Building The OHDSI Evidence Network Sprint, Final Session

OHDSI Community Call
Aug. 20, 2024 • 11 am ET

Upcoming Community Calls

Date	Topic
Aug. 20	Building The OHDSI Evidence Network Sprint – Open Discussion
Aug. 27	canceled due to ISPE 2024
Sept. 3	New Standardized Vocabularies Release
Sept. 10	Asia-Pacific Regional Updates
Sept. 17	TBA
Sept. 24	Recent OHDSI Publications

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Najia Ahmadi, Michele Zoch, Oya Guengoeze, Carlo Facchinello, Antonia Mondorf, Katharina Stratmann, Khader Musleh, Hans-Peter Erasmus, Jana Tchertov, Richard Gebler, Jannik Schaaf, Lena S. Frischen, Azadeh Nasirian, Jiabin Dai, Elisa Henke, Douglas Tremblay, Andrew Srisuwananukorn, Martin Bornhäuser, Christoph Röllig, Jan-Niklas Eckardt, Jan Moritz Middeke, Markus Wolfien and Martin Sedlmayr** on the publication of **How to customize common data models for rare diseases: an OMOP-based implementation and lessons learned** in the *Orphanet Journal of Rare Diseases*.

Ahmadi *et al.*
Orphanet Journal of Rare Diseases (2024) 19:298
https://doi.org/10.1186/s13023-024-03312-9

Orphanet Journal of Rare Diseases

RESEARCH Open Access

How to customize common data models for rare diseases: an OMOP-based implementation and lessons learned

Najia Ahmadi[1*], Michele Zoch[1], Oya Guengoeze[2], Carlo Facchinello[2], Antonia Mondorf[2], Katharina Stratmann[2], Khader Musleh[2], Hans-Peter Erasmus[2], Jana Tchertov[1], Richard Gebler[1], Jannik Schaaf[3], Lena S. Frischen[4], Azadeh Nasirian[7], Jiabin Dai[1], Elisa Henke[1], Douglas Tremblay[5], Andrew Srisuwananukorn[6], Martin Bornhäuser[8], Christoph Röllig[8], Jan-Niklas Eckardt[8,9], Jan Moritz Middeke[8,9], Markus Wolfien[1,10] and Martin Sedlmayr[1]

Abstract

Background Given the geographical sparsity of Rare Diseases (RDs), assembling a cohort is often a challenging task. Common data models (CDM) can harmonize disparate sources of data that can be the basis of decision support systems and artificial intelligence-based studies, leading to new insights in the field. This work is sought to support the design of large-scale multi-center studies for rare diseases.

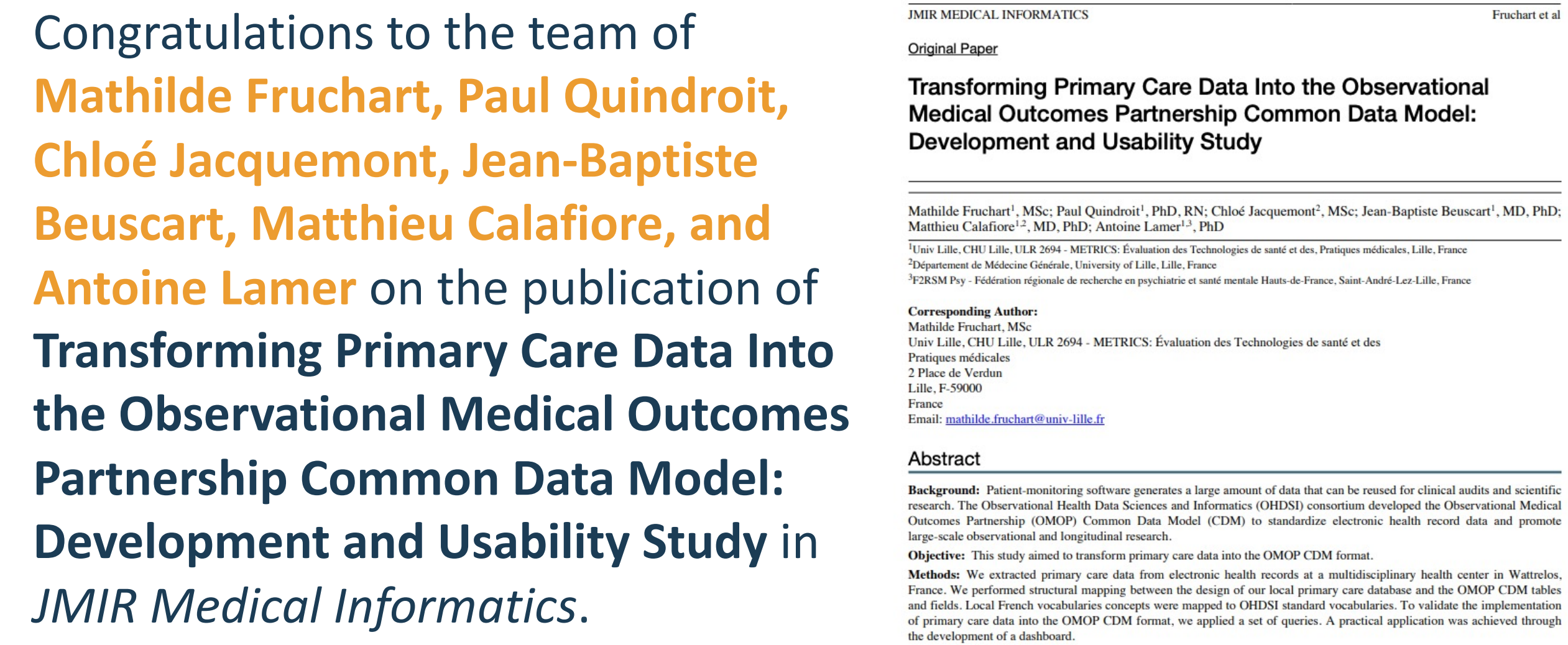
Methods In an interdisciplinary group, we derived a list of elements of RDs in three medical domains (endocrinology, gastroenterology, and pneumonology) according to specialist knowledge and clinical guidelines in an iterative process. We then defined a RDs data structure that matched all our data elements and built Extract, Transform, Load (ETL) processes to transfer the structure to a joint CDM. To ensure interoperability of our developed CDM and its subsequent usage for further RDs domains, we ultimately mapped it to Observational Medical Outcomes Partnership (OMOP) CDM. We then included a fourth domain, hematology, as a proof-of-concept and mapped an acute myeloid leukemia (AML) dataset to the developed CDM.

Results We have developed an OMOP-based rare diseases common data model (RD-CDM) using data elements from the three domains (endocrinology, gastroenterology, and pneumonology) and tested the CDM using data from the hematology domain. The total study cohort included 61,697 patients. After aligning our modules with those of Medical Informatics Initiative (MII) Core Dataset (CDS) modules, we leveraged its ETL process. This facilitated the seamless transfer of demographic information, diagnoses, procedures, laboratory results, and medication modules from our RD-CDM to the OMOP. For the phenotypes and genotypes, we developed a second ETL process. We finally derived lessons learned for customizing our RD-CDM for different RDs.

Discussion This work can serve as a blueprint for other domains as its modularized structure could be extended towards novel data types. An interdisciplinary group of stakeholders that are actively supporting the project's progress is necessary to reach a comprehensive CDM.

OHDSI Shoutouts!

Congratulations to the team of **Mathilde Fruchart, Paul Quindroit, Chloé Jacquemont, Jean-Baptiste Beuscart, Matthieu Calafiore, and Antoine Lamer** on the publication of **Transforming Primary Care Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study** in *JMIR Medical Informatics*.



JMIR MEDICAL INFORMATICS Fruchart et al

Original Paper

Transforming Primary Care Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study

Mathilde Fruchart[1], MSc; Paul Quindroit[1], PhD, RN; Chloé Jacquemont[2], MSc; Jean-Baptiste Beuscart[1], MD, PhD; Matthieu Calafiore[1,2], MD, PhD; Antoine Lamer[1,3], PhD

[1]Univ Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des Technologies de santé et des, Pratiques médicales, Lille, France
[2]Département de Médecine Générale, University of Lille, Lille, France
[3]F2RSM Psy - Fédération régionale de recherche en psychiatrie et santé mentale Hauts-de-France, Saint-André-Lez-Lille, France

Corresponding Author:
Mathilde Fruchart, MSc
Univ Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des Technologies de santé et des Pratiques médicales
2 Place de Verdun
Lille, F-59000
France
Email: mathilde.fruchart@univ-lille.fr

Abstract

Background: Patient-monitoring software generates a large amount of data that can be reused for clinical audits and scientific research. The Observational Health Data Sciences and Informatics (OHDSI) consortium developed the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to standardize electronic health record data and promote large-scale observational and longitudinal research.

Objective: This study aimed to transform primary care data into the OMOP CDM format.

Methods: We extracted primary care data from electronic health records at a multidisciplinary health center in Wattrelos, France. We performed structural mapping between the design of our local primary care database and the OMOP CDM tables and fields. Local French vocabularies concepts were mapped to OHDSI standard vocabularies. To validate the implementation of primary care data into the OMOP CDM format, we applied a set of queries. A practical application was achieved through the development of a dashboard.

OHDSI Shoutouts!

Congratulations to the team of **Nicola Barclay, Edward Burn, Antonella Delmestri, Talita Duarte-Salles, Asieh Golozar, Wai Yi Man, Eng Hooi Tan, Ilona Tietzova, OPTIMA Consortium, Daniel Prieto-Alhambra and Danielle Newby** on the publication of **Trends in incidence, prevalence, and survival of breast cancer in the United Kingdom from 2000 to 2021** in *Scientific Reports*.

scientific reports

Check for updates

OPEN

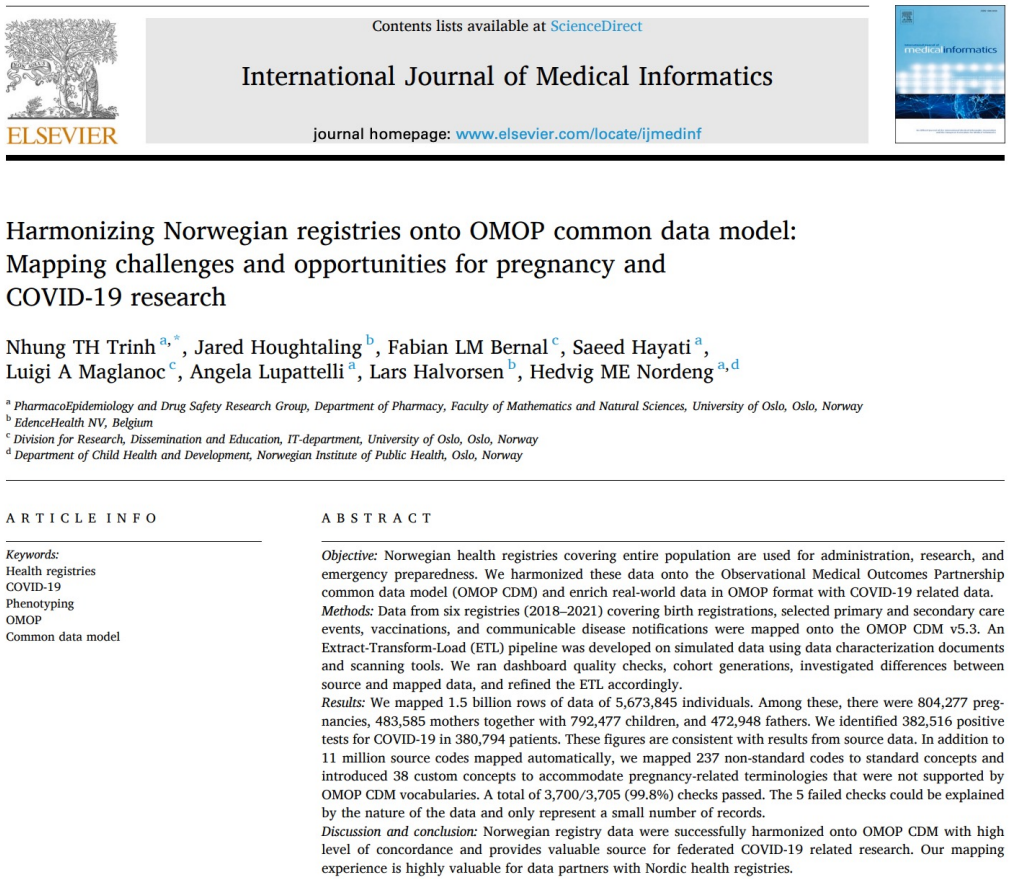
Trends in incidence, prevalence, and survival of breast cancer in the United Kingdom from 2000 to 2021

Nicola L. Barclay[1], Edward Burn[1], Antonella Delmestri[1], Talita Duarte-Salles[2,3], Asieh Golozar[4,5], Wai Yi Man[1], Eng Hooi Tan[1], Ilona Tietzova[6], OPTIMA Consortium*, Daniel Prieto-Alhambra[1,3]✉ & Danielle Newby[1]

Breast cancer is the most frequently diagnosed cancer in females globally. However, we know relatively little about trends in males. This study describes United Kingdom (UK) secular trends in breast cancer from 2000 to 2021 for both sexes. We describe a population-based cohort study using UK primary care Clinical Practice Research Datalink (CPRD) GOLD and Aurum databases. There were 5,848,436 eligible females and 5,539,681 males aged 18+ years, with ≥ one year of prior data availability in the study period. We estimated crude breast cancer incidence rates (IR), prevalence and survival probability at one-, five- and 10-years after diagnosis using the Kaplan–Meier method. Analyses were further stratified by age. Crude IR of breast cancer from 2000 to 2021 was 194.4 per 100,000 person-years for females and 1.16 for males. Crude prevalence in 2021 was 2.1% for females and 0.009% for males. Both sexes have seen around a 2.5-fold increase in prevalence across time. Incidence increased with age for both sexes, peaking in females aged 60–69 years and males 90+. There was a drop in incidence for females aged 70–79 years. From 2003–2019, incidence increased > twofold in younger females (aged 18–29: IR 2.12 in 2003 vs. 4.58 in 2018); decreased in females aged 50–69 years; and further declined from 2015 onwards in females aged 70–89 years. Survival probability for females after one-, five-, and ten-years after diagnosis was 95.1%, 80.2%, and 68.4%, and for males 92.9%, 69.0%, and 51.3%. Survival probability at one-year increased by 2.08% points, and survival at five years increased by 5.39% from 2000–2004 to 2015–2019 for females, particularly those aged 50–70 years. For males, there were no clear time-trends for short-term and long-term survival probability. Changes in incidence of breast cancer in females largely reflect the success of screening programmes, as rates rise and fall in synchronicity with ages of eligibility for such programmes. Overall survival from breast cancer for females has improved from 2000 to 2021, again reflecting the success of screening programmes, early diagnosis, and improvements in treatments. Male breast cancer patients have worse survival outcomes compared to females, highlighting the need to develop male-specific diagnosis and treatment strategies to improve long-term survival in line with females.

OHDSI Shoutouts!

Congratulations to the team of **Nhung TH Trinh, Jared Houghtaling, Fabian Bernal, Saeed Hayati, Luigi Maglanoc, Angela Lupattelli, Lars Halvorsen, and Hedvig Nordeng** on the publication of **Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research** in the *International Journal of Medical Informatics*.



International Journal of Medical Informatics 191 (2024) 105602

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research

Nhung TH Trinh [a,*], Jared Houghtaling [b], Fabian LM Bernal [c], Saeed Hayati [a], Luigi A Maglanoc [c], Angela Lupattelli [a], Lars Halvorsen [b], Hedvig ME Nordeng [a,d]

[a] PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway
[b] EdenceHealth NV, Belgium
[c] Division for Research, Dissemination and Education, IT-department, University of Oslo, Oslo, Norway
[d] Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway

ARTICLE INFO

Keywords:
Health registries
COVID-19
Phenotyping
OMOP
Common data model

ABSTRACT

Objective: Norwegian health registries covering entire population are used for administration, research, and emergency preparedness. We harmonized these data onto the Observational Medical Outcomes Partnership common data model (OMOP CDM) and enrich real-world data in OMOP format with COVID-19 related data.
Methods: Data from six registries (2018–2021) covering birth registrations, selected primary and secondary care events, vaccinations, and communicable disease notifications were mapped onto the OMOP CDM v5.3. An Extract-Transform-Load (ETL) pipeline was developed on simulated data using data characterization documents and scanning tools. We ran dashboard quality checks, cohort generations, investigated differences between source and mapped data, and refined the ETL accordingly.
Results: We mapped 1.5 billion rows of data of 5,673,845 individuals. Among these, there were 804,277 pregnancies, 483,585 mothers together with 792,477 children, and 472,948 fathers. We identified 382,516 positive tests for COVID-19 in 380,794 patients. These figures are consistent with results from source data. In addition to 11 million source codes mapped automatically, we mapped 237 non-standard codes to standard concepts and introduced 38 custom concepts to accommodate pregnancy-related terminologies that were not supported by OMOP CDM vocabularies. A total of 3,700/3,705 (99.8%) checks passed. The 5 failed checks could be explained by the nature of the data and only represent a small number of records.
Discussion and conclusion: Norwegian registry data were successfully harmonized onto OMOP CDM with high level of concordance and provides valuable source for federated COVID-19 related research. Our mapping experience is highly valuable for data partners with Nordic health registries.

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	1 pm	Common Data Model
Wednesday	4 pm	Joint Vulcan/OHDSI Meeting
Wednesday	7 pm	Medical Imaging
Thursday	9:30 am	Network Data Quality
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Steering Group
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Healthcare Systems Interest Group
Monday	10 am	CDM Survey Subgroup

2024 Global Symposium

Oct. 22-24 • Hyatt Regency Hotel, New Brunswick, N.J.

Registration is OPEN for the 2024 OHDSI Global Symposium. Collaborator Showcase notifications are taking place this week. Agendas and tutorial/workgroup schedules are posted.

Tuesday: Tutorials
Wednesday: Plenary/Showcase
Thursday: Workgroup Activities

ohdsi.org/OHDSI2024

2024 APAC Symposium

Dec. 4-8 • Marina Bay Sands & National University of Singapore (NUS)

Preliminary Dates To Know

Sept. 15: Collaborator Showcase Submission Deadline
Sept. 16-20: Collaborator Showcase Submission Review
Oct. 31: Notification of Acceptance

Symposium Agenda

Dec. 4: Tutorial at NUS
Dec. 5-6: Main Conference at Marina Bay Sands
Dec. 7-8: Datathon at NUS

Registration Information is coming soon!

ohdsi.org/APAC2024

The Center for Advanced Healthcare Research Informatics (CAHRI) at Tufts Medicine welcomes:

Ismail Gögenur, MD, DMSc

Dr. Gögenur is Chair of Surgery at the University of Copenhagen and Director of the Center for Surgical Science at Zealand University Hospital.

Andreas Weinberger Rosen, MD, MPM

Dr. Weinberger Rosen is a Researcher at the Center for Surgical Science at Zealand University Hospital.

'AI-augmented decision support for surgical oncology – The Danish experience'

August 29, 2024, 11am-12pm EST Virtually via Zoom

Please contact Marty Alvarez at malvarez2@tuftsmedicalcenter.org for calendar invite or questions.

TuftsMedicine
Tufts Medical Center

#OHDSISocialShowcase

MONDAY

Transforming Clinical Trial Data to the OMOP CDM

(Cynthia Sung, Mike Hamidi, Zhen Lin, Tom Walpole, Rebecca Baker, Melissa Cook, Shital Desai, Priya Gopal, Dan Hartley, Vojtech Huser, Priya Meghrajani, Tra Nguyen, Paul Orona, Katy Sadowski, Sebastiaan van Sandijk, Philip Solovyev, Ramona Walls, Kenneth J. Wilkins, Qi Yang, and the Clinical Trial Working Group)

Transformation of **clinical trial data** to the **OMOP CDM** will open new opportunities to **compare** observational real world data (**RWD**) and data from randomized controlled trials (**RCTs**)

Transforming Clinical Trial Data to the OMOP CDM

Cynthia Sung[1], Mike Hamidi[2]*, Zhen Lin[3]*, Tom Walpole[4]*, Rebecca Baker[5], Melissa Cook[6], Shital Desai[2], Priya Gopal[7], Dan Hartley[8], Vojtech Huser[9], Priya Meghrajani[10], Tra Nguyen[11], Paul Orona[12], Katy Sadowski[13], Sebastiaan van Sandijk[9], Philip Solovyev[9], Ramona Walls[8], Kenneth J. Wilkins[14], Qi Yang[15], and the Clinical Trial Working Group; *CTWG Co-leads

[1]Duke-NUS Medical School, [2]Independent Consultant, [3]Robot Bacon Corp, [4]ZS, [5]CDISC, [6]Essex Management, [7]Flatiron Health, [8]C-PATH, [9]Odysseus [10]Talosix, [11]Boston Medical Center, [12]Kaiser-Permanente, [13]TrialSpark, [14]National Institute of Diabetes, Digestive and Kidney Diseases, [15]IQVIA

Background

The OMOP Common Data Model (CDM) was originally designed to conduct analyses on observational data, rather than clinical trial data. Yet, a large volume of patient-level data exists from clinical trials that would be informative to compare to observational data. The Clinical Trial Working Group (CTWG) is developing a general guideline to facilitate Extract-Transform-Load (ETL) of CDISC SDTM data to the OMOP CDM and proposing conventions for handling unique aspects of clinical trial data.

Methods

The Critical Path Institute (C-PATH) provided deidentified data for TB-1015 (NCT02193776) *"A Phase 2 Open-Label, Partially Randomized Trial to Evaluate the Efficacy, Safety and Tolerability of Combinations of Bedaquiline, Moxifloxacin, PA-824 and Pyrazinamide in Adult Subjects with Newly Diagnosed, Drug-Sensitive or Multi Drug-Resistant Pulmonary Tuberculosis"*. The team benefitted from the diverse healthcare and informatics backgrounds of its members to understand the clinical context, CDISC standards and OMOP CDM conventions to guide the ETL process. Team members worked asynchronously to propose strategies for selecting the best standard concept. These proposals were deliberated upon during the biweekly meetings to reach consensus.

Results

Source SDTM Files to Target OMOP CDM Tables

SDTM file & description		OMOP CDM Table
dm	demographics	PERSON
sv	subject visits	VISIT_OCCURRENCE
ae	adverse events	CONDITION_OCCURENCE
ce	clinical events	
mh	medical history	
ex	drug exposure to experimental drugs	DRUG_EXPOSURE
cm	concomitant meds	
lb	laboratory test results	MEASUREMENT
ms	microbiology susceptibility findings	
mb	microbiology specimen	
vs	vital signs	
cm	concomitant meds	PROCEDURE_OCCURRENCE
pr	procedures	
ae	adverse events	OBSERVATION
dd	death details	

Selected Issues and Proposed Solutions

	Issue	Proposed Solution
CR.001	**USUBJID** is the identifier linking the same patient across all **SDTM** tables. Values often contain **text** characters.	Reference **USUBJID** to autogenerate an **integer** value for **PERSON_ID**. Record **USUBJID** in **PERSON_SOURCE_VALUE**
Handling Relative Dates		
CR.002	SDTM start of a trial **STDY = 1**; OMOP CDM requires calendar dates	Use the **study start date** as the reference start date (**RFSTDTC**) for all patients . This information can be found in the trial publication or the trial registry such as ClinicalTrials.gov
	STDY is a positive number	Add (**STDY**-1) to **RFSTDTC**
	STDY is a negative number (screening visit)	Add (**STDY**) to **RFSTDTC**
CR.004	Construct a unique number for cdm.**<table name>_OCCURRENCE_ID**	Concatenate **USUBJID** and **SEQ** to autogenerate an integer value for **<table name>_OCCURRENCE_ID**
CR.008	Construct a unique number for cdm.**VISIT_OCCURRENCE_ID**	Concatenate **USUBJID** and **SVSTDY** to autogenerate integer value for **VISIT_OCCURRENCE_ID**
Find closest matching Standard Concept id		Concatenate entries in these columns using **paste(. . .)**; can use Usagi to propose best matches to standard concept_ids
	lb.csv	'LBTEST','LBCAT','LBSPEC','LBSTRESU'
	cm.csv	'CMDECOD', 'CMDOSE', 'CMDOSU', 'CMROUTE'

Conclusion

We have reached consensus on several conventions and developed mapping guidelines for doing an ETL of CDISC SDTM data to the OMOP CDM. We welcome new members as we undertake the next stage of work to translate the rules and guidelines into SQL code, perform quality checks, and test the robustness of the methodology when applied to other clinical trials.

OHDSI

Join the Clinical Trial Working Group, which meets biweekly Friday at 11:30 ET.

#OHDSISocialShowcase

TUESDAY

OMOP Harmonization and Integration of Surgical Procedure Database into EHDEN with BC Platforms' Solutions

(Mai Nguyen, Stefano Gamage, Serena Ciaburri, Kalle Parn, Viktoria Sassi-Prantner, Hang T.T. Phan)

ENHANCING HEALTHCARE INSIGHTS

OMOP Harmonization and Integration of Surgical Procedure Database into EHDEN with BC Platforms' Solutions

BACKGROUND: Real-world data (RWD) offers valuable insights into healthcare patterns and outcomes, serving as a cornerstone for evidence-based decision-making in healthcare [1]. BC Platforms, an EHDEN certified SME and a global leader in building data networks for the life sciences industry, offers OMOP harmonization services. BC Platforms provided the OMOP solutions for the Data partner, who wanted to harmonize the clinical data of +750,000 patients collected between 2014 and 2022.

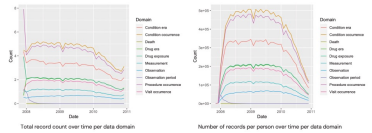
BC Platforms provides a comprehensive solution for OMOP harmonization:

1. Ensuring the capture of the information to the greatest extent possible from the data partner's database
2. Facilitating the seamless integration of healthcare records into global healthcare research initiatives

DESCRIPTIVE STATISTICS REPORT ON OMOP

CDM Inspection report for the Medicare Claims Synthetic Public Use Files database

EHDEN EUROPEAN HEALTH DATA & EVIDENCE NETWORK



Total record count over time per data domain

Number of records per person over time per data domain

METHODS

Figure 1: Schema of ETL flow for OMOP harmonization solution

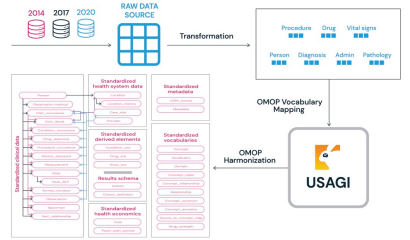
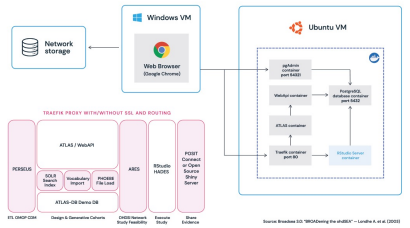


Figure 2: Portable docker solution for OMOP database and EHDEN integration in multi-layer virtual machine (VM) infrastructure



DISCUSSION:

- Understanding the raw data are required careful before implementing the harmonization solution
- Manual curation and translation to English were needed for free-text mapping to OMOP vocabularies
- Relationships between OMOP tables required manual adjustments to address issues related to the raw data

RESULTS

1. Encompassing the raw data, OMOP solutions with tables across multiple domains, and the development of ETL documents together with source code to facilitate the ETL of data into the final OMOP database
2. Facilitating seamless synchronization with OMOP data, empowering DPs to make informed decisions while ensuring compliance with privacy regulations
3. Enabling the seamless integration of surgical databases into healthcare research initiatives across Europe, fostering collaboration and facilitating evidence-based decision-making in the healthcare domain

References:
1. Dang A. Real-World Evidence: A Primer. Pharmaceut Med. 2023 Jan;37(1):25-36. doi: 10.1007/s40290-022-00456-6. Epub 2023 Jan 5. PMID: 36604368; PMCID: PMC9815890.
2. Londhe A. et al., Broadsea 3.0: "BROADening the ohdSEA", 2023, OHDSI work package

OHDSI

MAI NGUYEN, STEFANO GAMAGE, SERENA CIABURRI, KALLE PÄRN, VIKTORIA SASSI-PRANTNER, HANG T.T. PHAN

BC PLATFORMS AG, BLEICHERWEG 10,8002 ZÜRICH, SWITZERLAND

#OHDSISocialShowcase

WEDNESDAY

Creating clinically meaningful cancer groups from SNOMED for care systems and care quality research: a head and neck case study

(Hayley Fenton, Elisabeth Ross, Elin Hallan Naderi, Anne-Lore Bynens)

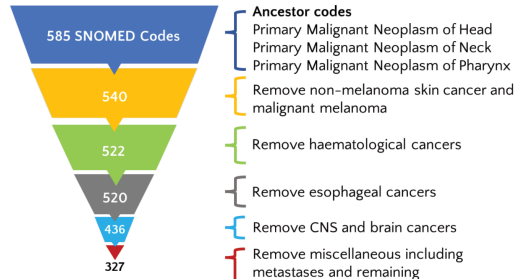
Identification and mapping of SNOMED concepts for head and neck cancer

A DigiONE initiative of DIGICORE
(The Digital Institute for Cancer Outcomes and Research)

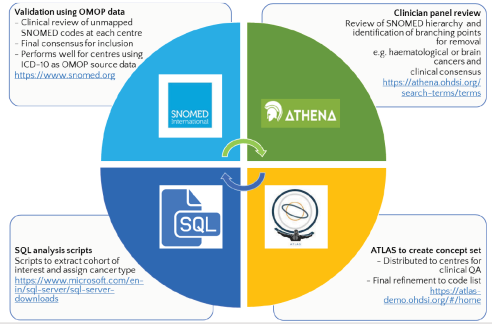
Creating clinically meaningful cancer groups from SNOMED for care systems and care quality research: a head and neck case study

Background Standardization of cancer data increases the potential to deliver data-driven care quality research and clinical dashboards for international collaborations in cancer care. However, to achieve a clinically meaningful and appropriate grouping of cancer indications within organ sites is challenging. This grouping can be described using ICD-10 hierarchies in Europe but alignment with the standard SNOMED concepts is complex. We present here a concept mapping case study from an international COVID-19 analysis that traces pandemic influence on cancer diagnoses and outcomes in hospitals in the UK, Norway and the Netherlands.

Results



Methodological Approach



Conclusion Our approach aimed at adaptation of coding hierarchies based on clinical consensus as an essential step in developing relevant national and international research protocols and data guides.

Hayley Fenton[1,2], Elisabeth Ross[3], Elin Hallan Naderi[3], Anne-Lore Bynens[4]

[1]Leeds Teaching Hospitals NHS Trust, UK
[2]IQVIA Oncology Evidence Network, UK
[3]Oslo University Hospital, Norway
[4]Maastricht University Medical Centre+, The Netherlands

#OHDSISocialShowcase

THURSDAY

Integrating NLP-derived results in the OMOP CDM

(Gabriel Maeztu, Mónica Arrúe, Mariona Forcada, María Quijada, Paula Chocrón, Miguel Ángel Mayer)

Integrating NLP-derived results in the OMOP CDM

Enhancing Real-World Clinical Data Analysis

Hospitals generate vast quantities of data daily, with **up to 80% of this data being collected in an unstructured** format such as free text. Converting this data into a structured format is essential to unlock its value. In this context, **integrating the output of NLP algorithms with the OMOP CDM** offers a compelling solution for enhancing the analysis of real-world clinical data using standard tools such as ATLAS or HADES.

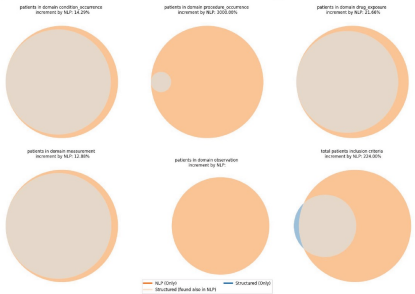
Methods

A study in the field of dermatology was conducted across four hospitals in Spain. We employed multiple NLP techniques to standardize clinical data from clinical notes into OMOP concept_ids. In this study, 132 relevant concepts were defined and extracted using NLP. The results of the NLP were stored in an expanded NLP schema[1] specifically tailored to efficiently store the NLP outputs. This schema was developed to address the complexities of NLP outputs and facilitate their transformation into the standard OMOP CDM v5.4 structures. Notably, OMOP CDM v5.4 includes a *'note_nlp'* table intended for storing NLP results; however, querying this table can often be complex and inefficient.

We finally compared in ATLAS the result of enriching the OMOP with NLP-derived data against structured data.

OMOP Domain	Structured registries	NLP registries	Increase
Condition	10393	295723	2746.25%
Procedure	84	1410	1578.57%
Drug	6970	104629	1398.99%
Measurement	6602	45182	584.18%
Observation	0	78099	+78099

Registries increment per OMOP domain by including NLP results



Patient increment per OMOP domain & inclusion criteria by including NLP results (number of patients masked per compliance reasons)

Results: Integration of NLP-derived results into the OMOP CDM led to significant enhancements in data richness. Firstly, we identified **224% more patients** across four different hospitals in Spain who met the inclusion criteria thanks to NLP-derived data. Moreover, the dataset incorporating NLP demonstrated a substantial increment in the proportion of records across different OMOP domains compared to the dataset without NLP. The structured inclusion of NLP-derived results facilitated more comprehensive analyses, enabling deeper insights into treatment patterns and patient outcomes.

[1] Extending the OMOP CDM to store the output of NLP pipelines. *M.Arrue et al.* OHDSI Global Symposium

Gabriel Maeztu gabriel.maeztu@iomed.health
Mónica Arrúe monica.arrue@iomed.health
Mariona Forcada mariona.forcada@iomed.health
María Quijada maria.quijada@iomed.health
Paula Chocrón paula.chocron@iomed.health
Miguel Ángel Mayer miguelangel.mayer@upf.edu
IOMED

#OHDSISocialShowcase

Power in numbers: overcoming the **scarcity of rare cancer** data by **harmonizing European sarcoma registries**

FRIDAY

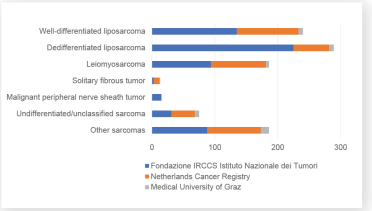
Standardizing European sarcoma registry data to the OMOP Common Data Model: the retroperitoneal sarcoma use case

(Peter Prinsen, Paolo Lasalvia, Roberto Lillini, Vittoria Ramella, Anna Alloni, Joanna Szkandera, Espen Enerly, Maaike van Swieten, Siri Larønningen, Julien Bollard, Audrey Pons, Thomas Gaudin, Claire Chemin-Airiau, Alric Sans, Jean-Yves Blay, Arnaud Malfilatre, Danielle Newby, Gijs Geleijnse, Annalisa Trama)

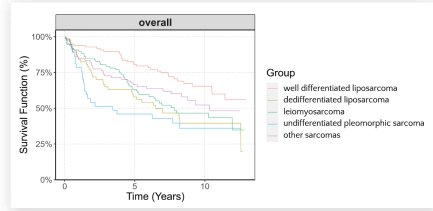
Standardizing European sarcoma registry data to the OMOP Common Data Model: the retroperitoneal sarcoma use case

Background: Research in rare cancers is hampered by low patient numbers, dispersed clinical data and tumor samples, and a limited number of experts per rare cancer diagnosis. The ERN EURACAN (European Reference Network for Rare Adult Solid Cancers) was established to bring together data and knowledge of European Healthcare professionals.

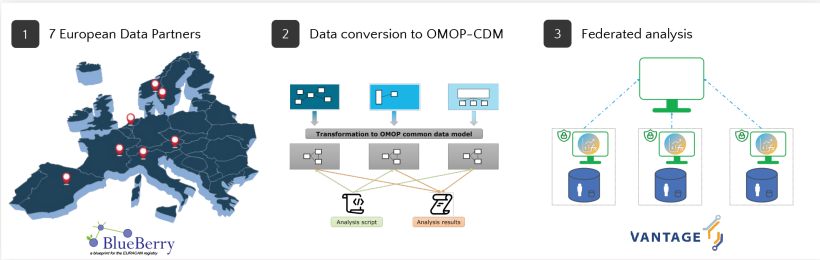
Result 1: # patients of each of the retroperitoneal sarcoma types in each registry



Result 2: Examples of survival curves generated with the CohortSurvival package.



Methods

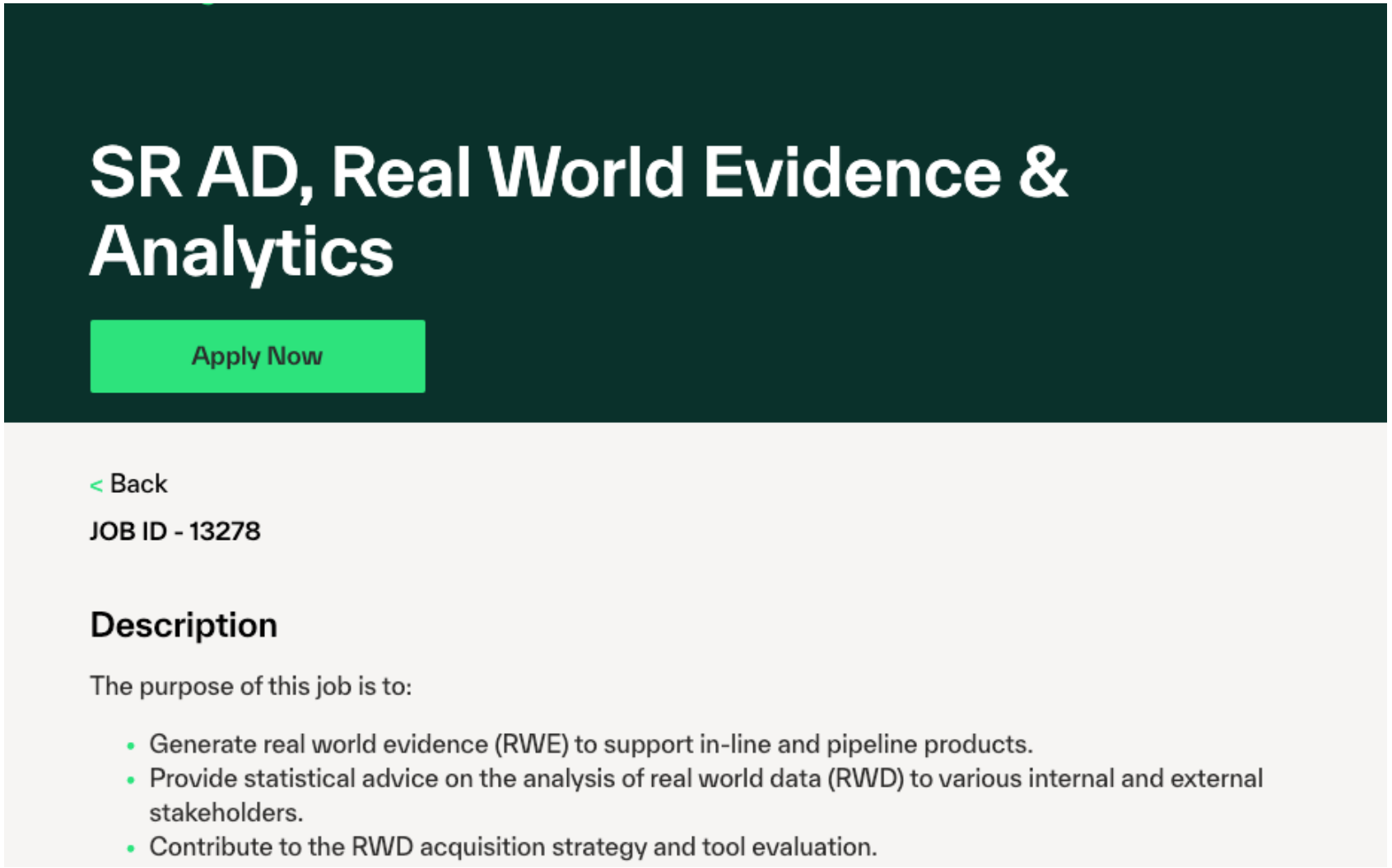


Limitation: Conversion of each of the data sources to the OMOP-CDM and setting up a node for federated analysis requires a lot of technical expertise. The OHDSI tools have a steep learning curve.

OHDSI

Peter Prinsen, Paolo Lasalvia, Roberto Lillini, Vittoria Ramella, Anna Alloni, Joanna Szkandera, Espen Enerly, Maaike van Swieten, Siri Larønningen, Julien Bollard, Audrey Pons, Thomas Gaudin, Claire Chemin-Airiau, Alric Sans, Jean-Yves Blay, Arnaud Malfilatre, Danielle Newby, Gijs Geleijnse, Annalisa Trama

Opening: Sr AD, Real World Evidence & Analytics Boehringer Ingelheim



Openings: Postdoctoral Fellow, Johns Hopkins Univ.

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

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

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?

Building The OHDSI Evidence Network

1. What are some examples of network studies completed previously?
2. How do I get IRB approval to participate?
3. Is there a deadline to join this initiative? It would be great to get some clarity about the timelines.
4. Is there an expectation to submit updated DbDiagnostic results on a regular basis? Is there a QA process and re-submission required after each run?

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