Building Up The OHDSI Evidence Network

(Session 1)

OHDSI Community Call

July 9, 2024 • 11 am ET

Upcoming Community Calls

Date	Topic
July 9	Building The OHDSI Evidence Network Sprint
July 16	HowOften Initiative & Early Results
July 23	Building The OHDSI Evidence Network Sprint
July 30	Advances in Patient-Level Prediction
Aug. 6	Building The OHDSI Evidence Network Sprint
Aug. 13	Global Symposium Tutorials
Aug. 20	Building The OHDSI Evidence Network Sprint
Aug. 27	canceled due to ISPE 2024
Sept. 3	New Standardized Vocabularies Release

July 16: HowOften Initiative & Early Results

Hsin Yi “Cindy” Chen

MD-PhD Student
Columbia University Department of Biomedical Informatics

Azza Shoaibi

Director, Observational Health Data Analytics
Janssen Research and Development

Elise Ruan

Clinical Informatics Fellow
NewYork-Presbyterian Hospital/Columbia University

George Hripcsak

Professor of Biomedical Informatics
Columbia University

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Jason Patterson and Nicholas Tatonetti** on the publication of **KG-LIME: predicting individualized risk of adverse drug events for multiple sclerosis disease-modifying therapy** in *JAMIA*.

OHDSI Shoutouts!

Congratulations to the team of **Yong Shang, Yu Tian, Kewei Lyu, Tianshu Zhou, Ping Zhang, Jianghua Chen, and Jingsong Li** on the publication of **Electronic Health Record–Oriented Knowledge Graph System for Collaborative Clinical Decision Support Using Multicenter Fragmented Medical Data: Design and Application Study** in the *Journal of Medical Internet Research*.

Journal of Medical Internet Research | Journal Information | Browse Journal | Sub

Published on 5.7.2024 in Vol 26 (2024)

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/54263, first published November 03, 2023.

Electronic Health Record–Oriented Knowledge Graph System for Collaborative Clinical Decision Support Using Multicenter Fragmented Medical Data: Design and Application Study

Yong Shang[1]; Yu Tian[2,3]; Kewei Lyu[2,3]; Tianshu Zhou[1]; Ping Zhang[4]; Jianghua Chen[4]; Jingsong Li[1]

Article | Authors | Cited by | Tweetations | Metrics

- Abstract
- Introduction
- Methods
- Results
- Discussion
- References
- Abbreviations
- Copyright

Abstract

Background:
The medical knowledge graph provides explainable decision support, helping clinicians with prompt diagnosis and treatment suggestions. However, in real-world clinical practice, patients visit different hospitals seeking various medical services, resulting in fragmented patient data across hospitals. With data security issues, data fragmentation limits the application of knowledge graphs because single-hospital data cannot provide complete evidence for generating precise decision support and comprehensive explanations. It is important to study new methods for knowledge graph systems to integrate into multicenter, information-sensitive medical environments, using fragmented patient records for decision support while maintaining data privacy and security.

Objective:
This study aims to propose an electronic health record (EHR)–oriented knowledge graph system for collaborative reasoning with multicenter fragmented patient medical data, all the while preserving data privacy.

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	12 pm	Generative AI and Analytics
Wednesday	7 am	Medical Imaging
Wednesday	9 am	Patient-Level Prediction
Wednesday	12 pm	Health Equity/OMOP + FHIR
Wednesday	2 pm	Natural Language Processig
Wednesday	4 pm	Vulcan/OHDSI
Thursday	9:30 am	Network Data Quality
Thursday	10:30 am	Evidence Network
Thursday	12 pm	Strategus Subgroup
Thursday	6 pm	Eyecare and Vision Research
Thursday	7 pm	Dentistry
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	GIS-Geographic Information System
Friday	11:30 am	Steering Group
Friday	11 pm	China Chapter
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Healthcare Systems Interest Group
Monday	11 am	Data Bricks User Group
Monday	2 pm	Electronic Animal Health Records

Health Equity / OMOP + FHIR Cross-Working Group Collaboration

Kelly Davison, RN, MN, MSc, CPHMN(C), CTSS
University of Victoria, British Columbia

HL7 Gender Harmony Project: Sex and Gender Representation

Wednesday, July 10 at 12pm ET (Meeting Link)

2024 OHDSI Europe Symposium

2024 OHDSI Europe Symposium

SESSION ONE

00:00 - Welcome to the European OHDSI Journey

Peter Rijnbeek, PhD, Chair, Department of Medical Informatics, Erasmus MC

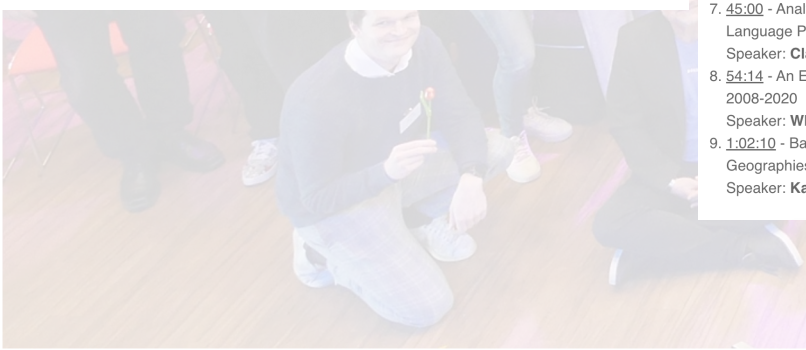
14:00 - Journey of OHDSI: Where Have We Been and Where We Can Go Together?

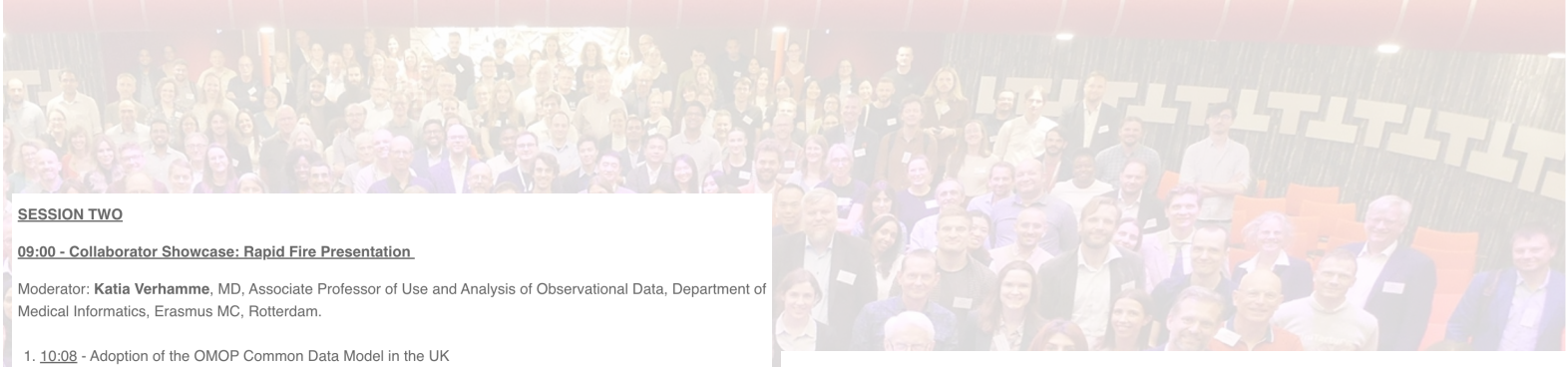
Patrick Ryan, PhD, Janssen Research and Development, Department of Biomedical Informatics, Columbia University Medical Center

42:45 - Selection of European Initiatives Using the OMOP CDM

Moderator: **Renske Los**, PhD, Assistant Professor of Medical Informatics, Department of Medical Informatics, Erasmus MC

1. 43:33 - OHDSI Europe National Nodes building opportunities through collaboration - **Renske Los**, Erasmus MC, The Netherlands
2. 1:05:00 - ONCOVALUE: Can Real-World Data Shape the Future of Health Technology Assessment in Oncology? - **Andreas Henriksen**, Copenhagen University Hospital, Denmark
3. 1:12:17 - DigiONE: technical challenges and solutions to the European cancer OMOP conversions from hospital EHR - **Piers Mahon**, DIGICORE, UK
4. 1:22:10 - The PHEMS Project: New Strategies in Health Data Sharing - **Sofia Bazakou**, The Hyve, The Netherlands
5. 1:28:00 - Ecraid: European Clinical Research Alliance on Infectious Diseases - **Ankur Krishnan**, Heidelberg University Hospital, Germany
6. 1:40:10 - PHederation - the Federated Network of Pulmonary Hypertension Registries - **Eva-Maria Didden**, Actelion, a Johnson & Johnson Company, Belgium
7. 1:48:05 - Q&A





SESSION TWO

09:00 - Collaborator Showcase: Rapid Fire Presentation

Moderator: **Katia Verhamme**, MD, Associate Professor of Use and Analysis of Observational Data, Department of Medical Informatics, Erasmus MC, Rotterdam.

1. 10:08 - Adoption of the OMOP Common Data Model in the UK
   Speaker: **Alex Knight**, Health Data Research UK
2. 15:39 - Piloting the Transformation of Multiple Sclerosis Real-World Data to the OMOP CDM: Lessons Learned
   Speaker: **Tina Parciak**, UHasselt
3. 21:45 - Annotation-preserving machine translation of English corpora to validate Dutch clinical concept extraction tools
   Speaker: **Tom Seinen**, Erasmus MC
4. 28:49 - Beyond Diagnostic Codes: A Weakly Supervised learning Framework for Accurate Multimorbidity Identification in Electronic Health Records
   Speaker: **Luz Saúde**, Portugal
5. 34:05 - OHDSI meets Flowise to Streamline Biomedical Data Discovery and Analysis
   Speaker: **João Almeida**
6. 38:16 - ReportGenerator: Automating study reports and visualization apps for DARWIN EU® research
   Speaker: **Cesar Barboza Gutierrez**, Erasmus MC
7. 45:00 - Analysis of Lung Cancer Patient Treatment with Immune Checkpoint Inhibitors Using Natural Language Processing for Data Extraction from Electronic Health Records
   Speaker: **Clara L. Oeste and Annelies Verbiest**, Lynxcare
8. 54:14 - An Exploration of Ovarian Cancer Therapy Sequence Utilization in Treatment-naive Women from 2008-2020
   Speaker: **Whitney Burton**, Taipei Medical University
9. 1:02:10 - Baseline Characterization and Treatment Pathways of Patients With Alport Syndrome Across Geographies: Exploring a Rare Disease in Multi-Database Retrospective Cohort Study
   Speaker: **Katrin Manlik**, Bayer AG

SESSION THREE

06:55 - Large Scale Evidence Generation in EHDEN and DARWIN EU®

Moderators: **Daniel Prieto-Alhambra and Katia Verhamme**, Department of Medical Informatics, Erasmus MC

1. 06:55 - Introduction to EHDEN - **Daniel Prieto Alhambra**
2. 10:15 - Predicting long term cancer survival for Health Technology Assessment: A Multinational Cohort Study Across Europe - **Jeremy Dietz**, National Institute for Health and Care Excellence
3. 20:05 - Trends over time in medicines with suggested shortages in Europe - **Marta Pineda-Moncusi**, Oxford University
4. 29:05 - Q&A
5. 35:10 - Introduction to DARWIN EU® - **Katia Verhamme**
6. 40:20 - DARWIN EU®- Trend of prescription opioid use in Europe - **Annika Jödicke**, Oxford University
7. 53:05 - DARWIN EU®- Treatments of multiple myeloma in Europe from 2012-2022: a population-based network cohort study - **Talita Duarte Salles**, Erasmus MC
8. 1:04:10 - Q&A

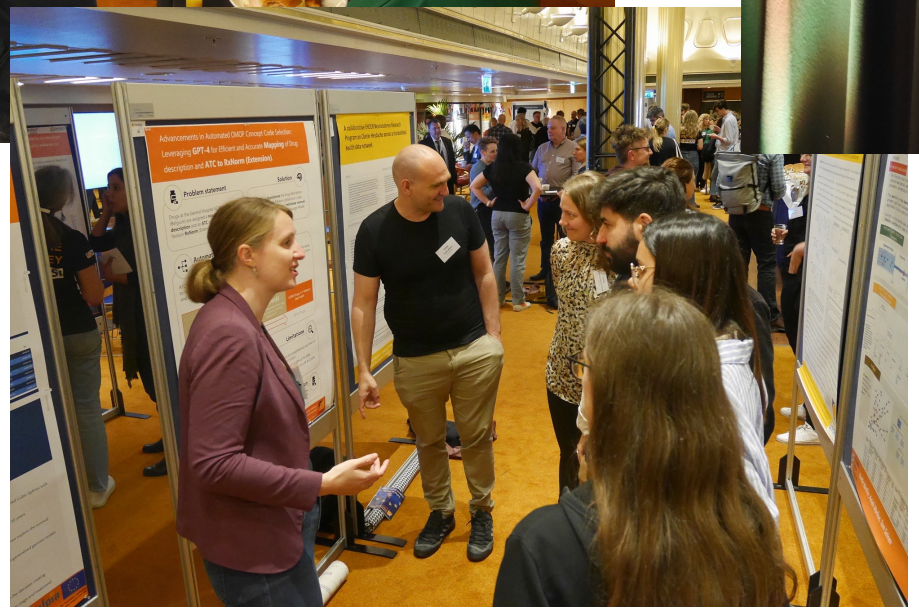
1:09:00 - What evidence are we going to showcase at OHDSI Europe in 2025?

Patrick Ryan, Johnson & Johnson, Columbia University

1:32:00 - Closure

Peter Rijnbeek, PhD, Chair, Department of Medical Informatics, Erasmus MC

2024 OHDSI Europe Symposium



2024 OHDSI Europe Symposium

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

Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University.

Join Us At The 2024 Global Symposium

Registration is now open for the 2024 Global Symposium, which will be held October 22-24 at the Hyatt Regency Hotel in New Brunswick, N.J., USA. Check out the event page for details on the collaborator showcase, tutorial offerings, workshop activities, and more!

2024 Global Symposium Homepage

OHDSI EUROPE
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Login

Home | Info | Past Events | National Nodes | Publications | Academy | MS Teams | Forum | Github | Photos | Contact

OHDSI Symposium 'Scaling up Reliable Evidence across Europe' June 3rd, 2024

The fifth European OHDSI Symposium called "Scaling up Reliable Evidence Across Europe" aimed to bring data partners, regulators, and researchers to collaborate and share results and ideas about the use of the OMOP-CDM in Europe.

The main symposium took place on Monday, **June 3rd 2024** at the ship SS Rotterdam. **Saturday June 1st**, and **Sunday June 2nd** were dedicated to workshop and workgroup meetings, held in the Education Centre of the Erasmus University Medical Center, Rotterdam.

The symposium had 360 participants, 25 plenary presentations, 130 posters, and 13 software demos.

- Event Page
- Workshop
- Workgroups
- Booklet

Video Recordings can be found on the OHDSI YouTube Channel.

- Posters
- Software demonstrations
- Sponsors
- Photos

Slides	Video	Description	Photo
		SESSION ONE **00:00 - Welcome to the European OHDSI Journey** **Peter Rijnbeek**, PhD, Chair, Department of Medical Informatics, Erasmus MC **14:00 - Journey of OHDSI: Where Have We Been and Where We Can Go Together?** **Patrick Ryan**, PhD, Janssen Research and Development, Department of Biomedical Informatics, Columbia University Medical Center	

Collaborator Spotlight: Sarah Seager

Sarah Seager is the Senior Director, Analytics & AI at IQVIA. She is an experienced technical senior leader who develops and leads analytical teams in the world of data science and advanced analytics.

In the latest edition of the Collaborator Spotlight, Sarah shares her thoughts on the current path of data management and analytics, why OMOP is an ideal common data model, recent advances around the European community, and how data is similar to another one of Sarah's passions, art.

ohdsi.org/spotlight-sarah-seager

Next CBER Best Seminar: July 17

Speaker: Yonas Ghebremichael-Weldeselassie, Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK

Topic: A modified self-controlled case series method for event-dependent exposures and high event-related mortality, with application to COVID-19 vaccine safety

Date/Time: Wednesday, July 17, 11 am ET

Upcoming Seminars

— July 17, 2024 (11 am) - Yonas Ghebremichael-Weldeselassie, Warwick Medical School

Topic: A modified self-controlled case series method for event-dependent exposures and high event-related mortality, with application to COVID-19 vaccine safety

Presenter: Yonas Ghebremichael-Weldeselassie, Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK

Watch This Seminar

Abstract:

We propose a modified self-controlled case series (SCCS) method to handle both event-dependent exposures and high event-related mortality. This development is motivated by an epidemiological study undertaken in France to quantify potential risks of cardiovascular events associated with COVID-19 vaccines. Event-dependence of vaccinations, and high event-related mortality, are likely to arise in other SCCS studies of COVID-19 vaccine safety. Using this case study and simulations to broaden its scope, we explore these features and the biases they may generate, implement the modified SCCS model, illustrate some of the properties of this model, and develop a new test for presence of a dose effect. The model we propose has wider application, notably when the event of interest is death.

Bio: Yonas Weldeselassie is a Lecturer of Statistics at School of Mathematics and Statistics, The Open University, UK. He graduated in statistics and demography from University of Asmara, Eritrea and went on to become an assistant lecturer in Mekelle University, Ethiopia, and then a Senior Research Fellow in Medical Statistics at Warwick Medical School, division of Population Evidence and Technologies. He earned a Msc in Biostatistics from Hasselt University, Belgium and PhD in statistics from the Open University, UK. After working as a research associate, on MRC project 'Software tools and online resources for the self-controlled case series method and its extensions', at the department of mathematics and statistics, the Open University since 2014, he joined Warwick Medical School in June 2017. His main research interest is in medical statistics specially in the methodological development and application of the self-controlled case series (SCCS) method. He published a book on SCCS with Paddy Farrington and Heather Whitaker, and he is currently working on early prediction of gestational diabetes mellitus.

ohdsi.org/cber-best-seminar-series

#OHDSI2024 Registration Is Open!

Registration is 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

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

Melissa Haendel, PhD

Director of Precision Health & Translational Informatics and the Sarah Graham Kenan Distinguished Professor in the Department of Genetics at The University of North Carolina at Chapel Hill and co-founder of the Monarch Initiative and the National Covid Cohort Collaborative

‘Journeys across the translational divide: making healthcare and basic research data interoperable’

July 25, 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

Additional technical data protection measures to improve the security of an OMOP/OHDSI infrastructure

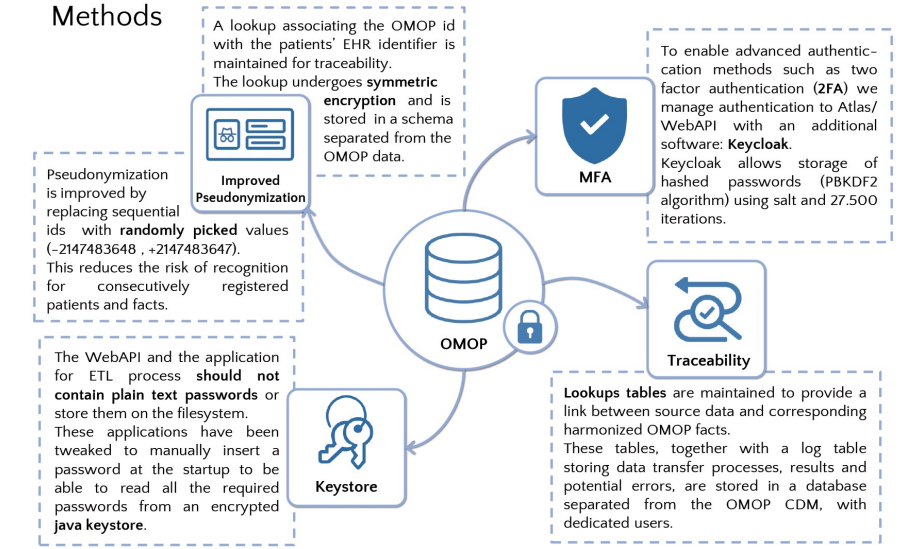
(Francesco Pozzoni, Matteo Gabetta, Mauro Bucalo, Anna Alloni, Giorgia Masina, Maurizio Pastore, Lucia Sacchi, Nicola Barbarini)

An improved OMOP infrastructure to provide greater security

Additional technical data protection measures to improve the security of an OMOP/OHDSI infrastructure

Background: According to the evaluation criteria by EDPB (European Data Protection Board), the implementation of an OMOP infrastructure is to be considered high risk due to its processing of sensitive data and to the large-scale data processing involved. We present additional security/privacy preserving measures identified and implemented to complement those already considered best practices within the OHDSI community.

Methods



Results: The described technical data protection measures have been included in the DPIA (Data protection impact assessment) template created within "OHDSI Italia" node, reviewed by DPOs from clinical centers. They have been included in the DPIAs related to the implementation of OMOP in several Italian centers. These measures have also been included in the BIOMERIS information security management system certified by ISO/IEC 27001:2022 for consulting services related to OMOP.

Francesco Pozzoni, Matteo Gabetta, Mauro Bucalo, Anna Alloni, Giorgia Masina, Maurizio Pastore, Lucia Sacchi, Nicola Barbarini

BIOMERIS

OHDSI

#OHDSISocialShowcase

TUESDAY

Forecasting the prescription rates of antibiotics in the UK between 2013 to 2023 incorporating the impact of COVID-19

(Yuchen Guo, Edward Burn, Daniel Prieto-Alhambra, Marta Pineda-Moncusí)

Using a term to model the impact of COVID-19 between January 2020 - December 2021 plus a 6-month tune period produced more accurate predictions when forecasting the use of antibiotics with ARIMA models

Forecasting the prescription rates of antibiotics in the UK between 2013 to 2023 incorporating the impact of COVID-19

Methods

- Study design: Retrospective analysis including all individuals in CPRD GOLD (UK) mapped to OMOP-CDM
- Exposure: monthly use of 6 antibiotics
- Outcome: incidence rates of antibiotic use
- Study period: Jul 2013 to May 2023
- Statistical methods: ARIMA and ARIMA with Exogenous Variable (ARIMAX) models. (**Figure 1**)

This study was run under the scenario that the available data ended in Nov 2022, and we forecasted the usage of antibiotics for the next 6-months. We have used the remaining data up to May 2023 to test the accuracy of the models pretending we received these subsequent data later.

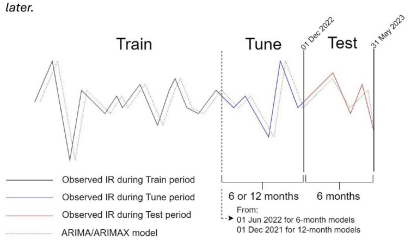


Figure 1. Graphical representation of the time periods where the ARIMA/ARIMAX models were trained, tuned and tested. ARIMAX includes a model term from Jan 2020 to Dec 2021 to account for the effect of the COVID-19 pandemic. Tune period: used to evaluate the model's predictive performance with accuracy metrics MAE (mean absolute error) and MAPE (mean absolute percentage error). Test period: used to evaluate the 6-month forecasted through the train and tune periods

After fitting the ARIMA models, the average ratio between the 6- and 12-month showed that using a tune period of 6-month were more accurate than using 12-month (average ratio was 0.66 for MAE and, 0.64 for MAPE). The test period metrics also favoured the 6-month models (average ratio was 0.96 for MAE and, 0.95 for MAPE).

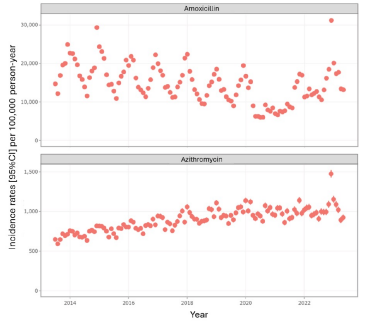


Figure 2. Incident use of amoxicillin and azithromycin in CPRD Gold between 2013 and 2023. We observed a seasonality pattern with peaks around Nov-March months in antibiotics with an indication for bacterial infections frequent in winter, such as the two presented in this figure.

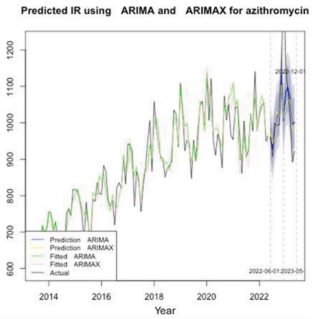


Figure 3. ARIMA and ARIMAX models for the use of azithromycin. When comparing ARIMA vs ARIMAX models, the tune period indicates that ARIMA may be more reliable (average ratio was 0.96 for MAE and, 0.95 for MAPE). However, when the test period was evaluated, the ratios favoured the ARIMAX (average of 1.08 for both, MAE and MAPE).

Yuchen Guo[1], Edward Burn[1], Daniel Prieto-Alhambra[1,2], Marta Pineda-Moncusí[1]
1 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, U.K.
2 Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

#OHDSISocialShowcase

WEDNESDAY

Challenges in harmonising data across multiple biobanks

(Karyn Mégy, Rebecca Akhanemhe, Ben Hollis, Ali Abbasi, Amanda O'Neill, Shikta Das, Stewart MacArthur, Sean O'Dell, Sebastian Wasilewski, Quanli Wang, Slavé Petrovski, Jen Harrow)

Challenges in harmonising data across multiple biobanks

Karyn Mégy[1], Rebecca Akhanemhe[1], Ben Hollis[1], Ali Abbasi[1], Amanda O'Neill[1], Shikta Das[1], Stewart MacArthur[1], Sean O'Dell[1], Sebastian Wasilewski[1], Quanli Wang[2], Slavé Petrovski[1], Jen Harrow[1].

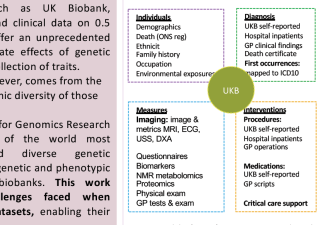
1. Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK. 2. Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Waltham, MA, USA

1. Background

Early-stage incorporation of human genomic data into the assessment of drug targets has been shown to significantly increase drug pipeline success rates.
Large biobanks such as UK Biobank, combining genetic and clinical data on 0.5 million individuals, offer an unprecedented opportunity to evaluate effects of genetic variants on a broad collection of traits.
Statistical power, however, comes from the size as well as the ethnic diversity of those biobanks.
AstraZeneca's Centre for Genomics Research is establishing one of the world most comprehensive and diverse genetic resource, combining genetic and phenotypic data for multiple biobanks. **This work describes the challenges faced when harmonizing such datasets,** enabling their cross analysis.

Different data sets, several data types, multiple standards

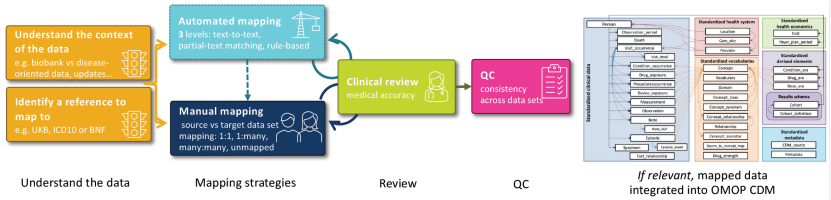
UK Biobank (UKB) is one of the golden standard, in terms of data diversity, sources but also coding systems. However, it is very reflective of the UK population and health care system. Biobanks from different countries will be using different coding system, different units (e.g. *Hba1c: mmol/mol vs.* %), medication names (e.g. *metformin vs. metformina*), and the local language, making comparison of those data sets challenging.



	Source of health data available in our cohorts					
	Hospital data	Primary care	Cancer data	Questions	Free text	And also....
UK Biobank	WHO ICD9 & 10	Read2 & 3	yes	formatted	-	Lab, proc.
US cohort #1	CM ICD9 & 10	CM ICD9 & 10	CM ICD9 & 10	-	-	Lab, proc.
UK cohort #1	WHO ICD9 & 10	Read2 & 3	yes	-	-	Lab, proc.
FinnGen	WHO ICD8, 9 & 10					
MCPS	-	-	WHO ICD10	WHO ICD10	yes	Lab.

ICD: International Classification of Diseases, versions 8, 9 and 10, in the WHO or US (CM) system
Snomed: Systematized Nomenclature of Medicine Clinical Terms ; **Lab & proc.**: laboratory results and procedures

2. Harmonisation process

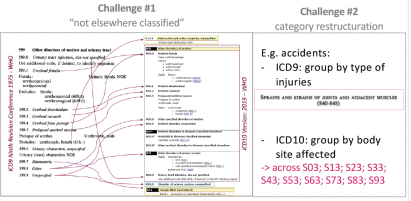


3. Results

E.g.1: mapping a small disease-oriented data set to a large biobank
We mapped one of our data sets, a small disease-oriented resource, to UK Biobank. An initial mapping was done manually, following the process described above, high-quality but timely. As a test, we then performed NLP (Natural Language Processing) on that same data set, reducing the mapping time from months to a week, however <50% of terms could be mapped.
=> NLP followed by manual mapping would be most efficient strategy in term of time and accuracy.
The final mapping will be transformed into the **OMOP common data model**

Approach	Description	Example
Text-to-text match	Exact Text match	'Age' = 'Age at recruitment'
Partial text match	**Step 1:** Sub-word matching, part of substring matches Field ID **Step 2:** Sub-word matching and matching all words present in column	'Supplement' = 'Supplements' 'weight loss' = 'loss in weight'
Rule-based match	Identify key terms and match to UKB Field ID **Layer 1:** Age or family history-based fields can be identified using 'age', 'fh', '<brother/mother>' terms **Layer 2:** Diseases and symptoms can be identified using ontologies and classified as cancer vs non-cancer **Layer 3:** Category and Question information to identify terms like self-reported, subsection etc	20002 (self-reported non-cancer)

E.g.2: mapping across ontologies, from UKB ICD9 to ICD10
UK Biobank diagnoses are encoded both in the ICD9 and in ICD10 WHO classifications. Following our harmonisation process, we have mapped the ICD9 terms present in the UKB data set to ICD10 and, according to the FAIR principles, are returning the results to UK Biobank so that they can be **shared with the community** (*manuscript in preparation*).
⇒ In total, we mapped 751 ICD9 codes to 573 unique ICD10 codes, with 85% having a 1:1 mapping.



4. Take home messages

- Harmonising data sets require to understand the data and adapt the strategy when needed.
- NLP followed by manual mapping is be most efficient strategy for mapping across cohorts
- Importance of return of data for use by the community (FAIR principle)
- Gap in OMOP: mapping of images & medications

Acknowledgements - We would like to thank the **participants and investigators in the UK Biobank** study who made this work possible.
We also acknowledge contribution from members of the **AstraZeneca Genomics Initiative** and the AZ IT Knowledge EngineeringTeam.

AstraZeneca

#OHDSISocialShowcase

THURSDAY

Predicting long term cancer survival for Health Technology Assessment: A multinational cohort study across Europe

(Jeremy Dietz, Ian Koblbauer, Ravinder Claire, James Love-Koh, Jamie Elvidge, Irene López-Sánchez, Laura Pérez-Crespo, Anna Palomar Cros, Asieh Golozar, Antonella Delmestri, George Corby, Marta Alcalde Herraiz, Abigail Robinson, Marti Catala Sabate, Edward Burn, Wai Yi Man, Xihang Chen, Miguel-Angel Mayer, Juan Manuel Ramirez Anguita, Maria Angeles Leis Machin, Nicola Symmers, Mahéva Vallet, Colin McLean, Peter Hall, Mees Mosseveld, Katia Verhamme, Espen Enerly, Peter Prinsen, Jelle Evers, Marek Oja, Raivo Kolde, Eric Fey, Tiago Taveira Gomes, Alberto Moreno Conde, Evelyne Fournier, Tommi Kauko, Rafael Marcos Gragera, Talita Duarte Salles, Dalia Dawoud, Danielle Newby)

NICE National Institute for Health and Care Excellence | UNIVERSITY OF OXFORD | EHDEN OHDSI

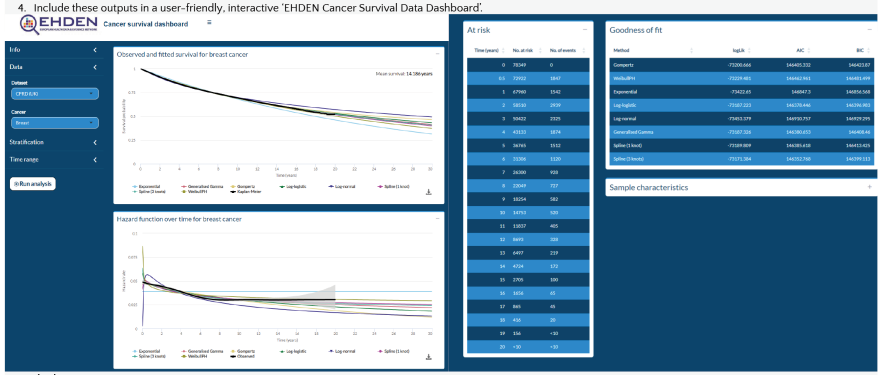
We have developed an **EHDEN Cancer Survival Dashboard**, allowing users to quickly examine **survival data** and explore **long-term projections**

Predicting long term cancer survival for Health Technology Assessment: A multinational cohort study across Europe

Introduction: In cancer health technology assessments (HTA), extrapolations are used to estimate survival beyond observed trial data. This is used in economic evaluations, which typically assess cost effectiveness over a lifetime. However, this is a key source of decision uncertainty because it involves forecasting the future based on shorter-term observed data. Real-world evidence can help address this uncertainty. For example, country-specific, real-world survival data for patients receiving the current standard of care could be used to validate the trial-based survival extrapolations typically used in HTA processes. Here, we describe an EHDEN use.

Aims: Using data from 14 databases mapped to the OMOP-CDM from the European Health Data & Evidence Network (EHDEN) from seven European countries (UK, Norway, Switzerland, Spain, the Netherlands, Finland and Portugal) we aimed to:

1. Develop and assess data quality of phenotypes for the identification of breast, pancreas, prostate, colorectal, lung, stomach, liver, and head and neck cancer.
2. Estimate overall survival of the studied cancers, stratifying by age and sex.
3. Fit parametric survival models to assess fit of distributions across observed data and allow for extrapolations of the studied cancers.
4. Include these outputs in a user-friendly, interactive 'EHDEN Cancer Survival Data Dashboard'.



Methods

- Inclusion criteria: Patients aged ≥18 with primary diagnoses of breast, pancreatic, prostate, colorectal, lung, stomach, liver, or head and neck cancer.
- Study period: 1st January 2000 to 31st December 2019, followed-up from diagnosis to death, database exit, or end of study.
- Curve fitting: Six standard (generalised gamma, lognormal, loglogistic, exponential, Weibull, Gompertz) and two flexible parametric models (Restricted Cubic Splines with one and three internal knots).
- Assessment: Akaike information criterion (AIC) and visual inspections used to assess model fit and validity of predicted survival.
- Outcome of Interest: Restricted mean survival time (RMST) difference between observed (Kaplan-Meier) and predicted (parametric) estimates at 10 years.

Results

- An interactive R Shiny cancer survival dashboard was developed. Users of the dashboard can select a database from a dropdown menu, and then explore the survival models fitted for the eight different cancers with age and sex adjustments/stratifications.
- Generally, flexible spline models performed best in terms of model fit (AIC) and performance (difference in observed versus predicted RMST).
- For standard parametric models, the best performing models differed by cancer and database however in general, lognormal, loglogistic, and generalised gamma models performing best across included cancers and databases.

Discussion

- This use case demonstrates a potential benefit of EHDEN and OHDSI tools to address priority areas of HTA agencies and industry.
- To our knowledge this is the first study to examine which distributions fit RWD best across multiple cancers, databases, and countries.
- This could have real benefits to HTA agencies as these results may be used to rule out implausible distributions and reduce uncertainty in the decision-making process (although this should be done on a case-by-case basis).
- Further research investigating generalisability of results for use in HTA is required, specifically regarding adjustment approaches (e.g. cancer stage and treatment).

Jeremy Dietz[1], Ian Koblbauer[2], Ravinder Claire[1], James Love-Koh[1], Jamie Elvidge[1], Irene López-Sánchez[3], Laura Pérez-Crespo[3], Anna Palomar Cros[3], Asieh Golozar[1,11], Antonella Delmestri[2], George Corby[4], Marta Alcalde Herraiz[4], Abigail Robinson[1], Marti Catala Sabate[5], Edward Burn[2], Wai Yi Man[2], Xihang Chen[2], Miguel-Angel Mayer[6], Juan Manuel Ramirez Anguita[6], Maria Angeles Leis Machin[6], Nicola Symmers[1], Mahéva Vallet[6], Colin McLean[6], Peter Hall[6], Mees Mosseveld[7], Katia Verhamme[7], Espen Enerly[8], Peter Prinsen[9], Jelle Evers[9], Marek Oja[10], Raivo Kolde[10], Eric Fey[11], Tiago Taveira Gomes[12], Alberto Moreno Conde[13], Evelyne Fournier[14], Tommi Kauko[15], Rafael Marcos Gragera[16], Talita Duarte Salles[17], Dalia Dawoud[1], Danielle Newby[1,*]
[1]National Institute for Health and Care Excellence, London, UK
[2]Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK
[3]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
[*]For a comprehensive affiliation list, please see separate sheet.

The European Health Data & Evidence Network has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

#OHDSISocialShowcase

FRIDAY

CDM Onboarding R package for data quality assessment

(Anne van Winzum, Maxim Moinat, Sofia Bazakou)

A comprehensive *report* that provides *insights* into the *completeness*, *transparency*, and *quality* of the Extract Transform Load (ETL) process.

CDM Onboarding R package for data quality assessment.

Data quality assessment of an observational health data set is an important aspect when deciding whether the data is suitable to answer a selected research question. Accordingly, the OHDSI (Observational Health Data Sciences and Informatics) community has developed a variety of tools to enable this need. Achilles [5] and the Data Quality Dashboard [3] are incorporated into the CDM Onboarding report.

The CDM Onboarding is an R package, developed by the DARWIN EU® Coordination Centre (CC) based on the EHDEN CDM Inspection report. The outcome is a comprehensive report that provides insights into the completeness, transparency, and quality of the Extraction, Transform, and Load (ETL) process step of the OHDSI journey. Additionally, at the DARWIN EU® onboarding process it helps with the evaluation of the new data partner's preparedness to join the DARWIN EU® data network and actively participate in research studies [1]. Building upon the CDM Inspection report foundation from the EHDEN consortium [2], the CDM Onboarding package integrates supplementary checks, enhancing the provision of valuable information. Noteworthy examples of these additional checks are highlighted below.

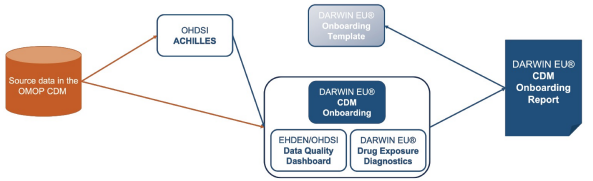


Figure 1: A schematic representation of the input for the CDM Onboarding report, a crucial part of the onboarding and refresh process for the DARWIN EU® Network Operations pillar.

Methods

1 Data Quality Dashboard

The results file extracted by the Data Quality Dashboard [3] on the OMOP CDM instance can now be provided to give an overview of the number of passed and failed checks and will be displayed in the report.

5. Data Quality Dashboard

DataQualityDashboard Version: 2.0.0
DataQualityDashboard executed at 2024-05-05 11:52:57 in 8 mins.

Table 40. Number of passed, failed and total DQD checks per category. For DQD v2, the checks with status 'NA' are not included.

Category	Pass	Fail	Total	%Pass
Plausibility	141	8	2,500	5.6%
Conformance	791	3	1,072	73.8%
Completeness	202	55	464	43.5%
Total	3,970	66	4,036	98.4%

2 Drug Exposure Diagnostics

The Drug Exposure Diagnostics is an R package [4], developed by DARWIN EU® CC and it is used to summarize ingredient-specific drug exposure data in the OMOP CDM. As an example, we have selected a set of eleven common ingredients with different ways of administration to run Drug Exposure Diagnostics on. The results are presented in the report as a summary (Table 1) and used upon onboarding for data quality checks and during study feasibility for more in-depth analysis. Indicative quality checks performed:

- Whether the routes correspond to the expectation. For example, Acetaminophen is expected to be given orally, whilst Albuterol is expected to be inhaled.
- Distribution of amount gives an idea of the strengths prescribed. For example, Acetaminophen is expected to be prescribed mainly as 500mg.
- Distribution of quantities, i.e. how many tablets are prescribed at a time. For example, for Albuterol we either expect a high number (puffs) or a low number (inhalers), but this should be consistent.
- Distribution of exposure days. An important quality check here is whether a duration is available, or always 1, indicating the exposure end date always equals the start date.

▼ Table 1. Example results for Drug Exposure Diagnostics runduring CdmOnboarding, for two of the eleven ingredients. Distributions are reported as median (q05-q95).

Ingredient	Number of records	Route (n,%)	Amount distrib. [null or missing]	Quantity distrib. [null or missing]	Exposure days distrib. [null or missing]	Neg. Days n (%)
acetaminophen	68266	Oral (65604, 96%) Rectal (2662, 3.9%) Intra-arterial (38, 0%) No matching concept (172, 0.1%)	500 (0-1000) [7659, 7.7%]	50 (14-250) [2252, 3.3%]	36 (1-372) [0, 0%]	5 (0%)
albuterol	67308	Respiratory tract (66308, 99.9%) Oral (50, 0.1%) Intra-arterial (11, 0%) No matching concept (479, 0%)	0.4 (0.2-1.8) [66354, 97.2%]	1 (1-9) [341, 0.5%]	1 (1-1) [0, 0%]	2 (0%)

3 Technical infrastructure

Additional assessments were incorporated to evaluate the technical infrastructure. Extracted information includes the indexes applied on the OMOP CDM instances, as well as those that are missing, which can aid in addressing performance issues. Furthermore, the report now lists all versions of the HADES and DARWIN EU packages, with missing packages highlighted.

4 Temporal details

Additional sections with temporal information were added; giving insights into overlapping observation periods, distribution of records over days of the week/month and time ranges of each domain. Figure 2 shows an example of the distribution of records over time. In general, less clinical events happen during the weekend, Death being an exception, and less events happen on the 31st of the month.

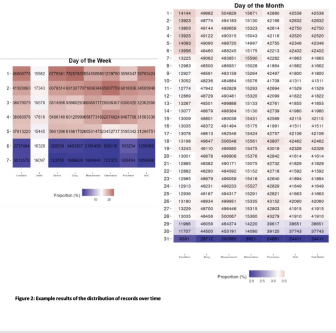


Figure 2: Example results of the distribution of records over time

With these new developments, the DARWIN EU® CC gets further insights into the quality of data converted to the OMOP CDM. This is already showing value in the onboarding of new DARWIN EU® data partners. In the future the plan is to make further developments in the CDM Onboarding package and integrate more R packages like the CDM Connector benchmark.

References
1.https://github.com/darwin-eu/CdmOnboarding
2.https://github.com/EHDEN/cdminspection
3.https://github.com/OHDSI/DataQualityDashboard
4.https://github.com/darwin-eu/DrugExposureDiagnostics
5.https://github.com/OHDSI/Achilles

Anne van Winzum, Maxim Moinat, Sofia Bazakou

Opening: Health Data Scientist, Erasmus MC

Erasmus MC

Jobs | About Erasmus MC | Research facilities | Career | NL / EN

Health Data Scientist

Would you like to be responsible for the creation and execution of study software to support our researchers? Join our team!

Closing date 18-07-2024 # 27.16.24.TG

36 hours

University of Applied Sciences

IT

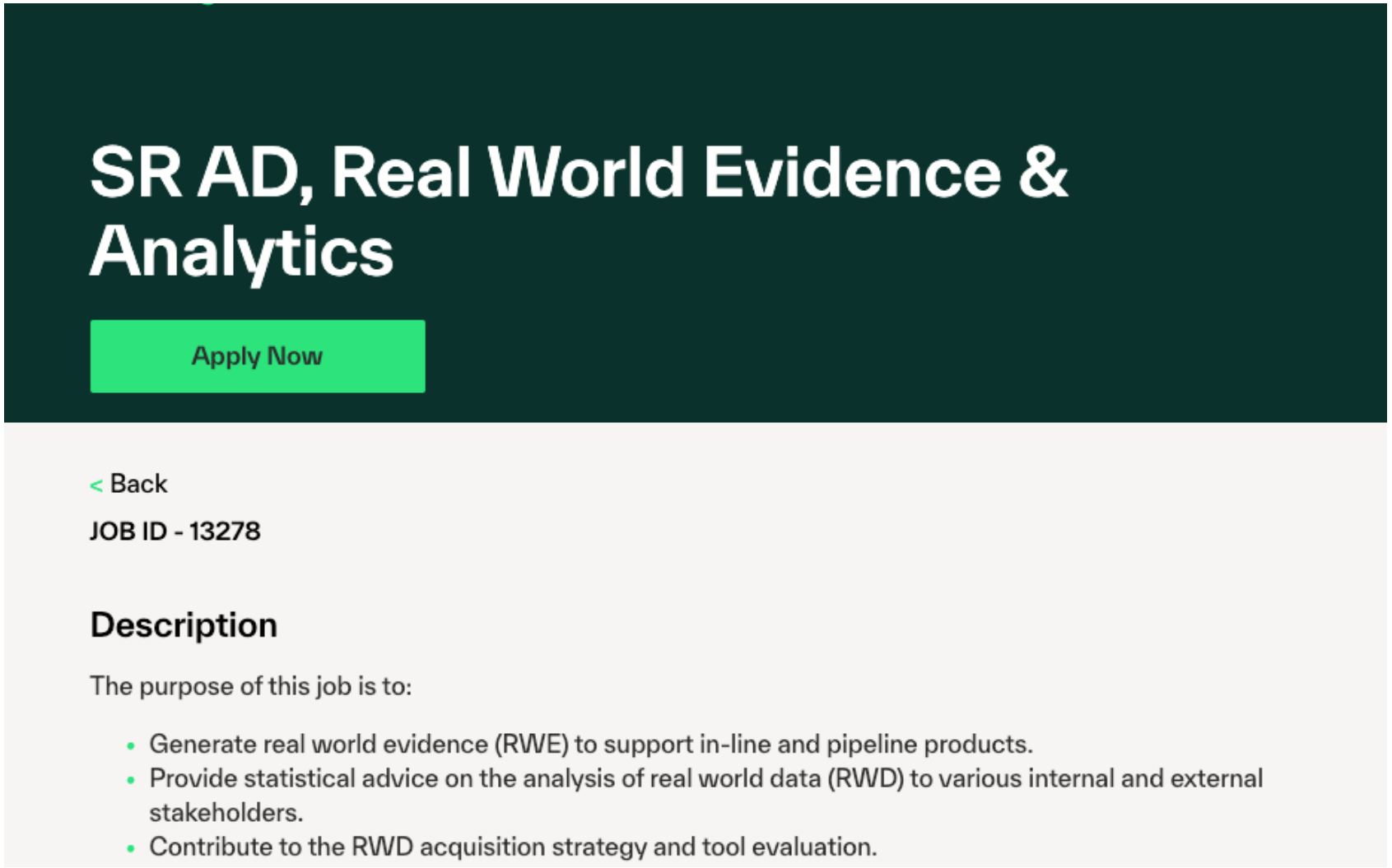
Apply

Job description

The Health Data Science group at the department of Medical Informatics is looking for an experienced R programmer to develop open-source analytics.

You will be responsible for the creation and execution of study software to support the work of epidemiologists. This will include designing, developing, documenting, and maintaining R code that will be executed against health data that is standardized to the OMOP Common Data Model (OMOP-CDM). This data model is maintained by the Observational Health Data Sciences and Informatics (OHDSI, www.ohdsi.org) initiative. The Department is leading the Data Analysis and Real World Interrogation Network coordination centre (DARWIN EU®) in which a large data network is created with data sources that are mapped to the OMOP-CDM. The aim of this network is to provide fast and reliable evidence to the European Medicines Agency. Your work will enable the use of health data at an unprecedented scale in Europe and will facilitate the execution of impactful studies to improve patient care. We recommend applicants to review the DARWIN EU ® software packages to get an understanding of the growing set of R packages.

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



Opening: Lead Director, RWE Distributed Research CVS Health

Lead Director, RWE Distributed Research

Apply

Hybrid

PA - Blue Bell
IL - Northbrook
CT - Hartford
RI - Woonsocket
AZ - Scottsdale

View All 7 Locations

Full time

Posted 6 Days Ago

R0268183

Bring your heart to CVS Health. Every one of us at CVS Health shares a single, clear purpose: Bringing our heart to every moment of your health. This purpose guides our commitment to deliver enhanced human-centric health care for a rapidly changing world. Anchored in our brand — with heart at its center — our purpose sends a personal message that how we deliver our services is just as important as what we deliver.

Our Heart At Work Behaviors™ support this purpose. We want everyone who works at CVS Health to feel empowered by the role they play in transforming our culture and accelerating our ability to innovate and deliver solutions to make health care more personal, convenient and affordable.

About us

CVS Health®

Our Work Experience is the combination of everything that's unique about us: our culture, our core values, our company meetings, our commitment to sustainability, our recognition programs, but most importantly, it's our people. Our

Read More

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?

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