Phenotype Aphril Week 4: Evaluating AMI Phenotypes

OHDSI Community Call

April 28, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Upcoming Community Calls

Date	Topic
Apr. 28	Phenotype Aphril, Week 4: Final Evaluation and Learnings
May 5	Europe Symposium Review/Phenotype Aphril Finale
May 12	Collaborator Showcase Brainstorm (Submission Deadline is June 5)
May 19	MEDS (Medical Event Data Standard) & Potential Collaborations with OHDSI
May 26	Workgroup Spotlight: Vocabulary and Evidence Network

Three Stages of The Journey

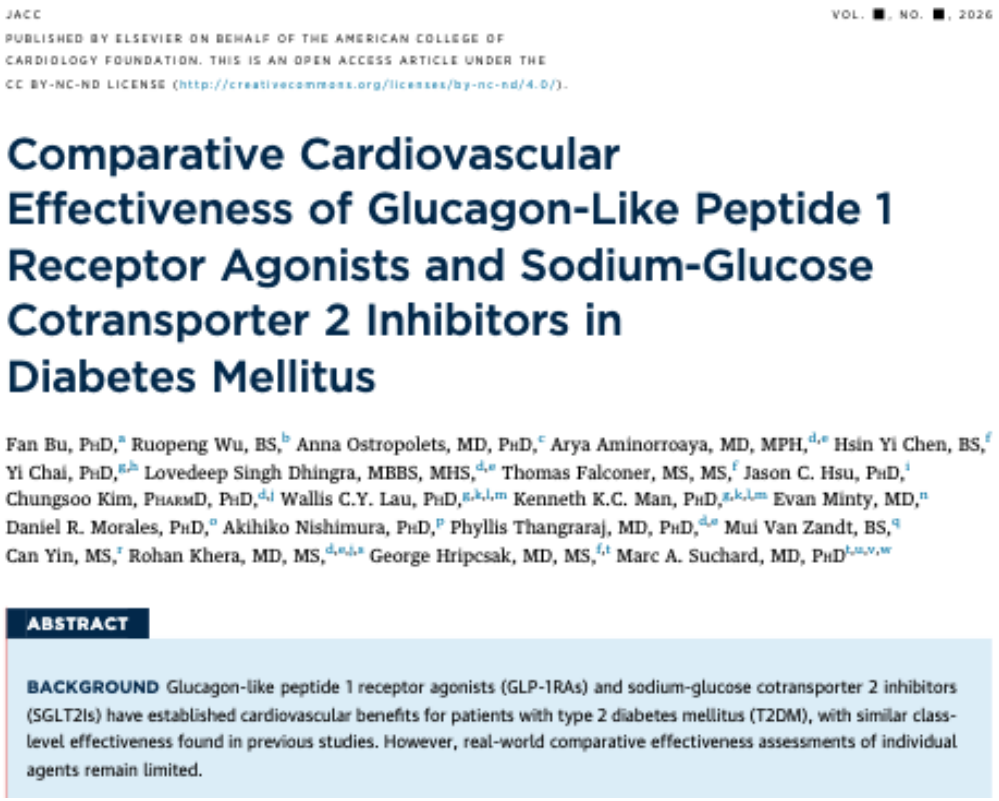
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Fan Bu, Ruopeng Wu, Anna Ostropolets, Arya Aminorroaya, Hsin Yi Chen, Yi Chai, Lovedeep Singh Dhingra, Thomas Falconer, Jason C Hsu, Chungsoo Kim, Wallis C Y Lau, Kenneth Man, Evan Minty, Daniel Morales, Akihiko Nishimura, Phyllis Thangraraj, Mui Van Zandt, Can Yin, Rohan Khera, George Hripcsak, and Marc Suchard** on the recent publication of **Comparative Cardiovascular Effectiveness of Glucagon-Like Peptide 1 Receptor Agonists and Sodium-Glucose Cotransporter 2 Inhibitors in Diabetes Mellitus** in *JACC*.



JACC
VOL. ■, NO. ■, 2026
PUBLISHED BY ELSEVIER ON BEHALF OF THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION. THIS IS AN OPEN ACCESS ARTICLE UNDER THE CC BY-NC-ND LICENSE (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Comparative Cardiovascular Effectiveness of Glucagon-Like Peptide 1 Receptor Agonists and Sodium-Glucose Cotransporter 2 Inhibitors in Diabetes Mellitus

Fan Bu, PhD,[a] Ruopeng Wu, BS,[b] Anna Ostropolets, MD, PhD,[c] Arya Aminorroaya, MD, MPH,[d,e] Hsin Yi Chen, BS,[f] Yi Chai, PhD,[g,h] Lovedeep Singh Dhingra, MBBS, MHS,[d,e] Thomas Falconer, MS, MS,[f] Jason C. Hsu, PhD,[i] Chungsoo Kim, PharmD, PhD,[d,i] Wallis C.Y. Lau, PhD,[g,k,l,m] Kenneth K.C. Man, PhD,[g,k,l,m] Evan Minty, MD,[n] Daniel R. Morales, PhD,[o] Akihiko Nishimura, PhD,[p] Phyllis Thangraraj, MD, PhD,[d,e] Mui Van Zandt, BS,[q] Can Yin, MS,[r] Rohan Khera, MD, MS,[d,e,j,s] George Hripcsak, MD, MS,[f,t] Marc A. Suchard, MD, PhD[t,u,v,w]

ABSTRACT

BACKGROUND Glucagon-like peptide 1 receptor agonists (GLP-1RAs) and sodium-glucose cotransporter 2 inhibitors (SGLT2Is) have established cardiovascular benefits for patients with type 2 diabetes mellitus (T2DM), with similar class-level effectiveness found in previous studies. However, real-world comparative effectiveness assessments of individual agents remain limited.

OHDSI Shoutouts!

Congratulations to the team of **Saketh Amasa, Vedant Agrawal, Teerth Patel, Apurvakumar Patel, Mert Karabacak, Isabelle Germano, and Konstantinos Margetis** on the recent publication of **Health literacy, healthcare access, and self-perception of health among intracranial tumor patients: an analysis of the National Institute of Health (NIH) "All of Us" research program** in the *Journal of Clinical Neuroscience*.


Journal of clinical neuroscience

Articles Publish About Contact Subscribe

ORIGINAL RESEARCH · Volume 150, 112028, August 2026 Download Full Issue

Health literacy, healthcare access, and self-perception of health among intracranial tumor patients: an analysis of the National Institute of Health (NIH) "All of Us" research program

Saketh Amasa [a] · Vedant Agrawal [a] · Teerth Y. Patel [a] · ... · Mert Karabacak [b] · Isabelle M. Germano [b] · Konstantinos Margetis [b] ... Show more

Affiliations & Notes Article Info

Get Access Cite Share Set Alert Get Rights Reprints

Abstract

Show Outline

Background and objective

Racial, ethnic, and socioeconomic disparities in healthcare outcomes are well documented; however, their influence on health literacy, patient-reported health status, and access to care among intracranial tumor (ICT) patients remains underexplored. This study aimed to characterize these disparities within a national cohort to inform equity-focused neurosurgical care strategies.

OHDSI Shoutouts!

Congratulations to the team of **Hiren N Parekh, Paul N Manson, Daniel J Lewis, Salih Colakoglu, and Sashank K Reddy** on the recent publication of **Risk factors and exploratory clustering of complications after reconstruction following Mohs surgery: A national NIH All of Us study** in the *Journal of Plastic, Reconstructive and Aesthetic Surgery*.

JPRAS BAPRAS

Submit Log in Register

Articles Publish About Contact Subscribe

Search for... Advanced search

FULL LENGTH ARTICLE · Volume 117, P42-54, June 2026

Download Full Issue

Risk factors and exploratory clustering of complications after reconstruction following Mohs surgery: A national NIH All of Us study

Hiren N. Parekh [a] · Paul N. Manson [b] · Daniel J. Lewis [c] · Salih Colakoglu [b] · Sashank K. Reddy [b]

Affiliations & Notes Article Info

Summary

Background

Mohs micrographic surgery (MMS) is an important surgical technique in the management of non-melanoma skin cancer; however, knowledge of the outcomes from Mohs reconstruction remains limited. The advent of nationwide databases coupled with advanced statistical and exploratory analytic approaches affords an opportunity to address this critical knowledge gap.

Methods

We conducted a retrospective cohort study of 6062 patients from the NIH All of Us Research Program who underwent reconstruction after MMS within 3 months of a skin cancer diagnosis. Demographics, comorbidities, and procedural data were extracted using OMOP fields. Complications within 30 days and 1 year were identified

OHDSI Shoutouts!

Congratulations to the team of **Peter J Leese, Tomas McIntee, Sydney E Browder, Mirjami Laivuori, Olamide Alabi, and Katharine L McGinigle** on the recent publication of **Extending the Observational Medical Outcomes Partnership Common Data Model to Support Observational Peripheral Vascular Disease Research** in the *Journal of Surgical Research.*

JSR Journal of Surgical Research

Submit Log in Register

Articles Publish Topics About Contact Subscribe

Search for... Advanced search

FULL LENGTH ARTICLE · Volume 322, P441-448, June 2026

Download Full Issue

Extending the Observational Medical Outcomes Partnership Common Data Model to Support Observational Peripheral Vascular Disease Research

Peter J. Leese, MSPH [a] · Tomas McIntee, PhD [a] · Sydney E. Browder, BS [b,c] · Mirjami Laivuori, MD, PhD [c,d] · Olamide Alabi, MD [e] · Katharine L. McGinigle, MD, MPH [c]

Affiliations & Notes Article Info

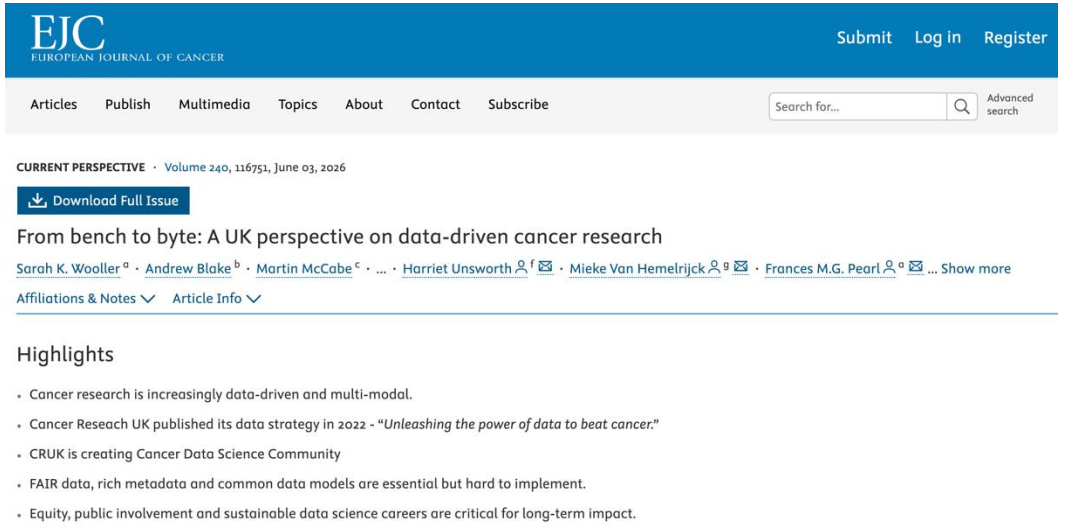
Abstract

Introduction

Peripheral artery disease (PAD) and chronic limb-threatening ischemia (CLTI) cause substantial morbidity and mortality, yet research progress is limited by fragmented, nonstandardized data. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) provides a standardized framework for electronic health record research but lacks domain-specific detail for peripheral vascular diseases. This study aimed to develop and test a vascular-specific OMOP CDM extension to

OHDSI Shoutouts!

Congratulations to the team of **Sarah K Wooller, Andrew Blake, Martin McCabe, Colin McLean, Gareth Price, Harriet Unsworth, Mieke Van Hemelrijck, and Frances M G Pearl** on the recent publication of **From bench to byte: A UK perspective on data-driven cancer research** in the *European Journal of Cancer.*



OHDSI Shoutouts!

Congratulations to the team of **Sima Mohammadi, Cori Campbell, Miriam C J M Sturkenboom, and Tiago A Vaz** on the recent publication of **A Systematic Review to Summarize and Critically Appraise Existing Phenotype Libraries Using Electronic Health Records** in *Pharmacoepidemiology & Drug Safety*.

Pharmacoepidemiology and Drug Safety

WILEY

REVIEW OPEN ACCESS

A Systematic Review to Summarize and Critically Appraise Existing Phenotype Libraries Using Electronic Health Records

Sima Mohammadi[1] | Cori Campbell[1,2] | Miriam C. J. M. Sturkenboom[1] | Tiago A. Vaz[1]

[1]Department of Data Science & Biostatistics, University Medical Center Utrecht (UMCU), Utrecht, the Netherlands | [2]Department of Health Services Research and Policy, The London School of Hygiene and Tropical Medicine, London, UK

Correspondence: Sima Mohammadi (s.mohammadi-3@umcutrecht.nl)

Received: 30 April 2025 | **Revised:** 31 December 2025 | **Accepted:** 9 April 2026

Keywords: algorithm | electronic health records | library | phenotype

ABSTRACT

Purpose: Pharmacoepidemiology and population health studies using electronic health care records (EHRs) must define study variables through available electronic data. These variables are operationalized through phenotypes, which are a defined set of criteria used to identify specific traits or medical conditions. There is diversity across phenotype libraries (collections of code lists or algorithms) which intend to standardize these sets of criteria. This review aimed to characterize the landscape of phenotype libraries and how phenotypes are constructed, validated, managed, and reused across research settings.

Methods: We conducted a systematic review of existing phenotype libraries to appraise their attributes. We systematically searched three databases (Scopus, PubMed, and Web of Science) up to November 2025 to identify studies on key characteristics of phenotype libraries. The search combined Medical Subject Headings (MeSH) terms related to "electronic health record," "phenotype algorithm," and "phenotype library". A structured hand search was performed to identify relevant web-based resources without accompanying publications to ensure comprehensive inclusion of libraries available to date. We extracted information on library size, vocabularies, phenotype construction methods, validation practices, management, and portability.

Results: Of 336 articles, 37 met eligibility criteria for full-text review, of which 25 were excluded because they were not EHR-based phenotype libraries (representing single algorithms, genomic resources, or study-specific phenotypes rather than reusable libraries), leaving 10 unique libraries described across 12 articles. A structured hand search identified seven more libraries. In total, 17 phenotype libraries met the inclusion criteria, including Education and Child Health Insights from Linked Data (ECHILD) Phenotype Code List Repository, Centralized Interactive Phenomics Resource (CIPHER), Chronic Condition Data Warehouse (CCW), ClinicalCodes Library, Clinical Classifications Software Refined (CCSR), ComPLy, CALIBER (Health Data Research UK (HDR UK) Phenotype Library or CALIBER), Jigsaw Algorithm Repository (JAR), Manitoba Centre for Health Policy (MCHP) Concept Dictionary, Open CodeLists, Observational Health Data Sciences and Informatics (OHDSI) ATLAS, PheCode, Phenotype KnowledgeBase (PheKB), Phenotype Execution and Modeling Architecture (PhEMA) Workbench, PheMap, Sharing and Reusing Computable Phenotype Definitions (SharePhe), Value Set Authority Center (VSAC). Libraries varied substantially in scope, size, and phenotype representation, including rule-based algorithms, probabilistic phenotypes, and standardized code groupings. Validation practices were heterogeneous and reported only for a subset of libraries. All the libraries utilized

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Wednesday	8 am	Psychiatry
Wednesday	10 am	Surgery and Perioperative Medicine
Wednesday	10 am	Women of OHDSI
Wednesday	1 pm	Latin America Chapter
Wednesday	7 pm	Medical Imaging
Thursday	6:30 am	India Community Call
Thursday	9 am	Phenotype Development and Evaluation
Thursday	10 am	GIS – Geographic Information System
Thursday	11 am	Perinatal and Reproductive Health
Thursday	7 pm	Dentistry
Friday	9 am	Waveform
Friday	10 am	Transplant
Friday	11:30 am	Steering Group
Monday	10 am	Healthcare Systems Interest Group
Tuesday	9 am	Data2Evidence

2026 OHDSI Global Symposium

Registration is OPEN for the **2026 OHDSI Global Symposium**, which will be held Oct. 20-22 in New Brunswick, N.J., USA.

Oct. 20: Tutorials
Oct. 21: Plenaries, Showcase
Oct. 22: Workgroup Activities

ohdsi.org/OHDSI2026

2026 OHDSI Global Symposium

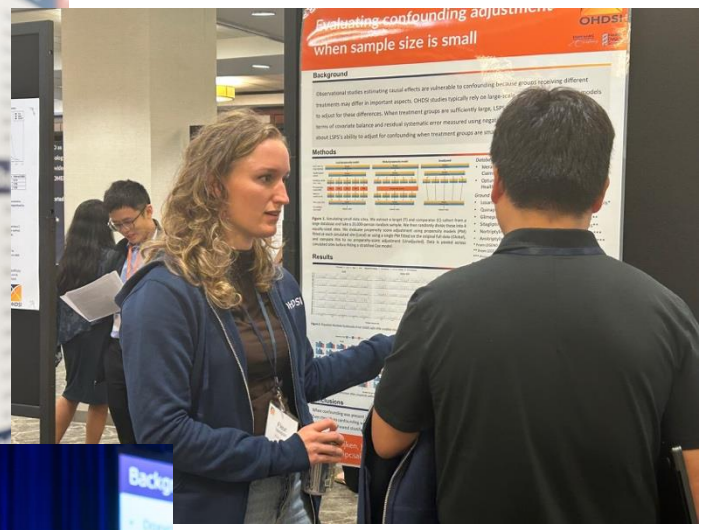
Start	End	Topic	Presenter/Lead
8:00 am	8:30 am	State of the Community	George Hripcsak
8:30 am	9:15 am	OHDSI Year In Review	Early-Stage Researcher WG
9:15 am	10:00 am	Collaborator Showcase: Posters and Demos (Session 1)	
10:00 am	11:00 am	Plenary 1: Federated Learning Meets Negative Control Calibration: Toward Reliable Multi-Site Evidence Generation	Yong Chen
11:00 am	12:00 pm	Plenary 2: Beyond the Defaults: How the OHDSI Community is Adapting, Extending, and Reimagining Its Tools	Scott Duvall
12:00 pm	1:00 pm	Network & Lunch	
1:00 pm	2:00 pm	Plenary 3: The role of national initiatives in supporting sustainability, collaboration, and growth of OHDSI	Ed Burn
2:00 pm	2:45 pm	Collaborator Showcase: Lightning Talks (Session 1)	5 presenters
2:45 pm	3:30 pm	Collaborator Showcase: Posters and Demos (Session 2)	
3:30 pm	4:15 pm	Collaborator Showcase: Posters and Demos (Session 3)	
4:15 pm	5:00 pm	Collaborator Showcase: Lightning Talks (Session 2)	5 presenters
5:00 pm	6:00 pm	Titan Awards, Closing	Patrick Ryan
6:00 pm	8:30 pm	Dinner on your own	
8:30 pm	11:30 pm	OHDSI Jam Session	Martijn Schuemie

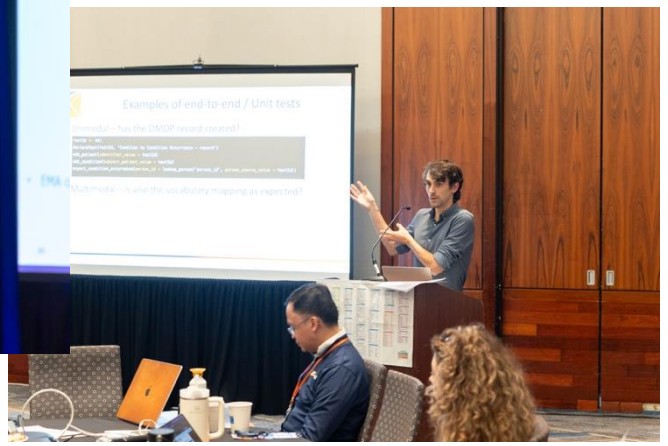
ohdsi.org/OHDSI2026

2026 OHDSI Global Symposium

The **call for participation** is open for the 2026 Global Symposium.

The submission deadline is June 5 at 8 pm ET.





38 DAYS LEFT!

ohdsi.org/OHDSI2026

2026 OHDSI Global Symposium

There are opportunities to be both a **sponsor** and an **exhibitor** at the Global Symposium.

Please reach out to **symposium@OHDSI.org** for more information.

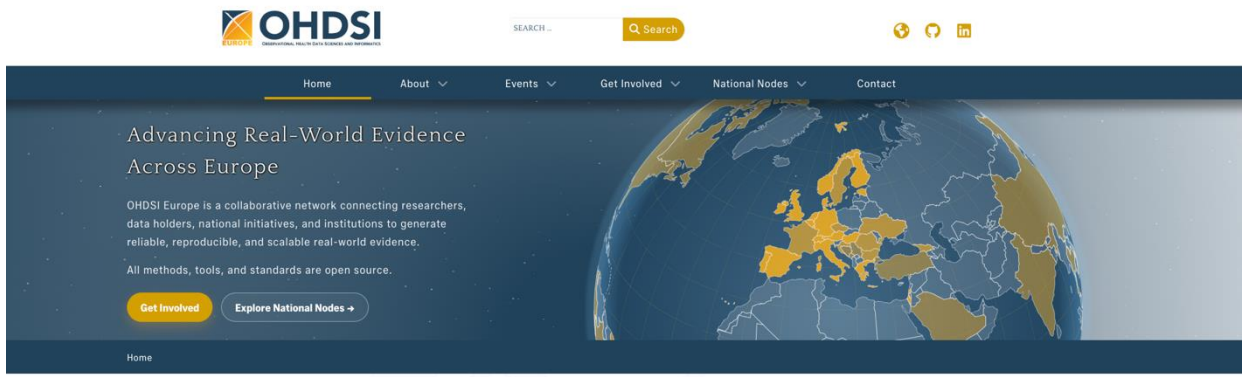
ohdsi.org/OHDSI2026

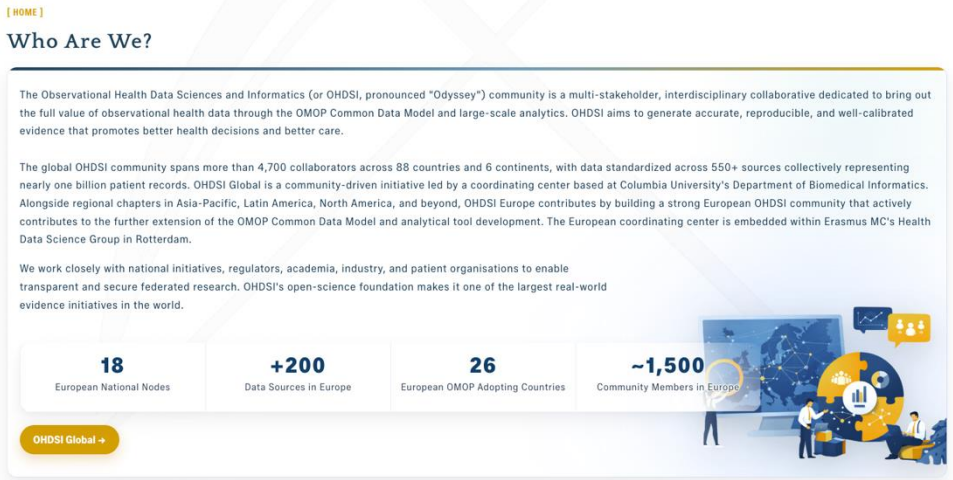
Europe Symposium Presentations Posted

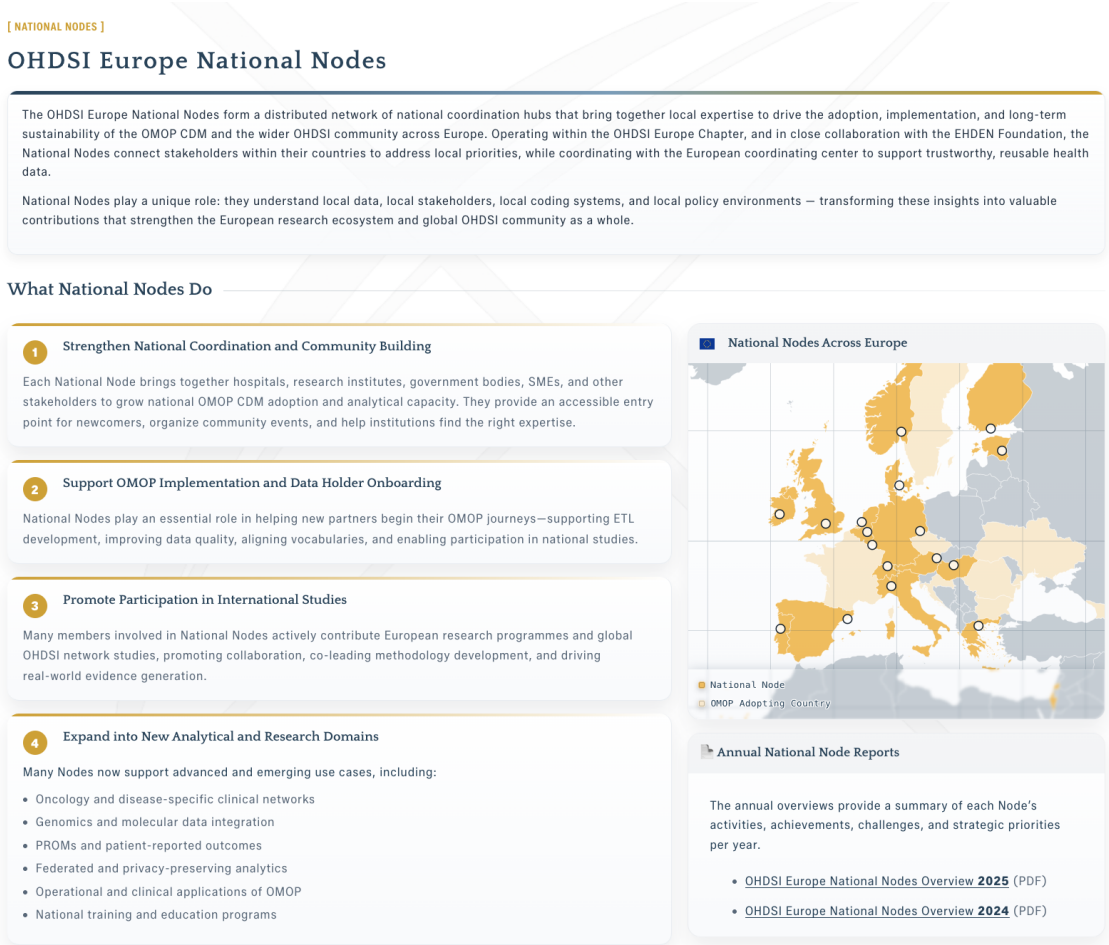
8:00	***Registration and Coffee***
9:00	**Welcome to OHDSI Europe** Dr. Renske Los & Dr. Aniek Markus, Department of Medical Informatics, *Erasmus MC*
9:05	**Journey of OHDSI** Prof. Peter Rijnbeek, Chair Department of Medical Informatics, *Erasmus MC*
9:30	**Collaborator Showcase - part 1** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC* **1.Extending FastOMOP to the OHDSI Application Layer** Niko Moeller-Grell, *King's College London*, UK **2.An Iterative Annotation Pipeline for Building Clinical Datasets and Training Information Extraction Models: The PREPARE Project** Erik Calcina, *Jožef Stefan Institute*, SLOVENIA **3.Completeness and characteristics of breastfeeding data in SIDIAP** Laura Granés, IDIAPJGol, SPAIN **4.Validation of OMOP-Based Secondary Healthcare Resource Use and Cost Estimates for Federated Health Economics Analyses in the UK** Gianluca Fabiano, *University of Oxford*, UK
10:00	**Speed networking**
10:15	***Coffee Break & posters National Nodes***
11:15	**Collaborator Showcase - part 2** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, Erasmus MC **1.Genetic Validation of Evidence-Based Phenotype Refinement in EHR Data** Mary Pat Reeves, *FinnGen, University of Helsinki*, FINLAND **2.Insights on Developing a Federated Machine Learning Prediction Model on Danish and Norwegian Colorectal Cancer Data** Samuel Wigvist, *Zealand University Hospital*, DENMARK **3.DARWIN EU® - Clozapine and the incidence of agranulocytosis over time** Dina Vojinovic, IQVIA, THE NETHERLANDS **4.Real-World Evidence on SGLT2 inhibitor utilization across cardiorenal phenotypes in Belgium: a federated OMOP-CDM and NLP-enabled hospital network study** Bart Verheyden, *AstraZeneca BeLux*, BELGIUM
11:45	**Dreaming about the OHDSI journey ahead** Dr. Patrick Ryan, Vice President, Observational Health Data Analytics, *Johnson & Johnson* & Dr. Renske Los, Department of Medical Informatics, *Erasmus MC*
12:15	***Lunch break & networking & posters/demo's*** *(Early investigator meeting - 13:00-13:45 Queen's Lounge)*
13:45	**From dreams to reality** Moderated by: Dr. Ilse Vermeulen, Department of Medical Informatics, *Erasmus MC* *Featuring OHDSI Titan Award winners:* Maxim Moinat, Department of Medical Informatics, *Erasmus MC* Anthony Sena, Global Epidemiology, *Johnson & Johnson* & Department of Medical Informatics, *Erasmus MC* Polina Talapova, *SciForce & Tufts Medical Center*
14:30	**Propositions for collaboration from the National Nodes** Ines Reinecke, *OHDSI Germany* & Aedin Culhane, *OHDSI Ireland*
14:45	***Coffee break & networking & posters/demo's***
16:15	**The OH!-Factor** *4 OHDSI OG's will take the stage for a secret mission*
17:00	**Closing** Prof. Peter Rijnbeek, Chair Department of Medical Informatics, Erasmus MC
17:15	***Networking reception***

ohdsi.org/Europe-symposium-2026

New OHDSI Europe Website







ohdsi-europe.org

Maternal Fellowship Deadline: May 15

The second **OHDSI Maternal Health Fellowship** is designed to train clinical investigators for improved maternal and neonatal care. This fellowship offers three key components: **Career Development, Practice, and Networking**.

Supported by both the OHDSI community and the NIH IMPROVE initiative, the program focuses on training clinical investigators in observational research methods to enable them to conduct reproducible research and generate real-world evidence.

First Latin America Symposium – July 30-31

UK Symposium Call for Abstracts Opens

HDR UK Event

OHDSI UK 2026

We're delighted to announce that OHDSI UK 2026 will be held on the 18th of September at the University of Nottingham. For the first time, there will also be an OMOP training day on the 17th of September.

Share this page

OHDSI (Observational Health Data Sciences and Informatics, pronounced "Odyssey") is an international community of stakeholders dedicated to unlocking the value of health data through large-scale analytics. OHDSI promotes open science and collaboration in health data research with a key focus on adoption of the OMOP Common Data Model, a global standard for harmonising data and facilitating federated analytics across institutions. **Find out more about OHDSI.**

Call for Abstracts

We invite you to submit an abstract for consideration at OHDSI UK 2026. Whether you wish to present a poster, software demo, or lightning talk, we welcome contributions from across the community. Abstract submission is available via this form, and the deadline is 1st May 2026. Please use this template to prepare your abstract and save it as a PDF, and start your file name with the surname (family name) of the presenting author.

Key dates:

Registration Opens: 20th April 2026

Registration Closes: 4th September 2026

Abstract Submission Opens: 20th March 2026

Abstract Submission Deadline: 1st May 2026

Training day: 17th September 2026

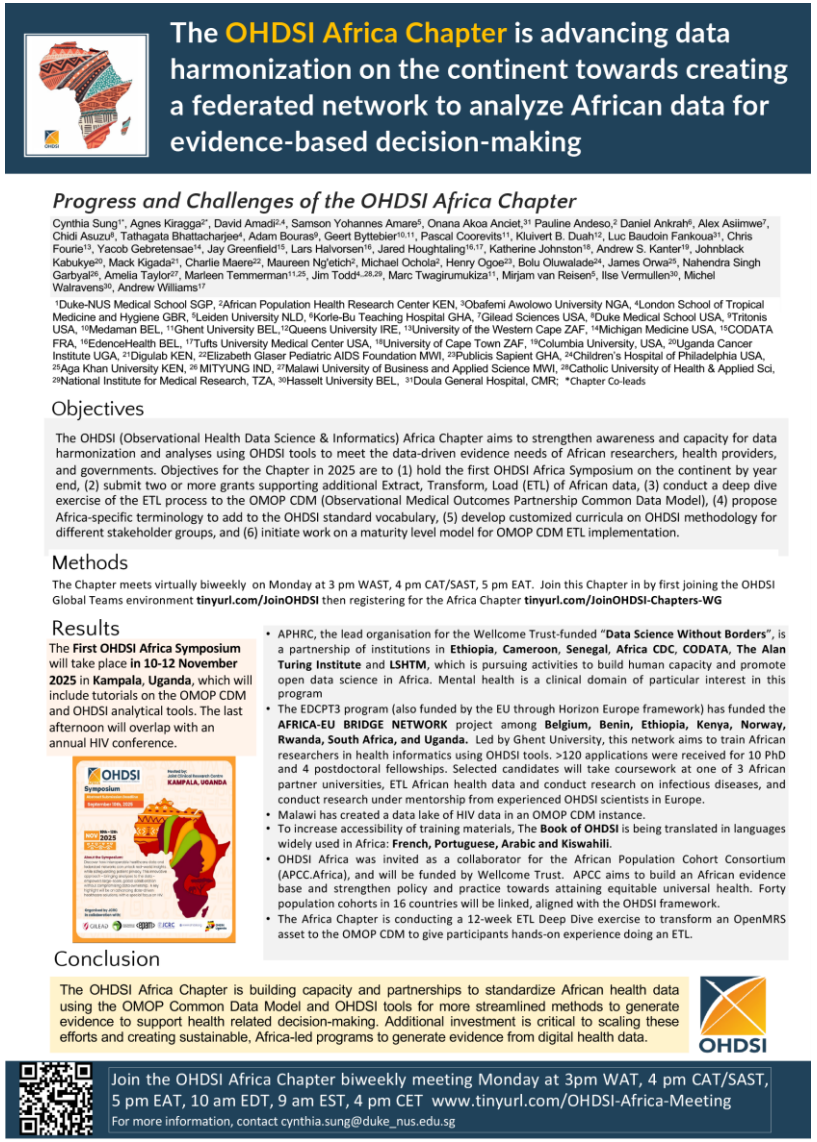
Symposium: 18th September 2026

#OHDSISocialShowcase This Week

Monday

Progress and Challenges of the OHDSI Africa Chapter

(Cynthia Sung, Agnes Kiragga, David Amadi, Samson Yohannes Amare, Onana Akoa Anciet, Pauline Andeso, Daniel Ankrah, Alex Asiimwe, Chidi Asuzu, Tathagata Bhattacharjee, Adam Bouras, Geert Byttebier, Pascal Coorevits, Kluivert B. Duah, Luc Baudoin Fankoua, Chris Fourie Yacob Gebretensae, Jay Greenfield, Lars Halvorsen, Jared Houghtaling , Katherine Johnston, Andrew S. Kanter, Johnblack Kabukye, Mack Kigada, Charlie Maere Maureen Ng'etich, Michael Ochola, Henry Ogoe, Bolu Oluwalade, James Orwa, Nahendra Singh Garbya, Amelia Taylor, Marleen Temmerman, Jim Todd Marc Twagirumukiza, Mirjam van Reisen, Ilse Vermullen, Michel Walravens, Andrew Williams)

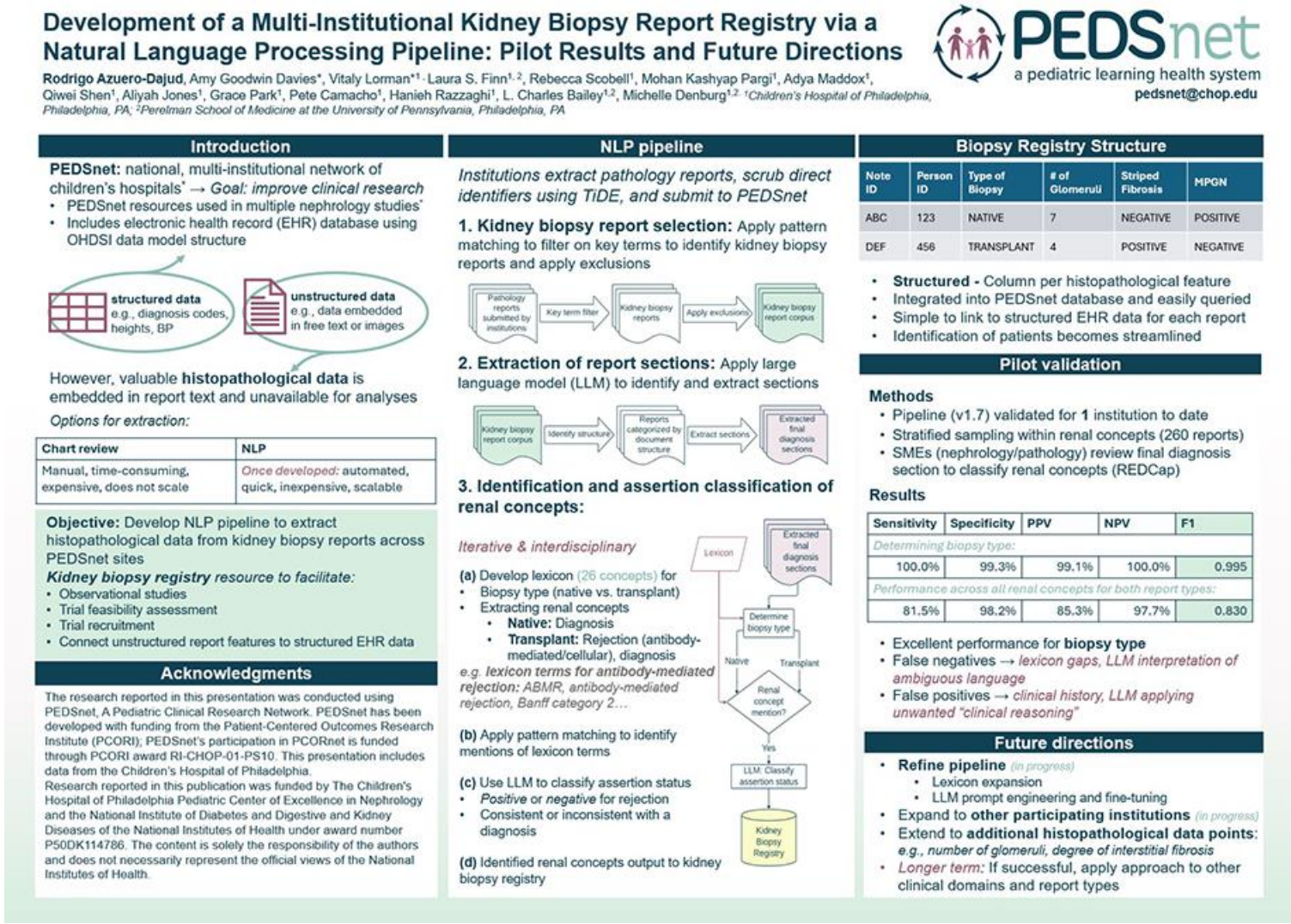


#OHDSISocialShowcase This Week

Tuesday

Development of a multi-institutional kidney biopsy report registry via a natural language processing pipeline

(Rodrigo Azuero-Dajud, Vitaly Lorman, Amy Goodwin Davies, Mohan Kashyap Pargi, Laura S. Finn, Rebecca Scobell, Adya Maddox, Qiwei Shen1, Aliyah Jones, Grace Park, Pete Camacho, Hanieh Razzaghi, Charles Bailey, Michelle Denburg)



Development of a Multi-Institutional Kidney Biopsy Report Registry via a Natural Language Processing Pipeline: Pilot Results and Future Directions

Rodrigo Azuero-Dajud, Amy Goodwin Davies*, Vitaly Lorman*[1] · Laura S. Finn[1, 2], Rebecca Scobell[1], Mohan Kashyap Pargi[1], Adya Maddox[1], Qiwei Shen[1], Aliyah Jones[1], Grace Park[1], Pete Camacho[1], Hanieh Razzaghi[1], L. Charles Bailey[1,2], Michelle Denburg[1,2] [1]Children's Hospital of Philadelphia, Philadelphia, PA; [2]Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA

PEDSnet a pediatric learning health system pedsnet@chop.edu

Introduction

PEDSnet: national, multi-institutional network of children's hospitals* → *Goal: improve clinical research*
- PEDSnet resources used in multiple nephrology studies*
- Includes electronic health record (EHR) database using OHDSI data model structure

structured data e.g., diagnosis codes, heights, BP

unstructured data e.g., data embedded in free text or images

However, valuable **histopathological data** is embedded in report text and unavailable for analyses

Options for extraction:

Chart review	NLP
Manual, time-consuming, expensive, does not scale	*Once developed:* automated, quick, inexpensive, scalable

Objective: Develop NLP pipeline to extract histopathological data from kidney biopsy reports across PEDSnet sites

***Kidney biopsy registry** resource to facilitate:*
- Observational studies
- Trial feasibility assessment
- Trial recruitment
- Connect unstructured report features to structured EHR data

Acknowledgments

The research reported in this presentation was conducted using PEDSnet, A Pediatric Clinical Research Network. PEDSnet has been developed with funding from the Patient-Centered Outcomes Research Institute (PCORI); PEDSnet's participation in PCORnet is funded through PCORI award RI-CHOP-01-PS10. This presentation includes data from the Children's Hospital of Philadelphia.
Research reported in this publication was funded by The Children's Hospital of Philadelphia Pediatric Center of Excellence in Nephrology and the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award number P50DK114786. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

NLP pipeline

Institutions extract pathology reports, scrub direct identifiers using TiDE, and submit to PEDSnet

1. Kidney biopsy report selection: Apply pattern matching to filter on key terms to identify kidney biopsy reports and apply exclusions

Pathology reports submitted by institutions → Key term filter → Kidney biopsy reports → Apply exclusions → Kidney biopsy report corpus

2. Extraction of report sections: Apply large language model (LLM) to identify and extract sections

Kidney biopsy report corpus → Identify structure → Reports categorized by document structure → Extract sections → Extracted final diagnosis sections

3. Identification and assertion classification of renal concepts:

Iterative & interdisciplinary

(a) Develop lexicon (26 concepts) for
- Biopsy type (native vs. transplant)
- Extracting renal concepts
  - **Native:** Diagnosis
  - **Transplant:** Rejection (antibody-mediated/cellular), diagnosis

e.g. lexicon terms for antibody-mediated rejection: ABMR, antibody-mediated rejection, Banff category 2…

(b) Apply pattern matching to identify mentions of lexicon terms

(c) Use LLM to classify assertion status
- *Positive* or *negative* for rejection
- Consistent or inconsistent with a diagnosis

(d) Identified renal concepts output to kidney biopsy registry

Lexicon → Extracted final diagnosis sections → Determine biopsy type → Native / Transplant → Renal concept mention? → Yes → LLM: Classify assertion status → Kidney Biopsy Registry

Biopsy Registry Structure

Note ID	Person ID	Type of Biopsy	# of Glomeruli	Striped Fibrosis	MPGN
ABC	123	NATIVE	7	NEGATIVE	POSITIVE
DEF	456	TRANSPLANT	4	POSITIVE	NEGATIVE

- **Structured -** Column per histopathological feature
- Integrated into PEDSnet database and easily queried
- Simple to link to structured EHR data for each report
- Identification of patients becomes streamlined

Pilot validation

Methods
- Pipeline (v1.7) validated for **1** institution to date
- Stratified sampling within renal concepts (260 reports)
- SMEs (nephrology/pathology) review final diagnosis section to classify renal concepts (REDCap)

Results

Sensitivity	Specificity	PPV	NPV	F1
Determining biopsy type:				
100.0%	99.3%	99.1%	100.0%	0.995
Performance across all renal concepts for both report types:				
81.5%	98.2%	85.3%	97.7%	0.830

- Excellent performance for **biopsy type**
- False negatives → *lexicon gaps, LLM interpretation of ambiguous language*
- False positives → *clinical history, LLM applying unwanted "clinical reasoning"*

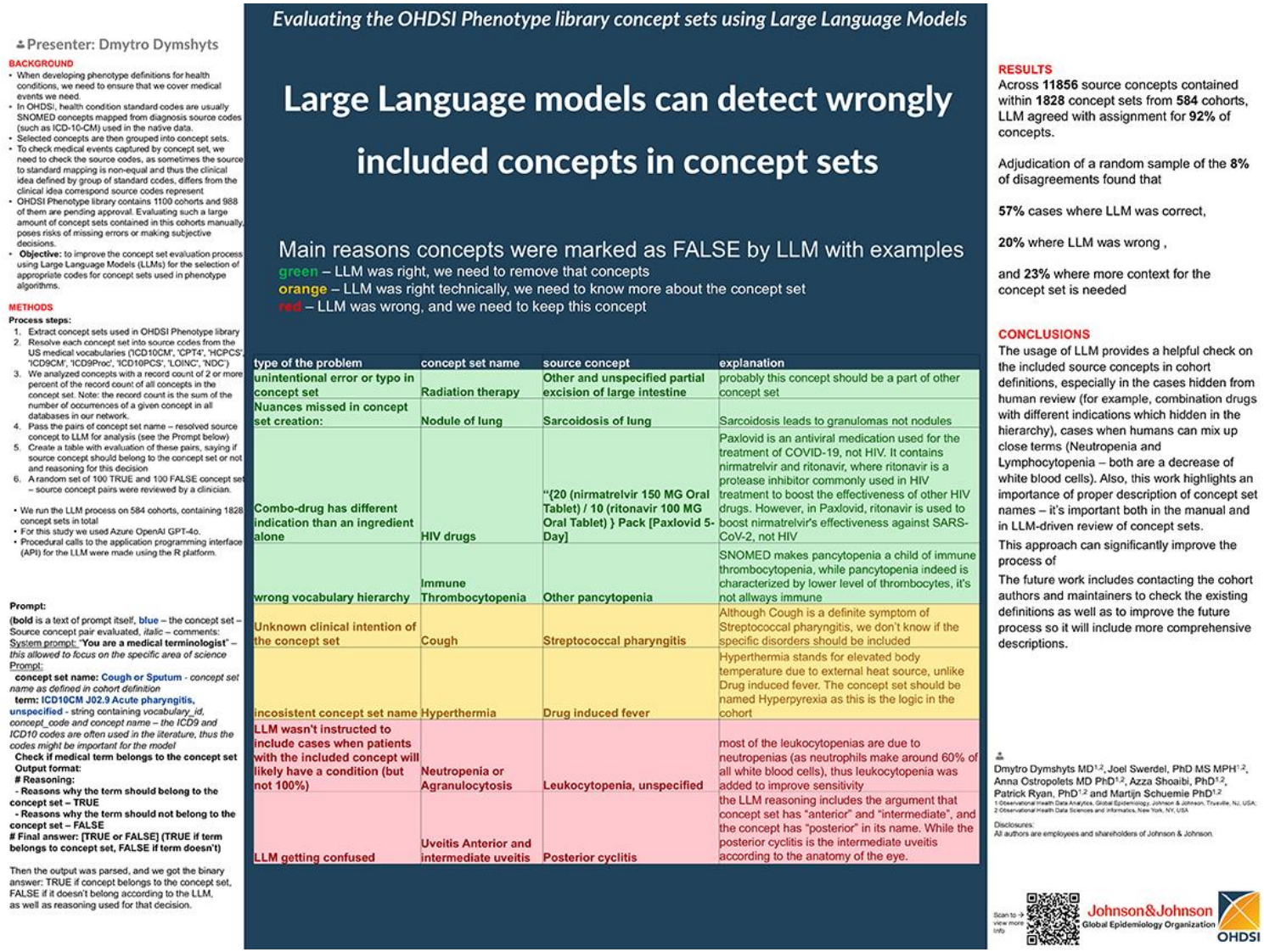
Future directions
- **Refine pipeline** *(in progress)*
  - Lexicon expansion
  - LLM prompt engineering and fine-tuning
- Expand to **other participating institutions** *(in progress)*
- Extend to **additional histopathological data points**: *e.g., number of glomeruli, degree of interstitial fibrosis*
- *Longer term:* If successful, apply approach to other clinical domains and report types

#OHDSISocialShowcase This Week

Wednesday

Evaluating the OHDSI Phenotype library concept sets using Large Language Models

(Dmytro Dymshyts, Joel Swerdel, Anna Ostropolets, Azza Shoaibi, Patrick Ryan, Martijn Schuemie)



Presenter: Dmytro Dymshyts

BACKGROUND

- When developing phenotype definitions for health conditions, we need to ensure that we cover medical events we need.
- In OHDSI, health condition standard codes are usually SNOMED concepts mapped from diagnosis source codes (such as ICD-10-CM) used in the native data.
- Selected concepts are then grouped into concept sets.
- To check medical events captured by concept set, we need to check the source codes, as sometimes the source to standard mapping is non-equal and thus the clinical idea defined by group of standard codes, differs from the clinical idea correspond source codes represent
- OHDSI Phenotype library contains 1100 cohorts and 988 of them are pending approval. Evaluating such a large amount of concept sets contained in this cohorts manually, poses risks of missing errors or making subjective decisions.
- **Objective:** to improve the concept set evaluation process using Large Language Models (LLMs) for the selection of appropriate codes for concept sets used in phenotype algorithms.

METHODS

Process steps:

1. Extract concept sets used in OHDSI Phenotype library
2. Resolve each concept set into source codes from the US medical vocabularies ('ICD10CM', 'CPT4', 'HCPCS', 'ICD9CM', 'ICD9Proc', 'ICD10PCS', 'LOINC', 'NDC')
3. We analyzed concepts with a record count of 2 or more percent of the record count of all concepts in the concept set. Note: the record count is the sum of the number of occurrences of a given concept in all databases in our network.
4. Pass the pairs of concept set name – resolved source concept to LLM for analysis (see the Prompt below)
5. Create a table with evaluation of these pairs, saying if source concept should belong to the concept set or not and reasoning for this decision
6. A random set of 100 TRUE and 100 FALSE concept set – source concept pairs were reviewed by a clinician.

- We run the LLM process on 584 cohorts, containing 1828 concept sets in total
- For this study we used Azure OpenAI GPT-4o.
- Procedural calls to the application programming interface (API) for the LLM were made using the R platform.

Prompt:

(**bold** is a text of prompt itself, blue – the concept set – Source concept pair evaluated, *italic* – comments:
System prompt: "**You are a medical terminologist**" – *this allowed to focus on the specific area of science*
Prompt:
concept set name: Cough or Sputum - *concept set name as defined in cohort definition*
term: ICD10CM J02.9 Acute pharyngitis, unspecified - string containing *vocabulary_id, concept_code and concept name – the ICD9 and ICD10 codes are often used in the literature, thus the codes might be important for the model*
Check if medical term belongs to the concept set
Output format:
 # Reasoning:
- Reasons why the term should belong to the concept set – TRUE
- Reasons why the term should not belong to the concept set – FALSE
 # Final answer: [TRUE or FALSE] (TRUE if term belongs to concept set, FALSE if term doesn't)

Then the output was parsed, and we got the binary answer: TRUE if concept belongs to the concept set, FALSE if it doesn't belong according to the LLM, as well as reasoning used for that decision.

Evaluating the OHDSI Phenotype library concept sets using Large Language Models

Large Language models can detect wrongly included concepts in concept sets

Main reasons concepts were marked as FALSE by LLM with examples
green – LLM was right, we need to remove that concepts
orange – LLM was right technically, we need to know more about the concept set
red – LLM was wrong, and we need to keep this concept

type of the problem	concept set name	source concept	explanation
unintentional error or typo in concept set	Radiation therapy	Other and unspecified partial excision of large intestine	probably this concept should be a part of other concept set
Nuances missed in concept set creation:	Nodule of lung	Sarcoidosis of lung	Sarcoidosis leads to granulomas not nodules
Combo-drug has different indication than an ingredient alone	HIV drugs	"{20 (nirmatrelvir 150 MG Oral Tablet) / 10 (ritonavir 100 MG Oral Tablet) } Pack [Paxlovid 5-Day]	Paxlovid is an antiviral medication used for the treatment of COVID-19, not HIV. It contains nirmatrelvir and ritonavir, where ritonavir is a protease inhibitor commonly used in HIV treatment to boost the effectiveness of other HIV drugs. However, in Paxlovid, ritonavir is used to boost nirmatrelvir's effectiveness against SARS-CoV-2, not HIV
wrong vocabulary hierarchy	Immune Thrombocytopenia	Other pancytopenia	SNOMED makes pancytopenia a child of immune thrombocytopenia, while pancytopenia indeed is characterized by lower level of thrombocytes, it's not allways immune
Unknown clinical intention of the concept set	Cough	Streptococcal pharyngitis	Although Cough is a definite symptom of Streptococcal pharyngitis, we don't know if the specific disorders should be included
incosistent concept set name	Hyperthermia	Drug induced fever	Hyperthermia stands for elevated body temperature due to external heat source, unlike Drug induced fever. The concept set should be named Hyperpyrexia as this is the logic in the cohort
LLM wasn't instructed to include cases when patients with the included concept will likely have a condition (but not 100%)	Neutropenia or Agranulocytosis	Leukocytopenia, unspecified	most of the leukocytopenias are due to neutropenias (as neutrophils make around 60% of all white blood cells), thus leukocytopenia was added to improve sensitivity
LLM getting confused	Uveitis Anterior and intermediate uveitis	Posterior cyclitis	the LLM reasoning includes the argument that concept set has "anterior" and "intermediate", and the concept has "posterior" in its name. While the posterior cyclitis is the intermediate uveitis according to the anatomy of the eye.

RESULTS

Across **11856** source concepts contained within **1828** concept sets from **584** cohorts, LLM agreed with assignment for **92%** of concepts.

Adjudication of a random sample of the **8%** of disagreements found that

57% cases where LLM was correct,

20% where LLM was wrong ,

and **23%** where more context for the concept set is needed

CONCLUSIONS

The usage of LLM provides a helpful check on the included source concepts in cohort definitions, especially in the cases hidden from human review (for example, combination drugs with different indications which hidden in the hierarchy), cases when humans can mix up close terms (Neutropenia and Lymphocytopenia – both are a decrease of white blood cells). Also, this work highlights an importance of proper description of concept set names – it's important both in the manual and in LLM-driven review of concept sets.

This approach can significantly improve the process of

The future work includes contacting the cohort authors and maintainers to check the existing definitions as well as to improve the future process so it will include more comprehensive descriptions.

Dmytro Dymshyts MD[1,2], Joel Swerdel, PhD MS MPH[1,2], Anna Ostropolets MD PhD[1,2], Azza Shoaibi, PhD[1,2], Patrick Ryan, PhD[1,2] and Martijn Schuemie PhD[1,2]
1 Observational Health Data Analytics, Global Epidemiology, Johnson & Johnson, Titusville, NJ, USA; 2 Observational Health Data Sciences and Informatics, New York, NY, USA

Disclosures:
All authors are employees and shareholders of Johnson & Johnson

Scan to view more info

Johnson&Johnson Global Epidemiology Organization

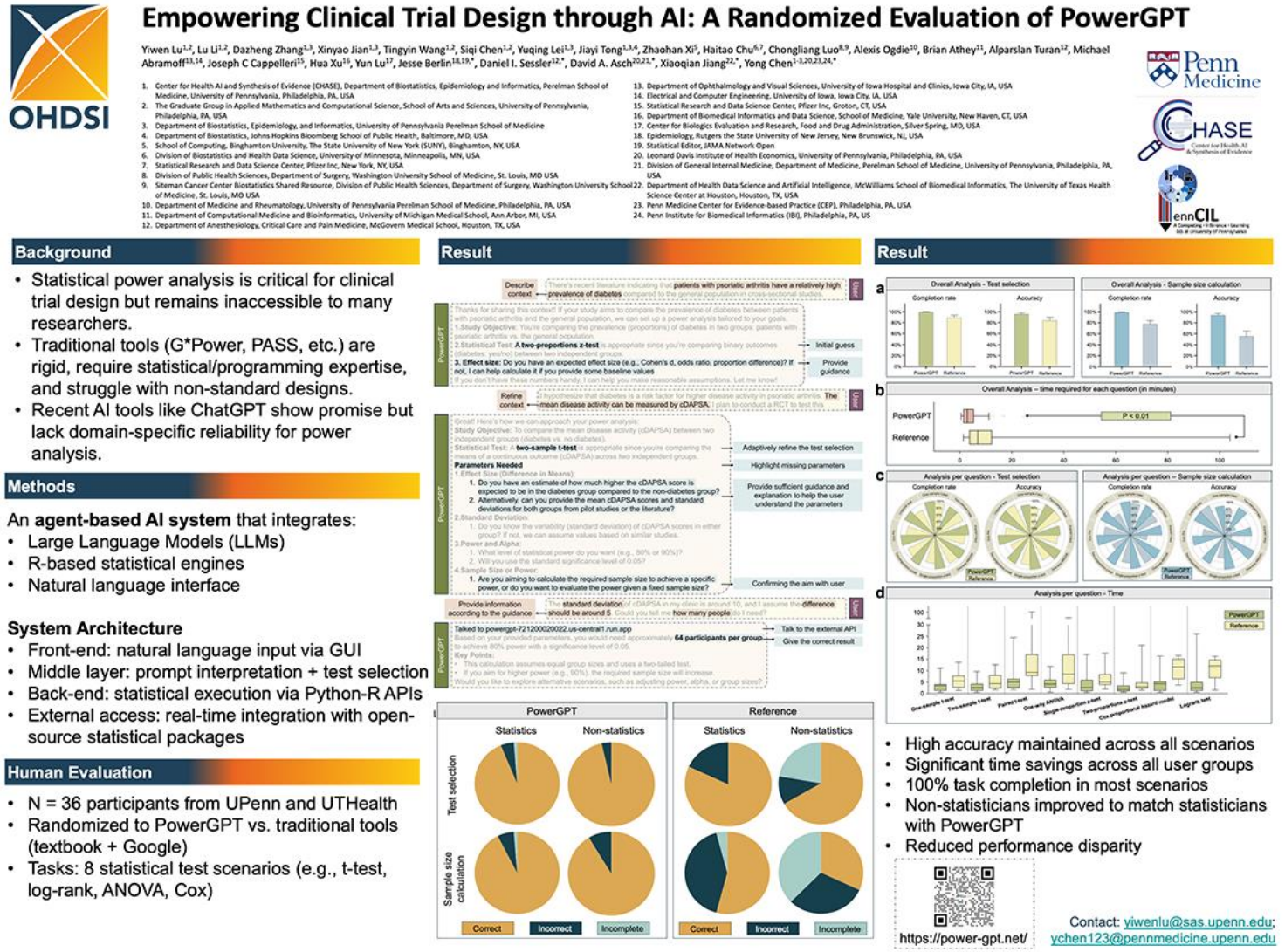
OHDSI

#OHDSISocialShowcase This Week

Thursday

Empowering Clinical Trial Design through AI: A Randomized Evaluation of PowerGPT

(Yiwen Lu, Lu Li, Dazheng Zhang, Xinyao Jian, Tingyin Wang, Siqi Chen, Yuqing Lei, Jessie Tong, Zhaohan Xi, Haitao Chu, Chongliang Luo, Alexis R. Ogdie, Brian Athey, Alparslan Turan, Michael Abramoff, Joseph C Cappelleri, Hua Xu, Yun Lu, Jesse Berlin, Daniel I. Sessler, David A. Asch, Xiaoqian Jiang, Yong Chen)

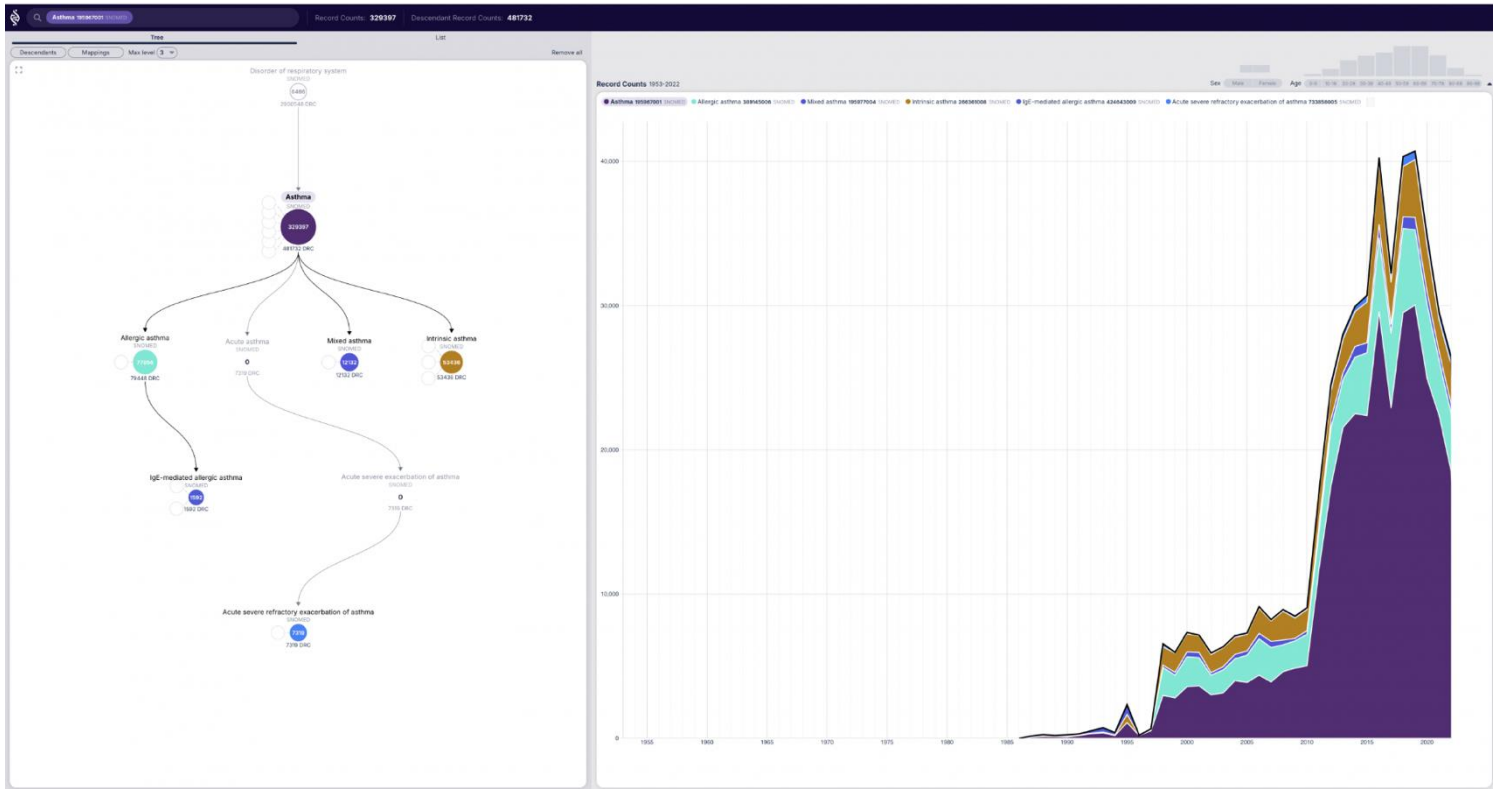


#OHDSISocialShowcase This Week

Friday

EHR Browser: A Web Tool to Explore OMOP-CDM Health Records by Concept Hierarchy, Mappings, and Temporal Trends

(Veronica Lorenzini, Javier Gracia-Tabuenca, Nicola Cerioli, FinnGen, Mary Pat Reeve)



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?

ATLAS Working Group OKR

Chris Knoll

27Apr 2026

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.

2026 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: Modernize application: JDK 21 Upgrade, permission revamp, code cleanup, remove obsolete code (Q1)
    2. KR 2: Introduce functions to support phenotype development and evaluation. (Q2-Q3)
    3. KR 3: Integrate Strategus designer. (Q4)
2. Objective 2: Continued Community Engagement
    1. KR1: Monthly WG meetings to connect with community
    2. KR2: OHDSI Tutorial for Atlas (Symposium 2026)
    3. KR3: F2F session at OHDSI Symposium to leverage GenAI to create useful applications using WebAPI.

2026 Meeting Schedule

- Monthly ATLAS WG Meetings 1st Thursday of the month at 10AM EST with the following aims
  - Highlight new features being built in the community
  - Provide updates on upcoming releases

OHDSI Workgroup
Objectives and Key Results (OKR)

Rehabilitation Workgroup

WG Name: Rehabilitation Workgroup
WG Leads: Esther Janssen & Ruud Selles

Mission statement

Promote better rehabilitation care
by leveraging the OHDSI collaborative to enable
large scale observational rehabilitation research

Two ongoing projects with a central role for OMOP-CDM

PREPARE REHAB

Umbrella
Revolutionising Stroke Care in Europe

- PREPARE is a HaDEA-Horizon European project or rehabilitation care (7 million Euro, nine countries)
- UMBRELLA is an Innovative Health Initiative (IHI) project on stroke care (26 million Euro, 20 public and private parties)
- Both generate (amongst others) federated data platform for sharing ML models, exploiting the OMOP CDM and OHDSI tools

WG Name: Rehabilitation Workgroup

WG Leads: Esther Janssen & Ruud Selles

1. Objective 1: further development of the OHDSI rehabilitation research learning community

2026 Key goals/results:

1. Establish a minimum of 3 workgroup meetings
2. Have at least 25 active working group members
3. Improve collaboration with associated OHDSI WGs to share learning experiences in similar challenges. (e.g., PROM and treatment mapping)

WG Name: Rehabilitation Workgroup

WG Leads: Esther Janssen & Ruud Selles

1. Objective 2: Address challenges and working towards best practices in using OMOP-CDM for rehabilitation research data

2026 Key goals/results:

1. Continue mapping rehabilitation-specific outcome data to the OMOP-CDM (e.g., PROMS)
2. Continue mapping rehabilitation-specific treatments to the OMOP-CDM (e.g., complex treatments, multidisciplinary treatments)
3. Share first experiences with cohort definitions and distributed analyses.

RAIIDIUS Symposium

June 9th, 2026 – New York, NY, USA

www.raiidius.org

RAIIDIUS

Responsible **AI** for **I**nfectious **D**isease **I**ntervention, **U**nderstanding, & **S**urveillance

An annual symposium series advancing responsible AI and informatics for infectious diseases.

ATTENDANCE IS FREE.

PROGRAM HIGHLIGHTS

FIRESIDE CHAT
Rochelle Walensky, MD, MPH
19th Director of US Centers for Disease Control
and Prevention (2021-2023)

WORKSHOP TUTORIAL
Introduction to OHDSI and Generating Real-World Evidence

ROUNDTABLE DISCUSSIONS

KEYNOTE ADDRESS & FIRESIDE CHAT
Akhila Kosaraju, MD
CEO and President of Phare Bio, a social venture applying AI to discover new antibiotics

PANEL DISCUSSIONS
Responsible AI & Informatics Across the STI Research & Practice Continuum

ORAL AND POSTER PRESENTATIONS

REGISTRATION AND ABSTRACT SUBMISSION ARE OPEN:
www.raiidius.org

RAIIDIUS Symposium

June 9th, 2026 – New York, NY, USA

www.raiidius.org

1. Please **share RAIIDIUS with others in your network** who may be interested.
2. If you can attend in person, please **register and submit an abstract**.
3. If interested in helping to organize an **OHDSI workgroup** focused on **infectious diseases and immunology,** please reach out to:

Harry Reyes Nieva, PhD, MAS
Division of Infectious Diseases
Columbia University Irving Medical Center
harry.reyes@columbia.edu

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-2026

Find your workgroup.

Fuel our mission.

ohdsi.org/workgroups