10-Minute Tutorials

OHDSI Community Call
Nov. 22, 2022 • 11 am ET

Upcoming OHDSI Community Calls

Date	Topic
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

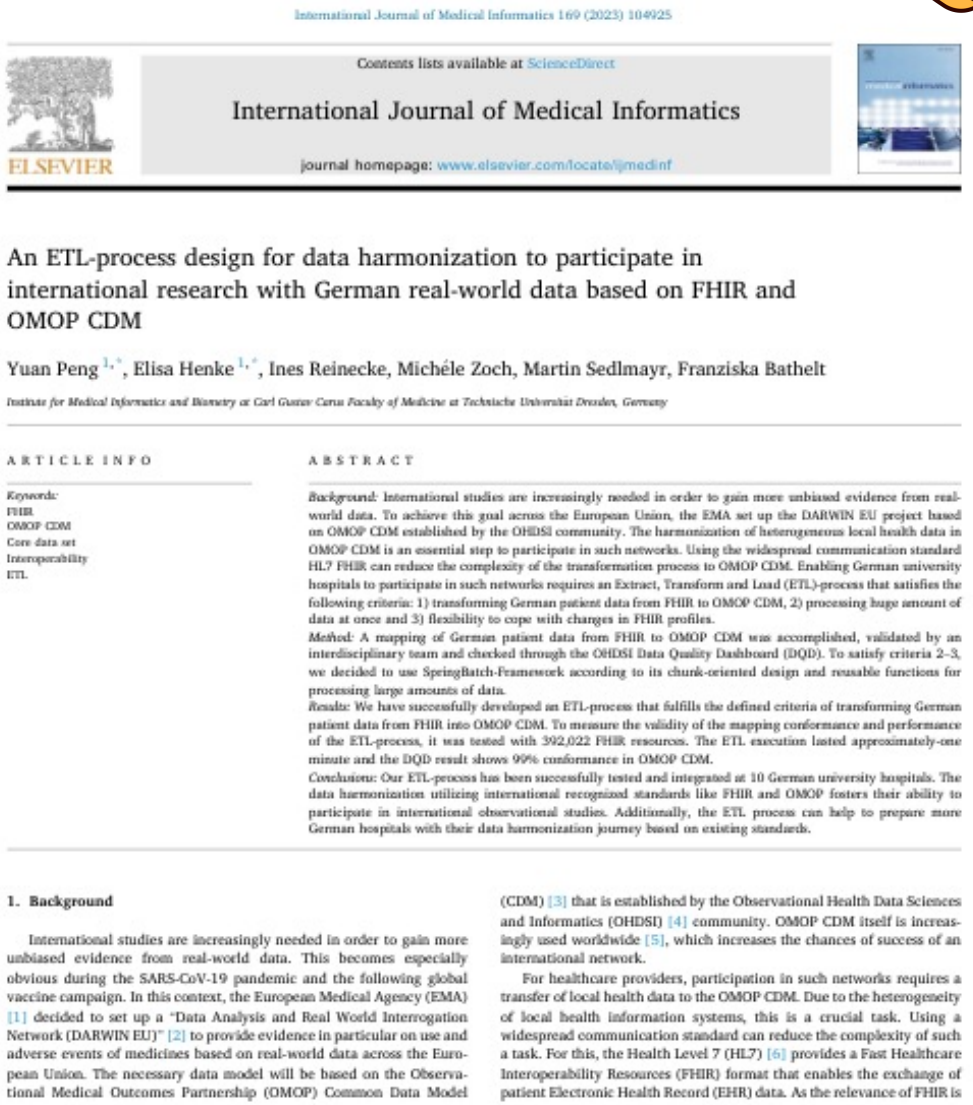
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Yuan Peng, Elisa Henke, Ines Reinecke, Michéle Zoch, Martin Sedlmayr, and Franziska Bathelt** on the publication of **An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM** in the International Journal of Medical Informatics.



International Journal of Medical Informatics 169 (2023) 104925

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM

Yuan Peng [1,*], Elisa Henke [1,*], Ines Reinecke, Michéle Zoch, Martin Sedlmayr, Franziska Bathelt

Institute for Medical Informatics and Biometry at Carl Gustav Carus Faculty of Medicine at Technische Universität Dresden, Germany

ARTICLE INFO

Keywords:
FHIR
OMOP CDM
Core data set
Interoperability
ETL

ABSTRACT

Background: International studies are increasingly needed in order to gain more unbiased evidence from real-world data. To achieve this goal across the European Union, the EMA set up the DARWIN EU project based on OMOP CDM established by the OHDSI community. The harmonization of heterogeneous local health data in OMOP CDM is an essential step to participate in such networks. Using the widespread communication standard HL7 FHIR can reduce the complexity of the transformation process to OMOP CDM. Enabling German university hospitals to participate in such networks requires an Extract, Transform and Load (ETL)-process that satisfies the following criteria: 1) transforming German patient data from FHIR to OMOP CDM, 2) processing huge amount of data at once and 3) flexibility to cope with changes in FHIR profiles.

Method: A mapping of German patient data from FHIR to OMOP CDM was accomplished, validated by an interdisciplinary team and checked through the OHDSI Data Quality Dashboard (DQD). To satisfy criteria 2–3, we decided to use SpringBatch-Framework according to its chunk-oriented design and reusable functions for processing large amounts of data.

Results: We have successfully developed an ETL-process that fulfills the defined criteria of transforming German patient data from FHIR into OMOP CDM. To measure the validity of the mapping conformance and performance of the ETL-process, it was tested with 392,022 FHIR resources. The ETL execution lasted approximately-one minute and the DQD result shows 99% conformance in OMOP CDM.

Conclusion: Our ETL-process has been successfully tested and integrated at 10 German university hospitals. The data harmonization utilizing international recognized standards like FHIR and OMOP fosters their ability to participate in international observational studies. Additionally, the ETL process can help to prepare more German hospitals with their data harmonization journey based on existing standards.

1. Background

International studies are increasingly needed in order to gain more unbiased evidence from real-world data. This becomes especially obvious during the SARS-CoV-19 pandemic and the following global vaccine campaign. In this context, the European Medical Agency (EMA) [1] decided to set up a "Data Analysis and Real World Interrogation Network (DARWIN EU)" [2] to provide evidence in particular on use and adverse events of medicines based on real-world data across the European Union. The necessary data model will be based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [3] that is established by the Observational Health Data Sciences and Informatics (OHDSI) [4] community. OMOP CDM itself is increasingly used worldwide [5], which increases the chances of success of an international network.

For healthcare providers, participation in such networks requires a transfer of local health data to the OMOP CDM. Due to the heterogeneity of local health information systems, this is a crucial task. Using a widespread communication standard can reduce the complexity of such a task. For this, the Health Level 7 (HL7) [6] provides a Fast Healthcare Interoperability Resources (FHIR) format that enables the exchange of patient Electronic Health Record (EHR) data. As the relevance of FHIR is

OHDSI Shoutouts!

Congratulations to the team of **Yerim Kim, Seung In Seo, Kyung Joo Lee, Jinseob Kim, Jong Jin Yoo, Won-Woo Seo, Hyung Seok Lee, Woon Geon Shin** on the publication of **Long-term use of proton-pump inhibitor on Alzheimer's disease: a real-world distributed network analysis of six observational Korean databases using a Common Data Model** in Therapeutic Advances in Neurological Disorders.

Therapeutic Advances in Neurological Disorders
Volume 15, January-December 2022
© The Author(s), 2022, Article Reuse Guidelines
https://doi.org/10.1177/17562864221135700

Original Research

Long-term use of proton-pump inhibitor on Alzheimer's disease: a real-world distributed network analysis of six observational Korean databases using a Common Data Model

Yerim Kim [1], Seung In Seo [2,3], Kyung Joo Lee [4], Jinseob Kim [5], Jong Jin Yoo [6], Won-Woo Seo [7], Hyung Seok Lee [8], and Woon Geon Shin [9,10]

Background: Dementia has a crucial impact on the quality of life of elderly patients and their caregivers. Proton-pump inhibitors (PPIs) are the most frequently prescribed treatment, but they have been shown to be associated with dementia. The data are inconsistent, however.

Objective: To investigate the association between PPIs use and Alzheimer's disease (AD) or all-cause dementia in six observational Korean databases using a Common Data Model (CDM) and to perform a distributed network analysis.

Methods: Subjects aged over 18 years between 1 January 2004 and 31 December 2020. Among 7,293,565 subjects from 6 cohorts, 41,670 patients met the eligibility criteria. A total of 2206 patients who were included in both cohorts or with a history of dementia were excluded. After propensity matching, 5699 propensity-matched pairs between the PPIs and histamine-2 receptor antagonist (H_2RA) users were included in this study. The primary outcome was the incidence of AD at least 365 days after drug exposure. The secondary outcome was the incidence of all-cause dementia at least 365 days after drug exposure.

Results: In the 1:1 propensity score matching, the risk of AD or all-cause dementia was not significantly different between the PPIs and H_2RA groups in all six databases. In the distributed network analysis, the long-term PPI users (≥365 days) were unassociated with AD [hazard ratio (HR) = 0.92, 95% confidence interval (CI) = 0.68–1.23; $I^2 = 0\%$] and all-cause dementia (HR = 1.04, 95% CI = 0.82–1.31; $I^2 = 0\%$) compared with H_2RA users.

Conclusion: In the distributed network analysis of six Korean hospital databases using Observational Medical Outcomes Partnership (OMOP)-CDM data, the long-term use of PPI was not associated with a statistically significantly increased risk of AD or all-cause dementia. Therefore, we suggest that physicians should not avoid these medications because of concern about dementia risk.

OHDSI Shoutouts!

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

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

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Common Data Model Vocabulary Subgroup
Tuesday	3 pm	OMOP CDM Oncology Outreach/Research Subgroup
Wednesday	10 am	FHIR and OMOP Digital Quality Measurements Subgroup (ZOOM)
Wednesday	11 am	Latin America
Wednesday	11 am	Open-Source Community
Thursday	10 am	Medical Devices
Friday	9 am	Pheontype Development and Evaluation
Friday	9 am	Geographic Information System (GIS) Development
Friday	10:15 am	Clinical Trials
Monday	10 am	Healthcare Systems Interest Group

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

OMOP CDM ERD Challenge

Patrick Ryan shared a recent forum post called “Introducing the OMOP CDM ER Diagram Challenge” and is calling for community submissions by Tuesday, Dec. 13.

The winner will be announced at the Dec. 20 community call!

Introducing the OMOP CDM ER diagram challenge! Submissions due 13Dec2022

General

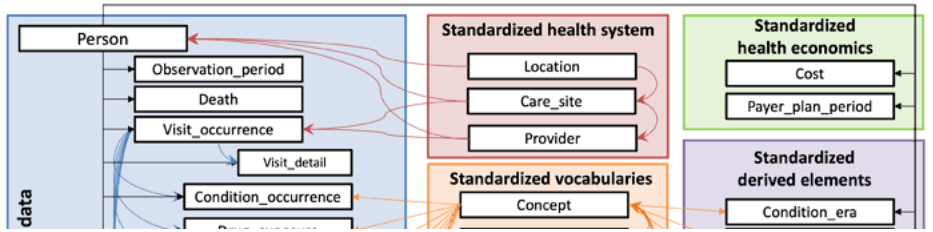
Patrick_Ryan 1 3d

Friends:

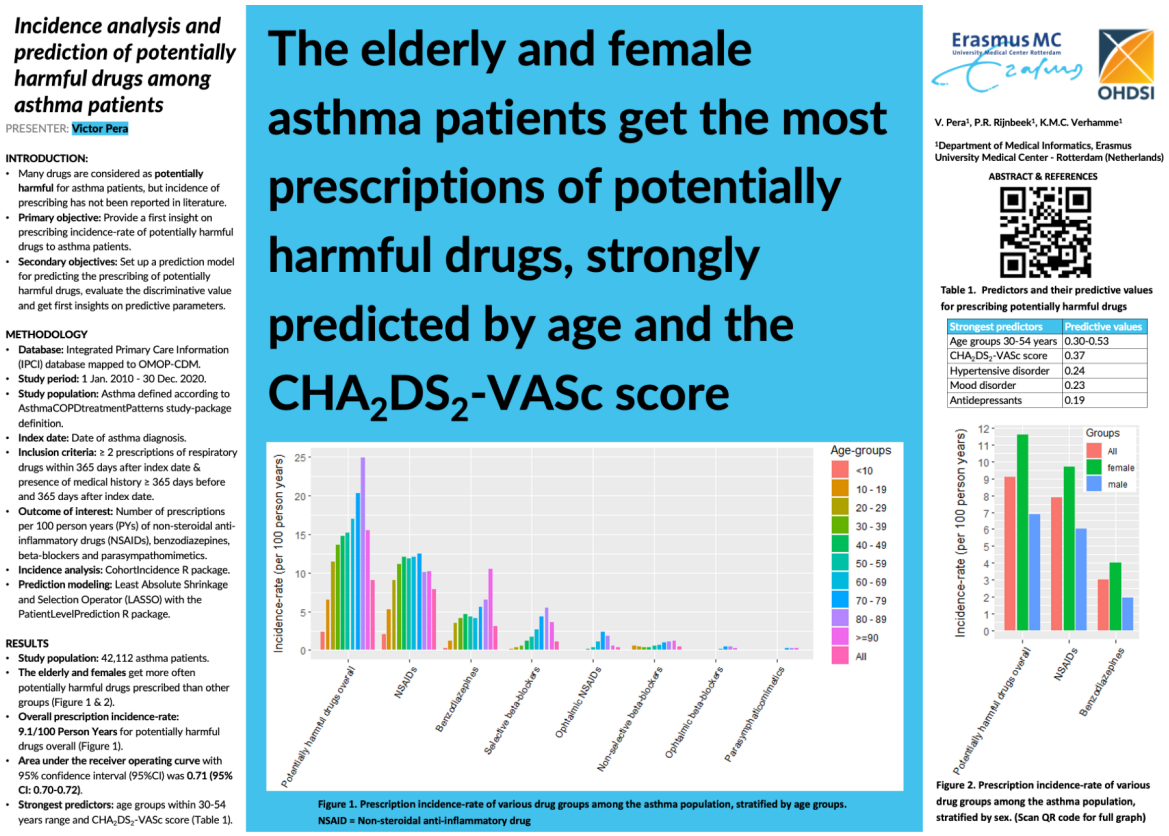
I’d like to announce the OMOP Common Data Model (CDM) Entity-Relationship Diagram (ERD) Challenge! Starting today, all members of the OHDSI community are welcome and encouraged to submit their entries of the best ERD for the OMOP CDM to this forum post (or to the CDM Workgroup teams site) by Tuesday, Dec 13. One winner will be selected by a committee from the CDM workgroup, and announced on OHDSI’s last community call of the year on Dec20, with their award-winning ERD being a gift to our entire community, posted on the official OMOP CDM git page, but also the winner receiving a special gift from the OHDSI community! This should be a fun activity for our community, particularly those of you helping drive our open community data standards, to learn and collaborate and contribute to a community resource that all of us can benefit in. So, please accept the OMOP CDM ERD Challenge and get diagraming!

Background:

The OMOP Common Data Model v5.4 serves our community well as an open community data standard to enable standardized analytics and large-scale evidence generation. The CDM Workgroup, under the amazing leadership of @clairblacketer , has done a tremendous job of stewarding this standard, improving our documentations and conventions, and providing reference implementations, all available here at: index.knit 3 . Our primary schematic that we use to describe the OMOP CDM is below:



#OHDSISocialShowcase This Week



Incidence analysis and prediction of potentially harmful drugs among asthma patients

PRESENTER: Victor Pera

INTRODUCTION:

- Many drugs are considered as **potentially harmful** for asthma patients, but incidence of prescribing has not been reported in literature.
- **Primary objective:** Provide a first insight on prescribing incidence-rate of potentially harmful drugs to asthma patients.
- **Secondary objectives:** Set up a prediction model for predicting the prescribing of potentially harmful drugs, evaluate the discriminative value and get first insights on predictive parameters.

METHODOLOGY

- **Database:** Integrated Primary Care Information (IPCI) database mapped to OMOP-CDM.
- **Study period:** 1 Jan. 2010 - 30 Dec. 2020.
- **Study population:** Asthma defined according to AsthmaCOPDtreatmentPatterns study-package definition.
- **Index date:** Date of asthma diagnosis.
- **Inclusion criteria:** ≥ 2 prescriptions of respiratory drugs within 365 days after index date & presence of medical history ≥ 365 days before and 365 days after index date.
- **Outcome of interest:** Number of prescriptions per 100 person years (PYs) of non-steroidal anti-inflammatory drugs (NSAIDs), benzodiazepines, beta-blockers and parasympathomimetics.
- **Incidence analysis:** CohortIncidence R package.
- **Prediction modeling:** Least Absolute Shrinkage and Selection Operator (LASSO) with the PatientLevelPrediction R package.

RESULTS

- **Study population:** 42,112 asthma patients.
- **The elderly and females** get more often potentially harmful drugs prescribed than other groups (Figure 1 & 2).
- **Overall prescription incidence-rate: 9.1/100 Person Years** for potentially harmful drugs overall (Figure 1).
- **Area under the receiver operating curve** with 95% confidence interval (95%CI) was **0.71 (95% CI: 0.70-0.72)**.
- **Strongest predictors:** age groups within 30-54 years range and CHA_2DS_2-VASc score (Table 1).

The elderly and female asthma patients get the most prescriptions of potentially harmful drugs, strongly predicted by age and the CHA_2DS_2-VASc score

Figure 1. Prescription incidence-rate of various drug groups among the asthma population, stratified by age groups.
NSAID = Non-steroidal anti-inflammatory drug

Erasmus MC University Medical Center Rotterdam

OHDSI

V. Pera[1], P.R. Rijnbeek[1], K.M.C. Verhamme[1]

[1]Department of Medical Informatics, Erasmus University Medical Center - Rotterdam (Netherlands)

ABSTRACT & REFERENCES

Table 1. Predictors and their predictive values for prescribing potentially harmful drugs

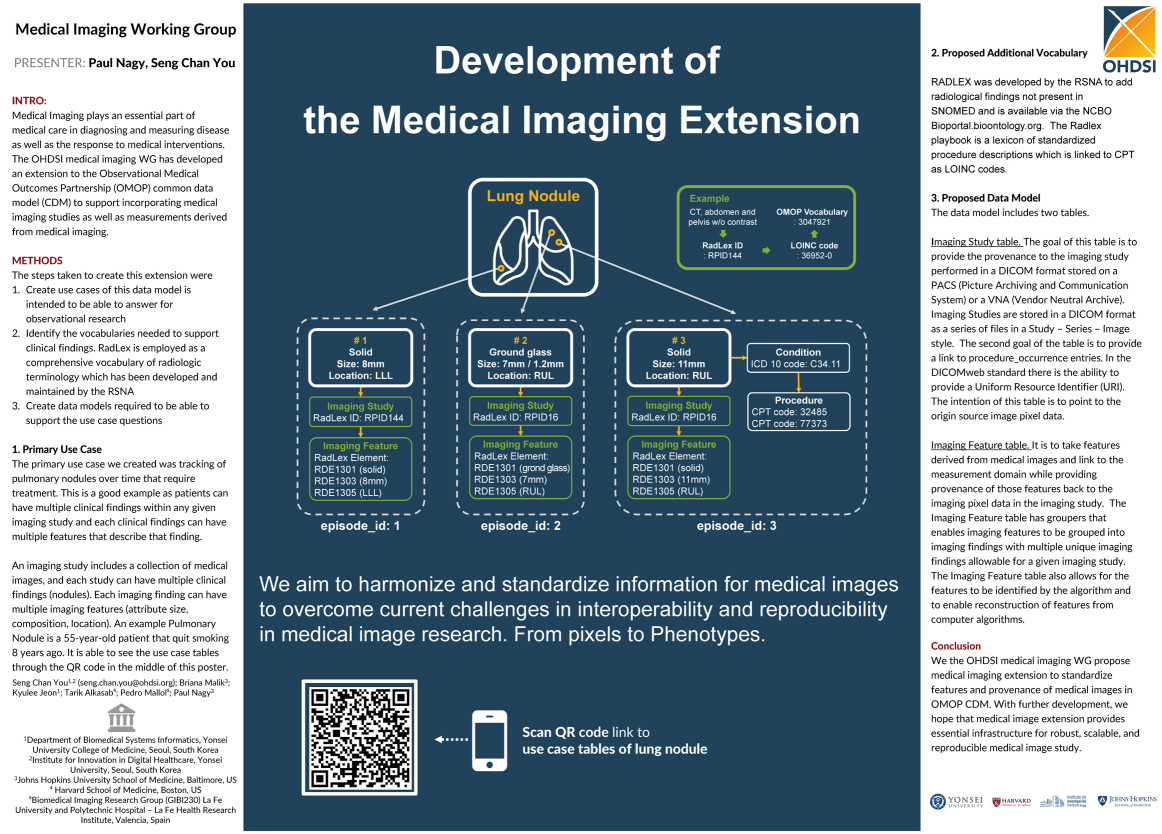
Strongest predictors	Predictive values
Age groups 30-54 years	0.30-0.53
CHA_2DS_2-VASc score	0.37
Hypertensive disorder	0.24
Mood disorder	0.23
Antidepressants	0.19

Figure 2. Prescription incidence-rate of various drug groups among the asthma population, stratified by sex. (Scan QR code for full graph)

MONDAY

Incidence analysis and prediction of potentially harmful drugs among asthma patients (Victor Pera, Peter Rijnbeek, Katia Verhamme)

#OHDSISocialShowcase This Week



Medical Imaging Working Group

PRESENTER: **Paul Nagy, Seng Chan You**

INTRO:
Medical Imaging plays an essential part of medical care in diagnosing and measuring disease as well as the response to medical interventions. The OHDSI medical imaging WG has developed an extension to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to support incorporating medical imaging studies as well as measurements derived from medical imaging.

METHODS
The steps taken to create this extension were
1. Create use cases of this data model is intended to be able to answer for observational research
2. Identify the vocabularies needed to support clinical findings. RadLex is employed as a comprehensive vocabulary of radiologic terminology which has been developed and maintained by the RSNA
3. Create data models required to be able to support the use case questions

1. Primary Use Case
The primary use case we created was tracking of pulmonary nodules over time that require treatment. This is a good example as patients can have multiple clinical findings within any given imaging study and each clinical findings can have multiple features that describe that finding.

An imaging study includes a collection of medical images, and each study can have multiple clinical findings (nodules). Each imaging finding can have multiple imaging features (attribute size, composition, location). An example Pulmonary Nodule is a 55-year-old patient that quit smoking 8 years ago. It is able to see the use case tables through the QR code in the middle of this poster.

Seng Chan You[1,2] (seng.chan.you@ohdsi.org); Briana Malik[3]; Kyulee Jeon[1]; Tarik Alkasab[4]; Pedro Mallol[5]; Paul Nagy[3]

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, South Korea
[2]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, South Korea
[3]Johns Hopkins University School of Medicine, Baltimore, US
[4] Harvard School of Medicine, Boston, US
[5]Biomedical Imaging Research Group (GIBI230) La Fe University and Polytechnic Hospital – La Fe Health Research Institute, Valencia, Spain

Development of the Medical Imaging Extension

We aim to harmonize and standardize information for medical images to overcome current challenges in interoperability and reproducibility in medical image research. From pixels to Phenotypes.

Scan QR code link to use case tables of lung nodule

2. Proposed Additional Vocabulary
RADLEX was developed by the RSNA to add radiological findings not present in SNOMED and is available via the NCBO Bioportal.bioontology.org. The Radlex playbook is a lexicon of standardized procedure descriptions which is linked to CPT as LOINC codes.

3. Proposed Data Model
The data model includes two tables.

Imaging Study table. The goal of this table is to provide the provenance to the imaging study performed in a DICOM format stored on a PACS (Picture Archiving and Communication System) or a VNA (Vendor Neutral Archive). Imaging Studies are stored in a DICOM format as a series of files in a Study – Series – Image style. The second goal of the table is to provide a link to procedure_occurrence entries. In the DICOMweb standard there is the ability to provide a Uniform Resource Identifier (URI). The intention of this table is to point to the origin source image pixel data.

Imaging Feature table. It is to take features derived from medical images and link to the measurement domain while providing provenance of those features back to the imaging pixel data in the imaging study. The Imaging Feature table has groupers that enables imaging features to be grouped into imaging findings with multiple unique imaging findings allowable for a given imaging study. The Imaging Feature table also allows for the features to be identified by the algorithm and to enable reconstruction of features from computer algorithms.

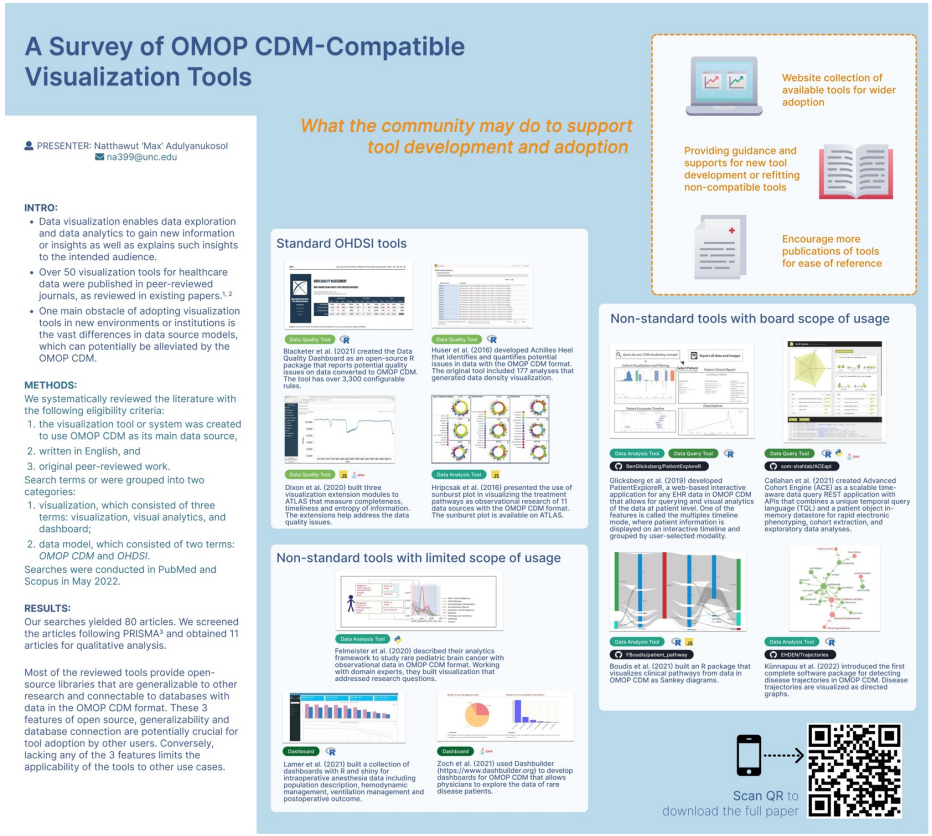
Conclusion
We the OHDSI medical imaging WG propose medical imaging extension to standardize features and provenance of medical images in OMOP CDM. With further development, we hope that medical image extension provides essential infrastructure for robust, scalable, and reproducible medical image study.

TUESDAY

Development of the Medical Imaging Extension for OMOP-CDM

(Briana Malik, Kyulee Jeon, Tarik Alkasab, Pedro Mallol, Seng Chan You, Paul Nagy)

#OHDSISocialShowcase This Week



WEDNESDAY A survey of OMOP CDM-compatible visualization tools & what the community may do to support tool development and adoption (Natthawut Adulyanukosol, David Gotz)

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?

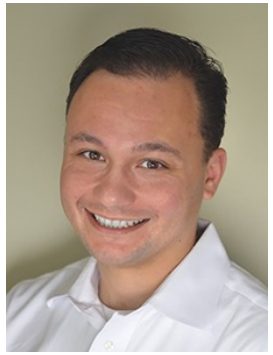
Nov. 22: 10-Minute Tutorials

PHOEBE 2.0

Anna Ostropolets
PhD Student, Columbia University

Automated Comparator Selection

Justin Bohn
Associate Director, Epidemiology at Janssen



Strategus

Anthony Sena
Associate Director, Observational Health Data Analytics at Janssen

Einstein-ATLAS

Selvin Soby
Director, Informatics & Data Analytics at Montefiore

Broadsea

Lee Evans
Founder, LTS Computing LLC