



Clinical Guideline Review, Session 2

OHDSI Community Call
Jan. 28, 2025 • 11 am ET



Upcoming Community Calls

Date	Topic
Jan. 28	Clinical Guideline Review, Session II
Feb. 4	First Week of 2025 Workgroup OKRs/Phenotype Phebruary
Feb. 11	Second Week of 2025 Workgroup OKRs/Phenotype Phebruary
Feb. 18	Third Week of 2025 Workgroup OKRs/Phenotype Phebruary
Feb. 25	Fourth Week of 2025 Workgroup OKRs/Phenotype Phebruary
Mar. 4	Vocabulary Release Update, Winter 2025



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





Workgroup OKRs

Each year, workgroup representatives join a February community call to present the mission, objectives and key results for their respective groups. These 2-4 minute presentations are recorded and posted on the Workgroups homepage on OHDSI.org.

Please choose a date to sign up for a February date – Feb. 25 is now closed.



- Already Signed Up:**
- Africa Chapter
 - CDM Survey Subgroup
 - Clinical Trials
 - Common Data Model
 - Data Bricks
 - Evidence Network
 - Eye Care and Vision Research
 - GIS - Geographic Information System
 - Health Equity
 - Health Systems Interest Group
 - Latin America
 - Medical Devices
 - Natural Language Processing
 - Oncology
 - Pregnancy and Reproductive Health
 - Psychiatry
 - Rare Disease
 - Rehabilitation
 - Steering
 - Surgery and Perioperative Medicine
 - Themis
 - Transplant
 - Vocabulary
 - Women of OHDSI



#OHDSISocialShowcase This Week

Monday

NCO-Calibrated DID Analysis: Addressing Unmeasured Confounding in Difference-in-Differences Analyses Using Negative Control Outcomes Experiments

(Dazheng Zhang, Bingyu Zhang, Huiyuan Wang, Charles J. Wolock, Yiwen Lu, Yong Chen)



NCO-Calibrated DID Analysis: Addressing Unmeasured Confounding in Difference-in-Differences Analyses Using Negative Control Outcomes Experiments

Dazheng Zhang^{a,b}, Bingyu Zhang^{a,c}, Huiyuan Wang^{a,b}, Charles J. Wolock^b, Yiwen Lu^{b,c}, Linbo Wang^d, Yong Chen^{a,b,c,e,f,g}

^a The Center for Health Analytics and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA
^b Department of Biostatistics, University of Pennsylvania, Philadelphia, PA, USA
^c Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA
^d Department of Statistical Sciences, University of Toronto, Ontario, Canada
^e Leonard Davis Institute of Health Economics, Philadelphia, PA, USA
^f Penn Medicine Center for Evidence-Based Practice (CEBP), Philadelphia, PA, USA
^g Penn Institute for Biomedical Informatics (PIBI), Philadelphia, PA, USA

^a co-first authors



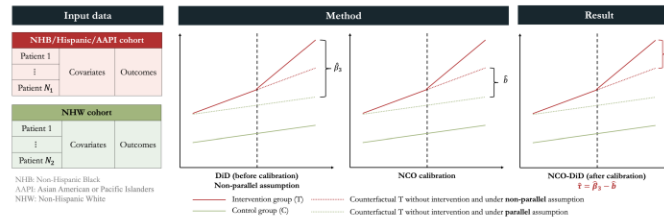
Background

- Difference-in-differences (DiD) analysis is a statistical method used to estimate causal effects by comparing changes in outcomes over time between an intervention group and a control group.
- A critical challenge in DiD analysis is the reliance on the parallel trends assumption, which assumes that in the absence of intervention, the intervention and control groups would, on average, have followed a parallel trajectory over time. The violation of such assumption can lead to diverging trends between the control and intervention groups and, if not accounted for, can introduce a systematic bias to the estimated effects of the intervention.
- Negative control outcome (NCO) experiments¹, which assume no intervention effect on the outcome, have been used to calibrate the systematic bias such as the unmeasured confounding bias. However, existing methods for NCO experiments are generally limited to the regression analysis rather than being adapted to the DiD framework.
- **Goal:** To develop an NCO-calibrated DiD method that addresses time-varying systematic bias from unmeasured confounding variables.

Method

- **Main steps of NCO-DiD**
- **Step 1**, we used the propensity score to match the treatment group to the control group. To implement the DiD method, we apply the log-linear model to the matched cohort:

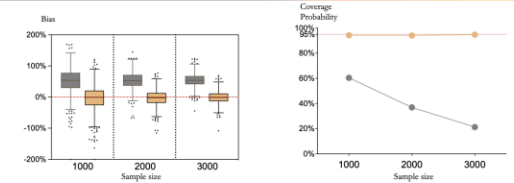
$$\log(E(Y|A, T)) = \beta_0 + \beta_1 A + \beta_2 T + \beta_3 AT,$$
 - β_0 is a constant, and β_1 , β_2 , and β_3 are coefficients of A , T , and their interaction.
 - β_3 represents the intervention effect in risk ratio (RR), which may be affected by systematic bias b .
- **Step 2**, we repeat this procedure using the NCOs, assuming that the intervention does not affect these outcomes. Applying this procedure to the NCOs provides an estimate \hat{b} of the systematic bias. If $b = 0$, this suggests that the parallel trends assumption holds. Based on \hat{b} , we derive a test statistic and corresponding two-sided test of the null hypothesis, $H_0: b = 0$.
- **Step 3**, we calibrate β_3 by subtracting the estimated bias, yielding the calibrated estimator $\hat{\tau} = \hat{\beta}_3 - \hat{b}$.



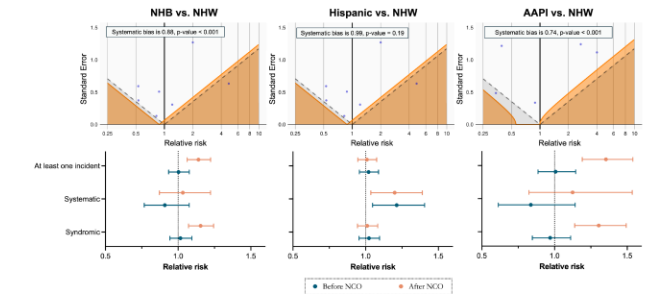
- **Simulation study:** We performed extensive simulation studies covering a wide range of settings and compared the estimates obtained from NCO-DiD (after calibration) and baseline DiD method (before calibration) in terms of the relative bias to the true value of parameters.

Contact: dazheng.zhang@pennmedicine.upenn.edu, ychen123@upenn.edu

Results



- The boxplots above display the simulation results. Across sample sizes ranging from 1,000 to 3,000, the proposed NCO-DiD method demonstrates higher accuracy, with a smaller relative bias to the true value and better coverage probability, compared to the baseline DiD method.
- **Data application:**
 - Forest plots below display comparisons between the racial/ethnic differences after COVID-19 infection in “long-covid” symptoms and conditions from baseline DiD method (before calibration) and proposed NCO-DiD (after calibration). Compared with baseline method, the proposed NCO-DiD method identifies racial/ethnic differences for racial/ethnic differences across all minor racial/ethnic groups (NHB, Hispanic, and AAPI).



Conclusions

- NCO-DiD is an effective causal inference framework to calibrate systematic bias from unmeasured confounding variables for DiD model.
- We illustrate the great potential of the proposed NCO-DiD method for OHDSI study in real-world evidence generation.

Reference

Schuemij MJ, Hripcak G, Ryan PB, et al. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. Proceedings of the National Academy of Sciences 2018; 115: 2571–2577.
Zhang, D., Zhang, B., Wu, ..., & Chen, Y. (2024). Racial/Ethnic Differences in Long-COVID-Associated Symptoms among Pediatrics Population: Findings from Difference-in-differences Analyses in RECOVER Program. Invited Revision at Nature Communications.



#OHDSISocialShowcase This Week

Tuesday

Improving Team Science Through “Thons” Reflections on the April Olympians Community Event

(Clair Blacketer, Melanie Philofsky, Evanette Burrows, Maxim Moinat, Katy Sadowski)

Improving Team Science through “Thons” Reflections on the April Olympians Community Event

PRESENTERS: Clair Blacketer
Melanie Philofsky

INTRO:

- Regular OHDSI community events serve as platforms for intense, team-driven science.
- In April 2024, the April Olympians event convened with the goal to develop an ETL convention library.
- Here we detail the methodology and insights gained on optimizing global team science

METHODS

Preparation Phase

- The CDM, DQ, and THEMIS working group leads held weekly one-hour planning sessions.
- Tasks divided into three teams



Hunters of Artemis



Writers of Apollo



Builders of Hephaestus

- Field guides, issue templates and documentation crafted for each team prior to kick-off.

Execution Phase

- One-hour kick off meeting and bi-weekly 30-minute check-ins to accommodate time zones.
- Tasks broken into 15-minute segments.
- Team leads prioritized rapid responses and communication.

RESULTS

- 80 github issues closed driven by 20 collaborators.



Figure 3. OHDSI Project Board

Comprehensive preparation and clear communication are critical factors in the success of community events



Take a picture to download the full paper

RESULTS CONT.

- Convention Library created



Figure 2. OHDSI DQC, EXCELISE conventions

Insights from Team Science

DO

- Prepare comprehensive materials
- Be responsive
- Break tasks into small chunks
- Test Permissions
- Empower participants
- Recognize contributions
- Regular read-outs

DON'T

- Use intimidating language
- Over-schedule meetings
- Create complex task descriptions

CONCLUSION

- The April Olympians event culminated in the creation of the THEMIS repository.
- The initiative provided valuable lessons in team science.

Clair Blacketer^{1,2*}, Melanie Philofsky^{3*}, Evanette Burrows⁴, Maxim Moinat⁵, Katy Sadowski⁶
¹Common Research & Development, Boston, MA; ²Department of Medical Informatics, Erasmus MC, Rotterdam, NL; ³EPAM Systems, Inc., Cambridge, MA, USA; ⁴Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT
^{*} See first authors

Johnson & Johnson

Boehringer Ingelheim <epam>

Erasmus MC

OHDSI

#OHDSISocialShowcase This Week

Wednesday

Clinically validated line of therapy (LoT) algorithm for patients with metastatic Non-Small Cell Lung Cancer (mNSCLC) can be implemented using systemic anti-cancer therapy (SACT) in OMOP database

(Joana Moreira, Fabian Acker, Jack Brewster, Anne-Lore Bynens, Susan Lara Cheeseman, Francesca Fusco, Åslaug Helland, Lizza Hendriks, Pooja Jain, Rosie McDonald, Sarah Seager, Andrea Wolf, Åsa Öjlert, Francesca Ogliari)

Clinically validated line of therapy (LoT) algorithm for patients with metastatic non-small cell lung cancer (mNSCLC) can be implemented using systemic anti-cancer therapy (SACT) in Observational Medical Outcomes Partnership (OMOP) database

J Moreira¹, F Acker², J Brewster³, A-L Bynens⁴, S Cheeseman⁵, F Fusco⁶, Å Helland⁷, L Hendriks⁸, P Bhatnagar⁹, R McDonald³, A Wolf⁸, J Yeap³, Å Öjlert⁷, F Ogliari^{10*}

*Joint last authors; **corresponding author. ¹IQVIA Ltd, Portugal. ²Goethe University Frankfurt, University Hospital, Germany. ³IQVIA Ltd, UK. ⁴Maastricht University Medical Center+, The Netherlands. ⁵Leeds Teaching Hospital NHS Trust, UK. ⁶IRCCS Regina Elena National Cancer Institute, Rome, Italy. ⁷Oslo University Hospital, Norway. ⁸University Hospital Frankfurt, University Cancer Center, Germany. ⁹IRCCS San Raffaele Scientific Institute, Milan, Italy.

CONCLUSION

- Consistent application of a LoT algorithm is crucial for describing treatment and prognosis by LoT in multi-center cancer research.
- DigiONE introduced a **clinician-developed LoT algorithm** to group SACT prescribed for mNSCLC.
- The LoT algorithm **has been tested on US and UK OMOP databases**. There is ongoing validation in other European centers to assess its generalizability.
- The algorithm can be shared with researchers in the OHDSI community once finalized and is **most applicable within Europe** where patients are managed similarly, and EMA approvals are practised.

DigiCore Corresponding author: Ogliari.Francesca@hr.it
OHDSI Global Symposium 2024, Oct 23-24, New Jersey

INTRODUCTION

- An accurate assignment of line of therapy (LoT) received is important in observational studies to assess response to therapy and prognosis, patient suitability for interventional trials, and for clinical audits¹.
- Currently, no common definition exists for LoT between hospitals and research groups which is straightforward to code.
- Here, Digital Oncology Network for Europe (DigiONE) introduces the approach to developing a clinically validated LoT algorithm specifically for mNSCLC and its key principles.
- The fundamental concept of LoT algorithm is that LoT advances when there is clinical progression of disease. However, since date of progression is typically manually inputted, which can vary in consistency across hospitals, this LoT algorithm infers disease progression based on drug-level data.
- The LoT algorithm developed considers SACT prescribed for mNSCLC with palliative intent in the real-world, including the use of any trial drugs.

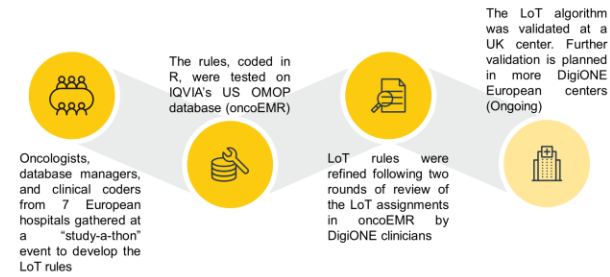
- The current DigiONE mNSCLC LoT algorithm rules (principles described in Table 1) were applied to oncoEMR in OMOP.
- Patients with another non-NSCLC primary malignancies were excluded to avoid capturing SACT prescribed for other malignancies.

Table 1. Principles of the DigiONE mNSCLC LoT algorithm

Rule	Rule definition
Start date of LoT	Earliest drug start date in the LoT. LoT may start before mNSCLC diagnosis due to early SACT initiation based on suspected metastases prior to confirmation from biopsy results
Grouping SACT into LoT	A LoT can consist of one or multiple regimens, and regimens may include one or multiple drugs with different start dates. Drugs that share the same start date are considered as a 'protocol'
Treatment changes that do not advance the LoT	If one or more drugs are stopped while other concurrently prescribed drugs continue If the dosage or administration route is changed, but the drug continues to be prescribed Switching between certain drugs which are presumed to be for toxicity reasons rather than for clinical progression of disease. Examples include • carboplatin and cisplatin • paclitaxel and nab-paclitaxel • pemetrexed, vinorelbine and gemcitabine with the same platinum-based therapy partner • PD1 and PDL1 inhibitors • first- and second-generation EGFR TKIs • targeted therapies that target the same mutations Stopping a drug for any duration if the same drug is initiated after the break
Treatment changes that do advance the LoT	Addition of a new drug that is not concurrently prescribed with other drugs, unless the drug is in the list of allowable switches due to changes for toxicity reasons
End of LoT	The latest end date of drugs prescribed within the LoT. If a patient has a date of death prior to the end date of the treatment in the database, the date of death is used as LoT end

METHODS

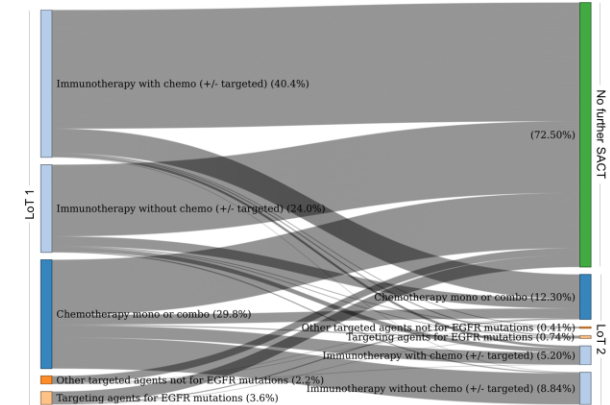
The DigiONE mNSCLC LoT algorithm development and validation involved four steps:



RESULTS

- In oncoEMR, there were 2,302 eligible patients of which 52.6% received a 1st LoT and 14.5% received a 2nd LoT for mNSCLC (Figure 1).
- Reasons a patient may not initiate SACT for mNSCLC include they are deemed too unfit for SACT and therefore receive best supportive care, patient refuses treatment plan, or the patient dies before treatment is initiated². This finding that approximately half of patients receive SACT for mNSCLC is aligned with clinical expectations³.

Figure 1. Sankey diagram of treatment changes from 1st to 2nd LoT in oncoEMR



References
 1. Sani K. S., Twelves C. Determining lines of therapy in patients with solid cancers: a proposed new systematic and comprehensive framework. Br J Cancer. 2021;125(2):155-63.
 2. Hendriks L. E., Kerr K. M., Menis J., Mok T. S., Neill U., Passaro A., et al. Non-oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2023;34(4):358-76.
 3. Hofmarcher T., Lindgren P., Wilking N. Systemic anti-cancer therapy patterns in advanced non-small cell lung cancer in Europe. J Cancer Policy. 2022;34:100362.

Acknowledgements
 This work is part of the DigiCore pan-European research network comprised of academic cancer centres and two industrial members (IQVIA and Illumina). The DigiONE project was funded by IQVIA and Illumina.



#OHDSISocialShowcase This Week

Thursday

Impact of phenotype error adjustment on background incidence of COVID-19 vaccine adverse events of special interest

(James Weaver, Patrick B. Ryan, Victoria Strauss, Marc A. Suchard, Joel Swerdel, Daniel Prieto-Alhambra)

Impact of phenotype error adjustment on background incidence of COVID-19 vaccine AESIs

PRESENTER: James Weaver

INTRO:

- Substantial background IR heterogeneity reported across age, sex, and database for COVID-19 vaccine AESIs[1].
- Unclear what proportion of observed IR heterogeneity is attributable AESI phenotype error (outcome misclassification e.g., low sensitivity).
- Assessed if adjusting AESI IRs for phenotype error reduced heterogeneity.
- Evaluated the impact of phenotype error adjustment on background incidence of 5 AESIs in 5 databases, stratified by age x sex.

METHODS:

- US databases:** Clinformatics®, Merative® CCAE, Merative® MDCCD, Merative® MDCR, Optum EHR®.
- AESIs:** inpatient AMI, appendicitis, DVT, non-hemorrhagic stroke, PE
- IR method:** outcomes/100,000 person-years at-risk stratified by age x sex during 3 years before COVID19 pandemic
- Validation:** internal probabilistic reference standard stratified by age x sex
- Simple adjustment:**
 $outcomes_{Adj_i} = (outcomes - (1-SP) * persons_{at-risk} / (SP - (1-SP)))$

Probabilistic adjustment (QBA principles)

Apply simple adjustment to 10,000 draws from non-symmetrical beta distribution specified where $\mu = \text{phenotype error point estimate}$ and $\sigma = \text{phenotype error SD across databases}$.

$IR_{Adj_i} = \text{median}(IR) [95\% SI]$

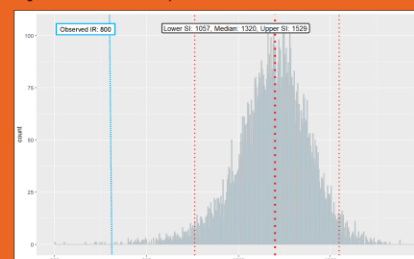
- Pooled IR adjustment**
 DerSimonian-Laird random effects meta-analysis

Adjusting background incidence of COVID-19 vaccine AESIs for phenotype error did not reduce heterogeneity across US data sources stratified by age x sex

Table 1: Age x sex stratum with greatest and least impact of simple phenotype error adjustment

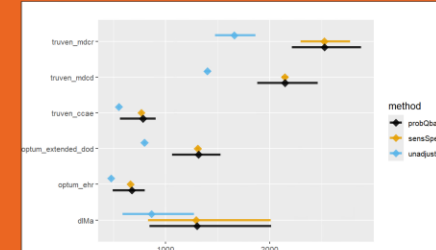
AESI	Stratum	SN	SP	IR obs.	IR adj.	Rel. diff.	EAME
Greatest impact							
AMI	F 18-34	0.0775	0.9999	111	113	10.27	1.01
Appendicitis	M 35-54	0.8889	0.9998	137	273	1.99	0.30
DVT	M 18-34	0.8971	0.9999	69	110	1.59	0.20
Non-hemorrhagic stroke	F 18-34	0.1281	0.9999	26	191	7.35	0.87
PE	M 18-34	0.3951	0.9999	33	74	2.24	0.35
Least impact							
AMI	M >=85	0.6429	0.9958	1224	70	0.06	1.24
Appendicitis	M 75-85	0.9235	0.9995	49	29	0.59	0.23
DVT	F >=85	0.8709	0.9997	707	174	0.25	0.61
Non-hemorrhagic stroke	F >=85	0.6703	0.9975	834	931	1.12	0.05
PE	M 75-84	0.8934	0.9965	469	341	0.73	0.14

Figure 1: Probabilistic AMI IR adjustment results for M 55-64 in Clinformatics®



Key - AMI: acute myocardial infarction, IR: incidence rate in events/100,000 person-years, M 55-64: males 55-64 years of age, Median: median incidence rate from incidence rate distribution, SI: simulation interval, y-axis: n simulations of corresponding x-axis values, x-axis: adjusted IR values from sensitivity and specificity distributions

Figure 2: Source and pooled AMI IR forest plot for M 55-64: unadjusted, simple adjusted, probabilistic adjustment



Key - truven_mdcr: Merative® MarketScan® Medicare Database, truven_mdod: Merative® MarketScan® Multi-state Medicaid database, truven_ccae: Merative® MarketScan® Commercial Database, optum_extended_dod: Optum's de-identified Clinformatics® Data Mart Database, optum_ehr: Optum® de-identified Electronic Health Record Dataset, probObs: probabilistic quantitative bias analysis adjustment, sensSpec: simple bias analysis using sensitivity and specificity

EVALUATION:

- Relative difference = IR_{Adj_i} / IR
- EAME = $(\text{abs}(\log(IR_{Adj_i} / IR)))$
- Pooled EAME: log relative pooled IR change after adjustment
- Relative difference indicates impact direction and multiplicative magnitude
- EAME indicates directionless linear impact magnitude
- T2 indicates pooled IR heterogeneity

RESULTS

- The greatest impact across AESIs was mostly in younger age strata (18-34) for rare events (AMI, non-hemorrhagic stroke) in Clinformatics®. The least impact was mostly in older age strata (>=75) in Optum EHR® [Table 1].
- The greatest impacts were associated with IR increase after adjustment. The least impacts were mostly associated with IR decrease after adjustment.
- Simple and probabilistic adjustment increased AMI IRs in all databases similarly among males 55-64y. This increase was also observed in the pooled analysis [Figure 1].
- The AMI meta-analysis T2 metric was greater in both adjusted analyses vs the unadjusted analysis. Meta-analysis results for other strata and AESIs were qualitatively similar.

CONCLUSION

Adjusting COVID-19 vaccine AESI IRs for phenotype error did not reduce observed heterogeneity

LIMITATIONS

- Results from 5 inpatient conditions studied in 5 US databases is not generalizable
- Further probabilistic results interpretation required
- More sources needed for PI assessment

James Weaver, Patrick B. Ryan, Victoria Strauss, Marc A. Suchard, Joel Swerdel, Daniel Prieto-Alhambra



Take a picture to view the abstract





#OHDSISocialShowcase This Week

Friday

End-to-End Implementation of a Workflow for Validating Semantic Mappings and Constructing Ontology Extensions

(Jared Houghtaling, Polina Talapova, Soojin Park, Harry Caulfield, Andrew Williams)

End-to-End Implementation of a Workflow for Validating Semantic Mappings and Constructing Ontology Extensions

Presenter: Jared Houghtaling

Intro:

The Bridge2AI for Clinical Care (B2AI for CC) research consortium aims to capture and consolidate risk in clinical data from the data contributing sites in order to support complex analytic processes in machine learning (ML) and artificial intelligence (AI), such as consolidation and analytic support is vital and requires a diversity of expertise and consortium-specific OMO concepts for interacting with multimodal (e.g. images, waveforms) files alongside OMO P-shaped datasets. In this work, we demonstrate a novel, cross-platform approach that provides a user-friendly endpoint (i.e. Google Sheets) for clinical experts to evaluate mapping representations using A Simple Standard for Sharing Ontology Mappings (SSSOM) format. The methodology is generalizable and allows for drafting, sharing, searching, and validating semantic mappings to standard OMO concepts that will be made available to the entire OHDSI community to facilitate reuse of validated mappings, automation of ETLs, and adherence to ETL conventions.

Methods:

- Data entry in SSSOM-specific collaborative spreadsheets**
We chose to design and create a simple Google Sheet environment to its familiar, ease-of-use, collaborative features, and App Script integrations.²
- Cross-platform data integration processes**
We created a Google AppScript to transfer mappings from the Google environment to GitHub at regular intervals to ensure a robust version history and to make those mappings readily available to GitHub Actions workflows.
- OHDSI Vocabulary Building Processes**
We coordinated execution using a GitHub action that (1) logs the mapping, (2) validates syntax and domain for, (3) evaluates if the user has been mapped terms and the existing B2AI term set, (4) assigns concept id values in appropriate 28+ range, (5) inserts new terms into consolidated vocabulary tables, and (6) commits the vocabulary back to GitHub for uptake.
- Ontology Validation and Cross-annotation**
Following validation, we migrated the data vocab reports to a public repository and placed it alongside scripts that enable data contributing sites to assign their local vocabulary tables.

7-Step Process:

- Mappings assigned to clinical experts, review polished by App Script
- Feedback provided to experts on mapping app each performance
- AppScript regularly commits mapping updates to GitHub repository
- Updates to mappings trigger vocabulary build: workflow
- GitHub Action executed on self-hosted runner deployed as ContainerApp in Azure cloud
- Self-hosted runner performs queries against B2AI vocabulary table to integrate mappings
- New version of B2AI ontology (data tables, pg_dump, roles) pushed to GitHub

CHO RUS/B2AI Mapping Repo: SSSOM Mappings (Extract, Broad, Narrow, Unlabeled) feed into Vocabulary Builder (SSSOM Instance, UI Assignment, Insert, Update, Merge, Validate Changes). This feeds into B2AI Ontology (Unlabeled, Narrow, Unlabeled, Unlabeled, Unlabeled, Unlabeled). The process also involves Formal Submissions of Ontology to OHDSI Vocabulary and Distribution to Data Contributing Sites.

SSSOM-Format Input for Validation:

course_code	predicate_id	course_vocab_id	course_description
B2AI: 1000245	skos:relatedMatch	B2AI	Agents Helix %
B2AI: 1000165	skos:relatedMatch	B2AI	Agents M20 M20
B2AI: 1000165	skos:relatedMatch	B2AI	Agents Helix Ho

OMOP-Format Output for Implementation and Analysis:

concept_id	concept_name	concept_id_1	concept_id_2
200000415	Agents Helix %	200000415	19027073
200000442	Agents M20 M20	200000442	19021129
200000297	Agents Helix Ho	200000297	19027073

Results & Discussion:

This tool, we have successfully applied the procedure described above to integrate more than 1000 unique concepts, representing more than 50,000 distinct rows of data across various vocabulary tables, into a B2AI-specific ontology. Now that the workflow is established, the ontology is growing organically as daily mapping updates or walkdatabases are captured and consolidated daily with minimal effort. We expect to define and implement robust protocols for contributing mappings and taking clinical experts in the validation process.

Note that we rely on the CONCEPT SYNONYM table to capture site-specific identifiers in source site descriptions. More specifically, we capture upstream representations of concepts at the site in the synonym table such that sites are able to link their source data descriptions directly on this table to ensure the B2AI concepts associated with their particular description. We are also pushing direct support to sites to incorporate these new terms into their existing Extract, Transform, and Load (ETL) processes locally, both semantically (i.e. updating vocabulary tables) and structurally (i.e. updating codebase).

Conclusions

The ontology pipeline we've established is a representative complete pathway to meet clinical experts and life operable mapping in clinical ops. The work builds on prior SSSOM entity standardization and makes use of cross-platform automation strategies to increase ease-of-use, transparency, and collaboration. Because of the open-source nature of the tooling, we expect that this workflow can serve as a model for other consortia, or individual institutions, that require standardized processes to make consistent or fill gaps in the OMO vocabulary in a robust and useable manner to dynamic table recently implemented the pipeline for the Geospatial Information Systems (GIS) working group, and expect to apply it to other use cases in the coming months.

Authors: Jared Houghtaling*, Polina Talapova*, Soojin Park†, J. Harry Caulfield‡, Andrew E. Williams*

*Tufts Medicine - Institute for Clinical Research and Health Policy Studies (ICRH-HP)
†Columbia University - Vagelos College of Physicians and Surgeons
‡Lawrence Berkeley National Laboratory - Berkeley Bioinformatics Open-Source Projects (BBOP)





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