Workgroup Updates + Phenotype Phebruary Report

OHDSI Community Call
Feb. 8, 2022 • 11 am ET

Future OHDSI Community Calls

Date	Topic
Feb. 8	Workgroup Updates (Healthcare Systems, Open Source Community), Phenotype Phebruary Report
Feb. 15	Workgroup Updates (Common Data Model, Data Quality), Phenotype Phebruary Report
Feb. 22	Workgroup Updates (ATLAS/WebAPI, Medical Imaging), Phenotype Phebruary Report
Mar. 1	Breakout Sessions (Characterization, Estimation, Prediction)
Mar. 8	CDM Workshop (Part 1)
Mar. 15	CDM Workshop (Part 2)
Mar. 22	OHDSI Vocabulary Journey
Mar. 29	Reproducibility

Future OHDSI Community Calls

Date	Topic
Feb. 8	Workgroup Updates (Healthcare Systems, Open Source Community), Phenotype Phebruary Report
Feb. 15	Workgroup Updates (Common Data Model, Data Quality), Phenotype Phebruary Report
Feb. 22	Workgroup Updates (ATLAS/WebAPI, Medical Imaging), Phenotype Phebruary Report
Mar. 1	Breakout Sessions (Characterization, Estimation, Prediction)
Mar. 8	CDM Workshop (Part 1)
Mar. 15	CDM Workshop (Part 2)
Mar. 22	OHDSI Vocabulary Journey
Mar. 29	Reproducibility

February 15 OHDSI Community Call

Common Data Model Workgroup Update

Clair Blacketer

Data Quality Dashboard Workgroup Update

Clair Blacketer

Phenotype Phebruary Update #2

Patrick Ryan

Three Stages of The Journey

Where Have We Been?

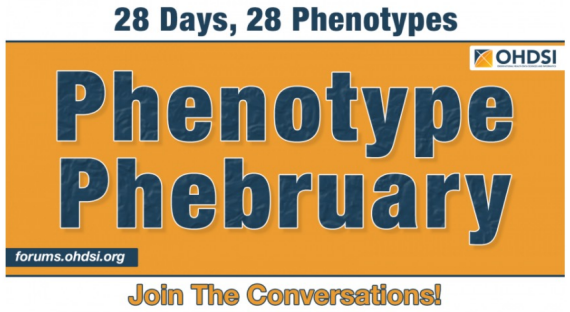
Where Are We Now?

Where Are We Going?

Phenotype Phebruary

Phenotype Phebruary Daily Updates

"Phenotype Phebruary" is a community-wide initiative to both develop and evaluate phenotypes for health outcomes that could be investigated by the community. Patrick Ryan introduced this initiative in both a video presentation and a forum post, and each of the conversations around the "28 phenotypes for 28 days" are being held within the OHDSI forums.



This page will provide direct links to each forum post, which is where conversations around each specific phenotype should be held.

Please be active in these discussions. What ways can you contribute?

1. Join the conversation

- Discussions will be here on forums.ohdsi.org
- Each day will be a new thread
  – Ex: Look for: "Phenotype Phebruary Day 1 – Type 2 diabetes mellitus"
- Explore the definitions and review the results provided
- Reply with your thoughts, reflections, insights and question

2. Evaluate the cohort definitions in your data

- Execute cohort definitions and CohortDiagnostics in your CDM
- Share insights you learn from your data on the forums
- Share results to compile across the network on data.ohdsi.org

3. Lead a discussion

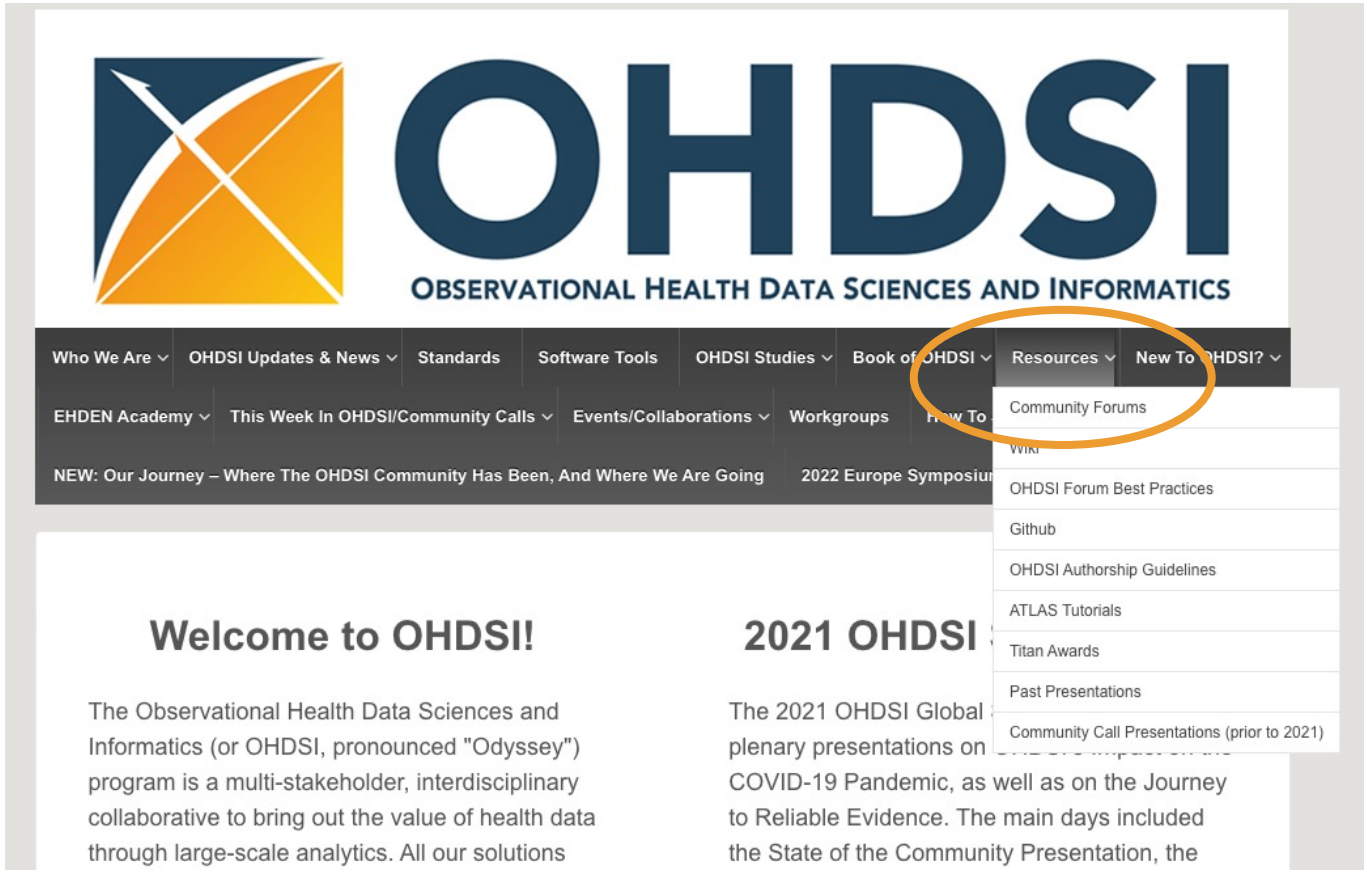
- Patrick will be leading the discussion for the first 7 days, but if others would like to similarly lead a phenotype development and evaluation activity, contact ryan@ohdsi.org or chat with him in OHDSI MSTeams, tell me your desired phenotype target and calendar date you want to commit to.

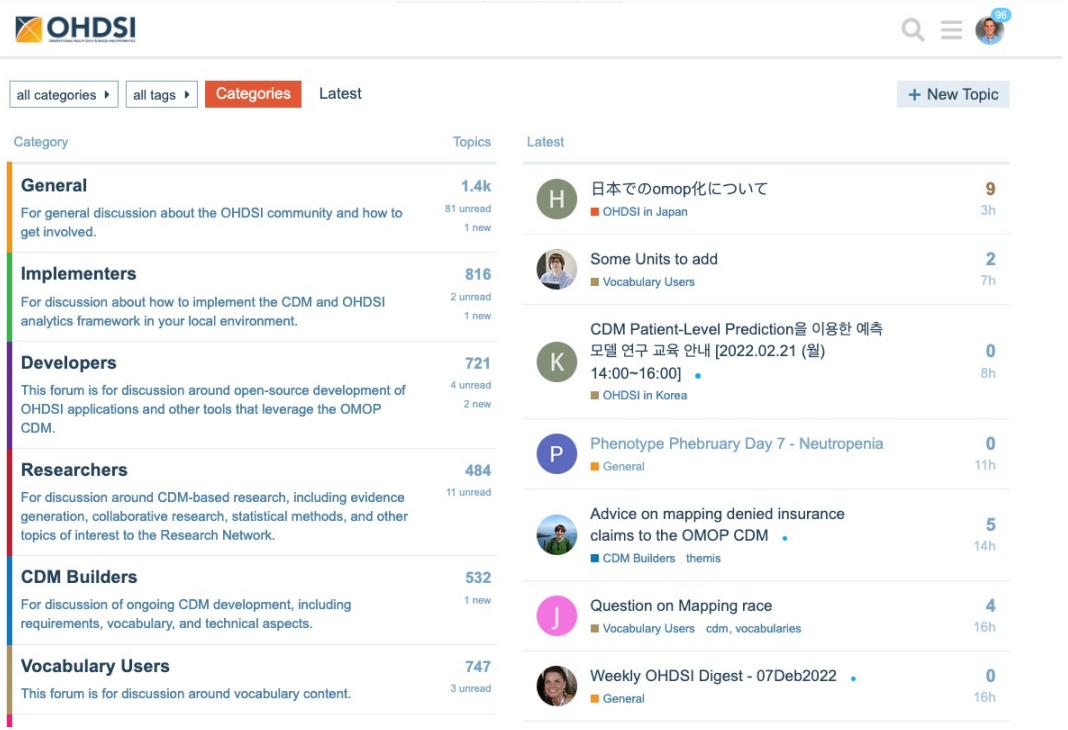
Daily Phenotype February Links

Feb. 1 • Type 2 Diabetes Mellitus
Feb. 2 • Type 1 Diabetes Mellitus
Feb. 3 • Atrial Fibrillation
Feb. 4 • Multiple Myeloma
Feb. 5 • Alzheimer's Disease
Feb. 6 • Hemorrhagic Events
Feb. 7 • Neuropenia
Feb. 8 •
Feb. 9 •
Feb. 10 •
Feb. 11 •
Feb. 12 •
Feb. 13 •
Feb. 14 •
Feb. 15 •
Feb. 16 •
Feb. 17 •
Feb. 18 •
Feb. 19 •
Feb. 20 •
Feb. 21 •
Feb. 22 •
Feb. 23 •
Feb. 24 •
Feb. 25 •
Feb. 26 •
Feb. 27 •
Feb. 28 •

https://www.ohdsi.org/phenotype-phebruary

Navigating OHDSI.org





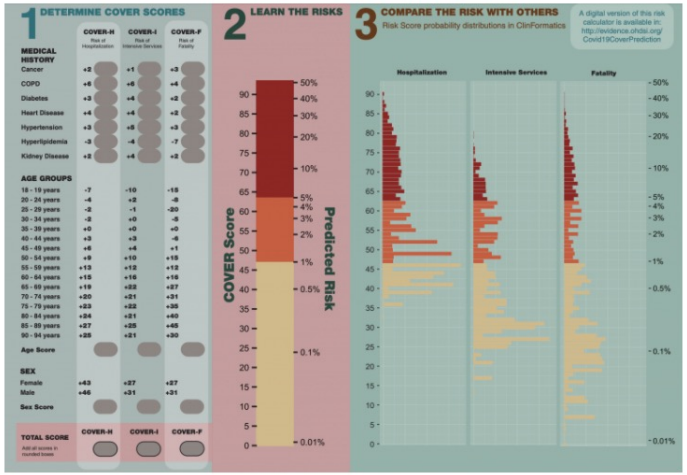
forums.ohdsi.org

COVER Model Insights

COVER Prediction Model Generated During COVID Study-A-Thon Published; Lead Authors Share Thoughts On Model, Impact

1) How were you able to develop a prediction model so early in the pandemic with such little data?

The amount of data needed to evaluate model performance reliably is much less than the amount needed to train a model. Early on in the pandemic we quickly reached the level needed for model evaluation, but model development would have been more problematic. Therefore, we decided to use a proxy disease (influenza) to preserve the COVID-19 data that we had available. Our assumption was that the people vulnerable to influenza would have similar characteristics as those vulnerable to COVID-19. The large amount of historic influenza cases allowed us to overcome the issues of model development with small data samples. After model training we evaluated the model on data from COVID-19 patients to evaluate model performance reliably.

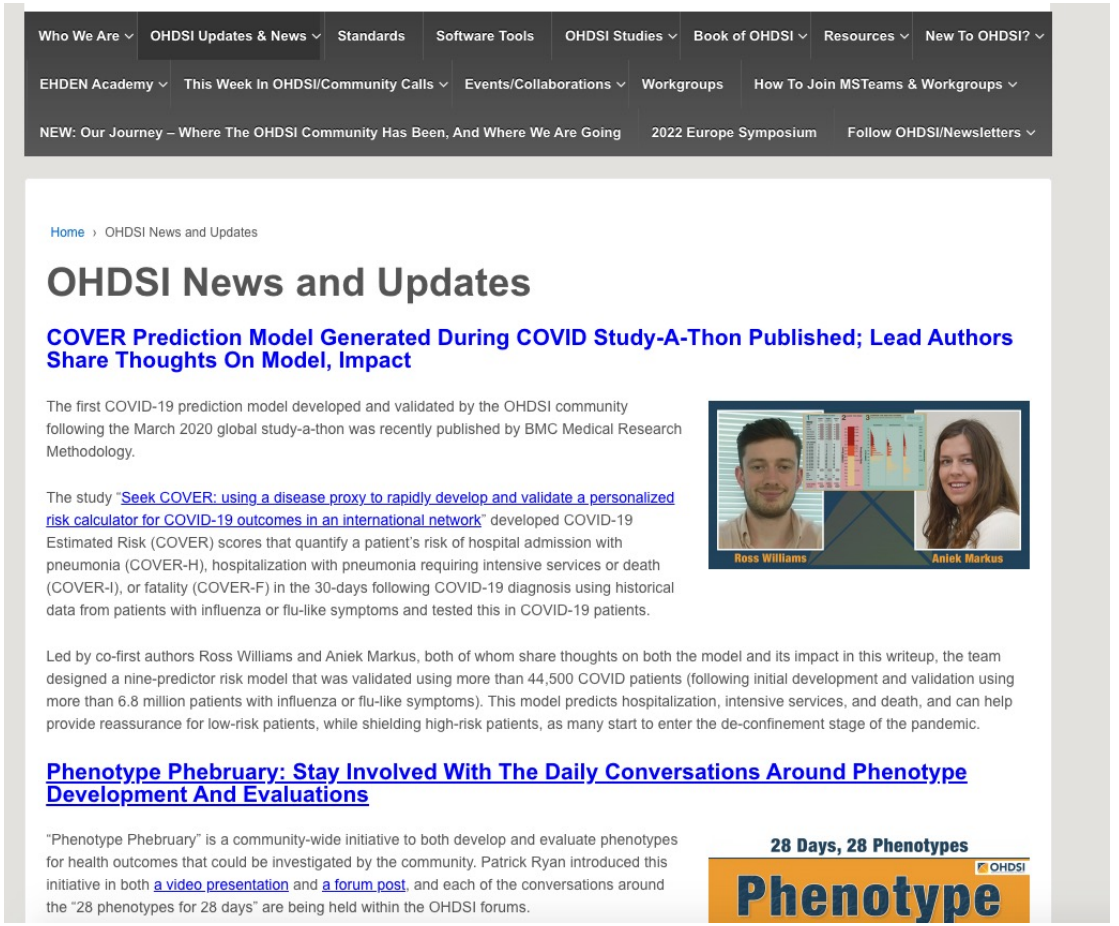


2) When the model was shared via preprint, are you aware of how it was used and what impact it had?

The COVER scores were used for strategic planning purposes by hospitals and regional governments as well as for risk assessment purposes by institutions planning their office work strategies.

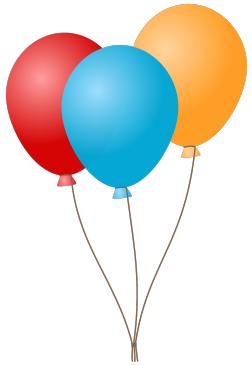
https://www.ohdsi.org/cover-prediction-model

Navigating OHDSI.org



www.ohdsi.org/ohdsi-news-updates/

OHDSI Shoutouts!



Any shoutouts from the community? Please share and help promote and celebrate OHDSI work!

Have a study published? Please send to sachson@ohdsi.org so we can share during this call and on our social channels.
Let's work together to promote the collaborative work happening in OHDSI!

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	Common Data Model Vocabulary Subgroup
Tuesday	2 pm	Health Equity
Wednesday	12 pm	FHIR and OMOP Terminologies Subgroup (Zoom)
Wednesday	2 pm	Natural Language Processing
Wednesday	7 pm	Medical Imaging
Thursday	10 am	Data Quality Dashboard
Friday	TBA	Education
Friday	10 am	Phenotype Development and Evaluation
Friday	11 pm	China Chapter
Monday	8 am	Early-Stage Researchers (Europe/Western Hemisphere)
Monday	10 am	Healthcare Special Interest Group

www.ohdsi.org/upcoming-working-group-calls

Get Access To Different Teams/WGs/Chapters

5. Select the workgroups you want to join (you can refer to the WIKI for work group objectives www.ohdsi.org/web/wiki/doku.php?id=projects:overview)

- [ ] ATLAS
- [ ] Clinical Trials
- [ ] Common Data Model
- [ ] Data Quality Dashboard Development
- [ ] Early-stage Researchers
- [ ] Education Work Group
- [ ] FHIR and OMOP
- [ ] Geographic Information System (GIS)
- [ ] HADES Health Analytics Data-to-Evidence Suite
- [ ] Healthcare Systems Interest Group (formerly EHR)
- [ ] Health Equity
- [ ] Latin America
- [ ] Medical Devices
- [ ] Medical Imaging
- [ ] Natural Language Processing
- [ ] OHDSI APAC
- [ ] OHDSI APAC Steering Committee
- [ ] OHDSI Steering Committee
- [ ] Oncology
- [ ] Open-source Community
- [ ] Phenotype Development and Evaluation
- [ ] Population-Level Effect Estimation / Patient-Level Prediction
- [ ] Psychiatry
- [ ] Registry (formerly UK Biobank)
- [ ] Surgery and Perioperative Medicine
- [ ] Vaccine Evidence
- [ ] Vaccine Vocabulary

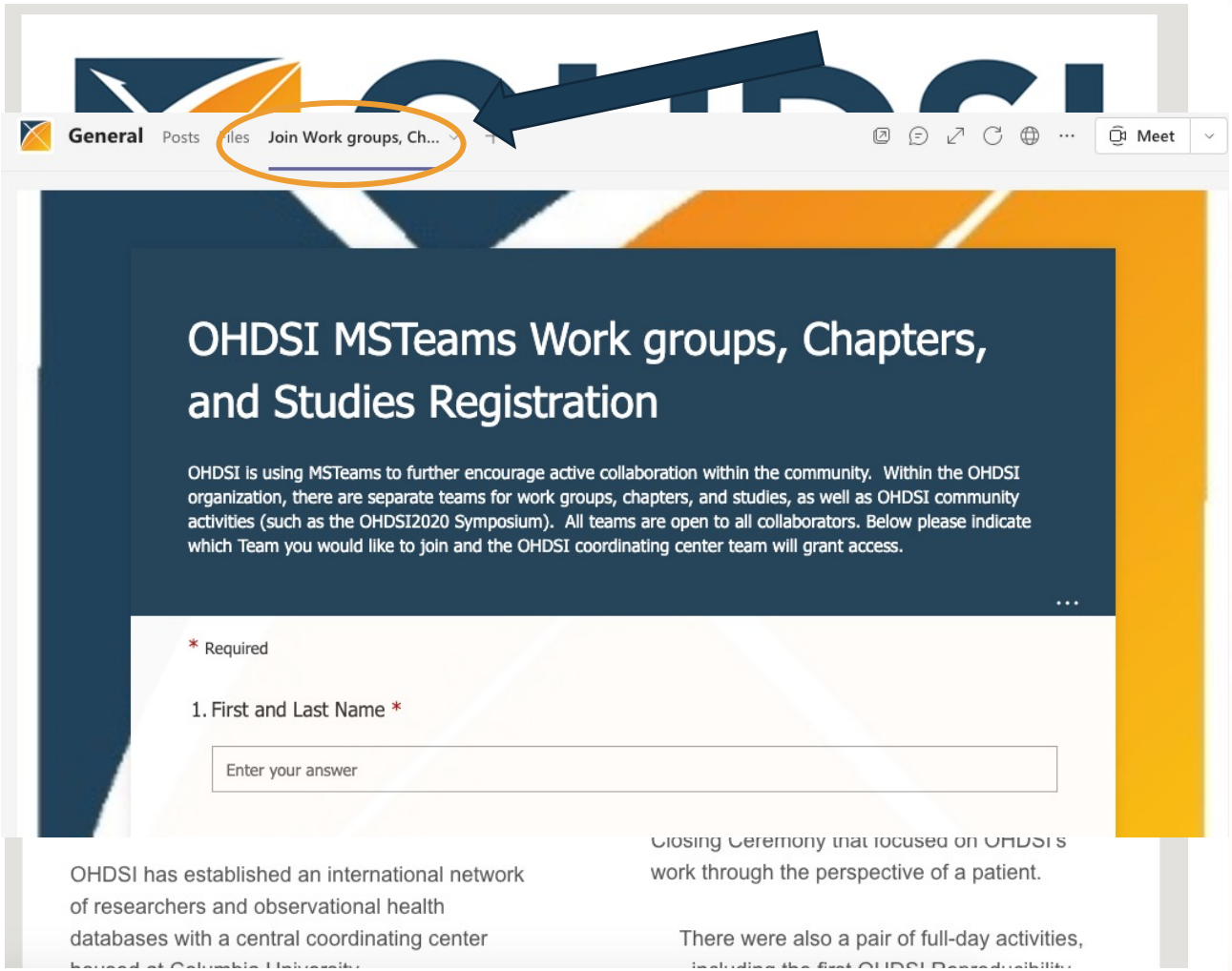
6. Select the chapter(s) you want to join

- [ ] Africa
- [ ] Australia
- [ ] China
- [ ] Europe
- [ ] Japan
- [ ] Korea
- [ ] Singapore
- [ ] Taiwan

7. Select the studies you want to join

- [ ] HERA-Health Equity Research Assessment
- [ ] PIONEER for Prostate Cancer (study-a-thon ended)
- [ ] SCYLLA (SARS-Cov-2 Large-scale Longitudinal Analyses)

Get Access To Different Teams/WGs/Chapters



5. Select the workgroups you want to join (you can refer to the WIKI for work group objectives www.ohdsi.org/web/wiki/doku.php?id=projects:overview)

- ☐ ATLAS
- ☐ Clinical Trials
- ☐ Common Data Model
- ☐ Data Quality Dashboard Development
- ☐ Early-stage Researchers
- ☐ Education Work Group
- ☐ FHIR and OMOP
- ☐ Geographic Information System (GIS)
- ☐ HADES Health Analytics Data-to-Evidence Suite
- ☐ Healthcare Systems Interest Group (formerly EHR)
- ☐ Health Equity
- ☐ Latin America
- ☐ Medical Devices
- ☐ Medical Imaging
- ☐ Natural Language Processing
- ☐ OHDSI APAC
- ☐ OHDSI APAC Steering Committee
- ☐ OHDSI Steering Committee
- ☐ Oncology
- ☐ Open-source Community
- ☐ Phenotype Development and Evaluation
- ☐ Population-Level Effect Estimation / Patient-Level Prediction
- ☐ Psychiatry
- ☐ Registry (formerly UK Biobank)
- ☐ Surgery and Perioperative Medicine
- ☐ Vaccine Evidence
- ☐ Vaccine Vocabulary

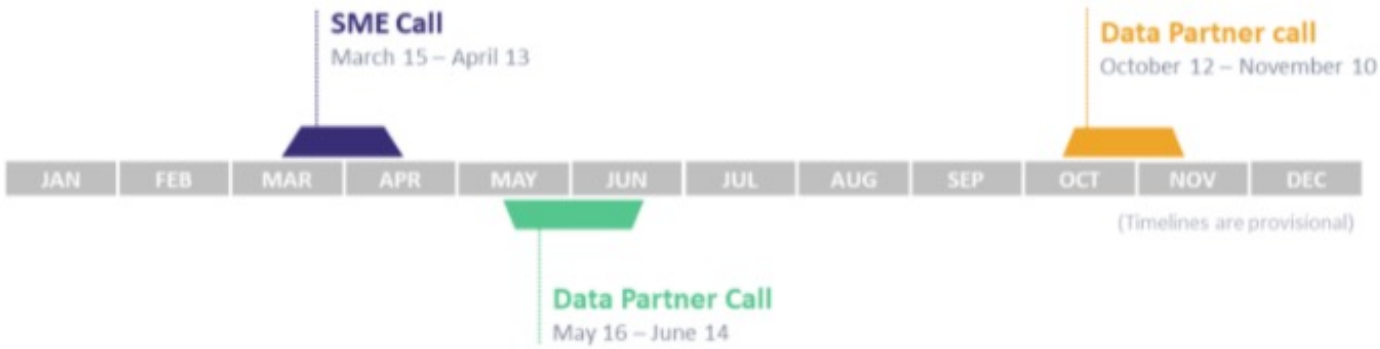
6. Select the chapter(s) you want to join

- ☐ Africa
- ☐ Australia
- ☐ China
- ☐ Europe
- ☐ Japan
- ☐ Korea
- ☐ Singapore
- ☐ Taiwan

7. Select the studies you want to join

- ☐ HERA-Health Equity Research Assessment
- ☐ PIONEER for Prostate Cancer (study-a-thon ended)
- ☐ SCYLLA (SARS-Cov-2 Large-scale Longitudinal Analyses)

EHDEN Data Partner/SME Calls for 2022



We are happy to share that throughout 2022, we will host 3 open calls:

- An open call for SMEs running from March 15 to April 13
- An open call for Data Partner running from May 16 to June 14
- A second open call for Data Partners running from October 12 to November 10

These timelines are provisional and more detailed information will become available closer to the opening of these calls. A general description of each of the calls is available via the **Open calls for SMEs** page and the **Open calls for Data Partners** page. More detailed information on each call will become available closer to the opening of the calls.

Ehden.eu

EHDEN Data Partner/SME Calls for 2022

Latest Edition of The Journey Newsletter

Community Updates

Where Have We Been

• **Hongfang Liu** and **Christopher Chute** led a presentation on "Extracting OHDSI Concepts from Clinical Narratives for COVID" during the Jan. 25 OHDSI Community Call. The full session, which included Q&A from the 240+ attendees, is available below, along with the slides from the presentation.

• **Patrick Ryan** opened the year with a discussion about "Where Can OHDSI Go In 2022," which focused on goals at both the community and workgroup levels. One of the potential focal points he discussed was engineering open science systems that build trust into the real-world evidence generation and dissemination process. The entire video presentation is available below.

Where Are We Now

• Today starts "Phenotype Phebruary," which will be a multi-platform community initiative to highlight the importance of, and develop at least 28 new, phenotypes. A detailed preview of the discussion will take place during the Feb. 1 community call; if you miss it live, please catch the recording at our Community Calls page, and then join our community in both our MSTeams environment and the OHDSI forums in this work.

• Several workgroups will detail their 2022 objectives and key results during the Tuesday community calls this month, while workgroup leads will connect in a leadership summit during the month to discuss best practices and drive greater collaboration and efficiency. You can learn more about each at our new workgroups page and request to join any of the workgroups or chapters.

Where Are We Going

• The first in-person OHDSI event since the start of the pandemic will take place June 24-26 during the OHDSI European Symposium. This will take place on the Steam Ship Rotterdam in the Netherlands, with the main symposium set for Friday, June 24, and two days of workshops and tutorials to follow. For more information or to register, please visit the symposium homepage.

OHDSI

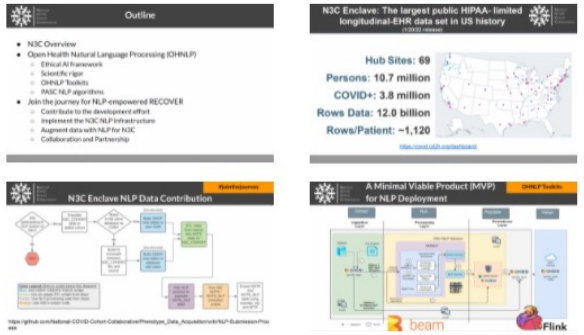
The Journey Newsletter (February 2022)

The OHDSI community is off and running in 2022. We had a terrific presentation on "Extracting OHDSI Concepts from Clinical Narratives for COVID" from our colleagues within the N3C, and we discussed 2022 goal-setting, both at the global level, as well as within individual workgroups. Several studies relating to OHDSI or OMOP were published. Check it all out in the latest edition of The Journey! #JoinTheJourney

February Update Podcast

Patrick Ryan and Craig Sachson discuss goal-setting for both the global community and individual workgroups for 2022, the N3C-led presentation on Extracting OHDSI Concepts from Clinical Narratives for COVID, and the upcoming concept of "Phenotype Phebruary," which will kick off during the Feb. 1 OHDSI Community Call.

Presentation: Extracting OHDSI Concepts from Clinical Narratives for COVID



During our Jan. 18 community call, Dr. Hongfang Liu (Mayo Clinic) and Dr. Christopher Chute (Johns Hopkins University) led a session on Extracting OHDSI Concepts from Clinical Narratives for COVID. Following the presentation (approximately 33 minutes), there is a Q&A session. You can access both the presentation and the slides below.

Video Presentation

Slides

Publications

Paris N, Lamer A, Parrot A. Transformation and Evaluation of the MIMIC Database in the OMOP Common Data Model: Development and Usability Study. JMIR Med Inform 2021;9(12):e30970. doi: 10.2196/30970. PMID: 34904958

Choi, S., Choi, S.J., Kim, J.K. *et al.* Preliminary feasibility assessment of CDM-based active surveillance using current status of medical device data in medical records and OMOP-CDM. Sci Rep 11, 24070 (2021). https://doi.org/10.1038/s41598-021-03332-6

Nestsiarovich, A., Reps, J.M., Matheny, M.E. *et al.* Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study. Transl Psychiatry 11, 642 (2021). https://doi.org/10.1038/s41398-021-01760-6

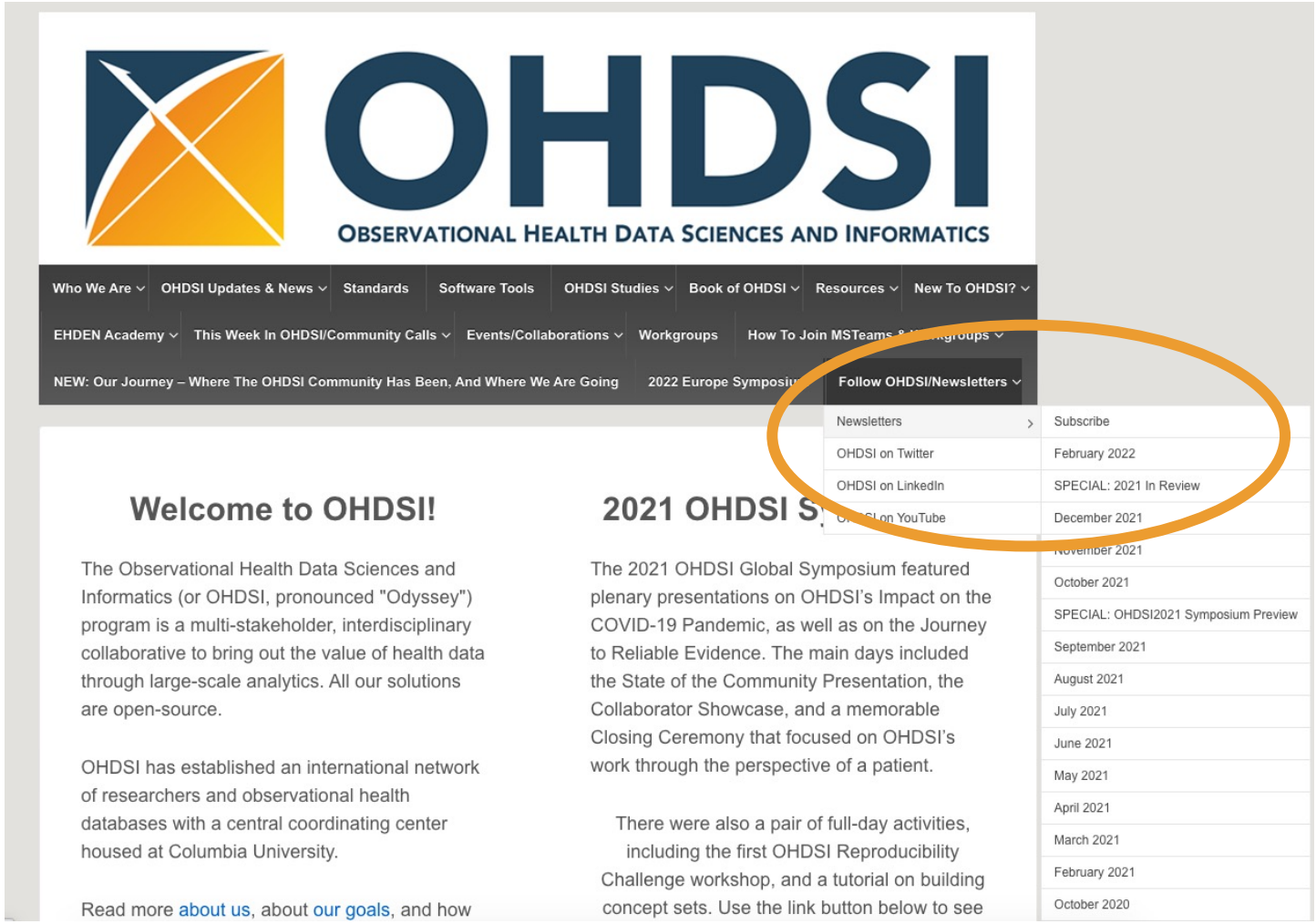
Reyes C, Pistillo A, Fernández-Bertolín S, et al. Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study. BMJ Open 2021;11:e057632. doi: 10.1136/bmjopen-2021-057632

Reps JM, Ryan P, Rijnbeek PR. Investigating the impact of development and internal validation design when training prognostic models using a retrospective cohort in big US observational healthcare data. BMJ Open 2021;11:e050146. doi: 10.1136/bmjopen-2021-050146

Daniel Morales, Anna Ostropolets, Lana Lai, Anthony Sena, Scott Duvall, Marc Suchard, Katia Verhamme, Peter Rijnbeek, Joe Posada, Waheed Ahmed, Thamer Alshammary, Heba Alghoul, Osaid Alser, Carlos Areia, Clair Blacketer, Ed Burn, Paula Casajust, Seng You, Dalia Dawoud, Asieh Golozar, Menchung Gong, Jitendra Jonnagaddala, Kristine Lynch, Michael Matheny, Evan Minty, Fredrik Nyberg, Albert Uribe, Martina Recalde, Christian Reich, Martijn Scheumie, Karishma Shah, Nigam Shah, Lisa Schilling, David Vizcaya, Lin Zhang, George Hripcsak, Patrick Ryan, Daniel Prieto-Alhambra, Talita Durate-Salles & Kristin Kostka (2022). Characteristics and outcomes of COVID-19 patients with and without asthma from the United States, South Korea, and Europe, Journal of Asthma, DOI: 10.1080/02770903.2021.2025392

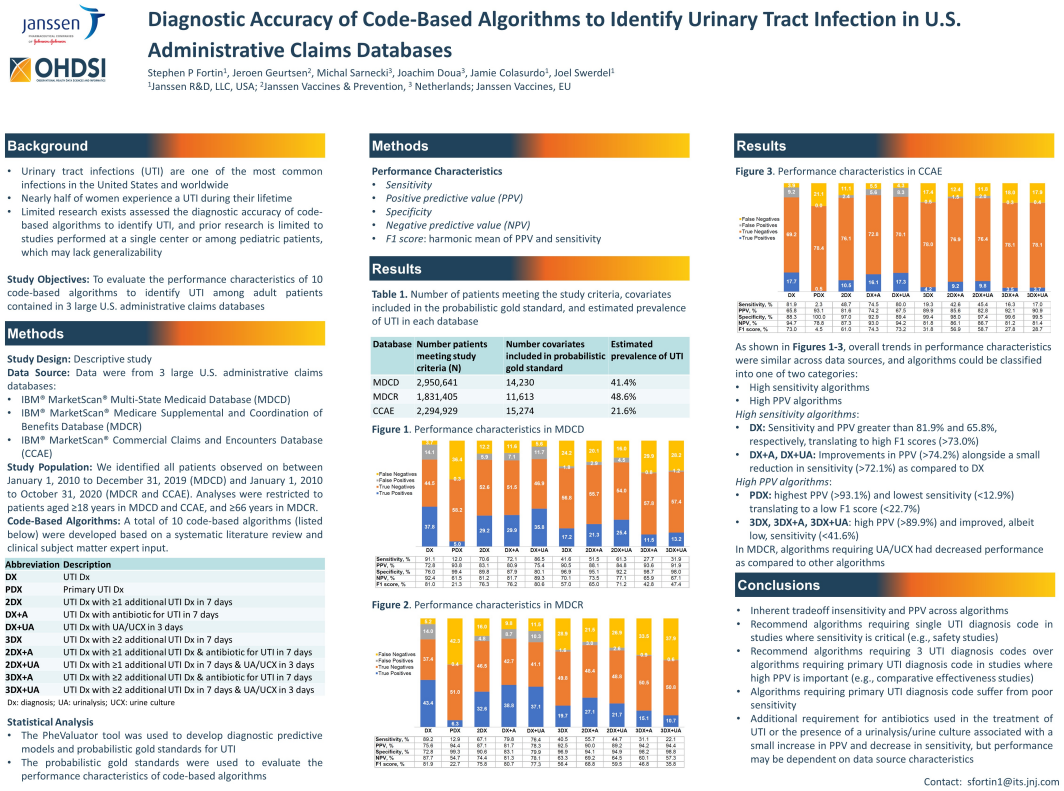
Ji Xiangmin, Cui Guimei, Xu Chengzhen, Hou Jie, Zhang Yunfei, Ren Yan. Combining a Pharmacological Network Model with a Bayesian Signal Detection Algorithm to Improve the Detection of Adverse Drug Events. Frontiers in Pharmacology, 2022. DOI: 10.3389/fphar.2021.773135.

Subscribe Or Find the OHDSI Newsletters



https://ohdsi.org/subscribe-to-our-newsletter

#OHDSISocialShowcase This Week



Diagnostic Accuracy of Code-Based Algorithms to Identify Urinary Tract Infection in U.S. Administrative Claims Databases

Stephen P Fortin[1], Jeroen Geurtsen[2], Michal Sarnecki[3], Joachim Doua[3], Jamie Colasurdo[1], Joel Swerdel[1]
[1]Janssen R&D, LLC, USA; [2]Janssen Vaccines & Prevention, [3] Netherlands; Janssen Vaccines, EU

Background

- Urinary tract infections (UTI) are one of the most common infections in the United States and worldwide
- Nearly half of women experience a UTI during their lifetime
- Limited research exists assessed the diagnostic accuracy of code-based algorithms to identify UTI, and prior research is limited to studies performed at a single center or among pediatric patients, which may lack generalizability

Study Objectives: To evaluate the performance characteristics of 10 code-based algorithms to identify UTI among adult patients contained in 3 large U.S. administrative claims databases

Methods

Study Design: Descriptive study
Data Source: Data were from 3 large U.S. administrative claims databases:

- IBM® MarketScan® Multi-State Medicaid Database (MDCD)
- IBM® MarketScan® Medicare Supplemental and Coordination of Benefits Database (MDCR)
- IBM® MarketScan® Commercial Claims and Encounters Database (CCAE)

Study Population: We identified all patients observed on between January 1, 2010 to December 31, 2019 (MDCD) and January 1, 2010 to October 31, 2020 (MDCR and CCAE). Analyses were restricted to patients aged ≥18 years in MDCD and CCAE, and ≥66 years in MDCR.
Code-Based Algorithms: A total of 10 code-based algorithms (listed below) were developed based on a systematic literature review and clinical subject matter expert input.

Abbreviation	Description
DX	UTI Dx
PDX	Primary UTI Dx
2DX	UTI Dx with ≥1 additional UTI Dx in 7 days
DX+A	UTI Dx with antibiotic for UTI in 7 days
DX+UA	UTI Dx with UA/UCX in 3 days
3DX	UTI Dx with ≥2 additional UTI Dx in 7 days
2DX+A	UTI Dx with ≥1 additional UTI Dx & antibiotic for UTI in 7 days
2DX+UA	UTI Dx with ≥1 additional UTI Dx in 7 days & UA/UCX in 3 days
3DX+A	UTI Dx with ≥2 additional UTI Dx & antibiotic for UTI in 7 days
3DX+UA	UTI Dx with ≥2 additional UTI Dx in 7 days & UA/UCX in 3 days

Dx: diagnosis; UA: urinalysis; UCX: urine culture

Statistical Analysis

- The PheValuator tool was used to develop diagnostic predictive models and probabilistic gold standards for UTI
- The probabilistic gold standards were used to evaluate the performance characteristics of code-based algorithms

Methods

Performance Characteristics

- *Sensitivity*
- *Positive predictive value (PPV)*
- *Specificity*
- *Negative predictive value (NPV)*
- *F1 score*: harmonic mean of PPV and sensitivity

Results

Table 1. Number of patients meeting the study criteria, covariates included in the probabilistic gold standard, and estimated prevalence of UTI in each database

Database	Number patients meeting study criteria (N)	Number covariates included in probabilistic gold standard	Estimated prevalence of UTI
MDCD	2,950,641	14,230	41.4%
MDCR	1,831,405	11,613	48.6%
CCAE	2,294,929	15,274	21.6%

Figure 1. Performance characteristics in MDCD

Figure 2. Performance characteristics in MDCR

Results

Figure 3. Performance characteristics in CCAE

As shown in **Figures 1-3**, overall trends in performance characteristics were similar across data sources, and algorithms could be classified into one of two categories:

- High sensitivity algorithms
- High PPV algorithms

High sensitivity algorithms:

- **DX:** Sensitivity and PPV greater than 81.9% and 65.8%, respectively, translating to high F1 scores (>73.0%)
- **DX+A, DX+UA:** Improvements in PPV (>74.2%) alongside a small reduction in sensitivity (>72.1%) as compared to DX

High PPV algorithms:

- **PDX:** highest PPV (>93.1%) and lowest sensitivity (<12.9%) translating to a low F1 score (<22.7%)
- **3DX, 3DX+A, 3DX+UA**: high PPV (>89.9%) and improved, albeit low, sensitivity (<41.6%)

In MDCR, algorithms requiring UA/UCX had decreased performance as compared to other algorithms

Conclusions

- Inherent tradeoff insensitivity and PPV across algorithms
- Recommend algorithms requiring single UTI diagnosis code in studies where sensitivity is critical (e.g., safety studies)
- Recommend algorithms requiring 3 UTI diagnosis codes over algorithms requiring primary UTI diagnosis code in studies where high PPV is important (e.g., comparative effectiveness studies)
- Algorithms requiring primary UTI diagnosis code suffer from poor sensitivity
- Additional requirement for antibiotics used in the treatment of UTI or the presence of a urinalysis/urine culture associated with a small increase in PPV and decrease in sensitivity, but performance may be dependent on data source characteristics

Contact: sfortin1@its.jnj.com

MONDAY

Diagnostic Accuracy of Code-Based Algorithms to Identify Urinary Tract Infection in U.S. Administrative Claims Databases

Authors: Stephen P Fortin, Jeroen Geurtsen, Michal Sarnecki, Joachim Doua, Jamie Colasurdo, Joel Swerdel

#OHDSISocialShowcase This Week

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

Assessing impact on change in incidence on calibration performance across external validation

Ross D. Williams MSc (1), Jenna M. Reps PhD(2), Peter R. Rijnbeek PhD (1)
(1) Erasmus University Medical Centre, Rotterdam; (2) Janssen Research and Development, Raritan, NJ, USA

Introduction

Increasingly external validation is being seen as the gold standard in assessing prediction model performance[1]. One of the ways to measure the performance of a prediction model is to assess the calibration. This measures the agreement between predicted and observed risks. Calibration is essential to aiding decision making as using a poorly calibrated model would result in missing people who need intervention (if under-estimated risk) or giving an intervention unnecessarily (if over-estimating risk). A common issue that occurs when performing external validation is a worsening of calibration performance[2]. A possible reason for this could be due to the change in event rate between the development and validation environment. If this is the case it could be possible to correct some miscalibration by adjusting the model bias based upon the known differential event rates without retraining the model.

Materials and Methods

We developed and externally validated models across a multitude of databases and problem settings. These databases included: CCAE, MDCD, MDCR, Optum claims and Optum EHR. The problems are specified elsewhere in two studies, one looking at hospitalization risk in covid-19 patients and another looking at predicting heart failure in type 2 diabetes patients [2,3]. All studies predict binary outcomes. We hypothesised that miscalibration is dependent on differential event rate. To test this, we compared the differential event rates (equation 1), to determine whether the model produced an over or underestimate of risk. We assessed this using the calibration-in-the-large and the intercept of the model. These are both metrics to assess calibration. Calibration-in-the-large checks the agreement between the mean predicted risk and the event rate, a value of 1 being optimal. The calibration intercept, obtained by fitting a linear model between the predicted and observed values, assesses whether the risks are over or underestimated. A value of zero being perfect and a negative value suggesting overestimation and positive value underestimation. A perfectly calibrated model would thus have a calibration-in-the-large of 1 and a calibration intercept of 0.

$$\text{(Eq. 1) } \delta\, Incidence = event\ rate_{external\ database} - event\ rate_{development\ database}$$

Results

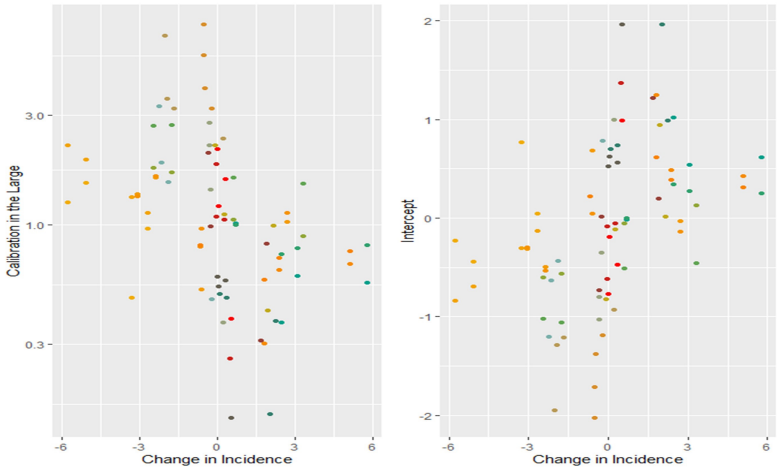


Figure 1 shows a relationship between the differential event rates and the calibration statistics. Calibration in the large and Intercept were both correlated using Pearsons test with a coefficient of 0.35 and 0.46 respectively.

Conclusion

The results show a relationship and suggest that using the differential event rate to create a correction factor for model recalibration is possible. The aim of future work is to extend it to include a correction factor based upon this relationship to provide a method of recalibration.

References
1. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016;69:245-247.
2. Williams RD, Markus AF, Yang C, et al. Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network. *medRxiv.* 2020:2020.2005.2026.20112649.
3. Williams RD, Reps JM, Kors JA, et al. Using the OHDSI network to develop and externally validate a patient-level prediction model for Heart Failure in Type II Diabetes Mellitus. *medRxiv.* 2021:2021.2004.2006.21254966.

Erasmus MC
Email: R.Williams@erasmusmc.nl

www.ehden.eu @IMI_EHDEN IMI_EHDEN github.com/EHDEN

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

TUESDAY

Assessing impact on change in incidence on calibration performance across external validation

Authors: Ross Williams, Jenna Reps, Peter Rijnbeek

#OHDSISocialShowcase This Week

A Journey through VA's uptake of the OMOP common data model.

PRESENTER: **Benjamin Viernes**

INTRO

- The Department of Veterans Affairs (VA) has a large, national community of users of electronic health record data for research and operational purposes to advance the health and healthcare of United States' Veterans.
- Here, we describe the outreach strategies VINCI employed to facilitate uptake and implementation of the OMOP CDM within the VA research community.

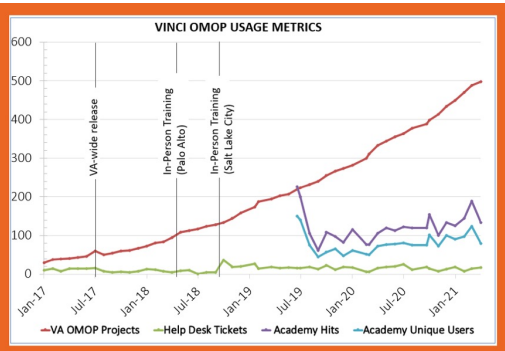
METHODS

We developed numerous outreach & education initiatives to support use and understanding:

* in-person, hands-on trainings
* virtual presentations,
* OMOP Academy, including online video tutorials
* transform documentation
* links to VA source data docs
* VA OMOP simulated data sandbox

We established and staffed a VA OMOP helpdesk:

* specific guidance to resolve an issue
* frequently asked question documentation
* additions and fix requests



The VA OMOP CDM, requested by over 500 research projects in 4 years, has reached wide-spread usage within the VA research ecosystem.

Take a picture to download the full paper

RESULTS

- Over 500 VA research projects were requested and approved for access since the first VA OMOP release in July 2017.
- Over 382 unique principal investigators requested access for VA research projects.
- Over the last 18 months, users of VA's OMOP Academy increased from approximately 50 to 100 users, monthly.
- Between January 2020 and April 2021, over 230 helpdesk tickets were received and answered.
- Since query logging began in June 2020, the VA OMOP database has been queried 9,141 times by 148 unique users.

Benjamin Viernes, Elizabeth E Hanchrow, Steven M Johnson, Elise Gatsby, Michael E Matheny, Daniel J Park, Jill M Whitaker, Scott L Duvall, Kristine E Lynch

VINCI

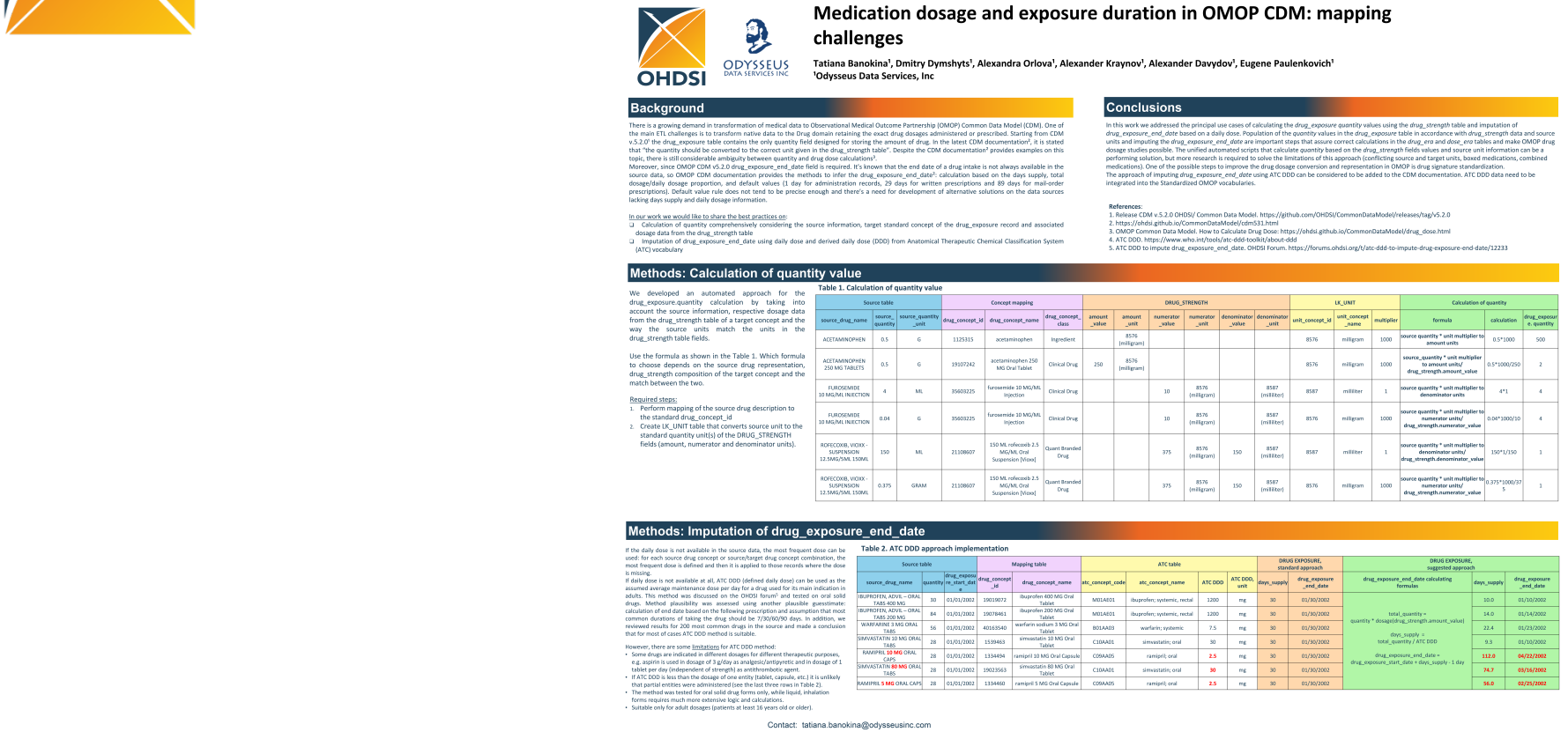
OHDSI

WEDNESDAY

A journey through VA's uptake of the OMOP common data model

Authors: Benjamin Viernes, Elizabeth E Hanchrow, Steven M Johnson, Elise Gatsby, Michael E Matheny, Daniel J Park, Jill M Whitaker, Scott L Duvall, Kristine E Lynch

#OHDSISocialShowcase This Week



Medication dosage and exposure duration in OMOP CDM: mapping challenges

Tatiana Banokina[1], Dmitry Dymshyts[1], Alexandra Orlova[1], Alexander Kraynov[1], Alexander Davydov[1], Eugene Paulenkovich[1]
[1]Odysseus Data Services, Inc

Background

There is a growing demand in transformation of medical data to Observational Medical Outcome Partnership (OMOP) Common Data Model (CDM). One of the main ETL challenges is to transform native data to the Drug domain retaining the exact drug dosages administered or prescribed. Starting from CDM v.5.2.0[1] the drug_exposure table contains the only quantity field designed for storing the amount of drug. In the latest CDM documentation[2], it is stated that "the quantity should be converted to the unit given in the drug_strength table". Despite the CDM documentation[2] provides examples on this topic, there is still considerable ambiguity between quantity and drug dose calculations[3].
Moreover, since OMOP CDM v5.2.0 drug_exposure_end_date field is required. It's known that the end date of a drug intake is not always available in the source data, so OMOP CDM documentation provides the methods to infer the drug_exposure_end_date[2] calculation based on the days supply, total dosage/daily dosage proportion, and default values (1 day for administration records, 29 days for written prescriptions and 89 days for mail-order prescriptions). Default value rule does not tend to be precise enough and there's a need for development of alternative solutions on the data sources lacking days supply and daily dosage information.

In our work we would like to share the best practices on:

- ❑ Calculation of quantity comprehensively considering the source information, target standard concept of the drug_exposure record and associated dosage data from the drug_strength table
- ❑ Imputation of drug_exposure_end_date using daily dose and derived daily dose (DDD) from Anatomical Therapeutic Chemical Classification System (ATC) vocabulary

Conclusions

In this work we addressed the principal use cases of calculating the *drug_exposure quantity* values using the *drug_strength* table and imputation of *drug_exposure_end_date* based on a daily dose. Population of the *quantity* values in the *drug_exposure* table in accordance with *drug_strength* data and source units and imputing the *drug_exposure_end_date* are important steps that assure correct calculations in the *drug_era* and *dose_era* tables and make OMOP drug dosage studies possible. The unified automated scripts that calculate *quantity* based on the *drug_strength* fields values and source unit information can be a performing solution, but more research is required to solve the limitations of this approach (conflicting source and target units, boxed medications, combined medications). One of the possible steps to improve the drug dosage conversion and representation in OMOP is drug signature standardization.
The approach of imputing *drug_exposure_end_date* using ATC DDD can be considered to be added to the CDM documentation. ATC DDD data need to be integrated into the Standardized OMOP vocabularies.

References:
1. Release CDM v.5.2.0 OHDSI/ Common Data Model. https://github.com/OHDSI/CommonDataModel/releases/tag/v5.2.0
2. https://ohdsi.github.io/CommonDataModel/cdm531.html
3. OMOP Common Data Model. How to Calculate Drug Dose: https://ohdsi.github.io/CommonDataModel/drug_dose.html
4. ATC DDD. https://www.who.int/tools/atc-ddd-toolkit/about-ddd
5. ATC DDD to impute drug_exposure_end_date. OHDSI Forum. https://forums.ohdsi.org/t/atc-ddd-to-impute-drug-exposure-end-date/12233

Methods: Calculation of quantity value

We developed an automated approach for the drug_exposure.quantity calculation by taking into account the source information, respective dosage data from the drug_strength table of a target concept and the way the source units match the units in the drug_strength table fields.

Use the formula as shown in the Table 1. Which formula to choose depends on the source drug representation, drug_strength composition of the target concept and the match between the two.

Required steps:

1. Perform mapping of the source drug description to the standard drug_concept_id
2. Create LK_UNIT table that converts source unit to the standard quantity unit(s) of the DRUG_STRENGTH fields (amount, numerator and denominator units).

Table 1. Calculation of quantity value

Source table			Concept mapping			DRUG_STRENGTH						LK_UNIT			Calculation of quantity		
source_drug_name	source_quantity	source_quantity_unit	drug_concept_id	drug_concept_name	drug_concept_class	amount_value	amount_unit	numerator_value	numerator_unit	denominator_value	denominator_unit	unit_concept_id	unit_concept_name	multiplier	formula	calculation	drug_exposure_quantity
ACETAMINOPHEN	0.5	G	1125315	acetaminophen	Ingredient		8576 (milligram)					8576	milligram	1000	source quantity * unit multiplier to amount units	0.5*1000	500
ACETAMINOPHEN 250 MG TABLETS	0.5	G	19107242	acetaminophen 250 MG Oral Tablet	Clinical Drug	250	8576 (milligram)					8576	milligram	1000	source_quantity * unit multiplier to amount units/ drug_strength.amount_value	0.5*1000/250	2
FUROSEMIDE 10 MG/ML INJECTION	4	ML	35603225	furosemide 10 MG/ML Injection	Clinical Drug			10	8576 (milligram)		8587 (milliliter)	8587	milliliter	1	source quantity * unit multiplier to denominator units	4*1	4
FUROSEMIDE 10 MG/ML INJECTION	0.04	G	35603225	furosemide 10 MG/ML Injection	Clinical Drug			10	8576 (milligram)		8587 (milliliter)	8576	milligram	1000	source quantity * unit multiplier to numerator units/ drug_strength.numerator_value	0.04*1000/10	4
ROFECOXIB, VIOXX - SUSPENSION 12.5MG/5ML 150ML	150	ML	21108607	150 ML rofecoxib 2.5 MG/ML Oral Suspension [Vioxx]	Quant Branded Drug			375	8576 (milligram)	150	8587 (milliliter)	8587	milliliter	1	source quantity * unit multiplier to denominator units/ drug_strength.denominator_value	150*1/150	1
ROFECOXIB, VIOXX - SUSPENSION 12.5MG/5ML 150ML	0.375	GRAM	21108607	150 ML rofecoxib 2.5 MG/ML Oral Suspension [Vioxx]	Quant Branded Drug			375	8576 (milligram)	150	8587 (milliliter)	8576	milligram	1000	source quantity * unit multiplier to numerator units/ drug_strength.numerator_value	0.375*1000/375	1

Methods: Imputation of drug_exposure_end_date

If the daily dose is not available in the source data, the most frequent dose can be used: for each source drug concept or source/target drug concept combination, the most frequent dose is defined and then it is applied to those records where the dose is missing.
If daily dose is not available at all, ATC DDD (defined daily dose) can be used as the assumed average maintenance dose per day for a drug used for its main indication in adults. This method was discussed on the OHDSI forum[2] and tested on oral solid drugs. Method plausibility was assessed using another plausible guesstimate: calculation of end date based on the following prescription and assumption that most common durations of taking the drug should be 7/30/60/90 days. In addition, we reviewed results for 200 most common drugs in the source and made a conclusion that for most of cases ATC DDD method is suitable.

However, there are some limitations for ATC DDD method:

- Some drugs are indicated in different dosages for different therapeutic purposes, e.g. aspirin is used in dosage of 3 g/day as analgesic/antipyretic and in dosage of 1 tablet per day (independent of strength) as antithrombotic agent.
- If ATC DDD is less than the dosage of one entity (tablet, capsule, etc.) it is unlikely that partial entities were administered (see the last three rows in Table 2).
- The method was tested for oral solid drug forms only, while liquid, inhalation forms requires much more extensive logic and calculations.
- Suitable only for adult dosages (patients at least 16 years old or older).

Table 2. ATC DDD approach implementation

Source table			Mapping table		ATC table				DRUG EXPOSURE, standard approach		DRUG EXPOSURE, suggested approach		
source_drug_name	quantity	drug_exposure_start_date	drug_concept_id	drug_concept_name	atc_concept_code	atc_concept_name	ATC DDD	ATC DDD, unit	days_supply	drug_exposure_end_date	drug_exposure_end_date calculating formulas	days_supply	drug_exposure_end_date
IBUPROFEN, ADVIL - ORAL TABS 400 MG	30	01/01/2002	19019072	ibuprofen 400 MG Oral Tablet	M01AE01	ibuprofen; systemic, rectal	1200	mg	30	01/30/2002	total_quantity = quantity * dosage(drug_strength.amount_value) / days_supply = total_quantity / ATC DDD drug_exposure_end_date = drug_exposure_start_date + days_supply - 1 day	10.0	01/10/2002
IBUPROFEN, ADVIL - ORAL TABS 200 MG	84	01/01/2002	19078461	ibuprofen 200 MG Oral Tablet	M01AE01	ibuprofen; systemic, rectal	1200	mg	30	01/30/2002		14.0	01/14/2002
WARFARINE 5 MG ORAL TABS	56	01/01/2002	40163540	warfarin sodium 5 MG Oral Tablet	B01AA03	warfarin; systemic	7.5	mg	30	01/30/2002		22.4	01/23/2002
SIMVASTATIN 10 MG ORAL TABS	28	01/01/2002	1539463	simvastatin 10 MG Oral Tablet	C10AA01	simvastatin; oral	30	mg	30	01/30/2002		9.3	01/10/2002
RAMIPRIL 10 MG ORAL CAPS	28	01/01/2002	1334494	ramipril 10 MG Oral Capsule	C09AA05	ramipril; oral	2.5	mg	30	01/30/2002		112.0	04/22/2002
SIMVASTATIN 80 MG ORAL TABS	28	01/01/2002	19033563	simvastatin 80 MG Oral Tablet	C10AA01	simvastatin; oral	30	mg	30	01/30/2002		74.7	03/16/2002
RAMIPRIL 5 MG ORAL CAPS	28	01/01/2002	1334460	ramipril 5 MG Oral Capsule	C09AA05	ramipril; oral	2.5	mg	30	01/30/2002		56.0	02/25/2002

Contact: tatiana.banokina@odysseusinc.com

THURSDAY

Medication dosage and exposure duration in OMOP CDM: mapping challenges

Authors: Tatiana Banokina, Dmitry Dymshyts, Alexandra Orlova, Alexander Kraynov, Alexander Davydov

#OHDSISocialShowcase This Week

Building Bridges with Julia
Using OHDSI R Packages in Julia

PRESENTER: Jacob S. Zelko

INTRODUCTION:

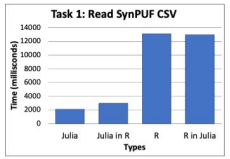
The amount of healthcare data is skyrocketing. Waiting on answers from this data just one day, can mean thousands of lives lost – as seen in the COVID pandemic. Beyond the cost of human lives, the financial costs of needed hardware for delivering these crucial answers are rising. Between the issues of financial costs and the urgency for rapid insights, there is clear need in the OHDSI community to encourage tooling that can bridge this problem.

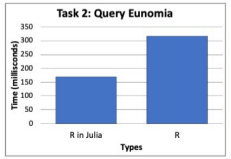
METHODS:

Demonstrated here is a possible solution, using a dynamic and high performance programming language, Julia[1], to interoperate with R to utilize OHDSI packages and perform basic procedures easily to produce performance gains. For this approach, I used the R package JuliaConnectoR[2] to use Julia tools from within R and the Julia package RCall[3] to call R tools from within Julia. Two basic exercises for benchmarking were done:

- Task 1: Read SynPUF[4] CSV
- Task 2: Query Eunomia[5]

RESULTS:





For Task 1, a ≈5x's speedup is seen from the Julia in R example over R's CSV reader.

For Task 2, only using R and R embedded in Julia was considered. Leveraging Julia, the R in Julia example is ≈2x's fast as the R implementation.

Task 1 Code: Reading SynPUF CSV

Figure A: Julia

```
# Using Julia's CSV reader
using CSV

# Reading in SynPUF data
data = CSV.File("synpuf.csv")
```

Figure B: R

```
# Reading raw SynPUF data
data <- read.csv("synpuf.csv")
```

Figure C: R in Julia

```
# Load RCall package
using RCall

# Read SynPUF data using R
data = R"read.csv(\"synpuf.csv\")"
```

Figure D: Julia in R

```
# Load the JuliaConnectoR package
library("JuliaConnectoR")

# Import Julia's CSV reader
jcsv <- juliaImport("CSV")

# Read in example SynPUF data
data <- jcsv$File("synpuf.csv")
```

Julia in R (Fig 1D), can provide a nearly 5x's speedup over the R CSV reader (Fig 1B).

Task 2 Code: Query Eunomia

Figure A: Querying in R

```
# Open connection to Eunomia
library('DatabaseConnector')
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
connection <- connect(connectionDetails)

# Create SQL Query
sql <- "
SELECT *
FROM @cdm.person
"

# Return people from SQL query
result <- renderTranslateQuerySql(connection, sql, cdm ="main")

# Make R data frame
data.frame(t(sapply(result,c)))
```

Figure B: Querying in Julia Using R

```
# Load RCall package
using RCall

# Get patients from Eunomia PERSON table
people = R"""

library('DatabaseConnector')
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
connection <- connect(connectionDetails)

sql <-  "
SELECT *
FROM @cdm.person
"
result <- renderTranslateQuerySql(connection, sql, cdm ="main")

"""

# Convert R 'list' to Julia DataFrame
people_data = rcopy(people)
```

Querying OHDSI's Eunomia package in Julia using R gives a 2x's speedup over the base R implementation

For more information, scan the QR code here!

Georgia Tech Research Institute
Problem. Solved.

SELECTED DISCUSSION TOPICS:

How were benchmarks made?

Table 1: Task Benchmarks

Task	Type	Time (ms)
Read SynPUF CSV	Julia	2114
Read SynPUF CSV	Julia in R	2960
Read SynPUF CSV	R	13100
Read SynPUF CSV	R in Julia	12980
Query Eunomia	R in Julia	169
Query Eunomia	R	317

In Table 1, the minimum time from 10 evaluations of the code created for each task was recorded. For the "Read SynPUF CSV" task, 150 MBs of SynPUF data were read. The benchmarking tools, BenchmarkTools.jl[6] for Julia and bench[7] for R was used to generate times.

Why Julia instead of language X?

- Interoperability with other languages
- High performance computing
- Understandable syntax.
- Emerging resources for OHDSI tasks (e.g. database interfaces, OMOP CDM, etc.).

What are future research directions?

Further directions for this research will be to

- Leverage existing OHDSI tools with Julia
- Identify improvements with Julia via language interoperability (i.e. R & Julia)
- Develop tooling for actual study to test how feasible it is to leverage Julia in an OHDSI network study design.

ACKNOWLEDGEMENTS

Thank you so much to the following people for their support in this endeavor!

- Charity Hilton and Jon Duke (Georgia Tech Research Institute).
- Dilum Aluthge and Clark C. Evans (JuliaHealth)
- Kristin Kostka (OHDSI Community).

REFERENCES:

1. Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B Shah. 2017. "Julia: A Fresh Approach to Numerical Computing." *SIAM review* 59(1): 65–98.
2. Lenz, Stefan, Maren Hackenberg, and Harald Binder. 2021. "The JuliaConnectoR: A Functionally Oriented Interface for Integrating Julia in R." *arXiv:2005.06334 [cs, stat]*. http://arxiv.org/abs/2005.06334 (August 13, 2021).
3. RCall.Jl. 2021. Julia Interop. Julia. https://github.com/JuliaInterop/RCall.jl (August 13, 2021).
4. "Medicare Claims Synthetic Public Use Files (SynPUFs) | CMS." https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs (August 13, 2021).
5. *Eunomia*. 2021. Observational Health Data Sciences and Informatics. R. https://github.com/OHDSI/Eunomia (August 13, 2021). Environments." *arXiv e-prints*.
6. Chen, Jiahao, and Jarrett Revels. 2016. "Robust Benchmarking in Noisy
7. High Precision Timing of R Expressions • Bench." https://bench.r-lib.org/ (August 13, 2021).

OHDSI

FRIDAY

There & Back Again: Using Julia to Augment OHDSI R Packages

Author: Jacob Zelko

Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Welcome To OHDSI Newcomers

Are there any new people to the OHDSI community call who would like to introduce themselves?

Please raise your hand and share why you are interested in joining the OHDSI community.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

February 8 OHDSI Community Call

Healthcare Systems Interest Group Update

Melanie Philofsky

Open Source Community Workgroup Update

Adam Black

Phenotype Phebruary Update #1

Patrick Ryan