Lessons Learned from the 2023 UK Studyathon

OHDSI Community Call
Jan. 23, 2024 • 11 am ET
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 23</td>
<td>2023 UK Study-A-Thon Lessons Learned</td>
</tr>
<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:

- FHIR + OMOP
- Generative AI and Analytics in Healthcare (GAIA)
- HADES
- Methods Research
- NLP
- Perinatal and Reproductive Health
- Registry
- Steering Group
- Health Equity
- Oncology
- CDM
- Medical Devices

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 *
   
   Enter your answer

2. Presenter Name *
   
   Enter your answer

3. Date to Present? *
   
   - Feb 6
   - Feb 13
   - Feb 20
   - Feb 27
# Upcoming Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 23</td>
<td>2023 UK Study-A-Thon Lessons Learned</td>
</tr>
<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>
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
Congratulations to the team of Joel Swerdel and Mitchell Conover on the publication of Comparing broad and narrow phenotype algorithms: differences in performance characteristics and immortal time incurred in the Journal of Pharmacy & Pharmaceutical Sciences.
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
<table>
<thead>
<tr>
<th>Date</th>
<th>Time (ET)</th>
<th>Meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday</td>
<td>12 pm</td>
<td>Common Data Model Vocabulary Subgroup</td>
</tr>
<tr>
<td>Wednesday</td>
<td>7 am</td>
<td>Medical Imaging</td>
</tr>
<tr>
<td>Wednesday</td>
<td>9 am</td>
<td>OMOP CDM Oncology – Outreach/Research Subgroup</td>
</tr>
<tr>
<td>Wednesday</td>
<td>12 pm</td>
<td>Latin America</td>
</tr>
<tr>
<td>Wednesday</td>
<td>4 pm</td>
<td>Vulcan/OHDSI Meeting</td>
</tr>
<tr>
<td>Thursday</td>
<td>9:30 am</td>
<td>Network Data Quality</td>
</tr>
<tr>
<td>Thursday</td>
<td>7 pm</td>
<td>Dentistry</td>
</tr>
<tr>
<td>Friday</td>
<td>9 am</td>
<td>Phenotype Development and Evaluation</td>
</tr>
<tr>
<td>Friday</td>
<td>10 am</td>
<td>GIS – Geographic Information System</td>
</tr>
<tr>
<td>Friday</td>
<td>11 am</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Friday</td>
<td>11:30 am</td>
<td>Steering Group</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Healthcare Systems Interest Group</td>
</tr>
</tbody>
</table>
2023: Outreach & Preparation

- **Rework Processes & Enable Contributors**
  - Refactor Oncology IWG processes and style of project management
  - Further leveraging Github for transparency and expanded avenues of contribution
  - Focus on asynchronous development

- **Landscaping & Outreach**
  - Leverage external resources and use cases
  - Receive feedback from community regarding hurdles, data sources, and use cases
  - Understand overlap and encourage collaboration

- **Identify & Inventory**
  - Build list of use cases as well as known gaps in conventions, vocabularies and desired source data to incorporate
  - Assess overlap and find common threads

- **Prioritize & Build Roadmap**
  - Prioritize effort around maturing model using active use cases and prevalence of experienced hurdles
  - Iteratively work in sprints towards continuously improving the standards

Community sentiment towards effort overwhelmingly positive:
- Viewed as worthwhile and impactful
- Many community members willing to contribute time and resources

Landscaping & outreach have shown significant overlap in:
- Implementation barriers and roadblocks experienced
- Data sources and variables of interest
- Use cases and interest in network research
The stage is set!

A GitHub Project has been created with:

- An inventory of the outstanding work identified by the outreach efforts
- Documentation on project processes, methods and contribution mechanisms
- Tasks broken down into smaller “chunks” to enable many small, and often asynchronous, contributions (rather than singular large bodies of work)
- Plan: complete as much as we can, prioritized by use cases, in preparation for a new stable release

Who should get involved?

To achieve international and source-agnostic interoperability, specifically the harmonization of diverse data representations, a diverse group stakeholders, data sources and contributors is required.

Let’s get to work!

An OHDSI forum post contains an overview of the effort and relevant links: t.ly/XbspZ

To supplement the documentation, there is a meeting scheduled tomorrow (Teams/Onc/Dev Vocab Subgroup) at this same time to give an overview and answer any questions. This will be recorded and linked in the documentation

Please see the “Getting Involved” section of the docs to #joinTheJourney
2024 Oxford Summer School: June 17-21

Oxford Summer School 2024: Real World Evidence using the OMOP Common Data Model

COURSE DIRECTORS
Daniel Prieto-Alhambra
Professor of Pharmaco- and Device Epidemiology

COURSE ADMINISTRATOR
Mahkameh Mafi
Personal Assistant to Professor Prieto-Alhambra
MONDAY

Mapping of Critical Care EHR Flowsheet data to the OMOP CDM via SSSOM

(Polina Talapova, Andrew Williams, Nicolas Matentzoglu, Anna Ostropolets, Michael Kallfelz)
Paving the way to estimate daily dose in OMOP CDM for Drug Utilisation Studies in DARWIN EU®

(Theresa Burkard, Kim Lopez-Güell, Artem Gorbachev, Annika M Jödicke, Nuria Mercadé-Besora, Talita Duarte-Salles, Maria de Ridder, Mees Mosseveld, Dani Prieto-Alhambra, Christian Reich, Marti Catala)
Generating Synthetic Electronic Health Records in OMOP using GPT

(Chao Pang, Xinzhuo Jiang, Nishanth Parameshwar Pavinkurve, Krishna S. Kalluri, Elise L. Minto, Karthik Natarajan)
Comparing concepts extracted from clinical Dutch text to conditions in the structured data

**INTRO**
Understanding valuable hidden information in clinical narratives is crucial for clinical research and practice. This study focuses on assessing the semantic similarity between coded conditions and extracted concepts using a Dutch concept extraction framework, contributing to bridging the gap for non-English language processing in healthcare.

**METHODS**
- Data: Integrated Primary Care Information (IPC) Dutch general practice EHR database.
- Concept extraction: MedDox with Dutch resources.
- Setup: We applied the concept extraction framework to clinical narratives related to commonly occurring ICD-10-coded conditions.
- Semantic Similarity: Donahue's Cosine formulated text and ontologies-based embeddings were used to calculate the similarity between coded conditions and extracted concepts.

**RESULTS**
- The information difference between coded conditions and their related clinical notes.

**THURSDAY**
Comparing concepts extracted from clinical Dutch text to conditions in the structured data

*(Tom M. Seinen, Jan A. Kors, Erik M. van Mulligen, Peter R. Rijnbeek)*

**Methods Extra**
- Concept extraction framework:
  - Preprocessing: Lowercase, Remove numbers
  - Spa's Dutch tokenization
  - Quick-UMLS concept extraction
  - EM blows up concept
  - Patient preferred terms
  - MedDox framework
  - Dutch context rules
- Annotation:
  - 3000 code occurrences
  - 250 different codes
- Embeddings:
  - Text-based embeddings (9)
  - Ontologies-based embeddings (4)
- Similarity threshold:
  - Similar concepts:
    - From 1 SD from the mean (1)
    - To the mean (1)
- Related concepts:
  - From 1 SD from the median
  - To 1 SD from the mean (1)
- Some numbers:
  - Codes occurring - 100K times
  - 317 different ICD-10 condition codes
  - 29 million condition occurrences
  - 115 million notes
  - 42 different extracted concepts

**Future steps**
- Concept extraction framework:
  - Adding more Dutch synonyms from UMLS
  - Comparison to other concept extraction frameworks
  - Try the same for another language
FRIDAY

Finding a constrained number of predictor phenotypes for multiple outcome prediction

(Jenna M Reps, Jenna Wong, Egill A. Fridgeirsson, Chungsoo Kim, Luis H. John, Ross D. Williams, Patrick Ryan)

It is possible to develop high performing models using the same small set of predictors

Title: Finding a constrained number of predictor phenotypes for multiple outcome prediction

Presenter: Jenna M. Reps

ABSTRACT:
- Can we perform a large-scale characterization study to identify a constrained set of predictors that generally discriminate whether a patient will develop a future outcome?
- If so, we can use these predictors to predict 100s or 1000s of outcomes.

METHODS
1. We used OMDP.COM databases (MDIC, MDCC, CFAE, JMDIC, Germany and Australia).
2. We investigated candidate covariates consisting of conditions/age groups grouped using the hierarchies that are recorded in the 5-year prior to target cohort.

RESULTS
The models using the constrained predictor sets (constrained LR/GBM) often performed similarly to the models that had thousands of Candidate predictors (best case LR) and did better than models trained using only agnostic predictors (worst case LR).

EXPLORE YOUR RISKS
View: WhatHappenToMe.org
# Opening: Research Information Specialist at UNC

<table>
<thead>
<tr>
<th>Department</th>
<th>TRCS Institute-429801</th>
</tr>
</thead>
<tbody>
<tr>
<td>Career Area</td>
<td>Information Technology</td>
</tr>
<tr>
<td>Posting Open Date</td>
<td>12/13/2023</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>Working Title</td>
<td>Research Informatics Specialist</td>
</tr>
<tr>
<td>Appointment Type</td>
<td>EHRA Non-Faculty</td>
</tr>
<tr>
<td>Position Number</td>
<td>20060002</td>
</tr>
<tr>
<td>Vacancy ID</td>
<td>N10007640</td>
</tr>
<tr>
<td>Full Time/Part Time</td>
<td>Full-Time Permanent</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.

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

**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.
- Demonstrable past experience in scoring 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 23: 2023 UK Study-a-Thon Lessons Learned 🙌

Dani Prieto-Alhambra
Professor of Pharmaco- and Device Epidemiology, Oxford University

Katherine Donegan
Head of Epidemiology, MHRA

Jennifer Lane
NIHR Clinical Lecturer in Trauma and Orthopaedic Surgery, Barts Bone and Joint Health, Queen Mary Univ. of London

Annika Jodicke
Senior Researcher in Pharmacoepidemiology, University of Oxford
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