an expert panel (representatives from academic research centers, professional organizations, federal agencies) to develop common data elements (CDEs) for maternal health
- This analysis examines the extent to which maternal health constructs prioritized by the expert panel are available in the OMOP-CDM

Methods

- Employed modified Delphi exercise (response rate 84%) to prioritize maternal health constructs grouped as biomedical and psychosocial (consensus if >75% "yes" votes)
- Used the search function in ATLAS and the Johns Hopkins Medicine OMOP-CDM instance (~3.1 million records between 7/1/2016 and 5/31/2024) to identify concepts that correspond to priority constructs from Delphi exercise
- Quantified the number of standard and valid concepts for each priority construct; examined the common types of domains and standard vocabularies used for mapping; and noted whether the identified standard and valid concepts yield descendant person counts
- Through reviews of distance=1 parent and children concepts, qualified the perceived ease of deriving accurate concept sets in ATLAS for our Delphi priority constructs

Results

- 36 of 267 biomedical and 23 of 194 psychosocial constructs were prioritized for inclusion in a minimum NICHD-endorsed dataset to be used by all maternal health researchers
- Of these, only 1 biomedical and 2 psychosocial constructs did not have a standard and valid concept in our OMOP instance
- Among priority constructs with standard and valid concepts in OMOP, 31 (55%) mapped to <10 and only 4 (7%) to ≥100 concepts
- Nine in 10 priority constructs mapped to observation domain concepts
- SNOMED and LOINC were the most frequent source vocabularies for our priority constructs
- Derivation of accurate concept sets is relatively easy for 24 biomedical and 17 psychosocial priority constructs (70% total), moderate for 11 biomedical and 4 psychosocial constructs (25% total), and difficult for 1 biomedical and 2 psychosocial constructs (5% total)

Conclusions

- Our study provides support for using OMOP-CDM data to conduct research in maternal health
- There is need to develop phenotypes for key maternal health constructs, going beyond those included in this analysis

Domains	Constructs	Delphi Vote	Standard & valid concepts[1]	Domains[2]	SNOMED	LOINC	ICD10CM	CPT4	RxNorm	OMOP Extension	DPC[3]	Ease of deriving accurate concept set[4]
			Available in OMOP-CDM		Vocabularies							
BIOMEDICAL												
Pregnancy Episode	Pregnancy status	97%	+	O, P	x	x					0	
	Gestational age at time of event	97%	++	C, M, O, P	x	x	x	x			✓	
	Plurality	97%	+	M, O		x					0	
	Pregnancy outcome	94%	++	C, M, O	x	x	x				✓	
	Estimated due date	94%	+	C, O	x	x		x			✓	
	Days postpartum at time of event	81%										*Can be derived w/ cohort definition
Delivery Episode	Mode of delivery	94%	++	C, M, O, P	x	x	x	x			✓	
	Date of delivery/end of pregnancy	93%	+	O	x						✓	
Maternal Health Conditions and Outcomes	Maternal death	97%	+	C, O	x	x	x				0	
	Causes of maternal death	94%	+	M, O		x	x				0	
	Gestational diabetes	93%	++	C, D, M, O, P	x	x	x	x	x		✓	
	Severe maternal morbidity	90%	+++	C, P	x	x	x	x			0	
	Gestational hypertension	90%	++	C, D, M, O, P	x	x	x	x	x	x	✓	
	Preeclampsia	90%	++	C, D, M, O, P	x	x	x	x	x	x	✓	
	Obstetric hemorrhage	90%	+++	C, D, M, O, P	x	x	x	x	x	x	✓	
	Eclampsia	87%	++	C, M, O, P	x	x	x	x		x	✓	
	Date of maternal death	84%	+	O	x						0	
	HELLP syndrome	81%	+	C, M, O	x	x					✓	
	Sepsis	81%	++	C, D, M, O, P	x	x	x	x	x		✓	
	Placental complications	77%	+++	C, D, M, O, P	x	x	x	x	x		✓	
Neonatal Characteristics and Outcomes	Neonatal death	97%	++	O	x						0	
	Causes of neonatal death	90%	+	C, O	x	x	x				0	
	Date of birth	84%	+	O	x						0	
	Neonatal birthweight	84%	++	C, M, O, P	x	x	x	x			✓	
	Neonatal sex assigned at birth	84%	+	O	x						0	
	Timing of neonatal death	77%	++	O	x						0	
Maternal Health History	Pregnancy history (GPA status)	97%	++	O	x						✓	
	Chronic (pre-gestational) diabetes	94%	++	C, D, M, O, P	x	x	x	x	x		✓	
	Chronic (pre-gestational) hypertension	94%	++	C, D, M, O, P	x	x	x	x	x	x	✓	
	Prior cesarean	87%	+	C, M, O, P	x	x	x	x			✓	
	Comorbidities	81%	+++	C, P	x	x	x	x			0	
Maternal Health Status Assessments	Pre-pregnancy weight	90%	+	M, O		x					0	
	Weight (current)	87%	+	M, O		x					✓	
	Height	81%	+	M, O		x					✓	
	Gestational weight gain	77%	+	C, M	x	x	x				0	
Care Encounters	ICU admission	77%	+	M, O, P	x	x		x			✓	
PSYCHOSOCIAL												
Mental Health	Depressive disorders	84%	+	C, M, O, P	x						✓	
Substance Use	Smoking/tobacco use	92%	+	C, D, M, O, P	x	x	x	x	x		✓	
	Alcohol use	88%	+	C, D, M, O, P	x	x	x	x	x		✓	
	Substance/drug use	84%	++	C, D, M, O, P	x	x	x	x	x		✓	
Infant Feeding	Human milk or breastfeeding	84%	++	C, M, O, P	x	x					0	
Violence	Intimate partner violence	80%	++	C, M, O, P	x	x					0	
Access to Medical Care	Access to health care	92%	++	O	x	x					0	
	Health insurance status prior to pregnancy	88%										*Potentially derived w/ cohort definition
	Health insurance status (current)	80%	++	O	x	x					0	
	Health insurance type	76%	++	O	x	x					0	
Economic Stability	Food security	88%	+	M, O, P	x	x					0	
	Transportation	80%	++	O	x	x					✓	
Patient Experience	Everyday experiences of discrimination	88%										
Demographics	Age	96%	+	O	x						✓	
	Educational attainment	96%	+	O		x					0	
	Ethnicity and race	92%	+	O		x					0	
	Sex assigned at birth	88%	+	O		x					0	
	Partnership/marital status	84%	+	O	x	x					0	
	Primary language	84%	+	O	x	x					0	
	Current place of residence	84%	+	O	x	x					0	
	Gender identity	80%	+	O		x					0	
	Birthplace (country)	77%	+	O	x	x					0	
	Disability status	76%	+	O		x					0	

Notes: [1]Number of standard and valid OMOP concepts denoted: + if <10; ++ if 10-99; +++ if ≥100; [2]Domains assessed include: condition (C); drug (D); measurement (M); observation (O); and procedure (P); [3]DPC, descendant person counts; [4]Perceived ease of deriving accurate concept sets in ATLAS categorized as follows: *easy (shown in green)*, if an exact standard and valid concept match was identified in OMOP; *moderate (shown in yellow)*, if related standard and valid concept matches were identified but they would require de-duplication or an OMOP extension; and *difficult (shown in red)*, if related standard and valid concept matches were not identified or they would require derivation via cohort definition terms.

Funding: Research supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number U24HD113136.

#OHDSISocialShowcase This Week

Wednesday

Design of Feedback Reports for Evaluating Data Fitness for Use in the Bridge2AI For Clinical Care Research Consortium

(Jared Houghtaling, Gilles Clermont, Andrew Williams)

Design of Feedback Reports for Evaluating Data Fitness for Use in the Bridge2AI For Clinical Care Research Consortium

Presenter: Jared Houghtaling

Intro:

The Bridge2AI for Clinical Care (B2AI For CC) research consortium consists of fifteen data contributing sites (DCS) led by various researcher teams based on expertise and objectives. Two of these teams, namely Standards and Data Acquisition, are responsible for providing guidance to DCS in their efforts to generate interoperable, multimodal data extracts. One critical tool in this guidance effort has been the creation and dissemination of formal reports that characterize submitted data extracts with respect to: (1) metadata (i.e. number of patients, size of files, etc.) about the delivery, (2) protected health information (PHI) and IRB compliance, (3) extent of cohort capture using validated definitions from the OHDSI Phenotype Library, (4) standard data quality and characterization checks (i.e. DQD[1] & Achilles[2]), and (5) fitness for use in the B2AI For CC consortium and comparisons with other data contributing sites. These reports serve both a prospective role as detailed instructions for sites to iteratively update their data extracts to suit the needs and requirements of the consortium, and a retrospective role to detail a history of the prior extracts those sites have delivered and the associated feedback they have received.

Methods:

a. *File Delivery Details*

Data contributing sites organized files (e.g. tabular, waveform, and imaging) in structured delivery packets; the AzureStor[3] package in R parsed these packets and retrieved metadata.

b. *Checks for PHI*

We executed a per-mode PHI scan using tooling developed within the B2AI For CC consortium,[4] and included details from these scans like pass/fail percentages and redacted contextual examples, as their own report section.

c. *Capture of Validated Phenotypes*

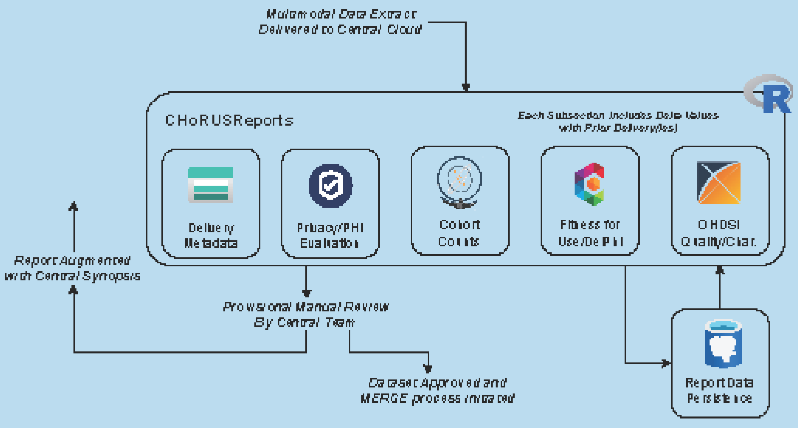
We used WebAPI to programmatically ingest validated cohort definitions from the OHDSI Phenotype Library, and then generated each of those cohorts on the resulting ingested data.

d. *Fitness for Use in B2AI For CC*

We included data characterization and data quality checks specific to the B2AI For CC consortium; these additional queries incorporate elements from a prioritization process[5] based on critical-care settings.

e. *Postgres-Based Report Data Persistence*

We stored R objects and report elements as large objects in the central database to calculate and persist changes in data content over time.



Feedback reports have helped to iteratively improve data contributions across this multi-partner consortium; ***emphasis on captured phenotypes as a metric for fitness for use*** has provided valuable context for granular quality and characterization details, and has accelerated targeted updates at sites.

Example Report Table Created From Data Packet Scan

category	file_count	person_count	person_delta	target_percent
ALL Data Modes	12,345	7,864	1,184	78.6
ANY Data Modes	21,546	9,423	2,000	94.3
OMOP Data	13	9,423	2,000	94.3
Imaging Data	17,583	8,654	1,578	86.5
Waveform Data	19,817	9,068	1,651	90.7
Note Data	2	9,253	1,889	92.5
Waveform Data	18,754	7,443	1,517	74.4
Note Data	14,576	6,846	1,254	68.5

Check out the full conference proceeding here:

Results & Discussion:

While the B2AI For CC consortium is still in nascent stages with regard to multimodal data standardization, ingestion, and consolidation, the feedback reports we present here have helped to accelerate iterative data contribution processes and to improve overall data quality. In contrast with the EHDEN consortium that aims to support large-scale federated analyses in which all data remains at each site, B2AI For CC is aggregating and curating a multi-site dataset centrally; the high level of detail, consortium specificity, and cross-site comparisons in reports for B2AI For CC relative to EHDEN[6] reflect this inherent advantage of broad data accessibility in a central location. Moreover, the reports we generate leverage and augment powerful OHDSI resources like the Phenotype Library in order to produce a phenotype-oriented overview of data fitness for use. These reports do not stand on their own; rather, they serve as a tool for the Standards and Data Acquisition teams to use during meetings with individual sites to review those sites' data packets and help to interpret and prioritize updates needed for subsequent deliveries.

Conclusions

- The work described here represents a first step toward a robust tool for enhancing consortium-wide data quality within B2AI For CC; we expect that the software we have developed[7] in support of this reporting functionality holds utility in the OHDSI community and other research consortia with similar objectives regarding data quality and interoperability.

- Much of the work presented here builds on the effort and dedication of so many others in the OHDSI community; we will continue to contribute to – and advocate for – open-source development of these powerful tools, and we plan to continue to share our efforts and experiences along the way.

Authors: Jared Houghtaling[a], Gilles Clermont[b], Andrew E. Williams[a]

[a] *Tufts Medicine - Institute for Clinical Research and Health Policy Studies (ICRHPS)*

[b] *University of Pittsburgh School of Medicine*

References:

TuftsMedicine

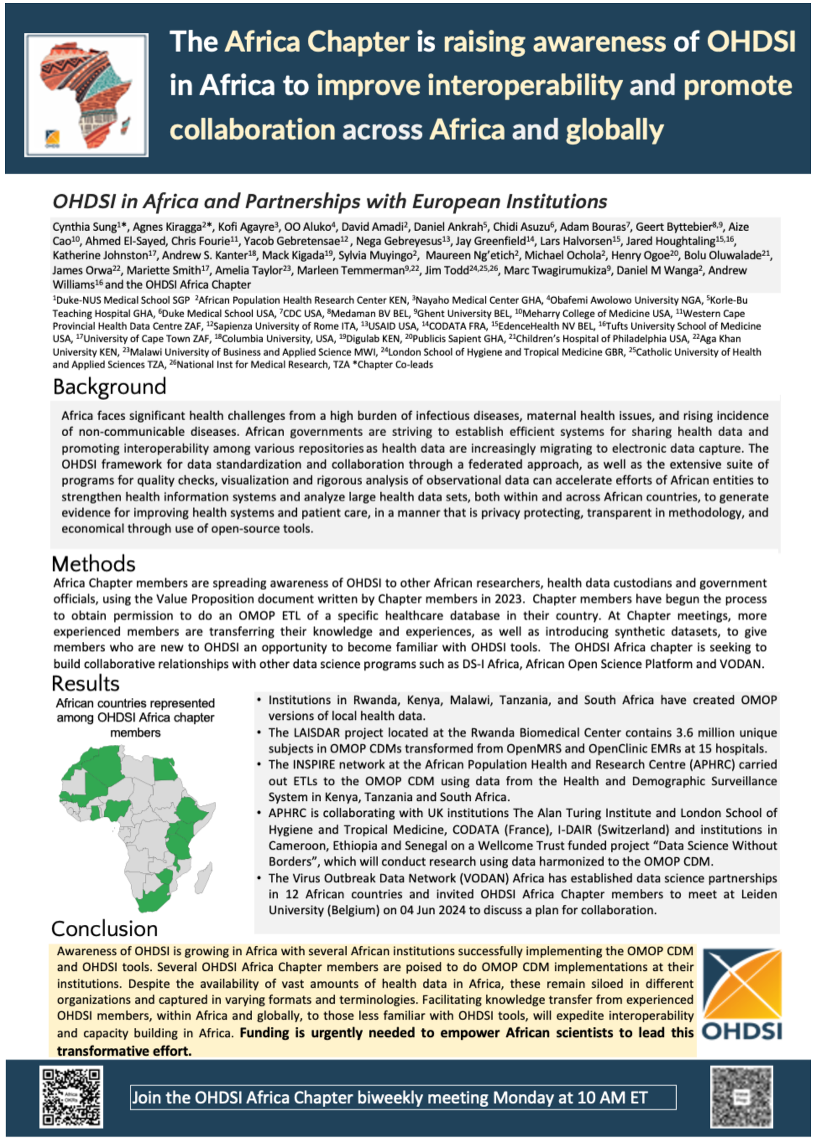
OHDSI

#OHDSISocialShowcase This Week

Thursday

OHDSI in Africa and Partnerships with European Institutions

(Cynthia Sung, Agnes Kiragga, Kofi Agayre, OO Aluko, David Amadi, Daniel Ankrah, Chidi Asuzu, Adam Bouras, Geert Byttebier, Aize Cao, Ahmed El-Sayed, Chris Fourie, Yacob Gebretensae , Nega Gebreyesus, Jay Greenfield, Lars Halvorsen, Jared Houghtaling, Katherine Johnston, Andrew S. Kanter, Mack Kigada, Sylvia Muyingo, Maureen Ng'etich, Michael Ochola, Henry Ogoe, Bolu Oluwalade, James Orwa, Mariette Smith, Amelia Taylor, Marleen Temmerman, Jim Todd, Marc Twagirumukiza, Daniel M Wanga, Andrew Williams)



The Africa Chapter is raising awareness of OHDSI in Africa to improve interoperability and promote collaboration across Africa and globally

OHDSI in Africa and Partnerships with European Institutions

Cynthia Sung[1*], Agnes Kiragga[2*], Kofi Agayre[3], OO Aluko[4], David Amadi[2], Daniel Ankrah[5], Chidi Asuzu[6], Adam Bouras[7], Geert Byttebier[8,9], Aize Cao[10], Ahmed El-Sayed, Chris Fourie[11], Yacob Gebretensae[12] , Nega Gebreyesus[13], Jay Greenfield[14], Lars Halvorsen[15], Jared Houghtaling[15,16], Katherine Johnston[17], Andrew S. Kanter[18], Mack Kigada[19], Sylvia Muyingo[2], Maureen Ng'etich[2], Michael Ochola[2], Henry Ogoe[20], Bolu Oluwalade[21], James Orwa[22], Mariette Smith[17], Amelia Taylor[23], Marleen Temmerman[9,22], Jim Todd[24,25,26], Marc Twagirumukiza[9], Daniel M Wanga[2], Andrew Williams[16] and the OHDSI Africa Chapter

[1]Duke-NUS Medical School SGP [2]African Population Health Research Center KEN, [3]Nayaho Medical Center GHA, [4]Obafemi Awolowo University NGA, [5]Korle-Bu Teaching Hospital GHA, [6]Duke Medical School USA, [7]CDC USA, [8]Medaman BV BEL, [9]Ghent University BEL, [10]Meharry College of Medicine USA, [11]Western Cape Provincial Health Data Centre ZAF, [12]Sapienza University of Rome ITA, [13]USAID USA, [14]CODATA FRA, [15]EdenceHealth NV BEL, [16]Tufts University School of Medicine USA, [17]University of Cape Town ZAF, [18]Columbia University, USA, [19]Digulab KEN, [20]Publicis Sapient GHA, [21]Children's Hospital of Philadelphia USA, [22]Aga Khan University KEN, [23]Malawi University of Business and Applied Science MWI, [24]London School of Hygiene and Tropical Medicine GBR, [25]Catholic University of Health and Applied Sciences TZA, [26]National Inst for Medical Research, TZA *Chapter Co-leads

Background

Africa faces significant health challenges from a high burden of infectious diseases, maternal health issues, and rising incidence of non-communicable diseases. African governments are striving to establish efficient systems for sharing health data and promoting interoperability among various repositories as health data are increasingly migrating to electronic data capture. The OHDSI framework for data standardization and collaboration through a federated approach, as well as the extensive suite of programs for quality checks, visualization and rigorous analysis of observational data can accelerate efforts of African entities to strengthen health information systems and analyze large health data sets, both within and across African countries, to generate evidence for improving health systems and patient care, in a manner that is privacy protecting, transparent in methodology, and economical through use of open-source tools.

Methods

Africa Chapter members are spreading awareness of OHDSI to other African researchers, health data custodians and government officials, using the Value Proposition document written by Chapter members in 2023. Chapter members have begun the process to obtain permission to do an OMOP ETL of a specific healthcare database in their country. At Chapter meetings, more experienced members are transferring their knowledge and experiences, as well as introducing synthetic datasets, to give members who are new to OHDSI an opportunity to become familiar with OHDSI tools. The OHDSI Africa chapter is seeking to build collaborative relationships with other data science programs such as DS-I Africa, African Open Science Platform and VODAN.

Results

African countries represented among OHDSI Africa chapter members

- Institutions in Rwanda, Kenya, Malawi, Tanzania, and South Africa have created OMOP versions of local health data.
- The LAISDAR project located at the Rwanda Biomedical Center contains 3.6 million unique subjects in OMOP CDMs transformed from OpenMRS and OpenClinic EMRs at 15 hospitals.
- The INSPIRE network at the African Population Health and Research Centre (APHRC) carried out ETLs to the OMOP CDM using data from the Health and Demographic Surveillance System in Kenya, Tanzania and South Africa.
- APHRC is collaborating with UK institutions The Alan Turing Institute and London School of Hygiene and Tropical Medicine, CODATA (France), I-DAIR (Switzerland) and institutions in Cameroon, Ethiopia and Senegal on a Wellcome Trust funded project "Data Science Without Borders", which will conduct research using data harmonized to the OMOP CDM.
- The Virus Outbreak Data Network (VODAN) Africa has established data science partnerships in 12 African countries and invited OHDSI Africa Chapter members to meet at Leiden University (Belgium) on 04 Jun 2024 to discuss a plan for collaboration.

Conclusion

Awareness of OHDSI is growing in Africa with several African institutions successfully implementing the OMOP CDM and OHDSI tools. Several OHDSI Africa Chapter members are poised to do OMOP CDM implementations at their institutions. Despite the availability of vast amounts of health data in Africa, these remain siloed in different organizations and captured in varying formats and terminologies. Facilitating knowledge transfer from experienced OHDSI members, within Africa and globally, to those less familiar with OHDSI tools, will expedite interoperability and capacity building in Africa. **Funding is urgently needed to empower African scientists to lead this transformative effort.**

OHDSI

Join the OHDSI Africa Chapter biweekly meeting Monday at 10 AM ET

#OHDSISocialShowcase This Week

Friday

Comorbidities among patients with Severe Maternal Morbidity: A comparison of conditions identified through active hospital-based surveillance versus OMOP CDM

(Carrie Wolfson, Benjamin Martin, Khyzer Aziz, Paul Nagy, Andreea Creanga)

Comorbidities among patients with Severe Maternal Morbidity:

A comparison of conditions identified through active hospital-based surveillance versus OMOP CDM

PRESENTER: **Carrie Wolfson**

INTRO

- Patient characteristics, especially presence of comorbid conditions, strongly affect the risk of SMM.
- This analyses compares prevalence of comorbidities and risk factors for SMM identified through surveillance, based on manual chart abstraction, to those from a cohort characterization exercise in the OMOP common data model

METHODS

1. Data sources: Maryland Maternal Health Innovation Program (MDMOM)'s SMM facility-based surveillance; EHR data structured using the OMOP CDM in the Johns Hopkins Health System
2. We computed the incidence of 24 comorbidities and pregnancy risk factors identified using both sources
3. Conditions and risk factors with <20% difference in prevalence between the 2 methods=aligned, 20-50% different=moderately aligned, >50% different=not aligned.

RESULTS:

Prevalence of comorbidities and risk factors identified among patients with SMM

	OMOP (n=1,014)	SMM Surveillance (n=205)	Alignment
Mental health conditions	35.8	36.6	Close
Anxiety	22.1	22.0	Close
Depression	15.2	20.0	Moderate
Bipolar disorder	2.5	2.0	Moderate
Obesity	32.6	39.0	Close
Prior cesarean	22.8	30.7	Close
HDP	18.8	21.5	Close
Asthma	17.2	17.1	Close
Chronic hypertension	14.7	17.1	Close
Hypothyroidism	6.7	5.9	Close
Sexually transmitted infections	6.7	5.9	Close
Fibroids	6.5	6.8	Close
Sickle cell	4.9	4.9	Close
Anemia	29.7	22.0	Moderate
Gestational diabetes	12.0	8.8	Moderate
Substance use	9.9	16.1	Moderate
Preexisting diabetes	5.0	7.8	Moderate
Twins or higher order	4.7	3.4	Moderate
Lupus	2.0	1.5	Moderate
Renal conditions	15.3	1.5	Not aligned
Elderly primigravida	10.5	5.9	Not aligned
Prior preterm delivery	4.6	15.1	Not aligned
Placental complication	2.1	25.4	Not aligned
Cardiovascular disease	1.2	10.7	Not aligned

Prevalence of **comorbidities** and **pregnancy risk factors** identified using **manual chart abstraction vs. OMOP-CDM** were in close or moderate alignment for a majority of SMM cases

Take a picture to download the full abstract

MDMOM's SMM facility-based surveillance

- Trained clinician abstractors in each hospital identified all cases that met SMM surveillance definition
- Abstractors reviewed EHR to document information about the patient and SMM event using a standardized electronic REDCap form
- Surveillance data captures data between July 2020 and December 2023

JHM OMOP

- EHR data includes records from patients with live birth deliveries in the Johns Hopkins Health System between July 2016 -May 2024
- SMM events are identified using the CDC algorithm of 21 indicator corresponding to ICD10-CM codes during delivery hospitalization applied to the JHM OMOP instance

Carrie Wolfson, Benjamin Martin, Khyzer Aziz, Paul Nagy, Andreea Creanga

JOHNS HOPKINS UNIVERSITY

OHDSI

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?

The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

Links are sent out weekly and available at:
ohdsi.org/community-calls-2025