# OHDSI 2025
# Software Demos

## OHDSI Community Call
## Nov. 11, 2025 • 11 am ET

# Upcoming Community Calls

| Date | Topic |
|---|---|
| Nov. 11 | Collaborator Showcase Demo Spotlight |
| Nov. 18 | DARWIN EU 2025 Update |
| Nov. 25 | Early-Stage Researcher Presentations |
| Dec. 2 | OHDSI/OMOP Research Spotlight |
| Dec. 9 | How Did OHDSI Do This Year? |
| Dec. 16 | Holiday Farewell To 2025 |

# DARWIN EU Update

# Three Stages of The Journey

# Where Have We Been?

## Where Are We Now?

## Where Are We Going?

# OHDSI Shoutouts!

Congratulations to the team of **Cindy Xinji Cai, Akihiko Nishimura, Sally Baxter, Kerry Goetz, Michelle Hribar, Brian Toy, Andrew Barkmeier, Sophia Wang, Swarup Swaminathan, Alexis Flowers, Eric Brown, Benjamin Xu, John Chen, Aiyin Chen, Theodore Leng, Michael Boland, Thamir Alshammari, Fan Bu, Thomas Falconer, Benjamin Martin, Erik Westlund, Nestoras Mathioudakis, Linying Zhang, Ruochong Fan, Adam Wilcox, Albert Lai, Jacqueline Stocking, Yangyiran Xie, Lok Hin Lee, David Dorr, Izabelle Humes, David McCoy, Mohammad Adibuzzaman, Raymond Areaux Jr, James Brash, Nicole Weiskopf, Hannah Morgan-Cooper, Priya Desai, Diep Tran, Zainab Rustam, Gina Zhu, Joel Swerdel, Anthony Sena, Paul Nagy, Marc Suchard, Martijn Schuemie, George Hripcsak, and Patrick Ryan** on the publication of **Semaglutide and diabetic retinopathy: an OHDSI network study** in *BMJ Open Diabetes Research and Care*.

## Semaglutide and diabetic retinopathy: an OHDSI network study

Cindy Xinji Cai [1,2] Akihiko Nishimura,[3] Sally Baxter,[4,5] Kerry Goetz,[6] Michelle Hribar,[6,7,8] Brian Toy,[9] Andrew Barkmeier,[10] Sophia Wang,[11] Swarup Swaminathan,[12] Alexis Flowers,[13] Eric Brown,[13] Benjamin Xu,[9] John Chen,[10] Aiyin Chen,[7,8] Theodore Leng,[11] Michael Boland,[14] Thamir Alshammari,[15,16] Fan Bu,[17] Thomas Falconer,[18] Benjamin Martin,[2] Erik Westlund,[3] Nestoras Mathioudakis,[19] Linying Zhang,[20] Ruochong Fan,[20] Adam Wilcox,[20] Albert Lai,[20] Jacqueline C Stocking [21] Yangyiran Xie,[13] Lok Hin Lee,[13] David Dorr,[8] Izabelle Humes,[22] David McCoy,[22] Mohammad Adibuzzaman,[22] Raymond Areaux Jr.,[23] James Brash,[24] Nicole Weiskopf,[8] Hannah Morgan-Cooper,[25] Priya Desai,[25] Diep Tran,[1] Zainab Rustam,[1] Gina Zhu,[1] Joel Swerdel,[26] Anthony Sena,[26,27] Paul Nagy,[2] Marc Suchard,[28,29] Martijn Schuemie,[28,30] George Hripcsak,[18] Patrick Ryan[18,30]

## ABSTRACT

**Introduction** Semaglutide, a glucagon-like peptide-1 receptor agonist (GLP-1RA) used to treat type 2 diabetes mellitus (T2D), has potential associations with higher rates of diabetic retinopathy (DR) complications including proliferative DR (PDR) and diabetic macular edema (DME). The purpose of this study was to determine whether an association exists between semaglutide and PDR and treatment-requiring DR/DME.

**Research design and methods** This was a retrospective cohort study of 14 databases (six administrative claims and eight electronic health records) in the Observational Health Data Sciences and Informatics Evidence Network. Adults with T2D on semaglutide, other GLP-1RA medications (dulaglutide, exenatide), or non-GLP-1RA medications (empagliflozin, sitagliptin, glipizide) from 1 December 2017 to 31 December 2023 were included. The association between semaglutide and PDR or treatment-requiring DR/DME was assessed using an active-comparator cohort design comparing new users of semaglutide as second-line T2D treatment to those on other GLP-1RAs and non-GLP-1RAs. Propensity score-adjusted Cox proportional hazards models were used to estimate hazard ratios (HRs). Network-wide HR estimates were generated using a random-effects meta-analysis.

### WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Semaglutide, a glucagon-like peptide-1 receptor agonist (GLP-1RA) used to treat type 2 diabetes mellitus, has been associated with higher rates of diabetic retinopathy complications.

### WHAT THIS STUDY ADDS

⇒ In this retrospective study of adults with type 2 diabetes mellitus across 14 databases in the Observational Health Data Sciences and Informatics Evidence Network, we did not identify an increased risk for proliferative diabetic retinopathy or treatment-requiring diabetic retinopathy comparing 810 390 new users of semaglutide with other GLP-1RAs and non-GLP-1RAs.

### HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ In contrast to some prior studies, real-world evidence does not suggest an increased risk of diabetic retinopathy complications with semaglutide. Patients with diabetes should still be monitored for vision-threatening complications.

# OHDSI Shoutouts!

👏

Congratulations to the team of **Thomas Datzmann, Caroline Lang, Falko Tesch, Melissa Spoden, Patrik Dröge, Franz Ehm, Ekkehard Schuler, Christos Krogias, Christian Günster, Jochen Schmitt, Christoph Gumbinger, and Jessica Barlinn** on the publication of **Evaluation of hybrid stroke quality indicators by integrating NIHSS and claims data for improved outcome prediction** in *Scientific Reports*.

## scientific reports

OPEN

### Evaluation of hybrid stroke quality indicators by integrating NIHSS and claims data for improved outcome prediction

Thomas Datzmann[1], Caroline Lang[1], Falko Tesch[1], Melissa Spoden[2], Patrik Dröge[2], Franz Ehm[1], Ekkehard Schuler[3], Christos Krogias[4], Christian Günster[2], Jochen Schmitt[1], Christoph Gumbinger[5,7] & Jessica Barlinn[6,7]

Accurately measuring the quality of stroke care based on claims data alone is challenging. Traditional outcome metrics, e.g. mortality rates, do not capture the effectiveness of critical stroke care processes. We aimed to develop hybrid quality indicators (QIs) by integrating clinical stroke severity data with claims data. Claims data were linked to patient-level clinical data from 15 hospitals (2017–2020) and harmonized in the Observational Medical Outcome Partnership (OMOP) data model. Inclusion criteria, outcomes and risk factors were developed by medical expert panels. We applied machine learning for modeling the outcomes *30-day-mortality*, *reinfarction within 90 days*, and *care degree increase within 180 days*. We compared extreme gradient boosting (XGBoost) models with and without the National Institutes of Health Stroke Scale (NIHSS) using Receiver-Operating-Characteristic-Area-Under-the-Curve (ROC-AUC) and Brier Score (BS). Hospitals were ranked according to the impact of each QI using Standardized Mortality Ratios (SMRs). The study included 9,348 ischemic (I63) and 1,554 hemorrhagic (I61) strokes, with NIHSS available for 5,012 patients. For all three outcomes, disease severity as measured by NIHSS was the most important determinant. The predictive power of the hybrid models was higher than that of models based on claims data alone. For SMR, the influence of NIHSS was greater than that of age, the most important variable in the claims data model. The results were consistent between the two entities, different outcomes, and sensitivity analyses. Including NIHSS information alongside claims data improves the risk adjustment of quality indicators.

**Keywords** Stroke, Health care quality assurance, Routinely collected health data, Health care quality indicators

# 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 | ATLAS/WebAPI |
| Tuesday | 12 pm | Generative AI and Analytics |
| Wednesday | 8 am | Psychiatry |
| Wednesday | 9 am | Patient-Level Prediction |
| Wednesday | 11 am | Common Data Model |
| Wednesday | 2 pm | Natural Language Proocessing |
| Wednesday | 7 pm | Eyecare and Vision Research |
| Wednesday | 7 pm | Medical Imaging |
| Thursday | 7 am | Europe Community Call |
| Thursday | 8 am | Medical Devices |
| Thursday | 9:30 am | Network Data Quality |
| Thursday | 10 am | Rare Diseases |
| Thursday | 10:30 am | Evidence Network |
| Thursday | 7 pm | Dentistry |
| Friday | 9 am | Phenotype Development and Evaluation |
| Friday | 10 am | GIS-Geographic Information System |
| Friday | 10 am | Transplant |
| Friday | 11 am | Clinical Trials |
| Friday | 11:30 am | Steering |
| Monday | 9 am | Vaccine Vocabulary |
| Monday | 10 am | Healthcare Systems |
| Monday | 11 am | Data Bricks User Group |
| Monday | 2 pm | Electronic Animal Health Records |
| Tuesday | 9 am | Data2Evidence |

# Africa Symposium: Nov. 10-12



**ohdsi.org/africa2025**

# India Symposium

# APAC Symposium: Dec. 6-7

Day 1 (December 6) – Tutorial at Room 102, Dongxia Yuan Building (Zheng-Cai Cuiju Teaching Building)

**Morning Session**

- 09:00-09:20 Introduction of OHDSI/OMOP
- 09:20-10:00 OMOP CDM and Vocabulary
- 10:00-10:30 OMOP Conversion Process
- 10:40-12:00 ETL Exercises

**Afternoon Session**

- 13:30-14:50 OHDSI Analyses: Building Cohorts & Hands-on
- 14:50-15:30 CohortDiagnostics and Population-Level Estimation
- 15:50-16:30 Interpreting Results

Day 2 (December 7) – Main conference at Room A100, 1F, Student Center

*Session 1 – From Global to Regional Impact: OHDSI across APAC & Africa*

- 09:00 – 09:15 Opening Speech
- 09:15 – 09:45 Keynote Speech from OHDSI Global
- 09:45 – 10:45 APAC Regional Chapter Updates
- 10:45 – 11:00 OHDSI Africa

Day 2 (December 7) – Main conference at Room A100, 1F, Student Center

*Session 1 – From Global to Regional Impact: OHDSI across APAC & Africa*

- 09:00 – 09:15 Opening Speech
- 09:15 – 09:45 Keynote Speech from OHDSI Global
- 09:45 – 10:45 APAC Regional Chapter Updates
- 10:45 – 11:00 OHDSI Africa

*Session 2 – From Research to Reflection: 2025 APAC Studies and Lessons Learned*

- 11:15 – 11:30 2025 APAC Study 1 by Fudan University
- 11:30 – 11:45 2025 APAC Study 2 by Peking University
- 11:45 – 12:00 2025 APAC Study 3 by University of Science and Technology of China (USTC)
- 12:00 – 12:10 Journal's Perspectives
- 12:10 – 12:30 Panel Discussion

*Session 3 – From Regional Insights to Local Challenges: Real-World Evidence and OHDSI/OMOP in China*

- 13:30 – 14:30 Collaborator Showcase: Lightning Talks
- 14:30 – 14:45 Real-World Evidence Talk 1
- 14:45 – 15:00 Real-World Evidence Talk 2
- 15:00 – 15:15 Real-World Evidence Talk 3
- 15:30 – 15:50 Real-World Evidence Using OHDSI/OMOP
- 15:50 – 16:10 Panel Discussion: Opportunities and Challenges Using OHDSI/OMOP for Real-World Evidence in China
- 16:10 – 16:50 Closing & Networking

**ohdsi.org/apac2025**

# Collaborator Spotlight: Andrew Kanter

# Latest OHDSI Newsletter is Available

## The Journey Newsletter (November 2025)

Last month, the OHDSI community came together for the 11th annual Global Symposium to explore how we can strengthen trust in science and expand global collaboration through network studies. The event celebrated shared progress, new partnerships, and the community's ongoing commitment to generating trustworthy, reproducible evidence around the world. Discover highlights and what's still ahead for OHDSI in our latest newsletter. #JoinTheJourney

## Podcast: Trust In Science, Building Collaboration

### On The Journey

Craig Sachson     Patrick Ryan

@OHDSI    www.OHDSI.org

In the latest On The Journey podcast, Patrick Ryan and Craig Sachson look back at the recent Global Symposium, focusing first on the connections made within the community during the showcase, tutorials and workshops. They discuss the scientific content, including topics around building trust in science and global collaboration in network studies. Then they look at how OHDSI can continue building with upcoming events in Africa and Asia-Pacific. *(If video does not appear, please click 'view this email in your browser.)*

## Community Updates

### Where Have We Been?
• The 2025 Global Symposium brought together more than 400 people for a three-day event (Oct. 7-9, New Brunswick, N.J.) that included a main conference, tutorials and workgroup activities. Learn more about the event throughout this newsletter.
• Congratulations to Michael Matheny (VA, Vanderbilt) on his recent election as a new member of the National Academy of Medicine (NAM).
• Congratulations to Early-Stage Researchers Workgroup lead and recent Columbia PhD graduate Harry Reyes Nieva on being named a 2025 STAT Wunderkind, an award given to the top 30 early career scientists in North America.

### Where Are We Now?
• The first-ever 2025 OHDSI Africa Symposium will be held Nov. 10-12 in Kampala, Uganda. Thank you to the symposium leadership team and the Africa Chapter for their leadership in putting together this event, which reached its maximum registration total.
• The first pan-Canadian event to bring stakeholders from across Canada who are engaged in OMOP CDM transformations and research will be held Nov. 17-18 in Toronto. The event will advance efforts to establish a Canadian OHDSI node and provide an opportunity to connect, collaborate, and gain insights into Canada's OMOP landscape.
• The OHDSI Sweden Symposium 2025 (Nov. 5) will bring together leaders, researchers, and innovators across healthcare, life science, policy and patient organizations to explore how the international data model OMOP and assets from the open-science community OHDSI can unlock the value of Sweden's health data.

### Where Are We Going?
• The 2025 OHDSI Asia-Pacific Symposium will be held Dec. 6-7 in Shanghai, China. Registration is open, and agenda details are available on the event homepage.
• Registration is now open for the 2026 Summer School in Observational Health Data Science & Informatics, AI, and Real World Evidence, which will be held June 22-26 at the Columbia University Department of Biomedical Informatics. Now in its second-year, the Columbia OHDSI Summer School provides health professionals, researchers, and industry practitioners with an immersive, hands-on training to working with real-world health data and generating real-world evidence (RWE). Participants will explore the types of

## Global Collaboration and Trust In Science Take Center Stage At OHDSI 2025

More than 400 collaborators convened in New Brunswick, N.J., for the 11th OHDSI Global Symposium, where the community explored how to strengthen trust in science through large-scale, collaborative studies. Plenary talks examined topics including why network studies are essential to evidence generation, and lessons learned from decades of FDA safety review. Researchers also presented new approaches to data standards, analytics, an applications of AI, showcasing the innovation and creativity driving OHDSI's work.

Beyond the research, the Symposium reaffirmed OHDSI's collaborative spirit. The "State of the Community" highlighted recent milestones and global leadership, while the Collaborator Showcase allowed teams to share emerging projects that will shape the next phase of OHDSI's journey. Posters and video from the event are available at the symposium homepage.

The Symposium also featured a full day of tutorials, including the always-popular *Introduction to OHDSI*, which welcomed newcomers into the community and provided a foundation for future collaboration. Other tutorials covered methods, tools, and use cases across the research lifecycle. In addition, a day of workgroup activities allowed community members to connec within their areas of expertise, share progress, and plan initiatives that will guide OHDSI's work in the coming year.

**Event Homepage: Talks, Slides**

**Collaborator Showcase: Posters, Software Demos**

## 2025 'Our Journey' Report Highlights Growing Global Community, Latest Advances

The 2025 edition of "Our Journey: Where the OHDSI community has been, and where we are going" was published prior to the Global Symposium and is available online. This annual publication provides an overview of all aspects of the OHDSI community, including its collaborators and activities, work in data, methods and open-source applications, and a list of nearly 900 peer-reviewed publications around the OMOP CDM or OHDSI tools/practices. This edition highlights the community events of 2025, testimonials from our global community, workgroups and regional teams, and plenty more.

**Our Journey PDF**

## October Publications

Paradinha R, Barros V, Almeida JR, Oliveira JL. A Semantic-Driven for Cohort Data Harmonisation into OMOP CDM Schema. Stud Health Technol Inform. 2025 Oct 2;332:190-194. doi: 10.3233/SHTI251524. PMID: 41041772.

Abedian S, Yesakov E, Ostrovskiy S, Hussein R. Integrating Garmin Wearable Data into FHIR-Based Health Systems for Improved Interoperability. Stud Health Technol Inform. 2025 Oct 2;332:185-189. doi: 10.3233/SHTI251523. PMID: 41041771.

Rajwa P, Borkowetz A, Abbott T, Alberti A, Beyer K, Bjartell A, Brash JT, Chilelli A, Davies E, Meulder B, Fazekas T, Golozar A, Hijazy A, Josefsson A, Kasivisvanathan V, Kolde R, Kotik D, Leapman MS, Miszczyk M, Nicoletti R, Prinsen P, Remmers S, Ribal MJ, Rivas JG, Rodriguez-Sanchez L, Roobol MJ, Smith E, Snijder R, Steinbeisser C, Stroomberg HV, Gandaglia G, Cornford P, Evans-Axelsson S, N'Dow J, Willemse PM; PIONEER Consortium. Observational Health Data Analysis of the Cardiovascular Adverse Events of Systemic Treatment in Patients with Metastatic Hormone-sensitive Prostate Cancer: Big Data Analytics Using the PIONEER Platform. Eur Urol Focus. 2025 Oct 3:S2405-4569(25)00251-2. doi: 10.1016/j.euf.2025.08.005. Epub ahead of print. PMID: 41046191.

Badri P, Hernández I, Long J, Amin M, Friesen R. Chronic orofacial pain and psychological distress: findings from a multidisciplinary university clinic. J Oral Facial Pain Headache. 2025 Sep;39(3):152-162. doi: 10.22514/jofph.2025.057. Epub 2025 Sep 12. PMID: 41070576; PMCID: PMC12520447.

Bohn J, Gilbert JP, Knoll C, Kern DM, Ryan PB. Large-scale Empirical Identification of Candidate Comparators for Pharmacoepidemiological Studies. Drug Saf. 2025 Nov;48(11):1229-1241. doi: 10.1007/s40264-025-01569-y. Epub 2025 Jun 4. PMID: 40467833; PMCID: PMC12515211.

Lee KH, Jang S, Kim GJ, Park S, Kim D, Kwon OJ, Lee JH, Kim YH. Large Language Models for Automating Clinical Trial Criteria Conversion to Observational Medical Outcomes Partnership Common Data Model Queries: Validation and Evaluation Study. JMIR Med Inform. 2025 Oct 16;13:e71252. doi: 10.2196/71252. PMID: 41100527; PMCID: PMC12530336.

Bellas L, Català M, Burn E, Guo Y, Du M, Verhamme K, Fridgeirsson E, Duarte-Salles T, Kauko T, Kronqvist E, Brash JT, Seager S, Prieto-Alhambra D, Jödicke AM, Prats-Uribe A. Secular Trends in the Use of Valproate-Containing

# Job Opening – Erasmus MC



All vacancies

New

## Health Data Scientist

•••••••

Drive innovation in healthcare at Erasmus MC. Develop open-source R tools for European research via DARWIN EU®. Work with top scientists to turn health data into actionable insights that improve lives.

### Job description

At Erasmus MC's Health Data Science group, you'll help shape the future of healthcare by developing open-source analytics that support large-scale epidemiological research. You'll work with health data standardized to the OMOP Common Data Model (OMOP-CDM), enabling studies across Europe through the DARWIN EU® network an initiative coordinated by Erasmus MC on behalf of the European Medicines Agency.

Your work will directly support researchers and regulators in generating real-world evidence to improve patient care. You'll collaborate with experts in epidemiology, data science, and software engineering, and contribute to a growing ecosystem of R packages that make health data analysis faster, more reliable, and reproducible.

Your responsibilities include:

- Developing and maintaining R packages for executing studies on OMOP-CDM formatted health data.
- Translating epidemiological study designs into executable software.
- Collaborating within the OHDSI community and contributing to open-source tools.
- Supporting the DARWIN EU® Coordination Centre by building scalable, reusable analytics.
- Applying software development best practices, including version control (GIT), documentation, and testing.

# Job Opening – Erasmus MC



Alle vacatures

Nieuw

## PhD student Medical Informatics

• • • • • •

Drive innovation at the intersection of AI, data science, and healthcare by developing cutting-edge methods for synthetic health data generation and analysis within a leading international research environment.

The Department of Medical Informatics at Erasmus MC invites applications for a PhD position at the forefront of health AI and data science. Join our innovative, multidisciplinary team engaged in international collaborations, including the Observational Health Data Sciences and Informatics (OHDSI) initiative, a global consortium dedicated to harnessing the value of health data via large-scale analytics, and the SYNTHIA project, a Europe-wide multidisciplinary initiative focused on advancing privacy-preserving healthcare solutions through the development and validation of tools and methods for synthetic data generation across diverse data types and clinical use cases.

As a PhD candidate, you will conduct research at the intersection of clinical data science and artificial intelligence, with a strong focus on the generation and evaluation of structured synthetic health data. You will also explore the application of large language models (LLMs) to healthcare data and terminology systems.

Your key research areas may include:

- Developing and evaluating methods for generating synthetic healthcare data from real-world primary care sources.
- Assessing the utility of synthetic data for population-level characterization and patient-level risk prediction.
- Advancing the extraction, retrieval, and standardization of information from clinical texts.

Throughout your research, you will address critical challenges in large-scale, federated observational studies and contribute to the advancement of multimodal synthetic data methodologies and analytic approaches in health data science.

# #OHDSISocialShowcase This Week

## Monday

**Leveraging OHDSI Tools for Advancing Federated Machine Learning in Clinical Prediction: Insights from the PREPARE project**

(Margarita Grammatikopoulou, Vasilis Alepopoulos, Giorgos Giannios, Spiros Nikolopoulos, PREPARE Project Group)



The PREPARE-Rehab project uses OHDSI tools to support privacy-preserving clinical prediction. A federated learning simulation framework has been developed and tested using synthetic data as part of the project

Leveraging OHDSI Tools for Advancing Federated Machine Learning in Clinical Prediction: Insights from the PREPARE project

**Background:** The PREPARE project aims to improve rehabilitation care for chronic disease patients using AI-driven patient stratification and outcome prediction. To protect patient privacy, it employs federated learning, which enables decentralized model training across clinical sites without sharing sensitive data. By integrating OHDSI's open-source tools, the project is developing a scalable, privacy-preserving pipeline for clinical prediction using real-world data from multiple conditions.

**Results**
- Prediction challenge:
  - "Identify new NSAID users at risk of gastrointestinal bleeding within 3 years of treatment initiation"
- The Federated Global model outperforms the Local Models but underperforms compared to the Global Model
- Synthetic dataset appears sufficiently homogeneous to predict distributions, leading to similar prediction outcomes for local and global models

**Conclusions**
- The PREPARE project successfully demonstrates the integration of OHDSI tools into a federated learning pipeline
- This approach enables privacy-preserving, large-scale model training across multiple clinical sites and institutions
- Future work will focus on:
  - refining FL aggregation methods,
  - expanding real-world validation,
  - and extending the methodology towards more complex models (e.g. Deep Learning, Ensemble models)
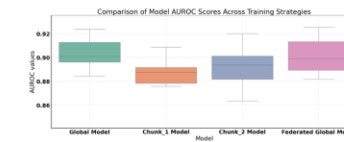
**Methods**
- ACHILLES and ARES are used to extract and visualize descriptive statistics from available data sources
- ATLAS facilitates cohort definition and the prediction task setup
- HADES provides all the packages necessary for the machine learning training process.
- Local model training is conducted on ARACHNE Nodes, with ARACHNE Central orchestrating the federated access
- PREPARE Middleware aggregates the local models into a single Global Model and provides it for inference through the dedicated PREPARE platform UI
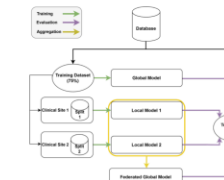
**Federated Learning Simulation**
1. The Eunomia dataset was split into 75% training and 25% testing sets
2. A Global Model was trained using the entire training set to serve as a baseline solution
3. The training set was further divided into two uneven partitions, simulating heterogeneous clinical data nodes
4. Each partition was used to train a Local Model independently, representing decentralized training on local data. Then, local models were aggregated to constitute the Federated Global Model
5. In total, all four models were evaluated on the initial 25% test set

**Limitations**
- The current work is based solely on synthetic datasets, which may lack the complexity and heterogeneity found in real-world data. To address this, publicly available real-world datasets such as MIMIC and the OMOPed CDM databases of PREPARE clinical partners will be used to assess performance in complex clinical settings.
- The pipeline currently utilizes the PatientLevelPrediction (PLP) package, which supports only a limited set of traditional machine learning models, potentially restricting performance in complex prediction tasks. To overcome this, the DeepPLP package by OHDSI will be considered, enabling the integration and customization of advanced deep learning architectures within the OHDSI framework.

Margarita Grammatikopoulou, Vasilis Alepopoulos, Giorgos Giannios, Spiros Nikolopoulos, and the PREPARE Project Group

# #OHDSISocialShowcase This Week

## Tuesday

**OHSIRIS - Open Health Data Space for Biomedical Research and development of Industrial Services based on OMOP**

(**Alberto Moreno**, Jesus Moreno, Francisco Nuñez, Miguel Giraldez, Samuel Salas, Angela Leis, Noelia García, Victoria Alonso, Raúl Martínez, Ibai Tamayo, Ariadna Curto, Jose Manuel Simarro)

# #OHDSISocialShowcase This Week

## Wednesday

### The birth of the Belgian FHIN-hospitals Industry Data Alliance (BELFHINDA)

(Karen Crabbé, Peter De Jaeger, Camille Deltomme, Kim Denturck, Aldo Elsen, Katoo Muylle)

---

**BELFHINDA : The Belgian FHIN-hospitals and Industry Data Alliance**

Feasibility assessment in non-small cell lung cancer (NSCLC) leveraging the OMOP CDM through an innovative public-private partnership in Belgium

**Background:** The Federated Health Innovation Network (FHIN) VZW is a collaborative initiative uniting 8 academic and non-academic hospitals in Belgium to facilitate data-driven care by leveraging data harmonization via the OMOP CDM, with a specific focus on prediction models in lung and prostate cancer [1]. The Belgian pharmaceutical Industry Association (pharma.be) aims to advance Belgian real-world data (RWD) research by establishing a Belgian Hospital Industry Data Alliance (BELHINDA). The collaboration with the FHIN VZW as BELFHINDA is a first example focusing on federated collaborative RWD in the lung cancer domain. Our aim is to present the principles, set-up and status of this innovative partnership.

**1 Agreement on principles**

Table 1. Principles of the partnership

| Principle | Outcome |
|---|---|
| Goal | Feasibility assessment to map the NSCLC patient journey across multiple centers using federated data analytics to inform data optimization strategies and study designs |
| Governance | ✓ 1 point of contact & 1 point of contract<br>✓ Streamlined approval procedure<br>✓ Privacy-by-design using federated data model |
| Setting | Non-competitive product agnostic in NSCLC |

**2 Governance workstream**

Figure 1. BELFHINDA governance
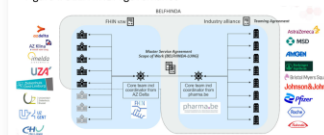
From 5*9 contracts in a traditional set-up to 3 contracts (teaming agreement, master service agreement and scope of work) with 15 partners including pharma.be and FHIN vzw in the BELFHINDA set-up.

**3 Content workstream**

2 scoping workshops

Table 2. Joint development of the content of the feasibility assessment

| Subject | Outcome |
|---|---|
| Study period | 1 Jan 2020 – 31 Dec 2024 |
| Population | Prevalent and incident stage I-IV NSCLC based on national cancer registry data |
| Datapoint list | 23 categories representing >2000 data points |
| Clinical definitions beyond data points | - Algorithms for treatment types, categories and outcomes based on Athena codes<br>- Treatment lines based on OHDSI Regimen Finder |
| Challenges | - Differences in data maturity between hospitals<br>- Data points from multiple sources (EHR, lab, pathology, etc.)<br>- Translating research questions into technical ETL challenges |

**Data quality metrics**

OHDSI Data Quality Dashboard — Check overall data completeness, conformance, and plausibility

Custom Data Quality Checks — Check data availability in electronic health records (% and ranges), availability in OMOP databases (% and ranges) and quality (outlier analysis) per data point or clinical definition

**Anticipated output**

- Data catalogue with all defined OMOP concept IDs
- Data Quality Dashboard per hospital partner
- Outcome custom data quality checks per hospital partner

**Conclusion:** Substantial simplification in governance and contracting set-up compared to traditional approaches was achieved. Requirements for a feasibility assessment with use case in NSCLC (data point list, data quality metrics, reporting format) were jointly developed. These efforts lay the foundation for a leading and sustainable framework for collaborative RWD research in Belgium.

Karen Crabbé[1], Siel Depestele[2], Camille Deltomme[2], Peter De Jaeger[2], Kim Denturck[2], Aldo Elsen[3] and Katoo Muylle[4] on behalf of BELFHINDA-Lung group
1 pharma.be, Belgium; 2 RADAR AZ Delta Roeselare, Belgium; 3 MSD Belgium; 4 AstraZeneca BeLux

Reference: 1. Deltomme C, Denturck K, De Jaeger P, et al. Data-driven medical practices and enhanced patient care through OMOP CDM and federated learning in 9 Belgian hospitals. Presented at OHDSI Europe 2024.

Acknowledgments: The BELFHINDA-Lung group: for the hospital partners (Annelies Verbiest, Reinier Wener, Stephanie Detailleur, Karel Van Brantegem, Bram De Caluwé, Hilde Philips, Sofie Vandemoortele, Josica Werbrouck, Lore Vermeylen, Atiqa Ahmad, Zarah Van Schoor, Pravesh Gurung) and for the industry partners (David Smeets, Katoesja vander Hoeven, Jens Tytgat, Bieke van der Veken, Géraldine Devos, Ann Curias, Catheline de Moyer, Bart Verheyden, Jo Jansen, Geert Van Gassen, Eline Darquennes, Jan Vertriest, Bart Vannieuwenhuyse, Larissa De Rop, Hilde Van Campenhout, Saartje Burgmans, Erwin Bruninx)

---

# #OHDSISocialShowcase This Week

## Thursday

# Harmonization of NAJS DWH to the OMOP CDM

(**Marko Čavlina**, Antea Jezidžić, Karlo Pintarić, Pero Ivanko)



Mapping **Croatian health data** from diverse sources **to OMOP CDM**

Title: *Harmonization of NAJS DWH to the OMOP CDM*

**Background:** Croatian National Public Health Information System is our main data source, a national-level coverage access to information from primary (visits, receipts, vaccinations and laboratory results), secondary conciliatory and hospital care, and a number of registries like Cancer Registry, Causes of Death Registry etc..
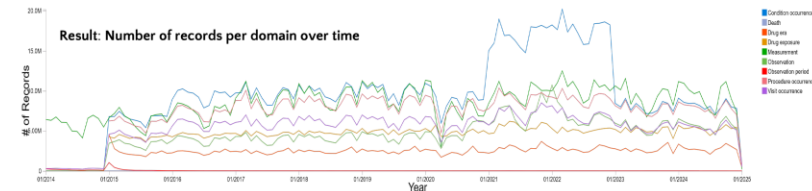
**Key challenges** in the ETL process included the lack of internal documentation, questionable mapping choices that required several revisions and hardware and logic scripts that were not optimized for manipulation of this amount of data. With a solid base mapped we were able to participate in over a dozen of network studies, and we have plans to intensify our participation and contribution in the future.
**Future efforts** include mapping emergency hospitalizations, drug exposure from hospital visits, and the full integration of additional registries, such as the Cancer Registry, Registry of Births and Pregnancies, and Diabetes Registry.
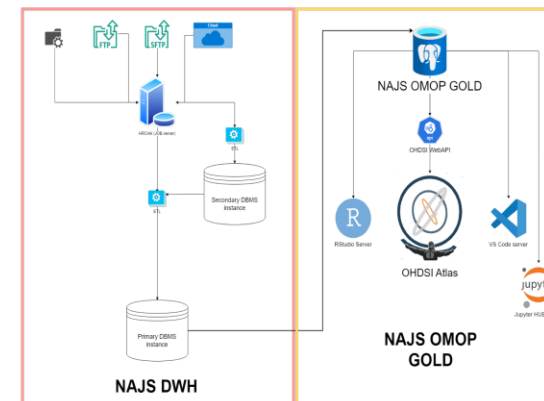
Start of data collection:
- 2014 for lab tests,
- 2015 for general practitioners,
- 2016 for secondary conciliatory care,
- 2017 for hospital records
- 2020 for vaccination records.

Marko Čavlina, Antea Jezidžić, Karlo Pintarić, Pero Ivanko

# 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?

# OHDSI2025 Software Demos

**Adil Ahmed**

*Senior Machine Learning Engineer, Montefiore*

Automated OMOP Concept Mapping Using Multi-Agent Large Language Models and Graph-Enhanced Semantic Retrieval

**Ioanna Nika**

*Data Scientist, Erasmus MC*

DarwinBenchmark: Evaluating cohort generation and analytics in OMOP CDM databases

**Cameron Atkins**

*Software Developer, Johnson & Johnson*

Enhancing Empirical Comparator Recommendations with User Specified Weights: Approach and Assessment

**Katy Sadowski**

*Senior Associate Director, Real World Evidence Analytics, Boehringer Ingelheim*

dqdbt: Continuous Data Quality Testing for OMOP ETL with dbt

**Robert Barrett**

*Research Assistant, Johns Hopkins University*

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

**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-2025**