Workgroup Spotlight: CDM and HADES

OHDSI Community Call
March 17, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Upcoming Community Calls

Date	Topic
Mar. 17	A Closer Look: Common Data Model, HADES
Mar. 24	OHDSI/OMOP Research Spotlight
Mar. 31	Kickoff to Phenotype April
Apr. 7	Phenotype April, Week 1
Apr. 14	Phenotype April, Week 2
Apr. 21	NO MEETING / EUROPE SYMPOSIUM
Apr. 28	Pheotype April, Week 4
May 5	Europe Symposium Review/Phenotype April Finale

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Anton Barchuk, Cesar Barboza, Julieta Politi, Berta Raventós, Peter Prinsen, Jelle Evers, Vincent Ky Ho, Michiel Aj Van de Sande, Eric Fey, Kimmo Porkka, Anna Hammais, Tiina Wahlfors, Tuomo Nieminen, Toni Lehtonen, Antonella Delmestri, Guillaume Verdy, Romain Griffier, Airam De Burgos-González, Ana Llorente-Garcia, Cristina Justo-Astorgano, Miguel-Angel Macia-Martinez, Olli Tenhunen, Anja Schiel, Alexandra Pacurariu, Ross Brennan, Ross Williams, Katia Verhamme, and Talita Duarte Salles** on the recent publication of **Characteristics, treatment and survival of patients with chondrosarcoma in five European countries: a DARWIN EU® cohort study** in *acta oncologica*.

ACTA ONCOLOGICA
2026, VOL. 65, 193–200
https://doi.org/10.2340/1651-226X.2026.45117

ORIGINAL ARTICLE

Characteristics, treatment and survival of patients with chondrosarcoma in five European countries: a DARWIN EU® cohort study

Anton Barchuk[a,b], Cesar Barboza[a], Julieta Politi[a], Berta Raventós[a], Peter Prinsen[c], Jelle Evers[c], Vincent K.Y. Ho[c], Michiel A.J. van de Sande[d], Eric Fey[e], Kimmo Porkka[e], Anna Hammais[f], Tiina Wahlfors[f], Tuomo Nieminen[f], Toni Lehtonen[f], Antonella Delmestri[g], Guillaume Verdy[h], Romain Griffier[h], Airam de Burgos-González[i], Ana Llorente-Garcia[i], Cristina Justo-Astorgano[i], Miguel-Angel Macia-Martinez[j], Anja Schiel[j], Olli Tenhunen[k,l], Alexandra Pacurariu[m], Ross Brennan[m], Ross Williams[a], Katia Verhamme[a], and Talita Duarte Salles[a,n]

[a]Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands; [b]University of Helsinki, Helsinki, Finland; [c]Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, The Netherlands; [d]Orthopedic Surgery, Leiden University Medical Center, Leiden, The Netherlands; [e]iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki and Helsinki University Central Hospital Cancer Center, Helsinki, Finland; [f]Finnish Institute for Health and Welfare (THL), Helsinki, Finland; [g]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, United Kingdom; [h]Public Health Department, Medical Information Service, University Hospital of Bordeaux, Bordeaux, France; [i]Agencia Española de Medicamentos y Productos Sanitarios, Madrid, Spain; [j]Norwegian Medical Products Agency (NOMA), Oslo, Norway; [k]Medical Research Center Oulu, Oulu University Hospital, University of Oulu, Oulu, Finland; [l]Finnish Medicines Agency, Helsinki, Finland; [m]Real World Evidence Workstream, European Medicines Agency, Amsterdam, The Netherlands; [n]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

ABSTRACT

Background and purpose: Chondrosarcoma is a rare bone malignancy with a poor response to systemic therapy in advanced stages. European-level epidemiological data remain scarce. This study aimed to characterise patient demographics, treatments and survival using real-world data to inform regulatory decisions about the feasibility and design of new trials for the systemic treatment of chondrosarcoma.

Patient/material and methods: This cohort study, part of the DARWIN EU® initiative, analysed data from six healthcare databases in Finland, France, the Netherlands, Spain and the UK. Patients diagnosed with chondrosarcoma between 2010 and 2022 were identified. Standardised analyses were performed within a federated network using the Observational Medical Outcomes Partnership (OMOP) Common Data Model.

Results: A total of 2,498 chondrosarcoma patient records were identified, covering at least 2,356 unique patients. Median age at diagnosis was 52–55 years, with a balanced sex distribution. Surgical treatment was the most common intervention, recorded in 15.2% to 88.9% of patients, depending on the database. Fewer than 5% received systemic anticancer therapy, and radiotherapy was reported in fewer than 7%. The 10-year overall survival (OS) ranged from 58% (95% confidence interval [CI]: 43–78) to 80% (95% CI: 78–82), with restricted mean survival between 7.4 and 8.7 years. In the Netherlands, patients with late-stage, metastatic or high-grade disease showed significantly poorer outcomes.

ARTICLE HISTORY
Received 4 December 2025
Accepted 12 February 2026
Published 10 March 2026

KEYWORDS
Chondrosarcoma; survival; federated analysis, rare cancers; Darwin EU

OHDSI Shoutouts!

Congratulations to the team of **Junmo Kim, Namkyeong Lee, Jiwon Kim, and Kwangsoo Kim** on the recent publication of **MedRep: medical concept representations for general electronic health record foundation models** in *JAMIA*.

JAMIA
A SCHOLARLY JOURNAL OF INFORMATICS IN HEALTH AND BIOMEDICINE

Issues | More Content | Submit | Purchase | Alerts | About | Journal of the Ameri

Article Contents

- Abstract
- Introduction
- Materials and methods
- Results
- Discussion
- Conclusion
- Author contributions
- Supplementary material
- Funding
- Conflicts of interest
- Data availability
- Code availability
- References

JOURNAL ARTICLE

MedRep: medical concept representations for general electronic health record foundation models Get access

Junmo Kim, BS, Namkyeong Lee, MS, Jiwon Kim, BS, Kwangsoo Kim, PhD

Journal of the American Medical Informatics Association, ocag032, https://doi.org/10.1093/jamia/ocag032

Published: 10 March 2026 **Article history**

Cite | Permissions | Share

Abstract

Objective

Traditional electronic health record (EHR) foundation models fail to process unseen medical codes, limiting generalizability across institutions with different vocabularies. To address this problem, we introduce medical concept representation (MedRep), standardized medical concept representations for EHR foundation models, enabling recognition of semantically similar concepts regardless of their specific IDs.

Materials and Methods

We utilized Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) vocabulary covering 7.5 million concepts from 66 medical vocabularies. MedRep integrates large language model–generated concept descriptions and OMOP graph ontology using graph contrastive learning with knowledge distillation. We evaluated MedRep-based models on MIMIC-IV

OHDSI Shoutouts!

Congratulations to the team of **Benjamin Martin, Will Kelly, Hannah Morgan-Cooper, Thomas Falconer, Elizabeth Park, Priya Desai, David Fiorentino, Lorinda Chung, Sean Yen, Zachary Wang, Didem Saygin, Michael George, Gowtham A. Rao, Joel Swerdel, Azza Shoaibi, and Christopher A. Mecoli** on the recent publication of **Identification of Adult Patients With Dermatomyositis Using Real-World Data Sources** in *Arthritis Care & Research*.

Arthritis Care & Research
AN OFFICIAL JOURNAL OF THE AMERICAN COLLEGE OF RHEUMATOLOGY

Original Article

Identification of Adult Patients With Dermatomyositis Using Real-World Data Sources

Benjamin Martin, Will Kelly, Hannah Morgan-Cooper, Thomas Falconer, Elizabeth Park, Priya Desai, David Fiorentino, Lorinda Chung, Sean Yen, Zachary Wang, Didem Saygin ... See all authors

First published: 12 August 2025 | https://doi.org/10.1002/acr.25625 | VIEW METRICS

This manuscript is the result of funding in whole or in part by the NIH. It is subject to the NIH Public Access Policy. Through acceptance of this federal funding, the NIH has been given a right to make this manuscript publicly available in PubMed Central upon the official date of publication, as defined by the NIH.

Supported in part by the Jerome L. Greene Foundation and the NIH (grants 5U24-HD-113136 and 5T15-LM-013979 to Dr Martin). Dr Saygin's work was supported by the Rheumatology Research Foundation Scientist Development Award, Rukel Family Foundation, Chicago Center on Musculoskeletal Pain Pilot Grant, and RUSH to Progress Pilot Award.

Drs Martin and Kelly are co-first authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25625).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25625.

OHDSI Shoutouts!

Congratulations to the team of **Matthew Spotnitz, John Giannini, Emily Clark, Yechiam Ostchega, Tamara Litwin, and Lew Berman** on the recent publication of **Assessing data quality of rheumatoid and psoriatic arthritis patients in the All of Us Research Program** in *JAMIA Open*.

JAMIA Open, 2026, **9(2)**, ooag028
https://doi.org/10.1093/jamiaopen/ooag028
Research and Applications

Research and Applications

Assessing data quality of rheumatoid and psoriatic arthritis patients in the *All of Us* Research Program

Matthew Spotnitz, MD, MPH*[,1], John Giannini, PhD[1], Emily Clark, MPH[2], Yechiam Ostchega, PhD, RN[1], Tamara R Litwin, PhD, MPH[1], Lew Berman, PhD, MS[1]

[1]All of United States Research Program, Office of the Director, National Institutes of Health, Bethesda, MD, United States, [2]GAP Solutions, Inc, Herndon, VA, United States

*Corresponding author: Matthew Spotnitz, MD, MPH, All of United States Research Program, Office of the Director, National Institutes of Health, 6710B Rockledge Drive, Bethesda, MD 20892, United States (matthew.spotnitz@nih.gov)

Abstract

Purpose: Rheumatoid and Psoriatic Arthritis (RA and PsA) are autoimmune diseases that cause debilitating joint pain. Disease-modifying antirheumatic drugs (DMARDs) are recommended for the treatment of both conditions. However, real-world evidence studies would help characterize compliance with these recommendations. The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) standardizes electronic health record (EHR) data, allowing for research that incorporates multiple data sources. We are interested in determining whether OMOP CDM data on RA and PsA are fit-for-use.

Methods: We selected diagnosis codes for RA and PsA that were the basis for each phenotype. We used a data quality checklist to evaluate 5 domains systematically: conformance, completeness, concordance, plausibility, and temporality.

Results: Most phenotype-defining ICD source codes mapped to SNOMED. Both cohorts had low concept prevalences. Most concept correlations were weak ($\rho \leq 0.5$). The relative distribution of DMARD ingredients in both cohorts was consistent with prior studies. The proportion of the RA and PsA cohorts that had data for timing between event calculations ranged from 13% to 85% and 16% to 81%, respectively. Despite variability in concept sequence analysis, symptomatic treatment concepts for RA and PsA were preceded by rheumatoid factor concepts, followed by DMARD therapy and disease diagnosis concepts.

Conclusion: We have shown a novel implementation of our data quality framework on autoimmune disease cohorts.

Lay Summary

We characterized the data quality of rheumatoid and psoriatic arthritis (RA and PsA) cohorts in the *All of Us* Research Program. Those data are mapped to the observational medical outcomes partnership common data model (OMOP CDM). We evaluated the cohort according to 5 data quality dimensions: conformance, completeness, concordance, plausibility and temporality. We found that most concepts mapped to SNOMED, however the concept prevalence was low for multiple variables. Also, many of the correlations between concepts were weak. Of the participants who had data that were available for analysis, their distribution of disease-modifying antirheumatic drugs (DMARDs) were plausible. The temporality analysis showed that the disease diagnosis codes lagged behind concepts for medications and laboratory tests. In conclusion, we have applied a data quality framework that has previously been used to evaluate oncology cohorts to autoimmune disease cohorts.

Key words: rheumatoid arthritis; psoriatic arthritis; precision medicine; electronic health record; data quality.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

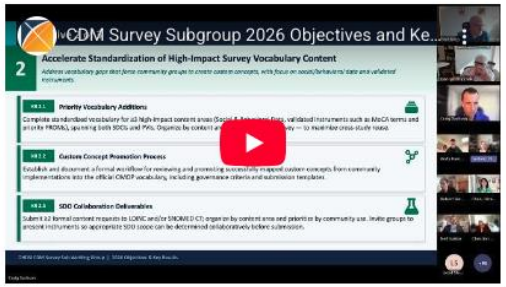
Date	Time (ET)	Meeting
Tuesday	12 pm	Vocabulary
Wednesday	7 am	Medical Imaging
Wednesday	8 am	Psychiatry
Wednesday	8 am	Medical Devices
Thursday	8 am	India Community Call
Thursday	9 am	Oncology Vocabulary/Development Subgroup
Thursday	12 pm	HADES
Thursday	7 pm	Dentistry
Friday	9 am	Waveform
Friday	10 am	Transplant
Friday	11:30 am	Steering
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	9 am	Data2Evidence

Workgroups Homepage

CDM Survey Subgroup

2026 OKR slides

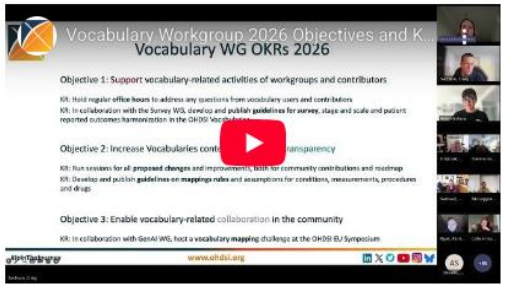
Nicole Gerlanc



CDM Vocabulary Subgroup

2026 OKR slides

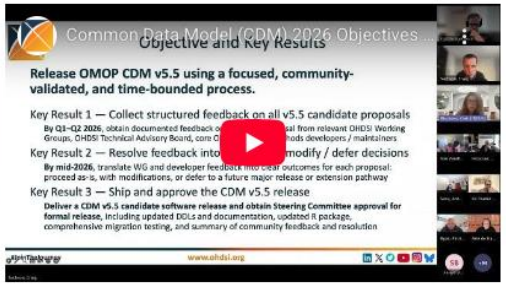
Anna Ostropolets



Common Data Model

2026 OKR slides

Clair Blacketer



Early-Stage Researchers

2026 OKR slides

Shounak Chattopadhyay

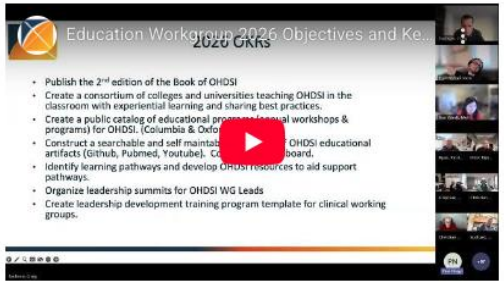
Ben Martin

Harry Reyes Nieva

Education

2026 OKR slides

Paul Nagy



ohdsi.org/workgroups

Europe Community Call Recording Posted

Implementing the European Health Data Space (EHDS): The role of OMOP & OHDSI in secondary use of health data

Prof. dr. ir. Peter R. Rijnbeek

Lead OHDSI Europe

Chair Department of Medical Informatics

Erasmus MC, The Netherlands

OHDSI Europe Community Call – March 12, 2026

Science Summit 2025

alongside the United Nations General Assembly (UNGA80)

Standardizing Health Data and Analytics to Accelerate Clinical Impact and Global Reach, September 18

Part 1: 8:30 to 10:30 EDT

George Hripcsak
Professor, Biomedical Informatics
Columbia University

Agnes Kiragga
Global Health Leader, Data Scientist Principal Investigator
African Population Health and Research Centre

Seng Chan You
Assistant Professor
Yonsei University College of Medicine

Nicole Pratt
Professor, Biostatistics and Pharmacoepidemiology
University of South Australia

Part 2: 11:00 to 13:00 EDT

Peter Rijnbeek
Professor, Medical Informatics
Erasmus University Medical Center

Katia Verhamme
Associate Professor of Use and Analysis of Observational Data
Erasmus University Medical Center

Julio Oliveira
CEO
Precision Data

Cynthia Sung
Adjunct Associate Professor
Duke-NUS Medical School
Centre of Regulatory Excellence

Patrick Ryan
VP Janssen Observational Health Data Analytics also Co-Founder OHDSI
OHDSI Observational Health Data Science and Informatics

OHDSI

Standardizing Health Data and Analytics to Accelerate Clinical Impact and Global Reach Part 1

1. Observational Health Data Science and Informatics (OHDSI): Inclusive and Collaborative Science. *George Hripcsak*
2. Promoting Data Harmonization and Data Science in Africa. *Agnes Kiragga*
3. Rapid Response to the Covid-19 Pandemic Using a National Scale Database. *Chan Seng You*
4. OHDSI in Asia and the Australian Health Data and Evidence Network *Nicole Pratt*

Session Recording: https://us02web.zoom.us/clips/share/VaVXhn_pQ9CTaoyaKaB-5A

Standardizing Health Data and Analytics to Accelerate Clinical Impact and Global Reach Part 2

1. Enabling Reliable Evidence Generation from Real-world Data in Europe. *Peter Rijnbeek*
2. DARWIN-EU® – Delivering Real World Evidence to Support Regulatory Decision-making by the European Medicines Agency. *Katia Verhamme*
3. OHDSI Adoption and Current Implementation Landscape in Latin America. *Julio Cesar Barbour Oliveira*
4. Learning Opportunities for OHDSI Skills Development. *Cynthia Sung*
5. Clinical and Public Health Impact of OHDSI. *Patrick Ryan*

Session Recording: https://us02web.zoom.us/clips/share/il_mzHJQSc-3uiD92ca_7Q

2026 Europe Symposium

The 2026 OHDSI Europe Symposium returns to Rotterdam next year and will be held **April 18-20.**

Registration is open on the **OHDSI & OHDSI Europe** web sites.

Time	Symposium Agenda - Monday April 20, 2026	Location
8:00	***Registration and Coffee***	Queen's Lounge
9:00	**Welcome to OHDSI Europe** Dr. Renske Los, Department of Medical Informatics, *Erasmus MC* Dr. Aniek Markus, Department of Medical Informatics, *Erasmus MC*	Theatre
9:05	**Journey of OHDSI** Prof. Peter Rijnbeek, Chair Department of Medical Informatics, *Erasmus MC*	Theatre
9:30	**Collaborator Showcase - part 1** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC*	Theatre
10:00	**Speed networking**	Theatre
10:15	***Coffee Break & posters National Nodes***	Queen's Lounge
11:15	**Collaborator Showcase - part 2** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC*	Theatre
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*	Theatre
12:15	***Lunch break & networking & posters/demo's*** *(Early investigator meeting - 13:00-13:45 Queen's Lounge)*	La Fontaine & Odyssee Room
13:45	**From dreams to reality** OHDSI Titan Award winners	Theatre
14:30	**Propositions for collaboration from the National Nodes** National Node leads	Theatre
14:45	***Coffee break & posters/demo's***	La Fontaine & Odyssee Room
16:15	**The OH Factor** To be announced	Theatre
17:00	**Closing**	Theatre
17:15	***Networking reception***	Queen's Lounge

2026 Global Symposium

The 2026 OHDSI Global Symposium will return to the Hyatt Regency Hotel in New Brunswick, N.J., on **Oct. 20-22.**

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
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

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.

#OHDSISocialShowcase This Week

Monday

Simplifying Research that Involves Multiple Care Sites: insights and implementation at the US VA

(Richard D. Boyce, Patrick R. Alba, Katherine R. Simon, Benjamin Viernes, William J. Obrien, Marc Suchard, Michael E. Matheny)

Title: Simplifying Research that Involves Multiple Care Sites: insights and implementation at the US Veterans Health Administration

PRESENTER: Patrick R. Alba

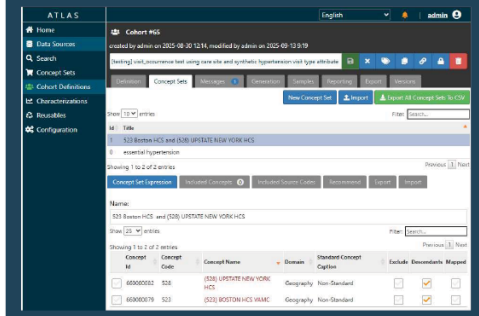
INTRO:

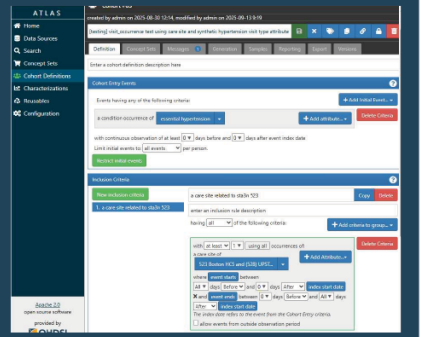
- The Veterans Health Administration (VHA) is the largest integrated health network in the United States, serving over 9 million patients at over 1,300 care facilities.
- Researchers within the VA or using the VA as a site in a network study might be interested in breaking down study results or differences amongst the sites.
- The VA Site Tracking (VAST) database is used for detailed tracking of patients care at both VHA care sites and care sites outside of the VHA network.
- These data are translated into facts within the VHA Corporate Data Warehouse and then into the OMOP CDM using the `fact_relationship` and `care_site` tables
- This approach leads to some issues:
  - The hierarchical organization of VHA care sites is obscured.
  - It is difficult for researchers to create cohorts that include criteria based on hierarchal care site relationships.
  - There is no representation of a patients care site history
  - It limits HADES and Atlas tool use

METHODS

1. We collected use cases from VHA researchers that require the use of VA care sites.
2. We studied the VAST and developed vocabulary and CDM extensions to fulfill the use cases.
3. We implemented and tested changes to Atlas and HADES to enable user friendly integration of care sites into concept set and cohort definitions

Do you have patients from multiple care-sites that have changed over time? Seamlessly define your cohorts using concepts and criteria using care_site_history and location_history!





Take a picture to view a short video motivating this work

RESULTS

- A VHA care site terminology with more than a half-million concepts .

Care site concept class	Concept count
VHA care sites	
Region and National	5
Veteran's integrated service network (VISN)	21
VA Medical Center or Health care center	145
Regional benefits office	52
VHA Institution	418902
Non-VHA care sites	
nursing facility	12706
dialysis center	7165
home health	28817
inpatient hospital	21283

- The terminology represents important hierarchical and logical relationships such the arrangement of institutions into VISNs.
- The new `care_site_history` and revised `location_history` tables enable cohort definitions to track patient movement within the health system.

FUTURE WORK

- Coordinate with the Vocabulary, CDM, and Atlas/WebAPI workgroups to translate these findings into OHDSI tools that benefit other researchers

RELATED CODE

- https://github.com/vinci-ohdsi/WebAPI
- https://github.com/vinci-ohdsi/Atlas-rc15
- https://github.com/vinci-ohdsi/circe-be
- https://github.com/vinci-ohdsi/CirceR
- https://github.com/vinci-ohdsi/Capr

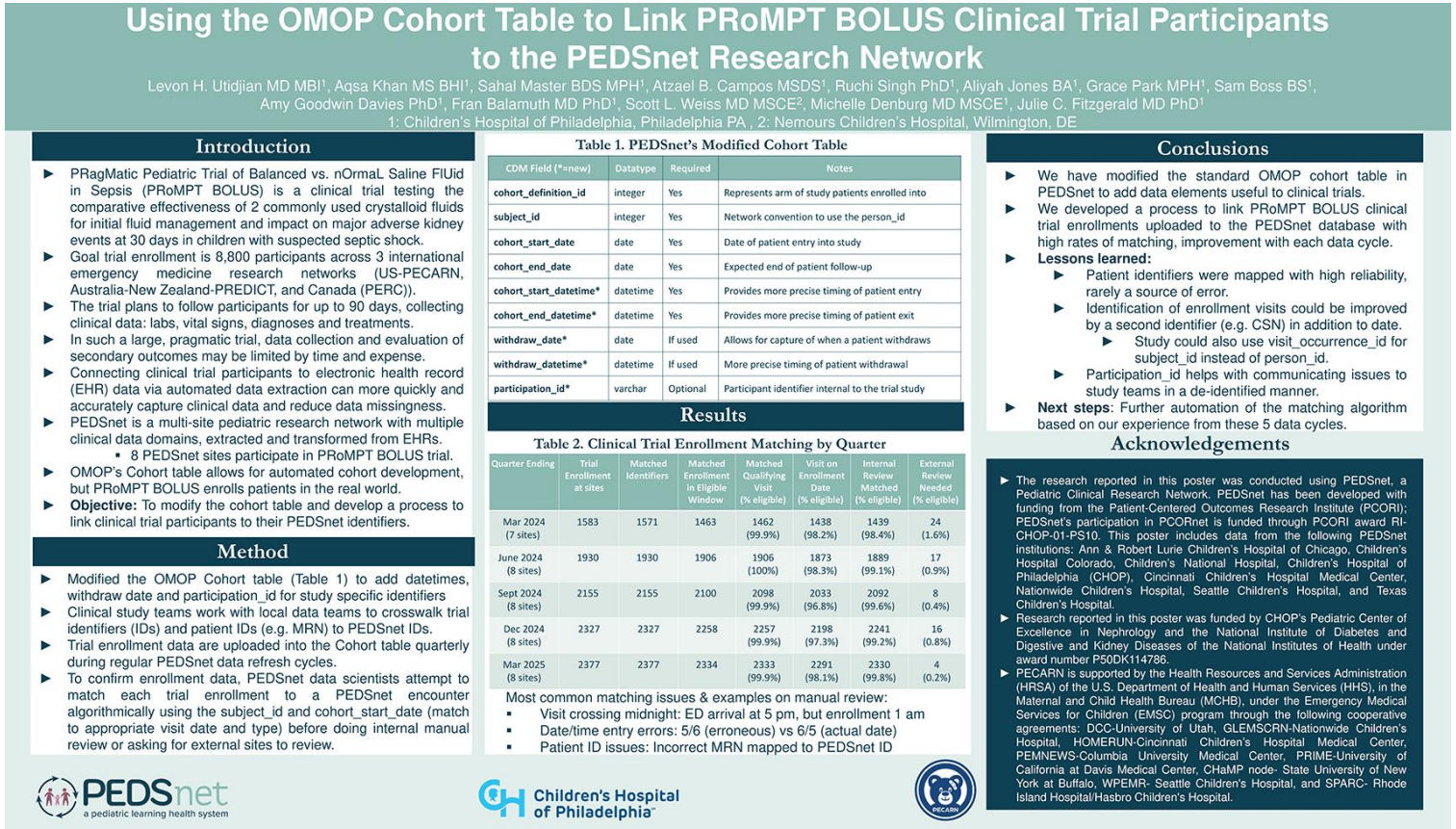
Richard D. Boyce, Patrick R. Alba, Katherine R. Simon, Benjamin Viernes, William J. O'Brien, Marc A. Suchard, Michael E. Matheny

#OHDSISocialShowcase This Week

Tuesday

Using the OMOP Cohort Table to Link PRoMPT BOLUS Clinical Trial Participants to the PEDSnet Research Network

(Levon H. Utidjian, Aqsa Khan, Sahal Master, Atzael B. Campos, Ruchi Singh, Aliyah Jones, Grace Park, Sam Boss, Amy Goodwin Davies, Fran Balamuth, Scott Weiss, Michelle Denburg, Julie C. Fitzgerald)



Using the OMOP Cohort Table to Link PRoMPT BOLUS Clinical Trial Participants to the PEDSnet Research Network

Levon H. Utidjian MD MBI[1], Aqsa Khan MS BHI[1], Sahal Master BDS MPH[1], Atzael B. Campos MSDS[1], Ruchi Singh PhD[1], Aliyah Jones BA[1], Grace Park MPH[1], Sam Boss BS[1], Amy Goodwin Davies PhD[1], Fran Balamuth MD PhD[1], Scott L. Weiss MD MSCE[2], Michelle Denburg MD MSCE[1], Julie C. Fitzgerald MD PhD[1]
1: Children's Hospital of Philadelphia, Philadelphia PA , 2: Nemours Children's Hospital, Wilmington, DE

Introduction

- PRagMatic Pediatric Trial of Balanced vs. nOrmaL Saline FIUid in Sepsis (PRoMPT BOLUS) is a clinical trial testing the comparative effectiveness of 2 commonly used crystalloid fluids for initial fluid management and impact on major adverse kidney events at 30 days in children with suspected septic shock.
- Goal trial enrollment is 8,800 participants across 3 international emergency medicine research networks (US-PECARN, Australia-New Zealand-PREDICT, and Canada (PERC)).
- The trial plans to follow participants for up to 90 days, collecting clinical data: labs, vital signs, diagnoses and treatments.
- In such a large, pragmatic trial, data collection and evaluation of secondary outcomes may be limited by time and expense.
- Connecting clinical trial participants to electronic health record (EHR) data via automated data extraction can more quickly and accurately capture clinical data and reduce data missingness.
- PEDSnet is a multi-site pediatric research network with multiple clinical data domains, extracted and transformed from EHRs.
  - 8 PEDSnet sites participate in PRoMPT BOLUS trial.
- OMOP's Cohort table allows for automated cohort development, but PRoMPT BOLUS enrolls patients in the real world.
- **Objective:** To modify the cohort table and develop a process to link clinical trial participants to their PEDSnet identifiers.

Method

- Modified the OMOP Cohort table (Table 1) to add datetimes, withdraw date and participation_id for study specific identifiers
- Clinical study teams work with local data teams to crosswalk trial identifiers (IDs) and patient IDs (e.g. MRN) to PEDSnet IDs.
- Trial enrollment data are uploaded into the Cohort table quarterly during regular PEDSnet data refresh cycles.
- To confirm enrollment data, PEDSnet data scientists attempt to match each trial enrollment to a PEDSnet encounter algorithmically using the subject_id and cohort_start_date (match to appropriate visit date and type) before doing internal manual review or asking for external sites to review.

Table 1. PEDSnet's Modified Cohort Table

CDM Field (*=new)	Datatype	Required	Notes
cohort_definition_id	integer	Yes	Represents arm of study patients enrolled into
subject_id	integer	Yes	Network convention to use the person_id
cohort_start_date	date	Yes	Date of patient entry into study
cohort_end_date	date	Yes	Expected end of patient follow-up
cohort_start_datetime*	datetime	Yes	Provides more precise timing of patient entry
cohort_end_datetime*	datetime	Yes	Provides more precise timing of patient exit
withdraw_date*	date	If used	Allows for capture of when a patient withdraws
withdraw_datetime*	datetime	If used	More precise timing of patient withdrawal
participation_id*	varchar	Optional	Participant identifier internal to the trial study

Results

Table 2. Clinical Trial Enrollment Matching by Quarter

Quarter Ending	Trial Enrollment at sites	Matched Identifiers	Matched Enrollment in Eligible Window	Matched Qualifying Visit (% eligible)	Visit on Enrollment Date (% eligible)	Internal Review Matched (% eligible)	External Review Needed (% eligible)
Mar 2024 (7 sites)	1583	1571	1463	1462 (99.9%)	1438 (98.2%)	1439 (98.4%)	24 (1.6%)
June 2024 (8 sites)	1930	1930	1906	1906 (100%)	1873 (98.3%)	1889 (99.1%)	17 (0.9%)
Sept 2024 (8 sites)	2155	2155	2100	2098 (99.9%)	2033 (96.8%)	2092 (99.6%)	8 (0.4%)
Dec 2024 (8 sites)	2327	2327	2258	2257 (99.9%)	2198 (97.3%)	2241 (99.2%)	16 (0.8%)
Mar 2025 (8 sites)	2377	2377	2334	2333 (99.9%)	2291 (98.1%)	2330 (99.8%)	4 (0.2%)

Most common matching issues & examples on manual review:

- Visit crossing midnight: ED arrival at 5 pm, but enrollment 1 am
- Date/time entry errors: 5/6 (erroneous) vs 6/5 (actual date)
- Patient ID issues: Incorrect MRN mapped to PEDSnet ID

Conclusions

- We have modified the standard OMOP cohort table in PEDSnet to add data elements useful to clinical trials.
- We developed a process to link PRoMPT BOLUS clinical trial enrollments uploaded to the PEDSnet database with high rates of matching, improvement with each data cycle.
- **Lessons learned:**
  - Patient identifiers were mapped with high reliability, rarely a source of error.
  - Identification of enrollment visits could be improved by a second identifier (e.g. CSN) in addition to date.
    - Study could also use visit_occurrence_id for subject_id instead of person_id.
  - Participation_id helps with communicating issues to study teams in a de-identified manner.
- **Next steps**: Further automation of the matching algorithm based on our experience from these 5 data cycles.

Acknowledgements

- The research reported in this poster 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 poster includes data from the following PEDSnet institutions: Ann & Robert Lurie Children's Hospital of Chicago, Children's Hospital Colorado, Children's National Hospital, Children's Hospital of Philadelphia (CHOP), Cincinnati Children's Hospital Medical Center, Nationwide Children's Hospital, Seattle Children's Hospital, and Texas Children's Hospital.
- Research reported in this poster was funded by CHOP's 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.
- PECARN is supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS), in the Maternal and Child Health Bureau (MCHB), under the Emergency Medical Services for Children (EMSC) program through the following cooperative agreements: DCC-University of Utah, GLEMSCRN-Nationwide Children's Hospital, HOMERUN-Cincinnati Children's Hospital Medical Center, PEMNEWS-Columbia University Medical Center, PRIME-University of California at Davis Medical Center, CHaMP node- State University of New York at Buffalo, WPEMR- Seattle Children's Hospital, and SPARC- Rhode Island Hospital/Hasbro Children's Hospital.

#OHDSISocialShowcase This Week

Wednesday

Scalable Big Data Workflow for OMOP CDM: Performance Optimization and Automated Quality Evaluation of Real-World Data

(Danilo Luis Cerqueira Dias, Ricardo Felix Monteiro Neto, Juliana Araújo Prata de Faria, Valentina Martufi, Julio Barbour Oliveira, Karine Brito Beck da Silva Magalhães, Roberto Perez Carreiro, Maurício L. Barreto, Elzo Pereira Pinto Junior, Pablo Ivan Pereira Ramos)

Scalable Big Data Workflow for OMOP CDM

Performance Optimization and Automated Quality Evaluation of Real-World Data

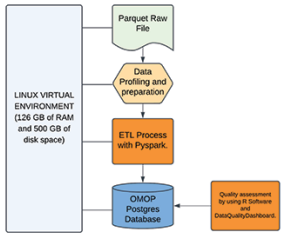
PRESENTER: **Valentina Martufi**

INTRODUCTION

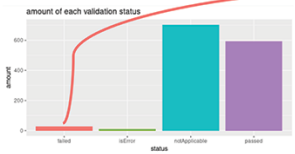
- ✓ Ensuring the **quality and interoperability** of **real-world, administrative** health data is critical to answer relevant public health questions and advancing evidence-based public health policies;
- ✓ We hereby present the development of a **computational infrastructure tailored to maintain data integrity while supporting federated analyses through the OMOP CDM,** using datasets from the Brazilian public health information system within CIDACS' secure and large-scale data environment.

METHODS

Data 'OMOPping' carried out within a TRE - ensuring strict governance, data privacy, and compliance with ethical and legal standards.

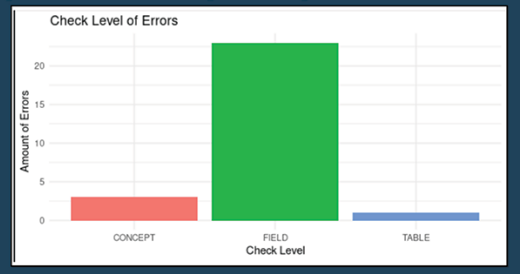


RESULTS



❑ DQD: 1,363 standard, automated validation tests, only 37 (2.71%) resulted in failures;

Inherent challenges in the application of the OMOP CDM were **effectively addressed** using **robust** and **scalable** technologies – **PySpark** for distributed data transformation, **PostgreSQL** for structured storage, **R software** for advanced analytics, and the **DataQualityDashboard** for systematic quality assurance.



Take a picture to **download** our ETL documentation

Acknowledgements:
This work was funded by the Gates Foundation

Contact email: valentina.martufi@fiocruz.br

AMMO BAR

Methods deets:

1. Computational infrastructure: high-performance Linux environment with 126 GB of RAM and 500 GB of Disk storage - for efficient large-scale data processing while maintaining a secure and controlled setting;
2. Adherence to OHDSI best practices, in three main stages:

i. Data profiling and ETL: Jupyter Notebooks and PySpark, distributed processing of over 24m health records – incl. mapping source variables to standardized OMOP concepts for semantic interoperability;
ii. Data Storage and Structuring: loaded into a PostgreSQL database structured according to OMOP CDM version 5.4;
iii. Data Quality Assessment: DQD R package

Results deets:

Primarily at the field level, e.g. contradictory dates or improperly formatted values; followed by concept-level inconsistencies related to vocabulary mapping.
Notably, the lowest number of errors was detected at the table level, indicating structural soundness across the CDM schema.
Inconsistencies thoroughly reviewed, traced back to origin in ETL pipeline, and corrected with targeted transformations, then fully reprocessed and redeployed data.
Challenges faced and strategies developed:
a. Explode (split) data where ICDs were grouped;
b. Enrich the data with PySpark for computational reproducibility;
c. Performance optimization with PySpark

Danilo Luis Cerqueira Dias, Ricardo Felix Monteiro Neto, Juliana Araújo Prata de Faria, Valentina Martufi, Julio Barbour Oliveira, Karine Brito Beck da Silva Magalhães, Roberto Perez Carreiro, Maurício L. Barreto, Elzo Pereira Pinto Junior, Pablo Ivan Pereira Ramos

#OHDSISocialShowcase This Week

Thursday

AutoSPICT – Identifying End of Life Care Needs using an electronic implementation of the SPICT™ Questionnaire

(Joseph S. Boyle, Mike O'Neil, Maria Liakata, Alison Q. Smithard)

OHDSI

AutoSPICT: Identifying End of Life Care Needs using an electronic implementation of the SPICT™ Questionnaire

Joseph S. Boyle[1,2,3], Mike O'Neil[3], Maria Liakata[1,4,5], Alison Smithard[1,6]

[1] Canon Medical Research Europe, [2] Queen Mary University of London, [3] NHS Nottingham Data Management Team, [4] University of Warwick, [5] The Alan Turing Institute, [6] The University of Edinburgh

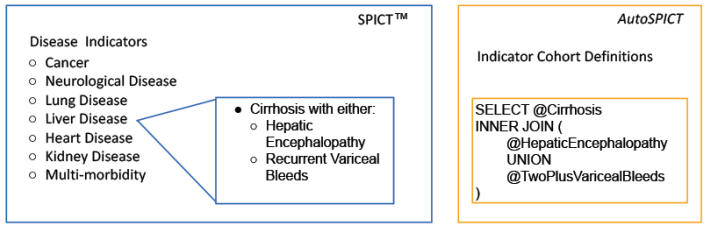
Background

SPICT™ is a screening tool used to determine if a seriously ill patient is indicated for palliative care[1,2]. The indicators cover the seven major terminal disease-trajectories, and the criteria to look for in each.

An automated implementation of the SPICT™ questionnaire could identify the cohort of patients with palliative needs and enable primary care practitioners to proactively screen the population[2]. In this work, we propose **AutoSPICT: a digital version of the popular *SPICT*™ palliative care screening tool.**

Methods

1. Map the definitions from SPICT™ to concepts
2. Create operational definitions for each criterion of each disease indicator
3. Program disease-specific cohorts, and union them to form the *AutoSPICT* cohort

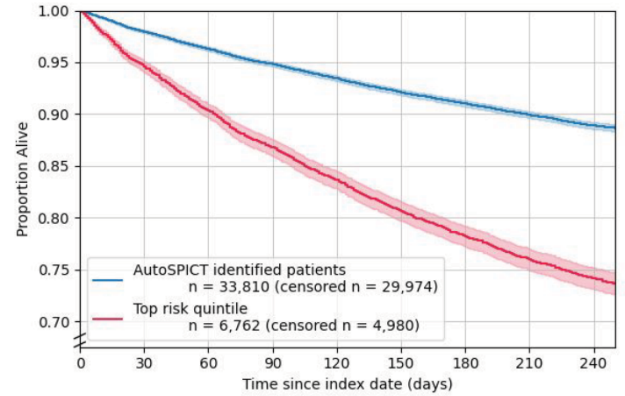


We evaluate AutoSPICT prospectively on it's capacity as a predictor of mortality (↑) using coded EHR data in a population of 1,402,251 patients registered with *NHS Nottingham and Nottinghamshire*, a care board in England.

We use patient-level logistic regression to predict the likelihood of 1-year mortality.

1. Highet G, Crawford D, Murray S, Boyd K. Development and evaluation of the Supportive and Palliative Care Indicators Tool (SPICT): a mixed-methods study. BMJ supportive & palliative care. 2013 Jul;4. https://doi.org/10.1136/ bmjspcare-2013-000488.
2. Mason B, Boyd K, Murray SA, Steyn J, Cormie P, Kendall M, et al. Developing a computerised search to help UK General Practices identify more patients for palliative care planning: a feasibility study. BMC Family Practice. 2015 Aug;16(1):99. https://doi.org/10.1186/s12875-015-0312-z.

Results



Prospective Kaplan-Meier survival curve showing the observed mortality rate of patients identified by AutoSPICT on the index date, January 1st 2025.

AutoSPICT identified patients in the top mortality risk quintile experienced a 26% 8 month mortality rate (n=1,782) during that period, giving this subgroup an estimated annual mortality rate of 36%

Conclusions

We implemented a set of existing palliative care indicators electronically, and found that the resulting screening tool was acceptable to GPs, and had validity as a predictor of prospective 10-month mortality.

The AutoSPICT screening tool is now a recommended tool for GPs in NHS Nottingham to use for detecting end of life care needs.

The concept relationships encoded in the CDM provided a valuable resource for code discovery and for mapping codes between the SNOMED-CT, ICD-10 and Read vocabularies occurring in our NHS database. Future work will investigate the use of AutoSPICT as a screening tool within primary care, and will aim to provide a set of portable disease-specific cohorts using standardised concepts.

Canon MEDICAL RESEARCH EUROPE LTD.

Queen Mary University of London

NHS Nottingham and Nottinghamshire

Correspondence to: j.s.boyle@qmul.ac.uk

#OHDSISocialShowcase This Week

Friday

dqdbt: Continuous Data Quality Testing for OMOP ETL with dbt

(Katy Sadowski, Lawrence Adams, Thomas Wylie)

What is dbt?

dbt is an open-source data transformation framework

- SQL-first approach to data modeling
- Native testing and documentation features
- Used by over 25,000 companies worldwide

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?

GIS Working Group: 2026 Objectives & Key Results (OKRs)

Objective: Release a stable "working beta" of the Gaia toolchain.
Key Result (Definition): Deployable version with no workflow breaking bugs and a stable PostgREST API layer.
Scope: Priority on core functionality (ingestion, pipeline execution); connectors are secondary.
Target Timeline: Early May 2026
(Rationale: Allows summer testing ahead of October OHDSI Global Symposium).

Objective: Ensure location_history and external_exposure tables meet forthcoming OHDSI extension requirements
Key Deliverables: Extension tables and associated documentation and tooling that meet formal requirements
Scope: DQD Checks - SQL/R checks for new tables, transformation examples incl documented workflow to external_exposure, analytical tools built on extension tables.
Target Timeline: June 2026

Objective: Demonstrate a reproducible workflow using GIS data within the OHDSI stack.
Key Result: Workflow established using CaprForExtensions and featureExtractionForExtensions packages
Dependency: Synthetic data (environmental and socioeconomic factors linked to respiratory disease) posted to GIS repo.
Target Timeline: Present at OHDSI Europe, Expand to Showcase at OHDSI Global

Objective: Validate the entire workflow with real-world data across multiple sites
Significance: Pilot completion is essential proof-of-value for a strong GIS manuscript
Potential Data Sources: CHoRUS (Bridge2AI): Centralized multi-site access, MUSC/VA Rural Program (Strong candidate),CODATA: Real-world sensor data (Dhaka/Douala)
Target Timeline: October 2026

Waveform Working Group: 2026 Objectives & Key Results (OKRs)

Objective: External evaluation of extension tables and further refinement.
Key Result (Definition): Tested and agreed upon specification for the extension tables, ideally with examples
Scope: Multiple partners with different objectives test out the proposed extension tables and try to identify issues/missingness.
Target Timeline: end March 2026
(*Rationale: Allows summer testing ahead of October OHDSI Global Symposium).*

Objective: Follow newly established conventions for formalizing OMOP Extensions
Key Deliverables: Extension tables and associated documentation and tooling that meet formal requirements
Scope: DQD Checks - SQL/R checks for new tables, transformation examples incl documented workflow to extension table, analytical tools built on extension tables, synthetic data
Target Timeline: June 2026

Objective: Demonstration of waveform extension table integration
Key Result: Workflow established using CaprForExtensions and featureExtractionForExtensions packages
Dependency: Synthetic data (environmental and socioeconomic factors linked to respiratory disease) posted to Waveform repo.
Target Timeline: Present at OHDSI Europe, Expand to Showcase at OHDSI Global

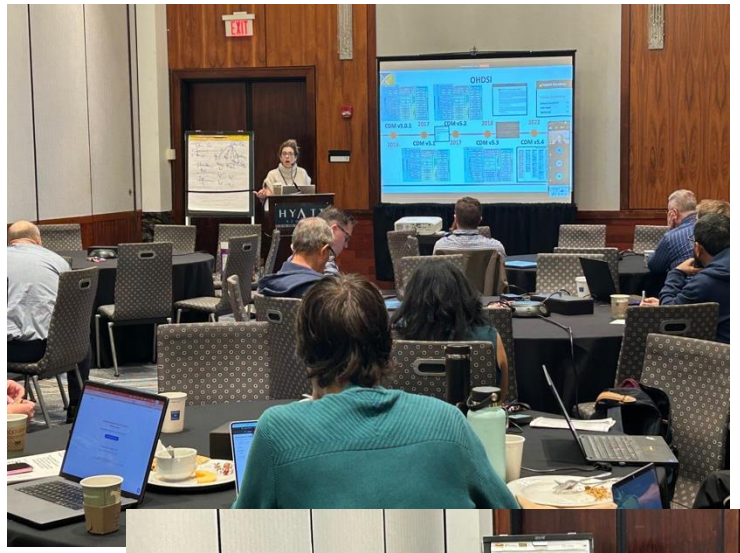
Objective: Initial multi-site demonstration of waveform data alongside OMOP in study
Significance: Pilot completion is essential proof-of-value for a strong Waveform WG manuscript
Potential Data Sources: CHoRUS (Bridge2AI): Centralized multi-site access, Indicate (federated intensive care data)
Target Timeline: October 2026

March 17: Workgroup Spotlight

Common Data Model

Clair Blacketer

Director, Johnson & Johnson



HADES

Anthony Sena

Director, Johnson & Johnson

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