Our Journey
Where The OHDSI Community Has Been
And Where We Are Going
2023 edition
Publication was written and designed by Craig Sachson. Editorial assistance by Patrick Ryan, Paul Nagy, George Hripcsak, Martijn Schuemie, Marc Suchard, Jody-Ann McLeggon, Jenna Reps, Peter Rijnbeek, Clair Blacketer, Anna Ostropolets, Mui Van Zandt, and other members of the OHDSI community. Photography shared by the OHDSI community unless specifically credited next to image. Printed by ABGPrint. Thank you to all members of the OHDSI community for all you have done towards improving global healthcare.
# Join The Journey

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

## Table Of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Letter To The Community</td>
<td>2</td>
</tr>
<tr>
<td>II. Mission, Values, And What We Do</td>
<td>5</td>
</tr>
<tr>
<td>How OHDSI Works</td>
<td>7</td>
</tr>
<tr>
<td>Columbia University as Coordinating Center</td>
<td>8</td>
</tr>
<tr>
<td>III. Collaborators</td>
<td>9</td>
</tr>
<tr>
<td>Map of Collaborators</td>
<td>10</td>
</tr>
<tr>
<td>Organizations Involved with OHDSI</td>
<td>12</td>
</tr>
<tr>
<td>Testimonials</td>
<td>14</td>
</tr>
<tr>
<td>The Titan Awards</td>
<td>16</td>
</tr>
<tr>
<td>IV. Collaborative Activities</td>
<td>19</td>
</tr>
<tr>
<td>Workgroups</td>
<td>20</td>
</tr>
<tr>
<td>Regional Chapters and National Nodes</td>
<td>22</td>
</tr>
<tr>
<td>Community Calls</td>
<td>23</td>
</tr>
<tr>
<td>Phenotype Phebruary</td>
<td>26</td>
</tr>
<tr>
<td>Save Our Sisyphus (SOS) Challenge</td>
<td>28</td>
</tr>
<tr>
<td>The Book of OHDSI</td>
<td>30</td>
</tr>
<tr>
<td>OHDSI and Large Community Initiatives</td>
<td>32</td>
</tr>
<tr>
<td>OHDSI Collaboration with FDA CBER BEST</td>
<td>33</td>
</tr>
<tr>
<td>Darwin EU® Initiative</td>
<td>34</td>
</tr>
<tr>
<td>EHDEN</td>
<td>35</td>
</tr>
<tr>
<td>Study-A-Thons and Other Events</td>
<td>36</td>
</tr>
<tr>
<td>OHDSI's COVID-19 Study-A-Thon</td>
<td>38</td>
</tr>
<tr>
<td>Support The Journey</td>
<td>39</td>
</tr>
<tr>
<td>Symposia Around The World</td>
<td>40</td>
</tr>
<tr>
<td>V. Data Standards</td>
<td>43</td>
</tr>
<tr>
<td>OMOP Common Data Model</td>
<td>44</td>
</tr>
<tr>
<td>OMOP CDM Data Sources</td>
<td>46</td>
</tr>
<tr>
<td>OHDSI Evidence Network</td>
<td>48</td>
</tr>
<tr>
<td>OHDSI Standardized Vocabularies</td>
<td>50</td>
</tr>
<tr>
<td>Vocabularies Improvement Initiative</td>
<td>52</td>
</tr>
<tr>
<td>VI. Open-Source Software</td>
<td>54</td>
</tr>
<tr>
<td>HADES Packages</td>
<td>55</td>
</tr>
<tr>
<td>Kheiron Contributor Cohort</td>
<td>58</td>
</tr>
<tr>
<td>Package Statuses/Maintainers</td>
<td>59</td>
</tr>
<tr>
<td>ATLAS</td>
<td>60</td>
</tr>
<tr>
<td>VII. Methods Research</td>
<td>61</td>
</tr>
<tr>
<td>Empirical Calibration</td>
<td>62</td>
</tr>
<tr>
<td>Principles of the LEGEND Project</td>
<td>63</td>
</tr>
<tr>
<td>The LEGEND Project</td>
<td>64</td>
</tr>
<tr>
<td>Patient-Level Prediction</td>
<td>66</td>
</tr>
<tr>
<td>Evidence Synthesis</td>
<td>70</td>
</tr>
<tr>
<td>VIII. Publications</td>
<td>71</td>
</tr>
<tr>
<td>Collaborations Graph</td>
<td>72</td>
</tr>
<tr>
<td>Community Dashboards</td>
<td>74</td>
</tr>
<tr>
<td>OHDSI Publications (2010 - August 2023)</td>
<td>75</td>
</tr>
<tr>
<td>IX. Join The Journey</td>
<td>101</td>
</tr>
<tr>
<td>Building Community, One Lego At A Time</td>
<td>102</td>
</tr>
<tr>
<td>Closing Letter</td>
<td>104</td>
</tr>
<tr>
<td>How Can You Join The Journey?</td>
<td>Inside Back Cover</td>
</tr>
</tbody>
</table>
Welcome to the third edition of Our Journey.

This book highlights the Observational Health Data Sciences and Informatics (OHDSI) journey from its inception in 2013—growing out of the Observational Medical Outcomes Partnership (OMOP)—to today. Our mission is to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

We have created a community of thousands of collaborators, a federated database with over 12 percent of the world’s population, models and standards for representing that population, and systematic research methods and tools that allow us to generate reliable, scalable evidence in health care. We have used these resources to influence medical decisions in areas like hypertension treatment and COVID-19, and our evidence has affected hundreds of millions of patients.

I would like to focus here on the scale of evidence. I believe that OHDSI’s focus on the reliability of evidence generated at large scale sets it apart from most other efforts. There are many data models, and now an increasing number of data networks, but they are just a means to an end. OHDSI is about generating evidence. Models and networks are built under the assumption that that is the hard part and if we can just improve access to the data, the rest will follow. Yet the medical literature remains sparse and unreliable. Huge databases like MarketScan have been around for decades, yet we have not even attempted to answer most medical questions that can be answered using those databases. Recent data networks pull data from electronic health records, making them more detailed, but the lack of forward motion on the old databases bodes poorly for full use of new databases and networks. The primary problem is not the data.

Reliability has been a key challenge, with some insisting that only randomized trials deliver evidence worthy of medical decisions. Unfortunately, this leaves the vast majority of medical decisions up to chance and bias, allowing us to pick whatever pays the most or costs the least or otherwise fits our agenda without regard for the actual health of patients. Most clinical research groups see themselves as pushing forward the reliability agenda and generating reliable research. What sets OHDSI apart is the breadth of its reliability agenda, embodied in its ten LEGEND Principles (see image above), which can be distilled into two commandments: verify and be open. Every step of the research process should be verified with diagnostics and every step of the research process must be made public: pre-specified protocol, source code, diagnostics, and results.

This brings us to scale. A byproduct of generating reliable evidence is making the process more systematic, which allows for larger-scale evidence generation. Here, OHDSI may be unique. I do not see other networks and initiatives pushing for scale. Large scale is only possible with systematic processes and extensive diagnostics, and these have been OHDSI’s focus. We need to incorporate existing knowledge into our analyses—knowledge of biases and confounders, knowledge of physiology, and previous evidence—but we still need to do it at large scale if we want to have a useful effect on medical practice.

We can scale in several ways. We can rely on our community to work in parallel, identifying solvable problems and carrying out the research needed to produce relevant evidence. We can pick areas of medicine with shared processes and biases, and carry out many studies in parallel within each area; LEGEND hypertension, LEGEND diabetes, and our depression drug side effect study are examples of this. We need to advance phenotyping so that each new clinical concept does not take weeks to define and verify.
WELCOME LETTER TO THE COMMUNITY

Since the last edition of Our Journey, large language models like ChatGPT have exploded onto the scene. I think for many, the answer to scale is clear: train large language models on all the medical literature and all the clinical databases, and then give proper prompts to answer all answerable medical questions. The problem goes back to reliability. Is it verified and open? Insofar as these models are black boxes, they are neither verified nor open.

We are learning how to ask questions so that the models reveal their causal argument, but that work is early and a way off from proving reliability. And early failures in having large language models produce large-scale evidence may reflect poorly on OHDSI's own large-scale effort. It is important for OHDSI to research the potential and limitations of such models and incorporate them into its large-scale efforts.

It is an exciting time to be doing observational medical research. Going large scale is an extraordinary challenge and opportunity, and I believe that OHDSI is the only group poised to do it.

- George Hripcsak

II. OHDSI Mission and Values
**OHDSI Mission**

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

**OHDSI Vision**

A world in which observational research produces a comprehensive understanding of health and disease.

**OHDSI Values**

**Innovation:** Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.

**Reproducibility:** Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.

**Community:** Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.

**Collaboration:** We work collectively to prioritize and address the real-world needs of our community’s participants.

**Openness:** We strive to make all our community’s proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.

**Beneficence:** We seek to protect the rights of individuals and organizations within our community at all times.

Observedational Health Data Sciences and Informatics (OHDSI, pronounced “Odyssey”) strives to promote better health decisions and care through globally standardized health data, continuously developing large-scale analytics and a spirit of collaboration though open science.

Founded in 2013, OHDSI is a growing collaborative of more than 3,700 researchers across disciplines (including biomedical informatics, epidemiology, statistics, computer science, health policy, clinical sciences), across stakeholders (including academia, industry, government and regulatory authorities, and health providers), and across geographies (including 83 countries and six continents). OHDSI also has established an international distributed data network that applies one open community data standard and collectively contains data for more than 950 million patients around the world, and has produced a suite of open-source software packages that enables the community to translate that data into reliable evidence.

OHDSI collaborates to establish open community data standards, develop open source software, conduct methodological research, and apply best practices across the OHDSI data network to generate clinical evidence. The OHDSI distributed data network is comprised of data partners who standardize their source data through an extract-transform-load (ETL) into the OMOP Common Data Model (CDM) and apply OHDSI open-source tools securely behind their own firewall.

OHDSI network studies involve researchers collaborating to design analyses.
OHDSI MISSION AND VALUES

The Department of Biomedical Informatics at Columbia University (DBMI) serves as the coordinating center for the OHDSI community. Located on the Columbia University Irving Medical Center campus, DBMI is both an academic department and an information services partner to NewYork-Presbyterian Hospital, a major healthcare provider in greater New York.

One of the oldest informatics departments in the nation, faculty and students at DBMI have set the path for design of clinical information systems, methodologies in clinical natural language processing, and machine learning over electronic health record data.

Faculty research includes the development and evaluation of innovative information technologies, which has led to enhancements in both health and healthcare. Both faculty and students work in a highly collaborative environment, applying informatics from the atomic level to global populations.

OHDSI’s research has been presented across various scientific societies, such as American Medical Informatics Association (AMIA), American Statistics Association (ASA/JSM), and International Society of Pharmacoepidemiology (ISPE), and published in top medical journals, including The Lancet, JAMA, BMJ, PNAS and JAMIA.

Our growing global community is always seeking new collaborators.

Please learn more about OHDSI through this publication and Join The Journey!
The OHDSI community brings together volunteers from around the world to establish open community data standards, develop open-source software, conduct methodological research, and apply scientific best practices to both answer public health questions and generate reliable clinical evidence.

Map of Collaborators

Our community is ALWAYS seeking new collaborators. Do you want to focus on data standards or methodological research? Are you passionate about open-source development or clinical applications? Do you have data that you want to be part of global network studies? Do you want to be part of a global community that truly values the benefits of open science? Add a dot to the map below and JOIN THE JOURNEY!

OHDSI By The Numbers

- 3,758 collaborators
- 83 countries
- 21 time zones
- 6 continents
- 1 community
OHDSI collaborators include a global community of collaborators. Many of the individuals represent organizations who contribute to and benefit from their participation in the OHDSI community. OHDSI is proud to collaborate with the more than 150 organizations who contribute to and benefit from their participation in the OHDSI community. OHDSI is a global community of collaborators. Many of the individuals represent organizations who contribute to and benefit from their participation in the OHDSI community. OHDSI is proud to collaborate with the more than 150 organizations who contribute to and benefit from their participation in the OHDSI community.
Testimonials From The OHDSI Community

What makes OHDSI unique is its way of conducting trustworthy research and taking care of every detail, from the research idea through validating the data and selecting the best methodological design. I would love to see more involvement of OHDSI in many conferences in the Middle East and west of Asia to make people aware of OHDSI and all tools, practices, and experiences. A main advantage of OHDSI that countries should be aware of is the community itself … the different expertise, sharing knowledge, working together, and helping each other are examples of the beauty of this community.

Thamir AlShammary
Advisor to the President of the Saudi Food and Drug Authority (SFDA)

Completing that [effectiveness/safety of famotidine as a COVID treatment] study in a span of 3 weeks, I experienced what blows every scientist’s mind. Once the clinical question was clear, OHDSI resources/tools/best practices enabled me to rapidly generate evidence that was critically needed in a structured, standardized, transparent and reliable manner that no other way can possibly do.

Azza Shoiabi
Associate Director with Janssen Research and Development, Inc.

As the COVID pandemic still looms large in our rear-view, a sense of urgency to generate and apply advanced scientific methods to public health issues remains palpable to me. I believe the OMOP + FHIR partnership provides a focused, high-ROI opportunity to create relevant informatics advances with broad impact that will allow us to better address immediate and future health crises, both local and global.

Davera Gabriel
Director of Terminology Management, Johns Hopkins University

The OHDSI community has always been very welcoming. This is such an ambitious group of skilled people, who always try to make time to guide newcomers; it is remarkable. Personally, I have grown to love the field of observational health research. It combines my passion for cleaning messy data and my background in medical science. OHDSI has even inspired me to go back to academia and pursue a (part-time) PhD in medical informatics!

Maxim Moinat
Scientific Researcher, Erasmus MC

Research allows me to exercise clinical thinking skills in ways college classes don’t. Research gives you a broader, deeper understanding of any field. Especially in OHDSI, research is interdisciplinary, so you are not only learning the science underlying certain diseases, but also how to code with data, how to standardize and improve quality of data, how to do predictive modeling and estimation, how to characterize populations, how to reproduce evidence. You’re developing a skill set. OHDSI changed the way I think about participants in healthcare: I’ve gained an overwhelming appreciation for everyone’s expertise.

Alexandey Davydov
Technical Team Lead for the Medical Vocabulary, Odysseus Data Services

I do know these people, the people of OHDSI who daily move the community forward. I know how they love to work together. I’ve seen the results of this work, which are impressive. I believe in OHDSI because I believe in the people, their commitment and devotion, and their effective collaboration that is gradually changing the world for the better.

Faaizah Arshad
Undergraduate, UCLA

I believe in the people, their commitment and devotion, and their effective collaboration that is gradually changing the world for the better.

Mengling ‘Mornin’ Feng
Assistant Professor, National University of Singapore

I met Rae Woong Park from Ajou University at a conference and learned that he was the pioneer in clinical data standardization in South Korea. We chatted for hours about how his team has successfully implemented the OMOP CDM as the standard data schema for research clinical data and how that may have accelerated the clinical research in Korea. Inspired by his success, I learned more about OHDSI and its platforms, tools and organizations. I began to advocate for OHDSI as the data standard and consortium that we should follow as a country to our Ministry of Health and major healthcare institutes.
The Titan Awards

To recognize OHDSI collaborators (or collaborating institutions) for their contributions towards OHDSI’s mission, the OHDSI Titan Awards were introduced at the 2018 Symposium and have been awarded each year since then.

Each year, community members nominate individuals or institutions they feel have made significant contributions towards advancing OHDSI’s mission, vision and values. Once nominations are submitted, the OHDSI Titan Award Committee selects the award winners, and the honorees are announced at the annual global symposium.

The award categories, as well as previous recipients, are listed here.

**Data Standards**

This Titan Award recognizes extraordinary contributions by an individual, organization, or team in development or evaluation in community data standards, including OMOP common data model and standardized vocabularies

- **2022** – Melanie Philofsky, Odysseus Data Services
- **2021** – Maxim Moinat, The Hyve/Erasmus University Medical Center
- **2020** – Clair Blacketer, Janssen Research and Development
- **2019** – Oncology Workgroup (Michael Gurley, Northwestern University; Rimma Belenkaya, Memorial Sloan Kettering Cancer Center; Robert Miller, Tufts CTSI)
- **2018** – Vocabulary team (Christian Reich, IQVIA; Anna Ostroplats, Columbia University; Dmitry Dymshyts, Odysseus Data Services)

**Methods Research**

This Titan Award recognizes extraordinary contributions by an individual, organization, or team in development or evaluation in analytical methods for clinical characterization, population-level effect estimation, or patient-level prediction

- **2022** – Fan Bu, University of California, Los Angeles
- **2021** – Yong Chen, University of Pennsylvania
- **2020** – Nicolas Thurin, Université de Bordeaux
- **2019** – Jenna Reps, Janssen Research and Development
- **2018** – Martijn Schuemie, Janssen Research and Development; Marc Suchard, University of California, Los Angeles

**Clinical Applications**

This Titan Award recognizes extraordinary contributions by an individual, organization, or team in generating clinical evidence that improves health by informing better health decisions and better care

- **2022** – Xintong Li, University of Oxford
- **2021** – Asieh Golozar, Odysseus Data Services
- **2020** – Jenny Lane, University of Oxford
- **2019** – Oxford Study-A-Thon (Dani Prieto-Alhambra, University of Oxford; Edward Burn, University of Oxford; Jamie Weaver, Janssen Research and Development; Ross Williams, Erasmus University Medical Center)
- **2018** – Seng Chan You, Ajou University

**Open-Source Development**

This Titan Award recognizes extraordinary contributions by an individual in design, development, testing, and deployment of open-source software to enable observational analyses

- **2022** – Egill Fridgeirsson, Erasmus MC/James Gilbert, Janssen Research and Development
- **2021** – Adam Black, Odysseus Data Services
- **2020** – Anthony Sena, Janssen Research and Development
- **2019** – Pavel Grafkin, Odysseus Data Services
- **2018** – Christopher Knoll, Janssen Research and Development

Congratulations to our 2023 Titan Award nominees!

The 2022 Titan Award honorees came together at the OHDSI Symposium. Oxford professor Dani Prieto-Alhambra accepted the award for his student, Xintong Li, who was unable to attend the event.
IV. Collaborative Activities

This Titan Award recognizes an individual, team, or organization for their contributions to ensuring the sustainability of the OHDSI community.

2022 – Craig Sachson, Columbia University
2021 – Faaizah Arshad, UCLA; Ross Williams, Erasmus University Medical Center
2020 – COVID-19 Support Team, Erasmus University Medical Center
2019 – James Wiggins, Amazon Web Services
2018 – Lee Evans, LTS Computing LLC

This Titan Award recognizes an individual for their leadership in advancing the OHDSI mission.

2022 – Paul Nagy, Johns Hopkins University
2021 – Mui Van Zandt, IQVIA
2020 – Dani Prieto-Alhambra, University of Oxford
2019 – Peter Rijnbeek, Erasmus University Medical Center
2018 – Rae Woong Park, Ajou University School of Medicine

This Titan Award recognizes an individual for their collaborative spirit in helping their fellow community members reach their goals.

2022 – Ajit Londhe, Boehringer Ingelheim
2021 – Erica Voss, Janssen Research and Development
2020 – Talita Duarte-Salles, IDIAPJGol
2019 – Andrew Williams, Tufts Medical Center
2018 – Kristin Kostka, Deloitte; Mui Van Zandt, IQVIA

Community Collaboration

Community Leadership

Community Support
Collaborative Activities

OHDSI Workgroups

OHDSI has a central mission to improve health globally, but there are countless areas where our community can be of service. Work around data, methods, open-source tools, and clinical applications are all pieces of the puzzle, and within OHDSI, there are opportunities to work in any or many of these areas. Our workgroups, led by the extraordinary leads shown on these pages, present opportunities for all community members to find a home for their talents and passions, and make meaningful contributions. We are always looking for new collaborators. See an area where you want to contribute? Please Join The Journey!

www.ohdsi.org/workgroups

Collaborative Activities

OHDSI Workgroups

Our workgroups hold meetings, share files, chat asynchronously and more in the OHDSI Microsoft Teams environment. Collaborators can request access to any workgroup through an online form available on both OHDSI.org and our main OHDSI Microsoft Teams environment (see QR codes, right).

Want to learn more? Check out our homepage: ohdsi.org/ohdsi-workgroups
Regional Chapters and National Nodes

An OHDSI regional chapter represents a group of OHDSI collaborators located in a geographic area who wish to hold local networking events and meetings to address problems specific to their geographic location. The OHDSI Europe Chapter, in collaboration with the EHDEN project, recently created National Nodes to facilitate national and international collaborations. An OHDSI Europe National Node is a collection of research institutes within a member country. The Node builds on the strengths of the stakeholders and scientific communities of that country. Each Node has a lead institution that oversees the work of that Node and assigns a lead and co-lead.

European National Nodes

<table>
<thead>
<tr>
<th>Regional Chapters</th>
<th>National Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>Belgium</td>
</tr>
<tr>
<td>Leads: Ahmed El Sayed, Cynthia Sung</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>Germany</td>
</tr>
<tr>
<td>Lead: Nicole Pratt</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>Greece</td>
</tr>
<tr>
<td>Lead: Hua Xu</td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>Lead: Peter Rijnbeek</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>Italy</td>
</tr>
<tr>
<td>Lead: Lakshmi Kubendran</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>The Netherlands</td>
</tr>
<tr>
<td>Lead: Tatsuo Hiramatsu</td>
<td></td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>Portugal</td>
</tr>
<tr>
<td>Lead: Seng Chan Yeu</td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>Spain</td>
</tr>
<tr>
<td>Lead: Mengling ‘Mornin’ Feng</td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Lead: Jason Hsu</td>
<td></td>
</tr>
</tbody>
</table>

The weekly OHDSI community call is where our global network gathers together to share research, discuss various topics around observational health, keep apprised on community updates, learn about recent OHDSI research, learn about open-source tools or best practices within the community, and plenty more. Our weekly calls, led by Craig Sachson, take place on Tuesdays at 11 am ET. They are both recorded and posted to both OHDSI.org and our YouTube channel.

These pages highlight many of the meeting topics from the last year; please check out ohdsi.org/community-calls to learn more about these interactive community gatherings.
What Is Happening In OHDSI?
Join Our Weekly Community Call and Find Out!

**Collaborative Activities**

**What Is Happening In OHDSI?**
Join Our Weekly Community Call and Find Out!

**April 4: Data Diagnostics (SOS Week 2)**
- **Clair Blacketer**
  - Associate Director, Janssen Research and Development, Inc.

**April 11: Phenotype Development (SOS Week 3)**
- **Aseh Golozar**
  - AS Global Lead of Data Science, Covisio Data Services, Inc.

**May 2: Network Execution (SOS Week 6)**
- **Jenna Reps**
  - Associate Director, Janssen Research & Development
- **Jack Brewster**
  - Data Scientist, Upshift

**May 9: Study Diagnostics (SOS Week 7)**
- **George Hripscak**
  - Professor of Preventive Medicine, Emory University

**May 16: Evidence Synthesis (SOS Week 8)**
- **Yong Chen**
  - Director, Center for Health Outcomes & Policy Studies, Department of Epidemiology & Biostatistics, University of Pennsylvania
- **Martijn Schuemie**
  - Research Fellow, Epidemiology, Delft University of Technology and Eindhoven University of Technology

**June 27: Recent OHDSI Publications**
- **Building the International OHDSI Community**
- **Catherine Carver**
  - Senior Director, Office of Science and Policy, National Institutes of Health
- **Xingtong Li**
  - Assistant Professor, Department of Biostatistics, University of Washington
- **Kim López Güell**
  - Assistant Professor, Department of Epidemiology, University of California, San Francisco
- **Daniel Morales**
  - Assistant Professor, Department of Biostatistics, University of California, San Francisco

**July 11: European Symposium Review**
- **Jensie Los**
  - Researcher, National Institute of Public Health and the Environment

**July 25: Asia-Pacific (APAC) Regional Updates**
- **Jaemin Yu**
  - Senior Manager, Korea Data Science
- **Junhui Li**
  - Data Scientist, China Data Science
- **Nobuki Pratt**
  - Senior Manager, Australia

**How Can You Join Our Calls?**
If you are a part of the OHDSI Teams environment, you will receive a weekly calendar invite that includes the upcoming agenda. If you don’t have access, the link is on our Community Calls page, which features all recordings and updates from past calls.

Weekly calls are currently held on Tuesdays at 11 am ET. Learn more at our website!

www.ohdsi.org/community-calls
Phenotype Phebruary

“Phenotype Phebruary” has been an inclusive community activity aimed at putting focused attention toward the science of phenotyping and stimulating collaborations to develop and evaluate phenotype algorithms across the OHDSI data network.

Introduced in 2022, this month-long event provides community members opportunities to engage in discussions about specific phenotypes along with associated methodological processes and technical topics, as well as to build cohort definitions using the community’s open-source standardized tools, such as ATLAS and CapR, and to evaluate those cohort definitions using other HADES packages, such as PheEvaluator and CohortDiagnostics.

During Phenotype Phebruary in 2023, we collaboratively identified 11 phenotypes to develop, evaluate and publicly discuss, and we hosted four community discussions related to phenotype development and evaluation best practices.

What We Accomplished Together

In 2023, we collaboratively identified 11 phenotypes that we aimed to develop, evaluate, and publicly discuss throughout our second annual “Phenotype Phebruary.” We deliberated upon four primary topics related to phenotype development and evaluation best practices.

Utilizing the OHDSI tool Atlas, various community members spearheaded the development of phenotype definitions for Acute Pancreatitis, Anaphylaxis, Appendicitis, and all the other phenotypes listed on the previous page.

Community members from diverse institutions globally utilized the OHDSI tools CohortDiagnostics and PheEvaluator to evaluate the newly proposed phenotypes implemented on multiple data sources.

During our February community meetings, we shared the progress and insights garnered through collaborative phenotyping efforts. We used this platform to discuss best practices, and the ‘Phenotype Development and Evaluation Workgroup’ opened its weekly calls to foster real-time collaboration in developing and evaluating phenotypes.

The finalized phenotype definitions were archived in the OHDSI Phenotype Library, available in a version-controlled format that supports liberal licensing, referencing, and reuse. The entire process was transparent and open, facilitated through a public forum that encouraged community collaboration and feedback.

The figure to the left demonstrates the workflow adhered to by community members while using OHDSI tools for the development, evaluation, and curation of phenotype definitions.

In Phenotype Phebruary 2023, a total of 32 collaborators from 11 organizations actively led the conversation. The developed algorithms were assessed using a total of 14 observational data sources from 6 countries, incorporating 8 electronic health records and six administrative claims data sources.
OHDSI’s central mission is to generate real-world evidence that positively impacts global health. Achieving that mission requires rigorous network studies and an open-science ecosystem that can build trust in the evidence generated through these collaborative studies.

The OHDSI community works hard to build both methodological best practices for network studies and the open-source tools to carry them forward, but that doesn’t mean the process is simple. In fact, it’s so challenging that it requires a team effort.

During the spring of 2023, the OHDSI community initiated the SOS Challenge, a global effort to design, implement, execute and ultimately disseminate four network studies. Two studies were featured weekly over the course of nine community calls in different time zones to be inclusive for all collaborators, while two other studies were run asynchronously. While doing this, OHDSI faculty provided focused sessions to teach each step of the network study journey. The SOS Challenge homepage has each tutorial video, as well as information on all four studies.

www.ohdsi.org/SOS-Challenge

The Process

1. Initiating A Network Study
2. Data Diagnostics
3. Phenotype Development
4. Phenotype Evaluation
5. Creating Analysis Specifications
6. Network Execution
7. Study Diagnostics
8. Evidence Synthesis
9. Interpreting The Results

Learn More
Want to learn more about any of these steps? Check out the homepage, which has all tutorial videos!
Collaborative Activities

The Book of OHDSI

Published in 2019, the Book of OHDSI (book.ohdsi.org) aims to be a central knowledge repository for OHDSI, and it focuses on describing the OHDSI community, OHDSI data standards, and OHDSI tools.

It is intended for both OHDSI newcomers and veterans alike, and aims to be practical, providing the necessary theory and subsequent instructions on how to design and implement research yourself. You will learn about the OMOP common data model and standard vocabularies, and how they can be used to standardize an observational healthcare database. You will learn about three analytic use cases for these data: characterization, population-level estimation, and patient-level prediction. You will read about OHDSI's open-source tools and how they can be applied to your data and how you can design and implement your own analyses following OHDSI's best practices.

Chapters on data quality, clinical validity, software validity, and method quality will explain how to establish the quality of the generated evidence. Lastly, you will learn how to use the OHDSI tools to execute these studies in a distributed research network.

The Book of OHDSI is available for free online in English, Korean, and Chinese, and can also be purchased through Amazon (all links on OHDSI.org).

Thank You To Our Book of OHDSI Contributors

Hamed Abedast
Brian Christian
Sergio Esteva
Mark Eaton
Chen Li
Ellen Palmer
Christian Reich
Izzy Saradikis
Sunah Song
Don Torok
Mike Warte

Mustafa Ascha
Gino Colt
Clark Evans
Greg Khebarov
David Madigan
Nirav Patil
Jenna Reps
Sarah Sarofim
Matthew Spotnitz
Kees van Bochove
Jamie Weaver
Mark Beno
Frank DeFalco
Thomas Falcóner
Kristin Kuoets
Sindhoosha Malay
Jose Posada
Piet Rijnbeek
Martijn Schuemie
Marc Schudde
Mai Van Zandt
James Wiggins
Clair Blacketer
Dana Dempster
George Hripcak
Bob Lanese
Wanda Lattimore
Harry Menegay
Nicole Pratt
Patrick Ryan
Sarah Seager
Joel Swardel
Erica Voss
Andrew Williams
David Blatt
Jon Duke
Vojtech Hauer
Ahikihoko Nishimura
Danie Prieto-Alhambra
Craig Sachson
Anthony Sana
Devlin Tien
Krystin Waite
Seng Chan You

What Will You Find In The Book of OHDSI?

Preface

Chapter 1: The OHDSI Community

Chapter 2: Open Science

Chapter 3: Open Standards

Chapter 4: The Common Data Model

Chapter 5: Defining a Cohort for Hypertension

Chapter 6: Defining Cohorts

Chapter 7: Data Analytics Use Cases

Chapter 8: OHDSI Analytics Tools

Chapter 9: SQL and R

Chapter 10: Defining Cohorts

Chapter 11: Cohort Characterization

Chapter 12: Cohort Method Design

Chapter 13: Data Extraction

Chapter 14: Characteristics of Health Care Databases

Chapter 15: Data Quality

Chapter 16: Character Validity

Chapter 17: Method Validity

Chapter 18: Method Validity

Chapter 19: Study Steps

Chapter 20: OHDSI Network Research

Appendix

IV. Evidence Quality

14 Evidence Quality

14.1 Attributes of Reliable Evidence

14.2 Development Strategy

14.3 Communicating Evidence Quality

15 Data Quality

15.1 Sources of Data Quality Problems

15.2 Data Quality in General

15.3 Study-Specific Checks

15.4 ACHILLES in Practice

15.5 Designing a Simple Study

15.6 Study-Specific Checks in Practice

15.7 Summary

15.8 Exercises

16 Clinical Validity

16.1 Characteristics of Health Care Databases

16.2 Summary

16.3 Source Record Verification

16.4 Phenokracker

16.5 Generalizability of the Evidence

16.6 Summary

17 Software Validity

17.1 Study-Care Validity

17.2 Methods Library Software Development Process

17.3 Methods Library

17.4 Summary

18 Method Validity

18.1 Design-Specific Diagnostics

18.2 Diagnostics for All Estimation

18.3 Method Validation in Practice

18.4 OHDSI Measurement Report

18.5 Summary

19 Study Steps

19.1 General Best Practice Guidelines

19.2 Study Steps in Detail

19.3 Summary

20 OHDSI Network Research

20.1 OHDSI as a Research Network

20.2 OHDSI's Cross-Project Tiers

20.3 Running an OHDSI Network Study

20.4 Look-Foward Using Network Study Automation

20.5 Best Practices for OHDSI Network Studies

20.6 Summary

Index

Appendix

A Glossary

B Cohort Definitions

C Negative Controls

D Protocol Template

E Web Resources and Bibliography

F Bibliography

G Index

OHDSI.org
OHDSI + Large Community Initiatives

OHDSI is proud to collaborate with large community initiatives around the world, to support the adoption of the OMOP Common Data Model and OHDSI tools, and to advance our shared interests in generating reliable evidence.

If your organization would like to collaborate with OHDSI, please reach out on our forums!

The All of Us Research Program collects a wide range of health information, including genetic, clinical, environmental, wearable and lifestyle data, from participants across diverse backgrounds. OMOP is central to the representation of this diverse dataset. Over 55 sites across the country are contributing EHR data are participants in OMOP format, which gets combined with other data types, allowing researchers better understand the causes of various diseases, develop more effective treatments, and tailor medical care to individual needs. Currently, the data set contains EHR data from over 250K participants, surveys from 400K participants and whole genome sequences on over 250K participants.

US FDA CBER Biologics Effectiveness and Safety (BEST) Initiative

Researchers within the OHDSI community, primarily comprising personnel from Columbia University, UCLA, Northeastern University and Johns Hopkins University, currently provide support to the U.S. Food and Drug Administration (FDA) Biologics Effectiveness and Safety (BEST) Initiative in its mission to conduct safety and effectiveness surveillance of biologic products (vaccines, blood and blood products, tissues and advanced therapeutics).

Specific means of FDA support through this grant include serving in a convening role to 1) develop methods related to using observational data from electronic health records and administrative claims to study the effectiveness and safety of biologics, 2) work collaboratively with FDA staff to plan, develop, coordinate and convene meetings and workshops, and 3) educate FDA staff and external stakeholders on the BEST-infrastructure, capabilities, and applications that serve FDA and stakeholder needs.

Vaccine Surveillance Methods Research

Factors Influencing Background Incidence Rate Calculation: Systematic Empirical Evaluation Across an International Network of Observational Databases

Factors influencing background incidence rate calculation are an important consideration in vaccine effectiveness studies. This study examined the impact of these factors across an international network of observational databases, with the aim of developing recommendations for choosing an appropriate background rate calculation method. Four common methods were considered: the global rate, the site-specific rate, the population-based rate, and the unadjusted rate. The study used a range of data sources and disease outcomes to evaluate the methods. The findings suggested that the choice of method can have a significant impact on the results, and that careful consideration is needed to select the most appropriate method for a given study.
Collaborative Activities

**DARWIN EU®**

The European Medicines Agency (EMA) and the European Medicines Regulatory Network established a coordination centre to provide timely and reliable evidence on the use, safety and effectiveness of medicines for human use, including vaccines, from real world healthcare databases across the European Union (EU). This capability is called the Data Analysis and Real World Interrogation Network (DARWIN EU®).

EMA is working with Erasmus University Medical Center Rotterdam to: 1) establish the DARWIN EU Coordination Centre, and support its work to build a distributed data network; 2) conduct scientific studies and answer research questions supporting regulatory decision-making by EMA’s scientific committees and the European Medicines regulatory network; and 3) maintain a catalogue of real world data sources for use in the regulatory context and their metadata. DARWIN EU® is using the OMOP common data model and OHDSI tools as part of its operations. OHDSI Titan Award winners Peter Rijnbeek (Executive Director, Technology Pillar Lead), Dani Prieto-Alhambra (Deputy Director and Development Lead), Maxim Moinat (Network Operations Pillar Lead), Ross Williams (Analytics Team Co-lead) as well as many other OHDSI collaborators from Erasmus MC, University of Oxford, Synapse, IQVIA, The Hyve, and Odysseus Data Services are participating.

Learn more at darwin-eu.org.

---

**The EH DEN Academy**

The EH DEN Academy serves as a free, publicly available online educational resource for anyone working in the domain of real-world data and real-world evidence. It continues to evolve as a valuable and highly-rated resource on tools, methods and skills for all those who generate and utilize data, work technically with it, e.g., ETL and mapping, and are involved in methodological development and the use of standardized tools.

The Academy currently counts more than 600 enrollees from 100 countries across the globe and offers the following 19 courses led by a range of industry experts:

- Getting Started
- EH DEN Foundation
- Patient Organisations: Introduction to Real World Data & Real World Research
- OMOP CDM and Standardised Vocabularies
- ATLAS
- Infrastructure
- Extract, Transform and Load
- Introduction to Usagi & Code Mappings for an ETL
- OHDSI in a Box
- Open Science & FAIR Principles
- Introduction to Data Quality
- Phenotype Definition, Characterisation and Evaluation
- Population-level Effect Estimation
- Patient-level Prediction
- R for Patient-level Prediction
- Applied Cost-Effectiveness Modeling with R
- Health Technology Assessment

Coming Soon:
- Assessing healthcare outcomes that matter to patients
- Assessing healthcare outcomes that matter to patients

---

**The EHDEN Portal**

To facilitate its European federated network and the research workflow from discovery to analyses within the Findable, Accessible, Interoperable and Reusable (FAIR) principles, EH DEN built the sociotechnical architecture: the EHDEN Portal. The Portal is a one-stop shop for researchers to identify relevant data partners for federated network studies on observational health data. It currently consists of 119 databases from 27 countries and more than 221 million health records that have been mapped to OMOP. Additionally, the Portal is a great starting point to learn about the OHDSI ecosystem through the informative courses available in the EH DEN Academy (see left).

The EHDEN Portal is the main tool for data partners and researchers, covering multiple aspects from account creation, single sign off authentication and approval to role assignment. The Portal plays a crucial role in the initial study set up phase, starting with a feasibility assessment to understand which data partners can provide relevant databases and are interested in contributing to a study. After data partners are identified, research queries and protocols can be shared prior to fully engaging, contracting and seeking approvals from data partners’ ethical boards. Therefore, the EHDEN Portal is the starting point for running rapid network analyses, enabling the open science community to generate high quality evidence.

---

**Collaborative Activities**

**The EH DEN Portal**

The EH DEN Portal is the main tool for data partners and researchers, covering multiple aspects from account creation, single sign off authentication and approval to role assignment. The Portal plays a crucial role in the initial study set up phase, starting with a feasibility assessment to understand which data partners can provide relevant databases and are interested in contributing to a study. After data partners are identified, research queries and protocols can be shared prior to fully engaging, contracting and seeking approvals from data partners’ ethical boards. Therefore, the EH DEN Portal is the starting point for running rapid network analyses, enabling the open science community to generate high quality evidence.

---

**The EH DEN Academy**

The EH DEN Academy serves as a free, publicly available online educational resource for anyone working in the domain of real-world data and real-world evidence. It continues to evolve as a valuable and highly-rated resource on tools, methods and skills for all those who generate and utilize data, work technically with it, e.g., ETL and mapping, and are involved in methodological development and the use of standardized tools.

The Academy currently counts more than 600 enrollees from 100 countries across the globe and offers the following 19 courses led by a range of industry experts:

- Getting Started
- EH DEN Foundation
- Patient Organisations: Introduction to Real World Data & Real World Research
- OMOP CDM and Standardised Vocabularies
- ATLAS
- Infrastructure
- Extract, Transform and Load
- Introduction to Usagi & Code Mappings for an ETL
- OHDSI in a Box
- Open Science & FAIR Principles
- Introduction to Data Quality
- Phenotype Definition, Characterisation and Evaluation
- Population-level Effect Estimation
- Patient-level Prediction
- R for Patient-level Prediction
- Applied Cost-Effectiveness Modeling with R
- Health Technology Assessment

Coming Soon:
- Assessing healthcare outcomes that matter to patients

---

**The EH DEN Portal**

To facilitate its European federated network and the research workflow from discovery to analyses within the Findable, Accessible, Interoperable and Reusable (FAIR) principles, EH DEN built the sociotechnical architecture: the EHDEN Portal. The Portal is a one-stop shop for researchers to identify relevant data partners for federated network studies on observational health data. It currently consists of 119 databases from 27 countries and more than 221 million health records that have been mapped to OMOP. Additionally, the Portal is a great starting point to learn about the OHDSI ecosystem through the informative courses available in the EH DEN Academy (see left).

The EHDEN Portal is the main tool for data partners and researchers, covering multiple aspects from account creation, single sign off authentication and approval to role assignment. The Portal plays a crucial role in the initial study set up phase, starting with a feasibility assessment to understand which data partners can provide relevant databases and are interested in contributing to a study. After data partners are identified, research queries and protocols can be shared prior to fully engaging, contracting and seeking approvals from data partners’ ethical boards. Therefore, the EHDEN Portal is the starting point for running rapid network analyses, enabling the open science community to generate high quality evidence.
OHDSI Study-A-Thons & Other Events

How does OHDSI go about empowering a community to collaboratively generate the evidence that promotes better health decisions and better care?

We do it by innovating on what it means to do collaborative research. The premise of the study-a-thon is simple: bring together a diverse group of researchers aligned on a common question and focus together on collaboratively designing research protocols, executing analyses across databases, and interpreting results over an intense but fun-filled few days.

OHDSI collaborators have held multiple study-a-thons on a wide array of topics, including orthopedic surgery, rheumatoid arthritis, colorectal cancer, cardiovascular prediction, prostate cancer, and COVID-19. Each event has demonstrated our collective ability to accomplish in a short time what may be unimaginable alone, and it has provided further reinforcement of the power of community and the value of multi-disciplinary collaboration.

Our most memorable study-a-thon happened at the beginning of the COVID pandemic. Read more about that on page 38.
COLLABORATIVE ACTIVITIES

88 Hours: OHDSI’s Signature Moment

The signature study-a-thon — and community event — for OHDSI came in March of 2020, when the COVID pandemic began to close down much of the world. When the planned European Symposium was cancelled, our global collaborators came together for four days of rigorous work about a disease with limited available data. These 88 hours set the foundation for years of COVID research, as well as ongoing research around vaccine surveillance.

Read the feature on this memorable community event by either scanning the QR code or visiting ohdsi.org/88-hours.

What You Should Know About
The 2020 OHDSI COVID-19 Study-A-Thon

- More than 330 people from across 30 countries (six continents) registered for the event.
- The event took place over 88 hours between March 26-29, and it was coordinated by the Erasmus University Medical Center.
- There were 17 concurrent channels on the overall Teams platform, and those channels hosted more than 100 collaborator calls.
- There were 12 global huddles, spaced out so collaborators from around the world would have a daily opportunity to hear about community progress.
- More than 10,000 publications were reviewed both prior and during the event.
- There were 13,000+ chat messages that helped design both 355 cohort definitions and nine protocols, as well as the release of 13 study packages.
- The closing call has been viewed almost 1,800 times since it was posted to YouTube.
- The OHDSI community has published multiple COVID-19 studies (including in Lancet Rheumatology, Nature Communications, Lancet Digital Health, and The BMJ), and it continues to do research in many areas of COVID-19 and vaccine surveillance.

Support The Journey

The OHDSI community is comprised of a global team of volunteers who collaborate together using open-source tools and shared best practices to support our shared mission of generating real-world evidence that promotes better health decisions and better care.

In order to foster growth in our community of nearly 3,500 volunteers across six continents, the OHDSI Coordinating Center at Columbia University has created a sponsorship program. This program allows both corporations and individuals to make meaningful contributions in support of OHDSI’s central coordinating activities. There are three levels of support, including donation amount and benefits to the sponsor, detailed below. Any level of support enhances both our community and our mission.

If you are interested, please reach out to sponsorship@ohdsi.org.

Gold Sponsorship • Donation Level: US $500k/year
- Your logo will be placed on our OHDSI Sponsors page (under Gold Level Sponsors heading) with link to your home page
- Sponsors Spotlight feature (Q&A with a member of your organization) placed on website and newsletter
- Monthly recognition on OHDSI Twitter (2800+ followers) and LinkedIn (5500+) pages
- Inclusion in “Thank You Sponsors” graphic in all OHDSI monthly newsletters (4200+ on mailing list)
- Listing in all OHDSI annual reports: Our Journey
- Recognition at all OHDSI in-person events

Silver Sponsorship • Donation Level: US $100k/year
- Your logo will be placed on our OHDSI Sponsors page (under Silver Level Sponsors heading) with link to your home page
- Use of OHDSI Silver Sponsor logo on your webpage
- Annual meeting with OHDSI leadership to learn about current and future initiatives, and participate in an OHDSI sponsor Q&A session
- Logo placement on title slide of OHDSI community call (average >=180 attendees per week)

Bronze Sponsorship • Donation Level: US $25k/year
- Your logo will be placed on our OHDSI Sponsors page (under Bronze Level Sponsors heading) with link to your home page
- Use of OHDSI Bronze Sponsor logo on your webpage
- Inclusion in “Thank You Sponsors” graphic in all OHDSI monthly newsletters (4200+ on mailing list)
- Listing in all OHDSI annual reports: Our Journey
- Recognition at all OHDSI in-person events

What You Should Know About
The 2020 OHDSI COVID-19 Study-A-Thon

• More than 330 people from across 30 countries (six continents) registered for the event.
• The event took place over 88 hours between March 26-29, and it was coordinated by the Erasmus University Medical Center.
• There were 17 concurrent channels on the overall Teams platform, and those channels hosted more than 100 collaborator calls.
• There were 12 global huddles, spaced out so collaborators from around the world would have a daily opportunity to hear about community progress.
• More than 10,000 publications were reviewed both prior and during the event.
• There were 13,000+ chat messages that helped design both 355 cohort definitions and nine protocols, as well as the release of 13 study packages.
• The closing call has been viewed almost 1,800 times since it was posted to YouTube.
• The OHDSI community has published multiple COVID-19 studies (including in Lancet Rheumatology, Nature Communications, Lancet Digital Health, and The BMJ), and it continues to do research in many areas of COVID-19 and vaccine surveillance.

Support The Journey

The OHDSI community is comprised of a global team of volunteers who collaborate together using open-source tools and shared best practices to support our shared mission of generating real-world evidence that promotes better health decisions and better care.

In order to foster growth in our community of nearly 3,500 volunteers across six continents, the OHDSI Coordinating Center at Columbia University has created a sponsorship program. This program allows both corporations and individuals to make meaningful contributions in support of OHDSI’s central coordinating activities. There are three levels of support, including donation amount and benefits to the sponsor, detailed below. Any level of support enhances both our community and our mission.

If you are interested, please reach out to sponsorship@ohdsi.org.

Gold Sponsorship • Donation Level: US $500k/year
• Your logo will be placed on our OHDSI Sponsors page (under Gold Level Sponsors heading) with link to your home page
• Use of OHDSI Gold Sponsor logo on your site
• Joint press release with OHDSI
• Annual meeting with OHDSI leadership to learn about current and future initiatives, and participate in an OHDSI sponsor Q&A session
• Weekly logo placement on title slide of OHDSI community call (average >=180 attendees per week)

Silver Sponsorship • Donation Level: US $100k/year
• Your logo will be placed on our OHDSI Sponsors page (under Silver Level Sponsors heading) with link to your home page
• Use of OHDSI Silver Sponsor logo on your webpage
• Quote for your press release
• Annual meeting with OHDSI leadership to learn about current and future initiatives, and participate in an OHDSI sponsor Q&A session
• Logo placement on monthly “Thank You Sponsors” slide during OHDSI community call

Bronze Sponsorship • Donation Level: US $25k/year
• Your logo will be placed on our OHDSI Sponsors page (under Bronze Level Sponsors heading) with link to your home page
• Use of OHDSI Bronze Sponsor logo on your webpage
• Inclusion in “Thank You Sponsors” graphic in all OHDSI monthly newsletters (4200+ on mailing list)
• Listing in all OHDSI annual reports: Our Journey
• Recognition at all OHDSI in-person events
There is nothing quite like an OHDSI symposium. Whether it is held in the U.S., Europe or Asia, our community has turned the symposium into one of the most anticipated events of the year. The pandemic forced a temporary shift to virtual symposia, but we have been thrilled to return to in-person gatherings this year, beginning with the European symposium in June.

The opportunity to learn from each other and connect as colleagues and friends is unmatched, and our most impactful scientific discoveries are shared at the symposia. We hope you can join us at a future event!

Oct. 20, 2015 • Washington, D.C.

Sept. 23-24, 2016 • Washington, D.C.

Oct. 18-20, 2017 • Bethesda, Md.

Mar. 23-24, 2018 • Rotterdam, Neth.

Oct. 11-13, 2018 • Bethesda, Md.

June 27-29, 2019 • Guangzhou, China

Sept. 15-17, 2019 • Bethesda, Md.

Dec. 12-14, 2019 • Gwangju, Korea

Mar. 29-30, 2019 • Rotterdam, Neth.

June 24-26, 2022 • Rotterdam, Neth.
V. Data Standards

The 2023 OHDSI Global Symposium was held Oct. 20-22 in East Brunswick, NJ., USA. It featured the most diverse agenda in our event history, as well as a record total of Collaborator Showcase submissions, including 137 posters and 24 software demos that were presented throughout the three-day event.

Talks from previous symposia, as well as presentations from weekly community calls, monthly video podcasts, features and more are available on our YouTube page: youtube.com/c/OHDSI.

Check Out #OHDSI2023

The 2023 OHDSI Global Symposium was held Oct. 20-22 in East Brunswick, NJ., USA. It featured the most diverse agenda in our event history, as well as a record total of Collaborator Showcase submissions, including 137 posters and 24 software demos that were presented throughout the three-day event.

Talks from previous symposia, as well as presentations from weekly community calls, monthly video podcasts, features and more are available on our YouTube page: youtube.com/c/OHDSI.
OMOP Common Data Model

The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is an open community data standard, designed to standardize the structure and content of observational data and to enable efficient analyses that can produce reliable evidence.

“OMOP CDM serves as the foundation of all our work in the OHDSI community, and I’m proud that our open community data standard has been so widely adopted and so extensively used to generate reliable evidence.”

- Clair Blacketer
2020 Titan Award for Data Standards recipient

OMOP CDM By The Numbers

37 tables
• 17 to standardize clinical data
• 10 to standardize vocabularies

394 fields
• 193 with _id to standardize identification
• 101 with _concept_id to standardize content
• 43 with _source_value to preserve original data

1 Open Community Data Standard
Join The Journey

OHDSI Evidence Network

OHDSI is proud to have a global community dedicated to generating real-world evidence and which recognizes the opportunity to collaborate together as part of a distributed network based on standardized data and standardized analytics.

The OHDSI Evidence Network consists of organizations equipped with access to one or more databases standardized to the OMOP CDM who express a keen interest in participating in OHDSI network studies. Collaboratively, OHDSI Evidence Network partners share aggregate summary statistics about their databases, which are used to support Database Diagnostics, helping identify databases within the network that are fit-for-use for particular research questions. Additionally, partners have the opportunity to opt in and contribute to network studies proposed by the OHDSI community.

The recent SOS challenge serves as a compelling demonstration of the OHDSI Evidence Network’s current capabilities and its promising future potential. We wholeheartedly encourage all organizations that are adopting the OMOP CDM and aspire to apply standardized analytics for the reliable generation of real-world evidence to become part of the OHDSI Evidence Network.

During the first community call of 2023, Patrick Ryan unveiled the strategic priorities for the OHDSI Community for the year. Among these, a key focus is on enhancing the transparency and maturity of the OHDSI network. To address this objective, we began by considering how network studies are currently conducted, recognizing the challenges and complexities faced by collaborating organizations when contributing to the body of evidence. This investigation led to the creation of Database Diagnostics, a tool designed to answer a critical question: when tackling a specific research inquiry, which data sources within the OHDSI Evidence Network are the most relevant and suitable for generating robust evidence?

This innovative approach leverages aggregated summary statistics from each data source, obtained through the open-source tool dbProfile. It evaluates data fitness-for-use across various dimensions, including patient demographics, domain coverage, longitudinal data availability, and the capture of target, comparator, and outcome variables. The overarching vision was to establish these database profiles as the foundation to enable the OHDSI Evidence Network.

A message from Common Data Model workgroup lead Clair Blacketer ...

On March 28, 2023, the OHDSI Global Community initiated the Save Our Sisyphus (SOS) Challenge, a groundbreaking opportunity for collaborative research involving simultaneous participation in four different network studies. What made it truly remarkable was that any organization interested in joining the OHDSI Evidence Network could contribute to these studies by sharing their database profiles for the data sources they had access to. These profiles were centrally aggregated at the OHDSI Central Coordinating Center, enabling us to empirically determine which of the four study questions each data source was best suited to address. This inaugural OHDSI Evidence Network endeavor encompassed 36 diverse data sources from 16 different organizations. Not only did this foster rapid evidence generation and collaboration during the SOS Challenge, but it also positioned us for future collaborations on additional network studies as part of the OHDSI Evidence Network.

If you are interested in becoming a part of the OHDSI Evidence Network and contributing to advancing evidence-based healthcare, please use the provided QR code to complete a brief form about your organization and your data source. A member of the OHDSI Network Data Quality Working Group will reach out to you to explore this exciting opportunity further!
## OHDSI Standardized Vocabularies

The OHDSI vocabularies allow organization and standardization of medical terms to be used across the various clinical domains of the OMOP common data model, and enables standardized analytics that leverage the knowledge base when constructing exposure and outcome phenotypes and other features within characterization, population-level effect estimation, and patient-level prediction studies.

You can download the OHDSI Standardized Vocabularies at athena.ohdsi.org.

This treemap shows all concepts in the OHDSI vocabularies, organized by domain (color) and vocabularies (boxes sized by the number of concepts).

## OHDSI Vocabularies By The Numbers

- 11,027,290 concepts
- 3,598,454 standard concepts
- 847,008 classification concepts
- 142 vocabularies
- 44 domains
- 82,142,038 concept relationships
- 87,967,689 ancestral relationships
- 4,673,156 concept synonyms

“If we really want to achieve global collaboration, we need more than just standardizing data format. We have to establish a shared understanding of data meaning and speak the same language when expressing clinical ideas. The OHDSI vocabularies is a community resource that makes it possible to work to reach this common goal.”

- Christian Reich

2018 Titan Award recipient for Data Standards
OHDSI Vocabularies

OHDSI Standardized Vocabularies are a pillar of our community’s mission to generate real-world evidence that benefits healthcare. In 2023, there was a global focus on improving our vocabularies in four areas: scalability, robustness, timeliness, and transparency. Our Vocabulary Team conducted a landscape assessment early in 2023 to evaluate community needs. We received an amazing response, with 188 collaborators from 144 institutions describing their needs and challenges, as well as information on the vocabularies’ use in 60 data sources across the network. Results were made publicly available.

Based on the results of the assessment (see below, left), we identified the following areas for improvement: (1) Roadmap and release schedule; (2) Documentation & guidelines (developer, end-user, community contribution); (3) Quality framework and QA system; (4) Common development infrastructure; (5) Better code and automation (machinery); and (6) Vocabulary versioning, one-stop-shop and self-service.

To help in prioritization of these tasks, we established a Committee (below) that meets monthly to govern Vocabularies improvement and maintenance activities.

One of the first activities within this initiative focused on an effort to develop a stable release schedule and transparent roadmap. Vocabularies are now released in a semi-annual cycle based on the most common data refresh cadence in the community. Releases happen in August and February and their current and upcoming content is posted on OHDSI Vocabulary v5.0 GitHub Wiki (below, top image).

Our August 2023 release was the first one where we incorporated community contribution. We created the pipeline to enable the OHDSI community an opportunity to propose changes to the existing content and add their new content to the Vocabularies in a timely and transparent fashion (below, bottom image). Four first community contributions (two new vocabularies, mapping fixes and new concepts in the existing Vocabularies) have been processed in August release and more contributions are coming our way.

Thank you to the Vocabulary Team (see image on page 52) for all of their hard work this year to enhance our standardized vocabularies!
VI. Open-Source Software

Open-Source Software

HADES

HADES is a set of open source R packages for large scale analytics, including population characterization, population-level causal effect estimation, and patient-level prediction.

The packages offer R functions that together can be used to perform an observational study through the full journey from data to evidence, including data manipulation, statistical modeling, and results generation with supporting statistics, tables and figures.

Each package includes functions for specifying and subsequently executing multiple analyses efficiently. HADES supports best practices for use of observational data as learned from previous and ongoing research, such as transparency, reproducibility, as well as measuring of the operating characteristics of methods in a particular context and subsequent empirical calibration of estimates produced by the methods.

Learn more about the individual HADES packages in this section.

Population-Level Estimation

CohortMethod

CohortMethod is an R package for performing new-user cohort studies in an observational database in the OMOP Common Data Model.

EvidenceSynthesis

This R package contains routines for combining causal effect estimates and study diagnostics across multiple data sites in a distributed study. This includes functions for performing meta-analysis and forest plots.

SelfControlledCaseSeries

SelfControlledCaseSeries is an R package for performing Self-Controlled Case Series (SCCS) analyses in an observational database in the OMOP Common Data Model.

SelfControlledCohort

This package provides a method to estimate risk by comparing time exposed with time unexposed among the exposed cohort.

Patient-Level Prediction/Characterization

PatientLevelPrediction

PatientLevelPrediction is an R package for building and validating patient-level predictive models using data in the OMOP Common Data Model format.

DeepPatientLevelPrediction

DeepPatientLevelPrediction is an R package for building and validating deep learning patient-level predictive models using data in the OMOP Common Data Model format and OHDSI PatientLevelPrediction framework.

EnsemblePatientLevelPrediction

EnsemblePatientLevelPrediction is an R package for building and validating ensemble patient-level predictive models using data in the OMOP Common Data Model format. The package expands the OHDSI R PatientLevelPrediction package to enable ensemble learning.

Characterization

Characterization is an R package for performing characterization of a target and a comparator cohort.
Cohort Construction

CAPR

The goal of Capr, pronounced ‘ka-pee’ like the edible flower, is to provide a language for expressing OHDSI cohort definitions in R code. OHDSI defines a cohort as “a set of persons who satisfy one or more inclusion criteria for a duration of time” and provides a standardized approach for defining them (Circe-be). Capr exposes the standardized approach to cohort building through a programmatic interface in R which is particularly helpful when creating a large number of similar cohorts. Capr version 2 introduces a new user interface designed for readability with the goal that Capr code being a human readable description of a cohort while also being executable on an OMOP Common Data Model.

CirceR

A R-wraper for Circe, a library for creating queries for the OMOP Common Data Model. These queries are used in cohort definitions (CohortExpression) as well as custom features (CriterialFeature). This package provides convenient wrappers for Circe functions, and includes the necessary Java dependencies.

CohortDiagnostics

CohortDiagnostics is an R utility package for the development and evaluation of phenotype algorithms for OMOP CDM compliant data sets. This package provides a standard, end to end, set of analytics for understanding patient capture including data generation and result exploration through an R Shiny interface. Analytics computed include cohort characteristics, record counts, index event misclassification, captured observation windows and basic incidence proportions for age, gender and calendar year. Through the identification of errors, CohortDiagnostics enables the comparision of multiple candidate cohort definitions across one or more data sources, facilitating reproducible research.

CohortExplorer

This software tool is designed to extract data from a randomized subset of individuals within a cohort and make it available for exploration in a ‘Shiny’ application environment. It retrieves date-stamped, event-level records from one or more data sources that represent patient data in the Observational Medical Outcomes Partnership (OMOP) CDM data model format. This tool features a user-friendly interface that enables users to efficiently explore the extracted profiles, thereby facilitating applications, such as reviewing structured profiles. The output of this R-package is a self-contained R shiny that contains persistent data for review.

PhenotypeLibrary

The OHDSI community has developed a publicly accessible, version-controlled Phenotype Library to guide real-world evidence towards the FAIR principles: Findability, Accessibility, Reproducibility, and Interoperability. This library aims to foster the submission and retrieval of high-quality cohort definitions, cataloging of metadata, attribution and promotion of discovery and reuse in scientific research. Within the OHDSI Phenotype Library (OHDSI PL), each entry represents a unique cohort definition identifiable by a stable, externally referenceable ID. Comprehensive metadata about each cohort definition is cataloged and made searchable for researchers.Content in the library is subject to version control, with each version is assigned a specific DOI.

Evidence Quality

Supporting Packages

Andromeda

AsynchroNouS Disk-based Representation of Massive Data (Andromeda): An R package for storing large data objects. Andromeda allow storing data objects on a local drive, while still making it possible to manipulate the data in an efficient manner.

BrokenAdaptiveRidge

BrokenAdaptiveRidge is an R package for performing L_0-based regressions using Cyclops.

Cyclops

Cyclops (Cyclic coordinate descent for Logistic, Poisson and survival analysis) is an R package for performing large scale regularized regressions.

DatabaseConnector

This R package provides function for connecting to various DBMSs. Together with the SqlRender package, the main goal of DatabaseConnector is to provide a uniform interface across database platforms: the same code should run and produce equivalent results, regardless of the database back end.

Evidence Quality

Data Quality Dashboard

DataQualityDashboard (DQD) is an R package for exposing and evaluating observational data quality. This package runs a series of data quality checks against an OMOP CDM instance. It systematically runs the checks, evaluates each check against a pre-specified threshold, and then communicates what was done in a transparent and easily understandable way.

Evidence Quality

EmpiricalCalibration

This R package contains routines for performing empirical calibration of observational study estimates. By using a set of negative control hypotheses we can estimate the empirical null distribution of a particular observational study setup. This empirical null distribution can be used to compute a calibrated p-value, which reflects the probability of observing an estimated effect size when the null hypothesis is true taking both random and systematic error into account, as described in the paper Interpreting observational studies: why empirical calibration is needed to correct p-values. Also supported is empirical calibration of confidence intervals, based on the results for a set of negative and positive controls, as described in the paper Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data.

Method Evaluation

This R package contains resources for the evaluation of the performance of methods that aim to estimate the magnitude (relative risk) of the effect of a drug on an outcome. These resources include reference sets for evaluating methods on real data, as well as functions for inserting simulated effects in real data based on negative control drug-outcome pairs. Further included are functions for the computation of the minimum detectable relative risks and functions for computing performance statistics such as predictive accuracy, error and bias.

Epistemology

Epistemology is a R package for generating features (covariates) for a cohort using data in the Common Data Model.

Hydra

An R package and Java library for hydrating package skeletons into executable R study packages based on specifications in JSON format.

IterativeHardThresholding

IterativeHardThresholding is an R package for performing L_0-based regressions using Cyclops.

OdhsiSharing

This is an R package for sharing data between OHDSI partners.
**Supporting Packages**

**OhdsiShinyModules**

OhdsiShinyModules is an R package containing shiny modules that can be used within shiny result interfaces. The OHDSI tools often provide shiny interfaces for viewing and exploring results. Many of these shiny apps have overlapping features. To ensure consistency we have created a repository containing useful shiny modules that can be used in multiple result explorers.

**ParallelLogger**

Support for parallel computation with progress bar, and option to stop or proceed on errors. Also provides logging to console and disk, and the logging persists in the parallel threads. Additional functions support function call automation with delayed execution (e.g. for executing functions in parallel).

**ResultModelManager**

RMM is an R package designed to handle common OHDSI results data management functions by providing a common API for data model migrations and definitions.

**RohdsiWebApi**

RohdsiWebApi is an R based interface to ‘WebApi’ (OHDSI RESTful services), and performs GET/PULL/POST/DELETE calls via the WebApi. All objects starting from R or output to R - are analysis ready R-objects like list and data.frame. The package handles the intermediary steps by converting R-objects toate time are converted from string to POSIXct.

This package makes reproducible research easier, by offering ability to retrieve detailed study specifications, transport study specifications from one instance to another, programmatically invoke the generation of a sequence of steps that are part of a study, manage running studies in batch mode. An example of a WebApi endpoint is “http://server.org:80/WebAPI”.

**ShinyAppBuilder**

Create shiny apps using modules from OhdsiShinyModules or custom modules.

**SqlRender**

This is an R package for rendering parameterized SQL, and translating it to different SQL dialects. SqlRender can also be used as a stand-alone Java library and a command-line executable.

The eight HADES packages shown above have been released on CRAN and have been downloaded more than 500,000 times.

**Kheiron Contributor Cohort**

The Kheiron Contributor Cohort has entered its second year, and 21 new members have been accepted into the leadership program with the aim of onboarding new software developers into the OHDSI open-source software community. The Kheiron faculty has grown and now includes Adam Black, Nate Buesgens, Clark Evans, Paul Nagy, Katy Sadowski, Anthony Sena, and Dan Smith. The leadership program kicks off each spring at the OHDSI DevCon event. Developers commit 10% of their time for a year to participate in the open-source journey, working closely with an experienced OHDSI developer who volunteers to assist their mentees in making meaningful contributions to the community.

The cohort participates in hands-on workshops, attends Technical Advisory Board meetings, and performs development work in HADES, vocabulary mapping, and more. Details about the next year’s Kheiron Cohort will be shared at DevCon 2024.

**Package Statuses (as of 26Sep2023)**

<table>
<thead>
<tr>
<th>Package</th>
<th>Version</th>
<th>Maintainer(s)</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>HADES</td>
<td>0.2.8</td>
<td>Frank DeFalco</td>
<td>CRAN</td>
</tr>
<tr>
<td>database</td>
<td>0.2.0</td>
<td>Adam Black</td>
<td>CRAN</td>
</tr>
<tr>
<td>data</td>
<td>0.2.0</td>
<td>Harry Schuemie</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataModelManager</td>
<td>0.2.0</td>
<td>Marc Suchard</td>
<td>CRAN</td>
</tr>
<tr>
<td>dataR</td>
<td>0.2.0</td>
<td>Martin Lavalle</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRformat</td>
<td>0.2.0</td>
<td>Janna Reps</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRpublish</td>
<td>0.2.0</td>
<td>Chris Knoll</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtemplate</td>
<td>0.2.0</td>
<td>Jamie Gilbert</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Goodhum Rao</td>
<td>CRAN</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Anthony Sena</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Harry Schuemie</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Marc Suchard</td>
<td>CRAN</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Katy Saltwick</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Egil Fridgeirsson</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Martin Schuemie</td>
<td>CRAN</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Janna Reps</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Frank DeFalco</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Anthony Sena</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Martin Schuemie</td>
<td>CRAN</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Anthony Sena</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Martin Schuemie</td>
<td>CRAN</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Janna Reps</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Peter Rijnbeek</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Joel Swerdel</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Joel Swerdel</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Janna Gilbert</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Goodhum Rao</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Jamie Gilbert</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Jenny Reps</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Peter Rijnbeek</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Katy Saltwick</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Marc Suchard</td>
<td>CRAN</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Martin Schuemie</td>
<td>CRAN</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Janna Reps</td>
<td>GitHub</td>
</tr>
<tr>
<td>dataRtransform</td>
<td>0.2.0</td>
<td>Martin Schuemie</td>
<td>CRAN</td>
</tr>
</tbody>
</table>

Curious about the current status of any HADES package? Check out the QR code.

**Open-Source Software**

The open-source tools that empower OHDSI research are not only available to the community, but they are DEVELOPED by the community. We thank the many developers and maintainers who empower our research initiatives around the world!
**Open-Source Software**

**ATLAS**

ATLAS is a free, publicly available, web-based tool developed by the OHDSI community that facilitates the design and execution of analyses on standardized, patient-level, observational data in the OMOP CDM format.

**Enabling A Journey From Data To Evidence**

“ATLAS makes it possible for everyone in the OHDSI community to collaboratively design high-quality observational studies and produce reproducible code that can be shared and executed on OMOP CDM databases around the world.”

- Christopher Knoll

2018 Titan Award for Open-Source Development recipient

Want to learn more about ATLAS?
Experience: atlas-demo.ohdsi.org
Download: github.com/ohdsi/atlas
Read: book.ohdsi.org
Train: academy.ehden.eu

---

**VII. Methods**

**Research**

This graphic is taken from the Suchard et al study, published in The Lancet, that is featured on page 65.
Empirical Calibration

Methodological research is a foundational aspect of OHDSI work. We seek to evaluate the performance of analytics methods so we understand when they can be appropriately applied and how confident we can be in the reliability of the evidence we generate. This research has provided the empirical evidence to allow OHDSI to establish best practices for the design and implementation of population-level effect estimation, as applied for safety surveillance and comparative effectiveness research.

Negative controls – exposure-outcome pairs with no causal relationship – offer a powerful diagnostic to evaluate the reliability of a population-level effect estimation study. By applying the same method on the same data to a large collection of negative controls, one can determine if there is systematic error in the analysis, whether due to selection bias, confounding, or measurement error. Empirical calibration is a statistical procedure developed by OHDSI collaborators to use the error distribution estimated from negative controls and correct the original study statistics – point estimates, confidence intervals, and p-values – to restore their nominal operating characteristics and allow for a more honest interpretation of what really has been learned from observational data.

Legend

LEGEND (Large-scale Evidence Generation and Evaluation across a Network of Databases) applies high-level analytics to perform observational research on hundreds of millions of patient records within OHDSI’s international database network. LEGEND is based on 10 guiding principles that were published in JAMIA (August, 2020) and are listed below.

1. LEGEND will generate evidence at a large scale. Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). Aim: Aims: avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.

2. Dissemination of the evidence will not depend on the estimated effects. All generated evidence is disseminated at once. Aim: Aims: avoids publication bias and enhances transparency.

3. LEGEND will generate evidence using a prespecified analysis design. All analyses, including the research questions that will be answered, will be decided prior to analysis execution. Aim: Aims: avoids P hacking.

4. LEGEND will generate evidence by consistently applying a systematic process across all research questions. This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. Aim: Aims: avoids P hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).

5. LEGEND will generate evidence using best practices. LEGEND answers each question using current best practices, including advance methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. Aim: Aims: minimizes bias.

6. LEGEND will include empirical evaluation through the use of control questions. Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our P values, effect estimates, and confidence intervals using empirical calibration. [7,8] Aim: Aims: enhances transparency on the uncertainty due to residual bias.

7. LEGEND will generate evidence using open-source software that is freely available to all. The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. Aim: Aims: enhances transparency and allows replication.

8. LEGEND will not be used to evaluate new methods. Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). Aim: Aims: avoids bias and improves interpretability.

9. LEGEND will generate evidence across a network of multiple databases. Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. Aim: Aims: enhances generalizability and uncovers potential between-site heterogeneity.

10. LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network. Not sharing data will ensure patient privacy, and comply with local data governance rules. Aim: Aims: privacy.

Journal of the American Medical Informatics Association, 27(8), 2020, 1281–1287
doi: 10.1093/amia/jaz033
Perspective

Principles of Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND)
Martin J. Schuemie &; 1, 2, Patrick B. Ryan &; 3, Nicole Pratt &; 4, RuJun Chen &; 5, Seng Chan You &; 6, Harlen M. Krumholz &; 7, David Madigan &; 8, George Hripcsak &; 9, and Marc A. Suchard &; 10
Methodologies Research

LEGEND in Action

LEGEND (Large-scale Evidence Generation and Evaluation Across a Network of Databases) principles have been applied to studying the effects of treatments for depression, hypertension, and COVID-19, and are being applied to Type 2 diabetes.

The clinical impact of LEGEND has already been observed, with important evidence that promotes better health decisions published in Lancet, JAMA Internal Medicine, and Hypertension.

First-line Thiazide Diuretic Users Experience 15% Fewer Adverse Cardiovascular Outcomes Than ACE Inhibitor Users

The 2017 ACC/AHA guidelines on antihypertensives recommend initiating hypertension (high blood pressure) treatment with prescription medications from any of five drug classes, including both thiazides and ACE inhibitors. Within the LEGEND project, ACE inhibitors produced both worse cardiovascular outcomes and worse side effects than thiazides.

First-line thiazide new-users experienced three major medical outcomes (heart attack, hospitalization for heart failure, and stroke) at an approximate 15% lower event rate than those who began treatment with an ACE inhibitor. Furthermore, among potential side effects associated with first-line hypertensive drugs, ACE inhibitor new-users experienced a higher rate of 19 potential side effects – and a lower rate of 2 – than thiazide diuretic new-users.

In spite of these differences, the majority of patients from this study who initiated treatment were prescribed ACE inhibitors (48%) over thiazides (17%); the results, however, indicate that over 3,100 major cardiovascular events could potentially have been avoided had those approximately 2.4 million ACE inhibitor new-users chosen a thiazide diuretic instead.

Filling The Evidence Gaps

The LEGEND project attempts to fill the evidence gaps in treatment choices that randomized controlled trials (RCTs) leave unanswered,” said lead author Marc A. Suchard, MD, PhD (University of California, Los Angeles). “We were able to compare all antihypertensive drug classes against each other at a massive scale and in a transparent and reproducible manner to study what patients worry about. Heart attack. Stroke. Heart failure. Drug safety. LEGEND synthesizes real-world evidence to determine how different drug classes impact the people who have to choose between them.”

“We did not execute our study to prove one particular drug class was most effective,” Suchard added. “Instead, we used the high-level analytics and best practices developed within OHDSI to study all of these drug classes against each other and openly report on all possible comparisons. Researchers can then interpret specific results in the context of their own research questions.”

The paper also reported that non-dihydropyridine calcium channel blockers proved inferior to the four other first-line antihypertensive drug classes recommended in the 2017 guidelines; other classes included are angiotensin receptor blockers and dihydropyridine calcium channel blockers. A LEGEND-ary Approach To Observational Science

LEGEND is a unique, sophisticated approach to using observational data in a way that is reliable, rich and relevant.” Suchard said. “With the availability of existing health data available, we can start to answer important clinical questions in a reproducible manner.”

The LEGEND Hypertension project used state-of-the-art causal methods to address both observed confounding and residual bias. Covering patients from July 1996 to March 2018, the study filled in evidence gaps that were unavailable for the 2017 ACC/AHA guidelines. The RCTs from those guidelines factored approximately 31,000 users of either thiazide diuretics or ACE inhibitors, far fewer than the approximately 3.2 million new-users available in the LEGEND project.

“LEGEND is a novel approach that could transform the way we use real-world evidence in healthcare,” said senior author Patrick Ryan, PhD, Adjunct Assistant Professor of Biomedical Informatics (Columbia University). “Rather than inefficiently conducting bespoke analyses one-question-one-method-one-database-at-a-time, leaving us vulnerable to various threats to scientific validity, LEGEND provides a systematic framework that can reproduce generate real-world evidence by applying advanced analytics across a network of disparate databases for a wide array of exposures and outcomes.”

Starting On The Most Popular Hypertension Drug Isn’t Most Effective, Per OHDSI’s LEGEND Study

Thiazide diuretics demonstrate better effectiveness and cause fewer side effects than ACE inhibitors as first-line antihypertensives, according to a report published Oct. 24, 2019, in The Lancet. The study factors insurance claim data and electronic health records from 4.9 million patients nine observational studies into the most comprehensive one ever on first-line antihypertensives, and it provides additional context to the 2017 guidelines for high blood pressure treatment developed by the American College of Cardiology (ACC) and American Heart Association (AHA).

Collaborators within the OHDSI network produced the paper “Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis” as part of the collaborative’s ongoing LEGEND (Large-Scale Evidence Generation and Evaluation Across a Network of Databases) project, which applies high-level analytics to perform observational research on hundreds of millions of patient records within OHDSI’s international database network.

OHDSI researchers believe LEGEND will continue to significantly enhance how real-world evidence is used to study important healthcare questions that impact millions of patients worldwide.
The Journey To Reliable Evidence

With Patient-Level Prediction

**Design and Extraction**
- **Study design**: Case-control prone to misclassification and should be avoided; use cohort design.

**Model Development**
- **Learning across datasets**: Models can be learned across datasets while maintaining privacy.

**Model Evaluation**
- **Model usability**: Simple score-based models are easier to apply and can be benchmarked against large-scale models.

**Applications**
- **Bipolar**
- **Mortality**
- **Infection**
- **MI, Stroke**
- **Fall Risk**

**Clear specification of the prediction task:**
- **Target Population**: patients at risk
- **Outcome**: medical event to predict
- **Time-at-risk (TAR)**: interval to predict if outcome will occur

**The patient-level prediction journey is more than just classification...**

**Loss to follow-up**
- Excluding non-outcomes lost to follow-up can bias the data

**Test/Train split**
- The design used to pick hyper-parameters and estimate internal validation matters, even with big p and big n data.

**Feature extraction**
- Feature lockback can make an impact on model performance if it is too short (<180 days)

**Sample size**
- Learning curves provide a way for model developers to determine whether they have sufficiently sized data.

**Network validation**
- The OHDSI network enables large-scale external validation and improves our understanding of models.

---

"Patient-level prediction can make a huge impact on the way we deliver medicine, but a lot more work is needed to ensure quality models are developed. OHDSI is leading research to establish best practices, answering important questions that will ensure future predictive models generate reliable evidence."

- Jenna Reps
2019 Titan Award recipient for Methodological Research
Developing Hundreds of Models
Using the Same Predictors
The OHDSI Patient-Level Prediction team performed a large-scale characterization to learn 67 phenotypes that generally make good ‘history of X’ predictors. These phenotypes can be found in the OHDSI phenotype library and corresponding R package.

Future Work
Research on federated learning, deep learning, Bayesian modelling and more can be implemented using this constrained predictor set.

New Resource
We have a model repository for uploading/downloading prediction models: delphi.ohdsi.org

Benchmark Tasks
We identified a set of 18 clinically useful and diverse prediction tasks to use when implementing large-scale empirical methods research.

Join the monthly PatientLevelPrediction workgroup call on the second Wednesday of each month at 9am ET/3pm CET within our OHDSI Teams environment.
One of the strengths of OHDSI is in its numbers: data from across the OHDSI network can contribute to our understanding of the effects of treatments. One challenge is that we cannot share patient-level data, only summary statistics. In the past, each site would produce an effect estimate (such as a hazard ratio) and its confidence interval, and we would use standard meta-analysis to combine the evidence across sites.

However, OHDSI research has shown that when the outcome is rare this approach can lead to severe bias due to violation of the normality assumption. We developed a new approach, where each site shares the shape of the likelihood curves as a set of points. Combining these shapes avoids the aforementioned bias while still maintaining patient privacy. Since this approach is always as good or better than standard meta-analysis, this is now the default in HADES, and is used in all OHDSI studies.

“Having a large network of databases means we can detect safety issues much earlier, when a single database is too small, but together we have enough power. However, we need to make sure to use the right methods when synthesizing evidence across the network.”

- Martijn Schuemie
2018 Titan Award recipient for Methodological Research
In this chapter, you will see both the depth and wide range of peer-reviewed publications that our community has produced over the last decade. How has OHDSI accomplished so much in so little time? We work together. This graphic highlights just how much our community collaborates to produce high-quality observational research.

Since our community writes many, MANY papers together, this graphic can’t have everybody in the perfect spot. But it clearly shows how the culture of ‘we’ over ‘me’ has powered OHDSI to incredible heights.
The OHDSI Community Dashboard is a tool to help us make forward the progress we are making toward this mission and the collective accomplishments and impact of our community. A goal of the dashboard is to help our community identify how members can see the OHDSI ecosystem as an interconnected system to make a larger impact. We hope you find these tools useful staying up to date with all the activities in OHDSI as well as finding new colleagues in our community to collaborate with. Dashboards are developed to represent the OHDSI community activities.

PubMed Publication Tracking highlights using the OMOP Common Data Model, OHDSI tools, or the OHDSI network. These publications represent scientific accomplishments across areas of data standards, methodological research, open-source development, and clinical applications.

The OHDSI Community Dashboard is a tool to help us make forward the progress we are making toward this mission and the collective accomplishments and impact of our community. A goal of the dashboard is to help our community identify how members can see the OHDSI ecosystem as an interconnected system to make a larger impact. We hope you find these tools useful staying up to date with all the activities in OHDSI as well as finding new colleagues in our community to collaborate with. Dashboards are developed to represent the OHDSI community activities.

PubMed Publication Tracking highlights using the OMOP Common Data Model, OHDSI tools, or the OHDSI network. These publications represent scientific accomplishments across areas of data standards, methodological research, open-source development, and clinical applications.


ODHS Publications


Building Community, One Lego At A Time

The term ‘community’ is defined in the Oxford Dictionary as ‘a feeling of fellowship with others, as a result of sharing common attitudes, interests, and goals.’ Improving health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care — the OHDSI mission — is not a one-person endeavor. It isn’t a one-company, one-country, one-stakeholder, one-discipline, one-anything endeavor. The challenge is too great. The stakes are too high.

Open science is a team effort, and the OHDSI community knows that success can only occur if we come together and build upon each other’s strengths and passions. This message was at the heart of the 2022 Global Symposium closing, as small sets of individual legos were passed out to the hundreds of collaborators who came together. Small sections were built, and they were nice. When all pieced together, they formed something nobody expected beforehand. That is the OHDSI belief: What would be unimaginable alone, we build together.
A favorite part of every OHDSI Symposium is the closing talk, given by Patrick Ryan. Naturally, we figured the appropriate way to end this annual report was a closing letter from Patrick.

Thank you for Joining The Journey with OHDSI!

George Hripcsak opened this year’s Our Journey by discussing the significance of scale and how OHDSI’s focus on reliable evidence generation through large-scale analytics sets us apart from others. It’s clear to me that medicine needs large-scale evidence to bridge the current knowledge gaps and guide better health decisions and better care. It’s also clear to me that this can only be accomplished through large-scale analyses atop large-scale data networks.

What’s most obvious to me is our critical need to continue to innovate on large-scale collaboration.

OHDSI has now grown to more than 3,700 collaborators in 83 countries. The OMOP Common Data Model has become the most widely used open community data standard for real-world evidence generation in the world, with more than 530 data sources converted. OHDSI’s open-source tools have been downloaded more than half a million times from CRAN. Our Global, European, and Asia-Pacific Symposia have brought together hundreds of researchers from all stakeholder communities who share the mission: ‘to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.’ The potential for what we can accomplish together is staggering, but realizing that full potential is extraordinarily hard. Open science collaboration, like what we are building together in OHDSI, is not natural. It’s much easier to focus narrowly on our individual responsibilities, perhaps expanding to the confines of our immediate team. It’s comfortable to understand your role within your local organization—whether you are in academia, government, industry, patient advocacy or another part of the health system—and much tougher to think about how to fit and contribute to a broader multi-disciplinary, cross-stakeholder, international effort.

When we first started OHDSI, I observed that a lot of people were working in their own silos, blissfully unaware of what others were doing. We worked to create a venue where barriers were broken down and trust was built, so that people would have a chance at seeing each other’s silos, albeit from a distance, and could benefit from learning from each other’s experiences. Of course, we really want to get to a place where people step out from their silos and create a new shared space to co-create together. Sharing not just tools, practices, and research, but sharing of yourself: listening to alternative perspectives, offering up your thoughts (even if only half-baked), asking questions, and trying new ideas with a reasonable expectation of failure. We do this with an eye toward contributing to a larger mission that can’t be accomplished alone.

Our community stands as a testament to the power of collaboration in the realm of healthcare. By coming together from diverse disciplines, backgrounds, and corners of the globe, we are pioneering the future of healthcare research, amplifying the impact of observational data, and have the very real opportunity to revolutionize patient care. The harmonization of disparate data sources into a common data model and the development of open-source standardized analytics exemplifies the spirit of collaborative science. We’ve seen firsthand the synergies that arise when statisticians converse with clinicians, when bioinformaticians collaborate with epidemiologists, and when data scientists team up with regulatory scientists. Our challenges are numerous, from ensuring data quality and interoperability to addressing bias and fostering reproducibility. Through large-scale collaboration, we can transform these challenges into opportunities.

Among other things, Jeff Bezos is famously known for the two-pizza rule: that teams should be limited to a size no
larger than can be fed by two pizzas. The general argument being that the smaller the team, the closer the collaboration between its individuals. Smaller teams can be more agile and stay ‘in-the-loop’ informally and focus more of their time on focusing on getting the work done. It’s been argued that one secret to Amazon’s success with the two-pizza team rule has been its scalability, because it’s easier to have a large set of small teams that can be autonomous and find a way to connect together laterally than one large hierarchy that tries to manage the connections top-down. Smaller teams can focus on specific tasks and work quickly, and each member of the team can feel a sense of ownership and accountability toward delivery. The trick is to make sure these lateral connections are as seamless as possible, ideally by establishing specific rules of engagement that allow separate parts to fit together. In software development, we sometimes refer to this as an interface, the point of connection between two components. An effective interface is one that enables continuous development and innovation on both sides, without causing disruption on either side.

At last year’s Global Symposium, I tried to draw the analogy using Legos, where the interface is the studs on top of one Lego brick and tubes on the bottom of another allow Legos to snap together. I argued that for OHDSI’s open-source development activities, we need to think of our HADES packages as Lego bricks that can snap together to enable building study packages that we can run across our data network. But we also need to think of workproducts of our workgroup activities as Lego bricks aimed at building toward something together.

Over the last year, I’ve been proud of the progress we’ve made toward large-scale collaboration. We’ve seen multiple HADES-wide releases that coordinated our open-source software development across our community of maintainers to support more seamless use by our community of researchers. We saw the community come together for the Sisyphus Challenge to design and execute multiple network studies over a few short months, all while generating the educational content to enable future studies to follow in the same footsteps. We’ve seen process improvements to the OHDSI Standardized Vocabularies development that has enabled community contributions from multiple organizations around the world. We’ve also seen more than 100 community contributions into OHDSI’s Phenotype Library, creating a shared resource that should accelerate future research by enabling re-use of standardized components.

This year, we saw the OHDSI community publish its largest network study to date, in terms of the number of data partners represented. “Contextualizing adverse events of special interest to characterize the baseline incidence in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study,” led by Erica Voss and published in eClinicalMedicine, included data across 11 countries and involved 60 co-authors across our community. Within our EHDEN community, there are plans actively underway to conduct a large-scale network study of drug utilization trends over time with aspirations to reach more than 100 databases.

None of these tasks should be considered easy, all of them push people out of their comfort zones. We’re still learning how to share openly – before, during, and after any activity – to avoid reinventing wheels and benefit from others’ contributions. We’re still learning how to develop our technical and organizational interfaces, how to break large problems down into smaller tasks and build them back up into even larger solutions.

While Bezos’ management edict is now folklore, I will argue that it’s a bit underspecified – there is no assertion of where the team needs to get their pizzas from. Earlier in 2023, YouTube personality Airrack partnered up with Pizza Hut to set a Guinness Book of World Records for largest pizza, at 13,990 square feet. According to the Guinness website, “the giant pizza was made using 13,653 pounds (6,192.8kg) of dough, 4,948 pounds (2,244.375kg) of sweet
marinara sauce, over 8,800 pounds (3991.6 kg) of cheese and around 630,496 pieces of pepperoni.” Now that’s large-scale! And perhaps perfect for the large-scale collaboration we need to realize our potential together.

The strength of OHDSI lies not only in our standardized data network, our open-source tools, or our methodological innovations – it lies in our people in our community. Together, we have the power to redefine the landscape of healthcare research, to shape policies, practices, and perceptions, and most importantly, to improve patient outcomes across the globe. As we continue on this journey together, I urge each one of you to embrace the spirit of large-scale collaboration. Remember, every conversation is an opportunity, every partnership a potential paradigm shift, and every collaborative endeavor a step closer to our shared dream of a healthier, brighter future for all.

-Patrick Ryan
How Can You Join The Journey?

Our community has set both the foundation and the highest of standards for global collaboration around observational research. We continue to make real differences in healthcare, and we are doing it through transparent and reproducible science. We also recognize that there is so much more to be done, and so much more that we can do.

If you are inspired by what you read in this book, if you want to learn more about methods research or open-source development, if you have a clinical question you believe needs answering, or if you want to join a community of people dedicated to the team sport of observational health data sciences and informatics, we have a place for you.

How can you get started?

**Join The OHDSI Forums ([forums.ohdsi.org](forums.ohdsi.org))**

Connect with other OHDSI collaborators on our community forums and start discussing how you can help us inform medical decision-making, or simply follow discussions that are interesting to you and learn about the work happening within our global community.

**Join Our Workgroups & MS Teams Environment ([ohdsi.org/ohdsi-workgroups](ohdsi.org/ohdsi-workgroups))**

Our workgroups present opportunities for all community members to find a home for their talents and passions, and a place to make meaningful contributions. We are always looking for new collaborators. Learn more by checking out the workgroups homepage. Our workgroups collaborate inside the OHDSI MS Teams environment; a form to join our Teams environment is available here: [bit.ly/Join-OHDSI-Teams](bit.ly/Join-OHDSI-Teams).

**Join Our Community Calls ([ohdsi.org/community-calls/](ohdsi.org/community-calls/))**

Join collaborators around the world each week during our OHDSI Community Call, held Tuesdays at 11 am ET within our Teams environment. Following weekly updates, we have a variety of call formats, including research presentations, workgroup updates, discussions, tutorials, debates and more. These calls are recorded, and you can access them (as well as the meeting link) at our Community Calls page.

**Continue To Learn About OHDSI**

Learn about OHDSI tools and research processes in a variety of ways.

- The OHDSI website keeps you informed of recent news, publications, upcoming studies and more, while providing all critical links needed to help with your journey: [ohdsi.org](ohdsi.org)

- The Book of OHDSI (which is also translated into both Korean and Chinese) is a community-developed resource with information for every step of your journey: [ohdsi.github.io/TheBookOfOhdsi](ohdsi.github.io/TheBookOfOhdsi)

- Check out the EHDEN Academy, a set of free, on-demand training and development courses. These are open to anybody, but we always encourage new OHDSI collaborators to use this resource to learn about best practices towards our mission of improving health by empowering a community to collaboratively generate evidence that promotes better health decisions and better care: [academy.ehden.eu](academy.ehden.eu)

- Check out the OHDSI YouTube page ([youtube.com/c/OHDSI](youtube.com/c/OHDSI)) for many community-developed learning resources, including tutorials, research presentations and more. Follow OHDSI on both Twitter (@OHDSI) and LinkedIn (OHDSI) to keep updated on community research and follow the #OHDSISocialShowcase to see the research shared at our annual symposia.

**Join The Journey**

Your journey with OHDSI has started. Your interest in our global community is the first step in making a difference in global health. There is no limit to the impact you can make, and you can do so in a supportive, positive and fun environment. We invite you to search our website, post to the forum, join us in Teams, check out our GitHub ([github.com/OHDSI](github.com/OHDSI)), or reach out to us over email ([contact@ohdsi.org](mailto:contact@ohdsi.org)).

Thank you for Joining The Journey with OHDSI!