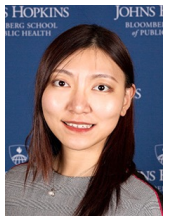
Workgroup Updates

OHDSI Community Call
Nov. 29, 2022 • 11 am ET

Upcoming OHDSI Community Calls

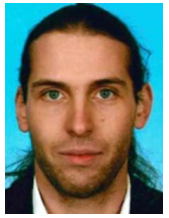
Date	Topic
Dec. 6	Fall Publications
Dec. 13	How Did OHDSI Do In 2022?
Dec. 20	Holiday-Themed Final Call of 2022

Dec. 6: Recent OHDSI Publications



Comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with different covid-19 vaccines: international network cohort study from five European countries and the US

Xintong Li PhD student, University of Oxford



Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond

Václav Papež Research Associate, UCL Institute of Health Informatics

Adjusting for indirectly measured confounding using large-scale propensity score

Linying Zhang PhD student, Columbia University

PheValuator 2.0: Methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation

Joel Swerdel Associate Director of Epidemiology Analytics, Janssen Research and Development

Integrating real-world data from Brazil and Pakistan into the OMOP common data model and standardized health analytics framework to characterize COVID-19 in the Global South

Sara Khalid Research Associate, UCL Institute of Health Informatics

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Nigel Hughes, Peter Rijnbeek, Kees van Bochove, Talita Duarte-Salles, Carl Steinbeisser, David Vizcaya, Dani Prieto-Alhambra, and Patrick Ryan** on the publication of **Evaluating a novel approach to stimulate open science collaborations: a case series of "study-a-thon" events within the OHDSI and European IMI communities** in JAMIA.

JAMIA Open, 5(4), 2022, 1–9
https://doi.org/10.1093/jamiaopen/ooac100
Research and Applications

Research and Applications

Evaluating a novel approach to stimulate open science collaborations: a case series of "study-a-thon" events within the OHDSI and European IMI communities

N. Hughes[1], P.R. Rijnbeek[2], K. van Bochove[3], T. Duarte-Salles[4], C. Steinbeisser[5], D. Vizcaya[6], D. Prieto-Alhambra[7], and P. Ryan[8]

[1]Epidemiology, Janssen R&D, Beerse, Belgium, [2]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands, [3]The Hyve, Utrecht, The Netherlands, [4]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain, [5]Steinbeisser Project Management, Munich, Germany, [6]Bayer Pharmaceuticals, Sant Joan Despi, Spain, [7]NDORMS, University of Oxford, Oxford, UK, [8]Epidemiology, Janssen R&D, Titusville, New Jersey, USA

Corresponding Author: Nigel Hughes, MSc, Janssen R&D, Turnhoutseweg 30, 2340 Beerse, Belgium; nhughes@its.jnj.com

Received 1 May 2022; Revised 25 October 2022; Editorial Decision 29 October 2022; Accepted 2 November 2022

ABSTRACT

Objective: We introduce and review the concept of a study-a-thon as a catalyst for open science in medicine, utilizing harmonized real world, observation health data, tools, skills, and methods to conduct network studies, generating insights for those wishing to use study-a-thons for future research.

Materials and Methods: A series of historical study-a-thons since 2017 to present were reviewed for thematic insights as to the opportunity to accelerate the research method to conduct studies across therapeutic areas. Review of publications and experience of the authors generated insights to illustrate the conduct of study-a-thons, key learning, and direction for those wishing to conduct future such study-a-thons.

Results: A review of six study-a-thons have provided insights into their scientific impact, and 13 areas of insights for those wishing to conduct future study-a-thons. Defining aspects of the study-a-thon method for rapid, collaborative research through network studies reinforce the need to clear scientific rationale, tools, skills, and methods being collaboratively to conduct a focused study. Well-characterized preparatory, execution and postevent phases, coalescing skills, experience, data, clinical input (ensuring representative clinical context to the research query), and well-defined, logical steps in conducting research via the study-a-thon method are critical.

Conclusions: A study-a-thon is a focused multiday research event generating reliable evidence on a specific medical topic across different countries and health systems. In a study-a-thon, a multidisciplinary team collaborate to create an accelerated contribution to scientific evidence and clinical practice. It critically accelerates the research process, without inhibiting the quality of the research output and evidence generation, through a reproducible process.

OHDSI Shoutouts!

Any shoutouts from the community? Please share and help promote and celebrate OHDSI work!

Have a study published? Please send to sachson@ohdsi.org so we can share during this call and on our social channels.
Let's work together to promote the collaborative work happening in 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
Wednesday	7 am	Medical Imaging
Wednesday	12 pm	FHIR and OMOP Terminologies Subgroup (ZOOM)
Thursday	10 am	Data Quality Dashboard
Thursday	12 pm	Population-Level Estimation
Thursday	1 pm	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	7 pm	Dentistry
Friday	9 am	Education
Friday	9 am	GIS – Geographic Information System
Monday	10 am	Healthcare Systems Interest Group
Tuesday	10 am	Common Data Model

ohdsi.org/upcoming-working-group-calls/

DARWIN EU® Welcomes First Data Partners

The EMA has selected the first 10 data partners to collaborate with DARWIN EU®, the Data Analysis and Real-World Interrogation Network. The data available to these partners will be used for studies to generate real-world evidence that will support scientific evaluations and regulatory decision making, and all have already been mapped to the OMOP CDM.

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Medicines | Human regulatory | Veterinary regulatory | Committees | News & events | Partners &

DARWIN EU® welcomes first data partners Share

News 23/11/2022

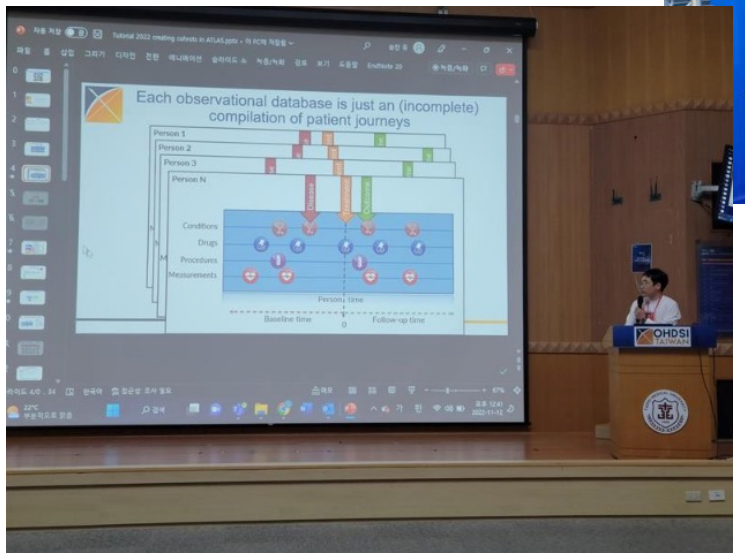
EMA has selected the first set of data partners to collaborate with DARWIN EU®, the Data Analysis and Real-World Interrogation Network. The data available to these partners will be used for studies to generate **real-world evidence** that will support scientific evaluations and regulatory decision making. Real-world evidence refers to information derived from analysis of real-world data, which is routinely collected data about a patient's health status or delivery of healthcare from a variety of sources other than traditional clinical trials.

The selected partners include both public and private institutions. The common feature is that they all have access to **real-world healthcare data** from one or more sources such as hospitals, primary care, health insurance, biobanks, or disease-specific patient registries. The data partners will provide the DARWIN EU® Coordination Centre with results of analyses of these data.

Next APAC Community Call

The next Asia-Pacific (APAC) community call will be Dec. 1 (Nov. 30 in the Western Hemisphere) and will provide a recap of the APAC Symposium.

The link to join these bi-weekly calls is on the APAC Community Calls page on OHDSI.org.



MEDINFO Deadline is Dec. 1

Call for speakers ends Thursday 1 December @ 11:59pm AEDT

Not long to go until our call for speakers for #MEDINFO23 closes on 1 December 2022 at 11:59pm AEDT!

If you have a new project, experience or innovation in digital health that you would like to share with a global community, then make sure you submit a paper for MedInfo 2023 by then.

The 19th edition of MedInfo is in the Land Down Under in Sydney, Australia, from 8-12 July 2023. Hosted by AIDH and IMIA, the conference brings together thousands of digital health leaders and practitioners at the forefront of healthcare.

Don't delay and make a submission today!

MORE INFO & APPLY

Join Anna Ostropolets' Dissertation Defense

OHDSI veteran and 2018 Titan Award winner **Anna Ostropolets** will defend her Columbia University dissertation Wed., Nov. 30, on **Generating Reliable and Responsive Observational Evidence: Reducing Pre-analysis Bias**. The open session will be at 10 am ET on Zoom.

Wednesday, Nov. 30, 10 am ET

OMOP CDM ERD Challenge

Patrick Ryan shared a recent forum post called “Introducing the OMOP CDM ER Diagram Challenge” and is calling for community submissions by Tuesday, Dec. 13.

The winner will be announced at the Dec. 20 community call!

Introducing the OMOP CDM ER diagram challenge! Submissions due 13Dec2022

General

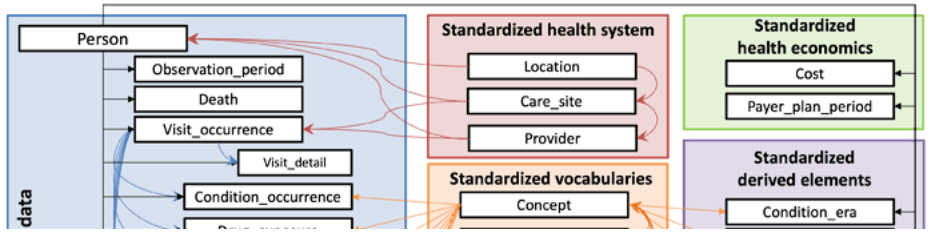
Patrick_Ryan 1 3d

Friends:

I'd like to announce the OMOP Common Data Model (CDM) Entity-Relationship Diagram (ERD) Challenge! Starting today, all members of the OHDSI community are welcome and encouraged to submit their entries of the best ERD for the OMOP CDM to this forum post (or to the CDM Workgroup teams site) by Tuesday, Dec 13. One winner will be selected by a committee from the CDM workgroup, and announced on OHDSI's last community call of the year on Dec20, with their award-winning ERD being a gift to our entire community, posted on the official OMOP CDM git page, but also the winner receiving a special gift from the OHDSI community! This should be a fun activity for our community, particularly those of you helping drive our open community data standards, to learn and collaborate and contribute to a community resource that all of us can benefit in. So, please accept the OMOP CDM ERD Challenge and get diagraming!

Background:

The OMOP Common Data Model v5.4 serves our community well as an open community data standard to enable standardized analytics and large-scale evidence generation. The CDM Workgroup, under the amazing leadership of @clairblacketer , has done a tremendous job of stewarding this standard, improving our documentations and conventions, and providing reference implementations, all available here at: index.knit 3 . Our primary schematic that we use to describe the OMOP CDM is below:



#OHDSISocialShowcase This Week

Machine Learning to Predict the Ischemic Stroke among Type 2 Diabetes Mellitus Patients using Taipei Medical University Clinical Research Database

PHAN THANH PHUC [1], PHUNG ANH NGUYEN [2] , JASON C. HSU [1,2,*]
[1] International PhD Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan;
[2] Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan;

Background

Ischemic stroke has been recognized as a clinically important complication of type 2 diabetes (T2DM) patients. Risk prediction models for DM complications/comorbidities have substantial capacity to support the decision-making process regarding the patient's clinical management.
This study aims to develop machine learning algorithms to predict the risk of ischemic stroke among T2DM patients using various predictors such as patients' characteristics, disease history, laboratory tests, and medication.

Methods

1. Data source and study population
The dataset was collected from the Taipei Medical University Clinical Research Database (TMUCRD) in this study. Index 2008 data as wash-out-period—newly diagnosed T2DM patients from 2009 to 2019 as our cohort study with [ICD9-CM] codes 250.xx, and [ICD10-CM] codes E11 .xx.
2. Outcome
All patients were monitored from the date of taking antidiabetic drugs to the date the patients were admitted to hospitals with ischemic stroke (ICD9-CM codes 433, 434, 436, and ICD10-CM codes I60, I61, I62) during a one-year follow-up.
3. Features
The features were collected, including (i) patient characteristics (i.e., age, sex), (ii) comorbidities (i.e., any diagnoses before the date of taking antidiabetic drugs), (iii) other medication uses, and (iv) laboratory exams (i.e., Glucose, HbA1C, etc.)..
4. Statistical analysis and Model development
The training set, containing the data of Taipei Medical University Hospital and Wang Fang Hospital. The testing set, including the data of Shuang Ho Hospital, was used to validate the models. The stratified 10-fold cross-validate was applied in the training set to assess different machine learning models' performance and general errors.
Machine learning techniques, such as Logistic Regression (LR), Linear Discriminant Analysis (DT), Gradient Boosting Machine (GBM) and Random Forest (RF), to develop the prediction models. The performance of the algorithms was measured by Area Under the Curve (AUC), sensitivity, specificity, and F1-score.

Results

Table 1. Patient baseline and characteristic

Feature	Overall (n=4,697)	Training cohort (n=4,697)	Testing cohort (n=4,582)
Ischemic stroke, No. (%), patient	217 (2.34)	89 (1.9%)	128 (2.8%)
Age, Mean (SD), y	61.8 (4)	60.7 (15.0)	62.9 (13.0)
Gender, No. (%)			
Female	4,305 (46.4)	2,182 (46.5%)	2,123 (46.3%)
Male	4,974 (53.6)	2,515 (53.5%)	2,459 (53.7%)
Visit per patient, Mean (SD)	217 (2.34)	24.1 (0.34)	23.3 (0.35)
Laboratory test, Mean (SD)	6.3 (0.08)	6.03 (0.07)	6.7 (0.09)
Medication, Mean (SD)	25 (0.2)	23 (0.2)	27 (0.2)

Table 2. Model performance evaluation

Model	AUC (CV)	AUC (Testing)	Sensitivity	Specificity	F1-score
Logistic Regression	0.88	0.85	0.819	0.748	0.16
LDA	0.88	0.85	0.721	0.802	0.161
GBM	0.91	0.85	0.744	0.813	0.164
Random Forest	0.93	0.84	0.811	0.692	0.133

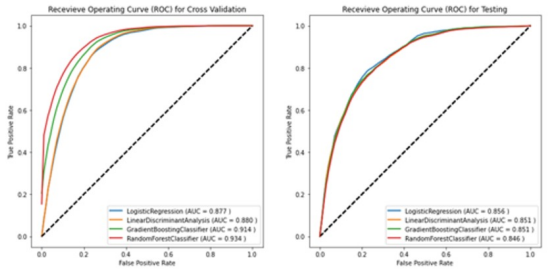


Figure 1. Receiver Operating Characteristic (ROC) Curve to evaluate the model performance

Conclusions

We successfully developed machine learning models to predict the risk of ischemic stroke among T2DM. Our model performance improved from Random Forest to Gradient Boosting Machine. The top three important features executed from our best model are antiplatelet agent, age, and prior stroke.
The strong association of diabetes with stroke has long been appreciated. To the best of our knowledge, there are limited studies in classifying and predicting ischemic stroke in the T2DM cohort by developing machine learning-based models. Therefore, our findings were essential to improve the accuracy of early detection, diagnosis, and prognosis of ischemic stroke to manage the risk of diabetes complications.

Contact: * **Jason C. Hsu**, International Ph.D. program in Biotech and Health Management, College of Management, Taipei Medical University, Taipei, Taiwan; 11F., No.172-1, Sec. 2, Keelung Rd., Daan Dist., Taipei City 106, Taiwan (R.O.C.);
E-mail: jasohsu@tmu.edu.tw

MONDAY

Machine Learning to Predict the Ischemic Stroke among Type 2 Diabetes Mellitus Patients using Taipei Medical University Clinical Research Database

(Phan Thanh Phuc, Phung Anh Nguyen, Jason C. Hsu)

#OHDSISocialShowcase This Week

Examining the Differences in Baseline Characteristics of One-code and Two-code Phenotype Algorithms

Presenters: Pranav Bhimani, Raechel Davis

BACKGROUND

- The guidance and implications regarding broad and narrow phenotype algorithm (PA) use remain unclear.
- Broad PAs requiring one diagnostic code identify a greater number of subjects, often producing higher sensitivities, albeit with lower positive predictive values (PPVs).
- Narrow PAs that require a second diagnostic code during some timeframe after the first diagnostic code, are often accompanied by lower sensitivities but produce higher PPVs.
- The objective of this study was to compare the similarity of baseline characteristics for phenotype algorithms requiring one and two diagnostic codes for health outcomes in therapeutic areas of neurology, immunology, oncology, and cardiology using six real world databases.

METHODS

- A network of six US observational databases that were transformed to the Observational Medical Outcomes Partnership (OMOP) Common Data Model version 5.3.1 were used [1].

Database	Years	Number of Persons (millions)	Median Follow-up (years)
CCAE	2000-2021	162	1.56
MDCD	2006-2020	33	1.52
MDCR	2000-2021	10	2.46
Optum DOD	2007-2021	92	1.48
Optum EHR	2007-2021	105	2.63
Pharmetrics	2013-2021	162	3.25

- PAs for associated outcomes within each therapeutic area were analyzed.

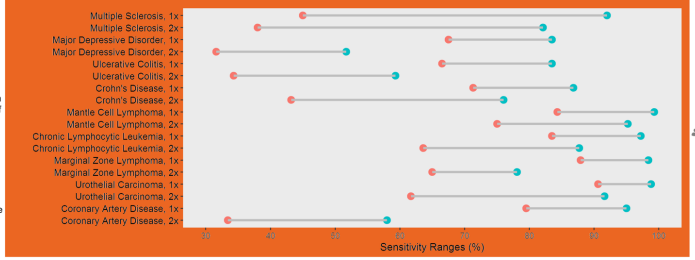
Therapeutic Area	Condition
Neurology	Multiple Sclerosis
	Major Depressive Disorder
Immunology	Ulcerative Colitis
	Crohn's Disease
Oncology	Mantle Cell Lymphoma
	Chronic Lymphocytic Leukemia
	Marginal Zone Lymphoma
	Urothelial Carcinoma
Cardiology	Coronary Artery Disease

- The ATLAS tool was used to create PAs and generate cohorts [2].
- The PheValuator method provided performance characteristics, i.e., sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), associated with each PA [3].
- The Cohort Diagnostics tool allowed for evaluation of PAs at a cohort-level, providing a comparison of baseline covariates between the one-code and two-code algorithms [4].
- SMD > 0.1 is used as an ad-hoc heuristic for what constituted a substantial difference in compared covariates [5].
- Baseline characteristics were compared by generating a proportion. The numerator is the number of covariates with an absolute standardized mean difference (SMD) > 0.1 and the denominator is the total number of covariates compared between phenotypes.

Figure 1. Variability in Baseline Characteristics Between One-code and Two-code Phenotype Algorithms

Comparing baseline covariates between broad and narrow phenotype algorithms provides a more complete understanding of algorithm differences.

Figure 2. Ranges of Sensitivities Across all Databases for Broad (One-code) and Narrow (Two-code) Phenotype Algorithms, red = lowest; blue = highest



RESULTS

- In six of the nine outcomes there was minimal variability in the comparison of baseline covariates.
- Comparisons between the Ulcerative Colitis and Crohn's Disease one-code and two-code PAs showed the greatest variability, while the Multiple Sclerosis comparison showed a moderate level of variability (Figure 1).
- Sensitivities of one-code algorithms were observed to be higher and less variable across databases than two-code algorithms (Figure 2).

View Results Interactively:
https://data.ohdsi.org/PhenotypeComparisons/

CONCLUSIONS

- Comparisons of baseline characteristics in 1-code and 2-code PAs in most (6 of the 9) outcomes showed minimal variability.
- For outcomes in specific therapeutic areas such as immunology, greater variability in baseline covariates may be present when comparing 1 and 2 code algorithms.
- Comparison of the similarity of baseline covariates between phenotype algorithms provides a more complete understanding of algorithm differences.

REFERENCES

1. Voss EA MR, Matcho A, Ma Q, Knoll C, Schuemie M, et al. . Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. . J Am Med Inform Assoc 2015;22(3):553-64.
2. Schuemie, M. and DeFalco F. (2019). OHDSI Analytics Tools. In Observational Health Data Sciences and Informatics The Book of OHDSI (pp. 109-123). Independently published.
3. Swerdel JN, Hripcsak G, Ryan PB. PheValuator: Development and evaluation of a phenotype algorithm evaluator. Journal of Biomedical Informatics. 2019;97:103258.
4. Rao G, Schuemie M, Ryan P, Weaver J, Gilbert J (2022). CohortDiagnostics: Diagnostics for OHDSI Cohorts. https://ohdsi.github.io/CohortDiagnostics, https://github.com/OHDSI/CohortDiagnostics.
5. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol. 2015;68(6):627-36.

Jill Hardin[1,2], Pranav Bhimani[1,2,3], Raechel Davis[1,2,4], Joel N. Swerdel[1,2]
[1]Janssen Research and Development, Titusville, NJ, USA; [2]Observational Health Data Sciences and Informatics (OHDSI), New York, NY; [3]Weill Cornell Graduate School, New York, NY; [4]Yale School of Public Health, New Haven, CT

TUESDAY

Examining Differences in Baseline Characteristics of Broad and Narrow Phenotype Algorithms (Jill Hardin, Pranav Bhimani, Raechel Davis, Joel Swerdel)

#OHDSISocialShowcase This Week

OMOP and FHIR Data Comparison

Spencer SooHoo[1], Andrey Soares[2], Rohith Mohan[1], Renier Estiandan[1], Ryan Hoffman[1], Shao Chi Huang[1], Brian Tep[1], David Kreda[3], Dan Gottlieb[3], Aaron Boussina[4], Paul Kingsbury[5], Lisa Schilling[2]

[1]Cedars-Sinai Medical Center
[2]University of Colorado Anschutz Medical Campus
[3]Harvard Medical School
[4]University of California, San Diego
[5]University of Southern California

Introduction

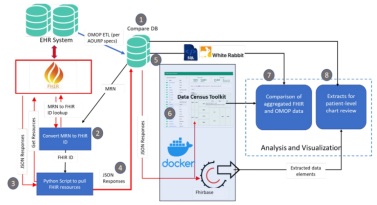
The All of Us Research Program (AoURP) has two methods for submitting EHR data. Health Provider Organizations (HPO) can submit participant data in file transfers modeled to the OMOP Common Data Model (CDM) or willing participants can directly submit their EHR data via the AoURP Sync4Science FHIR app.

Phase 1 of this project aimed to identify high-level differences between the OMOP and FHIR data submissions for key EHR domains.

Methods

- Cohort: OMOP and FHIR dataset for 5000 randomly selected patients who met the eligibility criteria: Age >= 18 years old on 1/1/2017 with at least 5 visits/encounters face-to-face or telehealth between 1/1/2018 and 9/1/2021. No "Break the Glass", "Confidential", or "No Research"
- Data Sources: OMOP data extracted from Epic's Clarity and Caboodle, and Cerner. FHIR data extracted from the institution's FHIR Endpoint API
- Tools: SQL scripts and White Rabbit for OMOP data and FHIR Data Census Toolkit, FHIRBase and Python scripts for parsing JSON resources for FHIR data
- Approaches: High-level characterization of the data models, comparing counts per cohort in specifically aligned OMOP categories and parameterized FHIR Resources, matching LOINC codes at patient level for labs, and manual chart review

High-level Process

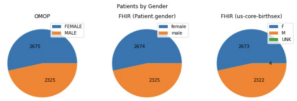


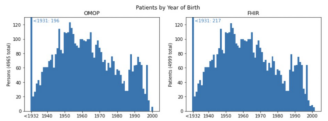
1. Generate an OMOP data mart (Compare DB) of 5000 randomly selected patients using the AoURP OMOP CDM ETL specifications
2. Use the MRNs from the patients as input for a script that makes FHIR requests to retrieve the FHIR ID for the patients
3. Use a Python script to get FHIR Resources and receive the JSON responses from the FHIR endpoint.
4. Store the JSON responses into NDJSON files
5. Load the NDJSON to the Compare DB
6. Use the Data Census Toolkit to produce high-level characterization of the FHIR data
7. Use SQL scripts and tools like OHDSI White Rabbit to generate high-level characterization for the OMOP data
8. Use FHIRBase to parse out discrete elements from the JSON responses that can be used for patient-level data comparisons

Results

Cedars-Sinai Medical Center

- Close agreement between the OMOP and FHIR demographic data at the cohort-level
- Differences between the OMOP Procedure_Occurrence and FHIR Procedure Resource
- OMOP pulls relevant encounter data while FHIR misses many encounter types and misclassifies many of the existing encounters
- Medications administered (as Procedure), were found in the OMOP data but not in FHIR
- FHIR Condition Resource entry included 3 pieces of information: a SNOMED-CT code indicating the condition, a LOINC code indicating condition status (active/inactive), and an ICD-10 code indicating the condition





OMOP Domain	OMOP Counts	FHIR Resource	FHIR Counts
Person	5,000	Patients	5,000
Condition_Occurrence	1,663,603	Conditions	36,196
Procedure_Occurrence	510,140	Procedures	37,649
Drug_Exposure	421,978	Medications	412,604
		Medication Request	412,604
Measurement	1,846,126	Observation (Lab)	2,074,025
Visit_Occurrence	668,529	Encounters	357,051
Death	107	Patients(deceased)	88

OMOP	FHIR	Chart Review
56 records including: • PICC line installation • US-guided needle biopsy • Wrist X-Ray • 22 evaluation and management records • Missing: repair of laceration/suture	4 records: • 2 internal codes for suture repair of laceration and PICC line installation • CPT for suture repair of laceration • 2 CPT for PICC line installation • CPT code for US-guided needle biopsy	• Confirmed suture repair • Confirmed ED visit evaluation • Confirmed wrist X-ray • Confirmed US-guided needle biopsy

University of Colorado-Anschutz Medical Campus

- OMOP Drug_Exposure concept types can be mapped to different FHIR Medication-related Resources
- When limited by type_concept_id, Drug_Exposure to MedicationRequest and Condition_Occurrence to Condition counts presented fairly close numbers
- MedicationRequest Resources include prescribed, hospital and clinic administered, and patient reported medications
- FHIR data provides more information about the immunization such as manufacturer, lot number, expiration date, site and performer

drug_type_concept_id	drug_type_concept_name	FHIR Resource Mapping
32818	EHR administration record	MedicationAdministration Immunization
32833	EHR order	MedicationRequest
32865	Patient self-report	MedicationStatement MedicationRequest (reported Boolean=true)
32825	EHR dispensing record	MedicationDispense
32821	EHR billing record	
32869	Pharmacy claim	
32817	EHR	

OMOP Domain	OMOP Counts	FHIR Resource	FHIR Counts
Person	5,000	Patient	5,000
Condition_Occurrence	749,101	Condition (FHIR Problem List only)	48,383
– EHR Problem List only	53,454	Observation (category=social-history)	8,560
		AllergyIntolerance	9,119
Procedure_Occurrence	725,043	Procedure	357,387
		Immunization	32,034
Drug_Exposure	823,269	Medication	403,589
– EHR Orders only	342,282	MedicationRequest	416,589
– Patient Reported only	45,205		
– Immunization (CVX)	9,723	Immunization	32,034
Measurement	4,949,581	Observation (category=laboratory)	1,737,419
		Observation (category=vital-signs)	1,300,767
Visit_Occurrence	197,293	Encounter (and type=finished)	401,269
(completed visits)			
Death	172	Patient (deceased)	-

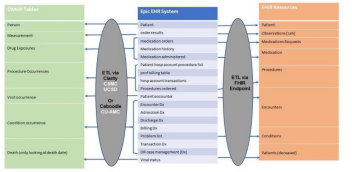
University of California, San Diego

- Patient demographics were generally consistent between OMOP and FHIR
- FHIR captured the deceased status, but OMOP ETL was omitting patient death when the death date was not recorded
- The overall counts of procedures and drug exposures differed substantially. FHIR API only contained items from the problem list and the encounter diagnoses
- Values from the OMOP Measurement table were overall consistent with those observed from the FHIR Observation Resource
- The overall count of encounters from FHIR vs the OMOP Visit_Occurrence table were also relatively close

OMOP Domain	OMOP Counts	FHIR Resource	FHIR Counts
Person	5,000	Patients	5,000
Condition_Occurrence	572,378	Conditions	71,462
		Conditions (problem list)	12,402
		Conditions (Encounter Diagnosis)	59,060
Procedure_Occurrence	378,598	Procedures	10,419
Drug_Exposure	699,866	Medications	
		Medication Requests	126,282
Measurement	1,403,686	Observation (Lab)	598,408
		Observation (Vital-Signs)	792,347
		Observation (Social History)	7,537
Visit_Occurrence	161,660	Encounters	112,304
Death	120	Patients(deceased)	128

Discussion & Conclusion

- Comparison of aggregate record counts for OMOP and FHIR initially appeared dissimilar.
- FHIR and OMOP data are not necessarily a one-to-one mapping
- Site-specific differences: 1) OMOP ETL conventions, 2) the parameters used for the FHIR requests, and 3) current FHIR conventions and site implementations
- Inability to perform identical comparisons across all sites
- Targeted chart review was invaluable in illuminating some of the differences in alignment between the OMOP and FHIR aggregate counts



We would like to thank the All of Us Research Program and Scripps Research for funding this project (grant 5U24OD23176) to characterize and compare data extracts from OMOP and FHIR.

WEDNESDAY

OMOP and FHIR Data Comparison (Spencer SooHoo, Andrey Soares, Rohith Mohan, Renier Estiandan, Ryan Hoffman, Shao Chi Huang, Brian Tep, David Kreda, Dan Gottlieb, Aaron Boussina, Paul Kingsbury, Lisa Schilling)

#OHDSISocialShowcase This Week

Analyzing the Use of Beers Criteria Guidelines for Older Adults through ATLAS Operationalization

Jacob P. Lombardi[1], Rohit P. Marwah[1], Krishi T. Akenapalli[1], Richard D. Boyce[2], Jonathan M. Raviotta[3], Sandra L. Kane-Gill[1], Steven M. Albert[4]

[1]School of Pharmacy, University of Pittsburgh; [2]Department of Biomedical Informatics, University of Pittsburgh; [3]School of Medicine, University of Pittsburgh; [4]Department of Behavioral and Community Health Sciences, Graduate School of Public Health, University of Pittsburgh

Background

- Overprescribing and misprescribing pose a great risk to geriatric patients and are a major focus geriatric pharmacy management
- The Beers criteria outlines medications that have potential to create overprescribing or misprescribing scenarios for older adults

Objective

- This project seeks to develop an operationalized form of the Beers criteria in ATLAS that shows the incidence of potentially inappropriate medications (PIMs) and acceptable uses according to the Beers Criteria in geriatric patient populations whose data is represented in the OHDSI common data model

Methods

- Our dataset contains outpatient drug claims from University of Pittsburgh Medical Center (UPMC) from 2015 to 2018
  - This dataset consists of 89,600 PACE enrollees and 557,000 non-PACE enrollees
  - PACE/PACENET is a PA prescription subsidy program with ~250,000 enrollees designed to protect enrollees from overutilization and misutilization of medications
  - In this dataset, those not enrolled in PACE received no prescription subsidy or overutilization review of their medication history

Process for Creating Cohort Definitions for Beers Criteria

- 2019 AGS Beers Criteria Guidelines[1] → List of medications that pose potential risk for geriatric patients
- Concept Set Development → Created concept sets in Atlas for medication classes, procedures, disease states, and lab values
- Cohort Definition Design → Characterized groups that are defined in the Beers Criteria as being at risk for adverse medication use
- Outcome: 39 cohort definitions mirroring 2019 Beers Guidelines

1. 2019 Updated AGS Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2019 Apr;67(4):674-694.

Results

- Our complete set of cohorts contained 39 individual medications or medication classes, split into two types
  - Cohorts with no appropriate use; i.e. no acceptable situation for use in a patient over 65 years old based on the Beers Criteria
  - Cohorts with some appropriate use; i.e. there are certain situations where a patient over 65 years old can appropriately receive the medication(s) according to the Beers Criteria
- All 39 cohorts were collated to compare how effectively PACE has enacted the Beers criteria as compared to patients not enrolled in PACE

	Number of patients receiving a potentially inappropriate medication	Number of total patients	Percentage of patients experiencing a potentially inappropriate medication from 2015-2018
PACE enrollees	24,463	89,600	27.30%
Non-PACE enrollees	126,748	557,000	22.76%

- Certain medications or medication classes were identified as having a higher incidence of potentially inappropriate use based upon Beers criteria guidelines

Medication / Class Name	Percentage of drug orders classified as PIM in PACE	Percentage of drug orders classified as PIM in non-PACE
Amiodarone	30.32%	34.36%
Antiparksonian Agents	45.87%	34.36%
Digoxin	10.49%	9.63%
Proton-pump inhibitors	19.65%	17.91%
Short-acting insulins	45.25%	42.62%

- Further analyses are needed to compare matched PACE and non-PACE patients, as the PACE population is older with a greater prevalence of comorbidities

Conclusions

- Our operationalization effectively captures the incidence of potentially inappropriate medications and acceptable uses according the the Beers criteria:
  - specific medications or medication classes that require more attention when treating older patients
  - demographic risk factors that leave certain patients more vulnerable to potentially inappropriate medications
  - comparisons between standard and focused geriatric care, providing a baseline for future use
- In the future, this design provides individual practitioners, hospitals, health systems, and even larger entities the ability to assess and inspect the pharmaceutical care provided to their geriatric patients. By producing these trends, adjustments can be made to improve health outcomes and limit medication side effects in older adults.

University of Pittsburgh

OHDSI

THURSDAY Analyzing the Use of Beers Criteria Guidelines through ATLAS Operationalization (Richard Boyce, Steven Albert, Jacob Lombardi, Krishi Akenapalli, Rohit Marwah)

#OHDSISocialShowcase This Week

Cohort Definition Validation in Atlas

Charity Hilton MS[1], Saul Crumpton MS[1], Jon Duke MD, MS[1,2]

[1]Georgia Tech Research Institute, [2]Georgia Institute of Technology

Background

OHDSI Atlas has long been an effective tool for developing rule-based cohort definitions in observational data. In the public version of Atlas, thousands of cohort definitions have been created. While patient record verification is a common method of cohort definition validation, it is not without difficulties, including but not limited to the need for clinical experts to access data, a tool to review all in-cohort patients, a method to gather review data, and a system of tabulation to determine in-cohort (case/no-case) participation or not[1].

Until now, there has not been an Atlas-based system for clinical expert review. For this effort, we introduce the Atlas Cohort Definition Validation tool (ACDV). This tool aims to solve some of the primary concerns around cohort definition validation, while having the chief benefit of being cohesively integrated into the OHDSI Atlas stack. Additionally, the tool allows for creation of more complex validation question sets, beyond the standard case/no-case assessment.

Figure 1: Question Set Creation

Methods

We designed and developed two modules around cohort definition validation. The first (1) allows for validation study creation and management, and the second (2) allows for validation of study questions for clinical reviewers in the Atlas Patient Profile tool.

The ACDV tool introduces a 'Validation' section to Atlas cohort definition creation, which allows for cohort managers to complete a cohort definition validation workflow. This workflow begins by the creation of question set. Question sets in the ACDV tool, shown in Figure 1, allow for common types of questions (including text, radio, checkbox, numbers, and dates). Multiple questions in a question set can be created and a case/no-case distinction can be selected at the question level. After a question set has been created, it can be linked to a cohort definition sample, this creates the validation study.

After a validation study is created, cohort managers can assign patients for review in the Atlas Patient Profile tool to clinical reviewers. Study questions are displayed to clinical reviewers at the patient level in a collapsible sidebar (see Figure 3). The study question set at the patient profile-level can be accessed via the Cohort Definition tool, the Patient Profile tool, or via a customized link. Once reviewers have viewed patient profiles and answered study questions, study results can be viewed by cohort managers in Atlas or exported to CSV (Figure 4).

Results

Primary development efforts of the ACDV tool are complete, and final modifications and integrations to the tool are being prepared for inclusion in an upcoming OHDSI release. We have validated the tool internally with a clinician-informaticist.

Figure 2: Annotation Study Manager View

Figure 3: Profile Level Validation

Conclusions

The Atlas Cohort Definition Validation tool will provide an integrated way for clinical chart reviewers to validate cohorts well beyond the question of cohort inclusion or not.

This tool will support research in the OHDSI community by living firmly within the active OHDSI Atlas ecosystem of tools. Additionally, this tool will continue the OHDSI legacy of open and community-driven tools to advance research in observational health data.

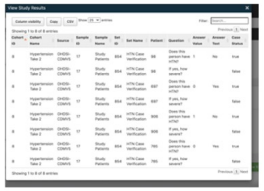


Figure 4: Study Results

Bibliography

1. Observational Health Data Sciences and Informatics. The Book of OHDSI; 2020. Available from: https://ohdsi.github.io/TheBookOfOhdsi/

FRIDAY Cohort Definition Validation in Atlas (Charity Hilton, Saul Crumpton, Jon Duke)

Opening: Northeastern University

Northeastern University invites applications for multiple tenured/ tenure-track faculty positions in support of an Impact Engine centered on large-scale observational health data science and informatics to start in the fall of 2023. These faculty will be core members of our Real-World Healthcare Navigator (RWHN) Impact Engine which aims to change how research is translated into clinical practice by establishing a sustainable service that leads the way in fully reproducing health studies.

Careers

Open Rank Professor of Large Scale Observational Data Science

Apply

Boston, MA (Main Campus)

Full time

Posted 14 Days Ago

R110388

About the Opportunity

About Northeastern

Founded in 1898, Northeastern is a global research university and the recognized leader in experience-driven lifelong learning. Our world-renowned experiential approach empowers our students, faculty, alumni, and partners to create impact far beyond the confines of discipline, degree, and campus.
Our locations—in Boston; Oakland; Arlington, Charlotte, North Carolina; London; Portland, Maine; Oakland; San Francisco; Seattle; Silicon Valley; Toronto; Vancouver; and the Massachusetts communities of Burlington and Nahant—are nodes in our growing global university system. Through this network, we expand opportunities for flexible, student-centered learning and collaborative, solutions-focused research.

Northeastern's comprehensive array of undergraduate and graduate programs— in a variety of on-campus and online formats—lead to degrees through the doctorate in nine colleges and schools. Among these, we offer more than 195 multi-discipline majors and degrees designed to prepare students for purposeful lives and careers.

Responsibilities

Responsibilities will include teaching undergraduate and graduate courses, conducting an independent and externally funded research program, and participating in departmental, college, and university service. Qualified candidates must have expertise in, or a demonstrated commitment to, working with diverse student populations and/or in a culturally diverse work and educational environment.

Qualifications

- PhD or equivalent in Statistics, Bioinformatics, Data Science, Epidemiology, Computer Science, Computer Engineering, or similar field.
- Expertise working with large relational databases (e.g., EHRs, Medicare) preferred.
- Advanced knowledge of analytic approaches including data wrangling, visualization, and machine learning preferred.
- Expertise in either R or Python.

Opening: Northeastern University

The OHDSI Center at the Roux Institute seeks a postdoctoral fellow to join their team focused on developing statistical methods and applying them to observational data from large-scale federated datasets (e.g. electronic health records and administrative claims data), with specific applications to the safety of biologics. This research will directly improve our ability to use real world data to characterize patient populations, construct population level estimates relating exposures to health outcomes, and to enhance clinical decision making through improved patient-level predictions.

Careers

Postdoctoral Research Fellow, Observational Health Data Science and Informatics

Apply

Portland, ME
Full time
Posted 30+ Days Ago
R109484

About the Opportunity

Job Summary:
The Observational Health Data Science and Informatics (OHDSI) Center housed within the Roux Institute at Northeastern University (NU) is a new administrative hub of the largest observational health research community in the world. The OHDSI Center @ the Roux Institute works to advance OHDSI's research mission of improving health by empowering a community that collaboratively generates evidence that promotes better health decisions and better care. As part of a multi-institution team, the OHDSI Center is participating in a project to provide support to the U.S. Food and Drug Administration's Biologics Effectiveness and Safety (BEST) program. The mission of the BEST program is to conduct safety and effectiveness surveillance of biologic products, such as vaccines, tissues, and advanced therapeutics.

Responsibilities:
This fellow will collaborate with a growing OHDSI Center team of faculty, staff, postdocs, and students to develop and test statistical methodologies related to the use of real world data (e.g. electronic health records and administrative claims data) to better analyze observational health studies. Fellows will be expected to publish and present their work in leading journals and conferences, participate in departmental seminars, and meet regularly with the OHDSI Center research team. There are also opportunities to mentor junior team members and provide educational support across the Center. The ideal candidate will have a strong statistical and computational background, experience processing and analyzing large datasets, and outstanding communication skills.

Qualifications:
- Ph.D. or equivalent degree in Biostatistics, Data Science, Statistics, Epidemiology, Computer Science, or related fields
- Research experience in observational health data theory/methods
- Demonstrated experience working with large observational health databases and/or medical claims data
- Advanced experience in statistical programming languages such as R or Python and familiarity with version control (i.e., Git/Github)
- Excellent writing and communication skills

Opening: FDA/CDER

FDA/CDER's Division of Hepatology and Nutrition is seeking a clinician with bioinformatics or biostatistics training to work with the Drug-Induced Liver Injury (DILI) Team to evaluate large datasets of liver-related data, collaborate on the Team's review of drugs with hepatotoxicity signals, and help develop informatics-based processes in DILI evaluation across the Agency.

Contact **Judy Racoosin** at **judith.racoosin@fda.hhs.gov** for information about the application process (that will be through USAJOBS).

Opening: Tufts Medicine

Andrew Williams recently announced two exciting new openings at Tufts Medicine.

1) Senior Project Manager for a multisite multiyear grant standardizing critical care EHR and waveform data. (CHoRUS Bridge2AI)

2) Lead software developer and research data warehouse manager for Tufts Medicine's OMOP instance and related services.

Remote work is possible for both positions.

1. Link for Senior Project Manager position: https://smrtr.io/bBVzh

2. Link for Lead Software Developer and Research Data Warehouse Manager position: https://jobs.smartrecruiters.com/TuftsMedicalCenter1/743999857980631-software-development-lead-res-g-c-ctsi

Andrew's email: awilliams15@tuftsmedicalcenter.org

Openings: Johns Hopkins University

Research Associate (Data Scientist/Statistical Engineer), Johns Hopkins inHealth and Biostatistics Center

- Execute OHDSI studies (e.g. for cohort characterizations and comparative effectiveness) on Johns Hopkins's EHR data to support clinicians;
- Collaborate with statisticians and clinicians to continuously integrate state-of-the-art statistical tools to the inHealth/OHDSI tool stack for deployment;
- Mentor trainees on data science and software development skills;
- Co-teach courses on observational health data analytics and data science skills at School of Medicine and Public Health;
- Facilitate adoption of the inHealth tools among the broader OHDSI community by contributing to OHDSI's Health Analytics Data-to-Evidence Suite.
- https://apply.interfolio.com/114436

Where Are We Going?

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

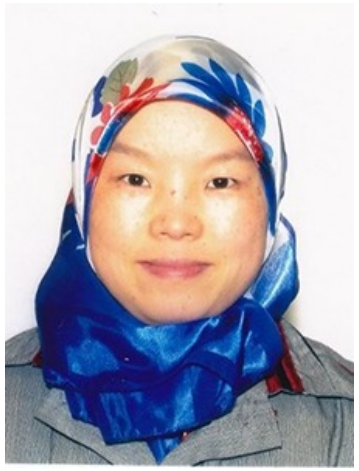
Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

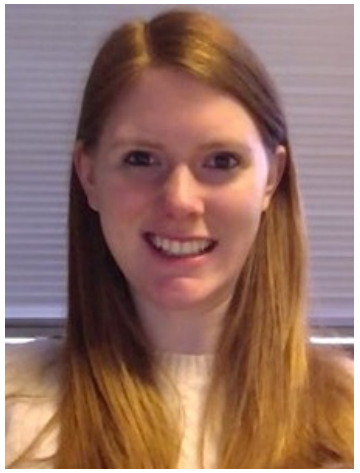
Nov. 29: Workgroup Updates



Medical Devices

Asiyah Lin

Data and technology advancement Scholar, NIH



Patient-Level Prediction

Jenna Reps

Director, Janssen R&D

Clinical Trials

Tom Walpole

Chief Technology Officer, Trials.ai

Psychiatry

Dmitry Dymshyts

Associate Director, Janssen R&D