Evidence Network/ Data Diagnostics

OHDSI Community Call
March 18, 2025 • 11 am ET

Upcoming Community Calls

Date	Topic
Mar. 18	OHDSI Evidence Network and Data Diagnostics Design
Mar. 25	Methods for Evaluating Data Fitness for Use
Apr. 1	Recent OHDSI Publications
Apr. 8	Strategus Update & Review
Apr. 15	Treatment Pathways
Apr. 22	Current Practices in Estimation and Prediction
Apr. 29	DevCon 2025 Review
May 6	Evidence Synthesis
May 13	Maternal Health Fellowship Review

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	CDM Vocabulary Subgroup
Tuesday	12 pm	Atlas
Tuesday	1 pm	Common Data Model
Wednesday	7 am	Medical Imaging
Wednesday	1 pm	Perinatal and Reproductive Health
Thursday	8 am	OHDSI India Community Call
Thursday	9 am	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	9 am	CDM Vocabulary Subgroup Office Hours
Thursday	11 am	Themis
Thursday	12 pm	HADES
Friday	10 am	Transplant
Friday	10 am	GIS - Geographic Information System
Friday	10:30 am	Open-Source Community
Friday	11:30 am	Steering
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Healthcare Systems Interest Group

Wednesday: Europe Student Career Webinar

Patrick Ryan Honored By Women in Pharma

OHDSI Europe Symposium - Save-the-date!

OHDSI BELGIUM

OHDSI Europe Symposium

5-7 July 2025

Registrations open

End of February 2025

Abstract submission deadline

31 March 2025

Notification of selection

5 May 2025

UHASSELT

OHDSI EUROPE

OHDSI

Old Prison - Hasselt University

Martelarenlaan, Hasselt - BELGIUM

Global Symposium: Oct. 7-9

The 2025 OHDSI Global Symposium will return to the Hyatt Regency Hotel in New Brunswick, N.J., on Oct. 7-9.

More information on the collaborator showcase will be coming soon.

The Center for Advanced Healthcare Research Informatics (CAHRI) at Tufts Medicine welcomes:

Hongfang Liu, PhD

D. Bradley McWilliams Chair Professor of Biomedical Informatics, Vice President of Learning Health System, University of Texas Health Science Center at Houston

'A Translational Science Framework in Advancing Healthcare AI'

March 27, 2025, 11am-12pm EST
Virtually via Zoom

Please contact Marty Alvarez at malvarez2@tuftsmedicalcenter.org for calendar invite or questions.

TuftsMedicine
Tufts Medical Center

#OHDSISocialShowcase This Week

Monday

Towards Reproducible Imaging Research: Implementation of DICOM to OMOP CDM

(Woo Yeon Park, Ben Martin, Gabriel Salvador, Blake Dewey, Teri Sippel Schmidt, Paul Nagy)

Towards Reproducible Imaging Research: Implementation of DICOM to OMOP CDM

PRESENTER: Jen Park

INTRO

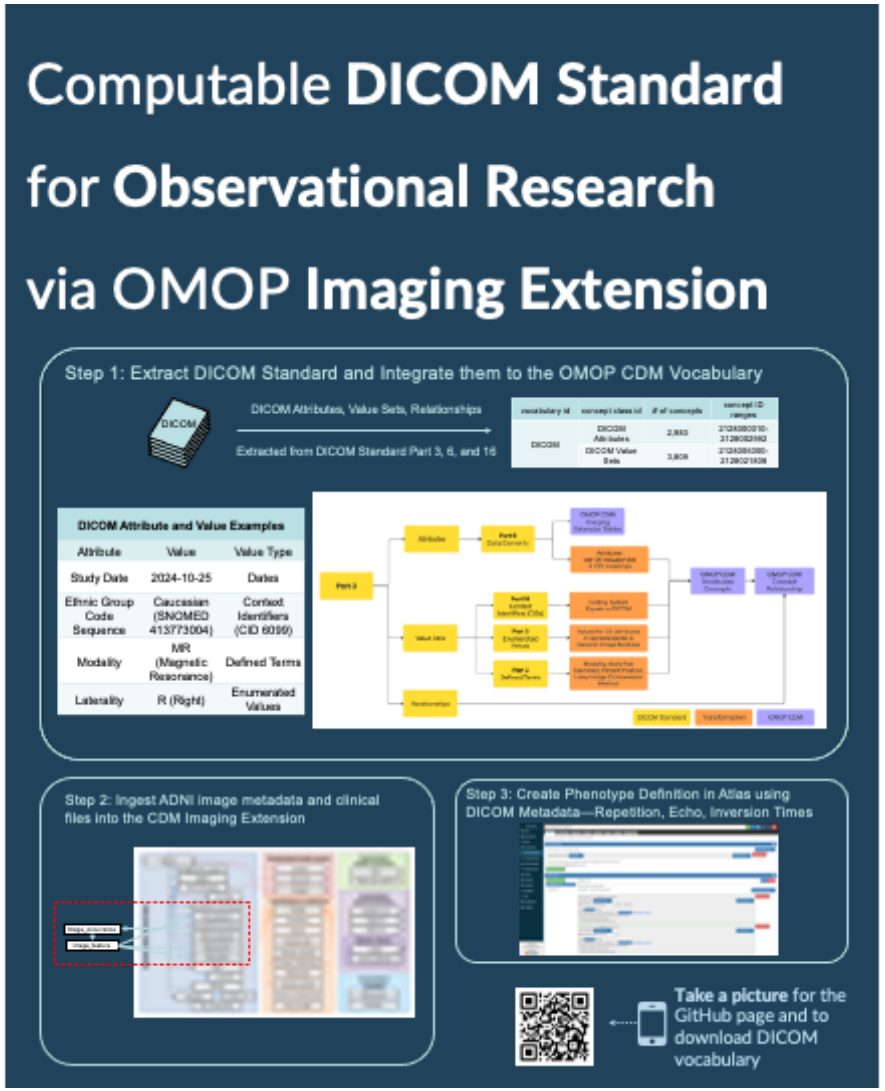
- Healthcare utilizes many forms of data such as structured, images, waveforms, and narrative texts. The multimodality of datasets imposes various challenges to researchers, such as data processing, knowledge abstraction, and reproducibility.
- This study aims to integrate DICOM terminologies to OMOP CDM vocabulary and demonstrate them using imaging extension tables and Alzheimer's Diseases Neuroimaging Initiative (ADNI) data.

METHODS

1. **Harvest Digital Images in Communications in Medicine (DICOM) Standard:** We harvested Parts 3, 6, and 16 of DICOM standards. These were then added as custom concepts in the OMOP CDM.
2. **Data Transformation and Ingestion:** We downloaded and transformed patient demographic and neuropsychiatric inventory files to update the Person and Measurement tables. After extracting DICOM metadata from ADNI images, we populated OMOP CDM and medical imaging extension tables using the DICOM concepts created in Step 1.
3. **Phenotype Definition in Atlas:** The cohort definition was recreated using Atlas. The criteria included having done a T1-weighted Brain MRI scan, evaluated neuropsychiatric inventory score (NPI), and gotten an Alzheimer's disease diagnosis.

RESULTS

- We identified 2,983 DICOM Attribute concepts, 3,809 Value Sets concepts, and 601,825 concept relationships from DICOM Standard.
- We ingested 545 ADNI DICOM studies, which included 4,756 DICOM series for 289 patients. The DICOM series metadata resulted in 296,396 elements, organized in the imaging extension tables and other clinical domain tables such as Measurement.



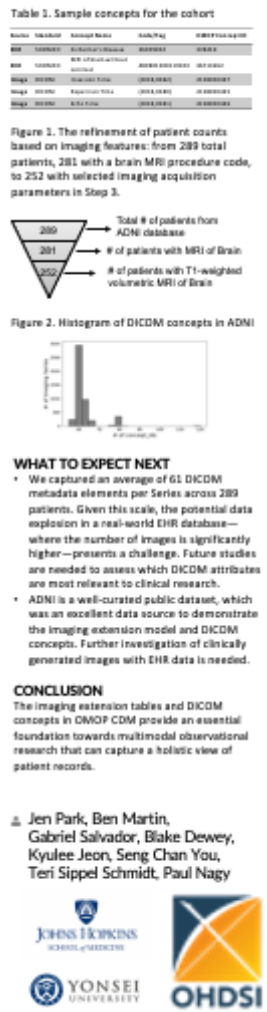


Table 1. Sample concepts for the cohort

Figure 1. The refinement of patient counts based on imaging features: from 289 total patients, 281 with a brain MRI procedure code, to 252 with selected imaging acquisition parameters in Step 3.

289 — Total # of patients from ADNI database
281 — # of patients with MRI of Brain
252 — # of patients with T1-weighted volumetric MRI of Brain

Figure 2. Histogram of DICOM concepts in ADNI

WHAT TO EXPECT NEXT

- We captured an average of 61 DICOM metadata elements per Series across 289 patients. Given this scale, the potential data explosion in a real-world EHR database—where the number of images is significantly higher—presents a challenge. Future studies are needed to assess which DICOM attributes are most relevant to clinical research.
- ADNI is a well-curated public dataset, which was an excellent data source to demonstrate the imaging extension model and DICOM concepts. Further investigation of clinically generated images with EHR data is needed.

CONCLUSION

The imaging extension tables and DICOM concepts in OMOP CDM provide an essential foundation towards multimodal observational research that can capture a holistic view of patient records.

Jen Park, Ben Martin, Gabriel Salvador, Blake Dewey, Kyulee Jeon, Seng Chan You, Teri Sippel Schmidt, Paul Nagy

Johns Hopkins School of Medicine · Yonsei University · OHDSI

#OHDSISocialShowcase This Week

Tuesday

Using Vaccine Ontology to Analyze and Integrate Vaccine Terms in N3C Dataset

(Yuanyi Pan, Jie Zheng,Yongqun Oliver He)

Using Vaccine Ontology to Analyze and Integrate Vaccine Terms in N3C Dataset

Yuanyi Pan, MD[1,2], Jie Zheng, PhD[2], Yongqun Oliver He, PhD[2], on Behalf of N3C

1 Guizhou University Medical college, Guiyang 550025, Guizhou Province, China;
2 University of Michigan Medical School, Ann Arbor, MI, USA.

Background

The National COVID Cohort Collaborative (N3C) dataset, one of the largest and most detailed collections of electronic health record (EHR) data related to COVID-19 patients, enabling COVID-19 vaccine studies with rich records. N3C employs the OMOP CDM as its basic infrastructure. However, the vast and heterogeneous nature of the N3C dataset presents significant challenges for integrating and analyzing specific vaccine terms. Leveraging CVX, RxNorm, and RxNorm Extension to standardize vaccine-related concepts lack robust semantic relations and proper hierarchy, leading to ineffective and discrepancy in the vaccine-related research.

The Vaccine Ontology (VO) is a biomedical ontology that organizes vaccine terms systematically, offering better coverage and structure. VO helps in consistently annotating and integrating vaccine-related data, making it a valuable tool for analyzing large, complex datasets like N3C. In this study, we explore how mapping vaccine records in N3C to VO can improve data classification and support advanced vaccine-related research.

Methods

1. Extraction of COVID-19 vaccine data from N3C

Based on the current designated hierarchy of Athena and standard vocabulary, we first used concept *'947817:covid-19 vaccines; systemic'* as the highest level ancestor term of COVID-19 vaccines terms to retrieve all COVID-19 vaccine records. The used SQL code is shown as follows.

```
WITH cte_covid_vaccine AS
( SELECT b.*
FROM concept_ancestor a JOIN concept b ON a.descendant_concept_id = b.concept_id
AND a.ancestor_concept_id = 947817 )
SELECT drug_concept_id, drug_concept_name, vocabulary_id, COUNT(*) AS counts
FROM drug_exposure a JOIN cte_covid_vaccine b ON a.drug_concept_id = b.concept_id
GROUP BY drug_concept_id, drug_concept_name, vocabulary_id;
```

In addition, we used wildcards to retrieve COVID-19 vaccine terms as a comparison to see if any records are uncaptured. The used code is shown as follows.

```
SELECT drug_concept_id, drug_concept_name, vocabulary_id, COUNT(*) AS counts
FROM drug_exposure a
WHERE drug_concept_name like '%COVID-19%vaccine%'
GROUP BY drug_concept_id, drug_concept_name, vocabulary_id;
```

All data were accessed on June 12, 2024 on N3C enclave. A data quality check was executed to check for missing value and outliers.

2. Mapped vaccine terms to VO

All collected vaccine terms were mapped to VO and then classified, analyzed based on VO pattern. One-to-one exact mapping was employed throughout the process, which means that for any single term, there is one and only one VO term mapped to it with the same gratuity and semantic content. No uphill or downhill mapping was allowed. So we might add new VO terms if necessary corresponding to a non existing vaccine term. To support terminology-specific annotations in VO, specific annotation properties, including 'RxNorm ID', 'RxNorm Extension ID', and 'OMOP concept ID', among others, were later added to VO to represent the corresponding content. The Robot tool and Protege-OWL editor were used to edit and display the terms.

Contact: yuanyp@umich.edu; yongqunh@med.umich.edu

Methods

1. Summary of vaccine records from N3C

'947817:covid-19 vaccines; systemic' : 25,835,254 rows records were extracted, including 17 distinct COVID-19 vaccine terms. All 17 terms were from RxNorm.

Using wildcards: 27,371,805 rows were extracted, including 36 different COVID-19 vaccine terms, including 31 terms of RxNorm, 3 of CVX and 1 of RxNorm Extension. (1,536,560 more rows)

Table 1 lists the top ten most frequently identified terms with detail. 'SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension' was the most frequent COVID-19 vaccine term in N3C.

No.	concept_id	concept_name	Vocabulary	Counts	VO_ID
1	37003436	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension	RxNorm	7,913,175	VO:0020221
2	1759206	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]	RxNorm	7,141,808	VO:0020222
3	37003518	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension	RxNorm	5,405,988	VO:0020206
4	779679	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension [Spikevax]	RxNorm	1,290,092	VO:0020207
5	1525538	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.05 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 OMICRON (BA.4/BA.5) 0.05 MG/ML Injectable Suspension	RxNorm	1,012,279	VO:0020217
6	702118	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.05 MG/ML Injectable Suspension	RxNorm	793,039	VO:0020216
7	724904	SARS-COV-2 (COVID-19) vaccine, UNSPECIFIED	CVX	703,050	VO:0006704
8	37003432	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein	RxNorm	650,716	VO:0020194
9	1525543	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.05 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 OMICRON (BA.4/BA.5) 0.05 MG/ML Injectable Suspension	RxNorm	533,834	VO:0020201
10	739906	SARS-COV-2 (COVID-19) vaccine, vector - Ad26 100000000000 UNT/ML Injectable Suspension	RxNorm	515,887	VO:0020227

Table 1. Top 10 RxNorm COVID-19 Vaccine Concepts Based on Record Counts

Results

2. VO-based analysis of N3C vaccine records after vaccine term mapping

Figure 1 shows how the VO represents the hierarchical structure of the vaccines with records in N3C. Our study found clearer relations among these vaccine terms.

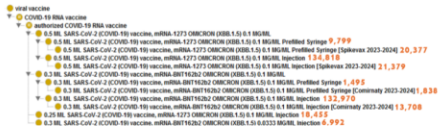


Figure 1. An example of VO hierarchical structure of the vaccines with N3C records.
Protégé-OWL editor was used for ontology visualization. The numbers represent the counts of vaccine records in N3C.

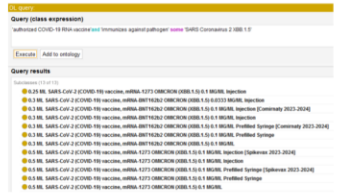


Figure 2. DL-query of XBB.1.5 containing COVID-19 vaccines

Overall, our DL-query identified 11 specific XBB.1.5-containing COVID-19 vaccines (Figure 2). Similarly, SPARQL can be used to perform such a query (data not shown). These queries demonstrate that the VO supports more advanced data analysis of N3C vaccine studies.

Conclusions

The vaccines recorded in the N3C dataset were mapped to and then analyzed using the VO. Our study shows that the VO improves semantic classification and applications of vaccine records in N3C, leading to more advanced data query and analysis.

#OHDSISocialShowcase This Week

Wednesday

Building OHDSI with Privacy Computing in Shanghai Medical College, Fudan University

(Changran Wang, Lei Liu, Feizhen Wu, Li Lin)

Building OHDSI with Privacy Computing in Shanghai Medical College, Fudan University

Wang Changran, Liu Lei, Wu Feizhen, Lin Li
Medical Science Data Center
Intelligent Medicine Institute
Shanghai Medical College
Fudan University

Background

The continuous advancement of global medical informatization has resulted in vast amounts of health data, reaching exabyte and zettabyte scales. However, these data are scattered across various institutions, hindering their orderly circulation. The integration of artificial intelligence (AI) with healthcare is emerging as a key driver in transforming medical technology. Health data are essential for clinical research, and multi-center research institutions that can securely aggregate data from various sources demonstrate superior efficiency. This approach facilitates the collection of extensive datasets, enabling deeper and more comprehensive data utilization.

Shanghai Medical College is positioned to seize this historic opportunity by participating in the pilot construction of high-level local universities in Shanghai. The institution aims to leverage AI in healthcare to enhance its strengths and address its weaknesses. By fostering innovation in clinical research and promoting multi-center research collaborations, Shanghai Medical College seeks to establish a robust multi-party data collaboration model, advancing clinical medical research. This integration will promote interdisciplinary cooperation and scientific innovation, accelerating the overall development of medical disciplines. Committed to becoming a top-tier domestic and world-class medical school, Shanghai Medical College will significantly contribute to public health improvements and the advancement of a healthy China and a healthy Shanghai.

Methods

Determine the quality management content of multi-center clinical research medical data, and confirm the multi-center clinical research medical data collection process and quality control content through literature review and expert consultation. On the basis of the investigation of the medical data quality management system, a standard framework for the construction of a multi-center medical data platform was constructed, and a multi-center clinical research data quality management system was established from multiple dimensions such as operating procedures, information collection, and quality control.

Based on the core technologies of privacy computing (federated learning, secure sandbox, multi-party secure computing, etc.), the privacy computing engine (PCP) is used to ensure the safe flow of data. Referring to the experience of multi-center clinical research projects and guided by the OHDSI-OMOP model, a multi-center clinical research approach is proposed, which closely combines the project lead and participating units, bringing together multi-party research data on the platform for joint application in research can increase the dimension and breadth of clinical research data, and relies on the project to build a data platform Cooperation in the cultivation of medical talents and the research and promotion of new technologies.

This project builds a clinical data collaborate platform (CDCP), improves data collection and governance capabilities, system design is shown as figure 1. According to the cooperation mode of the OHDSI model, create a unified medical terminology system by using web Protégé, accelerate the process of medical data circulation and application, and empower clinical research cooperation and medical data sharing with advanced technology. We also create a data security sharing mechanism, and improves the integration of data resources. Provide relevant platform support for "building multiple high-level multi-center clinical medical research projects".

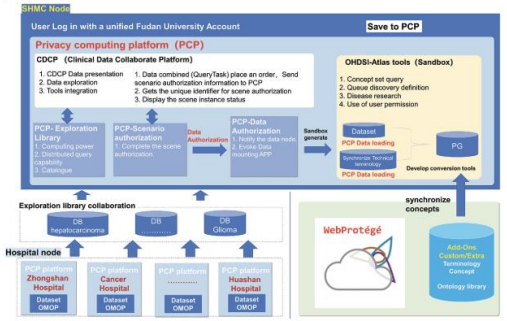


Figure 1: CDCP + PCP Design

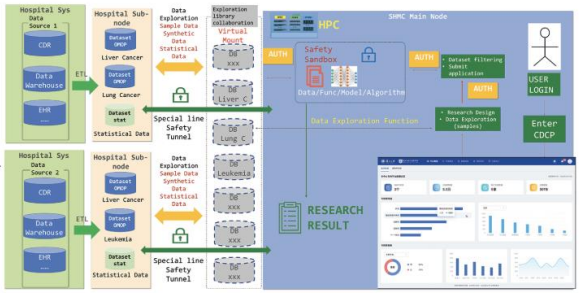


Figure 2: Conceptual Topology

Results

As a crucial support platform for multi-center clinical research, aligned with the actual needs of Shanghai Medical College at Fudan University and in accordance with relevant laws and regulations, this initiative provides robust medical data management support for multi-center research projects that adhere to medical ethics and pertinent regulations. A basic privacy computing platform is established within each hospital to host distributed databases, which undergo ETL processes to conform to the OHDSI format. The dataflow is shown as Figure 2.

This platform empowers multicenter clinical research at Fudan University. The fusion of multi-center data offers several key advantages to the medical research efforts at Shanghai Medical College. In recent years, multi-center clinical research has been increasingly conducted across various disease fields, with multiple research units and researchers collaboratively executing work based on the same design and objectives. These studies encompass clinical drug trials and more generalized clinical investigations, including prospective and retrospective studies.

The sandbox environment created on the main node HPC cluster allows researchers to access necessary data without viewing actual patient details, thus maintaining data privacy. This invisible data access, coupled with the ability to utilize OHDSI analytic tools, provides significant advantages. Researchers can perform complex analyses and derive insights without compromising patient confidentiality. This capability not only enhances the efficiency and scope of clinical research but also fosters collaboration across institutions, driving forward medical innovation and improving patient outcomes.

Conclusions

In response to the needs of Shanghai Medical College at Fudan University and in compliance with relevant laws and medical ethics, a robust support platform for multi-center clinical research has been established. This platform features a basic privacy computing infrastructure within hospitals to host distributed databases conforming to the OHDSI format, facilitating the secure management of multi-center research data. Currently, the system comprises three sub nodes, two disease categories, and four distributed databases.

The fusion of multi-center data within this framework offers significant advantages for clinical research, including enhanced data volume and diversity, which are crucial for comprehensive medical studies. Researchers can access necessary data through a secure sandbox environment on the main node HPC cluster, ensuring patient privacy while allowing the use of advanced OHDSI analytic tools. This approach has already proven beneficial in recent studies, which encompass clinical drug trials and broader clinical investigations. The ability to perform complex analyses without compromising patient confidentiality enhances the efficiency and scope of clinical research, fostering collaboration across institutions and driving medical innovation.

Currently, the platform is focused on three sub nodes and two disease categories, utilizing four distributed databases. However, the vision for the future is expansive. Plans are underway to increase the number of sub nodes to six, significantly broadening the data and research capabilities of the platform. Additionally, there is a strategic initiative to pilot collaborations with overseas institutions and other research entities, leveraging main nodes to conduct multi-center clinical research on a global scale.

This planned expansion will further enhance the platform's capacity for high-level research, fostering greater collaboration and innovation in the field of clinical medicine. By integrating advanced privacy computing technologies and adhering to stringent ethical standards, the platform aims to set a new benchmark in multi-center clinical research, ultimately contributing to improved patient outcomes and the advancement of medical science. This comprehensive approach ensures that the platform will remain at the forefront of clinical research, continually evolving to meet the growing needs of the medical community.

#OHDSISocialShowcase This Week

Thursday

Collaborative Population-adjusted Indirect Comparison with Multiple Single-arm Data Sources

(Yuru Zhu, Huiyuan Wang, Haitao Chu, Yong Chen)

Collaborative Population-adjusted Indirect Comparison with Multiple Single-arm Data

Yuru Zhu[a,b], Huiyuan Wang[a,b], Haitao Chu[c,d], and Yong Chen[a,b,e,f,g,h]

a The Center for Health Analytics and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA
b Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
c Global Biometrics and Data Management, Pfizer Inc., New York, NY, USA
d Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA
e Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA
f Leonard Davis Institute of Health Economics, Philadelphia, PA, USA
g Penn Medicine Center for Evidence-based Practice (CEP), Philadelphia, PA, USA
h Penn Institute for Biomedical Informatics (IBI), Philadelphia, PA, USA

Background

- **Treatment comparison** is critical in healthcare, with applications spanning drug development, public health policy, and precision medicine.
- While randomized controlled trials (RCTs) are the gold standard, they are often impractical due to strict eligibility criteria, ethical concerns, or the growing number of treatment options. Single-arm trials, where all participants receive the same treatment, offer a practical alternative, especially in rare diseases, early-phase drug development, and cases where comparator groups are infeasible.
- **Goal**: Propose a **communication-efficient** method to **estimate average treatment effects (ATEs) for all pairwise treatment comparisons in all combined populations** which consist of arbitrary number of the sub-populations for different sites in a **distributed research network** (DRN). Each site conducts a **single-arm study** on a unique treatment, with **only aggregated data allowed for sharing across sites** in the DRN.

Method

- We develop **doubly robust (DR)** and locally efficient estimators for ATEs across various target populations, using the **calibration weighting (CW) approach** (Hainmueller, 2012) to balance covariates across sites. These estimators are consistent if either of the working models for the propensity score and outcome is correctly specified, and can accommodate nonparametric methods.
- We provide **a lossless algorithm with 3 communications**: In a DRN with K sites,
  1. Within each site k, we regress the outcome on the covariates and obtain the fitted outcome model function $\hat{m}_k(\cdot)$, then send the summary statistics of covariates $\bar{\boldsymbol{g}}_1(X)$ and the fitted outcome model function $\hat{m}_k(\cdot)$ to the cloud server.
  2. The cloud server sends $\{\bar{\boldsymbol{g}}_1(X), \hat{m}_1(\cdot), \ldots, \bar{\boldsymbol{g}}_K(X), \hat{m}_K(\cdot)\}$ to all sites..
  3. Within each site k, obtain $\omega_{jk}(X_s)$ by calibration weighting for $j \neq k$ and $\omega_{kk}(X_s) = 1/n_k$, then calculate the aggregated data AD_k and send them to the cloud server.

$$\mathrm{AD}_k = \{A^1_{ki}, A^2_{ki}, i \in \{1, \cdots, K\}\}$$

$$A^1_{kj} = \frac{1}{n_k} \sum_{D_s=k} \hat{m}_j(X_s),\ A^2_{jk} = \sum_{D_s=j} \omega_{kj}(X_s)\{Y_s - \hat{m}_j(X_s)\}$$

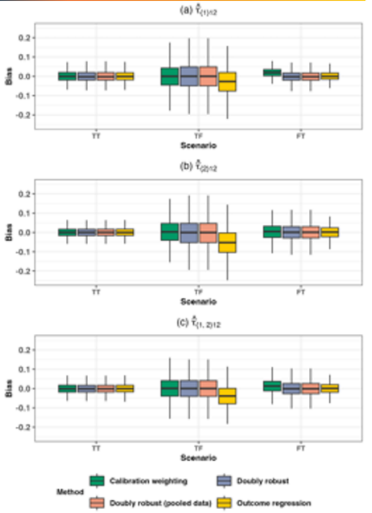
4. For $i, j = 1, \ldots, K$ and $\mathcal{L} \subseteq \{1, \ldots, K\}$, the cloud server calculates the doubly robust estimator $\hat{\tau}^{\mathrm{dr}}_{\mathcal{L}ij}$ of the ATEs and the estimate of its variance.

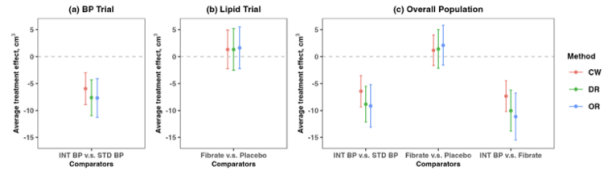
$$\hat{\tau}_{\mathcal{I}ij} = \sum_{l \in \mathcal{I}} \frac{n_l}{\sum_{l \in \mathcal{I}} n_l} (A^1_{ij} - A^1_{li} + A^2_{jl} - A^2_{il}).$$

Contact: Yuru.Zhu@Pennmedicine.upenn.edu, ychen123@pennmedicine.upenn.edu

Results

- The bias boxplots display simulation results. The results of the DR estimators calculated by aggregated data and by pooled data are exactly same. The DR estimators have very small bias when either outcome models or propensity score models are correctly specified.
- **Data application:** We use the data in ACCORD MIND trial to perform empirical analysis. We investigate the pairwise ATEs between four interventions (standard blood pressure (BP), intensive BP, lipid placebo and lipid fibrate) on the 40-month changes of total brain volume (TBV) from baseline in the BP trial, lipid trial and the overall population.
- At 40 months, TBV had declined more under intensive BP compared to the standard BP, with the DR estimate of ATE in the overall population being -8.8 [95% CI, -12.2 to -5.5] cm^3 ($P = 2\times10^{-7}$). Fibrate therapy had no effect on TBV declines compared with placebo. The conclusions are consistent with those in Williamson et al. (2014).





Conclusions

- Our proposed DR estimators for indirect treatment comparison in the DRN show robustness when either outcome models or propensity score models are correctly specified.
- The developed algorithm which only requires aggregated data has three communications and is lossless.

Reference

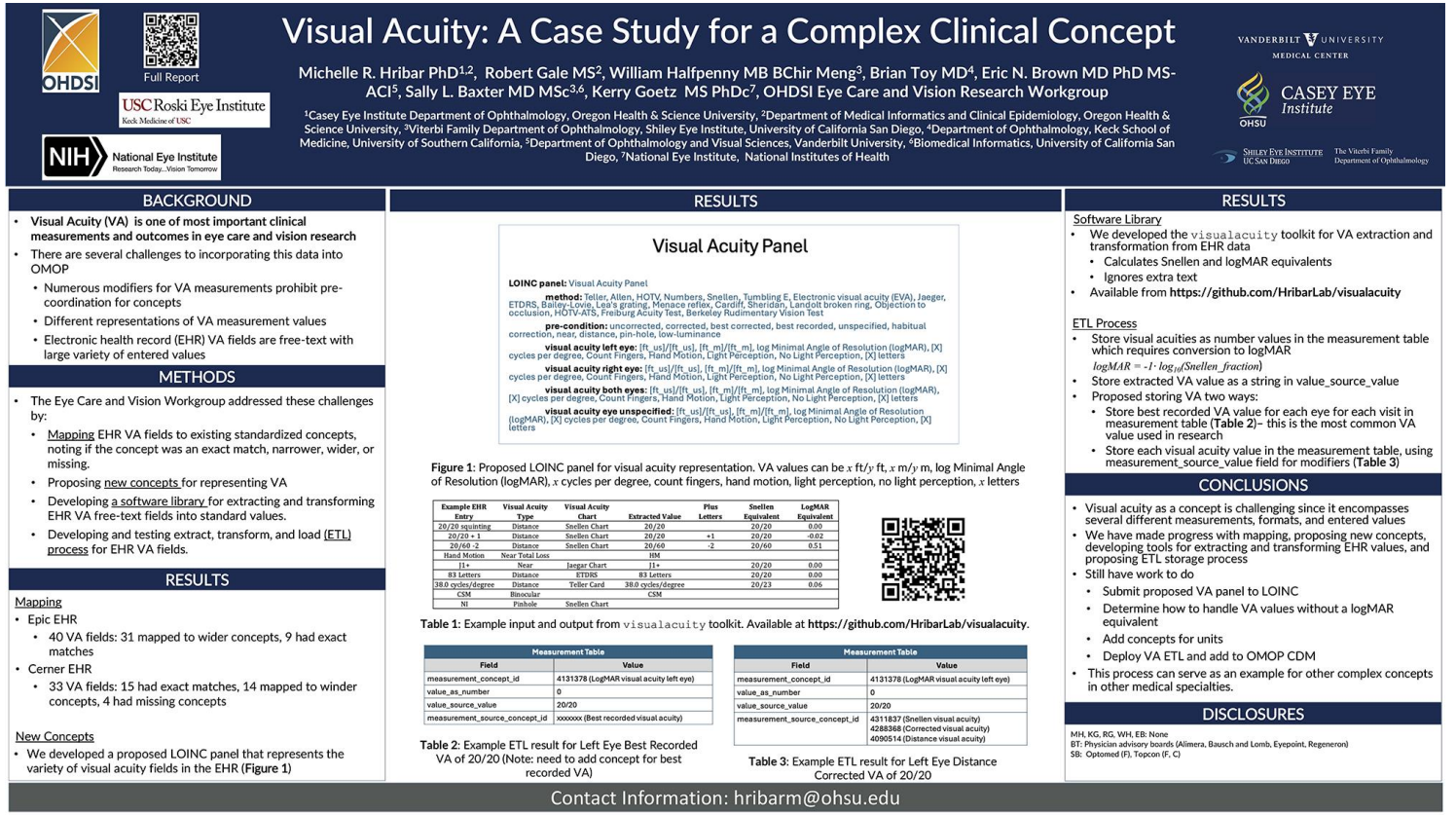
- Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political analysis, 20(1):25–46.
- Williamson, J. D. et al. (2014). Cognitive function and brain structure in persons with type 2 diabetes mellitus after intensive lowering of blood pressure and lipid levels: a randomized clinical trial. JAMA internal medicine, 174(3):324–333.

#OHDSISocialShowcase This Week

Friday

Visual Acuity: A Case Study for a Complex Clinical Concept

(Michelle R. Hribar, Robert Gale, William Halfpenny, Brian Toy, Eric N. Brown, Sally L. Baxter, Kerry Goetz, OHDSI Eye Care and Vision Research Workgroup)



Visual Acuity: A Case Study for a Complex Clinical Concept

Full Report

Michelle R. Hribar PhD[1,2], Robert Gale MS[2], William Halfpenny MB BChir Meng[3], Brian Toy MD[4], Eric N. Brown MD PhD MS-ACI[5], Sally L. Baxter MD MSc[3,6], Kerry Goetz MS PhDc[7], OHDSI Eye Care and Vision Research Workgroup

[1]Casey Eye Institute Department of Ophthalmology, Oregon Health & Science University, [2]Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, [3]Viterbi Family Department of Ophthalmology, Shiley Eye Institute, University of California San Diego, [4]Department of Ophthalmology, Keck School of Medicine, University of Southern California, [5]Department of Ophthalmology and Visual Sciences, Vanderbilt University, [6]Biomedical Informatics, University of California San Diego, [7]National Eye Institute, National Institutes of Health

BACKGROUND

- **Visual Acuity (VA) is one of most important clinical measurements and outcomes in eye care and vision research**
- There are several challenges to incorporating this data into OMOP
  - Numerous modifiers for VA measurements prohibit pre-coordination for concepts
  - Different representations of VA measurement values
  - Electronic health record (EHR) VA fields are free-text with large variety of entered values

METHODS

- The Eye Care and Vision Workgroup addressed these challenges by:
  - Mapping EHR VA fields to existing standardized concepts, noting if the concept was an exact match, narrower, wider, or missing.
  - Proposing new concepts for representing VA
  - Developing a software library for extracting and transforming EHR VA free-text fields into standard values.
  - Developing and testing extract, transform, and load (ETL) process for EHR VA fields.

RESULTS

Mapping

- Epic EHR
  - 40 VA fields: 31 mapped to wider concepts, 9 had exact matches
- Cerner EHR
  - 33 VA fields: 15 had exact matches, 14 mapped to winder concepts, 4 had missing concepts

New Concepts

- We developed a proposed LOINC panel that represents the variety of visual acuity fields in the EHR (**Figure 1**)

RESULTS

Visual Acuity Panel

LOINC panel: Visual Acuity Panel

method: Teller, Allen, HOTV, Numbers, Snellen, Tumbling E, Electronic visual acuity (EVA), Jaeger, ETDRS, Bailey-Lovie, Lea's grating, Menace reflex, Cardiff, Sheridan, Landolt broken ring, Objection to occlusion, HOTV-ATS, Freiburg Acuity Test, Berkeley Rudimentary Vision Test

pre-condition: uncorrected, corrected, best corrected, best recorded, unspecified, habitual correction, near, distance, pin-hole, low-luminance

visual acuity left eye: [ft_us]/[ft_us], [ft_m]/[ft_m], log Minimal Angle of Resolution (logMAR), [X] cycles per degree, Count Fingers, Hand Motion, Light Perception, No Light Perception, [X] letters

visual acuity right eye: [ft_us]/[ft_us], [ft_m]/[ft_m], log Minimal Angle of Resolution (logMAR), [X] cycles per degree, Count Fingers, Hand Motion, Light Perception, No Light Perception, [X] letters

visual acuity both eyes: [ft_us]/[ft_us], [ft_m]/[ft_m], log Minimal Angle of Resolution (logMAR), [X] cycles per degree, Count Fingers, Hand Motion, Light Perception, No Light Perception, [X] letters

visual acuity eye unspecified: [ft_us]/[ft_us], [ft_m]/[ft_m], log Minimal Angle of Resolution (logMAR), [X] cycles per degree, Count Fingers, Hand Motion, Light Perception, No Light Perception, [X] letters

Figure 1: Proposed LOINC panel for visual acuity representation. VA values can be x ft/y ft, x m/y m, log Minimal Angle of Resolution (logMAR), x cycles per degree, count fingers, hand motion, light perception, no light perception, x letters

Example EHR Entry	Visual Acuity Type	Visual Acuity Chart	Extracted Value	Plus Letters	Snellen Equivalent	LogMAR Equivalent
20/20 squinting	Distance	Snellen Chart	20/20		20/20	0.00
20/20 + 1	Distance	Snellen Chart	20/20	+1	20/20	-0.02
20/60 -2	Distance	Snellen Chart	20/60	-2	20/60	0.51
Hand Motion	Near Total Loss		HM			
J1+	Near	Jaegar Chart	J1+		20/20	0.00
83 Letters	Distance	ETDRS	83 Letters		20/20	0.00
38.0 cycles/degree	Distance	Teller Card	38.0 cycles/degree		20/23	0.06
CSM	Binocular		CSM			
NI	Pinhole	Snellen Chart				

Table 1: Example input and output from `visualacuity` toolkit. Available at **https://github.com/HribarLab/visualacuity**.

Measurement Table	
Field	Value
measurement_concept_id	4131378 (LogMAR visual acuity left eye)
value_as_number	0
value_source_value	20/20
measurement_source_concept_id	xxxxxxx (Best recorded visual acuity)

Table 2: Example ETL result for Left Eye Best Recorded VA of 20/20 (Note: need to add concept for best recorded VA)

Measurement Table	
Field	Value
measurement_concept_id	4131378 (LogMAR visual acuity left eye)
value_as_number	0
value_source_value	20/20
measurement_source_concept_id	4311837 (Snellen visual acuity) 4288368 (Corrected visual acuity) 4090514 (Distance visual acuity)

Table 3: Example ETL result for Left Eye Distance Corrected VA of 20/20

RESULTS

Software Library

- We developed the `visualacuity` toolkit for VA extraction and transformation from EHR data
  - Calculates Snellen and logMAR equivalents
  - Ignores extra text
- Available from **https://github.com/HribarLab/visualacuity**

ETL Process

- Store visual acuities as number values in the measurement table which requires conversion to logMAR
  $logMAR = -1 \cdot \log_{10}(Snellen\_fraction)$
- Store extracted VA value as a string in value_source_value
- Proposed storing VA two ways:
  - Store best recorded VA value for each eye for each visit in measurement table (**Table 2**)– this is the most common VA value used in research
  - Store each visual acuity value in the measurement table, using measurement_source_value field for modifiers (**Table 3**)

CONCLUSIONS

- Visual acuity as a concept is challenging since it encompasses several different measurements, formats, and entered values
- We have made progress with mapping, proposing new concepts, developing tools for extracting and transforming EHR values, and proposing ETL storage process
- Still have work to do
  - Submit proposed VA panel to LOINC
  - Determine how to handle VA values without a logMAR equivalent
  - Add concepts for units
  - Deploy VA ETL and add to OMOP CDM
- This process can serve as an example for other complex concepts in other medical specialties.

DISCLOSURES

MH, KG, RG, WH, EB: None
BT: Physician advisory boards (Alimera, Bausch and Lomb, Eyepoint, Regeneron)
SB: Optomed (F), Topcon (F, C)

Contact Information: hribarm@ohsu.edu

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?

The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

Links are sent out weekly and available at:
ohdsi.org/community-calls-2025