



2024 DevCon Review

OHDSI Community Call
May 7, 2024 • 11 am ET

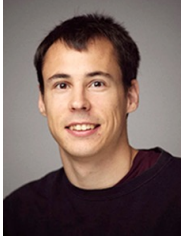


Upcoming Community Calls

| Date | Topic |
|---------|--|
| 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 |



May 14: 10-Minute Tutorials



Martí Català Sabaté

Medical Statistician/Data Scientist
University of Oxford

Drug Utilization



Kim López Güell

Dphil Student
University of Oxford

Cohort Survival



Maarten van Kessel

Software Developer
Erasmus MC

Treatment Patterns



Louisa Smith

Assistant Professor
Northeastern University

All of Us Research



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!



Congratulations to the team of **William G. Adams, Sarah Gasman, Ariel L. Beccia and Liza Fuentes** on the publication of **The Health Equity Explorer: An open-source resource for distributed health equity visualization and research across common data models** in the *Journal of Clinical and Translational Science*.

Journal of Clinical and Translational Science

www.cambridge.org/cts

Research Methods and Technology Research Article


Cite this article: Adams WG, Gasman S, Beccia AL, and Fuentes L. The Health Equity Explorer: An open-source resource for distributed health equity visualization and research across common data models. *Journal of Clinical and Translational Science* 8, e72, 1–9. doi: [10.1017/cts.2024.500](https://doi.org/10.1017/cts.2024.500)

Received: 26 December 2023
Revised: 1 March 2024
Accepted: 11 March 2024

Keywords: Common data models; distributed analytics; health equity; open-source software; R-Shiny; translational informatics

Corresponding author: W. G. Adams, Email: badams@bu.edu

The Health Equity Explorer: An open-source resource for distributed health equity visualization and research across common data models

William G. Adams^{1,2} , Sarah Gasman¹, Ariel L. Beccia³ and Liza Fuentes⁴

¹Department of Pediatrics, Boston Medical Center, Boston, MA, USA; ²Boston University Clinical and Translational Science Institute, Chobanian & Avedisian School of Medicine, Boston, MA, USA; ³Boston Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, MA, USA and ⁴Health Equity Accelerator, Boston Medical Center, Boston, MA, USA

Abstract

Introduction: There is an urgent need to address pervasive inequities in health and healthcare in the USA. Many areas of health inequity are well known, but there remain important unexplored areas, and for many populations in the USA, accessing data to visualize and monitor health equity is difficult. **Methods:** We describe the development and evaluation of an open-source, R-Shiny application, the “Health Equity Explorer (H2E),” designed to enable users to explore health equity data in a way that can be easily shared within and across common data models (CDMs). **Results:** We have developed a novel, scalable informatics tool to explore a wide variety of drivers of health, including patient-reported Social Determinants of Health (SDoH), using data in an OMOP CDM research data repository in a way that can be easily shared. We describe our development process, data schema, potential use cases, and pilot data for 705,686 people who attended our health system at least once since 2016. For this group, 996,382 unique observations for questions related to food and housing security were available for 324,630 patients (at least one answer for all 46% of patients) with 65,152 (20.1% of patients with at least one visit and answer) reporting food or housing insecurity at least once. **Conclusions:** H2E can be used to support dynamic and interactive explorations that include rich social and environmental data. The tool can support multiple CDMs and has the potential to support distributed health equity research and intervention on a national scale.

Introduction



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. 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]. 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 and report results from applying the same analyses to the Norwegian Linked Health Registries at University of Oslo, covering

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), 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. 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.

| | Vaccinated | | Unvaccinated | |
|----------|-------------|----------------|--------------|----------------|
| | Individuals | Long COVID | Individuals | Long COVID |
| Cohort 1 | 397 174 | 782 (0.19%) | 224 223 | 4113 (1.83%) |
| Cohort 2 | 434 723 | 2266 (0.52%) | 321 977 | 7009 (2.18%) |
| Cohort 3 | 263 957 | 2814 (1.07%) | 438 151 | 18 544 (4.23%) |
| Cohort 4 | 1 469 697 | 39 210 (2.67%) | 548 584 | 41 971 (7.64%) |

Data shown are n (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 and contribute to event counts multiple times.

Table: Diagnostic records from the Norwegian Linked Health Registries at University of Oslo for any long COVID symptoms between 90 and 365 days after SARS-CoV-2 infection, stratified by cohort and exposed status.

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 Articles page 225 See Online for appendix For the complete text see [https://www.thelancet.com/journal/2024/04/10/S2213-2600\(24\)00082-1](https://www.thelancet.com/journal/2024/04/10/S2213-2600(24)00082-1)



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





Upcoming Workgroup Calls



| Date | Time (ET) | Meeting |
|-----------|-----------|------------------------------------|
| Wednesday | 9 am | Patient-Level Prediction |
| Wednesday | 12 pm | Health Equity |
| Wednesday | 2 pm | Natural Language Processing |
| Wednesday | 3 pm | Joint Vulcan/OHDSI Meeting |
| Thursday | 9:30 am | Network Data Quality |
| Thursday | 10:30 am | Evidence Network |
| Thursday | 12 pm | Strategus HADES Subgroup |
| Thursday | 7 pm | Dentistry |
| Friday | 9 am | Phenotype Development & Evaluation |
| Friday | 10 am | GIS-Geographic Information System |
| Friday | 10:30 am | Clinical Trials |
| Friday | 11:30 am | Steering Group |
| Friday | 10 pm | China Chapter |
| Monday | 10 am | Africa Chapter |
| Monday | 11 am | Early-Stage Researchers |
| Tuesday | 9 am | OMOP CDM Oncology Genomic Subgroup |



Latest OHDSI Newsletter Is Available



On The Journey (May 2024)

CDM and Themis conventions and open-source software were both community focuses during April. Reflections on the April Olympians activity and DevCon can be found in this newsletter, both of which included numerous tutorials to aid in our global research mission. Registration for the 2024 OHDSI Global Symposium has opened, and details on the Collaborator Showcase, tutorials, workgroup activities and more are now available. [#JoinTheJourney](#)

Videocast: CDM, Open-Source Development



In the latest On The Journey videocast, Patrick Ryan and Craig Sachson reflect on two community activities held last month, April Olympians and DevCon. April Olympians brought the community together to focus on CDM/Themis conventions and documentation, while DevCon highlighted some of the recent breakthroughs from the open-source community. They also discuss the upcoming Europe Symposium, scheduled for June 1-3 in Rotterdam, Neth. *(If this video doesn't appear, please click 'View this email in your browser')*

Community Updates

Where Have We Been?

• The latest edition of the [CBER BEST Seminar Series](#) was held last month. Yong Chen shared a presentation on 'Real-World Effectiveness of BNT162b2 Against Infection and Severe Diseases in Children and Adolescents: causal inference under misclassification in treatment status.' You can [find the recording here](#).

• The April Olympians event brought together community members to identify all currently ratified CDM and THEMIS conventions for every CDM table and field, write clear documentation for each THEMIS convention, establish a repository for THEMIS conventions, update the CDM documentation to link to relevant THEMIS repository entries, and create CDM documentation related to expansion module efforts around the community. Learn more about this effort later in the newsletter.

• [DevCon 2024](#) served as an opportunity to connect our global open-source community and discuss ways we can collaborate and continue enhancing the future of OHDSI open-source software. The full agenda from the event is posted below, and recordings are available on the event homepage.

Where Are We Now?

• [Registration](#) is now open for the [2024 OHDSI Global Symposium](#), which will be held in person October 22-24 at the Hyatt Regency Hotel in New Brunswick, N.J., USA. The event will include a day of tutorials, a day of plenaries and the collaborator showcase, and a day of workgroup activities. Check out [the event homepage](#) for more information.

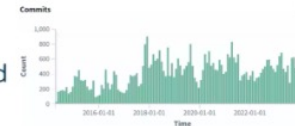
• Applications are now being accepted for the 2024 Maternal Health Data Science Fellowship, which is designed to empower clinical investigators to leverage emerging technologies for improved maternal and neonatal care while reducing morbidity and mortality. The program, which will include the components of career development, practice and networking, will train clinical investigators in observational research methods to enable them to conduct reproducible research and generate real-world evidence. [More information](#), including application details, are now available, [and the deadline to apply](#) is May 15, 2024.

DevCon 2024 Brought Together Open-Source Community, Envisioned Potential & Possibilities of OHDSI Software



OHDSI Open Source Ecosystem

- Software that supports the OMOP CDM
- In the Github OHDSI Organization (Last 10 yrs)
- 284 Repositories
- 611 Developers
- 1,084 Issue Submitters
- 10,812 Issues Submitted
- 50,575 commits
- 618,428 Files
- 80,361,640 lines of code added
- 45,717,357 lines of code refactored



The third annual OHDSI DevCon was held April 26, and it served as an opportunity to connect our global open-source community and discuss ways we can collaborate and continue enhancing the future of OHDSI open-source software.

The first session included a series of lightning talks from a developers' panel, well as a series of development updates from the DARWIN EU® initiative. The second session agenda included a series of updates on the OHDSI open-source ecosystem, as well as presentations around a tool for machine learning healthcare data mapping and management, and knowledge graphs using tools with example applications in drug surveillance and computational genotyping.

April Publications

van Baalen V, Didden EM, Rosenberg D, Bardenheuer K, van Speybroeck M, Brand M. [Increase transparency and reproducibility of real-world evidence in rare diseases through disease-specific Federated Data Networks](#). *Pharmacoepidemiol Drug Saf.* 2024 Apr;33(4):e5778. doi: 10.1002/pds.5778. PMID: 38556812.

Spotnitz M, Ekanayake CD, Ostropelets A, McKhann GM, Choi H, Ottman R, Neugut AI, Hripscak G, Natarajan K, Youngerman BE. [Use of Recommended Neurodiagnostic Evaluation Among Patients With Drug-Resistant Epilepsy](#). *JAMA Neurol.* 2024 Apr 1:e240551. doi: 10.1001/jamaneurol.2024.0551. Epub ahead of print. PMID: 38557864; PMCID: PMC10985618.

Rajwa P, Borkowetz A, Abbott T, Alberti A, Bjartell A, Brash JT, Campi R, Chilelli A, Conover M, Constantinovici N, Davies E, De Meulder B, Eid S, Gacci M, Golozar A, Hafeez H, Haque S, Hijazy A, Hulsen T, Josefsson A, Khalid S, Kolde R, Kotik D, Kurki S, Lambrecht M, Leung CH, Moreno J, Nicoletti R, Nieboer D, Oja M, Palanisamy S, Prinsen P, Reich C, Raffaele Resta G, Ribal MJ, Gómez Rivas J, Smith E, Snijder R, Steinbeisser C, Vandenberghe F, Cornford P, Evans-Axelsson S, N'Dow J, Willemse PM. [Research Protocol for an Observational Health Data Analysis on the Adverse Events of Systemic Treatment in Patients with Metastatic Hormone-sensitive Prostate Cancer: Big Data Analytics Using the PIONEER Platform](#). *Eur Urol Open Sci.* 2024 Mar 25;63:81-88. doi: 10.1016/j.euro.2024.02.019. PMID: 38572301; PMCID: PMC10987796.

Tsafnat G, Dunscombe R, Gabriel D, Grieve G, Reich C. [Converge or Collide? Making Sense of a Plethora of Open Data Standards in Health Care](#). *J Med Internet Res.* 2024 Apr 9;26:e55779. doi: 10.2196/55779. PMID: 38593431; PMCID: PMC11040436.

Lawlor A, Lin C, Gómez Rivas J, Ibáñez L, Abad López P, Willemse PP, Imran Omar M, Remmers S, Cornford P, Rajwa P, Nicoletti R, Gandaglia G, Yuen-Chun Teoh J, Moreno Sierra J, Golozar A, Bjartell A, Evans-Axelsson S, N'Dow J, Zong J, Ribal MJ, Roobol MJ, Van Hemelrijck M, Beyer K; PIONEER Consortium. [Predictive Models for Assessing Patients' Response to Treatment](#)



Latest OHDSI Newsletter Is Available



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

- Who We Are ▾
- Updates & News ▾
- Standards
- Software Tools ▾
- Network Studies ▾
- Community Forums ▾
- Education ▾
- New To OHDSI? ▾
- Community Calls ▾
- Past Events ▾
- Workgroups ▾
- 2023 'Our Journey' Annual Report
- This Week In OHDSI
- Support & Sponsorship
- CBER Best Seminars
- 2024 Europe Symposium
- 2024 Global Symposium
- Github
- YouTube
- Twitter
- LinkedIn
- Newsletters ▾

- Subscribe
- May 2024
- April 2024
- March 2024

Welcome to OHDSI!

Join Us At The 2024 Symposium



Collaborator Spotlight: Monste Camprubi



One of the major [EHDEN] accomplishments over the last year has been all the activities around evidence generation, with several studyathons organised. These have proven to be very informative activities for the participating data partners, many of whom were sharing standardised data for the first time.



ohdsi.org/spotlight-montse-camprubi



Announcing the Maternal Health Data Science Fellowship

Career Development



- Create evidence from real-world data
- Leverage standard data models for reproducible research
- Build skills on effective network studies

Practice



- Design effective observational research protocols
- Master OHDSI tools
- Write papers & grants

Networking



- Build relationships with mentors & fellow learners
- Coordinate with colleagues in the OHDSI data network, spanning 450 sites worldwide & 960 million unique patients

Want to build
your career?

Generate
reproducible
evidence by leading
multi-institutional
studies!

**Application deadline
extended to Wednesday,
May 22, 2024**



**Find out more and apply here
by May 15th, 2024 !**



Top 10 Reasons to Apply for the Maternal Health Data Science Fellowship

1. If you want to make an impact on a major public health issue that's complex to address
2. If you want to catalyze your career in maternal health research
3. If you want to learn how to conduct inter-institutional network studies
4. If you want to lead a publication on evidence generated via an OHDSI Network Study
5. If you want to be part of an active data network to create evidence at scale
6. If you want to become a leader in the OHDSI Community
7. If you want to learn how to do reproducible research
8. If you want personal career mentoring on publishing, writing grants, and research
9. If you want to learn team science
10. If you want to learn how to create validated cohorts



Symposium on Risks and Opportunities of AI in Pharmaceutical Medicine

2024 ANNUAL SYMPOSIUM

Generative AI in Medical Research & Drug Development: Hype or Reality?

June 10-11 | Boston, MA





RWE Workshop at AIME24: Call for Submissions!

Workshop: AI for Reliable and Equitable Real-World Evidence Generation in Medicine

<https://medicine.utah.edu/dbmi/aime/ai-reliable>

Organizing Committee

Linying Zhang
Adam Wilcox
Yves Lussier

Scientific Program Committee

Peter Rijnbeek Mattia Prosperi
Larry Han Xia Ning
Xiaoqian Jiang Yifan Peng

Opening Keynote

George Hripcsak

IMPORTANT DATES

May 31, 2024 | Submission Deadline

June 14, 2024 | Notice of Acceptance

July 12, 2024 | Workshop

AIME 2024
22nd International Conference on Artificial Intelligence in Medicine
Salt Lake City, Utah, USA, July 9-12
Hosted by the University of Utah



OHDSI Europe Symposium

Registration is 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



#OHDSI2024 Registration Is Open!

Registration is now OPEN for the 2024 OHDSI Global Symposium, which will be held Oct. 22-24 at the Hyatt Regency Hotel in New Brunswick, N.J., USA.

Tuesday: Tutorials

Wednesday: Plenary/Showcase

Thursday: Workgroup Activities



ohdsi.org/OHDSI2024



#OHDSISocialShowcase This Week

MONDAY

Jackalope Plus: AI-Enhanced Solution for Mapping Unmappable Concepts

(Denys Kaduk, Marta Vikhrak, Polina Talapova, Eduard Korchmar, Inna Ageeva, Max Ved)

Jackalope Plus:
AI-Enhanced
Solution for Mapping
Unmappable
Concepts

PRESENTERS:

Denys Kaduk

denys.kaduk@sciforce.tech



Polina Talapova

polina.talapova@sciforce.tech



INTRO

- Problem:** When converting to OMOP CDM, you risk losing crucial clinical and observational details, compromising analytics, predictive models and patient care outcomes.
- Old Solution:** 2022's Jackalope captured 'unmappable' data but required manual work.
- New Solution:** Introducing AI-Enhanced Jackalope - automatically capture every nuance, minimal manual input.
- Impact:** Missed details in mapping can compromise data analytics and skew evidence. Our goal is to make your OMOP CDM data rock-solid and reliable!

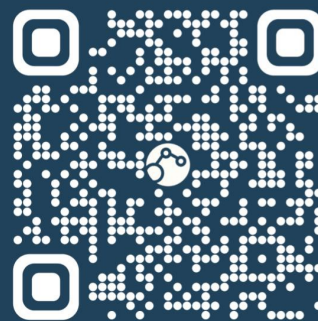
METHODS

- Core Technology:** Used GatorTron, an NLP model trained on diverse medical texts to extract semantic medical details and attributes.
- Integration:** POST requests moved expressions to Jackalope interface.
- Storage:** Inserted new terms and vocabularies into OMOP CDM via SQL or CSV.

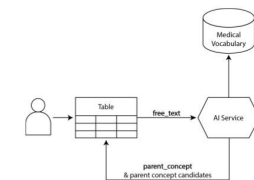
RESULTS

- Preliminary:** Tested 1-to-many ICD10CM codes for manual vs. model comparison.
- GatorTron-medium:** Achieved high semantic similarity scores, suggesting potential for full automation.
- UI Upgrades:** Enhanced Jackalope's interface for easier data processing.

Minimize the need for manual work while accurately capturing all details in the mapping of clinical and observational data.



Take a picture to download the full paper



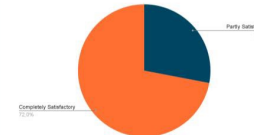
Semantic similarity score generated by model for search

| Name | Score |
|--|-------|
| Laceration of greater saphenous vein at lower leg level, left leg, initial encounter | 0.94 |
| Injury of greater saphenous vein at lower leg level | 0.94 |
| Injury of lesser saphenous vein at lower leg level | 0.94 |
| Laceration of left lower leg | 0.93 |
| Cocaine dependence with intoxication delirium | 0.97 |
| Cocaine intoxication delirium | 0.96 |
| Cocaine delirium | 0.96 |
| Cocaine intoxication | 0.94 |
| Cocaine dependence | 0.93 |

Final parent concept suggestions

| Source Name | Concept Name | Score |
|--|--|-------|
| Aphasia due to and following non-traumatic subarachnoid hemorrhage | Aphasia due to and following non-traumatic subarachnoid hemorrhage | 0.99 |
| Ataxia following nontraumatic subarachnoid hemorrhage | Ataxia due to and following non-traumatic subarachnoid hemorrhage | 0.98 |
| Laceration of greater saphenous vein at lower leg level, left leg, initial encounter | Injury of greater saphenous vein at lower leg level | 0.94 |
| Cocaine dependence with intoxication delirium | Cocaine dependence | 0.93 |
| | Cocaine intoxication delirium | 0.97 |

Results assessment



THE TEAM: Denys Kaduk, Polina Talapova, Maksym Trofymenko, Tetiana Nesmian, Marta Vikhrak, Eduard Korchmar, Lucy Kadets, Yuriy Yakubov, Eugene Kukharchuk, Kirill Barkalov, Inna Ageeva, Max Ved





#OHDSISocialShowcase This Week

TUESDAY

Framework and Implementation of an OMOP-Oriented Clinical Data Warehouse Using Databricks

(Jared Houghtaling, Kyrilo Simonov, Kyle Zollo-Venecek, Elina Hadelia, Manlik Kwong, Polina Talapova, Clark Evans, Robert Miller, Andrew E. Williams)

Tufts Research Data Warehouse (TRDW)

Framework and Implementation of an OMOP-Oriented Data Warehouse Using Databricks

Presenter: Jared Houghtaling

Intro:

- The Clinical and Translational Science Institute (CTSI) at Tufts Medical Center has established a Tufts Research Data Warehouse (TRDW) that incorporates rich observational health data from Tufts Medicine's three hospitals, 40-practice physician network, and home health care organization.

- Once fully implemented, the TRDW will facilitate:

- (1) handling of medical data requests in support of clinical research across Tufts & affiliates
- (2) participation in broad consortia – such as the OHDSI Data Network, Bridge2AI [1], ACT, TriNetX, N3C [2, 3], and CRITICAL [4]
- (3) advancement and utilization of open-source OHDSI tooling and methodologies
- (4) integration of unstructured and semi-structured data (e.g. waveforms, free text, images, flowcharts, genomic tumor profiles) with a structured relational data framework like the OMOP CDM

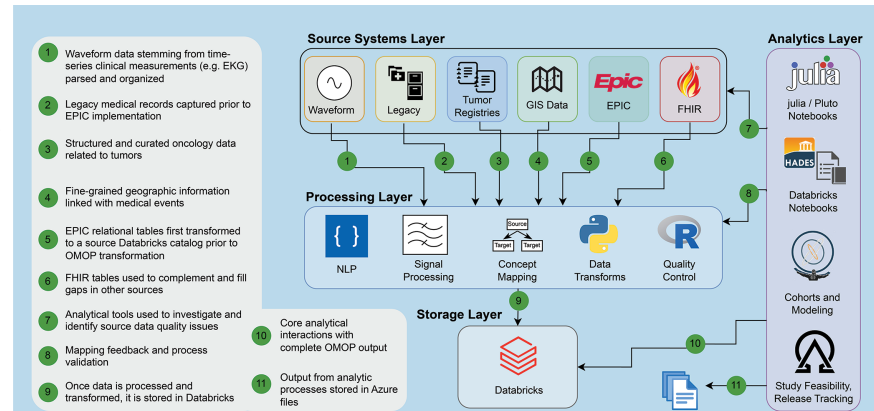
Methods:

- The TRDW is hosted on Azure cloud and incorporates a range of containerized services
- All services are managed via Terraform [6] and ephemeral processes are orchestrated using Prefect [7]

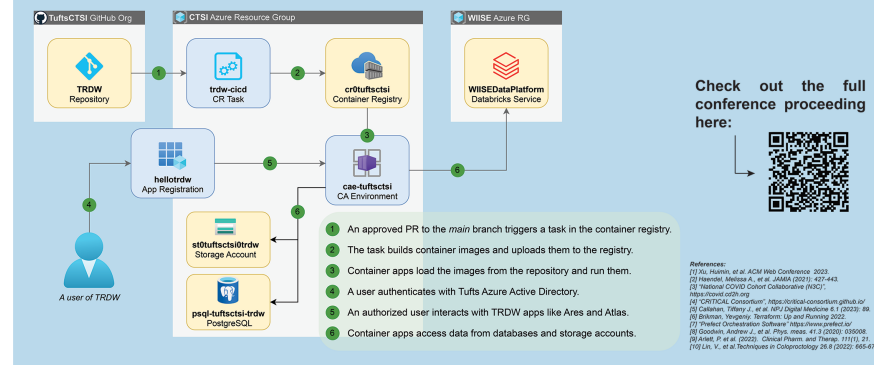
- GitHub is the first step in the continuous integration/continuous delivery (CI/CD) process
- Extract-Transform-Load (ETL) processes using a combination of languages (python, R, Julia) with a fundamental goal of producing a thoroughly documented, highly transparent, and easily maintainable codebase

- Transformations are stratified by OMOP table, and further sub stratified by data source
- Processing pipelines for waveform signals and free-text analysis can require significant computational time; these are currently executed *ad hoc* and independently from the other transformations

- Once transformed, we organize the various sources in an instance architecture, in which independent OMOP instances represent specific sources or combinations thereof



The TRDW enables **broad and deep phenotyping**, linking to knowledge graphs for translational research on **biological causes of disease**, and alignment of health systems' data definitions and data types with **diverse use cases**.



Results & Discussion:

- We are still in the proof-of-concept (POC) stage of defining and implementing a fully operational clinical data warehouse
- For authentication and authorization, we are leveraging newly offered features from Databricks (e.g. Delta Share) that enable highly granular permission structures for data access
- Data governance features, together with a modular notebook-based approach, provide users with configurable tooling that allows them to interact with precisely the data they request and are permitted to view
- Highly scalable framework in terms of data (e.g. quantity, diversity), computational performance, and services provisioning
- We are evaluating potential approaches, like Atrium DB proposed by Goodwin et al. [8], to integrate high-density data forms into our processing pipeline and subsequent informatics work.
- The TRDW currently provides support for a network of more than 50 researchers, and facilitates data deliveries for more than five research consortia.

Conclusions

- The TRDW currently serves as a sandbox for securely and efficiently interacting with multiple rich OMOP datasets as well as with diverse data types. We expect that in the months to come it will enable the construction and implementation of sophisticated statistical models based on multimodal data.

- Much of the work presented here builds on the effort and dedication of so many others in the OHDSI community; we will continue to contribute to - and advocate for - open-source development of these powerful tools, and we plan to continue to share our efforts and experiences along the way.

Authors: Jared Houghtaling^a, Kyrilo Simonov^a, Kyle Zollo-Venecek^a, Elina Hadelia^a, Manlik Kwong^a, Polina Talapova^a, Clark Evans^a, Robert Miller^a, Andrew E. Williams^a

^aTufts Medical Center - Clinical and Translational Sciences Institute (CTSI)



References:
[1] M. Hahn, et al. ACM Web Conference 2023.
[2] Hadelia, Elina A., et al. JAMA 2023; 329:443.
[3] National COVID Cohort Collaborative (N3C). <https://n3c.columbia.edu/>
[4] CRITICAL Consortium. <https://critical-consortium.github.io/>
[5] C. Talapova, et al. NPJ Digital Medicine 5:1 (2022); 50.
[6] Brannen, Virginia. Terraform: Up and Running 2023.
[7] Prefect Conductor Software. <https://www.prefect.io/>
[8] Goodwin, Andrew J., et al. Phys. med. 47:3 (2022); 0300006.
[9] Aker, P. et al. 2022. *Clinical Trials and Therap.* 11(1): 21.
[10] Li, Y., et al. *Techniques in Coloproctology* 26:8 (2022): 868-878.

#OHDSISocialShowcase This Week

WEDNESDAY

Mother-Infant Linked Data: Methodology, Case Studies, and Cohort Development for Investigating Prenatal Exposure and Neonatal Outcomes

(Jill Hardin, Alexis Krumme, David Kern, James Weaver, Clair Blacketer)

Mother-Infant Linked Data: Methodology, Case Studies, and Cohort Development for Investigating Prenatal Exposure and Neonatal Outcomes

BACKGROUND

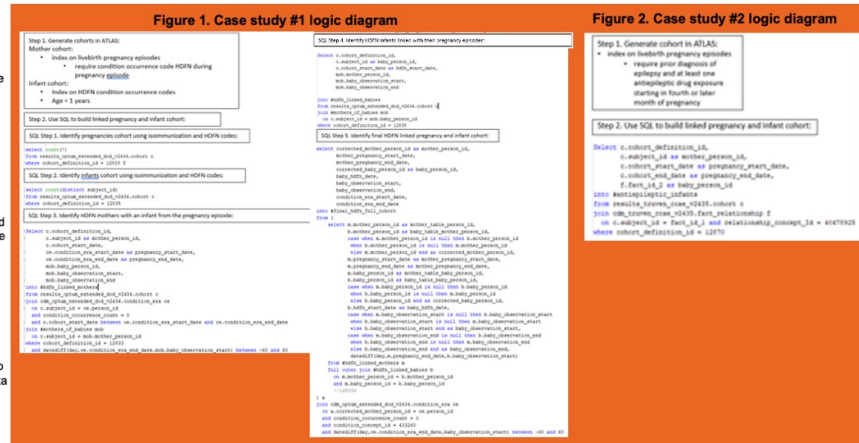
- Linkage of maternal and infant records from routinely collected healthcare data facilitates research on prenatal exposure, prenatal comorbidity, and infant health outcomes[1].
- A recent study developed a mother-infant linkage algorithm using two commercial claims databases[2].
- We present two case studies demonstrating the use of this linkage and provide a step-by-step methodological guide to develop linked mother and infant cohorts.
- For each case study, we provide SQL code to navigate the Common Data Model (CDM) [3] and explain the step-by-step process of extracting the mother-infant linkage data using the fact relationship table in combination with pregnancy cohorts built in ATLAS (accessible at <https://github.com/OHDSI/ATLAS>) [4].
- The objective of this study is to provide researchers reproducible code to access the mother infant linkage algorithm through two use cases which are used to explain the step-by-step method of extracting the mother infant linked data allowing researchers to take advantage of these useful resources.

METHODS

- Two US observational databases that were transformed to the Observational Medical Outcomes Partnership (OMOP) Common Data Model version 5.3.1 were used [3].
- The ATLAS tool [4] was used to develop cohorts by defining pregnancy episodes and identifying relevant exposures or diagnoses.
- Case study #1** identifies a cohort of infants affected by Hemolytic Disease of the Fetus and Newborn (HDFN), a condition where maternal and fetal red blood cell incompatibility can lead to maternal isoimmunization and subsequent anemia in the fetus and neonate. It uses pregnancy episodes with concurrent diagnostic codes for maternal or infant isoimmunization (the precursor to HDFN) or HDFN due to Rhesus alloantibodies (using SNOMED codes). This case study is unique in that the diagnosis can occur either during pregnancy or in the infant.
- Case study #2** focuses on exposure to antiepileptic seizure medications after the first trimester during pregnancy. It uses pregnancy episodes concurrent with antiepileptic seizure medications dispensing (using ATC3 ingredient-level drug concepts)

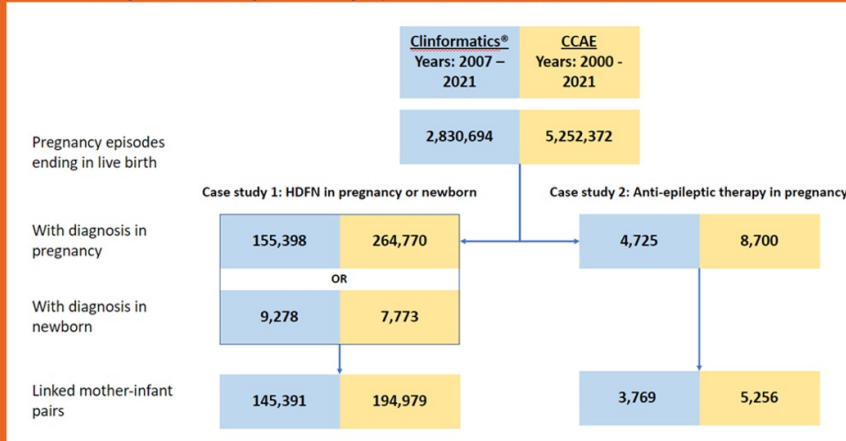
RESULTS

| Database | Years | # of persons (millions) | # of pregnancy episodes ending in a live birth | # of linked infants |
|----------------|-----------|-------------------------|--|---------------------|
| CCAE | 2000-2021 | 157 | 5,252,372 | 4,304,632 |
| Clinformatics® | 2007-2021 | 71 | 2,830,694 | 1,850,278 |



Standardized tools for pregnancy episode identification and mother-child linkage enable network studies on maternal and neonatal outcomes.

Figure 3. Attrition diagram illustrating steps for two case studies in two databases



RESULTS, cont.

- Note that the mother-infant relationship is not a one-to-one relationship as mothers can have multiple linked infants.
- Figure 1** illustrates the logic for building the isoimmunization or HDFN pregnancy and infant cohorts for case study 2, with the first step involving two ATLAS cohorts.
- Figure 3** illustrates that from the ATLAS cohort for isoimmunization or HDFN, there were 155,398 (Clinformatics®) and 264,770 (CCAE) pregnancy episodes ending in a live birth and 9,278 (Clinformatics®) and 7,773 (CCAE) infants. The third step involves executing SQL code using the ATLAS isoimmunization or HDFN pregnancy cohort and identified 141,361 (Clinformatics®) and 192,206 (CCAE) pregnancies linked to infant records. The fourth step links the isoimmunization or HDFN to the pregnancy episode and identifies 6,028 (Clinformatics®) and 4,828 (CCAE) infants. The final SQL step identifies the final isoimmunization or HDFN linked pregnancy and infant cohort and identifies 145,391 (Clinformatics®) and 194,979 (CCAE) episodes and infants.
- Figure 2** illustrates the logic for building the exposed infant cohort for case study 2, with the first step involving an ATLAS cohort. The second step involves executing SQL code using the ATLAS cohort and the fact relationship table to link exposed mothers and infants. After executing this code, the number of pregnancy episodes exposed to antiepileptic medication in the fourth month or later of pregnancy was 4,725 in Clinformatics® and 8,700 in CCAE, linked to 3,769 linked infants in Clinformatics® and 5,256 in CCAE.

CONCLUSIONS

- The study of perinatal exposures, maternal and neonatal outcomes, and subgroups, which are frequently constrained in smaller linked observational and registries, is also possible in large observational databases via mother infant linkage.
- This method requires fewer study resources than primary data collection, and this study demonstrated how to use existing tools in an OMOP CDM setting to increase the utility of this resource.

REFERENCES

- Margolis, A.V. and E.B. Andrews. The Safety of Medications in Pregnant Women: An Opportunity to Use Database Studies. *Pediatrics*, 2017, 140(1).
- Weaver, J., et al. Development and evaluation of an algorithm to link mothers and infants in two US commercial healthcare claims databases for pharmacogenetics research. *medRxiv*, 2022; p. 2022.12.13.22283418.
- Voss EA, MR, Malcho A, Ma G, Kroll C, Schumm M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc* 2015;22(3):553-64.
- Schumm M, and DeFalco F. (2019). OHDSI Analytics Tools. In *Observational Health Data Sciences and Informatics: The Book of OHDSI* (pp. 109-123). Independently published.

Jill Hardin^{1,2}, Alexis Krumme^{1,2}, David Kern^{1,2}, James Weaver^{1,2}, Clair Blacketer^{1,2}
¹Janssen Research and Development, Titusville, NJ, USA; ²Observational Health Data Sciences and Informatics (OHDSI), New York, NY;





#OHDSISocialShowcase This Week

THURSDAY

Integrating ATLAS
Cohorts with DICOM
Images and ECG
Waveforms to Enrich
Real-World Evidence
Research

(**Boudewijn Aasman**, Selvin Soby, Sudhakar Veeraraghavan, Erin M. Henninger, Chandra has Nelapatla, Manuel Wahle, Adil Ahmed, Pavel Goriacko, Parsa Mirhaji)



Integrating ATLAS Cohorts with DICOM Images and ECG
Waveforms to Enrich Real-World Evidence Research

CHDI

#OHDSISocialShowcase This Week

FRIDAY

Real-world Effectiveness of BNT162b2 in Children and Adolescents in Preventing Infection and Severe Diseases with SARS-CoV-2 During the Delta and Omicron Periods

(**Qiong Wu**, Jiayi Tong, Bingyu Zhang, Dazheng Zhang, Jie Xu, Yishan Shen, Lu Li, L. Charles Bailey, Jiang Bian, Dimitri A. Christakis, Megan L. Fitzgerald, Kathryn Hirabayashi, Ravi Jhaveri, Alka Khaitan, Tianchen Lyu, Suchitra Rao, Hanieh Razzaghi, Hayden T. Schwenk, Fei Wang, Margot I. Witvliet, Eric J. Tchetgen, Jeffrey S. Morris, Christopher B. Forrest, and Yong Chen)



Real-world Effectiveness of BNT162b2 Against Infection and Severe Diseases in Children and Adolescents: Target Trial Emulation

Qiong Wu^a, Jiayi Tong^b, Bingyu Zhang^b, Dazheng Zhang^b, Jie Xu^b, Yishan Shen^b, Lu Li^b, L. Charles Bailey^c, Jiang Bian^d, Dimitri A. Christakis^e, Megan L. Fitzgerald^e, Kathryn Hirabayashi^f, Ravi Jhaveri^g, Alka Khaitan^h, Tianchen Lyuⁱ, Suchitra Rao^j, Hanieh Razzaghi^k, Hayden T. Schwenk^l, Fei Wang^m, Margot I. Witvlietⁿ, Eric J. Tchetgen Tchetgen^o, Jeffrey S. Morris^p, Christopher B. Forrest^q, and Yong Chen^r
 a. The Center for Health Analytics and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA
 b. Department of Health Outcomes Biomedical Informatics, University of Florida, Gainesville, FL, USA
 c. Applied Clinical Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
 d. Center for Child Health, Behavior, and Development, Seattle Children's Research Institute, Seattle, WA, USA
 e. Department of Medicine, Grossman School of Medicine, New York University, New York, NY, USA
 f. Division of Pediatric Infectious Diseases, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
 g. Department of Pediatrics, Ryan White Center for Pediatric Infectious Diseases and Global Health, Indiana University School of Medicine, IN, USA
 h. Department of Pediatrics, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO, USA
 i. Department of Pediatrics, Stanford School of Medicine, Stanford, CA, USA
 j. Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA
 k. Department of Sociology, Social Work and Criminal Justice, Lamar University, Beaumont, TX, USA
 l. Department of Statistics and Data Science, The Wharton School, The University of Pennsylvania, PA, USA

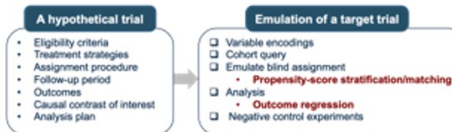


Background

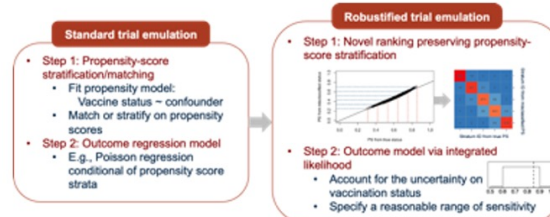
- Clinical Inquiry:**
 - Evaluating the efficacy of BNT162b2 within the U.S. pediatric population
- Knowledge gaps:**
 - Variant evolution:** Existing RCTs were conducted up until the prevalence of the Delta variant
 - Uncertain durability:** Research on the Omicron variant has primarily focused on short-term vaccine effects.
 - Scope limitation:** Few existing studies covered both hospitalized patients and those with mild or asymptomatic infections.
- Methodological Challenges:**
 - Target trial emulation** using electronic health record (EHR) data
 - Incomplete vaccination status** documentation within U.S. health systems

Methods

- Emulate three target trials to investigate the effectiveness of the BNT162b2 vaccine in preventing infection with various strains of the SARS-CoV-2 virus in children and adolescents in U.S.
 - Target trial 1** (Delta study in adolescents): adolescents aged 12-20 years during the period when the Delta variant was prevalent from July 1, 2021, to November 30, 2021.
 - Target trial 2** (Omicron study in children): children aged 5 to 11 years during the period when the Omicron variant was prevalent from January 1, 2022, to November 30, 2022.
 - Target trial 3** (Omicron study in adolescents): adolescents aged 12-20 years during the period when the Omicron variant was prevalent from January 1, 2022, to November 30, 2022.

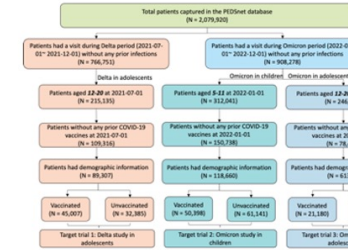


- A robustified trial emulation pipeline to account for incomplete vaccine records

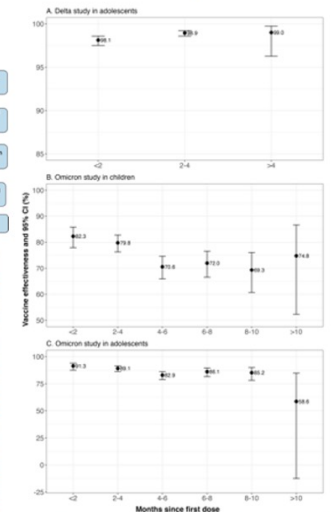


Results

- Data source:** EHR data from PEDSnet
- Eligibility criteria:**
 - In the age group at the study start
 - No previous COVID-19 vaccination
 - No previous SARS-CoV-2 infection
 - User of the healthcare system of PEDSnet (i.e., having a primary care visit in the past 18 months)
- Intervention:** BNT162b2 vaccine vs. no receipt of any type of COVID-19 vaccine
- Outcome:** documented SARS-CoV-2 infection (i.e., virus test/diagnosis), severity of COVID-19 infections, ICU admission with COVID-19
- Confounding variables:** demographic factors, clinical factors, and healthcare utilization factors (including the number of negative COVID-19 tests prior to the cohort entry).



| Vaccine effectiveness (in %) and 95% CI | |
|---|-------------------|
| Delta study in adolescents | |
| Documented infection | 98.4 (98.1, 98.7) |
| Mild COVID-19 | 99.0 (98.5, 99.3) |
| Moderate or severe COVID-19 | 98.7 (97.4, 99.3) |
| ICU admission with COVID-19 | 99.0 (92.5, 99.9) |
| Omicron study in children | |
| Documented infection | 74.3 (72.2, 76.2) |
| Mild COVID-19 | 73.5 (69.2, 77.1) |
| Moderate or severe COVID-19 | 75.5 (69.0, 81.0) |
| ICU admission with COVID-19 | 84.9 (64.8, 93.5) |
| Omicron study in adolescents | |
| Documented infection | 85.5 (83.8, 87.1) |
| Mild COVID-19 | 87.0 (83.5, 89.8) |
| Moderate or severe COVID-19 | 84.8 (77.3, 89.9) |
| ICU admission with COVID-19 | 91.5 (69.5, 97.6) |



Conclusions

- The research suggests a moderate effectiveness of the BNT162b2 vaccine for preventing infection and severe diseases of the SARS-CoV-2 Omicron variant and high effectiveness against the Delta variant based on a national pediatric cohort in the U.S.
- The novel trial emulation pipeline offers a new approach for assessing real-world effectiveness with incomplete vaccine records.



Contact: qiong.wu@penmedicine.upenn.edu and ychen123@penmedicine.upenn.edu



Opening: PhD Student, Erasmus MC

PhD student in Health Data Science

| Published | Deadline | Location |
|-----------|----------|-----------|
| 30 Apr | 13 May | Rotterdam |



JOB DESCRIPTION

The department of Medical Informatics is looking for a PhD student to work on cutting-edge health AI and data science topics. This research will be performed in close collaboration with the Observational Health Data Sciences and Informatics (OHDSI) initiative, which is a global, multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics (www.ohdsi.org) which develop frameworks to generate reliable real-world evidence.

As a PhD student, you will be responsible for the research on using federated data networks to improve best practices around the development and validation of prediction models. You will lead and contribute to projects conducting methodological research within the field of machine learning in healthcare. Research will be performed in clinical settings, feature engineering methods, deep-learning, and other advanced machine-learning methods to support personalized medicine. The research will focus on using large-scale federated data networks made possible by the standardization of health data to the OMOP Common Data Model. Impact assessments of these new approaches on patient care are part of the research agenda.

> Apply now
(7 days remaining)



Specifications

- 📄 PhD; Research, development, innovation
- 🏥 Health
- 🕒 max. 36 hours per week
- € €2789—€3536 per month
- 🎓 University graduate
- # 31.12.23.TG-P1507376-1

Employer



Erasmus MC (University Medical Center Rotterdam)

[Learn more about this employer](#)

Location



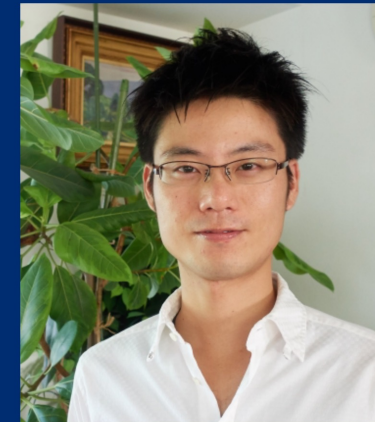
Openings: Postdoctoral Fellow, Johns Hopkins Univ.

PHARMACOEPIDEMIOLOGY POST-DOCTORAL TRAINING PROGRAM

Co-Directors: Caleb Alexander, MD, MS and Jodi Segal, MD, MPH

The **Pharmacoepidemiology Training Program** at the Johns Hopkins Bloomberg School of Public Health (BSPH) is currently **seeking to support postdoctoral fellows**. All supported trainees work with core faculty on existing or newly developed research projects on pharmacoepidemiology, so as to optimize the safe and effective use of medicines to treat heart, lung and blood diseases in the United States. |

Deadline for applications: rolling





Opening: Junior Research Software Engineer, Tufts



INFORMATICS

Research Services

COVID-19 Information and Resources

Data and Safety Monitoring Board (DSMB) Program

Center for Clinical Trials (CCT)

Program Evaluation

Qualitative and Mixed Methods Service

Clinical Trial Design Labs

Dissemination and Implementation (D&I) Core

Science Communications



“Our Informatics team can help you collect and manage research data, develop databases, and identify study participants. We’ll find the best data collection solution for your study. To get started, please submit a request below.”

William Harvey, MD, MSc, FACR
Co-Director, Informatics and Tufts Medical Center CMIO

Overview

We participate in development of a robust institutional informatics infrastructure, enabling research teams to maintain their focus on scientific discovery and analyses rather than on data wrangling. Our infrastructure and support systems are dynamic, to keep pace with the changing and interdependent fields of health informatics, bioinformatics, statistics, and data science; expandable, to accommodate new data types and analytic methods; and scalable, to support efficient and methodologically rigorous multisite/institution research. These defining traits allow us to elucidate novel methods and operational principles, harmonize datasets, and create pipelines for data sharing and analytics.



Opening: Research Assistant, University of Oxford



UK date and time: 23-April-2024 15:25

Applicant Options

- › [New Search](#)
- › [Login](#)
- › [Job Details](#)
- › [Help](#)
- › [Terms of Use & Privacy Policy](#)

Job Details

Research Assistant in Health Data Sciences

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD

We have an exciting opportunity for a Research Assistant in Health Data Sciences to join the Pharmacology and Device epidemiology research group led by Professor Daniel Prieto-Alhambra at the Botnar Research Centre, NDORMS, University of Oxford. The NDORMS Pharmacology and Device epidemiology research group is involved in a number of national and international studies exploring the conditions of use (adherence, compliance, off and on-label use) of a number of licensed drugs, devices, and vaccines for the prevention and treatment of human disease in 'real world' (routine practice) conditions.

As a Research Assistant in Health Data Sciences you will contribute to the programming of analytical pipelines for the analysis of routinely collected data mapped to the OMOP Common Data Model. You will analyse real world data to address regulatory questions related to the prevalence/incidence of disease, use of medicines/vaccines, and the risks or benefits of medicines/vaccines or devices. You will prepare analytical packages to run a number of pre-specified analyses, contribute to wider project planning, including ideas for new research projects and gather, analyse, and present scientific data from a variety of sources.

You will hold a relevant BA or MSc degree in Mathematics, Engineering, or a related field. Knowledge of medical statistics and experience analysing large datasets, experience in biostatistics and/or health data sciences and experience in the programming of R packages are essential. Experience in propensity scores, overlap weighting, inverse probability weighting and/or similar methods, expertise in pharmacology or vaccine epidemiology and experience of working with electronic medical records/routinely collected data are desirable.

This is a full-time fixed-term appointment for 2 years.

The closing date for this position is 12 noon on 10 May 2024. You will be required to upload a CV and supporting statement as part of your online application.

Contact Person : HR Team, NDORMS
Contact Phone :
Pay Scale : STANDARD GRADE 6
Salary (£) : £32,332 - £38,205 p.a

Vacancy ID : 172348
Closing Date & Time : 10-May-2024 12:00
Contact Email : hr@ndorms.ox.ac.uk



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?





May 7: DevCon Review



Paul Nagy

Johns Hopkins University

Topic: Open-Source Overview



Lee Evans

LTS Computing LLC

Topic: Broadsea Update



Vishnu Chandrabalan

Lancaster University

Topic: OHDSI/OMOP – The hard way is the easy way



Frank DeFalco

Janssen Research & Development

Topic: Technical Advisory Board (TAB) Update



Roger Carlson

Spectrum Health

Topic: Moving OMOP to the Cloud w/ DBT & Snowflake



Katy Sadowski

Boehringer Ingelheim

Topic: Kheiron Cohort Update



Adam Black

Erasmus MC

Topic: DARWIN EU® Developers Updates



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