Workgroup OKRs + Phenotype Phebruary, Session 4

OHDSI Community Call
Feb. 25, 2025 • 11 am ET

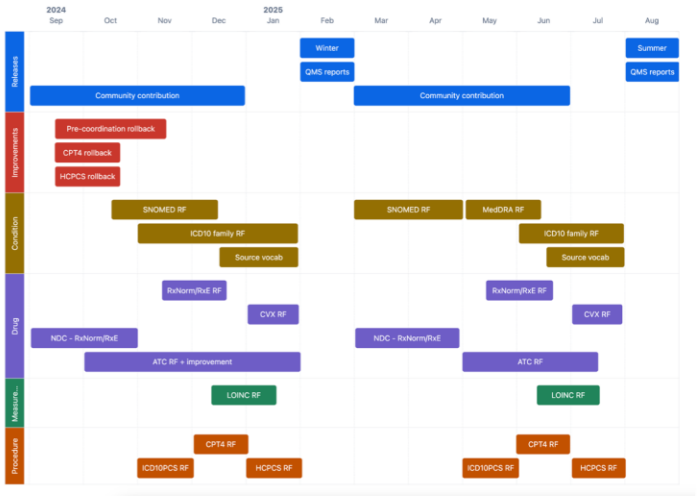
Upcoming Community Calls

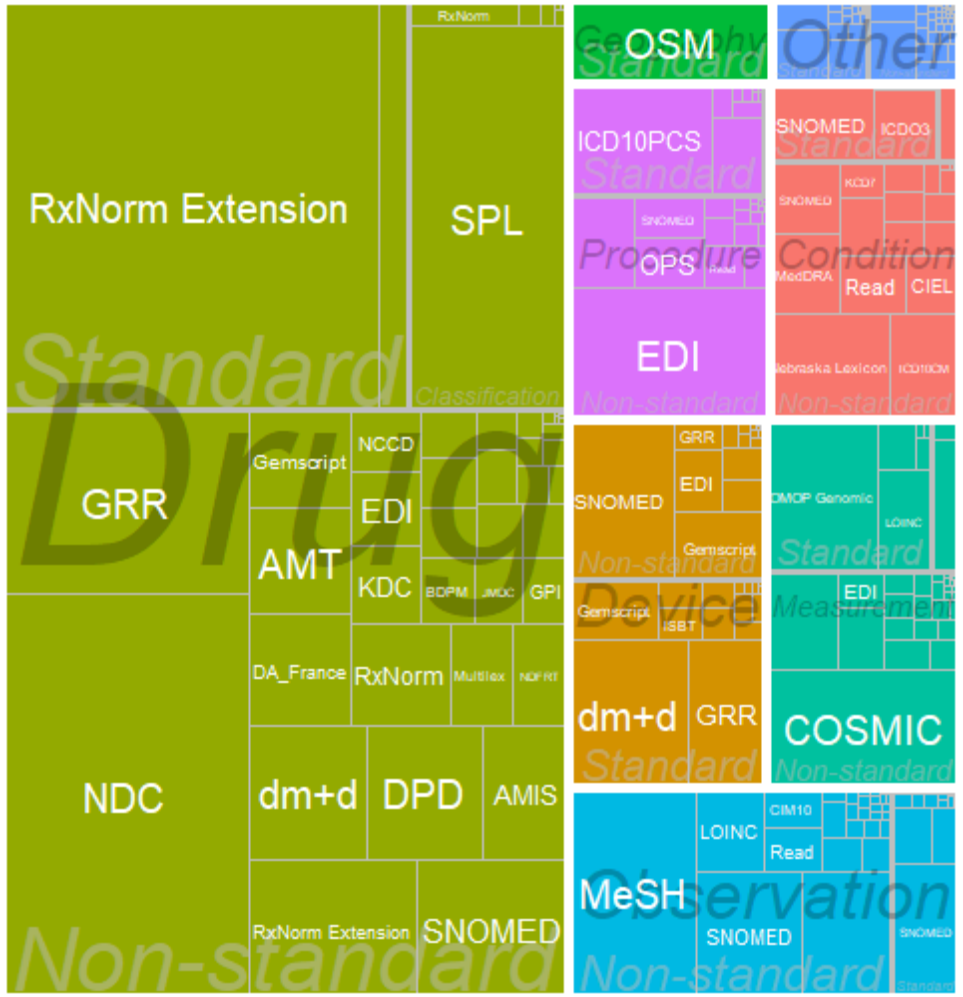
Date	Topic
Feb. 25	Fourth Week of 2025 Workgroup OKRs/Phenotype Phebruary
Mar. 4	Vocabulary Release Update, Winter 2025
Mar. 11	Book of OHDSI 2.0 Brainstorm and Planning Session
Mar. 18	March to Data Fitness
Mar. 25	TBA
Apr. 1	Recent OHDSI Publications

March 4 Community Call

Our Vocabulary Team will highlight the updates and changes from the Winter 2025 Refresh!

The roadmap for 2024/2025 was built using the insights from the landscape assessment to prioritize needs of the majority of the community and includes refreshes of the commonly used standard vocabularies such as SNOMED, CPT4 and LOINC, improvements in the mappings of the commonly used source terminologies and continuation of the work on a new approach to building drug classification to ATC (https://github.com/OHDSI/Vocabulary-v5.0/wiki/Vocab.-ATC).





Three Stages of The Journey

Where Have We Been?

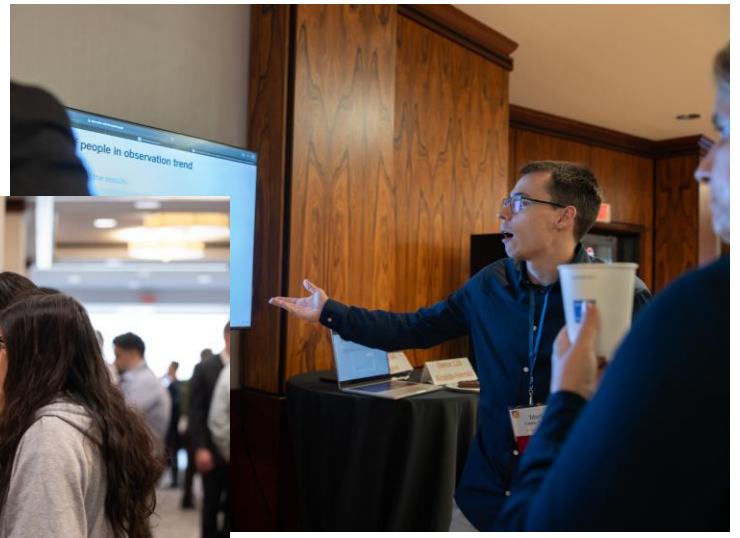
Where Are We Now?

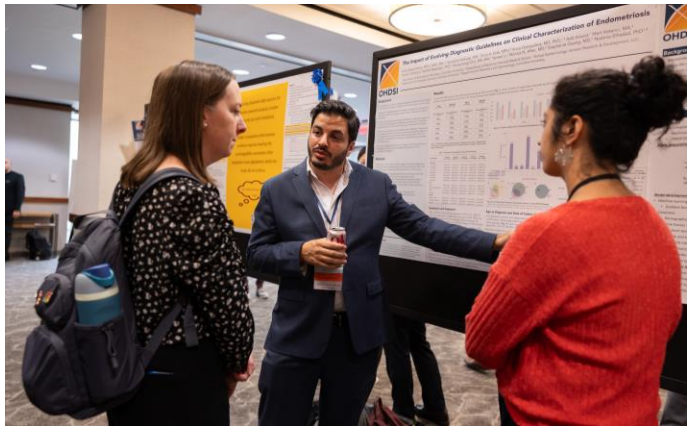
Where Are We Going?

Join The Scientific Review Committee

Please consider joining the Scientific Review Committee for OHDSI2025.

The deadline to sign up is March 3. Please use the form in the chat or on the community calls homepage.





#OHDSISocialShowcase This Week

Monday

Standardizing Rare Disease Patient Registry data to the OMOP-CDM

(Parag Shiralkar, Radhika Lakireddy, Sushma Ghanta, Sanket Kalyankar)

Standardizing Leigh Syndrome Patient Registry Data to the OMOP Common Data Model by Using Open Source R

Authors: Parag Shiralkar[1], Sanket Kalyankar[1], Jyoti Maini[1], Sushma Ghanta[1], Radhika Lakkireddy[1], Kasey Woleben[2], Sophia Zilber[2]

1- Sumptuous Data Sciences; 2 - Cure Mito Foundation

Background

OMOP (Observational Medical Outcomes Partnership)Common Data Model (CDM) was developed by the Observational Health Data Sciences and Informatics (OHDSI) community, which provides a standardized structure for organizing healthcare data and facilitating interoperability between different data sources. After standardization, data from disparate sources can be integrated and analyzed more efficiently, enabling comprehensive research and analysis.

Leigh syndrome is a rare, complex, and incurable early onset (typically infant or early childhood) mitochondrial disorder with both phenotypic and genetic heterogeneity, for which generating real-world evidence is challenging. Leigh syndrome global patient registry was developed by the Cure MIto Foundation, a leading patient advocacy organization dedicated to advancing research of Leigh syndrome and empowering and supporting affected families worldwide.

The registry includes two patient-reported surveys: a general survey utilizing the NIH Common Data Elements, and a Leigh Syndrome-specific survey and is hosted on a Coordination of Rare Diseases at Sanford (CoRDS) platform. Our objective is to document the process and outcomes of transforming registry data to the OMOP CDM and highlight challenges and our potential solutions.

Methods

The registry's raw data was scanned using White Rabbit, an application that evaluates the structure and contents of a database in preparation for developing an ETL (Extract, Transform, Load) process. ETL is used to effectively automate the transformations while preserving the quality and integrity of the data.

With the aid of the scan report produced by White Rabbit, we then created mapping specifications using Rabbit in a Hat, an application for interactively designing an ETL to the OMOP Common Data Model. Using OMOP CDM v5.4 IG, we applied mapping standards and used R programming to change the data through value standardization, de-duplication, and normalization to bring it into compliance with OMOP CDM criteria.

Additionally, with Athena's assistance, we gave concept ids and vocabulary to the data. Following data transformation, independent programming utilizing certain R packages was used for validation and quality assurance checks to guarantee data quality and conformance to OMOP CDM criteria.

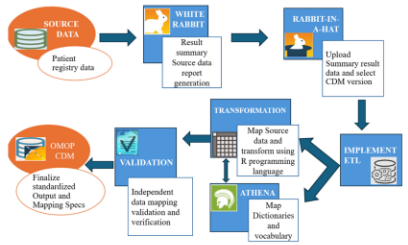
Contact:
parag.shiralkar@sumptuous-ds.com
sophia@curemito.org

Results

The entire registry dataset was evaluated, processed, and standardized to align with the requirements of the OMOP Common Data Model. The output we have developed was in compliance with OMOP CDM v5.4 Implementation guide.

Mapping source registry data to the OMOP CDM and aligning it with standardized vocabulary enhances data availability and usability for further disease research. This standardization eases seamless data integration, enabling robust real-world evidence generation. Our approach overcomes data heterogeneity challenges, paving the way for future research and improved patient outcomes. Additionally, converting source data into OMOP CDM generally enables mapping to other standards like SDTM, facilitating data availability for U.S. FDA submission purposes, although in our case data has already been mapped to SDTM previously.

Figure 1: Process flow of patient registry data transformation to OMOP



Conclusions

In this study, novel approaches for converting raw data from two registry databases to the OMOP CDM were successfully developed and applied with limited deviation and very few records being excluded from mapping. This approach addresses a need for combining real-world data of patients with rare diseases for the purpose of evidence generation and could serve as reference for future researchers wishing to undertake similar data mapping projects.

Acknowledgments: Vani Musikara, Alekhya Ettamsetti, Pallavi Bakare, Suraj Pujari, Danielle Boyce, MPH, DPA

Table 1: Mapping of Leigh syndrome patient registry data to OMOP domains.

Domain	Domain description	Leigh syndrome registry data mapped to domain
PERSON	Identifies each person with demographic details	Demographic information
VISIT	Captures details of healthcare encounters or instances of survey completion	Questionnaire date
CONDITION_OCCURRENCE	Records of conditions (diagnoses, symptoms) experienced by a person	Diagnosis, first concerns noticed, symptoms history
PROCEDURE_OCCURRENCE	Records of procedures	Type of genetic and diagnostic tests done
DEVICE_EXPOSURE	Records of medical device use	Devices such as feeding tubes, mobility devices
MEASUREMENT	Records of measurements or tests	Genetic testing results
OBSERVATION	Captures clinical facts about a person obtained in the context of examination, questioning or a procedure	Loss of milestones, caregiver burden, quality of life, family history, healthcare utilization
DEATH	Records of death and cause of death	Death information
SPECIMEN	Information about biological samples	Biospecimen information
LOCATION	Geographic information	Participant country or state
PROVIDER	Information about healthcare providers	Healthcare providers seen by participant
CDM_SOURCE	Contains detail about the source database and the process used to transform the data into the OMOP Common Data Mode	CDM Metadata information such as CDM source name, version of CDM, CDM holder name, CDM release date.

References

Zilber, S., Woleben, K., Johnson, S. C., et al. (2023). Leigh syndrome global patient registry: uniting patients and researchers worldwide. Orphanet Journal of Rare Diseases, 18, 264. https://doi.org/10.1186/s13023-023-02886-0

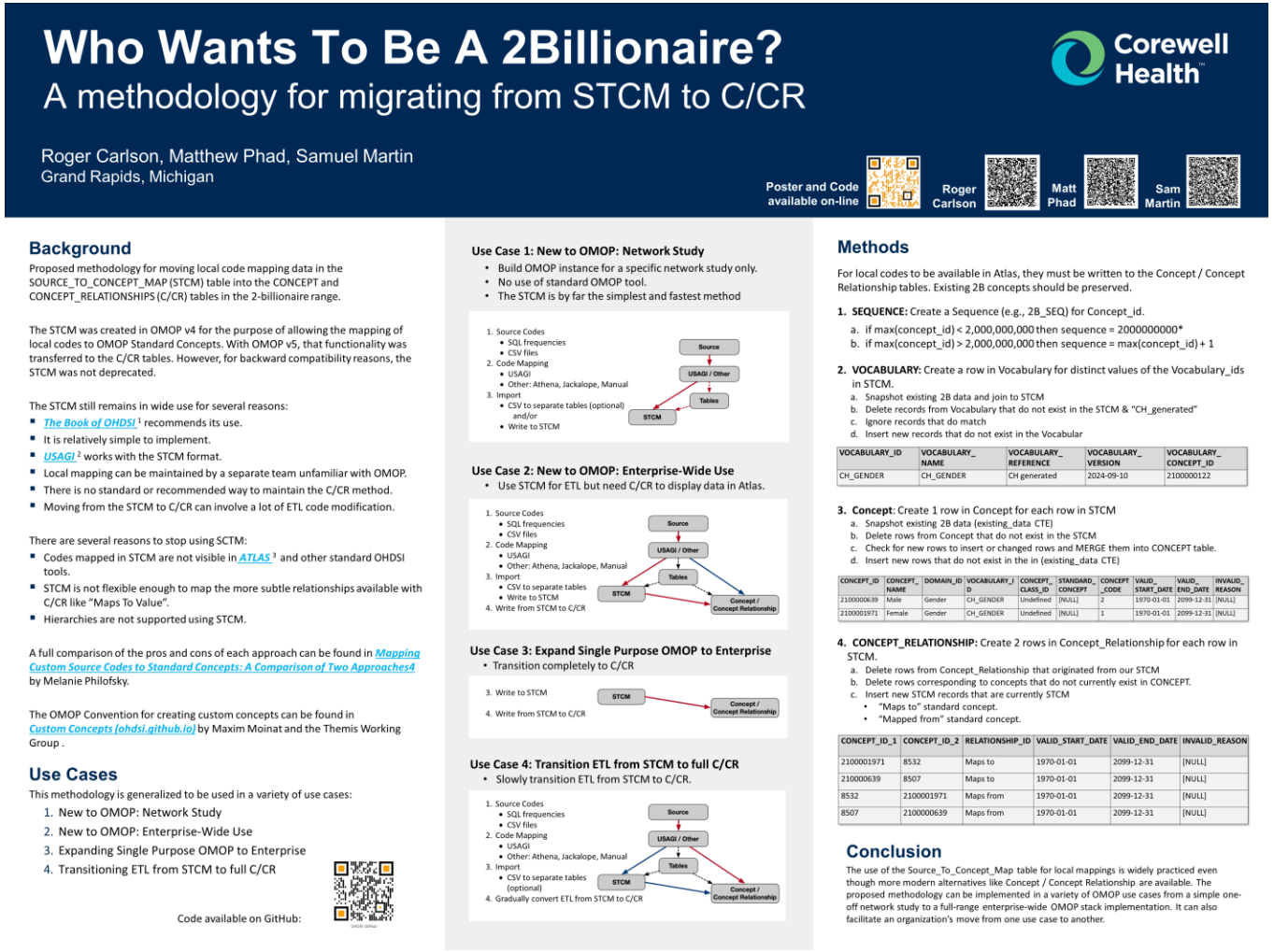
Shiralkar, P., Bakare, P., Woleben, K., & Zilber, S. (2024). Interoperability of Leigh syndrome patient registry data with regulatory submission standards. Journal of the Society for Clinical Data Management, 4(1). https://doi.org/10.47912/jscdm.244

#OHDSISocialShowcase This Week

Tuesday

Who Wants To Be A 2Billionaire? - A methodology for migrating from STCM to C/CR

(Roger Carlson, Matthew Phad, Samuel Martin)



Who Wants To Be A 2Billionaire?

A methodology for migrating from STCM to C/CR

Roger Carlson, Matthew Phad, Samuel Martin
Grand Rapids, Michigan

Corewell Health

Poster and Code available on-line | Roger Carlson | Matt Phad | Sam Martin

Background

Proposed methodology for moving local code mapping data in the SOURCE_TO_CONCEPT_MAP (STCM) table into the CONCEPT and CONCEPT_RELATIONSHIPS (C/CR) tables in the 2-billionaire range.

The STCM was created in OMOP v4 for the purpose of allowing the mapping of local codes to OMOP Standard Concepts. With OMOP v5, that functionality was transferred to the C/CR tables. However, for backward compatibility reasons, the STCM was not deprecated.

The STCM still remains in wide use for several reasons:

- *The Book of OHDSI*[1] recommends its use.
- It is relatively simple to implement.
- *USAGI*[2] works with the STCM format.
- Local mapping can be maintained by a separate team unfamiliar with OMOP.
- There is no standard or recommended way to maintain the C/CR method.
- Moving from the STCM to C/CR can involve a lot of ETL code modification.

There are several reasons to stop using SCTM:

- Codes mapped in STCM are not visible in *ATLAS*[3] and other standard OHDSI tools.
- STCM is not flexible enough to map the more subtle relationships available with C/CR like "Maps To Value".
- Hierarchies are not supported using STCM.

A full comparison of the pros and cons of each approach can be found in *Mapping Custom Source Codes to Standard Concepts: A Comparison of Two Approaches*[4] by Melanie Philofsky.

The OMOP Convention for creating custom concepts can be found in *Custom Concepts (ohdsi.github.io)* by Maxim Moinat and the Themis Working Group .

Use Cases

This methodology is generalized to be used in a variety of use cases:

1. New to OMOP: Network Study
2. New to OMOP: Enterprise-Wide Use
3. Expanding Single Purpose OMOP to Enterprise
4. Transitioning ETL from STCM to full C/CR

Code available on GitHub:

Use Case 1: New to OMOP: Network Study

- Build OMOP instance for a specific network study only.
- No use of standard OMOP tool.
- The STCM is by far the simplest and fastest method

Use Case 2: New to OMOP: Enterprise-Wide Use

- Use STCM for ETL but need C/CR to display data in Atlas.

Use Case 3: Expand Single Purpose OMOP to Enterprise

- Transition completely to C/CR

Use Case 4: Transition ETL from STCM to full C/CR

- Slowly transition ETL from STCM to C/CR.

Methods

For local codes to be available in Atlas, they must be written to the Concept / Concept Relationship tables. Existing 2B concepts should be preserved.

1. **SEQUENCE:** Create a Sequence (e.g., 2B_SEQ) for Concept_id.
   a. if max(concept_id) < 2,000,000,000 then sequence = 2000000000*
   b. if max(concept_id) > 2,000,000,000 then sequence = max(concept_id) + 1
2. **VOCABULARY:** Create a row in Vocabulary for distinct values of the Vocabulary_ids in STCM.
   a. Snapshot existing 2B data and join to STCM
   b. Delete records from Vocabulary that do not exist in the STCM & "CH_generated"
   c. Ignore records that do match
   d. Insert new records that do not exist in the Vocabular

VOCABULARY_ID	VOCABULARY_ NAME	VOCABULARY_ REFERENCE	VOCABULARY_ VERSION	VOCABULARY_ CONCEPT_ID
CH_GENDER	CH_GENDER	CH generated	2024-09-10	2100000122

3. **Concept**: Create 1 row in Concept for each row in STCM
   a. Snapshot existing 2B data (existing_data CTE)
   b. Delete rows from Concept that do not exist in the STCM
   c. Check for new rows to insert or changed rows and MERGE them into CONCEPT table.
   d. Insert new rows that do not exist in the in (existing_data CTE)

CONCEPT_ID	CONCEPT_ NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT _CODE	VALID_ START_DATE	VALID_ END_DATE	INVALID_ REASON
2100000639	Male	Gender	CH_GENDER	Undefined	[NULL]	2	1970-01-01	2099-12-31	[NULL]
2100001971	Female	Gender	CH_GENDER	Undefined	[NULL]	1	1970-01-01	2099-12-31	[NULL]

4. **CONCEPT_RELATIONSHIP:** Create 2 rows in Concept_Relationship for each row in STCM.
   a. Delete rows from Concept_Relationship that originated from our STCM
   b. Delete rows corresponding to concepts that do not currently exist in CONCEPT.
   c. Insert new STCM records that are currently STCM
      - "Maps to" standard concept.
      - "Mapped from" standard concept.

CONCEPT_ID_1	CONCEPT_ID_2	RELATIONSHIP_ID	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
2100001971	8532	Maps to	1970-01-01	2099-12-31	[NULL]
210000639	8507	Maps to	1970-01-01	2099-12-31	[NULL]
8532	2100001971	Maps from	1970-01-01	2099-12-31	[NULL]
8507	2100000639	Maps from	1970-01-01	2099-12-31	[NULL]

Conclusion

The use of the Source_To_Concept_Map table for local mappings is widely practiced even though more modern alternatives like Concept / Concept Relationship are available. The proposed methodology can be implemented in a variety of OMOP use cases from a simple one-off network study to a full-range enterprise-wide OMOP stack implementation. It can also facilitate an organization's move from one use case to another.

#OHDSISocialShowcase This Week

Wednesday

Leveraging the active comparator new user design to identify potential unknown benefits of canagliflozin

(Justin Bohn, Jamie P. Gilbert, Christopher Knoll, Zhong Yuan, David M. Kern, Patrick B. Ryan)

Leveraging the active comparator new user design to identify potential unknown benefits of canagliflozin

PRESENTER: **Justin Bohn**

INTRO

- The active comparator new user (ACNU) cohort design has emerged as a best practice for the estimation of drug effects from observational data
- We describe an attempt to use the ACNU design to screen for potential unknown benefits of the antidiabetic agent canagliflozin and compare our results with those obtained from a self-controlled cohort (SCC) design.

METHODS

- We generated new user cohorts for canagliflozin (a sodium-glucose transport protein 2 [SGLT2] inhibitor) and two frequently chosen active comparators, sitagliptin (a dipeptidyl peptidase-4 [DPP-4] inhibitor) and liraglutide (a glucagon-like peptide 1 [GLP-1] analogue), across five real-world databases
- Outcome cohorts were generated based on the first occurrence of each SNOMED condition, but analysis was limited to those with ≥ 300 events to reduce computation time
- Large-scale propensity score (LSPS) models were used to match new users of canagliflozin to each comparator in a 1:1 manner
- Cox proportional hazards models were used to estimate treatment effects conditioned on matched set
- Empirical calibration using negative control outcomes was used to account for residual systematic error and objective diagnostics determined whether a result would be eligible for review
- We used an ad-hoc criteria defining a potential benefit "signal" as an outcome with a **hazard ratio** (or incidence rate ratio in the SCC design) **≤ 0.7 and p-value ≤ 0.05** after calibration (without multiplicity adjustment) in **at least two databases**

RESULTS

- 1,850 outcomes were assessed across study designs (**Figure 2**, potential benefits only)
- Runtime for the ACNU design varied from roughly 1-4 days, depending on database, compared to 1-3 hours for the SCC
- 64 outcomes met the ad-hoc signaling criteria in the SCC design vs. 13 in either comparison of the ACNU design (**Table 1**)

Scan for Details

Scaling the **comparative cohort design** to accommodate **thousands of outcomes** is feasible and can complement traditional self-controlled designs in identifying of potential **unknown benefits**

Table 1

Outcome	# DBs with Signal in Design: A	B	C	DB-Specific Estimates: JMDC	Merative CCAE	Merative MDCD	Merative MDCR	Optum
Chronic kidney disease due to hypertension	0	4	1	A: n/a B: n/a C: n/a	A: 2.07 (1.03, 4.14) B: 0.68 (0.59, 0.78) C: 0.68 (0.58, 0.80)	A: 1.78 (1.05, 3.03) B: 0.69 (0.58, 0.82) C: 0.80 (0.66, 0.96)	A: 1.78 (1.00, 3.16) B: 0.70 (0.60, 0.81) C: 0.84 (0.69, 1.02)	A: n/a B: 0.69 (0.63, 0.75) C: 0.82 (0.72, 0.93)
Hydronephrosis	0	2	2	A: 1.02 (0.62, 1.67) B: n/a C: n/a	A: 0.79 (0.39, 1.58) B: 0.53 (0.45, 0.62) C: 0.56 (0.46, 0.68)	A: 1.39 (0.77, 2.52) B: n/a C: n/a	A: 1.54 (0.80, 2.96) B: n/a C: n/a	A: 0.83 (0.52, 1.31) B: 0.67 (0.56, 0.79) C: 0.64 (0.52, 0.79)
Hypertensive renal disease	0	3	0	A: 0.42 (0.08, 2.11) B: n/a C: n/a	A: 1.29 (0.65, 2.57) B: 0.71 (0.63, 0.81) C: 0.72 (0.61, 0.84)	A: 1.36 (0.81, 2.29) B: 0.69 (0.59, 0.82) C: 0.83 (0.69, 1.00)	A: 1.20 (0.69, 2.11) B: 0.69 (0.60, 0.79) C: 0.80 (0.66, 0.97)	A: 1.26 (0.82, 1.95) B: 0.70 (0.64, 0.77) C: 0.83 (0.73, 0.94)
Heart failure	0	3	0	A: 0.76 (0.61, 0.94) B: 1.04 (0.94, 1.14) C: n/a	A: 0.61 (0.41, 1.61) B: 0.75 (0.66, 0.86) C: 0.78 (0.67, 0.92)	A: 0.97 (0.58, 1.62) B: 0.68 (0.57, 0.81) C: 0.93 (0.77, 1.12)	A: 0.88 (0.50, 1.54) B: 0.67 (0.58, 0.77) C: 0.86 (0.70, 1.06)	A: 0.86 (0.55, 1.33) B: 0.69 (0.62, 0.76) C: 0.93 (0.81, 1.07)
Localized edema	0	3	0	A: 1.15 (0.77, 1.70) B: n/a C: n/a	A: 1.86 (0.94, 3.71) B: 0.69 (0.61, 0.78) C: 0.82 (0.71, 0.95)	A: 1.13 (0.68, 1.87) B: 0.77 (0.67, 0.88) C: 0.93 (0.80, 1.09)	A: 1.60 (0.89, 2.87) B: 0.69 (0.59, 0.80) C: n/a	A: n/a B: 0.69 (0.63, 0.76) C: 0.89 (0.79, 1.02)
Iron deficiency anemia	0	2	0	A: 1.10 (0.86, 1.40) B: 1.15 (1.02, 1.30) C: n/a	A: 0.85 (0.43, 1.69) B: 0.85 (0.75, 0.95) C: 0.88 (0.76, 1.02)	A: 0.96 (0.57, 1.61) B: 0.68 (0.57, 0.81) C: 0.91 (0.75, 1.11)	A: 1.09 (0.61, 1.93) B: n/a C: n/a	A: 1.05 (0.68, 1.63) B: 0.69 (0.62, 0.77) C: 0.87 (0.76, 1.01)
Anemia	0	2	0	A: 1.09 (0.87, 1.37) B: 1.03 (0.93, 1.15) C: n/a	A: 0.73 (0.37, 1.43) B: 0.78 (0.71, 0.86) C: 0.80 (0.70, 0.90)	A: 0.82 (0.50, 1.36) B: 0.72 (0.63, 0.83) C: 0.81 (0.70, 0.95)	A: 0.91 (0.52, 1.57) B: 0.67 (0.60, 0.76) C: 0.87 (0.74, 1.03)	A: 0.88 (0.57, 1.35) B: 0.70 (0.64, 0.76) C: 0.86 (0.76, 0.96)
Vomiting	0	0	2	A: 0.97 (0.78, 1.20) B: 1.06 (0.96, 1.17) C: n/a	A: 1.00 (0.50, 1.96) B: 0.93 (0.84, 1.03) C: 0.69 (0.61, 0.78)	A: 1.00 (0.61, 1.65) B: 0.96 (0.84, 1.10) C: 0.68 (0.59, 0.79)	A: 1.19 (0.68, 2.09) B: 0.92 (0.78, 1.08) C: 0.79 (0.64, 0.97)	A: 1.04 (0.67, 1.61) B: 0.91 (0.82, 1.00) C: 0.72 (0.64, 0.81)
Chronic heart failure	0	2	0	A: n/a B: n/a C: n/a	A: 0.98 (0.49, 1.98) B: 0.70 (0.59, 0.83) C: 0.71 (0.57, 0.88)	A: 1.44 (0.84, 2.45) B: n/a C: n/a	A: 1.12 (0.62, 2.01) B: n/a C: n/a	A: 1.02 (0.65, 1.59) B: 0.60 (0.53, 0.67) C: 0.82 (0.70, 0.96)

Design A: self-controlled cohort; Design B: active comparator new user (vs. sitagliptin); Design C: active comparator new user (vs. liraglutide); Results are given as incidence rate ratio [95% CI] for design A or hazard ratio [95% CI] for designs B and C. DB: database; JMDC: Japan Medical Data Center; CCAE: Commercial Claims and Encounters; MDCD: Multi-State Medicaid; MDCR: Medicare Supplemental and Coordination of Benefits; Optum® De-identified Clinformatics® Data Mart. Four additional outcomes (hemoglobin low, RBC count low, and hemoglobin level outside reference range, finding of vomiting) not included in table due to identical results to related outcomes in table.

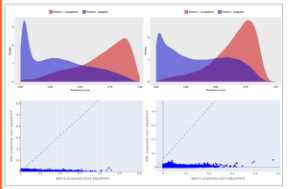


Figure 1. Preference score distributions and covariate balance plots for canagliflozin vs. sitagliptin (left column) and liraglutide (right column)

RESULTS (continued)a

- Upon manual review, all outcomes meeting the signaling criteria in the SCC design were deemed to have arisen due to protopathic bias
- Of the 13 outcomes that met the signaling criteria in either ACNU comparator (**Table 1**), clinical conditions related to kidney disease, heart failure, anemia, and vomiting arose favoring canagliflozin, although these signals were not observed using the SCC design

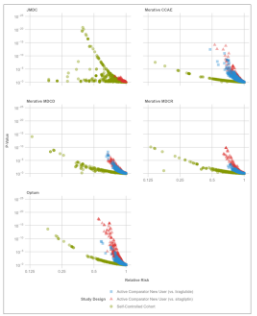


Figure 2. Results for 1,741 outcomes demonstrating a potential benefit of canagliflozin in at least one database and one study design

Justin Bohn, Jamie P. Gilbert, Christopher C. Knoll, Zhong Yuan, David M. Kern, Patrick B. Ryan

OHDSI

#OHDSISocialShowcase This Week

Thursday

Does the SARS-CoV-2 Infection Increase the Onset of New Mental Health Disorder? Findings from Difference-in-Differences Analyses Using an EHR-Based Cohort from the RECOVER Program

(Yiwen Lu, Jiayi Tong, Dazheng Zhang, Lu Li, Yuqing Lei, Ting Zhou, Jiajie Chen, Levon H Utidjian, Nathan J Blum, Kelly Kelleher, Kathleen Pajer, Raghuram Prasad, Josephine Elia, Christopher B. Forrest, Yong Chen)

OHDSI

RECOVER

Does the SARS-CoV-2 Infection Increase the Risk of Mental Health Disorder? Findings from Difference-in-Differences Analyses Using an EHR-Based Cohort from the RECOVER Program

Yiwen Lu[1,2,*], Jiayi Tong[1,3,*], Dazheng Zhang[1,3,*], Jiajie Chen[1,3], Lu Li[1,2], Yuqing Lei[1,3], Ting Zhou[1,3], Nathan J Blum[4], Kelly Kelleher[5], Kathleen Pajer[6], Levon H Utidjian[7], Raghuram Prasad[8,†], Josephine Elia[9,†], Christopher B. Forrest[10,†], Yong Chen[1-3, 11-13,†] on behalf of the RECOVER consortium

[1] Center for Health AI and Synthesis of Evidence (CHASE), Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[2] The Graduate Group in Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA
[3] Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine
[4] Division of Developmental and Behavioral Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[5] The Research Institute, Nationwide Children's Hospital, Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA
[6] University of Ottawa Department of Psychiatry, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada
[7] Applied Clinical Research Center, Children's Hospital of Philadelphia, Philadelphia, PA
[8] Department of Child and Adolescent Psychiatry, Children's Hospital of Philadelphia, Perelman School of Medicine, the University of Pennsylvania, Philadelphia, PA, USA
[9] Department of Pediatrics, Nemours Children's Health Delaware, Sydney Kimmel School of Medicine, Philadelphia, PA, US
10 Applied Clinical Research Center, Children's Hospital of Philadelphia, Department of Healthcare Management, Perelman School of Medicine, the University of Pennsylvania, Philadelphia, USA Philadelphia, PA
[11] Leonard Davis Institute of Health Economics, Philadelphia, PA, USA
[12] Penn Medicine Center for Evidence-based Practice (CEP), Philadelphia, PA, USA
[13] Penn Institute for Biomedical Informatics (IBI), Philadelphia, PA, USA

*Co-first author †Co-senior author

Penn Medicine

CHASE Center for Health AI & Synthesis of Evidence

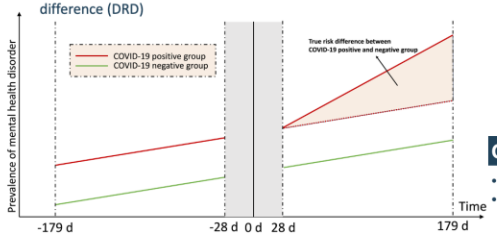
pennCIL A Computing · Inference · Learning lab at University of Pennsylvania

Background

- **PASC** is defined by the World Health Organization (WHO) as the persistence of at least one physical symptom for 12 weeks following initial testing without an alternative
- PASC impacts various organ systems, including **mental health**, with increased risks of anxiety, depression, and other mental health disorders documented in adults.
- **Our goals** are to:
  - Focus on pediatric population
  - Compare the frequency of mental health disorders among COVID-19-positive patients versus COVID-19-negative patients
  - Assess if COVID-19 infection heightens mental health disorder risks and account for pandemic-related factors.

Methods

- To assess differences in the prevalence of mental health conditions between COVID-19 positive and negative patients, we conducted a Difference-in-difference (DiD) analysis using a two-sample proportion test with stratified cohorts of children and adolescents.
  - Use **risk difference (RD)** as the comparative measure
  - Fit **large scale propensity score (LSPS) model** to control measured confounders
  - Assess covariate balance with a difference of standardized mean difference (SMD) of 0.1 or less
  - **Difference-in-differences (DiD)** analysis to control pre-COVID case/control differences
  - Linear regression model to estimate the difference of risk difference (DRD)



Results

- **Data source:** This retrospective cohort study is part of the NIH Researching COVID to Enhance Recovery (RECOVER) Initiative (https://recovercovid.org/), which aims to learn about the long-term effects of COVID-19
- **Cohort:** A total of 326,074 patients with SARS-CoV-2 infection and 887,314 patients without SARS-CoV-2 infection in the RECOVER program between March 2020 and October 2022 with at least 6 months of follow-up time
- **Stratification:** All analyses were stratified by age at the index date (Children: 5~11; adolescents: 12~20)

Conclusions

- This study reveals an increased risk of mental health disorders in children and adolescents following SARS-CoV-2 infection.
- Awareness of mental health complications in the post-acute phase of COVID-19 infection will enhance the timely diagnosis and treatment of these conditions.

Contact: yiwenlu@sas.upenn.edu; ychen123@pennmedicine.upenn.edu

#OHDSISocialShowcase This Week

Friday

dbt for OMOP Phase I: dbt-synthea

(Katy Sadowski, Vishnu Chandrabalan, Adam Bouras, Roger Carlson)

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?

Africa Chapter

Chapter Leads: Cynthia Sung, Agnes Kiragga

Purpose

- To strengthen awareness and capacity for data harmonization and analysis using OHDSI tools to meet the data-driven evidence needs of African researchers, health providers, and governments

Teams Meeting Link

Africa Chapter 2025 Objectives and Key Results

Objective 1	Key Result	OKR Lead(s)
Grant funding to expand data in OMOP CDM	• Submit 2 or more grants	Marc Twagirumukiza (Rwanda) Agnes Kiragga (Kenya, Uganda)
Objective 2	**Key Results**	**OKR Lead(s)**
OHDSI Africa Symposium on the Continent	• Hold Symposium before end of year: sometime during week of Nov 10 • Identify hosting site: Joint Clin Res Ctr Kampala	Alex Asiimwe (Uganda) Francis Kanyike (JCRC Uganda) David Muyomba (JCRC Uganda)
	• Publicity to achieve >100 attendees	APHRC (Kenya), JCRC (Uganda)
	• Find sponsors • Symposium content: overview, group activity; tutorials, leadership dvpt vision,	? members
Objective 3	**Key Results**	**OKR Lead(s)**
Deep dive on ETL/data transformation process	• Group hands-on ETL exercise over ~2 mo	Narem Singh (India)
	• Identify usable data sources on the continent	IQVIA, JCRC, HDSS sites

Africa Chapter 2025 Objectives and Key Results

Objective 4	Key Result	OKR Lead(s)
Propose Africa-specific terminology to add to	Compare local data dictionaries/sources to standardized vocabularies (e.g. CIEL -> ATHENA)	Andy Kanter (USA)
OHDSI Standard Vocabulary	• Identify at least 3 vocabularies to run concept prevalence studies • Create a process for moving the data from CIEL-ATHENA	Bolu Oluwalade (Nigeria), Katherine Johnston (S. Africa)
Objective 5	**Key Results**	OKR Lead(s)
Conduct learning sessions for different audiences	• Develop customized curricula for MOH vs doctors vs clinical/IT support staff • Hands on use cases relevant for the audience	David Amadi (Kenya) Lars Halvorsen (Norway, Belgium)
Objective 6	**Key Results**	**OKR Lead(s)**
Maturity model for ETL implementation	• Outline characteristics of 4-5 levels of maturity	Agnes Kiragga (Kenya), Adam Bouras (Morocco)

Ongoing Activities

- Translation of the Book of OHDSI: French, Arabic, Portuguese, Swahili - Michel Walvarens
- Data Science without Borders Program (Kenya, Cameroon, Senegal)
- BRIDGE PhD/postdoc Training (Rwanda, Kenya, Uganda, S. Africa, Benin, Norway, Belgium, Ethiopia)

Africa Chapter

Chapter Leads: Cynthia Sung, Agnes Kiragga

Biweekly Meeting – Monday 10 AM ET

4 pm WAST; 5 pm CAT/SAST; 6 pm EAT (during **EST**)

Mar 3

US switches to Daylight Savings Time Mar 9

3 pm WAST; 4 pm CAT/SAST; 5 pm EAT (during **EDT**)

Mar 17, Mar 31, Apr 14 . . .

Teams Meeting Link

OHDSI Rare Diseases Working Group

2025 OKR

WG leads:
Xiaoyan Wang PhD
Chunhua Weng PhD

Rare Diseases Working Group

Mission Statement:

To advance the understanding, diagnosis, and treatment of rare diseases through the development and application of open-source analytics tools, standardized data models, and collaborative research.

- Develop and refine methodologies for collaborative rare disease case aggregation within the OHDSI community.
- Support evidence generation over rare diseases within the OHDSI ecosystem.
- Facilitate collaboration across stakeholders, including researchers, clinicians, sponsors, industry partners, and patient advocacy groups, to drive impactful discoveries.

Rare Disease WG 2025 OKRs

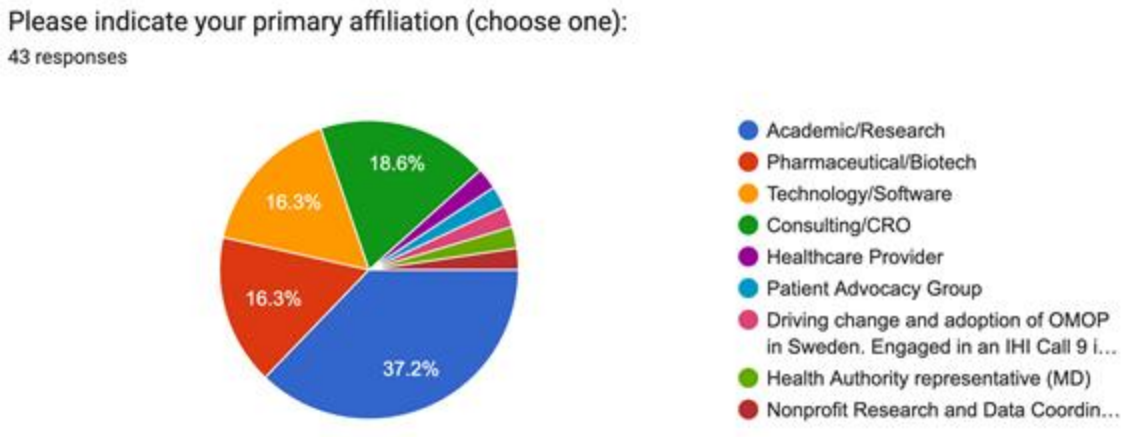
WG Leads: Xiaoyan Wang, Chunhua Weng

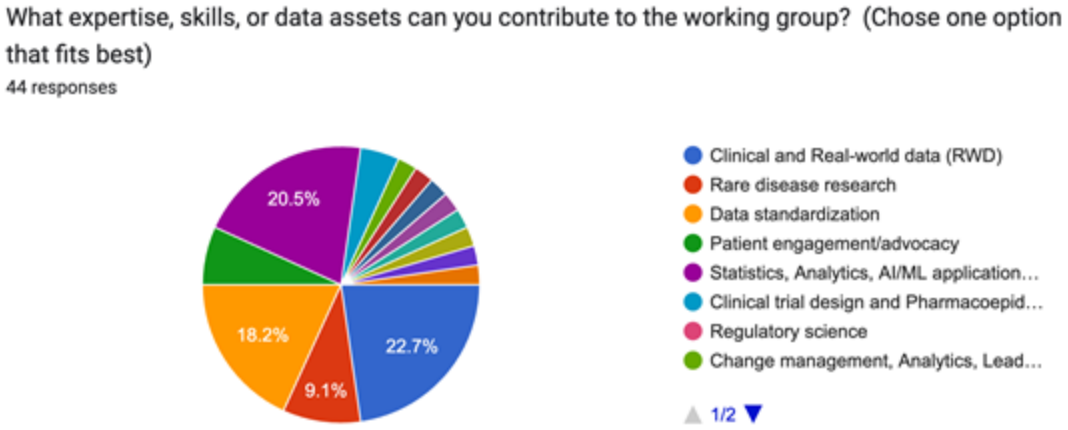
Objective 1: **Strengthen community engagement and collaboration**

KR1: Form a collaborative team of ~20 active members with expertise in rare disease research, data science, phenotyping and clinical study design.

KR2: Organize about three knowledge-sharing sessions to foster interdisciplinary collaboration.

KR3: Establish partnerships with key stakeholders, including researchers, clinicians, patient advocacy groups, regulatory agencies, and industry leaders, to enhance research impact.





Rare Disease WG 2025 OKR

WG Lead: Xiaoyan Wang, Chunhua Weng

- **Objective 2 : Enable scalable and reproducible rare disease analytics**
  - **KR1**: Design and initiate a multi-site rare disease study leveraging OHDSI's network and the OMOP CDM.
  - **KR2:** Develop a standardized approach for identifying and integrating hard-to-find rare disease cohorts across multiple data sources.
  - **KR3:** Release preliminary findings and methodological insights to guide future large-scale rare disease research.
- **Objective 3: Integrate Innovative AI Technologies for Rare Disease Research**
  - **KR1:** Evaluate and implement NLP and LLMs for rare disease cohort identification.
  - **KR2:** Develop and test the integration of knowledge graphs and AI-driven knowledge discovery tools for rare disease evidence generation.
  - **KR3:** Release an open-source framework or best practices document for leveraging AI in rare disease studies within the OHDSI community.

Rare Diseases WG Next Meeting

- To join, go to ohdsi.org -> workgroups tab ->
- Select Workgroup - Rare Diseases
- Meeting details:

  Time: 10-11am EST, Feb 28th, 2025

  Meeting Link: Join the meeting now

  Meeting ID: 210 677 103 238

  Passcode: Yw2GJ7pq

Agenda:

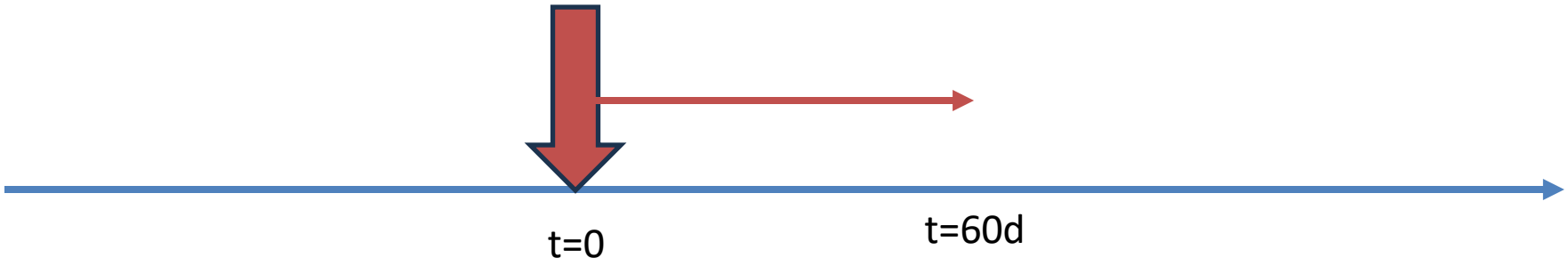
- Introduce the WG's objectives
- Discuss priority areas
- Identify opportunities for collaboration.

Surgery and Perioperative WG: 2025 OKR

Feb 25, 2025

Surgery and Perioperative WG

To collaboratively generate the observational health evidence needed to better understand the role of surgical and perioperative interventions in medical care.



Meetings: Last Wednesday of Every Month

7am PST / 8am MST / 10am EST / 3pm BST / 4pm CET.

Evidence Driven Data Standardization

- Objective: Convert existing surgical cohort assets from non-standard concepts to standard concepts
  - Key Results:
    - Validation completed through peer review of included concepts / source codes
    - Validation completed through population level validation (CohortDiagnostics)
- Objective: Improve vocabulary mappings in the procedural domain (in partnership with the OHDSI Vocabulary Team)
  - Key Results
    - Contribute to LLM enhanced / Manual Review of ICD 10 PCS → SNOMED mappings in the OHDSI vocabulary

Guideline Driven Evidence **Generation** & Dissemination

- Objective: Refine, improve, and disseminate surgical (and outcome) **cohorts**
- Key Results:
  - Complete of Target Surgery Translation to Standard Vocabulary Concepts, and validate. →Updates submitted to Phenotype Library
  - Completion of Outcome Cohort Review →Updates submitted to Phenotype Library
  - Complete phenotype description of at least one surgical cohort (as published paper)

Guideline Driven Evidence **Generation** & **Dissemination**

- Objective: Completion of Surgery Incidence Rate Study
  - Key Results
    - Complete specification and execution of Surgery Incidence Rate study
    - Publication of *at least two* Surgery: Outcome Incidence Rate Papers
- Objective: Specification / Planning of *at least one* surgery related effect estimation or prediction Network study
  - Key Results:
    - Background, proposed methodology, initial cohorts characterized by Jan 2026

Guideline Driven Evidence **Generation** & **Dissemination**

- Objective: Improve Engagement with the Surgery and Periop Medicine Clinical Community
  - Key Results
    - Acceptance of Clinical WG work in at least 3 surgery / perioperative conferences
    - Presentation of broader OHDSI Mission / Capabilities / Methods in at least one Surgery / Periop conference
    - Involvement of (at least) 3 new members from (at least) 3 new surgery / perioperative focused research groups

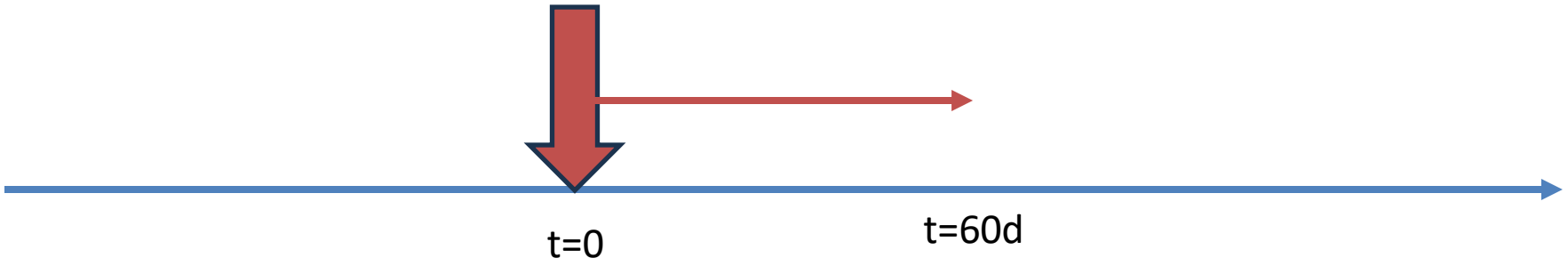
Collaborative Education / OHDSI Community Collaboration

- Objective: Promote Cross Work Group Collaboration
  - Key Results:
    - Establish 2 strategic Collaborations with other WG
    - 2 Joint meetings in 2025

Surgery and Perioperative WG

Questions? Interest?

minty@ohdsi.org



Meetings: Last Wednesday of Every Month

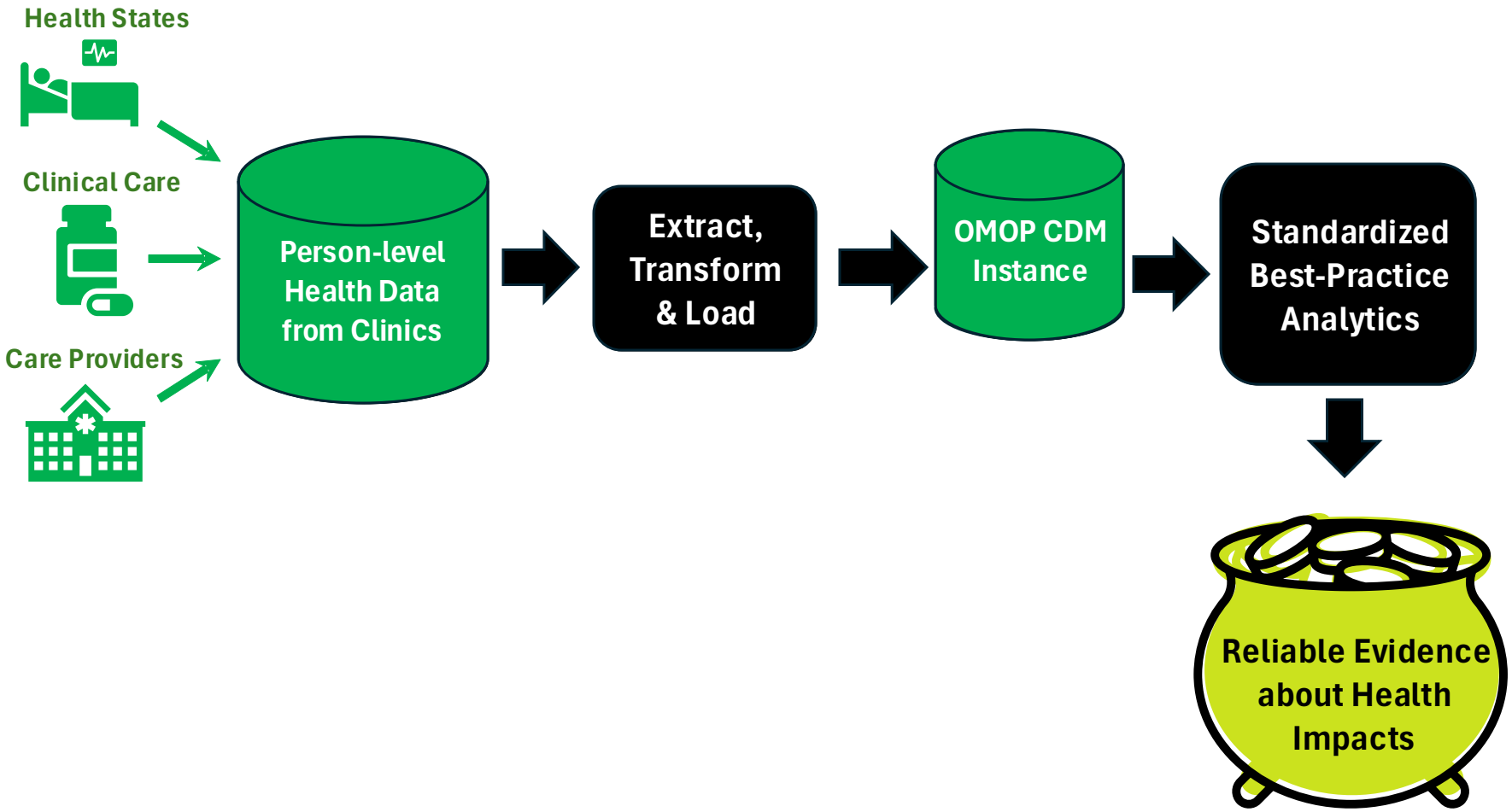
7am PST / 8am MST / 10am EST / 3pm BST / 4pm CET.

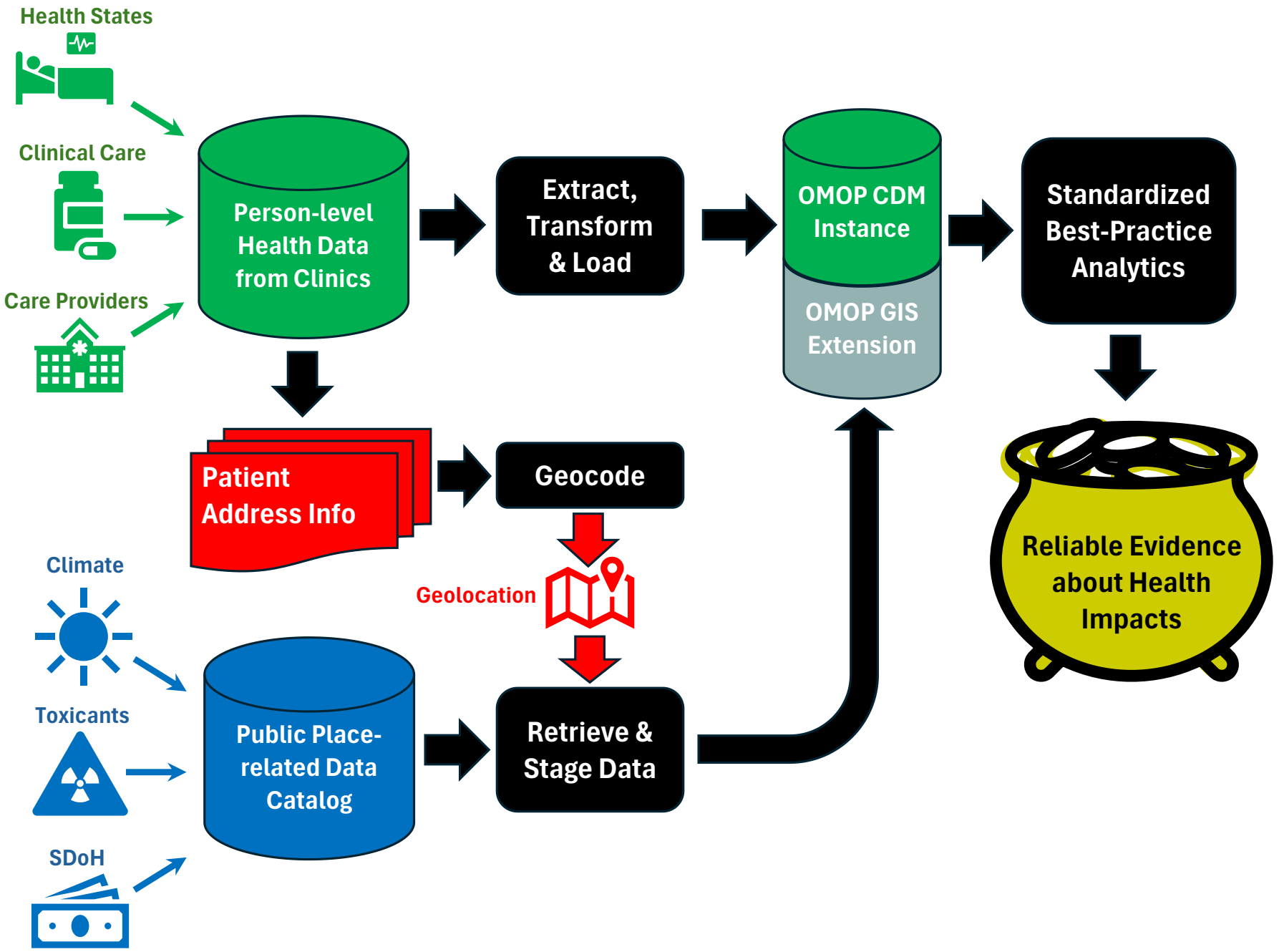
OHDSI GIS Workgroup

GIS Workgroup: Mission

Improve the health of populations by generating reliable evidence from integrated geospatial and person-level health data.

- Climate
- Toxicants
- Regional policies
- Social determinants of health
- …


Health States
Clinical Care
Care Providers
Person-level Health Data from Clinics
Extract, Transform & Load
OMOP CDM Instance
Standardized Best-Practice Analytics
Reliable Evidence about Health Impacts


Health States
Clinical Care
Care Providers
Person-level Health Data from Clinics
Extract, Transform & Load
OMOP CDM Instance
OMOP GIS Extension
Standardized Best-Practice Analytics
Reliable Evidence about Health Impacts
Patient Address Info
Geocode
Geolocation
Climate
Toxicants
SDoH
Public Place-related Data Catalog
Retrieve & Stage Data

GIS Workgroup: 2024 Accomplishments

- **Released GIS Extension to OMOP Schema**
  - https://ohdsi.github.io/GIS
- **Released draft vocabulary for standard representation of geospatial attributes that affect health:** https://github.com/OHDSI/GIS/tree/main/vocabularies
  - GIS Vocabulary
  - Exposome Vocabulary
  - SDoH Vocuablary
- Tested GAIA Core schema and code for ETLing geospatial data
- Developed flexible data cataloging approach
- Developed flexible geocoding approach
- Transitioned to a flexible use case-driven management structure
- Developed roadmaps for key infrastructure components

GIS Workgroup: 2025 Goals Infrastructure

- Roadmap for GAIA
- Roadmap for vocabulary
- Roadmap for cataloging
- Roadmap for geocoding
- Roadmap for containerization

GIS Workgroup: Use cases

- Climate and HIV mortality in Africa (Inspire)
- Temperature and health (Boston area test case for vocab validation)
- SDoH and intensive care (Bridge2AI)
- Geocoding accuracy (CLAD)
- Anaphalaxis incidence linkage to geospatial exposures (Linying Zhang)
- Highway bufferzone association with pulmonary outcomes
- Various informatics milestones relate to the toolchain

GIS Workgroup: Meeting times

- The primary meeting takes place on Fridays at 10 a.m. Eastern US

OHDSI Workgroup
Objectives and Key Results (OKR)

NLP Workgroup

Hua Xu

NLP Workgroup Purpose

NLP WG exists to promote the use of textual information in electronic health records (EHRs) , to facilitate the generation of evidence for observational studies.

- Develop standard representations for clinical text and NLP output data
- Build methods and tools to facilitate textual data processing
- Conduct cross-institutional studies and disseminate best practice of using textual data for real world evidence generation

NLP 2025 Objectives and Key Results

Objective 1: Conduct multi-site clinical studies that utilize both structured and textual data

Key results

1. Characterizing the Anticancer Treatment Trajectory and Pattern in Patients Receiving Chemotherapy (Oncology) – annotation guideline development
2. Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study (Psychiatry) – revising the protocol
3. Social Determinants of Health and Treatment Outcomes in Type 2 Diabetes: A Multi-Site Analysis of LEGEND-T2DM study – finalizing SDoH factors and subset of Treatment-Comparator-Outcome pairs

Objective 2: Open source LLMs for information extraction from clinical notes

Key results

1. Development of Kiwi - An LLM-based Clinical Information Extraction System
2. Expanding Kiwi for open-source community development

Objective 3: Knowledge dissemination

Key results

1. Monthly presentations on NLP advances (~ 30 attendees on average in 2024)
2. Book of OHDSI – NLP Chapter Textbook -> Cookbook

OHDSI Medical Device WG - Device ID Data subgroup

Subgroup Leads: Asiyah Lin

- **Year 2025 OKR: Deep dive into OHDSI vocab – device & procedure**
  - Complete the medical device terminology manuscript and submit for publication
  - Establish the medical device branch in OHDSI device vocab.

- **Achieved in 2024**
  - OHDSI annual symposium poster
  - Deepened collaboration with Korea group (Seng Chan You et al.) and Vocab WG (Alex Davydov, Oleg Zhuk)
  - Manuscript drafted (led by Seojeong Shin et al.)

OHDSI Medical Device WG
- Device-Generated Data subgroup

Subgroup Leads: Andrew Williams, Manlik Kwong

- **Year 2025 OKR:**
  - Develop standard strategy for managing and representing features waveform and other device-generated data.

- **Achieved in 2024**
  - Clarify OMOP Standard concept coverage gaps for features from 12-lead ECG Data and ICU monitor data
  - Demonstrate the strategy for supporting OHDSI standardized analytics across integrated EHR and waveform/"numerics" data

OHDSI Medical Device WG
- Device Adverse Event subgroup

Subgroup Leads: Michael Matheny

- **Year 2025 OKR:**

Develop plan for comparative effectiveness network study based on OHDSI site data availability & interest

- **Achieved in 2024**
  - Engaged Korea group for collaboration
  - Established the VA and Vanderbilt's capacity for medical device RWE research

Oncology Workgroup
2025 OKR

What have we done in 2024?

Goal: Enabling Observational Cancer Research

- Step 1.
  - Reviewed the Oncology use case at the OHDSI European symposium
  - Identified the use case requirement and surveyed 26 data partners

	Base Dx	Metastasis	Stage	Grade	Lymph nodes	Others (specify)	-Omics	Regimens	Radiation	Surgery	Extent	Dynamic	Episode of care	Death
Use case requirement	0.93	0.57	0.66	0.13	0	0	0.38	0.46	0.16	0.08	0.11	0.39	0.1	0.56
Vocab readiness	1	1	1	1	0.5	0.5	1	1	0.3	0.5	0.9	0.9	1	1
Model readiness	1	1	1	1	1	1	1	1	0.1	1	1	1	1	1
Available data/algorithm	0.77	0.65	0.79	0.69	0.48	0.58	0.40	0.69	0.50	0.62	0.46	0.35	0.31	0.69
Data Partners with data	20	17	20.5	18	12.5	15	10.5	18	13	16	12	9	8	18

- Step 2. Systematically assess the data readiness for oncology evidence generation

Oncology WG in 2025

Goal: Enabling Observational Cancer Research

Community support and collaboration	High quality oncology OHDSI data network	Evidence generation
1. Oncology wiki update 2. Oncology chapter in the book of OHDSI	1. Oncology data readiness assessment 2. Fixing the impediments 3. Use case driven vocab improvement	1. Guideline driven evidence generation 2. Oncology WG studies 3. Regular use case repo update

Themis working group

2025 OKRs

Themis Mission & Ethos

The goal of Themis is to provide conventions on how source data should be standardized to the OMOP CDM to support the OHDSI community to generate the evidence that promotes better health decisions and better care. When there is ambiguity on how data should be inserted into the CDM, Themis will examine the issue, create a convention and document it.

We follow the FAIR principles: Findable, Accessible, Interoperable and Reusable. *FAIRR - reproducible*

Themis makes decisions for the good of the whole community. We must compromise. We can always revisit and modify the convention. Don't let perfect be the enemy of great. And interoperability between different OMOP CDMs is great!

Themis 2024 accomplishments

- ✓ Establish a formal project management process
- ✓ Establish a repository for Themis conventions

2025 Objectives

Objectives:

1. Have one central location for OMOP CDM conventions

2. Succinct list of active issues

Objective 1 & key results

Objective #1: Have one central location for OMOP CDM conventions

Key results: Incorporate resources from other OHDSI standards into Themis documentation

- The Themis convention library will contain or link to formal CDM expansions
- The Themis convention library will link to domain specific WG documenatation

Objective 2 & key results

Objective #2: Succinct list of active issues

Key results: Clean up current GitHub issues

- Archive
- De-duplicate

Themis working group details

- Located in MS Teams
- Meetings: 1st & 3rd Thursday at 10 am Eastern Time
- All are welcome!
- #JoinTheJourney

CDM SURVEY SUB-WORK GROUP

MISSION STATEMENT

The CDM Survey Sub-Workgroup is a collaborative effort to unlock the potential of survey data within the Observational Health Data Sciences and Informatics (OHDSI) framework. We aim to develop a standardized approach for integrating survey data into the OMOP Common Data Model (CDM). This will be achieved through the development of standards, tools, and best practices to transform survey questions and responses into the CDM format. This, in turn, will empower researchers to conduct more robust analyses across diverse datasets, ultimately leading to richer insights and improved health outcomes.

OBJECTIVES AND KEY RESULTS 2024

- ✓ Establish the CDM Survey Subgroup
- ✓ Landscape Assessment
- ➢ Development and Testing – use cases and value proposition
- ❑ Find and Apply for Future Support

ESTABLISH STANDARDIZED MAPPING FOR HIGH-USE CLINICAL SURVEYS

- **Key Results**
  - Identify 3 to 5 high-use clinical surveys (e.g., PHQ-9, GAD-7, Asthma Control Test, SF-36, PROMIS)
  - Develop standardized mapping guidelines for each of these identified surveys
  - Collaborate with the OMOP Vocabulary Working Group to develop necessary vocabulary elements for these surveys in OMOP CDM where needed
  - Publish these standardized mappings on the OHDSI GitHub, ensuring accessibility for the community

ENHANCE COMMUNITY ENGAGEMENT AND CONTRIBUTION

- **Key Result**
  - Create a repository for survey mapping contributions from the community on the OHDSI GitHub
  - Host a webinar or workshop to discuss progress, share best practices, and gather feedback from the community
  - Develop and distribute a guide on how community members can contribute their survey mappings and experiences

FOSTER COLLABORATION WITH OTHER RELEVANT GROUPS AND INITIATIVES

- **Key Results**
  - Establish regular communication and joint meetings with the OMOP Vocabulary Working Group to address vocabulary needs for surveys
  - Initiate collaborations with groups such as the PhenX Toolkit community, HL7 Behavioral Health Project Group, Psychiatry Workgroup, etc. to align efforts and share resources
  - Present the workgroup's progress and findings at relevant conferences and community calls to raise awareness and gain insights

DEVELOP AND IMPLEMENT MAPPING TOOLS AND RESOURCES

- **Key Results**
  - Identify and evaluate existing tools for mapping survey data
  - Develop a user-friendly tool or set of tools that facilitate the mapping of survey data to OMOP
  - Pilot the tools with at least three different survey datasets and gather feedback for improvements
  - Document and publish a user guide for the tools, including case studies from the pilot

SECURE FUNDING AND RESOURCES FOR SUSTAINED EFFORT

- **Key Result**
  - Identify potential funding sources and grant opportunities to support the workgroup’s initiatives
  - Develop and submit grant proposals to secure funding
  - Establish partnerships with academic institutions and industry stakeholders to leverage additional resources and expertise

Nicole Gerlanc, PhD

Data Analyst Lead, Connect Study
Trans-Divisional Research Program
Division of Cancer Epidemiology and Genetics
National Cancer Institute

Email: nicole.gerlanc@nih.gov
CDM Survey Subgroup Wiki https://github.com/OHDSI/CdmSurveySubWg/wiki

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