2026 Global Symposium Rising Star Presentations

OHDSI Community Call

Nov. 17, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Upcoming Community Calls

Date	Topic
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 Farewell to 2026
Dec. 22	NO CALL – Holiday Break

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Severin Kohler, Falk Meyer-Eschenbach, Leon Schüßler, Maximilian Meixner, Florian Kärcher, Andreas Kling, Vivien Schubert, Dirk Meyer Zum Büschenfelde, Michael Marschollek, and Roland Eils** on the recent publication of **An interoperable and reusable research environment for digital medicine** in *Frontiers in Digital Medicine*.

frontiers in Digital Health

TYPE Original Research
PUBLISHED 28 October 2026
DOI 10.3389/fdgth.2026.1943167

Check for updates

OPEN ACCESS

EDITED BY
Shameer Khader,
Sanofi, France

REVIEWED BY
Shuxin Zhang,
Amsterdam University Medical Center, Netherlands
Afra Shamnath,
Columbia University, United States
Isaac Cano,
Hospital Clinic of Barcelona, Spain

*CORRESPONDENCE
Severin Kohler
severin.kohler@bih-charite.de

RECEIVED 20 July 2026
REVISED 13 September 2026
ACCEPTED 06 October 2026
PUBLISHED 28 October 2026

CITATION
Kohler S, Meyer-Eschenbach F, Schüßler L, Meixner M, Kärcher F, Kling A, Schubert V, Meyer zum Büschenfelde D, Marschollek M and Eils R (2026) An interoperable and reusable research environment for digital medicine. Front. Digit. Health 8:1943167. doi: 10.3389/fdgth.2026.1943167

COPYRIGHT
© 2026 Kohler, Meyer-Eschenbach, Schüßler, Meixner, Kärcher, Kling, Schubert, Meyer zum Büschenfelde, Marschollek and Eils. 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.

An interoperable and reusable research environment for digital medicine

Severin Kohler[1,2,3,4]*, Falk Meyer-Eschenbach[5,6], Leon Schüßler[1,7], Maximilian Meixner[1], Florian Kärcher[3], Andreas Kling[3], Vivien Schubert[1], Dirk Meyer zum Büschenfelde[1], Michael Marschollek[4] and Roland Eils[1,3,8]

[1]Center of Digital Health, Berlin Institute of Health at Charité—Universitätsmedizin Berlin, Berlin, Germany, [2]Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany, [3]Health Data Science Unit, Heidelberg University Hospital and BioQuant, Heidelberg, Germany, [4]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany, [5]Institute of Medical Informatics, Charité—Universitätsmedizin Berlin, Berlin, Germany, [6]Clinical Study Center, Berlin Institute of Health at Charité—Universitätsmedizin Berlin, Berlin, Germany, [7]Hochschule für Technik und Wirtschaft Berlin University of Applied Sciences, Berlin, Germany, [8]Intelligent Medicine Institute, Fudan University, Shanghai, China

Clinical research infrastructures remain fragmented and project-specific, with semantic inconsistency accumulating across studies and constantly increasing "interoperability debt". Repeated transformations between heterogeneous standards cause semantic drift and information loss. Prospective studies require duplicate data entry across clinical and research systems, increasing workload and introducing inconsistencies. Project-driven data models accumulate technical debt that undermines the sustainability and maintainability of research infrastructure across studies and funding cycles. Here, we explore how a reusable research environment can be built, and whether governing data semantics at the persistence layer provides a foundation that outlasts individual projects. We present an environment addressing these deficiencies at the architectural level. An openEHR-based clinical data repository with governed archetypes serves as the canonical persistence layer, integrating retrospective EHR data, prospective study data, patient-reported outcomes, and wearable sensor streams under a shared semantic model. Prospective data is captured directly through openEHR-based forms, which removes the separate capture system for study variables, while duplicate documentation with a hospital information system that does not run on openEHR remains. Reusable transformation pipelines support bidirectional FHIR interoperability and OMOP-based analytics. Federation

OHDSI Shoutouts!

Congratulations to the team of Cecilia Campanile, Alex E. Knight, Hiba Junaid, Jennifer C. E. Lane, Usama Rahman, Ben Eaton, Xavier L. Griffin, Andy South, Ana Cavalcante, Stefan Piatek, Baptiste Briot Ribeyre, Steve Harris, Timothy Howcroft, Vishnu Chandrabalan, Peter S. Hall, Mahéva Vallet, Marta Pasikowska, Nicola Symmers, Colin McLean, Spyro Nita, Daniel Dedman, Zara Cuccu, Stelios Theophanous, Geoff Hall, Edward Bolton, Shiren Patel, Daniel Key, Lydia Briggs, Prasiddha Khadka, David Seymour, Elin Rowlands, Danielle Newby, Daniel Prieto-Alhambra, Edward Burn, and Martí Català on the recent publication of **HEalth Data Research OMOP Network (HERON-UK), A Real-World Evidence Federated Network of Hospital and Primary Care Electronic Health Records Including 57 Million Individuals and 7 Data Partners in the United Kingdom** in *Pharmacoepidemiology and Drug Safety*.

Pharmacoepidemiology and Drug Safety

WILEY

REAL-WORLD DATA SOURCES FOR PHARMACOEPIDEMIOLOGIC RESEARCH

OPEN ACCESS

HEalth Data Research OMOP Network (HERON-UK), A Real-World Evidence Federated Network of Hospital and Primary Care Electronic Health Records Including 57 Million Individuals and 7 Data Partners in the United Kingdom

Cecilia Campanile[1] | Alex E. Knight[2] | Hiba Junaid[3,4,5] | Jennifer C. E. Lane[3,4] | Usama Rahman[3,4] | Ben Eaton[3,4] | Xavier L. Griffin[3,4] | Andy South[6] | Ana Cavalcante[6] | Stefan Piatek[6] | Baptiste Briot Ribeyre[6] | Steve Harris[6,7] | Timothy Howcroft[8] | Vishnu Chandrabalan[8] | Peter S. Hall[9] | Mahéva Vallet[9] | Marta Pasikowska[9] | Nicola Symmers[9] | Colin McLean[9] | Spyro Nita[9] | Daniel Dedman[10] | Zara Cuccu[10] | Stelios Theophanous[11] | Geoff Hall[11] | Edward Bolton[11] | Shiren Patel[12] | Daniel Key[12] | Lydia Briggs[12] | Prasiddha Khadka[12] | David Seymour[2] | Elin Rowlands[1] | Danielle Newby[1] | Daniel Prieto-Alhambra[1,13] | Edward Burn[1] | Martí Català[1]

[1]Health Data Sciences, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK | [2]Health Data Research UK (HDR UK), London, UK | [3]Barts Bone and Joint Health, Queen Mary University of London, London, UK | [4]Barts Health NHS Trust, NHS Trust, London, UK | [5]Barts Life Sciences, London, UK | [6]University College London Hospitals Biomedical Research Centre, London, UK | [7]Institute of Health Informatics, University College London, London, UK | [8]Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK | [9]DataLoch, Edinburgh Cancer Informatics, University of Edinburgh, Edinburgh, UK | [10]Clinical Practice Research Datalink (CPRD), Medicines and Healthcare Products Regulatory Agency (MHRA), London, UK | [11]Leeds Teaching Hospitals NHS Trust, London, UK | [12]Data Research, Innovation and Virtual Environments (DRIVE), Great Ormond Street Hospital for Children, London, UK | [13]Department of Medical Informatics, Erasmus MC University, Rotterdam, the Netherlands

Correspondence: Daniel Prieto-Alhambra (daniel.prietoalhambra@ndorms.ox.ac.uk)

Received: 21 April 2026 | **Revised:** 26 October 2026 | **Accepted:** 3 November 2026

ABSTRACT

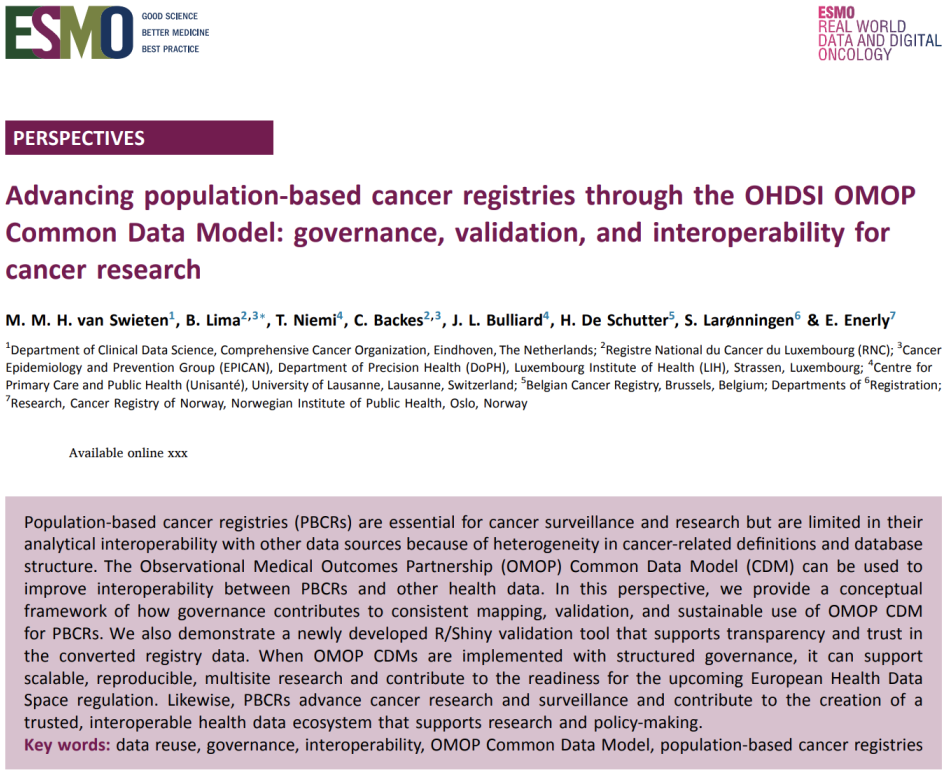
Introduction: The increasing availability of routinely collected healthcare data presents opportunities to generate real-world evidence (RWE) to inform regulatory, clinical, and policy decisions. However, fragmented data sources, heterogeneous data structures, and inconsistent coding systems hinder reliable, large-scale analyses. Federated data networks adopting a common data model, such as the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), offer a standardised, privacy-preserving approach to addressing these challenges.

Methods: We established a federated data network comprising seven UK-based partners, each maintaining routinely collected healthcare data mapped to the OMOP CDM. A central coordinating centre develops and distributes standardised analytical code for local execution within each partner's secure environment. Only aggregated, non-identifiable results are returned to the coordinating centre, enabling reproducible analyses while preserving data privacy. Partner c followed a three-stage process comprising database description, benchmarking, and comprehensive data characterisation to assess data quality, support study feasibility, and improve understanding of the available data.

Results: The network comprises five English National Health Service (NHS) trusts, one primary care data provider, and one Scottish NHS trust. Collectively, the participating databases contain records representing approximately 57.2 million individuals,

OHDSI Shoutouts!

Congratulations to the team of **M M H van Swieten, B Lima, T Niemi, C Backes, J L Bulliard, H De Schutter, S Larønningen, and E Enerly** on the recent publication of **Advancing population-based cancer registries through the OHDSI OMOP Common Data Model: governance, validation, and interoperability for cancer research** in *ESMO Real World Data and Digital Oncology*.



ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO REAL WORLD DATA AND DIGITAL ONCOLOGY

PERSPECTIVES

Advancing population-based cancer registries through the OHDSI OMOP Common Data Model: governance, validation, and interoperability for cancer research

M. M. H. van Swieten[1], B. Lima[2,3*], T. Niemi[4], C. Backes[2,3], J. L. Bulliard[4], H. De Schutter[5], S. Larønningen[6] & E. Enerly[7]

[1]Department of Clinical Data Science, Comprehensive Cancer Organization, Eindhoven, The Netherlands; [2]Registre National du Cancer du Luxembourg (RNC); [3]Cancer Epidemiology and Prevention Group (EPICAN), Department of Precision Health (DoPH), Luxembourg Institute of Health (LIH), Strassen, Luxembourg; [4]Centre for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland; [5]Belgian Cancer Registry, Brussels, Belgium; Departments of [6]Registration; [7]Research, Cancer Registry of Norway, Norwegian Institute of Public Health, Oslo, Norway

Available online xxx

Population-based cancer registries (PBCRs) are essential for cancer surveillance and research but are limited in their analytical interoperability with other data sources because of heterogeneity in cancer-related definitions and database structure. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) can be used to improve interoperability between PBCRs and other health data. In this perspective, we provide a conceptual framework of how governance contributes to consistent mapping, validation, and sustainable use of OMOP CDM for PBCRs. We also demonstrate a newly developed R/Shiny validation tool that supports transparency and trust in the converted registry data. When OMOP CDMs are implemented with structured governance, it can support scalable, reproducible, multisite research and contribute to the readiness for the upcoming European Health Data Space regulation. Likewise, PBCRs advance cancer research and surveillance and contribute to the creation of a trusted, interoperable health data ecosystem that supports research and policy-making.

Key words: data reuse, governance, interoperability, OMOP Common Data Model, population-based cancer registries

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls



Date	Time (ET)	Meeting
Tuesday	12 pm	Vocabulary
Tuesday	4 pm	Electronic Animal Health Records
Thursday	4 am	Waveform
Thursday	9 am	Africa Chapter
Thursday	10 am	GIS – Geographic Information System
Thursday	10:30 am	Evidence Network
Thursday	12 pm	HADES
Friday	10:30 am	Education
Friday	11:30 am	Steering
Tuesday	9 am	Oncology Genomic Subgroup
Tuesday	9 am	OpenEHR & OMOP
Tuesday	10 am	CDM Survey

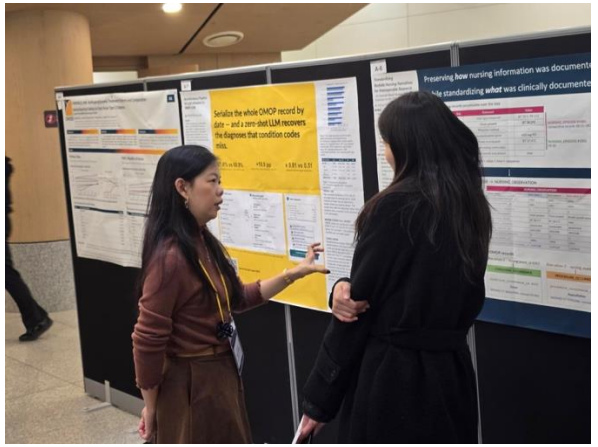
APAC Symposium













ohdsi.org/APAC2026

OHDSI/Medical AI Survey



Jiwon Um developed a brief survey focusing on OHDSI in the Era of Medical AI, and she presented preliminary results highlighting a diverse range of perspectives during the recent APAC Symposium.

A key recommendation emerging from the preliminary results is the need to establish a clear community roadmap for how and where AI can be responsibly applied across the journey from data to evidence, with particular emphasis on developing best practices for LLM-assisted data standardization.

If you have not yet participated, please take approximately 10 minutes to complete the survey in the next two weeks before the end of November.



Canada Symposium: Nov. 17-18

The 2026 OHDSI Canada Symposim begins today 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

The **OHDSI India Symposium** will be held in Dec. 11 in Chennai. This 1-day, hands-on symposium will focus on:

- Standardizing health data with the **OMOP CDM**
- Leveraging **real-world data** for research and better patient outcomes
- Interactive discussions led by **experts and practitioners**
- **Networking** with clinicians, researchers, data scientists, and healthcare innovators

Capacity is limited to 60 seats only. Register your interest today and be part of India's growing health data ecosystem.



#OHDSISocialShowcase This Week

Monday

Mind Meets Machines?: AI Assisted Semantic Mapping of Clinical Concepts

(Andrew S. Kanter, Jonathan Payne, Filipe Lopes, June Bronnert, Theresa Rihanek)

Mind Meets Machines?: AI Assisted Semantic Mapping of Clinical Concepts

PRESENTER: **Andrew S. Kanter**

Andrew S. Kanter[1], **Jonathan Payne**[2], **Filipe Lopes**[3], **June Bronnert**[4], **Theresa Rihanek**[4]
[1]Dept. of Biomedical Informatics, Columbia University; [2]Open Concept Lab; [3]Orango I/O (Health); [4]IMO Health

INTRO

Mapping local terms to standard codes is the slow, skilled prerequisite of every OMOP ETL, and trained terminologists are scarce.

Can AI map to reference codes reliably, and does a curated interface terminology help?

We tested mapping to ICD-11: heavily post-coordinated, the WHO standard for health statistics since 2022, and increasingly the target for national EHR dictionaries (e.g. Uganda, Malawi). CIEL interface terminology is in Athena, so its maps also reach standard OMOP concepts, improving the quality of real-world data for federated analytics.

METHODS

1. 100 high-frequency clinical terms, drawn at random from IMO Health's concept library.
2. Two authors mapped each to ICD-11 (MMS); a third resolved disagreements. 68 of the 100 selected concepts required post-coordination.
3. OCL Mapper pooled candidates from four algorithms: the curated **CIEL bridge**, **WHO's ICD-11 linearization search** and **autocode** APIs, and an **OCL ICD-11 agent** ("LLM as terminologist").
4. A RAG-based "LLM as judge" AI Assistant (Claude Haiku 4.5) picked a primary code and up to two alternates from the candidate pool, or rejected with rationale.
5. Ablation study: one fixed candidate pool; the AI sees only the candidates from each configuration's algorithms.
6. **Strict** = exact code or exact post-coordinated codes. **Loose** = clinically acceptable, judged by the mappers.

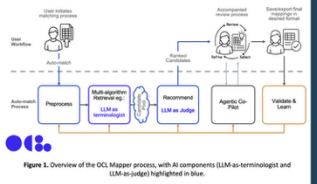
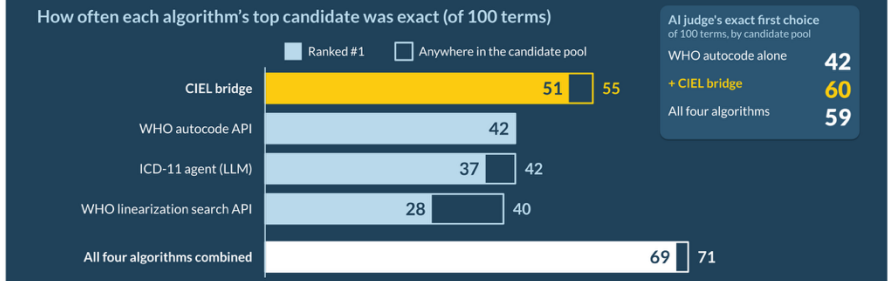


Figure 1. Overview of the OCL Mapper process, with AI components (LLM-as-terminologist and LLM-as-judge) highlighted in blue.

With AI as Judge, adding a curated terminology, the CIEL bridge, raised exact ICD-11 matches from 42 to 60 out of 100 complex clinical terms.

Stacking more algorithms on top added no exact matches (all four: 59)

How often each algorithm's top candidate was exact (of 100 terms)

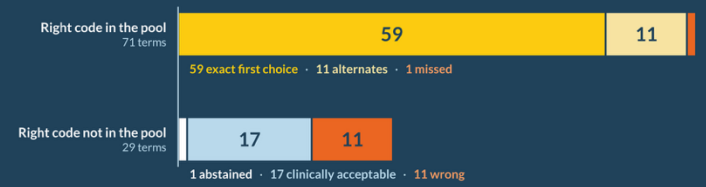


AI judge's exact first choice of 100 terms, by candidate pool

WHO autocode alone	42
+ CIEL bridge	60
All four algorithms	59

NOTE: CIEL alone found 14 correct codes that no other algorithm found. WHO autocode returns only a single code by design. WHO autocode alone: 42 exact. Autocode + CIEL bridge with the AI judge: 60 (right code in its pool for 63). All four algorithms: 59.

What the AI Assistant chose, with and without the right code in its pool



When the right code was available, the AI's primary/alternate selection held it 70 of 71 times (99%). When it was not, the AI still recommended a code 28 of 29 times. "Clinically acceptable" = judged by the mappers; the brief report scored exact matches only.

Scan to try the **OCL Mapper**
openconceptlab.org/tools/mapper

Scan to explore the **CIEL dictionary**
app.v3.openconceptlab.org/orgs/CIEL

OCL Mapper source code is available at: github.com/OpenConceptLab/oclmap

RESULTS

- The correct code was in the candidate pool for **71 of 100** terms. The AI's first choice was exact for **59 of those 71**, and its primary or alternate for **70**.
- The OCL AI Assistant's first choice was clinically acceptable (loose) for **80 of 100**.
- CIEL alone found **14** correct codes that no other algorithm found (Figure 2).
- Abstention depends on the pool: with only the two WHO APIs, the AI abstained 22 times, and **19 (86%)** truly had no correct code. With all four algorithms it abstained once.
- Post-coordination is the frontier: only **35 of 68** exactly right with all four algorithms.
- *Limits: n = 100 from one source; answer key by study authors. A larger multi-source study is in preparation.*



Figure 2. Candidate-pool coverage & unique contribution by algorithm. Green bar = CIEL's 14 rows found by no other algorithm. (UpSet plot, n = 100, strict matching).

DISCUSSION

- Mapping to ICD-11 and OMOP standards can be greatly simplified and improved through the use of AI-enhanced tools.
- Curated interface terminologies matter: the CIEL bridge was the strongest single source and found 14 codes no other algorithm did.
- OCL's AI Assistant picked the right code when it was available (70 of 71) and can flag thin candidate pools for human review.

REFERENCES:

1. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform. 2015;216:574–578. PMID: 26262116.
2. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inform Decis Mak. 2021;21(Suppl 6):206. doi:10.1186/s12911-021-01534-6.
3. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006;13(3):277–288. PMID: 16501181.
4. Observational Health Data Sciences and Informatics. The Book of OHDSI. 2021.

#OHDSISocialShowcase This Week

Tuesday

Accelerating Translational Oncology Research Through a Reproducible OMOP CDM: Lessons Learned from Fred Hutch Cancer Center

(Monica W. Gerber, Megan Stachura, Josh Southam, Sitapriya Moorthi, Amy Paguirigan)

***Accelerating Translational Oncology Research Through a Reproducible OMOP CDM:** Lessons Learned from Fred Hutch Cancer Center*

PRESENTERS: Monica Gerber & Megan Stachura

INTRO

- The Fred Hutch Cancer Center (FHCC) OMOP CDM was built on a new translational data platform (Clinical and Research Data System [CARDS]) utilizing Databricks in order to unify and modernize management of cancer-related care data derived from Fred Hutch's adult oncology program.
- A core goal of the platform is to use standard data models, open source tooling, and reproducible workflows, while being responsive to evolving research use cases.

METHODS

- We began by profiling user needs and calculating the percentage of use cases requiring data from distinct oncology data domains.
- To speed development, we created markdown templates for turning our use case analysis into key domain rules and research-driven requirements.
- We utilized Declarative Automation Bundles (DABs) on Databricks to set up our refresh pipeline, along with data quality monitoring with Data Quality Dashboard and the DQX Python library.
- We completed the FALCON network oncology readiness assessment to compare FHCC data to other cancer centers.

RESULTS

- By fall 2026, we've transformed Epic Clarity EHR data to 13 clinical and health system data tables.

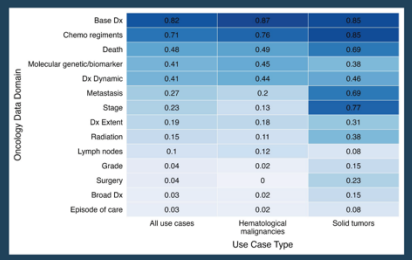


Figure 1. Proportion of use cases requiring data from each oncology data domain, by tumor type (all use cases, n = 76; hematological malignancies, n = 57; solid tumors, n = 15; other, n = 6). Boxes are shaded by proportion of uses cases, with darker colors representing higher proportions. The proportion is displayed within each box.

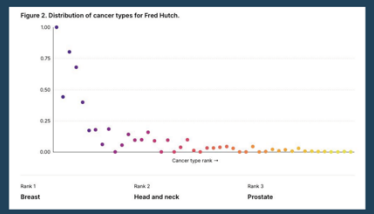


Figure 2. Distribution of cancer types at Fred Hutch. Cancer types are ordered along the x-axis by FALCON network-wide prevalence (most prevalent leftmost).

	Verification				Validation				Total			
	Pass	Fail	Total	Pass %	Pass	Fail	Total	Pass %	Pass	Fail	Total	Pass %
Plausibility	254	7	261	97%	234	8	242	97%	488	15	503	97%
Conformance	635	2	637	100%	66	0	66	100%	701	2	703	100%
Completeness	238	7	245	97%	12	5	17	71%	250	12	262	95%
Total	1,127	16	1,143	99%	312	13	325	96%	1,439	29	1,468	98%

Figure 3. Data Quality Dashboard (DQD) results from the 1,468 total checks applicable to to the FHCC OMOP CDM instance. The 29 failed checks are primarily related to (1) columns we have not yet fully populated, (2) plausible gender for conditions, (3) record completeness for persons, (4) plausibility of date sequences.

Figure 4. Key pillars of our ETL pipeline designed for reproducible, automated workflows to deliver reliable, high-quality data to meet oncology needs.

Declarative Automation Bundle (DAB) Infrastructure as Code (IaC) YAML
- Orchestrate jobs and dependencies
- Workspace-specific variables
- Alerts for job failure and data quality
- Dashboards
- Permission management for secrets

Data Transformation & Update
- Weekly full table overwrite to avoid incremental update complexity and meet user needs
- Deterministic, hash-based IDs for stability across refreshes
- Reusable skills and templates defined

Workspaces & Deployment
- Test and production workspaces
- Version control and branch standards
- GitHub Action for continuous integration and deployment (CI/CD) to validate and deploy bundle-defined resources

Data Quality
- DQX Python library to identify and quarantine records with data quality issues
- Data Quality Dashboard (DQD) for standard OMOP checks, with Databricks dashboard (managed through bundle) to visualize results

Visit our **GitHub repository**
https://github.com/FredHutch/fhcc-omop-template

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

RESULTS (cont'd)

- The FHCC OMOP CDM will need to integrate data from multiple sources:
  - EHR data meets critical data domains such as base diagnosis (82% of use cases) and chemotherapy regimens (71%).
  - Integration of cancer registry is necessary for histology and topography concepts.
  - 41% of use cases requiring genetic and biomarker data will require integration of data from diverse sources and mapping to OMOP genomic vocabulary.
- Team standards and coordination was essential for moving quickly, documenting decision making, and adapting to a new technical stack.
  - Our template GitHub repository is available for other health systems to "fork" and adapt for their own use.
- Our pipeline implementation on Databricks allows for secure, reproducible, automated workflows that meet oncology user needs.
  - DABs provide and infrastructure-as-code approach.
  - Deterministic ID generation was essential for consistency across releases and refreshes.
  - DQX monitoring defined in YAML increases confidence in data quality.

CONCLUSIONS

- Our development framework, building on a use case analysis and transparent documentation, was essential for creating a culture of data excellence in a landscape where data is the critical infrastructure for building high-stakes AI models.

Monica W. Gerber[1], Megan Stachura[1], Josh Southam[1], Sitapriya Moorthi[1], Christian Reich[2], Asieh Golozar[2], Amy Paguirigan[1]
[1] Fred Hutch Cancer Center, Translational Data Accelerator
[2] Nemesis Health

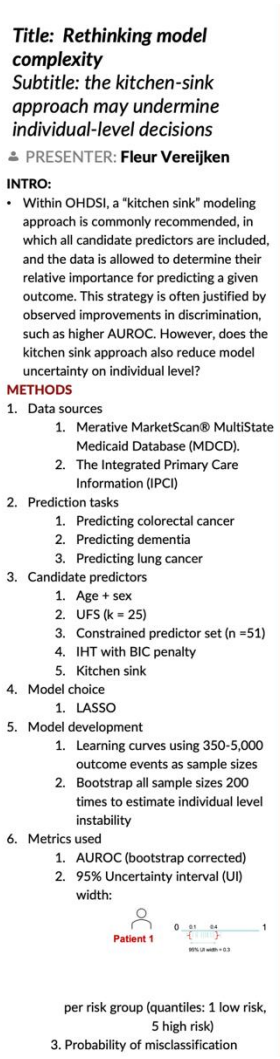
OHDSI
Fred Hutch Cancer Center

#OHDSISocialShowcase This Week

Wednesday

Rethinking Model Complexity in Clinical Prediction Models

(Fleur Vereijken, Jenna M. Reps, Peter Rijnbeek, Ross D. Williams)



Title: Rethinking model complexity
Subtitle: the kitchen-sink approach may undermine individual-level decisions

PRESENTER: **Fleur Vereijken**

INTRO:

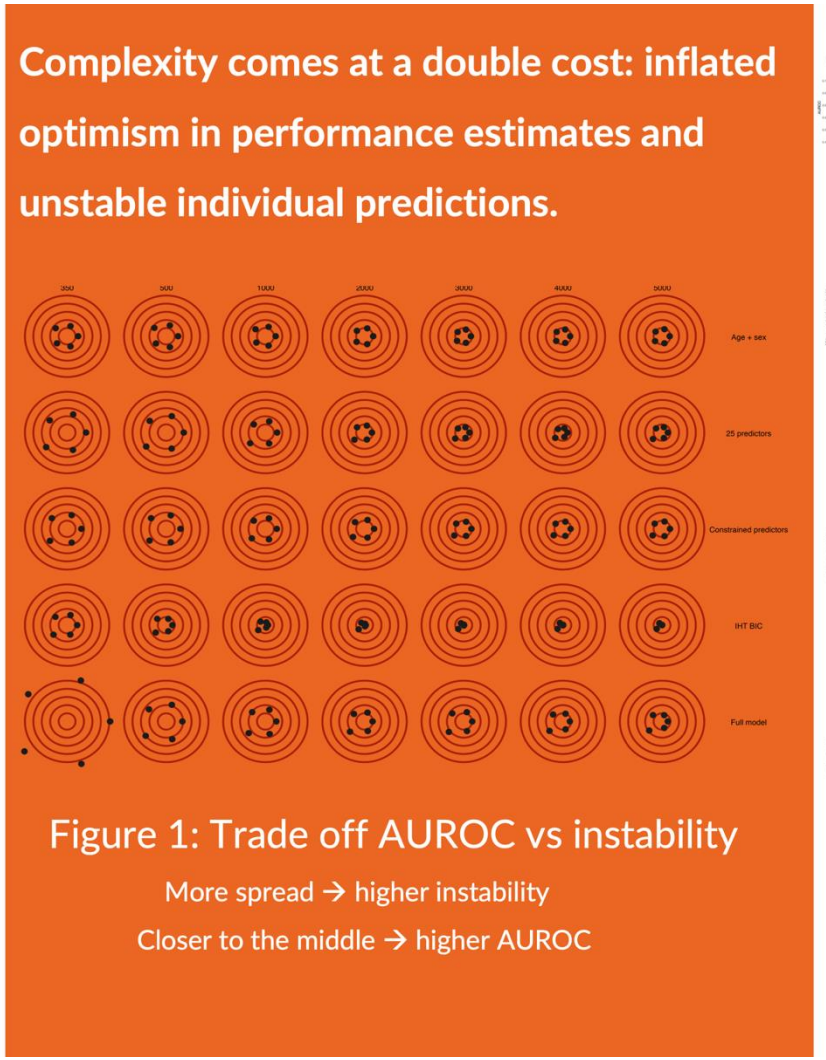
- Within OHDSI, a "kitchen sink" modeling approach is commonly recommended, in which all candidate predictors are included, and the data is allowed to determine their relative importance for predicting a given outcome. This strategy is often justified by observed improvements in discrimination, such as higher AUROC. However, does the kitchen sink approach also reduce model uncertainty on individual level?

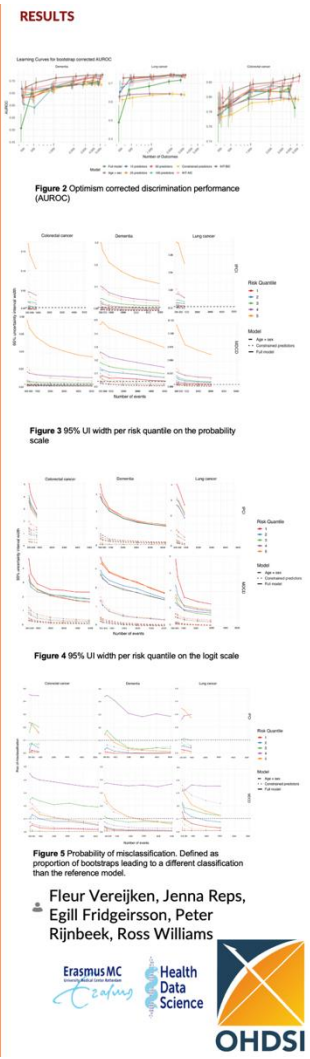
METHODS

1. Data sources
    1. Merative MarketScan® MultiState Medicaid Database (MDCD).
    2. The Integrated Primary Care Information (IPCI)
2. Prediction tasks
    1. Predicting colorectal cancer
    2. Predicting dementia
    3. Predicting lung cancer
3. Candidate predictors
    1. Age + sex
    2. UFS (k = 25)
    3. Constrained predictor set (n =51)
    4. IHT with BIC penalty
    5. Kitchen sink
4. Model choice
    1. LASSO
5. Model development
    1. Learning curves using 350-5,000 outcome events as sample sizes
    2. Bootstrap all sample sizes 200 times to estimate individual level instability
6. Metrics used
    1. AUROC (bootstrap corrected)
    2. 95% Uncertainty interval (UI) width:

per risk group (quantiles: 1 low risk, 5 high risk)

3. Probability of misclassification





#OHDSISocialShowcase This Week

Thursday

Congenital heart disease in pregnancy and severe maternal morbidity: a distributed data network study

(Elizabeth B. Sherwin, Benjamin Martin, Xiao Xu, Anna Booman, Alyssa Howren, Kimberlee McKay, Sadaf Kazi, Sean O'Reilly, Katie S. Kolla, Khyzer Aziz, Katherine Bianco, Stephanie A. Leonard)

Congenital heart disease in pregnancy and severe maternal morbidity

A distributed data network study

Authors: Elizabeth B. Sherwin[1], Benjamin Martin[2], Xiao Xu[3], Anna Booman[1], Alyssa Howren[1], Kimberlee McKay[4], Sadaf Kazi[5,6], Sean O'Reilly[2], Katie S. Kolla[7], Khyzer Aziz[2], Katherine Bianco[1], Stephanie A. Leonard[1]

INTRODUCTION

- **Background**: Pregnant people with congenital heart disease (CHD) are a growing patient population in obstetrics, yet evidence on the risk for severe maternal morbidity (SMM) has largely been limited to studies that lack specificity for CHD.
- **Study aim**: We conducted this study to demonstrate the utility of distributed data networks for obstetric research and to characterize pregnant patients with CHD and their risk of SMM.

METHODS

- **Study population:** Females age 12-55 who delivered at three U.S. academic healthcare institutions between 2015-2026
- **Exposure:** We used OMOP standard concepts to identify CHD from EHR data
- **Outcomes:** We used OMOP standard concepts to identify SMM from EHR data. Following CDC definitions, we identified SMM indicators occurring between one month prior to delivery through 6 weeks postpartum. We categorized outcomes as SMM, non-transfusion SMM, and cardiac SMM.
- **Analysis:** Analytical code was developed at the primary institution and executed locally at all sites. We analyzed incidence of the outcomes and estimated risk ratios and Wald 95% confidence intervals using contingency tables.

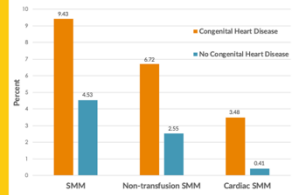
1 in 15 pregnant people with **congenital heart disease (CHD)** had **non-transfusion severe maternal morbidity (SMM)** compared to **1 in 39 people without CHD**. This study demonstrates the **feasibility** and **utility** of **OHDSI network studies for obstetric research.**

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

RESULTS

- Prevalence of CHD in pregnancy was 2.2% among 166,033 deliveries across three healthcare organizations.
- 19.3% of people with CHD had an acquired heart condition, compared to 2.1% of people without CHD.
- SMM occurred in 9.4% of deliveries to people with CHD compared to 4.5% among people without CHD (RR: 2.08; 95% CI: 1.88, 2.31), and non-transfusion SMM occurred in 6.7% vs 2.5% of deliveries (RR: 2.63; 95% CI: 2.33, 2.98), respectively (Figure 1).
- Cardiac SMM represented 51.8% of non-transfusion SMM among people with CHD compared to 16.2% among people without CHD

Figure 1. Incidence of SMM by congenital heart disease status.



AUTHOR AFFILIATIONS

[1]Stanford University School of Medicine, Stanford, CA
[2]Johns Hopkins University School of Medicine, Baltimore, MD
[3]Columbia University Irving Medical Center, New York, NY
[4]Avera Research Institute, Sioux Falls, SD
[5]National Center for Human Factors in Healthcare, MedStar Health Research Institute, Washington, DC
[6]Georgetown University Medical Center, Washington, DC
[7]Stanford Children's Health, Stanford, CA

Stanford MEDICINE Dunlevie Maternal-Fetal Medicine

PRIHSM PRomoting Improvement in Hemorrhage-related Severe Maternal Morbidity

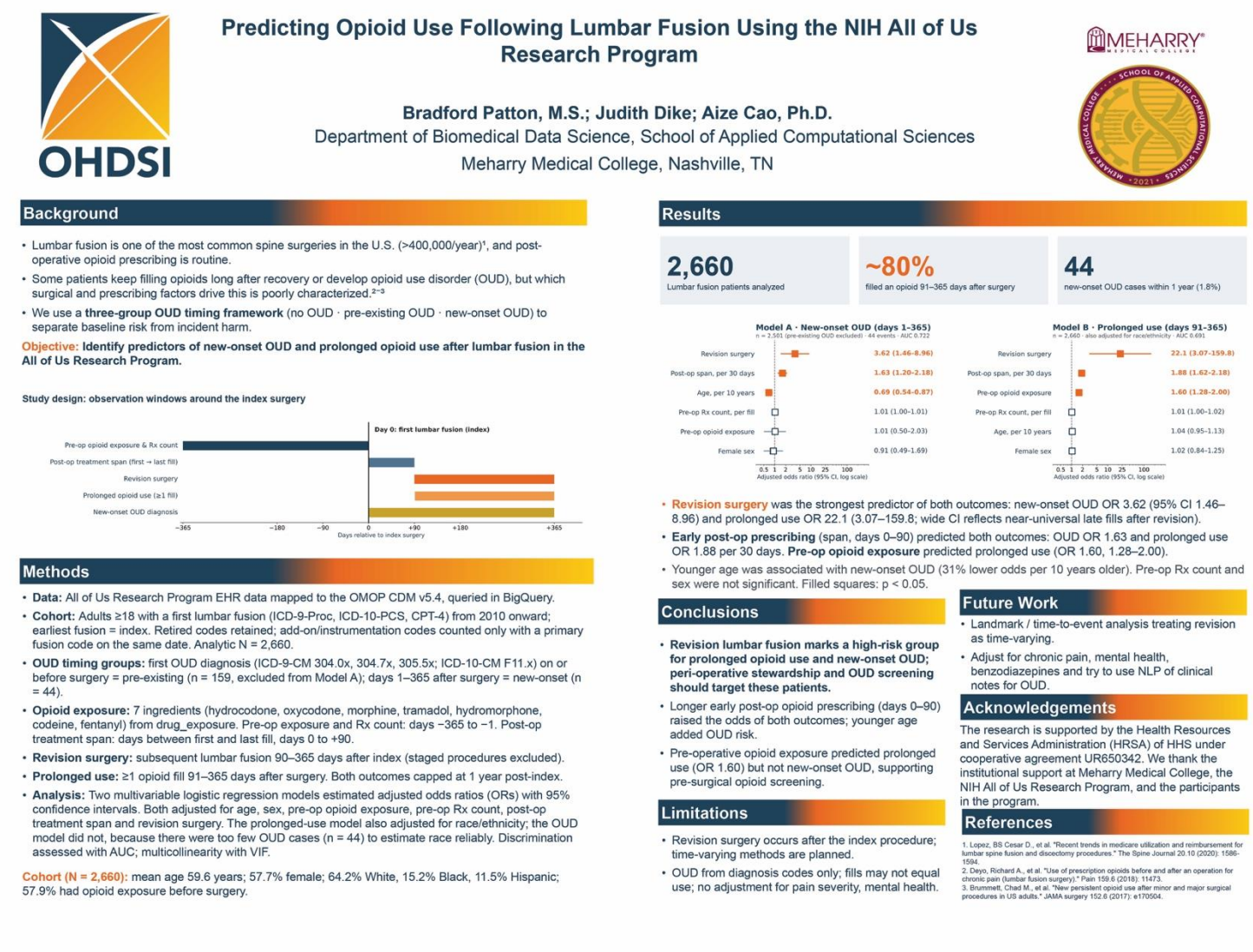
OHDSI

#OHDSISocialShowcase This Week

Friday

Predicting Opioid Use Following Lumbar Fusion Using the NIH All of Us Research Program

(Bradford Patton, Judith Dike, Aize Cao)



Predicting Opioid Use Following Lumbar Fusion Using the NIH All of Us Research Program

Bradford Patton, M.S.; Judith Dike; Aize Cao, Ph.D.
Department of Biomedical Data Science, School of Applied Computational Sciences
Meharry Medical College, Nashville, TN

Background

- Lumbar fusion is one of the most common spine surgeries in the U.S. (>400,000/year)[1], and post-operative opioid prescribing is routine.
- Some patients keep filling opioids long after recovery or develop opioid use disorder (OUD), but which surgical and prescribing factors drive this is poorly characterized.[2–3]
- We use a **three-group OUD timing framework** (no OUD · pre-existing OUD · new-onset OUD) to separate baseline risk from incident harm.

Objective: Identify predictors of new-onset OUD and prolonged opioid use after lumbar fusion in the All of Us Research Program.

Study design: observation windows around the index surgery

Methods

- **Data:** All of Us Research Program EHR data mapped to the OMOP CDM v5.4, queried in BigQuery.
- **Cohort:** Adults ≥18 with a first lumbar fusion (ICD-9-Proc, ICD-10-PCS, CPT-4) from 2010 onward; earliest fusion = index. Retired codes retained; add-on/instrumentation codes counted only with a primary fusion code on the same date. Analytic N = 2,660.
- **OUD timing groups:** first OUD diagnosis (ICD-9-CM 304.0x, 304.7x, 305.5x; ICD-10-CM F11.x) on or before surgery = pre-existing (n = 159, excluded from Model A); days 1–365 after surgery = new-onset (n = 44).
- **Opioid exposure:** 7 ingredients (hydrocodone, oxycodone, morphine, tramadol, hydromorphone, codeine, fentanyl) from drug_exposure. Pre-op exposure and Rx count: days −365 to −1. Post-op treatment span: days between first and last fill, days 0 to +90.
- **Revision surgery:** subsequent lumbar fusion 90–365 days after index (staged procedures excluded).
- **Prolonged use:** ≥1 opioid fill 91–365 days after surgery. Both outcomes capped at 1 year post-index.
- **Analysis:** Two multivariable logistic regression models estimated adjusted odds ratios (ORs) with 95% confidence intervals. Both adjusted for age, sex, pre-op opioid exposure, pre-op Rx count, post-op treatment span and revision surgery. The prolonged-use model also adjusted for race/ethnicity; the OUD model did not, because there were too few OUD cases (n = 44) to estimate race reliably. Discrimination assessed with AUC; multicollinearity with VIF.

Cohort (N = 2,660): mean age 59.6 years; 57.7% female; 64.2% White, 15.2% Black, 11.5% Hispanic; 57.9% had opioid exposure before surgery.

Results

2,660	~80%	44
Lumbar fusion patients analyzed	filled an opioid 91–365 days after surgery	new-onset OUD cases within 1 year (1.8%)

- **Revision surgery** was the strongest predictor of both outcomes: new-onset OUD OR 3.62 (95% CI 1.46–8.96) and prolonged use OR 22.1 (3.07–159.8; wide CI reflects near-universal late fills after revision).
- **Early post-op prescribing** (span, days 0–90) predicted both outcomes: OUD OR 1.63 and prolonged use OR 1.88 per 30 days. **Pre-op opioid exposure** predicted prolonged use (OR 1.60, 1.28–2.00).
- Younger age was associated with new-onset OUD (31% lower odds per 10 years older). Pre-op Rx count and sex were not significant. Filled squares: $p < 0.05$.

Conclusions

- **Revision lumbar fusion marks a high-risk group for prolonged opioid use and new-onset OUD; peri-operative stewardship and OUD screening should target these patients.**
- Longer early post-op opioid prescribing (days 0–90) raised the odds of both outcomes; younger age added OUD risk.
- Pre-operative opioid exposure predicted prolonged use (OR 1.60) but not new-onset OUD, supporting pre-surgical opioid screening.

Limitations

- Revision surgery occurs after the index procedure; time-varying methods are planned.
- OUD from diagnosis codes only; fills may not equal use; no adjustment for pain severity, mental health.

Future Work

- Landmark / time-to-event analysis treating revision as time-varying.
- Adjust for chronic pain, mental health, benzodiazepines and try to use NLP of clinical notes for OUD.

Acknowledgements

The research is supported by the Health Resources and Services Administration (HRSA) of HHS under cooperative agreement UR650342. We thank the institutional support at Meharry Medical College, the NIH All of Us Research Program, and the participants in the program.

References

1. Lopez, BS Cesar D., et al. "Recent trends in medicare utilization and reimbursement for lumbar spine fusion and discectomy procedures." The Spine Journal 20.10 (2020): 1586-1594.
2. Deyo, Richard A., et al. "Use of prescription opioids before and after an operation for chronic pain (lumbar fusion surgery)." Pain 159.6 (2018): 11473.
3. Brummett, Chad M., et al. "New persistent opioid use after minor and major surgical procedures in US adults." JAMA surgery 152.6 (2017): e170504.

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?

2026 Rising Star Presentations



Aniek Markus

Erasmus MC

DARWIN EU® – A Transportable Framework for Evaluating Pregnancy Episode Algorithms in Heterogeneous Real-World Databases



Hsin Yi “Cindy” Chen

Columbia University

Incorporating Inverse Probability of Censoring Weighting (IPCW) into the OHDSI CohortMethod Workflow



Nicholas Solages

Bascom Palmer Eye Institute

GLLaucoMed: A Secure LLM-Powered Agentic Workflow for Automated Medication Extraction from Free-Text Glaucoma Clinical Notes



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