Workgroup OKRs + Phenotype Phebruary Update #2

OHDSI Community Call
Feb. 13, 2024 • 11 am ET

OHDSI
Observational Health Data Sciences and Informatics

Upcoming Community Calls

Date	Topic
Feb. 13	Workgroup OKRs / Phenotype Phebruary Update 2
Feb. 20	Workgroup OKRs / Phenotype Phebruary Update 3
Feb. 27	Workgroup OKRs / Phenotype Phebruary Update 4
Mar. 5	New Vocabulary Release Update
Mar. 12	TBA
Mar. 19	NO MEETING
Mar. 26	Recent OHDSI Publications

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Xinyuan Zhang, Yixue Feng, Fang Li, Jin Ding, Danyal Tahseen, Ezekiel Hinojosa, Yong Chen, and Cui Tao** on the publication of **Evaluating MedDRA-to-ICD terminology mappings** in *BMC Medical Informatics and Decision Making.*

Zhang *et al.*
BMC Medical Informatics and Decision Making (2023) 23:299
https://doi.org/10.1186/s12911-023-02375-1

BMC Medical Informatics and Decision Making

RESEARCH Open Access

Check for updates

Evaluating MedDRA-to-ICD terminology mappings

Xinyuan Zhang[1], Yixue Feng[2], Fang Li[1], Jin Ding[1], Danyal Tahseen[3], Ezekiel Hinojosa[3], Yong Chen[4] and Cui Tao[1,5*]

From 8th-12th International Workshop on Vaccine and Drug Ontology Studies (VDOS-2019-2022)
Various locations. Various dates.

Abstract

Background In this era of big data, data harmonization is an important step to ensure reproducible, scalable, and collaborative research. Thus, terminology mapping is a necessary step to harmonize heterogeneous data. Take the Medical Dictionary for Regulatory Activities (MedDRA) and International Classification of Diseases (ICD) for example, the mapping between them is essential for drug safety and pharmacovigilance research. Our main objective is to provide a quantitative and qualitative analysis of the mapping status between MedDRA and ICD.

We focus on evaluating the current mapping status between MedDRA and ICD through the Unified Medical Language System (UMLS) and Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). We summarized the current mapping statistics and evaluated the quality of the current MedDRA-ICD mapping; for unmapped terms, we used our self-developed algorithm to rank the best possible mapping candidates for additional mapping coverage.

Results The identified MedDRA-ICD mapped pairs cover 27.23% of the overall MedDRA preferred terms (PT). The systematic quality analysis demonstrated that, among the mapped pairs provided by UMLS, only 51.44% are considered an exact match. For the 2400 sampled unmapped terms, 56 of the 2400 MedDRA Preferred Terms (PT) could have exact match terms from ICD.

OHDSI Shoutouts!

Congratulations to the team of **Tathagata Bhattacharjee, Sylvia Kiwuwa-Muyingo, Chifundo Kanjala, Mulalaqhooa L Maoyi, David Amadi, Michael Ochola, Damazo Kadengye, Arofan Gregory, Agnes Kiragga, Amelia Taylor, Jay Greenfield, Emma Slaymaker, Jim Todd, and the INSPIRE Network** on the publication of **INSPIRE datahub: a pan-African integrated suite of services for harmonising longitudinal population health data using OHDSI tools** in *Frontiers in Digital Health.*

frontiers | Frontiers in Digital Health

TYPE Methods
PUBLISHED 29 January 2024
DOI 10.3389/fdgth.2024.1329630

Check for updates

OPEN ACCESS

EDITED BY
Qingyu Chen,
National Library of Medicine (NIH),
United States

REVIEWED BY
Cheng Peng,
University of Florida, United States
Xia Jing,
Clemson University, United States

*CORRESPONDENCE
Tathagata Bhattacharjee
tathagata.bhattacharjee@lshtm.ac.uk
Sylvia Kiwuwa-Muyingo
smuyingo@aphrc.org

RECEIVED 31 October 2023
ACCEPTED 15 January 2024
PUBLISHED 29 January 2024

CITATION
Bhattacharjee T, Kiwuwa-Muyingo S, Kanjala C, Maoyi ML, Amadi D, Ochola M, Kadengye D, Gregory A, Kiragga A, Taylor A, Greenfield J, Slaymaker E, Todd J and INSPIRE Network (2024) INSPIRE datahub: a pan-African integrated suite of services for harmonising longitudinal population health data using OHDSI tools.
Front. Digit. Health 6:1329630.
doi: 10.3389/fdgth.2024.1329630

COPYRIGHT
© 2024 Bhattacharjee, Kiwuwa-Muyingo, Kanjala, Maoyi, Amadi, Ochola, Kadengye, Gregory, Kiragga, Taylor, Greenfield, Slaymaker, Todd and INSPIRE Network. This is an open-access article distributed under the

INSPIRE datahub: a pan-African integrated suite of services for harmonising longitudinal population health data using OHDSI tools

Tathagata Bhattacharjee[1*], Sylvia Kiwuwa-Muyingo[2*], Chifundo Kanjala[1,3], Molulaqhooa L. Maoyi[4], David Amadi[1], Michael Ochola[2], Damazo Kadengye[2,5], Arofan Gregory[6], Agnes Kiragga[2], Amelia Taylor[7], Jay Greenfield[5], Emma Slaymaker[1], Jim Todd[1] and INSPIRE Network[8]

[1]Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom, [2]African Population and Health Research Center (APHRC), Nairobi, Kenya, [3]UNICEF (Malawi), Lilongwe, Malawi, [4]South African Population Research Infrastructure Network (SAPRIN), South African Medical Research Council, Durban, South Africa, [5]Department of Economics and Statistics, Kabale University, Kabale, Uganda, [6]Committee on Data of the International Science Council (CODATA), Paris, France, [7]Malawi University of Business and Applied Sciences, Blantyre, Malawi, [8]Implementation Network for Sharing Population Information from Research Entities (INSPIRE Network), Nairobi, Kenya

Introduction: Population health data integration remains a critical challenge in low- and middle-income countries (LMIC), hindering the generation of actionable insights to inform policy and decision-making. This paper proposes a pan-African, Findable, Accessible, Interoperable, and Reusable (FAIR) research architecture and infrastructure named the INSPIRE datahub. This cloud-based Platform-as-a-Service (PaaS) and on-premises setup aims to enhance the discovery, integration, and analysis of clinical, population-based surveys, and other health data sources.

Congratulations, Dr. Chungsoo Kim!

Chungsoo Kim • 1st
PharmD, PhD in Biomedical Inf...
9h •

Every beginning has an end and every end has a new beginning.

I have graduated with a PhD from the Department of Biomedical Informatics at Ajou University(아주대학교). I am deeply grateful to everyone who collaborated closely and worked intensely with me (including my current/former lab friends and #OHDSI Folks). I couldn't have done it without your help and I will never forget my time in this lab.

I am very much looking forward to my new career, which is a #Postdoctoral Associate at Yale University/Yale New Haven Hospital Center for Outcomes Research and Evaluation (CORE) at Yale University School of Medicine.

Happy Lunar New Year and wish me luck!!

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	3 pm	OMOP CDM Oncology Outreach/Research Subgroup
Wednesday	9 am	Patient-Level Prediction
Wednesday	12 pm	Health Equity
Wednesday	2 pm	Natural Language Processing
Wednesday	3 pm	Vulcan/OHDSI Meeting (ZOOM)
Thursday	9 am	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	9:30 am	Themis
Thursday	12 pm	HADES
Thursday	7 pm	Dentistry
Friday	10 am	GIS – Geographic Information System
Friday	10:30 am	Open-Source Community
Friday	11:30 am	Clinical Trials
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Africa Chapter
Monday	11 am	Data Bricks User Group
Monday	2 pm	Electronic Animal Health Records

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

Scientific Review Committee

If you are interested in joining the Scientific Review Committee for the 2024 Global Symposium, you can sign up now.

The deadline to sign up is Feb. 16, and the first meeting will be held March 7.

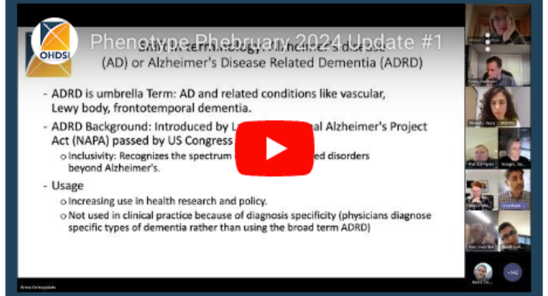
Phenotype Phebruary Homepage

Phenotype Phebruary 2024

"Phenotype Phebruary" is a community-wide initiative to advance the field of phenotyping in observational studies. The OHDSI community has engaged in Phenotype Phebruary in both 2022 and 2023, and this year the community set a goal to understand what is the current practices in the field and how much researchers introduce variability in the process of phenotype development and evaluation.

Under the leadership of **Azza Shoaibi, Anna Ostropolets, Gowtham Rao and James Weaver**, Phenotype Phebruary 2024 focuses on assessing consistency in phenotype definition components, phenotype representation structure, and phenotype validation methods. The month-long activity empowers OHDSI collaborators to engage with each other while advancing the science of phenotyping and gaining education and training around phenotype development and evaluation.

Feb. 6 Phenotype Phebruary Update



Throughout the month, collaborators will engage in a month-long study focused on assessing consistency in phenotype definitions and methods. The goal for this is to evaluate reporting patterns and consistency among reported phenotype algorithms for the same clinical phenotype across observational studies.

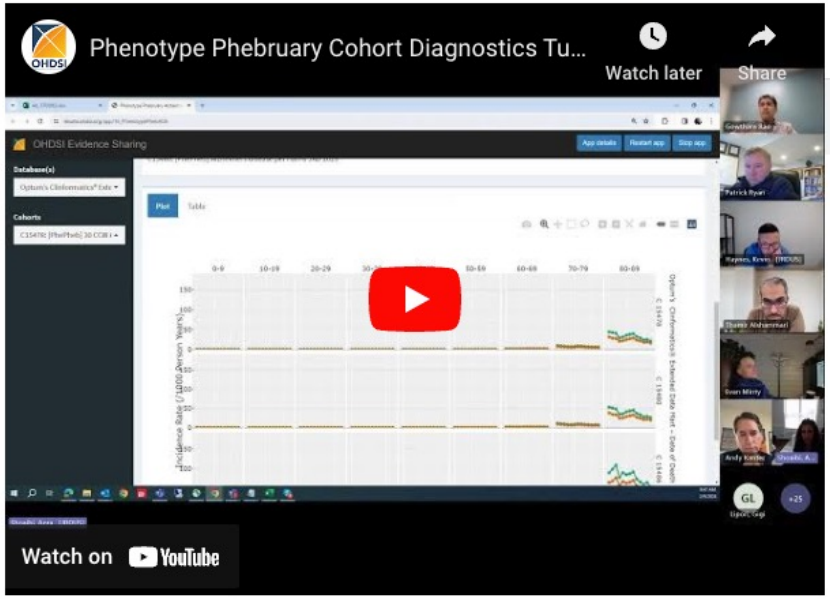
During the Phenotype Phebruary introductory call, community members voted to focus efforts on four specific phenotypes: Alzheimer's Disease, pulmonary hypertension, major depression disorder and prostate cancer). Each week, there will be systematic literature search and synthesis, replication using ATLAS and other OHDSI tools, and summarize variations in population characteristics like incidence rates.

There will be consistent updates on the forum post linked below, and weekly updates during February community calls. The working folder is accessible for anybody who wants to read about our community efforts. If you are interested in joining, please consider joining the Phenotype Development & Evaluation workgroup so you have edit access to the working folder. Please join our meetings and identify an area/task you would be interested in helping complete.

Forum Updates

Working Folder

Videos - Phenotype Phebruary Updates and Discussions



Playlist 5 Videos

- Cohort Diagnostics... 0:16
- ATLAS Demo (09Fe... 0:16
- Update #1 (06Feb2... 0:16
- Discussion (05Feb... 0:16
- Introductory Call (2... 0:16

ohdsi.org/phenotype-phebruary-2024

New Study: Deep Learning Comparison

Network Study: Deep Learning Comparison

Researchers patientprediction, networkstudy

It's been a while since we've seen lhjohn — their last post was 2 years ago.

lhjohn Henrik John 1 7d

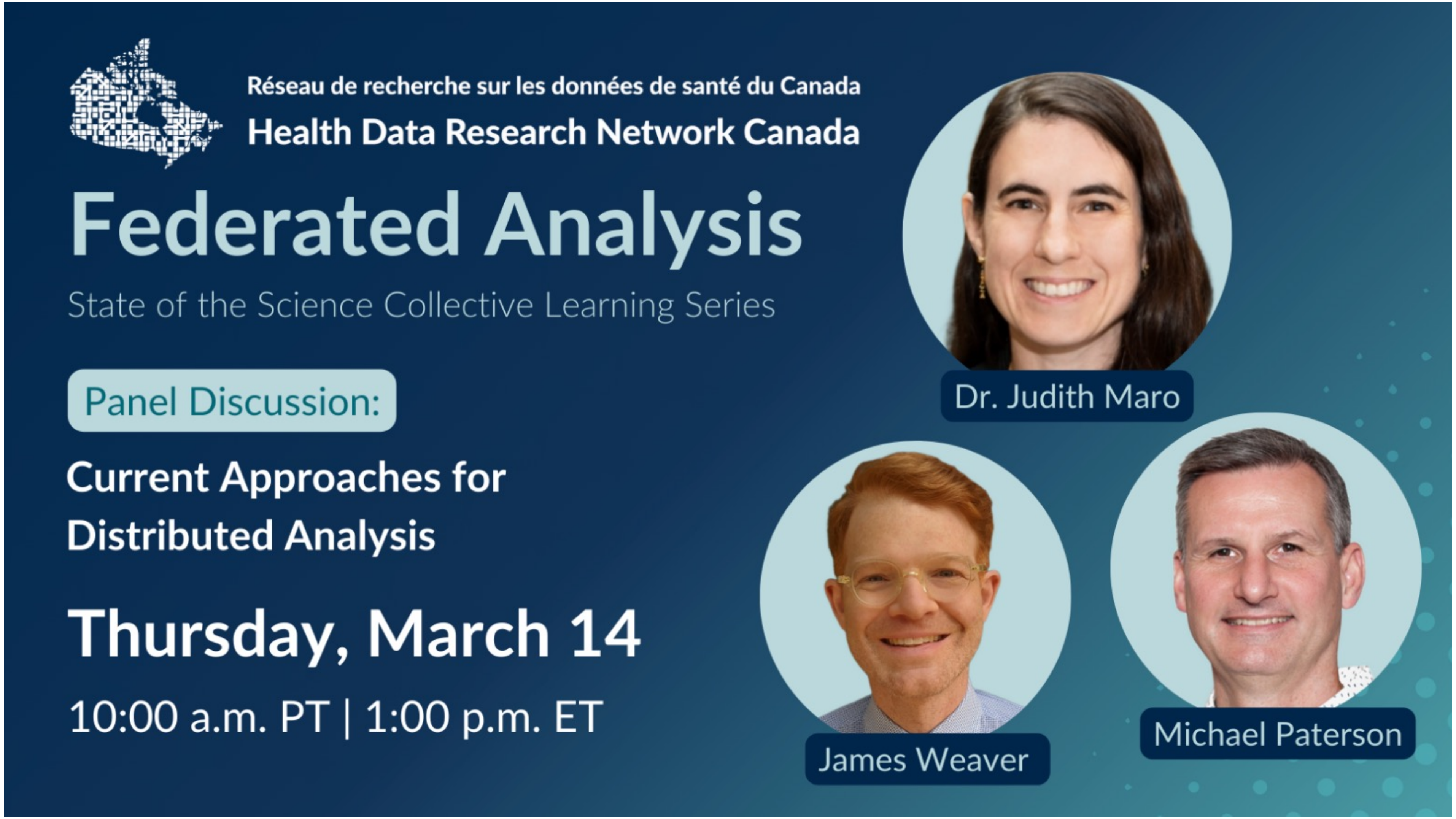
We are pleased to announce our network study **Deep Learning Comparison**.

Study leads: Henrik John (@lhjohn), Chungsoo Kim (@Chungsoo_Kim), Jenna Reps (@jennareps), and Egill Fridgeirsson (@egillax)
GitHub: Deep Learning Comparison - GitHub Repository
Protocol: Deep Learning Comparison - Protocol
Infrastructure: To execute the analysis an Nvidia GPU with CUDA support is required. We recommend a minimum of 12 GB video memory; more is preferred to speed up analysis.
Participant deadline: Please let us know before 1 March, if you are interested in joining the study.

Aim: Assess the value of deep learning methods over conventional methods for the development of clinical prediction models. The specific diseases under consideration are dementia in individuals over 55, lung cancer in those over 45, and bipolar disorder in patients misdiagnosed with major depressive disorder.

Rationale: Deep learning techniques have proven to be highly effective for prediction on unstructured data, such as image and text. However, when applied to structured, sparse, and high-dimensional healthcare data deep learning often yields results comparable to those of simpler, conventional prediction methods. In this study we develop and validate clinical prediction models using deep learning and conventional approaches to compare their discriminatory power and calibration on OMOP CDM data.

March 14: Current Approaches for Distributed Analysis



#OHDSISocialShowcase This Week

MONDAY

Toward a General-Purpose Geography-Focused OHDSI Infrastructure

(**Kyle Zollo-Venecek**, Robert Miller, William G. Adams, Jay Greenfield, Timothy B Norris, Polina Talapova, Maksym Trofymenko, Andrew Williams)

Gaia

Toward a General-Purpose Geography-Focused OHDSI Infrastructure

PRESENTER:
Kyle Zollo-Venecek
kylezollo@gmail.com

BACKGROUND

OHDSI studies typically overlook ***regional factors*** like ***poverty*** and the ***environment*** or rely on non-scalable solutions due to the lack of a geography-focused infrastructure compatible with OMOP CDM and OHDSI tools. While past one-off studies have produced laudable results leveraging spatial data with ad-hoc methods, the OHDSI GIS WG has made notable progress towards making place-based variables available in downstream OMOP analyses using a novel scalable and reproducible framework.

METHODS

The OHDSI GIS Workgroup is building support that complements prior OHDSI GIS efforts by enabling analysis of region attributes like poverty in conjunction with OMOPed clinical data. Development in the workgroup can be thought of as two segments: the ***foundational work*** for a geography-focused infrastructure and ***OHDSI integration***. ***Foundational work*** includes a data catalog for managing compatible external sources, a harmonized data model for staging geospatial data, and software tools for data ingestion and transformation. ***OHDSI integration*** has so far focused on the development of a just-in-time event table *exposure_occurrence*, which relates regional information like social determinants of health (SDOHs) and environmental pollutants to patients and their addresses over time.

RESULTS

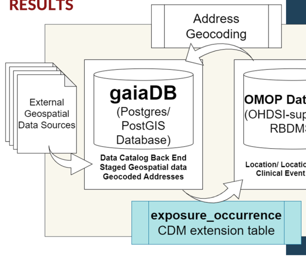
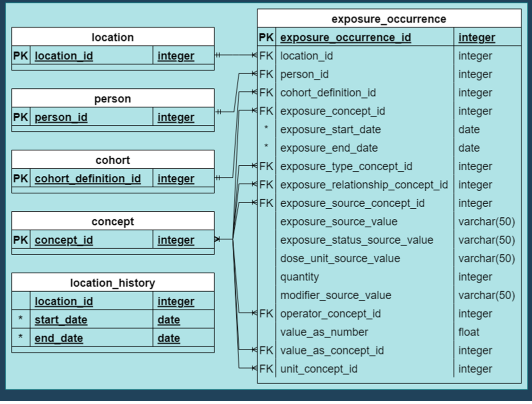


Fig 1. Simplified GIS workflow (above). Geospatial data sources are registered in the gaiaDB data catalog. OMOP addresses are geocoded and stored alongside harmonized geospatial data. A spatio-temporal join creates the exposure_occurrence table which is integrated back into an OMOP CDM.

Fig 2. The *exposure_occurrence* table and its relationships (right). *exposure_occurrence* relates patients to exposures by their address's overlap with a region (aka geo) associated with an exposure attribute. *exposure_occurrence* is available just-in-time for use in cohort definitions and as covariates

Gaia introduces a **geography-focused** infrastructure to the OMOP CDM including a **universal representation** for geospatial data, software tools for **data ingestion**, an OMOP-aligned **GIS vocabulary package**, and a new event table for capturing **person-level exposures**

View the poster and abstract submission



OHDSI GIS near-term objectives include:

- Collect and address more use cases
- Create and curate more data sources
- Integration with OHDSI tools
- FAIRification of the Gaia workflow
- Visualization tooling

We are seeking contributors with experience in:

- Geospatial visualization
- Geospatial statistical methods

Learn how you can get involved with OHDSI GIS WG, dig deeper into our example use case, or get started with *Gaia*

OHDSI GIS Vocabularies

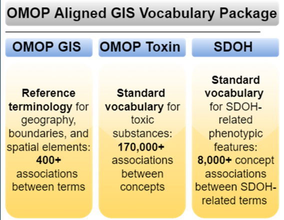


Fig 3. The OMOP Aligned GIS Vocabulary Package advances data-driven healthcare research by seamlessly integrating spatial, environmental, and societal determinants into unified data structures, catering to the evolving demands of the field

OHDSI GIS Demo Use Case: Social Determinants of Health (SDOH)

We can make a preliminary exploration of regional poverty and Type 2 Diabetes management using Gaia:

- Discover and transform registered external data sources using gaiaCatalog (Fig 4)
- Relate regional information to patients in OMOP with the exposure_occurrence table (Fig 2)
- Populate the cohort table using Atlas-style SQL queries that leverage the GIS Vocabularies (Fig 3)
- Explore occurrences of conditions within cohorts (Fig 5)

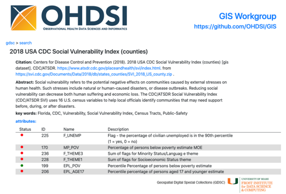


Fig 4. *gaiaCatalog displays* the CDC Social Vulnerability Index (SVI) by using the external data source's curated metadata to display information for discovering regional attributes and building layers

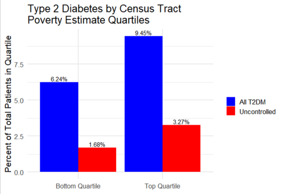


Fig 5. T2DM diagnoses across poverty quartiles. Preliminary exploration shows a roughly 51% increase between percent of total T2DM diagnoses between patients living in bottom and top quartile poverty regions, and over a 94% increase in uncontrolled or under-managed T2DM between the groups

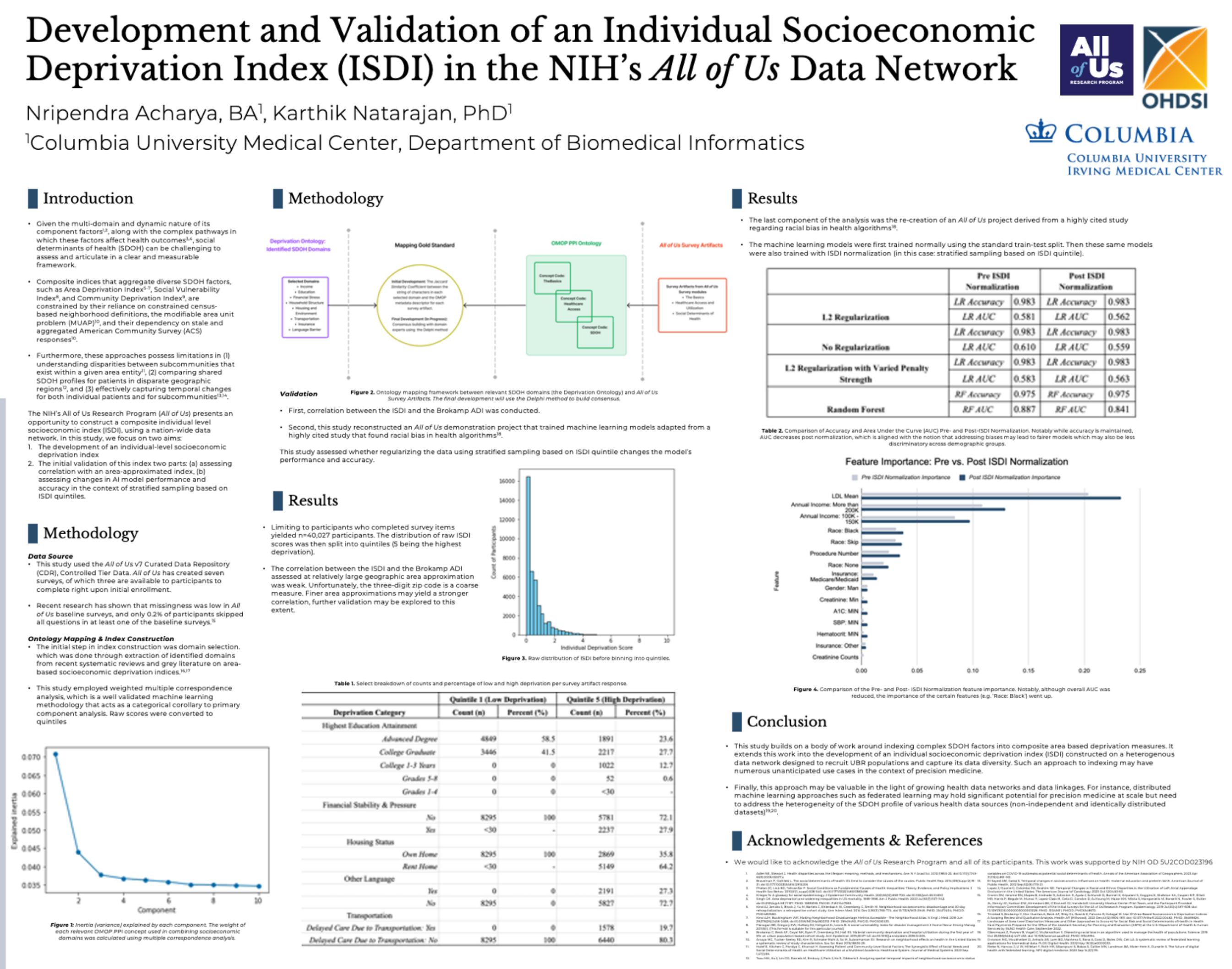
Kyle Zollo-Venecek,
Robert Miller,
William G. Adams,
Jay Greenfield,
Timothy B Norris,
Polina Talapova,
Maksym Trofymenko,
Andrew Williams

#OHDSISocialShowcase This Week

TUESDAY

The Development and Validation of an Individual-Level Socioeconomic Deprivation Index (ISDI) with OMOP in the NIH's All of Us Data Network

(Nripendra Acharya, Karthik Natarajan)



Development and Validation of an Individual Socioeconomic Deprivation Index (ISDI) in the NIH's *All of Us* Data Network

Nripendra Acharya, BA[1], Karthik Natarajan, PhD[1]
[1]Columbia University Medical Center, Department of Biomedical Informatics

Introduction

- Given the multi-domain and dynamic nature of its component factors[1,2], along with the complex pathways in which these factors affect health outcomes[3,4], social determinants of health (SDOH) can be challenging to assess and articulate in a clear and measurable framework.
- Composite indices that aggregate diverse SDOH factors, such as Area Deprivation Index[5-7], Social Vulnerability Index[8], and Community Deprivation Index[9], are constrained by their reliance on constrained census-based neighborhood definitions, the modifiable area unit problem (MUAP)[10], and their dependency on stale and aggregated American Community Survey (ACS) responses[10].
- Furthermore, these approaches possess limitations in (1) understanding disparities between subcommunities that exist within a given area entity[11], (2) comparing shared SDOH profiles for patients in disparate geographic regions[12], and (3) effectively capturing temporal changes for both individual patients and for subcommunities[13,14].

The NIH's All of Us Research Program (*All of Us*) presents an opportunity to construct a composite individual level socioeconomic index (ISDI), using a nation-wide data network. In this study, we focus on two aims:

1. The development of an individual-level socioeconomic deprivation index
2. The initial validation of this index two parts: (a) assessing correlation with an area-approximated index, (b) assessing changes in AI model performance and accuracy in the context of stratified sampling based on ISDI quintiles.

Methodology

Data Source

- This study used the *All of Us v7* Curated Data Repository (CDR), Controlled Tier Data. *All of Us* has created seven surveys, of which three are available to participants to complete right upon initial enrollment.
- Recent research has shown that missingness was low in *All of Us* baseline surveys, and only 0.2% of participants skipped all questions in at least one of the baseline surveys.[15]

Ontology Mapping & Index Construction

- The initial step in index construction was domain selection. which was done through extraction of identified domains from recent systematic reviews and grey literature on area-based socioeconomic deprivation indices.[16,17]
- This study employed weighted multiple correspondence analysis, which is a well validated machine learning methodology that acts as a categorical corollary to primary component analysis. Raw scores were converted to quintiles

Figure 1: Inertia (variance) explained by each component. The weight of each relevant OMOP PPI concept used in combining socioeconomic domains was calculated using multiple correspondence analysis.

Methodology

Figure 2. Ontology mapping framework between relevant SDOH domains (the Deprivation Ontology) and *All of Us Survey Artifacts. The final development will use the Delphi method to build consensus.*

Validation

- First, correlation between the ISDI and the Brokamp ADI was conducted.
- Second, this study reconstructed an *All of Us* demonstration project that trained machine learning models adapted from a highly cited study that found racial bias in health algorithms[18].

This study assessed whether regularizing the data using stratified sampling based on ISDI quintile changes the model's performance and accuracy.

Results

- Limiting to participants who completed survey items yielded n=40,027 participants. The distribution of raw ISDI scores was then split into quintiles (5 being the highest deprivation).
- The correlation between the ISDI and the Brokamp ADI assessed at relatively large geographic area approximation was weak. Unfortunately, the three-digit zip code is a coarse measure. Finer area approximations may yield a stronger correlation, further validation may be explored to this extent.

Figure 3. Raw distribution of ISDI before binning into quintiles.

Table 1. Select breakdown of counts and percentage of low and high deprivation per survey artifact response.

	Quintile 1 (Low Deprivation)		Quintile 5 (High Deprivation)	
Deprivation Category	Count (n)	Percent (%)	Count (n)	Percent (%)
Highest Education Attainment				
Advanced Degree	4849	58.5	1891	23.6
College Graduate	3446	41.5	2217	27.7
College 1-3 Years	0	0	1022	12.7
Grades 5-8	0	0	52	0.6
Grades 1-4	0	0	<30	-
Financial Stability & Pressure				
No	8295	100	5781	72.1
Yes	<30	-	2237	27.9
Housing Status				
Own Home	8295	100	2869	35.8
Rent Home	<30	-	5149	64.2
Other Language				
Yes	0	0	2191	27.3
No	8295	0	5827	72.7
Transportation				
Delayed Care Due to Transportation: Yes	0	0	1578	19.7
Delayed Care Due to Transportation: No	8295	100	6440	80.3

Results

- The last component of the analysis was the re-creation of an *All of Us* project derived from a highly cited study regarding racial bias in health algorithms[18].
- The machine learning models were first trained normally using the standard train-test split. Then these same models were also trained with ISDI normalization (in this case: stratified sampling based on ISDI quintile).

	Pre ISDI Normalization		Post ISDI Normalization	
L2 Regularization	*LR Accuracy*	0.983	*LR Accuracy*	0.983
	LR AUC	0.581	*LR AUC*	0.562
No Regularization	*LR Accuracy*	0.983	*LR Accuracy*	0.983
	LR AUC	0.610	*LR AUC*	0.559
L2 Regularization with Varied Penalty Strength	*LR Accuracy*	0.983	*LR Accuracy*	0.983
	LR AUC	0.583	*LR AUC*	0.563
Random Forest	*RF Accuracy*	0.975	*RF Accuracy*	0.975
	RF AUC	0.887	*RF AUC*	0.841

Table 2. Comparison of Accuracy and Area Under the Curve (AUC) Pre- and Post-ISDI Normalization. Notably while accuracy is maintained, AUC decreases post normalization, which is aligned with the notion that addressing biases may lead to fairer models which may also be less discriminatory across demographic groups.

Figure 4. Comparison of the Pre- and Post- ISDI Normalization feature importance. Notably, although overall AUC was reduced, the importance of the certain features (e.g. 'Race: Black') went up.

Conclusion

- This study builds on a body of work around indexing complex SDOH factors into composite area based deprivation measures. It extends this work into the development of an individual socioeconomic deprivation index (ISDI) constructed on a heterogenous data network designed to recruit UBR populations and capture its data diversity. Such an approach to indexing may have numerous unanticipated use cases in the context of precision medicine.
- Finally, this approach may be valuable in the light of growing health data networks and data linkages. For instance, distributed machine learning approaches such as federated learning may hold significant potential for precision medicine at scale but need to address the heterogeneity of the SDOH profile of various health data sources (non-independent and identically distributed datasets)[19,20].

Acknowledgements & References

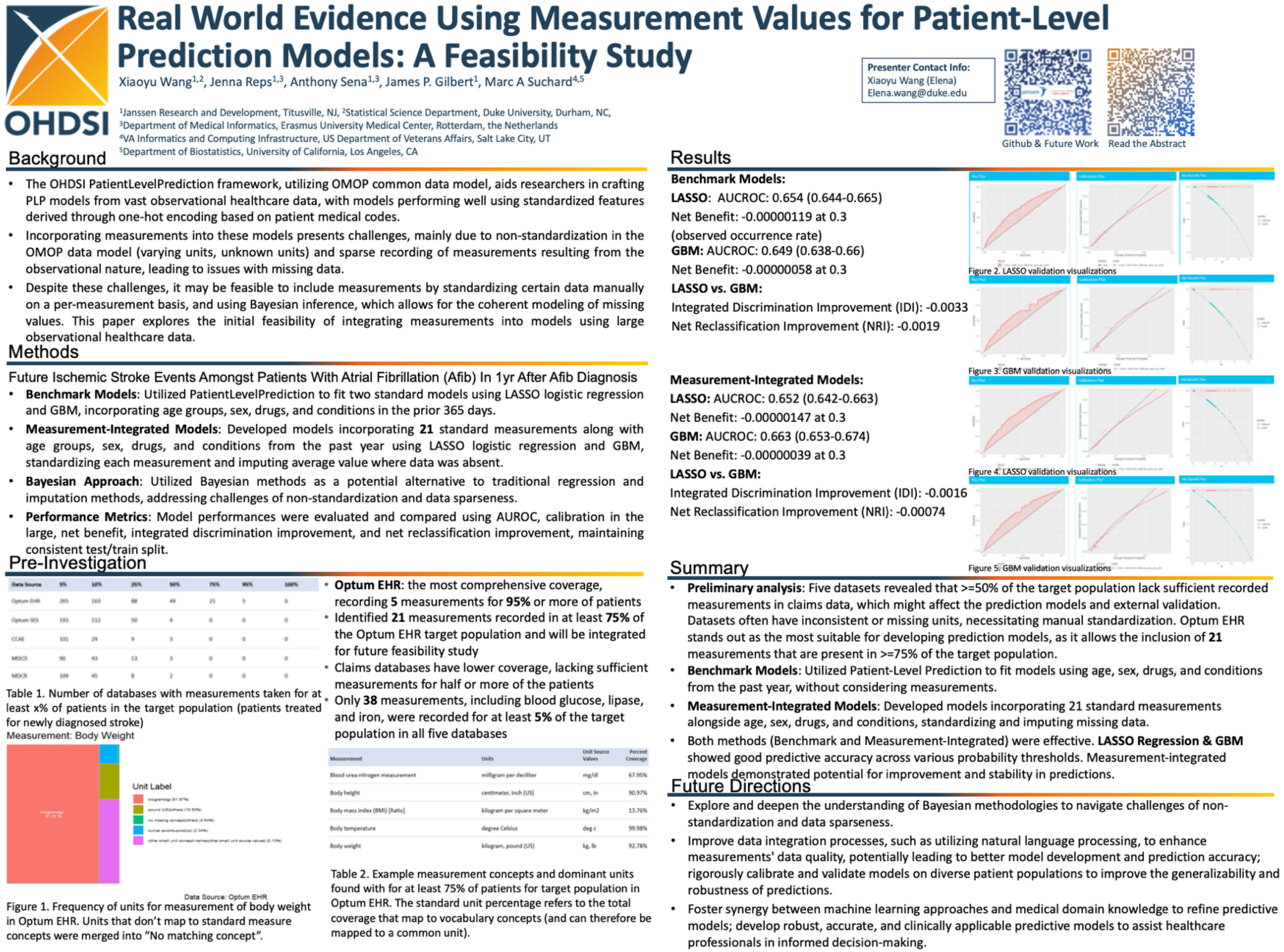
- We would like to acknowledge the *All of Us* Research Program and all of its participants. This work was supported by NIH OD 5U2COD023196

#OHDSISocialShowcase This Week

WEDNESDAY

Incorporating measurement values into patient-level prediction with missing entries: a feasibility study

(Xiaoyu Wang, Jenna Reps, Anthony Sena, James Gilbert, Marc Suchard)



Real World Evidence Using Measurement Values for Patient-Level Prediction Models: A Feasibility Study

Xiaoyu Wang[1,2], Jenna Reps[1,3], Anthony Sena[1,3], James P. Gilbert[1], Marc A Suchard[4,5]

[1]Janssen Research and Development, Titusville, NJ, [2]Statistical Science Department, Duke University, Durham, NC,
[3]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands
[4]VA Informatics and Computing Infrastructure, US Department of Veterans Affairs, Salt Lake City, UT
[5]Department of Biostatistics, University of California, Los Angeles, CA

Presenter Contact Info:
Xiaoyu Wang (Elena)
Elena.wang@duke.edu

Github & Future Work

Read the Abstract

Background

- The OHDSI PatientLevelPrediction framework, utilizing OMOP common data model, aids researchers in crafting PLP models from vast observational healthcare data, with models performing well using standardized features derived through one-hot encoding based on patient medical codes.
- Incorporating measurements into these models presents challenges, mainly due to non-standardization in the OMOP data model (varying units, unknown units) and sparse recording of measurements resulting from the observational nature, leading to issues with missing data.
- Despite these challenges, it may be feasible to include measurements by standardizing certain data manually on a per-measurement basis, and using Bayesian inference, which allows for the coherent modeling of missing values. This paper explores the initial feasibility of integrating measurements into models using large observational healthcare data.

Methods

Future Ischemic Stroke Events Amongst Patients With Atrial Fibrillation (Afib) In 1yr After Afib Diagnosis

- **Benchmark Models**: Utilized PatientLevelPrediction to fit two standard models using LASSO logistic regression and GBM, incorporating age groups, sex, drugs, and conditions in the prior 365 days.
- **Measurement-Integrated Models**: Developed models incorporating **21** standard measurements along with age groups, sex, drugs, and conditions from the past year using LASSO logistic regression and GBM, standardizing each measurement and imputing average value where data was absent.
- **Bayesian Approach**: Utilized Bayesian methods as a potential alternative to traditional regression and imputation methods, addressing challenges of non-standardization and data sparseness.
- **Performance Metrics**: Model performances were evaluated and compared using AUROC, calibration in the large, net benefit, integrated discrimination improvement, and net reclassification improvement, maintaining consistent test/train split.

Pre-Investigation

Data Source	5%	10%	25%	50%	75%	95%	100%
Optum EHR	265	163	88	49	21	5	0
Optum SES	193	132	50	4	0	0	0
CCAE	101	29	9	3	0	0	0
MDCD	90	43	13	3	0	0	0
MDCR	109	45	8	2	0	0	0

Table 1. Number of databases with measurements taken for at least x% of patients in the target population (patients treated for newly diagnosed stroke)

Measurement: Body Weight

Unit Label

Data Source: Optum EHR

Figure 1. Frequency of units for measurement of body weight in Optum EHR. Units that don't map to standard measure concepts were merged into "No matching concept".

- **Optum EHR**: the most comprehensive coverage, recording **5** measurements for **95%** or more of patients
- Identified **21** measurements recorded in at least **75%** of the Optum EHR target population and will be integrated for future feasibility study
- Claims databases have lower coverage, lacking sufficient measurements for half or more of the patients
- Only **38** measurements, including blood glucose, lipase, and iron, were recorded for at least **5%** of the target population in all five databases

Measurement	Units	Unit Source Values	Percent Coverage
Blood urea nitrogen measurement	milligram per deciliter	mg/dl	67.95%
Body height	centimeter, inch (US)	cm, in	90.97%
Body mass index (BMI) [Ratio]	kilogram per square meter	kg/m2	13.76%
Body temperature	degree Celsius	deg c	99.98%
Body weight	kilogram, pound (US)	kg, lb	92.78%

Table 2. Example measurement concepts and dominant units found with for at least 75% of patients for target population in Optum EHR. The standard unit percentage refers to the total coverage that map to vocabulary concepts (and can therefore be mapped to a common unit).

Results

Benchmark Models:

LASSO: AUCROC: 0.654 (0.644-0.665)

Net Benefit: -0.00000119 at 0.3

(observed occurrence rate)

GBM: AUCROC: 0.649 (0.638-0.66)

Net Benefit: -0.00000058 at 0.3

LASSO vs. GBM:

Integrated Discrimination Improvement (IDI): -0.0033

Net Reclassification Improvement (NRI): -0.0019

Figure 2. LASSO validation visualizations

Figure 3. GBM validation visualizations

Measurement-Integrated Models:

LASSO: AUCROC: 0.652 (0.642-0.663)

Net Benefit: -0.00000147 at 0.3

GBM: AUCROC: 0.663 (0.653-0.674)

Net Benefit: -0.00000039 at 0.3

LASSO vs. GBM:

Integrated Discrimination Improvement (IDI): -0.0016

Net Reclassification Improvement (NRI): -0.00074

Figure 4. LASSO validation visualizations

Figure 5. GBM validation visualizations

Summary

- **Preliminary analysis**: Five datasets revealed that >=50% of the target population lack sufficient recorded measurements in claims data, which might affect the prediction models and external validation. Datasets often have inconsistent or missing units, necessitating manual standardization. Optum EHR stands out as the most suitable for developing prediction models, as it allows the inclusion of **21** measurements that are present in >=75% of the target population.
- **Benchmark Models**: Utilized Patient-Level Prediction to fit models using age, sex, drugs, and conditions from the past year, without considering measurements.
- **Measurement-Integrated Models**: Developed models incorporating 21 standard measurements alongside age, sex, drugs, and conditions, standardizing and imputing missing data.
- Both methods (Benchmark and Measurement-Integrated) were effective. **LASSO Regression & GBM** showed good predictive accuracy across various probability thresholds. Measurement-integrated models demonstrated potential for improvement and stability in predictions.

Future Directions

- Explore and deepen the understanding of Bayesian methodologies to navigate challenges of non-standardization and data sparseness.
- Improve data integration processes, such as utilizing natural language processing, to enhance measurements' data quality, potentially leading to better model development and prediction accuracy; rigorously calibrate and validate models on diverse patient populations to improve the generalizability and robustness of predictions.
- Foster synergy between machine learning approaches and medical domain knowledge to refine predictive models; develop robust, accurate, and clinically applicable predictive models to assist healthcare professionals in informed decision-making.

#OHDSISocialShowcase This Week

THURSDAY

The Use of the Julia Programming Language for Global Health Informatics and Observational Health Research

(Jacob Zelko, Varshini Chinta, Malina Hy, Fareeda Abdelazeez)

The Use of the Julia Programming Language for Global Health Informatics and Observational Health Research

PRESENTER: Jacob S. Zelko

Northeastern University
Observational Health Data Sciences and Informatics Center

Introduction:

Performing real world data based studies continues to be increasingly vital for a variety of domains (1, 2). The OHDSI HADES ecosystem provides invaluable research tools written in languages such as R and Java that critical to such studies (3). Presented here is an emerging compliment to these tools created in the Julia language. With its focus on high performance computing, composability, and expressivity, Julia tools presents a potential avenue to rapidly process massive amounts of real world data.

Methods:

This poster describes a workflow to conduct an observational health research using tools from within the Julia sub-ecosystems alongside HADES tools:

JuliaHealth Ecosystem Tools:

- HealthSampleData.jl – Sample health data for a variety of health formats and use cases
- OHDSICohortExpressions.jl – OHDSI phenotype definition parser
- OMOPCDMCohortCreator.jl – Create cohorts from databases utilizing the OMOP CDM
- OHDSIAPI.jl – Julia interface to a variety of OHDSI web or API-based services

General Julia Ecosystem Tools:

- DBConnector.jl – Simplified interface that builds on Julia database packages adhering to DBInterface
- JSON3.jl – Julia JSON package focused on speed and slick struct mapping
- DataFrames.jl – Tools for working with tabular data in Julia
- FunSQL.jl – Julia library for compositional construction of SQL queries

Sample Phenotype Definition:

Definition: Simplified strep throat definition with the following characteristics:

Cohort Entry Events: People enter the cohort when observing any of the following:

- Condition occurrences of '[jz] Strep Throat Concepts'.

Cohort Exit: The person exits the cohort at the end of continuous observation.

Cohort Eras: Entry events will be combined into cohort eras if they are within 0 days of each other.

Concept Set Definition:

ID	NAME	DOMAIN	VOCAB	EXCL.	DESC.	MAP.
28060	Streptococcal sore throat	Condition	SNOMED	No	No	No

Preparing Synthetic OMOP CDM Data for Analysis 🤖

```
import HealthSampleData: Eunomia

Eunomia = Eunomia()
```

Step 1: Download and load synthetic OMOP CDM patient data set, *Eunomia*.

```
import DBConnector: DBConnection

conn = DBConnection(
    "sqlite", # SQL dialect
    db_path = eunomia # DB Location
)
```

Step 2: Create connection to SQLite instance of *Eunomia*.

```
import OMOPCDMCohortCreator as occ

occ.GenerateDatabaseDetails(
    :sqlite, # SQL dialect
    "main" # Schema name
)

occ.GenerateTables(conn)
```

Step 3: Create additional connection details

Finding Patients Who Have Had Strep Throat 🤒

```
import OHDSIAPI: ATLAS

using JSON3

cohort = ATLAS.get_atlas_cohort_definition("1787567")
  ▷ JSON3.read

cohort_expression = cohort[:items][1][:expression]
```

Step 1: Query ATLAS to get strep throat phenotype definition and read in cohort expression as JSON.

```
import DBInterface as DBI
import OHDSICohortExpressions: translate, Model

model = Model(cdm_version = v"5.3.1",
       cdm_schema = "main",
       vocabulary_schema = "main",
       results_schema = "main",
       target_schema = "main",
       target_table = "cohort"
)
```

Step 2: Staging creation of cohort tables for the OMOP CDM

```
sql = translate(cohort_expression,
       dialect = :sqlite,
       model = model,
       cohort_definition_id = 1
)
```

Step 3: Generating SQL statements for generating strep throat cohorts

```
for q in split(sql, ";")[1:end - 1]
       DBI.execute(conn, q)
end
```

Step 4: Executing SQL statements to build strep throat cohort

Characterizing Patients That Have Had Strep Throat

```
strep_patients = occ.GetPatientCohort(
       1,
       conn
).person_id
```

Step 1: Pull strep throat patients from cohort table

```
strep_patients_age_group = occ.GetPatientAgeGroup(
       strep_patients,
       conn
)

strep_patients_race = occ.GetPatientRace(
       strep_patients,
       conn
)

strep_patients_gender = occ.GetPatientGender(
       strep_patients,
       conn
)
```

```
import DataFrames as DF

strep_patients_characterized = DF.outerjoin(strep_patients_race,
                                 strep_patients_gender,
                                 strep_patients_age_group;
                                 on = :person_id,
                                 matchmissing = :equal)
```

Calculating Crude Prevalence Rates

```
strep_patients_characterized = strep_patients_characterized[:, DF.Not(:person_id)]

strep_patient_groups = DF.groupby(
       strep_patients_characterized,
       [
              :race_concept_id,
              :gender_concept_id,
              :age_group
       ]
)

strep_patient_groups = DF.combine(strep_patient_groups, DF.nrow => :count)
```

Final Results

```
audited_strep_df = occ.ExecuteAudit(strep_df; hitech = true)
```

Discussion:

FunSQL

Age groupings

[[b, b + 4] for in 0:5:119]

Next Steps:

Key Takeaways, Summarized:

Acknowledgments:

References:

Leeroy Jenkins, author2, author3, author4, author5, author6, author7, author42

OHDSI

#OHDSISocialShowcase This Week

FRIDAY

Analyzing a Tabloid Headline with Real-World Data: A Summer Intern's Investigation

(Delia Harms, Kristin Kostka)

***Analyzing a Tabloid Headline with Real-World Data:** A Summer Intern's Investigations*

PRESENTER: **Delia** Harms

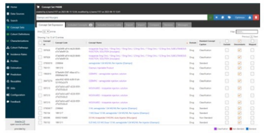
INTRO:

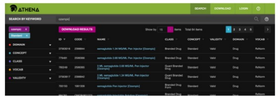
- **Headlines about new weight loss drugs** have been all over our computers recently. What are the real usage rates of these drugs and what populations are using them? Are they being used by the targeted populations?

What Is Ozempic and Why Is It Getting So Much Attention?
More people are turning to a diabetes medication to induce weight loss — but experts say it's not a miracle drug.

METHODS

1. **Environment:** Northeastern's OHDSI Lab (see poster #202) + ATHENA
2. **OMOP Database:** IQVIA PharMetrics Plus for Academics
3. **Study Population:** Data collected from 2017 - 2023
4. **Target Cohorts:** Two cohorts were compared, both containing persons aged 18 and over and without indication of Type II Diabetes. One cohort contained users of Ozempic or Mounjaro, which are approved for weight loss, while the other contained users of Saxenda or Wegovy, which are approved only for people with Type II Diabetes.
5. **Analysis:** Ozempic/Mounjaro – **69%** of users did not have a record of Type II Diabetes. Wegovy/Saxenda – **15.25%** of users had no record of Type II **Diabetes.**



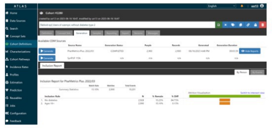


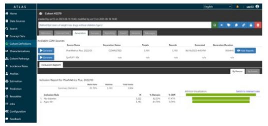
With **misinformation** on the rise, OHDSI tools help us **interrogate headlines** and understand **real-world prevalence** rates.

Take a picture to see training opportunities at the Roux Institute

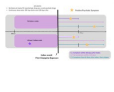
How it started...

RESULTS





How it went!



Cohort	Total Members	Percent Relapsed
Tobacco Users	432	13.7%
Tobacco Non-Users	1448	9.3%

Partnered with Psychiatric Nurse Practitioner to study the relationship between clozapine and tobacco use

Saw some OHDSI legends at AIPM 2023.

Connected with interns in other fields.

Delivered a talk on my undergrad summer experience with OHDSI.

Delia Harms, Kristin Kostka

Massachusetts Institute of Technology

Northeastern University Observational Health Data Sciences and Informatics Center

OHDSI

Opening: Three Positions at Gilead

About Us

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.

Sr. Director, Head of Data Office

Apply

Job Description:
As a Senior Director in our Data Office, you will play a pivotal role in shaping and executing our data strategy. In this leadership position, you will oversee and drive activities related to data sharing, governance, and access across the organization. Working closely with cross-functional teams, you will define and implement data acquisition policies and practices, ensuring the efficient and effective use of data to support our scientific and business objectives.

Director, Data Acquisition - Clinical Data Science

Apply

Director, Data Acquisition - Clinical Data Science

This role reports to the Head of Gilead data office, RWE Generation, Clinical Data Science and is based at different Gilead sites. This individual has responsibility for acquiring all data across clinical, development, medical affairs function and Gilead affiliates. This individual will work in close collaboration with the Development organization, Commercial, Procurement, Medical Affairs, IT, and other functions at Gilead in implementing data acquisition processes and is expected to operate with a "one Gilead" mindset & play a key role in the global Gilead Data Office set up.

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.

Postdoc/Senior Data Analyst Opening at WashU

PI: Linying Zhang, PhD

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.

- 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

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

Opening: Research Information Specialist at UNC

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

- Home
- Search Jobs
- Careers At Carolina Notifications
- Log In /Create Account
- ? Help
- Working at Carolina

Research Informatics Specialist

Bookmark this Posting | Print Preview | Apply for this Job

Please see Special Instructions for more details.

Working hours are Monday-Friday, 8:00 am – 6:00 pm EST with flexibility available within that window.

Posting Information

Posting Information

Department	TraCS Institute-429801
Career Area	Information Technology
Posting Open Date	12/13/2023
Application Deadline	01/30/2024
Open Until Filled	No
Position Type	Permanent Staff (EHRA NF)
Working Title	Research Informatics Specialist
Appointment Type	EHRA Non-Faculty
Position Number	20060002
Vacancy ID	NF0007640
Full Time/Part Time	Full-Time Permanent
FTE	1

Position Summary

Responsibilities include:
* Perform SQL-based programming against UNC's clinical data warehouse to identify patient cohorts and develop patient datasets.
* Consult with and collaborate with researchers to ensure programming work aligns with project needs.
* Develop ETL (extract, transform, and load) and data integration processes to support common data models (OMOP, PCORnet) using appropriate technologies (SQL, Python, or R).
* Carefully following UNC's regulatory and governance policy to ensure data integrity and security.
* In collaboration with IDSci team, identify potential enhancements in current workflows and data architecture.
* Implement quality assurance strategies, such as data validation and peer code review.
* Write and maintain up-to-date supporting documentation. Ensure code is well-commented and use GitLab/GitHub to manage code changes and track data lineage.
* Provide technical leadership and direction for assigned projects and/or data requests.

Minimum Education and Experience Requirements

Master's and 1-2 years' experience; or Bachelors and 2-4 years' experience; or will accept a combination of related education and experience in substitution.

Required Qualifications, Competencies, and Experience

This position requires two or more years of relevant work experience and:

* Expert-level knowledge of SQL programming, data modeling, and relational database systems such as Oracle, Microsoft SQL Server, MySQL, etc.
* Past experience working with health care data in an analytic capacity, particularly electronic health record and/or claims data.
* Demonstrable past experience in scoping technical projects in terms of length of time, competencies and cost. Individual will be expected to manage multiple projects at once while delivering high-quality work on time.
* Excellent written and oral business communication skills. Public speaking at meetings and conferences may be required. The ability to clearly convey technical concepts to non-technical clients is a must.

Opening: Data Steward at EBMD

Description

Are you looking for a job where you can make a difference and work in a non-profit?
Would you like to be a part of an ambitious and international organisation on the cutting edge of science?
Then this position might be right up your alley.

The EBMT is a non-profit medical and scientific organisation which hosts a unique patient registry providing a pool of data to perform studies and assess new trends.

OUR MISSION
Save and improve the lives of patients with blood-related disorders.

The Registry
Holding the **data of over half a million patients**, the EBMT registry is the **starting point for all studies** carried out through the EBMT working parties. The department focuses on data collection processes, data quality monitoring, and maintenance of the database.

YOUR MISSION
Responsible for collecting, collating, and evaluating issues and problems with data and enforcing data usage policies.

RESPONSIBILITIES AND TASKS

Data Stewardship:
- Design, implementation and testing of new data collection processes including data collection forms (DCFs) development.
- Take care of the mapping of new items from DCFs to the OMOP CDM
- Providing input on data quality reports
- Check and clean data on request and ad hoc.
- Data retrieval including designing data reports and data report running.
- Carry out computerized system validation activities.
- Supporting consolidation/harmonization of data
- Creating standard data definitions, and maintain a consistent use of data assets across the organization
- Documenting data policies and data standards

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?

Learn more about all of the OHDSI workgroups

ohdsi.org/workgroups

Common Data Model Workgroup

2024 OKR Update

Purpose

The CDM workgroup exists to maintain and improve the use of the OMOP Common Data Model to make it the premier observational health data model in the world. We ensure the integrity and usability of the OMOP CDM in relation to other working groups by providing guidance on data standardization best practices.

2024 Objectives and Key Results

Objective 1: Facilitate collaboration and alignment between the CDM and other OHDSI working groups

- Host a hack-a-thon to collaborate with THEMIS, DQD, and Vocabulary WGs, aligning the community on data standards, conventions, and evaluation
  - Clarify data standardization best practices and share with other workgroups
  - Document prior decisions made by the CDM WG

2024 Objectives and Key Results

Objective 2: Make the OMOP CDM the premier observation health data model by reducing technical debt and improving documentation

- Get the CDM package onto CRAN
- Clean up existing documentation and remove outdated documentation
  - Document the STEM table and clarify its usage
  - Remove CDM v6.0 from website
  - Write down add-on, extension, expansion information
  - Write down our maturity model

CRAN Achieved!

CommonDataModel: OMOP CDM DDL and Documentation Generator

Generates the scripts required to create an Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) database and associated documentation for written in parameterized Structured Query Language (SQL) to the other supported dialects.

Version:	0.2.0
Depends:	DatabaseConnector, SqlRender, rJava
Imports:	rmarkdown, stringr, DBI, dplyr, readr
Suggests:	knitr, testthat (≥ 3.0.0), RSQLite, withr
Published:	2024-02-07
Author:	Clair Blacketer [aut, cre]
Maintainer:	Clair Blacketer <mblacke at its.jnj.com>
License:	Apache License 2.0
NeedsCompilation:	no
Materials:	README
CRAN checks:	CommonDataModel results

Documentation:

Reference manual: CommonDataModel.pdf

Downloads:

Package source:	CommonDataModel_0.2.0.tar.gz
Windows binaries:	r-devel: CommonDataModel_0.2.0.zip, r-release: CommonDataModel_0.2.0.zip, r-oldrel: CommonDataModel_0.2.0.zip
macOS binaries:	r-release (arm64): CommonDataModel_0.2.0.tgz, r-oldrel (arm64): CommonDataModel_0.2.0.tgz, r-release (x86_64): CommonDataModel_0.2.0.tgz

Linking:

Please use the canonical form `https://CRAN.R-project.org/package=CommonDataModel` to link to this page.

Network Data Quality Workgroup

2024 OKR Update

Purpose

The Network Data Quality workgroup exists to recommend, enable, and develop best practices related to observational data quality at the level of a federated network.

2024 Objectives and Key Results

Objective 1: Improve Data Quality reporting for the OHDSI Community

- Complete information pages for all check types in the DQD – Q1
- Create a new Data Quality report that informs the user on how to interpret and remediate failing DQD checks
  - Draft of report in Q1,
  - Inform and work on refactor in Q2
- Create at least 1 new DQD check as identified by THEMIS

2024 Objectives and Key Results

Objective 2: Support and collaborate with THEMIS and CDM working groups

- Specify the requirements necessary to make THEMIS conventions assessable and reportable as data quality checks
- Create at least 1 new DQD check as identified by THEMIS

2024 Objectives and Key Results

Objective 3: Refine the approach for quantitatively assessing a OHDSI Network datasets' fitness for specific study questions.

- Define "fitness for use" in the context of an OHDSI network study
- Conduct an assessment of data diagnostics and provide a report of potential improvements (Q1)
- Prioritize improvements and implement the top prioritized features to data diagnostics (Q2 and beyond)

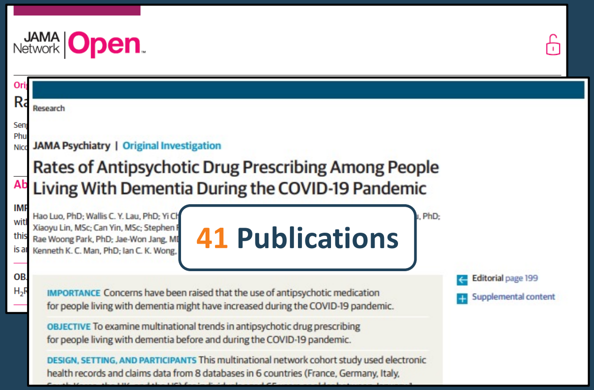
OHDSI APAC OKRs

Mui Van Zandt

2023 OHDSI APAC Key Results

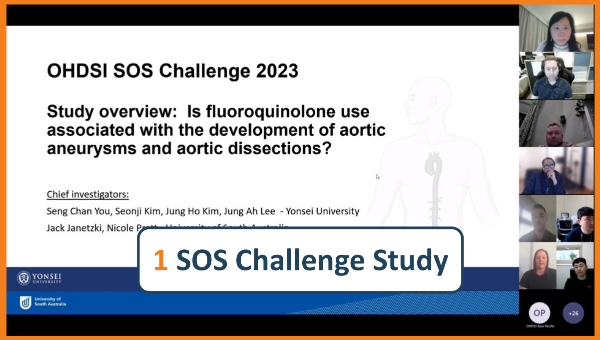
Research

Build research expertise and collaboration amongst the different chapters through publication



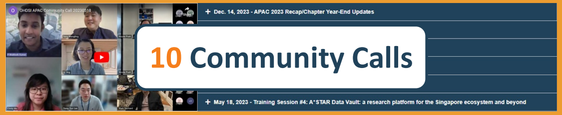
Training

Create an APAC training program to expand reach to the general community



Communication

Create collaboration activities that encourage collaborative generation and dissemination of evidence that promotes better health decisions and better care



2024 OHDSI APAC Goals

Research

Build research expertise and collaboration amongst the different chapters through publication

Milestones

- Conduct APAC SOS Challenge studies
- Replicate Cindy Kai's SOS Challenge study

Training

Create an APAC training program to expand reach to the general community

Milestones

- Host at least 2 in-person trainings in APAC
- Train community through APAC SOS Challenge studies

Communication

Create collaboration activities that encourage collaborative generation and dissemination of evidence that promotes better health decisions and better care

Milestones

- Host APAC symposium
- Distribute quarterly newsletters
- Host monthly community calls and scientific forums

Join Us

[Sign up](#) for the OHDSI APAC WG! ☑ OHDSI APAC

- APAC Community Calls
  - Every <u>third</u> Thursday, 12 p.m. Korea time

Date	Topic
Jan 18	APAC 2024 Kickoff
	Training Session #7 by Japan
	Eye Care and Vision Research WG Intro
Feb 15	New WG Intro: Evidence Translation, Industry
Mar 21	Vocabulary Contribution by Korea
	OHDSI Evidence Network
Apr 18	Newcomers Session

[Direct link to community calls](#)

- APAC Scientific Forum
  - Every <u>first</u> Thursday, 12 p.m. Korea time

Date	Topic
Feb 1	Survey Results and Plans for 2024
	APAC Study Updates
Mar 7	Perseus Intro & Demo
Apr 4	Genomic Data Mapping

[Direct link to scientific forums](#)

OHDSI Industry Working Group OKRs

Established: October 21st, 2023

OHDSI Industry Work Group

Who should attend?

This is an open meeting with a focus on those members of the OHDSI community who have ties and affiliations with the Pharma and Biotech industries and would like to work together to represent those interests more broadly within OHDSI.

- Foster a stronger collaboration between the life science, pharma, and biotech industries, and the OHDSI community.
- Identify and develop strategies to encourage the active participation of these industries in OHDSI studies and initiatives.
- Facilitate knowledge transfer, sharing industry expertise and learnings with the broader OHDSI community.
- Identify opportunities for mutual support, leveraging industry resources and capabilities to advance OHDSI's goals.
- Increase the visibility and understanding of OHDSI's initiatives within these industries, promoting active involvement and commitment.

2024 OHDSI Industry Working Group Goals

Data Marketplace

Develop OMOP data marketplace and supporting framework

Go-To-Market

OHDSI advocacy group

Collaboration

Use case identification and OHDSI working group partnerships

Sign up for the OHDSI Industry WG!

2024 OHDSI Industry Working Group Goals

Data Marketplace

Develop OMOP data marketplace and supporting framework

Milestones

- Create a catalogue of OMOP datasets that are open for industry sponsored studies
- Develop framework for interacting with the marketplace

Go-To-Market

Design structure and purpose of OHDSI advocacy group and ensure interoperability

Milestones

- Design structure and purpose of OHDSI 'advocacy group'
- Create recommendations for OHDSI/OMOP models to be regulatory/governmentally aligned – ISO standards etc

Collaboration

Collaboration with other working groups to build use cases specific to industry

Milestones

- Identify 2-3 use cases
- Identify work group partnerships for each use case

Eye Care and Vision Research Workgroup

- **Workgroup purpose:** The purpose of the Eye Care and Vision Research Workgroup is to advance the development and implementation of data standards in ophthalmology, optometry, and the vision sciences, and to support studies using observational ophthalmic data for generating insights to improve health and vision outcomes.

- **Workgroup past accomplishments:**
  - Published a gap analysis of two large, well-known EHR systems for eye care (Epic and Cerner).
  - In addition to standing monthly meetings, organized in-person meetings at major conferences.
  - Organized additional subgroups: retina, glaucoma, pediatrics, uveitis, imaging and ETLs.
  - Collaborated with Verana Health for OMOP transformation of the AAO IRIS Registry.
  - Partnered with the NIH Bridge2AI AI-READI project to map ophthalmic data elements; pilot public release in spring 2024.
  - Submitted retinal condition codes to SNOMED International.
  - Submitted glaucoma examination codes to SNOMED International.
  - Submitted uveitis phenotypes to HowOften.
  - Supported SOS Challenge project examining the risk of kidney injury associated with anti-VEGF.
  - Engaged with LOINC to develop framework for representing visual acuity data.
  - Started working on ETLs of ophthalmic data at several participating sites.

Eye Care and Vision Research Workgroup OKRs

Objective 1: Continue advancing data standards development around specific use cases

Key Result 1: Build upon prior success with developing tonometry-related concepts with the glaucoma subgroup and advance representation of additional concepts relevant for glaucoma research, including gonioscopy-related concepts and visual field concepts. Timeline – end of Q2 2024.

Key Result 2: Submit diabetic retinopathy phenotype-related concepts from the retina subgroup to SNOMED for subsequent incorporation into the CDM. Timeline – end of Q1 2024.

Key Result 3: Contribute to public release of pilot data from the AI-READI Bridge2AI project, which includes ophthalmic data element mapped to standard OMOP concepts. Timeline – end of Q2 2024.

Objective 2: Map common ophthalmic data elements at multiple institutions

Key Result 1: Submit visual acuity codes to LOINC using panel approach. Timeline – end of Q1 2024.

Key Result 2: Trial ETL processes at 3 institutions for visual acuity data. Timeline – end of Q4 2024.

Key Result 3: Trial ETL processes at 3 institutions for IOP data. Timeline – end of Q2 2024.

Objective 3: Develop long-term sustainability to workgroup efforts.

Key Result 1: Organize grant-writing committee to plan proposals. Timeline – end of Q2 2024.

Key Result 2: Submit grant for funding data network. Timeline – end of Q4 2024.

Surgery and Perioperative WG

Objectives and key results 2024

Feb 13 2024

OKR Themes

- WG Growth and Processes
- Cohorts & Characterization
- Community Events / Evidence Generation
- Perioperative Prediction

OKR Theme: WG Growth and Processes

- Strategic WG Growth
- Key Results
  - Involvement of (at least) 3 new members from (at least 3 different) surgical / perioperative science focused lab groups
  - Presentation of WG supported work in at least one surgical / perioperative medicine conference.
  - Presentation of broader OHDS mission and capabilities within at least one surgical / perioperative conference.
  - Establishing 2 strategic collaborations with other WGs: 2 joint meetings during 2024

Theme: Cohorts & Characterization

- Completion of HowOften Incidence Rate Characteriztion Studies
  - Key Results: Submission for publication
    - Surgical cohorts against post operative outcomes of interest
    - CRC against post operative outcomes of interest
    - Surgical cohorts with post op afib, ischemic stroke.
- Completion of Fragility Fracture Study thon
  - Key Result: Submission for publication

Theme: Cohorts & Characterization

- Exploration of proxies for pre / post operative functional outcomes within the OHDSI network
  - Key Results:
    - Compile list of existing proxies within the vocabulary, determine use in the network
    - Survey WG / OHDSI network with respect to existence of proxy data at their site.

Theme: Community Events (Cohorts & Characterization)

- Support one surgical cohort hack-a-thon
  - Key Results:
    - Complete the extension of HowOften Surgical Cohorts into standard OHDSI vocabulary representations
    - Creation of at least 3 denovo surgical cohorts during the hackathon.
    - Initiate 3 members new to cohort building in OHDSI into the cohort Building process

Theme: Perioperative Prediction

- Execution of a Perioperative Prediction study (jointly with PLP workgroup) using the Major Non-Cardiac Surgery cohort; other surgical risk cohorts, and outcomes of post operative interest.
  - Key Results:
    - Generation of study design, and github page ready to launch study
    - Execution of PLP network study on at least 3 OHDSI data sources.
    - Preparation of at least 1 draft manuscript (Q4 2024)

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