



April Olympians #3 / Tools to Evaluate ETL

OHDSI Community Call
April 16, 2024 • 11 am ET



Upcoming Community Calls

Date	Topic
April 16	April Olympians Update Presentation: Tools to Evaluate ETL
April 23	April Olympians Update Presentation: Themis & CDM Process Overview
April 30	April Olympians Update Presentation: What We Achieved & How You Can Use It
May 7	DevCon 2024 Review
May 14	10-Minute Tutorials
May 21	Open Studies in the OHDSI Community
May 28	Collaborator Showcase Brainstorm
June 4	NO CALL – EUROPEAN SYMPOSIUM
June 11	European Symposium Review
June 18	Application of LLMs In Evidence Generation Process
June 25	Recent OHDSI Publications



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!



Correspondence

Congratulations to the team of **Nhung TH Trinh, Annika M Jödicke, Martí Català, Núria Mercadé-Besora, Saeed Hayati, Angela Lupattelli, Daniel Prieto-Alhambra, and Hedvig ME Nordeng** on the publication of **Effectiveness of COVID-19 vaccines to prevent long COVID: data from Norway** in *The Lancet Respiratory Medicine*.

Effectiveness of COVID-19 vaccines to prevent long COVID: data from Norway

Our recent study using data from more than 20 million participants has shown that COVID-19 vaccines consistently prevent long COVID symptoms in adults, with meta-analytic calibrated subdistribution hazard ratio (sHRs) of 0.54 (95% CI 0.44-0.67) in CPRD GOLD, 0.48 (0.34-0.68) in CPRD AURUM, 0.71 (0.55-0.91) in SIDIAP, and 0.59 (0.40-0.87) in CORIVA.¹ In addition, when considering post-COVID thromboembolic and cardiovascular complications as outcomes of interest, recently published data have shown that vaccination with any COVID-19 first vaccine dose (ChAdOx1, BNT162b2, and mRNA-1273) is associated with reduced risk of post-acute heart failure (0.45 [0.38-0.53] 0-30 days after SARS-CoV-2 infection; 0.61 [0.51-0.73] 91-180 days after SARS-CoV-2 infection), venous thromboembolism (sHR 0.22 [95% CI 0.17-0.29] 0-30 days after SARS-CoV-2 infection; 0.53 [0.40-0.70] 91-180 days after SARS-CoV-2 infection), and arterial thrombosis (0.53 [0.44-0.63] 0-30 days after SARS-CoV-2 infection; 0.72 [0.58-0.88] 91-180 days after SARS-CoV-2 infection).² With the use of the Observational Medical Outcomes Partnership (OMOP) common data model (CDM), all our analyses were conducted across three European countries (Estonia, Spain, and the UK) without transferring patient data, using federated analyses similar to those used by the European Medicines Agency-funded Data Analysis and Real World Interrogation Network. Here, we show further reproducibility of our findings in the entire Norwegian population of approximately 5.4 million inhabitants. Data from six registries covering primary and secondary care, hospitalisations, vaccinations, communicable disease notifications, prescriptions, and sociodemographic factors between 2018 and 2021 were mapped to the OMOP CDM. Reproducing previous methods (appendix p 25 and our previous study),¹ we generated four study cohorts in line with the Norwegian vaccination campaign rollout between Jan 9, 2021, and Aug 6, 2021: people aged 75 years and older (cohort one); 65 years and older and clinically extremely vulnerable people, and those with underlying health conditions aged 18 years and older (cohort two); 18 years and older with underlying conditions (cohort three); and 18 years and older (cohort four; appendix p 3). We then applied the publicly available scripts to assess the effectiveness of COVID-19 vaccines to prevent long COVID and post-acute complications. A total of 2 364 651 vaccinated and 1 532 935 unvaccinated individuals in Norway were included (appendix p 18). Of the vaccinated individuals, 1576 (0.09%) developed at least one of the 25 WHO-listed symptoms recorded at between 90 and 365 days after the date of a COVID-19 positive test or diagnosis, with no record

of that symptom 180 days before SARS-Cov-2 infection, and were therefore identified as long COVID cases, compared with 2922 (0.17%) of the unvaccinated individuals (table). Background characteristics of the study population by cohorts are presented in the appendix (p 4). Adequate covariate balance between the vaccinated and unvaccinated groups was achieved after weighting, as shown in the appendix (p 6). Information regarding follow-up time and censoring information are summarised in the appendix (p 8). Overall, vaccination with any COVID-19 vaccine (namely BNT162b2, mRNA-1273, and ChAdOx1) reduced the risk of developing long COVID symptoms across all study cohorts: the meta-analytic (sHR) was 0.64 (95% CI 0.55-0.74) if vaccinated people were censored at the second vaccine dose (appendix p 19), in line with our previous study.¹ Sensitivity analyses without censoring on the second dose for vaccinated groups provided similar results: meta-analytic sHR was 0.55 (0.46-0.66; appendix p 20). Estimates obtained by Fine-Gray method and Cox regression are highly consistent (appendix p 9). As Norway suspended the use of ChAdOx1 vaccine on March 11, 2021, we did not perform comparative effectiveness analyses between BNT162b2 and ChAdOx1, as was done in the main Article.

Lancet Respir Med 2024
Published Online
April 10, 2024
[https://doi.org/10.1016/S2213-2600\(24\)00082-1](https://doi.org/10.1016/S2213-2600(24)00082-1)
See Online/Articles
[https://doi.org/10.1016/S2213-2600\(23\)00414-9](https://doi.org/10.1016/S2213-2600(23)00414-9)
See Online for appendix

For the scripts see <https://github.com/oxford-pharmacoep/LongCovidVaccineEffectiveness>
For more on post-acute complications see <https://github.com/oxford-pharmacops/vaccineEffectOnPostCovidCardiacThromboembolicEvents>

	Vaccinated			Unvaccinated		
	Individuals	COVID-19	Long COVID	Individuals	COVID-19	Long COVID
Cohort 1	397 174	782	168 (21.48%)	224 223	4133	751 (18.26%)
Cohort 2	434 723	3266	520 (15.92%)	321 977	7000	643 (9.19%)
Cohort 3	263 057	2814	370 (13.15%)	438 151	18 544	1267 (6.83%)
Cohort 4	1 469 637	39 210	518 (1.32%)	548 584	41 973	261 (0.62%)

Data shown are n or n (%). Exposure is any COVID-19 vaccine. Outcome is having at least one WHO-listed symptom 90 days or more following SARS-CoV-2 infection with no history of that symptom in the previous 180 days (long COVID). Cohort 1 includes individuals aged 75 years and older, cohort 2 includes individuals aged 65 years and older and clinically extremely vulnerable people, and those with underlying health conditions aged 18 years and older, cohort 3 includes individuals aged 18 years and older with underlying conditions, and cohort 4 includes individuals aged 18 years and older. Unvaccinated people can be included in different cohorts



OHDSI Shoutouts!



Congratulations to the team of **Guy Tsafnat, Rachel Dunscombe, Davera Gabriel, Grahame Grieve, and Christian Reich** on the publication of **Converge or Collide? Making Sense of a Plethora of Open Data Standards in Health Care** in the *Journal of Medical Internet Research*.

JOURNAL OF MEDICAL INTERNET RESEARCH

Tsafnat et al

Editorial

Converge or Collide? Making Sense of a Plethora of Open Data Standards in Health Care

Guy Tsafnat^{1,2,3}, PhD; Rachel Dunscombe^{4,5*}, MICT; Davera Gabriel^{1,3,6*}, RN; Grahame Grieve^{7,8*}, PhD; Christian Reich^{3,9*}, BSc, MD

¹Evidentli Pty Ltd, Surry Hills, Australia

²Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Macquarie Park, Australia

³OHDSI OMOP + FHIR Working Group

⁴openEHR International, St. Helens, United Kingdom

⁵Imperial College London, London, United Kingdom

⁶School of Medicine, Johns Hopkins University, Baltimore, MD, United States

⁷Health Level 7 International, Ann Arbor, MI, United States

⁸Health Intersections Pty Ltd, Melbourne, Australia

⁹Odysseus Data Services, Cambridge, MA, United States

*these authors contributed equally

Corresponding Author:

Guy Tsafnat, PhD
Evidentli Pty Ltd
50 Holt St
Suite 516
Surry Hills, 2010
Australia
Phone: 61 415481043
Email: guyt@evidentli.com

Abstract

Practitioners of digital health are familiar with disjointed data environments that often inhibit effective communication among different elements of the ecosystem. This fragmentation leads in turn to issues such as inconsistencies in services versus payments, wastage, and notably, care delivered being less than best-practice. Despite the long-standing recognition of interoperable data as a potential solution, efforts in achieving interoperability have been disjointed and inconsistent, resulting in numerous incompatible standards, despite the widespread agreement that fewer standards would enhance interoperability. This paper introduces a framework for understanding health care data needs, discussing the challenges and opportunities of open data standards in the field. It emphasizes the necessity of acknowledging diverse data standards, each catering to specific viewpoints and needs, while proposing a categorization of health care data into three domains, each with its distinct characteristics and challenges, along with outlining overarching design requirements applicable to all domains and specific requirements unique to each domain.

(*J Med Internet Res* 2024;26:e55779) doi: [10.2196/55779](https://doi.org/10.2196/55779)



OHDSI Shoutouts!



Congratulations to the team of **Ailbhe Lawlor, Carol Lin, Juan Gómez Rivas, Laura Ibáñez, Pablo Abad López, Peter-Paul Willemse, Muhammad Imran Omar, Sebastiaan Remmers, Philip Cornford, Pawel Rajwa, Rossella Nicoletti, Giorgio Gandaglia, Jeremy Yuen-Chun Teoh, Jesús Moreno Sierra, Asieh Golozar, Anders Bjartell, Susan Evans-Axelsson, James N'Dow, Jihong Zong, Maria J. Ribal, Monique J. Roobol, Mieke Van Hemelrijck, Katharina Beyer, on behalf of the PIONEER Consortium** on the publication of **Predictive Models for Assessing Patients' Response to Treatment in Metastatic Prostate Cancer: A Systematic Review in European Urology Open Science.**

EUROPEAN UROLOGY OPEN SCIENCE 63 (2024) 126–135

available at www.sciencedirect.com
journal homepage: www.eu-openscience.europeanurology.com



Review – Prostate Cancer

Predictive Models for Assessing Patients' Response to Treatment in Metastatic Prostate Cancer: A Systematic Review

Ailbhe Lawlor^{a,}, Carol Lin^b, Juan Gómez Rivas^c, Laura Ibáñez^c, Pablo Abad López^d, Peter-Paul Willemse^e, Muhammad Imran Omar^f, Sebastiaan Remmers^b, Philip Cornford^g, Pawel Rajwa^h, Rossella Nicoletti^{i,j}, Giorgio Gandaglia^{k,l}, Jeremy Yuen-Chun Teoh^j, Jesús Moreno Sierra^c, Asieh Golozar^{l,m}, Anders Bjartellⁿ, Susan Evans-Axelsson^o, James N'Dow^p, Jihong Zong^q, Maria J. Ribal^p, Monique J. Roobol^b, Mieke Van Hemelrijck^a, Katharina Beyer^b, on behalf of the PIONEER Consortium*

^aTranslational Oncology and Urology Research (TOUR), King's College London, London, UK; ^bDepartment of Urology, Erasmus MC Cancer Institute, Erasmus University Medical Centre, Rotterdam, The Netherlands; ^cDepartment of Urology, Health Research Institute, Hospital Clinico San Carlos, Madrid, Spain; ^dDepartment of Urology, Hospital Universitario La Paz, Madrid, Spain; ^eDepartment of Oncological Urology, University Medical Center, Utrecht Cancer Center, Utrecht, The Netherlands; ^fAcademic Urology Unit, University of Aberdeen, Aberdeen, UK; ^gLiverpool University Hospitals NHS Trust, Liverpool, UK; ^hDepartment of Urology, Medical University of Silesia, Zabrze, Poland; ⁱDepartment of Experimental and Clinical Biomedical Science, University of Florence, Florence, Italy; ^jS.H. Ho Urology Centre, Department of Surgery, The Chinese University of Hong Kong, Hong Kong, China; ^kDepartment of Urology and Division of Experimental Oncology, Urological Research Institute, IRCCS San Raffaele Hospital, Milan, Italy; ^lOHDSI Center, Northeastern University, Boston, MA, USA; ^mOdysseus Data Services, New York, NY, USA; ⁿDepartment of Translational Medicine, Lund University, Malmö, Sweden; ^oBayer AB, Medical Affairs Oncology, Stockholm, Sweden; ^pEuropean Association of Urology, Guidelines Office, Arnhem, The Netherlands; ^qBayer Healthcare, Global Medical Affairs Oncology, Whippany, NJ, USA

Article info

Article history:
15 January 2024
5 February 2024
Accepted March 20, 2024

Associate Editor:
Roderick van den Bergh

Keywords:
Adverse events
Disease progression
Metastatic prostate cancer
Overall survival
Predictive models
Treatment discontinuation
Treatment selection

Abstract

Background and objective: The treatment landscape of metastatic prostate cancer (mPCa) has evolved significantly over the past two decades. Despite this, the optimal therapy for patients with mPCa has not been determined. This systematic review identifies available predictive models that assess mPCa patients' response to treatment.

Methods: We critically reviewed MEDLINE and CENTRAL in December 2022 according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses statement. Only quantitative studies in English were included with no time restrictions. The quality of the included studies was assessed using the PROBAST tool. Data were extracted following the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews criteria.

Key findings and limitations: The search identified 616 citations, of which 15 studies were included in our review. Nine of the included studies were validated internally or externally. Only one study had a low risk of bias and a low risk concerning applicability. Many studies failed to detail model performance adequately, resulting in a



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





Upcoming Workgroup Calls



Date	Time (ET)	Meeting
Tuesday	12 pm	Common Data Model Vocabulary Subgroup
Tuesday	1 pm	Common Data Model
Wednesday	7 am	Medical Imaging
Wednesday	3 pm	Joint Vulcan/OHDS Meeting
Thursday	8 am	OHDSI India Community Call
Thursday	9 am	OMOP CDM Oncology Vocabulary/Development Subgroup
Thursday	9:30 am	Themis
Thursday	12 pm	HADES
Thursday	7 pm	Dentistry
Friday	10 am	GIS-Geographic Information System
Friday	10:30 am	Open-Source Community
Friday	11:30 am	Steering Group
Monday	9 am	Vaccine Vocabulary
Monday	4 pm	Eyecare & Vision Research
Tuesday	9 am	OMOP CDM Oncology Genomic Subgroup



Maternal Health Data Science Fellowship

This program is designed to empower clinical investigators to leverage emerging technologies for improved maternal and neonatal care while reducing morbidity and mortality.

Three main components of this program:

- 1) Career Development** (create evidence, leverage data models, build skills on network studies)
- 2) Practice** (design effective observational research protocols, master tools, write papers/grants)
- 3) Networking** (build relationships with mentors, learners, coordinate with global OHDSI collaborators)

Application deadline: May 15

Want to build your career?

Generate reproducible evidence by leading multi-institutional studies!

Learn more & apply!





Next CBER BEST Seminar: Apr. 17

2021 Titan Award honoree **Yong Chen** will lead the next CBER BEST Seminar on Wednesday, April 17 (11 am-12 pm).

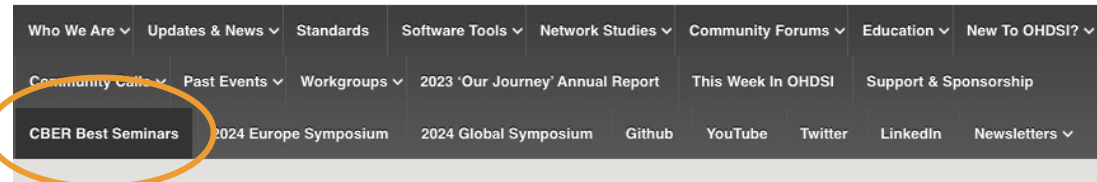
Topic: Real-World Effectiveness of BNT162b2 Against Infection and Severe Diseases in Children and Adolescents: causal inference under misclassification in treatment status.



ohdsi.org/cber-best-seminar-series



Next CBER BEST Seminar: Apr. 17



CBER BEST Seminar Series

The [CBER BEST Initiative](#) Seminar Series is designed to share and discuss recent research of relevance to ongoing and future surveillance activities of CBER regulated products, namely biologics. The series focuses on safety and effectiveness of biologics including vaccines, blood components, blood-derived products, tissues and advanced therapies. The seminars will provide information on characteristics of biologics, required infrastructure, study designs, and analytic methods utilized for pharmacovigilance and pharmacoepidemiologic studies of biologics. They will also cover information regarding potential data sources, informatics challenges and requirements, utilization of real-world data and evidence, and risk-benefit analysis for biologic products. The length of each session may vary, and the presenters will be invited from outside FDA.



Below you will find details of upcoming CBER BEST seminars, including virtual links that will be open to anybody who wishes to attend. Speakers who give their consent to be recorded will also have their presentations included on this page; you can find those sessions below the list of upcoming speakers.

CBER BEST Seminar Series

The [CBER BEST Initiative](#) Seminar Series is designed to share and discuss recent research of relevance to ongoing and future surveillance activities of CBER regulated products, namely biologics. The series focuses on safety and effectiveness of biologics including vaccines, blood components, blood-derived products, tissues and advanced therapies. The seminars will provide information on characteristics of biologics, required infrastructure, study designs, and analytic methods utilized for pharmacovigilance and pharmacoepidemiologic studies of biologics. They will also cover information regarding potential data sources, informatics challenges and requirements, utilization of real-world data and evidence, and risk-benefit analysis for biologic products. The length of each session may vary, and the presenters will be invited from outside FDA.



Below you will find details of upcoming CBER BEST seminars, including virtual links that will be open to anybody who wishes to attend. Speakers who give their consent to be recorded will also have their presentations included on this page; you can find those sessions below the list of upcoming speakers.

Upcoming Seminars

+ April 17 (11 am ET): Yong Chen, University of Pennsylvania

Previous Seminars

- + Jan. 17, 2024 · Anna Ostropelets, Odysseus Data Services
- + Dec. 6, 2023 · Jenny Sun, Pfizer
- + June 14, 2023 · Katsiaryna Bykov, Harvard Medical School
- + May 3, 2023 · Xintong LI and Daniel Prieto-Alhambra, University of Oxford, NDORMS
- + Apr. 12, 2023 · Kaatje Bollaerts, P-95
- + Mar. 22, 2023 · Martijn Schuemie, Janssen R&D
- + Feb. 8, 2023 · Fan Bu, UCLA

ohdsi.org/cber-best-seminar-series



DevCon 2024: April 26, 9 am-3 pm ET

Morning Agenda

9:00 am – Introduction

9:15 am – Developers Panel and Lightning Talks

(Katy Sadowski)

- *OHDSI/OMOP – The hard way is the easy way!* (Prof. Vishnu V Chandrabalan)
- *Moving OMOP to the Cloud With DBT and Snowflake* (Roger Carlson)
- *Use cases for ORMs in OMOP* (Dr. Georgina Kennedy)
- *Carrot: code-free OMOP ETL without full data access* (Dr. Sam Cox)

10:45 am – Darwin EU[®] Developers Update (Adam Black)

12:00 pm – Break

Afternoon Agenda

12:30 pm – OHDSI Ecosystem Updates

- TAB Update (Frank DeFalco)
- Strategus Update (Anthony Sena)
- Broadsea Update (Lee Evans)
- Kheiron Updates (Paul Nagy)

1:15 pm – JACKALOPE PLUS The Power of ML for Healthcare Data Mapping & Management (Denys Kaduk)

2:00 pm - An Introduction to Knowledge Graphs using PheKnowLator and OMOP2OBO with Example Applications in Drug Surveillance and Computational Phenotyping (Tiffany Callahan)



OHDSI Global Symposium

The **2024 OHDSI Global Symposium** will be held Oct. 22-24 at the Hyatt Regency Hotel in New Brunswick, NJ.

Tentative symposium format:

Oct. 22 – tutorials

Oct. 23 – plenaries, collaborator showcase

Oct. 24 – workgroup activities





OHDSI Europe Symposium

Registration is now OPEN for the **2024 OHDSI Europe Symposium**, which will be held June 1-3 in Rotterdam, Netherlands.

June 1 – tutorial/workshop
June 2 – tutorial/workshop
June 3 – main conference



ohdsi-europe.org



#OHDSISocialShowcase This Week

MONDAY

Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and Deidentified Clinical Dataset

(Stephanie Hong, Thomas Richards, Benjamin Amor, Tim Schwab, Philip Sparks, Maya Choudhury, Saad Ljazouli, Peter Leese, Amin Manna, Christophe Roeder, Tanner Zhang, Lisa Eskenazi, Bryan Laraway, James Cavallon, Eric Kim, Shijia Zhang, Emir Amaro Syailendra, Shawn O'Neil, Davera Gabriel, Sigfried Gold, Tricia Francis, Andrew Girvin, Emily Pfaff, Anita Walden, Harold Lehmann, Melissa Haendel, Ken Gersing, Christopher G Chute)

Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and Deidentified Clinical Dataset

PRESENTER: **Stephanie S. Hong**

INTRO:

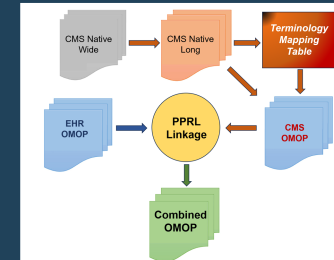
The National COVID Cohort Collaborative (N3C) data Enclave is a platform that provides researchers access to COVID-related patient EHR data in OMOP CDM format. It is the largest centralized repository of COVID-related Patient EMR data in U.S. *It is the largest OMOP instance to our knowledge.* CMS claims data is also transformed into OMOP CDM format using code map terminology translation. N3C COVID patient cohort is now linked to CMS claims data via Privacy Preserving Record Linkage (PPRL). As a result, N3C EHR datasets in OMOP CDM format are enriched with the following additional CMS claims data.

- Inpatient**
- Part D drug prescription**
- Part B**
- Long term care**
- Durable medical equipment**
- Outpatient,**
- Home health**
- Skilled nursing**
- Other services**
- Long Term Care**

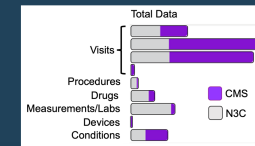
METHODS

1. CMS claim files in wide format are parsed and pivoted into long format. The clinical concept codes are organized into a condensed format per patient per visit for efficient data transformation.
2. The condensed dataset is then used by the Code Map service to generate the clinical concept translation table. The unified version of the OMOP vocabulary tables are used to perform the translation from the source code to OMOP concept IDs
3. The generated code map service table is used as input in the data pipeline to transform the CMS claims datasets into OMOP CDM format.
4. The data pipeline is built to generate CMS dataset in OMOP CDM format with N3C PPRL linkage.
5. N3C data is enriched with CMS data per PPRL-linked N3C patient. In cases where N3C person_id is duplicated, a Global ID is provided for each.

Workflow to generate **Combined OMOP** from EHR & CMS data



How much Medicare data?



How much Medicaid data?



Link to continuously updated view of N3C enriched dataset via PPRL linkage: <https://covid.cd2h.org/dashboard/public-health/pprl/1>

Putting the patient back together again

A timeline with no gaps, either overlapping or contiguous, is used to construct a "Macro Visit" akin to the N3C macro visit approach.



Take a picture to download the brief report

Contact: shong59@jh.edu

RESULT:

- 82 DTAs in N3C | 30 DTAs in N3C CMS PPRL linkage
- N3C total patients : 20,868,921
- N3C PPRL-linked CMS patients : 359,096
- Total rows of data in N3C: 28.3billion
- Total rows of data in CMS: 653,366,927
- N3C dataset enriched by CMS

Among the PPRL-linked patients, in average, additional concepts are available from CMS:

Claims	Domain	PPRL-Linked patient/ domain	Average # of additional concepts added per patient
Medicare	Condition	71%	78
Medicare	Procedure	60%	6.33
Medicare	Drug_exposure	75%	21.83
Medicare	Measurement	60%	16.48
Medicare	Observation	66.9%	8.6
Medicare	Device	47.6%	6.8
Medicaid	Condition	20.8%	33.9
Medicaid	Procedure	20.2%	23
Medicaid	Drug_exposure	21.9%	20.9
Medicaid	Measurement	17.8%	17.44
Medicaid	Observation	18.3%	6.68
Medicaid	Device	13.9%	6.3

Terminology Mapping: Terminology codes appear in multiple columns, i.e.col01 to col45. And some claim source files were over 4000 columns wide. The dataset is pivoted to condense format to generate the clinical concept translation table using OMOP vocabulary tables.



Claims data is reshaped to condensed format.

Source code	Code Column ID	Source code system	Mapped vocabulary id
E11	41	ICD10CM	ICD10CM
09D30Z	45	ICD10PCS	ICD10PCS

Terminology mapping table is used to create OMOP instance of claims data



Stephanie Hong, FAMA¹, Thomas Richards, MS², Benjamin Amor PhD², Tim Schwab³, Philip Sparks⁴, Maya Choudhury⁵, Saad Ljazouli, MS⁶, Peter Leese, MSPH⁷, Amin Manna, MEng⁸, Christophe Roeder⁹, MS¹⁰, Tanner Zhang, MD, MS¹¹, Lisa Eskemati¹², Bryan Laraway, MS¹³, Nirvisha Garara, Rasi Talluri, James Cavallon¹⁴, Eric Kim¹⁵, Shijia Zhang, MS¹⁶, Emir Amaro Syailendra, MD¹⁷, Shawn O'Neil, PhD¹⁸, Davera Gabriel, RN FHL7, FAMA¹⁹, Sigfried Gold, MS²⁰, Tricia Francis, MS²¹, Andrew Girvin, PhD²², Emily Pfaff, PhD, MS²³, Anita Walden, MS²⁴, Harold Lehmann, MD, PhD²⁵, Melissa Haendel, PhD²⁶, Ken Gersing MD²⁷, Christopher G Chute, MD DPH²⁸, on behalf of the N3C Consortium





#OHDSISocialShowcase This Week

TUESDAY

The Feasibility of Clinical Quality Language (CQL) Based Digital Quality Measures (dQMs) Implementation to OMOP CDM

(Emir Amaro Syailendra, Woo Yeon Park, Ben Hamlin, Paul Nagy)

The Feasibility of Clinical Quality Language (CQL) Based Digital Quality Measures (dQMs) Implementation to OMOP CDM

PRESENTER: Emir Syailendra

INTRO:

- Quality measurement is evolving to become more patient-focused in supporting quality improvement.
- Digital Quality Measures (dQMs) are FHIR CQL-based digitized measures created to improve quality assessment at the point of care.
- Given the interoperability of using OMOP CDM to apply an existing phenotype, applying dQMs is one of the possibilities for utilizing the OHDSI vast network of real-world data.
- This study aims to assess the feasibility of dQMs implementation on OMOP data by converting dQMs cohort definitions into Circe-compatible representations.

METHODS

- Three NCQA-HEDIS DQMs were chosen, the ADD-E, PND-E, and SPC. The measures were chosen as they have different elements within them. Concept sets and cohort definitions were created on Atlas. The cohort definitions include initial population, numerator, and denominator derived from the NCQA measures specifications.
- Johns Hopkins Hospital (JHH) OMOP CDM instance was used to apply the translated dQMs.

RESULTS

- Translating the DQMs to OMOP-based cohort definitions took 5 steps, which include creating and obtaining the DQMs, reviewing the DQMs' CQL code and value sets, creating the OMOP concept sets, defining the cohort definitions for the numerator and denominator, and executing also evaluating the cohort definitions to OMOP CDM database.

Translating CQL to OMOP CDM for Digital Quality Measures

CQL to Value Sets

Value Sets to OMOP Concepts Sets

OMOP Cohort Definitions

CQL example: define "Delivery with Gestational Age Assessment or Diagnosis during Measurement Period": (Perinatal."Delivery with Gestational Age Assessment or Diagnosis during Period" ("Measurement Period")) DeliveryInfo where Enrollment."Meets Health Plan Enrollment Criteria" ("Member Coverage", null, Interval[DeliveryInfo.deliveryDate e - 28 days, DeliveryInfo.deliveryDate], 0)

Value sets example:

Obtained Digital Quality Measure (DQM) definitions from HEDIS, a registered trademark of the National Committee for Quality Assurance (NCQA).

In this step, we converted the Value Sets to OMOP Concept Sets using ATLAS.

Each Concept Set represent a single idea, such as deliveries after 37 gestation weeks.

Certain inclusion or exclusion criteria, such as 28 days prior to delivery date, were input directly into the OMOP Cohort Definition in the next step.

The denominator and numerator of a metric is independently defined as a cohort in the ATLAS.

Each Cohort Definition may be composed of multiple concept sets created in the last step.

Compute the Digital Quality Measures on ATLAS based on OMOP Cohort Definitions created in the previous step.

Measures	Summary Definitions	Patient Count (JHH De-Id OMOP database)
Prenatal Depression Screening & Follow-up (PNE-D) Denominator 1	Deliveries during the measurement period where deliveries occurred at more than 37 weeks gestation, and not in hospice.	38,305 patients
Prenatal Depression Screening & Follow-up (PNE-D) Numerator 1	Deliveries in which members had a documented result for depression screening, using an age-appropriate standardized screening instrument, performed during pregnancy.	18,455 patients
Follow-Up Care for Children Prescribed ADHD Medication (ADD-E) Denominator 1	Children 6 years of age as of the start of the intake period to 12 years of age, having dispensed ADHD medication and had a negative medication history.	3593 patients
Follow-Up Care for Children Prescribed ADHD Medication (ADD-E) Numerator 1	An outpatient, intensive outpatient, partial hospitalization, observation, health and behavior assessment or intervention, community mental health center, telehealth or telephone follow-up visit with a practitioner with prescribing authority during the initiation phase.	108 patients

Learning Points and Limitations

- The current workflow started by manually mapping CQL to Value Sets using standard vocabularies, such as SNOMED, ICD10, LOINC codes. Then the Value Sets are converted into OMOP Concept Sets through inputting source codes. The ideal method is to computationally translate CQL to OMOP Concept Sets.
- The PND-E Numerator 2 required the sum of depression survey PHQ-9 to be greater than 10. However, the survey data was formatted in question level, and the total score could not be calculated on Atlas. This limitation prevented to complete this metric and should be addressed in the future study.
- The date range was another limitation as DQM often anchors measurement period end on the December 31st of the measurement year, but the Atlas inclusion criteria uses relative date range. This led the cohort definition to not exactly matching the DQM inclusion criteria.
- Access to the NCQA HEDIS DQM require a paid license.
- Future studies need to be done to compare the metric results from OMOP CDM to the CQL results based on institutional database.
- HEDIS is catered to US based healthcare quality, and applicability of the metrics should be evaluated for non-US implementation use cases.



Emir Amaro Syailendra, Woo Yeon Park, Ben Hamlin, Paul Nagy



contact: msyaille1@jhu.edu





#OHDSISocialShowcase This Week

WEDNESDAY

Using Cohort Diagnostics to Assess the Phenotypic Data Quality in All of Us Research Program

(Lina Sulieman, Karthik Natarajan)

Leveraging OHDSI Cohort diagnostic to assess and externally validate the quality of breast cancer phenotype in the *All of Us* Research Program, **expected trends in breast cancer**, including temporal trends, **were observed. General quality metrics miss metrics essential for phenotypic data credibility**

Using Cohort Diagnostics to Assess the Phenotypic Data Quality in *All of Us* Research Program

INTRODUCTION

- Quality of clinical data for research:
 - Influence the utility of the data
 - Affect reproducibility of clinical research
- Current quality metrics: general, not phenotype-specific
- Objective:** Utilizing OHDSI Cohort Diagnostics to assess the phenotypic data quality in the *All of Us* Research Program, focusing on breast cancer

METHODS

- Applying phenotype library on *All of Us* Research Program, March 2022 release:
 - Identified two breast cancer cohorts: Cohort-1: 4112853001 and Cohort-2: 4112853002
 - Overlaps between cohorts
 - Extracted the incident rates, time distributions, and covariant
 - External validation: Compared temporal trends in breast cancer cohorts in *All of Us* Research Program to multiple datasets in <https://data.ohdsi.org/CohortDiagnosticBreastCancer/>

RESULTS

- All of Us* dataset included 331,382 participants:
 - White: 55.81% participants
 - Female: 60% participants
- OHDSI breast cancer cohorts:
 - Cohort-1: 5905 participants
 - Cohort-2: 7410 participants
 - No participants younger than 30 years old (Fig 1)
 - Median days before/after index diagnosis of breast cancer: *All of Us* higher than other datasets (Fig 2)
 - Temporal trends (Table 1):
 - Height, weight: Increased by 0.15 in "Starts 31 to 365"
 - Hemoglobin, Neutrophils: doubled in "Starts 31 to 365"

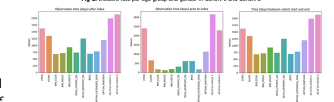
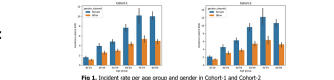


Table 1. Temporal Characteristics of time values for selected covariates breast cancer cohort overlaps.

Primary independent covariate of	Cohort-1	Cohort-2	Cohort-1 & Cohort-2	Cohort-1	Cohort-2	Cohort-1 & Cohort-2
	n	n	n	Q1	Q2	Q3
Height	5905	7410	13315	162.5	167.5	172.5
Weight	5905	7410	13315	65.0	70.0	75.0
Hemoglobin	5905	7410	13315	12.0	13.0	14.0
Neutrophils	5905	7410	13315	0.0	1.0	2.0

DISCUSSION

- Cohort diagnostic: assessing the quality of phenotypic data in the research datasets
- Observed data quality in breast cancer cohorts extracted from the *All of Us* Research Program:
 - Expected trends in age groups
 - Higher-than-expected incident rates in participants identified as Other than female
 - Sex or Gender in *All of Us*: different option, cautious when implementing phenotype library
 - Expected trends in temporal characteristics for breast cancer diagnosis:
 - Height, weight, Hemoglobin, Neutrophils: more labs were ordered after diagnosis. Measurements are taken periodically to assess the patient's prognosis and prepare for chemotherapy
 - Drugs to treat breast cancer usage: much higher than before diagnosis

Acknowledgment

- All of Us* Research Program Participants
- VUMC/Columbia Data and Research Centers

VANDERBILT UNIVERSITY MEDICAL CENTER
Department of Biomedical Informatics

All of Us
DATA & RESEARCH CENTER

Lina Sulieman, PhD¹; Karthik Natarajan, PhD²
¹Vanderbilt University Medical Center
²Columbia University Medical Center



#OHDSISocialShowcase This Week

THURSDAY

Demonstration of the OHDSI phenotype library

(Gowtham Rao)

Demonstration of the OHDSI phenotype library

PRESENTER: **Gowtham A Rao**

Who Cares?

Researchers: No more reinventing the wheel. Use peer-reviewed cohort definitions and get to your results faster.

Healthcare Policy Makers: Need credible data for policy? Access vetted, reliable cohort definitions here.

Data Scientists: Garbage in, garbage out. Use quality definitions to elevate your findings.

Funders: Maximize ROI. Your grants power more with reusable, standardized research assets.

Bottom Line: It's more than a library; it's a healthcare research game-changer. Community-built, peer-reviewed, and FAIR-compliant.

How We Keep It Real

Collected: Sourced from OHDSI forums. Email submissions to Rao. Attested: You say it's good; we take your word (initially).

Tested: Run through PhenotypeLibraryDiagnostics on OHDSI network.

Quality Checks: Automated + human oversight.

Peer Review: Open forum or workgroup discussions. Public scrutiny.

Infrastructure: Hosted on GitHub; R package. API-ready.

Versioning: DOIs for every release. Cite us.

Main Finding 1: FAIR^{ish}-Compliant Cohort Definition Repository

What: The OHDSI Phenotype Library (PL) is an open, version-controlled repository designed to guide real-world evidence towards the FAIR principles—Findability, Accessibility, Reproducibility, and Interoperability.

So What: This compliance ensures that each cohort definition is easily searchable, version-controlled, and standardized for use across different studies.

Now What: Researchers can leverage this library for more efficient and standardized cohort definitions in their observational research.

Main Finding 2: Comprehensive Metadata and Quality Assurance

What: The OHDSI PL collects extensive metadata for each cohort definition, which includes user-submitted, librarian-assigned, and computer-generated metadata.

So What: Such metadata enhances the understandability, searchability, and reliability of the cohort definitions. Peer review processes further assure quality.

Now What: Researchers can have confidence in using these cohort definitions and can easily navigate the library to find definitions that suit their study requirements.

Main Finding 3: Technical Infrastructure and Maintenance

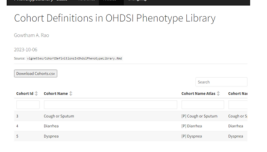
What: The OHDSI PL is hosted on GitHub and encapsulated within an R package, adhering to the OHDSI HADES ecosystem. It also follows a regular release cycle for updates.

So What: This setup ensures that the library is both accessible and maintained, with about 25 releases since the establishment of major version 3 in 2022.

Now What: Researchers can easily integrate the PhenotypeLibrary with other HADES packages and can expect the library to be updated and maintained regularly.



550 Cohorts
21 peer reviewed
18 contributors
Version 3.25.0





#OHDSISocialShowcase This Week

FRIDAY

Large variety Country size RWD data-lake

(**Guy Livne**, Keren Rosenstein, Atif Adam, Milou Brand, Nikolai Grewe, Ludovica Ancora, Nathan Japhet)



Kineret Israel health data-lake
Guy Livne

INTRO

15 years of FULL EHR data from Governmental hospitals, Birth to Death patient record.

METHODS

- Gathering Patient data from general medical center's legacy systems.
- Mapping ALL EHR data to standard OMOP concepts using machine learning pipelines and expert reviews.
- Conforming all to OMOP CDM v5.3 structure.
- Developing ETL processes for quarterly updates.
- Applying data de-identification and privacy rules.
- Enabling OHDSI analytic tools.

RESULTS

A nationwide, 15-year RWD available for collaborative research through Kineret's platform.

Salient features:

- **Longitudinal**, linked data across inpatient, outpatient, ER and specialized care.
- **Diverse population** covering all regions, ethnicities, religions, and socioeconomic strata in Israel.
- **Predefined datasets** for diseases like diabetes, heart failure, infections to accelerate research.
- **Professional team** to initiate studies in less than 3 months.



<https://kineret.health.gov.il/en>

Large VARIETY country size DATA-LAKE



Diverse population covering all regions, ethnicities, religions, and socioeconomic & lifestyle



Israel nation wide EHR data (15 years)

Combine and Encrypt



Combine with NON-OMOP data: DICOM, Sleep apnea, Intracardiac pressure, etc.



Establish OMOP CDM harmonization with OHDSI standard

Pre-define datasets:
diabetes, heart failure, Pregnancy & newborn

Secure cloud environment for Study & collaboration

Contact us:



Fast work process for the benefit of researches



Guy Livne, Dr. Nadav Rappoport, Nir Makover, Hadas Eshel-Geva, Hadar Kapach, Tomer Hadad, Yariv Alon, Naama Perry-Cohen.





Opening: Biomedical Informatics Data Scientist at Stanford



[Who We Are](#) ▾

[What We Offer](#) ▾

[How We Hire](#) ▾

[Career Areas](#) ▾

[Search](#) .

Biomedical Informatics Data Scientist

1.0 FTE • Full time • Day - 08 Hour • R2335119 • Hybrid • 84866 IT RESEARCH • Technology & Digital Solutions • 455 Broadway, REDWOOD CITY, California •

1.0 FTE Full time Day - 08 Hour R2335119 Hybrid 84866 IT RESEARCH Technology & Digital Solutions 455 Broadway, REDWOOD CITY, California

If you're ready to be part of our legacy of hope and innovation, we encourage you to take the first step and explore our current job openings. Your best is waiting to be discovered.

Day - 08 Hour (United States of America)

This is a [Stanford Health Care](#) job.

A Brief Overview

The Biomedical Informatics Data Scientist will partner with researchers and clinicians to enable effective and efficient use of data and resources available via Stanford's research clinical data repository (STARR) including the Electronic Health Records in the OMOP Common Data Model, radiology and cardiology imaging data and associated metadata, and new data types as they get integrated along with their databases and respective cohort query tools and interfaces e.g., OHDSI ATLAS. This individual will enable researchers to maximize their understanding, interpretation and use of these clinical and research tools for more informed and productive research, clinical trials, patient care and quality outcome projects.

Clean, extract, transform and analyze various kinds of clinical data to create analysis-ready datasets that follow the FAIR (Findable, Accessible, Interoperable and Re-usable) principles. Partner with researchers and clinicians to enable effective and efficient use of Stanford Clinical data and resources for the advancement of research and the educational mission.

Postdoc/Senior Data Analyst Opening at WashU

The Zhang Lab at Washington University School of Medicine in St. Louis has **one postdoc/senior data analyst position** to work on **causal machine learning** and **responsible AI** for reliable real-world evidence generation.



PI: Linying Zhang, PhD

- More details at <https://linyingzhang.com>
 - Postdoc:
<https://linyingzhang.com/files/Postdoc.pdf>
 - Data analyst:
<https://linyingzhang.com/files/Analyst.pdf>
- If interested, please send CV and cover letter to linyingz@wustl.edu





Director, RWE at Gilead

About Us



Director, RWE - Data Science - OHDSI

Apply

Gilead Sciences, Inc. is a biopharmaceutical company that has pursued and achieved breakthroughs in medicine for more than three decades, with the goal of creating a healthier world for all people. The company is committed to advancing innovative medicines to prevent and treat life-threatening diseases, including HIV, viral hepatitis and cancer. Gilead operates in more than 35 countries worldwide, with headquarters in Foster City, California.

Responsibilities:

Collaborate with researchers and data scientists to understand project requirements and translate them into OHDSI-compatible solutions. Work with databases, ensuring data integrity and optimization for OHDSI-related queries and analyses. Perform data analyses in OHDSI-related tools like ATLAS. Customize and extend OHDSI tools and applications to meet specific project needs. Collaborate with cross-functional teams to troubleshoot and resolve technical issues related to OHDSI implementations. Stay informed about OHDSI community updates, best practices, and emerging trends in observational health data research. Contribute to the development and documentation of data standards and conventions within the OHDSI community.



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?





April 16: Tools to Evaluate ETL



Frank DeFalco

Senior Director
Observational Health Data Analytics
Janssen Research & Development



Katy Sadowski

Senior Associate Director
Boehringer Ingelheim

April Olympians Week 3 Update



Clair Blacketer

Director
Janssen Research & Development



Melanie Philofsky

Senior Business Analyst and Project
Manager, Odysseus Data Services, Inc.



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