



Meet The Titans

OHDSI Community Call
Nov. 1, 2022 • 11 am ET



Upcoming OHDSI Community Calls

Date	Topic
Nov. 8	Collaborator Showcase Presentations
Nov. 15	Open Network Studies
Nov. 22	10-Minute Tutorials
Nov. 29	Workgroup Updates
Dec. 6	Fall Publications
Dec. 13	How Did We Do In 2022?
Dec. 20	Holiday-Themed Final Call of 2022



Upcoming OHDSI Community Calls

Date	Topic
Nov. 8	Collaborator Showcase Presentations
Nov. 15	Open Network Studies
Nov. 22	10-Minute Tutorials
Nov. 29	Workgroup Updates
Dec. 6	Fall Publications
Dec. 13	How Did We Do In 2022?
Dec. 20	Holiday-Themed Final Call of 2022

Best Community Contribution Awards

Data Standards

Multi-institutional collaborative research using ophthalmic medical image data standardized by Radiology Common Data Model (R-CDM)

PRESENTER: ChulHyoung Park

INTRO

- Unstructured data which is beyond the scope of OMOP-CDM standardization is difficult to be used for multi-institutional collaborative research.
- Radiology Common Data Model (R-CDM) has been developed to standardize the terminology and structure of medical imaging data, which is representative unstructured data.
- In this study, a multi-institutional collaborative research was conducted by establishing an R-CDM database that standardized ophthalmic medical imaging data at two tertiary hospitals in Korea.

METHODS

- Standardizing optical coherence tomography (OCT) data into R-CDM format
 - Aju University School of Medicine (AUSOM)
 - Taken with ZEISS medical device during Jan 2013 - Apr 2022
 - Seoul National University Bundang Hospital (SNUBH)
 - Taken with HEIDELBERG medical device during Jul 2006 - Aug 2019
 - Standardize OCT data into R-CDM format (Figure 1)

- Design study to analyze changes in retinal thickness due to chronic disease
 - Patient cohort with hypertension (HTN), patient cohort with diabetes mellitus (DM), normal comparator cohort were created. Design of the HTN and comparator cohort can be seen in Figure 2.
 - Gender and age of the patient cohort and the control cohort were matched by conducting 1:2 propensity score matching (PSM) method.
 - OCT data of the left eye, which was taken last during the period in which the patient was in the cohort, was used for analysis.

- OCT data extraction through interworking of R-CDM and OMOP-CDM

- By linking OMOP-CDM and R-CDM, an environment has been established to extract specific image data taken by a specific patient cohort.
- The previously set hypertensive, diabetic, and control cohorts were constructed through OMOP-CDM, and then the OCT data they took were extracted through R-CDM.

- Retinal thickness data extraction using OCR technique
 - From the OCT result sheet of AUSOM, data was extracted using the easyOCR package of python. From the OCT result sheet of SNUBH, data was extracted using the OCR machine learning model developed in-house.

Multi-institutional collaborative research using ophthalmic medical image data standardized by Radiology Common Data Model (R-CDM)

Radiology Occurrence table					
study_id	person_id	study_date	modality	manufacturer	protocol_concept_id
1164861	1164861	2012-02-29	OCT	ZEISS	4213040 (Optical coherence tomography)
5158120	5215813	2012-08-15	OCT	ZEISS	4213040
3002305	1564510	2013-04-19	OCT	ZEISS	4213040

Radiology Image table					
image_id	series_id	study_id	series_type	source_value	file_path
12163542	3752729	1164861	RNFL analysis report	E:\R6212199\OCT15.docx	
4539345	7827354	5158120	Macular cube analysis report	F:\R3935248\OCT082.docx	
7867688	7867321	3002305	GCPL analysis report	F:\R3235248\OCT080.docx	

Figure 1. R-CDM Standardized OCT data

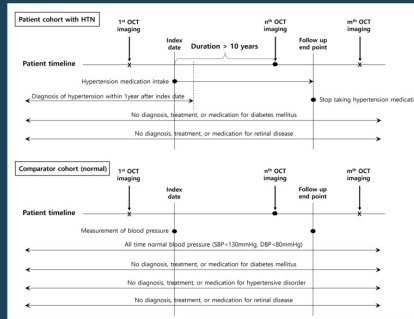


Figure 2. Design of the HTN and comparator cohort

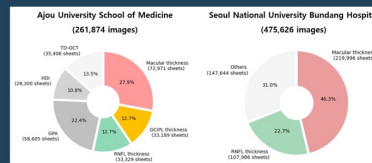


Figure 3. Composition of OCT data in each hospital

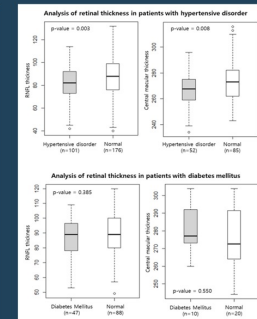


Figure 4. Differences of retinal thickness between cohorts

- The RNFL thickness and central macular thickness data of AUSOM and RNFL thickness data of SNUBH Hospital were successfully extracted and used for analysis.

RESULTS

1. Composition of R-CDM standardized OCT data

- 261,874 and 475,626 OCT data from AUSOM and SNUBH were standardized in R-CDM format.
- OCT data containing features of retinal thickness are central macula, GCPL, and retinal nerve fiber layer (RNFL) thickness reports, which are colored in red, yellow, and green, respectively (Figure 3).

2. Analysis of retinal thickness differences between cohorts (Figure 4)

2-1) Patient cohort with HTN VS comparator cohort

- The HTN cohort (101 patients) and control cohort (176 patients) each had an average RNFL thickness of 80.70µm, 86.80µm.
- The HTN cohort (52 patients) and control cohort (85 patients) each had an average central macular thickness of 265.73µm, 273.05µm.
- RNFL thickness, and Central macular thickness from hypertension cohort was significantly lower than that of the normal control cohort.

2-2) Patient cohort with DM VS comparator cohort

- There was no significant difference in RNFL thickness and central macular thickness between the DM cohort and the control cohort.

CONCLUSION

- In this study, OCT data of AUSOM and SNUBH were obtained for research purposes and standardized in the form of R-CDM.
- The retinal thickness was compared between the patients with chronic disease and the normal comparator cohort, and the retinal thickness was significantly lower in the patients with hypertension for more than 10 years.
- It is meaningful in that multi-institutional collaborative research which combines clinical and image data in various ways can be conducted very efficiently.

Chul Hyoung Park¹, Sang Jun Park², Da Yun Lee³, Seng Chan You¹, Su Ji Yeo⁴, Ki Hwang Lee⁴, Rae Woong Park^{1,5}

¹Department of Biomedical Informatics, Aju University School of Medicine

²Department of Ophthalmology, Seoul National University College of Medicine, Seoul National University Bundang Hospital

³Department of Biomedical Systems Informatics, Yonsei University College of Medicine

⁴Department of Ophthalmology, Aju University School of Medicine

⁵Department of Biomedical Sciences, Aju University Graduate School of Medicine



Analyzing the Effect of Hypertension on Retinal Thickness Using Radiology Common Data Model (R-CDM) (Chul Hyoung Park, Rae Woong Park, Sang Jun Park, Da Yun Lee, Seng Chan You, Ki Hwang Lee)

Best Community Contribution Awards

Methods Research

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

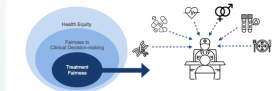


Abstract

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.

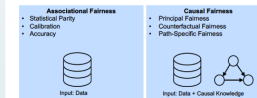
INTRODUCTION

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.



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.

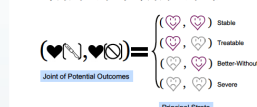
METHODS

1 Notations

- For the i -th patient,
- A_i : a sensitive attribute (e.g., sex)
 - D_i : a medical decision on treatment
 - $Y_i(0), Y_i(1)$: the potential outcome under no treatment and under treatment
 - H_i : principal strata, $H_i = (Y_i(0), Y_i(1))$
 - X_i : a vector of pre-treatment patient features.

2 Fairness Definitions

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



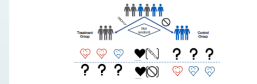
Principal fairness states that a decision is fair if patients who would benefit equally from the treatment have an equal probability of having the treatment regardless of the value of their sensitive attribute. The degree of violation is measured as $\Delta(D) = p(D_i = 1 | A_i = 1, H_i = h) - p(D_i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

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: $Y_i(0), Y_i(1) \perp A_i | X_i$
- Overlap: $0 < p(D_i = 1 | X_i = x) < 1 \forall x \in \mathcal{X}$
- Consistency: $Y_i(D_i) = Y_i$



3 Algorithm

Algorithm 1: Bayesian Principal Fairness Estimation Algorithm
Input: $D = \{D_i, A_i, X_i\}_{i=1}^n$
Output: $\Delta(D)$ (violation)
Estimate $q(A_i)$ with VI
Estimate $q(D_i)$ with VI
for $s = 1$ to S do
 Sample parameters from the posterior
 Estimate $q(A_i)$ with VI
 Estimate $q(D_i)$ with VI
 Estimate $q(Y_i(0), Y_i(1) | X_i, A_i, D_i)$ with VI
 Assign $H_i = (Y_i(0), Y_i(1))$
 Compute $\Delta(D)$ (violation)
end

SIMULATIONS

1 Setup

Simulate

$$A_i \sim \text{Bern}(0.5) \\ D_i \sim \text{Bern}(0.5) \\ Y_i(0) \sim \text{Bern}(e^{\beta_0} \theta_{Y_0} + \theta_0 | 0) \\ Y_i(1) \sim \text{Bern}(e^{\beta_1} \theta_{Y_1} + \theta_1 | 1),$$

where

$$\theta_{Y_0}, \theta_{Y_1} \sim N_m(0, 1), \theta_0 = -1.$$

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

$$D_i | H_i, A_i \sim \text{Bern}(p_{H,A})$$

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

Principal fairness

Fig 1. Associational fairness results from the simulation.

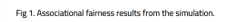


Fig 2(a). Principal fairness results from the simulation.

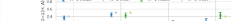


Fig 2(b). Proportion of principal strata from the simulation.



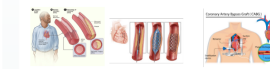
- The proposed algorithm is able to detect the unfair decision and estimate the level of unfairness correctly.
- The associational fairness metrics fail to detect the bias, deliver conflicting messages, and/or produce biased estimate of the degree of violation.
- A limitation that applies to all associational 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

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 3. Sex fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 4. Racial fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 5. Principal fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 6. Principal fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 7. Principal fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 8. Principal fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 9. Principal fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 10. Principal fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 11. Principal fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Fig 12. Principal fairness of revascularization in CAD.

(a) Statistical parity

(b) Calibration

(c) Accuracy

(d) Principal fairness

Assessing Racial Fairness of Dialysis Allocation in End-Stage Renal Disease (Linying Zhang, Lauren R. Richter, David M. Blei, Yixin Wang, Anna Ostropolets, Noemie Elhadad, George Hripcsak)



Best Community Contribution Awards



Cohort Definition Validation in Atlas

Charity Hilton MS¹, Saul Crumpton MS¹, Jon Duke MD, MS^{1,2}



¹Georgia Tech Research Institute, ²Georgia Institute of Technology

Open-Source Analytics

Background

OHDSI Atlas has long been an effective tool for developing rule-based cohort definitions in observational data. In the public version of Atlas, thousands of cohort definitions have been created. While patient record verification is a common method of cohort definition validation, it is not without difficulties, including but not limited to the need for clinical experts to access data, a tool to review all in-cohort patients, a method to gather review data, and a system of tabulation to determine in-cohort (case/no-case) participation or not¹.

Until now, there has not been an Atlas-based system for clinical expert review. For this effort, we introduce the Atlas Cohort Definition Validation tool (ACDV). This tool aims to solve some of the primary concerns around cohort definition validation, while having the chief benefit of being cohesively integrated into the OHDSI Atlas stack. Additionally, the tool allows for creation of more complex validation question sets, beyond the standard case/no-case assessment.



Figure 1: Question Set Creation

Methods

We designed and developed two modules around cohort definition validation. The first (1) allows for validation study creation and management, and the second (2) allows for validation of study questions for clinical reviewers in the Atlas Patient Profile tool.

The ACDV tool introduces a 'Validation' section to Atlas cohort definition creation, which allows for cohort managers to complete a cohort definition validation workflow. This workflow begins by the creation of question set. Question sets in the ACDV tool, shown in Figure 1, allow for common types of questions (including text, radio, checkbox, numbers, and dates). Multiple questions in a question set can be created and a case/no-case distinction can be selected at the question level. After a question set has been created, it can be linked to a cohort definition sample, this creates the validation study.

After a validation study is created, cohort managers can assign patients for review in the Atlas Patient Profile tool to clinical reviewers. Study questions are displayed to clinical reviewers at the patient level in a collapsible sidebar (see Figure 3). The study question set at the patient profile-level can be accessed via the Cohort Definition tool, the Patient Profile tool, or via a customized link. Once reviewers have viewed patient profiles and answered study questions, study results can be viewed by cohort managers in Atlas or exported to CSV (Figure 4).

Results

Primary development efforts of the ACDV tool are complete, and final modifications and integrations to the tool are being prepared for inclusion in an upcoming OHDSI release. We have validated the tool internally with a clinician-informaticist.



Figure 2: Annotation Study Manager View



Figure 3: Profile Level Validation

Conclusions

The Atlas Cohort Definition Validation tool will provide an integrated way for clinical chart reviewers to validate cohorts well beyond the question of cohort inclusion or not.

This tool will support research in the OHDSI community by living firmly within the active OHDSI Atlas ecosystem of tools. Additionally, this tool will continue the OHDSI legacy of open and community-driven tools to advance research in observational health data.

ID	Name	Status	Date
1	Study 1	Completed	2023-01-01
2	Study 2	In Progress	2023-01-02
3	Study 3	Pending Review	2023-01-03
4	Study 4	Completed	2023-01-04
5	Study 5	In Progress	2023-01-05
6	Study 6	Pending Review	2023-01-06
7	Study 7	Completed	2023-01-07
8	Study 8	In Progress	2023-01-08
9	Study 9	Pending Review	2023-01-09
10	Study 10	Completed	2023-01-10

Figure 4: Study Results

Bibliography

1. Observational Health Data Sciences and Informatics. The Book of OHDSI; 2020. Available from: <https://ohdsi.github.io/TheBookOfOhdsi/>

Cohort Definition Validation in Atlas (Charity Hilton, Saul Crumpton, Jon Duke)



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!






Congratulations to the team of **Philipp Wegner, Geena Mariya Jose, Vanessa Lage-Rupprecht, Sepehr Golriz Khatami, Bide Zhang, Stephan Springstubbe, Marc Jacobs, Thomas Linden, Cindy Ku, Bruce Schultz, Martin Hofmann-Apitius, Alpha Tom Kodamullil** for the **COPERIMOpus Consortium** on the publication of **Common data model for COVID-19 datasets** in *Bioinformatics*.

Bioinformatics, 2022, 1–3
<https://doi.org/10.1093/bioinformatics/btac651>
Advance Access Publication Date: 27 October 2022
Applications Note



Databases and ontologies

Common data model for COVID-19 datasets

Philipp Wegner ¹, Geena Mariya Jose², Vanessa Lage-Rupprecht¹, Sepehr Golriz Khatami¹, Bide Zhang¹, Stephan Springstubbe¹, Marc Jacobs ¹, Thomas Linden¹, Cindy Ku¹, Bruce Schultz¹, Martin Hofmann-Apitius ^{1,3,*} and Alpha Tom Kodamullil^{1,3,*}; and for the COPERIMOpus Consortium

¹Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, 53757, Germany,

²Causality Biomodels, Kinfra Hi-Tech Park, Cochin, Kerala 683503, India and ³Bonn-Aachen International Center for IT (B-IT), Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn 53115, Germany

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on January 21, 2022; revised on April 19, 2022; editorial decision on August 19, 2022; accepted on October 13, 2022

Abstract

Motivation: A global medical crisis like the coronavirus disease 2019 (COVID-19) pandemic requires interdisciplinary and highly collaborative research from all over the world. One of the key challenges for collaborative research is a lack of interoperability among various heterogeneous data sources. Interoperability, standardization and mapping of datasets are necessary for data analysis and applications in advanced algorithms such as developing personalized risk prediction modeling.

Results: To ensure the interoperability and compatibility among COVID-19 datasets, we present here a common data model (CDM) which has been built from 11 different COVID-19 datasets from various geographical locations. The current version of the CDM holds 4639 data variables related to COVID-19 such as basic patient information (*age, biological sex and diagnosis*) as well as disease-specific data variables, for example, *Anosmia* and *Dyspnea*. Each of the data variables in the data model is associated with specific data types, variable mappings, value ranges, data units and data encodings that could be used for standardizing any dataset. Moreover, the compatibility with established data standards like OMOP and FHIR makes the CDM a well-designed CDM for COVID-19 data interoperability.

Availability and implementation: The CDM is available in a public repo here: <https://github.com/Fraunhofer-SCAI-Applied-Semantics/COVID-19-Global-Model>.

Contact: alpha.tom.kodamullil@scai.fraunhofer.de

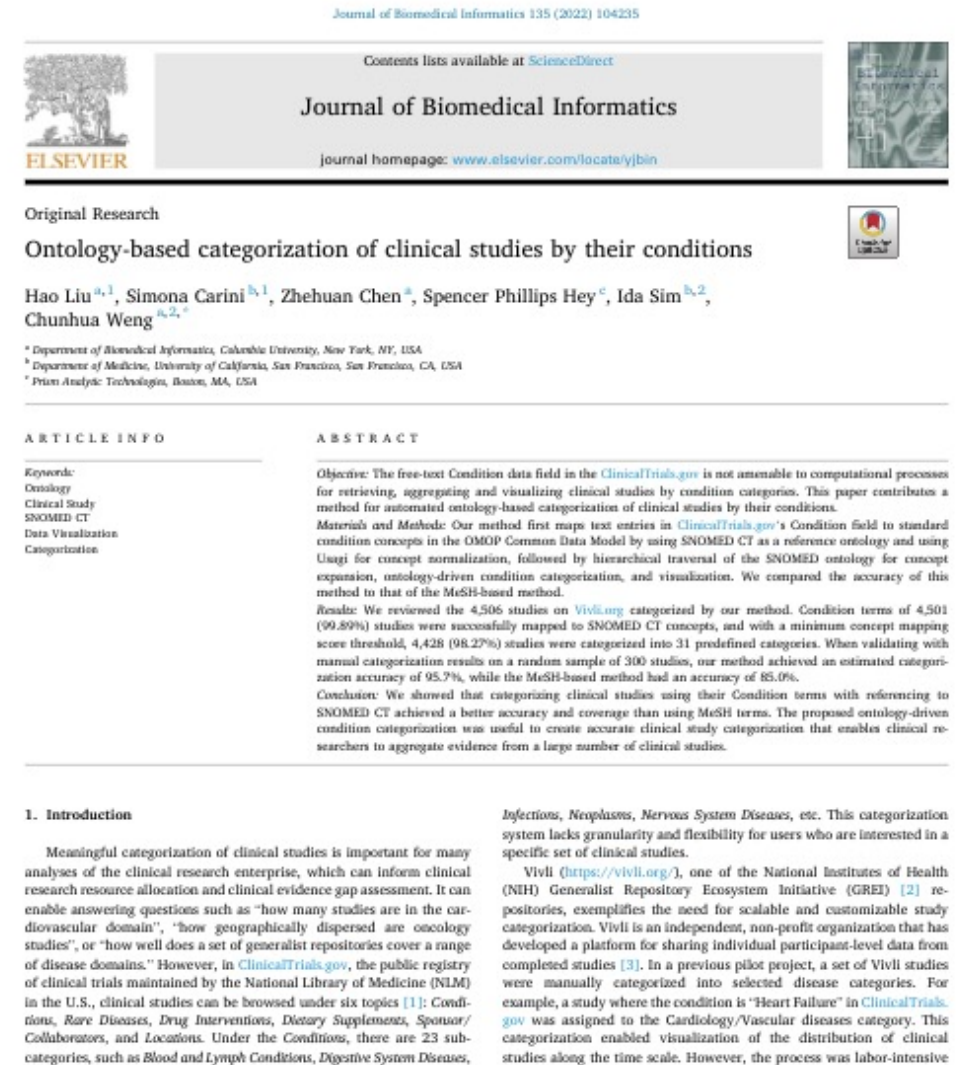
Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.



OHDSI Shoutouts!



Congratulations to the team of
**Hao Liu, Simona Carini,
Zhehuan Chen, Spencer Phillips
Hey, Ida Sim, and Chunhua
Weng** on the publication of
**Ontology-based categorization
of clinical studies by their
conditions in the Journal of
Biomedical Informatics.**





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
Wednesday	2 am	Population-Level Estimation
Wednesday	7 am	Medical Imaging
Wednesday	8 am	Psychiatry
Wednesday	9 am	ATLAS
Wednesday	12 pm	FHIR and OMOP Terminologies Subgroup (ZOOM)
Wednesday	12 pm	Health Equity
Wednesday	5 pm	FHIR and OMOP Data Model Harmonization Subgroup (ZOOM)
Thursday	12 pm	Population-Level Estimation
Thursday	1 pm	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	7 pm	Dentistry
Friday	9 am	GIS – Geographic Information System
Friday	9 am	Education
Monday	10 am	Healthcare Systems Interest Group
Tuesday	9 am	OMOP CDM Oncology Genomic Subgroup

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



OHDSI Dentistry Workgroup Objectives and Key Results (OKR)

Lead: Robert Koski



Workgroup Name: Dentistry Workgroup

Workgroup lead: Robert Koski

1. Expand the workgroup

1Q2023 Key Results:

1. Recruit a co-lead
2. Recruit at least three regularly attending members

2. Complete a scoping review of observational research in dentistry and map a dental use case to the OMOP-CDM

1Q2023 Key Results:

1. Refine and complete scoping review
2. Develop use case(s) further and share Miro implementation with OHDSI
3. Conduct gap analysis of use case mapping to determine future objectives
4. Submit scoping review for publication to JADA or JAMIA



Dentistry Work Group

Thursdays at 7PM ET on MS Teams



Latest Edition Of The Journey Newsletter



The Journey Newsletter (November 2022)

The 2022 OHDSI Symposium brought together more than 400 collaborators from around the world to share ideas, learn from each other, and have fun, all in the name of Building A Healthier World Together. All of the materials from the main conference and the full-day tutorial are now available in this newsletter, as well as plenty more updates from the OHDSI Community. [#JoinTheJourney](#)

November Video Podcast



In the latest On The Journey video, Patrick Ryan and Craig Sachson reflect on the OHDSI symposium weekend, including discussions on the presentations, collaborator showcase and weekend activities. (If the video does not appear, please click 'View this email in your browser')

Community Updates

Where Have We Been?

- The 2022 Symposium featured a plenary session on *Objective Diagnostics: A pathway to provably reliable evidence*, presentations on OHDSI support for regulatory authorities, a record-setting Collaborator Showcase with more than 120 posters and eight lightning talks, and a closing that focused on the path towards Building A Healthier World Together. All of these talks are now available [on our symposium homepage](#).
- [The Titan Awards](#) recognize OHDSI collaborators (or collaborating institutions) for their contributions towards OHDSI's mission, and [the 2022 honorees were announced](#) during the closing at the OHDSI Symposium.
- Volume 2 of *Our Journey: Where The OHDSI Has Been, And Where We Are Going* was introduced and distributed at the symposium. This book provides a high-level look at many aspects of OHDSI, including its mission, collaborators, data network, research, publications, and more.

Where Are We Now?

- Many of the 400+ collaborators at the symposium weekend were fairly new to the community and wanted to learn more, and 140 of them took part in the community's first full-day tutorial on [An Introductory Journey from Data to Evidence](#). There were eight sessions, and the videos and slides from each are now available on our tutorial homepage.
- Several OHDSI workgroups held meetings and activities during the symposium weekend, and more than 100 people connected for an all-hands meeting to discuss how workgroups could collaborate to address challenges around the community. If you are interested in joining the journey, [check out our workgroups page](#) and see where your interests and passions may align with ongoing OHDSI efforts.
- The Asia-Pacific (APAC) Symposium will be held Nov. 12-13 at the Taipei Medical University, although parts of the main conference on Nov. 13 will be streamed live. Day 1 will be focused on tutorials, while Day 2 will have talks and a collaborator showcase. More information and registration links [are available here](#).

Symposium Welcomes 400+ Collaborators In Hopes of Building A Healthier World Together



More than 400 community members from around the world connected Oct. 14-16 in Bethesda, Md., for the 2022 OHDSI Symposium, the first in-person global symposium since 2019. The weekend theme was "Building A Healthier World Together," and both the main conference and the weekend activities, including a full-day tutorial, highlighted the different ways OHDSI has impacted global healthcare, and the steps needed to be taken to build on that foundation.

October Publications

Kim Y, Seo SI, Lee KJ, et al. [Risks of long-term use of proton pump inhibitor on ischemic vascular events: A distributed network analysis of 5 real-world observational Korean databases using a common data model](#). *International Journal of Stroke*. 2022;0(0). doi: 10.1177/17474930221133219.

Nishimura Akihiko, Xie Junqing, Kostka Kristin, Duarte-Salles Talita, Fernández Bertolin Sergio, Aragón María, Blacketer Clair, Shoaibi Azza, DuVall Scott L., Lynch Kristine, Matheny Michael E., Falconer Thomas, Morales Daniel R., Conover Mitchell M., Chan You Seng, Pratt Nicole, Weaver James, Sena Anthony G., Schuermie Martijn J., Reps Jenna, Reich Christian, Rijnbeek Peter R., Ryan Patrick B., Hripcsak George, Prieto-Alhambra Daniel, Suchard Marc A. [International cohort study indicates no association between alpha-1 blockers and susceptibility to COVID-19 in benign prostatic hyperplasia patients](#). *Frontiers in Pharmacology*. Vol. 13. 2022. doi=10.3389/fphar.2022.945592. ISSN=1663-9812.


Fortin, S.P., Reps, J. & Ryan, P. [Adaptation and validation of a coding algorithm for the Charlson Comorbidity Index in administrative claims data using the SNOMED CT standardized vocabulary](#). *BMC Med Inform Decis Mak* 22, 261 (2022). <https://doi.org/10.1186/s12911-022-02006-1>.

Bae WK, Cho J, Kim S, Kim B, Baek H, Song W, Yoo S. [Coronary Artery Computed Tomography Angiography for Preventing Cardio-Cerebrovascular Disease: Observational Cohort Study Using the Observational Health Data Sciences and Informatics' Common Data Model](#). *JMIR Med Inform* 2022;10(10):e41503. doi: 10.2196/41503.

Vaclav Papez, Maxim Moinat, Erica A Voss, Sofia Bazakou, Anne Van Winzum, Alessia Peviani, Stefan Payralbe, Michael Kallfelz, Folkert W Asselbergs, Daniel Prieto-Alhambra, Richard J B Dobson, Spiros Denaxas, [Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond](#), *Journal of the American Medical Informatics Association*, 2022; ocac203, doi: 10.1093/jamia/ocac203.



Latest Edition Of The Journey Newsletter



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

[Who We Are](#) ▾[OHDSI Updates & News](#) ▾[Standards](#)[Software Tools](#)[OHDSI Studies](#) ▾[Book of OHDSI](#) ▾[Resources](#) ▾[New To OHDSI?](#) ▾[OHDSI Community Calls](#) ▾[Events & Past Collaborations](#) ▾[Workgroups](#) ▾[EHDEN Academy](#) ▾[This Week In OHDSI](#)[Our Journey \(Vol. 2\)](#)[Community Dashboard](#)[Support & Sponsorship](#) ▾[2022 Symposium](#) ▾[2022 APAC Symposium](#)[Newsletters](#) ▾[Follow OHDSI on Social](#) ▾

Subscribe

November 2022

October 2022

September 2022

August 2022

July 2022

June 2022

May 2022

Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data

Building the World

The 2022 OHDSI theme of "Building the World" and it featured



OHDSI 2022





OHDSI 2022

2022 OHDSI Symposium

Oct. 14-16 • Bethesda North Marriott Hotel & Conference Center



Thank you to everybody who joined us in Bethesda, Md., for the 2022 OHDSI Symposium. This page features all materials from the main conference, which focused on "Building A Healthier World Together." It includes videos of all the presentations, links to the collaborator showcase, and all slide decks from the speakers at the bottom of the page.

The weekend also featured an all-day tutorial on "An Introductory Journey from Data to Evidence." [Please visit our tutorial homepage](#) to visit those recordings and slides.

State of the Community

George Hripesak (Columbia University) opened the 2022 OHDSI Symposium with a reflection on the state of the OHDSI community. The talk provided a high-level overview of what OHDSI is all about, recent achievements, and how to become more active in the community. It uses the recent "Our Journey" publication as a guide, [which you can access here](#).

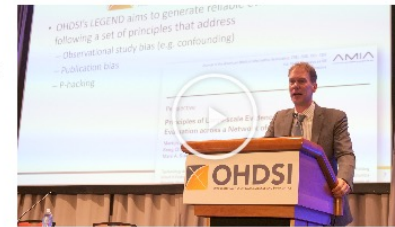
Following that talk, **Patricia Lloyd (US Food and Drug Administration)** gives a brief talk on safety monitoring of COVID-19 vaccines within the FDA BEST Initiative. OHDSI is proud to support FDA efforts in this area.



Plenary: Objective Diagnostics: A pathway to provably reliable evidence

The plenary presentation from the 2022 OHDSI Symposium was led by **Martijn Schuemie (Johnson & Johnson)** and focused on "Objective Diagnostics: A pathway to provably reliable evidence." **Patrick Ryan (Johnson & Johnson, Columbia University)** also took part in this session.

This session introduced a series of diagnostics that can be evaluated to determine database, phenotype, and analysis fitness-for-use for generating reliable evidence. The presentation demonstrates the empirical performance of these objective diagnostics across the LEGEND-HTN result set to illustrate how objective diagnostics can be used and how they improve the quality of evidence generated.



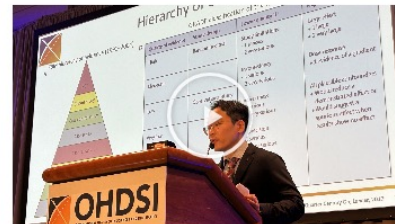
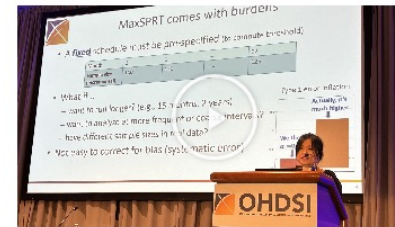
Presentations: OHDSI support for regulatory authorities

The 2022 OHDSI Symposium included a trio of presentations focused on OHDSI support for regulatory authorities. Each talk included an introduction from a regulatory agency representative.

Right – US FDA/CBER: Performance of vaccine safety surveillance methods (**Fan Bu, UCLA**)

Below, left – Korean National Institute of Food and Drug Safety Evaluation: Evolution of Evidence-Based Medicine: Why Do We Replicate Trials? (**Seng Chan You, Yonsei University**)

Below, right – European Medicines Agency: DARWIN-EU (**Peter Rijnbeek, Erasmus MC**)



Closing: Building A Healthier World Together

The 2022 Symposium Closing Talk was provided by **Patrick Ryan (Johnson & Johnson, Columbia University)**, who discussed the theme of Building A Healthier World Together, and how the OHDSI community can collaborate to achieve that goal. He focused on the scientific practices that can help this aim, but he also discussed the critical value of teamwork. That led to the first-ever OHDSI Lego Challenge, as well as a memorable lego stop-motion video (58:35 in the presentation).



This talk opened with the announcement of the 2022 Titan Award recipients. The Titan Awards are awarded annually to individuals or teams within the OHDSI community who have made significant contributions towards advancing OHDSI's mission, vision and values. They are both nominated and voted upon by community members, and they are announced at the global symposium. The recipients were:
Data Standards: **Melanie Philofsky, Odysseus Data Services**
Methodological Research: **Fan Bu, UCLA**
Open-Source Development: **Egill Fridgerisson, Erasmus MC** and **James Gilbert, Janssen Research and Development**
Clinical Applications: **Xintong Li, University of Oxford**
Community Collaboration: **Ajit Londhe, Boehringer Ingelheim**
Community Leadership: **Paul Nagy, Johns Hopkins University**
Community Support: **Craig Sachson, Columbia University**

Slidedecks From The Main Conference

Session I - State of the Community

- [State of the Community](#) (**George Hripesak, Columbia University**)
- [Safety Monitoring of COVID-19 Vaccines within the FDA BEST Initiative](#) (**Patricia Lloyd, US Food and Drug Administration**)

Session II - Plenary

- [Objective Diagnostics: A pathway to provably reliable evidence](#) (**Martijn Schuemie, Johnson & Johnson; Patrick Ryan, Johnson & Johnson/Columbia University**)

Session III - OHDSI support for regulatory authorities

- [US FDA/CBER: Performance of vaccine safety surveillance methods](#) (**Fan Bu, UCLA**)
- [Korean National Institute of Food and Drug Safety Evaluation: Evolution of Evidence-Based Medicine: Why Do We Replicate Trials?](#) (**Seng Chan You, Yonsei University**)
- [European Medicines Agency: DARWIN-EU](#) (**Peter Rijnbeek, Erasmus MC**)

Session IV - Lightning Talks

- [Disambiguation of ICDPC codes using free-text and active learning to improve concept mappings](#) (**Tom Seinen, Erasmus MC**)
- [OHDSI Phenotype Phebruary: lessons learned](#) (**Azza Shoaibi, Johnson & Johnson**)
- [Reduce, Reuse, & Recycle: Going Green with Atlas Reusables](#) (**Ajit Londhe, Amgen**)
- [Best practices for prognostic model development using observational health data: a scoring review](#) (**Cynthia Yang, Erasmus MC**)
- [Machine Learning for Predicting Patients at Risk of Prolonged Opioid Use Following Surgery](#) (**Behzad Naderalavjoud, Stanford University**)

ohdsi.org/ohdsi2022symposium



OHDSI 2022 Tutorial

OHDSI2022 Tutorial: An Introductory Journey From Data To Evidence

During the 2022 OHDSI Symposium, community leaders taught a full-day tutorial meant to introduce participants to the varying steps along the journey from data to evidence using the OMOP Common Data Model, OHDSI tools and scientific best practices.



In each 50-minute segment, the class learned the conceptual framing of the problem and approach to the solution. The class had the opportunity to gain hands-on exposure to design and implementation of analyses and interpretation of results. The course was motivated by a real use case: using observational data to generate evidence about the relationship between an exposure and outcome, and it highlighted how the suite of OHDSI tools and practices can enable such learning.

This class was designed for newcomers to the OHDSI community who were looking for a high-level summary across a wide range of topics covered within the OHDSI community. It was also designed for those in the OHDSI community who may be focused in one particular area of the journey, but who want exposure to the other areas, so they can better understand how their work contributes to be 'big picture.'

Videos and slides from the tutorial are all available on this webpage.

Tutorial Materials

1. Overview of the OHDSI Journey: Where are we going?

Faculty: Patrick Ryan

Video

Slides

3. Creating cohort definitions

Faculty: Asieh Golozar

Video

Slides

5. Characterization

Faculty: Kristin Kostka

Video

Slides

7. Prediction

Faculty: Jenna Reps

Video

Slides

2. OMOP Common Data Model & Vocabulary/ ETL a source database into OMOP CDM

Faculty: Clair Blacketer, Melanie Philofsky

Video

Slides

4. Phenotype evaluation

Faculty: Gowtham Rao

Video

Slides

6. Estimation

Faculty: Martijn Schuemie

Video

Slides

8. Where do we go from here?

Faculty: George Hripcsak

Video

ohdsi.org/ohdsi2022-tutorial



Healthcare System OMOP Adoption Survey

Do you represent a healthcare system that has adopted OMOP?

The Healthcare Systems Interest Group is gathering evidence to support additional healthcare systems' adoption decisions

We want to hear about the benefits your organization has realized

Please take our survey:

<https://bit.ly/OMOPAdopt>

John Methot, Melanie Philofsky, Brian J. Bush, Paul Nagy, Daniel Smith, Edward Smith



OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

#JoinTheJourney

69

OHDSI.org





2022 OHDSI APAC Symposium

Day 1 (Nov. 12) — Tutorial Workshop

8:30 – 9:00 • Registration
9:00 – 9:30 • Overview of the OHDSI Journey: where are we going
9:30 – 10:20 • OMOP Common Data Model and vocabulary
10:20 – 10:30 • Break
10:30 – 11:20 • ETL a source database into OMOP CDM
11:20 – 11:30 • Break
11:30 – 12:20 • Creating cohort definitions
12:20 – 13:30 • Lunch
13:30 – 14:20 • Phenotype evaluation
14:20 – 14:30 • Break
14:30 – 15:20 • Characterization
15:20 – 15:30 • Break
15:30 – 16:20 • Estimation
16:20 – 16:30 • Break
16:30 – 17:20 • Prediction
17:20 – 17:30 • Recap of the OHDSI Journey, where do we go from here

[Register for Day 1 Here](#)

Day 1 Registration Fees (In-Person)

International Student/Trainee: \$30
International Academia/Government: \$70
International Industry/Corporate: \$170
Local Registrants: Free

2022 APAC OHDSI Symposium

Nov. 12 - 13 • Taipei Medical University



Day 2 (Nov. 13) — Main Conference

08:00 – 09:00 • Registration & Light Breakfast
09:00 – 09:20 • Welcome Session
09:20 – 09:40 • Group Photo

Session 1: Envisioning of OHDSI Global & OHDSI APAC

09:40 – 10:00 • Keynote – OHDSI Global Presentation
10:00 – 10:20 • OHDSI APAC Introduction
10:20 – 10:30 • Break

Session 2: The Implication Experiences in OHDSI Region

10:30 – 11:30 • Researches in OHDSI APAC
11:30 – 11:45 • Researches using Taiwan National Data
11:45 – 12:00 • Researches using TMUCRD Data
12:00 – 13:00 • Lunch & Poster Presentation

Session 3: The Challenges of Research in OHDSI APAC

13:00 – 14:00 • Panel – Standardization & Common Data Models
14:00 – 15:00 • Panel – APAC Regional Adaption to Standardization
15:00 – 15:15 • Break
15:15 – 16:15 • Poster & Networking Session
16:15 – 17:00 • Closing Remarks

[Register for Day 2 Here](#)

Day 2 Registration Fees (In-Person)

International Student/Trainee: \$50
International Academia/Government: \$100
International Industry/Corporate: \$200
Local Registrant: Free

Day 2 Registration Fees (Virtual)

International Student/Trainee: \$25
International Academia/Government: \$50
International Industry/Corporate: \$100
Local Registrant: Free

ohdsi.org/2022apacsymposium



#OHDSISocialShowcase This Week

Representing and Utilizing Clinical Textual Data for Real World Studies: An OHDSI Approach
OHDSI Natural Language Processing Workgroup

PRESENTER: Vipina K. Keloth

INTRO

- Clinical documentation in electronic health records (EHRs) contains crucial narratives about patients and their care
- Clinical natural language processing (NLP) have evolved over the years to contribute to multiple types of real-world studies
- To promote the use of textual information in real world studies NLP WG has defined representations of textual data, developed NLP and ETL tools.

METHODS

Workflow:

- Executing the NLP systems to process the textual notes in the NOTE table
- Converting NLP system output into the NOTE_NLP table
- Transferring concepts from NOTE_NLP to individual clinical tables in CDM

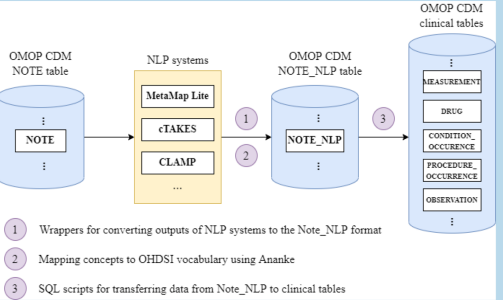
Tools:

- Wrappers for converting outputs of NLP systems to NOTE_NLP format
- Mapping concepts to OHDSI vocabulary using Ananke
- SQL scripts for transferring data from NOTE_NLP to clinical tables

CHALLENGES & LESSONS LEARNED

- Gaps in standardization of concepts extracted by NLP
- Challenges regarding the use of NLP systems
- Implementation issues to meet local application needs

A framework for representing and utilizing clinical textual data in the Note table for real-world evidence generation, applications and use cases, challenges faced and lessons learned



Scan here to read the full paper and access OHDSI NLP tools



Table: Healthcare systems that adopted OHDSI NLP solution

Healthcare Org.	NLP tools used	Applications and use cases
University of Utah Health	EasyCIE	Venous thromboembolism; pulmonary embolism
Columbia University Irving Medical Center	MedLEE, HealthTermIndex, and MedTagger	eMERGE phenotypic algorithms; infectious disease surveillance
Weill Cornell Medicine	RadText	Information extraction from radiology reports
University of Minnesota M Health Fairview	Locally trained NLP algorithms	COVID-19 sign/symptom extraction; dietary supplements information
UMass Memorial Health	cTAKES	Suicide prediction
University of Pittsburgh Medical Center	Locally trained NLP algorithms	Lifestyle-related Social Determinants of Health factors
Sydney Partnership for Health, Research, Education Enterprise	Luigi library; multiple spacy and Hugging Face models	Prevalence and impact of variation in clinical cancer care
Sema4 Mount Sinai Genomics Inc. (patients)	Locally developed NLP pipelines based on CLAMP	Extracting genetic variants, protein biomarkers, family medical history, diseases and procedures
Medical University of South Carolina	DECOVR (built on Apache UIMA); custom medicalspacy pipelines	Data Extraction for COVID-19 related information (specifically, symptom monitoring)

Vipina Keloth, Juan Banda, Michael Gurley, Paul Heider, Georgina Kennedy, Hongfang Liu, Feifan Liu, Timothy Miller, Karthik Natarajan, Olga Patterson, Yifan Peng, Ruth M. Reeves, Masoud Rouhizadeh, Jianlin Shi, Xiaoyan Wang, Yanshan Wang, Wei-Qi Wei, Andrew Williams, Rui Zhang, Rimma Belenkaya, Christian Reich, Clair Blacketer, Patrick Ryan, George Hripcsak, Noemie Elhadad, Hua Xu



MONDAY

Representing and Utilizing Clinical Textual Data for Real World Studies: An OHDSI Approach
(**Vipina Keloth**, Juan Banda, Michael Gurley, Paul Heider, Georgina Kennedy, Hongfang Liu, Feifan Liu, Timothy Miller, Karthik Natarajan, Olga Patterson, Yifan Peng, Ruth M. Reeves, Masoud Rouhizadeh, Jianlin Shi, Xiaoyan Wang, Yanshan Wang, Wei-Qi Wei, Andrew Williams, Rui Zhang, Rimma Belenkaya, Christian Reich, Clair Blacketer, Patrick Ryan, George Hripcsak, Noemie Elhadad, Hua Xu)



#OHDSISocialShowcase This Week

Explaining patient-level prediction models using permutation feature importance and SHAP

PRESENTER: Aniek Markus
(a.markus@erasmusmc.nl)
CO-AUTHORS: Egill A. Fridgeirsson, Jan A. Kors, Katia M.C. Verhamme, Peter R. Rijnbeek

INTRODUCTION:

- Feature (or variable) importance methods rank or measure the explanatory power of features and are often used to explain prediction models to end-users.
- Many methods to compute (global) feature importance exist, but clear guidance on which method is best to use is lacking.
- In this work, we compare which features are important according to permutation feature importance and Shapley Additive exPlanations (SHAP) for a given PLP model.

METHODS:

- We developed a prediction model on the Dutch IPCI database to answer the following question: "Among patients 60 years or older presenting at the general practitioner for an outpatient visit (target population), which patients will die (outcome) within 90 days (time-at-risk) after the visit?".
- We repeatedly computed permutation feature importance and SHAP on 2/3 of the training set and averaged the resulting importance values. We investigated the top 5 ranked features (using absolute values) and visualized the normalized feature importances.

RESULTS:

- The final prediction model showed good internal discrimination (AUC = 0.78).
- We found some overlapping features between methods as indicated by the numbers in brackets.
- However, there are also differences; some features included in the top 5 by one method are ranked much lower by another. In total 9 different features were identified by the three methods.
- Note these features should not be interpreted as 'risk factors', but only as features that are important for the given prediction model.

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 810168. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.



Feature importance methods show some agreement on the relative importance of features, but also large variation.

	Model coefficients	Permutation FI	SHAP
1.	Infective corneal ulcer previous 30 days (3x)	Infective corneal ulcer previous 30 days (3x)	Vascular disorder of pelvis previous 30 days (3x)
2.	Vascular disorder of pelvis previous 30 days (3x)	Vascular disorder of pelvis previous 30 days (3x)	Infective corneal ulcer previous 30 days (3x)
3.	Affective psychosis previous 30 days (2x)	Antibiotics and chemotherapeutics for dermatological use at day of visit	Affective psychosis previous 30 days (2x)
4.	Hypertensive disorder previous 30 days	Drug use temozolomide previous 365 days	Drug use goserelin previous 30 days
5.	Neoplasm of intrathoracic organs previous 30 days (2x)	Neoplasm of intrathoracic organs previous 30 days (2x)	Selective calcium channel blockers with direct cardiac effects use previous 365 days

Table 1: Top 5 ranked features in the model according to different feature importance methods (in brackets the number of times each feature was included in the top 5).

TAKE AWAYS:

- Feature importance methods vary in their intended behavior. For example, permutation feature importance explains model performance and model coefficients/SHAP explain model predictions. However, the effect of these differences (i.e. a different feature importance ranking) is not always clear to users of these methods.
- Knowing which feature importance method is best to use is important for reliable interpretation and presentation of prediction models developed within OHDSI.

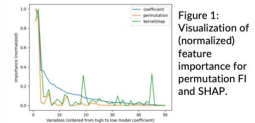


Figure 1: Visualization of (normalized) feature importance for permutation FI and SHAP.

	Permutation feature importance	SHAP
Interpretation	- Decrease in model performance. - Importance of all features together does not add up to the model performance (sum can be larger).	- Contribution to predicted output. - Importance of all features together adds up to the difference between the actual and average prediction.
Advantages	- Does not require retraining of the model. - Can be compared across problems because of using error ratio.	- 'Fair' distribution of importance by satisfying theoretical properties (efficiency, symmetry, dummy and linearity). - Allows contrastive explanation by comparing against subset of data.
Disadvantages	- Linked to error of the model, which might not be of interest (e.g. to investigate model robustness). - Extrapolates to unlikely unrealistic points (especially if features are correlated).	- Computationally expensive for large data and we have to rely on approximations. - Can give importance to features that are not used by the model and have no influence on spread of importance between correlated features.



Explaining patient-level prediction models using permutation feature importance and SHAP (Aniek F. Markus, Egill A. Fridgeirsson, Jan A. Kors, Katia M.C. Verhamme, Peter R. Rijnbeek)



#OHDSISocialShowcase This Week

PRESENTER:
EDUARD KORCHMAR

AUTHORED:
sciforce

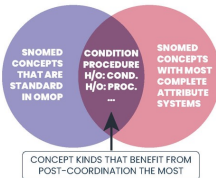
INTRO:

Using pre-coordinated standard concepts is a necessarily limited way of representing source data and real-world concepts include unrepresentable features

SNOMED CT, which is already a standard terminology in OMOP Vocabularies, has established post-coordination syntax and tools to process it

Parts of SNOMED standard in OMOP are also the ones with the best post-coordination potential (rich attribute system & maintained ontology)

SNOMED CONCEPTS IN OMOP CDM:



METHODS AND MATHS USED:

Limited but flexible implementation of OWL-like/SNOMED classifier

Solution for a special case of the subgraph isomorphism problem

BLAKE-2b for hash fingerprinting

SNOMED CT documentation

20 manually post-coordinated concepts as a test batch

TECHNOLOGIES USED:

OMOP CDM stack

Python 3 (pandas, networkx, Flask)

ANTLR4 for PCE parsing

SQLAlchemy to allow backend-independent SQL processing

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

Original SNOMED US release

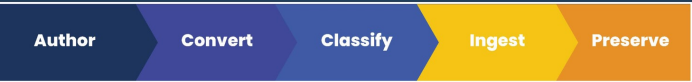


JACKALOPE: ENHANCING ETL COVERAGE THROUGH POST-COORDINATION

Some concepts can not be mapped precisely. But many of those can still be **post-coordinated**. With right tools, **full semantic meaning** of those concepts can be **stored, retrieved and used** by representing them as **post-coordinated expressions (PCE)**.

We envisioned and prototyped one such tool.

JACKALOPE TRANSFORMATION PIPELINE:



Concepts are presented as PCE. This is the only manual step, although there are tools to simplify it.

Expressions are converted to **Normal Form** as specified by SNOMED CT algorithms. At this point, pre- and post-coordination are equivalent.

An **algorithm**, analogous to SNOMED CT internal OWL reasoner, finds a place for the PCE in the **existing hierarchy**. For now, only child relations are built.

New concepts get added to the **local CDM instance**.

Their **hierarchy** is united with **SNOMED Vocabulary**, allowing use in **concept sets**.

Written PCE are stored for re-evaluation against newer SNOMED versions.

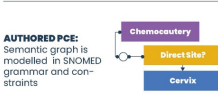
Hash of the **Normal Form** is used to identify PCE synonyms.



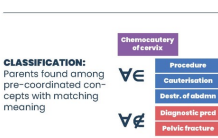
Full article, demonstration video and source code

DETAILED PIPELINE:

SOURCE CONCEPT:
Concept requiring post-coordination is identified



NORMAL FORM:
Semantic graph with redundant nodes, isomorphic with the synonymistic PCEs



OMOP INGESTION:
Changes made to local CDM instance and compatibility ensured

Changes made to local CDM instance and compatibility ensured

Changes made to local CDM instance and compatibility ensured



WEDNESDAY

Jackalope: A software tool for meaningful post-coordination for ETL purposes (Eduard Korchmar, Polina Talapova, Maria Kolesnyk, Denys Kaduk)



#OHDSISocialShowcase This Week



Clinical Sequelae of COVID-19 & Associated Healthcare Utilization: A Study Protocol

Ivan C.H. Lam^{1,2}, Yi Chai^{1,2}, Celine Sze Ling Chui^{2,3,4}, Eric Y.F. Wan^{1,4,5}, Xue Li^{1,4,5}, Carlos K.H. Wong^{1,4,5}, Hao Luo^{7,8,9}, Kenneth K.C. Man¹⁰, Xiaoyu Lin¹¹, Can Yin¹¹, Jing Li¹¹, Mui Van Zandt¹¹, Christian Reich¹¹, Katherine Duszynski¹², Nicole Pratt¹², Ian C.K. Wong^{1,4,10,*}

1. Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong (HKU), Hong Kong Special Administrative Region (HKSAR), China 2. School of Nursing, Li Ka Shing Faculty of Medicine, HKU, HKSAR, China 3. School of Public Health, Li Ka Shing Faculty of Medicine, HKU, HKSAR, China 4. Laboratory of Data Discovery for Health (D24H), Hong Kong Science and Technology Park, Sha Tin, HKSAR, China 5. Department of Family Medicine and Primary Care, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, HKU, HKSAR, China 6. Department of Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, HKU, HKSAR, China 7. Department of Social Work and Social Administration, HKU, HKSAR, China 8. Department of Computer Science, HKU, HKSAR, China 9. Sau Po Centre on Ageing HKU, HKSAR, China 10. Centre for Medicines Optimisation Research and Education, Research Department of Practice and Policy, School of Pharmacy, University College London, United Kingdom 11. IQVIA, Durham, NC, USA 12. Quality Use of Medicines and Pharmacy Research Centre, UniSA Clinical and Health Sciences, University of South Australia, Adelaide, SA, Australia.

* Co-first author *Corresponding author

HKU
Med

Background

COVID-19 infection caused by the SARS-CoV-2 virus is associated with various adverse outcomes impacting multiple organ systems.^{1,2} The rehabilitation and follow-up services for patients recovering from COVID-19 is anticipated to incur extra utilization of healthcare resource. As the COVID-19 pandemic progresses, evidence to facilitate a better understanding of the risk of long-term adverse clinical outcomes and associated utilization of healthcare resources will be of great public health interest to assess the impact of COVID-19.

Objectives

1. To monitor and evaluate the short- (6 months), medium- (12 months), long-term (3 years) mortality, morbidities and associated healthcare resources utilization following COVID-19 infection.
2. To investigate morbidities of COVID-19 infection in specific populations, including children, older adults, and patients with multi-morbidity.

Methods

- Electronic health records from six healthcare databases were mapped to the Observational Medical Outcomes Partnership (OMOP) common data model. (Table 1)
- A retrospective cohort study will be conducted using data from December 1st, 2019 to June 30th, 2023 (subject to data availability). Patients with COVID-19 diagnosed with or received a positive screening test results for COVID-19 between December 1st 2019 and June 30th 2022 will be matched to non-COVID-19 subjects by propensity score conditioned on the probability of COVID-19 infection with the aim to balance the baseline characteristics across the study cohorts. Large scale regularized regression will be applied to estimate the propensity score based on patients' demographics. Subjects with COVID-19 will be matched to up to 10 comparators with similar propensity score, using the caliper width of 0.05. Non-COVID-19 subjects infected with COVID-19 during the observation period will be censored from the study.
- All subjects will be monitored for first diagnosis of cardiovascular, hematological, respiratory, neurological, psychiatric, immunological, endocrine, malignant, dermatological or gastrointestinal disorders during the observation period. The number of general practitioner visits, accident and emergency department attendances, length of stay in general ward and intensive care unit during hospital admission will be reported and compared between COVID-19 and non-COVID-19 subjects.
- The hazard ratio (HR) and 95% confidence interval (CI) of each outcome will be estimated using Cox proportional hazard regressions. A random-effects meta-analysis model will be applied to combine results obtained across multiple databases.

Table 1. Electronic health records consisted in databases mapped to the Observational Medical Outcomes Partnership (OMOP) common data model

Database	Electronic health records
US Open-claim IQVIA	Pre-adjudicated health insurance claims collected from general practitioners and specialists
Germany DA IQVIA ^a , France LPD IQVIA ^b , Italy LPD IQVIA ^c , UK IMRD IQVIA ^c	Proprietary practice management software used by general practitioners and selected specialists
Hong Kong Hospital Authority (HA)	Patient records from general practitioners
	Patient records of public hospital and ambulatory clinics in Hong Kong

^a DA = Disease Analyser; ^b LPD = Longitudinal Patient Database; ^c IMRD = IQVIA Medical Research Data

Results

A total of 69,921 patients with COVID-19 between 1st December 2019 and 1st December 2020 were included in the preliminary analysis conducted on the France, Italy LPD and Hong Kong HA Databases. (Table 2) Patients with COVID-19 were observed to incur an increased incidence of myocarditis, pericarditis, myocardial infarction, pulmonary embolism, immune and idiopathic thrombocytopenia for their matched controls over a 12 months observation period. A comparable incidence of encephalitis and encephalomyelitis, Bell's palsy, Guillain-Barre Syndrome and Ischemic stroke was observed between the COVID and non-COVID-19 subjects over the same observation period. (Figure 1)

Table 2. Number of subjects, follow-up time (years) of target and comparator after propensity score matching

Datasets	Target		Comparator	
	Subjects	Follow-up time (years)	Subjects	Follow-up time (years)
France LPD	52,790	44,137	366,275	286,216
Italy LPD	16,841	13,948	123,820	102,765
Hong Kong HA	290	262	2,610	2,466
Overall	69,921	58,348	492,705	391,448

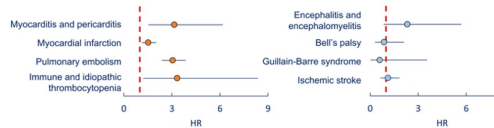


Figure 1. Hazard ratio (HR) and 95% confidence interval (CI) of clinical sequelae between COVID-19 and non-COVID-19 patients

Conclusions

Preliminary results of this study indicated an increased risk of cardiovascular and hematologic sequelae following COVID-19 infection. To our knowledge, this will be the largest observational study using multi-national healthcare databases to report the medium to long-term adverse outcomes of COVID-19 infection. The study will generate robust evidence to evaluate the adverse clinical outcomes of COVID-19 infection. Information on healthcare resources will inform the policy makers when budgeting future healthcare resources together with, an understanding of the breadth and duration of the long-term effect of COVID-19.

Acknowledgement

This work was supported by the Research Grants Council of Hong Kong under the Collaborative Research Fund Scheme (C7154-20G).

References

1. Leung TTM, Chan AYL, Chan EW, et al. Short- and potential long-term adverse health outcomes of COVID-19: a rapid review. *Emerg Microbes Infect.* Dec 2020;9(1):2190-2199. doi:10.1080/22221751.2020.1825914
2. Daugherty SE, Guo Y, Heath K, et al. Risk of clinical sequelae after the acute phase of SARS-CoV-2 infection: retrospective cohort study. *BMJ.* May 19 2021;373:n1098. doi:10.1136/bmj.n1098

Contact: wongick@hku.hk (Ian C.K. Wong), lami@hku.hk (Ivan C.H. Lam)

Clinical Sequelae of COVID-19 & Associated Healthcare Utilization: A Study Protocol (**Ivan Chun Hang Lam**, Yi Chai, Celine Sze Ling Chui, Eric Yuk Fai Wan, Xue Li, Carlos King Ho Wong, Hao Luo, Kenneth Keng Cheung Man, Xiaoyu Lin, Can Yin, Jing Li, Mui Van Zandt, Christian Reich, Katherine Duszynski, Nicole Pratt, Ian Chi Kei Wong)





Openings



FDA/CDER's Division of Hepatology and Nutrition is seeking a clinician with bioinformatics or biostatistics training to work with the Drug-Induced Liver Injury (DILI) Team to evaluate large datasets of liver-related data, collaborate on the Team's review of drugs with hepatotoxicity signals, and help develop informatics-based processes in DILI evaluation across the Agency.

Contact **Judy Racoosin** at judith.racoosin@fda.hhs.gov for information about the application process (that will be through USAJOBS).



Openings

Andrew Williams recently announced two exciting new openings at Tufts Medicine.

1) Senior Project Manager for a multisite multiyear grant standardizing critical care EHR and waveform data. (CHoRUS Bridge2AI)

2) Lead software developer and research data warehouse manager for Tufts Medicine's OMOP instance and related services.

Remote work is possible for both positions.



1. Link for Senior Project Manager position: <https://smrtr.io/bBVzh>
2. Link for Lead Software Developer and Research Data Warehouse Manager position: <https://jobs.smartrecruiters.com/TuftsMedicalCenter1/743999857980631-software-development-lead-res-g-c-ctsi>

Andrew's email:
awilliams15@tuftsmedicalcenter.org



Openings

Research Associate (Data Scientist/Statistical Engineer), Johns Hopkins inHealth and Biostatistics Center

- Execute OHDSI studies (e.g. for cohort characterizations and comparative effectiveness) on Johns Hopkins's EHR data to support clinicians;
- Collaborate with statisticians and clinicians to continuously integrate state-of-the-art statistical tools to the inHealth/OHDSI tool stack for deployment;
- Mentor trainees on data science and software development skills;
- Co-teach courses on observational health data analytics and data science skills at School of Medicine and Public Health;
- Facilitate adoption of the inHealth tools among the broader OHDSI community by contributing to OHDSI's [Health Analytics Data-to-Evidence Suite](#).
- <https://apply.interfolio.com/114436>



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?

