Week 4
Workgroup 2023 OKRs and Phenotype Phebruary Updates

OHDSI Community Call
Feb. 28, 2023 • 11 am ET

Upcoming OHDSI Community Calls

Date	Topic
Mar. 7	Save Our Sisyphus (SOS) Research Idea Presentations
Mar. 14	OHDSI Debates
Mar. 21	Recent Publications
Mar. 28	SOS Week 1 Tutorial: Initiating A Network Study
Apr. 4	SOS Week 2 Tutorial: Data Diagnostics
Apr. 11	SOS Week 3 Tutorial: Phenotype Development
Apr. 18	SOS Week 4 Tutorial: Phenotype Evaluation
Apr. 25	SOS Week 5 Tutorial: Creating Analysis Specifications

Three Stages of The Journey

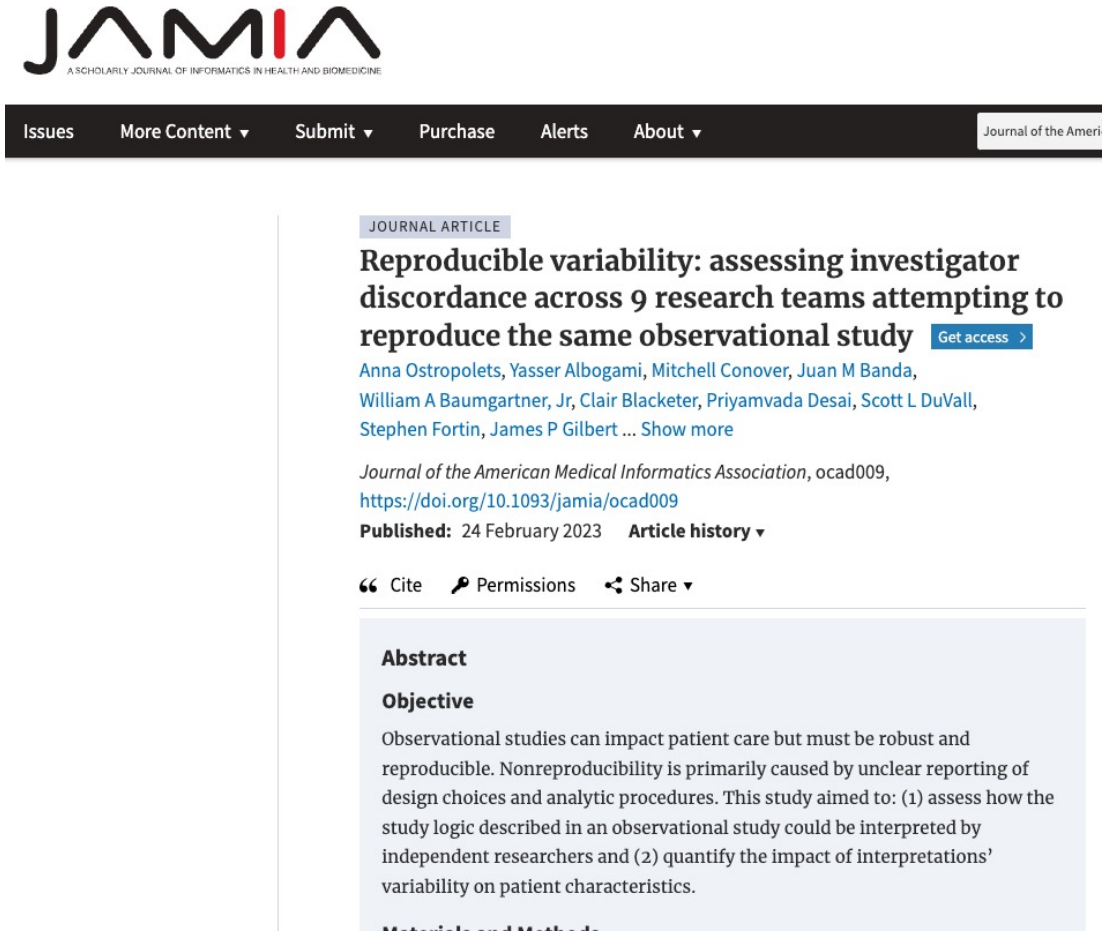
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Anna Ostropolets, Yasser Albogami, Mitchell Conover, Juan Banda, William Baumgartner, Clair Blacketer, Priyamvada Desai, Scott DuVall, Stephen Fortin, James Gilbert, Asieh Golozar, Joshua Ide, Andrew Kanter, David Kern, Chungsoo Kim, Lana Lai, Chenyu Li, Feifan Liu, Kristine Lynch, Evan Minty, Maria Inês Neves, Ding Quan Ng, Tontel Obene, Victor Pera, Nicole Pratt, Gowtham Rao, Nadav Rappoport, Ines Reinecke, Paola Saroufim, Azza Shoaibi, Katherine Simon, Marc Suchard, Joel Swerdel, Erica Voss, James Weaver, Linying Zhang, George Hripcsak, and Patrick Ryan** on the publication of **Reproducible variability: assessing investigator discordance across 9 research teams attempting to reproduce the same observational study** in JAMIA.



OHDSI Shoutouts!

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

Do you have anything you want to share? Please send to sachson@ohdsi.org so we can highlight 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	3 pm	OMOP CDM Oncology Outreach/Research Subgroup
Wednesday	2 am	Methods Research
Wednesday	8 am	Psychiatry
Wednesday	9 am	ATLAS
Wednesday	11 am	Open-Source Community
Wednesday	12 pm	Health Equity
Thursday	12 pm	Methods Research
Thursday	1 pm	OMOP CDM Oncology Vocabulary/Development Subgroup
Friday	9 am	GIS – Geographic Information System Development
Friday	1 pm	Clinical Trials
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Africa Chapter

ohdsi.org/workgroups

Spotlight: Faaizah Arshad

Get to know **Faaizah Arshad** in the latest collaborator spotlight.

- UCLA psychology major
- first undergraduate to present during symposium plenary
- co-founded the Early-Stage Researchers WG
- honored with 2021 Titan Award for Community Support

ohdsi.org/spotlight-faaizah-arshad

OHDSI HADES releases: SqlRender 1.12.1

SqlRender 1.12.1 | Reference | Articles | SqlDeveloper | Changelog | HADES

SqlRender 1.12.1

Bugfixes:

1. Fixed translation of `WITH ... INSERT` on Snowflake.
2. Fixed translation of some functions on Snowflake casting to `NUMERIC` instead of `FLOAT`.

SqlRender 1.12.0 2023-01-26

Changes:

1. Adding translation of `TRY_CAST()`.
2. The `loadRenderTranslateSql()` function now also looks in the `sql` folder of the package, so SQL files no longer have to be in the `sql/sql_server` subfolder.
3. Ensuring result of `YEAR()`, `MONTH()`, `DAY()`, and `DATEPART()` equivalents return integers on SQLite.
4. Ensuring interval is integer on BigQuery.

SqlRender 1.11.1 2023-01-11

Contents

1.12.1
1.12.0
1.11.1
1.11.0
1.10.0
1.9.2
1.9.1
1.9.0
1.8.3
1.8.2
1.8.1
1.8.0
1.7.0
1.6.8
1.6.7
1.6.6
1.6.5

Inaugural Lecture for Professor Peter Rijnbeek

Inaugural Lecture dr.ir. P.R. (Peter) Rijnbeek

Scalable Evidence

Date Friday 3 Mar 2023, 16:00 - 17:00

Inaugural Lecture

The rector magnificus of the Erasmus University Rotterdam announces that at Erasmus MC - Faculty of Erasmus University Rotterdam

Dr.ir. P.R. (Peter) Rijnbeek

appointed as professor of Medical Informatics will publicly accept his appointment on Friday 3 March 2023 with an inaugural lecture entitled:

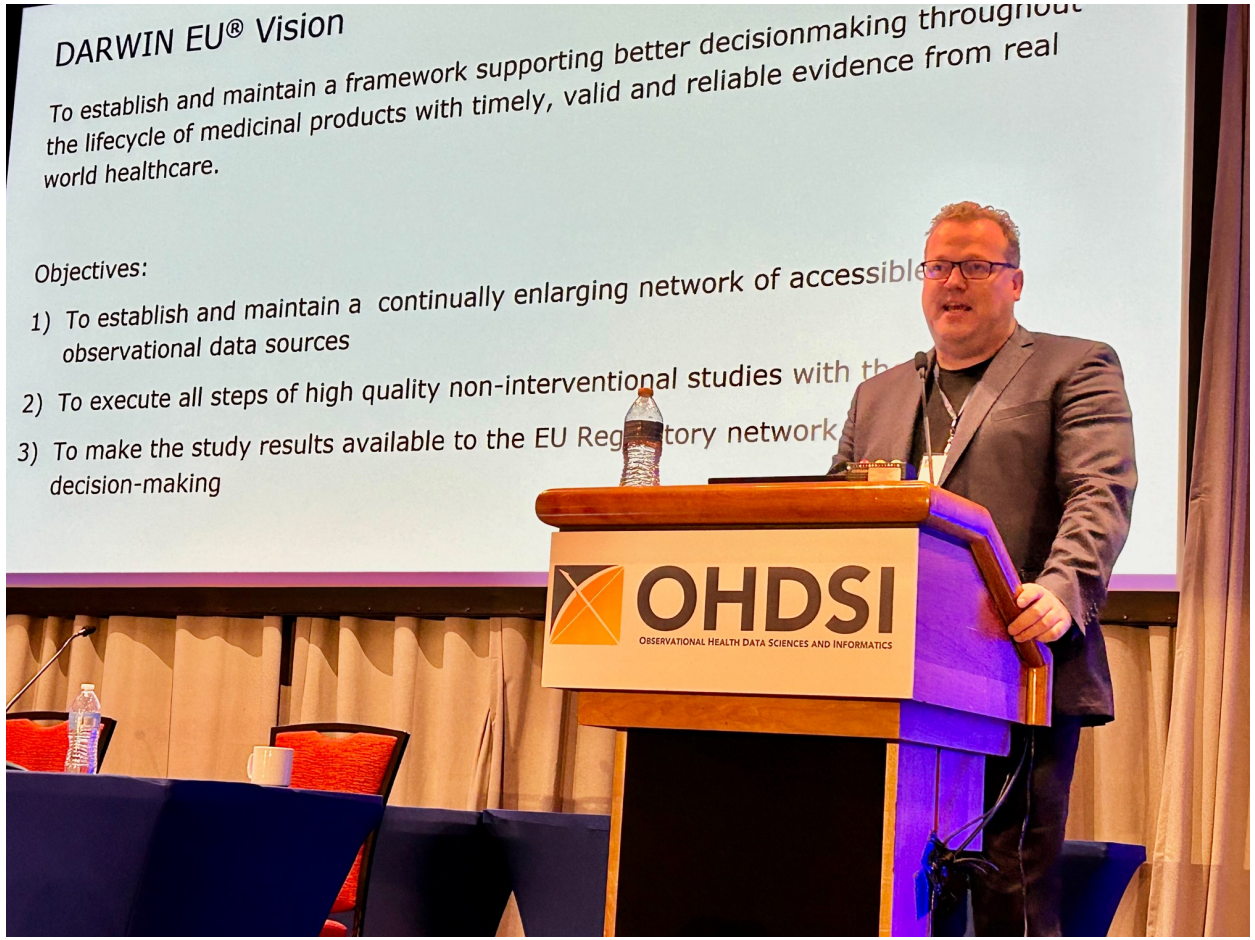
Scalable Evidence

Professors are invited to participate with a gown in the academic procession. We kindly request them to be present from 15:30 hrs on the first floor of the Erasmus (A) Building, near the rector's room.

The ceremony will start promptly at 16.00 hrs in the Aula of the university (Erasmus building) Burgemeester Oudlaan 50 Rotterdam.

If you are unable to attend, you can also attend the inaugural lecture via livestream: https://eur.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=3092bc5b-4ff3-43c9-8073-af5b00a4d282

The reception will take place in the same building afterwards. The rector magnificus invites you to attend this ceremony and the reception.



2023 AMIA Symposium Call For Participation

AMIA 2023 Annual Symposium Call for Participation

We invite you to contribute your best work for presentation at the AMIA 2023 Annual Symposium – the leading symposium for the science and practice of health and biomedical informatics. The AMIA 2023 Annual Symposium showcases submissions from scientists, clinicians, trainees, educators, policy makers, administrators, industry professionals, and technologists from around the world.

The AMIA 2023 Annual Symposium will consider submissions of the following types:

- Paper, Student Paper
- Podium Abstract
- Poster, Panel
- Informatics Debate
- Systems Demonstration
- Workshop

Proposals

Proposals are now being accepted.

Deadline: Mar. 8, 2023

Submit now

2023 DevCon: April 21

OHDSI DevCon 2022 Welcomes & Mentors New Contributors To Our Open-Source Environment

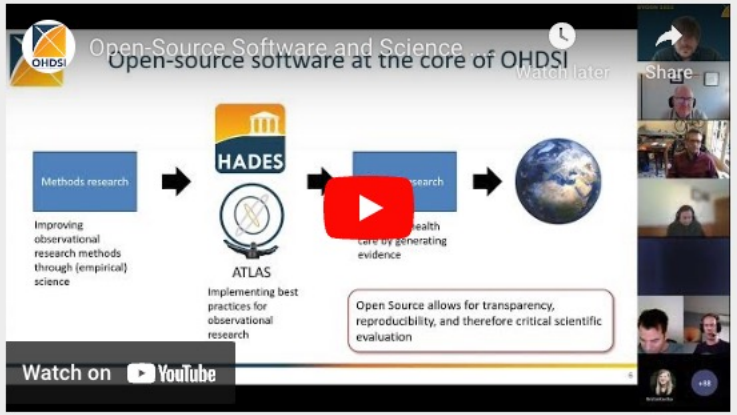
Watch All Eight Workshops, Talks & The Panel From DevCon Below

The Open-Source Community hosted the first Dev Con on Friday, April 22 as a way of accepting and mentoring new contributors to our environment. Organized by **Paul Nagy** and **Adam Black**, the event included eight workshops, talks and a panel discussion to both welcome and engage both current and future developers within OHDSI.

All videos from this session have or will be uploaded to this page. A big announcement from DevCon was the formation of the Khieron Contributor Cohort, which will help onboard and mentor open-source developers in the community. If you are interested in joining the effort, please fill out the application.

To learn more about the Khieron Contributor Cohort, please check out the State of the Open Source Community presentation below.

OHDSI DevCon Keynote



Martijn Schuemie provided the keynote address during DevCon 2022, entitled "Open-Source Software and Science … Obviously." His slides are available here.

Click Here To Apply For The 2022-23 Khieron Contributor Cohort

Join The Open-Source Community Workgroup in MS Teams

Workshops

ATLAS
(Anthony Sena)

HADES Introduction
(Adam Black)

WebAPI
(Anthony Sena)

Cohort Diagnostics
(James Gilbert)

European Symposium: July 1-3, 2023

APAC Symposium: July 13-14, 2023

2023 APAC Symposium

July 13-14 • University of New South Wales • Sydney, Australia

We are excited to announce that the 2023 OHDSI APAC Symposium will be held in Sydney, Australia at the University of New South Wales! Agenda and registration details are coming soon so please stay tuned! Meanwhile, here are some important dates for you to save to your calendar:

Collaboration Showcase submissions open: Feb. 13

Collaboration Showcase submissions deadline: March 31

Symposium Day 1, main conference: July 13

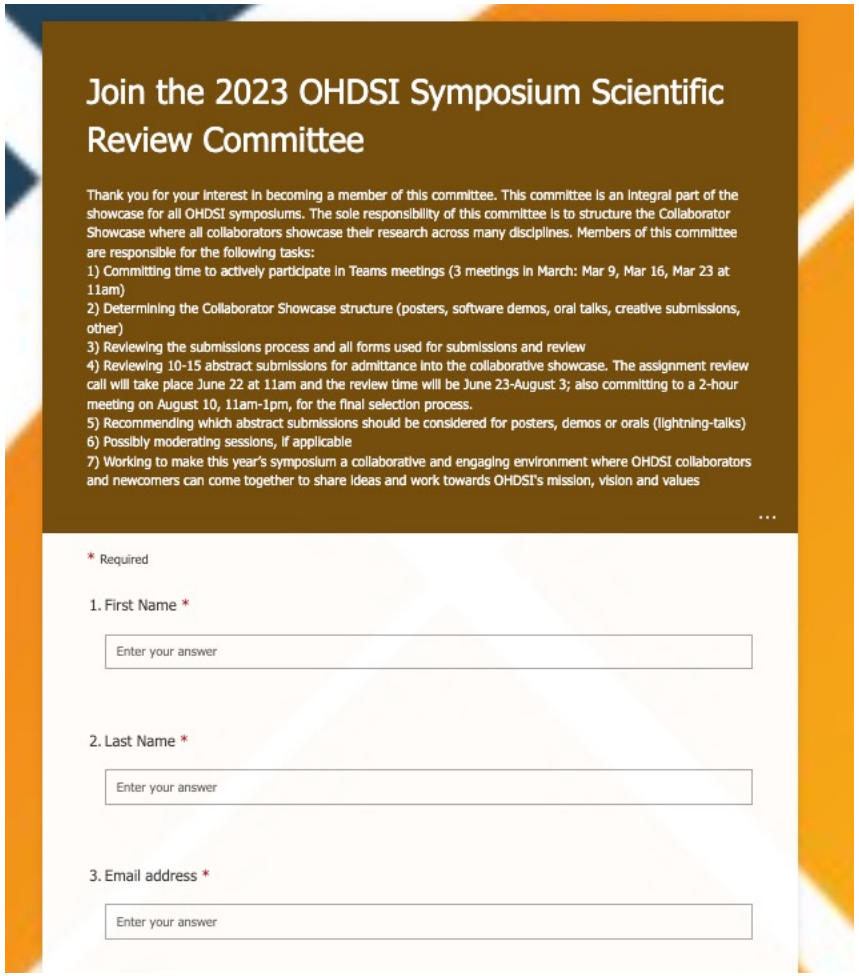
Symposium Day 2, tutorials: July 14

Global Symposium: Oct. 20-22, 2023

Hilton East Brunswick Hotel & Executive Meeting Center • East Brunswick, N.J.

Join The #OHDSI2023 Scientific Review Committee

We are looking for collaborators to join the OHDSI2023 scientific review committee. **Elisse Katzman** has opened the signup form to join the committee, and the first meeting is scheduled for March 9. **The deadline is Feb. 28.**



bit.ly/OHDSI2023ScientificReview

Do You Know Of A Collaboration Opportunity?

We are trying to keep the community updated on all collaboration opportunities, both inside AND outside of OHDSI activities. **Marty Alvarez** of Tufts University is doing a fantastic job of compiling them each week so we know what is on the horizon, and we are working on a format to post these for the community.

In the meantime, if you know of any upcoming opportunities (grants, conferences, calls for papers, etc.) that you think should be considered for this list, please send them to **Marta.Alvarez@tuftsmedicine.org**.

Thank you Marty!

Job Opening

Back to Search

Open Rank- Tenure Track of Internal Medicine in Translational Informatics

Posting Number	req23346
Employment Type	Faculty
Faculty Type	Open Rank
Hiring Department	IM Translations Informatics (852T)
Academic Location	School of Medicine
Benefits Eligible	The University of New Mexico provides a comprehensive package of benefits including medical, dental, vision, and life insurance. In addition, UNM offers educational benefits through the tuition remission and dependent education programs. See the Benefits home page for more information.
Position Summary	The University of New Mexico, Health Sciences Center, Department of Internal Medicine, seeks a faculty member to join the Division of Translational Informatics. This position is at the Open rank and Tenure track. While the focus of the position is research-oriented, optionally, the position affords the opportunity for the candidate to have a joint clinical appointment for part-time clinical service with the University of New Mexico, and/or the Raymond G. Murphy VA Medical Center. Salary will be commensurate with experience and education.

Job Opening

Software Dev Analyst II - Res - G&C - CTSI

Job ID: REF9053H
Date posted: 2/20/2023

Employment Type: Full Time
Shift: Days
Location: Boston, MA

PRINCIPAL DUTIES AND ESSENTIAL FUNCTIONS:

Responsible for executing software development initiatives.

Implementation

- Collaborate with various stakeholders to understand requirements and design solutions
- Evaluate options and develop technical design
- Develop solution using appropriate programming language and/or technical tools
- Complete thorough testing of solution
- Provide input to the development of integrated test plan
- Execute integrated test plan
- Provide input to the development of LIVE plan
- Support LIVE activities

Ongoing Enhancements and Support

- Build enhancements to current functionality using appropriate programming language and/or technical tools
- Perform detailed testing of software updates and upgrades
- Communicate in a friendly and professional manner, share the ideas, solutions, the approach, risks, and impacts, set appropriate expectations for the development timeline
- Participate in after-hours on call support rotation for one or more applications which generate Incidents outside of business hours.
- Participate in cross-training, as a trainer and a learner, for personal development and to ensure adequate secondary coverage on all applications

Job Opening

COLUMBIA | COLUMBIA UNIVERSITY DEPARTMENT OF BIOMEDICAL INFORMATICS

DBMI Home | News & Events | Research | People | Prospective Students | Academics | Resources

Tenure Track Faculty

#105752

Description

The Department of Biomedical Informatics (DBMI) of Columbia University seeks exceptional junior-level faculty members in the tenure track.

The positions are open to researchers interested in developing and applying informatics theory and achieving tangible benefits to health care and biology. Three particular foci are (1) machine learning for healthcare and health-related data science, (2) health information technology-based interventions to improve health care and the health of individuals and populations, and (3) translational bioinformatics.

Job Opening

Job Details

Database Programmer

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD

We are seeking to appoint a highly qualified and dedicated Database Programmer to join the Health Data Sciences research group led by Professor Daniel Prieto-Alhambra at the Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), Oxford.

You will join an outstanding, multi-disciplinary and friendly Group of motivated and cutting-edge researchers and to contribute to clinical research by providing technical knowledge, software engineering expertise and data insight.

As a Database Programmer you will Develop new database applications for big clinical data to meet project requirements and deadlines, provide software feedback and carry out software improvement, extension, integration and further development on existing code. You will contribute to the harmonisation, curation, and processing of large clinical datasets and develop code to validate, test, document and maintain database applications. You will also represent the project, team, and the University in collaboration meetings, conferences and at external meetings.

You will have a Degree in computer science, software engineering, health informatics or an equivalent combination of training and professional experience. Proven understanding and experience in one or more RDBMSs and SQL dialects (e.g. PostgreSQL), excellent skills in at least one high level programming language (e.g. Python, C#, C++) and excellent analytical and problem-solving skills with great attention to detail are essential. Experience in common data models (CDMs) and in the extract, transform, and load (ETL) process, knowledge of R and/or RStudio and working experience in a research environment are desirable.

This is a full-time fixed-term appointment for 2 years.

The closing date for this position is 12 noon on Monday 27 February 2023. You will be required to upload a CV and supporting statement as part of your online application.

Contact Person : HR Team, NDRMS
Contact Phone :
Vacancy ID : 163066
Closing Date & Time : 27-Feb-2023 12:00

Janssen R&D Summer Internships

General Administration

Epidemiology Graduate Intern

General Administration

OHDA Graduate Intern

General Administration

OHDA Undergraduate Intern

General Administration

Data Science RWE for R&D Summer Intern

General Administration

Data Science RWE DevCon Summer Intern

#OHDSISocialShowcase This Week

HERMES:
A Health Resources Econometric Analysis Tool

PRESENTER: Kyungseon Choi
Contact: kyungseon.choi@khu.ac.kr

INTRO:

- Do you want to compare which cohort had more medical expenses or health resource utilization with OMOP-CDM?
- You can estimate and compare the medical expenses and health resource utilization for disease or patients using HERMES on your cohorts.
- Estimating the economic burden through healthcare cost analysis is important to properly distribute the limited healthcare resources.
- However, there are huddles in estimation the unbiased precise healthcare costs using OMOP-CDM due to "**Zero-cost**" and "**Skewed Data**".

METHODS:

- To adjust positive skewness and zero cost by econometric model and estimate precise healthcare costs, we reviewed literature related to healthcare cost analysis.
- We structed an algorithm using an econometric model based on a previously well-established method (Manning, et al (2001). J Health Econ. 20(4):461-94).
- To verify the algorithm and R functions, we conducted an empirical study on patients with exudative age-related macular degeneration (AMD).
- For cross-validation, we compared the cost analysis method and the estimate of the reimbursement cost with the previous study conducted from claims data in South Korea for a similar period.
- During the empirical study and cross-validation, the results were confirmed by health economics experts and ophthalmologists.

ACKNOWLEDGMENT:

- This research was supported by KHIDI and a grant (21153MFDS601) from Ministry of Food and Drug Safety in 2022.

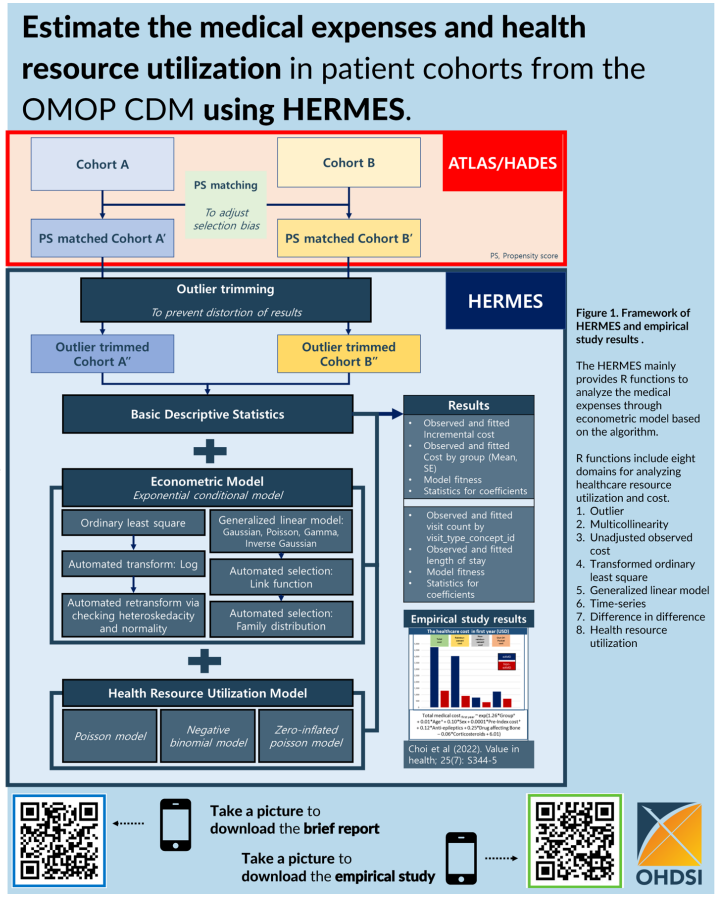


Figure 1. Framework of HERMES and empirical study results .

The HERMES mainly provides R functions to analyze the medical expenses through econometric model based on the algorithm.

R functions include eight domains for analyzing healthcare resource utilization and cost.

1. Outlier
2. Multicollinearity
3. Unadjusted observed cost
4. Transformed ordinary least square
5. Generalized linear model
6. Time-series
7. Difference in difference
8. Health resource utilization

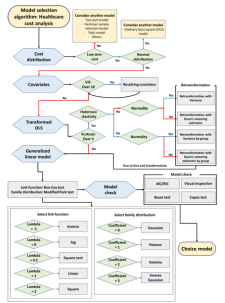


Figure 2. Full algorithm for HERMES for a healthcare cost analysis in OMOP-CDM

CONCLUSIONS:

- The HERMES can estimate the costs categorized by reimbursement, non-reimbursement, payers', and patients' costs. The results provide economic, clinical, and policy implications and help stakeholders to understand the economic impact of certain diseases or interventions.
- We conducted an empirical study to validate the usability of HERMES with OMOP CDM cost data. As a result, it was possible to derive similar results in the reimbursement cost compared to the results of previous studies using a population based national claims data (Kim et al. (2019). BMC Health Serv. Res. 19: 828).
- HERMES not only helps to select an appropriate econometric model according to the algorithm by identifying the characteristics of cost data, but also provides various tools related to estimating economic burden such as time series, difference in difference, and health resources utilization model.
- With the expectation that disease burden research through CDM will be further progressed in the future, the HERMES will contribute to healthcare cost research using OMOP-CDM.

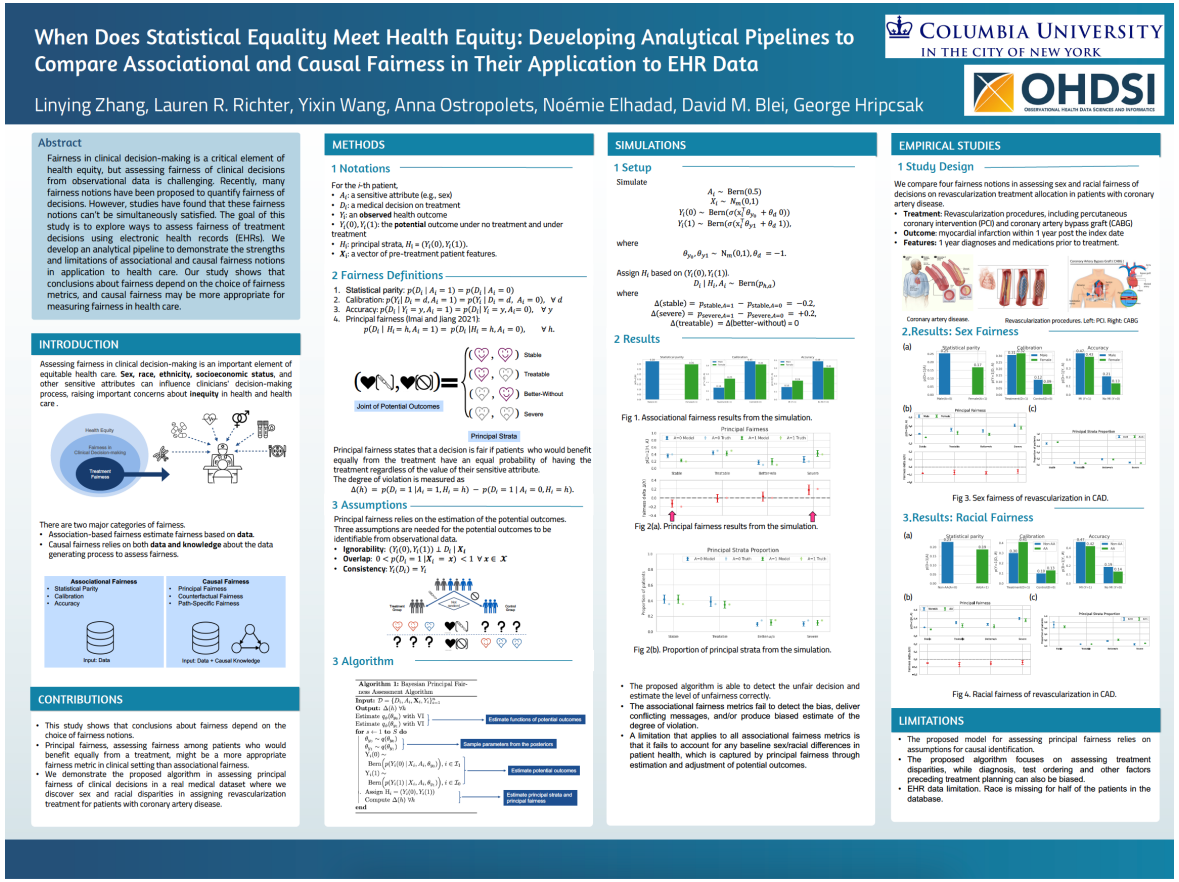
Kyungseon Choi, Sang Jun Park, Sola Han, Siin Kim, Hae Sun Suh

KYUNG HEE UNIVERSITY
SEOUL NATIONAL UNIVERSITY

MONDAY

HERMES: A Health Resources Econometric Analysis Tool (Kyungseon Choi, Sang Jun Park, Sola Han, Siin Kim, Hae Sun Suh)

#OHDSISocialShowcase This Week



When Does Statistical Equality Meet Health Equity: Developing Analytical Pipelines to Compare Associational and Causal Fairness in Their Application to EHR Data

Linying Zhang, Lauren R. Richter, Yixin Wang, Anna Ostropolets, Noémie Elhadad, David M. Blei, George Hripcsak

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

OHDSI

Abstract

Fairness in clinical decision-making is a critical element of health equity, but assessing fairness of clinical decisions from observational data is challenging. Recently, many fairness notions have been proposed to quantify fairness of decisions. However, studies have found that these fairness notions can't be simultaneously satisfied. The goal of this study is to explore ways to assess fairness of treatment decisions using electronic health records (EHRs). We develop an analytical pipeline to demonstrate the strengths and limitations of associational and causal fairness notions in application to health care. Our study shows that conclusions about fairness depend on the choice of fairness metrics, and causal fairness may be more appropriate for measuring fairness in health care.

INTRODUCTION

Assessing fairness in clinical decision-making is an important element of equitable health care. **Sex, race, ethnicity, socioeconomic status**, and other sensitive attributes can influence clinicians' decision-making process, raising important concerns about **inequity** in health and health care.

There are two major categories of fairness.
- Association-based fairness estimate fairness based on **data**.
- Causal fairness relies on both **data and knowledge** about the data generating process to assess fairness.

Associational Fairness: Statistical Parity, Calibration, Accuracy — Input: Data

Causal Fairness: Principal Fairness, Counterfactual Fairness, Path-Specific Fairness — Input: Data + Causal Knowledge

CONTRIBUTIONS

- This study shows that conclusions about fairness depend on the choice of fairness notions.
- Principal fairness, assessing fairness among patients who would benefit equally from a treatment, might be a more appropriate fairness metric in clinical setting than associational fairness.
- We demonstrate the proposed algorithm in assessing principal fairness of clinical decisions in a real medical dataset where we discover sex and racial disparities in assigning revascularization treatment for patients with coronary artery disease.

METHODS

1 Notations

For the i-th patient,
- A_i: a sensitive attribute (e.g., sex)
- D_i: a medical decision on treatment
- Y_i: an **observed** health outcome
- $Y_i(0), Y_i(1)$: the **potential** outcome under no treatment and under treatment
- H_i: principal strata, $H_i = (Y_i(0), Y_i(1))$.
- $\boldsymbol{X}_i$: a vector of pre-treatment patient features.

2 Fairness Definitions

1. Statistical parity: $p(D_i \mid A_i = 1) = p(D_i \mid A_i = 0)$
2. Calibration: $p(Y_i \mid D_i = d, A_i = 1) = p(Y_i \mid D_i = d,\ A_i = 0),\ \forall\, d$
3. Accuracy: $p(D_i \mid Y_i = y, A_i = 1) = p(D_i \mid Y_i = y, A_i = 0),\ \forall\, y$
4. Principal fairness (Imai and Jiang 2021): $p(D_i \mid H_i = h, A_i = 1) = p(D_i \mid H_i = h, A_i = 0),\ \forall\, h.$

Joint of Potential Outcomes → Principal Strata: Stable, Treatable, Better-Without, Severe

Principal fairness states that a decision is fair if patients who would benefit equally from the treatment have an equal probability of having the treatment regardless of the value of their sensitive attribute.
The degree of violation is measured as
$\Delta(h) = p(D_i = 1 \mid A_i = 1, H_i = h) - p(D_i = 1 \mid A_i = 0, H_i = h).$

3 Assumptions

Principal fairness relies on the estimation of the potential outcomes. Three assumptions are needed for the potential outcomes to be identifiable from observational data.
- **Ignorability**: $(Y_i(0), Y_i(1)) \perp D_i \mid \boldsymbol{X}_i$
- **Overlap**: $0 < p(D_i = 1 \mid \boldsymbol{X}_i = \boldsymbol{x}) < 1\ \forall\, \boldsymbol{x} \in \boldsymbol{X}$
- **Consistency**: $Y_i(D_i) = Y_i$

3 Algorithm

Algorithm 1: Bayesian Principal Fairness Assessment Algorithm

Inputs: $\mathcal{D} = \{D_i, A_i, \boldsymbol{X}_i, Y_i\}_{i=1}^n$
Output: $\Delta(h)$ $\forall h$
Estimate $q_\phi(\theta_{y_0})$ with VI → Estimate functions of potential outcomes
Estimate $q_\phi(\theta_{y_1})$ with VI
for $s \leftarrow 1$ **to** S **do**
  $\theta_{y_0} \sim q(\theta_{y_0})$
  $\theta_{y_1} \sim q(\theta_{y_1})$ → Sample parameters from the posteriors
  $Y_i(0) \sim$ $\mathrm{Bern}\left(\sigma(Y_i(0) \mid \boldsymbol{X}_i, A_i, \theta_{y_0})\right), i \in \mathcal{I}_1$
  $Y_i(1) \sim$ $\mathrm{Bern}\left(\sigma(Y_i(1) \mid \boldsymbol{X}_i, A_i, \theta_{y_1})\right), i \in \mathcal{I}_2$ → Estimate potential outcomes
  Assign $H_i = (Y_i(0), Y_i(1))$
  Compute $\Delta(h)$ $\forall h$ → Estimate principal strata and principal fairness
end

SIMULATIONS

1 Setup

Simulate

$$A_i \sim \mathrm{Bern}(0.5)$$
$$X_i \sim N_m(0,1)$$
$$Y_i(0) \sim \mathrm{Bern}(\sigma(\boldsymbol{x}_i^\top \theta_{y_0} + \theta_\delta\, 0))$$
$$Y_i(1) \sim \mathrm{Bern}(\sigma(\boldsymbol{x}_i^\top \theta_{y_1} + \theta_\delta\, 1)),$$

where

$$\theta_{y_0}, \theta_{y_1} \sim N_m(0,1), \theta_\delta = -1.$$

Assign H_i based on $(Y_i(0), Y_i(1))$.

$$D_i \mid H_i, A_i \sim \mathrm{Bern}(p_{h,a})$$

where

$$\Delta(\text{stable}) = p_{\text{stable},a=1} - p_{\text{stable},a=0} = -0.2,$$
$$\Delta(\text{severe}) = p_{\text{severe},a=1} - p_{\text{severe},a=0} = +0.2,$$
$$\Delta(\text{treatable}) = \Delta(\text{better-without}) = 0$$

2 Results

Fig 1. Associational fairness results from the simulation.

Fig 2(a). Principal fairness results from the simulation.

Fig 2(b). Proportion of principal strata from the simulation.

- The proposed algorithm is able to detect the unfair decision and estimate the level of unfairness correctly.
- The associational fairness metrics fail to detect the bias, deliver conflicting messages, and/or produce biased estimate of the degree of violation.
- A limitation that applies to all associational fairness metrics is that it fails to account for any baseline sex/racial differences in patient health, which is captured by principal fairness through estimation and adjustment of potential outcomes.

EMPIRICAL STUDIES

1 Study Design

We compare four fairness notions in assessing sex and racial fairness of decisions on revascularization treatment allocation in patients with coronary artery disease.
- **Treatment**: Revascularization procedures, including percutaneous coronary intervention (PCI) and coronary artery bypass graft (CABG)
- **Outcome**: myocardial infarction within 1 year post the index date
- **Features**: 1 year diagnoses and medications prior to treatment.

Coronary artery disease. Revascularization procedures. Left: PCI. Right: CABG

2.Results: Sex Fairness

Fig 3. Sex fairness of revascularization in CAD.

3.Results: Racial Fairness

Fig 4. Racial fairness of revascularization in CAD.

LIMITATIONS

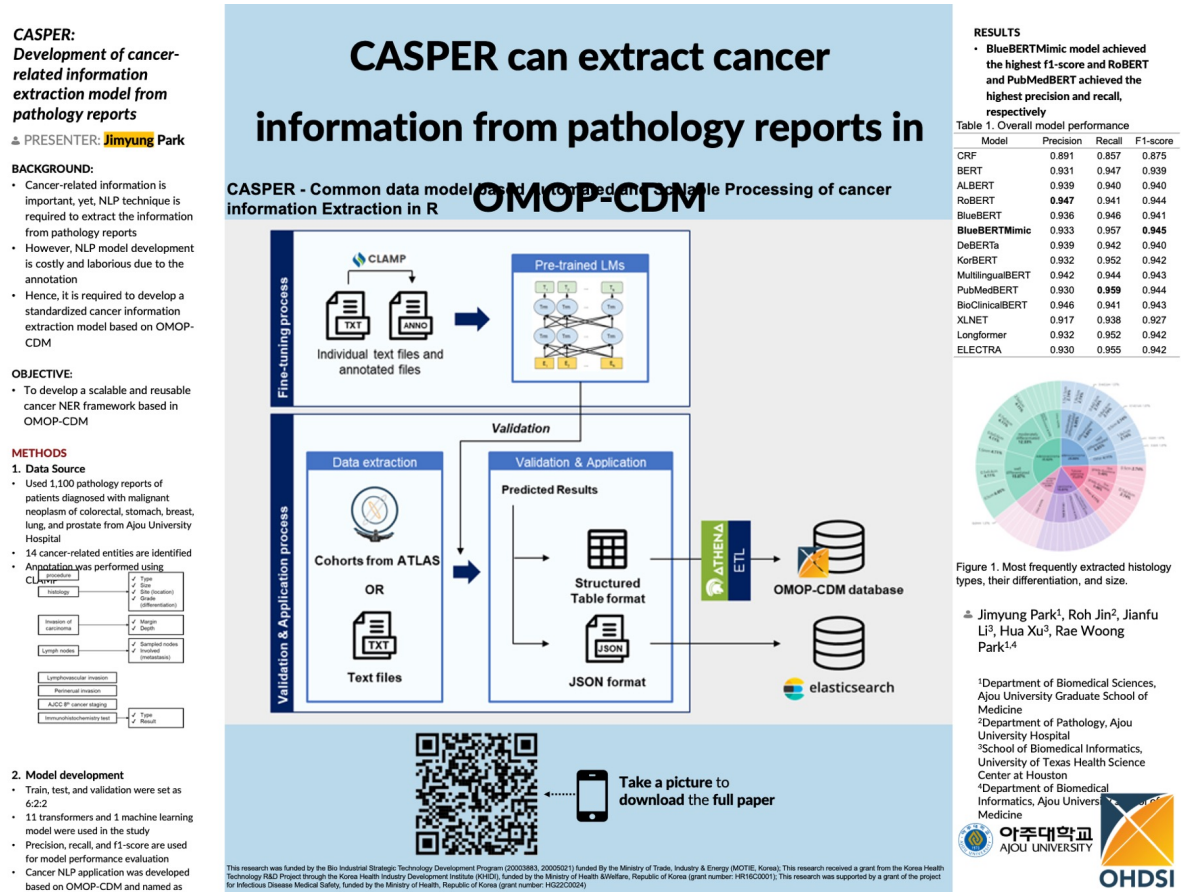
- The proposed model for assessing principal fairness relies on assumptions for causal identification.
- The proposed algorithm focuses on assessing treatment disparities, while diagnosis, test ordering and other factors preceding treatment planning can also be biased.
- EHR data limitation. Race is missing for half of the patients in the database.

TUESDAY

Assessing Racial Fairness of Dialysis Allocation in End-Stage Renal Disease

(presenter: Linying Zhang, Lauren R. Richter, David M. Blei, Yixin Wang, Anna Ostropolets, Noemie Elhadad, George Hripcsak)

#OHDSISocialShowcase This Week



CASPER:
Development of cancer-related information extraction model from pathology reports

PRESENTER: **Jimyung Park**

BACKGROUND:
- Cancer-related information is important, yet, NLP technique is required to extract the information from pathology reports
- However, NLP model development is costly and laborious due to the annotation
- Hence, it is required to develop a standardized cancer information extraction model based on OMOP-CDM

OBJECTIVE:
- To develop a scalable and reusable cancer NER framework based in OMOP-CDM

METHODS

1. Data Source
- Used 1,100 pathology reports of patients diagnosed with malignant neoplasm of colorectal, stomach, breast, lung, and prostate from Ajou University Hospital
- 14 cancer-related entities are identified
- Annotation was performed using CLAMP

2. Model development
- Train, test, and validation were set as 6:2:2
- 11 transformers and 1 machine learning model were used in the study
- Precision, recall, and f1-score are used for model performance evaluation
- Cancer NLP application was developed based on OMOP-CDM and named as CASPER - Common data model based Automated and Scalable Processing of cancer information Extraction in R

CASPER can extract cancer information from pathology reports in OMOP-CDM

Fine-tuning process: CLAMP, TXT, ANNO, Individual text files and annotated files, Pre-trained LMs

Validation

Validation & Application process: Data extraction, Cohorts from ATLAS, OR, TXT, Text files; Validation & Application, Predicted Results, Structured Table format, JSON, JSON format; ATHENA ETL, OMOP-CDM database, elasticsearch

Take a picture to download the full paper

This research was funded by the Bio Industrial Strategic Technology Development Program (20003883, 20005021) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea); This research received a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health &Welfare, Republic of Korea (grant number: HR16C0001); This research was supported by a grant of the project for Infectious Disease Medical Safety, funded by the Ministry of Health, Republic of Korea (grant number: HG22C0024)

RESULTS
- **BlueBERTMimic model achieved the highest f1-score and RoBERT and PubMedBERT achieved the highest precision and recall, respectively**

Table 1. Overall model performance

Model	Precision	Recall	F1-score
CRF	0.891	0.857	0.875
BERT	0.931	0.947	0.939
ALBERT	0.939	0.940	0.940
RoBERT	**0.947**	0.941	0.944
BlueBERT	0.936	0.946	0.941
BlueBERTMimic	0.933	0.957	**0.945**
DeBERTa	0.939	0.942	0.940
KorBERT	0.932	0.952	0.942
MultilingualBERT	0.942	0.944	0.943
PubMedBERT	0.930	**0.959**	0.944
BioClinicalBERT	0.946	0.941	0.943
XLNET	0.917	0.938	0.927
Longformer	0.932	0.952	0.942
ELECTRA	0.930	0.955	0.942

Figure 1. Most frequently extracted histology types, their differentiation, and size.

Jimyung Park[1], Roh Jin[2], Jianfu Li[3], Hua Xu[3], Rae Woong Park[1,4]

[1]Department of Biomedical Sciences, Ajou University Graduate School of Medicine
[2]Department of Pathology, Ajou University Hospital
[3]School of Biomedical Informatics, University of Texas Health Science Center at Houston
[4]Department of Biomedical Informatics, Ajou University School of Medicine

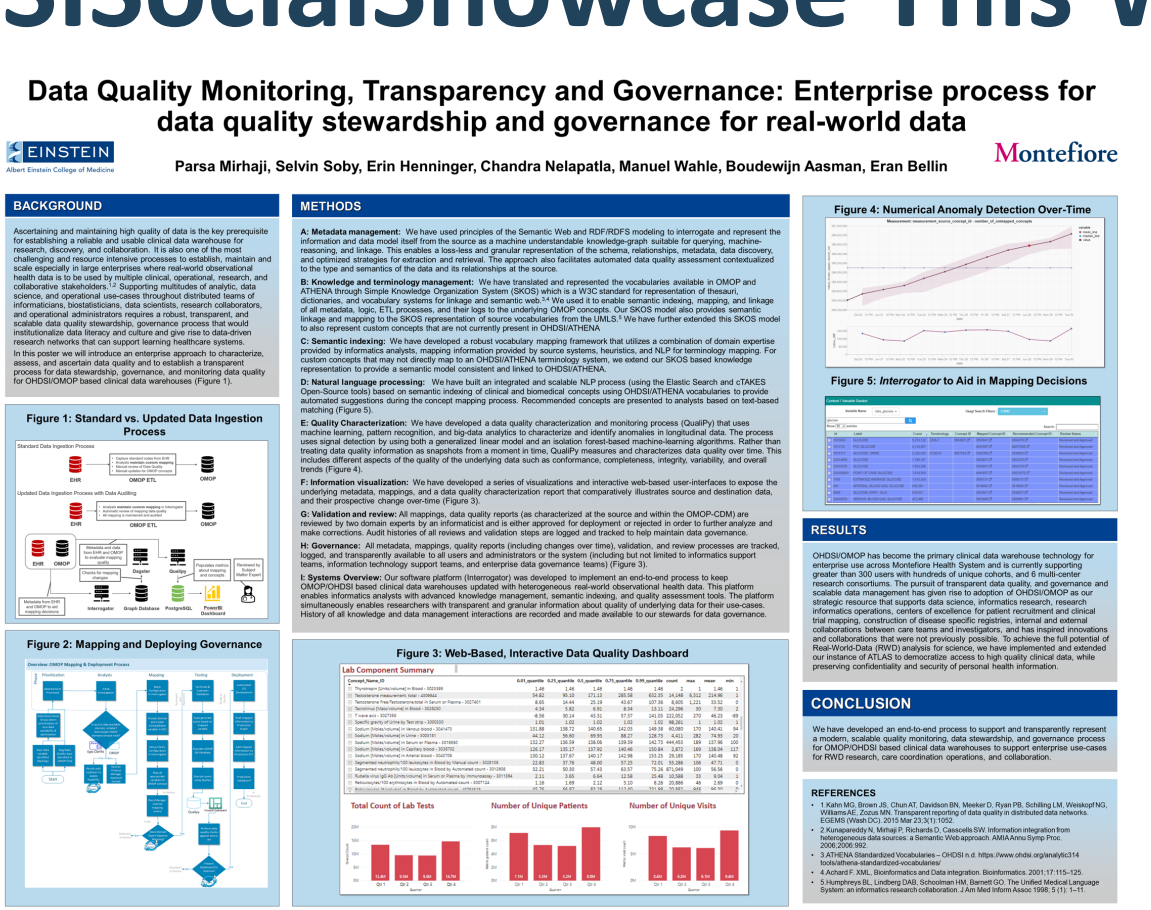
아주대학교 AJOU UNIVERSITY

OHDSI

WEDNESDAY

Preliminary Analysis of Self-Reported COVID-19 Vaccination Side Effects on Twitter (Nishanth Pavinkurve, Maura Beaton, Tilly Seesillapachi, Xinzhuo Jiang, Hua Xu, Karthik Natarajan)

#OHDSISocialShowcase This Week



Data Quality Monitoring, Transparency and Governance: Enterprise process for data quality stewardship and governance for real-world data

EINSTEIN Albert Einstein College of Medicine

Parsa Mirhaji, Selvin Soby, Erin Henninger, Chandra Nelapatla, Manuel Wahle, Boudewijn Aasman, Eran Bellin

Montefiore

BACKGROUND

Ascertaining and maintaining high quality of data is the key prerequisite for establishing a reliable and usable clinical data warehouse for research, discovery, and collaboration. It is also one of the most challenging and resource intensive processes to establish, maintain and scale especially in large enterprises where real-world observational health data is to be used by multiple clinical, operational, research, and collaborative stakeholders.[1,2] Supporting multitudes of analytic, data science, and operational use-cases throughout distributed teams of informaticians, biostatisticians, data scientists, research collaborators, and operational administrators requires a robust, transparent, and scalable data quality stewardship, governance process that would institutionalize data literacy and culture and give rise to data-driven research networks that can support learning healthcare systems.

In this poster we will introduce an enterprise approach to characterize, assess, and ascertain data quality and to establish a transparent process for data stewardship, governance, and monitoring data quality for OHDSI/OMOP based clinical data warehouses (Figure 1).

Figure 1: Standard vs. Updated Data Ingestion Process

Figure 2: Mapping and Deploying Governance

METHODS

A: Metadata management: We have used principles of the Semantic Web and RDF/RDFS modeling to interrogate and represent the information and data model itself from the source as a machine understandable knowledge-graph suitable for querying, machine-reasoning, and linkage. This enables a loss-less and granular representation of the schema, relationships, metadata, data discovery, and optimized strategies for extraction and retrieval. The approach also facilitates automated data quality assessment contextualized to the type and semantics of the data and its relationships at the source.

B: Knowledge and terminology management: We have translated and represented the vocabularies available in OMOP and ATHENA through Simple Knowledge Organization System (SKOS) which is a W3C standard for representation of thesauri, dictionaries, and vocabulary systems for linkage and semantic web.[3,4] We used it to enable semantic indexing, mapping, and linkage of all metadata, logic, ETL processes, and their logs to the underlying OMOP concepts. Our SKOS model also provides semantic linkage and mapping to the SKOS representation of source vocabularies from the UMLS.[5] We have further extended this SKOS model to also represent custom concepts that are not currently present in OHDSI/ATHENA

C: Semantic indexing: We have developed a robust vocabulary mapping framework that utilizes a combination of domain expertise provided by informatics analysts, mapping information provided by source systems, heuristics, and NLP for terminology mapping. For custom concepts that may not directly map to an OHDSI/ATHENA terminology system, we extend our SKOS based knowledge representation to provide a semantic model consistent and linked to OHDSI/ATHENA.

D: Natural language processing: We have built an integrated and scalable NLP process (using the Elastic Search and cTAKES Open-Source tools) based on semantic indexing of clinical and biomedical concepts using OHDSI/ATHENA vocabularies to provide automated suggestions during the concept mapping process. Recommended concepts are presented to analysts based on text-based matching (Figure 5).

E: Quality Characterization: We have developed a data quality characterization and monitoring process (QualiPy) that uses machine learning, pattern recognition, and big-data analytics to characterize and identify anomalies in longitudinal data. The process uses signal detection by using both a generalized linear model and an isolation forest-based machine-learning algorithms. Rather than treating data quality information as snapshots from a moment in time, QualiPy measures and characterizes data quality over time. This includes different aspects of the quality of the underlying data such as conformance, completeness, integrity, variability, and overall trends (Figure 4).

F: Information visualization: We have developed a series of visualizations and interactive web-based user-interfaces to expose the underlying metadata, mappings, and a data quality characterization report that comparatively illustrates source and destination data, and their prospective change over-time (Figure 3).

G: Validation and review: All mappings, data quality reports (as characterized at the source and within the OMOP-CDM) are reviewed by two domain experts by an informaticist and is either approved for deployment or rejected in order to further analyze and make corrections. Audit histories of all reviews and validation steps are logged and tracked to help maintain data governance.

H: Governance: All metadata, mappings, quality reports (including changes over time), validation, and review processes are tracked, logged, and transparently available to all users and administrators or the system (including but not limited to informatics support teams, information technology support teams, and enterprise data governance teams) (Figure 3).

I: Systems Overview: Our software platform (Interrogator) was developed to implement an end-to-end process to keep OMOP/OHDSI based clinical data warehouses updated with heterogeneous real-world observational health data. This platform enables informatics analysts with advanced knowledge management, semantic indexing, and quality assessment tools. The platform simultaneously enables researchers with transparent and granular information about quality of underlying data for their use-cases. History of all knowledge and data management interactions are recorded and made available to our stewards for data governance.

Figure 3: Web-Based, Interactive Data Quality Dashboard

Figure 4: Numerical Anomaly Detection Over-Time

Figure 5: *Interrogator* to Aid in Mapping Decisions

RESULTS

OHDSI/OMOP has become the primary clinical data warehouse technology for enterprise use across Montefiore Health System and is currently supporting greater than 300 users with hundreds of unique cohorts, and 6 multi-center research consortiums. The pursuit of transparent data quality, and governance and scalable data management has given rise to adoption of OHDSI/OMOP as our strategic resource that supports data science, informatics research, research informatics operations, centers of excellence for patient recruitment and clinical trial mapping, construction of disease specific registries, internal and external collaborations between care teams and investigators, and has inspired innovations and collaborations that were not previously possible. To achieve the full potential of Real-World-Data (RWD) analysis for science, we have implemented and extended our instance of ATLAS to democratize access to high quality clinical data, while preserving confidentiality and security of personal health information.

CONCLUSION

We have developed an end-to-end process to support and transparently represent a modern, scalable quality monitoring, data stewardship, and governance process for OMOP/OHDSI based clinical data warehouses to support enterprise use-cases for RWD research, care coordination operations, and collaboration.

REFERENCES

- 1.Kahn MG, Brown JS, Chun AT, Davidson BN, Meeker D, Ryan PB, Schilling LM, Weiskopf NG, Williams AE, Zozus MN. Transparent reporting of data quality in distributed data networks. EGEMS (Wash DC). 2015 Mar 23;3(1):1052.
- 2.Kunaparaddy N, Mirhaji P, Richards D, Casscells SW. Information integration from heterogeneous data sources: a Semantic Web approach. AMIA Annu Symp Proc. 2006;2006:992.
- 3.ATHENA Standardized Vocabularies – OHDSI n.d. https://www.ohdsi.org/analytic314 tools/athena-standardized-vocabularies/
- 4.Achard F. XML, Bioinformatics and Data integration. Bioinformatics. 2001;17:115–125.
- 5.Humphreys BL, Lindberg DAB, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc 1998; 5 (1): 1–11.

THURSDAY Data Quality Monitoring, Transparency and Governance: Enterprise process for data quality stewardship and governance for real-world data (Parsa Mirhaji, Selvin Soby, Erin Henninger, Chandra Nelapatla, Manuel Wahle, Boudewijn Aasman, Eran Belin)

#OHDSISocialShowcase This Week

Federated Patient-Level Prediction

1) integrated a common data model and federated learning process
2) federated prediction models with a PLP package

Byungjin Choi *, Dong Yun Lee †, Chungsoo Kim *, Jimyung Park *, Rae Woong Park*, †

INTRO

- Classic machine learning can learn only from centralized data, so the multi-institutional data use is limited.
- Federated learning is a method of developing a multi-institutional model by sharing only weights without sharing data.
- No common pipeline for feature extraction, so researchers at individual institutions had to manually extract
- Our aim is creating framework that unifies clinical data extraction and federated learning

METHODS

1. Training phase

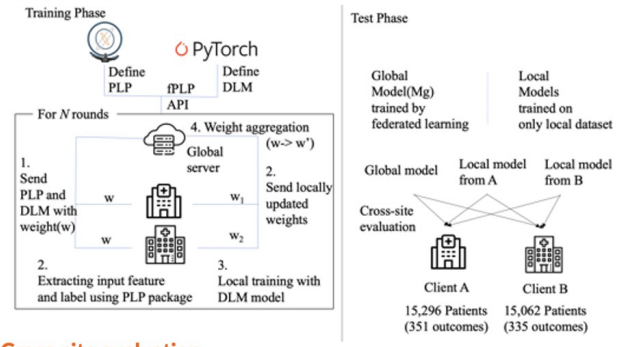
- Specifying cohort and feature settings using ATLAS. Settings are sent to client server
- Using defined PLP settings, labelled dataset were extracted from local OMOP CDM database
- Federated learning started. In each round, the server sends the global model weight to individual clients, global model weights are trained with local cohort and features. Then, only updated weights are sent to the global server. Global server aggregates client weights and makes up new global weights. This process is repeated for a predefined number of rounds by the Flower package.

2. Test Phase

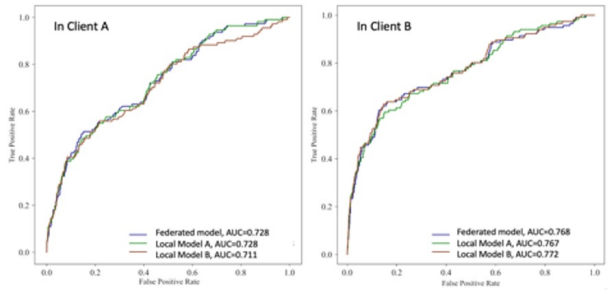
- In the test phase, the global model created through federated learning and all client models trained only with individual client data is distributed to every client server
- In each local server, the global model and all local models are validated on the each client test dataset. This process is called cross-site evaluation.

Federated Patient-Level Prediction : Federated learning library integrated with Patient-level prediction

- **Study Overview**



- **Cross-site evaluation**



3. Proof-of-concept

- For proof-of-concept, we simulated two clients with the Ajou University School of Medicine(AUSOM) CDM database
- Develops federated prediction model of acute kidney injury after coronary intervention .

RESULTS

- In the test dataset of client A, the global model, local model A, local model B showed AUC 0.728, 0.728, 0.711. And in the test dataset of client B, global model, local model A, local model B showed AUC 0.768, 0.767, 0.772
- The local model showed an average AUC reduction of 0.011 in cross-site evaluation. In contrast, the federated learning model shows an performance decrease of AUC 0.002
- This suggests that the federated model has generalizability while showing performance very close to the models trained and tested in same clients
- In further, we plan to develop and validate a federated model in a real-world multi-institutional environment

QR code to **Github page**

*Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Gyeonggi-do, Republic of Korea

†Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Gyeonggi-do, Republic of Korea

FRIDAY Real world prescribing patterns of dupilumab for atopic dermatitis (Torunn Sivesind, Grace Bosma, Camille Hochheimer, Lisa Schilling, Robert Dellavalle)

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?

Join Our Workgroups

OHDSI workgroups are always seeking new collaborators. If you are interested in **Joining The Journey** with any of our workgroups, **please visit our sign-up page (see link below or QR code)** and join our global collaborators in the mission to generate the real-world evidence that promotes better health decisions and better care.

ohdsi.org/workgroups

OHDSI Workgroup Objectives and Key Results (OKR)

2023 Update

Clinical Trials Workgroup leads: Mike Hamidi, Zhen Lin

CTWG Purpose

Objective: To allow adequate representation of clinical trial data represented as CDISC SDTM in OMOP.

Approach: We advocate minimum changes to the OMOP CDM and Standardized Vocabularies because we want to ensure minimum impact on OHDSI tools like Atlas, whilst providing a value-add SDTM-to-OMOP conversion with minimum data loss. We have proposed conventions introducing new concepts and modifiers, but no new CDM tables; and providing guidance for ETL developers where appropriate. Our proposals were originally built on OMOP CDM v6 and the Oncology extension, with v5.3 backward compatibility. In a new v5.4 the additions from the Oncology extension became standard, which made our changes minimal, thereby, making our proposals fully compatible with v5.4.

https://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:clinicalstudy CTWG Overview

CTWG Accomplishments

2023: Evaluating a single Vivli clinical study and developing initial high-level conceptual mappings between SDTM-to-OMOP (Persons, Procedure_Occurrence, etc.)

2022: CTWG was given access to 20 Vivli clinical study packages in the SDTM format. The CTWG team is doing an inventory of those study packages in order to prioritize SDTM-to-OMOP mappings. The existing CTWG guidance topics will be further assessed, and new ones identified where necessary.

2021: CTWG did an assessment of clinical trial data providers where SDTM data could be accessed. This eventually led to discussions with Vivli (i.e., general data usage agreements and platform feasibility evaluation).

2020: Used a synthetic representation of the CDISC SDTM data via PHUSE Test Data. Initial guidance topics were codified but require further testing with diverse real world SDTM data. The CTWG proposals submitted to the OHDSI community in July 2020.

CTWG Challenges

- Constraints by WG in accessing clinical study data within the Vivli environment (i.e., limited number of team members)
- Vivli environment time constraints (i.e., free access for one year, then pay for access)
- Installing needed software in the Vivli environment
- Working with obfuscated study data
- Pivoting strategy from mapping-to-execution to simply conceptual mapping guidance
- Access to less restricted SDTM study sources

CTWG OKR

Objective: To define the conceptual mappings and guidance to support CDISC SDTM-to-OMOP conversion

- Key Result #1: Identify >=3 real-world SDTM clinical studies
- Key Result #2: Develop conceptual SDTM-to-OMOP mapping specifications using a prioritized set of common SDTM domains (adverse events, vital signs, demographics, concomitant medications, laboratory test results, medical history, and procedures)
- Key Result #3: Publish draft SDTM-to-OMOP guidance by Q1 2024
  - Conceptual mappings on key domains of interest
  - Topic based best practices format
  - Identified gaps, issues, and challenges

CTWG Ask

- Additional sources of real-world clinical studies in SDTM format
- Any volunteers to support SDTM-to-OMOP high-level concept mappings
- Any organization active working on SDTM-to-OMOP conversions that have lessons learned outcomes

WG Name: OHDSI Vaccine Vocabulary WG

WG Lead: Asiyah Lin & Yongqun “Oliver” He

Objective 1 : Build up a consensus model of vaccines for OHDSI needs.

Key results:

1. Summarize or compare current models of vaccine representations in different standards such as Rx-Norm, Rx-Norm extension, SNOMED, CVX, and the Vaccine Ontology (VO). Timeline: 1Q2023.
2. Develop a consensus model of vaccines for OHDSI needs. Timeline: 1Q2023.

WG Name: OHDSI Vaccine Vocabulary WG

WG Lead: Asiyah Lin & Yongqun “Oliver” He

Objective 2 : Leverage existing works to map different vaccine representations using the consensus model developed in Objective 1.

Key results:

1. Identify methodology to achieve accurate mapping. Timeline: 2Q2023.
2. Use the identified method to establish vaccine term mapping for OMOP use. Timeline: 2-3Q2023

WG Name: OHDSI Vaccine Vocabulary WG

WG Lead: Asiyah Lin & Yongqun “Oliver” He

Objective 3 : Incorporate and evaluate the Objective 2 mapping results to OMOP vocabulary.

Key results:

1. Incorporate the Objective 2 mapping results to OMOP vocabulary. Timeline: 2-4Q2023
2. Evaluate the Objective 2 mapping results to OMOP vocabulary. Timeline: 3-4Q2023
3. Present at the OHDSI symposium. Timeline: 4Q2023.

WG Name: OHDSI Medical Device WG

WG Lead: Asiyah Lin & subgroup leaders

Objective 1 : Expand the leadership team and establish collaborations across OHDSI and beyond

Key results:

1. 1Q2023 : Establish subgroups (device generated data, device data and device adverse events) and leadership teams.
2. 1Q2023 : Respond to FDA medical device active surveillance RFI by Mar. 30, 2023.
3. 2-3Q2023: Develop activities to establish collaborations with other related WG or efforts: Sugery WG and Ehden
4. 3Q2023: Plan Think-a-thon or Hackathon at the OHDSI annual symposium

WG Name: OHDSI Medical Device WG - Device Data subgroup

Subgroup Lead: Anthony Molinaro & Carrie Bosela

Objective 2 : Enable the device standardization efforts to be interoperable with OMOP to support large scale device data analysis

Key results:

1.1-2Q 2023: Explore current OHDSI datasets for device data coverage.

2.1-2Q 2023 Explore and evaluate by extending OMOP by adding a device table

3.2-3Q2023: Explore tools and method to include device data in OMOP vocabulary

WG Name: OHDSI Medical Device WG -Device Generated Data subgroup

Subgroup Lead: Andrew Williams, Manlik Kwong

Objective 3: Develop standard strategy for managing and representing features waveform and other device-generated data:

1. Clarify OMOP Standard concept coverage gaps for features from 12-lead ECG Data and ICU monitor data
2. Develop strategy for addressing concept gaps
3. Test previously developed strategy for mapping covered concepts using MIMIC-4 Waveform Database waveform and "numerics" data

WG Name: OHDSI Medical Device WG - Device Adverse Event subgroup

Subgroup Lead: vacant

Objective 4: Establish the subgroup, identify leaders, and develop OKR

Key results:

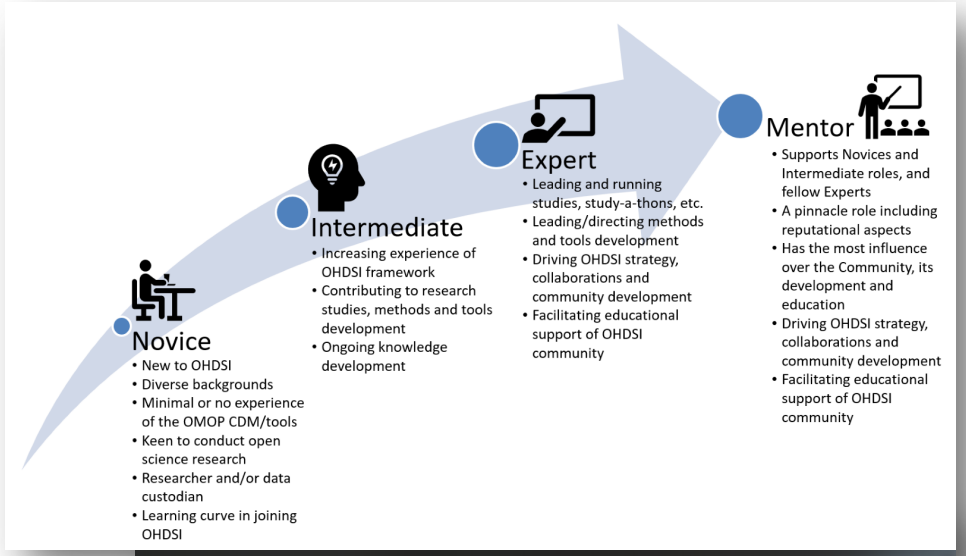
1. 1Q2023 : Identify leader for this group.
2. 2Q2023: develop OKR

Education WG Update: Purpose

- Education WG exists to support the community to address the learning curve from novice to mentors within the OHDSI research framework of tools, skills and methods utilised for quality observational research using OMOP CDM-mapped datasets

- Through guidance, signposting, materials and collaboration, the WG aims to support community members through relevant learning pathways

- We meet 4th Friday of the month

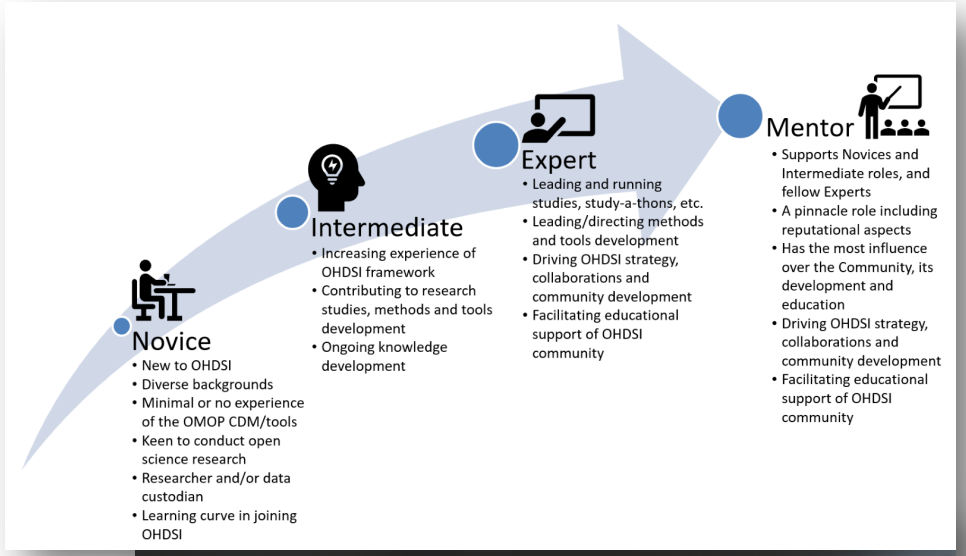


Education WG Update: OKR1

- For the first half of 2023, optimise the use and understanding of standard terminology within the OHDSI community, as measured by (Lead - Kristin):

a. Promotion and user rates of an updated glossary of terms visa OHDSI.eu, EHDEN Academy, et al

b. Uptake of the directory by OHDSI training providers

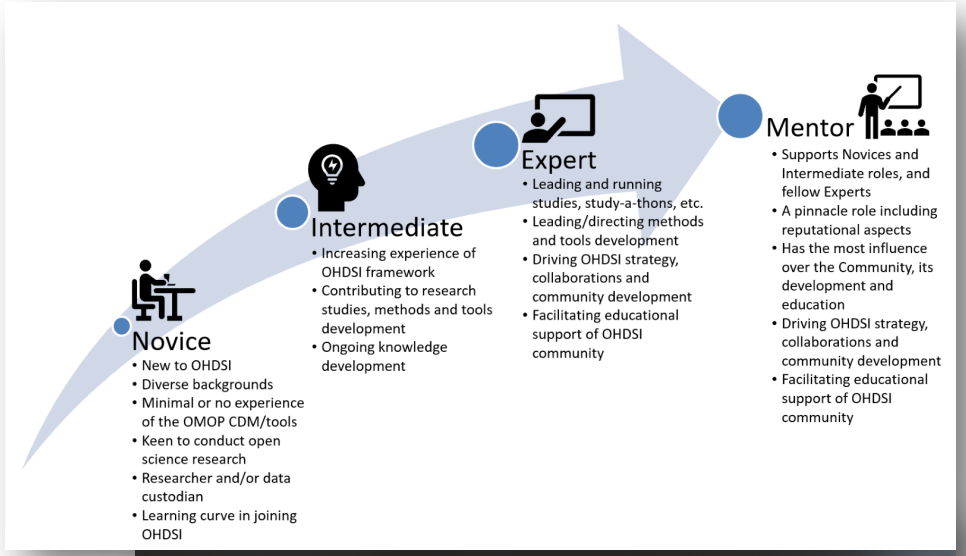


Education WG Update: OKR2

- For the first half of 2023, increase the exposure of training and education facilities within the OHDSI community, supporting colleagues needing to upskill and learn, as measured by (Lead - Paul):

a. Promotion and user rates of a new training and education directory via OHDSI.eu, EHDEN Academy, et al

b. Increase in uptake of materials and courses from directory providers above planned 2023 activity

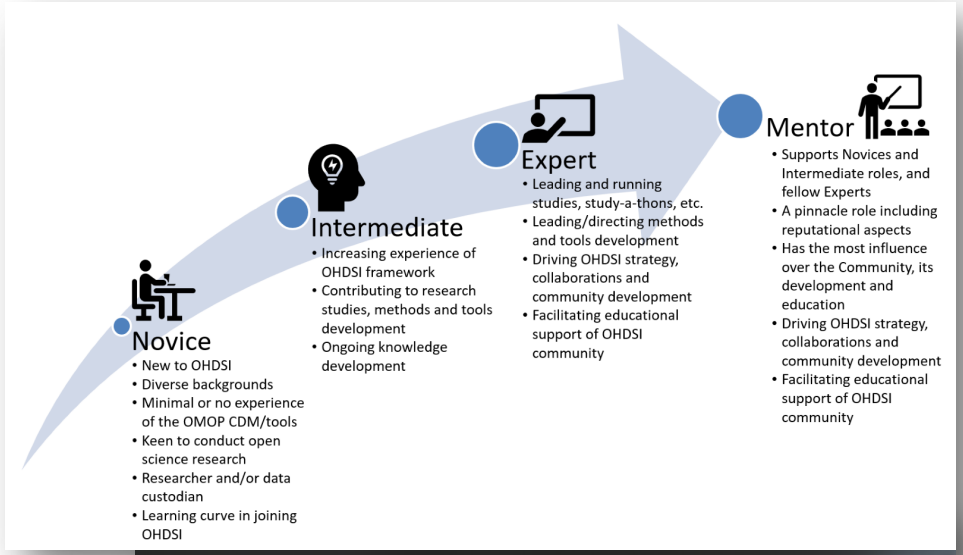


Education WG Update: OKR3

- For this year, focus the OHDSI community on a common view of what is required to effectively participate in the community and research, as measured by (Lead – Nige):

a. Development and uptake of a basic learning pathway – what is required learning across the study workflow and open science approach, with community input – launched c. Q2

b. Implementation of relevant learning blocks via training and education providers, prompted within the community and the directory, and their uptake



FHIR+ OMOP WG Purpose

To facilitate the collaboration between OHDSI and HL7 agreed by both parties in 2021. The work group will develop and validate standard transformation specifications and canonical maps between data conformant to FHIR to OMOP CDM, and from OMOP CDM to FHIR.

FHIR + OMOP WG Accomplishments

- Convened 3 broad-based community meetings engaging stakeholders.
- Formed 4 subgroups, each meeting on a weekly or bi-weekly basis composed of community members from the HL7 and OHDSI communities.
- Participation in 2 HL7 FHIR Connectathons demonstrating Oncology Use Case transformation from FHIR to OMOP.
- Convened community calls / exploration of 2 community generated use-cases: Digital Quality Measurements & Oncology.
- Developed system architecture & functional requirements for Digital Quality Measurement validation using OMOP & FHIR.

- Developed set of requirements for OMOP / FHIR Harmonization required to support Oncology Use Case.
- Developed proposal / approach for utilization of OMOP Vocabulary on FHIR (FHIR Extension).
- Identified semantic and structural patterns required for model harmonization between FHIR & OMOP CDM.
- Collaborated with Vulcan FHIR to OMOP project.
- Convened day-long workshop at OHDSI Symposium

FHIR + OMOP 2023 Objectives

- Consolidate the 4 subgroups into one working group and synthesize outputs from the subgroups and prior HL7 IGs work into a draft specification transforming OMOP v5.4 to FHIR R5 for core EMR data elements
- Develop draft (i.e. for broader consultation) specification (FHIR extension) for hosting OMOP Vocabulary on a FHIR Terminology Server
- Convene one (or more) Hack- / Transform-athon meeting(s) to validate and improve generated specifications

OHDSI Medical Imaging Working Group

From pixels to Phenotypes

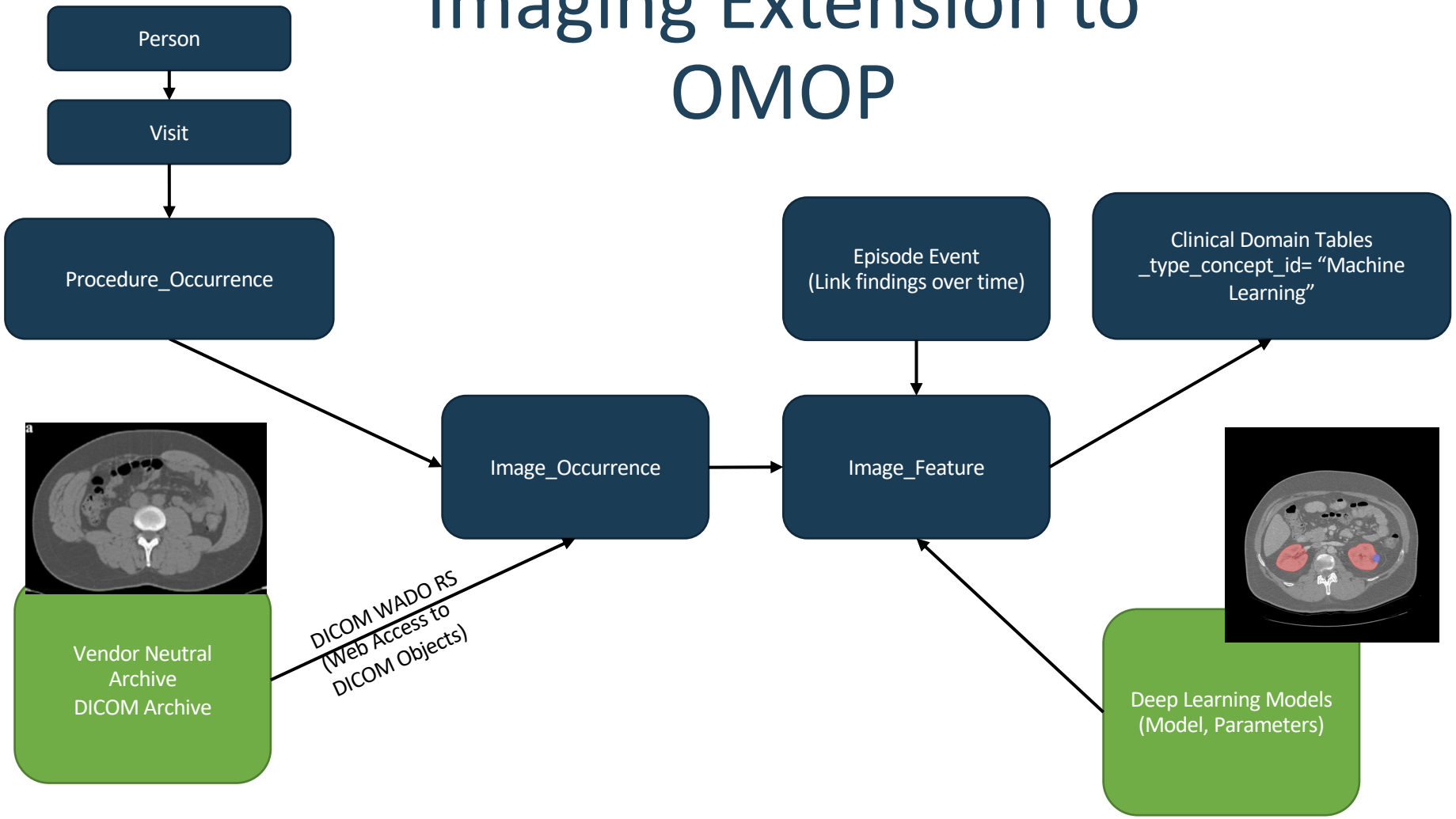
WG co-leads Seng Chan You and Paul Nagy

Wednesdays every 2 weeks at 7 AM / 7 PM

Imaging WG Goals

1. Extension to perform cohort definitions in OHDSI for medical imaging research studies.
2. Extension to bring features derived from medical images into the OMOP data model while maintaining provenance.
3. Create refence implementations of infrastructure for reproducible research on medical images.

Imaging Extension to OMOP



OKR #1

- Objective: Have CDM group approve the model into the base OMOP model Q3
- Key Results:
  - Publish a draft data model for the imaging extension Q1
  - Have Radlex and DICOM vocabularies added to the OMOP vocabulary Q2

OKR #2

- Objective: Conduct a network study based on the imaging extension Q4
- Key Results:
  - Have at least two reference implementations of this extension Q3
    1. Demonstration of cohort discovery
    2. Demonstration of imaging feature provenance
    3. Demonstration of combining EHR features in DCNN model building
    4. Demonstration of a network study
  - Write a roadmap how to implement imaging CDM and conduct network study

PRHeG (pronounced “preg”): Perinatal and Reproductive Health Group

Alison Callahan

Perinatal and Reproductive Health Group WG Purpose

The Perinatal and Reproductive Health Group (PRHeG) workgroup exists to develop tools and standards for perinatal and reproductive health research, to foster collaborative studies within the OHDSI network and advance research in the field.

Perinatal and Reproductive Health Group WG OKRs

1. Improve capture and representation of pregnancy and reproductive health data in the OMOP CDM

1. Complete a landscape assessment of how different institutions represent and organize pregnancy and reproductive health data
2. Produce a report summarizing PRHeG's consensus on best practices for doing perinatal and reproductive health research using multisite data in the OMOP CDM

2. Create an OHDSI data network of partners interested in perinatal and reproductive health research

1. Conduct at least 10 group meetings that include representatives from at least 10 different institutions on topics relevant to OKRs 1.1, 1.2
2. Pilot an initial research project with at least 5 institutions

PatientLevelPrediction (PLP)

Purpose and 2023 OKRs

We aim to establish a standardized process for developing accurate and well-calibrated patient-centered predictive models

The main research focusses are:

- Do methods research into best practices for prediction model development
- Apply our data, tools and framework to develop new clinically useful prediction models or validate existing ones
- Run network studies for methods research and clinical model development

Next meeting: Wednesday 8th March @ 9am ET

Objective: We should meet f2f to help further collaboration

Key result:

Organise work group meetings at:

1. European OHDSI Symposium
2. OHDSI Global Symposium
3. OHDSI APAC Symposium

Objective: We want set of benchmark problems

Key Results:

1. Have a moment in every workgroup meeting to discuss potential models
2. Identify 5-10 prediction tasks of interest
3. Add existing prediction models for the tasks of interest into DELPHI to make benchmarking easy

Objective: We would like to investigate learning models for rare outcomes

Key Results

1. Perform large scale study creating learning curves for stacker ensembles on new data
2. Perform large scale study creating learning curves for transfer learning on new data
3. Publish a paper comparing local model fitting, stacker ensemble and transfer learning on new data with rare outcomes

Objective: We want to better understand external validation

Key results

1. Develop tools to estimate external validation performance
2. Develop tools to understand external validation performance

Objective: We want to be able to locally update models

Key Results

1. Provide methods within the package to update models locally
2. Compare local to general models in terms of performance
3. Publish a paper on a framework for updating models locally
4. Develop a process for monitoring in situ model performance

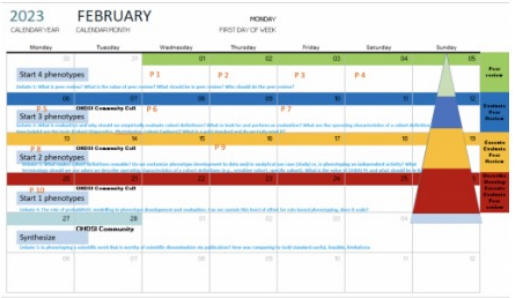
Objective: We want to be able to stratify PLP based on risk of outcome

Key results:

1. Add tools to be able to identify subgroups with different risks
2. Apply existing method for counterfactual deep learning as an OHDSI network study
3. Produce a paper looking at counterfactual prediction

Phenotype Phebruary Homepage

Phenotype Phebruary 2023: How To Join The Effort

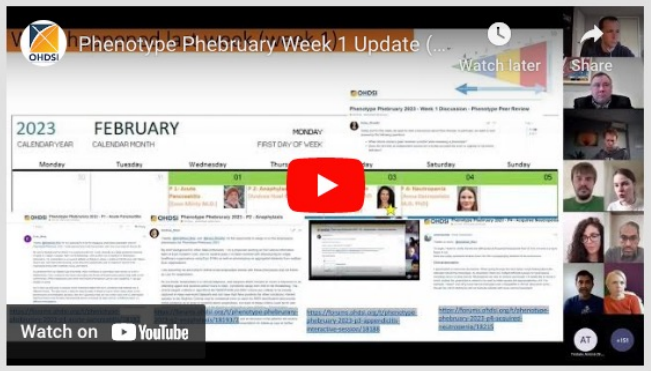


The schedule to the left lists the phenotypes that will be investigated throughout the month, along with the respective leads and reviewers. Check for updates to this graphic as more people join the effort. The graphic to the right highlights the four debates/discussions around phenotyping that are happening this month. Please use the forum links below to join any of these activities.

"Phenotype Phebruary" is a community-wide initiative to both develop and evaluate phenotypes for health outcomes that could be investigated by the community.

This is the second year of Phenotype Phebruary in the OHDSI community (look back at Year 1 here). It was introduced during the Jan. 31 community call (watch here), and will go on throughout the month. This year, the leadership team of **Gowtham Rao** and **Azza Shoaibi** helped identify 10 phenotypes that are being investigated throughout the month. If you would like to join the discussions around any of the phenotypes, please visit the appropriate links below, which will take you to the proper threads on the OHDSI forums.

Week 1 Update



Join Our Community Efforts Around Any Of These Phenotypes

(when phenotype threads get initiated, they will be added to the chart below)

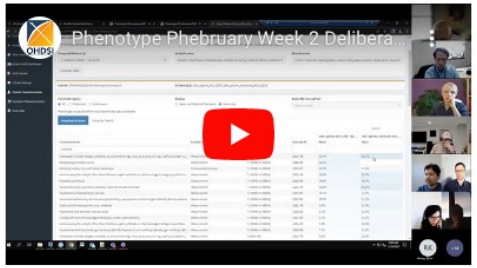
Announcements and Meeting/Workshop Links	Acute Pancreatitis	Anaphylaxis	Appendicitis
Acquired Neutropenia	Systemic Lupus Erythematosus	Acute Hepatic Failure	Idiopathic Inflammatory Myopathies
Parkinson's Disease	ST Elevation Myocardial Infarction	Neonatal Hypoxic Ischemic Encephalopathy	Neurofibromatosis type 1 with Optical Pathway Glioma

Join Our Community Discussions Around These Phenotype Phebruary Topics

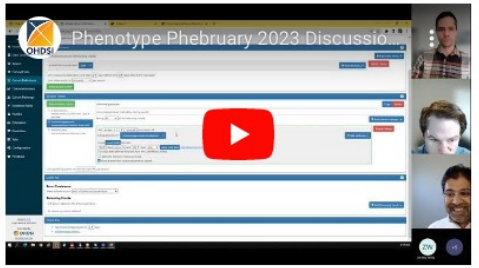
(when phenotype threads get initiated, they will be added to the chart below)

Phenotype Peer Review	Chart review gold standard validation vs innovative methods like PheValuator
What makes cohort definitions reusable, and what is the value of the OHDSI Phenotype Library? What should be in it?	The role of probabilistic modeling in phenotype development and evaluation

Phenotype Phebruary Videos



(Feb. 10) Week 2 of Phenotype Phebruary concluded with this OHDSI Phenotype Development and Evaluation workgroup meeting. In this session, the workgroup assigned leads to each phenotype that are



(Feb. 8) Christopher Mecoli, MD, and team demonstrated progress in the development of a cohort definition for Inflammatory Dermatomyositis at Johns Hopkins University. The team discussed

ohdsi.org/phenotype-phebruary-2023

Phenotype Phebruary Thank You!