Extracting OHDSI Concepts from Clinical Narratives for COVID

OHDSI Community Call

Jan. 25, 2022 • 11 am ET

Future OHDSI Community Calls

Date	Topic
Jan. 25	Extracting OHDSI Concepts from Clinical Narratives for COVID
Feb. 1	Introduction to Phenotype Phebruary
Feb. 8	Phenotype Phebruary Report #1, Workgroup Updates
Feb. 15	Phenotype Phebruary Report #2, Workgroup Updates
Feb. 22	Phenotype Phebruary Report #3, Workgroup Updates

Future OHDSI Community Calls

Date	Topic
Jan. 25	Extracting OHDSI Concepts from Clinical Narratives for COVID
Feb. 1	Introduction to Phenotype Phebruary
Feb. 8	Phenotype Phebruary Report #1, Workgroup Updates
Feb. 15	Phenotype Phebruary Report #2, Workgroup Updates
Feb. 22	Phenotype Phebruary Report #3, Workgroup Updates

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to co-authors **ChulHyoung Park, Seng Chan You, Hokyun Jeon, Chang Won Jeong, Jin Wook Choi, and Rae Woong Park** on the study "**Development and Validation of the Radiology Common Data Model (R-CDM) for the International Standardization of Medical Imaging Data**" which was recently published in the Yonsei Medical Journal.

Original Article
Yonsei Med J 2022 Jan;63 Suppl:S74-83
https://doi.org/10.3349/ymj.2022.63.S74
Yonsei Medical Journal YMJ
pISSN: 0513-5796 · eISSN: 1976-2437

Development and Validation of the Radiology Common Data Model (R-CDM) for the International Standardization of Medical Imaging Data

ChulHyoung Park[1*], Seng Chan You[2*], Hokyun Jeon[1], Chang Won Jeong[3], Jin Wook Choi[4], and Rae Woong Park[1,5]

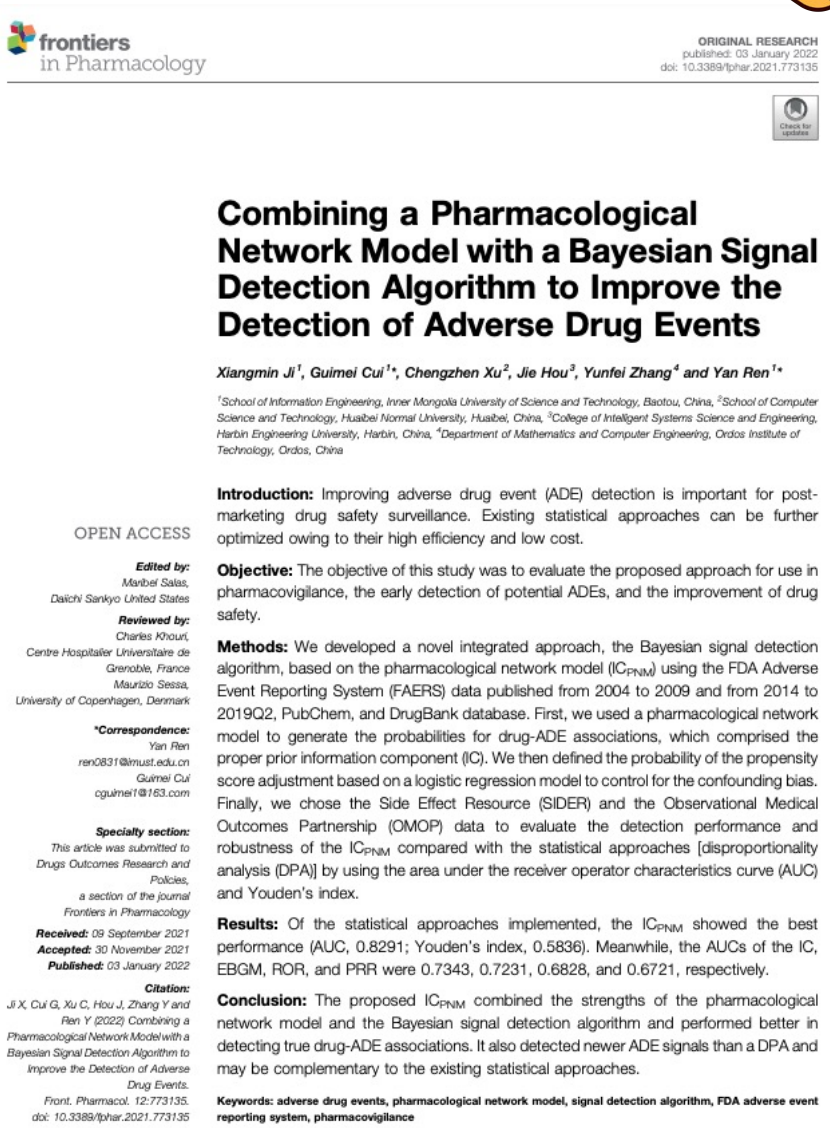
[1]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon;
[2]Department of Preventive Medicine, Yonsei University College of Medicine, Seoul;
[3]Medical Convergence Research Center, Wonkwang University, Iksan;
[4]Department of Radiology, Ajou University Medical Center, Suwon;
[5]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea.

Purpose: Digital Imaging and Communications in Medicine (DICOM), a standard file format for medical imaging data, contains metadata describing each file. However, metadata are often incomplete, and there is no standardized format for recording metadata, leading to inefficiency during the metadata-based data retrieval process. Here, we propose a novel standardization method for DICOM metadata termed the Radiology Common Data Model (R-CDM).
Materials and Methods: R-CDM was designed to be compatible with Health Level Seven International (HL7)/Fast Healthcare Interoperability Resources (FHIR) and linked with the Observational Medical Outcomes Partnership (OMOP)-CDM to achieve a seamless link between clinical data and medical imaging data. The terminology system was standardized using the RadLex playbook, a comprehensive lexicon of radiology. As a proof of concept, the R-CDM conversion process was conducted with 41.7 TB of data from the Ajou University Hospital. The R-CDM database visualizer was developed to visualize the main characteristics of the R-CDM database.
Results: Information from 2801360 cases and 87203226 DICOM files was organized into two tables constituting the R-CDM. Information on imaging device and image resolution was recorded with more than 99.9% accuracy. Furthermore, OMOP-CDM and R-CDM were linked to efficiently extract specific types of images from specific patient cohorts.
Conclusion: R-CDM standardizes the structure and terminology for recording medical imaging data to eliminate incomplete and unstandardized information. Successful standardization was achieved by the extract, transform, and load process and image classifier. We hope that the R-CDM will contribute to deep learning research in the medical imaging field by enabling the securement of large-scale medical imaging data from multinational institutions.

Key Words: Metadata, standardization, radiology information system

OHDSI Shoutouts!

Congratulations to co-authors **Xiangmin Ji, Guimei Cui, Chengzhen Xu, Jie Hou, Yunfei Zhang, and Yan Ren** on the study "**Combining a Pharmacological Network Model with a Bayesian Signal Detection Algorithm to Improve the Detection of Adverse Drug Events**" which was recently published in Frontiers of Pharmacology.



frontiers in Pharmacology

ORIGINAL RESEARCH
published: 03 January 2022
doi: 10.3389/fphar.2021.773135

Combining a Pharmacological Network Model with a Bayesian Signal Detection Algorithm to Improve the Detection of Adverse Drug Events

Xiangmin Ji[1], Guimei Cui[1], Chengzhen Xu[2], Jie Hou[3], Yunfei Zhang[4] and Yan Ren[1]**

[1]School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, China, [2]School of Computer Science and Technology, Huaibei Normal University, Huaibei, China, [3]College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, China, [4]Department of Mathematics and Computer Engineering, Ordos Institute of Technology, Ordos, China

Introduction: Improving adverse drug event (ADE) detection is important for post-marketing drug safety surveillance. Existing statistical approaches can be further optimized owing to their high efficiency and low cost.

Objective: The objective of this study was to evaluate the proposed approach for use in pharmacovigilance, the early detection of potential ADEs, and the improvement of drug safety.

Methods: We developed a novel integrated approach, the Bayesian signal detection algorithm, based on the pharmacological network model (IC_{PNM}) using the FDA Adverse Event Reporting System (FAERS) data published from 2004 to 2009 and from 2014 to 2019Q2, PubChem, and DrugBank database. First, we used a pharmacological network model to generate the probabilities for drug-ADE associations, which comprised the proper prior information component (IC). We then defined the probability of the propensity score adjustment based on a logistic regression model to control for the confounding bias. Finally, we chose the Side Effect Resource (SIDER) and the Observational Medical Outcomes Partnership (OMOP) data to evaluate the detection performance and robustness of the IC_{PNM} compared with the statistical approaches [disproportionality analysis (DPA)] by using the area under the receiver operator characteristics curve (AUC) and Youden's index.

Results: Of the statistical approaches implemented, the IC_{PNM} showed the best performance (AUC, 0.8291; Youden's index, 0.5836). Meanwhile, the AUCs of the IC, EBGM, ROR, and PRR were 0.7343, 0.7231, 0.6828, and 0.6721, respectively.

Conclusion: The proposed IC_{PNM} combined the strengths of the pharmacological network model and the Bayesian signal detection algorithm and performed better in detecting true drug-ADE associations. It also detected newer ADE signals than a DPA and may be complementary to the existing statistical approaches.

Keywords: adverse drug events, pharmacological network model, signal detection algorithm, FDA adverse event reporting system, pharmacovigilance

OPEN ACCESS

Edited by:
Maribel Salas,
Daiichi Sankyo United States

Reviewed by:
Charles Khouri,
Centre Hospitalier Universitaire de Grenoble, France
Maurizio Sessa,
University of Copenhagen, Denmark

****Correspondence:***
Yan Ren
ren0831@imust.edu.cn
Guimei Cui
cguimei1@163.com

Specialty section:
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

Received: *09 September 2021*
Accepted: *30 November 2021*
Published: *03 January 2022*

Citation:
Ji X, Cui G, Xu C, Hou J, Zhang Y and Ren Y (2022) Combining a Pharmacological Network Model with a Bayesian Signal Detection Algorithm to Improve the Detection of Adverse Drug Events. Front. Pharmacol. 12:773135. doi: 10.3389/fphar.2021.773135

OHDSI Shoutouts!

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

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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Common Data Model — Vocabulary Subgroup
Tuesday	2 pm	Health Equity
Wednesday	7 am	Medical Imaging
Wednesday	10 am	Data Quality Dashboard
Wednesday	11:30 am	Latin America
Wednesday	12 pm	FHIR and OMOP Terminologies Subgroup
Thursday	10 am	Medical Devices
Friday	10 am	Phenotype Development and Evaluation
Monday	10 am	Healthcare Systems (formerly EHR)
Monday	10 am	GIS-Geographic Information System
Tuesday	10 am	Common Data Model

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

Get Access To Different Teams/WGs/Chapters

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

- [ ] ATLAS
- [ ] Clinical Trials
- [ ] Common Data Model
- [ ] Data Quality Dashboard Development
- [ ] Early-stage Researchers
- [ ] Education Work Group
- [ ] Electronic Health Record (EHR) ETL
- [ ] Geographic Information System (GIS)
- [ ] HADES Health Analytics Data-to-Evidence Suite
- [ ] Health Equity
- [ ] Latin America
- [ ] Medical Devices
- [ ] Natural Language Processing
- [ ] OHDSI APAC
- [ ] OHDSI APAC Steering Committee
- [ ] OHDSI Steering Committee
- [ ] Oncology
- [ ] Patient-Generated Health Data
- [ ] Pharmacovigiliance Evidence Investigation
- [ ] Phenotype Development and Evaluation
- [ ] Population-Level Effect Estimation / Patient-Level Prediction
- [ ] Psychiatry
- [ ] Registry (formerly UK Biobank)
- [ ] Surgery and Perioperative Medicine
- [ ] Vaccine Safety
- [ ] Vaccine Vocabulary
- [ ] Women of OHDSI

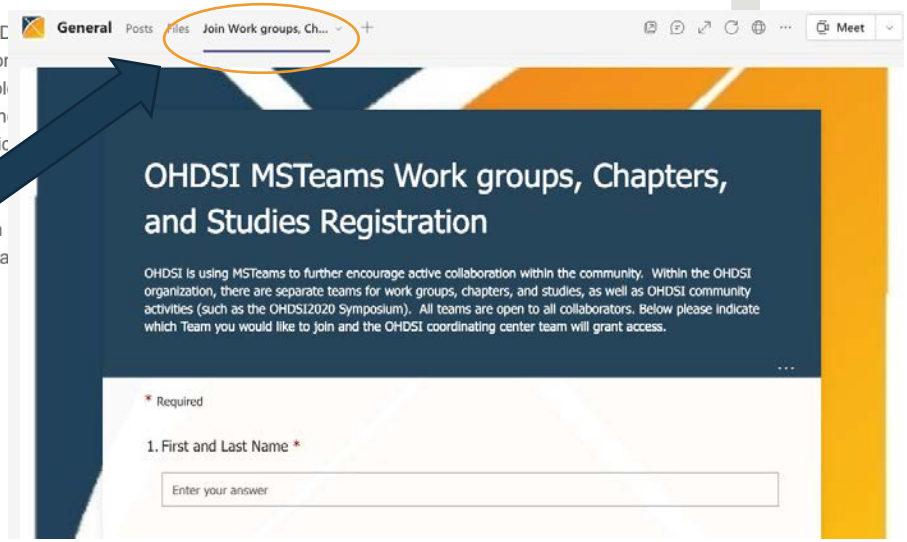
6. Select the chapter(s) you want to join

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

7. Select the studies you want to join

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

Get Access To Different Teams/WGs/Chapters



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

- [ ] ATLAS
- [ ] Clinical Trials
- [ ] Common Data Model
- [ ] Data Quality Dashboard Development
- [ ] Early-stage Researchers
- [ ] Education Work Group
- [ ] Electronic Health Record (EHR) ETL
- [ ] Geographic Information System (GIS)
- [ ] HADES Health Analytics Data-to-Evidence Suite
- [ ] Health Equity
- [ ] Latin America
- [ ] Medical Devices
- [ ] Natural Language Processing
- [ ] OHDSI APAC
- [ ] OHDSI APAC Steering Committee
- [ ] OHDSI Steering Committee
- [ ] Oncology
- [ ] Patient-Generated Health Data
- [ ] Pharmacovigiliance Evidence Investigation
- [ ] Phenotype Development and Evaluation
- [ ] Population-Level Effect Estimation / Patient-Level Prediction
- [ ] Psychiatry
- [ ] Registry (formerly UK Biobank)
- [ ] Surgery and Perioperative Medicine
- [ ] Vaccine Safety
- [ ] Vaccine Vocabulary
- [ ] Women of OHDSI

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

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

7. Select the studies you want to join

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

New Workgroups Page on OHDSI.org

ohdsi.org/ohdsi-workgroups

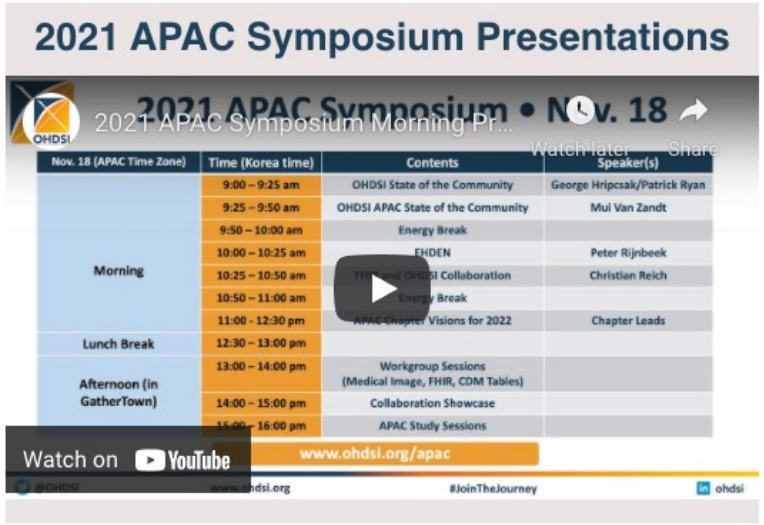
Next APAC Call: Jan. 26, 10 pm ET

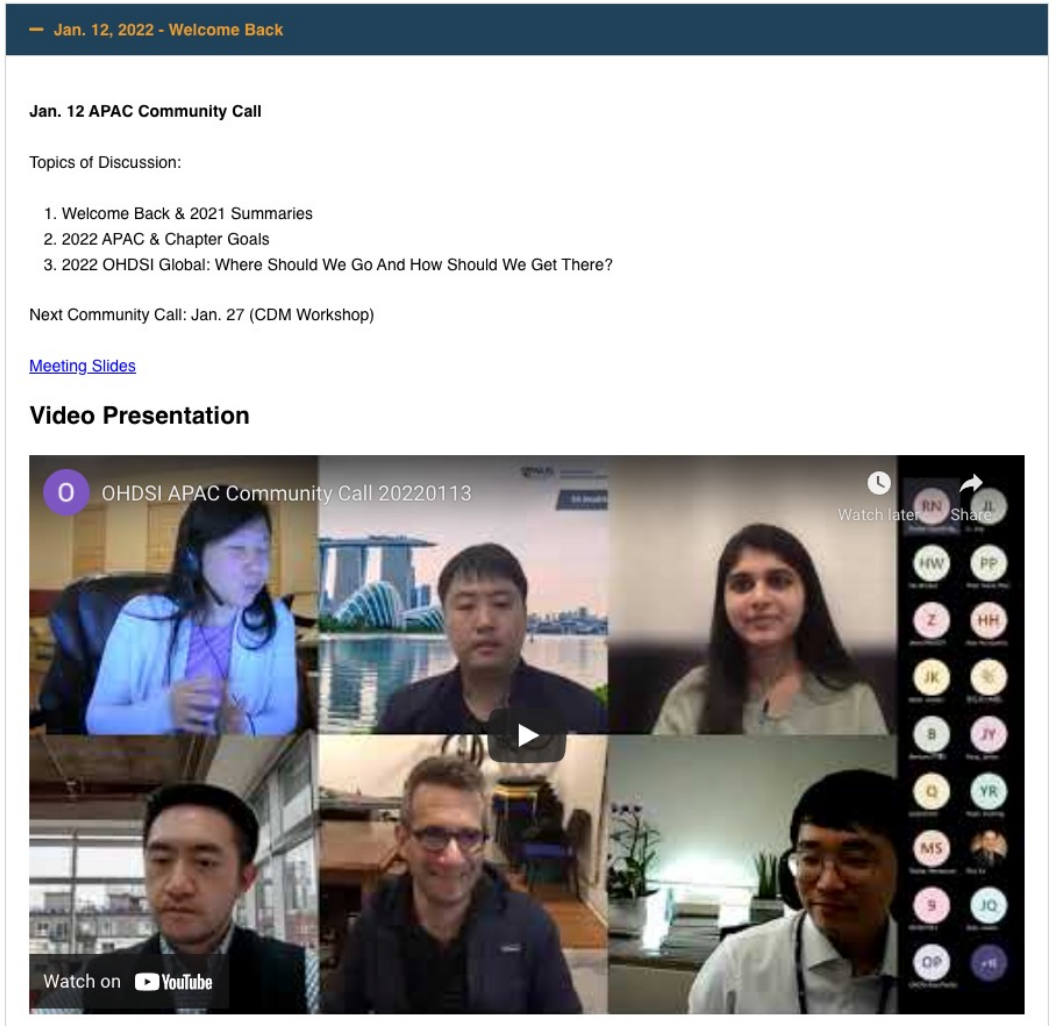
OHDSI APAC - Our Asia-Pacific Community

OHDSI is a global, multi-stakeholder, interdisciplinary and open-science network that collaborates to bring out the value of health data through large-scale analytics. Our Asia-Pacific (APAC) community comprises six regional chapters (Australia, China, Japan, Singapore, South Korea, Taiwan) and has led important OHDSI initiatives around the world.

The 2021 OHDSI Symposium was held Nov. 18, 2021, and it featured a set of morning presentations, which are available to the right. These presentations focused on the State of the Global and APAC Community, the EHDEN Consortium and the FHIR/OHDSI collaboration.

The afternoon featured several collaborative activities, including study sessions, workgroup meetings and a collaborator showcase.





ohdsi.org/apac

EHDEN 3-Year Report

Building on success: EHDEN completes its first three years!

17th January 2022

EHDEN

EUROPEAN HEALTH DATA & EVIDENCE NETWORK

enters its fourth year!

www.ehden.eu/ehden-enters-its-fourth-year

New Dates For The 2022 European Symposium

www.ohdsi-europe.org/symposium-2022

Openings!

← Back

2022 Observational Health Data Analytics Internship Program

Apply

Share Job

Location Titusville, New Jersey; Raritan, New Jersey; Horsham, Pennsylvania
Category General Administration
Req ID: 2105993940W

Get future jobs matching this search **Login** or **Register**

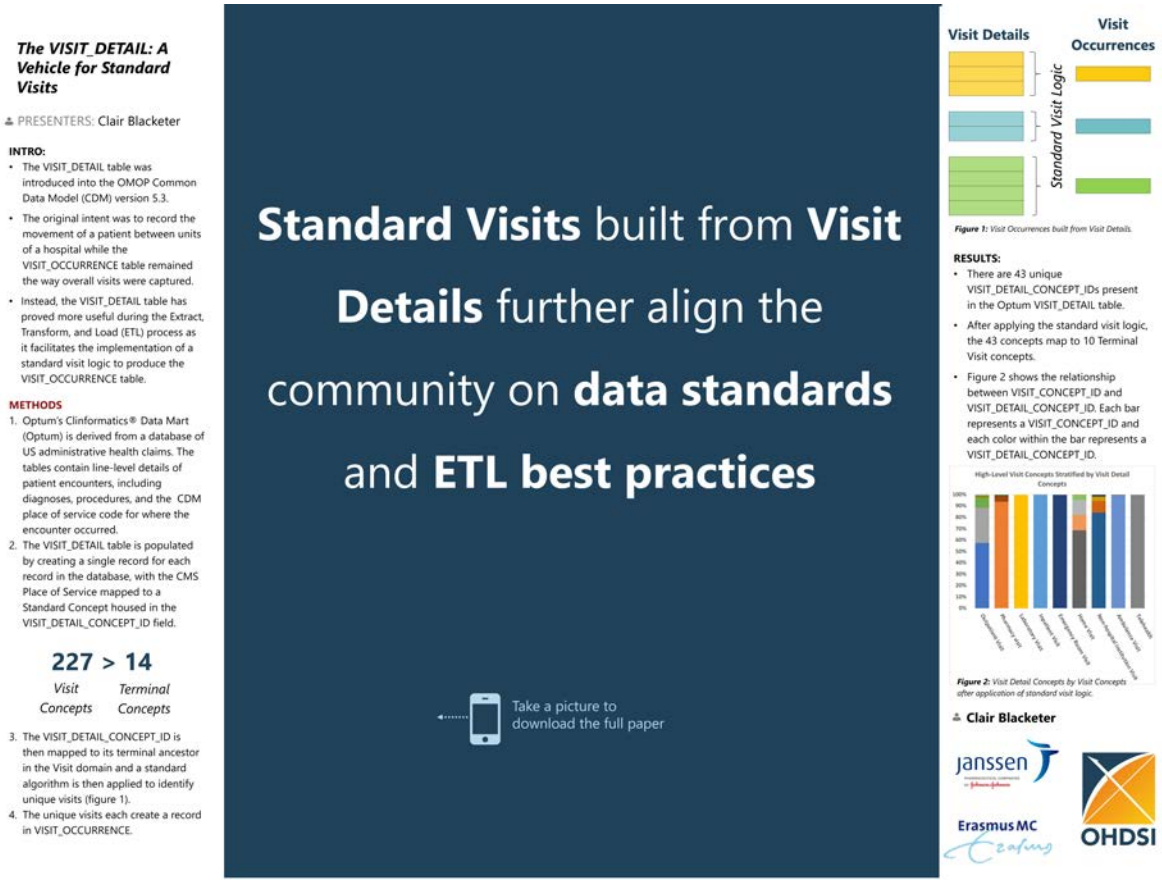
Job Description

At Johnson & Johnson, we use technology and the power of collaboration to discover new ways to prevent and overcome the world's most significant healthcare challenges. Our Corporate, Consumer Health, Medical Devices, and Pharmaceutical teams leverage data, real-world insights, and creative minds to make life-changing healthcare products and medicines. We're disrupting outdated healthcare ecosystems and infusing them with transformative ideas to help people thrive throughout every stage of their lives. With a reach of more than a billion people every day, there's no limit to the impact you can make here. Are you ready to reimagine healthcare?

Here, your career breakthroughs will change the future of health, in all the best ways. And you'll change, too. You'll be inspired, and you'll inspire people across the world to change how they care for themselves and those they love. Amplify your impact. Join us!

Janssen R&D Epidemiology is hiring an undergraduate or graduate level summer, intern. Location: Remote or In-person (Raritan, NJ; Titusville, NJ; Horsham, PA)

#OHDSISocialShowcase This Week



The VISIT_DETAIL: A Vehicle for Standard Visits

PRESENTERS: Clair Blacketer

INTRO:

- The VISIT_DETAIL table was introduced into the OMOP Common Data Model (CDM) version 5.3.
- The original intent was to record the movement of a patient between units of a hospital while the VISIT_OCCURRENCE table remained the way overall visits were captured.
- Instead, the VISIT_DETAIL table has proved more useful during the Extract, Transform, and Load (ETL) process as it facilitates the implementation of a standard visit logic to produce the VISIT_OCCURRENCE table.

METHODS

1. Optum's Clinformatics® Data Mart (Optum) is derived from a database of US administrative health claims. The tables contain line-level details of patient encounters, including diagnoses, procedures, and the CDM place of service code for where the encounter occurred.
2. The VISIT_DETAIL table is populated by creating a single record for each record in the database, with the CMS Place of Service mapped to a Standard Concept housed in the VISIT_DETAIL_CONCEPT_ID field.

227 > 14

Visit Concepts *Terminal Concepts*

3. The VISIT_DETAIL_CONCEPT_ID is then mapped to its terminal ancestor in the Visit domain and a standard algorithm is then applied to identify unique visits (figure 1).
4. The unique visits each create a record in VISIT_OCCURRENCE.

Standard Visits built from **Visit Details** further align the community on **data standards** and **ETL best practices**

Take a picture to download the full paper

Figure 1: Visit Occurrences built from Visit Details.

RESULTS:

- There are 43 unique VISIT_DETAIL_CONCEPT_IDs present in the Optum VISIT_DETAIL table.
- After applying the standard visit logic, the 43 concepts map to 10 Terminal Visit concepts.
- Figure 2 shows the relationship between VISIT_CONCEPT_ID and VISIT_DETAIL_CONCEPT_ID. Each bar represents a VISIT_CONCEPT_ID and each color within the bar represents a VISIT_DETAIL_CONCEPT_ID.

Figure 2: Visit Detail Concepts by Visit Concepts after application of standard visit logic.

Clair Blacketer

janssen

Erasmus MC

OHDSI

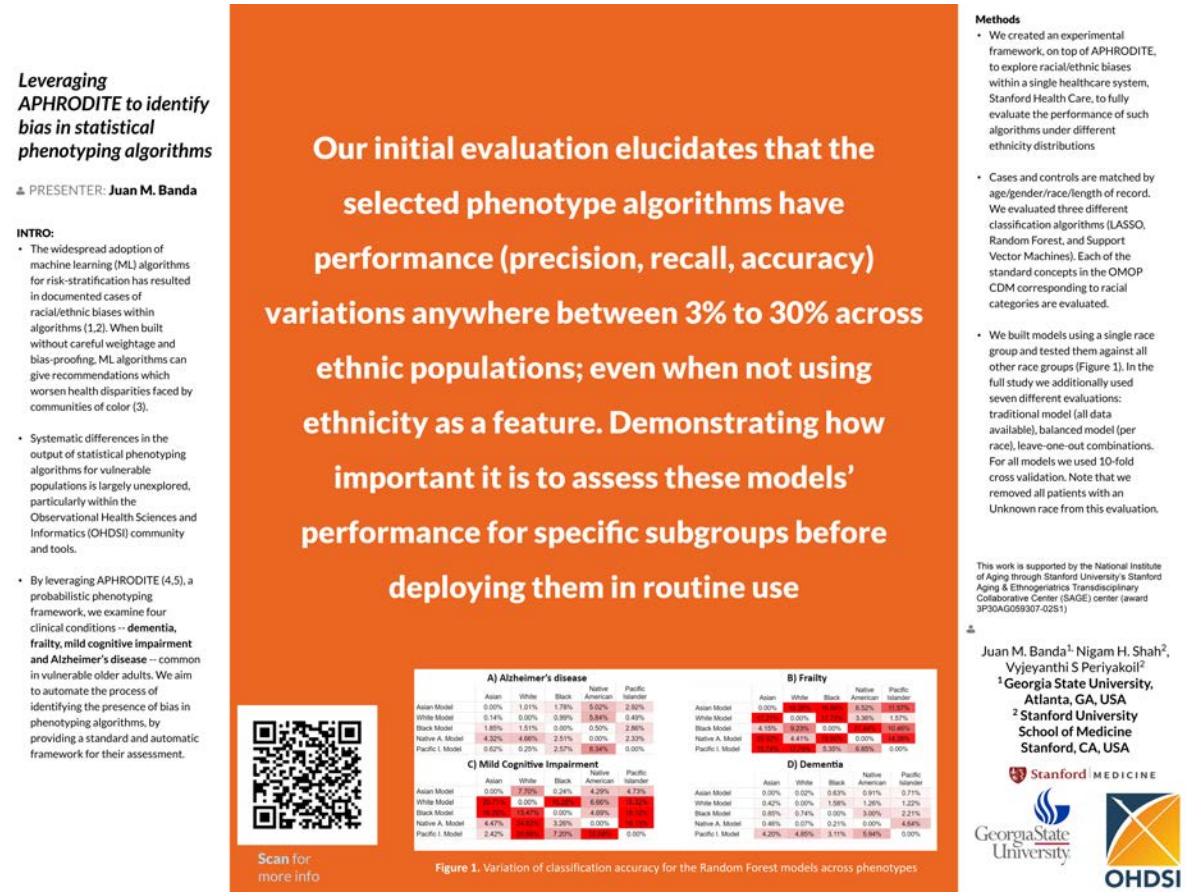
MONDAY

The VISIT_DETAIL: A Vehicle for Standard Visits

Author: Clair Blacketer

#OHDSISocialShowcase This Week

Lightning Talk!



Leveraging APHRODITE to identify bias in statistical phenotyping algorithms

PRESENTER: **Juan M. Banda**

INTRO:

- The widespread adoption of machine learning (ML) algorithms for risk-stratification has resulted in documented cases of racial/ethnic biases within algorithms (1,2). When built without careful weightage and bias-proofing, ML algorithms can give recommendations which worsen health disparities faced by communities of color (3).
- Systematic differences in the output of statistical phenotyping algorithms for vulnerable populations is largely unexplored, particularly within the Observational Health Sciences and Informatics (OHDSI) community and tools.
- By leveraging APHRODITE (4,5), a probabilistic phenotyping framework, we examine four clinical conditions -- **dementia, frailty, mild cognitive impairment and Alzheimer's disease** -- common in vulnerable older adults. We aim to automate the process of identifying the presence of bias in phenotyping algorithms, by providing a standard and automatic framework for their assessment.

Our initial evaluation elucidates that the selected phenotype algorithms have performance (precision, recall, accuracy) variations anywhere between 3% to 30% across ethnic populations; even when not using ethnicity as a feature. Demonstrating how important it is to assess these models' performance for specific subgroups before deploying them in routine use

Scan for more info

A) Alzheimer's disease

	Asian	White	Black	Native American	Pacific Islander
Asian Model	0.00%	1.01%	1.78%	5.02%	2.92%
White Model	0.14%	0.00%	0.99%	5.84%	0.49%
Black Model	1.85%	1.51%	0.00%	0.50%	2.86%
Native A. Model	4.32%	4.66%	2.51%	0.00%	2.33%
Pacific I. Model	0.62%	0.25%	2.57%	8.34%	0.00%

B) Frailty

	Asian	White	Black	Native American	Pacific Islander
Asian Model	0.00%	[illegible]	[illegible]	8.52%	11.97%
White Model	[illegible]	0.00%	[illegible]	3.36%	1.57%
Black Model	4.15%	9.27%	0.00%	0.00%	10.46%
Native A. Model	[illegible]	4.41%	[illegible]	0.00%	14.29%
Pacific I. Model	[illegible]	[illegible]	5.35%	6.85%	0.00%

C) Mild Cognitive Impairment

	Asian	White	Black	Native American	Pacific Islander
Asian Model	0.00%	7.50%	0.24%	4.29%	4.73%
White Model	[illegible]	0.00%	[illegible]	6.66%	15.32%
Black Model	[illegible]	13.47%	0.00%	4.69%	16.16%
Native A. Model	4.47%	[illegible]	3.26%	0.00%	[illegible]
Pacific I. Model	2.42%	[illegible]	7.20%	[illegible]	0.00%

D) Dementia

	Asian	White	Black	Native American	Pacific Islander
Asian Model	0.00%	0.02%	0.63%	0.91%	0.71%
White Model	0.42%	0.00%	1.58%	1.26%	1.22%
Black Model	0.85%	0.74%	0.00%	3.00%	2.21%
Native A. Model	0.48%	0.07%	0.21%	0.00%	4.64%
Pacific I. Model	4.20%	4.85%	3.11%	5.94%	0.00%

Figure 1. Variation of classification accuracy for the Random Forest models across phenotypes

Methods

- We created an experimental framework, on top of APHRODITE, to explore racial/ethnic biases within a single healthcare system, Stanford Health Care, to fully evaluate the performance of such algorithms under different ethnicity distributions
- Cases and controls are matched by age/gender/race/length of record. We evaluated three different classification algorithms (LASSO, Random Forest, and Support Vector Machines). Each of the standard concepts in the OMOP CDM corresponding to racial categories are evaluated.
- We built models using a single race group and tested them against all other race groups (Figure 1). In the full study we additionally used seven different evaluations: traditional model (all data available), balanced model (per race), leave-one-out combinations. For all models we used 10-fold cross validation. Note that we removed all patients with an Unknown race from this evaluation.

This work is supported by the National Institute of Aging through Stanford University's Stanford Aging & Ethnogeriatrics Transdisciplinary Collaborative Center (SAGE) center (award 3P30AG059307-02S1)

Juan M. Banda[1], Nigam H. Shah[2], Vyjeyanthi S Periyakoil[2]
[1] Georgia State University, Atlanta, GA, USA
[2] Stanford University School of Medicine Stanford, CA, USA

Stanford MEDICINE

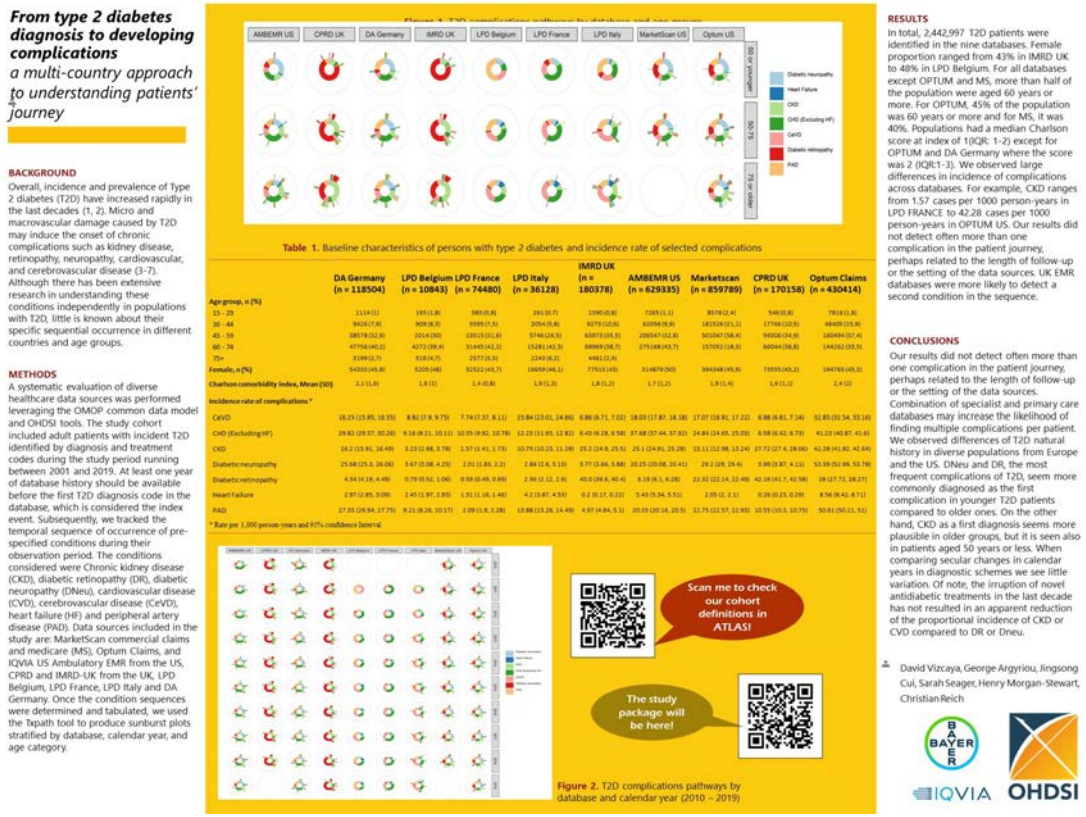
Georgia State University

OHDSI

TUESDAY Leveraging APHRODITE to identify bias in statistical phenotyping algorithms

Authors: Juan M. Banda (presenter), Nigam H. Shah, Vyjeyanthi S Periyakoil

#OHDSISocialShowcase This Week



From type 2 diabetes diagnosis to developing complications
a multi-country approach to understanding patients' journey

BACKGROUND

Overall, incidence and prevalence of Type 2 diabetes (T2D) have increased rapidly in the last decades (1, 2). Micro and macrovascular damage caused by T2D may induce the onset of chronic complications such as kidney disease, retinopathy, neuropathy, cardiovascular, and cerebrovascular disease (3-7). Although there has been extensive research in understanding these conditions independently in populations with T2D, little is known about their specific sequential occurrence in different countries and age groups.

METHODS

A systematic evaluation of diverse healthcare data sources was performed leveraging the OMOP common data model and OHDSI tools. The study cohort included adult patients with incident T2D identified by diagnosis and treatment codes during the study period running between 2001 and 2019. At least one year of database history should be available before the first T2D diagnosis code in the database, which is considered the index event. Subsequently, we tracked the temporal sequence of occurrence of pre-specified conditions during their observation period. The conditions considered were Chronic kidney disease (CKD), diabetic retinopathy (DR), diabetic neuropathy (DNeu), cardiovascular disease (CVD), cerebrovascular disease (CeVD), heart failure (HF) and peripheral artery disease (PAD). Data sources included in the study are: MarketScan commercial claims and medicare (MS), Optum Claims, and IQVIA US Ambulatory EMR from the US, CPRD and IMRD-UK from the UK, LPD Belgium, LPD France, LPD Italy and DA Germany. Once the condition sequences were determined and tabulated, we used the Txpath tool to produce sunburst plots stratified by database, calendar year, and age category.

Table 1. Baseline characteristics of persons with type 2 diabetes and incidence rate of selected complications

	DA Germany (n = 118504)	LPD Belgium (n = 10843)	LPD France (n = 74480)	LPD Italy (n = 36128)	IMRD UK (n = 180378)	AMBEMR US (n = 629335)	Marketscan (n = 859789)	CPRD UK (n = 170158)	Optum Claims (n = 430414)
Age group, n (%)									
15 - 29	1114 (1)	193 (1.8)	583 (0.8)	261 (0.7)	1590 (0.8)	7265 (1.1)	8578 (2.4)	546 (0.8)	7818 (1.8)
30 - 44	9426 (7.9)	909 (8.3)	5999 (7.5)	2094 (5.8)	9273 (5.0)	61096 (9.9)	181924 (21.1)	17746 (10.5)	48409 (11.9)
45 - 59	38578 (32.6)	3014 (28)	23515 (31.6)	5744 (24.5)	63973 (35.5)	208647 (32.8)	501047 (58.4)	59306 (34.9)	160494 (37.4)
60 - 74	47756 (40.2)	4272 (39.4)	31445 (42.2)	5281 (42.3)	69969 (38.7)	275188 (43.7)	157052 (18.3)	66044 (38.8)	144262 (33.5)
75+	3199 (2.7)	516 (4.7)	2577 (3.5)	2243 (6.2)	4482 (2.4)				
Female, n (%)	54333 (45.8)	5205 (48)	32521 (43.7)	16658 (46.1)	77519 (43)	314679 (50)	394348 (45.9)	73555 (43.2)	194769 (45.3)
Charlson comorbidity index, Mean (SD)	2.1 (1.6)	1.6 (1)	1.4 (0.8)	1.9 (1.3)	1.8 (1.2)	1.7 (1.2)	1.9 (1.4)	1.6 (1.1)	2.4 (2)
Incidence rate of complications *									
CeVD	16.25 (15.95, 16.55)	8.82 (7.9, 9.75)	7.74 (7.37, 8.11)	23.84 (23.01, 24.66)	6.86 (6.71, 7.02)	18.03 (17.87, 18.18)	17.07 (16.91, 17.22)	8.98 (8.81, 7.14)	52.85 (52.54, 53.16)
CHD (Excluding HF)	29.82 (29.57, 30.26)	9.16 (8.21, 10.11)	10.35 (9.92, 10.78)	12.23 (11.85, 12.62)	6.43 (6.28, 6.58)	37.68 (37.44, 37.92)	24.84 (24.65, 25.03)	6.58 (6.42, 6.73)	41.23 (40.87, 41.6)
CKD	16.2 (15.91, 16.48)	3.23 (2.68, 3.78)	1.57 (1.41, 1.73)	10.75 (10.23, 11.28)	25.2 (24.8, 25.5)	25.1 (24.91, 25.28)	13.11 (12.98, 13.24)	27.72 (27.4, 28.04)	42.28 (41.92, 42.64)
Diabetic neuropathy	25.68 (25.3, 26.06)	3.67 (3.08, 4.25)	2.01 (1.83, 2.2)	2.86 (2.6, 3.13)	3.77 (3.66, 3.88)	20.25 (20.08, 20.41)	29.2 (29, 29.4)	3.99 (3.87, 4.11)	53.39 (52.96, 53.74)
Diabetic retinopathy	4.54 (4.38, 4.48)	0.79 (0.52, 1.06)	0.59 (0.49, 0.69)	2.56 (2.12, 2.8)	40.0 (39.6, 40.4)	9.19 (8.1, 9.28)	22.32 (22.14, 22.48)	42.16 (41.7, 42.58)	28 (27.72, 28.27)
Heart Failure	2.97 (2.85, 3.08)	2.45 (1.97, 2.93)	1.31 (1.16, 1.46)	4.2 (3.87, 4.53)	0.2 (0.17, 0.22)	5.43 (5.34, 5.51)	2.05 (2, 2.1)	0.26 (0.23, 0.29)	8.56 (8.42, 8.71)
PAD	27.35 (26.94, 27.75)	9.21 (8.26, 10.17)	2.09 (1.9, 2.28)	13.88 (13.26, 14.49)	4.97 (4.84, 5.1)	20.39 (20.16, 20.5)	22.75 (22.57, 22.93)	10.55 (10.3, 10.75)	50.61 (50.21, 51)

* Rate per 1,000 person-years and 95% confidence interval

Figure 2. T2D complications pathways by database and calendar year (2010 – 2019)

RESULTS

In total, 2,442,997 T2D patients were identified in the nine databases. Female proportion ranged from 43% in IMRD UK to 48% in LPD Belgium. For all databases except OPTUM and MS, more than half of the population were aged 60 years or more. For OPTUM, 45% of the population was 60 years or more and for MS, it was 40%. Populations had a median Charlson score at index of 1(IQR: 1-2) except for OPTUM and DA Germany where the score was 2 (IQR:1-3). We observed large differences in incidence of complications across databases. For example, CKD ranges from 1.57 cases per 1000 person-years in LPD FRANCE to 42.28 cases per 1000 person-years in OPTUM US. Our results did not detect often more than one complication in the patient journey, perhaps related to the length of follow-up or the setting of the data sources. UK EMR databases were more likely to detect a second condition in the sequence.

CONCLUSIONS

Our results did not detect often more than one complication in the patient journey, perhaps related to the length of follow-up or the setting of the data sources. Combination of specialist and primary care databases may increase the likelihood of finding multiple complications per patient. We observed differences of T2D natural history in diverse populations from Europe and the US. DNeu and DR, the most frequent complications of T2D, seem more commonly diagnosed as the first complication in younger T2D patients compared to older ones. On the other hand, CKD as a first diagnosis seems more plausible in older groups, but it is seen also in patients aged 50 years or less. When comparing secular changes in calendar years in diagnostic schemes we see little variation. Of note, the irruption of novel antidiabetic treatments in the last decade has not resulted in an apparent reduction of the proportional incidence of CKD or CVD compared to DR or Dneu.

David Vizcaya, George Argyriou, Jingsong Cui, Sarah Seager, Henry Morgan-Stewart, Christian Reich

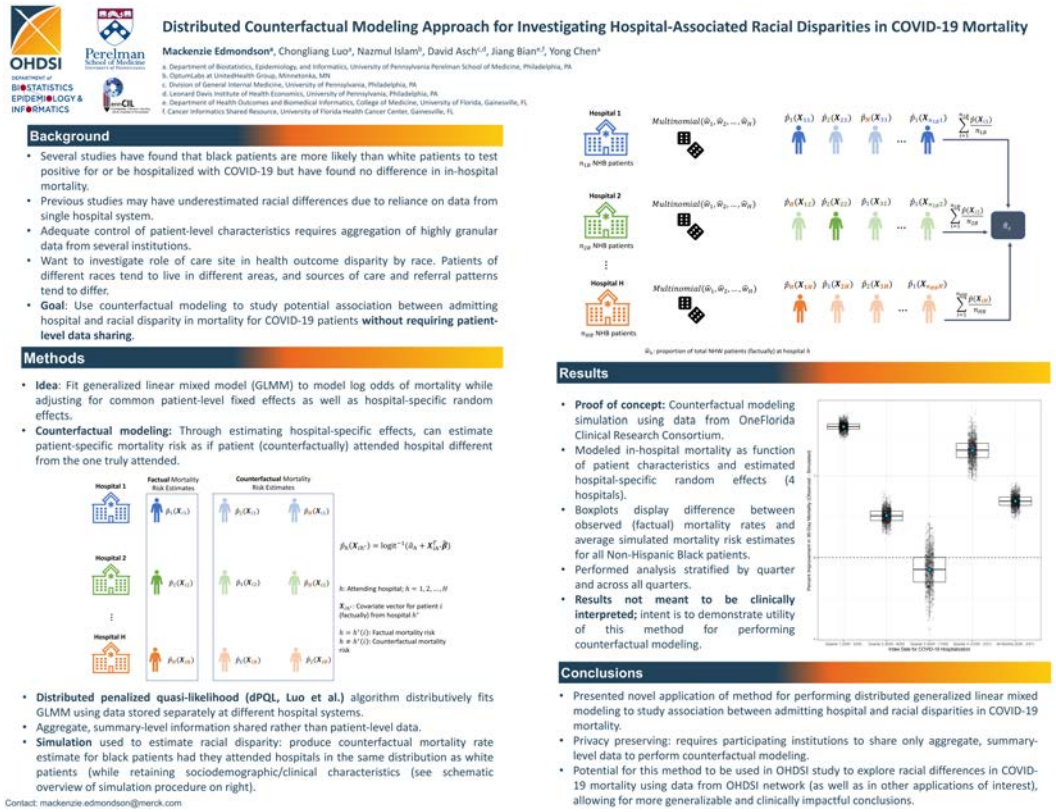
BAYER IQVIA OHDSI

WEDNESDAY

From type 2 diabetes diagnosis to developing complications: a multi-country approach to understanding patients journey

Authors: David Vizcaya, George Argyriou, Jingsong Cui, Sarah Seager, Henry Morgan-Stewart, Christian Reich

#OHDSISocialShowcase This Week



Distributed Counterfactual Modeling Approach for Investigating Hospital-Associated Racial Disparities in COVID-19 Mortality

Mackenzie Edmondson[a], Chongliang Luo[a], Nazmul Islam[b], David Asch[c,d], Jiang Bian[e,f], Yong Chen[a]

a. Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
b. OptumLabs at UnitedHealth Group, Minnetonka, MN
c. Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA
d. Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA
e. Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL
f. Cancer Informatics Shared Resource, University of Florida Health Cancer Center, Gainesville, FL

Background

- Several studies have found that black patients are more likely than white patients to test positive for or be hospitalized with COVID-19 but have found no difference in in-hospital mortality.
- Previous studies may have underestimated racial differences due to reliance on data from single hospital system.
- Adequate control of patient-level characteristics requires aggregation of highly granular data from several institutions.
- Want to investigate role of care site in health outcome disparity by race. Patients of different races tend to live in different areas, and sources of care and referral patterns tend to differ.
- **Goal**: Use counterfactual modeling to study potential association between admitting hospital and racial disparity in mortality for COVID-19 patients **without requiring patient-level data sharing**.

Methods

- **Idea**: Fit generalized linear mixed model (GLMM) to model log odds of mortality while adjusting for common patient-level fixed effects as well as hospital-specific random effects.
- **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.

- **Distributed penalized quasi-likelihood (dPQL, Luo et al.)** algorithm distributively fits GLMM using data stored separately at different hospital systems.
- Aggregate, summary-level information shared rather than patient-level data.
- **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: mackenzie.edmondson@merck.com

Results

- **Proof of concept:** Counterfactual modeling simulation using data from OneFlorida Clinical Research Consortium.
- Modeled in-hospital mortality as function of patient characteristics and estimated hospital-specific random effects (4 hospitals).
- Boxplots display difference between observed (factual) mortality rates and average simulated mortality risk estimates for all Non-Hispanic Black patients.
- Performed analysis stratified by quarter and across all quarters.
- **Results not meant to be clinically interpreted;** intent is to demonstrate utility of this method for performing counterfactual modeling.

Conclusions

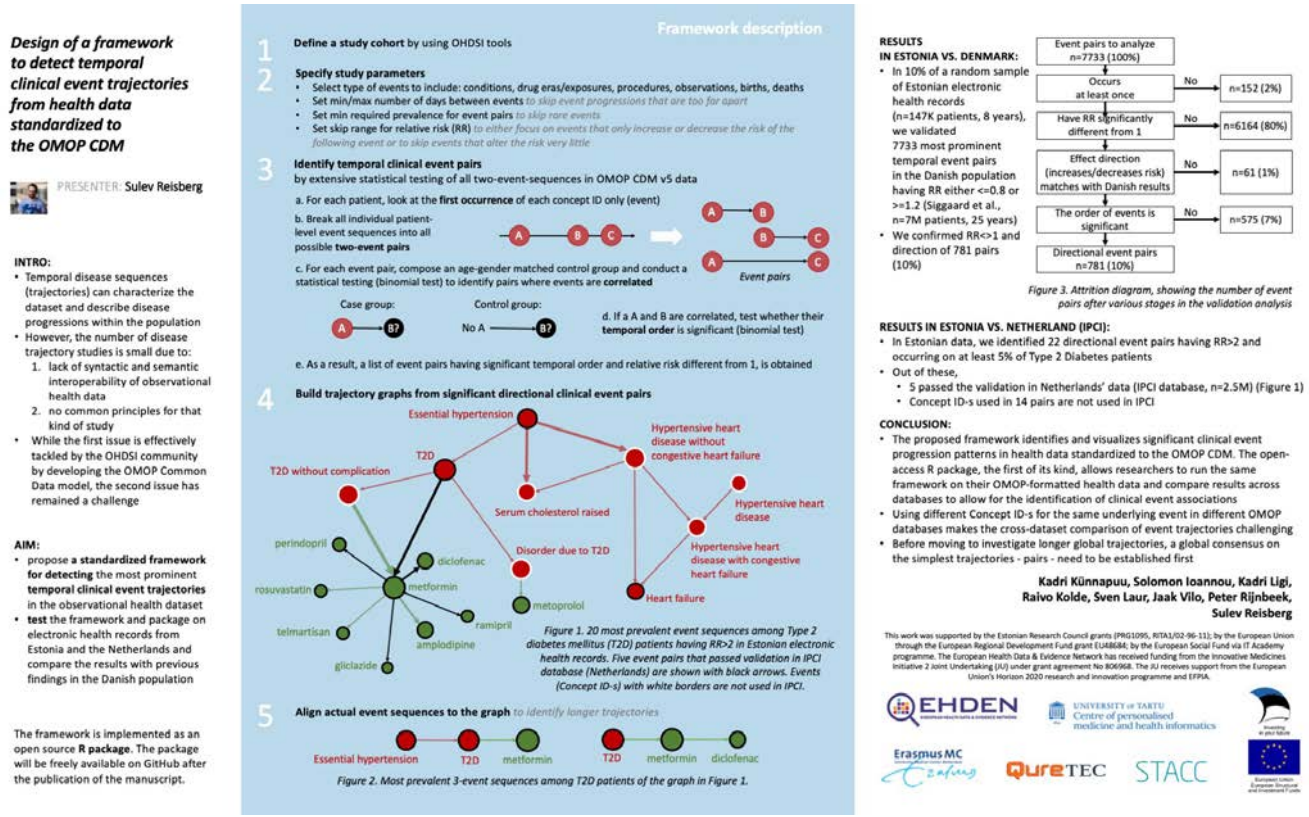
- Presented novel application of method for performing distributed generalized linear mixed modeling to study association between admitting hospital and racial disparities in COVID-19 mortality.
- Privacy preserving: requires participating institutions to share only aggregate, summary-level data to perform counterfactual modeling.
- Potential for this method to be used in OHDSI study to explore racial differences in COVID-19 mortality using data from OHDSI network (as well as in other applications of interest), allowing for more generalizable and clinically impactful conclusions.

THURSDAY

Distributed Counterfactual Modeling Approach for Investigating Hospital-Associated Racial Disparities in COVID-19 Mortality

Authors: Mackenzie Edmondson, Chongliang Luo, Nazmul Islam, David Asch, Jiang Bian, Yong Chen

#OHDSISocialShowcase This Week



Design of a framework to detect temporal clinical event trajectories from health data standardized to the OMOP CDM

PRESENTER: Sulev Reisberg

INTRO:
- Temporal disease sequences (trajectories) can characterize the dataset and describe disease progressions within the population
- However, the number of disease trajectory studies is small due to:
  1. lack of syntactic and semantic interoperability of observational health data
  2. no common principles for that kind of study
- While the first issue is effectively tackled by the OHDSI community by developing the OMOP Common Data model, the second issue has remained a challenge

AIM:
- propose **a standardized framework for detecting** the most prominent **temporal clinical event trajectories** in the observational health dataset
- **test** the framework and package on electronic health records from Estonia and the Netherlands and compare the results with previous findings in the Danish population

The framework is implemented as an open source **R package**. The package will be freely available on GitHub after the publication of the manuscript.

Framework description

1 **Define a study cohort** by using OHDSI tools

2 **Specify study parameters**
- Select type of events to include: conditions, drug eras/exposures, procedures, observations, births, deaths
- Set min/max number of days between events *to skip event progressions that are too far apart*
- Set min required prevalence for event pairs *to skip rare events*
- Set skip range for relative risk (RR) *to either focus on events that only increase or decrease the risk of the following event or to skip events that alter the risk very little*

3 **Identify temporal clinical event pairs** by extensive statistical testing of all two-event-sequences in OMOP CDM v5 data

a. For each patient, look at the **first occurrence** of each concept ID only (event)

b. Break all individual patient-level event sequences into all possible **two-event pairs**

Event pairs

c. For each event pair, compose an age-gender matched control group and conduct a statistical testing (binomial test) to identify pairs where events are **correlated**

Case group: Control group:

d. If a A and B are correlated, test whether their **temporal order** is significant (binomial test)

e. As a result, a list of event pairs having significant temporal order and relative risk different from 1, is obtained

4 **Build trajectory graphs from significant directional clinical event pairs**

Figure 1. 20 most prevalent event sequences among Type 2 diabetes mellitus (T2D) patients having RR>2 in Estonian electronic health records. Five event pairs that passed validation in IPCI database (Netherlands) are shown with black arrows. Events (Concept ID-s) with white borders are not used in IPCI.

5 **Align actual event sequences to the graph** *to identify longer trajectories*

Figure 2. Most prevalent 3-event sequences among T2D patients of the graph in Figure 1.

RESULTS IN ESTONIA VS. DENMARK:
- In 10% of a random sample of Estonian electronic health records (n=147K patients, 8 years), we validated 7733 most prominent temporal event pairs in the Danish population having RR either <=0.8 or >=1.2 (Siggaard et al., n=7M patients, 25 years)
- We confirmed RR<>1 and direction of 781 pairs (10%)

Figure 3. Attrition diagram, showing the number of event pairs after various stages in the validation analysis

RESULTS IN ESTONIA VS. NETHERLAND (IPCI):
- In Estonian data, we identified 22 directional event pairs having RR>2 and occurring on at least 5% of Type 2 Diabetes patients
- Out of these,
  - 5 passed the validation in Netherlands' data (IPCI database, n=2.5M) (Figure 1)
  - Concept ID-s used in 14 pairs are not used in IPCI

CONCLUSION:
- The proposed framework identifies and visualizes significant clinical event progression patterns in health data standardized to the OMOP CDM. The open-access R package, the first of its kind, allows researchers to run the same framework on their OMOP-formatted health data and compare results across databases to allow for the identification of clinical event associations
- Using different Concept ID-s for the same underlying event in different OMOP databases makes the cross-dataset comparison of event trajectories challenging
- Before moving to investigate longer global trajectories, a global consensus on the simplest trajectories - pairs - need to be established first

Kadri Künnapuu, Solomon Ioannou, Kadri Ligi, Raivo Kolde, Sven Laur, Jaak Vilo, Peter Rijnbeek, Sulev Reisberg

This work was supported by the Estonian Research Council grants (PRG1095, RITA1/02-96-11); by the European Union through the European Regional Development Fund grant EU48684; by the European Social Fund via IT Academy programme. The European Health Data & Evidence Network has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

EHDEN · University of Tartu Centre of personalised medicine and health informatics · Erasmus MC · QureTEC · STACC

FRIDAY

Design of a framework to detect temporal clinical event trajectories from health data standardized to the OMOP CDM

Authors: Kadri Kunnapuu, Solomon Ioannou, Kadri Ligi, Raivo Kolde, Sven Laur, Jaak Vilo, Peter Rijnbeek, Sulev Reisberg

Where Are We Going?

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

Welcome To OHDSI Newcomers

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

Please raise your hand,

and we will call on three people.

Three Stages of The Journey

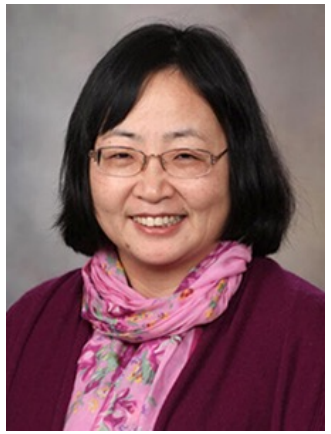
Where Have We Been?

Where Are We Now?

Where Are We Going?

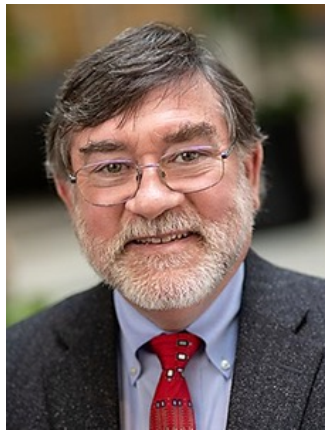
January 25 OHDSI Community Call

Extracting OHDSI Concepts from Clinical Narratives for COVID



Dr. Hongfang Liu

Professor of Biomedical Informatics, Mayo Clinic



Dr. Christopher G. Chute

Bloomberg Distinguished Professor of Health Informatics;
Professor of Medicine, Internal Medicine, Johns Hopkins University