Welcome to OHDSI: What We Do, How You Can Make an Impact

OHDSI Community Call

Oct. 27, 2026 • 11 am ET

Upcoming Community Calls

Date	Topic
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
Dec. 8	How Did OHDSI Do This Year?
Dec. 15	Holiday-Themed Fairwell to 2026

Three Stages of The Journey

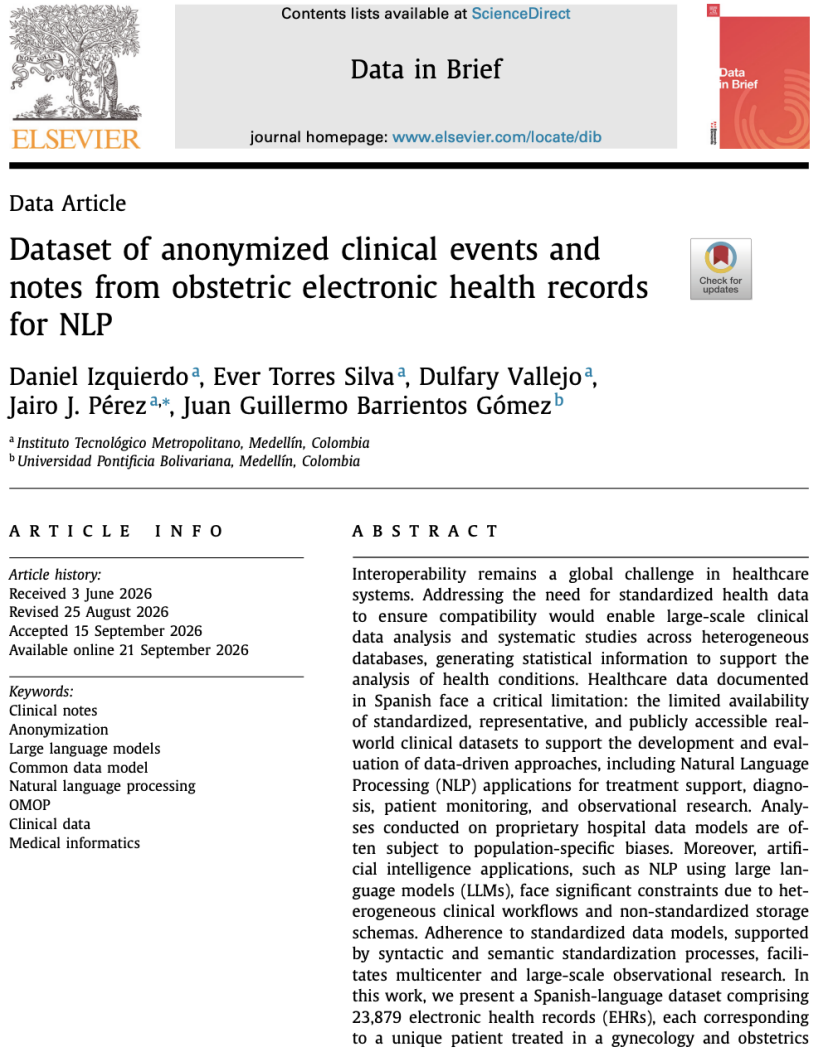
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Daniel Izquierdo, Ever Torres Silva, Dulfary Vallejo, Jairo J Pérez, and Juan Guillermo Barrientos Gómez** on the recent publication of **Dataset of anonymized clinical events and notes from obstetric electronic health records for NLP** in *Data in Brief*.



Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of anonymized clinical events and notes from obstetric electronic health records for NLP

Daniel Izquierdo[a], Ever Torres Silva[a], Dulfary Vallejo[a], Jairo J. Pérez[a,*], Juan Guillermo Barrientos Gómez[b]

[a] *Instituto Tecnológico Metropolitano, Medellín, Colombia*
[b] *Universidad Pontificia Bolivariana, Medellín, Colombia*

ARTICLE INFO

Article history:
Received 3 June 2026
Revised 25 August 2026
Accepted 15 September 2026
Available online 21 September 2026

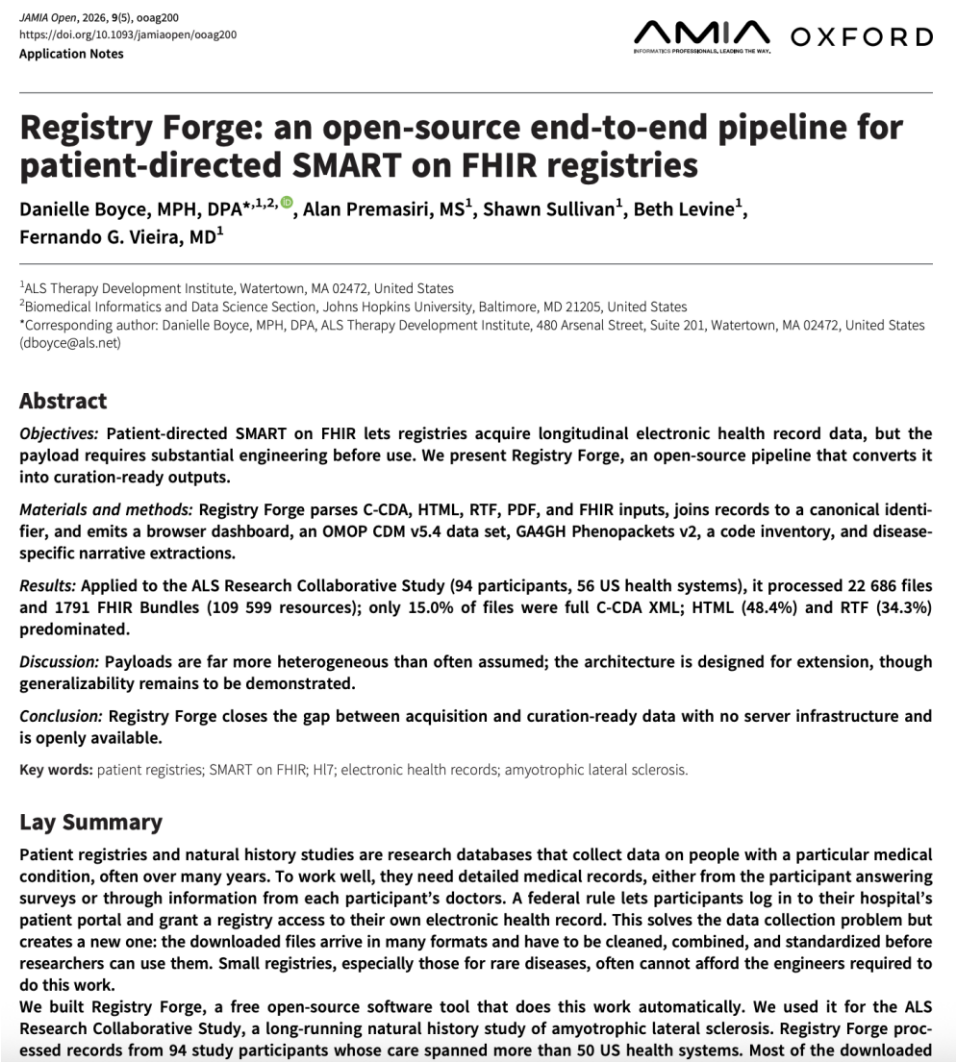
Keywords:
Clinical notes
Anonymization
Large language models
Common data model
Natural language processing
OMOP
Clinical data
Medical informatics

ABSTRACT

Interoperability remains a global challenge in healthcare systems. Addressing the need for standardized health data to ensure compatibility would enable large-scale clinical data analysis and systematic studies across heterogeneous databases, generating statistical information to support the analysis of health conditions. Healthcare data documented in Spanish face a critical limitation: the limited availability of standardized, representative, and publicly accessible real-world clinical datasets to support the development and evaluation of data-driven approaches, including Natural Language Processing (NLP) applications for treatment support, diagnosis, patient monitoring, and observational research. Analyses conducted on proprietary hospital data models are often subject to population-specific biases. Moreover, artificial intelligence applications, such as NLP using large language models (LLMs), face significant constraints due to heterogeneous clinical workflows and non-standardized storage schemas. Adherence to standardized data models, supported by syntactic and semantic standardization processes, facilitates multicenter and large-scale observational research. In this work, we present a Spanish-language dataset comprising 23,879 electronic health records (EHRs), each corresponding to a unique patient treated in a gynecology and obstetrics

OHDSI Shoutouts!

Congratulations to the team of **Danielle Boyce, Alan Premasiri, Shawn Sullivan, Beth Levine, and Fernando G Vieira** on the recent publication of **Registry Forge: an open-source end-to-end pipeline for patient-directed SMART on FHIR registries** in *JAMIA Open*.



JAMIA Open, 2026, 9(5), ooag200
https://doi.org/10.1093/jamiaopen/ooag200
Application Notes

AMIA OXFORD

Registry Forge: an open-source end-to-end pipeline for patient-directed SMART on FHIR registries

Danielle Boyce, MPH, DPA*[,1,2], Alan Premasiri, MS[1], Shawn Sullivan[1], Beth Levine[1], Fernando G. Vieira, MD[1]

[1]ALS Therapy Development Institute, Watertown, MA 02472, United States
[2]Biomedical Informatics and Data Science Section, Johns Hopkins University, Baltimore, MD 21205, United States
*Corresponding author: Danielle Boyce, MPH, DPA, ALS Therapy Development Institute, 480 Arsenal Street, Suite 201, Watertown, MA 02472, United States (dboyce@als.net)

Abstract

Objectives: **Patient-directed SMART on FHIR lets registries acquire longitudinal electronic health record data, but the payload requires substantial engineering before use. We present Registry Forge, an open-source pipeline that converts it into curation-ready outputs.**

Materials and methods: **Registry Forge parses C-CDA, HTML, RTF, PDF, and FHIR inputs, joins records to a canonical identifier, and emits a browser dashboard, an OMOP CDM v5.4 data set, GA4GH Phenopackets v2, a code inventory, and disease-specific narrative extractions.**

Results: **Applied to the ALS Research Collaborative Study (94 participants, 56 US health systems), it processed 22 686 files and 1791 FHIR Bundles (109 599 resources); only 15.0% of files were full C-CDA XML; HTML (48.4%) and RTF (34.3%) predominated.**

Discussion: **Payloads are far more heterogeneous than often assumed; the architecture is designed for extension, though generalizability remains to be demonstrated.**

Conclusion: **Registry Forge closes the gap between acquisition and curation-ready data with no server infrastructure and is openly available.**

Key words: patient registries; SMART on FHIR; HI7; electronic health records; amyotrophic lateral sclerosis.

Lay Summary

Patient registries and natural history studies are research databases that collect data on people with a particular medical condition, often over many years. To work well, they need detailed medical records, either from the participant answering surveys or through information from each participant's doctors. A federal rule lets participants log in to their hospital's patient portal and grant a registry access to their own electronic health record. This solves the data collection problem but creates a new one: the downloaded files arrive in many formats and have to be cleaned, combined, and standardized before researchers can use them. Small registries, especially those for rare diseases, often cannot afford the engineers required to do this work.
We built Registry Forge, a free open-source software tool that does this work automatically. We used it for the ALS Research Collaborative Study, a long-running natural history study of amyotrophic lateral sclerosis. Registry Forge processed records from 94 study participants whose care spanned more than 50 US health systems. Most of the downloaded

OHDSI Shoutouts!

Congratulations to the team of **Annelies Verbiest, Hans Prenen, Annelies Janssens, Christof Vulsteke, Danielle Delombaerde, Laura Tack, Philip R Debruyne, Iege Bassez, Clara L Oeste, Agnes Brandtmüller, Carl Baxter, Pieterjan Debie, and Jan Vertriest** on the recent publication of **Real-world utilization of pembrolizumab dosing schedules in Belgium: a multi-center retrospective study using a common data model and natural language processing** in the *Journal of Comparative Effectiveness Research*.



Research Article

For reprint orders, please contact: info@becaris.com

Real-world utilization of pembrolizumab dosing schedules in Belgium: a multi-center retrospective study using a common data model and natural language processing

Journal of Comparative Effectiveness Research

Annelies Verbiest*[,1], Hans Prenen[1,2], Annelies Janssens[3], Christof Vulsteke[4,5], Danielle Delombaerde[4,5], Laura Tack[6], Philip R Debruyne[6,7,8], Iege Bassez[9], Clara L Oeste[9], Agnes Brandtmüller[10], Carl Baxter[11], Pieterjan Debie[12] & Jan Vertriest[12]

[1]Department of Oncology, Multidisciplinary Oncological Center Antwerp, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium
[2]Center for Oncological Research (CORE), Integrated Personalized & Precision Oncology Network (IPPON), University of Antwerp, Antwerp, Belgium
[3]Department of Thoracic Oncology, Multidisciplinary Oncological Center Antwerp, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium
[4]Integrated Cancer Center Ghent, Maria Middelares General Hospital, Buitenring-Sint-Denijs 30, 9000 Ghent, Belgium
[5]Center for Oncological Research (CORE), University of Antwerp, Campus Drie Eiken T.430, Universiteitsplein 1, 2610 Wilrijk, Belgium
[6]Kortrijk Cancer Centre, AZ Groeninge General Hospital, Kennedylaan 4, 8500 Kortrijk, Belgium
[7]Medical Technology Research Centre (MTRC), School of Life Sciences, Anglia Ruskin University, Cambridge, UK
[8]School of Nursing & Midwifery, University of Plymouth, Plymouth, UK
[9]LynxCare Inc., Leuven, Belgium
[10]MS, MSD, Budapest, Hungary
[11]Center for Observational & Real-World Evidence, MSD (UK) Limited, London, UK
[12]MSD Belgium, Vorstlaan 25 Blvd du Souverain, 1170 Brussels, Belgium
*Author for correspondence: Tel.: +32 3 821 32 50; Annelies.Verbiest@uza.be

Aim: Pembrolizumab's six-week extended dosing (ED), approved as an alternative to the standard three-week schedule (SD), offers flexibility in treatment plans. **Materials & methods:** In this retrospective study across four Belgian hospitals, we described real-world pembrolizumab utilization using a pipeline that processed unstructured data with natural language processing (NLP) and mapped both structured and unstructured data to an OMOP-CDM database. Physicians and data analysts collaboratively developed the project-specific data dictionary and performed validation. **Results:** A total of 303 patients were included, 56% with non-small cell lung cancer (NSCLC). Most patients (72%) received monotherapy, with 89% starting on SD. For NSCLC first-line monotherapy, the median time on treatment (rwTOT) was 106 days in the 'SD only' cohort and 666 days in the 'SD to ED' cohort, while the median time to next treatment (rwTTNT) was 280 days for 'SD only' (median rwTTNT not reached for 'SD to ED'). Dosing cohort

OHDSI Shoutouts!

Congratulations to the team of **Giovanna Nicora, Francesca Negri, Valentina Tibollo, Alberto Malovini, Arianna Dagliati, Lucia Sacchi, Federico Sottotetti, Paola Baiardi, and Riccardo Bellazzi** on the recent publication of **Cross-continental generalizability of EHR foundation model: predicting length of stay in an Italian hospital using US-trained embeddings** in *Frontiers in Digital Health*.

frontiers in Digital Health

TYPE Original Research
PUBLISHED 09 October 2026
DOI 10.3389/fdgth.2026.1956993

Check for updates

OPEN ACCESS

EDITED BY
Selene Tomassini,
University of Trento, Italy

REVIEWED BY
Dohyoung Rim,
Rowan, Republic of Korea
Su Xian,
University of Washington, United States

*CORRESPONDENCE
Giovanna Nicora
giovanna.nicora@unipv.it

RECEIVED 03 August 2026
REVISED 25 September 2026
ACCEPTED 30 September 2026
PUBLISHED 09 October 2026

CITATION
Nicora G, Negri F, Tibollo V, Malovini A, Dagliati A, Sacchi L, Sottotetti F, Baiardi P and Bellazzi R (2026) Cross-continental generalizability of EHR foundation model: predicting length of stay in an Italian hospital using US-trained embeddings.
Front. Digit. Health 8:1956993.
doi: 10.3389/fdgth.2026.1956993

COPYRIGHT
© 2026 Nicora, Negri, Tibollo, Malovini, Dagliati, Sacchi, Sottotetti, Baiardi and Bellazzi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.

Cross-continental generalizability of EHR foundation model: predicting length of stay in an Italian hospital using US-trained embeddings

Giovanna Nicora[1]*, Francesca Negri[1], Valentina Tibollo[2], Alberto Malovini[2], Arianna Dagliati[1], Lucia Sacchi[1], Federico Sottotetti[2], Paola Baiardi[2] and Riccardo Bellazzi[1,2]

[1]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy, [2]IRCCS ICS Maugeri Hospital, Pavia, Italy

Background: Clinical Foundations Models (FMs) hold the potential to exhibit generalist medical AI capabilities, and the ability to reduce the need for explicit feature engineering. However, their evaluation often relies on locally-limited datasets, raising questions regarding their generalizability across diverse global healthcare systems, populations, and clinical practices. This study assesses the cross-continental transferability of a US-trained FM, CLMBR-T-base, applied to a large, distinct, Italian hospital cohort to predict Length of Stay (LoS).
Methods: We utilized a retrospective EHR dataset from IRCCS Maugeri, Pavia, Italy, comprising 99,884 admissions for 63,030 unique patients (2012–2025). To ensure interoperability, raw data were harmonized using the OMOP Common Data Model and converted to the Medical Event Data Standard (MEDS). Patient embeddings were generated zero-shot using the pretrained CLMBR-T-Base model. We benchmarked supervised regression models (Gradient Boosting, Multilayer Perceptron) trained on these embeddings against traditional count-based baselines to predict log-transformed LoS. We built different models trained on the complete hospital admissions set and stratified across four clinical macro-categories (Surgical, Rehabilitation, Oncology, Generic). Embedding cohesion was validated using latent space analysis (HDBSCAN clustering).

OHDSI Shoutouts!

Congratulations to the team of **Ying Xiao, Georgina Kennedy, Rebecca L Baker, Kari J Brewer Savannah, Kristy K Brock, Elizabeth L Covington, Peter E Gabriel, Asieh Golozar, Michael Gurley, Chen Hu, Denise Manfredi, Martha M Matuszak, Katherine Mayo, Jennifer Moughan, Elizabeth A O'Meara, Christian Reich, Lois Holloway, and Charles S Mayo** on the recent publication of **Harmonizing Radiotherapy Data Standards: A Comparative Analysis of OMOP, CDISC, FHIR, NAACCR, and OBO Foundry with an O3-Anchored Mapping Framework** in the *International Journal of Radiation Oncology*Biology*Physics.*



International Journal of Radiation Oncology*Biology*Physics

Date: Available online 23 October 2026

In Press, Journal Pre-proof

View accessibility information

Published by

Research article Get rights and content

Physics Contribution

Harmonizing Radiotherapy Data Standards: A Comparative Analysis of OMOP, CDISC, FHIR, NAACCR, and OBO Foundry with an O3-Anchored Mapping Framework

Ying Xiao PhD [1], Georgina Kennedy [2], Rebecca L Baker MS [3], Kari J Brewer Savannah PhD [4], Kristy K Brock PhD [4], Elizabeth L Covington PhD [5], Peter E Gabriel MD [1], Asieh Golozar PhD [6], Michael Gurley [7], Chen Hu PhD [8], Denise Manfredi [9], Martha M Matuszak PhD [5], Katherine Mayo PhD [10], Jennifer Moughan [8,9], Elizabeth A O'Meara [1], Christian Reich MD, PhD [6], Lois Holloway DPhil [11], Charles S Mayo MD [5] Show more

Cite Add to Mendeley Share 10.1016/j.ijrobp.2026.10.018

Access through Columbia University Purchase PDF Other access options

Article preview

Abstract

Purpose

This Special Report analyzes current radiotherapy data standards, determines which radiotherapy data elements should propagate between them, and identifies the barriers to that propagation to improve interoperability and data utilization in clinical care and research.

Methods

We compared five data standards—OMOP, CDISC SDTM, FHIR, NAACCR, and the Operational Ontology for Oncology (O3)—and one ontology framework, the OBO Foundry, for radiotherapy data representation, strengths, limitations, and implementation maturity, and examined harmonization initiatives including NRG's O3-based conversion frameworks, CodeX mCODE, and ESTRO and IAEA efforts. We then scored a 90-item benchmark drawn from radiotherapy operational and trial-reporting requirements, partitioned the radiotherapy items by whether and where they should propagate, and matched each barrier to a remediation route and owning body.

Article preview

Abstract

Introduction

Section snippets

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	7 am	Medical Imaging
Wednesday	10 am	Women of OHDSI
Wednesday	10 am	Common Data Model
Wednesday	11:30 am	Surgery and Perioperative Medicine
Wednesday	12 pm	Latin America Chapter
Thursday	7:30 am	India Community Call
Thursday	9 am	Phenotype Development and Evaluation
Thursday	10 am	GIS - Geographic Information System
Thursday	11 am	GIS-Geographic Information System
Thursday	7 pm	Dentistry
Friday	9 am	Waveform
Friday	10 am	Transplant
Friday	11:30 am	Steering
Monday	10 am	Healthcare Systems Interest Group
Tuesday	9 am	Data2Evidence

Congratulations to our 2026 Titan Winners!

Masha Khitrun

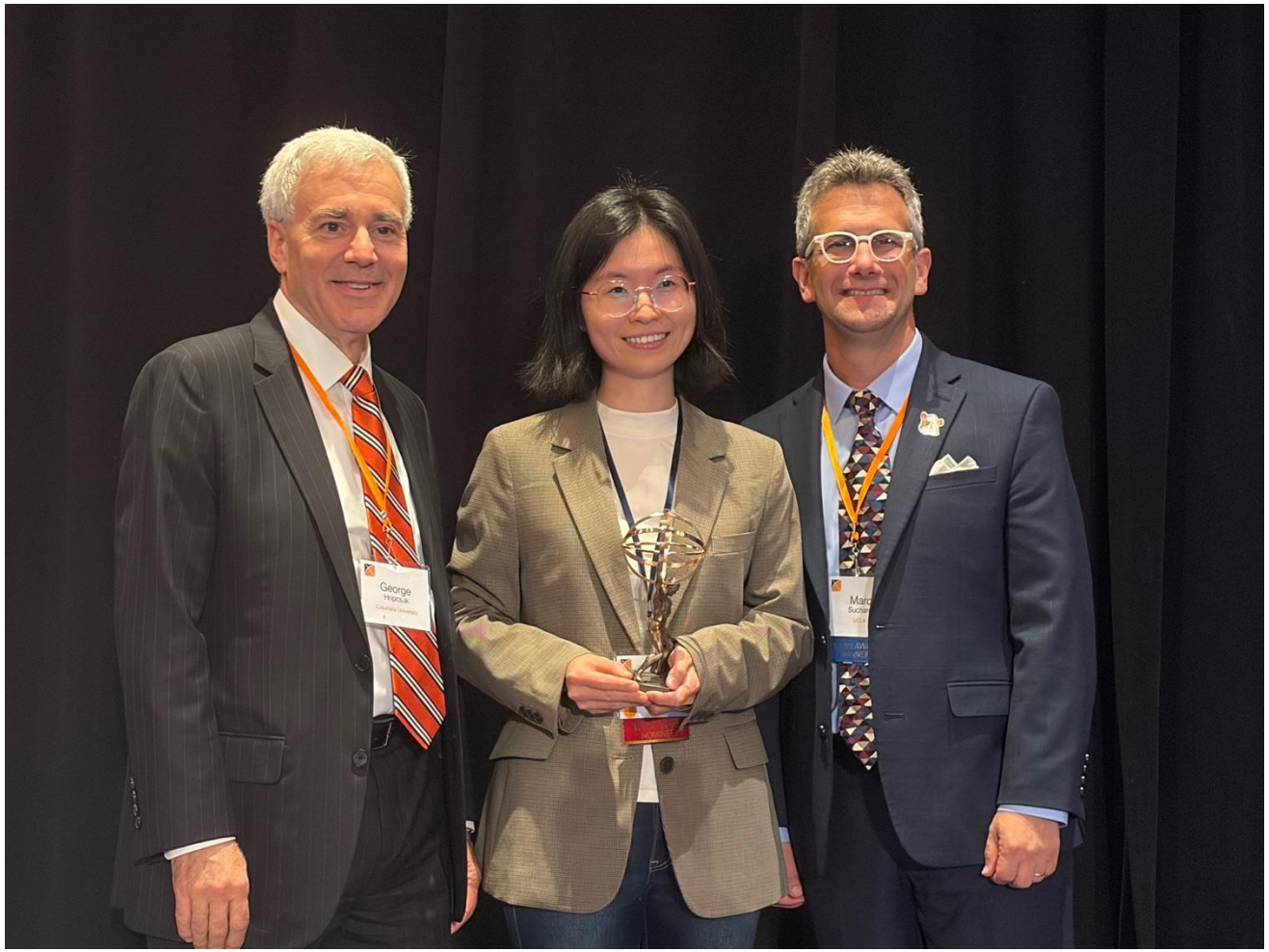
Data Standards



Congratulations to our 2026 Titan Winners!

Bingyu Zhang

Methods Research



Congratulations to our 2026 Titan Winners!

Richard Boyce

Open-Source Analytics



Congratulations to our 2026 Titan Winners!

Eyecare & Vision Research WG

Clinical Applications



Congratulations to our 2026 Titan Winners!

Andrew Kanter

Community Collaboration



Congratulations to our 2026 Titan Winners!

Renske Los

Community Support



Congratulations to our 2026 Titan Winners!

Jason Hsu

Community Leadership



OHDSI 2026 Materials



OHDSI 2026 Photo Sharing

Please share your photos from the Global Symposium!

By submitting photos, you grant permission for OHDSI to use them in upcoming community communications, recaps, and promotional materials.



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.



India Symposium: Dec. 11



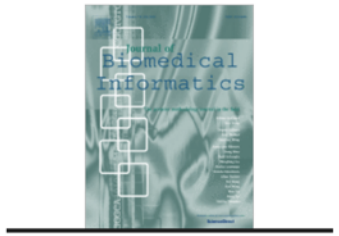
OHDSI INDIA

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.

Call for Papers: Agentic AI for RWE Life Cycle Support

The Journal of Biomedical Informatics has opened a call for papers focused on Agentic AI For RWE Life-Cycle Support — scaling trustworthy evaluation for treatment safety and effectiveness.



Journal of Biomedical Informatics
Date: Available online 8 June 2026
Article: 105062
In Press, Journal Pre-proof

Published by
ELSEVIER

Editorial • Full text access Get rights and content ↗

Editorial

Call for papers: agentic AI for real-world evidence life cycle support – scaling trustworthy evaluation of treatment safety and effectiveness

Show more

View PDF Cite Add to Mendeley Share 10.1016/j.jbi.2026.105062 ↗

Article

Guest editors:

Richard Boyce (Contact) rdb20@pitt.edu; Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA.

Scott Alexander Malec, smalec@salud.unm.edu; Translational Informatics Division, School of Medicine, University of New Mexico, Albuquerque, NM, USA.

Chunhua Weng. cw2384@cumc.columbia.edu; Biomedical Informatics, Data Science Institute, Columbia University, New York, NY, USA.

Outline

Declaration of competing interest

References

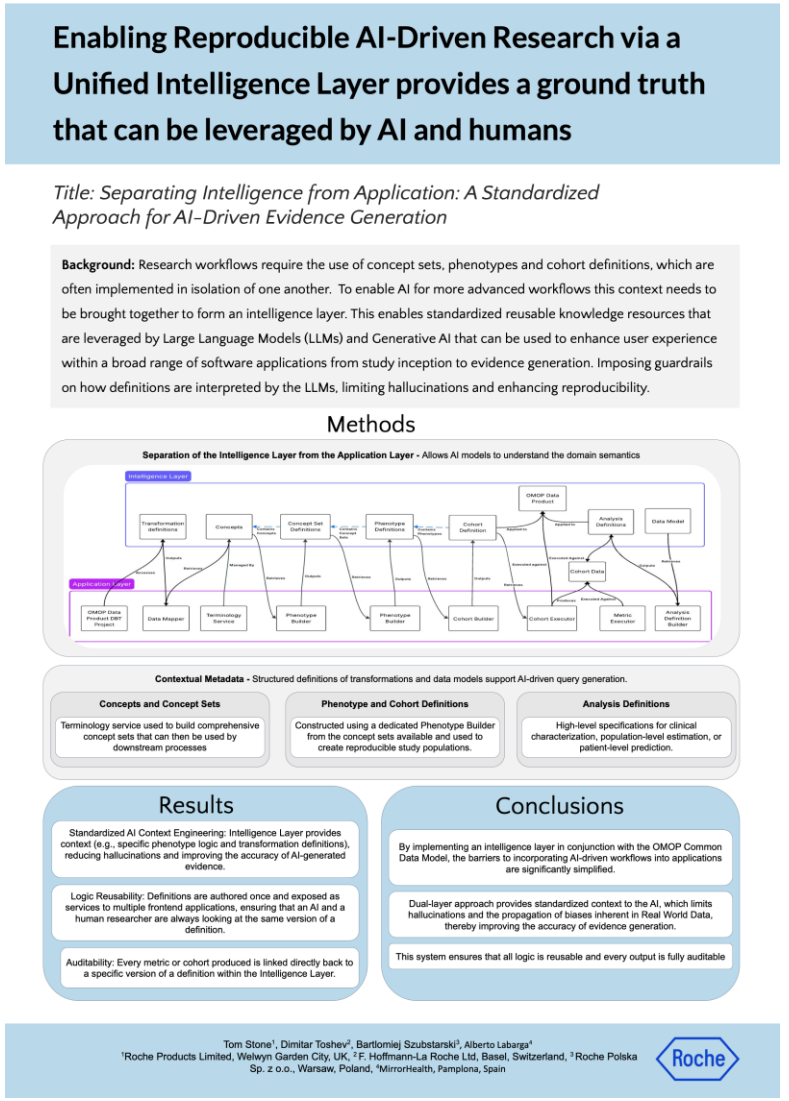
Submissions will be accepted from June 15 through November 15, 2026, with expected publication date of September 15, 2027.

#OHDSISocialShowcase This Week

Monday

Separating Intelligence from Application: A Standardized Approach for AI-Driven Evidence Generation

(Tom Stone, Dimitar Toshev, Bartlomiej Szubstarski, Alberto Labarga)



Enabling Reproducible AI-Driven Research via a Unified Intelligence Layer provides a ground truth that can be leveraged by AI and humans

Title: Separating Intelligence from Application: A Standardized Approach for AI-Driven Evidence Generation

Background: Research workflows require the use of concept sets, phenotypes and cohort definitions, which are often implemented in isolation of one another. To enable AI for more advanced workflows this context needs to be brought together to form an intelligence layer. This enables standardized reusable knowledge resources that are leveraged by Large Language Models (LLMs) and Generative AI that can be used to enhance user experience within a broad range of software applications from study inception to evidence generation. Imposing guardrails on how definitions are interpreted by the LLMs, limiting hallucinations and enhancing reproducibility.

Methods

Separation of the Intelligence Layer from the Application Layer - Allows AI models to understand the domain semantics

Contextual Metadata - Structured definitions of transformations and data models support AI-driven query generation.

Concepts and Concept Sets

Terminology service used to build comprehensive concept sets that can then be used by downstream processes

Phenotype and Cohort Definitions

Constructed using a dedicated Phenotype Builder from the concept sets available and used to create reproducible study populations.

Analysis Definitions

High-level specifications for clinical characterization, population-level estimation, or patient-level prediction.

Results

Standardized AI Context Engineering: Intelligence Layer provides context (e.g., specific phenotype logic and transformation definitions), reducing hallucinations and improving the accuracy of AI-generated evidence.

Logic Reusability: Definitions are authored once and exposed as services to multiple frontend applications, ensuring that an AI and a human researcher are always looking at the same version of a definition.

Auditability: Every metric or cohort produced is linked directly back to a specific version of a definition within the Intelligence Layer.

Conclusions

By implementing an intelligence layer in conjunction with the OMOP Common Data Model, the barriers to incorporating AI-driven workflows into applications are significantly simplified.

Dual-layer approach provides standardized context to the AI, which limits hallucinations and the propagation of biases inherent in Real World Data, thereby improving the accuracy of evidence generation.

This system ensures that all logic is reusable and every output is fully auditable

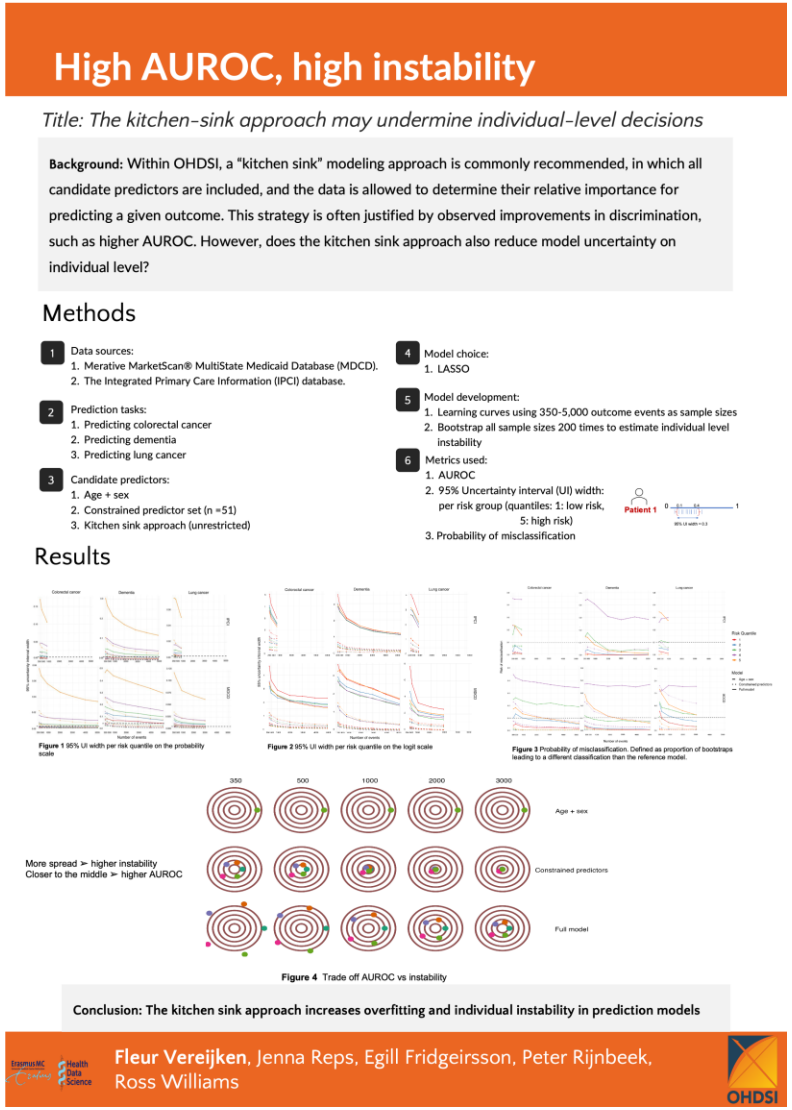
Tom Stone[1], Dimitar Toshev[2], Bartlomiej Szubstarski[3], Alberto Labarga[4]
[1]Roche Products Limited, Welwyn Garden City, UK, [2]F. Hoffmann-La Roche Ltd, Basel, Switzerland, [3]Roche Polska Sp. z o.o., Warsaw, Poland, [4]MirrorHealth, Pamplona, Spain

#OHDSISocialShowcase This Week

Tuesday

High AUROC, Low stability: the kitchen-sink approach may undermine individual-level decisions

(Fleur Vereijken, Jenna Reps, Egill Fridgeirsson, Peter Rijnbeek, Ross Williams)

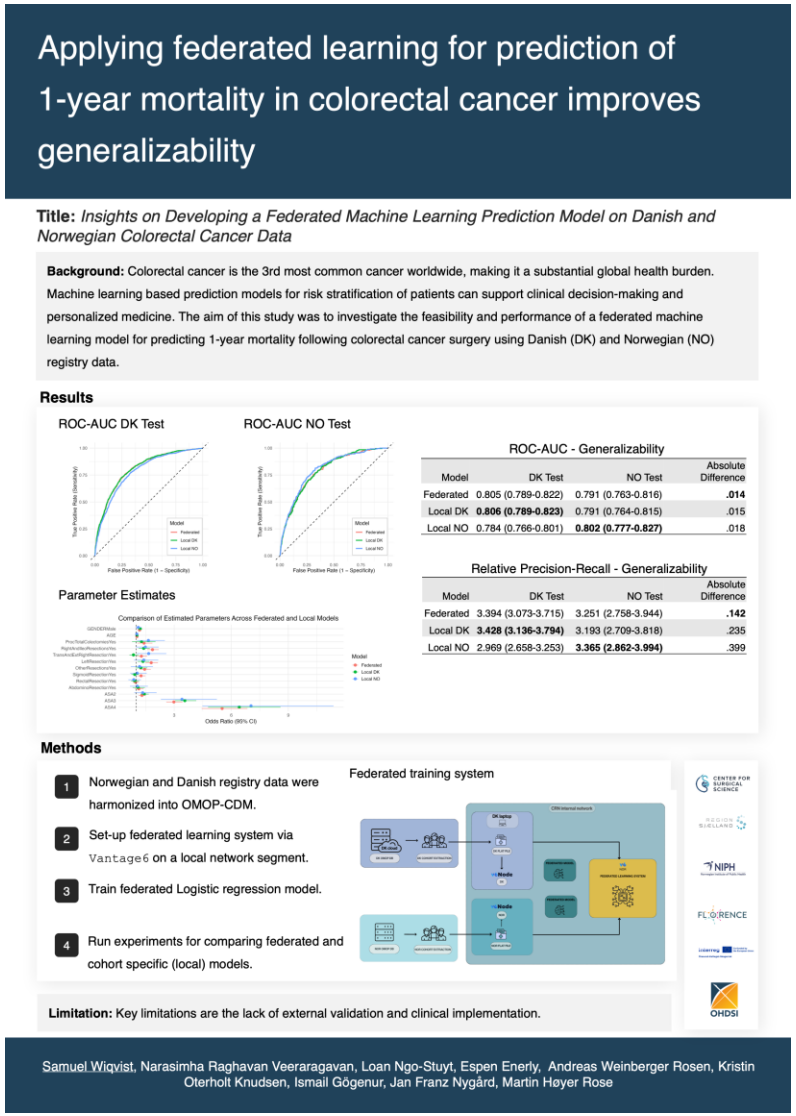


#OHDSISocialShowcase This Week

Wednesday

Insights on Developing a Federated Machine Learning Prediction Model on Danish and Norwegian Colorectal Cancer Data

(Samuel Wiqvist, Narasimha Raghavan Veeraragavan, Loan Ngo-Stuyt, Espen Enerly, Andreas Weinberger Rosen, Kristin Oterholt Knudsen, Ismail Gögenur, Jan Franz Nygård, Martin Høyer Rose)



#OHDSISocialShowcase This Week

Thursday

Sex and Race Differences in Type 2 Diabetes Drug Initiation After Metformin

(Michelle Jin, Clair Blacketer, Fan Bu, Talita Duarte-Salles, Wallis CY Lau, Kenneth KC Man, Evan Minty, Anna Ostropolets, Mui Van Zandt, Can Yin, George Hripcsak, Marc A Suchard)

Sex and Race Differences in Type 2 Diabetes Drug Initiation After Metformin

Michelle Jin[1], Clair Blacketer[2,3], Fan Bu[4], Talita Duarte-Salles[3,5], Wallis CY Lau[6], Kenneth KC Man[6], Evan Minty[7,9], Anna Ostropolets[2], Mui Van Zandt[9], Can Yin[9], George Hripcsak[10,11], Marc A Suchard[12,13]

[1]Computational Biology, University of California, Los Angeles, [2]Johnson and Johnson, Raritan, NJ, [3]Department of Medical Informatics, Erasmus University Medical Center, [4]Department of Biostatistics, University of Michigan, [5]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, [6]Research Department of Practice and Policy, UCL School of Pharmacy, [7]O'Brien Institute for Public Health, University of Calgary, [8]Department of Biomedical Informatics and Data Science, John Hopkins University, [9]Real World Solutions, IQVIA, [10]Department of Biomedical Informatics, Columbia University Medical Center, [11]New York-Presbyterian Hospital, [12]Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, [13]VA Informatics and Computing Infrastructure, US Department of Veterans Affairs

Digital Materials

Background

- Treatment escalation after metformin in type 2 diabetes varies across patient demographics and healthcare settings.
- Evidence on sex and race differences is inconsistent: some studies report lower use of newer cardioprotective agents (GLP1RAs and SGLT2Is) among women and racial minorities, while others find minimal or no differences.
- Patients with established cardiovascular disease may also initiate cardioprotective therapies at lower rates despite guideline recommendations.
- **Knowledge gap:** Few studies have systematically examined post-metformin treatment initiation across multiple drug classes, demographic groups, and healthcare settings using a common framework.

Study objective: Using multinational real-world data from the **LEGEND-T2DM** network, we characterized initiation of four major post-metformin drug classes across **sex, race, and clinical subgroups** and healthcare settings.

Methods

STUDY DESIGN & DATA SOURCES

Federated retrospective study within LEGEND-T2DM, using 10 multinational databases standardized to the OMOP CDMI v5.3.1

- **5 administrative claims + 5 electronic health record databases**

STUDY POPULATION & THERAPIES

Study Population: Adults ≥18 years with T2DM initiating a post-metformin AHA, with:

- ≥365 days prior observation
- No prior use of study post-metformin therapies
- ≤30 days of prior insulin use

Therapies: Four major drug classes

1. DPP4 inhibitors (DPP4Is)
2. GLP1 receptor agonists (GLP1RAs)
3. SGLT2 inhibitors (SGLT2Is
4. Sulfonylureas (SUs)

ANALYTICAL APPROACH

Subgroup Analyses: Initiation patterns were evaluated by **sex, race** (Black vs. all, where available), and **cardiovascular risk** at three levels:

1. Pooled across databases
2. Within individual databases
3. By individual drug

Temporal Trends: Drug-class initiation was assessed from 2010–2021 **by sex** using:

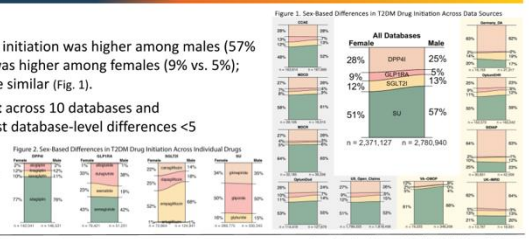
- **Absolute incidence rates** (new users per 1,000 person-years)
- **Proportional initiation** (share of total post-metformin initiations)
- **Log-linear regression** to estimate average annual percent change (AAPC)

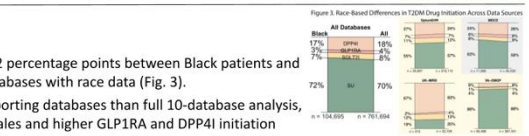
Contact: Michelle Jin | University of California, Los Angeles | mjin2027@g.ucla.edu

Results

1. **Sex Differences**

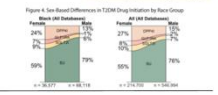
- **Modest class-level differences:** SU initiation was higher among males (57% vs. 51%), while GLP1RA initiation was higher among females (9% vs. 5%); DPP4I and SGLT2I proportions were similar (Fig. 1).
- **Patterns were generally consistent** across 10 databases and cardiovascular-risk strata, with most database-level differences <5 percentage points.
- **Greater individual-drug variation,** particularly for SGLT2Is and GLP1RAs (Fig. 2).





2. **Race Differences**

- **Minimal class-level differences**: ≤2 percentage points between Black patients and the overall population across 4 databases with race data (Fig. 3).
- **Greater sex differences** in race-reporting databases than full 10-database analysis, with higher SU initiation among males and higher GLP1RA and DPP4I initiation among females (Fig 4).
- **Drug-level variation** was greater for SGLT2Is and SUs, although interpretation is limited by incomplete and heterogeneous race capture.





3. **Temporal Trends, 2010–2021**

- **Absolute initiation increase**d across all classes, with fastest growth in GLP1RAs and SGLT2Is (Fig. 5).
- **Treatment composition shifted** toward GLP1RAs and SGLT2Is, while DPP4I and SU shares declined (Fig. 6).
- **Sex-specific trajectories diverged**: SGLT2I initiation increased faster among males, while GLP1RA initiation increased faster among females.

*** Data source type:** Gray = claims; Yellow = EHR

Conclusions

Overall demographic differences were modest, with relatively similar post-metformin treatment initiation patterns across the multinational population.

- **Sex-based differences were modest but consistent** across databases, with more pronounced differences within the subset of databases with available race data.
- **Limited race-based differences**, although interpretation is constrained by incomplete race capture across databases.
- **Treatment patterns evolved over time**, with increasing uptake of newer cardioprotective therapies and diverging sex-specific trajectories for GLP1RAs and SGLT2Is.

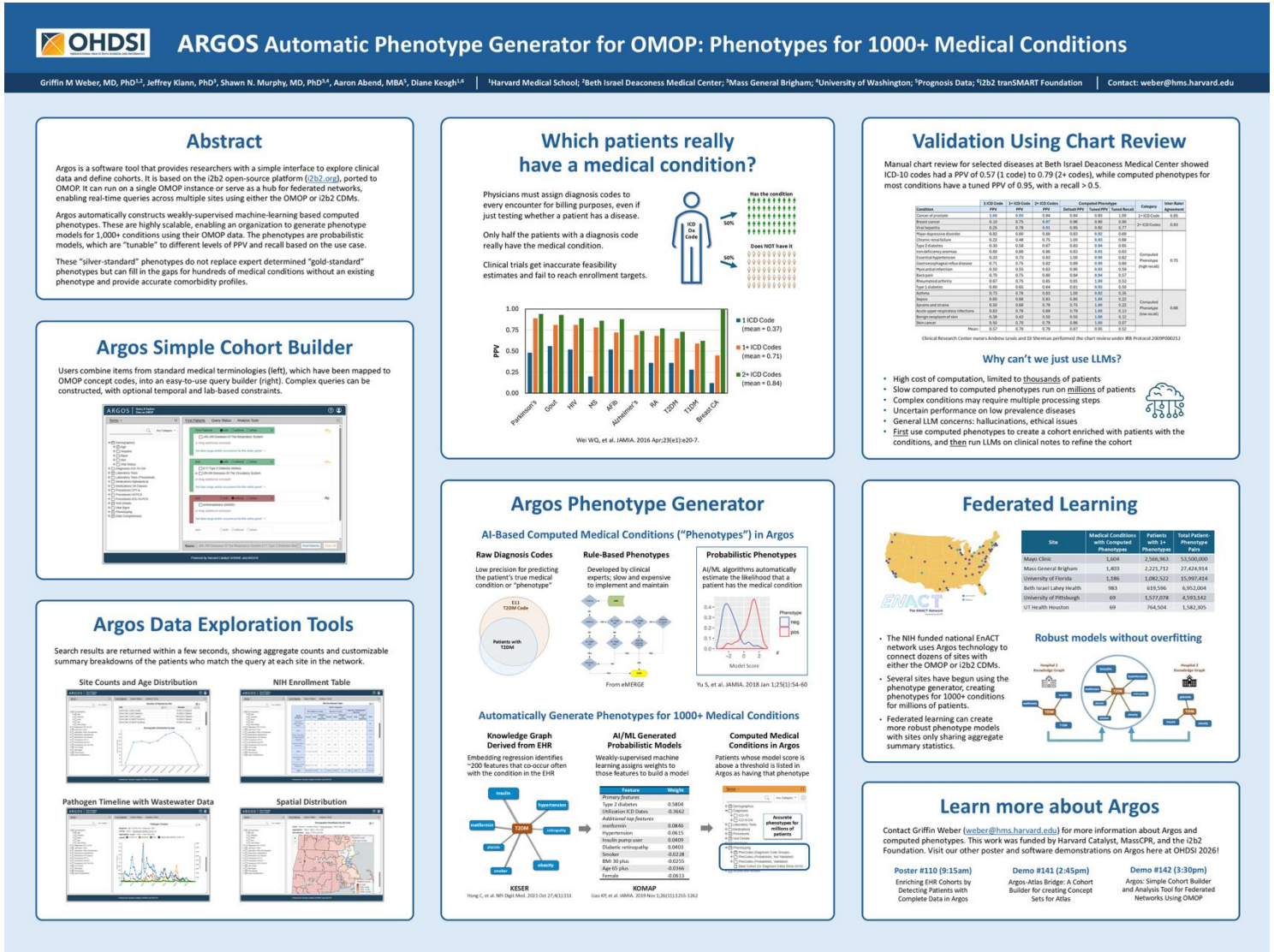
Modest Variation across databases and subgroups may reflect differences in populations, data sources, and healthcare systems. Findings should be interpreted cautiously given the descriptive analysis and incomplete race capture, highlighting the value of federated OMOP networks and the need for richer data to understand treatment selection.

#OHDSISocialShowcase This Week

Friday

Argos Automatic Phenotype Generator for OMOP: Phenotypes for 1000+ Medical Conditions

(Griffin M Weber, Jeffrey Klann, Shawn N. Murphy, Aaron Abend, Diane Keogh)



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?



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