Reproducibility

OHDSI Community Call
March 29, 2022 • 11 am ET
# April OHDSI Community Calls

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 5</td>
<td>Name That Result</td>
</tr>
<tr>
<td>April 12</td>
<td>OHDSI Coordinating Center</td>
</tr>
<tr>
<td>April 19</td>
<td>DARWIN EU</td>
</tr>
<tr>
<td>April 26</td>
<td>Open-Source Community</td>
</tr>
<tr>
<td>Date</td>
<td>Topic</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>April 5</td>
<td>Name That Result</td>
</tr>
<tr>
<td>April 12</td>
<td>OHDSI Coordinating Center</td>
</tr>
<tr>
<td>April 19</td>
<td>DARWIN EU</td>
</tr>
<tr>
<td>April 26</td>
<td>Open-Source Community</td>
</tr>
</tbody>
</table>
Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?
Congratulations to the team of Yuan Lu, Mui Van Zandt, Yun Liu, Jing Li, Xialin Wang, Yong Chen, Zhengfeng Chen, Jaehyeong Cho, Sreemane Raaj Dorajoo, Mengling Feng, Min-Huei Hsu, Jason C. Hsu, Usman Iqbal, Jitendra Jonnagaddala, Yu-Chuan Li, Siaw-Teng Liaw, Hong-Seok Lim, Kee Yuan Ngiam, Phung-Anh Nguyen, Rae Woong Park, Nicole Pratt, Christian Reich, Sang Youl Rhee, Selva Muthu Kumaran Sathappan, Seo Jeong Shin, Hui Xing Tan, Seng Chan You, Xin Zhang, Harlan M. Krumholz, Marc A. Suchard, and Hua Xu on the publication of Analysis of Dual Combination Therapies Used in Treatment of Hypertension in a Multinational Cohort in JAMA Network Open.
OHDSI Shoutouts!

Congratulations to both Aki Nishimura and Marc Suchard on the publication of Prior-preconditioned conjugate gradient method for accelerated Gibbs sampling in ‘large n & large p’ Bayesian sparse regression in the Journal of the American Statistical Association.

Prior-preconditioned conjugate gradient method for accelerated Gibbs sampling in ‘large n & large p’ Bayesian sparse regression

Akihiko Nishimura\(^a\) and Marc A. Suchard\(^b\)

\(^a\) Department of Biostatistics, Johns Hopkins University; \(^b\) Department of Biostatistics, Biostatistics, and Human Genetics, University of California - Los Angeles

ABSTRACT

In a modern observational study based on healthcare databases, the number of observations and of predictors typically range in the order of \(10^5 \sim 10^7\) and of \(10^8 \sim 10^9\). Despite the large sample size, data rarely provide sufficient information to reliably estimate such a large number of parameters. Sparse regression techniques provide potential solutions, one notable approach being the Bayesian method based on shrinkage priors. In the ‘large n & large p’ setting, however, the required posterior computation encounters a bottleneck at repeated sampling from a high-dimensional Gaussian distribution, whose precision matrix \(\Omega\) is expensive to compute and factorize. In this article, we present a novel algorithm to speed up this bottleneck based on the following observation: we can cheaply generate a random vector \(\mathbf{b}\) such that the solution to the linear system \(\mathbf{b}^T \mathbf{b} \approx \mathbf{b}\) has the desired Gaussian distribution. We can then solve the linear system by the conjugate gradient (CG) algorithm through matrix-vector multiplications by \(\mathbf{b}\); this involves no explicit factorization or calculation of \(\mathbf{b}\) itself. Rapid convergence of CG in this context is guaranteed by the theory of prior-preconditioning we develop. We apply our algorithm to a clinically relevant large-scale observational study with \(n = 72,489\) patients and \(p = 22,175\) clinical covariates, designed to assess the relative risk of adverse events from two alternative blood anti-coagulants. Our algorithm demonstrates an order of magnitude speed-up in posterior inference, in our case cutting the computation time from two weeks to less than a day.

ARTICLE HISTORY

Received 11 February 2020
Revised 3 March 2022
Accepted 18 March 2022

KEYWORDS

Big Data, Conjugate gradient, Markov chain Monte Carlo, numerical linear algebra, sparse matrix, variable selection
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

<table>
<thead>
<tr>
<th>Date</th>
<th>Time (ET)</th>
<th>Meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday</td>
<td>10 am</td>
<td>FHIR and OMOP Digital Quality Measurements Subgroup (ZOOM)</td>
</tr>
<tr>
<td>Wednesday</td>
<td>11 am</td>
<td>GIS – Geographic Information Systems Development</td>
</tr>
<tr>
<td>Thursday</td>
<td>12 pm</td>
<td>FHIR and OMOP Oncology Subgroup</td>
</tr>
<tr>
<td>Thursday</td>
<td>1 pm</td>
<td>OMOP CDM Oncology Vocabulary Subgroup</td>
</tr>
<tr>
<td>Thursday</td>
<td>6 pm</td>
<td>FHIR and OMOP Terminologies Subgroup (ZOOM)</td>
</tr>
<tr>
<td>Friday</td>
<td>10:30 am</td>
<td>Clinical Trials</td>
</tr>
<tr>
<td>Tuesday</td>
<td>10 am</td>
<td>Common Data Model</td>
</tr>
</tbody>
</table>

[www.ohdsi.org/upcoming-working-group-calls](http://www.ohdsi.org/upcoming-working-group-calls)
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)

   - ATLAS
   - Clinical Trials
   - Common Data Model
   - Data Quality Dashboard Development
   - Early-stage Researchers
   - Education Work Group
   - 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

6. Select the chapter(s) you want to join
   - Africa
   - Australia
   - China
   - Europe
   - Japan
   - Korea
   - Singapore
   - Taiwan

7. Select the studies you want to join
   - HIEA-Health Equity Research Assessment
   - PIONEER for Prostate Cancer (study-a-thin-ended)
   - SCYLLA (SARs-Cov-2 Large-scale, Longitudinal Analysis)
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)

- ATLAS
- Clinical Trials
- Common Data Model
- Data Quality Dashboard Development
- Early-stage Researchers
- Education Work Group
- FHRE 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

6. Select the chapter(s) you want to join
- Africa
- Australia
- China
- Europe
- Japan
- Korea
- Singapore
- Taiwan

7. Select the studies you want to join
- HERA-Health Equity Research Assessment
- PIONEER for Prostate Cancer (study-a-then-ended)
- SCYLLA (SARS-CoV-2 Large-scale Longitudinal Analyses)
Registration is OPEN for #OHDSI2022!

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

The main conference at the 2022 OHDSI Symposium returns Friday, Oct. 14, at the Bethesda North Marriott Hotel & Conference Center. The conference will include the collaborator showcase; the submission deadline for that is Friday, June 24.

Oct. 15: Full-Day Tutorial

An Introductory Journey From Data To Evidence

This tutorial is meant to introduce newer members of the community to steps along the journey from data to evidence using the OMOP Common Data Model, OHDSI tools and scientific best practices.
Saturday: Full-Day Tutorial

An Introductory Journey From Data To Evidence

This tutorial is meant to introduce newer members of the community to steps along the journey from data to evidence using the OMOP Common Data Model, OHDSI tools and scientific best practices.

Agenda

8:30 am • Overview of the OHDSI Journey: where are we going? — Patrick Ryan
9 am • OMOP Common Data Model and vocabulary — Clair Blacketer
9:50 am • Energy break
10 am • ETL a source database into OMOP CDM — TBA
10:50 am • Energy break
11 am • Creating cohort definitions — Asieh Golozar
11:50 am • Lunch break
12:30 pm • Phenotype evaluation — Gowtham Rao
1:20 pm • Energy break
1:30 pm • Characterization — Kristin Kostka
2:20 pm • Energy break
2:30 pm • Estimation — Martijn Schuemie
3:20 pm • Energy break
3:30 pm • Prediction — Jenna Reps
4:20 pm • Recap of the OHDSI Journey, where do we go from here? — George Hripcsak
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.

Read more about us, about our goals, and how you can help support the OHDSI community.

symposium@ohdsi.org
OHDSI Vocabulary Journey

The March 22 OHDSI Community Call provided an in-depth look at the OHDSI vocabulary, from how it is developed, to how it can be utilized, and where it should grow from here. Three leaders from the vocabulary work group joined to present a trio of topics for this session:

“A peek into the OHDSI vocabulary engine room”
— Michael Kallfelz, Physician Executive & Odysseus Data Services

“Fun things you can learn with the OHDSI standardized vocabularies”
— Patrick Ryan, Vice President, Observational Health Data Analytics • Janssen Research & Development; Adjunct Assistant Professor • Columbia University

“Time for reflection • Where are we? Where should we be?”
— Christian Reich, Vice President, RWE Systems • IQVIA

The slides from this presentation are available here.

Both Reich and Anna Astropoulos led the Standardized Vocabularies chapter of the Book of OHDSI, which was first released in 2019 and has been updated online since. The introductory paragraph for that chapter is below. You can read the whole chapter here.

The OMOP Standardized Vocabularies, often referred to simply as “the Vocabulary”, are a foundational part of the OHDSI research network, and an integral part of the Common Data Model (CDM). They allow standardization of methods, definitions and results by defining the content of the data, paving the way for true remote (behind the firewall) network research and analytics. Usually, finding and interpreting the content of observational healthcare data, whether it is structured data using coding schemes or laid down in free text, is passed all the way through to the researcher, who is faced with a myriad of different ways to describe clinical events. OHDSI requires harmonization not only to a standardized format, but also to a rigorous standard content.
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!

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

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!

bit.ly/OHDSIDev22
# DevCon Agenda

<table>
<thead>
<tr>
<th>Time (ET)</th>
<th>Track 1</th>
<th>Track 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 am</td>
<td>ATLAS <em>(Anthony Sena)</em></td>
<td>HADES Introduction <em>(Adam Black)</em></td>
</tr>
<tr>
<td>8:30 am</td>
<td>WebAPI <em>(Anthony Sena)</em></td>
<td>CohortDiagnostics <em>(James Gilbert)</em></td>
</tr>
<tr>
<td>9 am</td>
<td>White Rabbit/Rabbit In A Hat <em>(Maxim Moinat)</em></td>
<td>Patient-Level Prediction <em>(Jenna Reps)</em></td>
</tr>
<tr>
<td>9:30 am</td>
<td>Data Quality Dashboard <em>(Clair Blacketer)</em></td>
<td>Cyclops <em>(Marc Suchard)</em></td>
</tr>
<tr>
<td>10 am</td>
<td>State of OHDSI Development <em>(Adam Black and Paul Nagy)</em></td>
<td></td>
</tr>
<tr>
<td>10:20 am</td>
<td>Keynote <em>(Martijn Schuemie)</em></td>
<td></td>
</tr>
</tbody>
</table>
| 11 am     | Panel Discussion (Putting The Pieces Together)  
Lee Evans - Broadsea (OHDSI)  
Cory Stevenson - OHDSI on Azure (Microsoft)  
Paul Sexson - OHDSI In A Box (AWS)  
OHDSI on GCP | |
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 at 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
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?
March 29: Reproducibility

Anna Ostropolets
PhD Student • Columbia University

Topic: The OHDSI2021 Reproducibility Challenge

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

Topic: Developing Reproducible Studies

Asieh Golozar
Vice President, Global Head of Data Science • Odysseus Data Services, Inc.

Topic: The Reproducibility Service, via the OHDSI Center at the Roux Institute
Polina Talapova, MD