Tools for Adoption of OHDSI Data Standards

OHDSI Community Call
Nov. 9, 2021 • 11 am ET

Remaining 2021 OHDSI Community Calls

Date	Topic
Nov. 9	Demos: Tools for Adoption of OHDSI Data Standards
Nov. 16	Open Network Studies
Nov. 23	History of OHDSI
Nov. 30	Collaborator Showcase Presentations
Dec. 7	How Did We Do This Year? Final OKR Review
Dec. 14	Holiday-Themed Final Meeting Of 2021

Remaining 2021 OHDSI Community Calls

Date	Topic
Nov. 9	Demos: Tools for Adoption of OHDSI Data Standards
Nov. 16	Open Network Studies
Nov. 23	History of OHDSI
Nov. 30	Collaborator Showcase Presentations
Dec. 7	How Did We Do This Year? Final OKR Review
Dec. 14	Holiday-Themed Final Meeting Of 2021

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to **Patricia Biedermann, Rose Ong, Alexander Davydov, Alexandra Orlova, Philip Solovyev, Hong Sun, Graham Wetherill, Monika Brand and Eva-Maria Didden** for the publication of **"Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases"** in BMC Medical Research Methodology.

Biedermann *et al. BMC Med Res Methodol* (2021) 21:238
https://doi.org/10.1186/s12874-021-01434-3

BMC Medical Research Methodology

RESEARCH Open Access

Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases

Patricia Biedermann[1], Rose Ong[1], Alexander Davydov[2], Alexandra Orlova[2], Philip Solovyev[2], Hong Sun[1], Graham Wetherill[3], Monika Brand[1] and Eva-Maria Didden[1*]

Abstract

Background: The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) can be used to transform observational health data to a common format. CDM transformation allows for analysis across disparate databases for the generation of new, real-word evidence, which is especially important in rare disease where data are limited. Pulmonary hypertension (PH) is a progressive, life-threatening disease, with rare subgroups such as pulmonary arterial hypertension (PAH), for which generating real-world evidence is challenging. Our objective is to document the process and outcomes of transforming registry data in PH to the OMOP CDM, and highlight challenges and our potential solutions.

Methods: Three observational studies were transformed from the Clinical Data Interchange Standards Consortium study data tabulation model (SDTM) to OMOP CDM format. OPUS was a prospective, multi-centre registry (2014–2020) and OrPHeUS was a retrospective, multi-centre chart review (2013–2017); both enrolled patients newly treated with macitentan in the US. EXPOSURE is a prospective, multi-centre cohort study (2017–ongoing) of patients newly treated with selexipag or any PAH-specific therapy in Europe and Canada. OMOP CDM version 5.3.1 with recent OMOP CDM vocabulary was used. Imputation rules were defined and applied for missing dates to avoid exclusion of data. Custom target concepts were introduced when existing concepts did not provide sufficient granularity.

Results: Of the 6622 patients in the three registry studies, records were mapped for 6457. Custom target concepts were introduced for PAH subgroups (by combining SNOMED concepts or creating custom concepts) and World Health Organization functional class. Per the OMOP CDM convention, records about the absence of an event, or the lack of information, were not mapped. Excluding these non-event records, 4% (OPUS), 2% (OrPHeUS) and 1% (EXPOSURE) of records were not mapped.

Conclusions: SDTM data from three registries were transformed to the OMOP CDM with limited exclusion of data and deviation from the SDTM database content. Future researchers can apply our strategy and methods in different disease areas, with tailoring as necessary. Mapping registry data to the OMOP CDM facilitates more efficient collaborations between researchers and establishment of federated data networks, which is an unmet need in rare diseases.

OHDSI Shoutouts!

The Hyve recently presented their **FAIRplus recipe** on how to make published observational research data more FAIR.

The recipe describes the work done on making data on the website of the Covid-19 EHDEN/OHDSI studyathon, held in March 2020, more findable and interoperable.

www.thehyve.nl/articles/fairify-published-observational-studies-and-databases

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

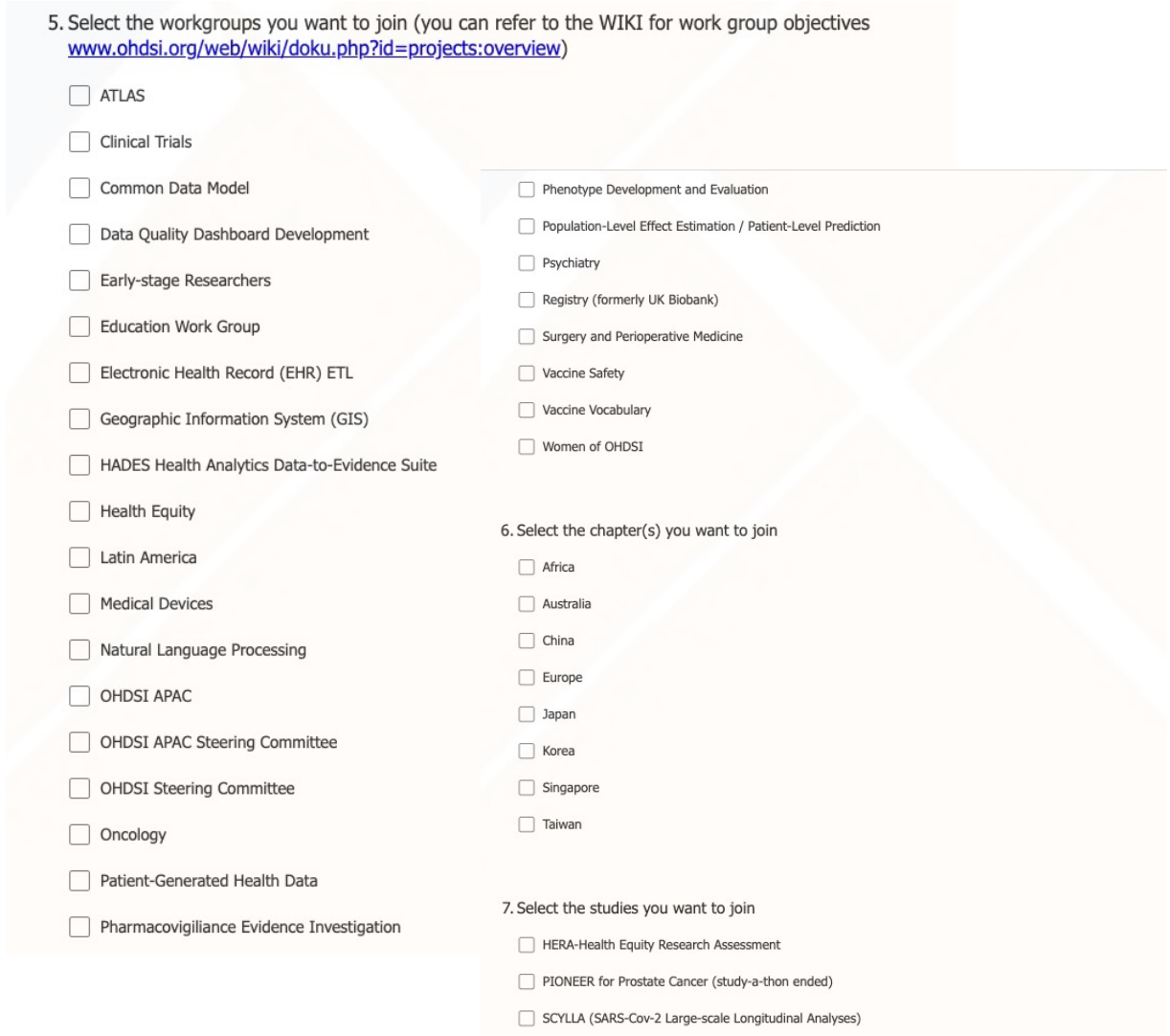
Where Are We Going?

Upcoming Workgroup Calls

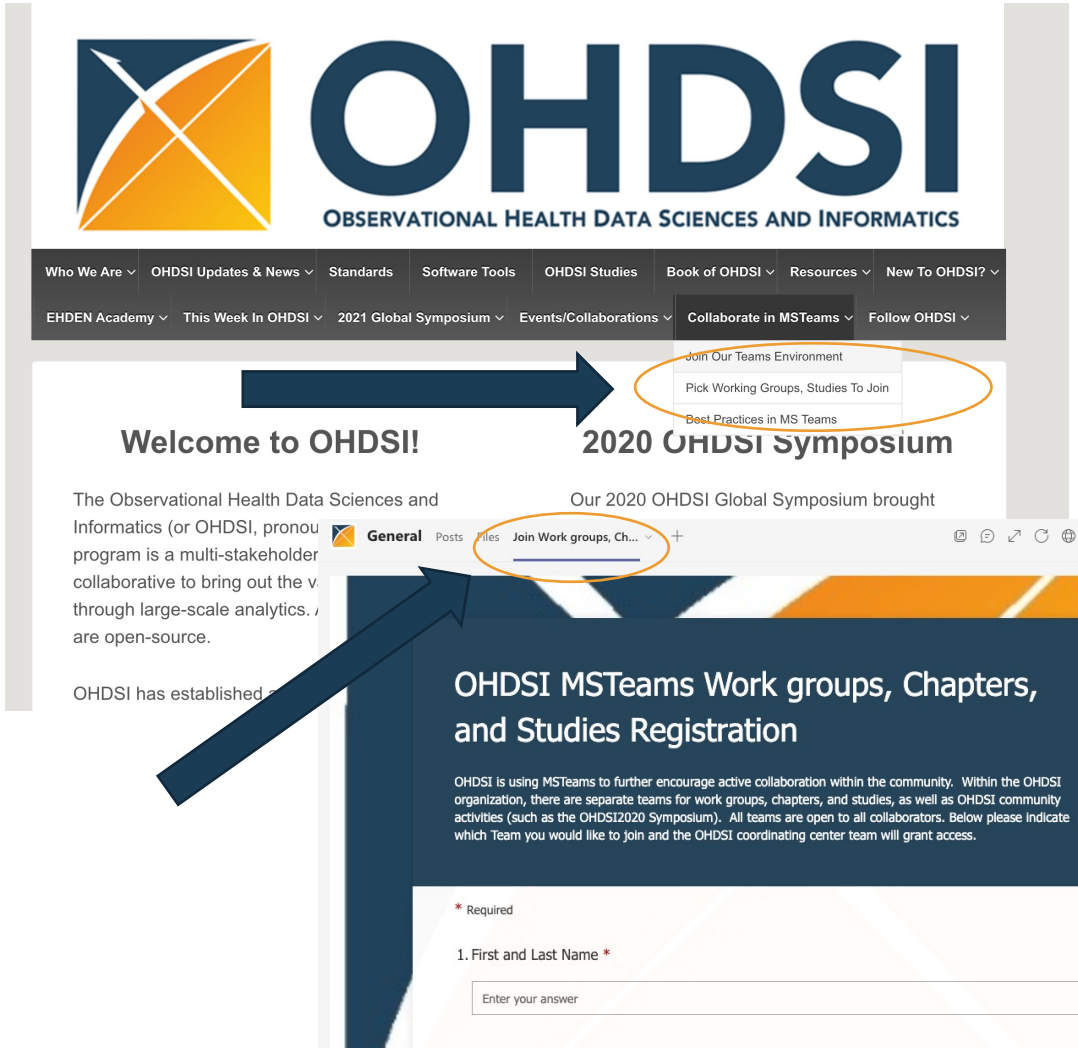
Date	Time (ET)	Meeting
Tuesday	12 pm	Common Data Model – Vocabulary Subgroup
Wednesday	10 am	FHIR and OMOP – Digital Quality Measurements Subgroup (ZOOM)
Wednesday	2 pm	Natural Language Processing
Thursday	8 am	Psychiatry
Thursday	1 pm	OMOP CDM Oncology – CDM/Vocabulary Subgroup
Friday	10 am	Electronic Health Record
Friday	10 am	Phenotype Development and Evaluation
Friday	10:30 am	Clinical Trials
Friday	11 pm	China Chapter
Monday	10 am	GIS-Geographic Information System General Meeting
Tuesday	9 am	OMOP CDM Oncology – Genomic Subgroup

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

Get Access To Different Teams/WGs/Chapters



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)

2021 APAC Symposium • Nov. 18

Nov. 18 (APAC Time Zone)	Time (Korea time)	Contents	Speaker(s)
Morning	9:00 – 9:25 am	OHDSI State of the Community	George Hripcsak/Patrick Ryan
	9:25 – 9:50 am	OHDSI APAC State of the Community	Mui Van Zandt
	9:50 – 10:00 am	Energy Break	
	10:00 – 10:25 am	EHDEN	Peter Rijnbeek
	10:25 – 10:50 am	FHIR and OHDSI Collaboration	Christian Reich
	10:50 – 11:00 am	Energy Break	
	11:00 - 12:30 pm	APAC Chapter Visions for 2022	Chapter Leads
Lunch Break	12:30 – 13:00 pm		
Afternoon (in GatherTown)	13:00 – 14:00 pm	Workgroup Sessions (Medical Image, FHIR, CDM Tables)	
	14:00 – 15:00 pm	Collaboration Showcase	
	15:00 – 16:00 pm	APAC Study Sessions	

www.ohdsi.org/apac

2021 APAC Symposium • Nov. 18

WG – Medical Imaging

Seng Chan You
Assist. Professor
Yonsei University Health System

WG – FHIR Collaboration

Christian Reich
VP IQVIA, OHDSI founder

Adam Chee
Chief of Smart Health Leadership Centre, NUS

WG – CDM v5.4

Clair Blacketer
Assoc. Director Janssen

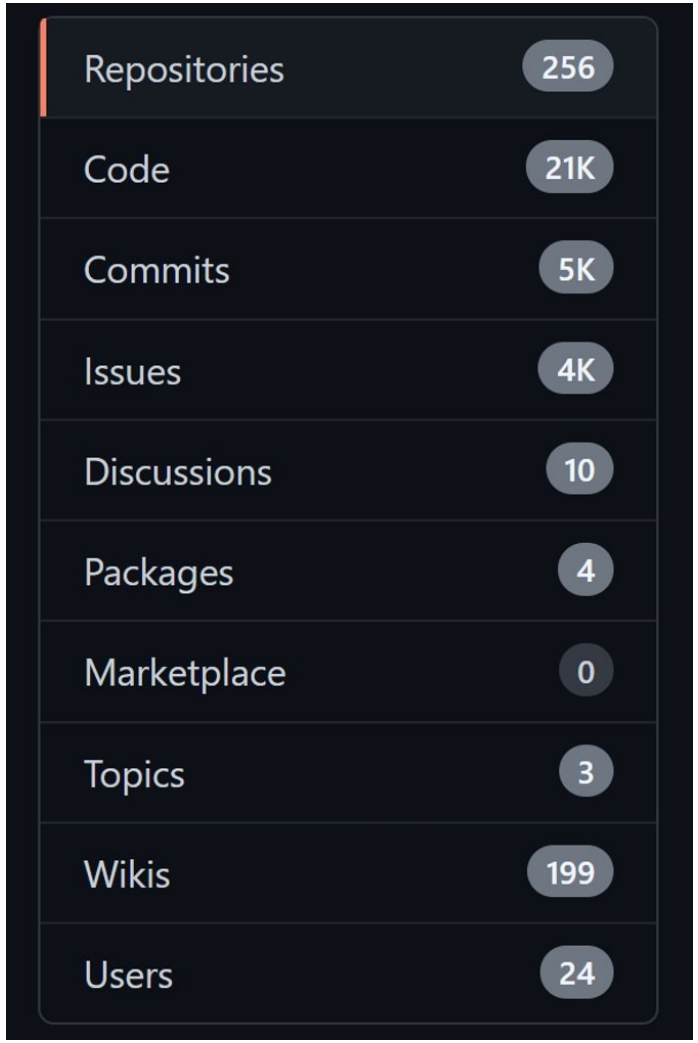
APAC Study Session

Marc Suchard
Professor, UCLA

www.ohdsi.org/apac

EHDEN Data Partner Call Concludes Nov. 15

Open-Source Governance Workshop • Nov 29, 9 am



OHDSI Open-Source Governance Workshop

Nov. 29, 2021
9 am – 1 pm

Open-Source Governance Workshop • Nov 29, 9 am

The operating code for how our organization comes together and builds things.

- Software
- Book of OHDSI
- Common data model
- Conferences
- Studyathons

OHDSI Open-Source Governance

“We want our technology stack to be easy to install, use, integrate, and contribute to.”

- Would like to encourage more contributors
- Would like to see more organizations
- Would like to reduce technical debt of support
- Would like to learn how to better recognize those that contribute through code
- We would like to increase the velocity of our development cycle

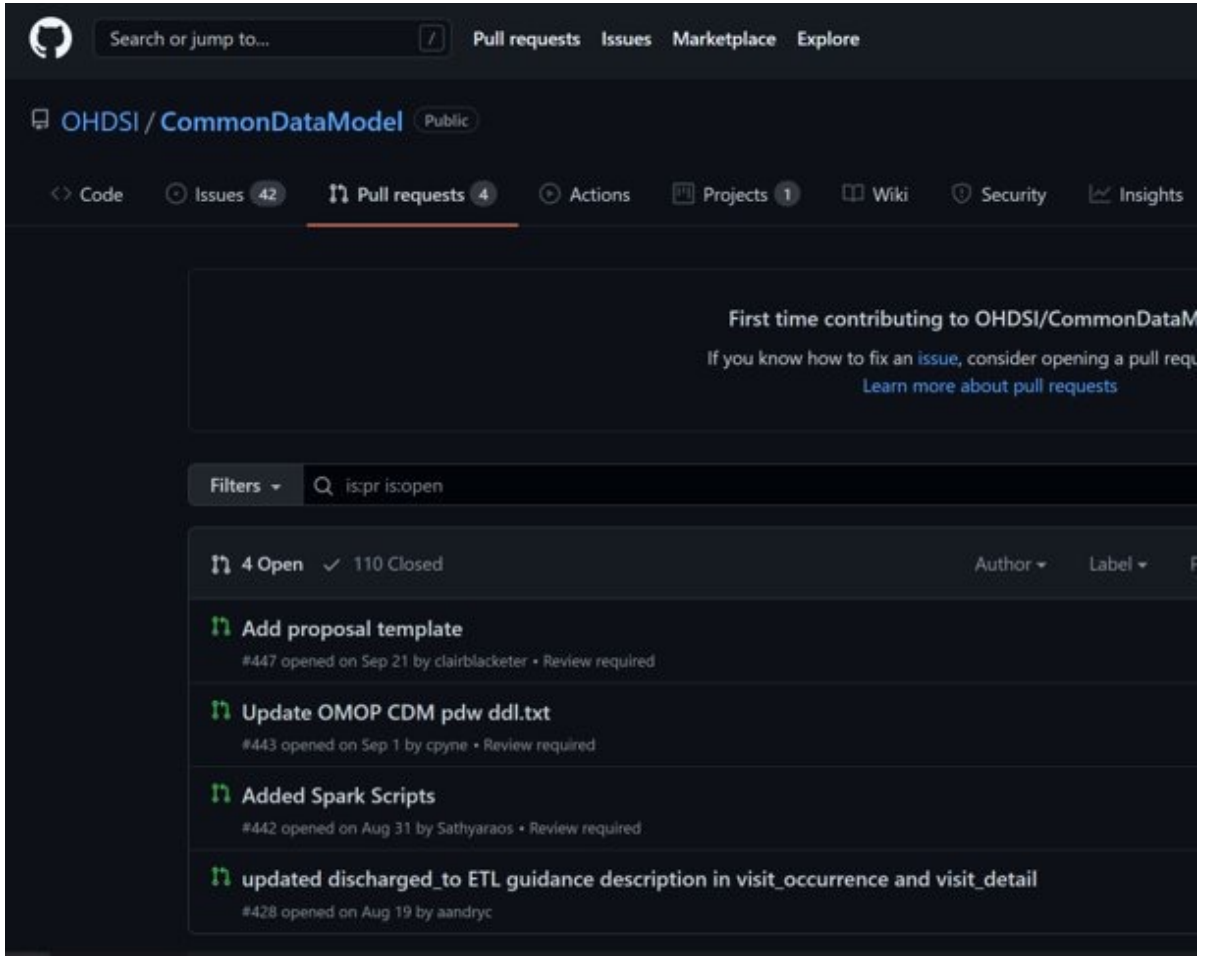
Open-Source Governance Workshop • Nov 29, 9 am

Package statuses

The table below lists, for each of the HADES packages, the following details:

- **Version**: The latest released version.
- **Maintainer(s)**: The persons responsible for the package. Only maintainers can create releases.
- **Availability**: whether the package can be installed from CRAN (using `install.packages()`), or needs to be installed from GitHub (using `remotes::install_github()`).
- **Open issues**: The number of open issues.
- **Open pull-requests**: The number of open pull-requests.
- **Build status**: Whether the package passes R Check (including unit tests) on several different operating systems, and using various database platforms. The status shown is of the the latest push, including those to develop branches. Released packages (in the master branch) by definition have successful builds, and are not shown here.
- **Coverage**: The percentage of lines of code that is covered by the unit tests in the master branch (i.e. in the latest released version).

Package	Version	Maintainer(s)	Availability	Open issues	Open pull-requests	Build status	Coverage
Andromeda	v0.5.0	Adam Black	CRAN	7	0	R-CMD-check passing	codecov 87%
BigKnn	v1.0.1	Martijn Schuemie	GitHub	1	0	R-CMD-check passing	codecov 96%
CirceR	v1.1.1	Chris Knoll	GitHub	1	0	R-CMD-check passing	codecov 87%
CohortDiagnostics	v2.1.3	Gowtham Rao	GitHub	22	1	R-CMD-check passing	codecov 82%
CohortMethod	v4.2.1	Martijn Schuemie	GitHub	6	0	R-CMD-check passing	codecov 86%
Cyclops	v3.1.2	Marc Suchard	CRAN	8	0	R-CMD-check passing	codecov 67%
DatabaseConnector	v4.0.2	Martijn Schuemie	CRAN	10	0	R-CMD-check failing	codecov 40%
EmpiricalCalibration	v3.0.0	Martijn Schuemie	CRAN	0	0	R-CMD-check passing	codecov 89%
Eunomia	v1.0.1	Frank DeFalco	CRAN	2	0	R-CMD-check failing	codecov 72%
EvidenceSynthesis	v0.2.3	Martijn Schuemie	CRAN	1	0	R-CMD-check passing	codecov 41%
FeatureExtraction	v3.1.1	Anthony Sena	GitHub	33	1	R-CMD-check passing	codecov 16%
Hydra	v0.2.0	Martijn Schuemie	GitHub	9	0	R-CMD-check passing	codecov 0%
MethodEvaluation	v2.1.0	Martijn Schuemie	GitHub	3	0	R-CMD-check passing	codecov 0%
OhdsiSharing	v0.2.2	Lee Evans	GitHub	1	0	R-CMD-check passing	codecov 0%
ParallelLogger	v2.0.2	Martijn Schuemie	CRAN	4	0	R-CMD-check passing	codecov 54%
PatientLevelPrediction	v4.3.10	Jenna Reps & Peter Rijnbeek	GitHub	33	0	R-CMD-check passing	codecov 14%
ROhdsiWebApi	v1.2.0	Gowtham Rao	GitHub	11	2	R-CMD-check passing	codecov 82%



Roles in an open-source project

- Coders,
- architects,
- reviewers,
- leaders,
- organizers,
- supporters,
- helpers,
- questioners,
- explainers,
- moderators

- Writing code
- Promotion of amazing people and projects
- Leading contributor communities
- Technical projection leadership
- Managing communication
- Updating documentation
- Answering questions form community members

The 4 Pillars of Successful Open-Source Communities (maximilianmichels.com)

Open-Source Governance Workshop • Nov 29, 9 am

Learning from large open-source communities

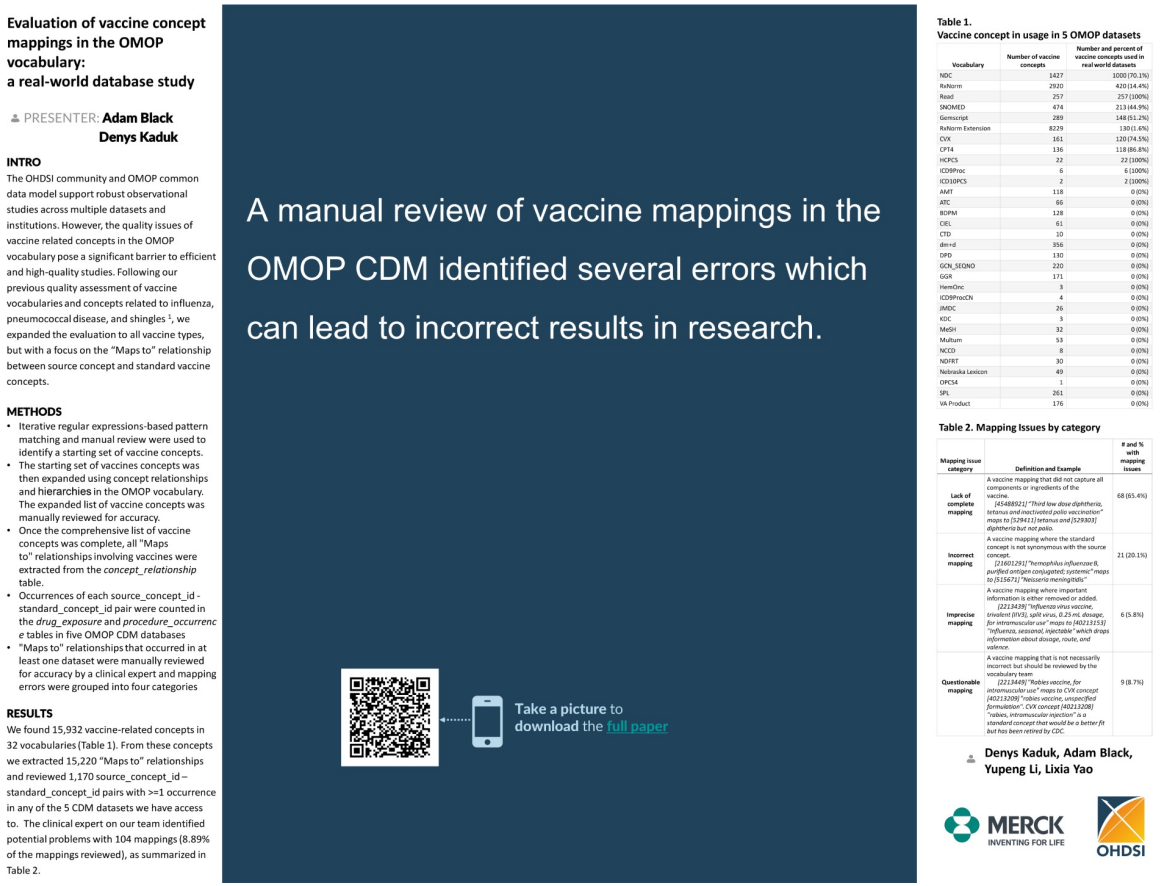
Open-Source Program Office @ JHU

Stephen Walli. Azure Office of the CTO, Microsoft

Using the Johns Hopkins OSPO as a coordination point, half-day of tutorial/workshop (virtual) and discussion for OHDSI audience

1. The software engineering economics of consumption + contribution, and of project production.
2. The nature of success (a healthy contribution flow) and building on-ramps for users, developers, and contributors.
3. The care and feeding of culture and governance over the long term.

#OHDSISocialShowcase This Week



Evaluation of vaccine concept mappings in the OMOP vocabulary: a real-world database study

PRESENTER: **Adam Black**
Denys Kaduk

INTRO

The OHDSI community and OMOP common data model support robust observational studies across multiple datasets and institutions. However, the quality issues of vaccine related concepts in the OMOP vocabulary pose a significant barrier to efficient and high-quality studies. Following our previous quality assessment of vaccine vocabularies and concepts related to influenza, pneumococcal disease, and shingles [1], we expanded the evaluation to all vaccine types, but with a focus on the "Maps to" relationship between source concept and standard vaccine concepts.

METHODS

- Iterative regular expressions-based pattern matching and manual review were used to identify a starting set of vaccine concepts.
- The starting set of vaccines concepts was then expanded using concept relationships and hierarchies in the OMOP vocabulary. The expanded list of vaccine concepts was manually reviewed for accuracy.
- Once the comprehensive list of vaccine concepts was complete, all "Maps to" relationships involving vaccines were extracted from the *concept_relationship* table.
- Occurrences of each source_concept_id - standard_concept_id pair were counted in the *drug_exposure* and *procedure_occurrence* tables in five OMOP CDM databases
- "Maps to" relationships that occurred in at least one dataset were manually reviewed for accuracy by a clinical expert and mapping errors were grouped into four categories

RESULTS

We found 15,932 vaccine-related concepts in 32 vocabularies (Table 1). From these concepts we extracted 15,220 "Maps to" relationships and reviewed 1,170 source_concept_id – standard_concept_id pairs with >=1 occurrence in any of the 5 CDM datasets we have access to. The clinical expert on our team identified potential problems with 104 mappings (8.89% of the mappings reviewed), as summarized in Table 2.

A manual review of vaccine mappings in the OMOP CDM identified several errors which can lead to incorrect results in research.

Take a picture to download the full paper

Table 1.
Vaccine concept in usage in 5 OMOP datasets

Vocabulary	Number of vaccine concepts	Number and percent of vaccine concepts used in real world datasets
NDC	1427	1000 (70.1%)
RxNorm	2920	420 (14.4%)
Read	257	257 (100%)
SNOMED	474	213 (44.9%)
Gemscript	289	148 (51.2%)
RxNorm Extension	8229	130 (1.6%)
CVX	161	120 (74.5%)
CPT4	136	118 (86.8%)
HCPCS	22	22 (100%)
ICD9Proc	6	6 (100%)
ICD10PCS	2	2 (100%)
AMT	118	0 (0%)
ATC	66	0 (0%)
BDPM	128	0 (0%)
CIEL	61	0 (0%)
CTD	10	0 (0%)
dm+d	356	0 (0%)
DPD	130	0 (0%)
GCN_SEQNO	220	0 (0%)
GGR	171	0 (0%)
HemOnc	3	0 (0%)
ICD9ProcCN	4	0 (0%)
JMDC	26	0 (0%)
KDC	3	0 (0%)
MeSH	32	0 (0%)
Multum	53	0 (0%)
NCCD	8	0 (0%)
NDFRT	30	0 (0%)
Nebraska Lexicon	49	0 (0%)
OPCS4	1	0 (0%)
SPL	261	0 (0%)
VA Product	176	0 (0%)

Table 2. Mapping Issues by category

Mapping issue category	Definition and Example	# and % with mapping issues
Lack of complete mapping	A vaccine mapping that did not capture all components or ingredients of the vaccine. *[45488921] "Third low dose diphtheria, tetanus and inactivated polio vaccination" maps to [529411] tetanus and [529303] diphtheria but not polio.*	68 (65.4%)
Incorrect mapping	A vaccine mapping where the standard concept is not synonymous with the source concept. *[2160129] "hemophilus influenzae B, purified antigen conjugated; systemic" maps to [51671] "Neisseria meningitidis"*	21 (20.1%)
Imprecise mapping	A vaccine mapping where important information is either removed or added. *[2213429] "influenza virus vaccine, trivalent (IIV3), split virus, 0.25 mL dosage, for intramuscular use" maps to [40213153] "influenza, seasonal, injectable" which drops information about dosage, route, and valence.*	6 (5.8%)
Questionable mapping	A vaccine mapping that is not necessarily incorrect but should be reviewed by the vocabulary team *[2213449] "Rabies vaccine, for intramuscular use" maps to CVX concept [40213208] "rabies vaccine, unspecified formulation". CVX concept [40213208] "rabies, intramuscular injection" is a standard concept that would be a better fit but has been retired by CDC.*	9 (8.7%)

Denys Kaduk, Adam Black, Yupeng Li, Lixia Yao

MERCK INVENTING FOR LIFE

OHDSI

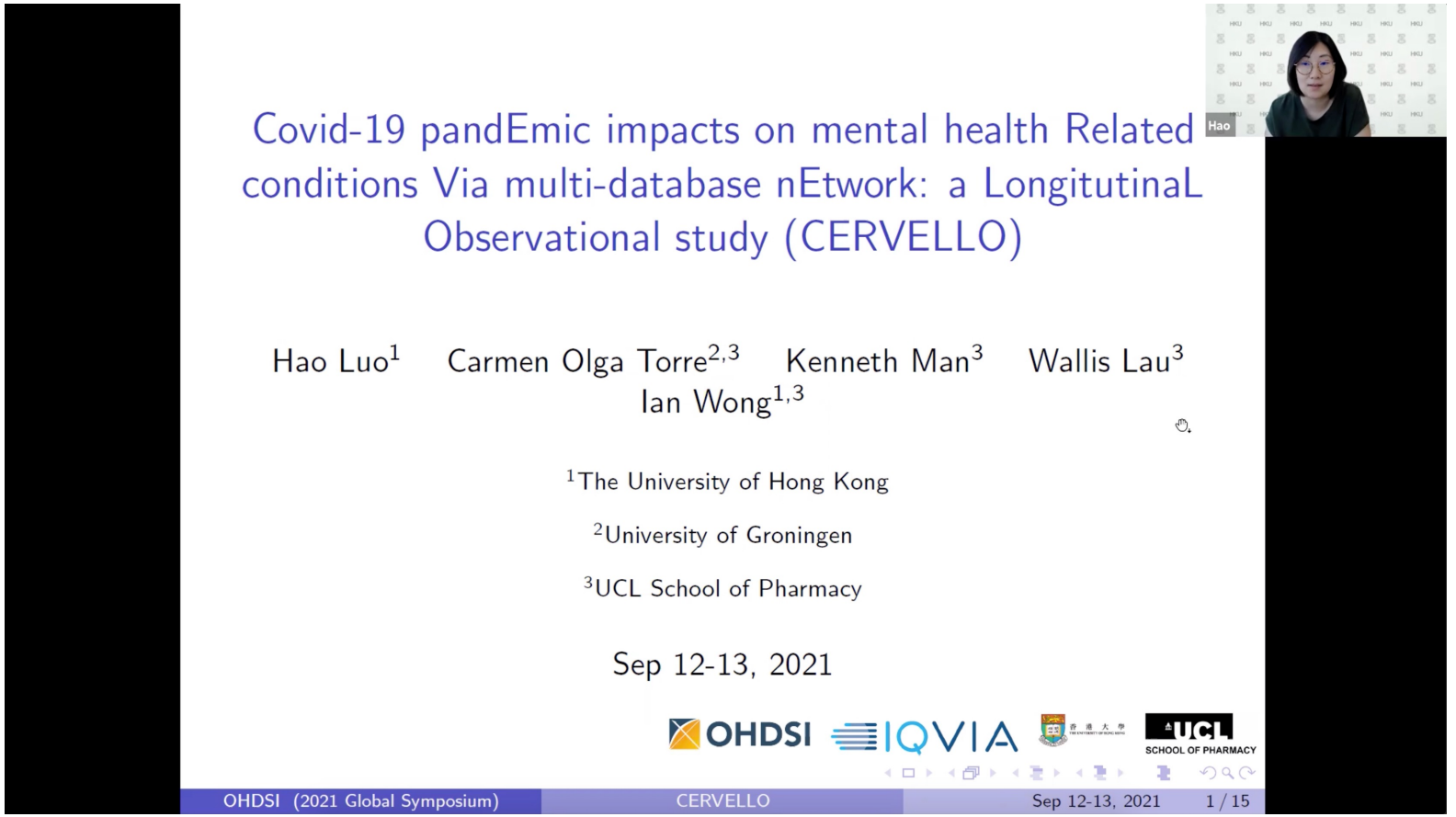
MONDAY

Evaluation of vaccine concept mappings in OMOP vocabulary: a real-world database study

Authors: Denys Kaduk, Adam Black, Yupeng Li, Lixia Yao

#OHDSISocialShowcase This Week

Lightning Talk!

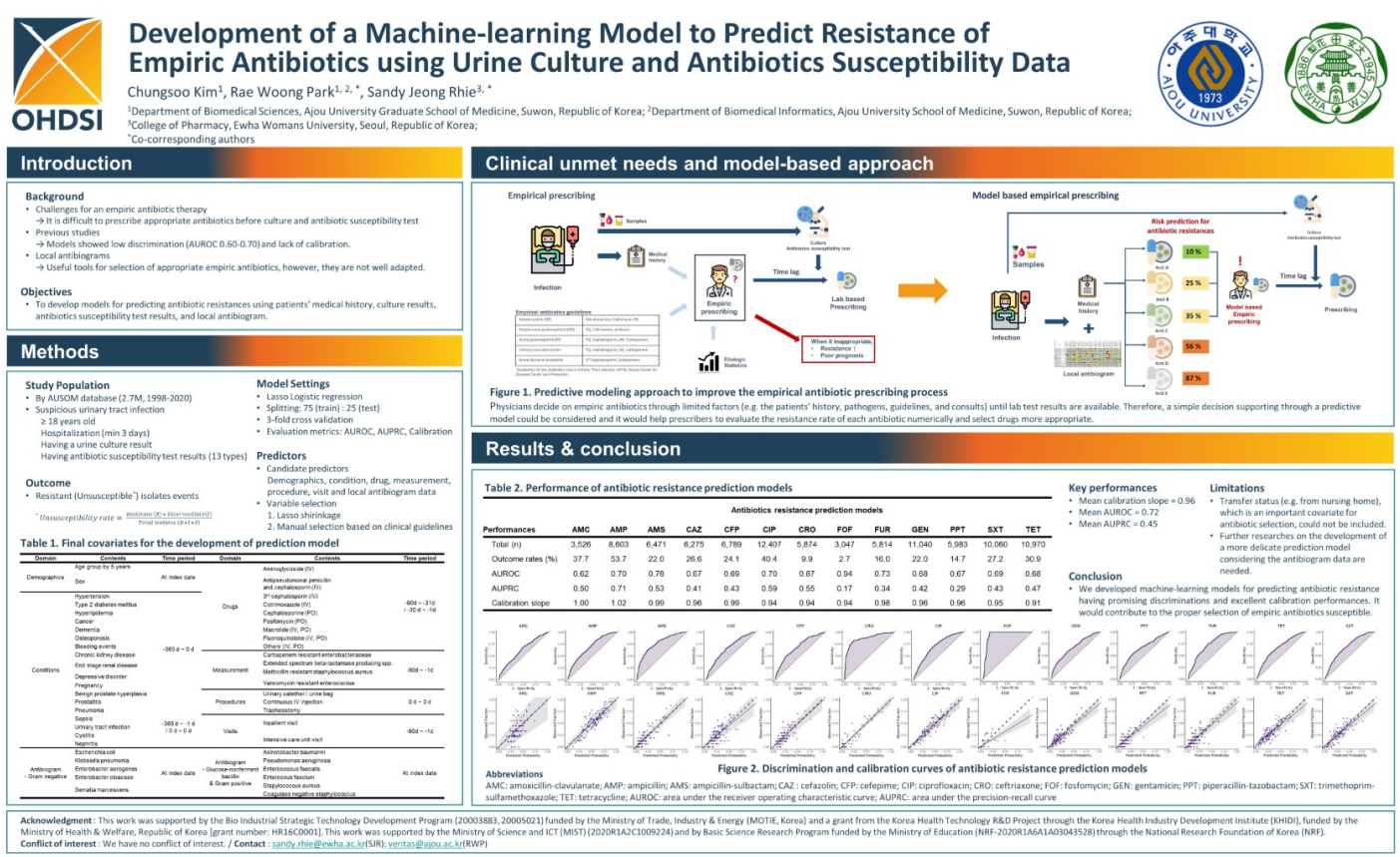


TUESDAY

Covid-19 Pandemic impacts on mental health Related conditions Via multi-database network: a Longitudinal Observational study (CERVELLO)

Authors: Carmen Olga Torre, Kenneth Man, Hao Luo (presenter), Wallis Lau, Ian Wong

#OHDSISocialShowcase This Week



Development of a Machine-learning Model to Predict Resistance of Empiric Antibiotics using Urine Culture and Antibiotics Susceptibility Data

Chungsoo Kim[1], Rae Woong Park[1, 2, *], Sandy Jeong Rhie[3, *]

[1]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Republic of Korea; [2]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea; [3]College of Pharmacy, Ewha Womans University, Seoul, Republic of Korea;

*Co-corresponding authors

Introduction

Background

- Challenges for an empiric antibiotic therapy
  → It is difficult to prescribe appropriate antibiotics before culture and antibiotic susceptibility test
- Previous studies
  → Models showed low discrimination (AUROC 0.60-0.70) and lack of calibration.
- Local antibiograms
  → Useful tools for selection of appropriate empiric antibiotics, however, they are not well adapted.

Objectives

- To develop models for predicting antibiotic resistances using patients' medical history, culture results, antibiotics susceptibility test results, and local antibiogram.

Methods

Study Population

- By AUSOM database (2.7M, 1998-2020)
- Suspicious urinary tract infection
  ≥ 18 years old
  Hospitalization (min 3 days)
  Having a urine culture result
  Having antibiotic susceptibility test results (13 types)

Outcome

- Resistant (Unsusceptible*) isolates events

$$^{*}Unsusceptibility\ rate = \frac{Resistant\ (R) + Intermediate\ (I)}{Total\ isolates\ (R+I+S)}$$

Model Settings

- Lasso Logistic regression
- Splitting: 75 (train) : 25 (test)
- 3-fold cross validation
- Evaluation metrics: AUROC, AUPRC, Calibration

Predictors

- Candidate predictors
  Demographics, condition, drug, measurement, procedure, visit and local antibiogram data
- Variable selection
  1. Lasso shrinkage
  2. Manual selection based on clinical guidelines

Table 1. Final covariates for the development of prediction model

Domain	Contents	Time period	Domain	Contents	Time period
Demographics	Age group by 5 years	At index date	Drugs	Aminoglycoside (IV)	-90d ~ -31d / -30d ~ -1d
	Sex			Antipseudomonal penicillin and cephalosporin (IV)	
Conditions	Hypertension	-365d ~ 0 d		3rd cephalosporin (IV)	
	Type 2 diabetes mellitus			Cotrimoxazole (IV)	
	Hyperlipidemia			Cephalosporine (PO)	
	Cancer			Fosfomycin (PO)	
	Dementia			Macrolide (IV, PO)	
	Osteoporosis			Fluoroquinolone (IV, PO)	
	Bleeding events			Others (IV, PO)	
	Chronic kidney disease		Measurement	Carbapenem resistant enterobacteriaceae	-90d ~ -1d
	End stage renal disease			Extended spectrum beta-lactamase producing spp.	
	Depressive disorder			Methicillin resistant staphylococcus aureus	
	Pregnancy			Vancomycin resistant enterococcus	
	Benign prostate hyperplasia		Procedures	Urinary catether / urine bag	0 d ~ 0 d
	Prostatitis			Continuous IV injection	
	Pneumonia			Tracheostomy	
	Sepsis	-365 d ~ -1 d / 0 d ~ 0 d	Visits	Inpatient visit	-90d ~ -1d
	Urinary tract infection			Intensive care unit visit	
	Cystitis				
	Nephritis				
Antibiogram - Gram negative	Escherichia coli	At index date	Antibiogram - Glucose-nonferment bacilli & Gram positive	Acinetobacter baumannii	At index date
	Klebsiella pneumonia			Pseudomonas aeruginosa	
	Enterobacter aerogenes			Enterococcus faecalis	
	Enterobacter cloacae			Enterococcus faecium	
	Serratia marcescens			Staphylococcus aureus	
				Coagulase negative staphylococcus	

Clinical unmet needs and model-based approach

Figure 1. Predictive modeling approach to improve the empirical antibiotic prescribing process

Physicians decide on empiric antibiotics through limited factors (e.g. the patients' history, pathogens, guidelines, and consults) until lab test results are available. Therefore, a simple decision supporting through a predictive model could be considered and it would help prescribers to evaluate the resistance rate of each antibiotic numerically and select drugs more appropriate.

Results & conclusion

Table 2. Performance of antibiotic resistance prediction models

Antibiotics resistance prediction models														
Performances	AMC	AMP	AMS	CAZ	CFP	CIP	CRO	FOF	FUR	GEN	PPT	SXT	TET
Total (n)	3,526	8,603	6,471	6,275	6,789	12,407	5,874	3,047	5,814	11,040	5,983	10,060	10,970
Outcome rates (%)	37.7	53.7	22.0	26.6	24.1	40.4	9.9	2.7	16.0	22.0	14.7	27.2	30.9
AUROC	0.62	0.70	0.78	0.67	0.69	0.70	0.67	0.94	0.73	0.68	0.67	0.69	0.68
AUPRC	0.50	0.71	0.53	0.41	0.43	0.59	0.55	0.17	0.34	0.42	0.29	0.43	0.47
Calibration slope	1.00	1.02	0.99	0.96	0.99	0.94	0.94	0.94	0.98	0.96	0.96	0.95	0.91

Key performances

- Mean calibration slope = 0.96
- Mean AUROC = 0.72
- Mean AUPRC = 0.45

Limitations

- Transfer status (e.g. from nursing home), which is an important covariate for antibiotic selection, could not be included.
- Further researches on the development of a more delicate prediction model considering the antibiogram data are needed.

Conclusion

- We developed machine-learning models for predicting antibiotic resistance having promising discriminations and excellent calibration performances. It would contribute to the proper selection of empiric antibiotics susceptible.

Figure 2. Discrimination and calibration curves of antibiotic resistance prediction models

Abbreviations

AMC: amoxicillin-clavulanate; AMP: ampicillin; AMS: ampicillin-sulbactam; CAZ : cefazolin; CFP: cefepime; CIP: ciprofloxacin; CRO: ceftriaxone; FOF: fosfomycin; GEN: gentamicin; PPT: piperacillin-tazobactam; SXT: trimethoprim-sulfamethoxazole; TET: tetracycline; AUROC: area under the receiver operating characteristic curve; AUPRC: area under the precision-recall curve

Acknowledgment : This work was supported by the Bio Industrial Strategic Technology Development Program (20003883, 20005021) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea [grant number: HR16C0001]. This work was supported by the Ministry of Science and ICT (MIST) (2020R1A2C1009224) and by Basic Science Research Program funded by the Ministry of Education (NRF-2020R1A6A1A03043528) through the National Research Foundation of Korea (NRF).

Conflict of interest : We have no conflict of interest. / **Contact** : sandy.rhie@ewha.ac.kr(SJR); veritas@ajou.ac.kr(RWP)

WEDNESDAY

Development of a Machine-learning Model to Predict Antibiotic Resistance using Urine Culture and Antibiotic Susceptibility Data

Authors: Chungsoo Kim, Sandy Jeong Rhie, Rae Woong Park

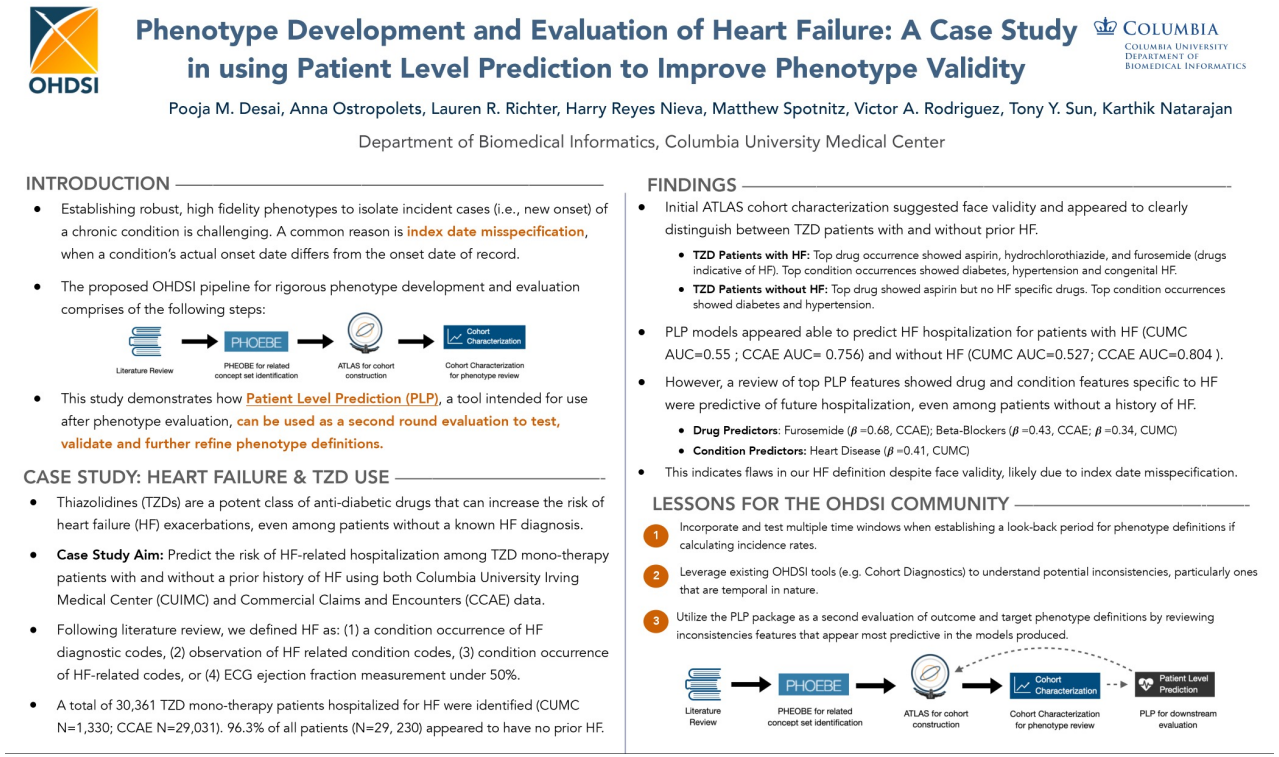
#OHDSISocialShowcase This Week

THURSDAY

ARES: A Research Exploration System

Authors: Frank J DeFalco, Alan Andryc, Anthony Molinaro, Clair Blacketer

#OHDSISocialShowcase This Week



Phenotype Development and Evaluation of Heart Failure: A Case Study in using Patient Level Prediction to Improve Phenotype Validity

Pooja M. Desai, Anna Ostropolets, Lauren R. Richter, Harry Reyes Nieva, Matthew Spotnitz, Victor A. Rodriguez, Tony Y. Sun, Karthik Natarajan

Department of Biomedical Informatics, Columbia University Medical Center

INTRODUCTION

- Establishing robust, high fidelity phenotypes to isolate incident cases (i.e., new onset) of a chronic condition is challenging. A common reason is **index date misspecification**, when a condition's actual onset date differs from the onset date of record.
- The proposed OHDSI pipeline for rigorous phenotype development and evaluation comprises of the following steps:

- This study demonstrates how Patient Level Prediction (PLP), a tool intended for use after phenotype evaluation, **can be used as a second round evaluation to test, validate and further refine phenotype definitions.**

CASE STUDY: HEART FAILURE & TZD USE

- Thiazolidines (TZDs) are a potent class of anti-diabetic drugs that can increase the risk of heart failure (HF) exacerbations, even among patients without a known HF diagnosis.
- **Case Study Aim:** Predict the risk of HF-related hospitalization among TZD mono-therapy patients with and without a prior history of HF using both Columbia University Irving Medical Center (CUIMC) and Commercial Claims and Encounters (CCAE) data.
- Following literature review, we defined HF as: (1) a condition occurrence of HF diagnostic codes, (2) observation of HF related condition codes, (3) condition occurrence of HF-related codes, or (4) ECG ejection fraction measurement under 50%.
- A total of 30,361 TZD mono-therapy patients hospitalized for HF were identified (CUMC N=1,330; CCAE N=29,031). 96.3% of all patients (N=29, 230) appeared to have no prior HF.

FINDINGS

- Initial ATLAS cohort characterization suggested face validity and appeared to clearly distinguish between TZD patients with and without prior HF.
  - **TZD Patients with HF:** Top drug occurrence showed aspirin, hydrochlorothiazide, and furosemide (drugs indicative of HF). Top condition occurrences showed diabetes, hypertension and congenital HF.
  - **TZD Patients without HF:** Top drug showed aspirin but no HF specific drugs. Top condition occurrences showed diabetes and hypertension.
- PLP models appeared able to predict HF hospitalization for patients with HF (CUMC AUC=0.55 ; CCAE AUC= 0.756) and without HF (CUMC AUC=0.527; CCAE AUC=0.804).
- However, a review of top PLP features showed drug and condition features specific to HF were predictive of future hospitalization, even among patients without a history of HF.
  - **Drug Predictors**: Furosemide (β =0.68, CCAE); Beta-Blockers (β =0.43, CCAE; β =0.34, CUMC)
  - **Condition Predictors:** Heart Disease (β =0.41, CUMC)
- This indicates flaws in our HF definition despite face validity, likely due to index date misspecification.

LESSONS FOR THE OHDSI COMMUNITY

1. Incorporate and test multiple time windows when establishing a look-back period for phenotype definitions if calculating incidence rates.
2. Leverage existing OHDSI tools (e.g. Cohort Diagnostics) to understand potential inconsistencies, particularly ones that are temporal in nature.
3. Utilize the PLP package as a second evaluation of outcome and target phenotype definitions by reviewing inconsistencies features that appear most predictive in the models produced.

FRIDAY

Phenotype Development and Evaluation of Heart Failure: A Case Study in using Patient Level Prediction to Improve Phenotype Validity

Authors: Pooja M. Desai, Anna Ostropolets, Lauren R. Richter, Harry Reyes Nieva, Matthew Spotnitz, Victor A. Rodriguez, Tony Y. Sun, Karthik Natarajan

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?

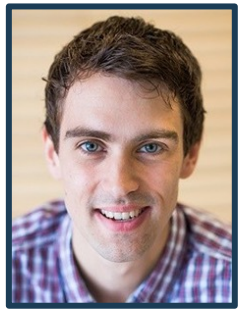
Nov. 9: Tools For Adoption Of OHDSI Data Standards

Clair Blacketer

Introduction To CDM 5.4

Michael Kallfelz

OHDSI Vocabularies



Maxim Moinat

ETL Inspection Report