



Falcon Symposium & Studyathon Review

OHDSI Community Call
Oct. 6, 2026 • 11 am ET



Upcoming Community Calls

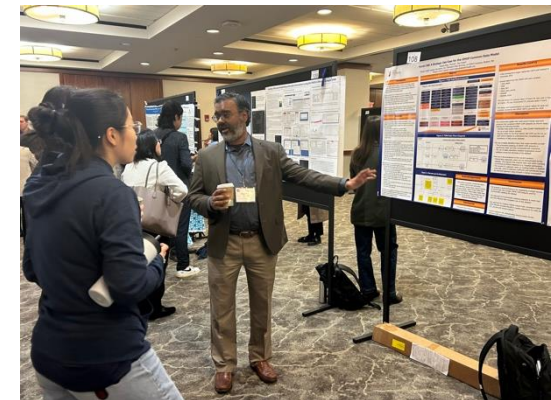
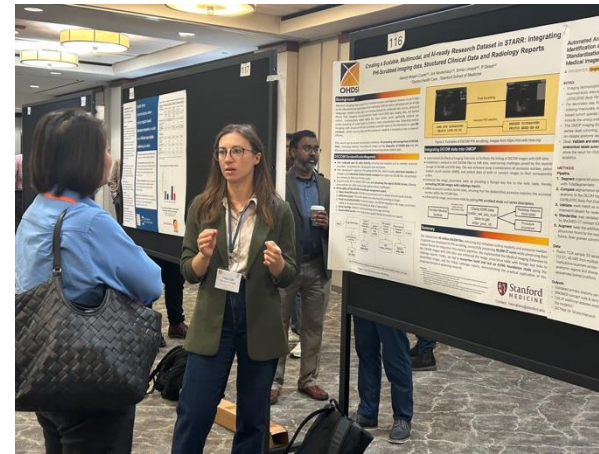
Date	Topic
Oct. 6	FALCON Symposium Report
Oct. 13	Global Symposium Poster Preview / Final Logistics
Oct. 20	NO CALL – Global Symposium
Oct. 27	Welcome to OHDSI: What We Do, How Can You Make an Impact
Nov. 3	Meet the Titans
Nov. 10	Collaborator Showcase Honoree Presentations
Nov. 17	Early-Stage Researcher Presentations
Nov. 24	Software Demos
Dec. 1	OHDSI/OMOP Research Spotlight



Symposium Poster/Demo Preview

During the Oct. 13 community call, Collaborator Showcase presenters can use 60 seconds to share why OHDSI Symposium attendees should visit their poster or software demo.

We want to hear from you!





Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?



OHDSI Shoutouts!



Congratulations to the team of **Yuqing Feng, Martin Yi, Catherine C. Aftandilian, Melissa P. Beauchemin, Christine Cunningham, David DeStephano, Rhea K. Khurana, Emily Larimer, Jennifer Oberg, Benjamin L. May, Keith Morse, Karthik Natarajan, David Noyd, Victoria Soucek, Yujie Chen, Adam Yan, Lin Lawrence Guo, and Lillian Sung** on the recent publication of **Multi-Center Evaluation of Opioid Administration or Prescription in Pediatric Oncology Patients Using the OMOP CDM** in *Cancer Medicine*.

Cancer Medicine

WILEY

Cancer Medicine

RESEARCH ARTICLE OPEN ACCESS

Multi-Center Evaluation of Opioid Administration or Prescription in Pediatric Oncology Patients Using the OMOP CDM

Yuqing Feng¹ | Martin Yi¹ | Catherine C. Aftandilian² | Melissa P. Beauchemin³ | Christine Cunningham² | David DeStephano² | Rhea K. Khurana³ | Emily Larimer⁴ | Jennifer Oberg⁵ | Benjamin L. May⁵ | Keith Morse⁶ | Karthik Natarajan⁷ | David Noyd⁴ | Victoria Soucek⁴ | Yujie Chen¹ | Adam Yan^{1,8} | Lin Lawrence Guo¹ | Lillian Sung^{1,8}

¹Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada | ²Department of Pediatrics, Stanford University School of Medicine, Palo Alto, California, USA | ³Columbia University School of Nursing/Herbert Irving Cancer Center, New York, New York, USA | ⁴Ben Towne Center for Childhood Cancer and Blood Disorders Research, Department of Pediatrics, Seattle Children's Hospital, University of Washington, Seattle, Washington, USA | ⁵Department of Pediatrics, Columbia University Irving Medical Center, New York, New York, USA | ⁶Stanford Center for Biomedical Informatics Research, Stanford University, Palo Alto, California, USA | ⁷Data Sciences Unit, Columbia University, New York, New York, USA | ⁸Division of Haematology/Oncology, The Hospital for Sick Children, Toronto, Ontario, Canada

Correspondence: Lillian Sung (lillian.sung@sickkids.ca)

Received: 22 December 2025 | Revised: 11 September 2026 | Accepted: 17 September 2026

Keywords: OMOP common data model | oncology | opioid | pediatric

ABSTRACT

Purpose: Opioid use in pediatric oncology is poorly characterized. The primary objective was to describe opioid administration or prescription patterns in pediatric oncology patients.

Methods: In this multi-center retrospective study conducted across three sites, we included patients aged 0–19 years with a malignant neoplastic disease who received their first chemotherapy between January 2019 and March 2024. Data were extracted from each site's Observational Medical Outcomes Partnership Common Data Model. We described patient characteristics based on: (1) first opioid administration or prescription, overall and for each specific opioid; and (2) all opioid administrations or prescriptions, overall and for each specific opioid. We also evaluated variability in opioid and naloxone use.

Results: Of 2903 pediatric patients with cancer, 2707 (93.2%) were administered or prescribed at least one opioid. Among these patients, fentanyl was the most common (92.4%), followed by morphine (68.7%), hydromorphone (55.3%), and oxycodone (38.6%). Across centers, the proportion of patients receiving fentanyl ranged from 90.9% to 94.9%, morphine from 34.0% to 89.2%, oxycodone from 0.9% to 76.1%, and naloxone from 2.4% to 29.5%. Overall, 10.9% were administered or prescribed naloxone, but among 169 methadone patients, 43.5% were administered or prescribed it. Of the 153,849 total opioid administrations or prescriptions, 22.4% occurred more than 12 months after initiating chemotherapy.

Conclusions: Among pediatric patients with cancer who receive chemotherapy, 93.2% were administered or prescribed at least one opioid; fentanyl was the most common. There was heterogeneity in opioid choice and naloxone use across sites. These findings lay preliminary groundwork for understanding opioid use patterns and identifying potential areas for practice improvement in pediatric oncology.



OHDSI Shoutouts!



Congratulations to the team of **Elif Inan-Eroglu, Daloha Rodriguez-Molina, Khaled Abdelgawwad, Carina Dinkel-Keuthage, Montse Soriano Gabarró, Pareen Vora, Gianmario Candore, David Vizcaya, and Katrin Manlik** on the recent publication of **Advancing multi-database cohort studies: lessons learned from a studyathon in a pharmaceutical company** in *JAMIA Open*.

JAMIA Open, 2026, 9(5), ooag189
<https://doi.org/10.1093/jamiaopen/ooag189>
Case Report

AMIA OXFORD
ADVANCING MEDICAL RESEARCH, IMPROVING THE WORLD

Advancing multi-database cohort studies: lessons learned from a studyathon in a pharmaceutical company

Elif Inan-Eroglu, PhD^{1,2}, Daloha Rodriguez-Molina, PhD^{1,2}, Khaled Abdelgawwad, MSc^{1,2}, Carina Dinkel-Keuthage, MSc^{1,2}, Montse Soriano Gabarró, MD, MSc^{1,3,4}, Pareen Vora, PhD^{1,2}, Gianmario Candore, MSc^{1,2}, David Vizcaya, PhD^{2,4}, Katrin Manlik, MSc^{*1,2}

¹Pharmaceuticals, Global Medical & Evidence, Bayer AG, 13353 Berlin, Germany
²Pharmaceuticals, Global Medical & Evidence, Bayer Hispania SL, 08970 Sant Joan Despí, Spain
³Current affiliation: Independent Researcher, 14482 Potsdam, Germany
⁴Current affiliation: Alexion Pharmaceuticals SL, 08029 Barcelona, Spain
^{*}Corresponding author: Katrin Manlik, MSc, Pharmaceuticals, Global Medical & Evidence, Bayer AG, Müllerstraße 178, 13353 Berlin, Germany (katrin.manlik@bayer.com)

Abstract

Objectives: To evaluate the feasibility and applicability of the studyathon format within a pharmaceutical company setting for generating reliable real-world evidence rapidly for a business-relevant research question.

Materials and methods: Conducted at Bayer, the studyathon assessed recurrent events following a first-ever transient ischemic attack. It involved an 8-week preparatory phase, a 1-week execution phase, and a 3-week post-event phase, using multiple real-world data sources.

Results: The studyathon concluded successfully with a conference abstract submission. Key learnings included the importance of defining precise research questions, organizational alignment, thorough preparation, and leveraging diverse expertise and skill sets.

Discussion: The studyathon format proved suited for a pharmaceutical setting and revealed insights into its application. Our learnings highlight differences from multi-stakeholder community studyathons.

Conclusion: The studyathon demonstrated its potential as a valuable research approach within a pharmaceutical company. It improved study timelines and cross-functional collaboration while preserving scientific rigor.

Key words: studyathon; real-world data; real-world evidence; OHDSI; OMOP common data model.

Lay Summary

In July 2024, Bayer conducted an innovative research event called a “studyathon” to rapidly generate evidence from routine healthcare data on recurrent events following a first-ever transient ischemic attack. This event brought together a team of 18 experts to work intensively over a short period, using standardized data and analytical tools. The studyathon was structured into 3 phases: preparation, execution, and post-event activities. It successfully resulted in a conference abstract submission and provided valuable lessons on organizing such events within a pharmaceutical company. Key insights included the importance of having a clear research question, aligning organizational goals, and involving a diverse group of experts. This approach not only generated evidence in a timely manner but also improved teamwork across different functions within the company. The studyathon format shows promise as a new way for pharmaceutical companies to conduct internal research efficiently and effectively.



OHDSI Shoutouts!



Congratulations to the team of **Francis Kanyike, Annet Nanungi, Harriet Dickinson, Atif Adam, James Brash, Thu Do, Caroline Otike, Michael Bogart, Alex Asiimwe, Mui Van Zandt, and Cissy Kityo Mutuluuza** on the recent publication of **From fragmentation to federation: multi-partner OMOP implementation in Uganda enabling global real-world evidence generation in *Frontiers in Digital Health*.**



frontiers
in Digital Health

TYPE Original Research
PUBLISHED 17 September 2026
DOI 10.3389/fdgth.2026.1907568



OPEN ACCESS

EDITED BY
Lowryanne Vick,
Office of Connected Care, United States

REVIEWED BY
Achilles Kwanuka,
Medical Research Council/Uganda Virus
Research Institute and London School of
Hygiene & Tropical Medicine Uganda
Research Unit, Uganda
Masoud Mahundi,
University of Dar es Salaam, Tanzania

*CORRESPONDENCE
Francis Kanyike
✉ fkanyike@jcrc.org.ug

[†]PRESENT ADDRESS
Boehringer Ingelheim Ltd., London,
United Kingdom

RECEIVED 12 June 2026
REVISED 14 August 2026
ACCEPTED 31 August 2026
PUBLISHED 17 September 2026

CITATION
Kanyike F, Nanungi A, Dickinson H,
Adam A, Brash J, Do T, Otike C,
Bogart M, Asiimwe A, Zandt MV and
Mutuluuza CK (2026) From
fragmentation to federation: multi-
partner OMOP implementation in
Uganda enabling global real-world
evidence generation.
Front. Digit. Health 8:1907568.
doi: 10.3389/fdgth.2026.1907568

COPYRIGHT
© 2026 Kanyike, Nanungi, Dickinson,
Adam, Brash, Do, Otike, Bogart, Asiimwe,
Zandt and Mutuluuza. This is an open-
access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

From fragmentation to federation: multi-partner OMOP implementation in Uganda enabling global real-world evidence generation

Francis Kanyike^{1*}, Annet Nanungi¹, Harriet Dickinson^{2†},
Atif Adam³, James Brash³, Thu Do⁴, Caroline Otike¹,
Michael Bogart⁴, Alex Asiimwe⁴, Mui Van Zandt³ and
Cissy Kityo Mutuluuza¹

¹Joint Clinical Research Centre (JCRC), Kampala, Uganda, ²Gilead Sciences, London, United Kingdom,
³IQVIA Inc, Durham, NC, United States, ⁴Gilead Sciences, Foster City, CA, United States

Background: Incorporating geographically diverse populations into real-world evidence studies is essential for generating globally applicable insights. However, many African healthcare institutions operate highly customized electronic medical record systems that lack interoperability with international data networks. This paper describes a collaborative implementation to transform legacy electronic medical record data into the OMOP Common Data Model at the Joint Clinical Research Centre (JCRC) in Uganda, enabling participation in federated real-world evidence networks.

Methods: A tri-partite collaboration was established between JCRC, IQVIA, and Gilead Sciences. The project was executed in three phases: (1) data assessment and architecture design, (2) extract, transform, and load execution and custom terminology mapping, and (3) validation of the transformed data through a cohort characterization study. JCRC's Integrated Clinical Enterprise Application (ICEA) database, containing over 48,000 patient records, was assessed and transformed to OMOP CDM v5.4. Following conversion, a descriptive observational study characterized four cohorts: all patients, people with HIV (PWH), hepatitis B virus (HBV)-infected, and hepatitis C virus (HCV)-infected patients.

Results: The source database comprised 168 tables and 1,894 fields with no standardized medical coding systems—100% custom terminology—requiring manual mapping. Over 100 h of collaborative mapping achieved 95% data capture for core domains. A total of 48,748 unique patients were successfully mapped to the OMOP CDM. The cohort characterization demonstrated the utility of the transformed data, revealing: HIV prevalence of 38.1% ($n = 18,588$);



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	9 am	Tidy R Programming
Thursday	9 am	European Community Call
Thursday	10 am	Rare Diseases
Thursday	10 am	Africa Chapter
Thursday	10 am	GIS-Geographic Information Systems
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Friday	11 pm	China Chapter
Monday	9 am	Vaccine Vocabulary
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	10 am	CDM Survey



Congratulations to our 2026 Titan Nominees!

ASSURE team, David Bard, Robert Barrett, Clair Blacketer, Rich Boyce, Danielle Boyce, Evanette Burrows, Cindy Cai, Carrot and Bunny Team, Cindy Chen, Pascal Coorevits, Data4Life, Scott Duvall, Noémie Elhadad, Lee Evans, Eye Care and Vision Research WG, FastOMOP, Mornin Feng, Davera Gabriel, Josh Gagne, Sarah Gasman, Lars Halvorsen, Peter Hoffmann, Christian Högberg, Michelle Hribar, Jason Hsu, Ben Illigens, Andy Kanter, Elisse Katzman, Masha Khitrin, Mack Kigada, Chungsoo Kim, Chris Knoll, Haeun (Hannah) Lee, Gaeun Lee, Renske Los, Benjamin Martin, Valentina Martufi, Mind Meets Machine, Evan Minty, Karthik Natarajan, Bill O'Brien, Sean O'Reilly, OHDSI UK organizers, Julio Oliveria, Anna Ostropolets, Woo Yeon (Jen) Park, Rowan Parry, Paul Petraro, Melanie Philofsky, Hanieh Razzaghi, Jenna Reys, Harry Reyes Nieva, Marti Sabate, Craig Sachson, Alexander Saelmans, Martijn Schuemie, Sarah Seager, Anthony Sena, Cynthia Sung, Jim Todd, Mui Van Zandt, Annelies Verbiest, Fleur Vereijken, Katia Verhamme, Ilse Vermeulen, David Vizcaya, Tom White, Qi Yang, Seng Chan You, Bingyu Zhang

October Newsletter is Available



The Journey Newsletter (October 2026)

Welcome to October, which means it's officially **OHDSI Global Symposium month!** We are just weeks away from gathering in New Brunswick, NJ, Oct. 20–22, and we can't wait to reconnect with our global community in person. In this edition, we dive into the major oncology enhancements featured in the latest vocabulary refresh, share key updates on OMOP v5.5, and spotlight community leaders Asieh Golozar and Rupa Makadia. Along with the latest community updates, publications, and presentations, there is plenty to explore as we count down the days to symposium week.

Podcast: Oncology/Vocab Refresh, OMOP 5.5, Upcoming Global Symposium



On The Journey

In the October 2026 On The Journey podcast, Patrick Ryan and Craig Sachson discussed the latest oncology enhancements that highlighted the latest refresh of the OHDSI standardized vocabularies. They also discussed OMOP v5.5, and specifically what has been added and who should be considered making the upgrade. They also share final thoughts on the upcoming global symposium, being held Oct. 20-22 in New Brunswick, N.J. (If video does not appear, please click 'View this email in your browser')

Community Updates

Where Have We Been?

- The Medical Imaging workgroup announced the formation of the OHDSI Imaging Evidence Network during a community call spotlight session earlier this month. If you are interested, there will be both a tutorial and a workgroup activity focused on imaging during the Global Symposium.
- The fourth annual [UK Symposium](#) was held September 18 at the University of Nottingham, preceded by an OMOP Training Day on September 17. Thank you to the leadership team for putting together this terrific event.
- The FALCON (Federated Alliance for Large-scale Cancer Observational Network) Research Network, which is dedicated to advancing real-world oncology evidence through federated health analyses, hosted its 2026 Symposium and Bladder Cancer Studyathon this week at Dunden in Antwerp, Belgium (see below). This intensive four-day event opened with a general symposium followed by a three-day collaborative working session for researchers, clinicians, and data scientists to generate high-quality clinical evidence on the OMOP CDM.



Where Are We Now?

- [Secure your spot](#) now for the **2026 OHDSI Global Symposium**, coming Oct. 20-22 in New Brunswick, N.J. The event will include [a day of tutorials](#), [the main conference](#) featuring plenaries and the collaborator showcase, and [a day of workgroup activities](#). Information on hotel booking, hosting an exhibit or serving as a sponsor is also on [the event homepage](#).
- The second OMOP School will be held Oct. 13-16 in Uppsala, Sweden, and it reached full capacity with 22 participants coming from five countries.
- DARWIN EU, now in its fourth year, delivers real-world evidence from across Europe on diseases, populations and the uses and performance of medicines using the OMOP CDM and OHDSI open-source tools and practices. Please join our Oct. 6 community call to learn about recent accomplishments and upcoming goals for this initiative.

My Journey: Asieh Golozar



In the latest installment of our "My Journey" series, Asieh Golozar (Founder and CEO, Nemesis Health) shares how a chance discovery during a 2016 interview led her into the global OHDSI network. Asieh discusses her pride in leading the OHDSI Oncology Workgroup, moving the needle on cancer research through OMOP, and why genuine collaboration allows researchers to overcome individual limitations to create real-world clinical impact. (If video does not appear, please click 'View this email in your browser')

Latest Vocabulary Refresh Brings Major Oncology Enhancements, Rare Disease Coverage, and More



Why download the release

- **Expanded SNOMED content:**
+7k terms for diseases and procedures
and +32k new hierarchical relationships
- **More measurements in LOINC:**
+4k new labs and improved grouping
- **More drugs and drug mappings:**
+500 RxNorm codes and +20k mappings
- **Rare disease coverage:**
Mondo and Orphanet with 31k of new rare disease terms
- **Oncology harmonization:**
Major refreshes and improvements
- New **metadata** tables in the bundle

The standardized vocabularies are the engine that powers OHDSI research, ensuring health data from across the globe speaks the same universal language. In our latest release, the vocabulary team rolled out substantial upgrades across standard sources like SNOMED, LOINC, and RxNorm, alongside expanded coverage for rare diseases. These ongoing improvements make it easier for our community to turn messy, fragmented electronic health records into clean, reliable observational research.

Spotlight: Rupa Makadia



What excites me most is how OHDSI has made large-scale observational research both more standardized and more open to participation. When I first became involved, there were only a few core workgroups. Today, the community spans an extraordinary range of clinical specialties and methodological areas, and each year it contributes more science, more tools, and more expertise to the ecosystem.



Rupa Makadia is a Director in Observational Health Data Sciences at Johnson & Johnson, based in Raritan, New Jersey. Her work focuses on generating real-world evidence to evaluate the safety and effectiveness of medical products, with particular emphasis on oncology and pregnancy research.

She leads the Perinatal and Reproductive Health Workgroup, fondly known as PRHeG, within the OHDSI community and contributes to methods development and network studies. Rupa holds a PhD in Biomedical Informatics from Rutgers University, an MS in Biostatistics from California State University, East Bay, and a BS in Biology from Cal Poly, San Luis Obispo. A longtime OHDSI member, she has helped advance cross-network initiatives and research throughout the community.

Rupa discusses her career journey and research interests, upcoming initiatives for the Perinatal and Reproductive Health Workgroup, the OHDSI approach and more in the latest edition of the Collaborator Spotlight.

Spotlight: Rupa Makadia



Congratulations to our 2026 Titan Nominees!

ASSURE team, David Bard, Robert Barrett, Clair Blacketer, Rich Boyce, Danielle Boyce, Evanette Burrows, Cindy Cai, Carrot and Bunny Team, Cindy Chen, Pascal Coorevits, Data4Life, Scott Duval, Noémie Elhadad, Lee Evans, Eye Care and Vision Research WG, FastOMOP, Morrin Feng, Davera Gabriel, Josh Gagne, Sarah Gasman, Lars Halvorsen, Peter Hoffmann, Christian Höglberg, Michelle Hribar, Jason Hsu, Ben Illigens, Andy Kanter, Elisse Katzman, Mascha Khitrun, Mack Kigada, Chungsoo Kim, Chris Knoll, Haeun (Hannah) Lee, Gaeun Lee, Renske Los, Benjamin Martin, Valentina Martuli, Mind Meets Machine, Evan Minty, Karthik Natarajan, Bill O'Brien, Sean O'Reilly, OHDSI UK organizers, Julio Oliveria, Anna Ostropolets, Woo Yeon (Jen) Park, Rowan Parry, Paul Petraro, Melanie Philofsky, Hanieh Razzaghi, Jenna Reys, Harry Reyes Nieva, Marti Sabate, Craig Sachson, Alexander Saelmann, Martijn Schuermie, Sarah Seager, Anthony Sena, Cynthia Sung, Jim Todd, Mui Van Zandt, Annelies Verbiest, Fleur Verrijken, Katia Verhamme, Ilse Vermeulen, David Vitzaya, Tom White, Qi Yang, Seng Chan You, Bingyu Zhang

September Publications

Ross E, Bouissou O, Helland Å, Faxvaag A. [From documentation to discovery: clinicians' perspectives on the generation, usability and standardization of real-world data](#). Int J Med Inform. 2026 Sep 1217:106501. doi: 10.1016/j.ijmedinf.2026.106501. Epub 2026 May 20. PMID: 42176598.

Lourengo Santos R, Cruz-Correia RJ. [An HL7 FHIR IG for lifestyle medicine in learning health systems: Multi-vendor wearable interoperability with documented terminology gaps](#). Int J Med Inform. 2026 Sep 1217:106465. doi: 10.1016/j.ijmedinf.2026.106465. Epub 2026 May 5. PMID: 42176600.

Haderlein TP, Der-Martirosian C, Bensken WP, Schuff R. [OMOP common data model transformation: leveraging a nationwide community-based health care database to support AI/ML research](#). J Am Med Inform Assoc. 2026 Sep 1;33(9):1785-1789. doi: 10.1093/jamia/ocag087. PMID: 42319147; PMCID: PMC13580724.

Patterson J, Minto E, Beaton M, Anand A, Velez M, Harris P, Hripscak G, Natarajan K. [A comparison of Fast Healthcare Interoperability Resources and Observational Medical Outcomes Partnership electronic health record data within the All of Us Research Program](#). J Am Med Inform Assoc. 2026 Sep 1;33(9):1593-1604. doi: 10.1093/jamia/ocag095. Erratum in: J Am Med Inform Assoc. 2026 Aug 17;ocag134. doi: 10.1093/jamia/ocag134. PMID: 42334204; PMCID: PMC13580739.

Sherwin EB, Martin B, Xu X, Booman A, Howren A, McKay K, Kazi S, O'Reilly S, Kolla KS, Aziz K, Bianco K, Leonard SA. [Congenital heart disease in pregnancy and severe maternal morbidity: a distributed data network study](#). Pregnancy (Hoboken). 2026 Sep;2(5):e70370. doi: 10.1002/pmf2.70370. Epub 2026 Aug 4. PMID: 42569246; PMCID: PMC13449514.

Hoppe JN, Lee J, Akinyemiju TF, Bitterman DS, Budro M, Green MF, Huser V, Lipson JA, Mishra S, Nussbaum DP, Patel JN, Petkov VI, Rasmussen KM, Sirintrapun SJ, Spears PA, Van Sandik S, Meyer AM, Topaloglu U, Warner JL. [The GENIE Data Model: A Solid Tumor, Person-Centric Framework for Scalable, Harmonized Precision Oncology Data Collection](#). Cancer Res Commun. 2026 Sep 1;6(9):2242-2255. doi: 10.1158/2767-9764.CRC-26-0148. PMID: 42658634; PMCID: PMC13598700.

Bellucci G, Gu W, Rose PW, Baranzini SE. [An agentic AI framework connecting language models to electronic health records and a biomedical knowledge graph for real-world evidence](#). Front Artif Intell. 2026 Sep 4;9:188353. doi: 10.3389/frai.2026.188353. PMID: 42761240; PMCID: PMC13585945.

Yu EH, Lee HJ, Han YM, Joo JK. [Association Between Endometriosis and Autoimmune Thyroid Disease Using a Multicenter Observational Medical Outcomes Partnership \(OMOP\) Common Data Model](#). J Clin Med. 2026 Sep 4;9:188353. doi: 10.3389/jcm15176919. PMID: 42739921; PMCID: PMC13566532.

mailchi.mp/ohdsi/october2026

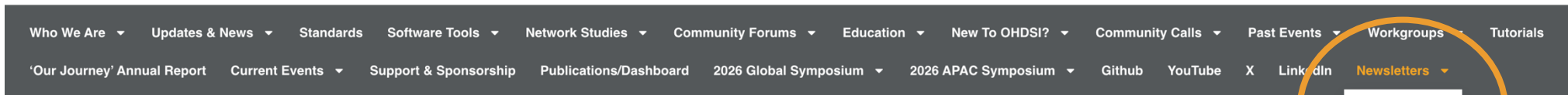
#JoinTheJourney

www.ohdsi.org





October Newsletter is Available



Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University.

Learn more [about our community](#), about [our mission, vision and values](#), and how you can [help support the OHDSI community](#).

[Join the Journey](#)

Join us at the 2026 OHDSI Global Symposium

Registration is OPEN for the 2026 OHDSI Global Symposium, which will be held Oct. 20-22 in New Brunswick, NJ. This event unites hundreds of collaborators to showcase scientific innovations and build bridges for future research. Together, we are advancing our mission to generate real-world evidence that informs better health decisions and improves patient care.

[Global Symposium Homepage](#)

[Register Me for OHDSI 2026](#)

What OHDSI is all about

OHDSI's mission

To improve health by empowering a

Latest News

• [Weekly OHDSI Digest – October 5, 2026](#)

[mailchi.mp/ohdsi/october2026](https://us4.campaign-archive.com/home/?u=9402e3dd40021ebda4b15a4b7&id=af9e5362f5)



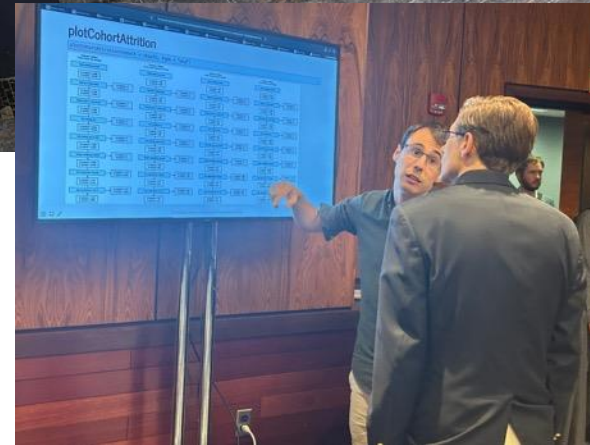
2026 OHDSI Global Symposium

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**
- **Using OMOP Model in Registry Context & Clinical Trials Standardization Context: Conventions, Past Use Cases, SDTM & Regulatory Consideration, Challenges**



2026 Global Symposium Agenda

Start	End	Topic	Presenter/Lead
7:15 am	8:00 am	Newcomer Welcome	Paul Nagy/Marc Suchard
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	11:00 pm	OHDSI Jam Session	Martijn Schuemie



2026 Symposium Workgroup Activities

Session 1 (8 am – 10 am): Asia-Pacific Chapter; Early-Stage Researchers; Eyecare & Vision Research; Gen AI, Phenotyping, Vocabularies: AI-assisted Phenotyping Connectathon (1 of 2); Geographic Information System; HADES Hackathon (1 of 3); Health Equity (1 of 2); Industry (1 of 2); Oncology (1 of 2); Tidy R programming with OMOP (1 of 2)

Session 2 (10:30 am – 12:30 pm): Gen AI, Phenotyping, Vocabularies: AI-assisted Phenotyping Connectathon (2 of 2); HADES Hackathon (2 of 3); Health Equity (2 of 2); Industry (2 of 2); Medical Imaging; Oncology (2 of 2); Perinatal and Reproductive Health; Rare Disease; Tidy R programming with OMOP (2 of 2); Waveform

Session 3 (1:30 pm – 3:30 pm): ATLAS/WebAPI; CDM Survey; Dentistry; Evidence Network Data Partners; GIS/Waveform Cross Pollination Meeting-Extension Table Status, Integration, and Approaches; HADES Hackathon (3 of 3); Health Economics & Value Assessment; Natural Language Processing (NLP); Surgery and Perioperative Medicine; Women of OHDSI

Session 4 (3:30 pm – 5 pm): Workgroup Summary Session (with prizes!)



AI-assisted Phenotyping Connectathon

Joint OHDSI Working Group Workshop at the Global Symposium

THURSDAY, OCTOBER 22

8:00 AM – 12:30 PM | In person

Connecting phenotypers, evaluators, and tool developers

Explore how emerging tools, including Generative AI and agentic solutions, can support activities across the phenotype development and evaluation lifecycle.



2026 OHDSI Global Symposium

Oct. 20-22 • New Brunswick, N.J. • Hyatt Regency Hotel

Explore

Real-world use cases and challenges across the phenotype lifecycle

Evaluate

Tool usability, accountability, deployability, strengths, and limitations

Build

A community-driven tool–use case matrix with examples and metadata

Who should join?

Phenotype developers and evaluators • Clinicians and domain experts • Researchers • Generative AI and tool developers • Anyone interested in advancing phenotyping within OHDSI

JOIN US

Register for the symposium and workshop

[View full announcement](#)



SCAN FOR DETAILS



APAC Symposium: Nov. 13-15

The 2026 APAC Symposium will be held Nov. 13-15 at Yonsei University in Seoul, South Korea.

Nov. 13: Main Conference

Nov. 14: Tutorials, Datathon

Nov. 15: Datathon Team Activities

Registration is open! All details can be found on the event homepage.



ohdsi.org/APAC2026



Canada Symposium: Nov. 17-18

Registration is now open for the 2026 OHDSI Canada Symposium, held Nov. 17-18 in Toronto, Ont.

Nov. 17: Workshops

Track A: From Source Data to OMOP

Track B: From OMOP to Evidence

Nov. 18: Symposium

The main symposium day will bring the Canadian OHDSI community together to share experiences, showcase emerging work, and explore opportunities for collaboration.



JOIN US IN TORONTO THIS NOVEMBER

2026 OHDSI Canada Symposium

DATE
Nov 17–18, 2026

LOCATION
Toronto, ON

Join the OHDSI Canada community in Toronto for two days of hands-on workshops, engaging talks, strategic discussions, and networking.

Workshop Day

Practical sessions on vocabulary mapping, ETL, and using OMOP and OHDSI to answer real-world research questions.

Symposium Day

Presentations, strategic discussions, and opportunities to connect with colleagues from across the OHDSI Canada community.

India Symposium: Dec. 11



OHDSI INDIA SYMPOSIUM 2026

11 DECEMBER 2026

CHENNAI

PARTNERING FOR IMPACT, GROWTH, AND VISIBILITY

COLLABORATE. INNOVATE. TRANSFORM.

A STRONG COMMUNITY | **DATA-DRIVEN INSIGHTS** | **REAL-WORLD IMPACT** | **GLOBAL VISION, LOCAL ACTION**

Be a Part of the Change



WHY TO ATTEND?

- **UNDERSTAND INDIA'S HEALTHCARE DATA LANDSCAPE**
Learn about the current state of healthcare data in India, emerging national priorities, and how global standards are shaping the future of healthcare and research.
- **EXPLORE OMOP AND HEALTHCARE DATA INITIATIVES**
Discover the latest developments in OMOP, Real-World Evidence (RWE), FHIR, AI, and other data standardization initiatives being implemented across India.
- **LEARN HOW YOUR ORGANIZATION CAN PARTICIPATE**
Understand how hospitals, medical colleges, and healthcare organizations can contribute to national data initiatives, multicenter research, training programs, and collaborative projects through OHDSI India.
- **LEARN FROM EXPERTS**
Gain insights from national and international experts on healthcare data standards, OMOP, FHIR, Real-World Evidence, AI, and observational research.
- **NETWORK WITH THE HEALTHCARE COMMUNITY**
Connect with clinicians, researchers, hospitals, academia, industry partners, and global OHDSI leaders to build meaningful collaborations
- **SHOWCASE YOURWORK**
SHOWCASE YOURWORK Present your research through posters, lightning talks, or case studies and receive valuable feedback from the community
- **BUILD FUTURE COLLABORATIONS**
Identify opportunities to participate in multicenter research, training programs, industry collaborations, and Real-World Evidence projects.



Job Openings at Illinois-Chicago



Director of Informatics and Data Science Operations and Development

Chicago, IL, United States | 1042586

[Apply Now](#)[Share](#)[Save Job](#)

Posted on: 9/22/2026 - Application Deadline: 10/15/2026

The Director of Informatics and Data Science Operations and Development provides strategic leadership and operational management for the Informatics and Data Science Program within the Center for Clinical and Translational Science (CCTS). Funded through the Clinical and Translational Science Awards Program (CTSA), the Program advances the national mission of accelerating the translation of scientific discoveries into improved health outcomes. This is achieved through the development of robust informatics infrastructure, the promotion of interdisciplinary team science, the enablement of high-quality and interoperable data integration, and the support of innovative, rigorous, and reproducible clinical and translational research.

This position bridges advanced technical expertise with real-world health implementation by identifying translational bottlenecks, developing novel methodological and analytic approaches, and advancing informatics initiatives aligned with institutional strategy and national CTSA priorities. In collaboration with Data Science/Informatics Director(s) this role contributes to shaping the scientific direction and growth of informatics-driven research initiatives across the CCTS and the UIC health sciences colleges, strengthening the University's impact within the national translational science ecosystem. The role manages multidisciplinary teams responsible for the design, development, implementation, and oversight of innovative informatics and data science solutions that advance clinical and translational research across the full research continuum.

This is an academic professional, non-faculty position responsible for advancing the institution's clinical and translational research informatics capabilities through strategic management, operational oversight, and collaboration with faculty and institutional leaders.

Responsibilities:

Operational Leadership and Team Management
Provide strategic leadership and day-to-day management for the CCTS Informatics/Data Science team.
Supervise, mentor, and develop informatics specialists and staff, including performance management, coaching, and professional development planning.
Foster a collaborative, accountable, and high-performing team culture.
Establish and enforce team norms, standard operating procedures, and quality standards in alignment with Center policies, Campus guidelines, and state/federal regulations.
Oversee resource allocation and project prioritization to ensure timely and high-quality delivery of services.
Assign personnel to projects; guide development of project plans and timelines; monitor progress and ensure achievement of strategic objectives.
Track and report scientific and operational progress, resource utilization, performance metrics, and quality indicators.
Identify operational or scientific barriers and implement solutions to ensure successful project outcomes.

Scientific and Technical Leadership
Advise Center leadership on innovative informatics strategies that align with local and national research priorities and sponsor requirements.
Lead needs assessments with investigators to identify opportunities for clinical information systems and data services.
Monitor emerging trends in biomedical informatics, clinical data science, AI, and translational research.
Oversee the design, development, and implementation of clinical informatics solutions that support research, quality improvement, and health intervention initiatives.
Ensure scientific rigor, regulatory compliance, and reproducibility across projects.

Strategic Partnerships and Collaboration
Lead collaborative informatics and data science initiatives with internal and external stakeholders to achieve objectives outlined in the CCTS grant.
Partner with the Center and Program leadership to represent CCTS in national networks and collaboratives, including the Clinical and Translational Science Awards Program Consortium.
Strengthen institutional partnerships with UI Health IT, UIC's Advanced Cyberinfrastructure for Education and Research (ACER), and the AI Health4All Center.
Serve as liaison between campus units and CCTS leadership to ensure alignment of informatics and data science strategies with institutional priorities.
Identify, evaluate, and recommend collaborative opportunities through engagement with CTSA hubs, Enterprise Committees, Working Groups, and professional associations.
Develop coalitions that expand institutional capabilities and enhance national visibility.



Postdoctoral Research Associate (Hribar Lab) - Ophthalmology & Visual Sciences

Chicago, IL, United States | 1042846

[Apply Now](#)[Share](#)[Save Job](#)

Posted on: 9/15/2026 - Application Deadline: 10/30/2026

Position Summary

The Postdoctoral Research Associate will work on the standardization and integration of electronic health record (EHR) and imaging data in ophthalmology. The focus will be on standardizing to the OMOP common data model to support federated learning and data networks. The position will have the opportunity to work collaboratively with data and AI models at UIC as well as other sites in the OHDSI Eye Care and Vision Research workgroup. In addition, the Postdoc will be using the AI-READI dataset as a test set for the models and federated network.

Duties & Responsibilities

- Perform other related duties and participate in special projects as assigned.

Minimum Qualifications

- PhD, MD, or equivalent terminal degree required
- Informatics, data science, or computer science training and/or experience

The University of Illinois System is an equal opportunity, affirmative action employer dedicated to building a community of excellence, equity, and diversity. We are committed to fostering an inclusive environment and we welcome applications from qualified individuals of all backgrounds and identities. The University of Illinois System does not discriminate against any applicant or employee based on their real or perceived race, color, religion, sex, national origin, ancestry, age, order of protection status, genetic information, marital status, disability, sexual orientation including gender identity, unfavorable discharge from the military or status as a protected veteran and will comply with all federal and state nondiscrimination, equal opportunity and affirmative action laws, orders and regulations. Visit [University of Illinois Non-Discrimination Statement](#)

The University of Illinois System conducts background checks on all job candidates upon acceptance of a contingent offer of employment. Convictions are not a bar to employment. Background checks will be performed in compliance with state and federal law.

The University of Illinois System requires candidates selected for hire to disclose any documented finding of sexual misconduct or sexual harassment and to authorize inquiries to current and former employers regarding findings of sexual misconduct or sexual harassment. For more information, visit [Policy on Consideration of Sexual Misconduct in Prior Employment](#).

As a qualifying federal contractor, the University of Illinois System uses E-Verify to verify employment eligibility.

Artificial Intelligence (AI) tools may be used in some portions of the candidate review process for this position; however, all employment decisions will be made by a person.

The university provides accommodations to applicants and employees. [Request an Accommodation](#)

Preferred Qualifications



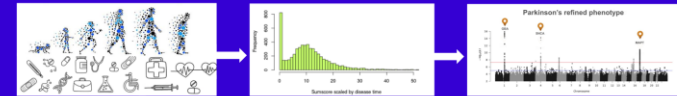
#OHDSISocialShowcase This Week

Monday

Genetic Validation of Evidence-Based Phenotype Refinement in EHR Data: Complementing OHDSI Tools with Biological Signal

(**Marika Kaakinen**, Javier Gracia-Tabuenca, Dawit A. Yohannes, Mary Pat Reeve)

Genetic Signal as an External Benchmark for Phenotype Algorithms FinnGen Phenotype Team



Golden genetic loci let us quantify phenotype algorithm improvement - no chart review needed

- GWAS provides two independent measures of refinement: p-value and effect size
- Weighting evidence across data domains adapts naturally to changing standard-of-care
- Cleaner cohorts improve trajectory, progression, and age-of-onset analyses

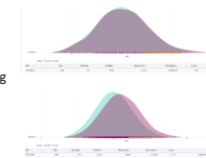
1 Unsupervised EHR code enrichments enable a scaled evidence score.



We weight evidence rather than collapse it, preserving the signal gradient that LASSO-style approaches discard. CodeWAS is similar to CohortDiagnostics Characterization but we use birth year and sex matched controls. Birth year matched controls help adjust for changes in standard-of-care over decades.

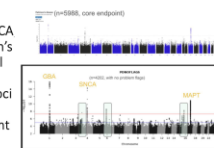
2 Polygenic risk scores (PGS) can also validate phenotyping algorithms

Low-evidence individuals show no polygenic risk for Parkinson's disease, confirming they are not true cases. Retained individuals show strong PGS association.



3 Genetic association studies independently quantitate phenotype improvement with effect size and p-value

Known loci from disease meta-analyses - GBA, SNCA, MAPT for Parkinson's - serve as biological anchors. Signal strength at these loci measures cohort quality, independent of any coding practice.



4 PhenotypeScoring: automating evidence-based cohort refinement on the OMOP CDM (CohortOperations tool)

Starting from CodeWAS enrichments, users select and group features, build a scoring algorithm, and flag low-evidence individuals for exclusion



Marika Kaakinen¹, Javier Gracia-Tabuenca^{1,2}, Dawit A. Yohannes¹, FinnGen¹, Mary Pat Reeve^{1,3}
¹ FinnGen, University of Helsinki, Helsinki, Finland; ² Tampere University, Tampere, Finland; ³ Broad Institute of MIT and Harvard, Cambridge, MA, USA

FinnGen consists of 520,000 genotyped individuals linked to national health registries on the OMOP CDM, providing decades of longitudinal EHR data across 4,000+ disease endpoints. On average, each person has 1,020 events including diagnoses, procedures, medications and lab tests. The combination of whole-population registry coverage and universal genotyping makes it uniquely suited for phenotype-GWAS feedback loops.



(**Karthik Prathaban**, Douwe J. Spaanderman, Farhan Akram, Stefan Klein, Martijn P.A. Starmans)

k.prathaban@erasmusmc.nl



#OHDSISocialShowcase This Week

Wednesday

From Observational Data to Synthetic Cohorts

(Alberto Labarga)

From Observational Data to Synthetic Cohorts

Alberto Labarga
MirrorHealth SL, Pamplona, Spain

The generation of high-quality synthetic health data has become a key enabler for collaborative research, method development, and education within the Observational Health Data Sciences and Informatics (OHDSI) community. Synthetic datasets can mitigate privacy risks, facilitate data sharing across institutions, and support reproducible research workflows without exposing sensitive patient-level information.

In this work, we present a novel synthetic data generation pipeline that leverages existing OMOP Common Data Model (CDM)² databases to automatically learn debiased graph representations of clinical events and their temporal and semantic relationships. These learned representations are then translated into executable Synthea³ modules, enabling scalable and transparent generation of synthetic cohorts that preserve clinically meaningful structure while explicitly addressing bias and fairness concerns.



OMOP-CDM

Extract data into a graph

First, standardized OMOP-CDM tables (e.g., PERSON, CONDITION, OCCURRENCE, DRUG_EXPOSURE, VISIT_OCCURRENCE, etc.) are transformed into an event-centric graph representation¹. Nodes correspond to clinical concepts (conditions, drugs, procedures) and patient states, while edges encode observed co-occurrence, temporal ordering, and transition probabilities derived from the source database.

Knowledge graph

Knowledge graph consolidation

Second, a debiasing step is applied to the learned graph to mitigate known sources of observational bias, including demographic imbalance, care-setting-specific practice patterns, and spurious correlations driven by data sparsity or coding artifacts. This is achieved through a combination of constraint-based normalization, re-weighting strategies, and graph pruning informed by domain-agnostic fairness criteria and configurable demographic stratifications.

Synthea module

Expose graph as Synthea module

Third, the resulting debiased graph is automatically translated into Synthea disease modules, preserving state transitions, conditional logic, and temporal dynamics.

Summary

The proposed pipeline consists of five main stages: (1) representation learning from OMOP-CDM data, (2) debiased graph construction, and (3) Synthea module instantiation, (4) synthetic data generation (5) Synthea to OMOP conversion. The pipeline is designed to be modular, reproducible, and compatible with standard OHDSI tooling, enabling integration into existing observational research workflows. A fully open-source Python implementation orchestrates this process, generating executable Synthea modules and subsequently producing synthetic patient-level data that can be exported directly into OMOP-CDM-compliant datasets.



Figure 1. Synthetic data generation pipeline

Conclusion

This work introduces a principled, data-driven approach for generating synthetic health data that bridges real-world observational datasets and rule-based simulation frameworks. By learning debiased graph representations from OMOP-CDM databases and translating them into executable Synthea modules, the proposed pipeline combines the realism of empirical data with the interpretability and controllability of simulation models.

Results

We evaluated the pipeline using several real-world OMOP-CDM databases covering diverse patient populations and care settings. The generated synthetic datasets were assessed along three dimensions: structural fidelity, statistical similarity, and bias attenuation.⁴

Structural evaluation demonstrated that the synthetic data preserved key OMOP-CDM characteristics, including visit structures, concept hierarchies, and longitudinal patient trajectories. Statistical comparisons showed that marginal and joint distributions of major clinical events closely approximated those of the source data, while avoiding direct record-level replication.

Importantly, bias-focused analyses indicated a measurable reduction in demographic and utilization imbalances when compared with naive data-driven generation approaches. The resulting Synthea modules were human-readable, version-controlled, and reusable, supporting transparency and downstream validation by domain experts.

References

- [1] OHDSI. The Book of OHDSI. <https://ohdsi.github.io/TheBookOfOhdsi/>
- [2] Wainowski J, et al. Synthea™: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association*.
- [3] Kang, Mengjia, et al. "Evolution of a graph model for the OMOP common data model." *Applied Clinical Informatics* 15.05 (2024)
- [4] Kiran, A., P. Rubini, and S. Saravana Kumar. Comprehensive review of privacy, utility and fairness offered by synthetic data. *IEEE Access* (2025)



mirrorhealth.eu



#OHDSISocialShowcase This Week

Thursday

Automated Extraction of Genomic Entities from Spanish and Catalan Clinical Notes using Active Learning

(**Antonio Arcas**, Ricard García-Isern, Angela Leis, Juan Manuel Ramírez-Anguita, Carla Arxé, Jan Carreras, Miguel-Angel Mayer)

Creating an automatic genetic entity extraction pipeline for clinical notes in using Active Learning

Automated Extraction of Genomic Entities from Spanish and Catalan Clinical Notes using Active Learning

Motivation: There are no open-source NLP tools capable of extracting genetic entities from clinical notes in Spanish and Catalan. Integration of genomic data into the OMOP CDM is key to linking genetic biomarkers with phenotypes and adverse drug reactions.

Results: The configurations for the final iteration model achieved the following averaged Macro F1-scores over 5-fold cross-validation:

Config	LR	WD	Warmup	Class Weights	F1
C15	3e-05	0.1	0.2	(0.1, 10.0, 10.0)	0.9408
C12	3e-05	0.01	0.2	(0.5, 5.0, 5.0)	0.9331
C16	3e-05	0.1	0.2	(0.5, 5.0, 5.0)	0.9319
C13	3e-05	0.1	0.1	(0.1, 10.0, 10.0)	0.9305
C7	2e-05	0.1	0.2	(0.1, 10.0, 10.0)	0.9279
C11	3e-05	0.01	0.2	(0.1, 10.0, 10.0)	0.9276
C6	2e-05	0.1	0.1	(0.5, 5.0, 5.0)	0.9265
C8	2e-05	0.1	0.2	(0.5, 5.0, 5.0)	0.9265
C4	2e-05	0.01	0.2	(0.5, 5.0, 5.0)	0.9254
C1	2e-05	0.01	0.1	(0.1, 10.0, 10.0)	0.9249

More results:

- Prototype used for the first iteration of Active Learning labeled genes in 16% of notes.
- Higher learning rates were generally more favorable.
- 491 error records were collected across all 16 configurations and folds.
- FN account for 317 cases (64.6%), FP for 159 (32.4%), and Boundary Errors for only 15 (3.1%).

GENE 1:

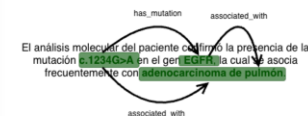
ADENGA: Comentat amb AP, NOS fet de la ultima biopsia pero per error informatic no estava permit. ~~BRAS~~ MUTAT p.G12V~~~BRAS~~ MUTAT c.35G>T, p.(Gly12Val) 36.6% -A mes ~~BRAS~~ MUTAT c.743G>A, p.(Arg248Gln) 51.5% /FBXW7: MUTAT c.139AG>A, p.(Arg46His) 19.2%

Methods

- 1 **Fine-tuned transformer** model roberta-base-biomedical-clinical-es with a CRF classification head with penalty weights to mitigate class imbalance.
- 2 **Bootstrap Active Learning:** used a fine-tuned baby model trained with 50 labeled clinical notes to pre-annotate a larger dataset of 500 clinical notes. This iterative process helps with annotation's cold start.
- 3 **Hyperparameter tuning:** A grid search approach was used to test 16 different configurations regarding: learning rate, weight decay, warmup ratio and class weights.

Under development:

- Expansion to other entities such as mutation and phenotype, transforming the model from a simple gene tagger into a multi-class extractor capable of parsing the full genomic context of a clinical note.
- Adding a Relation Extraction model to link genes and mutations to phenotypes.



We present a useful method for creating genomic NLP tools in low-resource settings, successfully overcoming the obstacles related to data scarcity and extreme class imbalance inherent in clinical narratives. We are developing the first open dataset and model for genetic NERC in Spanish and Catalan.



Antonio Arcas, Ricard García-Isern, Angela Leis, Juan Manuel Ramírez-Anguita, Carla Arxé Jan Carreras, Miguel-Angel Mayer



EUROPEAN OHDSI Symposium 18-20 April 2026 Rotterdam



#OHDSISocialShowcase This Week

Friday

Characterisation of Spanish OMOP CDM Databases Through a Study-a-thon: A Federated Descriptive Study within the OHDSI Europe Spain

(**Angela Leis**, Talita Duarte-Salles, Miguel Angel Mayer, Juan Manuel Ramírez-Anguita, Agustina Giuliodori, Carla Arxé, Alfons Hervàs, Javier Gorricho, Ibai Tamayo, Jose Arco Martín, Julián Librero, Cristina Justo, Ana Llorente, Hermenegildo Martínez-Alcalá, Ainhoa García, Miguel Gil, Anna Palomar, Laura Granés, Juan Luis Cruz, Noelia García, Paula Rubio, Jaime Meléndez-Cardiel, Javier González-Galindo, Carlos Tellería-Orriols, Alberto Moreno, Jesus Moreno, Claudia García, Adrià Ricarte, Natalia López, Maria Soler, Rafael Álvarez, Susanna Aussó, Zaira Benitez, Pilar Saez, Marc Fradera, Martina Pérez, Marc Caravaca-Rodriguez, Pau Pericàs, Joan Vicenç Cladera, Jordi Rodeiro-Boliart, Álvaro Martínez, Carlos López)

OHDSI Spain Study-a-thon: Enabling Federated Research through the Characterisation of 15 OMOP CDM Databases covering more than 55 million Patients

A Federated Characterization of the OHDSI Spanish Node: A Study-a-thon to Map the National Real-World Data Landscape

Background: Spain has a decentralised health system with rich RWD assets. While many sites have mapped to OMOP CDM, there is no consolidated inventory. Within Spain, the OHDSI Spanish Node, coordinated by Hospital del Mar and IDIAP Jordi Gol, has emerged as a collaborative forum to align methods, share ETL and vocabulary expertise, and promote re-use of open analytics. The OHDSI Spanish Node organised a Study-a-thon (Nov. 2025) to create a federated, consistent characterisation of these databases to improve feasibility assessments and multicenter collaboration.

Results

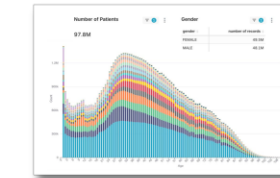
Scale: The node integrates over 15 institutions across 7 autonomous communities.

Population: Coverage ranges from hospital-specific data (0.39M) to national primary care data (BIFAP: 30.3M).

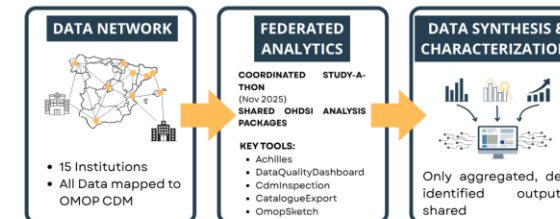
Harmonization: Each site executed a common analysis pipeline (Achilles, DQD, OmopSketch) to generate:

- Standardised Metadata Snapshots.
- Temporal trends by clinical domain.
- Clinical domain coverage and data quality summaries.

Institution	Type of data	Patients
Añeta Santa Ana Hospital Universitario de Navarra	Hospital	0.39M
Servicio Navarro de Salud (BASOSNA)	Hospital and Primary care	1.51M
BIFAP: Agencia Española de Medicamentos y Productos Sanitarios	Primary care	30.3M
Fundació Institut Universitari per a la recerca i l'innovació (FIDUS)	Primary care and hospital	8.89M
Jordi Gol - IDIAP (IDEG)	Hospital and Primary care	1.79M
Hospital del Mar and Hospital del Mar Research Institute	Hospital	2.29M
Hospital Universitario 12 de Octubre	Hospital	0.5M
Hospital Universitario de Bellvitge	Hospital	2.1M
Hospital Universitario Val d'Hebron	Hospital	1.1M
Hospital Universitario Virgen Macarena	Hospital	1.1M
Instituto d'Investigació i Innovació Parc Taulí (I3)	Hospital	1.1M
Instituto Aragonés de Ciencias de la Salud (IACS)	Hospital and Primary care	2.39M
Instituto de Recerca San Josep de Olot	Hospital	0.7M
Instituto d'Investigació Sanitaria de les Illes Balears (IISBa)	Hospital and Primary care	1.89M
Instituto de Investigación Sanitaria La Fe de Valencia (IIS La Fe)	Hospital	2.48M



Methods



Conclusions

This ongoing study has established a structured baseline for the Spanish Node, fostering synergies between national (OmicSpace, OHSIRIS, VHTeDades) and international (OHDSI, EHden and DARWIN EU) initiatives. Next steps: to publish the full sites characterisation and the expansion of the federated network to support multi-site epidemiologic studies.



Angela Leis¹, Talita Duarte-Salles², Miguel Angel Mayer³, Juan Manuel Ramírez-Anguita⁴, Agustina Giuliodori⁵, Carla Arxé⁶, Alfons Hervàs⁷, Javier Gorricho⁸, Ibai Tamayo⁹, Jose Arco Martín¹⁰, Julián Librero¹¹, Cristina Justo¹², Ana Llorente¹³, Hermenegildo Martínez-Alcalá¹⁴, Ainhoa García¹⁵, Miguel Gil¹⁶, Anna Palomar¹⁷, Laura Granés¹⁸, Juan Luis Cruz¹⁹, Noelia García²⁰, Paula Rubio²¹, Jaime Meléndez-Cardiel²², Javier González-Galindo²³, Carlos Tellería-Orriols²⁴, Alberto Moreno²⁵, Jesus Moreno²⁶, Claudia García²⁷, Adrià Ricarte²⁸, Natalia López²⁹, Maria Soler³⁰, Rafael Álvarez³¹, Susanna Aussó³², Zaira Benitez³³, Pilar Saez³⁴, Marc Fradera³⁵, Martina Pérez³⁶, Marc Caravaca-Rodriguez³⁷, Pau Pericàs³⁸, Joan Vicenç Cladera³⁹, Jordi Rodeiro-Boliart⁴⁰, Álvaro Martínez⁴¹, Carlos López⁴²

¹Hospital del Mar and Hospital del Mar Research Institute, Barcelona; ²IDEG Jordi Gol i Guàrdia, Barcelona; ³Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ⁴Servicio Navarro de Salud, Navarra; ⁵Fundació Institut Universitari per a la recerca i l'innovació, Barcelona; ⁶Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ⁷Servicio Navarro de Salud, Navarra; ⁸Jordi Gol - IDIAP, Barcelona; ⁹Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ¹⁰Servicio Navarro de Salud, Navarra; ¹¹Fundació Institut Universitari per a la recerca i l'innovació, Barcelona; ¹²Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ¹³Servicio Navarro de Salud, Navarra; ¹⁴Jordi Gol - IDIAP, Barcelona; ¹⁵Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ¹⁶Servicio Navarro de Salud, Navarra; ¹⁷Fundació Institut Universitari per a la recerca i l'innovació, Barcelona; ¹⁸Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ¹⁹Servicio Navarro de Salud, Navarra; ²⁰Jordi Gol - IDIAP, Barcelona; ²¹Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ²²Servicio Navarro de Salud, Navarra; ²³Fundació Institut Universitari per a la recerca i l'innovació, Barcelona; ²⁴Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ²⁵Servicio Navarro de Salud, Navarra; ²⁶Jordi Gol - IDIAP, Barcelona; ²⁷Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ²⁸Servicio Navarro de Salud, Navarra; ²⁹Fundació Institut Universitari per a la recerca i l'innovació, Barcelona; ³⁰Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ³¹Servicio Navarro de Salud, Navarra; ³²Jordi Gol - IDIAP, Barcelona; ³³Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ³⁴Servicio Navarro de Salud, Navarra; ³⁵Fundació Institut Universitari per a la recerca i l'innovació, Barcelona; ³⁶Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ³⁷Servicio Navarro de Salud, Navarra; ³⁸Jordi Gol - IDIAP, Barcelona; ³⁹Idiixa Santa Ana Hospital Universitario de Navarra, Navarra; ⁴⁰Servicio Navarro de Salud, Navarra; ⁴¹Fundació Institut Universitari per a la recerca i l'innovació, Barcelona; ⁴²Idiixa Santa Ana Hospital Universitario de Navarra, Navarra.

EUROPEAN OHDSI Symposium 19-20 April 2026
Barcelona





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?



Oct. 6: FALCON Symposium Report

The FALCON (Federated Alliance for Large-scale Cancer Observational Network) Research Network is advancing real-world oncology evidence through privacy-preserving, federated health data analyses.

Last week, our community gathered in Antwerp, Belgium, for a four-day collaborative Symposium on Bladder Cancer.

Join us for next week's community call to explore key takeaways, research outputs, and next steps from this global collaboration.





**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