MEDS (Medical Event Data Standard) & Potential Collaborations with OHDSI

OHDSI Community Call
May 19, 2026 • 11 am ET

OHDSI
Observational Health Data Sciences and Informatics

Upcoming Community Calls

Date	Topic
May 19	MEDS (Medical Event Data Standard) & Potential Collaborations with OHDSI
May 26	Workgroup Spotlight: Vocabulary and Evidence Network
June 2	LLM Research Around The World, Session 1
June 9	LLM Research Around The World, Session 2
June 16	LLM Research Around The World, Session 3
June 23	CANCELLED: OHDSI Summer School at Columbia University
June 30	OMOP & OHDSI Research Spotlight

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Elisabeth Mayrhuber, Philip Stampfer, Sai Pavan Kumar Veeranki, Lukas Steininger, Stephan Winkler** on the recent publication of **Automatic ETL Pipeline Generation for Mapping Heterogeneous Clinical Data into the OMOP Common Data Model** in *Volume 33 of Studies in Health Technology and Informatics: dHealth 2026*.

dHealth 2026 139
G. Schreier et al. (Eds.)

doi:10.3233/SHTI260071

Automatic ETL Pipeline Generation for Mapping Heterogeneous Clinical Data into the OMOP Common Data Model

Elisabeth MAYRHUBER[a,1], Philip STAMPFER[b], Sai Pavan Kumar VEERANKI[c], Lukas STEININGER[a] and Stephan WINKLER[a]
[a] *University of Applied Sciences Upper Austria, Hagenberg, Austria*
[b] *Joanneum Research Forschungsgesellschaft mbH, Graz, Austria*
[c] *Steiermärkische Krankenanstaltengesellschaft mbH, Graz, Austria*

Abstract. The OMOP Common Data Model (CDM) standardizes heterogeneous clinical data for large-scale research, yet ETL development remains complex and manually intensive. We present an automated, mapping-driven ETL pipeline that converts CSV-based clinical datasets into OMOP CDM v5.4 using Python-based metadata processing and dbt-generated SQL models. The architecture separates raw, staging, intermediate, and OMOP layers in PostgreSQL and incorporates Rabbit-in-a-Hat–derived mappings and robust handling of CSV formats. By generating all transformation logic programmatically, the pipeline improves reproducibility, transparency, and maintainability. Only mapping definitions require manual creation; all downstream models are synthesized automatically, enabling efficient reuse across datasets, institutions, and domains.

Keywords. Medical Informatics, Data Warehousing, Electronic Health Records, Data Transformation

OHDSI Shoutouts!

Congratulations to the team of **Ariadna Pérez Garriga, Philipp Honrath, Stefan Wolking, Beatrice Coldewey, Susann A. Bozkir, Nils Freyer, Patrick May, Yvonne Weber, Rainer Röhrig, Myriam Lipprandt** on the recent publication of **Flexible Data Integration for Genomics-Driven Decision Support in Rare Genetic Epilepsy** in *Volume 33 of Studies in Health Technology and Informatics: dHealth 2026*.

dHealth 2026 89
G. Schreier et al. (Eds.)

doi:10.3233/SHTI260061

Flexible Data Integration for Genomics-Driven Decision Support in Rare Genetic Epilepsy

Ariadna PÉREZ GARRIGA[a], Philipp HONRATH[b], Stefan WOLKING[b], Beatrice COLDEWEY[a], Susann A. BOZKIR[a], Nils FREYER[a], Patrick MAY[c], Yvonne WEBER[b], Rainer RÖHRIG[a] and Myriam LIPPRANDT[a, 1]
[a]*Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*
[b]*Section of Epileptology, Department of Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany*
[c]*Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg*

Abstract. Background: Precision medicine for complex diseases like epilepsy requires integrating heterogeneous clinical and genomic data but interpreting numerous disease-associated genes remains challenging. **Objectives:** How can data from disparate biomedical sources be organized efficiently and flexibly to support precision medicine in early project stages in genetic epilepsy? **Methods:** We applied a three-step system design approach tailored for academic medical research, considering project requirements, available resources, and technology selection. **Results**: The EAV-hybrid model accommodated diverse clinical and genetic data while preserving flexibility for future expansion. Integration with cBioPortal enabled intuitive visualization and interpretation. The design supports future migration to standard CDMs such as OMOP or i2b2. **Conclusion:** A flexible, metadata-driven EAV-hybrid model supports rapid prototyping and structured data integration in early-stage precision medicine projects, providing an infrastructure for molecular boards and clinical decision-making for genetic epilepsies.

Keywords. Data Management, Data Standardization, Common Data Model, Precision Medicine, Epilepsy

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Vocabulary
Wednesday	10 am	Common Data Model
Wednesday	8 am	Psychiatry
Thursday	9 am	Oncology Vocabulary/Development Subgroup
Thursday	10 am	GIS - Geographic Information System
Thursday	10 am	Africa Chapter (ZOOM)
Thursday	11 am	Themis
Thursday	12 pm	Medical Devices
Thursday	12 pm	HADES
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering Group
Monday	9 am	Vaccine Vocabulary
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	9 am	Open EHR and OMOP
Tuesday	10 am	CDM Survey

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.

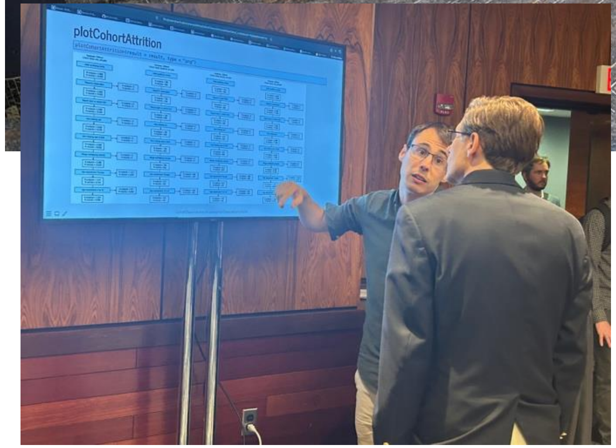
17 Days Remaining!

ohdsi.org/OHDSI2026

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 Symposium Tutorials – Session 1

- An Introduction to the Journey from Data to Evidence Using OHDSI
- An Introduction to ATLAS
- Bringing FAIR to Imaging Research with the Medical Imaging OMOP Extension
- Complex Phenotyping at Scale with and without LLMs Using PhenotypeR
- OHDSI Leadership Storytelling Workshop
- Mastering OMOP: Transforming EHR Data with Practical Strategies, Best Practices, and OHDSI Integration

2026 Symposium Tutorials – Session 2

- Building and Using the OHDSI Evidence Network: From Data Partner to Federated Study Execution
- From Multi-Modal Data to Real-World Evidence: Hands-on with the Data2Evidence Platform for OMOP Data Curation and Analytics
- Integrating Geospatial Data Into OMOP CDM
- Introduction to OHDSI Phenotype Development & Evaluation
- OHDSI Standardized Vocabularies on FHIR: A Deep Dive Using the Echidna Terminology Server
- Using OMOP Model in Registry Context & Clinical Trials Standardization Context: Conventions, Past Use Cases, SDTM & Regulatory Consideration, Challenges

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

Opening: Data Science Engineer

Position Summary

The Department of Biomedical Informatics at Columbia University is seeking a highly motivated data science engineer to support large-scale observational research within the OHDSI (Observational Health Data Sciences and Informatics) network. This role will focus on the design, implementation, and execution of distributed network studies using electronic health record (EHR) and administrative claims data to generate real-world evidence.

The successful candidate will contribute to characterization, population-level estimation (causal inference), and patient-level prediction analyses across multi-institutional data networks. This position offers a unique opportunity to work at the intersection of biomedical informatics, data science, and clinical research within a leading academic medical center.

This position is a full-time two-year position with a possibility of an extension, contingent on available funding.

Responsibilities

Key Responsibilities

- Design and implement observational network studies using distributed EHR and administrative claims data
- Conduct large-scale characterization, comparative effectiveness and safety estimation, and patient-level prediction analyses
- Develop reproducible analytic pipelines using R and SQL in relational database environments
- Apply and evaluate methods from causal inference (e.g., confounding control, bias assessment, sensitivity analyses)
- Apply machine learning approaches for predictive modeling using high-dimensional healthcare data
- Work with standardized data representations, including the OMOP Common Data Model and standardized clinical vocabularies for conditions, drugs, procedures, and measurements
- Collaborate with interdisciplinary teams including clinicians, statisticians, data engineers, and informaticians
- Contribute to scholarly outputs including manuscripts, presentations, and open-source analytic tools
- Support transparent, reproducible, and scalable research practices across distributed data networks

CAREERS AT COLUMBIA UNIVERSITY

Search All Jobs | Featured Career Areas | Temporary Opportunities | My Application | My Profile

Data Science Engineer

- 557250
- Columbia University Medical Center
- Biomedical Informatics
- Full Time
- Opening on: May 14 2026
- Grade 106

Add to favorites | View favorites

- Job Type: Officer of Administration
- Bargaining Unit:
- Regular/Temporary: Regular
- End Date if Temporary:
- Hours Per Week: 35
- Standard Work Schedule: Monday - Friday
- Building: PH-20
- Salary Range: $160,000 - $180,000

The salary of the finalist selected for this role will be set based on a variety of factors, including but not limited to departmental budgets, qualifications, experience, education, licenses, specialty, and training. The above hiring range represents the University's good faith and reasonable estimate of the range of possible compensation at the time of posting.

Maternal Fellowship Deadline: May 31

2026 Maternal Health Fellowship

Career Development

- Create evidence from real-world data
- Leverage standard data models for reproducible research
- Build skills on effective network studies

Practice

- Design effective observational research protocols
- Master OHDSI tools
- Write papers & grants

Networking

- Build relationships with mentors & fellow learners
- Coordinate with colleagues in the OHDSI data network, spanning 450 sites worldwide & 960 million unique patients

Want to build your career?

Generate reproducible evidence by leading multi-institutional studies!

Find out more and apply via this link by May 31st, 2026 !

First Latin America Symposium – July 30-31

Registration is open for the first OHDSI Latin America Symposium, taking place July 30-31 in Salvador, Brazil.

Day 1

Strategic panels with government, academia and industry

Thursday, July 30, 2026

Opening and keynote
Common Data Model for Health Equity: the Role of Latin America.

Panel 1 — Health data interoperability and standards
Panelists from the Ministry of Health, Bahia State Health Department, PAHO and Latin American Governments.

Panel 2 — The power of administrative data for health research
Panelists from the Ministry of Health, CONASS, Fiocruz, Latin American Governments, Industry and OHDSI Global.

Panel 3 — The future of interoperability in healthcare in Latin America
A public–private debate.
Panelists from the Ministry of Health, CONASS, Fiocruz, private hospitals and Latin American Governments.

Day 2

Hands-on workshops and scientific collaboration

Friday, July 31, 2026

Introductory OMOP CDM workshops
- Introduction to OMOP
- Building cohorts with OHDSI tools

Parallel tracks of specialized workshops
- ETL to OMOP
- Scientific collaboration

Closing
Future perspectives and next steps for the OHDSI Latin America community.

ohdsilatam.org

Columbia DBMI Summer School

The 2026 Summer School in Observational Health Data Science & Informatics, AI, and Real World Evidence

COLUMBIA | COLUMBIA UNIVERSITY DEPARTMENT OF BIOMEDICAL INFORMATICS

OHDSI

June 22–26, 2026, Columbia Biomedical Informatics

The Columbia OHDSI Summer School provides health professionals, researchers, and industry practitioners with an immersive, hands-on training to working with real-world health data and generating real-world evidence (RWE). Participants will explore the types of healthcare data captured during routine clinical care—such as electronic health records and administrative claims—and learn how to standardize these data using the OMOP Common Data Model to support collaborative, distributed research as part of a data network.

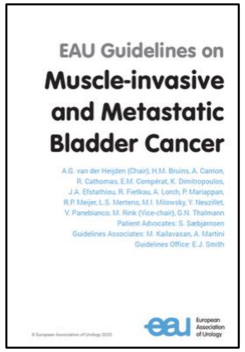
Over the course of the week, participants will engage with three real-world analytic use cases:

- **Clinical characterization** – using descriptive epidemiology to study disease natural history and treatment patterns
- **Population-level estimation** – applying causal inference to assess drug safety and comparative effectiveness
- **Patient-level prediction** – leveraging machine learning for early disease detection and precision medicine

Participants will be guided through the full RWE study lifecycle: from designing observational studies tailored to each use case, to applying open-source tools form the OHDSI community, and executing analyses across real-world data sources.

The curriculum combines foundational lectures on analytical methods with hands-on, interactive, faculty-led group exercises. In addition, participants will have dedicated time to develop and advance their own study concepts with personalized feedback and mentoring.

FALCON Bladder: RWE to influence Bladder Cancer Guideline

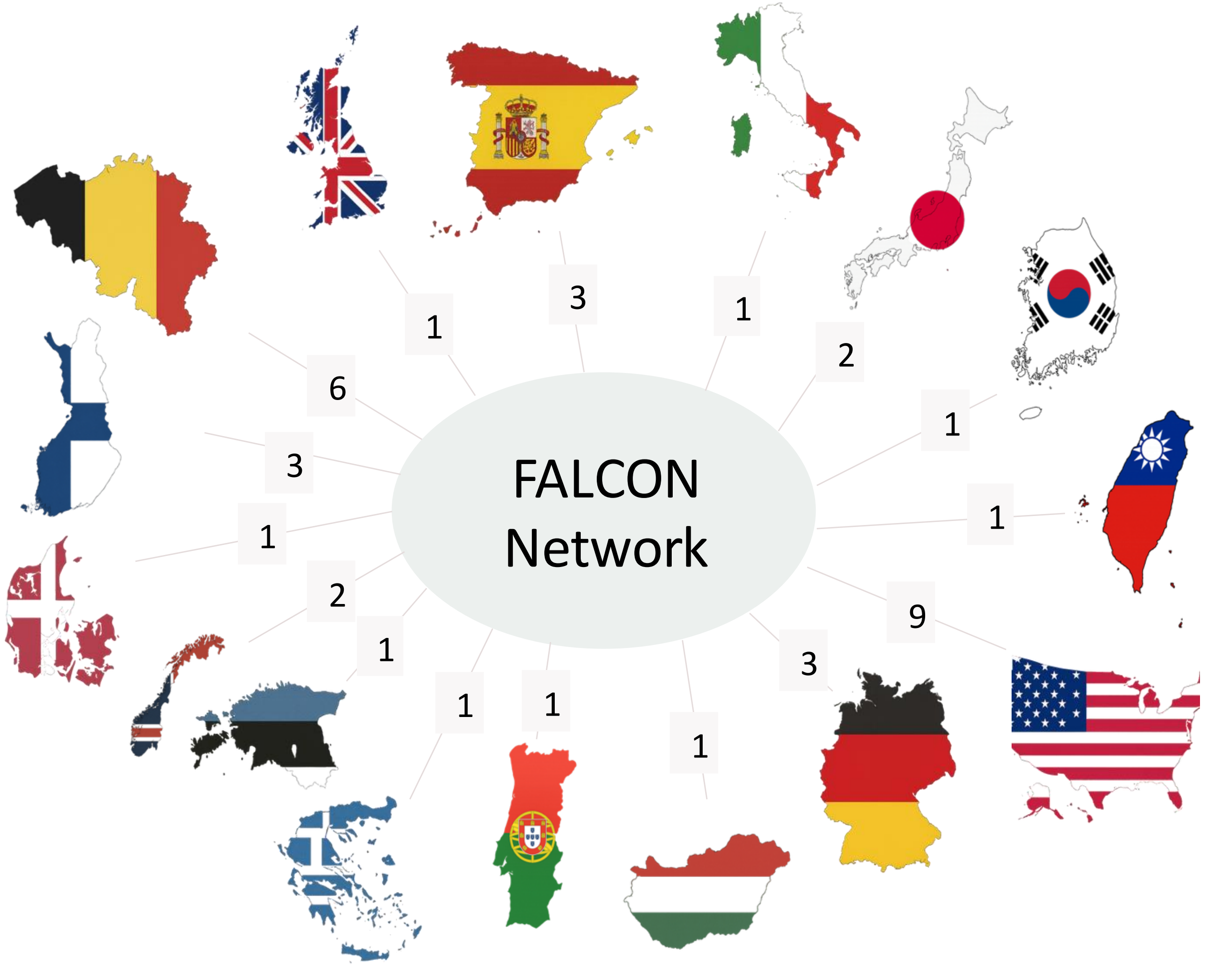


EAU Guidelines on
Muscle-invasive and Metastatic Bladder Cancer

A.G. van der Heijden (Chair), H.M. Bruins, A. Carrion, R. Cathomas, E.M. Compérat, K. Dimitropoulos, J.A. Efstathiou, R. Fietkau, A. Lorch, P. Mariappan, R.P. Meijer, L.S. Mertens, M.I. Milowsky, Y. Neuzillet, V. Panebianco, M. Rink (Vice-chair), G.N. Thalmann
Patient Advocates: S. Sæbjørnsen
Guidelines Associates: M. Kailavasan, A. Martini
Guidelines Office: E.J. Smith

European Association of Urology

Guideline-relevant RWE



16 countries, 37 sites

2318P FALCON: A novel high-quality cancer network for RWE, Golozar, A. et al. Annals of Oncology, Volume 36, S1273

FALCON Symposium & Bladder Cancer Studyathon

Sep 28 – Oct 1, 2026 Dunden, Antwerp, Belgium

Sep 28: Symposium — Public afternoon with oncologists, data scientists, regulators, and hospital leaders. Short talks and a network drink.

Sep 29 – Oct 1: Studyathon — Three hands-on days: cohort building, analytics, and abstract drafting.

Study related materials:

- All study related materials are here: GDE2025 - Bladder cancer treatment | OHDSI | Microsoft Teams
- GitHub repo: FALCON Bladder Git Repo

Register now

Details, agenda updates, venue information, and registration are available on the FALCON website.

www.falcon-network.org

#OHDSISocialShowcase This Week

Monday

Bridging Standards: Transforming Consolidated Clinical Document Architecture (C-CDA) Data via Health Information Networks to OMOP

(**Xiaohan Tanner Zhang**, Chris Roeder, Stephanie Hong, Thanaphop Na Nakhonphanom, Adam Lee, Richard Moffitt, Josh Lemieux, James Cavallon, Monique Bangudi, Lakshmi Anandan, Rob Schuff, Bill Hogan, Chris Chute, Emily Pfaff, Melissa Haendel)

PRESENTER: Tanner Zhang MD, MS

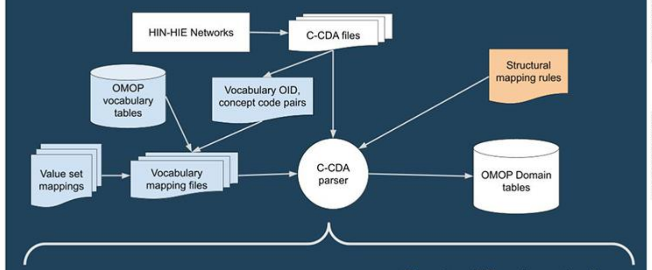
INTRODUCTION

- **The Problem:** The All of Us (AoU) program collects patient data from many sources to populate an OMOP data warehouse for use by researchers. The Center for Linkage and Acquisition of Data (CLAD) program acquires additional data through many avenues including HIN-HIE networks where much of it requires parsing the HL7 C-CDA format. The format is complex, and the data are often inconsistent, presenting a challenge for harmonizing and mapping structure and vocabulary.
- **The Goal:** Create an effective and transparent mechanism for transforming high-volumes of C-CDA data into the OMOP CDM to enable critical research (phenotyping, comparative studies, etc.)

METHODS

- Approach: A rule-based, multi-stage ETL pipeline using Python and SQL, scaled using PySpark.
- Core Components:
  - Vocabulary Mapping: Standardizes terminologies using custom maps, OID translation, and advanced code/format normalization.
  - Structural Mapping: rule driven process uses XPath to parse C-CDA structure with transparent mapping.
- Key Graphics:
  - The ETL Flowchart visually illustrates this two-part method.
  - An abbreviated set of mapping rules for the observation table shows the source for observation_id, person_id and observation_concept_id in the C-CDA document drawing from an observation element within a results section.
- Vocabulary Nuance:
  A custom C-CDA Value Set Mapping Table was created to accurately handle non-standard HL7-specific and EPIC internal vocabularies, increasing semantic precision.
- **Advanced Data Cleaning:**
  - Code System Normalization: Standardizes vocabulary names (e.g., 'CPT' vs. 'CPT4').
  - Pattern Recognition: Uses regex to automatically correct misclassified codes (e.g., ICD-9 vs. ICD-10).
  - Format Normalization: Standardizes various NDC and ICD code formats.
- **Flexible Structural Rules:**
  Concise rules show mapping from C-CDA to OMOP.
  The rules use distinct field types for adaptability: FIELD (direct copy), DATE (parses dates), DERIVED (handles complex transformations), etc.
- **Key Takeaway:**
  The successful pilot offers a flexible and viable mechanism to convert C-CDA to OMOP, providing a reusable framework for broader OHDSI initiatives.
- **Source Code Availability**
  A PyPi package for structural mapping is in development.

Bridging Standards: Transforming Consolidated Clinical Document Architecture (C-CDA) Data via Health Information Networks to OMOP



Simplified C-CDA Document Results Section (input)

```
<section>
  <templateId  root="2.16.840.1.113883.10.20.22.2.3.1"/>
  <code  code="30954-2"  codeSystem="2.16.840.1.113883.6.1"/>
  <entry  typeCode="DRIV">
    <organizer  classCode="BATTERY"  moodCode="EVN">
      <templateId  root="2.16.840.1.113883.10.20.22.4.1"/>
      <id  root="7d5a02b0-67a4-11db-bd13-0800200c9a66"/>
      <code xsi:type="CE" code="43789009"
            codeSystem="2.16.840.1.113883.6.96"/>
      <statusCode  code="completed"/>
      <component>
       <observation  classCode="OBS"  moodCode="EVN">
         <templateId  root="2.16.840.1.113883.10.20.22.4.2"/>
         <id  root="107c2dc0-67a5-11db-bd13-0800200c9a66"/>
         <code xsi:type="CE"  code="30313-1"
               codeSystem="2.16.840.1.113883.6.1"> </code>
         <statusCode code="completed"/>
         <effectiveTime value="20120810"/>
         <value xsi:type="PQ" value="10.2"  unit="g/dl"/>
```

```
def codemap_xwalk_concept_id(args_dict):
  id_value = _codemap_xwalk(args_dict['vocabulary_oid'],
                            args_dict['concept_code'],
                            'target_concept_id',
                            args_dict['default'])
  if id_value is not None:
    return int32(id_value)
  else:
    return None
```

Simplified OMOP Measurement table (output)

measurement_id	person_id	measurement_concept_id	value_as_number
(hash of) 107c2dc0-67a5-11db -bd13-0800200c9a66	(not shown)	3002173	10.2

Structural Mapping rules for Measurement table

```
'Measurement': {
  'root': {
    'config_type': 'ROOT',
    'expected_domain_id': Measurement',
    'element':
     ("/ClinicalDocument/component/"
      "structuredBody/component/section/"
      "/entry/organizer/component/observation")
   },
  'measurement_id_root': {
    'config_type': 'FIELD',
    'element': id',
    'attribute': 'root' },
  'measurement_id_extension': {
    'config_type': 'FIELD',
    'element': 'id',
    'attribute': 'extension' },
  'measurement_id': {
    'config_type': 'HASH',
    'fields' : ['measurement_id_root', measurement_id_extension'],
    'order': 1 },
  'person_id': {
    'config_type': 'FK',
    'FK': 'person_id',
    'order': 2 },
  'measurement_concept_code': {
    'config_type': 'FIELD',
    'element': "code" ,
    'attribute': "code" },
  'measurement_concept_codeSystem': {
    'config_type': 'FIELD',
    'element': "code",
    'attribute': "codeSystem" },
  'measurement_concept_id': {
    'config_type': 'DERIVED',
    'FUNCTION': VT.codemap_xwalk_concept_id,
    'argument_names': {
      'concept_code': measurement_concept_code',
      'vocabulary_oid': 'measurement_concept_codeSystem'
    },
    'order': 3 },
  'value_as_number': {
    'config_type': 'FIELD',
    'data_type': 'FLOAT',
    'element': 'value' ,
    'attribute': "value",
    'order': 9
  },
}
```

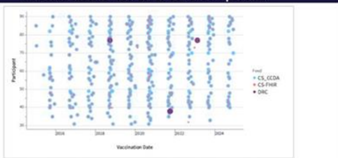
Take a picture to **download** the **full paper**

RESULTS

Data Source	OMOP Table	Total Records	Mapping Rate	Notes
Site B	Drug Exposure	52,823	100%	Complete mapping achieved.
Site B	Measurement	5,329,582	100%	Complete mapping achieved.
Site B	Condition	235,216	100%	Complete mapping achieved.
Site B	Procedure	193,634	100%	Complete mapping achieved.
Site B	Visit Occurrence	903,854	96.20%	High mapping success.
Site B	Observation	529,567	6.80%	Unmapped due to non-standard encodings.

Data Source	OMOP Table	Total Records	Mapping Rate	Notes
Site A	Drug Exposure	2,020,862	100%	Complete mapping achieved.
Site A	Measurement	842,392	100%	Complete mapping achieved.
Site A	Condition	49,610	100%	Complete mapping achieved.
Site A	Procedure	383,513	100%	Complete mapping achieved.
Site A	Observation	163,124	99.96%	Minor unmapped due to data quality.
Site A	Visit Occurrence	708,297	0.01%	Significant mapping limitation identified.

Cedars Sinai FHIR vs Cedars Sinai CCDA vs DRC: Influenza vaccine records for 60 random patients



Based on the relative lack of DRC data shown here, we assume that the Cedars Sinai submission to the DRC may not contain vaccine data. **Cedars Sinai CCDA and FHIR feeds clearly add significant numbers of vaccine records to the data already held by the DRC.** CCDA and FHIR mostly overlap here, with a few exceptions--e.g., patient 70, whose vaccines only appear in the CCDA feed.

LIMITATIONS

- Mapping non-standard vocabularies can require extensive manual effort.
- Identifiers within the documents can have varying quality. On-going work addresses these issues with data providers, and investigates ways to mitigate them.
- On-going work retrieves and counts concepts and values directly and simply from the C-CDA documents for comparison in data quality efforts that validate the resulting OMOP tables contain what the documents do.

Xiaohan Tanner Zhang MD, MS, Chris Roeder MS, Stephanie Hong MS, FAMIA, Thanaphop Na Nakhonphanom MS, MMedSc, MD, Adam Lee PhD, Richard Moffitt PhD, Josh Lemieux BA, James Cavallon BS, Monique Bangudi MPH, Lakshmi Anandan MPH, Rob Schuff MS, Bill Hogan MD, MS, Chris Chute MD, DrPH, Emily Pfaff MS, PhD, Melissa Haendel PhD

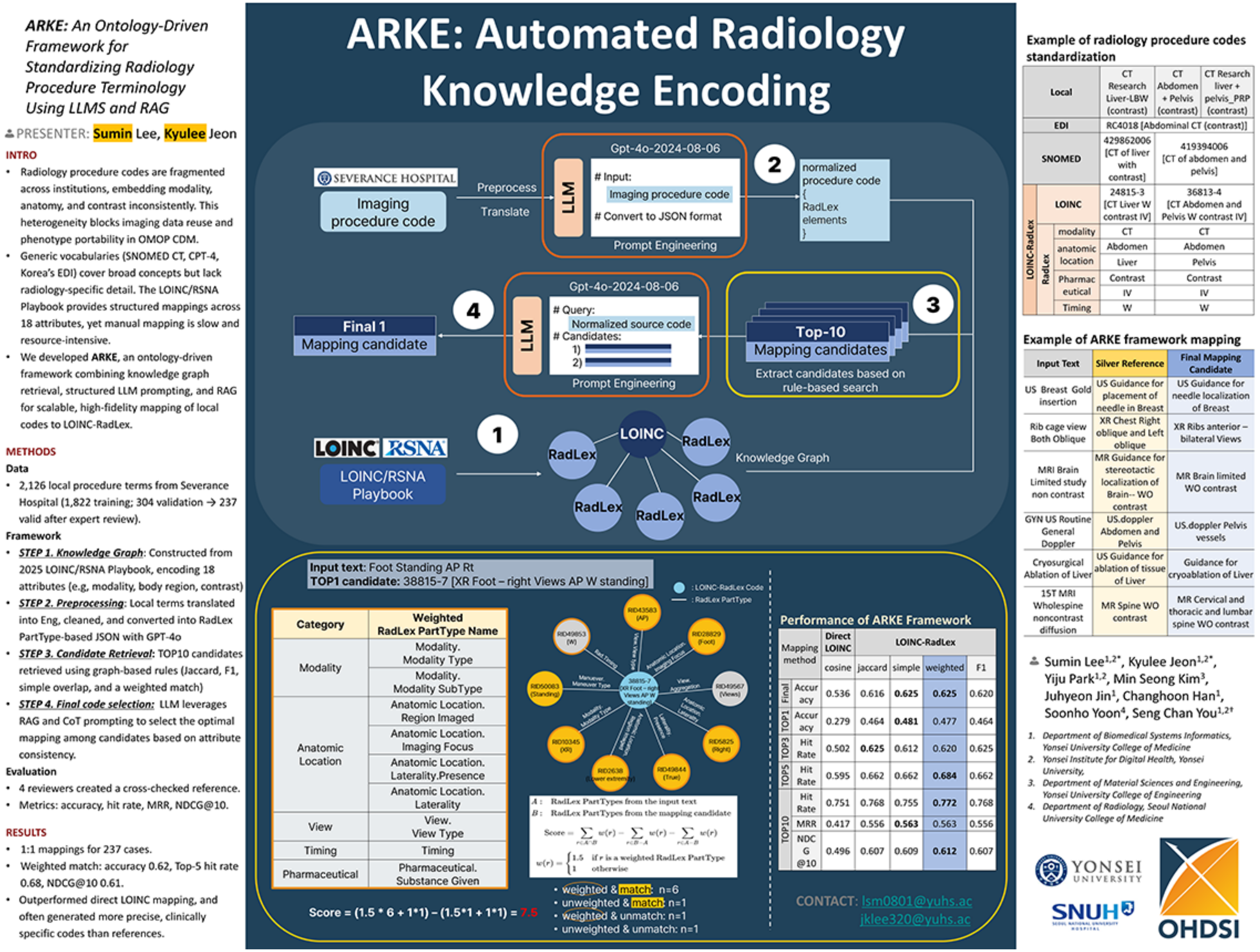
• Please contact Melissa Haendel, PhD for future collaboration work. This work was funded by NIH AoU CLAD program.

#OHDSISocialShowcase This Week

Tuesday

ARKE: An Ontology-Driven Framework for Standardizing Radiology Procedure Terminology Using LLMs and RAG

(Sumin Lee, Kyulee Jeon, Yiju Park, Min Seong Kim, Juhyeon Jin, Changhoon Han, Soonho Yoon, Seng Chan You)

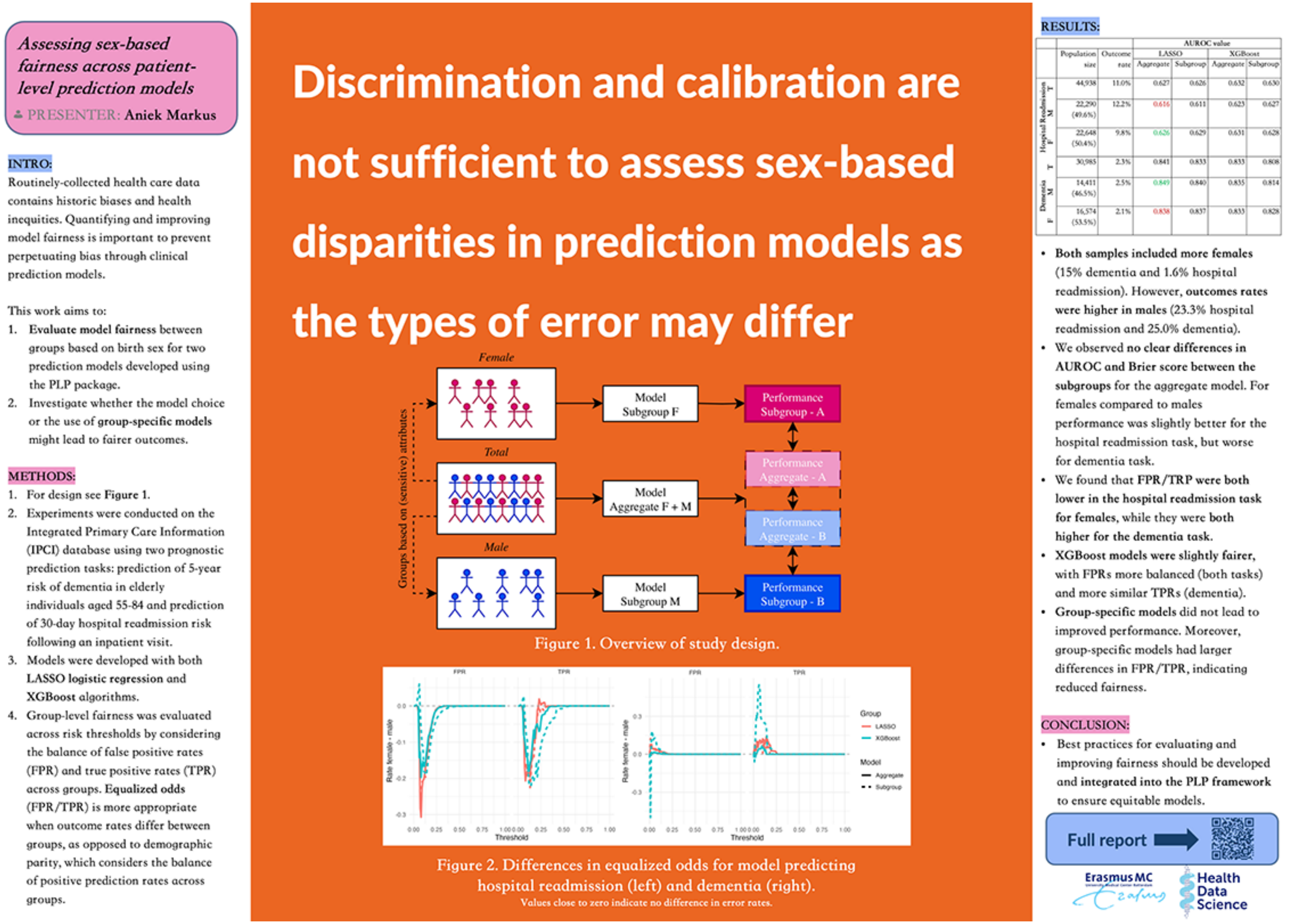


#OHDSISocialShowcase This Week

Wednesday

Assessing sex-based fairness across patient-level prediction models

(Aniek Markus)



Assessing sex-based fairness across patient-level prediction models

PRESENTER: Aniek Markus

INTRO:

Routinely-collected health care data contains historic biases and health inequities. Quantifying and improving model fairness is important to prevent perpetuating bias through clinical prediction models.

This work aims to:

1. **Evaluate model fairness** between groups based on birth sex for two prediction models developed using the PLP package.
2. Investigate whether the model choice or the use of **group-specific models** might lead to fairer outcomes.

METHODS:

1. For design see **Figure 1**.
2. Experiments were conducted on the Integrated Primary Care Information (IPCI) database using two prognostic prediction tasks: prediction of 5-year risk of dementia in elderly individuals aged 55-84 and prediction of 30-day hospital readmission risk following an inpatient visit.
3. Models were developed with both **LASSO logistic regression** and **XGBoost** algorithms.
4. Group-level fairness was evaluated across risk thresholds by considering the balance of false positive rates (FPR) and true positive rates (TPR) across groups. **Equalized odds** (FPR/TPR) is more appropriate when outcome rates differ between groups, as opposed to demographic parity, which considers the balance of positive prediction rates across groups.

Discrimination and calibration are not sufficient to assess sex-based disparities in prediction models as the types of error may differ

Figure 1. Overview of study design.

Figure 2. Differences in equalized odds for model predicting hospital readmission (left) and dementia (right).

Values close to zero indicate no difference in error rates.

RESULTS:

		Population size	Outcome rate	AUROC value LASSO Aggregate	LASSO Subgroup	XGBoost Aggregate	XGBoost Subgroup
Hospital Readmission	T	44,938	11.0%	0.627	0.626	0.632	0.630
	M	22,290 (49.6%)	12.2%	0.616	0.611	0.623	0.627
	F	22,648 (50.4%)	9.8%	0.626	0.629	0.631	0.628
Dementia	T	30,985	2.3%	0.841	0.833	0.833	0.808
	M	14,411 (46.5%)	2.5%	0.849	0.840	0.835	0.814
	F	16,574 (53.5%)	2.1%	0.838	0.837	0.833	0.828

- **Both samples included more females** (15% dementia and 1.6% hospital readmission). However, **outcomes rates were higher in males** (23.3% hospital readmission and 25.0% dementia).
- We observed **no clear differences in AUROC and Brier score between the subgroups** for the aggregate model. For females compared to males performance was slightly better for the hospital readmission task, but worse for dementia task.
- **We found that FPR/TRP were both lower in the hospital readmission task for females, while they were both higher for the dementia task.**
- **XGBoost models were slightly fairer,** with FPRs more balanced (both tasks) and more similar TPRs (dementia).
- **Group-specific models did not lead to** improved performance. Moreover, group-specific models had larger differences in FPR/TPR, indicating reduced fairness.

CONCLUSION:

- Best practices for evaluating and improving fairness should be developed and integrated into the **PLP framework** to ensure equitable models.

Full report

Erasmus MC

Health Data Science

#JoinTheJourney

www.ohdsi.org

#OHDSISocialShowcase This Week

Thursday

Cancer Treatment Guidelines Need a Reality Check — RWE Can Help

(Asieh Golozar, Patrick Alba, Henry Morgan Stewart, Dan Smith, Salma Rachidi, Eric Fey, Valtteri Nieminen, Alexey Ryzhenkov, Tommi Kauko, Pia Tajanen-Doumbouya, Mikael Högerman, Sampo Kukkurainen, Harri Rantala, Åslaug Helland, Åsa Ôjlert, Wei Hai Deng, Zarah Van Schoor, Jonas Minne, Stelios Theophanous, Geoff Hall, John Methot, Thejas Bharadwaj, Roshanthi Weerasinghe, Andrew Nute, Martin Koch, Ines Reinecke, Katja Hoffmann, Jasmin Carus, Stefan Bartels, Michael Franz, Fabian Prasser, Jonathan Jeutner, Álvaro MartÌnez Pèrez, Carlos López Gómez, Ben Gerber, Jung Ae Lee, Thamir Alshammari, Jared Houghtaling, Paul Nagy, Ben Martin, Raivo Kolde, Marek Oja, Sirli Tamm, Miguel-Angel Mayer, Juan Manuel Ramirez-Anguita, Angela Leis, Sam T Patnoe, Deran A Mckeen, Patricia L Mabry, Seng Chan You, Subin Kim, Chang Jun Koc, Jason C. Hsu, Phung Anh Nguyen, Nguyen Thi Kim Hien, Phan Thanh Phuc, Qi Yang, Kees Ebben, Maaike van Swieten, Jelle Evers, Agnes Moesgåard Eschen, Andreas Bjerrum, Irina Veytsman, Dalia Mobarek, Espen Enerly, Joelle Thonnard, Emmanuel Seront, Cedric van Marcke, Dries Hens, Clara L. Oeste, Mikaela Bruhammar, Loretta Zsuzsa Kiss, Mészáros Ágota, Zsolt István Bagyura, Ruochong Fan, Linying Zhang , Alberto Moreno Conde, Jesus Moreno Conde, Tomoni Kimura, Justin Matthew Petucci, Matteo Pontuni, Pantelis Natsiavas, Chytas Achilleas, Rekkas Alexandros, Farmaki Anastasia, Elad Sharon, Kimmo Porkka, Annelies Verbiest, Otto Ettala, Christian Reich)

Cancer Treatment Guidelines Need a Reality Check – RWE Can Help

ASIEH GOLOZAR

INTRODUCTION

- develop a scalable, systematic method for incorporating real-world evidence into clinical guideline development, demonstrated using the 2024 European Association of Urology (EAU) guideline on metastatic bladder cancer.

METHODS

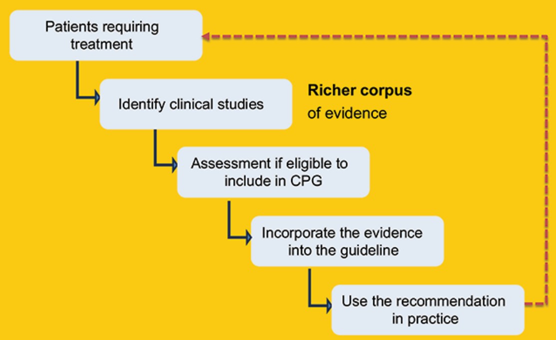
The methodology is based on three key components:

(1) Breaking down guideline recommendations into structured decision nodes with clearly defined populations, criteria, outcomes, and recommended treatments.

(2) Matching these questions against data from research networks for assessing feasibility and validity.

(3) Conducting targeted studies to fill significant evidence gaps or identify inconsistencies between guidelines and practice using high-quality RWE.

FALCON-Bladder : the first federated network to systematically incorporate RWE into clinical guidelines

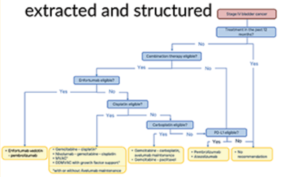


1. **Relevance:** Are there real-world patients who fit the criteria for each treatment recommendation?
2. **Adherence:** To which degree are clinical guideline recommendations applied in practice?
3. **Generalizability:** Do the recommended treatments achieve the desired outcomes in diverse patient populations?
4. **Unmet need:** Are there gaps in guidelines where RWE can improve recommendations?

Take a picture to **download** the **full paper**

RESULTS

- Guideline decision nodes were extracted and structured



FALCON-Bladder as of now

- 31 data partners from 9 countries participated.
- All sites underwent data readiness assessment
- Identified issues addressed through targeted vocabulary updates or local ETL patches.
- Data partners passing the assessment executed a standardized analytics.
- Results submitted to ASCO 2026

Asieh Golozar , Patrick Alba, Henry Morgan Stewart, Dan Smith, Salma Rachidi, Eric Fey, Valtteri Nieminen, Alexey Ryzhenkov, Tommi Kauko, Pia Tajanen-Doumbouya, Mikael Högerman, Sampo Kukkurainen, Harri Rantala, Åslaug Helland, Åsa Ôjlert, Wei Hai Deng, Zarah Van Schoor, Jonas Minne, Stelios Theophanous, Geoff Hall, John Methot, Thejas Bharadwaj, Roshanthi Weerasinghe, Andrew Nute, Martin Koch, Ines Reinecke, Katja Hoffmann, Jasmin Carus, Stefan Bartels, Michael Franz, Fabian Prasser, Jonathan Jeutner, Álvaro MartÌnez Pèrez, Carlos López Gómez, Ben Gerber, Jung Ae Lee, Thamir Alshammari, Jared Houghtaling, Paul Nagy, Ben Martin, Raivo Kolde, Marek Oja, Sirli Tamm, Miguel-Angel Mayer, Juan Manuel Ramirez-Anguita, Angela Leis, Sam T Patnoe, Deran A Mckeen, Patricia L Mabry, Seng Chan You, Subin Kim, Chang Jun Koc, Jason C. Hsu, Phung Anh Nguyen, Nguyen Thi Kim Hien, Phan Thanh Phuc, Qi Yang, Kees Ebben, Maaike van Swieten, Jelle Evers, Agnes Moesgåard Eschen, Andreas Bjerrum, Irina Veytsman, Dalia Mobarek, Espen Enerly, Joelle Thonnard, Emmanuel Seront, Cedric van Marcke, Dries Hens, Clara L. Oeste, Mikaela Bruhammar, Loretta Zsuzsa Kiss, Mészáros Ágota, Zsolt István Bagyura, Ruochong Fan, Linying Zhang , Alberto Moreno Conde, Jesus Moreno Conde, Tomoni Kimura, Justin Matthew Petucci, Matteo Pontuni, Pantelis Natsiavas, Chytas Achilleas, Rekkas Alexandros, Farmaki Anastasia, Elad Sharon, Kimmo Porkka, Annelies Verbiest, Otto Ettala, Christian Reich

OHDSI

#OHDSISocialShowcase This Week

Friday

ExCITE: A Containerized Open-Source Platform Integrating EHR, FHIR, and OMOP for Biomedical Informatics Education

(Robert Barrett, Hayden Spence, Benjamin Martin, Teri Sippel Schmidt, Saptarshi Purkayastha, Paul Nagy)

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?

PLP WG OKRs 2026

Ross D. Williams (Co-Lead)
Erasmus Mc

Improve training and documentation for PLP:

- Training on adding existing models to the plp/omop format
- Training on using cohort covariates / constrained predictors
- Add markdowns and youtube videos

Perform research on rare outcomes

- Add bootstrapping to the plp package
- Add benchmark tasks on paediatric population

Updating prediction models

- Validation of implemented models over time in observational data
- Diagnose where and why changes in performance over time occur
- Create a framework for when to update models
- Understanding changes in performance in subgroups
- Better understanding fine tuning of models and how this impacts updating and fairness

Improve our understanding of fairness and bias assessment

- Which fairness and bias metrics do we recommend using
- How do different use cases create different ethical challenges
- Practical advice on solving detected bias/unfairness in models
- Understand sample size needed for assessment in sub groups and how metrics decay with sample size
- Framework for when to split up and create separate models, tradeoff with overheads for implementation

Run a methods research network study

- If we fit full models vs subgroup models how does this impact performance
- Characterisation of the different patient populations

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