



Workgroup OKRs + Phenotype Phebruary, Session 3

OHDSI Community Call
Feb. 18, 2025 • 11 am ET



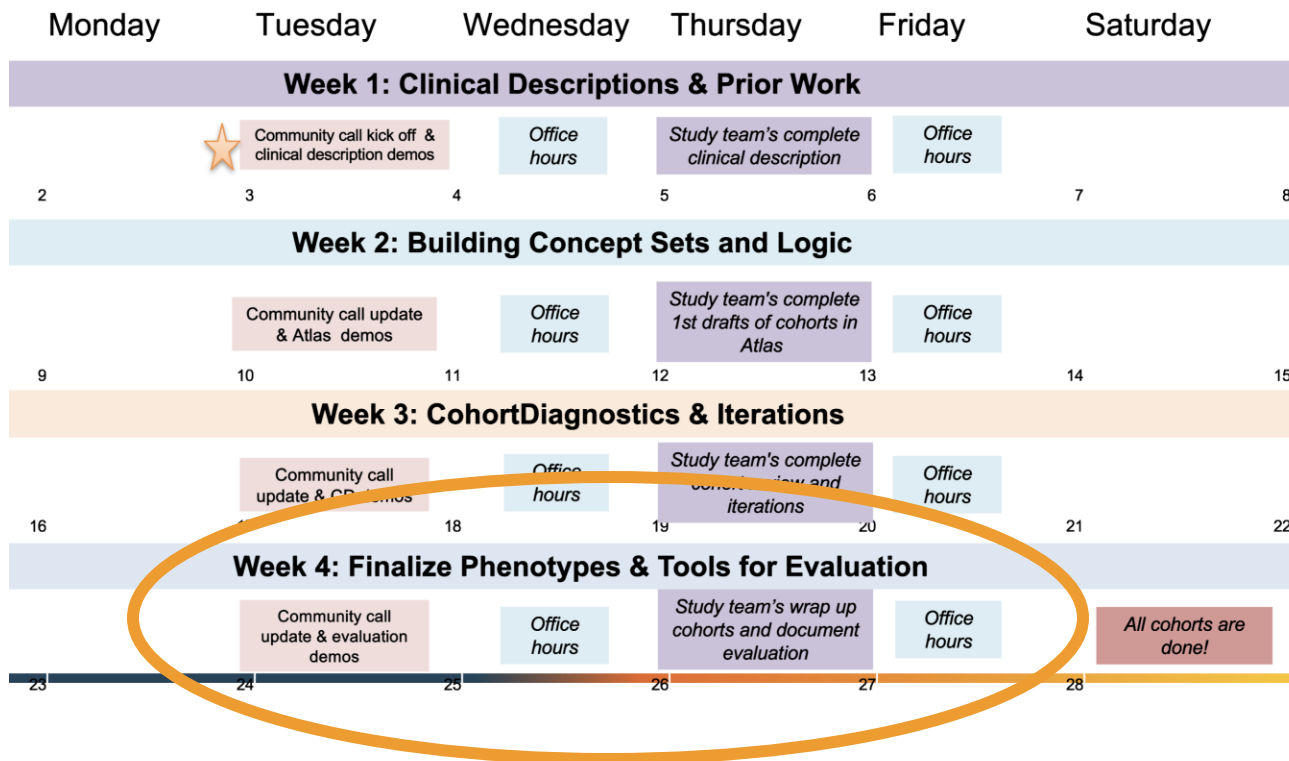
Upcoming Community Calls

Date	Topic
Feb. 18	Third Week of 2025 Workgroup OKRs/Phenotype Phebruary
Feb. 25	Fourth Week of 2025 Workgroup OKRs/Phenotype Phebruary
Mar. 4	Vocabulary Release Update, Winter 2025
Mar. 11	Book of OHDSI 2.0 Brainstorm and Planning Session



Feb. 25 Community Call

Phenotype Phebruary 2025 Calendar



Workgroup OKRs:

Africa Chapter

CDM Survey

Clinical Trials

GIS

Health Equity

Medical Device

Natural Language Processing

Oncology

Rare Disease

Surgery and Perioperative

Medicine

Themis



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!



Congratulations to **Phuc Phan Thanh** on successfully completing his PhD journey at Taipei Medical University.





OHDSI Shoutouts!



Congratulations to
Anthony Sena on
successfully earning his
Master of Science in
Analytics at Georgia Tech





Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?



iCAN mNSCLC Studyathon 2025




 March 25-28, Helsinki, Finland

Exploring the Real-World Treatment Landscape of mNSCLC

In this studyathon, we will characterize real-world treatment patterns of metastatic NSCLC, with a focus on the adoption and impact of immune checkpoint inhibitors (ICIs) across different regions.

 **Study GitHub Repository:** <https://github.com/ohdsi-studies/MNSCLCStudyathon>

 **If you're interested in contributing, please reach out:**

-  **Asieh Golozar** – golozar@ohdsi.org
-  **Kimmo Porkka** – kimmo.porkka@helsinki.fi
-  **Eric Fey** – eric.fey@hus.fi



Global Symposium: Oct. 7-9

The 2025 OHDSI Global Symposium will return to the Hyatt Regency Hotel in New Brunswick, N.J., on Oct. 7-9, 2025.

More details will be shared when available.



#OHDSISocialShowcase This Week

Monday

Going global, redeeming the local: an innovative approach to implement the OMOP CDM in two countries of the Global South

(Valentina Martufi, Emma Kalk, Enny S. Cruz, Juliana Araújo Prata de Faria, Adalton do Anjos Fonseca, Maurício L. Barreto, Maria Yury Travassos Ichihara, Jessica Gammon, Nicki Tiffin, Chris Fourie, Danilo Luis Cerqueira Dias, Denise Moraes Pimenta, Tsaone Tamuhla, Andrew Boule, Themba Mutemaringa, Juan-Paul Hynek, Muzzammil Ismail, Julio Barbour Oliveira, Ricardo Felix Monteiro Neto, Júlia Pescarini, Fernanda Revoredo de Sousa, Marianne Costa e Silva Lage, Adam Loff, Melvin Moodley, Elzo Pereira Pinto Junior)

Going global, redeeming the local: an innovative approach to implement the OMOP CDM in two countries of the Global South

Valentina Martufi

INTRO
 Today more than ever, health research has an impact beyond local boundaries.
 The OMOP CDM has facilitated the elaboration of federated analyses across different countries, expanding the knowledge base for the improvement of global health.

Uptake has predominantly been in the Global North.
 → The OMOP CDM has been strongly modelled to the socio-economic and health systems context of Global North countries
 → Global South countries remain significantly underrepresented OHDSI open-source development community



(classic) METHODS
 ✓ Cohort study
 ✓ Occurrence of gestational syphilis and tuberculosis in pregnant women
 ✓ In Brazil and the Western Cape Province, South Africa
 ✓ 2013-2018
 ✓ (partially) implemented on synthetic databases

Analyses will be performed to understand the relationship between syphilis and TB during pregnancy: health outcomes for the mother and baby, and socioeconomic variables.

(innovative) METHODS
 Differences between the variables held by the two collaborating institutions + These may also not be fully contemplated by the OMOP CDM = preconception with living, relevant context-specific information in the federated analyses

Proposed alternative approach:
A. Main set of analyses will contemplate the variables included in the OMOP CDM with data from both institutions;
B. Additional set of analyses applied only to CIDACS data (mapped to the OMOP CDM), considering additional significant variables that characterize Brazil-specific risk or protective factors for health outcomes, such as participation in social welfare programs (e.g. the Bolsa Familia conditional-cash transfer policy) and race;
C. Additional set of analyses applied only to PHDC data (mapped to the OMOP CDM), considering additional relevant variables such as contact with health service providers and laboratory results.

RESULTS
From main analyses (A):
 → enriched understanding of these two globally relevant infectious diseases, contextualizing their behavior in two Global South countries that hold many similarities.

From additional sets of analyses (B and C):
 a. assess and measure the relevance of the additional variables employed in each of them, compared to set A;
 b. highlight the importance of collecting specific types of data (eg. related to socio-economic characterization of individuals or their context) within health information systems, to provide more comprehensive panorama of health risk factors and outcomes; and
 c. provide evidence to make the case to include such variables in the OMOP CDM (in the cases when they are not), so as to make it more inclusive and truly globally representative.

References:
 1. Martufi V, et al. The Impact of COVID-19 on a Brazilian Network. *PLoS ONE* 2021;16(10):1-12. <https://doi.org/10.1371/journal.pone.0249701>
 2. Digital Health: Resilience - Introduction. <https://www.who.int/digitalhealth/evidence/resilience-introduction>
 3. <https://www.who.int/digitalhealth/evidence/resilience-introduction>
 4. <https://www.who.int/digitalhealth/evidence/resilience-introduction>
 5. <https://www.who.int/digitalhealth/evidence/resilience-introduction>
 6. <https://www.who.int/digitalhealth/evidence/resilience-introduction>
 7. <https://www.who.int/digitalhealth/evidence/resilience-introduction>
 8. <https://www.who.int/digitalhealth/evidence/resilience-introduction>
 9. <https://www.who.int/digitalhealth/evidence/resilience-introduction>
 10. <https://www.who.int/digitalhealth/evidence/resilience-introduction>

A common data model to study infectious diseases affecting pregnancy, representative of the Global South

We hope our experience will promote the valorization of socioeconomic information, including employment and income, education, etc., as well as be an incentive to apply the OMOP CDM to generate health evidence in other countries in the Global South.

Acknowledgments:
 This collaboration was funded by the Bill & Melinda Gates Foundation

Contact email: valentina.martufi@fiocruz.br

Find out more about the PHDC
 Find out more about CIDACS' 100m Brazilians cohort

Partner institutions:
 CIDACS-IGM/FIOCRUZ-BA, WC-PHDC, SANBI-UWC, Precision Data

Participating institutions:
 FIOCRUZ Bahia, CIDACS, Instituto Cochrane, SANBI, UWC

AMMO BAR

Why synthetic databases?
 generated from the databases stored and curated by CIDACS and the PHDC, they allow for the utilization of cloud-based resources provided by the OHDSI community whilst ensuring the protection of the data from these two populations.

What we will study
 To explore the relationship between syphilis, prematurity/low birth weight and infant mortality, in addition to standard regression methods, we will use statistical approaches to decompose the direct and indirect effects of syphilis during pregnancy and infant mortality, such as inverse probability weighting based on the counterfactual structure and marginal models. Finally, to assess the impact of TB on birth outcomes, we will apply logistic regression adjusted for selected variables.

What data we will use to study this
 Data sources
 In Brazil variables are being mapped from the following databases: the live-births information system (SINASC), the mortality information system (SIM), the single registry for social welfare programs (CadÚnico), the gestational syphilis and tuberculosis compulsory notification systems (SINAN), and the hospitalizations information system (SIH). In South Africa, the PHDC is mapping variables from multiple electronic health data sources used to generate an individual-level population database, namely: laboratory, administrative, pharmacy and disease-specific registers.

Why Brazil and South Africa
 two Global South countries that hold many similarities – including global economic standing, colonial history, and stark levels of racial, socioeconomic and health inequalities – as well as many differences – such as health systems organization, post-colonial dynamics, and epidemiological profile.

Gestational syphilis variables being mapped to the OMOP CDM

Variable	Source	Mapping Status
Gestational syphilis	SINAN	Completed
Gestational syphilis (clinical)	SINAN	In Progress
Gestational syphilis (laboratory)	SINAN	In Progress
Gestational syphilis (treatment)	SINAN	In Progress
Gestational syphilis (diagnosis)	SINAN	In Progress

Tuberculosis during pregnancy variables being mapped to the OMOP CDM

Variable	Source	Mapping Status
Tuberculosis during pregnancy	SIH	Completed
Tuberculosis during pregnancy (clinical)	SIH	In Progress
Tuberculosis during pregnancy (laboratory)	SIH	In Progress
Tuberculosis during pregnancy (treatment)	SIH	In Progress
Tuberculosis during pregnancy (diagnosis)	SIH	In Progress

Valentina Martufi, Emma Kalk, Enny S. Cruz, Juliana Araújo Prata de Faria, Adalton do Anjos Fonseca, Maurício L. Barreto, Maria Yury Travassos Ichihara, Jessica Gammon, Nicki Tiffin, Chris Fourie, Danilo Luis Cerqueira Dias, Denise Moraes Pimenta, Tsaone Tamuhla, Andrew Boule, Themba Mutemaringa, Juan-Paul Hynek, Muzzammil Ismail, Julio Barbour Oliveira, Ricardo Felix Monteiro Neto, Júlia Pescarini, Fernanda Revoredo de Sousa, Marianne Costa e Silva Lage, Adam Loff, Melvin Moodley, Elzo Pereira Pinto Junior



#OHDSISocialShowcase This Week

Tuesday

A Computable Phenotype for HSV Anterior Uveitis: Operationalizing the SUN Classification Criteria

(**Brian Toy**, Edward Lee, Andrew Kim, Edmund Tsui, Jessica Shantha, Kareem Moussa, Karen Armbrust, William Rojas Carabali, Rupesh Agrawal, Kiana Tavakoli)



A Computable Phenotype for HSV Anterior Uveitis: Operationalizing the SUN Classification Criteria employing the OMOP Common Data Model

OHDSI Eyecare and Vision Workgroup - Uveitis Subgroup
Andrew Kim¹, Edward Lee¹, Lingling Huang², Karen Armbrust³, Kiana Tavakoli², Edmund Tsui⁴, Jessica Shantha⁵, Kareem Moussa⁶, William Rojas Carabali⁷,
Rupesh Agrawal⁷, Lola Solebo⁸, Brian Toy¹

¹Roski Eye Institute Keck School of Medicine USC, ²Slit Eye Institute University of California San Diego, ³Minneapolis VA Health Care System University of Minnesota, ⁴UCLA Stein Eye Institute David Geffen School of Medicine at UCLA, ⁵EL Proctor Foundation University of California, San Francisco, ⁶Department of Ophthalmology & Vision Science University of California Davis, ⁷Lee Kong Chian School of Medicine Nanyang Technological University Singapore, ⁸University College London



Abstract

Data models can enhance observational research for rare diseases like uveitis. This study operationalized the Standardization of Uveitis Nomenclature (SUN) classification criteria for herpes simplex virus (HSV) anterior uveitis into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Using the Observational Health Data Sciences and Informatics (OHDSI) ATLAS tool, we constructed a computable phenotype, applied it to two electronic health records (EHR) systems for internal validation, and to Optum's Clinformatics® claims database for external validation. In EHR manual chart reviews, we identified 5,404 and 38 patients at two sites, achieving 70% and 90% accuracy, respectively, in a randomized validation sample. From the claims database, 10,858 (2.7%) patients met all criteria, consistent with previous literature. The most discriminatory inclusion criterion was a history of HSV keratitis. Our results indicate that OMOP can accurately identify HSV anterior uveitis, suggesting potential utility for operationalizing other uveitis subtypes in observational research.

Background

Classification Criteria for Herpes Simplex Virus Anterior Uveitis

THE STANDARDIZATION OF UVEITIS NOMENCLATURE (SUN) WORKING GROUP^{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100}

- SUN Working Group:** An international collaborative initiative focused on creating uniform terminology and classification for uveitis.
- Development of Criteria:**
- Classification criteria established for 25 common uveitis entities
 - Methodological approach included:
 - Informatics-based case collection.
 - Consensus techniques for case selection.
 - Machine learning for rule development.
 - HSV Anterior Uveitis Classification Criteria were developed in 2021 by the SUN Working Group (Fig 1)
- Implementation:**
- These criteria have the potential to be useful in the identification of patients for observational research
 - Real-life application can be labor-intensive, as evidenced by a study that identified 1,143 patients with HSV Anterior Uveitis from all patients who attended the University of Colorado Hospital Uveitis service from 2013-2020 through manual chart review (Mudie et. al, 2022)
 - Current unmet need for efficient application of SUN criteria in real-world data.
 - We proposed to create a computable phenotype using the OMOP Common Data Model to streamline data integration and analysis across multiple source databases

Contact: Brian.Toy@med.usc.edu

Methods

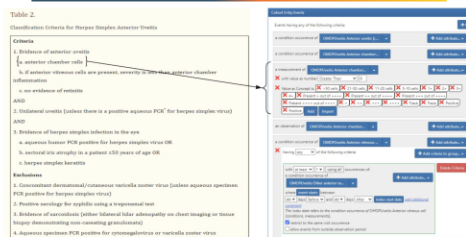


Fig 1. Each of the SUN criteria was transformed to OMOP concept sets. Logic statements were constructed. Here you can see how criteria 1A of "anterior chamber cells" was operationalized in the OHDSI ATLAS tool

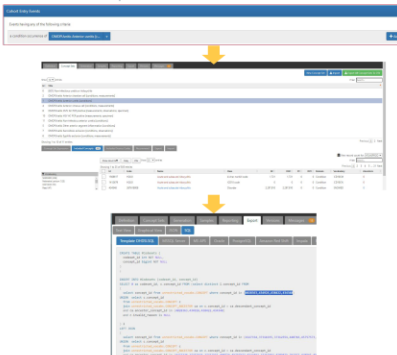


Fig 2. The SUN criteria were mapped to standardized concepts, comprising specific conditions, measurements, or observations. Descendants of these concepts, as well as suggested concepts, were searched for using the OHDSI Athena and ATLAS tools and added to the corresponding concept sets. SQL code was generated based on our mapped concept sets and applied to existing databases.

Results: Internal and External validation

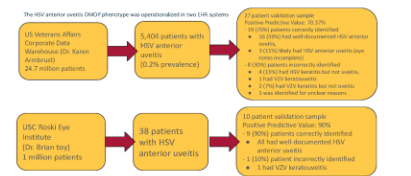


Fig 3. Internal EHR Chart Review Validation

- Optum Clinformatics Claims Database External Validation Cohort Overview:**
- 399,914 patients initially met the entry criteria of Uveitis
 - 10,858 patients (2.7%) were identified to have HSV Anterior Uveitis after incl/excl
 - 55.8% female, 44.2% male
 - 86.7% of patients > 40 years old
- In a paper by Zhang et al. published in 2020
- Inclusion in the sample required index uveitis diagnosis based on ICD-9 codes
 - 18,523 patients initially met the entry criteria
 - 442 patients (2.4%) were identified to have HSV Anterior Uveitis based on ICD-9 codes

Conclusions

- Significant Findings:**
- Iridocyclitis and Acute and Subacute Iridocyclitis conditions were the standard concepts that were most prevalent to identify uveitis patients from both claims and EHR databases (SUN HSV anterior uveitis criterion #1)
 - HSV keratitis condition was the most common standard concept to identify evidence of herpes simplex infection in the eye (SUN HSV anterior uveitis criterion #3)
 - Validation of the selected cohorts from EHR data was confirmed by chart review from 2 EHR systems
 - Validation of selected cohorts from claims data was confirmed by comparison with other existing literature

- Limitations:**
- Lack of representation of PCR results and sectoral iris atrophy highlights a need to improve ophthalmic data standards which support comprehensive definitions that fully leverage the SUN classification criteria

Disclosures/Acknowledgments

No relevant conflicts of interest.
BT: Physician advisory boards (Allergan, Bausch and Lomb, Eyepoint, Regeneron), Funding (NIH/NEI K23EY032985)

References

Mudie L, Reddy A, K, Pinnell J, Peers R, Kim E, Cole K, & Paley A. G. (2022). Evaluation of the sun classification criteria for Uveitis in an academic uveitis practice. American Journal of Ophthalmology, 241, 57-63. <https://doi.org/10.1016/j.ajo.2022.04.007>

THE STANDARDIZATION OF UVEITIS NOMENCLATURE (SUN) WORKING GROUP (2021). Classification criteria for herpes simplex virus anterior uveitis. American Journal of Ophthalmology, 228, 231-236. <https://doi.org/10.1016/j.ajo.2021.03.018>

Zhang, X, Amin, S, Lung, K, L, Seabury, S, Rao, N, & Toy, B. C. (2020). Incidence, prevalence, and risk factors of infectious uveitis and keratitis in the United States: A claims-based analysis. PLOS ONE, 15(8). <https://doi.org/10.1371/journal.pone.0237995>



#OHDSISocialShowcase This Week

Wednesday

Comparative Evaluation of Methods for Defining Observation Periods in Healthcare Databases and Their Impact on Incidence Rate Estimates

(Clair Blacketer, Patrick Ryan, Martijn Schuemie, Peter Rijnbeek)

Comparative Evaluation of Methods for Defining Observation Periods: Impact on Incidence Rate Estimates and Suggestions for the Future

INTRO:

- Defining the Observation Period table when standardizing data is inherently challenging.
- Significant discrepancies exist between EHR (encounter-based) and claims data (enrollment-based) in current methods for defining or inferring observation periods.
- This study tests various definitions of observation periods to determine their impact on evidence generated from real-world health data.

METHODS

- This study includes 11 databases standardized to the OMOP Common Data Model (CDM) version 5.3 or higher, including six enrollment-based databases (Mimvie*, MarketScan Commercial Database, Market* MarketScan* HMO State Medicaid Database, Market* MarketScan* Medicare Database, Optum's Clinformatics* Considered Data Mart, JMC, and IQVIA* Adjusted Health Plan Claims Data.) and five encounter-based databases (Upstream* Co-located Red Electronic Health Records, IQVIA* Longitudinal Patient Database in Australia, IQVIA* Disease Analyzer France, IQVIA* Disease Analyzer Germany, and Premier).
- We replicate the study Characterizing the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study by Li et al in the 11 databases using the following methods for defining the observation periods.
 - Persistence + Surveillance (P+S)** - A set number of days strings health events together into eras of persistent observation. A surveillance window is added to denote the time when the person is still considered under observation.



Figure 1: Visual representation of the persistence and surveillance approach


- Age + Gender** - Dynamic P+S windows are applied based on the persons' age and gender at the time of each health encounter.
 - Min/Max** - The earliest and latest events are used to define the start and end of observable time.
 - Enrollment time (enrollment-based databases only)** - Used as gold standard
- The methods a, b, and c are applied to the health encounters in all databases across two groups: **medical events only**, and **medical + pharmacy events (all events)**.
 - We calculate the mean squared error (MSE) between the IIRs from the different observation period definitions and those generated using enrollment time in the enrollment-based databases.

RESULTS

- The persistence + surveillance method showed a decreasing trend in IIRs across both enrollment- and encounter-based databases (Figures 2-4).
- There is a clear inverse relationship between persistence and surveillance observed in Figure 2 when the IIRs are compared with those generated using enrollment time.
- The IIRs for encounter-based databases are highly dependent on the total amount of follow-up time available, e.g. 2 years vs. 20 years, as well as the data capture processes.
- The min/max method, shown in Figure 5 as persistence 9999 and surveillance 0, was found to be less reliable when compared to the gold standard.

The definition of the Observation Period has the potential to influence study outcomes, making it a crucial component of transparent and reproducible research.

To improve data fitness-for-use, the observation period should be assessed to ensure that the data capture process during this time supports answering the clinical question reliably.



Take a picture to download the short report

RESULTS CONT.

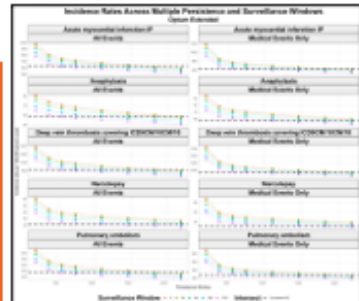


Figure 2: Incidence rates in Optum Extended across multiple persistence and surveillance windows, for outcomes, and two event types

- P+S IR patterns differ based on the clinical context of the encounter-based data, whether it primary includes hospital encounters or primary-care encounter.
- The age + gender approach, while showing some alignment with the gold standard, did not perform as well, particularly for rare outcomes like narcolepsy and anaphylaxis.

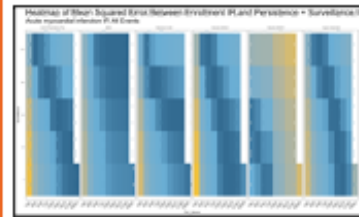


Figure 3: Heatmap of Mean Squared Error Between Enrollment P+S and Persistence + Surveillance IR using all months for the outcome acute myocardial infarction by the min/max approach expressed as persistence 9999 and surveillance 0.

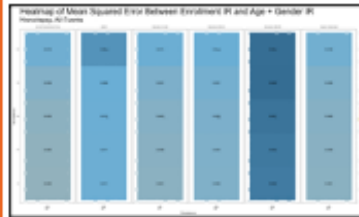


Figure 4: Heatmap of Mean Squared Error Between Enrollment P+S and Age + Gender IR using all months for the outcome acute myocardial infarction.

Clair Blacketer^{1,2}, Patrick Ryan^{1,2}, Peter Rijnbeek^{1,2}, Martijn Schuemie^{1,2}
¹University Research & Development, Statens Serum Institut, Department of Medical Informatics, Denmark
²Department of Biomedical Informatics, Columbia University, New York, NY, USA, Department of Biostatistics, University of California, Los Angeles, CA, USA





#OHDSISocialShowcase This Week

Thursday

How Often: Large Scale Incidence Rate Calculation of Health Outcomes for Drugs Nested by Indication

(Hsin Yi Chen, Christopher Knoll, Elise Ruan, Adam Black, Sarah Seager, Patrick Ryan, George Hripcsak)



How Often: Characterizing Heterogeneity in Drug-Outcome Incidence Rate Estimates Attributed to Drug Indication

Hsin Yi Chen, BS¹, Christopher Knoll, BS², Elise Ruan, MD, MPH¹, Adam Black, BA³, Sarah Seager, BA⁴, Patrick Ryan, PhD², George Hripcsak, MD, MS¹
¹Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, ²Janssen Research & Development, Titusville, NJ, ³Odyssey Data Services, Cambridge, MA, ⁴IQVIA Real World Solutions, Cambridge, MA

Background

- Drug labels list potential adverse reactions, but real-world incidence is less understood
- **Why Incidence Rates?** For most drug-reaction combinations, we do not know the attributable risk of the drug for the reaction. In such cases, incidence rates provide an upper bound to the potential impact of the drug and inform the clinician of the likelihood of needing to address the reaction, be it caused by the drug or not. If the incidence rate is low enough, then the reaction may not be of concern
- Heterogeneity in incidence rates estimates can be influenced by factors such as age, sex, calendar time, and differences in databases¹
- An important aspect that has not been thoroughly explored is the influence of drug indication on the incidence rate estimates of adverse events
- Thus, we aim to quantify the sources of variability that exist in incidence rate estimates in observational data, with a specific focus on how drug indication affects incidence rate variability

Methods

- We conducted an observational cohort study using 13 healthcare databases from various countries and populations (e.g., US, Belgium, Italy, Australia, France) in October 2023
- **Study Design:**
 - Target cohorts: first occurrence of drug exposure
 - Outcome cohorts: 73 different outcomes (defined in the OHDSI phenotype library²)
 - Time at risk: 1 day to 365 days after cohort start (Intent to treat analysis)
 - Stratifications: age and gender
- We calculated incidence rates for each drug-outcome pair and performed random-effects meta-analyses to pool results across databases
- For each drug class, we conducted variance components analysis to quantify the magnitude of the incidence rate heterogeneity that can be attributed to drug indications

Contact: hc3292@cumc.columbia.edu

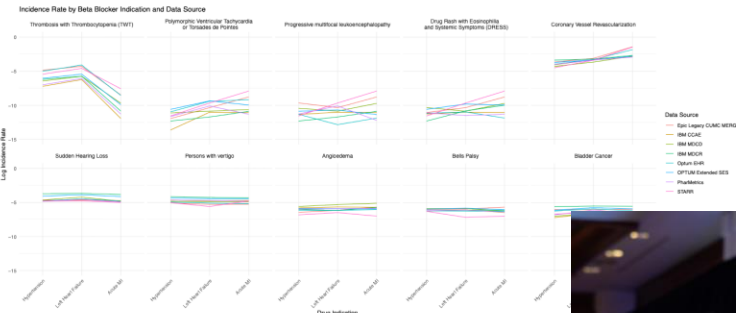


Figure 1. Outcomes with the 5 highest (top row) and 5 lowest (bottom row) indication variance components after adjusting for age, gender, and database. The five outcomes with the highest variance components are the outcomes where indications matter "most" after adjusting for database, age, and gender. Note: we only plot the databases where there is data available.

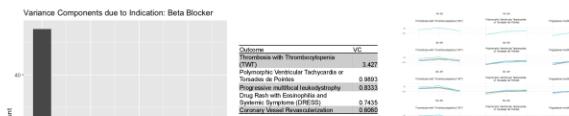


Figure 2. The distribution of indication variance components for beta blockers (adjusting also for age, gender, and database) among all 73 outcomes.

Drug class	Indications	Median VC
Beta Blockers	1) Essential Hypertension, 2) Left Heart Failure, 3) Acute Myocardial Infarction	0.1013
GLP-1 antagonists	1) Type 2 Diabetes Mellitus, 2) Left Heart Failure	0.2642
GLP-1 antagonists	1) Type 2 Diabetes Mellitus, 2) Obesity	<0.001
Cephalosporins	1) Urinary Tract Infection, 2) Acute Typical Pneumonia	0.0397
Fluoroquinolones	1) Urinary Tract Infection, 2) Acute Typical Pneumonia	0.0983
Trimethoprim	1) Urinary tract infection, 2) Pneumonia	0.4887
JAK inhibitors	1) Rhegmatoid Arthritis, 2) Ulcerative Colitis	0.0283 *
TNF-alpha inhibitors	1) Psoriasis, 2) Rheumatoid Arthritis, 3) Ulcerative Colitis, 4) Psoriatic Arthritis, 5) Crohn's Disease	0.0332

Table 1. 8 drug classes in our study had more than 1 indication. Of these 8 drug classes, trimethoprim had the highest heterogeneity measured by median VC (variance component) between indications among outcomes. Variance components are adjusted for age, gender, and indication.

* For JAK inhibitors, we did not adjust for age and gender, as we had <1000 patients at risk in these strata.

Results

- 8 different drug classes had at least 2 indications. After accounting for database, age, and gender differences, median variance components (VC) of incidence rates attributed to different drug indications ranged from 0.0005 (GLP-1 antagonists) to 0.4887 (trimethoprim)
- In contrast, median variance components attributed to age deciles ranged from 0.1815 (TNF-alpha inhibitors) to 1.52 (trimethoprim), and median VC attributed to sex ranged from 0.0155 (SGLT-2 inhibitors) to 0.1219 (trimethoprim)



Figure 3. The top 5 "most different" outcomes (bottom) and age (top). The difference between stratifications. Note: we only plot the databases stratified by age (top) and sex (bottom).



#OHDSISocialShowcase This Week

Friday

FinOMOP Swarm Learning - Deep learning for patient-specific modelling of Acute Myeloid Leukemia

(**Salma Rachidi**, Hartmut Schultze, Vytis Vadoklis, Perre Gustafsson, Johansson Markus, Kauko Tommi, Anna Hammals, Kukkurainen Sampo, Niemelä Sami, Tuomas Hakala, Alexey Ryzhenkov, Valteri Nieminen, Tomi Mäkelä, Oscar Brück, Joachim Schultze, Tarja Laitinen, Arho Virkki, Kimmo Porkka, Eric Fey)

FinOMOP Swarm Learning:
Deep learning for patient-specific modelling of Acute Myeloid Leukemia based on OMOP

PRESENTER: Eric Fey

INTRO:

- Deep learning is powerful but data hungry!
- Data sharing hindered by privacy concerns
- Federated learning hindered by heterogeneous data formats
- **Solution: Use OMOP and learn models at the edge without sharing source data using Swarm learning (SL)**

METHODS

1. OMOP + Swarm Learning => Joint model



2. Cohort: AML use case

All patients with AML diagnoses and at least 3 blast measurements within 21 days of diagnosis

3. Data

Endpoint: Overall patient survival
Features: Blood count measurements, timepoints up to 21 days after diagnosis

4. Model architecture v1:



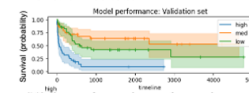
RESULTS

Developed FinOMOP-SL framework:

- GDPR & Findata compliant implementation at three sites: Helsinki, Turku, Tampere
- Integrating OMOP & swarm learning
- Audited by national data authority Findata (<https://findata.fi/en/>)
- FinOMOP swarm network established

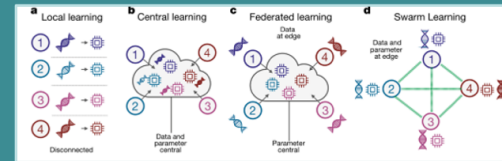
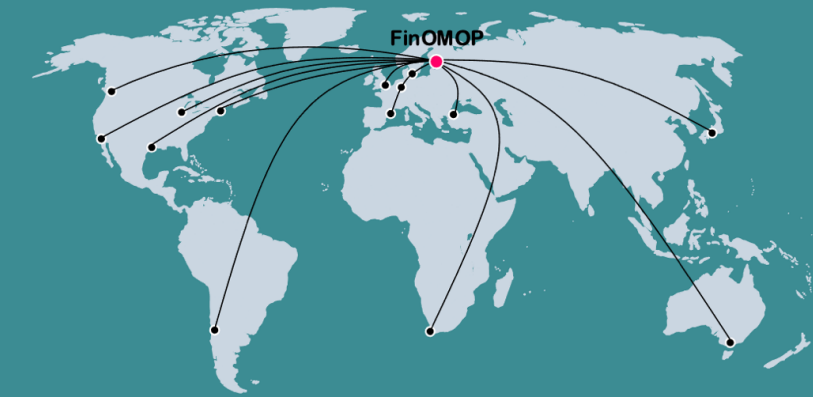
Proof of concept use case AML

- Framework for longitudinal & missing data
- Predictive power model v1, so far only trained on HUS data:



	High	Mid	Low
AT risk	2	2	0
Censored	2	2	0
Events	28	28	28
AT risk	15	9	2
Censored	11	11	11
Events	11	11	11
AT risk	10	5	2
Censored	16	11	11
Events	18	18	18

Build deep, predictive models for precision medicine together in global networks.



- Limited dataset
- Bias
- Low accuracy
- Disconnected

- + Enlarged dataset
- + Better accuracy
- Aggregation
- Monopolisation

- + Enlarged dataset
- + Better accuracy
- + No data movement
- + No central custodian
- + Shared insights

AMMO BAR

CONCLUSIONS

- Routine short-term follow up data - blood count measurements (21 days) - can predict long-term prognosis in AML
- Powerful OMOP-based framework for federated training of predictive models
- General applicability: Treatment responses, survival, adverse events, optimal treatments, ...

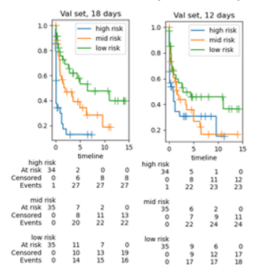
Next:

- **Go global! Contact Eric if interested.**
- Estimate and include pCR & RFS
- Integrate genomic data -> see poster #9
- Incorporate traditional diagnostic procedures and risk classification

Data availability HUS & VARHA (Turku)



Performance based on timepoint availability



Eric Fey^{1,2} (eric.fey@hus.fi), Salma Rachidi¹, Alexey Ryzhenkov¹, Valteri Nieminen¹, Tomi Mäkelä¹, Oscar Brück^{1,3}, Kimmo Porkka^{1,3}
¹HUS Helsinki University Hospital, Finland
²ICAN Digital Precision Cancer Medicine Flagship, University of Helsinki, Finland





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?

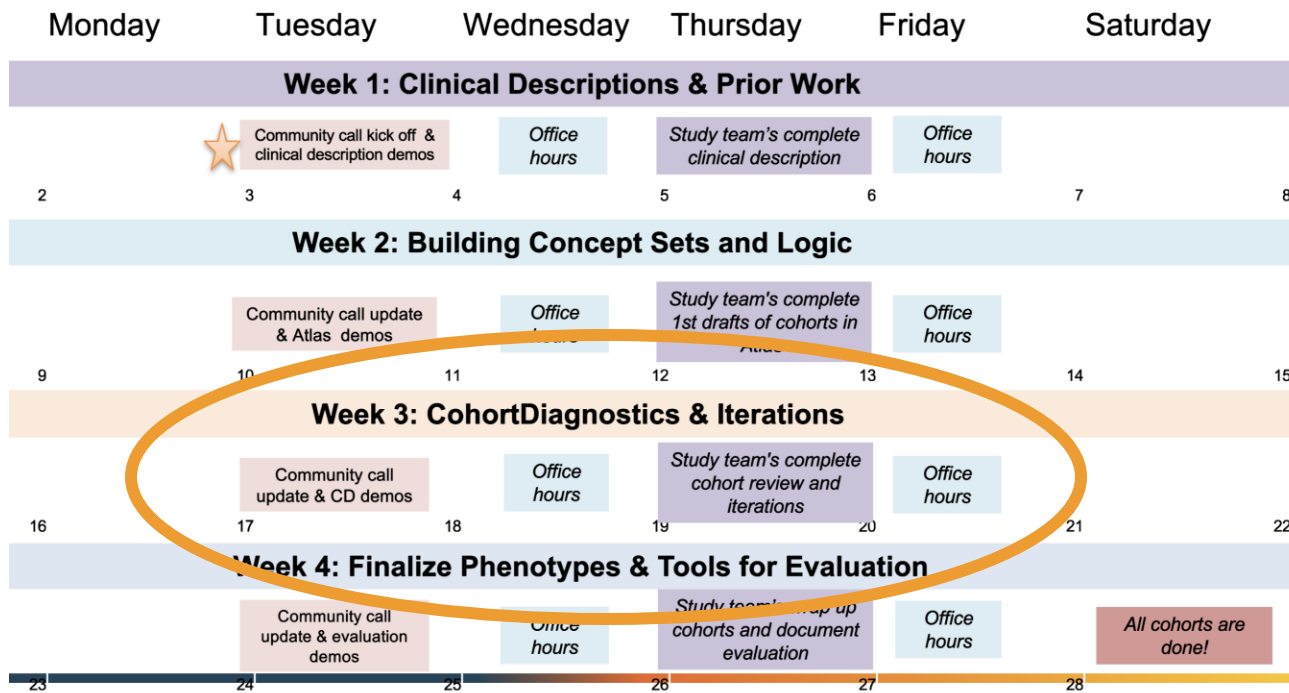
Where Are We Going?





Feb. 18 Community Call

Phenotype February 2025 Calendar



Workgroup OKRs:

Databricks

Dentistry

Eye Care and Vision Research

GenAI

HADES

Latin America

Medical Imaging

OHDSI APAC

Psychiatry

Transplant



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