Workgroup OKRs + Phenotype Phebruary Update #3

OHDSI Community Call
Feb. 20, 2024 • 11 am ET
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<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>
<tr>
<td>Mar. 5</td>
<td>New Vocabulary Release Update</td>
</tr>
<tr>
<td>Mar. 12</td>
<td>TBA</td>
</tr>
<tr>
<td>Mar. 19</td>
<td>NO MEETING</td>
</tr>
<tr>
<td>Mar. 26</td>
<td>Recent OHDSI Publications</td>
</tr>
<tr>
<td><strong>coming in April</strong></td>
<td><strong>CDM Month</strong></td>
</tr>
</tbody>
</table>
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
OHDSI Shoutouts!

Congratulations to the team of Martin Boeker, Daniela Zöller, Romina Blasini, Philipp Macho, Sven Helfer, Max Behrens, Hans-Ulrich Prokosch and Christian Gulden on the publication of Effectiveness of IT-supported patient recruitment: study protocol for an interrupted time series study at ten German university hospitals in Trials.

**STUDY PROTOCOL**

Effectiveness of IT-supported patient recruitment: study protocol for an interrupted time series study at ten German university hospitals

Martin Boeker¹ ², Daniela Zöller¹, Romina Blasini³, Philipp Macho³, Sven Helfer³, Max Behrens³, Hans-Ulrich Prokosch⁴ and Christian Gulden⁵

Abstract

**Background** As part of the German Medical Informatics Initiative, the MIRACUM project establishes data integration centers across ten German university hospitals. The embedded MIRACUM Use Case “Working in Care - IT Support for Patient Recruitment” aims to support the recruitment into clinical trials by automatically querying the repositories for patients satisfying eligibility criteria and presenting them as screening candidates. The objective of this study is to investigate whether the developed recruitment tool has a positive effect on study recruitment within a multicenter environment by increasing the number of participants. Its secondary objective is the measurement of organizational burden and user satisfaction of the provided IT solution.

**Methods** The study uses an interrupted Time Series Design with a duration of 15 months. All trials start in the control phase of randomized length with regular recruitment and change to the intervention phase with additional IT support. The intervention consists of the application of a recruitment support system which uses patient data collected in general care for screening according to specific criteria. The inclusion and exclusion criteria of all selected trials are translated into a machine-readable format using the OHDSI ATLAS tool. All patient data from the data integration centers is regularly checked against these criteria. The primary outcome is the number of participants recruited per trial and week standardized by the targeted number of participants per week and the expected recruitment duration of the specific trial. Secondary outcomes are usability, usefulness, and efficacy of the recruitment support. Sample size calculation based on simple parallel group assumption can demonstrate an effect size of δ=0.57 on a significance level of 5% and a power of 80% with a total number of 100 trials (10 per site). Data describing the included trials and the recruitment process is collected at each site. The primary analysis will be conducted using linear mixed models with the actual recruitment number per week and trial standardized by the expected recruitment number per week and trial as the dependent variables.
OHDSI Shoutouts!

Congratulations to the team of Moshe Zisser and Dvir Aran on the publication of Transformer-based time-to-event prediction for chronic kidney disease deterioration in Frontiers in Digital Health.
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>Tuesday</td>
<td>1 pm</td>
<td>Common Data Model</td>
</tr>
<tr>
<td>Wednesday</td>
<td>7 am</td>
<td>Medical Imaging</td>
</tr>
<tr>
<td>Wednesday</td>
<td>1 pm</td>
<td>Perinatal &amp; Reproductive Health</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3 pm</td>
<td>Vulcan/OHDSI Meeting (ZOOM)</td>
</tr>
<tr>
<td>Thursday</td>
<td>8 am</td>
<td>OHDSI India Community Call</td>
</tr>
<tr>
<td>Thursday</td>
<td>9 am</td>
<td>Medical Devices</td>
</tr>
<tr>
<td>Thursday</td>
<td>9:30 am</td>
<td>Data Network 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 &amp; Evaluation</td>
</tr>
<tr>
<td>Friday</td>
<td>10:30 am</td>
<td>GIS – Geographic Information System</td>
</tr>
<tr>
<td>Friday</td>
<td>11:30 am</td>
<td>Steering Group</td>
</tr>
<tr>
<td>Monday</td>
<td>9 am</td>
<td>Vaccine Vocabulary</td>
</tr>
<tr>
<td>Monday</td>
<td>10 am</td>
<td>Healthcare Systems</td>
</tr>
<tr>
<td>Monday</td>
<td>11 am</td>
<td>Data Bricks User Group</td>
</tr>
<tr>
<td>Monday</td>
<td>4 pm</td>
<td>Eyecare &amp; Vision Research</td>
</tr>
<tr>
<td>Tuesday</td>
<td>9 am</td>
<td>OMOP CDM Oncology Genomic Subgroup</td>
</tr>
</tbody>
</table>
OHDSI Global Symposium

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

Tentative symposium format:

- **Oct. 22** – tutorials/workshops
- **Oct. 23** – main conference
- **Oct. 24** – workgroup activities
Registration is now OPEN for the 2024 OHDSI Europe Symposium, which will be held June 1-3 in Rotterdam, Netherlands.

June 1 – tutorial/workshop
June 2 – tutorial/workshop
June 3 – main conference
Challenges and opportunities in adopting OMOP-CDM in Brazilian healthcare: a report from Hospital Israelita Albert Einstein

Maria Tereza Fernandes Abrahão, Uri Adrian Prync Flato, Mateus de Lima Freitas, Diogo Patrão, Amanda Gomes Rabelo, Cesar Augusto Madid Truyts, Gabriela Chiuffa Tunes, Etienne Duin, Gabriel Mesquita de Souza, Soraya Yukari Aashiro, Adriano José Pereira, Edson Amaro Jr.

Background
Brazil has one of the largest and most complex public health systems in the world, with concerns related to discrepancy in data quality, lack of interoperability among federated states and lack of informatization (paper based medical records). Broader use of real-world data and classification standards are particularly challenging, for instance, since health conditions and related services or procedures are not linked. Implementing a common standardized data model seen as an opportunity to integrate a multifaceted healthcare system in our country. This report describes our approach to overcoming data format and ontology challenges during the implementation of an OHDSI-OMOP database in a not-for-profit quaternary hospital in Brazil.

Methods
A descriptive case study of the OMOP-CDM implementation in a Brazilian hospital was performed from 2018 to 2023 after implementation of Cerner Electronic Health Record platform (2017). The local HIAE clinical data vocabulary was translated into English and imported into USAGI tool (V1.4.3) for mapping standard vocabulary OMOP by multidisciplinary groups. Additionally, we describe the direct conversion processes from the information contained in the Electronic Medical Record (EHR) and other internal databases. Data mapping, extraction, transform and load (ETL) processes were developed to enable vocabularies embedded in our EHR. The descriptions of the steps taken highlight the challenges observed, and the actions adopted.

Results
A dedicated team was established to implement OMOP-CDM, a project requiring expertise in mapping, transformation, and data quality assessment. The team included a project manager, data scientists, engineers, clinical consultants, and analyst consultants. The lack of standardized internal vocabulary and barriers related to Portuguese language hindered data interoperability. To address this, efforts were made to convert local codes into standardized vocabularies using the USAGI tool. Figure 1: overview of data flow from source database to OMOP tables Cerner HIAE. The total amount of data mapped per domain and the number of records with mapped terms are presented in Table 1: OMOP-CDM-HIAE mapping vocabularies, April 2023.

Conclusions
The value of this submission is the representation of a new energy in adoption of the OMOP-CDM and commitment to the open science principles. The feasibility of OMOP-CDM implementation in HIAE may be attributed to three key factors:

- presence of committed personnel
- adequate computer resources
- institutional and funding support

Potential impacts within the institution may go beyond research:
- exploring new uses in the management system of healthcare networks
- clinical assistance of rare conditions
- commercial partnerships.

Contact: maria.abrahao@einstein.br
Developing a pregnancy algorithm in ATLAS: Applying start date offset

**PRESENTERS:** Rupa Makadia

**INTRODUCTION:**
- Pregnancy is a complex physiological process with numerous factors that influence maternal and fetal outcomes. Understanding these factors and their impact on pregnancy outcomes requires the analysis of large-scale observational data. To facilitate such analyses, the development of pregnancy algorithms that are reliable, standardized, and portable has become crucial.
- Multiple algorithms to infer pregnancy episodes have been published, Match et al. which uses the pregnancy outcome to begin the assessment of a viable pregnancy episode and estimates the start date of pregnancy, was chosen. The development of pregnancy markers.
- We aimed to reproduce the logic from the Match et al. algorithm in ATLAS. Implementation required a new feature to be developed in the ATLAS/CRCE specification: the ability to specify a date offset to be added to a start or end date.

**METHODS:**
- A cohort of pregnancy episodes was created using ATLAS based on the logic in Match et al. algorithm (Figure 1).
- The database used in this study IBM MarketScan® Database Commercial Claims (CCAE).
- The cohort was compared to a cohort of pregnancy episodes identified via an SQL-based implementation of the pregnancy algorithm based on the number of persons, episodes, and episodes that match start and end dates.
- Patient profiles were reviewed to assess gaps in implementation of the algorithm using the two approaches.
- At a high level, the algorithm was developed using the following logic:
  1. Creating concept sets for each start marker.
  2. Adding date offset to set start markers in the index criteria (Figure 1).
  3. Defining the start marker by grouping higher order start markers to set the prior time (based on marker type).
  4. Adding criteria to specify at least 2 pregnancy episodes for the pregnancy episode to remove erroneous episodes based on outcome.
  5. Criteria is set to the outcome.
  6. Applying persistence maximum length of pregnancy dependent on outcome.

**RESULTS:**
- Table 1 shows the concordance of start and end dates of pregnancy episodes by outcome. Live birth outcomes match on end date and start date greater than 99%. Both outcomes of ectopic pregnancy and abortions does poorly with end and start date matching.
- Figure 2 shows the person distribution by outcome of the ATLAS based algorithm compared to the SQL algorithm. All outcomes have at least 90% of the correct persons identified with varying levels of missing (or incorrectly classified) persons (2.3-6.7%). All cohort-based algorithms also identify extra persons that do not appear in the SQL algorithm ranging from 1%-20%.

**CONCLUSIONS:**
- Refinement of the start marker selection needs to occur to minimize the use of the incorrect marker and misclassification of the outcomes.
- The Match algorithm uses a hierarchical approach to identify outcomes and pregnancy periods, live birth (1st in hierarchy) does better than other outcomes.
- The appearance of extra persons in each outcome needs to be evaluated for reliability. A selection of persons are likely misclassified for that outcome type.
- Alternative approaches to reduce the number of misclassifications have been explored in ATLAS.
- The adoption of standardized frameworks and open-source tools enhances the scalability and applicability of pregnancy algorithm with the goal of improving in prenatal care and maternal fetal health outcomes.

Rupa Makadia, Christopher Knoll, Patrick Ryan

**Developing a pregnancy algorithm in ATLAS will promote consistency, transparency, and collaboration among OHDSI researchers.**
Creating parsimonious patient-level prediction models using feature selection

(Aniek F. Markus, Egill A. Fridgeirsson, Ross D. Williams)
From Complexity to Clarity: Reproducible and Scalable Phenotype Development and application of LLM in a support role

**INTRO**
Pre-existing issues in depth understanding of the disease, its presentation, diagnosis, management and prognosis, and variability with medical practice across geographies and settings as well as reporting patterns to payers. Not everybody has that.

**Direct approaches result in complex, lengthy, and often not reproducible definitions with hard-to-assess performance metrics.**

Here, we introduce a structured and comprehensive process that guides researchers through the design process, addressing the limitations and arbitrariness of defining phenotype algorithms.

**METHODS**
We developed a three-step process for developing what we call a specified phenotype:

1. **Step 1:** Define the optimization requirements associated with each phenotype dimension (Table). The four requirements for optimization of the performance characteristics are:
   - Sensitivity
   - Specificity
   - Index date
   - Cohort end date

2. **Step 2:** Build a metatheme with for index criteria, inclusion/exclusion terms, and entity/relationship mappings based on the optimization requirements above.

3. **Step 3:** Fill in the details of the criteria and conditions, particularly the concepts.

**RESULTS**
We created a Standard Operating process to following step 1-3. We also created prompt-engendered GPT interaction to support this process.

**DISCUSSION**
By adopting this structured approach, we reduced variability and complexity of phenotype development while improving transparency, reproducibility, and efficiency.

It will be particularly useful to the analyst without a medical background. The incorporation of advanced language models enhances the process by automating certain aspects, reducing subjectivity, and facilitating the creation of concept sets.

The optimization requirement associated with each phenotype dimension.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Disease presentation</th>
<th>Record capture</th>
<th>Disease pattern</th>
<th>Use case context</th>
<th>Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Under-coded</td>
<td>Non-recurrent</td>
<td>Severe</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Over-coded</td>
<td>Reoccurring</td>
<td>Grade/Stage</td>
<td>optimize</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimize</td>
<td>optimize</td>
<td>optimize</td>
<td></td>
</tr>
<tr>
<td>Flavor</td>
<td>Incident</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prevent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intended use</td>
<td>Exclusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inclusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Target</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baseline characteristic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow-up characteristic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The proposed standard approach for computable phenotype development provides a systematic and transparent framework for overcoming the challenges associated with subjective and labor-intensive methods.

GPT-4 can be incorporated into the process, greatly improving researchers’ ability in performing the above steps.

(Asieh Golozar, Albert Prats Uribe, Tom Seinen, Dani Prieto-Alhambra, Peter Rijnbeek, Christian Reich)
Establishing and Operating the OHDSI Dentistry Workgroup: A Model for Other Disciplines

**INTRODUCTION**

Dentistry Workgroup Mission

To understand how dentistry can leverage observational research to improve oral health outcomes and further investigate the links between oral health and systemic disease.

A new trend in OHDSI Workgroups

In the past, most OHDSI workgroups have been focused on the data model and the research conducted within the scope of their specific field. While these workgroups have contributed significantly to conduct observational research, however, more recent trends show that the OHDSI workgroups are becoming more use case specific. This project showcases how the Dentistry Workgroup has developed and how other medical specialties can organize to adopt the OMOP-CDM for their own use cases.

**METHODS**

How to start a workgroup

1. Connect with OHDSI Global Symposium Event Planner & Teams Manager to announce interest in creating a workgroup.
2. Submit Objectives and Key Results to the OHDSI Steering Committee for consideration.
3. Once accepted, decide on a meeting schedule and announce the new workgroup in the OHDSI Forums and the OHDSI Community Call (coordinate with OHDSI Director of Communications).
4. Start meeting!

Meeting Structure

Weekly workgroup meetings to coordinate efforts, announce updates, provide educational opportunities, and a place for members to network and socialize. Formal agenda with meeting minutes.

Weekly activity report meetings to execute tasks for ongoing projects and lines of effort. Task based and geared toward productivity.

**RESULTS**

- Accomplishments since inception: 50 regularly attending members, 25 Teams/Contributors
- Three accepted posters at the 2023 OHDSI Global Symposium
- One hypothetical use case mapped, three near-use cases developed
- Ongoing discussions with the American Dental Association Standards Committee on Dental Informatics to begin development of a standard for common data models in dentistry.
- Coordinating with OHDSI Workgroups to further develop observational research capabilities in dentistry.

Top 5 Tips for running a workgroup

1. Define Clear Goals and Objectives
2. Promote communication and collaboration
3. Be inclusive and flexible in meeting logistics
4. Leverage OHDSI resources
5. Involve guests to bring a fresh perspective

**PATTERNS OF INFLUENCE**

Focuses on developing the sources of motivation and abilities for the organization, the team, and the individual.

**Organization (OHDSI)**

- Does OHDSI have the tools, infrastructure, and community available to facilitate the Dentistry WG's purpose?
- Does OHDSI want to support the Dentistry WG’s mission?

**Team (Dentistry Workgroup)**

- Does the team have the skills necessary to execute the workgroup’s mission?
- Can the team be appropriately incentivized to carry out the workgroup’s mission?

**Individual/Workgroup Member**

- Do the members understand the group's purpose and role in it?
- Is the workgroup beneficial to the member (career, education, network?)

Danielle Boyce (UCL), Robert Koski (UCL), Brock Johnson (UCL)

Johns Hopkins University School of Medicine

Tufts University School of Medicine

#OHDSI

www.ohdsi.org

#JoinTheJourney
Opening: Three Positions 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.
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
  - Postdoc: https://linyingzhang.com/files/Postdoc.pdf
  - Data analyst: https://linyingzhang.com/files/Analyst.pdf
- If interested, please send CV and cover letter to linyingz@wustl.edu
Opening: Epidemiology UX/Web Design Intern at J&J

Career Programs

Epidemiology UX/Web Design Intern

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

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 of 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

Posting Information

<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>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.

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 developing technical projects in terms of length of time, competencies and cost. Individual will be expected to manage multiple projects at once while delivering high-quality work on time.
* Excellent written and oral business communication skills. Public speaking at meetings and conferences may be required. The ability to clearly convey technical concepts to non-technical clients is a must.
Opening: Data Steward at EBMD

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

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

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

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

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

RESPONSIBILITIES AND TASKS

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

Any other announcements of upcoming work, events, deadlines, etc?
Three Stages of The Journey

Where Have We Been?
Where Are We Now?
Where Are We Going?
Learn more about all of the OHDSI workgroups

ohdsi.org/workgroups
Themis Mission & Ethos

The goal of Themis is to provide conventions on how source data should be standardized to the OMOP CDM to support the OHDSI community to generate the evidence that promotes better health decisions and better care. When there is ambiguity on how data should be inserted into the CDM, Themis will examine the issue, create a convention and document it.

We follow the FAIR principles: Findable, Accessible, Interoperable and Reusable.

Themis makes decisions for the good of the whole community. We must compromise. We can always revisit and modify the convention. Don’t let perfect be the enemy of great. And interoperability between different OMOP CDMs is great!
Themis 2023 accomplishments

- To re-establish the Themis WG
- Publish guidelines for the creation, nomination and adjudication process for Themis convention approval
- Completed some Themis work
2024 Objectives

Objectives:
1. Establish a formal project management process
2. Establish a repository for Themis conventions
Objective #1 & key results

**Objective #1:** Establish a formal project management process

**Key results:**
- Adopt a project management tool
- Implement the project management tool
- Bring a Themis issue through the whole ratification cycle utilizing the new project management tool
Objective #2 & key results

**Objective #2:** Establish a repository for Themis conventions

**Key results:**
- Identify current conventions
- Create a formal template for ratified conventions
- Create a repository for conventions
- Create a formal process to bring one ratified convention through to publication
- Publish one ratified convention
Themis working group details

- Located as a sub-group of the CDM WG in MS Teams
- Meetings: 1<sup>st</sup> & 3<sup>rd</sup> Thursday at 9:30am Eastern Time
- All are welcome!
- #JoinTheJourney
Mission statement

Support healthcare systems on their OHDSI journey
2024 Objectives

Objectives:
1. Establish a repository for community contributed, semantic mappings (Examples include: Lab results, units, payers)
2. Provide support for transforming source EHR data to the CDM
3. Establish a methodology to transition from the use of the STCM to C/CR
Objective & key result #1

Objective #1: Establish a repository for community contributed, semantic mappings for use by others in their ETL

Key results:

• Establish the format for contributions
  • Including documentation
• Successful upload of a contribution to the repository
• Successful use of semantic mapping contribution in an ETL
Objective & key result #2

Objective #2: Provide support for transforming source EHR data to the CDM

Key results:

• “Office hours” style agendas during our regularly scheduled meetings 10 times a year.

• Monitor & answer questions on the CDM Builders, Implementers, and Uncategorized forums related to source data or the ETL process
Objective & key result #3

Objective #3: Establish a methodology to transition from the use of the STCM to C/CR

Key results:
• Create a wiki on the why and how to move from STCM to C/CR
• Create scripts to move STCM mappings to C/CR
HSIG meeting details

- Located in MS Teams
- Meetings: Every other Monday at 9:00am Eastern Time
- All are welcome!
- #JoinTheJourney
Gen AI and Foundational Models Workgroup mission

To advance healthcare research and improve patient outcomes through the innovative application of generative AI and foundational models.
GenAI and FMs Workgroup objectives and key results

• Promote awareness and collaboration in GenAI and FM research
  • Maintain a comprehensive directory of ongoing methods research
  • Have at least 6 presentations of ongoing methods research (i.e. work that hasn't been published yet)
  • Average attendance of meetings >= 20 researchers

• Establish best practices and standards
  • Aid the formulation and adoption of the MEDS standard developed by Stanford
  • Draft best practices document
Natural Language Processing Workgroup
Oncology Workgroup

Enabling Observational Cancer Research

Outreach & Research WG
2nd Tuesday 3-4 pm EST
4th Wednesday 9-10 am EST

Vocab & Development Subgroup
1st Thursday 1-2 pm EST
3rd Thursday 9-10 am EST

-Omics subgroup
2nd & 4th Tuesday 3-4 pm EST

@OHDSI
www.ohdsi.org
#JoinTheJourney
Oncology WG 2024 OKR

• **Outreach & Research Subgroup**
  • Support dissemination and adoption of the Oncology conventions
    • 20% increase in the number of data partners adopting the conventions
    • Conduct three oncology studies

• **Vocab & Development Subgroup**
  • Enable and facilitate community contribution to Oncology Vocabulary and ETL
    • 25% increase in the number of contributors
    • Development of onboarding materials and tutorials for new collaborators

• **-Omics Subgroup**
  • Improve awareness and use of –Omics vocabulary
    • Publish the OMOP Genomic vocabulary paper
    • Adoptions and use of OMOP Genomic vocabulary by least five sites
WG Name: 2023 OHDSI Vaccine Vocabulary WG
WG Lead: Asiyash Lin & Yongqun “Oliver” He

2023 OHDSI Vaccine Vocabulary WG Achievements:

1. We built up a consensus of using the Vaccine Ontology (VO) to map and incorporate different vaccine terminologies and provide additional features.
2. Developed an effective pipeline by combining automatic term mapping with manual review, verification, and annotations.
3. Finished the mapping and incorporation of all vaccine terms (254 CVX vaccine terms) in CVX to VO using our pipeline.
4. The work was presented in the OHDSI 2023 symposium.

Citation: Yuanyi Pan, Warren Manuel, Rashmie Abeysinghe, Xubing Hao, Alexander Davydov, Qi Yang, Asiyah Yu Lin, Licong Cui, Yongqun Oliver He. Harmonization of OMOP vaccine-related vocabularies through the Vaccine Ontology. OHDSI 2023 Global Symposium, East Brunswick, New Jersey, Oct. 20-22, 2023. Four pages brief report and poster presentation.
Poster URL: https://www.ohdsi.org/2023showcase-27/
Objective 1: Map vaccine terms in two OMOP vaccine standards (i.e., RxNORM and RxNORM extension) to the Vaccine Ontology (VO) and add new VO terms if no mapping is found.

Key results:

1. Map/add vaccine terms in RxNORM to VO. Timeline: 1Q2024.
2. Map/add vaccine terms in RxNORM extension to VO. Timeline: 2-3Q2024.
Objective 2: Prepare an **OMOP submission** for the inclusion of VO to OMOP vocabulary as a **non-standard vaccine vocabulary**.

**Key results:**

1. Preparation for the submission. Timeline: 2Q2024.
2. Fill out forms and submit for VO inclusion to OMOP. Timeline: 2-3Q2024.
Objective 3: **Use case study and presentation.**

Key results:

1. Develop and evaluate the usages of the VO vaccine classification for OMOP EHR data analysis using IQVIA and N3C data sources. Timeline: 2-4Q2024
2. Submit abstract and present the results in **OHDSI 2024 symposium.** Timeline: 3-4Q2024
3. Prepare a journal article. Timeline: 3-4Q2024.
PLP Workgroup Details

Who are we?
We are a group of researchers interested in best practices for developing and implementing healthcare prediction models using observational data.

When do we meet?
Workgroup call is on the 2nd Wednesday of each month (next meeting is March 13th).

How to get involved?
Join the group and start posting on Teams plus join a workgroup call.
https://ohdsi.org/ > Workgroups > Join Our Workgroups > tick ‘PLP: Patient-Level Prediction’
Objective 1: Perform a deep learning network study

Key results:
• To have results back from 5 different data partners.
• Article submitted to high impact journal.
• Poster/talk at an OHDSI symposium.

See the study:
Code: https://github.com/ohdsi-studies/DeepLearningComparison
Objective 2: Improve software accessibility and simplicity

Key results:

• Get PLP into CRAN.
• Finish and publish PLP cheat sheet.
• Provide thorough documentation on going from "I can program R" to "I have put together a network study package”.
• 1+ new members run a PLP network study.
• Provide thorough documentation and example converting an externally developed, non-OMOP model into PLP/OMOP format.
Objective 3: Prospective clinical evaluation of model effectiveness/impact at 1+ site

Key results:

• Have 1+ model implemented in a pilot in 2024.
• Presentation on learnings/challenges of the pilot implementation.
Objective 4: Better publicize Delphi

Key results:
• Submit Delphi paper.
• Discuss on Tuesday weekly meetings.
• Add models form 10+ researchers in the database.
ATLAS Mission

• The ATLAS workgroup will provide a forum for the OHDSI community of developers that are interested in improving the open-source software solutions: ATLAS & WebAPI. These tools aim to provide capabilities to design standardized analytics to execute on the OMOP Common Data Model.
2024 OKRs: Atlas Working Group

1. Objective 1: Develop Atlas with standardized analytics that codify scientific best practices into consistent, reproducible and efficient processes
   1. KR 1: Define and publish a set of strategic objectives for ATLAS to clearly define the direction of the platform moving forward to aid in prioritization of future development efforts.
   2. KR 2: Define and publish a roadmap for ATLAS to align future work to the strategic objectives from KR1
   3. KR 3: Design and Execute a Strategus Study in Atlas via Arachne or some container solution capable of running Strategus

2. Objective 2: Resolve current technical issues in Atlas
   1. KR1: Eliminate ‘idle in transaction’ sessions in the database (Q2)
   2. KR2: Improve performance of permission checks in UI (Q1)
   3. KR3: Ensure cohort generation and caching functions correctly in concurrent operations. (Q2)
2024 Meeting Schedule

• Monthly ATLAS WG Meetings 1st Monday of the month at 9AM EST with the following aims
  • Highlight new features being built in the community
  • Provide updates on upcoming releases

• Weekly Developer Meetings - Tuesday, 8:30AM EST
  • Review open pull requests for Atlas/WebAPI
  • Triage/Review issues for upcoming releases
Objectives

- Educational Materials for Developers
- Host Developer Community Events
- Promote Standard Interfaces and Processes
- Measure the Health of OHDSI Software

Key Results

- Host OHDSI software training sessions for Kheiron mentorship program
- Host annual virtual OHDSI developer conference (DEVCON)
- Run monthly meetings of the OHDSI Technology Advisory Board (TAB)
Psychiatry WG 2024 OKR’s

- Psychiatry WG OKR’s
  - Further expansion of clinical vocabulary related to Psychiatry.
    - Scales and questionaries ontology
    - Improve it by classifying what this instrument (assessment/measurement) is property to be used for (what diagnosis)
  - Landscape analysis of the data currently available in OMOP format
  - Broader comparison study: comparative analysis of patients with affective disorders who used Lithium vs other standard treatment of affective disorders
    - The study is a part of Characterization of Psychiatry-specific datasets
  - To reform how the WG is run (similarly to GIS or Oncology workgroups)
    - Establishing github routine
    - Proper communication to the broader community
Landscape analysis

Landscape Analysis cohort to be defined as patients with a Psychiatry related condition or diagnosis (stratified by ICD10 subchapters):

- Demographic information of Psychiatry related patients
- Visit type and treatment setting of Psychiatry related patients
- Availability of structured and unstructured data
- Count of patients that fall into subsets of psychiatric condition (Bipolar Disorder, Suicidality, MDD, etc...)
- List of available Psychometric scales (PHQ-9, AUDIT, GAD-7, etc...)
NLP Workgroup Purpose

NLP WG exists to promote the use of textual information in electronic health records (EHRs) by developing software tools and methods to represent and utilize textual data thereby facilitating the generation of evidence for observational studies.
Objective 1: Knowledge dissemination – Contribute a chapter on NLP in the Book of OHDSI

Key results
1. Deliver the initial draft of the chapter; Timeline: 1Q2024

Objective 2: ETL for textual data representation and normalization

Key results
1. Presentation of the validation of Note_NLP proposal to CDM WG and incorporate changes to the next CDM release; Timeline: 1-4Q2024

Objective 3: Conduct multi-site clinical studies that utilize both structured and textual data

Key results
1. Oncology NLP study; Timeline: 2-4Q2023, 1-4Q2024
2. Psychiatry NLP study; Timeline: 2-4Q2023, 1-4Q2024
3. T2DM-SDoH-NLP study; Timeline: 2-4Q2023, 1-4Q2024
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