Network Studies

(and possible a software demo!)

OHDSI Community Call

Nov. 15, 2022 • 11 am ET

Upcoming OHDSI Community Calls

Date	Topic
Nov. 22	10-Minute Tutorials
Nov. 29	Workgroup Updates
Dec. 6	Fall Publications
Dec. 13	How Did We Do In 2022?
Dec. 20	Holiday-Themed Final Call of 2022

Three Stages of The Journey

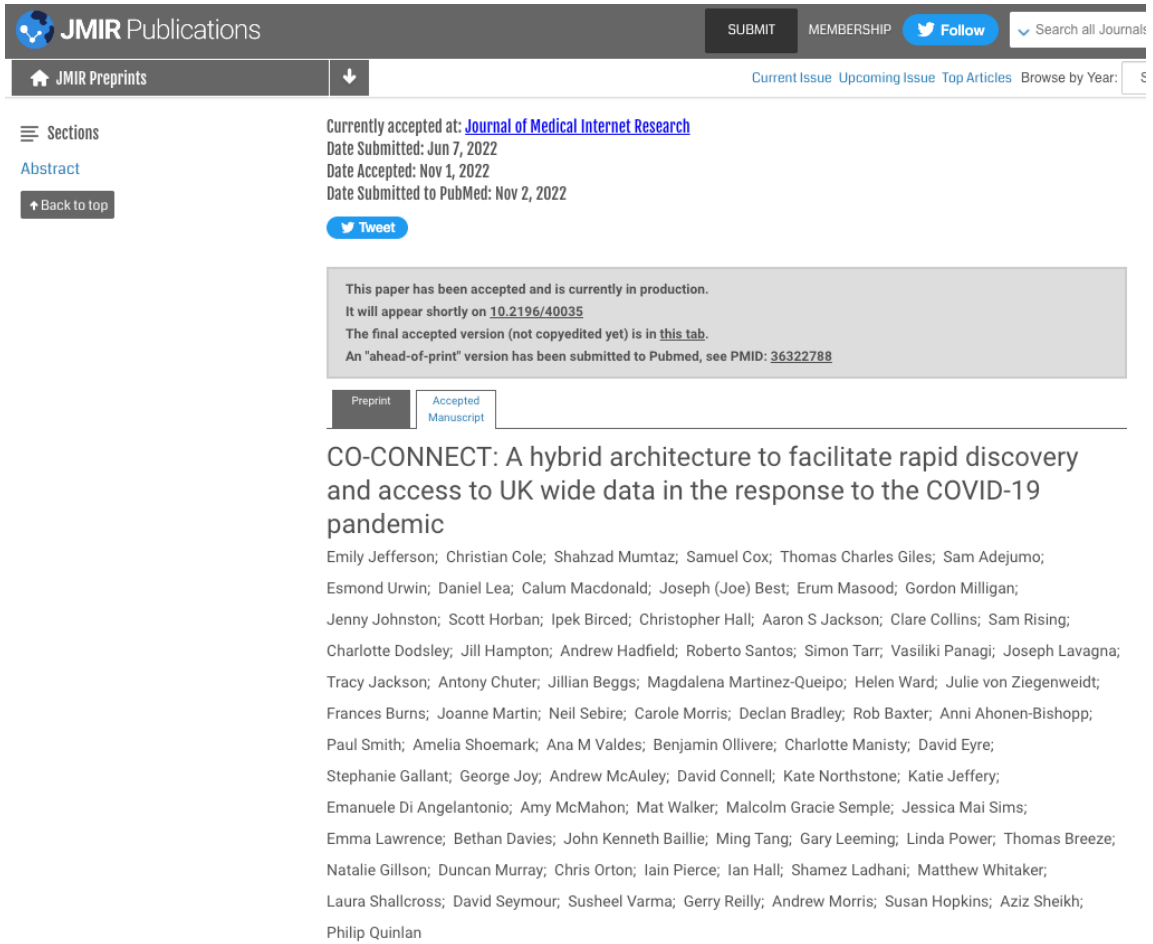
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of Emily Jefferson, Christian Cole, Shahzad Mumtaz, Sam Cox, Tom Giles, Samuel Adejumo, Esmond Urwin, Daniel Lea, Calum Macdonald, Joseph Best, Erum Masood, Gordon Milligan, Jenny Johnston, Scott Horban, Ipek Birced, Christopher Hall, Aaron S Jackson, Clare Collins, Sam Rising, Charlotte Dodsley, Jill Hampton, Andrew Hadfield, Roberto Santos, Simon Tarr, Vasiliki Panagi, Joseph Lavagna, Tracy Jackson, Antony Chuter, Jillian Beggs, Magdalena Martinez-Queipo, Helen Ward, Julie von Ziegenweidt, Frances Burns, Joanne Martin, Neil Sebire, Carole Morris, Declan Bradley, Rob Baxter, Anni Ahonen-Bishopp, Amelia Shoemark, Ana M Valdes, Benjamin Ollivere, Charlotte Manisty, David Eyre, Stephanie Gallant, George Joy, Andrew McAuley, David W Connell, Kate Northstone, Katie Jeffery, Emanuele Di Angelantonio, Amy McMahon, Mat Walker, Malcolm Gracie Semple, Jessica Mai Sims, Emma Lawrence, Bethan Davies, John Kenneth Baillie, Ming Tang, Gary Leeming, Linda Power, Thomas Breeze, Natalie Gilson, Paul Smith, Duncan Murray, Chris Orton, Iain Pierce, Ian Hall, Shamez Ladhani, Matthew Whitaker, Laura Shallcross, David Seymour, Susheel Varma, Gerry Reilly, Andrew Morris, Susan Hopkins, Aziz Sheikh, and Philip Quinlan on the publication of **CO-CONNECT: A hybrid architecture to facilitate rapid discovery and access to UK wide data in the response to the COVID-19 pandemic** in the Journal of Medical Internet Research.



JMIR Publications

SUBMIT MEMBERSHIP Follow Search all Journals

JMIR Preprints

Current Issue Upcoming Issue Top Articles Browse by Year:

Sections

Abstract

Back to top

Currently accepted at: Journal of Medical Internet Research
Date Submitted: Jun 7, 2022
Date Accepted: Nov 1, 2022
Date Submitted to PubMed: Nov 2, 2022

Tweet

This paper has been accepted and is currently in production.
It will appear shortly on 10.2196/40035
The final accepted version (not copyedited yet) is in this tab.
An "ahead-of-print" version has been submitted to Pubmed, see PMID: 36322788

Preprint | Accepted Manuscript

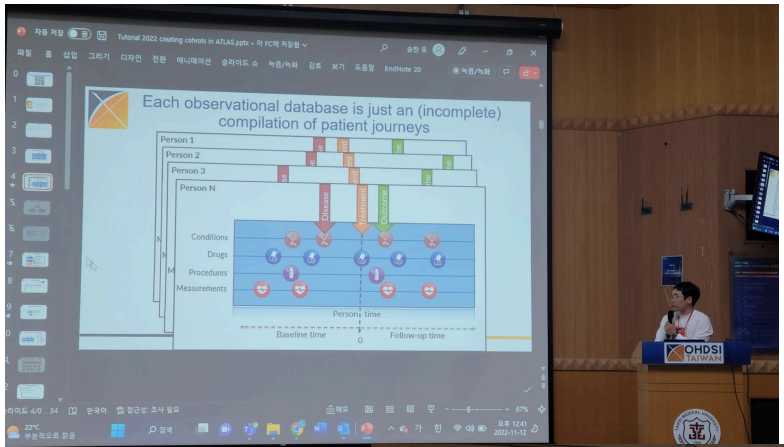
CO-CONNECT: A hybrid architecture to facilitate rapid discovery and access to UK wide data in the response to the COVID-19 pandemic

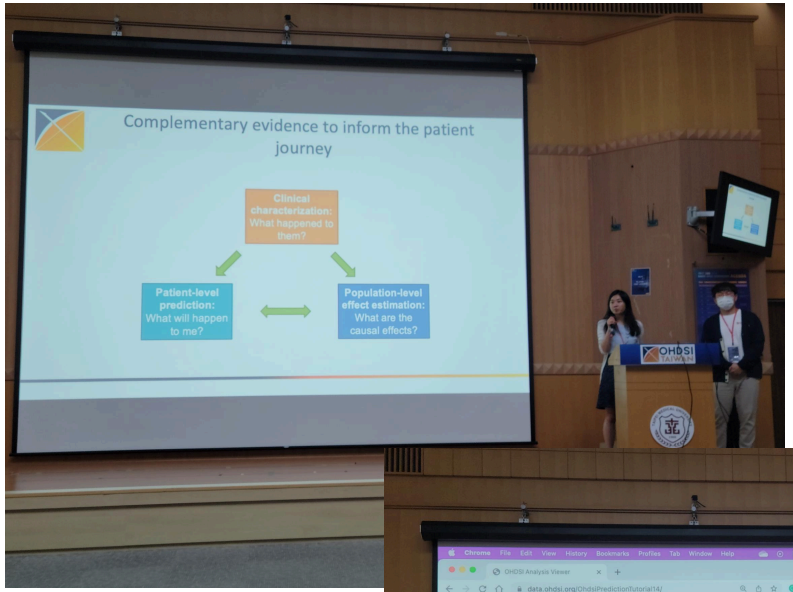
Emily Jefferson; Christian Cole; Shahzad Mumtaz; Samuel Cox; Thomas Charles Giles; Sam Adejumo; Esmond Urwin; Daniel Lea; Calum Macdonald; Joseph (Joe) Best; Erum Masood; Gordon Milligan; Jenny Johnston; Scott Horban; Ipek Birced; Christopher Hall; Aaron S Jackson; Clare Collins; Sam Rising; Charlotte Dodsley; Jill Hampton; Andrew Hadfield; Roberto Santos; Simon Tarr; Vasiliki Panagi; Joseph Lavagna; Tracy Jackson; Antony Chuter; Jillian Beggs; Magdalena Martinez-Queipo; Helen Ward; Julie von Ziegenweidt; Frances Burns; Joanne Martin; Neil Sebire; Carole Morris; Declan Bradley; Rob Baxter; Anni Ahonen-Bishopp; Paul Smith; Amelia Shoemark; Ana M Valdes; Benjamin Ollivere; Charlotte Manisty; David Eyre; Stephanie Gallant; George Joy; Andrew McAuley; David Connell; Kate Northstone; Katie Jeffery; Emanuele Di Angelantonio; Amy McMahon; Mat Walker; Malcolm Gracie Semple; Jessica Mai Sims; Emma Lawrence; Bethan Davies; John Kenneth Baillie; Ming Tang; Gary Leeming; Linda Power; Thomas Breeze; Natalie Gillson; Duncan Murray; Chris Orton; Iain Pierce; Ian Hall; Shamez Ladhani; Matthew Whitaker; Laura Shallcross; David Seymour; Susheel Varma; Gerry Reilly; Andrew Morris; Susan Hopkins; Aziz Sheikh; Philip Quinlan

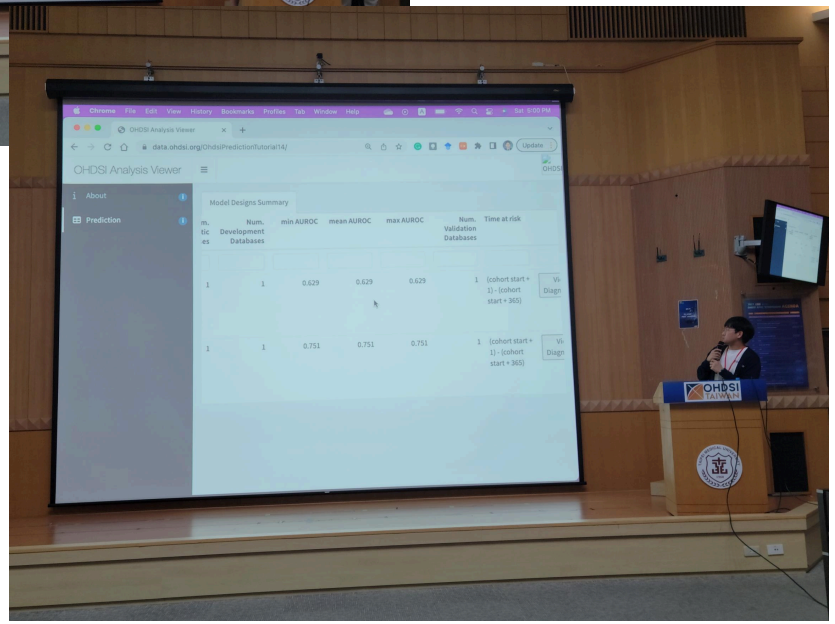
2022 Asia-Pacific (APAC) Symposium

2022 Asia-Pacific (APAC) Symposium

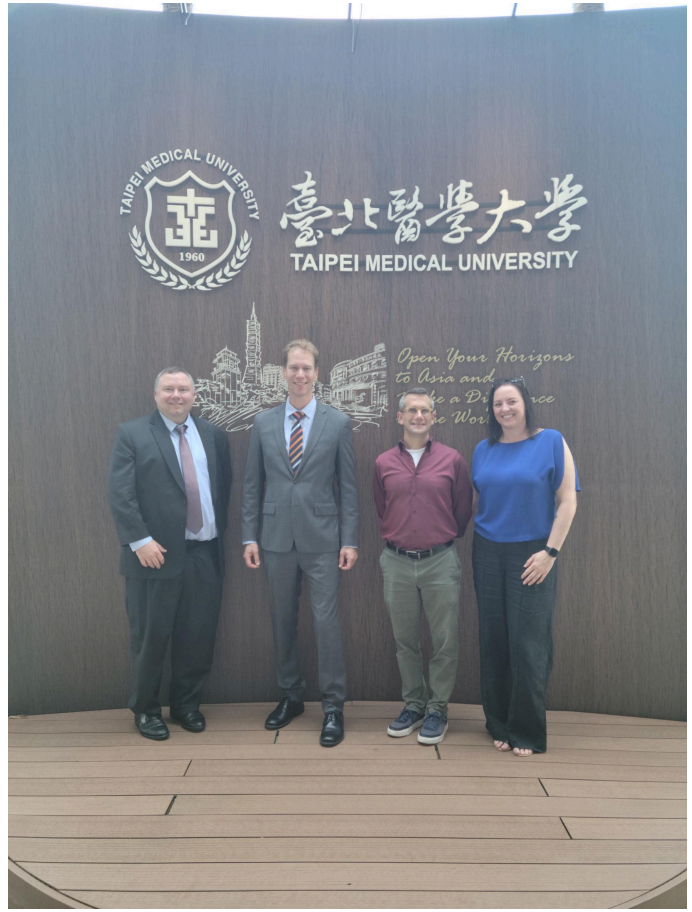
2022 Asia-Pacific (APAC) Symposium







2022 Asia-Pacific (APAC) Symposium



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	1 pm	Common Data Model
Wednesday	9 am	FHIR and OMOP Data Model Harmonization Subgroup (ZOOM)
Wednesday	12 pm	Health Equity Journal Club
Wednesday	2 pm	FHIR and OMOP Terminologies Subgroup (ZOOM)
Wednesday	7 pm	Medical Imaging
Thursday	10 am	Data Quality Dashboard Development
Thursday	12 pm	HADES
Thursday	1 pm	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	7 pm	Dentistry
Friday	9 am	Geographic Information System (GIS)
Monday	10 am	Healthcare Special Interest Group
Tuesday	9 am	OMOP CDM Oncology

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

Open-Source Community WG Meeting

Please join the Nov. 23 Open-Source Community WG meeting, which will include a presentation from 2018 Titan Award winner **Lee Evans** focused on **OHDSI open source software: continuous integration, automated testing, & test database infrastructure**.

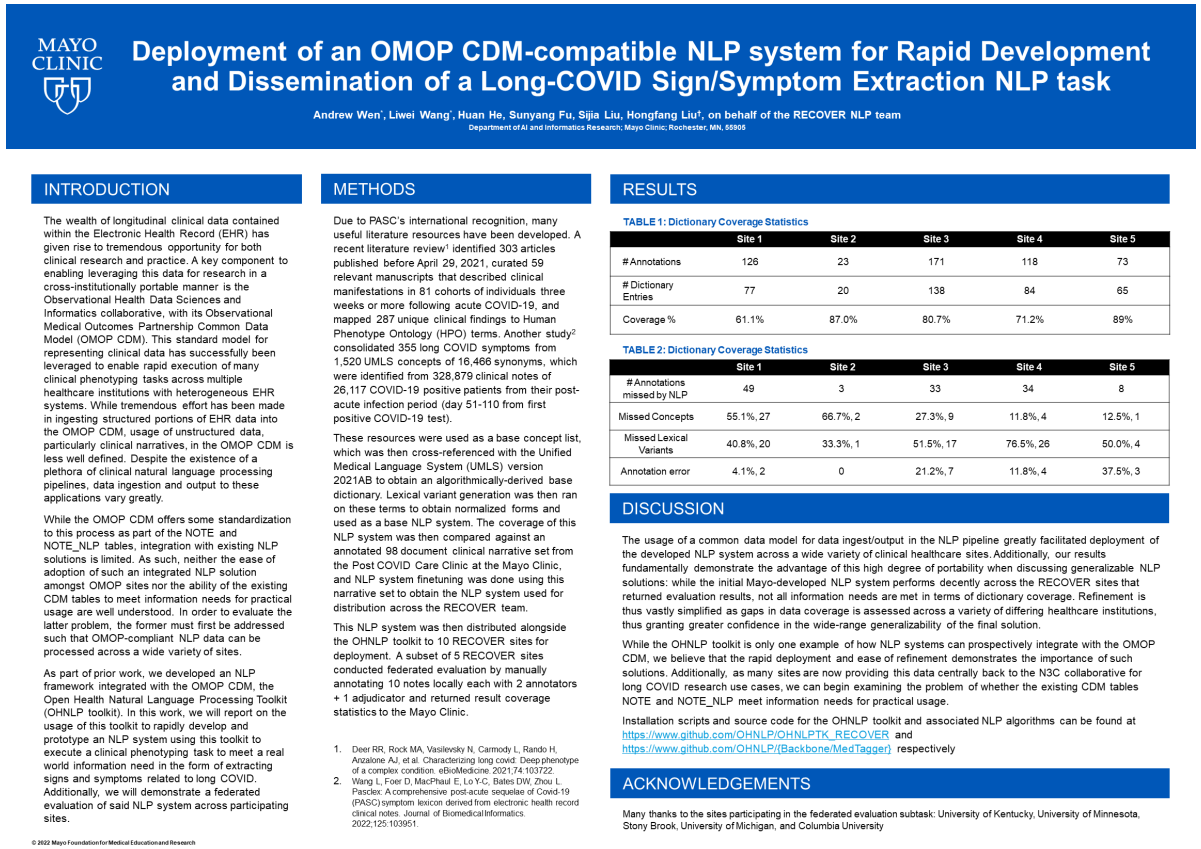
Wednesday, Nov. 23 , 11 am ET

Join Anna Ostropolets' Dissertation Defense

OHDSI veteran and 2018 Titan Award winner **Anna Ostropolets** will defend her Columbia University dissertation Wed., Nov. 30, on **Generating Reliable and Responsive Observational Evidence: Reducing Pre-analysis Bias**. The open session will be at 10 am ET on Zoom.

Wednesday, Nov. 30, 10 am ET

#OHDSISocialShowcase This Week



MAYO CLINIC

Deployment of an OMOP CDM-compatible NLP system for Rapid Development and Dissemination of a Long-COVID Sign/Symptom Extraction NLP task

Andrew Wen*, Liwei Wang*, Huan He, Sunyang Fu, Sijia Liu, Hongfang Liu†, on behalf of the RECOVER NLP team
Department of AI and Informatics Research; Mayo Clinic; Rochester, MN, 55905

INTRODUCTION

The wealth of longitudinal clinical data contained within the Electronic Health Record (EHR) has given rise to tremendous opportunity for both clinical research and practice. A key component to enabling leveraging this data for research in a cross-institutionally portable manner is the Observational Health Data Sciences and Informatics collaborative, with its Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). This standard model for representing clinical data has successfully been leveraged to enable rapid execution of many clinical phenotyping tasks across multiple healthcare institutions with heterogeneous EHR systems. While tremendous effort has been made in ingesting structured portions of EHR data into the OMOP CDM, usage of unstructured data, particularly clinical narratives, in the OMOP CDM is less well defined. Despite the existence of a plethora of clinical natural language processing pipelines, data ingestion and output to these applications vary greatly.

While the OMOP CDM offers some standardization to this process as part of the NOTE and NOTE_NLP tables, integration with existing NLP solutions is limited. As such, neither the ease of adoption of such an integrated NLP solution amongst OMOP sites nor the ability of the existing CDM tables to meet information needs for practical usage are well understood. In order to evaluate the latter problem, the former must first be addressed such that OMOP-compliant NLP data can be processed across a wide variety of sites.

As part of prior work, we developed an NLP framework integrated with the OMOP CDM, the Open Health Natural Language Processing Toolkit (OHNLP toolkit). In this work, we will report on the usage of this toolkit to rapidly develop and prototype an NLP system using this toolkit to execute a clinical phenotyping task to meet a real world information need in the form of extracting signs and symptoms related to long COVID. Additionally, we will demonstrate a federated evaluation of said NLP system across participating sites.

METHODS

Due to PASC's international recognition, many useful literature resources have been developed. A recent literature review[1] identified 303 articles published before April 29, 2021, curated 59 relevant manuscripts that described clinical manifestations in 81 cohorts of individuals three weeks or more following acute COVID-19, and mapped 287 unique clinical findings to Human Phenotype Ontology (HPO) terms. Another study[2] consolidated 355 long COVID symptoms from 1,520 UMLS concepts of 16,466 synonyms, which were identified from 328,879 clinical notes of 26,117 COVID-19 positive patients from their post-acute infection period (day 51-110 from first positive COVID-19 test).

These resources were used as a base concept list, which was then cross-referenced with the Unified Medical Language System (UMLS) version 2021AB to obtain an algorithmically-derived base dictionary. Lexical variant generation was then ran on these terms to obtain normalized forms and used as a base NLP system. The coverage of this NLP system was then compared against an annotated 98 document clinical narrative set from the Post COVID Care Clinic at the Mayo Clinic, and NLP system finetuning was done using this narrative set to obtain the NLP system used for distribution across the RECOVER team.

This NLP system was then distributed alongside the OHNLP toolkit to 10 RECOVER sites for deployment. A subset of 5 RECOVER sites conducted federated evaluation by manually annotating 10 notes locally each with 2 annotators + 1 adjudicator and returned result coverage statistics to the Mayo Clinic.

1. Deer RR, Rock MA, Vasilevsky N, Carmody L, Rando H, Anzalone AJ, et al. Characterizing long covid: Deep phenotype of a complex condition. eBioMedicine. 2021;74:103722.
2. Wang L, Foer D, MacPhaul E, Lo Y-C, Bates DW, Zhou L. Pasclex: A comprehensive post-acute sequelae of Covid-19 (PASC) symptom lexicon derived from electronic health record clinical notes. Journal of Biomedical Informatics. 2022;125:103951.

RESULTS

TABLE 1: Dictionary Coverage Statistics

	Site 1	Site 2	Site 3	Site 4	Site 5
# Annotations	126	23	171	118	73
# Dictionary Entries	77	20	138	84	65
Coverage %	61.1%	87.0%	80.7%	71.2%	89%

TABLE 2: Dictionary Coverage Statistics

	Site 1	Site 2	Site 3	Site 4	Site 5
# Annotations missed by NLP	49	3	33	34	8
Missed Concepts	55.1%, 27	66.7%, 2	27.3%, 9	11.8%, 4	12.5%, 1
Missed Lexical Variants	40.8%, 20	33.3%, 1	51.5%, 17	76.5%, 26	50.0%, 4
Annotation error	4.1%, 2	0	21.2%, 7	11.8%, 4	37.5%, 3

DISCUSSION

The usage of a common data model for data ingest/output in the NLP pipeline greatly facilitated deployment of the developed NLP system across a wide variety of clinical healthcare sites. Additionally, our results fundamentally demonstrate the advantage of this high degree of portability when discussing generalizable NLP solutions: while the initial Mayo-developed NLP system performs decently across the RECOVER sites that returned evaluation results, not all information needs are met in terms of dictionary coverage. Refinement is thus vastly simplified as gaps in data coverage is assessed across a variety of differing healthcare institutions, thus granting greater confidence in the wide-range generalizability of the final solution.

While the OHNLP toolkit is only one example of how NLP systems can prospectively integrate with the OMOP CDM, we believe that the rapid deployment and ease of refinement demonstrates the importance of such solutions. Additionally, as many sites are now providing this data centrally back to the N3C collaborative for long COVID research use cases, we can begin examining the problem of whether the existing CDM tables NOTE and NOTE_NLP meet information needs for practical usage.

Installation scripts and source code for the OHNLP toolkit and associated NLP algorithms can be found at https://www.github.com/OHNLP/OHNLPTK_RECOVER and https://www.github.com/OHNLP/{Backbone/MedTagger} respectively

ACKNOWLEDGEMENTS

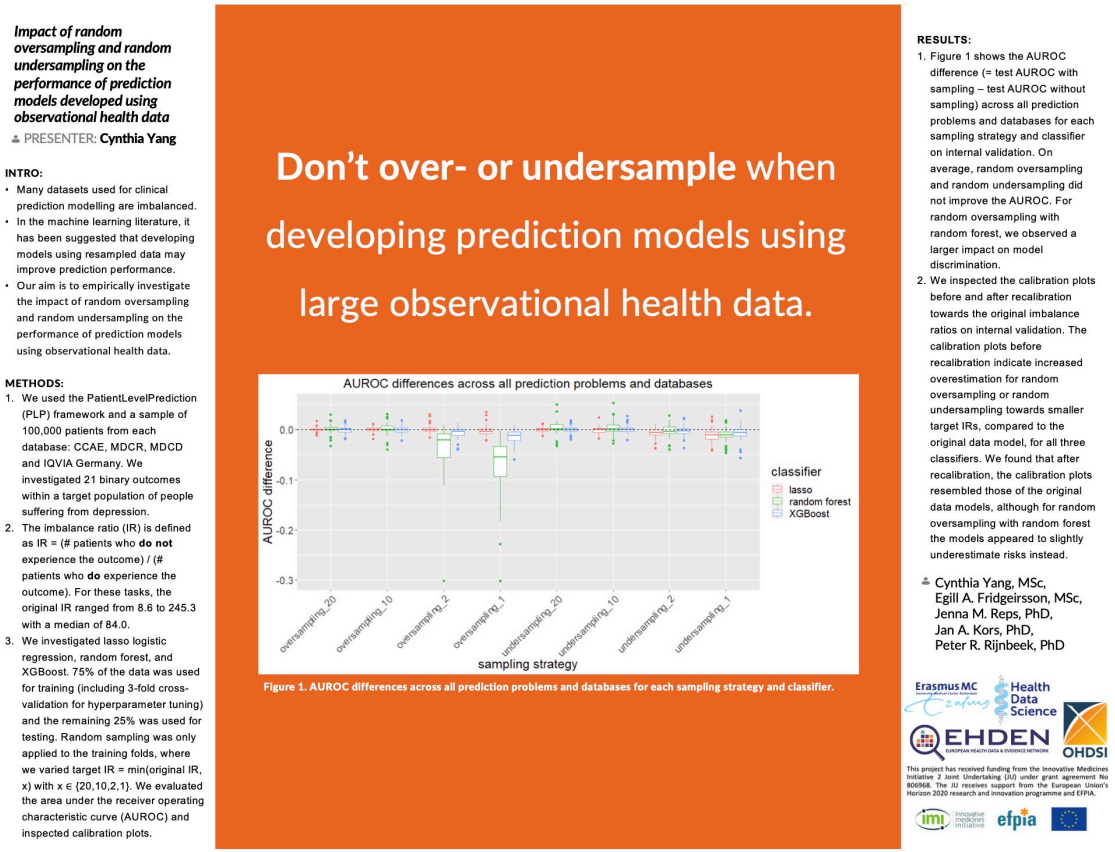
Many thanks to the sites participating in the federated evaluation subtask: University of Kentucky, University of Minnesota, Stony Brook, University of Michigan, and Columbia University

MONDAY

Deployment of an OMOP CDM-compatible NLP system for Rapid Development and Dissemination of a Long-COVID Extraction NLP task

(Andrew Wen, Liwei Wang, Huan He, Sunyang Fu, Sijia Liu, Hongfang Liu)

#OHDSISocialShowcase This Week



Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data

PRESENTER: **Cynthia Yang**

INTRO:

- Many datasets used for clinical prediction modelling are imbalanced.
- In the machine learning literature, it has been suggested that developing models using resampled data may improve prediction performance.
- Our aim is to empirically investigate the impact of random oversampling and random undersampling on the performance of prediction models using observational health data.

METHODS:

1. We used the PatientLevelPrediction (PLP) framework and a sample of 100,000 patients from each database: CCAE, MDCR, MDCD and IQVIA Germany. We investigated 21 binary outcomes within a target population of people suffering from depression.
2. The imbalance ratio (IR) is defined as IR = (# patients who **do not** experience the outcome) / (# patients who **do** experience the outcome). For these tasks, the original IR ranged from 8.6 to 245.3 with a median of 84.0.
3. We investigated lasso logistic regression, random forest, and XGBoost. 75% of the data was used for training (including 3-fold cross-validation for hyperparameter tuning) and the remaining 25% was used for testing. Random sampling was only applied to the training folds, where we varied target IR = min(original IR, x) with x ∈ {20,10,2,1}. We evaluated the area under the receiver operating characteristic curve (AUROC) and inspected calibration plots.

Don't over- or undersample when developing prediction models using large observational health data.

Figure 1. AUROC differences across all prediction problems and databases for each sampling strategy and classifier.

RESULTS:

1. Figure 1 shows the AUROC difference (= test AUROC with sampling – test AUROC without sampling) across all prediction problems and databases for each sampling strategy and classifier on internal validation. On average, random oversampling and random undersampling did not improve the AUROC. For random oversampling with random forest, we observed a larger impact on model discrimination.
2. We inspected the calibration plots before and after recalibration towards the original imbalance ratios on internal validation. The calibration plots before recalibration indicate increased overestimation for random oversampling or random undersampling towards smaller target IRs, compared to the original data model, for all three classifiers. We found that after recalibration, the calibration plots resembled those of the original data models, although for random oversampling with random forest the models appeared to slightly underestimate risks instead.

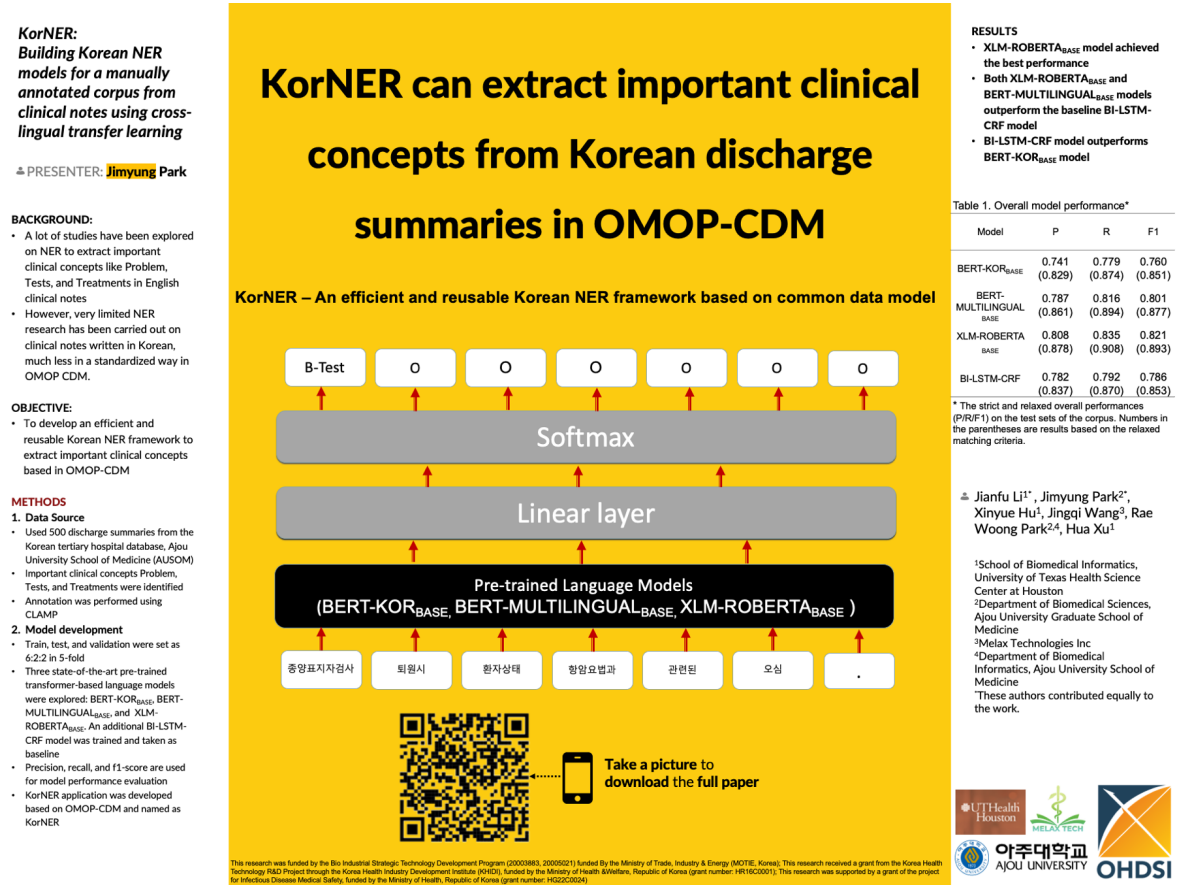
Cynthia Yang, MSc,
Egill A. Fridgeirsson, MSc,
Jenna M. Reps, PhD,
Jan A. Kors, PhD,
Peter R. Rijnbeek, PhD

This project 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.

TUESDAY

Impact of random oversampling and random undersampling on the development and validation of prediction models using observational health data (Cynthia Yang, Egill A. Fridgeirsson, Jan A. Kors, Jenna M. Reps, Peter R. Rijnbeek)

#OHDSISocialShowcase This Week



KorNER: Building Korean NER models for a manually annotated corpus from clinical notes using cross-lingual transfer learning

PRESENTER: Jimyung Park

BACKGROUND:
- A lot of studies have been explored on NER to extract important clinical concepts like Problem, Tests, and Treatments in English clinical notes
- However, very limited NER research has been carried out on clinical notes written in Korean, much less in a standardized way in OMOP CDM.

OBJECTIVE:
- To develop an efficient and reusable Korean NER framework to extract important clinical concepts based in OMOP-CDM

METHODS
1. **Data Source**
   - Used 500 discharge summaries from the Korean tertiary hospital database, Ajou University School of Medicine (AUSOM)
   - Important clinical concepts Problem, Tests, and Treatments were identified
   - Annotation was performed using CLAMP
2. **Model development**
   - Train, test, and validation were set as 6:2:2 in 5-fold
   - Three state-of-the-art pre-trained transformer-based language models were explored: BERT-KOR$_{BASE}$, BERT-MULTILINGUAL$_{BASE}$, and XLM-ROBERTA$_{BASE}$. An additional BI-LSTM-CRF model was trained and taken as baseline
   - Precision, recall, and f1-score are used for model performance evaluation
   - KorNER application was developed based on OMOP-CDM and named as KorNER

KorNER can extract important clinical concepts from Korean discharge summaries in OMOP-CDM

KorNER – An efficient and reusable Korean NER framework based on common data model

Take a picture to download the full paper

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

RESULTS
- XLM-ROBERTA$_{BASE}$ model achieved the best performance
- Both XLM-ROBERTA$_{BASE}$ and BERT-MULTILINGUAL$_{BASE}$ models outperform the baseline BI-LSTM-CRF model
- BI-LSTM-CRF model outperforms BERT-KOR$_{BASE}$ model

Table 1. Overall model performance*

Model	P	R	F1
BERT-KOR$_{BASE}$	0.741 (0.829)	0.779 (0.874)	0.760 (0.851)
BERT-MULTILINGUAL$_{BASE}$	0.787 (0.861)	0.816 (0.894)	0.801 (0.877)
XLM-ROBERTA$_{BASE}$	0.808 (0.878)	0.835 (0.908)	0.821 (0.893)
BI-LSTM-CRF	0.782 (0.837)	0.792 (0.870)	0.786 (0.853)

* The strict and relaxed overall performances (P/R/F1) on the test sets of the corpus. Numbers in the parentheses are results based on the relaxed matching criteria.

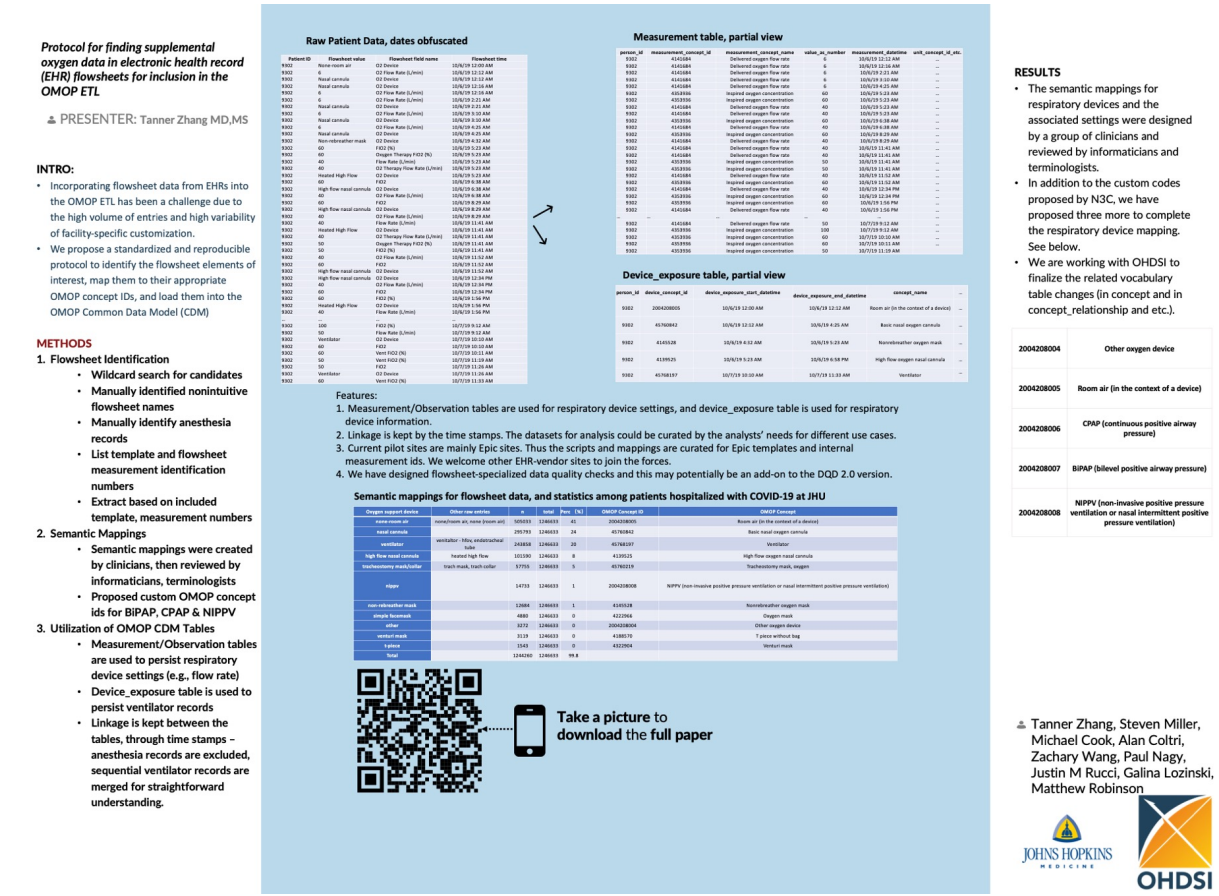
Jianfu Li[1*], Jimyung Park[2*], Xinyue Hu[1], Jingqi Wang[3], Rae Woong Park[2,4], Hua Xu[1]

[1]School of Biomedical Informatics, University of Texas Health Science Center at Houston
[2]Department of Biomedical Sciences, Ajou University Graduate School of Medicine
[3]Melax Technologies Inc
[4]Department of Biomedical Informatics, Ajou University School of Medicine
*These authors contributed equally to the work.

WEDNESDAY

Building Korean NER models for a manually annotated corpus from clinical notes using cross-lingual transfer learning (Jianfu Li, Jimyung Park, Xinyue Hu, Jingqi Wang, Rae Woong Park, Hua Xu)

#OHDSISocialShowcase This Week



Protocol for finding supplemental oxygen data in electronic health record (EHR) flowsheets for inclusion in the OMOP ETL

PRESENTER: Tanner Zhang MD,MS

INTRO:

- Incorporating flowsheet data from EHRs into the OMOP ETL has been a challenge due to the high volume of entries and high variability of facility-specific customization.
- We propose a standardized and reproducible protocol to identify the flowsheet elements of interest, map them to their appropriate OMOP concept IDs, and load them into the OMOP Common Data Model (CDM)

METHODS

1. Flowsheet Identification
   - Wildcard search for candidates
   - Manually identified nonintuitive flowsheet names
   - Manually identify anesthesia records
   - List template and flowsheet measurement identification numbers
   - Extract based on included template, measurement numbers
2. Semantic Mappings
   - Semantic mappings were created by clinicians, then reviewed by informaticians, terminologists
   - Proposed custom OMOP concept ids for BiPAP, CPAP & NIPPV
3. Utilization of OMOP CDM Tables
   - Measurement/Observation tables are used to persist respiratory device settings (e.g., flow rate)
   - Device_exposure table is used to persist ventilator records
   - Linkage is kept between the tables, through time stamps – anesthesia records are excluded, sequential ventilator records are merged for straightforward understanding.

Raw Patient Data, dates obfuscated

Patient ID	Flowsheet value	Flowsheet field name	Flowsheet time
9302	None-room air	O2 Device	10/6/19 12:00 AM
9302	6	O2 Flow Rate (L/min)	10/6/19 12:12 AM
9302	Nasal cannula	O2 Device	10/6/19 12:12 AM
9302	Nasal cannula	O2 Device	10/6/19 12:16 AM
9302	6	O2 Flow Rate (L/min)	10/6/19 12:16 AM
9302	6	O2 Flow Rate (L/min)	10/6/19 2:21 AM
9302	Nasal cannula	O2 Device	10/6/19 2:21 AM
9302	6	O2 Flow Rate (L/min)	10/6/19 3:10 AM
9302	Nasal cannula	O2 Device	10/6/19 3:10 AM
9302	6	O2 Flow Rate (L/min)	10/6/19 4:25 AM
9302	Nasal cannula	O2 Device	10/6/19 4:25 AM
9302	Non-rebreather mask	O2 Device	10/6/19 4:32 AM
9302	60	FiO2 (%)	10/6/19 5:23 AM
9302	60	Oxygen Therapy FiO2 (%)	10/6/19 5:23 AM
9302	40	Flow Rate (L/min)	10/6/19 5:23 AM
9302	40	O2 Therapy Flow Rate (L/min)	10/6/19 5:23 AM
9302	Heated High Flow	O2 Device	10/6/19 5:38 AM
9302	40	FiO2	10/6/19 5:38 AM
9302	High flow nasal cannula	O2 Device	10/6/19 5:38 AM
9302	40	O2 Flow Rate (L/min)	10/6/19 8:38 AM
9302	60	FiO2	10/6/19 8:29 AM
9302	High flow nasal cannula	O2 Device	10/6/19 8:29 AM
9302	40	O2 Flow Rate (L/min)	10/6/19 8:29 AM
9302	60	Flow Rate (L/min)	10/6/19 11:41 AM
9302	Heated High Flow	O2 Device	10/6/19 11:41 AM
9302	40	O2 Therapy Flow Rate (L/min)	10/6/19 11:41 AM
9302	50	Oxygen Therapy FiO2 (%)	10/6/19 11:41 AM
9302	50	FiO2 (%)	10/6/19 11:41 AM
9302	40	O2 Flow Rate (L/min)	10/6/19 11:52 AM
9302	60	FiO2	10/6/19 11:52 AM
9302	High flow nasal cannula	O2 Device	10/6/19 11:52 AM
9302	High flow nasal cannula	O2 Device	10/6/19 12:34 PM
9302	40	O2 Flow Rate (L/min)	10/6/19 12:34 PM
9302	60	FiO2	10/6/19 12:34 PM
9302	60	FiO2 (%)	10/6/19 1:56 PM
9302	Heated High Flow	O2 Device	10/6/19 1:56 PM
9302	40	Flow Rate (L/min)	10/6/19 1:56 PM
...	...	...	...
9302	100	FiO2 (%)	10/7/19 9:12 AM
9302	50	Flow Rate (L/min)	10/7/19 9:12 AM
9302	Ventilator	O2 Device	10/7/19 10:10 AM
9302	60	FiO2	10/7/19 10:10 AM
9302	60	Vent FiO2 (%)	10/7/19 10:11 AM
9302	50	Vent FiO2 (%)	10/7/19 11:19 AM
9302	50	FiO2	10/7/19 11:26 AM
9302	Ventilator	O2 Device	10/7/19 11:26 AM
9302	60	Vent FiO2 (%)	10/7/19 11:33 AM

Measurement table, partial view

person_id	measurement_concept_id	measurement_concept_name	value_as_number	measurement_datetime	unit_concept_id_etc.
9302	4141684	Delivered oxygen flow rate	6	10/6/19 12:12 AM	...
9302	4141684	Delivered oxygen flow rate	6	10/6/19 12:16 AM	...
9302	4141684	Delivered oxygen flow rate	6	10/6/19 2:21 AM	...
9302	4141684	Delivered oxygen flow rate	6	10/6/19 3:10 AM	...
9302	4141684	Delivered oxygen flow rate	6	10/6/19 4:25 AM	...
9302	4353936	Inspired oxygen concentration	60	10/6/19 5:23 AM	...
9302	4353936	Inspired oxygen concentration	60	10/6/19 5:23 AM	...
9302	4141684	Delivered oxygen flow rate	40	10/6/19 5:23 AM	...
9302	4141684	Delivered oxygen flow rate	40	10/6/19 5:23 AM	...
9302	4353936	Inspired oxygen concentration	60	10/6/19 6:38 AM	...
9302	4141684	Delivered oxygen flow rate	40	10/6/19 6:38 AM	...
9302	4353936	Inspired oxygen concentration	60	10/6/19 8:29 AM	...
9302	4141684	Delivered oxygen flow rate	40	10/6/19 8:29 AM	...
9302	4141684	Delivered oxygen flow rate	40	10/6/19 11:41 AM	...
9302	4353936	Inspired oxygen concentration	50	10/6/19 11:41 AM	...
9302	4353936	Inspired oxygen concentration	50	10/6/19 11:41 AM	...
9302	4141684	Delivered oxygen flow rate	40	10/6/19 11:52 AM	...
9302	4353936	Inspired oxygen concentration	60	10/6/19 11:52 AM	...
9302	4141684	Delivered oxygen flow rate	40	10/6/19 12:34 PM	...
9302	4353936	Inspired oxygen concentration	60	10/6/19 12:34 PM	...
9302	4353936	Inspired oxygen concentration	60	10/6/19 1:56 PM	...
9302	4141684	Delivered oxygen flow rate	40	10/6/19 1:56 PM	...
...	...	...	...	...	...
9302	4141684	Delivered oxygen flow rate	50	10/7/19 9:12 AM	...
9302	4353936	Inspired oxygen concentration	100	10/7/19 9:12 AM	...
9302	4353936	Inspired oxygen concentration	60	10/7/19 10:10 AM	...
9302	4353936	Inspired oxygen concentration	60	10/7/19 10:11 AM	...
9302	4353936	Inspired oxygen concentration	50	10/7/19 11:19 AM	...

Device_exposure table, partial view

person_id	device_concept_id	device_exposure_start_datetime	device_exposure_end_datetime	concept_name	...
9302	2004208005	10/6/19 12:00 AM	10/6/19 12:12 AM	Room air (in the context of a device)	...
9302	45760842	10/6/19 12:12 AM	10/6/19 4:25 AM	Basic nasal oxygen cannula	...
9302	4145528	10/6/19 4:32 AM	10/6/19 5:23 AM	Nonrebreather oxygen mask	...
9302	4139525	10/6/19 5:23 AM	10/6/19 6:58 PM	High flow oxygen nasal cannula	...
9302	45768197	10/7/19 10:10 AM	10/7/19 11:33 AM	Ventilator	...

Features:

1. Measurement/Observation tables are used for respiratory device settings, and device_exposure table is used for respiratory device information.
2. Linkage is kept by the time stamps. The datasets for analysis could be curated by the analysts' needs for different use cases.
3. Current pilot sites are mainly Epic sites. Thus the scripts and mappings are curated for Epic templates and internal measurement ids. We welcome other EHR-vendor sites to join the forces.
4. We have designed flowsheet-specialized data quality checks and this may potentially be an add-on to the DQD 2.0 version.

Semantic mappings for flowsheet data, and statistics among patients hospitalized with COVID-19 at JHU

Oxygen support device	Other raw entries	n	total	Perc (%)	OMOP Concept ID	OMOP Concept
none-room air	none/room air, none (room air)	505033	1246633	41	2004208005	Room air (in the context of a device)
nasal cannula		295793	1246633	24	45760842	Basic nasal oxygen cannula
ventilator	ventilator - hfnv, endotracheal tube	241858	1246633	20	45768197	Ventilator
high flow nasal cannula	heated high flow	101590	1246633	8	4139525	High flow oxygen nasal cannula
tracheostomy mask/collar	trach mask, trach collar	57755	1246633	5	45760219	Tracheostomy mask, oxygen
nippv		14733	1246633	1	2004208008	NIPPV (non-invasive positive pressure ventilation or nasal intermittent positive pressure ventilation)
non-rebreather mask		12684	1246633	1	4145528	Nonrebreather oxygen mask
simple facemask		4880	1246633	0	4222968	Oxygen mask
other		3272	1246633	0	2004208004	Other oxygen device
venturi mask		3139	1246633	0	4188570	T piece without bag
t-piece		1543	1246633	0	4322904	Venturi mask
Total		1246288	1246633	99.8		

Take a picture to download the full paper

RESULTS

- The semantic mappings for respiratory devices and the associated settings were designed by a group of clinicians and reviewed by informaticians and terminologists.
- In addition to the custom codes proposed by N3C, we have proposed three more to complete the respiratory device mapping. See below.
- We are working with OHDSI to finalize the related vocabulary table changes (in concept and in concept_relationship and etc.).

2004208004	Other oxygen device
2004208005	Room air (in the context of a device)
2004208006	CPAP (continuous positive airway pressure)
2004208007	BiPAP (bilevel positive airway pressure)
2004208008	NIPPV (non-invasive positive pressure ventilation or nasal intermittent positive pressure ventilation)

Tanner Zhang, Steven Miller, Michael Cook, Alan Coltri, Zachary Wang, Paul Nagy, Justin M Rucci, Galina Lozinski, Matthew Robinson

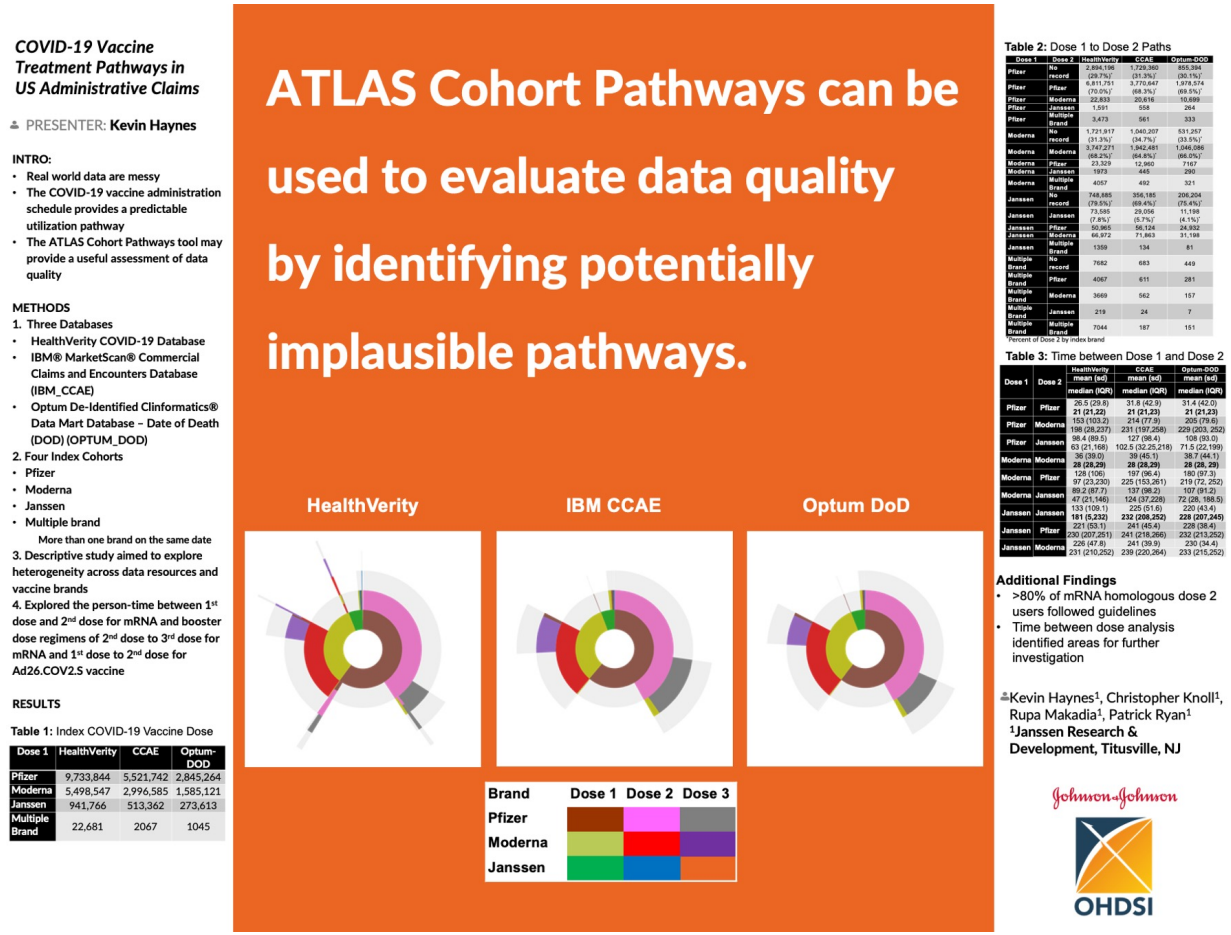
JOHNS HOPKINS MEDICINE

OHDSI

THURSDAY

Protocol for finding supplemental oxygen data in electronic health record (EHR) flowsheets for inclusion in the OMOP ETL (Tanner Zhang, Steven Miller, Michael Cook, Alan Coltri, Zachary Wang, Paul Nagy, Justin Rucci, Galina Lozinski, Matthew Robinson)

#OHDSISocialShowcase This Week



COVID-19 Vaccine Treatment Pathways in US Administrative Claims

PRESENTER: **Kevin Haynes**

INTRO:
- Real world data are messy
- The COVID-19 vaccine administration schedule provides a predictable utilization pathway
- The ATLAS Cohort Pathways tool may provide a useful assessment of data quality

METHODS
1. Three Databases
   - HealthVerity COVID-19 Database
   - IBM® MarketScan® Commercial Claims and Encounters Database (IBM_CCAE)
   - Optum De-Identified Clinformatics® Data Mart Database – Date of Death (DOD) (OPTUM_DOD)
2. Four Index Cohorts
   - Pfizer
   - Moderna
   - Janssen
   - Multiple brand

     More than one brand on the same date
3. Descriptive study aimed to explore heterogeneity across data resources and vaccine brands
4. Explored the person-time between 1st dose and 2nd dose for mRNA and booster dose regimens of 2nd dose to 3rd dose for mRNA and 1st dose to 2nd dose for Ad26.COV2.S vaccine

RESULTS

Table 1: Index COVID-19 Vaccine Dose

Dose 1	HealthVerity	CCAE	Optum-DOD
Pfizer	9,733,844	5,521,742	2,845,264
Moderna	5,498,547	2,996,585	1,585,121
Janssen	941,766	513,362	273,613
Multiple Brand	22,681	2067	1045

ATLAS Cohort Pathways can be used to evaluate data quality by identifying potentially implausible pathways.

HealthVerity | IBM CCAE | Optum DoD

Brand	Dose 1	Dose 2	Dose 3
Pfizer			
Moderna			
Janssen			

Table 2: Dose 1 to Dose 2 Paths

Dose 1	Dose 2	HealthVerity	CCAE	Optum-DOD
Pfizer	No record	2,894,196 (29.7%)	1,729,360 (31.3%)	855,394 (30.1%)
Pfizer	Pfizer	6,811,751 (70.0%)	3,770,647 (68.3%)	1,978,574 (69.5%)
Pfizer	Moderna	22,833	20,616	10,699
Pfizer	Janssen	1,591	558	264
Pfizer	Multiple Brand	3,473	661	333
Moderna	No record	1,721,917 (31.3%)	1,040,207 (34.7%)	531,267 (33.5%)
Moderna	Moderna	3,747,271 (68.2%)	1,942,461 (64.8%)	1,046,086 (66.0%)
Moderna	Pfizer	23,329	12,960	7167
Moderna	Janssen	1973	445	290
Moderna	Multiple Brand	4057	492	321
Janssen	No record	748,585 (79.5%)	356,185 (69.4%)	206,204 (75.4%)
Janssen	Janssen	73,585 (7.8%)	28,056 (5.7%)	11,198 (4.1%)
Janssen	Pfizer	50,965	56,124	24,932
Janssen	Moderna	66,972	71,863	31,198
Janssen	Multiple Brand	1359	134	81
Multiple Brand	No record	7682	683	449
Multiple Brand	Pfizer	4067	611	281
Multiple Brand	Moderna	3669	562	157
Multiple Brand	Janssen	219	24	7
Multiple Brand	Multiple Brand	7044	187	151

Percent of Dose 2 by index brand

Table 3: Time between Dose 1 and Dose 2

Dose 1	Dose 2	HealthVerity mean (sd) median (IQR)	CCAE mean (sd) median (IQR)	Optum-DOD mean (sd) median (IQR)
Pfizer	Pfizer	26.5 (29.8) **21 (21,22)**	31.8 (42.9) **21 (21,23)**	31.4 (42.0) **21 (21,23)**
Pfizer	Moderna	153 (103.2) 198 (28,237)	214 (77.9) 231 (197,258)	205 (79.6) 229 (203, 252)
Pfizer	Janssen	98.4 (89.5) 63 (21,168)	127 (98.4) 102.5 (32.25,218)	108 (93.0) 71.5 (22,199)
Moderna	Moderna	36 (39.0) **28 (28,29)**	39 (45.1) **28 (28,29)**	38.7 (44.1) **28 (28, 29)**
Moderna	Pfizer	128 (106) 97 (23,230)	197 (96.4) 225 (153,261)	180 (97.3) 219 (72, 252)
Moderna	Janssen	89.2 (87.7) 47 (21,146)	137 (98.2) 124 (37,228)	107 (91.2) 72 (28, 188.5)
Janssen	Janssen	133 (109.1) **181 (5,232)**	225 (51.6) **232 (208,252)**	220 (43.4) **228 (207,245)**
Janssen	Pfizer	221 (53.1) 230 (207,251)	241 (45.4) 241 (218,266)	228 (38.4) 232 (213,252)
Janssen	Moderna	226 (47.8) 231 (210,252)	241 (39.9) 239 (220,264)	230 (34.4) 233 (215,252)

Additional Findings
- >80% of mRNA homologous dose 2 users followed guidelines
- Time between dose analysis identified areas for further investigation

Kevin Haynes[1], Christopher Knoll[1], Rupa Makadia[1], Patrick Ryan[1]
[1]**Janssen Research & Development, Titusville, NJ**

Johnson & Johnson

OHDSI

FRIDAY

COVID-19 Vaccine Administration Pathways in US Administrative Claims

(Kevin Haynes, Christopher Knoll, Rupa Makadia, Patrick Ryan)

Opening: Northeastern University

Northeastern University invites applications for multiple tenured/ tenure-track faculty positions in support of an Impact Engine centered on large-scale observational health data science and informatics to start in the fall of 2023. These faculty will be core members of our Real-World Healthcare Navigator (RWHN) Impact Engine which aims to change how research is translated into clinical practice by establishing a sustainable service that leads the way in fully reproducing health studies.

Careers

Open Rank Professor of Large Scale Observational Data Science

Apply

Boston, MA (Main Campus)

Full time

Posted 14 Days Ago

R110388

About the Opportunity

About Northeastern

Founded in 1898, Northeastern is a global research university and the recognized leader in experience-driven lifelong learning. Our world-renowned experiential approach empowers our students, faculty, alumni, and partners to create impact far beyond the confines of discipline, degree, and campus.
Our locations—in Boston; Oakland; Arlington, Charlotte, North Carolina; London; Portland, Maine; Oakland; San Francisco; Seattle; Silicon Valley; Toronto; Vancouver; and the Massachusetts communities of Burlington and Nahant—are nodes in our growing global university system. Through this network, we expand opportunities for flexible, student-centered learning and collaborative, solutions-focused research.

Northeastern's comprehensive array of undergraduate and graduate programs— in a variety of on-campus and online formats—lead to degrees through the doctorate in nine colleges and schools. Among these, we offer more than 195 multi-discipline majors and degrees designed to prepare students for purposeful lives and careers.

Responsibilities

Responsibilities will include teaching undergraduate and graduate courses, conducting an independent and externally funded research program, and participating in departmental, college, and university service. Qualified candidates must have expertise in, or a demonstrated commitment to, working with diverse student populations and/or in a culturally diverse work and educational environment.

Qualifications

- PhD or equivalent in Statistics, Bioinformatics, Data Science, Epidemiology, Computer Science, Computer Engineering, or similar field.
- Expertise working with large relational databases (e.g., EHRs, Medicare) preferred.
- Advanced knowledge of analytic approaches including data wrangling, visualization, and machine learning preferred.
- Expertise in either R or Python.

Opening: Northeastern University

The OHDSI Center at the Roux Institute seeks a postdoctoral fellow to join their team focused on developing statistical methods and applying them to observational data from large-scale federated datasets (e.g. electronic health records and administrative claims data), with specific applications to the safety of biologics. This research will directly improve our ability to use real world data to characterize patient populations, construct population level estimates relating exposures to health outcomes, and to enhance clinical decision making through improved patient-level predictions.

Careers

Postdoctoral Research Fellow, Observational Health Data Science and Informatics

Apply

Portland, ME
Full time
Posted 30+ Days Ago
R109484

About the Opportunity

Job Summary:
The Observational Health Data Science and Informatics (OHDSI) Center housed within the Roux Institute at Northeastern University (NU) is a new administrative hub of the largest observational health research community in the world. The OHDSI Center @ the Roux Institute works to advance OHDSI's research mission of improving health by empowering a community that collaboratively generates evidence that promotes better health decisions and better care. As part of a multi-institution team, the OHDSI Center is participating in a project to provide support to the U.S. Food and Drug Administration's Biologics Effectiveness and Safety (BEST) program. The mission of the BEST program is to conduct safety and effectiveness surveillance of biologic products, such as vaccines, tissues, and advanced therapeutics.

Responsibilities:
This fellow will collaborate with a growing OHDSI Center team of faculty, staff, postdocs, and students to develop and test statistical methodologies related to the use of real world data (e.g. electronic health records and administrative claims data) to better analyze observational health studies. Fellows will be expected to publish and present their work in leading journals and conferences, participate in departmental seminars, and meet regularly with the OHDSI Center research team. There are also opportunities to mentor junior team members and provide educational support across the Center. The ideal candidate will have a strong statistical and computational background, experience processing and analyzing large datasets, and outstanding communication skills.

Qualifications:
- Ph.D. or equivalent degree in Biostatistics, Data Science, Statistics, Epidemiology, Computer Science, or related fields
- Research experience in observational health data theory/methods
- Demonstrated experience working with large observational health databases and/or medical claims data
- Advanced experience in statistical programming languages such as R or Python and familiarity with version control (i.e., Git/Github)
- Excellent writing and communication skills

Opening: FDA/CDER

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

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

Opening: Tufts Medicine

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

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

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

Remote work is possible for both positions.

1. Link for Senior Project Manager position: https://smrtr.io/bBVzh

2. Link for Lead Software Developer and Research Data Warehouse Manager position: https://jobs.smartrecruiters.com/TuftsMedicalCenter1/743999857980631-software-development-lead-res-g-c-ctsi

Andrew's email: awilliams15@tuftsmedicalcenter.org

Openings: Johns Hopkins University

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

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

Where Are We Going?

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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Nov. 15: Open Network Studies

Expanding maternal and infant data from EHRs for pregnancy research

- Safety and Effectiveness of Anti-Hypertensive Medications in Pregnancy
- Project to Characterize Anti-Hypertensive, Anti-Coagulant, Anti-Diabetic and Antibiotic Medication Usage During Pregnancy and Postpartum

Alison Callahan (Instructor, Medicine • Stanford University)
Stephanie Leonard (Instructor, Obstetrics & Gynecology • Stanford University)
Louisa Smith (Assistant Professor, Health Sciences • Northeastern University)

Relative Risk of Cervical Neoplasms Associated with Copper and Levonorgestrel Secreting Intrauterine Devices: Real World Evidence from the OHDSI Network

Matthew Spotnitz (Postdoctoral Research Fellow • Columbia University)