Innovation and Education Brainstorm

OHDSI Community Call

Jan. 27, 2026 • 11 am ET

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Upcoming Community Calls

Date	Topic
Jan. 27	Education and Innovation Brainstorm
Feb. 3	2026 Workgroup Objectives & Key Results, Part 1
Feb 10	2026 Workgroup Objectives & Key Results, Part 2

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **George Hripcsak, Marc Suchard, Martijn Schuemie and Patrick Ryan** on the recent publication of **Trust in Observational Research** in the *Journal of American College of Cardiology (JACC).*

ARTICLE IN PRESS

JACC VOL. ■, NO. ■, 2026

VIEWPOINT

Trust in Observational Research

George Hripcsak, MD, MS,[a] Marc A. Suchard, MD, PhD,[b,c] Martijn J. Schuemie, PhD,[d] Patrick B. Ryan, PhD[d]

Observational research promises to add evidence where randomized trials are underpowered, unethical, unrealistic, or—most often—simply too resource intensive to meet the huge demand. Clinical guidelines that are intended to be evidence-based are forced to rely on expert opinion,[1] and of the tens of thousands of side effects that could result from each of the thousands of drugs on the market, only a fraction of these potential side effects has been formally studied. The emergence of enormous databases, increasing computing power, and new analytic methods should propel observational research forward, but researchers and the public have become increasingly aware of the unreliability of published real-world evidence.[2,3]

For observational research to reach its potential, we must recognize where the challenges are and take concrete steps to address them. We focus here on observational research that tests clinical hypotheses with an intent to publish in top clinical journals and affect the care of millions of patients. We recognize that other observational research is hypothesis generating and may be treated differently.

Observational research can be roughly divided into several components: picking an appropriate and well-formulated hypothesis, developing a proper study design, accessing data that are relevant and accurate, executing the study rigorously, and interpreting the results correctly. Several initiatives have identified concrete steps we can take and criteria we can use to improve reliability. For example, the Observational Health Data Sciences and Informatics (OHDSI) Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) framework[4] identifies 10 criteria for improving the reliability of observational research and was used in published cardiovascular safety and effectiveness studies.[5] The criteria can be grouped as openness and verification, shown in Figure 1. Openness is well-known but poorly followed. The study protocol must be prespecified and published, all study software and clinical definitions must be made available,

> *"The study protocol must be prespecified and published, all study software and clinical definitions must be made available, all diagnostics must be shared before unblinding the results, and all results must be published in some format."*

all diagnostics must be shared before unblinding the results, and all results must be published in some format. Verification is through formal diagnostics that assess every component of the process, including data quality, accuracy of the definitions of outcomes and key covariates, analytic diagnostics such as achieved balance in confounding adjustment, a test for consistency among multiple databases, and large-scale negative control subjects to substantiate the claim of minimal bias. The Sentinel Initiative provides an overlapping set of system requirements, focusing on openness, data properties, and standardized tools.[6] It has been applied extensively in the work of the U.S.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Wednesday	9 am	Oncology Outreach/Research Subgroup
Wednesday	10 am	Surgery and Perioperative Medicine
Wednesday	10 am	Women of OHDSI
Wednesday	12 pm	Latin America
Thursday	9 am	Africa Chapter
Thursday	11 am	Perinatal and Reproductive Health
Friday	10 am	GIS – Geographic Information System
Friday	11:30 am	Steering
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Healthcare Systems Interest Group
Tuesday	10 am	CDM Survey

2026 Global Symposium

2026 OHDSI Global Symposium Call for Plenary Sessions

Symposium plenaries provide opportunities to share innovative, community-developed content to empower researchers to generate reliable real-world evidence. The community is currently seeking proposals for our #OHDSI2026 plenaries. These sessions will be 60 minutes in duration and must touch on at least two of following pillars of our community:

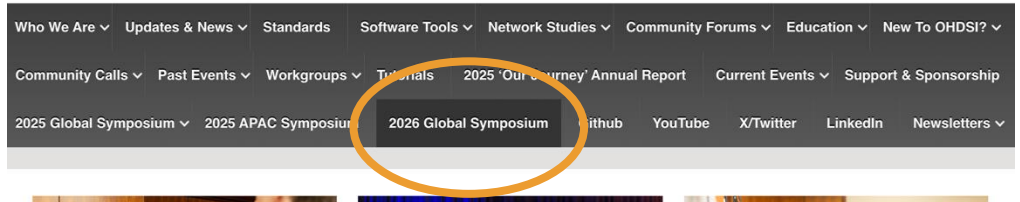
- Open community data standards
- Methodological research
- Open-source development
- Clinical applications

Plenary sessions must also involve three or more on-stage participants across at least two organizations. Sessions may include a combination of keynote talks, panel discussions, interactive activities, and more. We strongly encourage using multiple formats and synthesizing completed research, current perspectives and future calls-to-action to maximize community engagement.

The deadline for proposal submissions is January 30, 2026. Please use the link below to submit your proposal by answering the following questions:

- Name(s) of plenary session organizers:
- Your email address(es):
- Short (2,500 character max) description / abstract of your proposed session:
- Which pillars are you targeting:
- One sentence “pitch” of your session to excite the community:
- Names and roles of individuals who have tentatively agreed to participate in your session:

Deadline to submit proposals for #OHDSI2026 plenaries or tutorials is Jan. 30, 2026!



2026 OHDSI Global Symposium

Oct. 20-22 • New Brunswick, N.J. • Hyatt Regency Hotel

2026 OHDSI Global Symposium Call for Tutorials

Tutorial sessions aim to deliver educational content, led by community members who wish to train our global collaborators on scientific, technical, and other skills that can support advancing OHDSI’s mission and the effective use of real-world data and the generation and dissemination of reliable real-world evidence. Examples of prior tutorials offered are provided here: https://www.ohdsi.org/tutorials.

Tutorial sessions are 4 hours in duration. Registrants for your tutorial will be requested to pay a registration fee. The fees will be used to offset the costs of the symposium and other OHDSI expenses. Sessions may include a combination of talks, interactive activities, and more. We strongly encourage using multiple formats to maximize community engagement. Your session must include at least three people from at least two different organizations.

The deadline for tutorial proposal submissions is January 30, 2026. Please use the link below to submit your proposal by answering the following questions:

- Name(s) of tutorial session organizers:
- Your email address(es):
- Short (2,500 character) description / abstract of your proposed session:
- Names and roles of individuals who have tentatively agreed to participate in your session:

2026 Europe Symposium

The 2026 OHDSI Europe Symposium returns to Rotterdam next year and will be held **April 18-20.**

The deadline for abstract submissions will be Feb. 6, 2026.

2026 Europe Symposium

Time	Symposium Agenda - Monday April 20, 2026	Location
8:00	***Registration and Coffee***	Queen's Lounge
9:00	**Welcome to OHDSI Europe** Dr. Renske Los, Department of Medical Informatics, *Erasmus MC* Dr. Aniek Markus, Department of Medical Informatics, *Erasmus MC*	Theatre
9:05	**Journey of OHDSI** Prof. Peter Rijnbeek, Chair Department of Medical Informatics, *Erasmus MC*	Theatre
9:30	**Collaborator Showcase - part 1** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC*	
10:00	**Speed networking**	
10:15	***Coffee Break & posters National Nodes***	Que
11:15	**Collaborator Showcase - part 2** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC*	
11:45	**Dreaming about the OHDSI journey ahead** Dr. Patrick Ryan, Vice President, Observational Health Data Analytics, *Johnson & Johnson* Dr. Renske Los, Department of Medical Informatics, *Erasmus MC*	
12:15	***Lunch break & networking & posters/demo's*** *(Early investigator meeting - 13:00-13:45 Queen's Lounge)*	La Fontaine & Odyssee Room
13:45	**From dreams to reality** OHDSI Titan Award winners	Theatre
14:30	**Propositions for collaboration from the National Nodes** National Node leads	Theatre
14:45	***Coffee break & posters/demo's***	La Fontaine & Odyssee Room
16:15	**The OH Factor** To be announced	Theatre
17:00	**Closing**	Theatre
17:15	***Networking reception***	Queen's Lounge

ATHENA Survey

Athena user survey

Help us understand how to make Athena better

When you submit this form, it will not automatically collect your details like name and email address unless you provide it yourself.

* Required

1. If you are open to follow-up about your feedback, please provide your email address

Enter your answer

2. How do you use athena.ohdsi.org? *

- [ ] Search concepts
- [ ] Download current version of vocabularies
- [ ] Download previous version of vocabularies
- [ ] Other

3. What do you athena.ohdsi.org for? *

- [ ] ETL data
- [ ] Search concepts to create mappings
- [ ] Search concepts for concept sets (value sets, code lists)
- [ ] Translate concepts to other languages/find translations
- [ ] Use as knowledge graph outside of OMOP CDM
- [ ] Other

4. On average, how often do you access athena.ohdsi.org? *

- ◯ Every day
- ◯ Once a week
- ◯ A few times a month
- ◯ Once a month
- ◯ Once 6 month or less
- ◯ Other

5. If there was an Athena API, how would you use it and what would you use it for? *

Enter your answer

6. Anything else you'd like to tell us? *

Enter your answer

You can print a copy of your answer after you submit

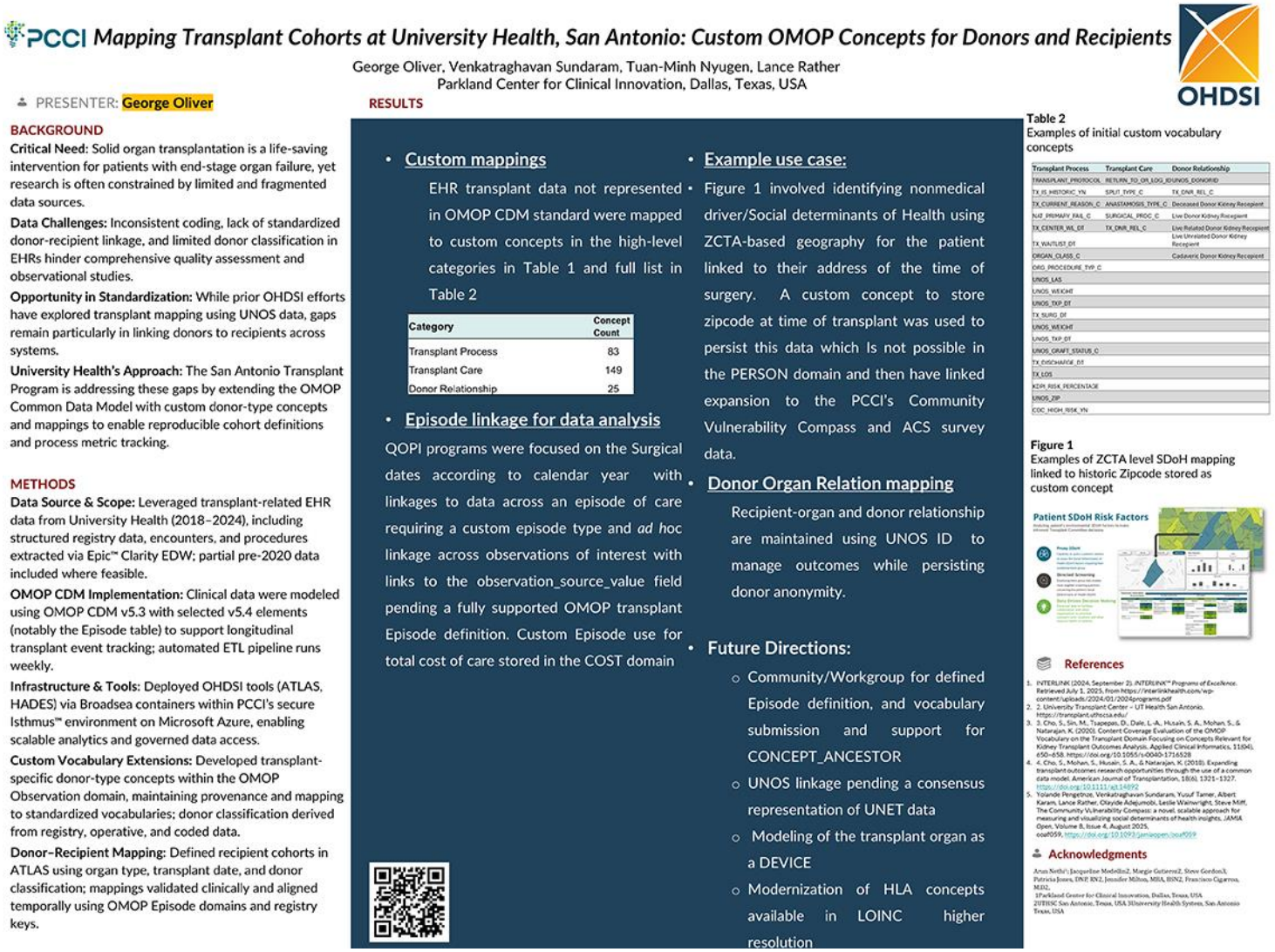
Submit

#OHDSISocialShowcase This Week

Monday

Mapping Transplant Cohorts at University Health, San Antonio: Custom OMOP Concepts for Donors and Recipients

(George "Holt" Oliver, Venkatraghavan Sundaram, Tuan-Minh Nguyen, Steve Gordon, Jacqueline Medellin, Margie Gutierez, Patricia Jones, Jennifer Milton, Francisco Cigarroa, Lance Rather)



PCCI *Mapping Transplant Cohorts at University Health, San Antonio: Custom OMOP Concepts for Donors and Recipients*

George Oliver, Venkatraghavan Sundaram, Tuan-Minh Nyugen, Lance Rather
Parkland Center for Clinical Innovation, Dallas, Texas, USA

OHDSI

PRESENTER: George Oliver

BACKGROUND

Critical Need: Solid organ transplantation is a life-saving intervention for patients with end-stage organ failure, yet research is often constrained by limited and fragmented data sources.

Data Challenges: Inconsistent coding, lack of standardized donor-recipient linkage, and limited donor classification in EHRs hinder comprehensive quality assessment and observational studies.

Opportunity in Standardization: While prior OHDSI efforts have explored transplant mapping using UNOS data, gaps remain particularly in linking donors to recipients across systems.

University Health's Approach: The San Antonio Transplant Program is addressing these gaps by extending the OMOP Common Data Model with custom donor-type concepts and mappings to enable reproducible cohort definitions and process metric tracking.

METHODS

Data Source & Scope: Leveraged transplant-related EHR data from University Health (2018–2024), including structured registry data, encounters, and procedures extracted via Epic™ Clarity EDW; partial pre-2020 data included where feasible.

OMOP CDM Implementation: Clinical data were modeled using OMOP CDM v5.3 with selected v5.4 elements (notably the Episode table) to support longitudinal transplant event tracking; automated ETL pipeline runs weekly.

Infrastructure & Tools: Deployed OHDSI tools (ATLAS, HADES) via Broadsea containers within PCCI's secure Isthmus™ environment on Microsoft Azure, enabling scalable analytics and governed data access.

Custom Vocabulary Extensions: Developed transplant-specific donor-type concepts within the OMOP Observation domain, maintaining provenance and mapping to standardized vocabularies; donor classification derived from registry, operative, and coded data.

Donor–Recipient Mapping: Defined recipient cohorts in ATLAS using organ type, transplant date, and donor classification; mappings validated clinically and aligned temporally using OMOP Episode domains and registry keys.

RESULTS

- **Custom mappings**

  EHR transplant data not represented in OMOP CDM standard were mapped to custom concepts in the high-level categories in Table 1 and full list in Table 2

Category	Concept Count
Transplant Process	83
Transplant Care	149
Donor Relationship	25

- **Episode linkage for data analysis**

QOPI programs were focused on the Surgical dates according to calendar year with linkages to data across an episode of care requiring a custom episode type and *ad hoc* linkage across observations of interest with links to the observation_source_value field pending a fully supported OMOP transplant Episode definition. Custom Episode use for total cost of care stored in the COST domain

- **Example use case:**

Figure 1 involved identifying nonmedical driver/Social determinants of Health using ZCTA-based geography for the patient linked to their address of the time of surgery. A custom concept to store zipcode at time of transplant was used to persist this data which Is not possible in the PERSON domain and then have linked expansion to the PCCI's Community Vulnerability Compass and ACS survey data.

- **Donor Organ Relation mapping**

Recipient-organ and donor relationship are maintained using UNOS ID to manage outcomes while persisting donor anonymity.

- **Future Directions:**
  - Community/Workgroup for defined Episode definition, and vocabulary submission and support for CONCEPT_ANCESTOR
  - UNOS linkage pending a consensus representation of UNET data
  - Modeling of the transplant organ as a DEVICE
  - Modernization of HLA concepts available in LOINC higher resolution

Table 2
Examples of initial custom vocabulary concepts

Transplant Process	Transplant Care	Donor Relationship
TRANSPLANT_PROTOCOL	RETURN_TO_OR_LOG_IDUNOS_DONORID	
TX_IS_HISTORIC_YN	SPLIT_TYPE_C	TX_DNR_REL_C
TX_CURRENT_REASON_C	ANASTAMOSIS_TYPE_C	Deceased Donor Kidney Recepient
NUI_PRIMARY_FAIL_C	SURGICAL_PROC_C	Live Donor Kidney Recepient
TX_CENTER_WL_DT	TX_DNR_REL_C	Live Related Donor Kidney Recepient
TX_WAITLIST_DT		Live Unrelated Donor Kidney Recepient
ORGAN_CLASS_C		Cadaveric Donor Kidney Recepient
ORG_PROCEDURE_TYP_C		
UNOS_LAS		
UNOS_WEIGHT		
UNOS_TXP_DT		
TX_SURG_DT		
UNOS_WEIGHT		
UNOS_TXP_DT		
UNOS_GRAFT_STATUS_C		
TX_DISCHARGE_DT		
TX_LOS		
KDPI_RISK_PERCENTAGE		
UNOS_ZIP		
CDC_HIGH_RISK_YN		

Figure 1
Examples of ZCTA level SDoH mapping linked to historic Zipcode stored as custom concept

References

1. INTERLINK (2024, September 2). INTERLINK™ Programs of Excellence. Retrieved July 1, 2025, from https://interlinkhealth.com/wp-content/uploads/2024/01/2024programs.pdf
2. 2. University Transplant Center – UT Health San Antonio. https://transplant.uthscsa.edu/
3. 3. Cho, S., Sin, M., Tsapepas, D., Dale, L.-A., Husain, S. A., Mohan, S., & Natarajan, K. (2020). Content Coverage Evaluation of the OMOP Vocabulary on the Transplant Domain Focusing on Concepts Relevant for Kidney Transplant Outcomes Analysis. Applied Clinical Informatics, 11(04), 650–658. https://doi.org/10.1055/s-0040-1716528
4. 4. Cho, S., Mohan, S., Husain, S. A., & Natarajan, K. (2018). Expanding transplant outcomes research opportunities through the use of a common data model. American Journal of Transplantation, 18(6), 1321–1327. https://doi.org/10.1111/ajt.14692
5. Yolande Pengetnze, Venkatraghavan Sundaram, Yusuf Tamer, Albert Karam, Lance Rather, Olayide Adejumobi, Leslie Wainwright, Steve Mitt, The Community Vulnerability Compass: a novel, scalable approach for measuring and visualizing social determinants of health insights, JAMIA Open, Volume 8, Issue 4, August 2025, ooaf059, https://doi.org/10.1093/jamiaopen/ooaf059

Acknowledgments

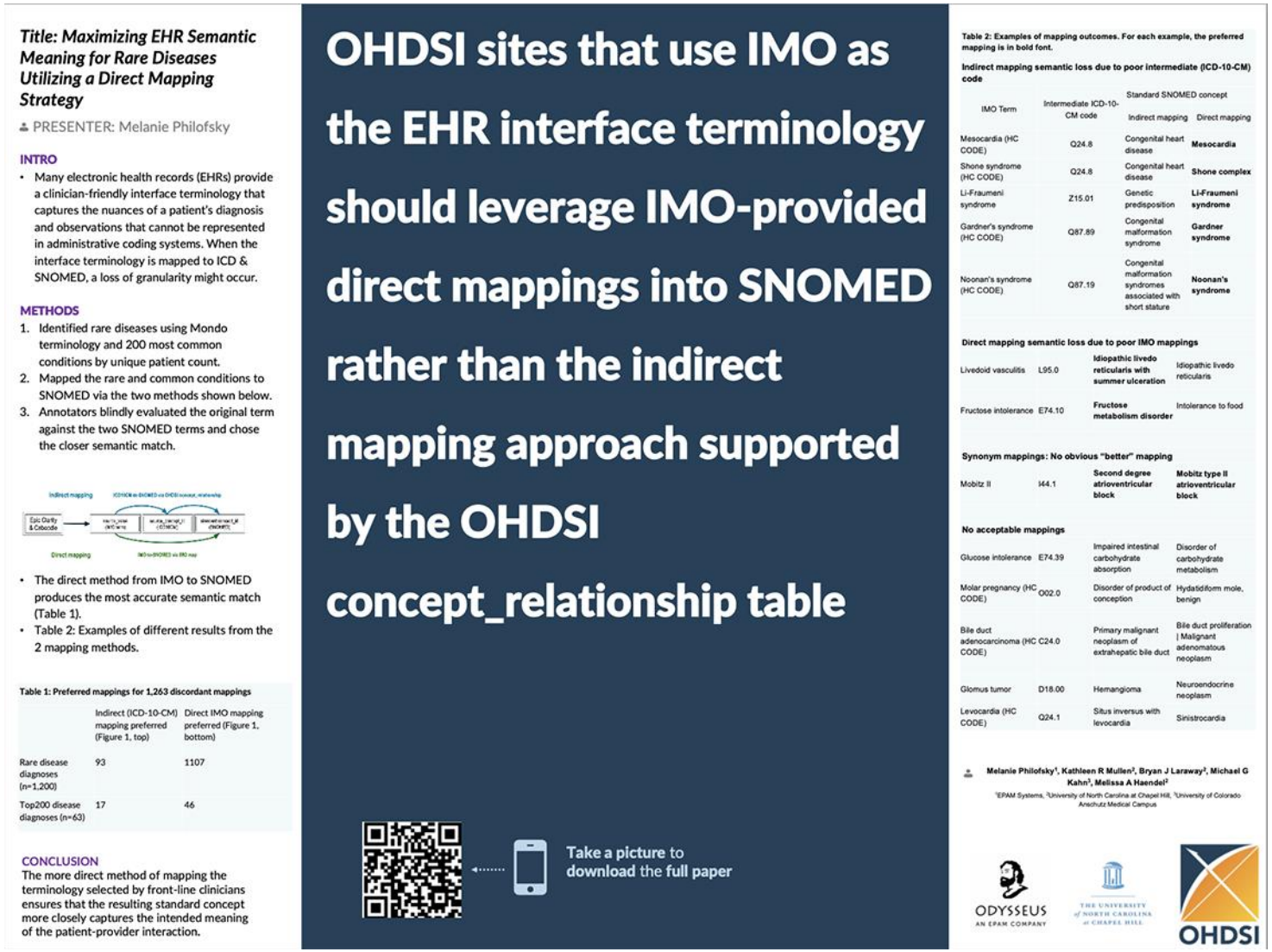
Arun Nethi[1], Jacqueline Medellin[2], Margie Gutierrez[2], Steve Gordon[3], Patricia Jones, DNP, RN[2], Jennifer Milton, MBA, BSN[2], Francisco Cigarroa, M.D.[2],
[1]Parkland Center for Clinical Innovation, Dallas, Texas, USA
[2]UTHSC San Antonio, Texas, USA [3]University Health System, San Antonio Texas, USA

#OHDSISocialShowcase This Week

Tuesday

Maximizing EHR Semantic Meaning for Rare Diseases Utilizing a Direct Mapping Strategy

(Melanie Philofsky, Kathleen R Mullen, Bryan J Laraway, Michael G Kahn, Melissa A Haendel)



Title: Maximizing EHR Semantic Meaning for Rare Diseases Utilizing a Direct Mapping Strategy

PRESENTER: Melanie Philofsky

INTRO

- Many electronic health records (EHRs) provide a clinician-friendly interface terminology that captures the nuances of a patient's diagnosis and observations that cannot be represented in administrative coding systems. When the interface terminology is mapped to ICD & SNOMED, a loss of granularity might occur.

METHODS

1. Identified rare diseases using Mondo terminology and 200 most common conditions by unique patient count.
2. Mapped the rare and common conditions to SNOMED via the two methods shown below.
3. Annotators blindly evaluated the original term against the two SNOMED terms and chose the closer semantic match.

- The direct method from IMO to SNOMED produces the most accurate semantic match (Table 1).
- Table 2: Examples of different results from the 2 mapping methods.

Table 1: Preferred mappings for 1,263 discordant mappings

	Indirect (ICD-10-CM) mapping preferred (Figure 1, top)	Direct IMO mapping preferred (Figure 1, bottom)
Rare disease diagnoses (n=1,200)	93	1107
Top200 disease diagnoses (n=63)	17	46

CONCLUSION

The more direct method of mapping the terminology selected by front-line clinicians ensures that the resulting standard concept more closely captures the intended meaning of the patient-provider interaction.

OHDSI sites that use IMO as the EHR interface terminology should leverage IMO-provided direct mappings into SNOMED rather than the indirect mapping approach supported by the OHDSI concept_relationship table

Take a picture to download the full paper

Table 2: Examples of mapping outcomes. For each example, the preferred mapping is in bold font.

Indirect mapping semantic loss due to poor intermediate (ICD-10-CM) code

IMO Term	Intermediate ICD-10-CM code	Standard SNOMED concept: Indirect mapping	Standard SNOMED concept: Direct mapping
Mesocardia (HC CODE)	Q24.8	Congenital heart disease	**Mesocardia**
Shone syndrome (HC CODE)	Q24.8	Congenital heart disease	**Shone complex**
Li-Fraumeni syndrome	Z15.01	Genetic predisposition	**Li-Fraumeni syndrome**
Gardner's syndrome (HC CODE)	Q87.89	Congenital malformation syndrome	**Gardner syndrome**
Noonan's syndrome (HC CODE)	Q87.19	Congenital malformation syndromes associated with short stature	**Noonan's syndrome**

Direct mapping semantic loss due to poor IMO mappings

Livedoid vasculitis	L95.0	**Idiopathic livedo reticularis with summer ulceration**	Idiopathic livedo reticularis
Fructose intolerance	E74.10	**Fructose metabolism disorder**	Intolerance to food

Synonym mappings: No obvious "better" mapping

Mobitz II	I44.1	**Second degree atrioventricular block**	**Mobitz type II atrioventricular block**

No acceptable mappings

Glucose intolerance	E74.39	Impaired intestinal carbohydrate absorption	Disorder of carbohydrate metabolism
Molar pregnancy (HC CODE)	O02.0	Disorder of product of conception	Hydatidiform mole, benign
Bile duct adenocarcinoma (HC CODE)	C24.0	Primary malignant neoplasm of extrahepatic bile duct	Bile duct proliferation | Malignant adenomatous neoplasm
Glomus tumor	D18.00	Hemangioma	Neuroendocrine neoplasm
Levocardia (HC CODE)	Q24.1	Situs inversus with levocardia	Sinistrocardia

Melanie Philofsky[1], Kathleen R Mullen[2], Bryan J Laraway[2], Michael G Kahn[3], Melissa A Haendel[2]
[1]EPAM Systems, [2]University of North Carolina at Chapel Hill, [3]University of Colorado Anschutz Medical Campus

ODYSSEUS AN EPAM COMPANY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

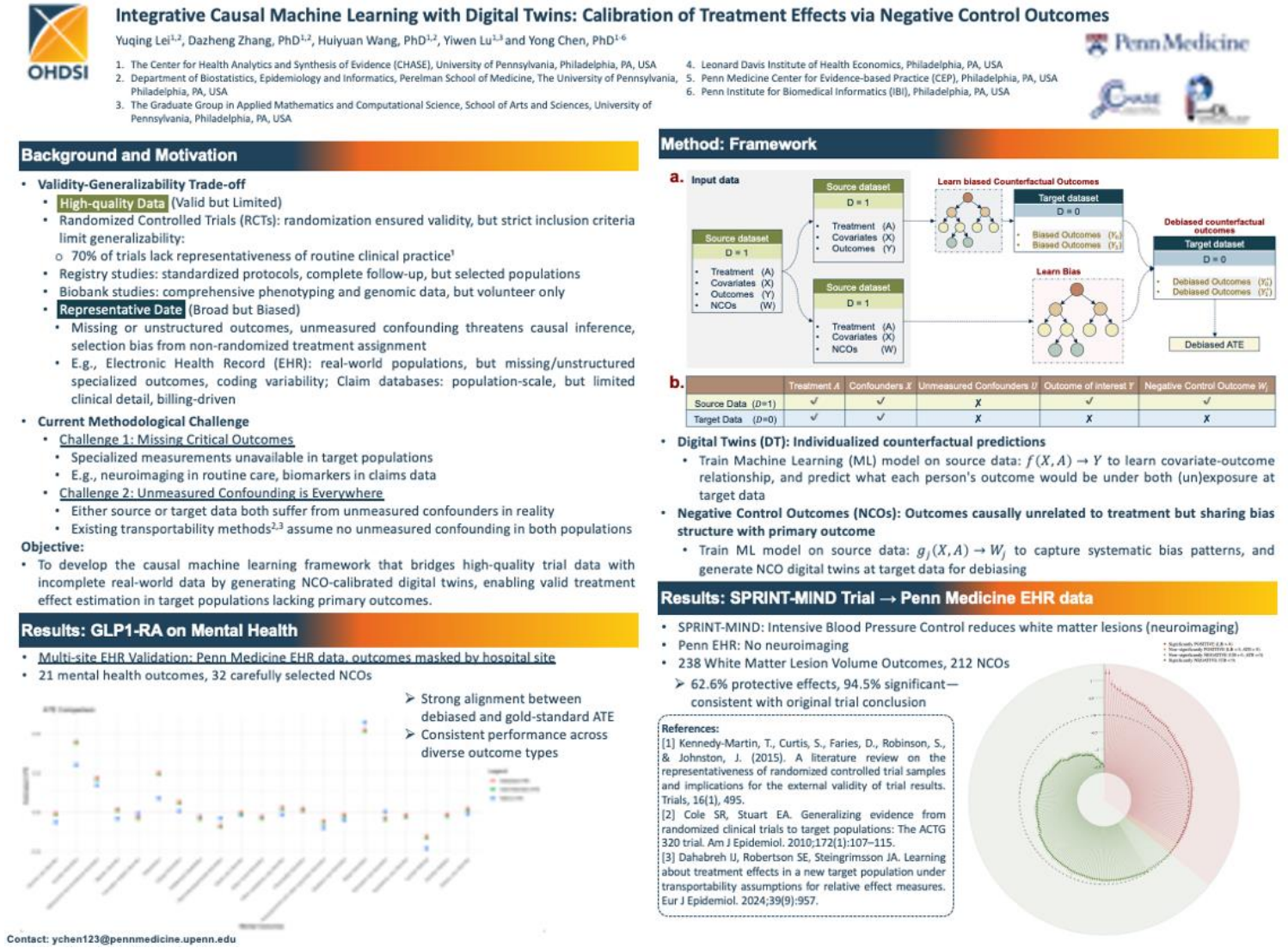
OHDSI

#OHDSISocialShowcase This Week

Wednesday

Integrative Causal Machine Learning with Digital Twins: Calibration of Treatment Effects via Negative Control Outcomes

(Yuqing Lei, Huiyuan Wang, Dazheng Zhang, Yiwen Lu, Yong Chen)



Integrative Causal Machine Learning with Digital Twins: Calibration of Treatment Effects via Negative Control Outcomes

Yuqing Lei[1,2], Dazheng Zhang, PhD[1,2], Huiyuan Wang, PhD[1,2], Yiwen Lu[1,3] and Yong Chen, PhD[1-6]

1. The Center for Health Analytics and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA
2. Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA, USA
3. The Graduate Group in Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA
4. Leonard Davis Institute of Health Economics, Philadelphia, PA, USA
5. Penn Medicine Center for Evidence-based Practice (CEP), Philadelphia, PA, USA
6. Penn Institute for Biomedical Informatics (IBI), Philadelphia, PA, USA

Background and Motivation

- **Validity-Generalizability Trade-off**
  - High-quality Data (Valid but Limited)
  - Randomized Controlled Trials (RCTs): randomization ensured validity, but strict inclusion criteria limit generalizability:
    - 70% of trials lack representativeness of routine clinical practice[1]
  - Registry studies: standardized protocols, complete follow-up, but selected populations
  - Biobank studies: comprehensive phenotyping and genomic data, but volunteer only
  - Representative Date (Broad but Biased)
    - Missing or unstructured outcomes, unmeasured confounding threatens causal inference, selection bias from non-randomized treatment assignment
    - E.g., Electronic Health Record (EHR): real-world populations, but missing/unstructured specialized outcomes, coding variability; Claim databases: population-scale, but limited clinical detail, billing-driven
- **Current Methodological Challenge**
  - Challenge 1: Missing Critical Outcomes
    - Specialized measurements unavailable in target populations
    - E.g., neuroimaging in routine care, biomarkers in claims data
  - Challenge 2: Unmeasured Confounding is Everywhere
    - Either source or target data both suffer from unmeasured confounders in reality
    - Existing transportability methods[2,3] assume no unmeasured confounding in both populations

Objective:

- To develop the causal machine learning framework that bridges high-quality trial data with incomplete real-world data by generating NCO-calibrated digital twins, enabling valid treatment effect estimation in target populations lacking primary outcomes.

Results: GLP1-RA on Mental Health

- Multi-site EHR Validation: Penn Medicine EHR data, outcomes masked by hospital site
- 21 mental health outcomes, 32 carefully selected NCOs

➢ Strong alignment between debiased and gold-standard ATE
➢ Consistent performance across diverse outcome types

Contact: ychen123@pennmedicine.upenn.edu

Method: Framework

a. Input data

b.

	Treatment A	Confounders X	Unmeasured Confounders U	Outcome of interest Y	Negative Control Outcome W_j
Source Data (D=1)	✓	✓	✗	✓	✓
Target Data (D=0)	✓	✓	✗	✗	✗

- **Digital Twins (DT): Individualized counterfactual predictions**
  - Train Machine Learning (ML) model on source data: $f(X, A) \rightarrow Y$ to learn covariate-outcome relationship, and predict what each person's outcome would be under both (un)exposure at target data
- **Negative Control Outcomes (NCOs): Outcomes causally unrelated to treatment but sharing bias structure with primary outcome**
  - Train ML model on source data: $g_j(X, A) \rightarrow W_j$ to capture systematic bias patterns, and generate NCO digital twins at target data for debiasing

Results: SPRINT-MIND Trial → Penn Medicine EHR data

- SPRINT-MIND: Intensive Blood Pressure Control reduces white matter lesions (neuroimaging)
- Penn EHR: No neuroimaging
- 238 White Matter Lesion Volume Outcomes, 212 NCOs
  - ➢ 62.6% protective effects, 94.5% significant—consistent with original trial conclusion

References:
[1] Kennedy-Martin, T., Curtis, S., Faries, D., Robinson, S., & Johnston, J. (2015). A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials, 16(1), 495.
[2] Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. Am J Epidemiol. 2010;172(1):107–115.
[3] Dahabreh IJ, Robertson SE, Steingrimsson JA. Learning about treatment effects in a new target population under transportability assumptions for relative effect measures. Eur J Epidemiol. 2024;39(9):957.

#OHDSISocialShowcase This Week

Thursday

Toward Accurate Identification of Fontan and TGA in OMOP CDM: Registry-Anchored Algorithm Validation

(Seohu Lee, Suhyun Kim, Haeun Lee, Jong M Ko, Woo Young Park, Kwangsoo Kim, Sang Yun Lee, Ari Cedars)

Toward Accurate Identification of Fontan and TGA in OMOP CDM Registry-Anchored Algorithm Validation

PRESENTER: Seohu Lee

INTRO:

- **Why it matters:** Adult Congenital Heart Disease (ACHD) is rare and complex, requiring multicenter data to achieve adequate sample size and enhance generalizability of findings.
- **Problem:** Existing CDM-based algorithms often fail to capture surgically repaired ACHD patients (Fontan, D-TGA) because procedures performed in early childhood are frequently absent or incompletely carried over into adult EHR/CDM records.
- **Aim:** Refine OMOP CDM algorithms for surgical ACHD phenotypes and validate them against registry and billing data.

METHODS

a. **Data Sources**
- Internal validation: JHH OMOP CDM (2.3M patients, 2016-2024)
- External validation: SNUH OMOP CDM (3.8M patients, 2004–2024) linked to Korean claims and registry

b. **ACHD Subtypes Defined**
- Fontan circulation
- D-TGA with arterial switch
- D-TGA with atrial switch

c. **Algorithm Design**
1. Define algorithms using condition + procedure codes with explicit exclusions
2. Extract patient-level cohorts
3. Link to claims/registry for gold standard validation
- Phenotype algorithms were developed and Implemented in SQL using OHDSI libraries

d. **Validation**
- Gold standards = Korean insurance claims + Korean fontan registry
- Metrics: Precision, Recall, F1-score

CDM-based phenotyping works, but we must **go beyond structured data** to capture **surgical history**

Table1. Fontan and D-TGA Phenotype Definitions

Subtype	Phenotype Criteria	OMOP Concept IDs
Fontan	Procedure occurrences of 'Fontan'	2107268, 2107269, 2107270
	OR	
	Procedure occurrences of 'Glenn'	2107355, 4051948, 2107356, 40491942
	NOT	*NOT*
	Condition occurrences of 'Ebstein's anomaly'	35210812, 4069182
D-TGA (Arterial Switch)	Condition occurrences of 'Discordant ventriculoarterial connection'	432431, 35210794
	AND	
	Procedure occurrences of 'Repair of TGA (anatomic / arterial switch)'	2107375, 2107377
D-TGA (Atrial Switch)	Condition occurrences of 'Discordant ventriculoarterial connection'	432431, 35210794
	AND	
	Procedure occurrences of 'Repair of TGA (non-anatomic / atrial switch)'	2107361, 2107356, 40491942

Table 2. Patient Counts for Fontan and D-TGA Phenotypes at JHH and SNUH

Subtype	JHH (All)	JHH (Age ≥ 18)	SNUH (All)	SNUH (Age ≥ 18)
Fontan	135	2	297	29
D-TGA with Arterial Switch	77	10	126	0
D-TGA with Atrial Switch	35	1	134	0

Table 3. Performance Metrics for Fontan and D-TGA (Atrial Switch) in Claims and Registry Data at SNUH

Data Source	Subtype	Precision (%)	Recall (%)	F1-score (%)
Claims	Fontan	81.8	72.8	77.0
	D-TGA	95.5	37.3	53.8
Registry	Fontan	62.6	44.8	52.3
	D-TGA	62.7	30.4	41

Take a picture to download the full paper

Key Takeaways

- **Clinical impact**: Incomplete surgical history leads to under-identification of ACHD adults, limiting multicenter studies.
- **Adult cases are rare in CDM**: Most cases were pediatric, reflecting difficulty tracing childhood surgeries into adult records.
- **Structured data is insufficient**: Missing childhood surgical records and vocabulary gaps both reduce phenotyping accuracy.
- **Cross-institution feasibility proven**: JHH and SNUH algorithms scale but need multimodal inputs (echo, clinical notes).

Future Directions

- **Integrate multimodal data**: Expand phenotyping by incorporating echocardiography reports and narrative clinical notes, using LLM-based pipelines.
- **Enhance external validation**: Benchmark via the Alliance for Adult Research in Congenital Cardiology (AARCC) and international ACHD networks

Seohu Lee, Suhyun Kim, Haeun Lee, Jong M Ko, Woo Young Park, Kwangsoo Kim, Sang Yun Lee, Ari Cedars

JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

SNUH SEOUL NATIONAL UNIVERSITY HOSPITAL

OHDSI

#OHDSISocialShowcase This Week

Friday

DarwinBenchmark: Evaluating cohort generation and analytics in OMOP CDM databases

(Ioanna Nika, Maxim Moniat, Guido van Leeuwen, Ross Williams)

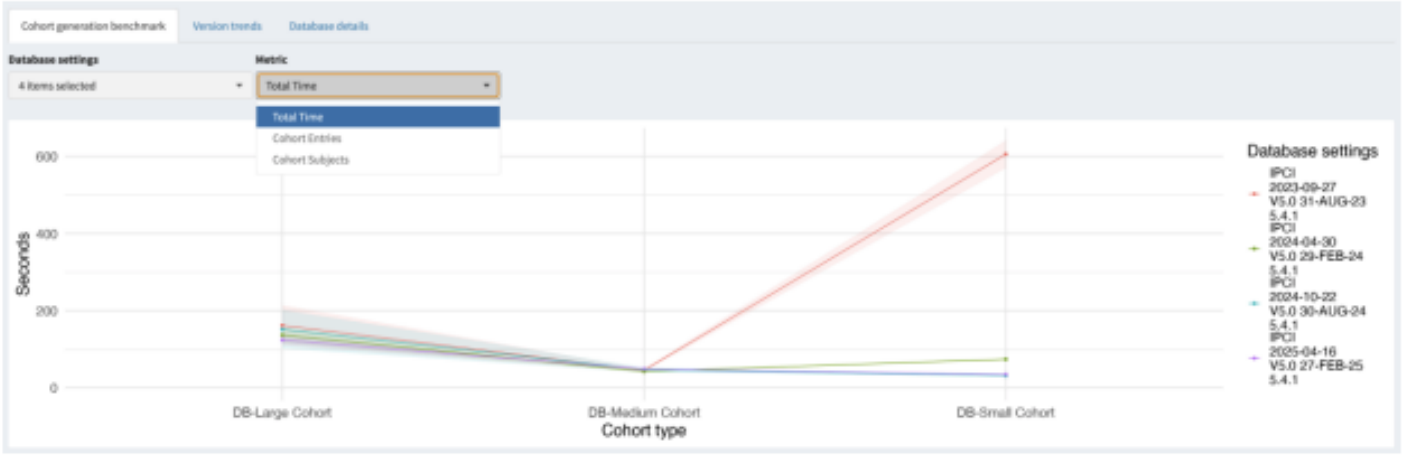
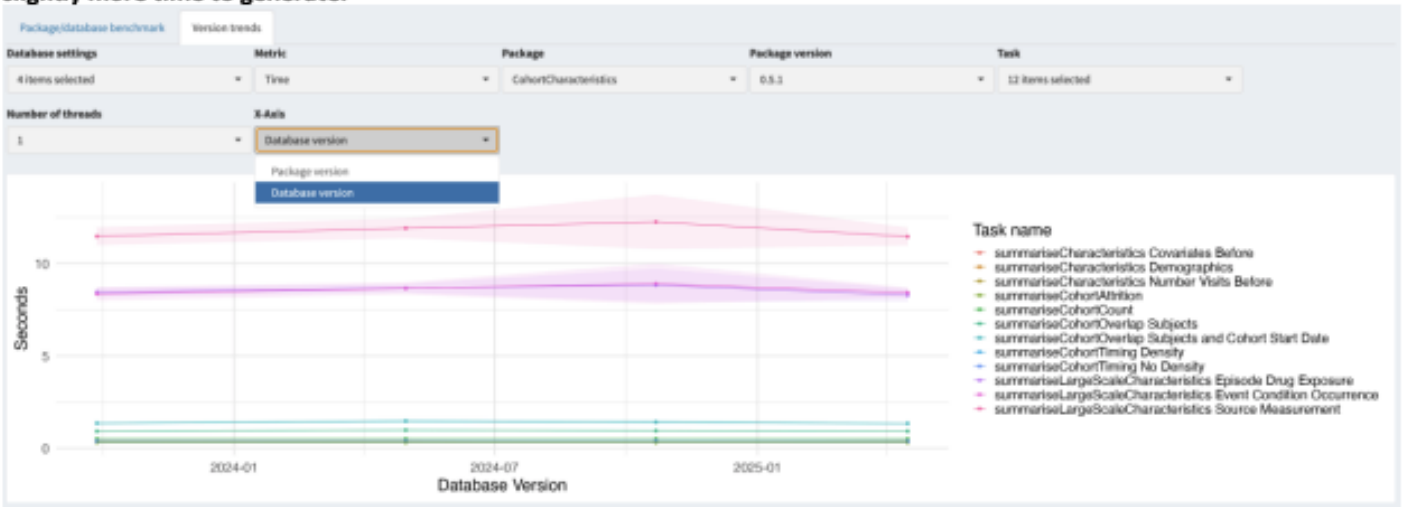


Figure 2. User interface for the Cohort Generation Benchmark. Despite larger data volumes, recent releases generate the most constrained cohort faster. Runtimes for other cohorts remain consistent across versions. The largest cohort requires slightly more time to generate.



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