Where Are We Going in 2023?

OHDSI Community Call

Jan. 10, 2022 • 11 am ET

Upcoming OHDSI Community Calls

Date	Topic
Jan. 10	Where Can OHDSI Go in 2023?
Jan. 17	OHDSI Speed Dating
Jan. 24	Collaborations For Strategic Priorities
Jan. 31	Introduction to Phenotype Phebruary
Feb. 7	Phenotype Phebruary Weekly Update + Workgroup Plans for 2023
Feb. 14	Phenotype Phebruary Weekly Update + Workgroup Plans for 2023
Feb. 21	Phenotype Phebruary Weekly Update + Workgroup Plans for 2023
Feb. 28	Phenotype Phebruary Weekly Update + Workgroup Plans for 2023

Upcoming OHDSI Community Calls

Date	Topic
Jan. 10	Where Can OHDSI Go in 2023?
Jan. 17	OHDSI Speed Dating
Jan. 24	Collaborations For Strategic Priorities
Jan. 31	Introduction to Phenotype Phebruary
Feb. 7	Phenotype Phebruary Weekly Update + Workgroup Plans for 2023
Feb. 14	Phenotype Phebruary Weekly Update + Workgroup Plans for 2023
Feb. 21	Phenotype Phebruary Weekly Update + Workgroup Plans for 2023
Feb. 28	Phenotype Phebruary Weekly Update + Workgroup Plans for 2023

Three Stages of The Journey

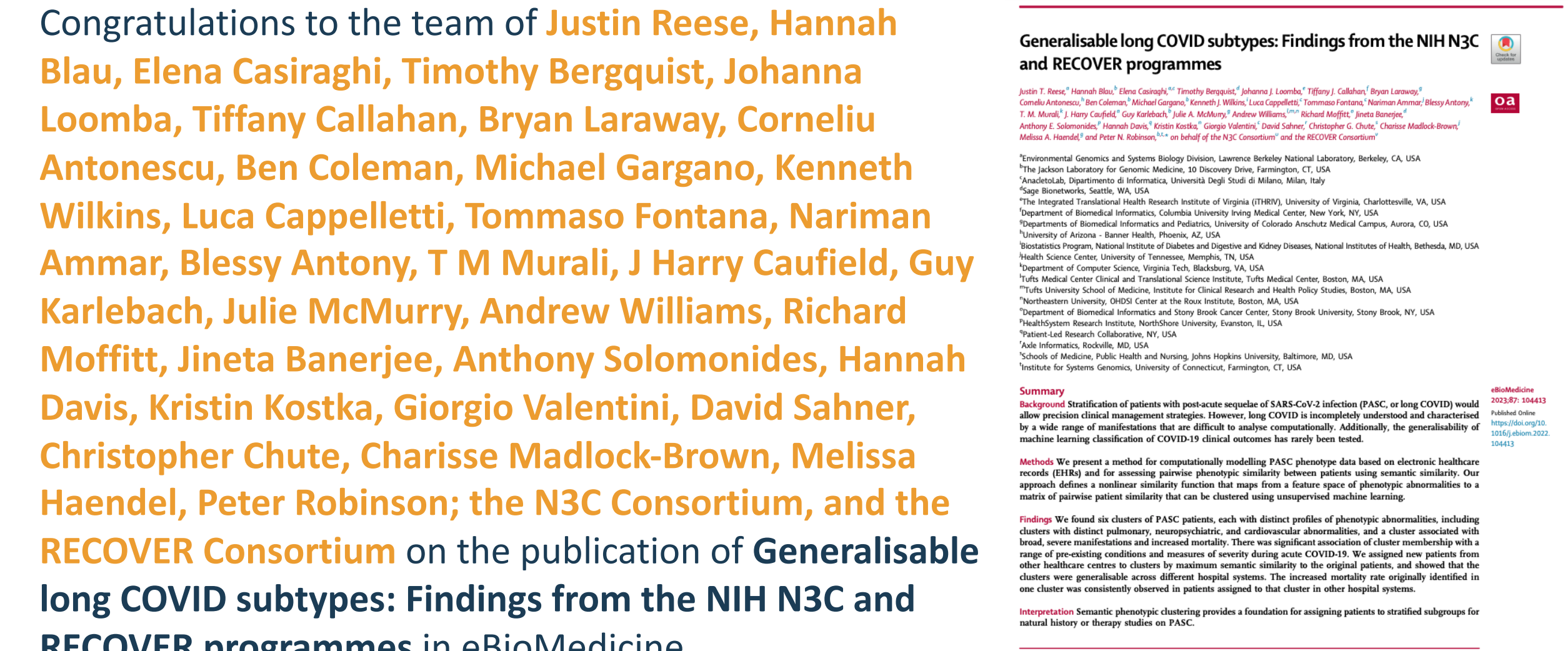
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Justin Reese, Hannah Blau, Elena Casiraghi, Timothy Bergquist, Johanna Loomba, Tiffany Callahan, Bryan Laraway, Corneliu Antonescu, Ben Coleman, Michael Gargano, Kenneth Wilkins, Luca Cappelletti, Tommaso Fontana, Nariman Ammar, Blessy Antony, T M Murali, J Harry Caufield, Guy Karlebach, Julie McMurry, Andrew Williams, Richard Moffitt, Jineta Banerjee, Anthony Solomonides, Hannah Davis, Kristin Kostka, Giorgio Valentini, David Sahner, Christopher Chute, Charisse Madlock-Brown, Melissa Haendel, Peter Robinson; the N3C Consortium, and the RECOVER Consortium** on the publication of **Generalisable long COVID subtypes: Findings from the NIH N3C and RECOVER programmes** in eBioMedicine.



Generalisable long COVID subtypes: Findings from the NIH N3C and RECOVER programmes

Justin T. Reese,[a] Hannah Blau,[b] Elena Casiraghi,[a,c] Timothy Bergquist,[d] Johanna J. Loomba,[e] Tiffany J. Callahan,[f] Bryan Laraway,[g] Corneliu Antonescu,[h] Ben Coleman,[b] Michael Gargano,[b] Kenneth J. Wilkins,[i] Luca Cappelletti,[c] Tommaso Fontana,[c] Nariman Ammar,[j] Blessy Antony,[k] T. M. Murali,[k] J. Harry Caufield,[a] Guy Karlebach,[b] Julie A. McMurry,[g] Andrew Williams,[l,m,n] Richard Moffitt,[o] Jineta Banerjee,[d] Anthony E. Solomonides,[p] Hannah Davis,[q] Kristin Kostka,[n] Giorgio Valentini,[c] David Sahner,[r] Christopher G. Chute,[s] Charisse Madlock-Brown,[j] Melissa A. Haendel,[g] and Peter N. Robinson,[b,t,*] on behalf of the N3C Consortium[v] and the RECOVER Consortium[v]

[a]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[b]The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, USA
[c]AnacletoLab, Dipartimento di Informatica, Università Degli Studi di Milano, Milan, Italy
[d]Sage Bionetworks, Seattle, WA, USA
[e]The Integrated Translational Health Research Institute of Virginia (iTHRIV), University of Virginia, Charlottesville, VA, USA
[f]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA
[g]Departments of Biomedical Informatics and Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
[h]University of Arizona - Banner Health, Phoenix, AZ, USA
[i]Biostatistics Program, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA
[j]Health Science Center, University of Tennessee, Memphis, TN, USA
[k]Department of Computer Science, Virginia Tech, Blacksburg, VA, USA
[l]Tufts Medical Center Clinical and Translational Science Institute, Tufts Medical Center, Boston, MA, USA
[m]Tufts University School of Medicine, Institute for Clinical Research and Health Policy Studies, Boston, MA, USA
[n]Northeastern University, OHDSI Center at the Roux Institute, Boston, MA, USA
[o]Department of Biomedical Informatics and Stony Brook Cancer Center, Stony Brook University, Stony Brook, NY, USA
[p]HealthSystem Research Institute, NorthShore University, Evanston, IL, USA
[q]Patient-Led Research Collaborative, NY, USA
[r]Axle Informatics, Rockville, MD, USA
[s]Schools of Medicine, Public Health and Nursing, Johns Hopkins University, Baltimore, MD, USA
[t]Institute for Systems Genomics, University of Connecticut, Farmington, CT, USA

Summary

Background Stratification of patients with post-acute sequelae of SARS-CoV-2 infection (PASC, or long COVID) would allow precision clinical management strategies. However, long COVID is incompletely understood and characterised by a wide range of manifestations that are difficult to analyse computationally. Additionally, the generalisability of machine learning classification of COVID-19 clinical outcomes has rarely been tested.

Methods We present a method for computationally modelling PASC phenotype data based on electronic healthcare records (EHRs) and for assessing pairwise phenotypic similarity between patients using semantic similarity. Our approach defines a nonlinear similarity function that maps from a feature space of phenotypic abnormalities to a matrix of pairwise patient similarity that can be clustered using unsupervised machine learning.

Findings We found six clusters of PASC patients, each with distinct profiles of phenotypic abnormalities, including clusters with distinct pulmonary, neuropsychiatric, and cardiovascular abnormalities, and a cluster associated with broad, severe manifestations and increased mortality. There was significant association of cluster membership with a range of pre-existing conditions and measures of severity during acute COVID-19. We assigned new patients from other healthcare centres to clusters by maximum semantic similarity to the original patients, and showed that the clusters were generalisable across different hospital systems. The increased mortality rate originally identified in one cluster was consistently observed in patients assigned to that cluster in other hospital systems.

Interpretation Semantic phenotypic clustering provides a foundation for assigning patients to stratified subgroups for natural history or therapy studies on PASC.

eBioMedicine 2023;87: 104413
Published Online https://doi.org/10.1016/j.ebiom.2022.104413

OHDSI Shoutouts!

Congratulations to the team of **Xiang Cheng, Meiling Cheng, Liyi Yu and Xuan Xiao** on the publication of **iADRGSE: A Graph-Embedding and Self-Attention Encoding for Identifying Adverse Drug Reaction in the Earlier Phase of Drug Development** in the International Journal of Molecular Sciences.

Article

iADRGSE: A Graph-Embedding and Self-Attention Encoding for Identifying Adverse Drug Reaction in the Earlier Phase of Drug Development

Xiang Cheng, Meiling Cheng, Liyi Yu and Xuan Xiao *

Department of Computer, Jingdezhen Ceramic University, Jingdezhen 333403, China
* Correspondence: xiaoxuan@jcu.edu.cn or jdzxiaoxuan@163.com; Tel.: +86-0798-8485-288

Abstract: Adverse drug reactions (ADRs) are a major issue to be addressed by the pharmaceutical industry. Early and accurate detection of potential ADRs contributes to enhancing drug safety and reducing financial expenses. The majority of the approaches that have been employed to identify ADRs are limited to determining whether a drug exhibits an ADR, rather than identifying the exact type of ADR. By introducing the "multi-level feature-fusion deep-learning model", a new predictor, called iADRGSE, has been developed, which can be used to identify adverse drug reactions at the early stage of drug discovery. iADRGSE integrates a self-attentive module and a graph-network module that can extract one-dimensional sub-structure sequence information and two-dimensional chemical-structure graph information of drug molecules. As a demonstration, cross-validation and independent testing were performed with iADRGSE on a dataset of ADRs classified into 27 categories, based on SOC (system organ classification). In addition, experiments comparing iADRGSE with approaches such as NPF were conducted on the OMOP dataset, using the jackknife test method. Experiments show that iADRGSE was superior to existing state-of-the-art predictors.

Keywords: adverse drug reactions; graph isomorphism network; self-attention; multi-label learning

check for updates

Citation: Cheng, X.; Cheng, M.; Yu, L.; Xiao, X. iADRGSE: A Graph-Embedding and Self-Attention Encoding for Identifying Adverse Drug Reaction in the Earlier Phase of Drug Development. *Int. J. Mol. Sci.* **2022**, 23, 16216. https://doi.org/10.3390/ijms232416216

Academic Editor: Francisco Torrens

Received: 22 November 2022
Accepted: 16 December 2022
Published: 19 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Adverse drug reactions (ADRs) or side effects are substantially harmful or distressing reactions, and are described as adverse responses to drugs beyond their anticipated therapeutic effects [1]. In the United States, it is estimated that ADRs result in over 100,000 patient deaths per year [2] and the cost of ADRs-related morbidity was USD 528.4 billion in 2016 [3]. The process of drug-development involves a lot of monetary resources because it involves a lot of clinical trials and tests [4]. Many ADRs are not detected in the early stages of drug development, owing to restricted trial samples and time [5]. Thus, ADRs not only jeopardize patient health but also result in wasted healthcare costs, and are considered as a major global public-health problem. Traditional laboratory experiments to identify potential ADRs are not merely cumbersome and low cost-effective, but also less effective in the earlier phase. In recent years, algorithms in silico have been employed to speed up the prediction process and reduce drug-development costs.

Among the existing studies, some utilize data mining to analyze potential ADRs from large amounts of data and various sources of information; others adopts machine learning methods to predict ADRs.

The available databases of ADRs have some limitations at present. The data collected by the spontaneous reporting systems (SRS) and FDA Adverse Event Reporting System (FAERS) are not comprehensive enough, and there are problems such as repeated declaration. Drugs in the Side Effect Resource (SIDER) are limited to FDA-approved drugs only. The content of the European Medicines Agency (EMA) and other large-scale databases is complicated, and has no special retrieval of ADRs, which cause a lot of inconvenience for the use of data. Considering the limitations of the existing database, some researchers

OHDSI Shoutouts!

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

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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

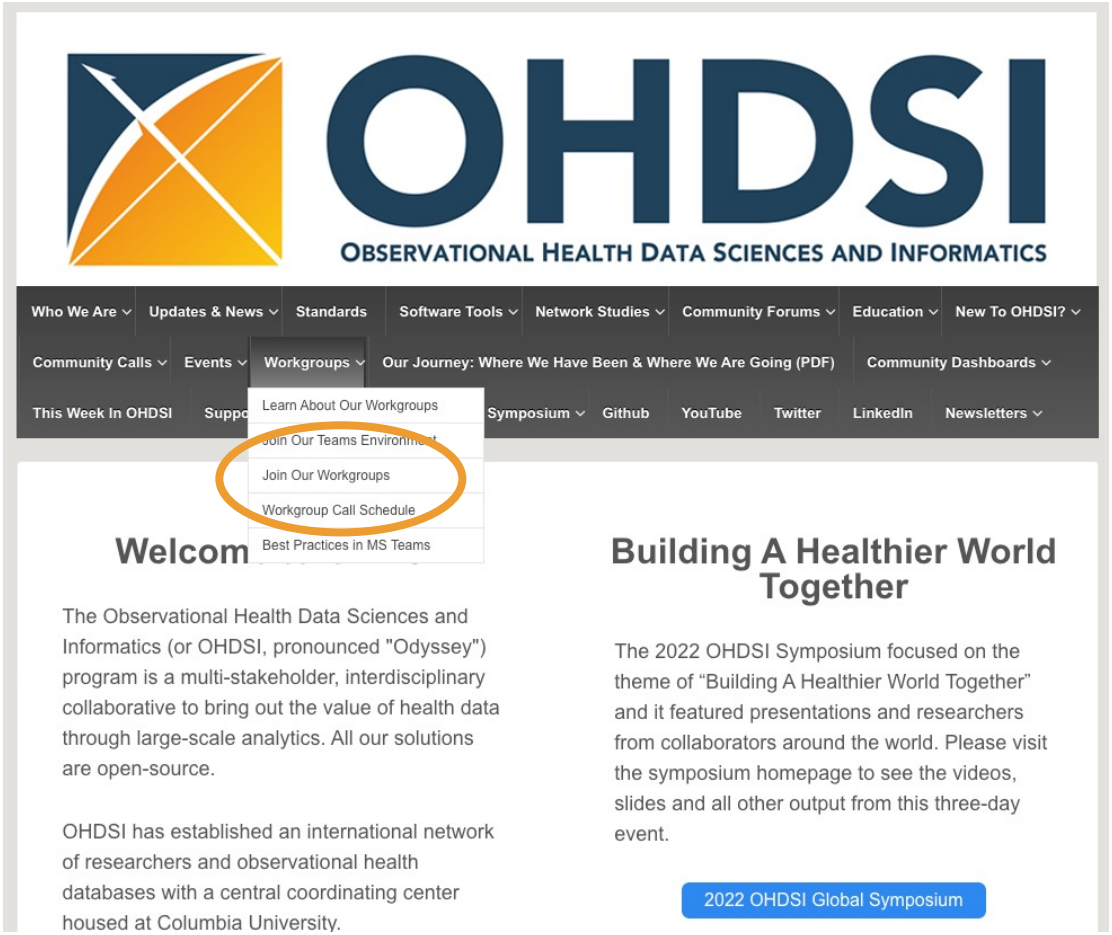
Where Are We Going?

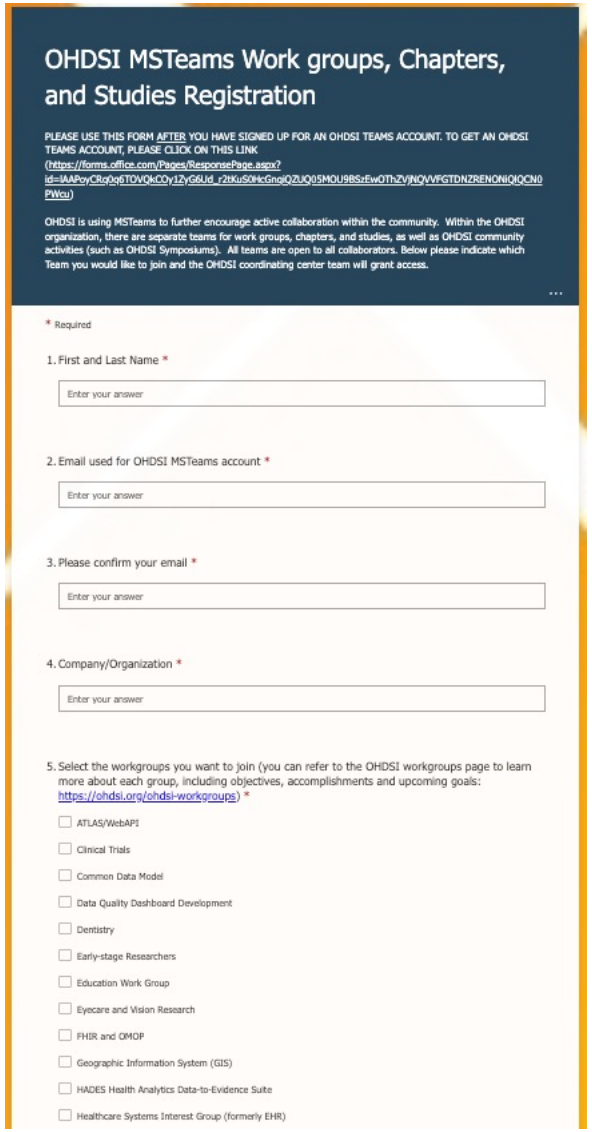
Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Common Data Model Vocabulary Subgroup
Tuesday	3 pm	OMOP CDM Oncology – Outreach/Research Subgroup
Tuesday	6 pm	Eyecare and Vision Research
Wednesday	9 am	Patient-Level Prediction
Wednesday	2 pm	Natural Language Processing
Wednesday	7 pm	Medical Imaging
Thursday	10 am	Data Quality Dashboard
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development and Evaluation
Friday	9 am	GIS – Geographic Information System General
Friday	1 pm	Clinical Trials
Friday	11 pm	China Chapter
Monday	10 am	Healthcare Systems Interest Group

ohdsi.org/workgroups

Upcoming Workgroup Calls





ohdsi.org/workgroups

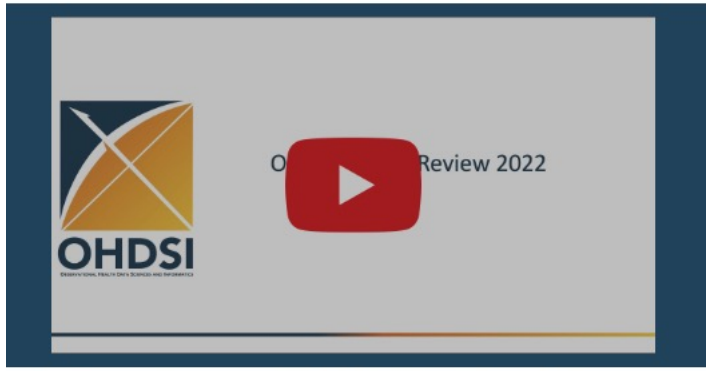
Interested in presenting to the CDM WG?

Enhancing workgroup collaboration has been a focus over the last year.

Clair Blacketer has a few thoughts for any workgroup that is considering presenting to the CDM WG.

January Newsletter Is Available

OHDSI 2022 In Review



Patrick Ryan presented a comprehensive look back at the activities, publications, open-source developments and more from the OHDSI community throughout 2022 during a December community call.

Collaborator Spotlight: Thamir AlShammary

Thamir AlShammary, an advisor to the President of the Saudi Food and Drug Authority (SFDA), has been an active contributor to the OHDSI community for several years. He collaborates in several workgroups, including Population-Level Estimation, Health Equity and the recently-completed Vaccine Evidence WG, and has been a contributor in several important network studies.

He discusses his background and journey into OHDSI, and why OHDSI can be a difference maker in generating trustworthy evidence, in the latest edition of the Collaborator Spotlight.

Spotlight: Thamir AlShammary

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

The Journey Newsletter (January 2023)

2022 was a memorable and productive year for the community. We produced a record-setting 111 publications and saw 738 new co-authors contributing to publications, returned to in-person symposia in three different continents, and made significant progress in work around data standards, methodological research, open-source development and clinical applications. That work will help create the foundation for what we can do together in 2023! **#JoinTheJourney**

January Update Podcast

In the latest On The Journey video, Patrick Ryan and Craig Sachson discuss the winning submissions in the OMOP Common Data Model (CDM) Entity-Relationship Diagram (ERD) Challenge, and then they reflect on some of the numerous activities, accomplishments and publications from the community in 2022.

Community Updates

Where Have We Been?

- 2022 saw a return to in-person events, several new workgroups and the Kheiron Cohort, community activities like Phenotype Phebruary, DevCon and the Early-Stage Researchers Career Speaker Series, and plenty more. Patrick Ryan provided a review of the year, which you can watch later in the newsletter.
- The OHDSI community produced a record-setting 111 publications in 2022, and have now seen 2,057 authors participate in OHDSI-related papers since our inception. You can read any of these studies via the new Publications Dashboard, developed by Paul Nagy, Star Liu and their team this past year.
- The 2022 Asia-Pacific (APAC) Symposium took place in November, and all talks and slides from both the main conference and the full-day tutorial are now available at the APAC Symposium homepage.

Where Are We Now?

- The open-source tools that empower OHDSI's global research initiatives are not only available to the community, but they are also developed by the community. Leaders from around the world have developed tools that provide the foundation for OHDSI collaborators to engage in robust, reliable and reproducible observational health research. Many of these developers have provided "10-Minute Tutorials" that can help aid your understanding of how to use these tools in research.
- The EHDEN Consortium announced that 22 data partners from across 13 countries have been selected from the final open call to join the EHDEN data network. The data partners from this call represent almost 200 million patient records, originating from various care settings, adding to the approximately 630 million records with the 166 partners working with EHDEN from the prior six calls.

Where Are We Going?

- Community calls will resume Jan. 10 (11 am ET) and continue each Tuesday as a way to connect as a community, share updates, learn from each other, and continue to move forward together. A new call invite will go out the week of January 2, but you can also find the new link on our community calls page.
- The #OHDSISocialShowcase will continue in 2023, as research from the OHDSI Symposium is featured each weekday on the OHDSI Twitter and LinkedIn accounts. Please follow us, learn about these new developments in our community, and share with your networks!

How Can You Join The Journey?

- The OHDSI research community strives to promote better health decisions and care through globally standardized health data, continuously developing large-scale analytics and a spirit of collaboration though open science. We are proud to have more than 3,200 collaborators across six continents, as well as health records for about 928 million unique patients from around the world. We are always looking for new collaborators, so if this sounds exciting to you, please read about how you can Join The Journey!

Oxford Real World Evidence Summer School

Oxford Summer School 2023: Real World Evidence using the OMOP Common Data Model

COURSE DIRECTORS

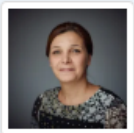
Daniel Prieto-Alhambra

Professor of Pharmaco- and Device Epidemiology

COURSE ADMINISTRATOR

Mahkameh Mafi

Personal Assistant to Professor Prieto-Alhambra



OTHER COURSES

Statistics: Designing clinical research and biostatistics

Brief Description:

Our Real World Evidence Summer School will provide participants with the tools and concepts necessary to plan and execute Real World Evidence studies, with a focus on the use of the OMOP common data model. The course will have morning lectures followed by afternoon practicals where concepts discussed in the morning will be put in practice with hands-on sessions. Practical sessions will have two tracks: a) for those interested in the design of studies and use of existing analytical and data curation tools; and b) for more advanced data scientists and programmers interested in the development or modification of analytical code using R.

Registration: It is now open

Venue: Lady Margaret Hall Talbot Hall Theatre, Norham Gardens, Oxford OX2 6QA

Date: 19th- 23rd June 2023

For booking please use **Booking information**

Please see the Preliminary Programme here

AUDIENCE:

Pharmacists, clinicians, academics (including statisticians, epidemiologists, and related MSc/PhD students); Industry (pharmacy or device) or Regulatory staff with an interest in the use of routinely collected data for research.

LEARNING GOALS:

Collaborator Spotlight: Thamir AlShammary

Thamir AlShammary, an advisor to the President of the Saudi Food and Drug Authority (SFDA), has been an active contributor to the OHDSI community for several years. He discusses his background, his journey into OHDSI and the impact he has seen, and why OHDSI can be a difference maker in generating trustworthy evidence, tools and best practices within the community, in the latest edition of the Collaborator Spotlight.

ohdsi.org/spotlight-thamir-alshammary

#OHDSISocialShowcase This Week

Analysis of Influencing Factors of Mortality in COVID-19 Patients: A Retrospective Cohort Study

Do Duy Khang; Phung-Anh Nguyen; Chang-I Chen; Chung-Chien Huang; Carlos Shu-Kei Lam; Noi Yar; Christine Y. Lu; Chi-Tsun Cheng; Jason C. Hsu*
Taipei Medical University, Taiwan

Abstract

Background
Coronavirus Disease (COVID-19) has spread rapidly around the world since the end of 2019. Because of its high incidence and high mortality, it is currently the most concerned health issue in the world. Clinically, avoiding mortality or severe illness is the main goal of Covid-19 treatment. Previous studies of factors influencing death of COVID-19 patients have shown that older age or certain comorbidities may increase the risk of severe illness in people with COVID-19, and some of these conditions may be fatal.1,2 In particular, cancer patients are particularly vulnerable to health consequences after infection, including increased risk of life-threatening infections and interruption of cancer or normal treatment.3 A comprehensive understanding of the factors affecting the mortality of Covid-19 cases and timely implementation of appropriate improvement strategies is one of the most important issues in clinical disease treatment.

Objectives
The purpose of this study was to explore the main influencing factors leading to Covid-19-related mortality and to provide clinical treatment recommendations based on the findings. This study used Taipei Medical University Clinical Research Database (TMUCRD) with data from 3 hospitals in Taiwan as the data source, the data were mapped to OHDSI OMOP CDM. It is expected to be developed into a multinational cooperative research using OHDSI tools and OMOP CDM in the future as well.

Methods
This study is a retrospective observational study. We obtained data from the TMUCRD, which collects three hospital electronic medical records in northern Taiwan. This study obtained 2021.01.01-2021.09.30 inpatients infected by Covid-19 from TMUCRD as the main study cohort. Patients who have not visited three hospitals in the past or who were younger than 20 years were excluded. The patient's first day of hospitalization was the index date, and the mortality was the main outcome. Covariates include demographic characteristics, health status, selected comorbidities and selected medications. Logistic regression with univariate and multivariate analysis method was used to estimate the association of each influencing factor with outcome. In addition, we also used the Cox regression model to conducted a further overall and stratified analysis about the associations between specific influencing factors and mortality among COVID patients.

Results
Totally 713 inpatient patients were included in this study. Uni-variable analysis showed that males, elderly, high CCI scores, co-morbidities such as congestive heart failure, cerebrovascular disease, dementia, chronic obstructive pulmonary disease, diabetes, renal disease, cancer, hypertension, hyperlipidemia, anemia, parkinson's disease, osteoporosis, etc., as well as the use of clonazepam sleeping pills, have a higher risk of mortality. However, after adjustment for other factors, only the following were statistically significant: older age and use of the clonazepam sleeping drug (OR= 4.358; 95% CI: 1.693-11.221; p-value=0.002). The results of the Cox regression regarding clonazepam use found that, overall, clonazepam sleeping pills significantly increase the mortality risk of COVID patients (HR=1.995; 95% CI: 1.007-3.954; p-value=0.048). Especially, when the patient's age was less than 65 years old, 0≤CCI<3, and no depression, the patients who used clonazepam sleeping pills had a higher risk of mortality than those who did not use clonazepam sleeping pills.

Methods

This study is a retrospective observational study. We obtained data from the Taipei Medical University Clinical Research Database (TMUCRD), which collects three hospital electronic medical records in northern Taiwan. This study obtained 2021.01.01-2021.09.30 inpatients infected by Covid-19 from TMUCRD as the main study cohort. Exclusion criteria included patients admitted to the ICU immediately after admission, deaths within 24 hours of admission, and cases under the age of 20. The patient's first day of hospitalization was the index date, and the mortality was the main outcome. Covariates include demographic characteristics, health status, selected comorbidities and selected medications. Logistic regression with univariate and multivariate analysis method was used to estimate the association of each influencing factor with outcome. In addition, we also used the Cox regression model to conducted a further overall and stratified analysis about the associations between clonazepam use and mortality among COVID patients.

Results

Table 1. Results of Uni-variable and Multi-variable Logistics Regression

Variables	Uni-variable Logistics Regression				Multi-variable Logistics Regression			
	OR	Lower 95% CI	Upper 95% CI	*p*-value	OR	Lower 95% CI	Upper 95% CI	*p*-value
Age					1.059	1.041	1.078	**<0.001**
Age <65								
65=<AGE=<85	1.959	1.161	3.305	**0.012**				
AGE >85	2.754	2.574	7.36	**<0.001**				
CCI score					1.177	0.858	1.614	0.315
CCI=0								
0=<CCI<3	5.605	3.271	9.603	**<0.001**				
CC=<3	11.22	6.06	20.804	**<0.001**				
Benzodiazepine derivatives (all)	1.436	0.774	2.3665	0.252	1.221	0.626	2.381	0.559
Benzodiazepine derivatives (clonazepam)	3.969	1.822	8.646	**0.001**	4.358	1.693	11.221	**0.002**
Benzodiazepine derivatives (N05BA)	1.409	0.617	3.219	0.416	0.954	0.352	2.588	0.926
Benzodiazepine derivatives (N05CD)	1.464	0.668	3.209	0.341	0.578	0.229	1.462	0.247

Table 2 Results of the associations between clonazepam use and mortality among COVID patients by overall and stratified analysis

Variables	N	Multi-variable COX Regression			
		HR	Lower 95% CI	Upper 95% CI	*p*-value
Overall Analysis	718	1.995	1.007	3.954	**0.048**
Stratified Analysis					
Age					
Age <65	300	11.340	2.179	59.005	**0.004**
CCI score					
0=<CCI<3	238	7.171	1.218	42.216	**0.029**
Comorbidity					
Congestive heart failure (CHF)					
No	75	3.193	1.449	7.036	**0.004**
Diabetes mellitus (DM)					
No	147	3.203	1.355	7.573	**0.008**
Cancer					
Yes	4	11.267	1.264	100.443	**0.030**
Hypertension					
No	255	3.655	1.5	8.906	**0.004**
Hyperlipidemia					
No	163	3.584	1.469	8.742	**0.005**
Depression					
No	33	2.213	1.097	4.46	**0.026**
Parkinson's disease					
No	33	2.313	1.09	4.909	**0.029**
Osteoporosis					
No	47	2.196	1.088	4.434	**0.028**

Results

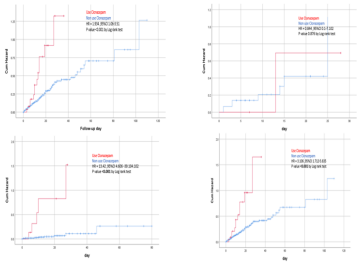


Figure 1. Kaplan-Meier curve for the mortality in COVID-19 patients between clonazepam users and non-users: (a) Overall population; (b) Age <65; (c) patients with depression; (d) patients without depression.

Conclusions

Our results suggest that male, elderly COVID-19 inpatients are at higher risk of mortality. Clonazepam sleeping pills users have significantly higher risk of mortality than non-users significantly. Patients younger than 65 years old, 0≤CCI<3, and without depression should avoid the use of clonazepam sleeping pills to reduce the risk of mortality.

References
1.Murthy S, Gomersall CD, Fowler RA. Care for critically ill patients with COVID-19. Jama. 2020;323(15):1499-1500.
2.Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. Jama. 2020;323(13):1239-1242.
3.Desai A, Sachdeva S, Parekh T, Desai R. COVID-19 and cancer: lessons from a pooled meta-analysis. JCO global oncology. 2020;6.

Contact: jasonhsu@tmu.edu.tw

TUESDAY Analysis of Influencing Factors of Mortality in COVID-19 Patients: A Retrospective Cohort Study (Do Duy Khang, Phung-Anh Nguyen, Chang-I Chen, Chung-Chien Huang, Carlos Shu-Kei Lam, Noi Yar, Christine Y. Lu,Chi-Tsun Cheng, Jason C. Hsu)

#OHDSISocialShowcase This Week

The manifold presentations of PROMS and questionnaires: *patient-reported outcomes in OMOP use cases*

PRESENTER: **Sebastiaan van Sandijk**

INTRODUCTION

- **PROMS and QoL questionnaires are essential in various use cases,** like HTA (Health Technology Assessment), drug approval and drug safety analyses, clinical guideline development, etc.
- Currently, however, OMOP-CDM and OHDSI tools **do not really support PROMS** data or standardized analytics with this type of information.
- **Guidelines or conventions** for adding PROMs data to OMOP-CDM do not yet exist. This (on-going) work aims to propose some recommendations.

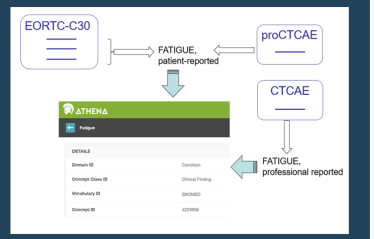
METHODS

1. Use-case driven approach working with data partners in the EHDEN network that are in the process of converting their data.
2. Inventory of common PROMS and questionnaires as well as their use cases – why and what in OMOP?
3. Inventory of recurring issues as well as often-used solutions to address these issues.
4. Discussion of potential best practice recommendations are discussed for selected PROMS / questionnaires, focusing on the quick wins.
5. (future step) Test the proposed recommendations in selected focused case studies, with a few data partners, to determine value added and usability of the approach or recommendations

PROMS and questionnaires are not all the same; implications and requirements vary – for clinical practice and OMOP use cases.

Start from what is in OMOP and examine how (validated) questionnaire scores relate to clinical standard concepts.

Conventions needed for Provenance and Context of Use!



Take a picture to download the full paper

RESULTS AND RECOMMENDATIONS

- Need to disambiguate surveys questionnaires, registries, Minimum or Recommended Data Sets, cohort studies, screenings, intake forms, etc. These all (can) have questionnaires or PROMS, each with their own implications.
- Common issues: "lack of standardization in use"; "standard / local (questions)"; "validated / non-validated"; "disease-specific / generic"; "quality of implementation"; "frequency"; "data types" (scores, strings, ordinal); "business rules"
- "Provenance" and "context of use" should be better represented in OMOP – **and used in analytics!**

DISCUSSION AND CONCLUSION

- With our use-case driven approach we expect to be able to formulate some recommendations for PROMS data in OMOP - for selected use cases.
- Initial focus will be on validated questionnaires with defined business rules – like EORTC-C30, (pro)CTCAE, and some semantically overlapping ICHOM Standard Sets.

Sebastiaan van Sandijk[1]*, Peter Prinsen[2], Mieke van Hemelrijck[3], Michael Kallfelz[1], Dalia Dawoud[4]

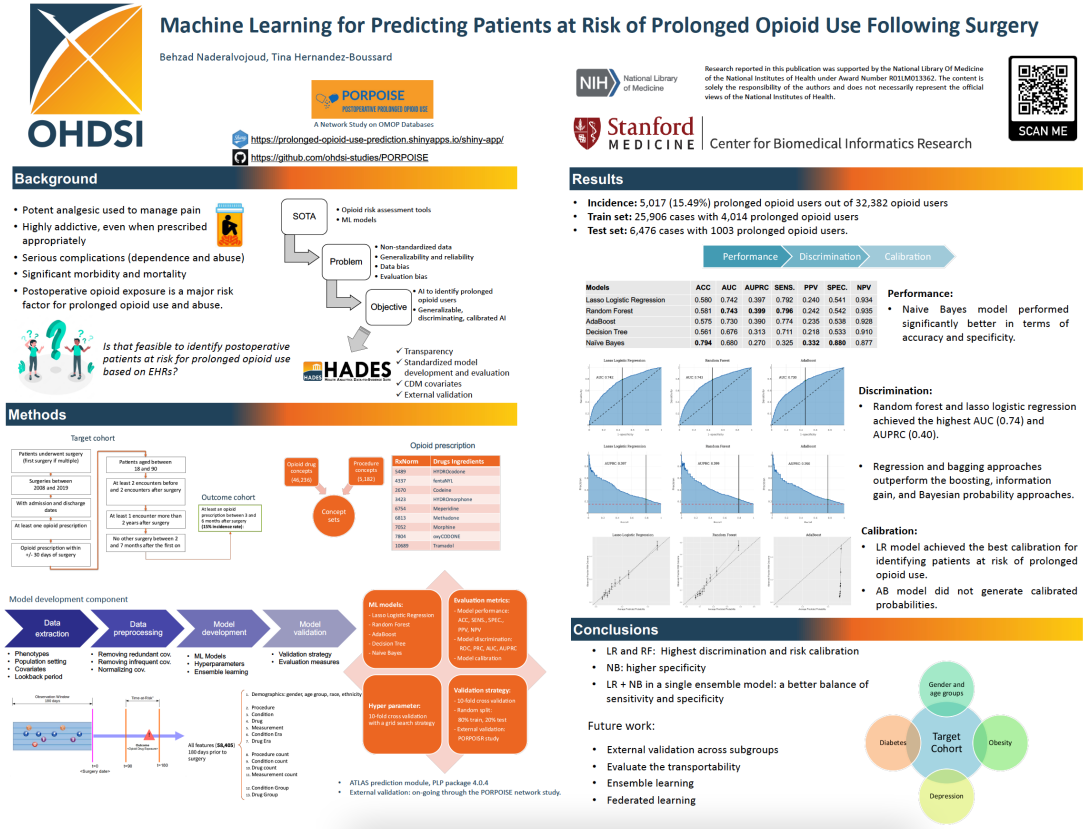
[1] Odysseus Data Services
[2] IKNL, [3] EORTC/Pioneer, [4] NICE
*astiaan.van.sandijk@odysseusinc.com

ODYSSEUS DATA SERVICES INC · IKNL integraal kankercentrum Nederland · EORTC · EHDEN · OHDSI

WEDNESDAY

The manifold presentations of PROMS and questionnaires: patient-reported outcomes in OMOP use cases (Sebastiaan van Sandijk, Peter Prinsen, Mieke van Hemelrijck, Michael Kallfelz, Dalia Dawoud)

#OHDSISocialShowcase This Week



THURSDAY Machine Learning for Predicting Patients at Risk of Prolonged Opioid Use Following Surgery (Behzad Naderalvojoud, Tina Hernandez-Boussard)

#OHDSISocialShowcase This Week

Post-Stroke Prediction on Cognitive Impairment Development: A Machine Learning Approach

Muhammad Solihuddin Muhtar[1], Faizul Hasan[2], Alex P.A. Nguyen[1], Jason C. Hsu[1], Hsiao-Yean Chiu[2], Min-Huei Hsu[1]

[1]Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei, Taiwan
[2]School of Nursing, College of Nursing, Taipei Medical University, Taipei, Taiwan.

Background

Cognitive impairment following stroke has wide prevalence ranging from 25% to 81%. Further, stroke and the subtypes, including ischemic stroke, transient ischemic attack and intracerebral hemorrhage, significantly increase the long-term risk of dementia after 5 and 10 years. The incidence rate of post-stroke dementia increases yearly, though the relative risk gradually decreases. The study aims to predict the dementia development one year after stroke diagnose (index date).

Methods

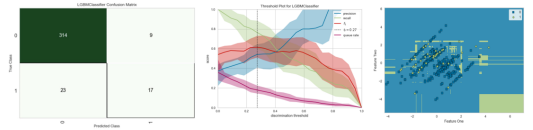
The study conducted on TMUCRD from January 2004 to September 2017. The inclusion, exclusion and outcome criteria are selected based on ICD9 and ICD10 codes. We include all patient with history of stroke, insomnia, cognitive impairment and other codes related with the diseases (362.3, 433.x1, 434.x1, 436, 431.x, 430.x, 435.x, H34.1, I63.x, I64.x, I61.x, I60.x, G45.x). We exclude psychiatric disorder, sleep apnea, traumatic brain injury, cancer, Parkinson's disease, and cognitive impairment from the outcome (300.4, 296.2-296.3, 300, 293.84, 296.4-296.7, 295, 327.23, 800-804, 850.0, 850.1, 850.5, 850.9, 854.0, 959.01, 199.1, 332.0, F34.1, F32.9, F41. 9, F31, F31.x, F32, F32.x, F33, F33.x, F20. 9, G47. 33, S02.0, S02.1, S02.8, S02.91, S04.02, S04.03, S04.04, S06, S07.1, T74.4, S09.90, C80.1, G20). The outcome are mild cognitive impairment, Senile dementia, uncomplicated, Senile dementia with delusional or depressive features, Senile dementia with delirium, Dementia in conditions classified elsewhere, Alzheimer's disease, Frontotemporal dementia, and Senile degeneration of brain (331.83, 290.0, 290.1, 290.2, 290.3, 294.1, 331.0, 331.1, 331.2, G31.84, R41.89, R41.84, R41.83, R41.82, R41.81, F03.90, F03). The patient with outcome at least one year after stroke index date are labelled by 1, and the rest without outcome labelled by 0. We use pycaret library to compare the performance of many different machine learning algorithms.

Results

Preliminary result performed on the holdout data (453 out of 4935 patients). LightGBM algorithm gives the best AUC metrics on the 10 fold cross validation training. The scores are 0.82 for accuracy, 0.81 for AUC, 0.17 for precision, 0.32 for recall, and 0.21 for F1. In the current model, we only use ICD and gender/age for the features.

Results

We can see the performance in Figure 1. The true positive is still less than the false negative. This is to be expected since we use common features between two labels. The Figure 2 shows the optimal threshold is pretty low, 0.27, much lower than default 0.5 for binary classification. Figure 3 shows the distribution of the data based on labels. We can see that some positive labels are overlapping with the negative ones (have the same features).



Conclusions

The current model is able to determine whether a patient will develop cognitive impairment in the next year or not, though the probability is still very low, by using only gender, age and ICD code. Further features engineering will be conducted to improve the performance, such as adding medication or demographic features, especially to increase the true positive and to reduce the false negative numbers. Some hyperparameters may need to be adjusted to obtain better metrics, since the current model still uses the default pycaret parameters. We plan to run it on the CDM once the final result is reasonable.

References

1. del Ser, T., Barba, R., Morin, M. M., Domingo, J., Cemillan, C., Pondal, M., & Vivancos, J. (2005). Evolution of cognitive impairment after stroke and risk factors for delayed progression. Stroke, 36(12), 2670-2675
2. Li, C. H., Chang, Y. H., Chou, M. C., Chen, C. H., Ho, B. L., Hsieh, S. W., & Yang, Y. H. (2019). Factors of post-stroke dementia: A nationwide cohort study in Taiwan. Geriatr Gerontol Int, 19(8),815-822. doi:10.1111/ggi.13725

Contact: jasonhsu@tmu.edu.tw

Acknowledgement
We appreciate the technical assistance provided by the staff of Taipei Medical University's Office of Information Technology.

FRIDAY One-year Post-Stroke Prediction on Cognitive Impairment: A Machine Learning Approach (Muhammad Solihuddin Muhtar, Faizul Hasan, Alex P.A. Nguyen, Jason C. Hsu, Hsiao-Yean Chiu)

#OHDSISocialShowcase This Week

Describing treatment with antidiabetics in patients with T2D and moderate to severe CKD across a network of OMOP databases

PRESENTER: Martin Lavallee

INTRODUCTION

- Chronic kidney disease (CKD) is a common complication of type 2 diabetes (T2D) (1)
- Tight glycemic control is instrumental in preventing development and progression of CKD (2)
- Despite availability of several anti-glycemics in the US and EU, little evidence is available on their real-world utilization for persons with T2D and CKD (3,4).

Objective: Describe demographics and clinical characteristics, post-index drug utilization, treatment patterns and time to discontinuation for patients with CKD and T2D who initiate one of 6 anti-glycemic drug classes.

METHODS

- **Data:** UK General Practitioner electronic health record EHR (CPRD Gold and CPRD AURUM), US nationwide claims (Optum and Truven Marketscan [CCAE and MDCR]) and US nationwide EHR (Optum)
- **Study Population:** Patients with CKD and T2D who are new-users of either SGLT2 inhibitors, GLP1-RA, Sulfonylurea, DPP4 inhibitors, insulin, or metformin.
- **Cohort Diagnostics:** used to assess the fitness of the drug cohorts (5).
- **Covariate Analysis:** Described patient demographics and clinical characteristics of patients at baseline defined as one year prior to index. Described frequency of medications used at 0-90 , 91-183 , 184 –365 and 366- 730 days post index using Feature Extraction (6).
- **Treatment Patterns:** Depicted using an adaptation of TreatmentPatterns (7). Time to discontinuation estimated using kaplan-meier curves.

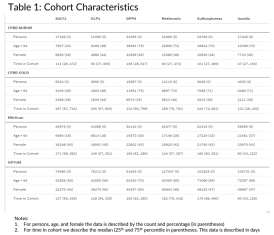
RESULTS

We assessed a total of 167.8 million patients across the 4 databases. The study time frame was from January 1st, 2012, to December 31st, 2020, for each database.

DISCUSSION

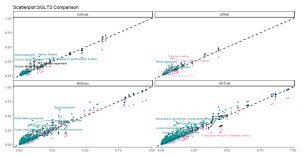
- Minor differences in clinical characteristics and post-index utilization were observed across six drug classes.
- New users of SGLT2 and GLP1 are on average prescribed more to younger and male populations.
- Patients stayed on medications on average longer in the UK than in the US. US patients on average tended to switch between medications.

Table 1: Cohort Characteristics



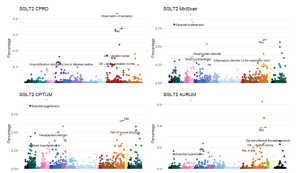
Large-scale characterization of six anti-glycemic drug classes enables **detection of treatment variation** across **geographies** and **settings**

Figure 1: Characteristics of new users of SGLT2 compared to other drugs in study (metformin, glp1, dpp4, sulfonylureas, and insulin)



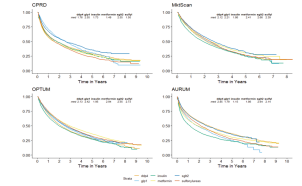
Notes: The condition characteristics between sglt2 and insulin populations show the most difference. This makes sense sence insulin is typically used as a last line of therapy. CPRD shows little deviation in terms of prevalence rates this may be an artifact of diagnosis coding.

Figure 2: Prevalence of conditions of SGLT2



Notes: The Manhattan plots show comorbidity profiles for the SGLT2 patients, giving a rough sense of which conditions are highly expressed across the databases. Except for CPRD, T2D is highly expressed in the databases. The variation is likely due to differences in coding. The other condition to stand out is hypertension in the US databases. It seems that SGLT2 patient profiles are largely consistent across databases.

Figure 3: Time to discontinuation of anti-glycemic monotherapy



Notes: The KM plots show the median time to discontinuation for each drug class in each database. Insulin usually has the shortest median time to discontinuation where as the longest varies by database (metformin for OPTUM, SGLT2 for MktScan, DPP4 for AURUM and SGLT2 for CPRD).

References

1. Nature Reviews Nephrology. 2016;12(5):73-81.
2. Diabetes care. 2018;41(12):2669-701.
3. Journal of Korean medical science. 2021;36(6):e328-e328.
4. Diabetes, Obesity and Metabolism. 2018;20(2):427-34.
5. CohortDiagnostics: Diagnostics for OHDSI Cohorts; 2022.
6. FeatureExtraction: Generating Features for a Cohort [Internet]; 2021. Available from: https://github.com/OHDSI/FeatureExtraction
7. Computer Methods and Programs in Biomedicine. 2022;221:107001.

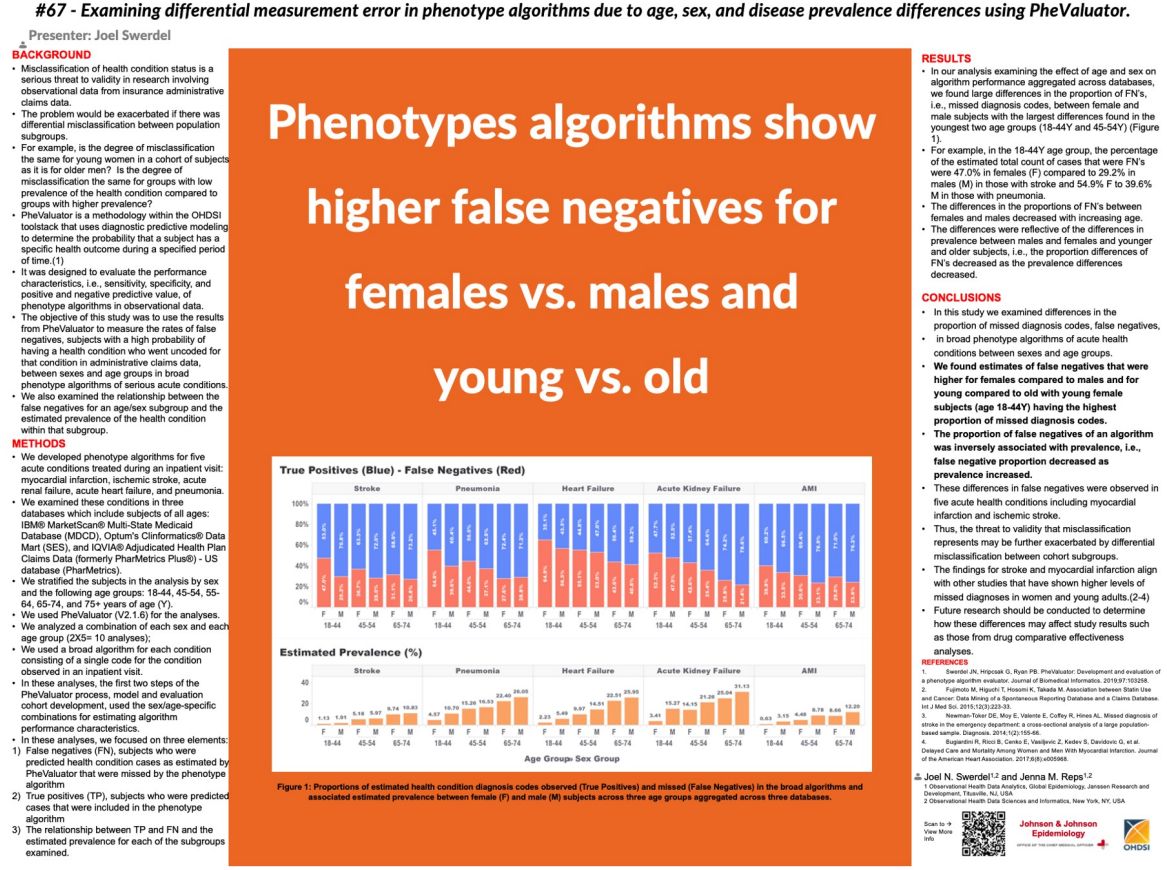
David Vizcaya, Ron Herrera, Nikolaus Oberprieler, Glen James, Darya Kosareva, Asieh Golozar

ODYSSEUS DATA SERVICES INC | BAYER | OHDSI

MONDAY

Describing treatment with antidiabetics in patients with T2D and moderate to severe CKD across a network of OMOP databases (Martin Lavallee, David Vizcaya, Ron Herrera, Niki Oberprieler, Glen James, Darya Kosareva, Asieh Golozar)

#OHDSISocialShowcase This Week



#67 - Examining differential measurement error in phenotype algorithms due to age, sex, and disease prevalence differences using PheValuator.

Presenter: Joel Swerdel

BACKGROUND

- Misclassification of health condition status is a serious threat to validity in research involving observational data from insurance administrative claims data.
- The problem would be exacerbated if there was differential misclassification between population subgroups.
- For example, is the degree of misclassification the same for young women in a cohort of subjects as it is for older men? Is the degree of misclassification the same for groups with low prevalence of the health condition compared to groups with higher prevalence?
- PheValuator is a methodology within the OHDSI toolstack that uses diagnostic predictive modeling to determine the probability that a subject has a specific health outcome during a specified period of time.(1)
- It was designed to evaluate the performance characteristics, i.e., sensitivity, specificity, and positive and negative predictive value, of phenotype algorithms in observational data.
- The objective of this study was to use the results from PheValuator to measure the rates of false negatives, subjects with a high probability of having a health condition who went uncoded for that condition in administrative claims data, between sexes and age groups in broad phenotype algorithms of serious acute conditions.
- We also examined the relationship between the false negatives for an age/sex subgroup and the estimated prevalence of the health condition within that subgroup.

METHODS

- We developed phenotype algorithms for five acute conditions treated during an inpatient visit: myocardial infarction, ischemic stroke, acute renal failure, acute heart failure, and pneumonia.
- We examined these conditions in three databases which include subjects of all ages: IBM® MarketScan® Multi-State Medicaid Database (MDCD), Optum's Clinformatics® Data Mart (SES), and IQVIA® Adjudicated Health Plan Claims Data (formerly PharMetrics Plus®) - US database (PharMetrics).
- We stratified the subjects in the analysis by sex and the following age groups: 18-44, 45-54, 55-64, 65-74, and 75+ years of age (Y).
- We used PheValuator (V2.1.6) for the analyses.
- We analyzed a combination of each sex and each age group (2X5= 10 analyses);
- We used a broad algorithm for each condition consisting of a single code for the condition observed in an inpatient visit.
- In these analyses, the first two steps of the PheValuator process, model and evaluation cohort development, used the sex/age-specific combinations for estimating algorithm performance characteristics.
- In these analyses, we focused on three elements:
1) False negatives (FN), subjects who were predicted health condition cases as estimated by PheValuator that were missed by the phenotype algorithm
2) True positives (TP), subjects who were predicted cases that were included in the phenotype algorithm
3) The relationship between TP and FN and the estimated prevalence for each of the subgroups examined.

Phenotypes algorithms show higher false negatives for females vs. males and young vs. old

Figure 1: Proportions of estimated health condition diagnosis codes observed (True Positives) and missed (False Negatives) in the broad algorithms and associated estimated prevalence between female (F) and male (M) subjects across three age groups aggregated across three databases.

RESULTS

- In our analysis examining the effect of age and sex on algorithm performance aggregated across databases, we found large differences in the proportion of FN's, i.e., missed diagnosis codes, between female and male subjects with the largest differences found in the youngest two age groups (18-44Y and 45-54Y) (Figure 1).
- For example, in the 18-44Y age group, the percentage of the estimated total count of cases that were FN's were 47.0% in females (F) compared to 29.2% in males (M) in those with stroke and 54.9% F to 39.6% M in those with pneumonia.
- The differences in the proportions of FN's between females and males decreased with increasing age.
- The differences were reflective of the differences in prevalence between males and females and younger and older subjects, i.e., the proportion differences of FN's decreased as the prevalence differences decreased.

CONCLUSIONS

- In this study we examined differences in the proportion of missed diagnosis codes, false negatives,
- in broad phenotype algorithms of acute health conditions between sexes and age groups.
- **We found estimates of false negatives that were higher for females compared to males and for young compared to old with young female subjects (age 18-44Y) having the highest proportion of missed diagnosis codes.**
- **The proportion of false negatives of an algorithm was inversely associated with prevalence, i.e., false negative proportion decreased as prevalence increased.**
- These differences in false negatives were observed in five acute health conditions including myocardial infarction and ischemic stroke.
- Thus, the threat to validity that misclassification represents may be further exacerbated by differential misclassification between cohort subgroups.
- The findings for stroke and myocardial infarction align with other studies that have shown higher levels of missed diagnoses in women and young adults.(2-4)
- Future research should be conducted to determine how these differences may affect study results such as those from drug comparative effectiveness analyses.

REFERENCES

1. Swerdel JN, Hripcsak G, Ryan PB. PheValuator: Development and evaluation of a phenotype algorithm evaluator. Journal of Biomedical Informatics. 2019;97:103258.
2. Fujimoto M, Higuchi T, Hosomi K, Takada M. Association between Statin Use and Cancer: Data Mining of a Spontaneous Reporting Database and a Claims Database. Int J Med Sci. 2015;12(3):223-33.
3. Newman-Toker DE, Moy E, Valente E, Coffey R, Hines AL. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. Diagnosis. 2014;1(2):155-66.
4. Bugiardini R, Ricci B, Cenko E, Vasiljevic Z, Kedev S, Davidovic G, et al. Delayed Care and Mortality Among Women and Men With Myocardial Infarction. Journal of the American Heart Association. 2017;6(8):e005968.

Joel N. Swerdel[1,2] and Jenna M. Reps[1,2]
1 Observational Health Data Analytics, Global Epidemiology, Janssen Research and Development, Titusville, NJ, USA
2 Observational Health Data Sciences and Informatics, New York, NY, USA

Scan to View More Info

Johnson & Johnson Epidemiology

OHDSI

TUESDAY Examining differential measurement error in phenotype algorithms due to age, sex, and disease prevalence differences using PheValuator (Joel Swerdel, Jenna Reps)

#OHDSISocialShowcase This Week

Lowering the OMOP ETL Barrier for Clinical Registries

PRESENTER: **Smith Heavner**

INTRODUCTION

Goal: To systematically collect anecdotal data of how clinicians use existing drugs in new ways to treat diseases with limited or no treatment options.
Aim: To simplify ETL process and create pathway for real-world data to be made available in CURE ID
Method: Data harmonization using OMOP CDM
Impact: Secure deidentified data elements for assessing the effectiveness of repurposed treatments for diseases of high unmet clinical need.

METHODS

With funding from HHS Assistant Secretary for Planning and Evaluation:

1. Developed tools and resources to disseminate harmonization methods developed by Johns Hopkins University
2. Expand CURE ID from physician entered reports to automated EHR data collection
3. Promote conversion of non-common data model EHR systems to OMOP standards
4. Developed minimal dataset for drug repurposing research in COVID-19 as a use case

RESULTS

- Possible to lower OMOP ETL barrier
  - Default configuration transformations
  - Support and feedback from experienced site
- Project continues with site implementing OMOP using Perseus

Pilot site implemented OMOP in less than 200 hours

DataQualityDashboard Version: 1.4.1
Results generated at 2022-07-07 21:20:09 in 11 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	1982	31	2013	98%	287	0	287	100%	2269	31	2300	99%
Conformance	656	23	679	97%	104	0	104	100%	760	23	783	97%
Completeness	381	5	386	99%	11	4	15	73%	392	9	401	98%
Total	3019	59	3078	98%	402	4	406	99%	3421	63	3484	**98%**

- **Streamlined process for future sites**

Take a picture to **download** the **full paper**

CONCLUSION:

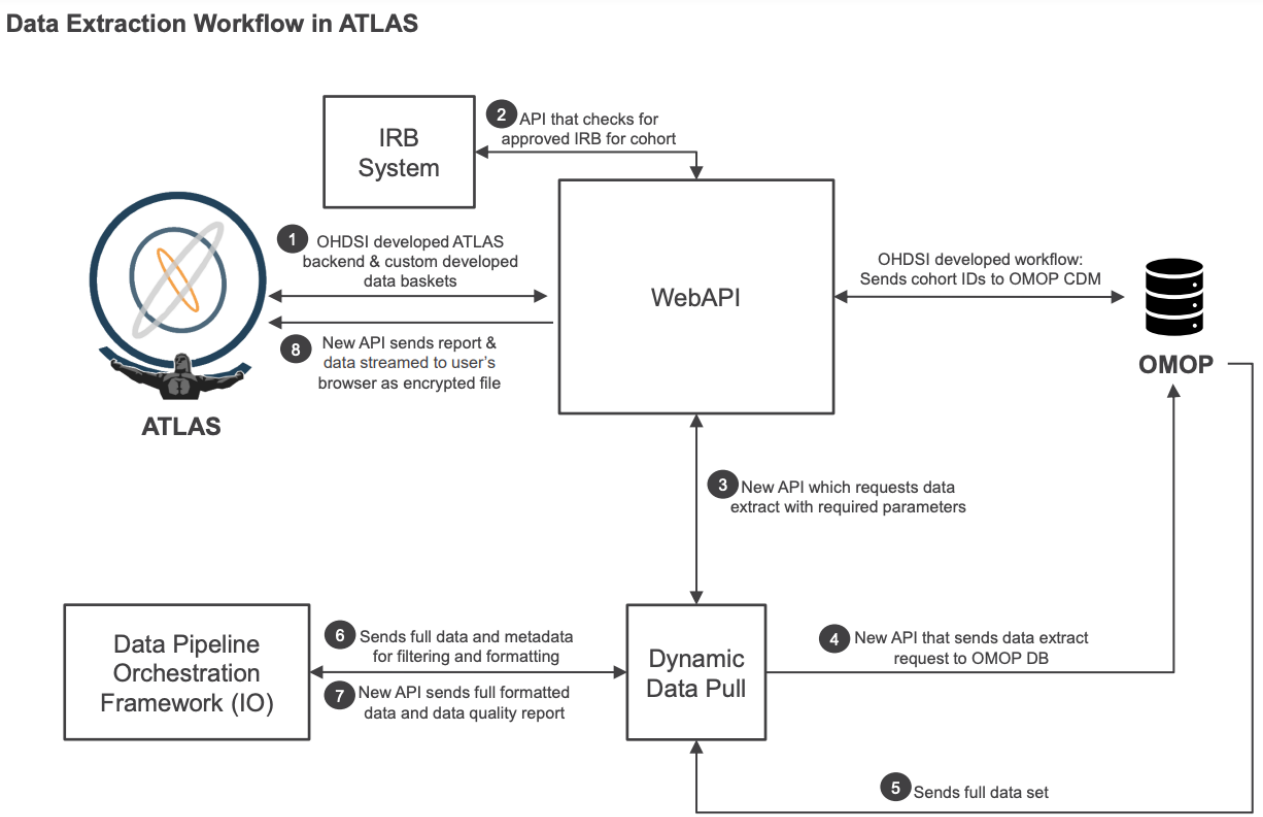
It is feasible to reduce the ETL implementation time by providing default configuration transformations along with assistance and feedback on the process. Further reduction in the person-hours required to perform an OMOP ETL will be evaluated with the Perseus web based OHDSI ETL project and cloud provider deployments of Atlas and the DQD. Our goal is to increase the adoption of OMOP in sites with fewer resources and enable wider participation in high-quality clinical registries with sufficient patient numbers and data variables to perform appropriate observational research techniques to control for potential confounders (e.g., propensity score matching).

Smith Heavner[1,2], Trayson Llano[3], Zachary Wang[4], Marco Schito[1], Heather Stone[5], Pam Dasher[1], Tresha Russel[6], Vishakha Kumar[6], Ben Saeks[4], Michael Cooke[4], Rahul Kashyap[6], Matt Robinson[4], Paul Nagy[4]
1. CURE Drug Repurposing Collaboratory, Critical Path Institute 2. Clemson University, Department of Public Health Sciences 3. Prisma Health 4. Johns Hopkins University 5. Food and Drug Administration 6. Society for Critical Care Medicine

WEDNESDAY

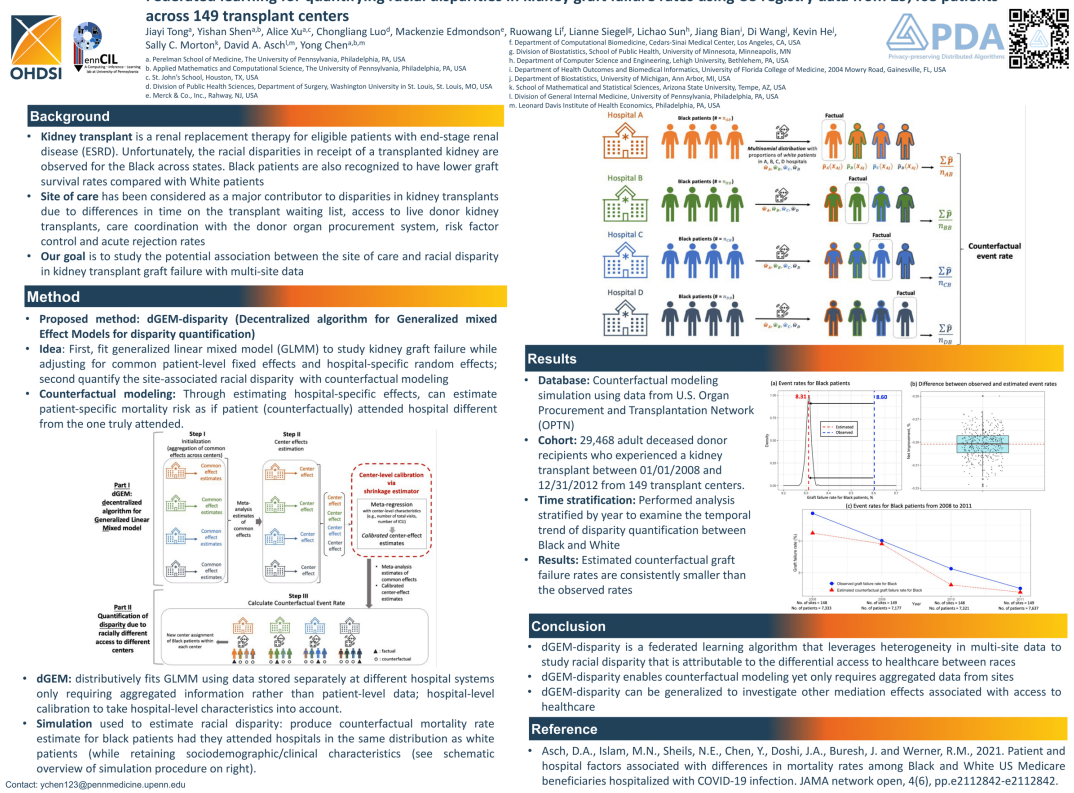
Lowering the OMOP ETL Barrier for Clinical Registries (Smith Heavner, Trayson Llano, Zachary Wang, Marco Schito, Heather Stone, Pam Dasher, Tresha Russel, Vishakha Kumar, Ben Saeks, Michael Cooke, Rahul Kashyap, Matthew Robinson, Paul Nagy)

#OHDSISocialShowcase This Week



THURSDAY **Einstein-ATLAS: Leveraging OHDSI/ATLAS and Open-Source Development to Support Translational Research, Data Science, and Regulatory Compliance** (Parsa Mirhaji, Selvin Soby, Erin Henninger, Chandra Nelapatla, Manuel Wahle, Boudewijn Aasman, Eran Bellin)

#OHDSISocialShowcase This Week



Federated learning for quantifying racial disparities in kidney graft failure rates using US registry data from 29,468 patients across 149 transplant centers

Jiayi Tong[a], Yishan Shen[a,b], Alice Xu[a,c], Chongliang Luo[d], Mackenzie Edmondson[e], Ruowang Li[f], Lianne Siegel[g], Lichao Sun[h], Jiang Bian[i], Di Wang[j], Kevin He[l], Sally C. Morton[k], David A. Asch[l,m], Yong Chen[a,b,m]

a. Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA, USA
b. Applied Mathematics and Computational Science, The University of Pennsylvania, Philadelphia, PA, USA
c. St. John's School, Houston, TX, USA
d. Division of Public Health Sciences, Department of Surgery, Washington University in St. Louis, St. Louis, MO, USA
e. Merck & Co., Inc., Rahway, NJ, USA
f. Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA
g. Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN
h. Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA
i. Department of Health Outcomes and Biomedical Informatics, University of Florida College of Medicine, 2004 Mowry Road, Gainesville, FL, USA
j. Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
k. School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, USA
l. Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA, USA
m. Leonard Davis Institute of Health Economics, Philadelphia, PA, USA

Background

- **Kidney transplant** is a renal replacement therapy for eligible patients with end-stage renal disease (ESRD). Unfortunately, the racial disparities in receipt of a transplanted kidney are observed for the Black across states. Black patients are also recognized to have lower graft survival rates compared with White patients
- **Site of care** has been considered as a major contributor to disparities in kidney transplants due to differences in time on the transplant waiting list, access to live donor kidney transplants, care coordination with the donor organ procurement system, risk factor control and acute rejection rates
- **Our goal** is to study the potential association between the site of care and racial disparity in kidney transplant graft failure with multi-site data

Method

- **Proposed method: dGEM-disparity (Decentralized algorithm for Generalized mixed Effect Models for disparity quantification)**
- **Idea**: First, fit generalized linear mixed model (GLMM) to study kidney graft failure while adjusting for common patient-level fixed effects and hospital-specific random effects; second quantify the site-associated racial disparity with counterfactual modeling
- **Counterfactual modeling:** Through estimating hospital-specific effects, can estimate patient-specific mortality risk as if patient (counterfactually) attended hospital different from the one truly attended.

- **dGEM:** distributively fits GLMM using data stored separately at different hospital systems only requiring aggregated information rather than patient-level data; hospital-level calibration to take hospital-level characteristics into account.
- **Simulation** used to estimate racial disparity: produce counterfactual mortality rate estimate for black patients had they attended hospitals in the same distribution as white patients (while retaining sociodemographic/clinical characteristics (see schematic overview of simulation procedure on right).

Contact: ychen123@pennmedicine.upenn.edu

Results

- **Database:** Counterfactual modeling simulation using data from U.S. Organ Procurement and Transplantation Network (OPTN)
- **Cohort:** 29,468 adult deceased donor recipients who experienced a kidney transplant between 01/01/2008 and 12/31/2012 from 149 transplant centers.
- **Time stratification:** Performed analysis stratified by year to examine the temporal trend of disparity quantification between Black and White
- **Results:** Estimated counterfactual graft failure rates are consistently smaller than the observed rates

Conclusion

- dGEM-disparity is a federated learning algorithm that leverages heterogeneity in multi-site data to study racial disparity that is attributable to the differential access to healthcare between races
- dGEM-disparity enables counterfactual modeling yet only requires aggregated data from sites
- dGEM-disparity can be generalized to investigate other mediation effects associated with access to healthcare

Reference

- Asch, D.A., Islam, M.N., Sheils, N.E., Chen, Y., Doshi, J.A., Buresh, J. and Werner, R.M., 2021. Patient and hospital factors associated with differences in mortality rates among Black and White US Medicare beneficiaries hospitalized with COVID-19 infection. JAMA network open, 4(6), pp.e2112842-e2112842.

FRIDAY

Federated learning for quantifying racial disparities in kidney graft failure rates using US registry data from 29,468 patients across 149 transplant centers (Jiayi Tong, Yishan Shen, Alice Xu, Chongliang Luo, Mackenzie Edmondson, Ruowang Li, Di Wang, Kevin He, David A. Asch, Yong Chen)

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?