Workgroup Updates

OHDSI Community Call
April 19, 2022 • 11 am ET

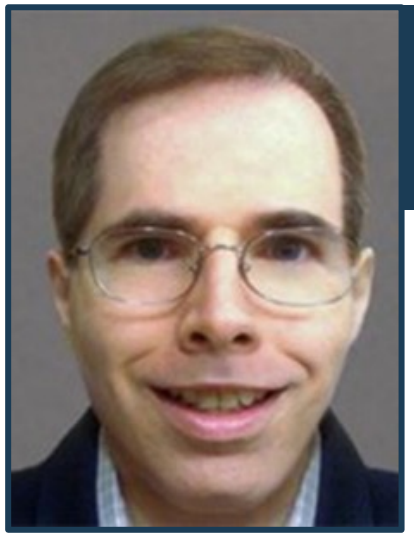
Upcoming OHDSI Community Calls

Date	Topic
April 26	Open-Source Community
May 3	DARWIN EU
May 10	Mother's Day-Themed Breakouts
May 17	OHDSI Debates
May 24	Open Studies
May 31	Workgroup OKRs

Upcoming OHDSI Community Calls

Date	Topic
April 26	Open-Source Community
May 3	DARWIN EU
May 10	Mother's Day-Themed Breakouts
May 17	OHDSI Debates
May 24	Open Studies
May 31	Workgroup OKRs

April 26 Community Call: Open Source Community



Panel Discussion Review

Lee Evans
Owner •
LTS Computing LLC

Keynote Summation

Martijn Schuemie
Research Fellow, Epidemiology Analytics •
Janssen Research and Development

State Of Open-Source Community

Paul Nagy
Associate Professor •
Johns Hopkins School of Medicine

State Of Open-Source Community

Adam Black
Data Sciences •
Odysseus Data Services

Three Stages of The Journey

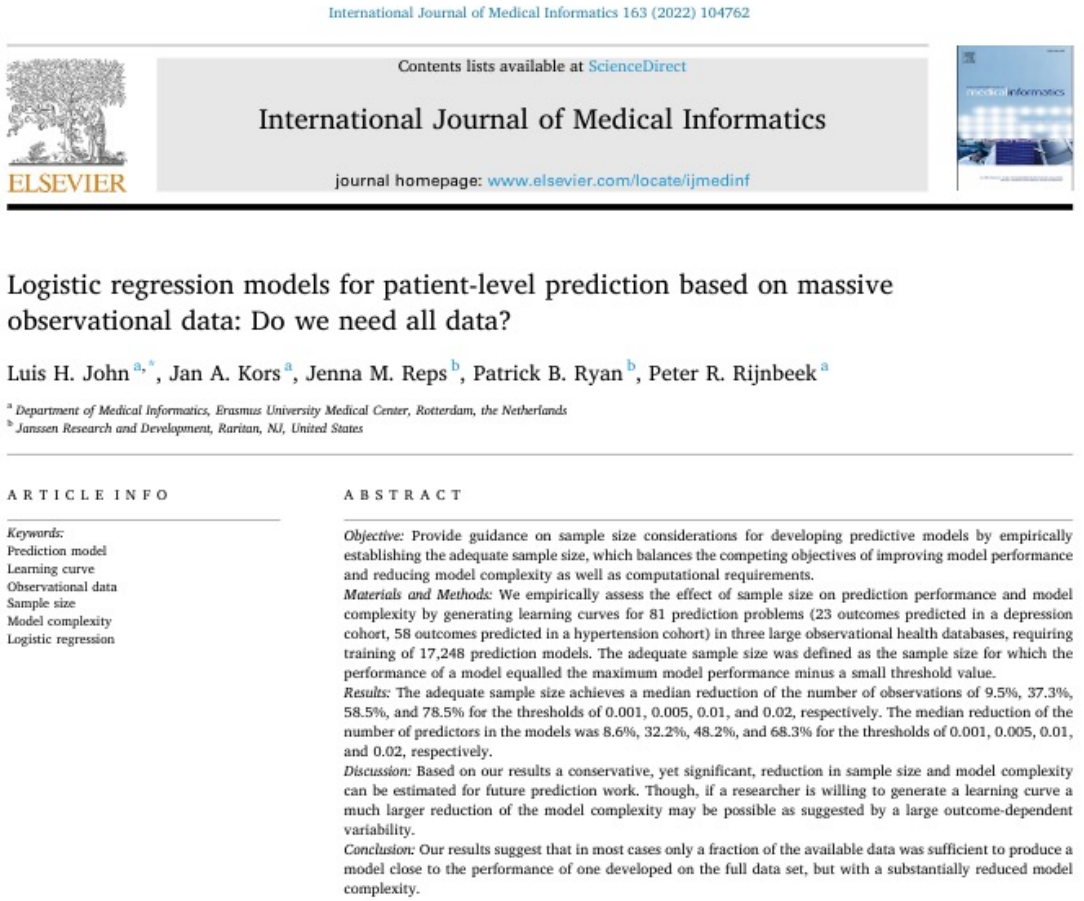
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Luis John, Jan Kors, Jenna Reps, Patrick Ryan, and Peter Rijnbeek** on the publication of "**Logistic regression models for patient-level prediction based on massive observational data: Do we need all data?**" in the International Journal of Medical Informatics.



International Journal of Medical Informatics 163 (2022) 104762

Contents lists available at ScienceDirect

International Journal of Medical Informatics

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

Logistic regression models for patient-level prediction based on massive observational data: Do we need all data?

Luis H. John [a,*], Jan A. Kors [a], Jenna M. Reps [b], Patrick B. Ryan [b], Peter R. Rijnbeek [a]

[a] Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands
[b] Janssen Research and Development, Raritan, NJ, United States

ARTICLE INFO

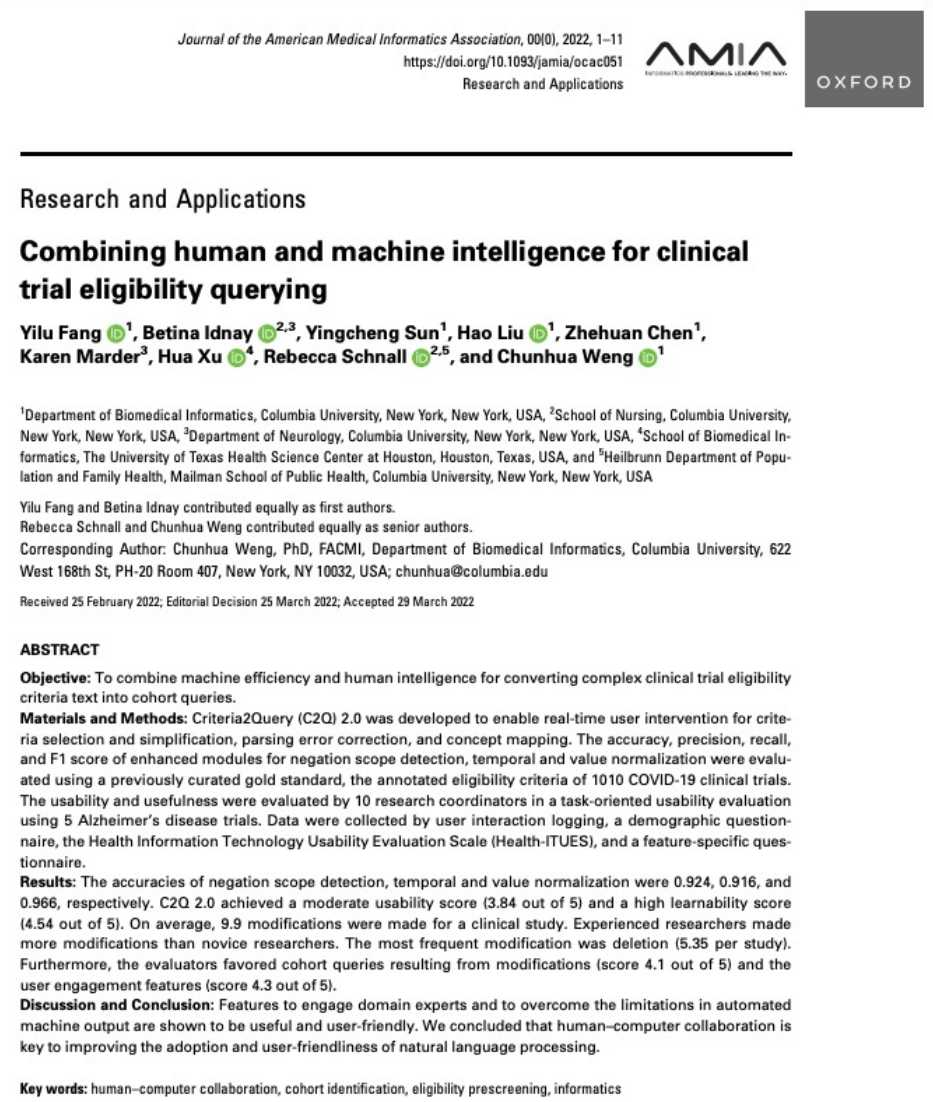
Keywords:
Prediction model
Learning curve
Observational data
Sample size
Model complexity
Logistic regression

ABSTRACT

Objective: Provide guidance on sample size considerations for developing predictive models by empirically establishing the adequate sample size, which balances the competing objectives of improving model performance and reducing model complexity as well as computational requirements.
Materials and Methods: We empirically assess the effect of sample size on prediction performance and model complexity by generating learning curves for 81 prediction problems (23 outcomes predicted in a depression cohort, 58 outcomes predicted in a hypertension cohort) in three large observational health databases, requiring training of 17,248 prediction models. The adequate sample size was defined as the sample size for which the performance of a model equalled the maximum model performance minus a small threshold value.
Results: The adequate sample size achieves a median reduction of the number of observations of 9.5%, 37.3%, 58.5%, and 78.5% for the thresholds of 0.001, 0.005, 0.01, and 0.02, respectively. The median reduction of the number of predictors in the models was 8.6%, 32.2%, 48.2%, and 68.3% for the thresholds of 0.001, 0.005, 0.01, and 0.02, respectively.
Discussion: Based on our results a conservative, yet significant, reduction in sample size and model complexity can be estimated for future prediction work. Though, if a researcher is willing to generate a learning curve a much larger reduction of the model complexity may be possible as suggested by a large outcome-dependent variability.
Conclusion: Our results suggest that in most cases only a fraction of the available data was sufficient to produce a model close to the performance of one developed on the full data set, but with a substantially reduced model complexity.

OHDSI Shoutouts!

Congratulations to the team of **Yilu Fang, Betina Idnay, Yingcheng Sun, Hao Liu, Zhehuan Chen, Karen Marder, Hua Xu, Rebecca Schnall, and Chunhua Weng** on the publication of "**Combining human and machine intelligence for clinical trial eligibility querying**" in JAMIA.



Journal of the American Medical Informatics Association, 00(0), 2022, 1–11
https://doi.org/10.1093/jamia/ocac051
Research and Applications

Research and Applications

Combining human and machine intelligence for clinical trial eligibility querying

Yilu Fang [1], Betina Idnay [2,3], Yingcheng Sun [1], Hao Liu [1], Zhehuan Chen [1], Karen Marder [3], Hua Xu [4], Rebecca Schnall [2,5], and Chunhua Weng [1]

[1]Department of Biomedical Informatics, Columbia University, New York, New York, USA, [2]School of Nursing, Columbia University, New York, New York, USA, [3]Department of Neurology, Columbia University, New York, New York, USA, [4]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, USA, and [5]Heilbrunn Department of Population and Family Health, Mailman School of Public Health, Columbia University, New York, New York, USA

Yilu Fang and Betina Idnay contributed equally as first authors.
Rebecca Schnall and Chunhua Weng contributed equally as senior authors.
Corresponding Author: Chunhua Weng, PhD, FACMI, Department of Biomedical Informatics, Columbia University, 622 West 168th St, PH-20 Room 407, New York, NY 10032, USA; chunhua@columbia.edu

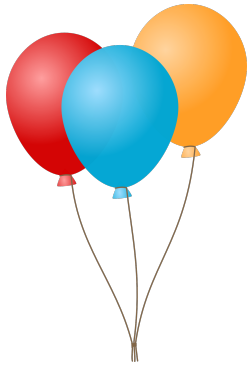
Received 25 February 2022; Editorial Decision 25 March 2022; Accepted 29 March 2022

ABSTRACT

Objective: To combine machine efficiency and human intelligence for converting complex clinical trial eligibility criteria text into cohort queries.
Materials and Methods: Criteria2Query (C2Q) 2.0 was developed to enable real-time user intervention for criteria selection and simplification, parsing error correction, and concept mapping. The accuracy, precision, recall, and F1 score of enhanced modules for negation scope detection, temporal and value normalization were evaluated using a previously curated gold standard, the annotated eligibility criteria of 1010 COVID-19 clinical trials. The usability and usefulness were evaluated by 10 research coordinators in a task-oriented usability evaluation using 5 Alzheimer's disease trials. Data were collected by user interaction logging, a demographic questionnaire, the Health Information Technology Usability Evaluation Scale (Health-ITUES), and a feature-specific questionnaire.
Results: The accuracies of negation scope detection, temporal and value normalization were 0.924, 0.916, and 0.966, respectively. C2Q 2.0 achieved a moderate usability score (3.84 out of 5) and a high learnability score (4.54 out of 5). On average, 9.9 modifications were made for a clinical study. Experienced researchers made more modifications than novice researchers. The most frequent modification was deletion (5.35 per study). Furthermore, the evaluators favored cohort queries resulting from modifications (score 4.1 out of 5) and the user engagement features (score 4.3 out of 5).
Discussion and Conclusion: Features to engage domain experts and to overcome the limitations in automated machine output are shown to be useful and user-friendly. We concluded that human–computer collaboration is key to improving the adoption and user-friendliness of natural language processing.

Key words: human–computer collaboration, cohort identification, eligibility prescreening, informatics

OHDSI Shoutouts!



Any shoutouts from the community? Please share and help promote and celebrate OHDSI work!

Have a study published? Please send to sachson@ohdsi.org so we can share during this call and on our social channels.
Let's work together to promote the collaborative work happening in OHDSI!

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

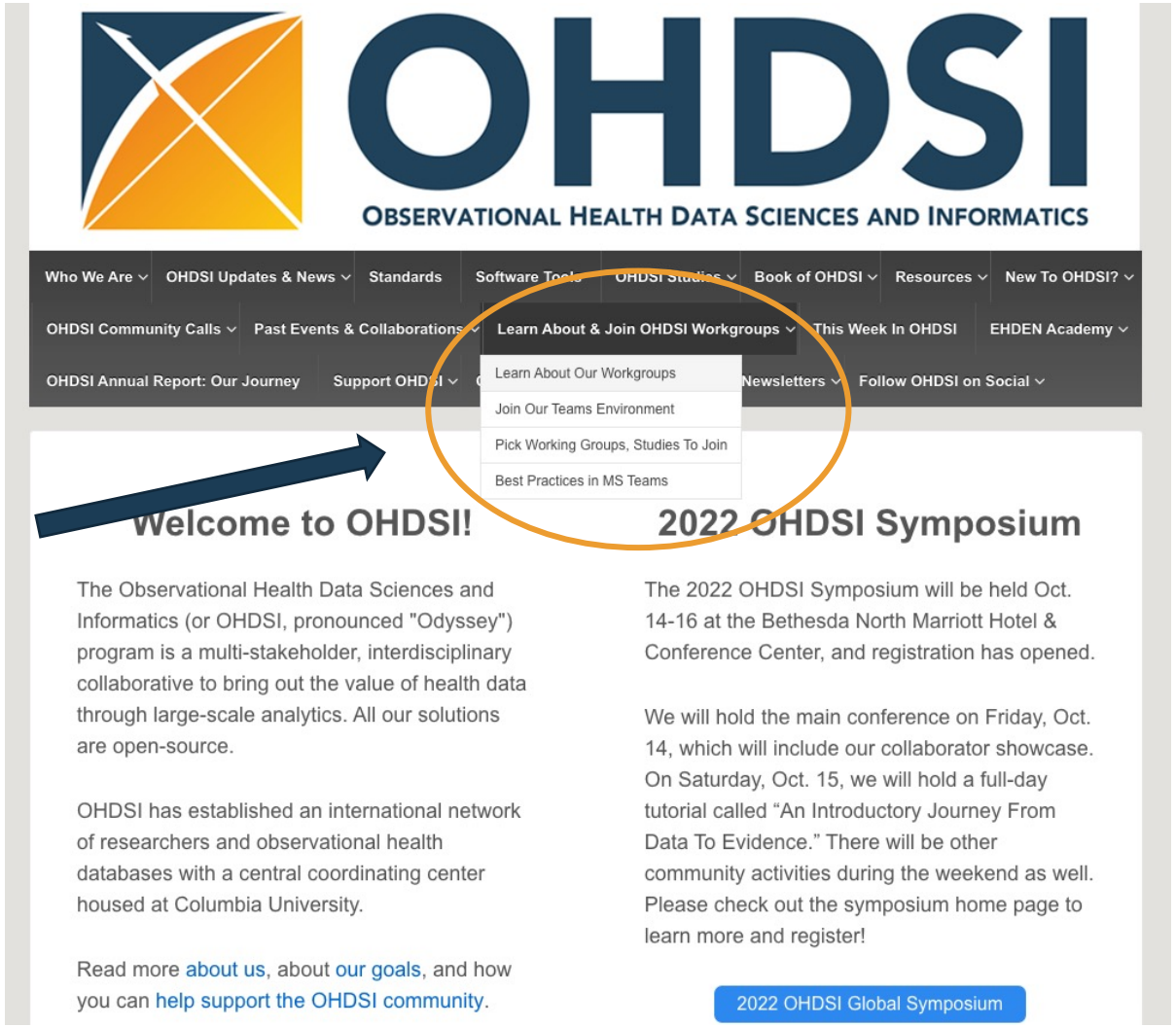
Where Are We Going?

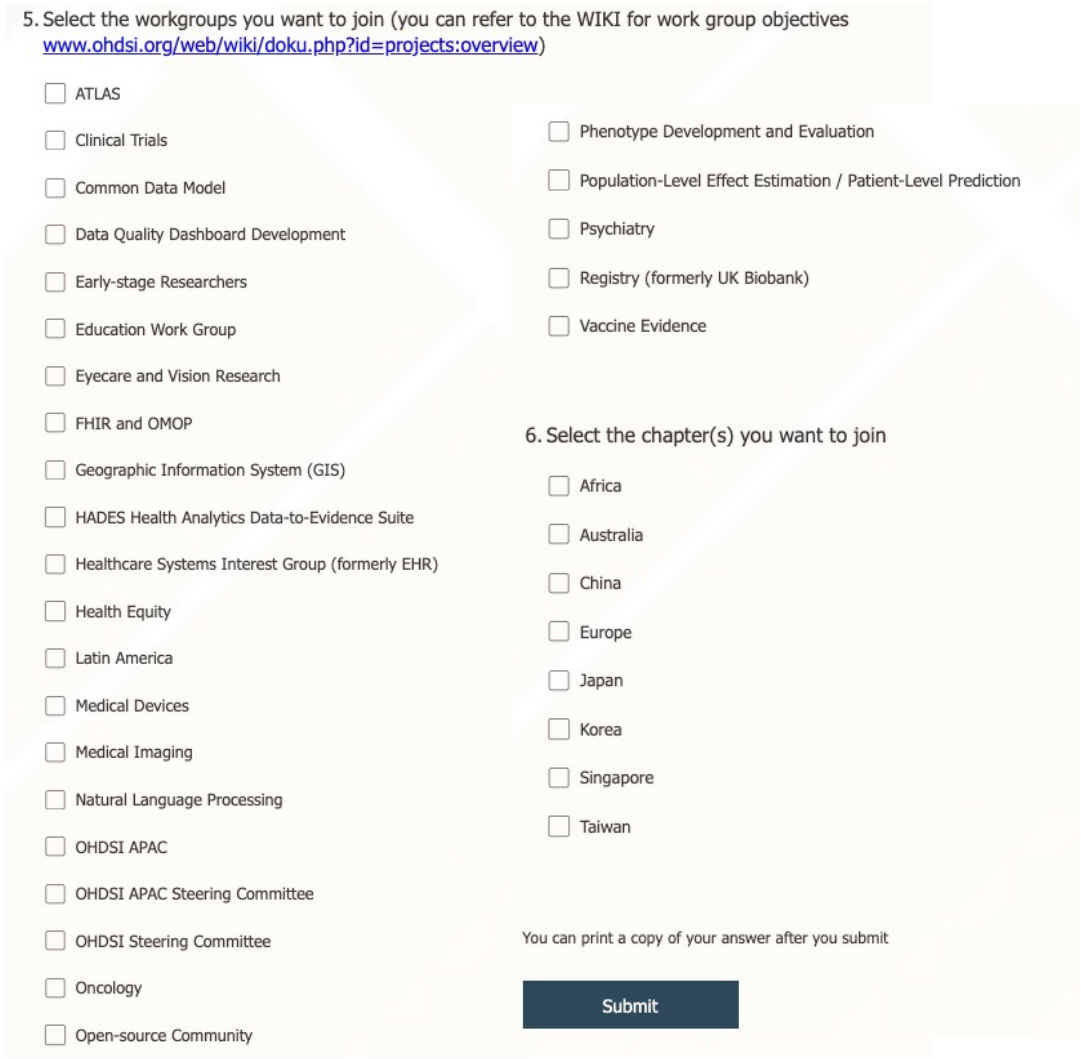
Upcoming Workgroup Calls

www.ohdsi.org/upcoming-working-group-calls

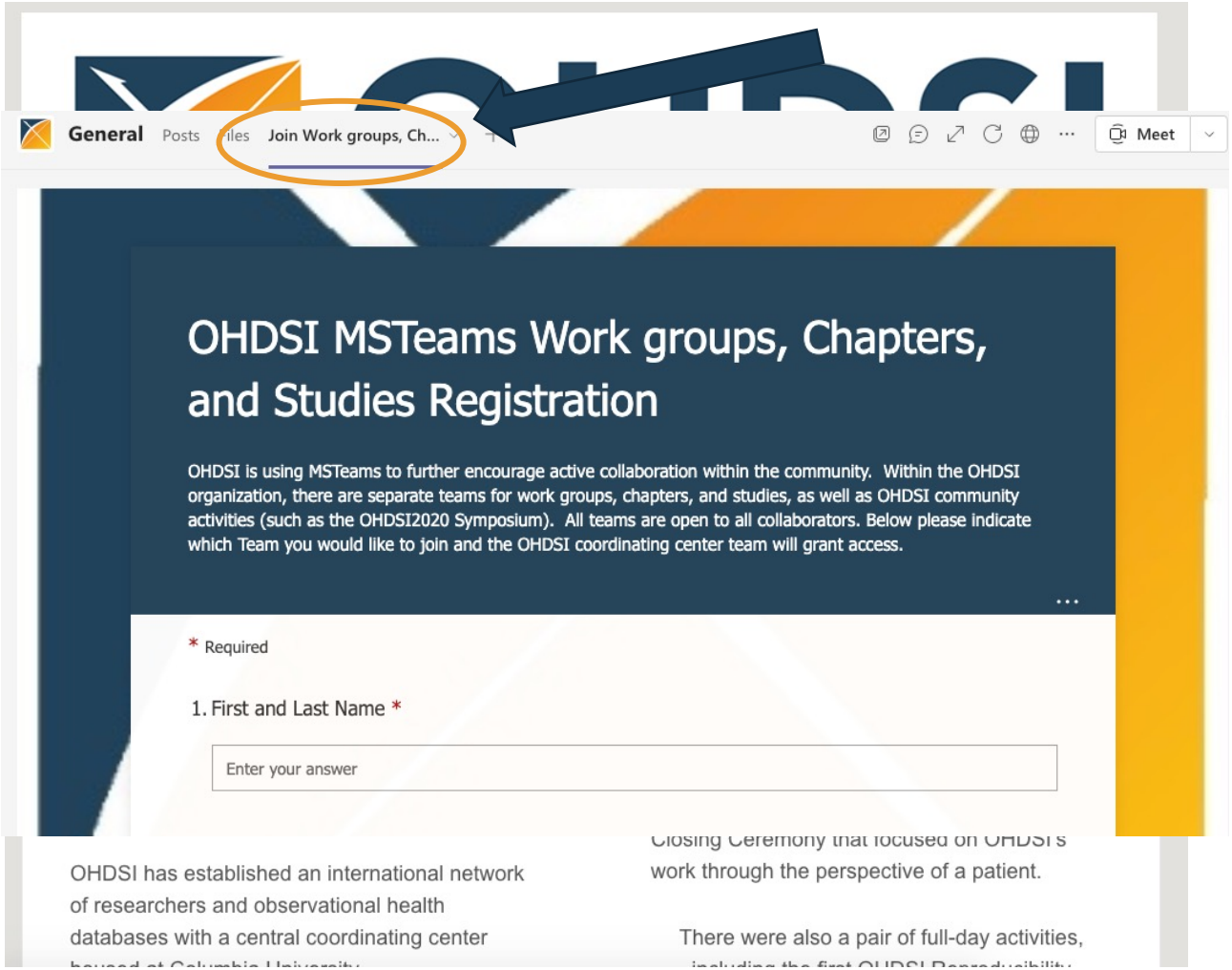
Date	Time (ET)	Meeting
Tuesday	1 pm	Common Data Model
Wednesday	7 am	Medical Imaging
Wednesday	9 am	Africa Chapter
Wednesday	9 am	FHIR and OMOP Data Model Harmonization Subgroup (ZOOM)
Wednesday	11 am	Open-Source Community
Wednesday	12 pm	Health Equity Journal Club
Wednesday	12 pm	FHIR and OMOP Terminologies Subgroup (ZOOM)
Thursday	10 am	Data Quality Dashboard
Thursday	12 pm	HADES
Thursday	12 pm	FHIR and OMOP Oncology Subgroup
Thursday	1 pm	OMOP CDM Oncology Vocabulary Subgroup
Friday	9 am	Education
Friday	9 am	GIS – Geographic Information System General
Friday	10 am	Phenotype Development and Evaluation
Friday	11:30 am	Steering Group
Monday	10 am	Healthcare Systems Interest Group
Tuesday	9 am	OMOP CDM Oncology Genomic Subgroup

Get Access To Different Teams/WGs/Chapters





Get Access To Different Teams/WGs/Chapters



5. Select the workgroups you want to join (you can refer to the WIKI for work group objectives www.ohdsi.org/web/wiki/doku.php?id=projects:overview)

- ☐ ATLAS
- ☐ Clinical Trials
- ☐ Common Data Model
- ☐ Data Quality Dashboard Development
- ☐ Early-stage Researchers
- ☐ Education Work Group
- ☐ Eyecare and Vision Research
- ☐ FHIR and OMOP
- ☐ Geographic Information System (GIS)
- ☐ HADES Health Analytics Data-to-Evidence Suite
- ☐ Healthcare Systems Interest Group (formerly EHR)
- ☐ Health Equity
- ☐ Latin America
- ☐ Medical Devices
- ☐ Medical Imaging
- ☐ Natural Language Processing
- ☐ OHDSI APAC
- ☐ OHDSI APAC Steering Committee
- ☐ OHDSI Steering Committee
- ☐ Oncology
- ☐ Open-source Community
- ☐ Phenotype Development and Evaluation
- ☐ Population-Level Effect Estimation / Patient-Level Prediction
- ☐ Psychiatry
- ☐ Registry (formerly UK Biobank)
- ☐ Vaccine Evidence

6. Select the chapter(s) you want to join

- ☐ Africa
- ☐ Australia
- ☐ China
- ☐ Europe
- ☐ Japan
- ☐ Korea
- ☐ Singapore
- ☐ Taiwan

You can print a copy of your answer after you submit

Submit

OHDSI Dev Con

April 22, 2022 (8 am – 12 pm)

The Open-Source Community is hosting the first **Dev Con** as a way of accepting and mentoring new contributors to our environment. We are planning multiple workshops, talks and a panel discussion to both welcome and engage both current and future developers within OHDSI.

Don't miss this opportunity! Use the link at the bottom to register!

Time	Topic
8 am	Open-Source Workshops
10 am	State of the OHDSI Community (Paul Nagy, Adam Black)
10:20 am	Keynote – Grand Vision for OHDSI Software Ecosystem (Martijn Schuemie)
11 am	Industry Panel Discussion (How Do/Should We Connect It All Together?)

bit.ly/OHDSIDev22

Are You Interested In …

- participating with an OHDSI project team?
- seeing 'under the hood' of the OHDSI engine?
- being mentored by professional developers?

Use This Link To Register Today!

DevCon Agenda

Time (ET)	Track 1	Track 2
8 am	ATLAS **(Anthony Sena)**	HADES Introduction **(Adam Black)**
8:30 am	WebAPI **(Anthony Sena)**	CohortDiagnostics **(James Gilbert)**
9 am	White Rabbit/Rabbit In A Hat **(Maxim Moinat)**	Patient-Level Prediction **(Jenna Reps)**
9:30 am	Data Quality Dashboard **(Clair Blacketer)**	Cyclops **(Marc Suchard)**
10 am	State of OHDSI Development **(Adam Black and Paul Nagy)**	
10:20 am	Keynote **(Martijn Schuemie)**	
11 am	Panel Discussion (Putting The Pieces Together) **Lee Evans** - Broadsea (OHDSI) **Cory Stevenson** - OHDSI on Azure (Microsoft) **James Wiggins** – OHDSI on AWS (Amazon) **Vivian Neilley** - OHDSI on Google Cloud	

2022 OHDSI Symposium

Registration is OPEN for #OHDSI2022!

The 2022 OHDSI Symposium will be held Oct. 14-16 at the Bethesda North Marriott Hotel & Conference Center.





An Introductory Journey From Data To Evidence

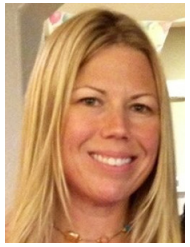
OHDSI2022 Tutorial • Saturday, Oct. 15 • Bethesda, Md.

The OHDSI Journey: Where Are We Going?

Patrick Ryan

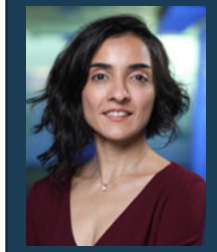
OMOP Common Data Model and Vocabulary

Clair Blacketer



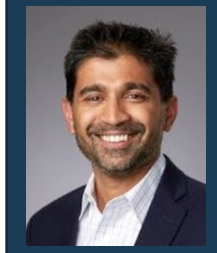
ETL – A Source Database Into OMOP CDM

Melanie Philofsky



Creating Cohort Definitions

Asieh Golozar



Phenotype Evaluations

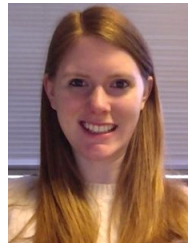
Gowtham Rao

Characterization

Kristin Kostka

Estimation

Martijn Schuemie



Prediction

Jenna Reps

The OHDSI Journey: Where Do We Go From Here?

George Hripcsak

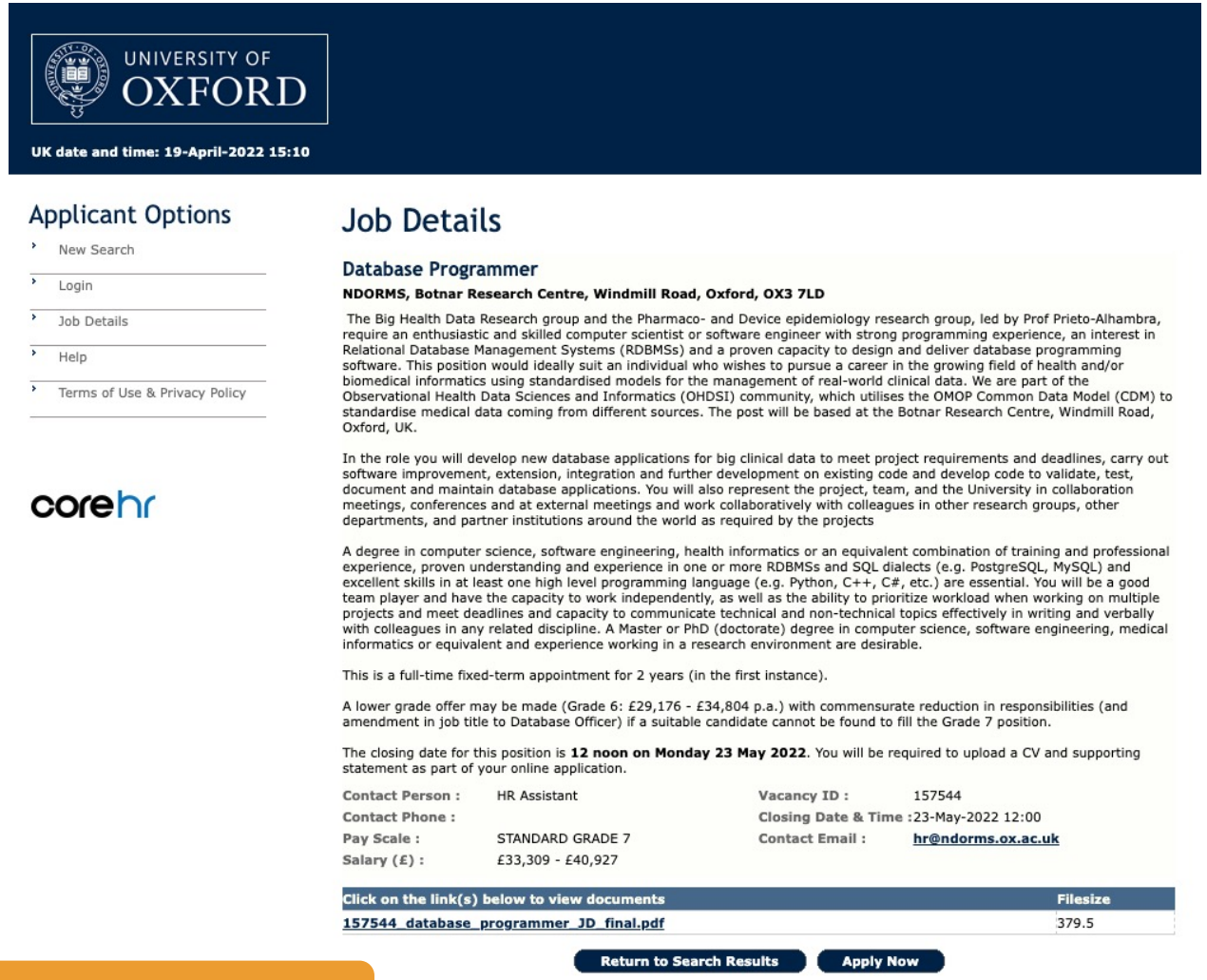
2022 European Symposium

www.ohdsi-europe.org/symposium-2022

Opening at Oxford

Dani Prieto-Alhambra's team at Oxford is recruiting a Database Programmer to join the team.

This position will contribute to the standardization and curation of large real world data from the UK and collaborate with the OHDSI, EHDEN and OPTIMA Oncology teams.



UNIVERSITY OF OXFORD

UK date and time: 19-April-2022 15:10

Applicant Options

- New Search
- Login
- Job Details
- Help
- Terms of Use & Privacy Policy

corehr

Job Details

Database Programmer

NDORMS, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD

The Big Health Data Research group and the Pharmaco- and Device epidemiology research group, led by Prof Prieto-Alhambra, require an enthusiastic and skilled computer scientist or software engineer with strong programming experience, an interest in Relational Database Management Systems (RDBMSs) and a proven capacity to design and deliver database programming software. This position would ideally suit an individual who wishes to pursue a career in the growing field of health and/or biomedical informatics using standardised models for the management of real-world clinical data. We are part of the Observational Health Data Sciences and Informatics (OHDSI) community, which utilises the OMOP Common Data Model (CDM) to standardise medical data coming from different sources. The post will be based at the Botnar Research Centre, Windmill Road, Oxford, UK.

In the role you will develop new database applications for big clinical data to meet project requirements and deadlines, carry out software improvement, extension, integration and further development on existing code and develop code to validate, test, document and maintain database applications. You will also represent the project, team, and the University in collaboration meetings, conferences and at external meetings and work collaboratively with colleagues in other research groups, other departments, and partner institutions around the world as required by the projects

A degree in computer science, software engineering, health informatics or an equivalent combination of training and professional experience, proven understanding and experience in one or more RDBMSs and SQL dialects (e.g. PostgreSQL, MySQL) and excellent skills in at least one high level programming language (e.g. Python, C++, C#, etc.) are essential. You will be a good team player and have the capacity to work independently, as well as the ability to prioritize workload when working on multiple projects and meet deadlines and capacity to communicate technical and non-technical topics effectively in writing and verbally with colleagues in any related discipline. A Master or PhD (doctorate) degree in computer science, software engineering, medical informatics or equivalent and experience working in a research environment are desirable.

This is a full-time fixed-term appointment for 2 years (in the first instance).

A lower grade offer may be made (Grade 6: £29,176 - £34,804 p.a.) with commensurate reduction in responsibilities (and amendment in job title to Database Officer) if a suitable candidate cannot be found to fill the Grade 7 position.

The closing date for this position is **12 noon on Monday 23 May 2022**. You will be required to upload a CV and supporting statement as part of your online application.

Contact Person :	HR Assistant	**Vacancy ID :**	157544
Contact Phone :		**Closing Date & Time :**	23-May-2022 12:00
Pay Scale :	STANDARD GRADE 7	**Contact Email :**	hr@ndorms.ox.ac.uk
Salary (£) :	£33,309 - £40,927		

Click on the link(s) below to view documents	Filesize
157544_database_programmer_JD_final.pdf	379.5

Return to Search Results | Apply Now

Deadline: May 23, 2022

Next CBER Best Seminar

Topic CBER BEST Seminar Series - Addressing Selection and Confounding Bias in Test-Negative Study Designs for Flu and COVID-19 Monitoring

Description: The test-negative design (TND) has become a standard approach to evaluate vaccine effectiveness against the risk of acquiring infectious diseases such as Influenza, Rotavirus, Dengue fever and more recently COVID-19 in real world settings. Despite the TND's potential to reduce unobserved differences in healthcare seeking behavior (HSB) between vaccinated and unvaccinated subjects, substantial variability in unobserved HSB may remain among study participants. As latent HSB is likely also a strong predictor of selection into the TND sample, confounding bias of the vaccine's causal effect by latent HSB may be induced by collider stratification bias resulting from the TND.

Speakers

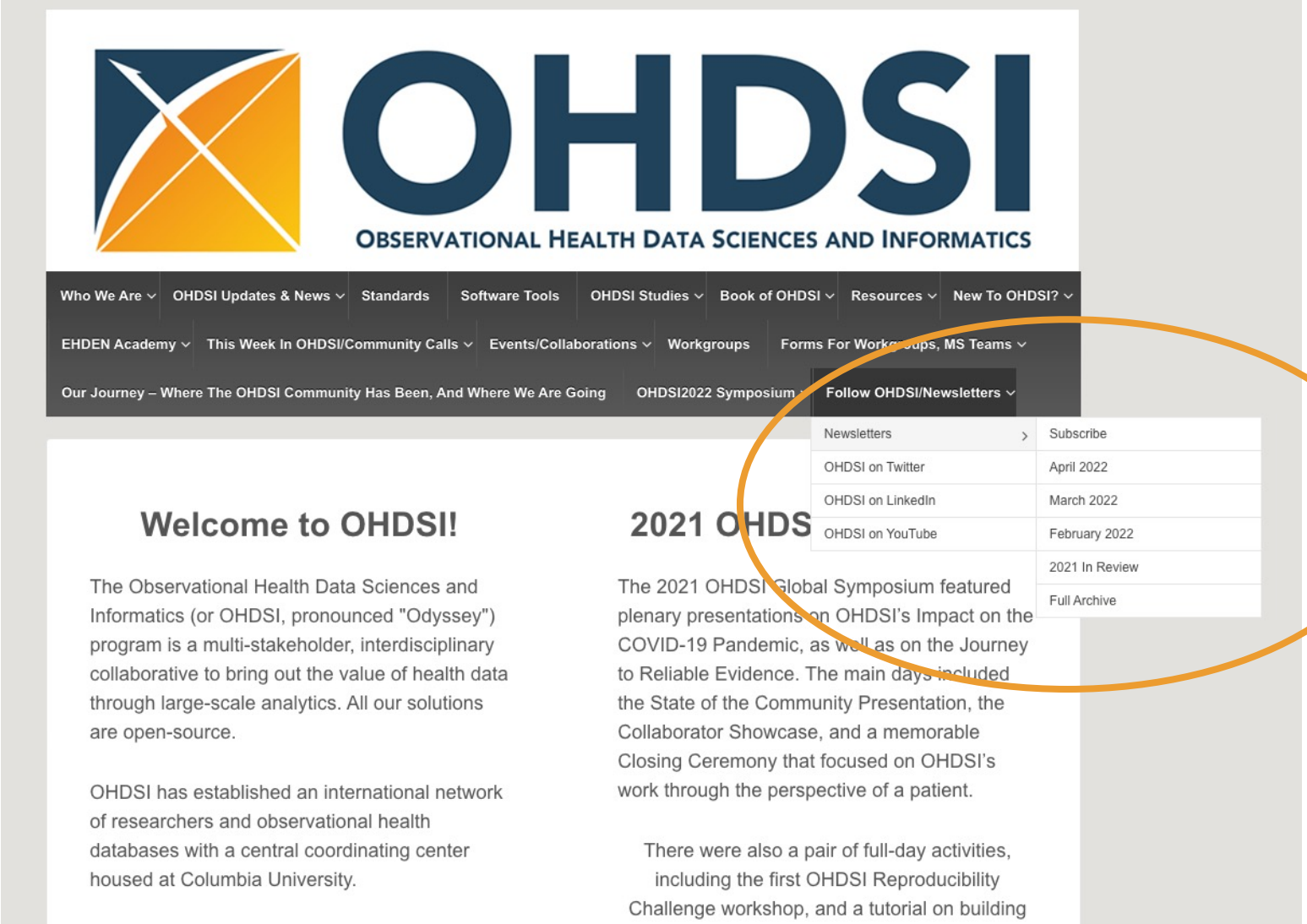
Dr. Eric Tchetgen Tchetgen

Luddy Family President's Distinguished Professor @Wharton School of the University of Pennsylvania

Eric J. Tchetgen Tchetgen is the Luddy Family President's Distinguished Professor at the Wharton School of the University of Pennsylvania. Professor Tchetgen Tchetgen comes to the University of Pennsylvania from Harvard University, where he has served since 2008 as Professor of Biostatistics and Epidemiologic Methods with joint appointments in the departments of Biostatistics and Epidemiology at the T.H. Chan School of Public Health. He researches infectious diseases, including HIV/AIDS, and the role of genetic and social factors in the patterns, causes, and effects of public health. Professor Tchetgen Tchetgen has received grants from the National Institutes of Health and the Centers for Disease Control. He completed his Ph.D. in Biostatistics at Harvard University in 2006 under the supervision of Professor James M. Robins. He received his B.S. in Electrical Engineering from Yale University in 1999.

Wed., April 27, 11 am ET

Latest OHDSI Newsletter Is Out



Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

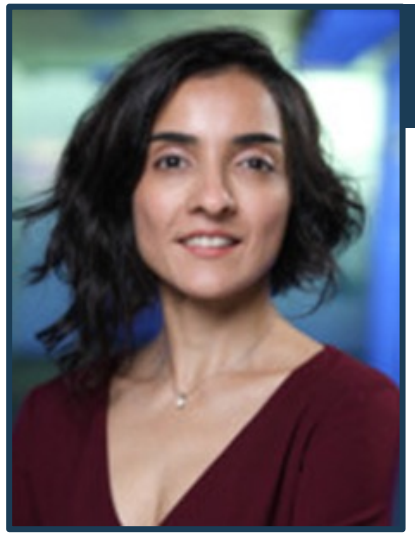
April 19 Community Call: Workgroup Updates

Eye Care and Vision Research

Sally Baxter
Assistant Professor, Ophthalmology • UCSD

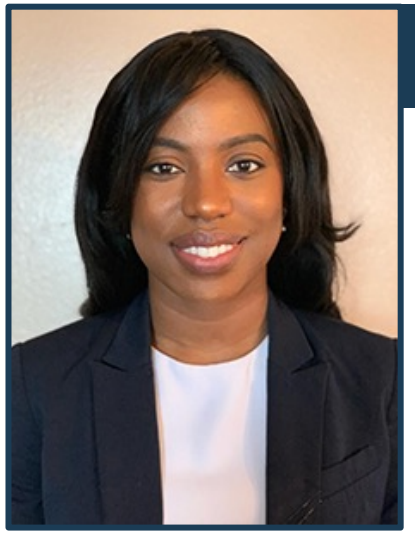
FHIR & OMOP

Christian Reich
Vice President, RWE Systems • IQVIA



Oncology

Asieh Golozar
VP, Global Head of Data Science • Odysseus Data Services



Steering Group

Jody-Ann McLeggon
Program Manager • Columbia University