Workgroup OKRs + Phenotype Phebruary Update #1

OHDSI Community Call
Feb. 6, 2024 • 11 am ET

Upcoming Community Calls

Date	Topic
Feb. 6	Workgroup OKRs / Phenotype Phebruary Update 1
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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Woo Yeon Park, Kyulee Jeon, Teri Sippel Schmidt, Haridimos Kondylakis, Tarik Alkasab, Blake E. Dewey, Seng Chan You & Paul Nagy** on the publication of **Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension** in the *Journal of Imaging Informatics in Medicine.*

Journal of Imaging Informatics in Medicine
https://doi.org/10.1007/s10278-024-00982-6

Check for updates

Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension

Woo Yeon Park[1] · Kyulee Jeon[2,3] · Teri Sippel Schmidt[1] · Haridimos Kondylakis[4] · Tarik Alkasab[5] · Blake E. Dewey[6] · Seng Chan You[2,3] · Paul Nagy[1]

Received: 4 September 2023 / Revised: 10 November 2023 / Accepted: 14 November 2023

Abstract
The rapid growth of artificial intelligence (AI) and deep learning techniques require access to large inter-institutional cohorts of data to enable the development of robust models, e.g., targeting the identification of disease biomarkers and quantifying disease progression and treatment efficacy. The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) has been designed to accommodate a harmonized representation of observational healthcare data. This study proposes the Medical Imaging CDM (MI-CDM) extension, adding two new tables and two vocabularies to the OMOP CDM to address the structural and semantic requirements to support imaging research. The tables provide the capabilities of linking DICOM data sources as well as tracking the provenance of imaging features derived from those images. The implementation of the extension enables phenotype definitions using imaging features and expanding standardized computable imaging biomarkers. This proposal offers a comprehensive and unified approach for conducting imaging research and outcome studies utilizing imaging features.

Keywords Data collection [MeSH] · Data standardization · Observational research · Data integration · Multimodal data analysis

OHDSI Shoutouts!

Collaborators from both the Columbia University Department of Biomedical Informatics and the Johnson & Johnson Observational Health Data Analytics team held a three-day studyathon this past weekend with a focus on women's health initiatives, specifically endometriosis and polycystic ovary syndrome.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

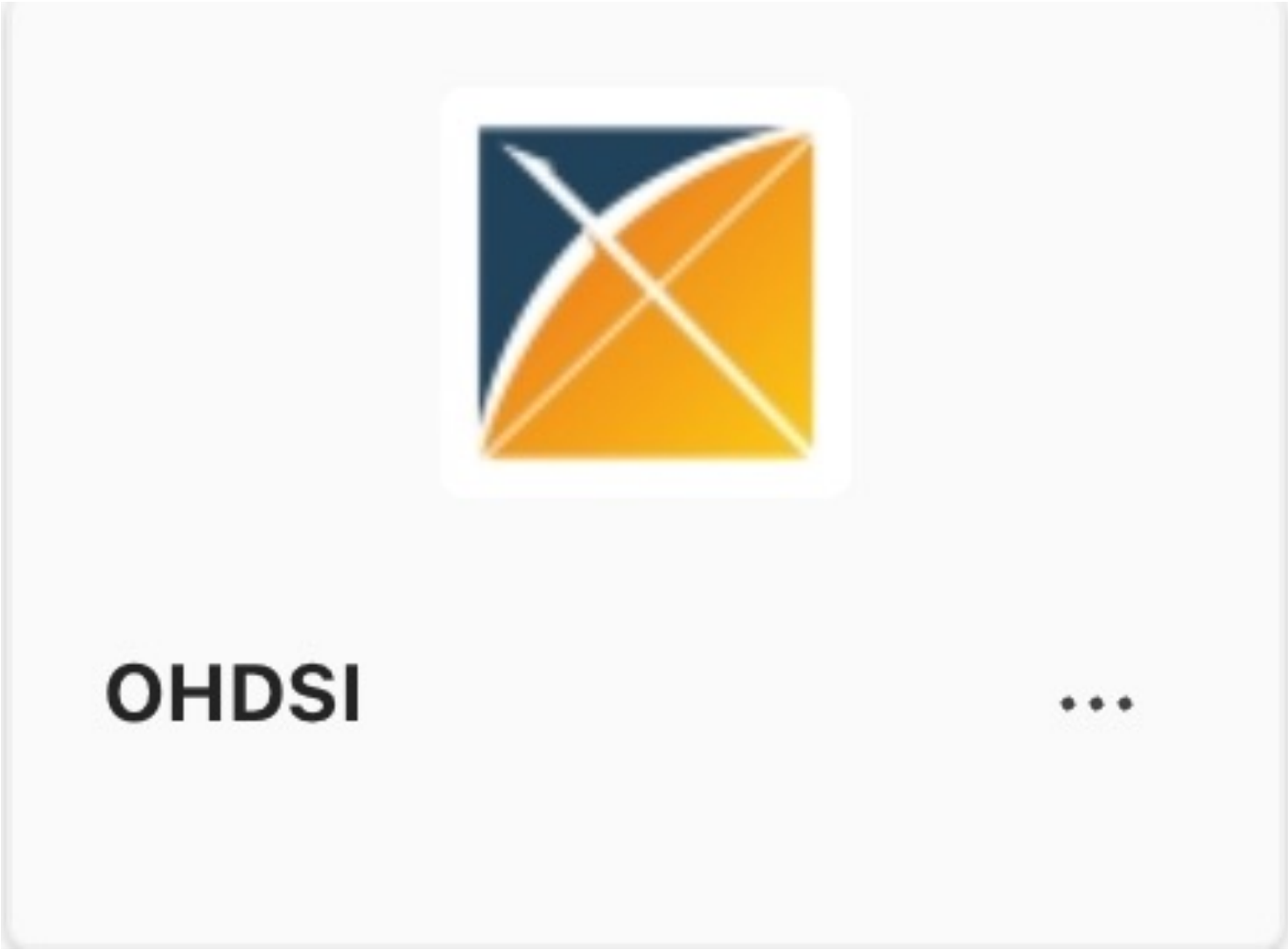
Date	Time (ET)	Meeting
Tuesday	12 pm	Generative AI and Analytics
Tuesday	12 pm	Common Data Model Vocabulary Subgroup
Wednesday	8 am	Psychiatry
Wednesday	3 pm	Vulcan/OHDSI Meeting (ZOOM)
Wednesday	7 pm	Medical Imaging
Thursday	9:30 am	Network Data Quality
Thursday	12 pm	Strategus Subgroup
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development & Evaluation
Friday	9 am	GIS – Geographic Information System
Friday	10 pm	China Chapter
Monday	10 am	Healthcare Systems Interest Group
Monday	11 am	Early-Stage Researchers
Monday	4 pm	Eyecare & Vision Research
Tuesday	9 am	OMOP CDM Oncology Genomic Subgroup

Join The Scientific Review Committee!

Would you be interested in helping shape #OHDSI2024?

We are actively seeking collaborators to join the scientific review committee. Please fill out the form in the chat or the community call page. Meetings will begin March 7, and the abstract review process will take place June 27-Aug. 9.

MS Teams Update



Spotlight: Kerry Goetz

Kerry Goetz is the Associate Director for the National Eye Institute's Office of Data Science and Health Informatics at the US National Institutes of Health. In this capacity she is responsible for advancing data management and sharing strategies to make NEI data FAIR (Fully AI-Ready & Findable, Accessible, Interoperable, and Reusable). For over a decade, Kerry has been leading the eyeGENE Program, a controlled access resource with imaging, data, samples, and a participant registry for rare eye conditions. Kerry has also been entrenched in standards development for over 15 years.

Kerry co-leads the Eye Care and Vision Research Observational Health Data Sciences and Informatics Working Group, is a member of the American Academy of Ophthalmology Standards Working Group, and also works to aligning imaging standards and health data to enable groundbreaking research. She has been attending OHDSI meetings for many years but didn't know how to truly get connected since she didn't have access to any OMOP'd data. The NIH clinical center operates in a much different capacity. However, after connecting with other like-minded collaborators like Sally Baxter and Michelle Hribar, there was momentum to create a Eye Care and Vision Health Working Group.

In Kerry's spare time, she enjoys traveling, snowboarding, camping, hiking, and biking and spending time with family, her two collies, or her Girl Scout Troop. She is also a PhD Candidate at George Mason University, studying Health Services Research with a Knowledge Discovery and Health Informatics Concentration. She discusses her career journey, evidence gaps around vision research, how OHDSI impacts her PhD journey, and more in the latest collaborator spotlight.

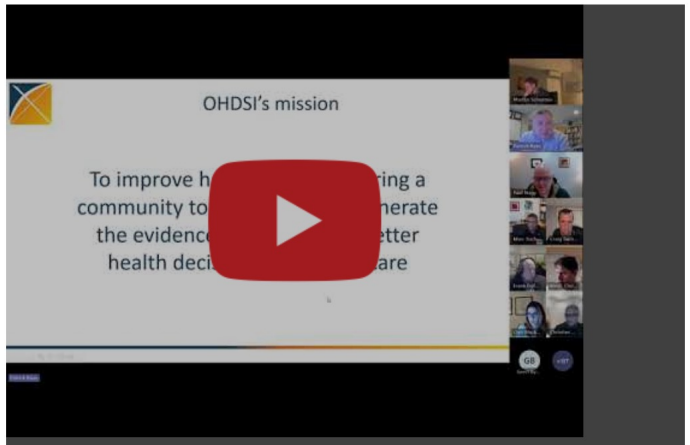
ohdsi.org/spotlight-kerry-goetz

February Newsletter is Available

The Journey Newsletter (February 2024)

The third year of Phenotype Phebruary has arrived! This newsletter looks at what OHDSI can achieve together in 2024, with a strong focus on evidence dissemination. We also preview Phenotype Phebruary and share how you can get involved. Check out more community updates, an impressive 16 publications related to OHDSI/OMOP, the latest community spotlight, and plenty more! **#JoinTheJourney**

Video: Evidence Dissemination Stands As Major Focus For Community in 2024



Patrick Ryan led the first community call of the year for a discussion on what OHDSI can achieve together in 2024.

Phenotype Phebruary 2024
A collaborative study

Goal: 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.

Month-long collaborative study focused on assessing consistency in :

- Phenotype definition Components
- Phenotype representation structure
- Phenotype validation methods

https://forums.ohdsi.org/t/ohdsi-phenotype-phebruary-2024/20940

Phenotype Phebruary Will Focus On Understanding Current Practices, Evaluating Four Community-Selected Phenotypes

"Phenotype Phebruary" is a community-wide initiative that began in 2022 and serves to both develop and evaluate phenotypes for health outcomes that could be investigated by the community.

The third edition of Phenotype Phebruary has begun, and the goals for this year's activity are to inspire community engagement and collaboration, advance the science of phenotyping, and educate/train on the process of phenotype development and evaluation. Anybody in the community, regardless of your background or focus, can positively impact our efforts by joining the activity (see sign-up link below).

During the Jan. 30 Introduction to Phenotype Phebruary call, the community voted on four phenotypes to focus on throughout the month (Alzheimer's, 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.

mailchi.mp/ohdsi/february2024

January Publications

Reich C, Ostropolets A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, Dymshyts D, Hripcsak G. OHDSI Standardized Vocabularies-a large-scale centralized reference ontology for international data harmonization. J Am Med Inform Assoc. 2024 Jan 4:ocad247. doi: 10.1093/jamia/ocad247. Epub ahead of print. PMID: 38175665.

Català M, Mercadé-Besora N, Kolde R, Trinh NTH, Roel E, Burn E, Rathod-Mistry T, Kostka K, Man WY, Delmestri A, Nordeng HME, Uusküla A, Duarte-Salles T, Prieto-Alhambra D, Jödicke AM. The effectiveness of COVID-19 vaccines to prevent long COVID symptoms: staggered cohort study of data from the UK, Spain, and Estonia. Lancet Respir Med. 2024 Jan 11:S2213-2600(23)00414-9. doi: 10.1016/S2213-2600(23)00414-9. Epub ahead of print. PMID: 38219763.

Khan H, Mosa ASM, Paka V, Rana MKZ, Mandhadi V, Islam S, Xu H, McClay JC, Sarker S, Rao P, Waitman LR. Mapping Clinical Documents to the *Logical Observation Identifiers, Names and Codes* (LOINC) Document Ontology using Electronic Health Record Systems Structured Metadata. AMIA Annu Symp Proc. 2024 Jan 11;2023:1017-1026. PMID: 38222329; PMCID: PMC10785913.

Zuo X, Zhou Y, Duke J, Hripcsak G, Shah N, Banda JM, Reeves R, Miller T, Waitman LR, Natarajan K, Xu H. Standardizing Multi-site Clinical Note Titles to LOINC Document Ontology: A Transformer-based Approach. AMIA Annu Symp Proc. 2024 Jan 11;2023:834-843. PMID: 38222429; PMCID: PMC10785935.

Hall ES, Melton GB, Payne PRO, Dorr DA, Vawdrey DK. How Are Leading Research Institutions Engaging with Data Sharing Tools and Programs? AMIA Annu Symp Proc. 2024 Jan 11;2023:397-406. PMID: 38222386; PMCID: PMC10785902.

Lyu T, Liang C. Computational Phenotyping of OMOP CDM Normalized EHR for Prenatal and Postpartum Episodes: An Informatics Framework and Clinical Implementation on All of Us. AMIA Annu Symp Proc. 2024 Jan 11;2023:1096-1104. PMID: 38222375; PMCID: PMC10785883.

HADES Development Updates: CohortMethod 5.2.1

CohortMethod 5.2.1 | Reference | Articles | Changelog | HADES

CohortMethod

R-CMD-check passing | codecov 89%

CohortMethod is part of HADES.

Introduction

CohortMethod is an R package for performing new-user cohort studies in an observational database in the OMOP Common Data Model.

Features

- Extracts the necessary data from a database in OMOP Common Data Model format.
- Uses a large set of covariates for both the propensity and outcome model, including for example all drugs, diagnoses, procedures, as well as age, comorbidity indexes, etc.
- Large scale regularized regression to fit the propensity and outcome models.
- Includes function for trimming, stratifying, matching, and weighting on propensity scores.
- Includes diagnostic functions, including propensity score distribution plots and plots showing covariate balance before and after matching and/or trimming.
- Supported outcome models are (conditional) logistic regression, (conditional) Poisson regression, and (conditional) Cox regression.

Links

Browse source code

Report a bug

Ask a question

License

Apache License 2.0

Citation

Citing CohortMethod

Developers

Martijn Schuemie
Author, maintainer

Marc Suchard
Author

Patrick Ryan
Author

HADES Development Updates: FeatureExtraction 3.4.0

FeatureExtraction 3.4.0 Reference Articles Changelog HADES

FeatureExtraction

R-CMD-check passing codecov 93%

FeatureExtraction is part of HADES.

Introduction

An R package for generating features (covariates) for a cohort using data in the Common Data Model.

Features

- Takes a cohort as input.
- Generates baseline features for that cohort.
- Default covariates include all drugs, diagnoses, procedures, as well as age, comorbidity indexes, etc.
- Support for creating custom covariates.
- Generate paper-ready summary table of select population characteristics.

Technology

FeatureExtraction is an R package, with some functions implemented in C++.

Links

Browse source code

Report a bug

Ask a question

License

Apache License 2.0

Citation

Citing FeatureExtraction

Developers

Martijn Schuemie
Author

Marc Suchard
Author

Patrick Ryan
Author

Jenna Reps
Author

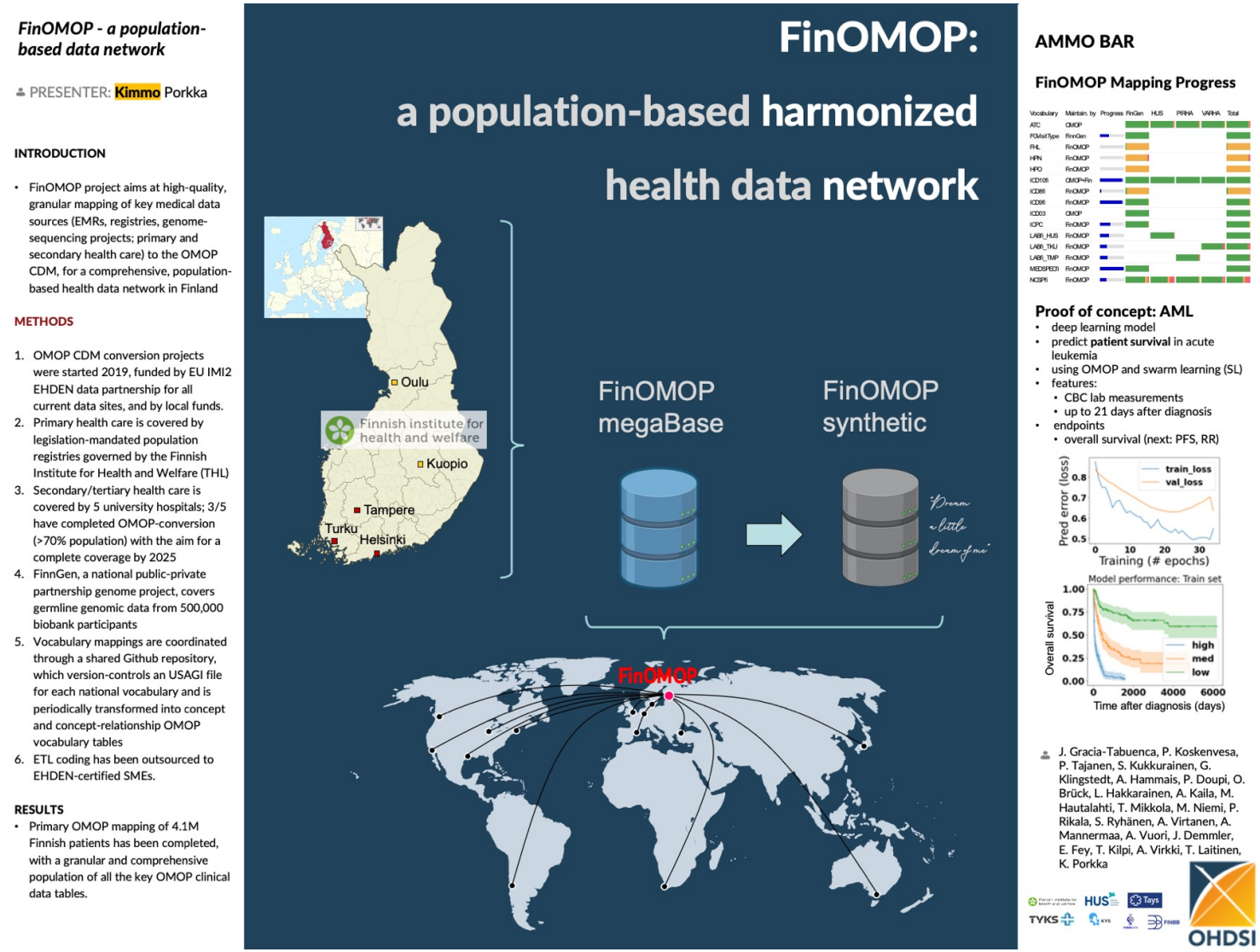
Anthony Sena

#OHDSISocialShowcase This Week

MONDAY

FinOMOP - a population-based data network

(Javier Gracia-Tabuenca, Perttu Koskenvesa, Pia Tajanen, Sampo Kukkurainen, Gustav Klingstedt, Anna Hammais, Persephone Doupi, Oscar Brück, Leena Hakkarainen, Annu Kaila, Marco Hautalahti, Toni Mikkola, Marianna Niemi, Pasi Rikala, Simo Ryhänen, Anna Virtanen, Arto Mannermaa, Arto Vuori, Joanne Demmler, Eric Fey, Terhi Kilpi, Arho Virkki, Tarja Laitinen, Kimmo Porkka)

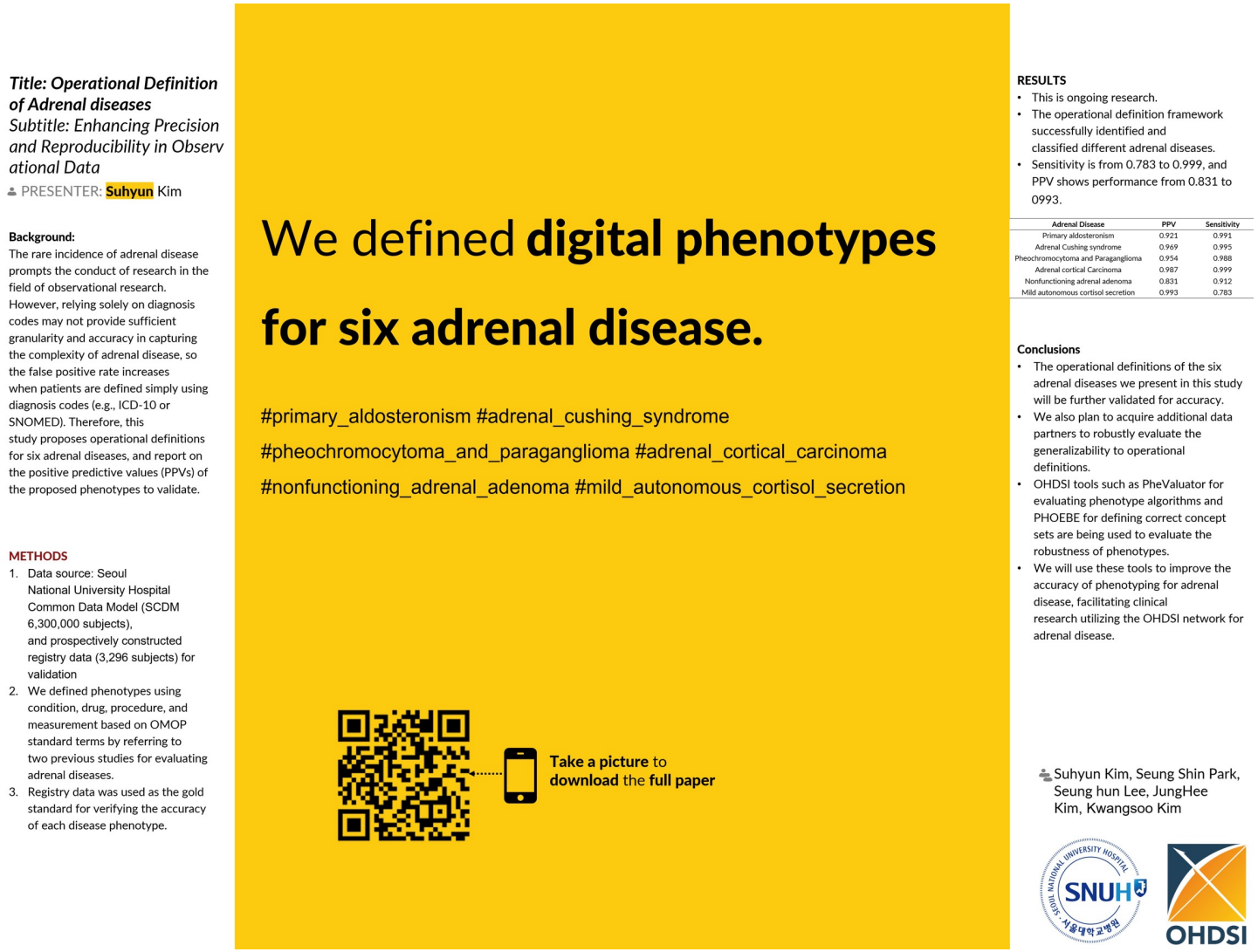


#OHDSISocialShowcase This Week

TUESDAY

Operational Definition of Adrenal diseases: Enhancing Precision and Reproducibility in Observational Data

(Suhyun Kim, Seung Shin Park, Seung hun Lee, Kwangsoo Kim, JungHee Kim)



#OHDSISocialShowcase This Week

WEDNESDAY

Validating a clinical informatics consulting service using negative control reference sets

(Michael Jackson, Saurabh Gombar, Raj Manickam, Robert Brown, Ramya Tekumalla, Yen Low)

Evaluating confounding adjustment when sample size is small

Martijn Schuemie[1,2], Marc A. Suchard[2], Akihiko Nishimura[3], Linying Zhang[4], George Hripcsak[4]

[1] Observational Health Data Analytics, Johnson & Johnson, [2] Department of Biostatistics, University of California, Los Angeles, [3] Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, [4] Department of Biomedical Informatics, Columbia University Medical Center

Read the abstract

Background

Observational studies estimating causal effects are vulnerable to confounding because groups receiving different treatments may differ in important aspects. OHDSI studies typically rely on large-scale propensity score (LSPS) models to adjust for these differences. When treatment groups are sufficiently large, LSPS has proven to work well, both in terms of covariate balance and residual systematic error measured using negative controls. However, little is known about LSPS's ability to adjust for confounding when treatment groups are small. To complicate matters, prior research shows that our ability to measure covariate balance — using the standardized difference of means (SDM) — degrades when sample size is limited.

Methods

To measure performance of LSPS under small sample sizes, we take a large study population and randomly divide it into smaller partitions to simulate different data sites, as shown in **Figure 1**.

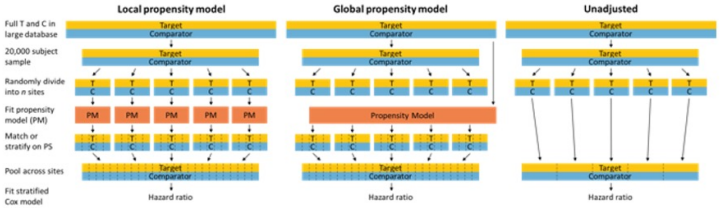


Figure 1. Simulating small data sites. We extract a target (T) and comparator (C) cohort from a large database and take a 20,000-person random sample. We then randomly divide these into *n* equally-sized sites. We evaluate propensity score adjustment using propensity models (PM) fitted at each simulated site (Local) or using a single PM fitted on the original full data (Global), and compare this to no propensity-score adjustment (Unadjusted). Data is pooled across simulated sites before fitting a stratified Cox model.

Ground truth

- Lisinopril vs hydrochlorothiazide (HCTZ), with 76 negative controls*
- Lisinopril vs metoprolol, with 76 negative controls*
- Sitagliptin vs glimepiride, with 94 negative controls**
- Sitagliptin vs liraglutide, with 94 negative controls**

* From LEGEND-HTN
** From LEGEND-T2DM

Databases

- Merative MarketScan MDCD
- Metarive MarketScan MDCR
- Optum® de-identified Electronic Health Record dataset (Optum EHR).

Metrics

- Expected Absolute Systematic Error (EASE) is computed by first fitting a Gaussian distribution to the estimated negative control hazard ratios, and then taking the expected absolute value of that distribution.
- Maximum standardized difference of mean (SDM) is computed by dividing the difference between the mean in T and C by the standard deviation for each covariate and taking the maximum of the absolute value.

Sampled sites

- 5 sites of 4,000 persons
- 10 sites of 2,000 persons
- 20 sites of 1,000 persons
- 40 sites of 500 persons
- 80 sites of 250 persons
- 160 sites of 125 persons

Contact: schuemie@ohdsi.org

Results

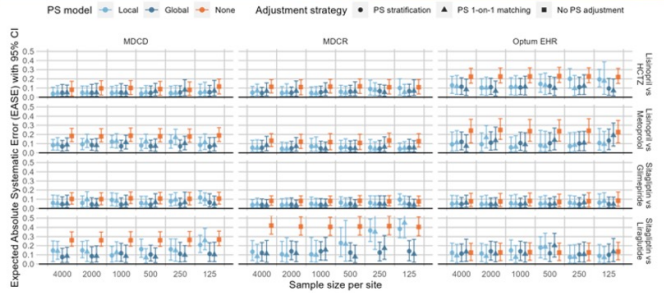


Figure 2. Expected Absolute Systematic Error (EASE) with 95% credible intervals per sample size.

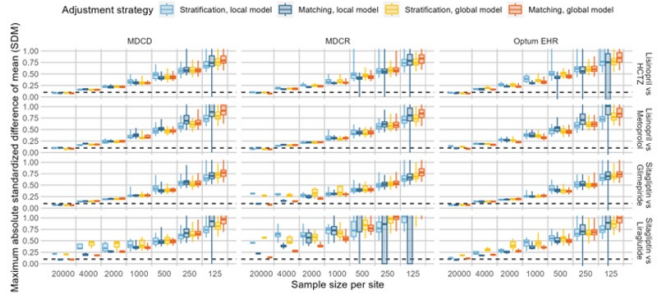


Figure 3. Maximum absolute SDM per sample size. Max SDM is computed at each site, resulting in a distribution characterized by box plots. A max SDM below 0.1 (dashed line) is considered to indicate balance.

Conclusions

- **Several target-comparator-database combinations already show little confounding in the unadjusted analyses** as measured by EASE. Here, locally-fitted propensity models did not make systematic error worse but also had no opportunity to improve.
- When confounding was detected in the unadjusted analysis, **LSPS was able to adjust for confounding at all but the smallest sample sizes**. No breakdown in performance as measured by EASE was observed when sample size >= 1,000. In many cases, sample size >= 250 was sufficient.
- Even though no confounding was observed (after adjustment) in most situations, max SDM always suggested large imbalance, meaning **our balance metric does not function when sample size is small** (*n* <= 4,000).

#OHDSISocialShowcase This Week

THURSDAY

Impact of concomitant use of proton pump inhibitors and clopidogrel on cardiovascular adverse outcomes - A multicenter study using common data model

(Seonji Kim, Kyung Joo Lee, Seng Chan You, Seung In Seo)

Using a Continuous Quality Improvement (CQI) Approach for Gap Analysis of OHDSI/ATLAS as An Enterprise Self-Service Analytics Platform by Academic Medical Centers

Albert Einstein College of Medicine

Selvin Soby[1], Pavel Goriacko[2], Jimmy John[1], Pavan Parimi[1], Erin M. Henninger[1], Parsa Mirhaji[2]
[1] Montefiore Medicine, [2] Albert Einstein College of Medicine at Montefiore

Montefiore

BACKGROUND

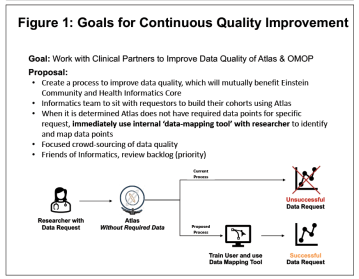
Informatics departments at large academic medical centers generally have two approaches to supplying observational health research data to investigators: custom queries developed manually by expert data engineers and data analysts or providing self-service analytic tools.[1] These tools ideally should engage users in systematic cohort identification and computable phenotyping activity in order to formulate a query based on the combination of clinical events and their temporal relationships, inclusion and exclusion criteria and other patient or population level characteristics.[2]

Custom queries require intimate understanding of the underlying data ecosystems by trained data experts, intensive communication between investigators and analysts, expose confidentiality and security of protected health information in unnecessarily, and are difficult to audit, trace, or reproduce. Self-service tools on the other hand require deep understanding of biomedical informatics standards and terminology systems, computable phenotyping, and training and onboarding process that is usually a barrier to general investigators.[3]

This project outlines a systematic research-question intake process that allows front-line researchers and clinical investigators to work directly with OHDSI Atlas system in collaboration with informatics analysts that support cohort studies and analytics while collecting direct user feedback about usability, user experience, challenges and short comings, and important feature requests from a non-informatics researchers' perspective. The current institutional perception is that Atlas' self-service capabilities, while powerful, are meant for advanced users with a formal informatics and data science background. Subsequently, this makes it difficult or impractical to use for general researchers who are interested in prep-for research or simple cohort-based studies in a local setting.

We aim to systematically guide users and provide just-in-time training in the context of user-requested projects, building their research question and guiding them to their analysis and providing applied hands-on experience to the researcher on how Atlas can help serve their real-world data research needs, while collecting direct feedback on usability, user experience, roadblocks to completion of self-service projects, potential novel optimizations to enable local users in a self-service mode to drive their projects to completion independently in a low-touch training environment. An issue tracker system is developed to log, classify, prioritize issues, and track resolution status of all projects to ensure that over time ATLAS becomes seen internally as a powerful analytics tool for all real-world data users and projects within across collaborating institutions.

Figure 1: Goals for Continuous Quality Improvement



METHODS

We created a new intake process for the research data request process at Albert Einstein's College of Medicine that utilizes a formal ITSM methodology and tool. This process includes the following steps:

A. Create an inventory of requested projects and feasibility of completing using Atlas

We looked at all channels of incoming data requests and consolidated them into our request management software, Atlassian's Jira Service Manager. The informatics team assessed each project's feasibility using the OHDSI/Atlas tool with respect to available data and cohort requirements. Each project was then tagged with pertinent information about the research question and design methodologies. Any specific gaps in our OHDSI/Atlas technology or availability of data in the OMOP-CDM was noted as items to clarify with researchers or to resolve before meetings.

B. Logistics regarding scheduling meeting with researcher

For projects that have been evaluated and identified as good candidates for OHDSI/Atlas implementation, informatics analysts reached out to the requestor with an initial analysis and request to meet. Meeting requests are coordinated using an online appointment manager linked to the team's and researcher's calendar availability to speed up scheduling the initial project review session.

C. Preparatory work in advance of meetings

Prior to the initial meeting, an informatics analyst created a draft cohort definition based on the observational research question. During the initial project review session, the researcher and the informatics analyst reviewed the creation of the cohort definition together to clarify any details, examine potential alternatives, and review differences on the cohorts and research questions. Once the cohort definition was created and generated, subsequent characterization and/or data extraction requirements are completed using the extended cohort extraction tools.

D. Issues discovered and follow-up

Any issues that were discovered, including missing data in OMOP-CDM, or limitations of the Atlas cohort workflow, usability issues, complaints, errors due to misinterpretation or misunderstanding of the tools and interfaces were all noted as build fixes. These were prioritized and entered into a custom developed issue tracker application to be shared with the product teams which were reviewed and resolved. Many issues related to data availability could be resolved using our mapping and data management workflows, since the requesting researcher is usually a subject matter expert for the requested data domains. We follow a bi-weekly release cycle to resolve and update usability issues, add new features, and improve data availability. The newly available data and build fixes were communicated to the institution's OHDSI/Atlas user community using a home page dedicated and dedicated content management system.

Figure 2: Request Triage Process

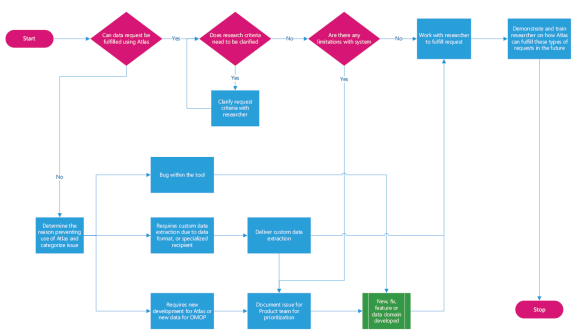


Figure 3: Summary of Active Requests by Status

Status Name	Description
Triage	Requests identified as projects which can be fulfilled using Atlas.
Evaluation	Waiting on researcher to clarify request, further investigation is needed to understand the underlying research question or problem before completing the build.
On Hold	Awaiting mapping or development to complete data query in Atlas.
In Progress	Engaged with researchers, scoped out initial requirements, negotiated a plan and timeline, and started the build in Atlas.
PI Review	Review results with investigator.
Complete	Data results provided to investigator and validation received the data provided satisfied their request.

RESULTS

As of Sept 2023, there were 48 data-request projects in review, including data needed for grant submissions, clinical trial site feasibility questionnaires, and quality improvement projects. The current project statuses include triage, evaluation, in progress, PI review, and completed. We have identified more than 37 high priority issues preventing from a truly self-service utilization of OHDSI/ATLAS by general users. However, most issues found with the inability to complete a data request in a self-service mode can be attributed to the following 4 categories: 1) source data not available yet in our OMOP-CDM instance, 2) the Atlas user interface is not intuitive and understandable to design the cohort, 3) user has trouble finding specific concepts using OMOP as terminology system and existing search and navigation process, and 4) projects require information available in non-discrete sources such as clinical text.

The informatics team is addressing these issues by working with the product teams to facilitate the data availability in OMOP and by improving the Atlas user interface to make it more intuitive. Additionally, bi-weekly summaries of data requests and their statuses are communicated to key institutional stakeholders via emailed report. This ensures transparency and accountability in the data-request process.

CONCLUSION

The new data-request review process has been well-received by researchers and the institution leadership. It has increased efficiency and collaboration between researchers and informatics analysts. However, there are still some manual steps involved, which will be automated as Atlas services are scaled up. The goal is to provide Atlas as an end-to-end self-service research data tool for researchers.

The Informatics team is committed to continuously improving the process and ensuring that researchers have the data they need to conduct high-quality research.

REFERENCES

1. Dash, B., & Ansari, M. F. (2022). Self-service analytics for data-driven decision making during COVID-19 pandemic: An organization's best defense. *Academia Letters*, 2.
2. Hellerstein, J. M., Heer, J., & Kandel, S. (2018). Self-Service Data Preparation: Research to Practice. *IEEE Data Eng. Bull.*, *41*(2), 23-34.
3. Lennerholt, C., Van Laere, J., & Söderström, E. (2018). Implementation challenges of self service business intelligence: A literature review.
4. Mirhaji, P., Soby, S., Henninger, E., Nelapatla, C., Wahle, M., Aasman, B., Belin, E., Leveraging OHDSI/ATLAS and Open-Source Development to Support Translational Research, Data Science, and Regulatory Compliance. Presented at 2022 OHDSI Symposium in Baltimore, MD.
5. Dash, B., & Ansari, M. F. (2022). Self-service analytics for data-driven decision making during COVID-19 pandemic: An organization's best defense. *Academia Letters*, 2.
6. Hellerstein, J. M., Heer, J., & Kandel, S. (2018). Self-Service Data Preparation: Research to Practice. *IEEE Data Eng. Bull.*, *41*(2), 23-34.
7. Lennerholt, C., Van Laere, J., & Söderström, E. (2018). Implementation challenges of self service business intelligence: A literature review.
8. Mirhaji, P., Soby, S., Henninger, E., Nelapatla, C., Wahle, M., Aasman, B., Belin, E., Leveraging OHDSI/ATLAS and Open-Source Development to Support Translational Research, Data Science, and Regulatory Compliance. Presented at 2022 OHDSI Symposium in Baltimore, MD.

#OHDSISocialShowcase This Week

FRIDAY

Leveraging the OMOP Common Data Model to Support Distributed Health Equity Research

(Sarah Gasman, William G. Adams)

Health Equity Explorer

- Translational informatics tool (R Shiny app) to support self-service exploration and visualization of:
  - *Any* computable health outcome
  - For a *broad range* of demographic, social, environmental and clinical drivers of health (SEDoH)
  - Via graphs, tables, maps, and statistical analysis (R)
  - Using *open-use software*, shared analytic code, common data models (OMOP) and a common data mart

Boston Medical Center — Exceptional Care. Without Exception.
Health Equity Accelerator — Boston Medical Center
Appsilon

Three Openings at Gilead

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.

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.

(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

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?

Methods Research Workgroup

Methods Workgroup mission

Empower real-world evidence generation through **collaborative innovation** in **statistical and computational methods**

Methods Workgroup objectives and key results

- Promote awareness and collaboration in methods research
  - Maintain a comprehensive directory of ongoing methods research Martijn
  - Have at least 6 presentations of ongoing methods research (i.e. work that hasn't been published yet) Martijn
  - Average attendance of meetings >= 20 researchers Martijn
- Align research topics with the needs of the OHDSI community
  - Perform a survey to elicit community needs Linying

Key result lead

HADES

HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

HADES mission

Enable the OHDSI community to **perform observational research** following **OHDSI best practices** for characterization, population-level estimation, and patient-level prediction by providing a **cohesive set of open-source analytic software**.

HADES objectives and key results

- More user involvement
  - Get post-mortems on 2 network studies, analyzing what worked well and what didn't. Martijn
- Document interfaces
  - Fully document Strategus inputs JPG, Chris
  - Fully document results schema Hayden
  - Manifest describing rules for database platform support Katy, Hayden
- Better testing
  - Establish an automated procedure for determining which packages uses which testing servers Martijn
- Strategus
  - Get Strategus into HADES Anthony, JPG
  - Less painful installation of Strategus Anthony
  - Containers + execution engine Anthony, Evan

Key result lead

Perinatal & Reproductive Health (PRHeG) 2024 OKRs

- Our **purpose** is to develop tools and standards for pregnancy and reproductive health research to foster collaborative studies within the OHDSI network, and advance research in this field generally.
- **Objectives and key results**
  - Develop and share phenotypes for key perinatal and reproductive health factors to enable research using diverse data sources.
  - Provide training and education to perinatal and reproductive health researchers interested in OHDSI projects.
  - Extend existing work to identify pregnancy episodes in data in the OMOP CDM using additional data sources e.g. EHR data.
  - Improve the transformation of data from pregnancy-specific EHR modules into the OMOP CDM.

Registry Workgroup

Tina Parciak

OKR 2024 of the Registry WG

Objectives:

Our workgroup wants to

- Build a network of registry stakeholders (data owners, ETL, project managers...) of existing or emerging OMOP-ed registry datasets to
- Support on-going or new initiatives in transforming registry data to the OMOP CDM.
- We want to enable this support through accessible documentation on GitHub, result dissemination on conferences, CC or in journals and dedicated workgroup discussions.

OKR 2024 of the Registry WG

Key results:

- Overview: Differences between EHR data vs. Registry (“curated”) data
- Overview: challenges in mapping registry (“curated”) data to the OMOP CDM
  - Generate list of challenges
  - Prioritise items
  - One challenge per 1-2 workgroup meetings
- Documentation of challenge, discussed solutions, recommendations (e.g. for changes in the vocabulary) or conventions for transformations
  - GitHub page
  - Manuscript for journal and/or OHDSI conference

Steering Workgroup

co-leads: Patrick Ryan, George Hripcsak

Purpose: Steering WG exists to support the community and its leaders in collaboratively generating the evidence that promotes better health decisions and better care, by identifying, organizing, and guiding collaborative activities, facilitating communications across the community, providing input to operations of the OHDSI Central Coordinating Center, and building consensus on the vision for where the OHDSI community should go together.

Objective 1: Empower workgroups to contribute to collaboratively generating the evidence that promotes better health decisions and better care

Key results:

1. 100% of active workgroups have defined purpose and 2024 OKRs that are communicated to broader community to promote focus and encourage contributions; Timeline: 1Q2024
2. 1 Workgroup Leader Summit convened to ensure appropriate communication across workgroups; Timeline: 1Q2024

Objective 2: Create collaboration activities that encourage collaborative generation and dissemination of the evidence that promotes better health decisions and better care

Key results:

1. OHDSI2024 Global Symposium scheduled with location/dates announced; Timeline: 1Q2024
2. 3 community activities with >30 collaborators participating: 1- Phenotype Phebruary , timeline: Feb2024;

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