March Madness & April Olympians

OHDSI Community Call
March 12, 2024 • 11 am ET

Upcoming Community Calls

Date	Topic
Mar. 12	March Madness & April Olympians
Mar. 19	NO MEETING
Mar. 26	Recent OHDSI Publications
coming in April	*CDM Month (hear more about this during today's call)*

March 26: Recent OHDSI Publications

Tathagata Bhattacharjee • University of London

INSPIRE datahub: a pan-African integrated suite of services for harmonising longitudinal population health data using OHDSI tools • *Frontiers in Digital Health*

Sulev Resiberg • University of Tartu

Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned • *JAMIA Open*

Fan Bu • University of Michigan

Bayesian safety surveillance with adaptive bias correction • *Statistics in Medicine*

Jen Wooyeon Park • Johns Hopkins University

Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension • *Journal of Imaging Informatics in Medicine*

Christian Reich • Odysseus

OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization • *JAMIA*

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	Registry
Tuesday	6 pm	Eyecare & Vision Research
Wednesday	9 am	Patient-Level Prediction
Wednesday	12 pm	Health Equity
Wednesday	3 pm	Vulcan/OHDSI
Wednesday	2 pm	Natural Language Processing
Thursday	9:30 am	Network Data Quality
Thursday	12 pm	Medical Devices
Thursday	12 pm	Strategus HADES Subgroup
Thursday	7 pm	Dentistry
Friday	10 am	GIS – Geographic Information System
Friday	10:30 am	Open-Source Community
Friday	11:30 am	Clinical Trials
Friday	11:30 am	Steering Group
Monday	10 pm	Africa Chapter
Monday	11 am	Data Bricks User Group
Monday	2 pm	Electronic Animal Health Records

Evidence Network WG Meeting
14 March 2024 10:30am EST

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Methods for Mapping GIS Data

Kyle Zollo-Venecek

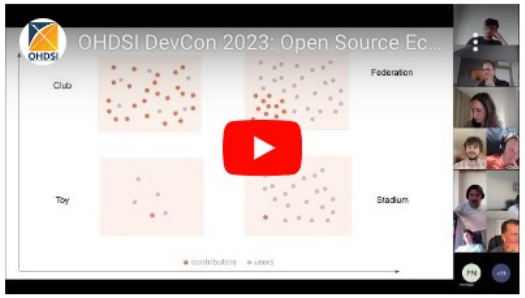
DevCon 2024: April 26, 9 am-3 pm ET

The third annual **OHDSI DevCon** will be held virtually on Friday, April 26, from 9 am-3 pm ET.

Join leaders from our Open-Source Community for a day to both welcome and inform both new and veteran developers within the OHDSI Community.

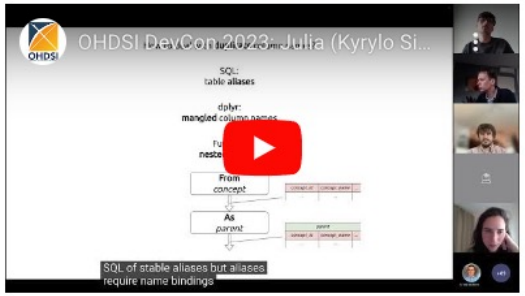
DevCon 2023 Presentations

Open-Source Economics (Adam Black, Clark Evans)



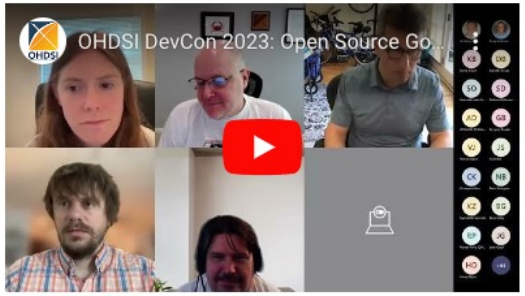
Darwin EU (Ed Burn, Berta Raventós)

Julia (Kyrylo Simonov, Jacob Zelko)



HADES (Anthony Sena, Jenna Reps)

Open-Source Governance (Paul Nagy, Robert Miller, Lee Evans)



Kheiron Cohort Testimonial (Katy Sadowski)

OHDSI Global Symposium

The **2024 OHDSI Global Symposium** will be held Oct. 22-24 at the Hyatt Regency Hotel in New Brunswick, NJ.

Tentative symposium format:

Oct. 22 – tutorials

Oct. 23 – plenaries, collaborator showcase

Oct. 24 – workgroup activities

OHDSI Europe Symposium

Registration is now OPEN for the **2024 OHDSI Europe Symposium**, which will be held June 1-3 in Rotterdam, Netherlands.

June 1 – tutorial/workshop
June 2 – tutorial/workshop
June 3 – main conference

ohdsi-europe.org

ETL-Synthea Package v2.0.0 Release

v2.0.0 Latest

Compare

burrowse released this 3 days ago · v2.0.0 · 583fed8

This update includes the following:

- Update of the synthea package to support synthea 3.1.0 and 3.20
- Location and caresite ETL support
- Update to the package to split intermediate steps and event table loading **@katy-sadowski** **@NACHC-CAD**
- Updates to export to SQLite to support OMOP v5.3 and v5.4 **@mccullen**
- Update to github.io documentation
- Improved version support for CreateCDMIndexAndConstraintScripts, backupCDM, createPrunedTables, getEventConceptId, pruneCDM, restoreCDMTables scripts
- Resolution of formatting and lintr issues
- Removal of redundant joins from allvisittable query
- Addition of the synthea version to the source description for debugging

Contributors

mccullen, katy-sadowski, and NACHC-CAD

Acute ST-Elevation Myocardial Infarction (STEMI) Network Study

OHDSI **Join Our CVD OHDSI Network Study on Acute STEMI!**

General networkstudy

atifadam 6d

Hello OHDSI Community,

We (@mirza_khan @mbrand @atifadam) are thrilled to announce an exciting opportunity for collaboration in new network study focusing on Acute ST-Elevation Myocardial Infarction (STEMI). This study promises to deepen our understanding of STEMI patients' characteristics and identify incidence rates across multiple real-world data datasets.

Why This Study Matters:

- Acute myocardial infarction is a leading cause of hospital admission in the U.S., with STEMI being a critical subtype requiring immediate attention.
- Accurate, scalable, and generalizable identification and characterization of STEMI in multi-country real-world data has numerous benefits, including informing resource allocation, promoting use of effective therapies and interventions, and improving cardiovascular health.

Study Highlights:

- A cohort study using administrative claims or EHR data mapped to the OMOP CDM across the OHDSI network.
- Aims to understand patient characteristics and incidence rates of acute STEMI.
- Utilizes various standardized analytics in the OHDSI community, including the Strategus pipeline and HADES library.

We Need Your Participation!

- Your expertise and data can significantly contribute to this large-scale study.
- Collaboration is key to achieving a comprehensive and diverse understanding of STEMI across different populations and healthcare systems.

Mar 5

1 / 2

Mar 6

6d ago

Webinar: Jackalope Plus (March 14)

SciForce invite you to join to a free webinar

JACKALOPE PLUS:

The Power of ML for Healthcare Data Mapping & Management

SPEAKER:

Denys Kaduk (MD, Data Scientist at Medical Team, SciForce)

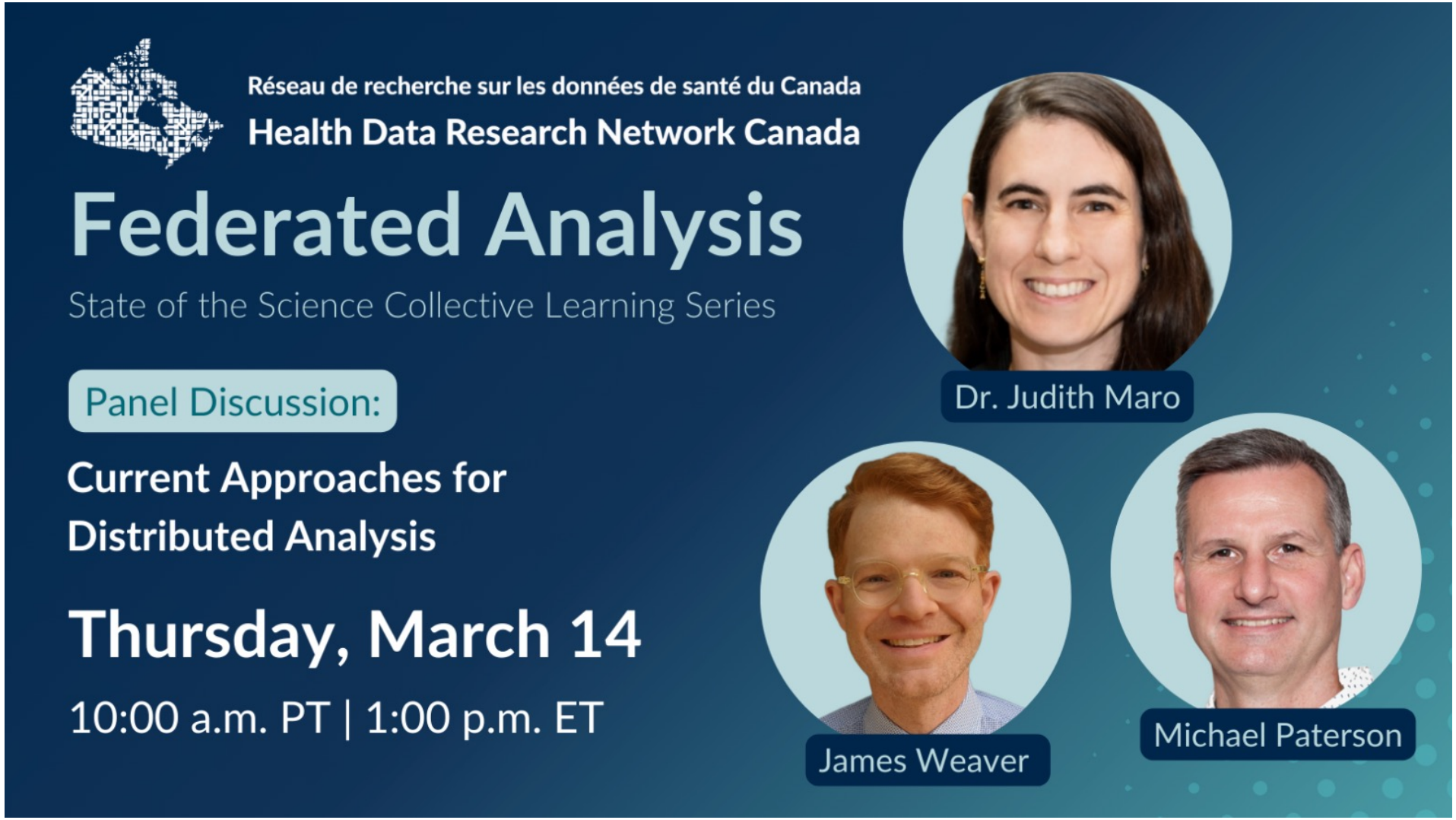
With its advanced ML algorithms, **Jackalope aims to revolutionize the field of data management and mapping**. This tool facilitates a seamless transition to OHDSI's OMOP CDM, ensuring greater efficiency.

JOIN ONLINE:

Save the date: **14th of March, 5 pm, GMT + 2**

To find out more about the event and reserve your spot you can here **https://www.sciforce.tech/**

March 14: Current Approaches for Distributed Analysis



#OHDSISocialShowcase This Week

MONDAY

Bladder cancer - a quality benchmark utilizing FHIR and OMOP

(Andries Clinckaert, Valerie Vandeweerd, Murat Akand, Charlotte De Vlieghere, Bart Vannieuwenhuyse, Michel Van Speybroek, Frank Van der Aa, Martine Lewi, Christos Chatzichristos)

Bladder cancer - a quality benchmark utilizing FHIR and OMOP

Andries Clinckaert[1,2], Valerie Vandeweerd[3], Murat Akand[1], Charlotte De Vlieghere[2], Bart Vannieuwenhuyse [3], Michel Van Speybroek [4], Frank Van der Aa [1,6], Martine Lewi [5], Christos Chatzichristos [3,7]

[1] Department of Urology, University Hospitals Leuven, Belgium
[2] Tiro Health, Belgium
[3] Clinical Innovation Research & Development Janssen Pharmaceutica, Beerse, Belgium
[4] EMEA IT Data Science, Janssen Pharmaceutica, Beerse, Belgium
[5] Global Commercial Data Science, J&J Innovative Medicine, Raritan, USA
[6] Department of Development and Regeneration, KU Leuven, Belgium
[7] Biomed Stadius, Department of Electrical Engineering, KU Leuven, Belgium

Background

- The field of oncology relies heavily on high quality and granular patient level data.
- One of the main challenges for multi-center benchmarks and potentially analysis across different centers is the quality of the data.

- Transurethral resection of bladder tumor (TURBT) is a treatment for non-muscle invasive bladder cancer (NMIBC).

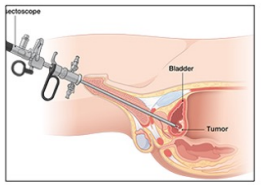


Fig. 1. Representation of TURBT [1]

- Details about the surgery, such as macroscopic aspects of the tumor and its localization and perio-operative complications are reported in the medical report.
- Following surgery, a multidisciplinary tumor meeting (MDT) takes place a few days later.
- Our quality benchmark platform uses data from the structured forms following surgery and MDT.

Methods

Quality indicators (Qis) - FHIR:
- Qis can serve as a valuable mitigation strategy.
- Focus specifically on the Re-TURBT for patients who had no detrusor muscle present.

$$\%\,re-TURBT = \frac{All\ TURBTs\ for\ no\ DM*\ AND\ subsequent\ TURBT\ within\ 6\ weeks}{All\ TURBTs\ for\ no\ DM*}$$

- Only patients where no detrusor muscle was found in the resection specimen are included.
- The tumor cannot be Ta or Tis grade (known at time of MDT, after the initial TURBT).
- Each sequence of a TURBT procedure followed by an MDT is counted.

Longitudinal Patient Trajectory (LPT) – OMOP:
- By leveraging the temporal and relational aspects of the OMOP data, we examined the sequence of events within the disease trajectory.

References

- http://www.urologynorth.ca/turbt.html

Results

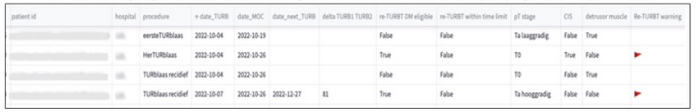
Quality indicators:
- 3% of patients who should undergo re-TURBT had a structured report.
- Doubled interval (84 days) – 35%.
- Tripled interval (126) -61%.

- Small final cohort of 31 patients.
- **Too strict inclusion or incorrectly / partially used reporting forms.**

Longitudinal Patient Trajectory (LPT):
- Instances of pathology results available without previous procedure documentation.
- Procedure reported but subsequently no pathology data reported.
- Incomplete understanding of patient's journey.
- Gaps in post-procedure monitoring of data capture.

Conclusions

- Incomplete documentation and inconsistent data export can significantly impact data quality.
- Implement mitigation strategies, **data quality dashboard extended with disease specific rules:**
  - A warning is generated if a TURBT procedure is not followed by a T-stage in one week.
  - A warning is generated when a T-stage is recorded without accompanying a TURBT procedure.



- By leveraging FHIR and OMOP in BC EHR data, we can ensure comprehensive data inclusion, and improve data registration and quality.
- By leveraging standardized data protocols healthcare organizations can optimize data quality and enhance decision-making.

Acknowledgements

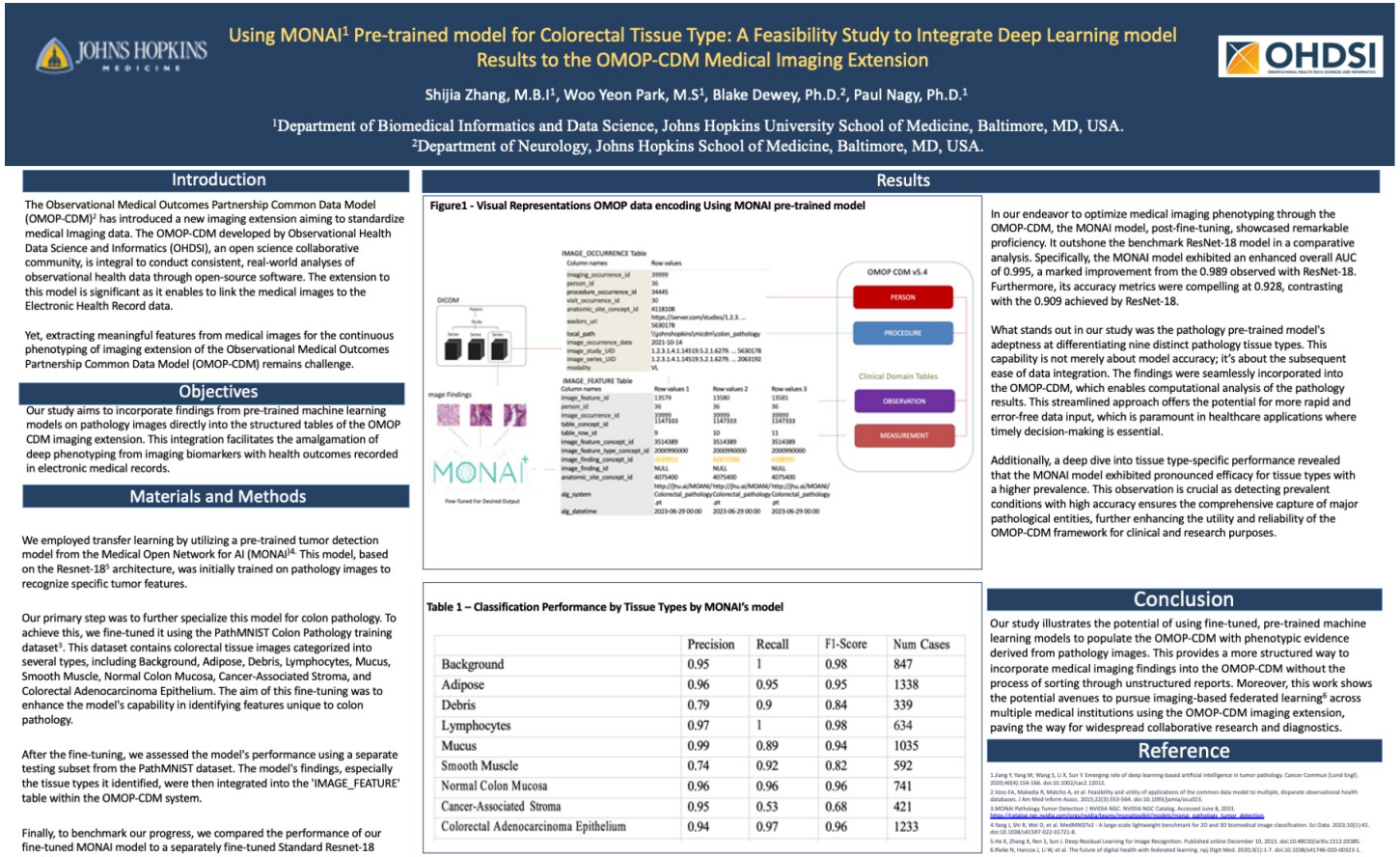
The current research was implemented within Athena project, which is funded by Flanders Innovation & Entrepreneurship (VLAIO). Project number: HNC.2019.2528.

#OHDSISocialShowcase This Week

TUESDAY

Using MONAI Pre-Trained Models for Colorectal Tissue Type Phenotyping: A Feasibility Study to Integrate Deep Learning Model Results using the Medical Extension OMOP CDM

(Shijia Zhang, Woo Yeon Park, Blake Dewey, Paul Nagy)



Using MONAI[1] Pre-trained model for Colorectal Tissue Type: A Feasibility Study to Integrate Deep Learning model Results to the OMOP-CDM Medical Imaging Extension

Shijia Zhang, M.B.I[1], Woo Yeon Park, M.S[1], Blake Dewey, Ph.D.[2], Paul Nagy, Ph.D.[1]

[1]Department of Biomedical Informatics and Data Science, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
[2]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA.

Introduction

The Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM)[2] has introduced a new imaging extension aiming to standardize medical Imaging data. The OMOP-CDM developed by Observational Health Data Science and Informatics (OHDSI), an open science collaborative community, is integral to conduct consistent, real-world analyses of observational health data through open-source software. The extension to this model is significant as it enables to link the medical images to the Electronic Health Record data.

Yet, extracting meaningful features from medical images for the continuous phenotyping of imaging extension of the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) remains challenge.

Objectives

Our study aims to incorporate findings from pre-trained machine learning models on pathology images directly into the structured tables of the OMOP CDM imaging extension. This integration facilitates the amalgamation of deep phenotyping from imaging biomarkers with health outcomes recorded in electronic medical records.

Materials and Methods

We employed transfer learning by utilizing a pre-trained tumor detection model from the Medical Open Network for AI (MONAI)[3,4]. This model, based on the Resnet-18[5] architecture, was initially trained on pathology images to recognize specific tumor features.

Our primary step was to further specialize this model for colon pathology. To achieve this, we fine-tuned it using the PathMNIST Colon Pathology training dataset[3]. This dataset contains colorectal tissue images categorized into several types, including Background, Adipose, Debris, Lymphocytes, Mucus, Smooth Muscle, Normal Colon Mucosa, Cancer-Associated Stroma, and Colorectal Adenocarcinoma Epithelium. The aim of this fine-tuning was to enhance the model's capability in identifying features unique to colon pathology.

After the fine-tuning, we assessed the model's performance using a separate testing subset from the PathMNIST dataset. The model's findings, especially the tissue types it identified, were then integrated into the 'IMAGE_FEATURE' table within the OMOP-CDM system.

Finally, to benchmark our progress, we compared the performance of our fine-tuned MONAI model to a separately fine-tuned Standard Resnet-18 model.

Results

Figure1 - Visual Representations OMOP data encoding Using MONAI pre-trained model

In our endeavor to optimize medical imaging phenotyping through the OMOP-CDM, the MONAI model, post-fine-tuning, showcased remarkable proficiency. It outshone the benchmark ResNet-18 model in a comparative analysis. Specifically, the MONAI model exhibited an enhanced overall AUC of 0.995, a marked improvement from the 0.989 observed with ResNet-18. Furthermore, its accuracy metrics were compelling at 0.928, contrasting with the 0.909 achieved by ResNet-18.

What stands out in our study was the pathology pre-trained model's adeptness at differentiating nine distinct pathology tissue types. This capability is not merely about model accuracy; it's about the subsequent ease of data integration. The findings were seamlessly incorporated into the OMOP-CDM, which enables computational analysis of the pathology results. This streamlined approach offers the potential for more rapid and error-free data input, which is paramount in healthcare applications where timely decision-making is essential.

Additionally, a deep dive into tissue type-specific performance revealed that the MONAI model exhibited pronounced efficacy for tissue types with a higher prevalence. This observation is crucial as detecting prevalent conditions with high accuracy ensures the comprehensive capture of major pathological entities, further enhancing the utility and reliability of the OMOP-CDM framework for clinical and research purposes.

Table 1 – Classification Performance by Tissue Types by MONAI's model

	Precision	Recall	F1-Score	Num Cases
Background	0.95	1	0.98	847
Adipose	0.96	0.95	0.95	1338
Debris	0.79	0.9	0.84	339
Lymphocytes	0.97	1	0.98	634
Mucus	0.99	0.89	0.94	1035
Smooth Muscle	0.74	0.92	0.82	592
Normal Colon Mucosa	0.96	0.96	0.96	741
Cancer-Associated Stroma	0.95	0.53	0.68	421
Colorectal Adenocarcinoma Epithelium	0.94	0.97	0.96	1233

Conclusion

Our study illustrates the potential of using fine-tuned, pre-trained machine learning models to populate the OMOP-CDM with phenotypic evidence derived from pathology images. This provides a more structured way to incorporate medical imaging findings into the OMOP-CDM without the process of sorting through unstructured reports. Moreover, this work shows the potential avenues to pursue imaging-based federated learning[6] across multiple medical institutions using the OMOP-CDM imaging extension, paving the way for widespread collaborative research and diagnostics.

Reference

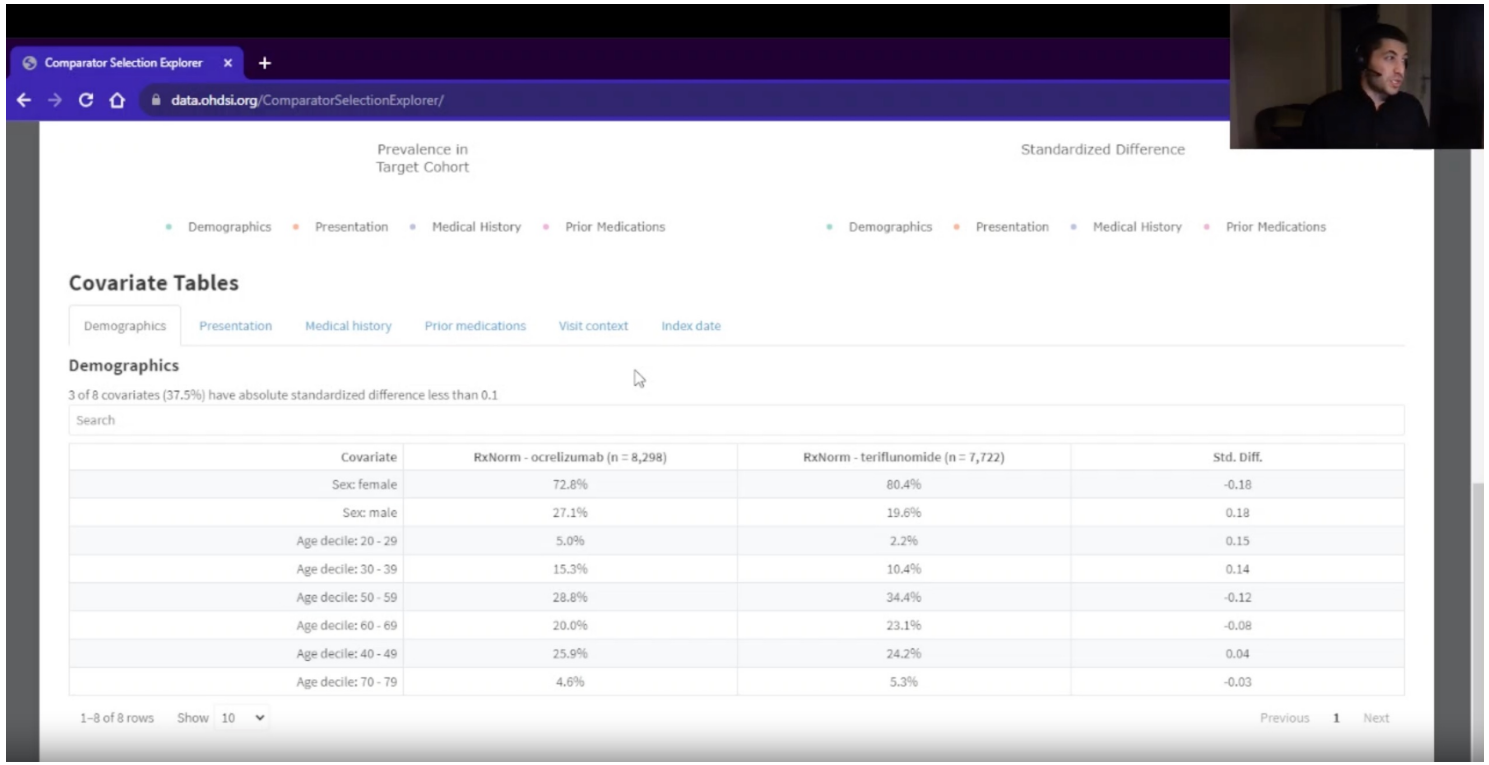
1.Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Commun (Lond Engl). 2020;40(4):154-166. doi:10.1002/cac2.12012.
2.Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc. 2015;22(3):553-564. doi:10.1093/jamia/ocu023.
3.MONAI Pathology Tumor Detection | NVIDIA NGC. NVIDIA NGC Catalog. Accessed June 8, 2023. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/monaihealth/models/monai_pathology_tumor_detection
4.Yang J, Shi R, Wei D, et al. MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification. Sci Data. 2023;10(1):41. doi:10.1038/s41597-022-01721-8.
5.He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Published online December 10, 2015. doi:10.48550/arXiv.1512.03385.
6.Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. npj Digit Med. 2020;3(1):1-7. doi:10.1038/s41746-020-00323-1.

#OHDSISocialShowcase This Week

WEDNESDAY

A tool for empirically identifying and reviewing candidate comparators for Pharmacoepidemiological studies

(Justin Bohn, Jamie P. Gilbert, Christopher Knoll, David M. Kern, Patrick B. Ryan)



#OHDSISocialShowcase This Week

THURSDAY

The necessity of validity diagnostics when drawing causal inferences from observational data

(James Weaver, Erica A Voss, Guy Cafri, Kathleen Beyrau, Michelle Nashleanas, Robert Suruki)

Title: The necessity of validity diagnostics when making causal inferences from observational data

Subtitle: An immunology safety study use-case

PRESENTER: **James Weaver**

INTRO:

- Autoimmune disorders may have primary manifestations such as joint pain and bowel inflammation but can also have secondary manifestations such as non-infectious uveitis (NIU)
- A potential safety concern was identified from spontaneous reports for NIU following exposure to Remicade®, a biologic therapy with multiple indications for which alternative therapies are available
- We assessed the risk of NIU among patients exposed to Remicade® within 4 indicated populations to reduce confounding by underlying disease state

METHODS:

1. Indication populations: inflammatory bowel conditions [IBD], psoriatic conditions, rheumatoid arthritis [RA], ankylosing spondylitis
2. Compared Remicade® to indication-specific alternative therapies
3. Four analysis strategies (2 propensity score [PS] methods x 2 time-at-risk [TAR] periods)
4. Five observational databases
5. 80 analyses total, 20 designated primary
6. Comprehensive NIU phenotyping
7. Four validity diagnostics: >0 outcomes in target and comparator TAR, >35% in equipoise, absolute standardized mean difference [ASMD] <0.1 for all covariates after PS matching, expected absolute systematic error [EASE] <0.25

Validity diagnostics should always be applied to determine which analyses will produce **reliable evidence** to support causal inferences

Indication	Database	Target exposure	Comparator exposure	Diagnostic #1		Diagnostic #2	Diagnostic #3	Diagnostic #4	Passed
				T events	C events	% equipoise	ASMD (max)	EASE	
AS	AMB EMR	Remicade®	certolizumab pegol, golimumab, ixekizumab, secukinumab	0	0	0.531	0.302	0.976	1
AS	Pharmetrics Plus	Remicade®	certolizumab pegol, golimumab, ixekizumab, secukinumab	6	9	0.405	0.157	0.044	3
AS	Optum® EHR	Remicade®	certolizumab pegol, golimumab, ixekizumab, secukinumab	1	6	0.638	0.218	0.233	3
AS	Clinformatics®	Remicade®	certolizumab pegol, golimumab, ixekizumab, secukinumab	3	4	0.417	0.272	0.172	3
AS	IBM CCAE	Remicade®	certolizumab pegol, golimumab, ixekizumab, secukinumab	2	6	0.434	0.239	0.180	3
IBD	AMB EMR	Remicade®	golimumab, certolizumab pegol, ustekinumab, vedolizumab	3	1	0.422	0.09	0.284	3
IBD	Pharmetrics Plus	Remicade®	golimumab, certolizumab pegol, ustekinumab, vedolizumab	12	42	0.431	0.047	0.074	4
IBD	Optum® EHR	Remicade®	golimumab, certolizumab pegol, ustekinumab, vedolizumab	4	7	0.480	0.055	0.087	4
IBD	Clinformatics®	Remicade®	golimumab, certolizumab pegol, ustekinumab, vedolizumab	6	15	0.412	0.105	0.040	3
IBD	IBM CCAE	Remicade®	golimumab, certolizumab pegol, ustekinumab, vedolizumab	10	18	0.387	0.071	0.107	4
PP-PsA	AMB EMR	Remicade®	golimumab, certolizumab pegol, guselkumab, risankizumab, tildrakizumab, brodalumab, ixekizumab, secukinumab, ustekinumab	0	0	0.254	0.145	0.344	0
PP-PsA	Pharmetrics Plus	Remicade®	golimumab, certolizumab pegol, guselkumab, risankizumab, tildrakizumab, brodalumab, ixekizumab, secukinumab, ustekinumab	4	10	0.155	0.132	0.178	2
PP-PsA	Optum® EHR	Remicade®	golimumab, certolizumab pegol, guselkumab, risankizumab, tildrakizumab, brodalumab, ixekizumab, secukinumab, ustekinumab	6	3	0.306	0.099	0.246	3
PP-PsA	Clinformatics®	Remicade®	golimumab, certolizumab pegol, guselkumab, risankizumab, tildrakizumab, brodalumab, ixekizumab, secukinumab, ustekinumab	0	7	0.171	0.199	0.110	1
PP-PsA	IBM CCAE	Remicade®	golimumab, certolizumab pegol, guselkumab, risankizumab, tildrakizumab, brodalumab, ixekizumab, secukinumab, ustekinumab	3	9	0.147	0.167	0.010	2
RA	AMB EMR	Remicade® & methotrexate	certolizumab pegol, tocilizumab	1	0	0.445	0.127	0.307	1
RA	Pharmetrics Plus	Remicade® & methotrexate	certolizumab pegol, tocilizumab	2	6	0.352	0.179	0.158	3
RA	Optum® EHR	Remicade® & methotrexate	certolizumab pegol, tocilizumab	5	4	0.558	0.097	0.141	4
RA	Clinformatics®	Remicade® & methotrexate	certolizumab pegol, tocilizumab	2	8	0.363	0.252	0.034	3
RA	IBM CCAE	Remicade® & methotrexate	certolizumab pegol, tocilizumab	4	8	0.508	0.151	0.070	3

Table 1 - 20 primary analyses comparing target vs comparator cohorts by indication population by database under the primary analysis settings: 1:10 propensity score matching during on-treatment time-at-risk. Shaded blue rows indicate which comparisons passed 4 validity diagnostics in each database. The 3 databases that passed diagnostics in the IBD population contributed to the meta-analysis (see Results)

Key - AMB EMR: IQVIA Ambulatory Electronic Medical Records, Pharmetrics Plus: IQVIA Adjudicated Health Plan Claims Data, Optum® EHR : Optum® de-identified Electronic Health Record Dataset, Clinformatics®: Optum's de-identified Clinformatics® Data Mart Database, IBM CCAE: IBM® MarketScan® Commercial Claims and Encounters Database, ITT: intent-to-treat, AS: ankylosing spondylitis, IBD: inflammatory bowel diseases (Crohn's disease or ulcerative colitis), RA: rheumatoid arthritis, T: target cohort, C: comparator cohort, ASMD: absolute standardized mean difference, EASE: expected absolute systematic error

Take a picture to view the abstract

METHODS, cont'd.

- Dersimonian-Laird (DL) random-effects meta-analysis within indication where >1 database passed all four diagnostics

RESULTS

- NIU definition: 2x code OR 1x code in ophthalmology setting
- 19/80 (24%) passed diagnostics, 4/20 (20%) were primary
- Most analyses that did not pass diagnostics were from failed covariate balance (ASMD<0.1)
- Among patients with IBD, pooled hazard ratio (HR) 0.75 (95% confidence interval [CI] 0.38-1.40)
- Among patients with RA HR 1.23, 95% CI 0.14-10.47)

LIMITATIONS

- Alternative PS balancing methods like inverse probability weighting not explored
- Outcome misclassification measured but not corrected
- 0 outcome count prevented calculating standard errors needed for DL meta-analysis; novel Bayesian meta-analysis methods can account for this

DISCUSSION

- Validity diagnostics applied to a specific research question in a heterogeneous, real-world setting indicated that effect estimates from many analyses were unreliable and should not be interpreted as causal
- Clinically, if an increased risk of NIU exists, it is unlikely to be greater than 40%

James Weaver, Erica A. Voss, Guy Cafri, Kathleen Beyrau, Michelle Nashleanas, Robert Suruki

janssen | Johnson & Johnson

OHDSI

#OHDSISocialShowcase This Week

FRIDAY

Identification of HIV positive individuals across multiple datasets

(Craig Mayer)

Identification of HIV positive individuals across multiple datasets

PRESENTER: **Craig Mayer**

INTRO:

- Condition attribution is key for appropriate patient capture
- Attribution may depend on amount and type of data available

METHODS

1. Analyzed 3 datasets
    1. All of Us (EHR and Research)
    2. UK Biobank (Research)
    3. Clinical Practice Research Datalink (EHR)
2. Analyzed 3 domains
    1. Condition
    2. Observation
        1. Self-reported
    3. Measurement
        1. Confirmatory tests
        2. No screening tests
3. Found patients from each domain and measured patient capture and crossover between domains

RESULTS

For All of Us, each HIV attribution method produced additional cases.

For UK Biobank, all HIV positive individuals were captured through self-reporting. The limited measurement and condition data captured no additional HIV positive individuals.

For CPRD, all HIV positive individuals were captured via condition. No additional HIV positive individuals were found via the measurement data.

Ideal HIV attribution method may vary by type of dataset

For robust case capture, HIV attribution may require attribution methods from multiple domains.

HIV Counts by Domain

Domain	AoU	UKBB	CPRD
Condition	5,185	214	50,374
Observation/Self-reported	1,686	484	X
Measurement	3,925	18	2,806

Case crossover by domain

Domains	AoU	UKBB	CPRD
Condition and Observation	1,031	194	x
Condition and Measurement	2,403	18	2,806
Measurement and Observation	575	18	x
Condition, Measurement and Observation	550	18	x

Dataset Descriptions

All of Us:

Combination of research and EHR. Attribution of HIV through EHR condition and measurement. Self-reported data through survey questionnaire for personal medical history. Ongoing data collection with participants increasing.

UK Biobank:

Combination of EHR and research. Partial imported EHR with limited bio sample measurement data. Self-reported through survey questionnaire. Less than 10,000 participants have biospecimen leading to limited HIV capture through measurement. No EHR measurement data. Participants over 50 years old.

CPRD:

Primary care EHR data with both condition and measurement. Data ranging back decades with a cutoff of June 2021. Includes large percentage of UK population

Participant Counts:

	Total Participants
AoU	413,457
UKBB	502,390
CPRD	49,102,289

Acknowledgement
This work was supported in part by the Lister Hill National Center for Biomedical Communications of the National Library of Medicine (NLM), and in part by the Office of AIDS Research (OAR), National Institutes of Health

NIH U.S. National Library of Medicine

OHDSI

Opening: Biomedical Informatics Data Scientist at Stanford

Stanford MEDICINE Health Care

Who We Are | What We Offer | How We Hire | Career Areas | Search

Biomedical Informatics Data Scientist

1.0 FTE • Full time • Day - 08 Hour • R2335119 • Hybrid • 84866 IT RESEARCH • Technology & Digital Solutions • 455 Broadway,REDWOOD CITY,California •

1.0 FTE Full time Day - 08 Hour R2335119 Hybrid 84866 IT RESEARCH Technology & Digital Solutions 455 Broadway,REDWOOD CITY,California

If you're ready to be part of our legacy of hope and innovation, we encourage you to take the first step and explore our current job openings. Your best is waiting to be discovered.

Day - 08 Hour (United States of America)

This is a Stanford Health Care job.

A Brief Overview

The Biomedical Informatics Data Scientist will partner with researchers and clinicians to enable effective and efficient use of data and resources available via Stanford's research clinical data repository (STARR) including the Electronic Health Records in the OMOP Common Data Model, radiology and cardiology imaging data and associated metadata, and new data types as they get integrated along with their databases and respective cohort query tools and interfaces e.g., OHDSI ATLAS. This individual will enable researchers to maximize their understanding, interpretation and use of these clinical and research tools for more informed and productive research, clinical trials, patient care and quality outcome projects.

Clean, extract, transform and analyze various kinds of clinical data to create analysis-ready datasets that follow the FAIR (Findable, Accessible, Interoperable and Re-usable) principles. Partner with researchers and clinicians to enable effective and efficient use of Stanford Clinical data and resources for the advancement of research and the educational mission.

(I²DB)

Postdoc/Senior Data Analyst Opening at WashU

The Zhang Lab at Washington University School of Medicine in St. Louis has **one postdoct/senior data analyst position** to work on **causal machine learning** and **responsible AI** for reliable real-world evidence generation.

PI: Linying Zhang, PhD

- More details at https://linyingzhang.com
  - Postdoc: https://linyingzhang.com/files/Postdoc.pdf
  - Data analyst: https://linyingzhang.com/files/Analyst.pdf
- If interested, please send CV and cover letter to linyingz@wustl.edu

Washington University School of Medicine in St. Louis

Opening: Epidemiology UX/Web Design Intern at J&J

Career Programs

Epidemiology UX/Web Design Intern

JOB TITLE	Epidemiology UX/Web Design Intern
FUNCTION	Career Programs
SUB FUNCTION	Non-LDP Intern/Co-Op
LOCATION	Raritan, New Jersey, United States
DATE POSTED	Jan 19 2024
REQUISITION NUMBER	2406163977W

DESCRIPTION

Janssen Research & Development, L.L.C., a division of Johnson & Johnson's Family of Companies is recruiting for Epidemiology UX/Web Design Intern. This position is a member of the Observational Health Data Analytics (OHDA) team. OHDA's mission is to improve the lives individuals and quality of healthcare by efficiently generating real-world evidence from the world's observational health data, transparently disseminating evidence-based insights to real-world decision-makers, and objectively advancing the science and technology behind reliab

Apply Now

Director, RWE at Gilead

Director, RWE - Data Science - OHDSI

Apply

Responsibilities:

Collaborate with researchers and data scientists to understand project requirements and translate them into OHDSI-compatible solutions. Work with databases, ensuring data integrity and optimization for OHDSI-related queries and analyses. Perform data analyses in OHDSI-related tools like ATLAS. Customize and extend OHDSI tools and applications to meet specific project needs. Collaborate with cross-functional teams to troubleshoot and resolve technical issues related to OHDSI implementations. Stay informed about OHDSI community updates, best practices, and emerging trends in observational health data research. Contribute to the development and documentation of data standards and conventions within the OHDSI community.

About Us

GILEAD

Gilead Sciences, Inc. is a biopharmaceutical company that has pursued and achieved breakthroughs in medicine for more than three decades, with the goal of creating a healthier world for all people. The company is committed to advancing innovative medicines to prevent and treat life-threatening diseases, including HIV, viral hepatitis and cancer. Gilead operates in more than 35 countries worldwide, with headquarters in Foster City, California.

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