Introduction to Phenotype Phebruary III

OHDSI Community Call
Jan. 30, 2024 • 11 am ET
## Upcoming Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 30</td>
<td>Phenotype Phebruary Introduction</td>
</tr>
<tr>
<td>Feb. 6</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 1</td>
</tr>
<tr>
<td>Feb. 13</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 2</td>
</tr>
<tr>
<td>Feb. 20</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 3</td>
</tr>
<tr>
<td>Feb. 27</td>
<td>Workgroup OKRs / Phenotype Phebruary Update 4</td>
</tr>
</tbody>
</table>
WG Leads: Please Sign Up For OKR Announcements

Currently Signed Up:

- Africa Chapter
- APAC
- CDM
- CDM Vocabulary Subgroup
- Dentistry
- Electronic Animal H.R.
- Eyecare & Vision Research
- FHIR + OMOP
- Generative AI and Analytics in Healthcare (GAIA)
- HADES
- Health Equity
- Healthcare Systems
- Medical Devices
- Methods Research
- NLP
- Network Data Quality
- Oncology
- Patient-Level Prediction
- Perinatal and Reproductive Health
- Phenotype Development & Evaluation
- Registry
- Steering Group
- Themis
- Vaccine Vocabulary

2024 Workgroup OKR Announcements

In order to highlight different initiatives and opportunities throughout the environment, workgroups will share their 2024 Objectives and Key Results (OKRs) during February community calls. These will be 2-3 minute presentations that will be posted to the OHDSI workgroup page. If you choose to include slides, please send them to Craig Sachson by 5 pm ET the day before your selected community call.

1. Workgroup Name *
   
   [Input Field]

2. Presenter Name *
   
   [Input Field]

3. Date to Present? *
   - [ ] Feb 6
   - [ ] Feb 13
   - [ ] Feb 20
   - [ ] Feb 27

@OHDSI www.ohdsi.org #JoinTheJourney
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
Congratulations to the team of Soobeen Seol, Jimyung Park, Chungsoo Kim, Dong Yun Lee, and Rae Woong Park on the publication of RHEA: Real-World Observational Health Data Exploration Application in Volume 310 of Studies in Health Technology and Informatics.
Congratulations to the team of Sujin Gan, Chungsoo Kim, Dong Yun Lee, and Rae Woong Park on the publication of Prediction Models for Readmission Using Home Healthcare Notes and OMOP-CDM in Volume 310 of Studies in Health Technology and Informatics.

Prediction Models for Readmission Using Home Healthcare Notes and OMOP-CDM

Sujin Gan¹, Chungsoo Kim¹, Dong Yun Lee², and Rae Woong Park¹

¹Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea
²Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

ORCID ID: Sujin Gan https://orcid.org/0000-0003-3326-8428, Chungsoo Kim https://orcid.org/0000-0003-1802-1777, Dong Yun Lee https://orcid.org/0000-0002-3678-9862, Rae, Woong Park https://orcid.org/0000-0003-4989-3287

Abstract. This study developed readmission prediction models using Home Healthcare (HHC) documents via natural language processing (NLP). An electronic health record of Ajou University Hospital was used to develop prediction models (A reference model using only structured data, and an NLP-enriched model with structured and unstructured data). Among 573 patients, 65 were readmitted to the hospital. Five topics were extracted from HHC documents and improved the model performance (AUROC 0.740).

Keywords. Readmission, home healthcare, machine learning, prediction

1. Introduction

Readmission is an indicator of inpatient care quality and a major contributor to growing healthcare costs [1]. Therefore, identifying patients at high risk for readmission is crucial to reduce the likelihood of readmission. Home healthcare (HHC) is provided to discharged patients. Thus, HHC documents, which include post-discharge information from medical procedures to patient complaints, may contain hidden risk factors. Machine learning-based prediction models have been developed to assist in the identification of readmission risk factors, but they have not been applied to HHC documents.
OHDSI Shoutouts!

Congratulations to the team of Scott DuVall, Craig Parker, Amanda Shields, Patrick Alba, Julie Lynch, Michael Matheny, and Aaron Kamauu on the publication of Toward Real-World Reproducibility: Verifying Value Sets for Clinical Research in Volume 310 of Studies in Health Technology and Informatics.

Abstract. Standardized operational definitions are an important tool to improve reproducibility of research using secondary real-world healthcare data. This approach was leveraged for studies evaluating the effectiveness of AZD7442, an SARS-CoV-2 spike protein, which received FDA Emergency Use Authorization (EUA) in December 2021 for COVID-19 pre-exposure prophylaxis (PEP) in patients with moderate to severe immunocompromising (IC) medical conditions. These operational definitions included value sets, which, where possible, were drawn from common use (e.g., Charlson Comorbidity Index [1-3], Value Set Authority Center [4]), validated constructs (e.g., eMERGE Phenotype Knowledgebase [5]), and other published resources [6,7]. However, as terminologies and dictionaries are constantly evolving,
OHDSI Shoutouts!

Congratulations to the team of Piper Ranallo, Bronwyn Southwell, Christopher Tignanelli, Steven G. Johnson, Richard Krueger, Tess Sevareid-Groth, Adam Carvel, and Genevieve B. Melton on the publication of Promoting Learning Health System Cycles by Optimizing EHR Data Clinical Concept Encoding Processes in Volume 310 of Studies in Health Technology and Informatics.
OHDSI Shoutouts!

Congratulations to the team of ChulHyoung Park, Sang Jun Park, Da Yun Lee, Seng Chan You, Kihwang Lee, and Rae Woong Park on the publication of Multi-Institutional Collaborative Research Using Ophthalmic Medical Image Data Standardized by Radiology Common Data Model (R-CDM) in Volume 310 of Studies in Health Technology and Informatics.
OHDSI Shoutouts!

Congratulations to the team of Seol Whan Oh, Soo Jeong Ko, Yun Seon Im, Surin Jung, Bo Yeon Choi, Jae Yoon Kim, Sunghyeon Park, Wona Choi, and In Young Choi on the publication of Development of Integrated Data Quality Management System for Observational Medical Outcomes Partnership Common Data Model in Volume 310 of Studies in Health Technology and Informatics.
Congratulations to the team of Jiyun Cha, Eun Kyoung Ahn, Young-Heum Yoon, and Man Young Park on the publication of Feasibility of Applying the OMOP Common Data Model to Traditional Eastern Asian Medicine Dataset in Volume 310 of Studies in Health Technology and Informatics.
OHDSI Shoutouts!

Congratulations to the team of Najia Ahmadi, Quang Vu Nguyen, Martin Sedlmayr, and Markus Wolfien on the publication of a comparative patient-level prediction study in OMOP CDM: applicative potential and insights from synthetic data in Scientific Reports.
OHDSI Shoutouts!

Congratulations to Ines Reinecke on successfully defending her doctoral thesis at Technische Universitat Dresden.

Dr. Reinecke’s dissertation was titled "Secondary use of German inpatient medication data in international studies while preserving semantics.” Congrats!
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
# Upcoming Workgroup Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Time (ET)</th>
<th>Meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday</td>
<td>10 am</td>
<td>Surgery and Perioperative Medicine</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3 pm</td>
<td>Vulcan/OHDSI Meeting (ZOOM)</td>
</tr>
<tr>
<td>Thursday</td>
<td>9:30 am</td>
<td>Themis</td>
</tr>
<tr>
<td>Thursday</td>
<td>11 am</td>
<td>Industry</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>Medical Devices</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>Methods Research</td>
</tr>
<tr>
<td>Thursday</td>
<td>1 pm</td>
<td>OMOP CDM Oncology Vocabulary/Development Subgroup</td>
</tr>
<tr>
<td>Thursday</td>
<td>7 pm</td>
<td>Dentistry</td>
</tr>
<tr>
<td>Friday</td>
<td>10 am</td>
<td>GIS – Geographic Information System</td>
</tr>
<tr>
<td>Friday</td>
<td>11:30 am</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Monday</td>
<td>9 am</td>
<td>Vaccine Vocabulary</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Africa Chapter</td>
</tr>
<tr>
<td>Tuesday</td>
<td>9 am</td>
<td>ATLAS/WebAPI</td>
</tr>
<tr>
<td>Tuesday</td>
<td>10 am</td>
<td>Common Data Model</td>
</tr>
</tbody>
</table>
OHDSI Evidence Network

OHDSI is proud to have a global community dedicated to generating real-world evidence and which recognizes the opportunity to collaborate together as part of a distributed network based on standardized data and standardized analytics. The OHDSI Evidence Network consists of organizations equipped with access to one or more databases standardized to the OMOP CDM who express a keen interest in participating in OHDSI network studies. Collaboratively, OHDSI Evidence Network partners share aggregate summary statistics about their databases, which are used to support Database Diagnostics, helping identify databases within the network that are fit-for-use for particular research questions. Additionally, partners have the opportunity to opt in and contribute to network studies proposed by the OHDSI community.

The recent SOS challenge serves as a compelling demonstration of the OHDSI Evidence Network’s current capabilities and its promising future potential. We wholeheartedly encourage all organizations that are adopting the OMOP CDM and aspire to apply standardized analytics for the reliable generation of real-world evidence to become part of the OHDSI Evidence Network.

During the first community call of 2023, Patrick Ryan unveiled the strategic priorities for the OHDSI Community for the year. Among these, a key focus is on enhancing the transparency and maturity of the OHDSI network.

To address this objective, we began by considering how network studies are currently conducted, recognizing the challenges and complexities faced by collaborating organizations when contributing to the body of evidence. This investigation led to the creation of Database Diagnostics, a tool designed to answer a critical question: when tackling a specific research inquiry, which data sources within the OHDSI Evidence Network are the most relevant and suitable for generating robust evidence?

This innovative approach leverages aggregated summary statistics from each data source, obtained through the open-source tool dbProfile. It evaluates databases’ fitness-for-use across various dimensions, including patient demographics, domain coverage, longitudinal data availability, and the capture of target, comparator, and outcome variables. The overarching vision was to establish these database profiles as the foundation to enable the OHDSI Evidence Network.

On March 28, 2023, the OHDSI Global Community initiated the Save Our Syrius (SOS) Challenge, a groundbreaking opportunity for collaborative research involving simultaneous participation in four different network studies. What made it truly remarkable was that any organization interested in joining the OHDSI Evidence Network could contribute to these studies by sharing their database profiles for the data sources they had access to. These profiles were centrally aggregated at the OHDSI Central Coordinating Center, enabling us to empirically determine which of the four study questions each data source was best suited to address. This inaugural OHDSI Evidence Network endeavor encompassed 36 diverse data sources from 16 different organizations. Not only did this foster rapid evidence generation and collaboration during the SOS Challenge, but it also positioned us for future collaborations on additional network studies as part of the OHDSI Evidence Network.

If you are interested in becoming a part of the OHDSI Evidence Network, and contributing to advancing evidence-based healthcare, please use the provided QR code to complete a brief form about your organization and your data source. A member of the OHDSI Network Data Quality Working Group will reach out to you to explore this exciting opportunity further!
SelfControlledCaseSeries

Introduction

SelfControlledCaseSeries is an R package for performing Self-Controlled Case Series (SCCS) analyses in an observational database in the OMOP Common Data Model.

Features

- Extracts the necessary data from a database in OMOP Common Data Model format.
- Optionally add seasonality using a spline function.
- Optionally add age using a spline function.
- Optionally add calendar time using a spline function.
- Optionally correct for event-dependent censoring of the observation period.
- Optionally add many covariates in one analysis (e.g., all drugs).
- Options for constructing different types of covariates and risk windows, including pre-exposure windows (to capture contraindications).
- Optionally use regularization on all covariates except the outcome of interest.
MONDAY
Transforming the Optum® Enriched Oncology module to OMOP CDM

(Dmitry Dymshyts, Clair Blacketer)

Title: Transforming the Optum® Enriched Oncology module to OMOP CDM

INTRO
- The Optum® Enriched Oncology Data set is a group of tables that can supplement the Optum® de-identified Electronic Health Record dataset.
- It includes specific oncology concepts important for understanding the progression of the disease, which often not available in structured formats, particularly the tumor, node, and metastasis (TNM) values, stage information and biomarkers.
- As of 2022, there are approximately 1.9 million patients with at least one solid tumor ICD-9 or ICD-10 diagnosis included in the data set.

METHODS
1. The data mapping was done with help of the OHDSI tools White Rabbit and Rabbit-in-a-hat.
2. The concept mapping was done by a semantic analysis of source concepts and with help of OHDSI Usagi tool**. Data and concepts mapping is based on the OHDSI Oncology Working group guidelines***.
3. The data cleaning was done by removing clinically impossible events, for example, the same cancer being "in situ" and "invasive" at the same day.

RESULTS
- The Optum EHR data now contains important characteristics of neoplastic disorders: staging, grading, histology, behavior, TNM stage, genetic markers, tumor size, etc. See example in Figure 1.
- This enables observational research on a very specific group of patients or detect a specific outcome. This is important in development of novel targeted therapies and better understanding of how-existing drugs affect patients.

* https://github.com/OHDSI/WhiteRabbit/
** https://github.com/OHDSI/Usagi/
*** https://github.com/OHDSI/UsagiWG

The OHDSI Standardized Vocabularies are robust enough to support transformation of oncology data, including staging, grading, histology and genetic markers of neoplastic disorders.

Highlights
- Data mapping: straightforward and source data fits well into the OMOP CDM in general.
- Concept mapping:
  - The following categories of concepts were mapped:
    - Coadjutor: combination of histology, morphology and behavior (adenocarcinoma of breast in situ)
    - Biomarkers: HER2 receptor tyrosine kinase 2 (HER2 or HER2/neu) etc.
    - TNM stage
  - Other cancer modifiers: grade, metastases location, summary stage, advanced/localized, evaluation systems (Breast Stage, Duke/Mallower Stage, ECOC performance status, FIGO Stage, Gerson, Choson scoring, tumor size, etc.
- Relatively small number of patients have cancer modifiers documented (see Figure 1). 
  - Only about 4% of patients have Estrogenreceptor status with only 10% of the data.
  - But the proportion between Positive and Negative status is the same reported in the literature, same applicable for the ERBB2 and Progesterone receptor measurements.

Limitations/needs for future development:
- Treatment response:
  - Good / incomplete / complete / partial / complete pathological / other
  - Progression concepts in the Episode domain
- Tumor progression will be mapped to the Episode table in the next iteration.
- Line of therapy to be mapped to the Hem/Onc vocabulary
- Data cleaning algorithms to be developed

@OHDSI
www.ohdsi.org
#OHDSISocialShowcase
#JoinTheJourney
TUESDAY

Community Contribution to the OHDSI Vocabularies: moving towards collaborative shared resource

(Oleg Zhuk, Anna Ostropolets, Alexander Davydov, Christian Reich)

Fix your own Vocabulary:

- Templates and guidelines let you
- add your source vocabulary
- add missing concepts
- fix domains, concept names and mappings

QC checklists help you don’t do anything wrong

Talk to us if you have more complex use cases!

Published guidelines cover the following use cases (Table 1).

Table 1. Supported contribution use cases

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Adding new non-standard concept(s) to an existing vocabulary</td>
</tr>
<tr>
<td>T2</td>
<td>Adding new synonym(s) to an existing concept(s)</td>
</tr>
<tr>
<td>T3</td>
<td>Adding a mapping to an existing concept</td>
</tr>
<tr>
<td>T4</td>
<td>Adding a new vocabulary as non-standard with mappings (full or partial) to a standard vocabulary</td>
</tr>
<tr>
<td>T5</td>
<td>Modifying attributes of an existing concept(s)</td>
</tr>
<tr>
<td>T6</td>
<td>Modifying mapping for an existing concept</td>
</tr>
<tr>
<td>T7</td>
<td>Promoting non-standard concepts to standard</td>
</tr>
</tbody>
</table>

Depending on the use case, the contributor needs to submit the following tables:
- concept_manual, concept_relationship_manual, concept_converse_manual, metadata, checklist
- The Vocabulary layer stores metadata about the contribution. Their format is compatible with SSDOM. Date of submission, reviewer status and contact details (email, name, organisation) are collected per contribution and other parameters are collected per relationship (Table 2).

Table 2. Metadata collected per relationship

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicate_id</td>
<td>type of matching (is-a, has-a, has-set, 3-way)</td>
</tr>
<tr>
<td>mapping_source</td>
<td>if applicable</td>
</tr>
<tr>
<td>mapping_justification</td>
<td>how the mapping was formed at, ManualMappingFunction, LexicalMatching, etc.</td>
</tr>
<tr>
<td>mapping_tool</td>
<td>if applicable</td>
</tr>
</tbody>
</table>

We are collecting more complex use cases to develop infrastructure enabling shared vocabulary development by other community members (e.g. Table). Talk to us if you want to support a vocabulary!
Evaluating confounding adjustment when sample size is small

Martijn Schuemie, Marc A. Suchard, Akihiko Nishimura, Linying Zhang, George Hripcsak

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 (LPS) models to adjust for these differences. When treatment groups are sufficiently large, LPSs have proven to work well, both in terms of covariate balance and residual systematic error measured using negative controls. However, little is known about LPS’s ability to adjust for confounding when treatment groups are small. To complicate matters, prior research showed 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 LPS under small sample sizes, we take a large study population and randomly divide it into smaller partitions to simulate different data sizes, as shown in Figure 1.

- Ground truth
  - Lisinopril vs hydrochlorothiazide (HCTZ), with 76 negative controls
  - Lisinopril vs metoprolol, with 76 negative controls
  - Statin vs glipizide, with 94 negative controls
  - Statin vs linsitide, with 94 negative controls

- 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.

Results

- Figure 1: Simulating small data sizes. We extract a target (T) and comparator (C) cohort from a large database and take a 30,000-patient random sample. We then randomly divide these into 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.

- 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 the plot. A max SDM below 0.1 (blended 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 analyses, LPSs 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 \( \geq 1,000\). In many cases, sample size \( \geq 250\) was sufficient.
- Even though no confounding was observed (after adjustment) in most situations, max SDM always suggested large imbalances, meaning our balance metric does not function when sample size is small (e.g., 400).

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

(Selvin Soby, Pavel Goriacko, Jimmy John, Pavan Parimi, Erin M. Henninger, Parsa Mirhaji)

**BACKGROUND**

Many academic health systems attempt to improve overall care quality through the use of clinical quality improvement (CQI) programs. CQI programs have been used to improve the quality of healthcare delivery. These programs aim to improve outcomes by identifying and addressing gaps in care processes. OHDSI/ATLAS is an enterprise self-service analytics platform by academic medical centers. The study aimed to identify areas where CQI could be applied to improve the quality of healthcare delivery.

**METHODS**

A CQI approach was used to analyze the quality of healthcare delivery. The study identified areas where CQI could be applied to improve the quality of healthcare delivery. The study used a mixed-methods approach, including surveys, focus groups, and interviews with healthcare providers.

**RESULTS**

The study found that CQI could be applied to improve the quality of healthcare delivery. The study identified areas where CQI could be applied to improve the quality of healthcare delivery. The study used a mixed-methods approach, including surveys, focus groups, and interviews with healthcare providers.

**CONCLUSION**

The study found that CQI could be applied to improve the quality of healthcare delivery. The study identified areas where CQI could be applied to improve the quality of healthcare delivery. The study used a mixed-methods approach, including surveys, focus groups, and interviews with healthcare providers.
Risk of aortic aneurysm or dissection following exposure to fluoroquinolone antibiotics

**INTRODUCTION**

- Fluoroquinolone (FQ) antibiotics are a broad-spectrum class of antibiotics. While FQ antibiotics are generally well tolerated, a meta-analysis of four observational studies found that FQ antibiotics were associated with an increased risk of aortic diseases compared to other antibiotics (adjusted odds ratio 2.30; 95% CI 1.65-3.28).
- However, the quality of this study was rated as moderate and the results from later studies were conflicting.
- We aimed to determine whether exposure to FQ are associated with aortic aneurysm or dissection after initiating treatment.

**METHODS**

- This study is being conducted through international collaboration through OHDSI network in Seoul SingPhe (SDS) challenge.
- The study utilized data from January 2015 to December 2019.
- We used a new user comparative cohort design including all patients diagnosed with urinary tract infection (UTI) and initiating either treatment with systemic FQ or a comparison group selected via a nearest neighbor matching algorithm (PTM: cephalosporin (CP) in outpatient setting.
- Primary outcome is any composite outcomes of hospitalization due to aortic aneurysm or aortic dissection within 60 days.
- The suite of diagnostics is employed to evaluate balance, empirical equipoise, and systematic bias. Only results passing the predefined suite of diagnostics are used for the meta-analysis.

**RESULTS**

- Among 14 databases, results from 7 and 9 databases passed diagnostics for comparison with TMF and CP respectively.
- Overlap of preference scores (PS) distributions across databases is heterogeneous across data sources.
- In results form US, more PS overlap is identified in comparison of FQ with TMF than comparison with CP, while vice versa in results from non-US databases.
- In meta-analysis using only results passing diagnostics, the hazard ratio of primary outcome was 0.91 (95% CI: 0.73-1.15) in comparison with TMF and 1.01 (95% CI: 0.82-1.22) indicating no different risk of primary outcome between treatment groups.

**CONCLUSION**

- We found that use of fluoroquinolone in patients with UTI did not result in different risk of following aortic dissection or aneurysm within 60 days compared with use of TMF or CP.
- Studies identifying comparative risk of aortic diseases following fluoroquinolone exposure from a single data source or single health system may be vulnerable to bias as there is substantial heterogeneity in preference of treatment.

---

**FRIDAY**

**Risk of aortic aneurysm or dissection following exposure to fluoroquinolone antibiotics**

(Jack L. Janetzki, Jung Ho Kim, Jung Ah Lee, Seng Chan You, Nicole L. Pratt)
The Zhang Lab at Washington University School of Medicine in St. Louis has **one postdoc/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](https://linyingzhang.com)
  - Postdoc: [https://linyingzhang.com/files/Postdoc.pdf](https://linyingzhang.com/files/Postdoc.pdf)
  - Data analyst: [https://linyingzhang.com/files/Analyst.pdf](https://linyingzhang.com/files/Analyst.pdf)
- If interested, please send CV and cover letter to linyingz@wustl.edu
### Job Overview

**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, LLC., 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 reliable evidence.
Opening: Research Information Specialist at UNC

Research Informatics Specialist

Position Information

<table>
<thead>
<tr>
<th>Department</th>
<th>TraCS Institute-429801</th>
</tr>
</thead>
<tbody>
<tr>
<td>Career Area</td>
<td>Information Technology</td>
</tr>
<tr>
<td>Posting Open Date</td>
<td>12/15/2023</td>
</tr>
<tr>
<td>Application Deadline</td>
<td>01/30/2024</td>
</tr>
<tr>
<td>Open Until Filled</td>
<td>No</td>
</tr>
<tr>
<td>Position Type</td>
<td>Permanent Staff (EHRA NF)</td>
</tr>
<tr>
<td>Position Number</td>
<td>20060002</td>
</tr>
<tr>
<td>Vacancy ID</td>
<td>N0007640</td>
</tr>
<tr>
<td>Full Time/Part Time</td>
<td>Full-Time Permanent</td>
</tr>
<tr>
<td>FTE</td>
<td>1</td>
</tr>
</tbody>
</table>

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

* 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 scope 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?
Jan 30: Introduction to Phenotype Phebruary

Azza Shoaibi
Director, Observational Health Data Analytics
Janssen Research and Development

Anna Ostropolets
Associate Director, Observational Health Data Analytics
Janssen Research and Development

Jamie Weaver
Associate Director, Observational Health Data Analytics
Janssen Research and Development
The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

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