Our Journey
Where The OHDSI Community Has Been And Where We Are Going
2022 edition
Publication was written and designed by Craig Sachson. Editorial assistance by Patrick Ryan, Kristin Kostka, George Hripcsak, Martijn Schuemie, Marc Suchard, Jody-Ann McLeggon, Jenna Reps, Peter Rijnbeek, Henrik John, Mui Van Zandt and Chungsoo Kim. 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.
<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>Research Flow</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</td>
<td>21</td>
</tr>
<tr>
<td>Community Calls</td>
<td>22</td>
</tr>
<tr>
<td>Study-A-Thons and Other Events</td>
<td>24</td>
</tr>
<tr>
<td>88 Hours: Our COVID-19 Study-A-Thon</td>
<td>26</td>
</tr>
<tr>
<td>The Book of OHDSI</td>
<td>30</td>
</tr>
<tr>
<td>Symposia Around The World</td>
<td>32</td>
</tr>
<tr>
<td>2021 OHDSI Symposium</td>
<td>34</td>
</tr>
<tr>
<td>EHDEN Academy</td>
<td>36</td>
</tr>
<tr>
<td>Darwin EU® Initiative</td>
<td>37</td>
</tr>
<tr>
<td>SNOMED Partnership</td>
<td>38</td>
</tr>
<tr>
<td>HL7 Collaboration</td>
<td>39</td>
</tr>
<tr>
<td>OHDSI and Large Community Initiatives</td>
<td>40</td>
</tr>
<tr>
<td>V. Data Standards</td>
<td>41</td>
</tr>
<tr>
<td>OMOP Common Data Model</td>
<td>42</td>
</tr>
<tr>
<td>Data Partners</td>
<td>44</td>
</tr>
<tr>
<td>Vocabularies</td>
<td>46</td>
</tr>
<tr>
<td>VI. Open-Source Software</td>
<td>49</td>
</tr>
<tr>
<td>HADES Packages</td>
<td>50</td>
</tr>
<tr>
<td>Kheiron Contributor Cohort</td>
<td>51</td>
</tr>
<tr>
<td>HADES Feature/Package Maintainers</td>
<td>53</td>
</tr>
<tr>
<td>ATLAS</td>
<td>54</td>
</tr>
<tr>
<td>VII. Methods Research</td>
<td>55</td>
</tr>
<tr>
<td>Empirical Calibration</td>
<td>56</td>
</tr>
<tr>
<td>Principles of the LEGEND Project</td>
<td>57</td>
</tr>
<tr>
<td>The LEGEND Project</td>
<td>58</td>
</tr>
<tr>
<td>Patient-Level Prediction Models</td>
<td>60</td>
</tr>
<tr>
<td>VIII. Publications</td>
<td>63</td>
</tr>
<tr>
<td>Collaborations Graph</td>
<td>64</td>
</tr>
<tr>
<td>Community Dashboards</td>
<td>66</td>
</tr>
<tr>
<td>OHDSI Publications (2010 - September 2022)</td>
<td>67</td>
</tr>
<tr>
<td>IX. Join The Journey</td>
<td>89</td>
</tr>
<tr>
<td>Cheers From Our Global Community</td>
<td>90</td>
</tr>
<tr>
<td>Closing Letter</td>
<td>91</td>
</tr>
<tr>
<td>How Can You Join The Journey?..... Inside Back Cover</td>
<td></td>
</tr>
</tbody>
</table>
Welcome to the second 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, which we repeat often, 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 approximately 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 evidence in health care.

We have turned that system on and begun to produce impactful evidence. We have gotten to the point that policy makers like the European Medicines Agency and medical influencers as represented in journals like Circulation, Hypertension, JAMA, Lancet, and BMJ have specifically called out OHDSI and its studies and have demonstrated OHDSI’s effect on hundreds of millions of people. We’ve proven what is possible and helped answer some important questions, but we are still just at the beginning of our journey.
With a systematic approach to observational research and an extensive international data network, we can contribute much more broadly across medicine.

How have we gotten here, and how can we continue to contribute?

**Leadership is the foundation of any initiative, and OHDSI is blessed with many leaders.** OHDSI has done a good job in finding leaders rather than bosses. The mark of a successful group is that it seems to lead itself (see, for example, the Tao of Leadership). Leaders inspire, set examples, and give credit. Put another way, quoting informatician Paul Clayton, we encourage leaders who are nice, bright, and hardworking. **Our leaders are a diverse group, pulling from nations around the world; pulling from industry, academia, and government; and pulling from all career stages from undergraduate to senior figures.**

To foster our open-science community and to improve transparency, **OHDSI has shifted most of its work from informal arrangements to formal workgroups.** OHDSI workgroup leaders have donated their time, creativity, and skills to advance observational research, and OHDSI has more consciously moved to support them through workgroup leadership summits and workshops. Thanks to Paul Nagy and colleagues for being leaders among leaders.

OHDSI recognizes leadership through its annual Titan Awards, which go to those in the community who have contributed greatly in the previous year. The awards include a specific award for Community Leadership, but all our Titan winners are leaders, and our leadership bench includes many more than we can give awards to.
Welcome Letter to the Community

To keep OHDSI healthy, we must expand our leadership as OHDSI grows. OHDSI is therefore looking to find and grow new leaders, especially from its junior ranks. We are an open-science initiative and everyone is welcome.

Leadership also refers to OHDSI’s mission-driven responsibility in the field, which is to pull research to be more rigorous and trusted. I use “pull” on purpose, as research must be pulled forward. You pull towards yourself and push away from yourself, and OHDSI must set the standard for rigorous research, encouraging the broader research community to do the same.

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.
Observational 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 2014, OHDSI is a growing collaborative of more than 3,200 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 80 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 928 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...
with pre-specified protocol and analysis code which can be executed across the OHDSI data network, allowing aggregate summary statistics (but no patient-level data) to be shared and collectively interpreted and disseminated.

**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!**
III.

OHDSI Collaborators
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.

OHDSI By The Numbers

- 3,266 collaborators
- 80 countries
- 21 time zones
- 6 continents
- 1 community
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 Collaborators

Organizations Involved With OHDSI

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 400 organizations listed below, and looks forward to other organizations joining the journey as well.

2Ca-Braga • Aarhus University • AbbVie • Advocate Aurora Health • Agenzia Di Tutela Della Salute Della Provincia Di Bergamo • Ajou University Hospital • Akrivia Health • All Of Us Research Program • Allscripts • AMC Medical Research BV • Amgen • Andrija Štampar School Of Public Health • APDP Diabetes Portugal • Arcadia Inc • ARS Toscana • Asan Hospital • ASCO CancerLinQ • Asociación Instituto De Investigación Sanitaria Biocruces Bizkaia • Assistance Publique - Hopitaux De Paris / Aphp • Assistance Publique Hopitaux De Marseille • Astellas Pharma • AstraZeneca • ASU • AU-EPBRN • AUS Dept of Veterans Affairs • AWS • Az Delta Vzw • Az Klin • Azienda Ospedaliera Nazionale Ss. Antonio E Biagio E Cesare Arigo Alessandria • Azienda Ospedaliera Universitaria (Aou) Di Modena • Azienda Ospedaliera Universitaria Integrita Verona • Azienda Unità Sanitaria Locale-Ircses In Reggio Emilia • B2I Healthcare • Barts Health NHS Trust • Bayer AG • BCB Medical Oy • Beijing Safe House • Ben-Gurion University • Berlin Institute of Health • Bill & Melinda Gates Foundation • Boehringer Ingelheim • Booz Allen Hamilton • Bordeaux Hospital • Boston Medical Center • Bradford Teaching Hospitals NHS Foundation Trust • Brazilian MOH • Brown University • Buccehon Hospital • Buddhmed Technologies • Caliber • Cancerdatanet Gmbh • Carilion Clinic • Carnegie Melon in Qatar • Case Western CICB • Catholic University of Korea Seoul St. Mary’s Hospital • Catholic University of Korea Yeouido St. Mary’s Hospital • CDPHP • CEEISCAT (Catalonia) • Cegeedem Health Data • Centre Hospitalier Universitaire De Lille • Centre Hospitalier Universitaire De Toulouse • Cerner • Cha University Bundang Medical Center • Charité - Universitätsmedizin Berlin • CHCO (USA) • Cherokee Health Systems • Children’s National • CHLA (USA) • Chonnam National University Hospital • CHOP (USA) • CHU Montpellier • Clinica Alemana de Santiago • Clinical Center of Serbia • Clinical Centre of Nis • Cognizant • Columbia University • Columbia University Irving Medical Center • CRHFEI • CSS Denmark • Daegu Catholic University Hospital • Data Integration Centre University Hospital Carl Gustav Carus Dresden • data4life • Databricks • Datasus Ambulatory • DFCI • DHS Los Angeles • DNAnexus • Dongguk University Ilsan Hospital • Dresden University Of Technology • DRG • Drug Safety Research Unit • Duke • Eau Claire Cooperative Health Center • EBM (EU) • EGCU • EISBM (Europe) • Eli Lilly & Company • Ephir Inc. • Erasmus MC • European Medicines Agency • Evidera • Evidnet • Ewha Womans University Mokdong Hospital • FIBH12O • FinnGen • Flatiron • Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico • Fondazione IRCCS Istituto Nazionale Dei Tumori • Fondazione IRCCS Policlinico “San Matteo” • Fondazione Poliambulanza • Fred Hutch Cancer Center • Freyr Ltd • Fudan University • Fujitsu • Fundacio Institut D’Inves72tiagas Mèdiques • Fundación Rioja Salud • FUS • GA4GH • Gacheon Gil Hospital • Galilee Medical Center • Gangbuk Samsung Hospital • Gangdong Sacred Heart Hospital • Gangnam Severance Hospital • Geisinger • General Hospital Of Kavala • Georgetown/MedStar Health • Getrude’s Children Hospital • Glmed Learning Health • Google • Great Ormond Street Hospital NHS Foundation Trust • GlaxoSmithKline • Georgia Tech Research Institute • George Washington University • Hanover Medical School (Germany) • Hanyang University Hospital • Harvard • Harvey Walsh Ltd • Hasselt University • HealthVerity • Hebe Mental Health Center • Helix • Helsinki UH CCC Hematology • Hierarchia D.O.O. • On Behalf Of University Hospital Centre Zagreb • Health Insurance Review and Assessment Service • HL7 • HM Hospitals • HMAR • Hospital District Of Southwest Finland (Varsinai-Suomen Sairaanhoitopiiri) • Hulafe (Spain) • Hus Datalake Ecareforme Poc • Hwasun Chonnam National University Hospital • IBM T.J. Watson Research Center • Indiana University School Of Medicine At Mount Sinai • ICON • ICVS (Portugal) • IDIAP/JGOL / SIDIAP • Idival • IMASIS • Imperial College Of Science Technology And Medicine • Incheon Sejong Hospital • Indian Society for Clinical Research • Indiana University
School Of Medicine • Inha University Hospital • Innovative Medical Research SA • Inova Health • Institute of Applied Biosciences • Int’l Uni of Health And Welfare • Integraal Kankercentrum Nederland • Intermountain Healthcare • IQVIA • IRST (Italy) • Istanbul Universitesi • Istanbul University-Cerrahpasa • Janssen R&D • Janssen Scientific Affairs • Jayne Koskinas Ted Giovanis Foundation • Jiangxi Province • Johns Hopkins University • Johnson & Johnson • Juntendo Uni SOM • Kangwon National University Hospital • Karolinska Institutet • Keck Medicine (USC) • Khoo Teck Puat Hospital • KI Research Institute • King Saud University Medical City • King’s College London • Kliničko-Bolnički Centar Zvezdara • Knight Cancer Institute • Konkuk University Medical Hospital • Konyang University Hospital • Korea Advanced Inst of Sci and Tech • Korea University Anam Hospital • Korea University Ansan Hospital • Korea University Guro Hospital • Kyoto University • Kyunghee University Hospital • Kyunghee Medical Center • Kyungpook National University Hospital • Kyushu University Hospital, Japan • Leeds Teaching Hospitals NHS Trust • Leiden MC • LIH (Luxembourg) • Loyola University (NOLA) • LTS Computing LLC • Lundbeck • Lynxcare Clinical Informatics NV • M2GEN • MaineHealth • Marina Salud S.A. • Mass General Brigham • Mayo Clinic • MDV (Japan) • Medaman BV • mederrata • Medibloc • Merck • Microsoft • MIT • MITRE • Momentum AD • Montefiore/AECOM • MS Urban Research Center • MSFP-gGmbH • MSKCC • MSU (MT) • MTPPI • MUVienna • MUSC / HSSC • Myongji Hospital • Nanfang Hospital • National Cancer Center • National Cancer Hospital East • National Health Insurance Corporation Ilsan Hospital • National Institute of Public Health (Japan) • National University Hospital of Hospital (SG_NUH) • NCQA • Nemours • NHIRD • NICE • Northshore • Northwell Health • Northwestern Med • Novartis • Novo Nordisk Inc. • NYU Langone • Odysseus Data Services • OHSU • Okayama University • Oklahoma U • Optimum Patient Care Limited • OSU Medical Center • Outcomes Insights • Oxford • Pareto Intelligence • Paxata • Pedianet • PEDSnet • Peking Union Medical College Hospital • Penn State • PhysioNet • Pirkammaa Hospital District • Plateforme De Données De Santé • Policlinico San Donato S.P.A. • Portuguese Institute of Oncology of Porto • Premier HealthCare • PSMAR (Barcelona) • PSSJD • Pusan National University Hospital • Queen Mary University Of London • RCGP (UK) • Regeneron • Regenstrief Institute • Reliant Medical Group • Roche • Rush UMC • Rutgers • RWJ Barnabas • Sage Bionetworks • SAIL Databank • Samsung Seoul Hospital • Sanford Health • Sanofi • Saudi FDA • SBU (USA) • Semantix • Semmelweis Egyetem • Seoul National University Bundang Hospital • Seoul National University Hospital • SERSMAS & FIIBAP • Severance Hospital • Shuanghe Hospital • Siemens Health Services • SIMG (Italy) • SNOMED CT • Snowflake • Soochunhyang University Hospital • Spectrum Health • Spok • St. Luke’s (Idaho) • Stanford University • Stichting Integraal Kankercentrum Nederland • STIZON • Sydney LHD • Taipei Medical University Affiliated Hospital • Taipei Municipal Wanfang Hospital • Takeda • Technical University Sofia • The Hyve • The Roux Institute at Northeastern • The University Court Of The University Of Edinburgh • Tokyo University • Tianjin Anding Hospital • tranSMART • TrialSpark • Tufts • Tulane • U Copenhagen • U Dundee • U Gothenburg • U Hong Kong • U IL Chicago • U Minho • U São Paulo Medical School • U South Australia • U Tartu • U Tsukuba • U Utah • U Witwatersrand • UA-Birmingham • UAkansas • UBuffalo • UCHealth • UCHealth-Anschutz Medical Campus • UCalgary • UChicago • UCinncinati • UCL (UK) • UCLA • UCSF • UFHealth • UH Geneva • UHG (USA) • UIO • University of Iowa • UK Biobank • UK-CRIS • UKentucky • UKER • Ulsan University Hospital • U Mass Memorial MC • UMC New Orleans • UMessina • University of Miami • University of Michigan • UMHealth School of Dentistry • University of Minnesota • University of Mississippi MC • UNC Chapel Hill • Unidade Local De Saúde De Matoiosinhos Epe • Université De Bordeaux • Université De Genève • University College London Hospitals NHS Foundation Trust • University of Pécs • UNMC • UNew Mexico • UNSW Medicine Australia • UPennsylvania • UPittsburgh • URochester • US Department of Veterans Affairs • US Department of Defense • US Food & Drug Administration • US National Cancer Institute • US National Institutes of Health • US National Library of Medicine • USAID • USC (LA) • UTexas-Austin • UTexas-Houston • UTHCS-Houston • UTMC • UVirginia • UW Washington (Seattle) • UW Madison • Vall D’Hebrón Hospital Campus • Vanderbilt • VCU • Veradigm • Vertex • Vivante Health Software • Vrije Universiteit Amsterdam • Wake Forest • Wanfang Hospital • Washington University • WashU St Louis • Weill Cornell Medical Center • WHO Uppsala Monitoring Centre • Winship Cancer Institute of Emory University • WMichigan USOM • Wonju Severance Hospital • Wonkwang University Hospital • WVU • Yale • Yongin Severance Hospital • Yonsei University • ZO (Belgium) • ZS Associates
There is a lot of energy and good will in the community. It is open, inclusive, and extremely diverse. I get my energy from the community. The enthusiasm and drive to do the right thing and to improve human lives and the possibility to work with such a diverse group of individuals and learn from them excites me and pushes me to do more. I am extremely proud to be a part of this effort.

Asieh Golozar  
*VP, Global Head of Data Science at Odysseus Data Services, Inc. and Professor of the Practice & Director of Clinical Research at the OHDSI Center, Northeastern University*

The values, goals, and people of OHDSI I find incredibly inspiring. The people in OHDSI are highly talented, practical, and collegial. They are always welcoming new people on the journey. OHDSI is a multidisciplinary community pushing the boundaries of computational observational research so there is always something new to learn. The values of being transparent in our methods and in our software resonate with me in trying to make a lasting impact. I get energized from going to OHDSI meetings because I always meet new people, learn new things, and am part of making a difference in improving healthcare.

Paul Nagy  
*Program Director for Graduate Training in Biomedical Informatics and Data Science and Deputy Director of the Johns Hopkins Medicine Technology Innovation Center*

I often sit in those teleconferences (often late at night or in the wee hours of the morning here in Australia) and marvel at the discoveries and new insights that are shared so openly, the light bulb moments that lead to fabulous discoveries! The work that has been generated in LEGEND and EUMAEUS is important clinically. It can help update clinical guidelines and provide robust evidence for medicine regulators – but for me these landmark studies have also provided critical insights into which methodologies are appropriate under which conditions – especially the value of empirical calibration!

Nicole Pratt  
*Professor at University of South Australia*

There is a saying in Chinese — *酒香不怕巷子深* — which could be translated into ‘Good wine needs no bush’. I really think the reason is because OHDSI has great methodologies, tools, and community support, which are making more and more people adopt OMOP and join the community.

Jing Li  
*Associate Director of Data Science at IQVIA*
Although 2020 was a challenging year for most, due to COVID-19, I will remember it for another significant experience as well: the beginning of my collaboration with the OHDSI community. I had the pleasure to participate in scientific research, symposiums, and other collaborative ventures and meet several Titans of the community in person.

ETHONs are a perfect example of how the collective effort of motivated and proactive people can contribute to a greater cause. Thanks to ETHONs and the cooperation with the University Clinical Center of Serbia, I had the opportunity to get better acquainted with the OHDSI toolset and the open-source community around it. In addition, observational studies in which I participated gave me and my colleagues valuable insights and new perspectives for future projects. Through efforts like these, OHDSI enables underrepresented populations to be exposed and contribute to scientific and medical evidence.

After a short time in the community, I felt really comfortable within it and I think it’s very easy to advance in it. OHDSI is a community of visionary, forward-looking and persistent people, and their efforts motivate me to improve my skills and contribute more to the community.

**Filip Maljković**
*Lead Programmer at Heliant*

In OHDSI I have found an interdisciplinary community dedicated to making the best use of routinely collected healthcare data to find better treatments for diseases that have greatly affected some of my closest friends and family members. My journey in OHDSI has taken me from being an inspired spectator to co-leading the Open Source Community workgroup with Paul Nagy. Together we are growing OHDSI’s open-source software ecosystem and the next generation of OHDSI developers.

Contributing to OHDSI open-source software has been a longtime goal for me and now that it’s part of my day job I can say that contributing to open source is more rewarding than I ever thought it could be. The problem with OHDSI is that there are too many amazing projects and people to keep up with, and that’s a good problem to have.

**Adam Black**
*Data Scientist at Odysseus Data Services, Inc.*

**Aniek Markus**
*PhD Candidate, Erasmus MC*
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 will be awarded for a fifth consecutive year at the 2022 Symposium.

Annually, community members are invited to 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 symposium.

The award categories, as well as all 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

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 Ostropolets, Columbia University; Dmitry Dymshyts, Odysseus Data Services)

Methodological 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

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
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

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

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

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

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

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

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.

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
IV.

Collaborative Activities
**OHDSI Workgroups**

OHDSI’s central mission is to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care. We work towards that goal in the areas of data standards, methodological research, open-source analytics development, and clinical applications.

There are currently 27 active working groups that present opportunities for all community members to find a home for their talents and passions, and to make meaningful contributions. We are always looking for new collaborators.

See an area where you want to contribute? Please **Join The Journey!**

<table>
<thead>
<tr>
<th>Collaborative Activities</th>
<th>OHDSI Workgroups</th>
<th>COllabOraTiVe aCTiViTies</th>
</tr>
</thead>
</table>
| **APAC (Asia-Pacific)**  | Current Participants: 289  
Lead: Mui Van Zandt | **Common Data Model**  
Current Participants: 596  
Lead: Clair Blacketer |
| **ATLAS/WebAPI**         | Current Participants: 226  
Lead: Anthony Sena | **Data Quality Dashboard Development**  
Current Participants: 260  
Lead: Clair Blacketer |
| **Clinical Trials**      | Current Participants: 252  
Leads: Mike Hamidi, Lin Zhen |
| **Education**            | Current Participants: 116  
Lead: Nigel Hughes | **Eye Care & Vision Research**  
Current Participants: 40  
Leads: Sally Baxter, Kerry Goetz |
| **Geographic Information System (GIS)** | Current Participants: 122  
Leads: Robert Miller, Andrew Williams | **HADES (Health Analytics Data-to-Evidence Suite)**  
Current Participants: 262  
Lead: Martijn Schuemie |
| **Healthcare Systems**   | Current Participants: 430  
Lead: Melanie Philofsky | **Health Equity**  
Current Participants: 201  
Lead: Jake Gillberg |
| **Latin America**        | Current Participants: 48  
Lead: Jose Posada | **Medical Devices**  
Current Participants: 130  
Leads: Vojtech Huser, Asiyah Lin |
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.

Want to learn more? Check out our homepage: ohdsi.org/ohdsi-workgroups

**Medical Imaging**
- Current Participants: 114
- Leads: Paul Nagy, Seng Chan You

**Natural Language Processing**
- Current Participants: 379
- Lead: Hua Xa

**Oncology**
- Current Participants: 241
- Lead: Asieh Golozar

**Open-Source Community**
- Current Participants: 118
- Leads: Adam Black, Paul Nagy

**Patient-Level Prediction**
- Current Participants: 355
- Leads: Jenna Reps, Ross Williams

**Phenotype Development & Evaluation**
- Current Participants: 249
- Lead: Gowtham Rao

**Population-Level Effect Estimation**
- Current Participants: 355
- Leads: Martijn Schuemie, Marc Suchard

**Psychiatry**
- Current Participants: 115
- Leads: Dmitry Dymshyts, Andrew Williams

**Registry**
- Current Participants: 115
- Lead: Tina Parciak

**Surgery and Perioperative Medicine**
- Current Participants: 37
- Lead: Evan Minty

**Vaccine Vocabulary**
- Current Participants: 76
- Lead: Adam Black

**Steering Group**
- Current Participants: 70
- Lead: Patrick Ryan

**OHDSI Regional Chapters**
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.

- **Africa**
  - Current Participants: 66
  - Lead: Nega Gebreyesus

- **Australia**
  - Current Participants: 74
  - Lead: Nicole Pratt

- **China**
  - Current Participants: 228
  - Lead: Hua Xu

- **Europe**
  - Current Participants: 321
  - Lead: Peter Rijnbeek

- **India**
  - Current Participants: new
  - Lead: Swetha Kiranmayi Jakkova

- **Japan**
  - Current Participants: 49
  - Lead: Tatsuo Hiramatsu

- **Korea**
  - Current Participants: 55
  - Lead: Seng Chan You

- **Singapore**
  - Current Participants: 58
  - Lead: Mengling Feng

- **Taiwan**
  - Current Participants: 71
  - Lead: Jason Hsu
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, and plenty more. Our weekly calls take place on Tuesdays at 11 am ET and are led by Craig Sachson, and they are both recorded and posted to both OHDSI.org and within our Teams environment.

These pages highlight just a few of the meeting topics from 2022; please check out ohdsi.org/community-calls to learn more about these interactive community gatherings.
Collaborative Activities

May 24: Network Studies

- IBD characterization
  Presenter: Chen Yongbo

- Characterization of Health by OHDSI Asia-Pacific chapter to identify Temporal Effect of the Pandemic (CHAPTER) Study
  Presenter: Song Chao Nuo

- Applying the Decentralized Generalized Linear Mixed Effects Model (dGEM) for Hospital Profiling of COVID-19 Mortality Data across OHDSI Network
  Presenter: Mike Tong

- Real world safety of treatments for multiple sclerosis
  Presenter: Nihar Patel

- Comparison of mortality, morbidity & healthcare resources utilization between patients with and without a diagnosis of COVID-19
  Presenter: Nemanja Vukic

- Quality assessment of CDM databases across the OHDSI-AP network
  Presenter: Fangqian Lin

June 14: OHDSI Publications in 2022

- Learning patient-level prediction models across multiple healthcare databases: evaluation of ensembles for increased model transportability
  Presenter: Jiajia He

- Analysis of Dual Combination Therapies Used in Treatment of Hypertension in a Multinational Cohort
  Presenter: Xin Li

- Factors influencing background incidence Rate Calculation: Systematic Empirical Evaluation Across an International Network of Observational Databases
  Presenter: Anvis Dhupar

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

- Prior-Preconditioned Conjugate Gradient Method for Accelerated Gibbs Sampling in “large n, Large p” Bayesian Sparse Regression
  Presenter: Ali Malek

June 21: 10-Minute Tutorials

- PheValuator
  Presenter: Joel Sweedel • Associate Director, Johnson & Johnson

- PheKnowLater
  Presenter: Tiffany Callahan • Postdoctoral Research Fellow, Columbia Univ.

- Patient-Level Prediction
  Presenter: Jiajia He • Associate Director, Johnson & Johnson

- CAPR
  Presenter: Martin Lavallee • PhD Student, Virginia Commonwealth Univ.

July 19 Community Call: Workgroup Updates

- Early-Stage Researchers
  Faazia Ashraf
  Undergraduate Student • UCLA

- Phenotype Development & Evaluation
  Gowtham Rao
  Senior Director • Johnson and Johnson

- NLP
  Hu Xia
  Associate Dean for Innovation, Professor • Univ. of Texas Health Science Center

- HADES
  Martijn Schuermie
  Research Fellow, Epidemiology Analytics • Janssen R&D

Aug. 2 Community Call: Building Organizational Support Within Your Community

- Greg Klebanov
  CodeTR • Cyberspace Data Services, Inc.

- Keran Moll
  Director, HIRer Real World Data & Analytics Research • Regenstrief

- Ajit Donde
  Senior Manager, Center for Observational Research • AMGEN

- Paul Nagy
  Program Director for Graduate Training in Biomedical Informatics and Data Science • Johns Hopkins University

Aug. 9: Asia-Pacific (APAC) Regional Updates

- Jason Hsu
  Taiwan Chapter

- Nicole Pratt
  Australia Chapter

- Seng Chan You
  Vietnam Chapter

- Mengling Feng
  Singapore Chapter

- Tatsuo Hiramatsu
  Japan Chapter

- Lei Liu
  China Chapter

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
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.
2018 Oxford study-a-thon, which focused on knee replacement surgery

2022 Center for Surgical Science analyses in colorectal cancer, held in Kusadasi, Turkey

2022 Korea Datathon, which focused on several research questions
The time was originally meant for highlighting OHDSI capabilities, not testing them.

The hours were meant for sharing global research, not sharing in global research.

The Observational Health Data Sciences and Informatics (OHDSI) community held a COVID-19 global, virtual study-a-thon March 26-29, 2020, believing that a network of people who valued both collaboration and open science could make a meaningful impact on the current global pandemic.

How? Nobody was quite sure in the moment, but they were confident they would figure it out.

“We chose an ambitious path and relied on our community and infrastructure to lead the way,” said Patrick Ryan, Vice President of Observational Health Data Analytics at Janssen Research and Development. “In simple terms, efforts within our community over the past 88 months set the foundation for OHDSI’s most important and impactful 88 hours.”

The OHDSI community, by definition, is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. In plainer terms, it’s a community of people who volunteer their time and talents for the shared goal of improving healthcare through observational research.

A global network of OHDSI colleagues planned to celebrate recent research initiatives and discuss future efforts during the annual European Symposium at Oxford University in late March of 2020. The symposium was canceled due to the rapidly spreading COVID-19 virus; in its place, the organizing committee planned a study-a-thon, which OHDSI has experienced significant success with several times over.

The twists?

The COVID-19 data was limited (a significant issue for an observational data science network), the needs were immediate, and everybody was staying home.

Those factors would be a hard stop to most, but the virtual OHDSI community has thrived on overcoming obstacles, and there was never a more crucial time to do so again.

Daniel Prieto-Alhambra, Professor of Pharmacoe- and Device Epidemiology at Oxford, remembers his OHDSI conversion occurring during one of the afore-mentioned study-a-thon events in 2018. This one had nothing to do with viruses or antibodies; it was about the safety profile of varied knee replacement procedures and ultimately produced a paper published in Lancet Rheumatology.

While that data didn’t affect COVID studies 15 months later, the impact of the event stayed with Prieto-Alhambra. He presented on it during the 2019 U.S. Symposium, led another one in Barcelona to focus on rheumatoid arthritis and volunteered to host the global community for the 2020 European Symposium.

“We were thrilled to bring the OHDSI community to Oxford, and we were excited about some new aspects, including new tutorials,” Prieto-Alhambra said. “It was crushing to cancel it in the moment, but we quickly looked ahead and saw an opportunity to make the most of our time and talents. From that moment, we never looked back.”

88 hours.

That was the time between the global kickoff and closing calls, both of which have combined for more than 2,300 views on YouTube (the entire set of calls and presentations is available at the OHDSI COVID-19 research page). More than 330 people from at least 30 nations registered to collaborate in the event, offering their services in areas like literature review, protocol development, study execution, etc.

Peter Rijnbeek, Associate Professor Health Data Science at the Erasmus University Medical Center in the Netherlands, has a history of bringing together leaders in observational health data science. He hosted the 2019 OHDSI European Symposium, and is leading the recently created EHDEN consortium, which is building a large-scale, federated network of European data sources for the discovery and generation of real-world evidence.

He took a leadership role once again; his Erasmus team set up the Microsoft Teams virtual platform and created 17 different teams that held varied roles throughout the event.

This setup, for example, enabled a group focused on phenotype development to work collaboratively, while also having the ability to connect with teams inside the characterization, estimation and prediction groups as well. When needed, there were support teams for literature review, data support, study design and more.

Your standard study-a-thon might just send various groups to different areas within a shared space. During these 88 hours, that ‘space’ might have had collaborators from both hemispheres working simultaneously at different points of
a 13-hour time period. From breakfast in one part of the world to dinner in another, determined volunteers didn’t stop working together to seek answers during a global crisis.

“OHDSI has always been about people working together to solve common goals, and I am proud our team helped to make this event possible,” Rijnbeek said. “We brought the OHDSI world to Erasmus MC in person last year, but it was even more important to bring them together virtually right now.”

88 hours.

It is unrealistic to think OHDSI’s monumental goals could be accomplished in such a limited time. Early work needed to be done to develop an infrastructure for both the meetings and the OHDSI technical platforms, which happened mainly due to the sustained efforts of Lee Evans, Anthony Sena, and James Wiggins. Beyond that, many of the prioritized questions that would become the primary focus of the four days were determined beforehand.

Community involvement was sought in suggesting such questions, but a group that truly believes in collaborative open science knew this was a time to reach outside the circle. Stakeholders around the world reached out to national governments, public health agencies, and health-related institutions to learn what the most critical questions were right now. That feedback, as well as a literature review process that began days before the study-a-thon, helped the core team provide a framework for the four days.

There was a clear desire to create a multi-nation characterization study of COVID-positive patients, even if the data size was more limited at the moment. Facing overwhelmed medical facilities globally were becoming abundantly clear.

Preliminary work with data was necessary as well. Christian Reich led the vocabulary team to develop COVID-related updates on the standardized vocabularies, while Kristin Kostka and Greg Klebanov were among many collaborators working with different sites on either data conversion or analysis support. Seng Chan You and Rae Woong Park collaborated with the South Korean HIRA, which worked with OHDSI to run packages against a more robust set of COVID data than anywhere in the United States. A handful of American institutions, including Columbia and Stanford, signed on to provide deidentified COVID data as well.

“The data owners chose to donate their data for use in these critical studies simply because they want to help,” Kostka said. “They share our belief in the power of the OHDSI community, and because of that trust, we are able to generate the world’s largest observational studies to help inform decision-making in this major public health issue. I think that’s the coolest thing imaginable, and I am so proud to be part of this effort.”

Laying the groundwork was the necessary warmup for the sprint that was to come — and the marathon that would follow.

88 hours.

It began Thursday, March 26, at 7 am in Oxford, as Prieto-Alhambra welcomed an international community of people

continued on next page
to this unique and critical initiative. A panel including Ryan, Rijnbeek and George Hripcsak — chair of the Department of Biomedical Informatics at Columbia University, the coordinating center for OHDSI — discussed the long journey from the formation of OHDSI to this moment, and what they believed could be accomplished over four days.

Subgroup calls immediately followed to set the course for their respective work plans. Teams within characterization, estimation and prediction studies discussed study questions, varied responsibilities, and timetables over the four days; those timetables were dependent on the phenotype group, which had to develop standard cohorts that could be used within all studies.

It was the ultimate team environment.

And the clock was now ticking.

88 hours.

Leadership from institutions including Oxford, Erasmus, Columbia, UCLA, Ajou University, Janssen Research and Development, and IQVIA helped put this event in motion, but OHDSI empowers collaborators at different stages of their own journey to make important contributions.

Jennifer Lane, an orthopedic surgeon pursuing her PhD at Oxford, led the literature review efforts and co-authored the manuscript for the largest safety profile on hydroxychloroquine ever executed. Ed Burn, a recent PhD graduate from Oxford, led the characterization team; he had also served as lead author for the Lancet Rheumatology paper on knee replacement.

Ross Williams, Cynthia Yang and Aniek Markus are each PhD students at Erasmus, and they worked on co-authoring a prediction study that could help critical hospitalization and triage decisions healthcare workers are making daily.

Anna Ostropolets, a PhD student at Columbia, shared in the leadership of the phenotype team and presented on the 114 validated & reviewed cohorts developed and distributed by the team during the closing call.

Many others within academia contributed to the initiative, while global stakeholders from both industry and healthcare agencies provided critical efforts, ranging from protocol design to data support.

“The OHDSI community has an open approach to everything,” said Lane, co-lead author of the hydroxychloroquine study, which would eventually be published by Lancet Rheumatology but made its immediate impact as a preprint. “It is based upon clear communication, that all contributions are valuable. Everyone is playing to their strengths, which means that the combined effort is precise in many areas that would be incredibly difficult or impossible within one research group or institution. I have met people who will shape the way I work in the future, both through their leadership and their willingness to help me learn novel research approaches.”

Many registrants were newcomers to the OHDSI process who found the idea of a COVID-19 study-a-thon either inspirational and interesting. Their contributions may have been more limited than others over the 88 hours. Some from that group quickly found their footing in the community afterwards and joined studies either developed or brainstormed over the four days.
Each person who takes that step strengthens the community.

88 hours.
You’ve seen that number before? OK, here are a few new ones.

Between the 12 global huddles, there were more than 100 collaborator calls and 13,000 chat messages over 17 concurrent channels (different teams). More than 10,000 publications were reviewed and 355 cohort definitions were defined to lead to the drafting of nine protocols and the release of 13 study packages.

“The real-world evidence we are generating to inform decision-making in this pandemic is the most important thing to come from these four days,” Ryan said. “Reflecting on what a community of volunteers achieved in this collaborative setting is humbling. We had a shared goal that mattered to everybody, but OHDSI has a way of attracting good people that you enjoy being around. I don’t take that for granted. The people that make up our community are our greatest strength.”

It’s easy to have that positive feeling on Day 1, or as you reach the close, but to have it in the middle of a four-day marathon is a testament to the energy created organically. The Friday night chat messages and Saturday morning team calls mattered — in that short a time, it all matters — and maintaining focus and enthusiasm powered the process from start to finish.

The 88th hour.
The global closing call was broadcast live to a global audience and provided a series of presentations about how OHDSI arrived at this moment. It was an opportunity to celebrate shared efforts, announce study designs and preliminary findings, and plan for the future.

When Prieto-Alhambra signed off for the final time, COVID-19 did not go away.

OHDSI won’t either.
The efforts continued immediately. As protocols continue to be designed or improved, data partners work to run studies and generate evidence. The first manuscript was submitted for peer review two weeks after the final signoff, and more followed.

Generating real-world evidence to improve healthcare has been the OHDSI mission since it officially formed in 2014. This has been a passion project for a global community that expands in both people and analytic capability each year.

Nobody saw this moment coming. But it did, and OHDSI was more ready for it than even the most optimistic collaborator could have imagined.

There were critical discoveries in the first six years, and there are many more to come — including some that will aid global efforts against COVID-19 in the near future.

But those 88 hours stand as a defining moment for OHDSI, and they are a glimpse of this community’s potential on the journey ahead.

by Craig Sachson
published April 17, 2020
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 validity 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).
# Join The Journey OHDSI.org

## 1 The OHDSI Community

1.1 The Journey from Data to Evidence ....... 3
1.2 Observational Medical Outcomes Partnership .... 5
1.3 OHDSI as an Open-Science Collaborative .... 6
1.4 OHDSI’s Progress ....... 7
1.5 Collaborating in OHDSI ....... 9
1.6 Summary ....... 9

## 2 Where To Begin ....... 11

2.1 Join The Journey ....... 11
2.2 Where You Fit In ....... 18
2.3 Summary ....... 20

## 3 Open Science ....... 21

3.1 Open Science ....... 21
3.2 Open Science In Action: the Study-a-Thon .... 23
3.3 Open Standards ....... 23
3.4 Open Source ....... 24
3.5 Open Data ....... 24
3.6 Open Discourse ....... 25
3.7 OHDSI and the FAIR Guiding Principles .... 25

## II. Uniform Data Representation

4 The Common Data Model ....... 31

4.1 Design Principles ....... 33
4.2 Data Model Conventions ....... 34
4.3 CDM Standardized Tables ....... 39
4.4 Additional Information ....... 51
4.5 Summary ....... 51
4.6 Exercises ....... 52

5 Standardized Vocabularies ....... 55

5.1 Why Vocabularies, and Why Standardizing .... 55
5.2 Concepts ....... 58
5.3 Relationships ....... 65
5.4 Hierarchy ....... 68
5.5 Internal Reference Tables ....... 70
5.6 Special Situations ....... 70
5.7 Summary ....... 72
5.8 Exercises ....... 73

## 6 Extract Transform Load ....... 75

6.1 Introduction ....... 75
6.2 Step 1: Design the ETL ....... 75
6.3 Step 2: Create the Code Mappings ....... 84
6.4 Step 3: Implement the ETL ....... 93
6.5 Step 4: Quality Control ....... 94
6.6 ETL Conventions and THEMIS ....... 95
6.7 CDM and ETL Maintenance ....... 96
6.8 Final Thoughts on ETL ....... 97
6.9 Summary ....... 97
6.10 Exercises ....... 98

## III. Data Analytics

7 Data Analytics Use Cases ....... 103

7.1 Characterization ....... 103
7.2 Population-Level Estimation ....... 104
7.3 Patient-Level Prediction ....... 105

8 OHDSI Analytics Tools ....... 109

8.1 Analysis Implementation ....... 109
8.2 Analysis Strategies ....... 110
8.3 ATLAS ....... 111
8.4 Methods Library ....... 114
8.5 Development Strategies ....... 121
8.6 Summary ....... 122

9 SQL and R ....... 125

9.1 SqInder ....... 126
9.2 DatabaseConnector ....... 135
9.3 Querying the CDM ....... 137
9.4 Using the Vocabulary When Querying ....... 141
9.5 QueryLibrary ....... 142
9.6 Designing a Simple Study ....... 143
9.7 Implementing the Study Using SQL and R ....... 143
9.8 Summary ....... 149
9.9 Exercises ....... 149

10 Defining Cohorts ....... 151

10.1 What Is A Cohort? ....... 152
10.2 Rule-Based Cohort Definitions ....... 153
10.3 Concept Sets ....... 155
10.4 Probabilistic Cohort Definitions ....... 156
10.5 Cohort Definition Validity ....... 156
10.6 Defining a Cohort for Hypertension ....... 157
10.7 Implementing a Cohort Using ATLAS ....... 158
10.8 Implementing the Cohort Using SQL ....... 168
10.9 Summary ....... 175
10.10 Exercises ....... 176

11 Characterization ....... 177

11.1 Database Level Characterization ....... 178
11.2 Cohort Characterization ....... 178
11.3 Treatment Pathways ....... 178
11.4 Incidence ....... 179
11.5 Characterizing Hypertensive Persons ....... 180
11.6 Database Characterization in ATLAS ....... 181
11.7 Cohort Characterization in ATLAS ....... 183
11.8 Cohort Characterization in R ....... 191
11.9 Cohort Pathways in ATLAS ....... 194
11.10 Incidence Analysis in ATLAS ....... 199
11.11 Summary ....... 202
11.12 Exercises ....... 203

12 Population-Level Estimation ....... 205

12.1 The Cohort Method Design ....... 206
12.2 The Self-Controlled Cohort Design ....... 210
12.3 The Case-Control Design ....... 210
12.4 The Case-Crossover Design ....... 211
12.5 The Self-Controlled Case Series Design ....... 212
12.6 Designing A Hypertension Study ....... 213
12.7 Implementing the Study Using ATLAS ....... 215
12.8 Implementing the Study Using R ....... 228
12.9 Study Outputs ....... 237
12.10 Summary ....... 242
12.11 Exercises ....... 242

13 Patient-Level Prediction ....... 245

13.1 The Prediction Problem ....... 246
13.2 Data Extraction ....... 248
13.3 Fitting The Model ....... 249
13.4 Evaluating Prediction Models ....... 254
13.5 Designing a Patient-Level Prediction Study ....... 258
13.6 Implementing the Study in ATLAS ....... 261

14 Evidence Quality ....... 293

14.1 Attributes of Reliable Evidence ....... 293
14.2 Understanding Evidence Quality ....... 295
14.3 Communicating Evidence Quality ....... 296
14.4 Summary ....... 296

15 Data Quality ....... 297

15.1 Sources of Data Quality Problems ....... 298
15.2 Data Quality in General ....... 298
15.3 Study-Specific Checks ....... 303
15.4 ACHILLES in Practice ....... 305
15.5 Data Quality Dashboard in Practice ....... 307
15.6 Study-Specific Checks in Practice ....... 308
15.7 Summary ....... 311
15.8 Exercises ....... 311

16 Clinical Validity ....... 313

16.1 Characteristics of Health Care Databases ....... 313
16.2 Cohort Validation ....... 314
16.3 Source Record Verification ....... 317
16.4 PhEvaluato ....... 320
16.5 Generalizability of the Evidence ....... 330
16.6 Summary ....... 331

17 Software Validity ....... 333

17.1 Study Code Validity ....... 333
17.2 Methods Library Software Development Process ....... 335
17.3 Methods Library ....... 338
17.4 Summary ....... 339

18 Method Validation ....... 341

18.1 Design-Specific Diagnostics ....... 341
18.2 Diagnostics for All Estimation ....... 342
18.3 Method Validation in Practice ....... 349
18.4 OHDSI Methods Benchmark ....... 357
18.5 Summary ....... 358

V. OHDSI Studies

19 Study Steps ....... 363

19.1 General Best Practice Guidelines ....... 364
19.2 Study Steps in Detail ....... 367
19.3 Summary ....... 373

20 OHDSI Network Research ....... 375

20.1 OHDSI as a Research Network ....... 375
20.2 OHDSI Network Studies ....... 376
20.3 Running an OHDSI Network Study ....... 380
20.4 Look Forward: Using Network Study Automation ....... 383
20.5 Best Practices for OHDSI Network Studies ....... 384
20.6 Summary ....... 386

Appendix

A Glossary ....... 387
B Cohort Definitions ....... 391
C Negative Controls ....... 409
D Protocol Template ....... 413
E Suggested Answers ....... 415
F Bibliography ....... 445
G Index ....... 455
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.
C O L L A B O R A T I V E  A C T I V I T I E S

Oct. 11-13, 2018 • Bethesda, Md.

June 27-29, 2019 • Guangzhou, China

Dec. 12-14, 2019 • Gwangju, Korea

Sept. 15-17, 2019 • Bethesda, Md.

June 24-26, 2022 • Rotterdam, Neth.
The 2021 OHDSI Global Symposium featured plenary presentations on both OHDSI’s Impact on the COVID-19 Pandemic (Day 1), and On the Journey to Reliable Evidence (Day 2). The main days included the State of the Community Presentation, the Collaborator Showcase, and a memorable Closing Ceremony that focused on OHDSI’s work through the perspective of a patient. There were also a pair of activities, including the first OHDSI Reproducibility Challenge workshop, and also a full-day tutorial on building concept sets.

If you missed it, or wish to watch any of the presentations again, they are all available at ohdsi.org/2021-ohdsi-global-symposium.

More Highlights from OHDSI2021
- The Closing Ceremony (Patrick Ryan, Jamie Weaver) highlighted the value of observational health research, as a former Titan Award recipient shared his story about appreciating OHDSI’s work from the patient’s perspective.
- The collaborator showcase received a record number of submissions, and following a peer-review process, there were more than 100 submissions of posters, software demos, and lightning talks; check out page 31 for more on the collaborator showcase.
- The State of the Community (George Hripcsak) highlighted the mission, development and direction of the OHDSI community.
2021 Lightning Talks

Data Quality Dashboard Used to Improve Data Quality in the EHDEN Network
Presenter: Clair Blacketer, Erica Voss

Beyond Clinical: Integrating Research Assay Data into the Observational Health Data Sciences and Informatics Common Data Model (OHDSI CDM) through the Surgical Critical Care Initiative (SC2i)
Presenter: Chandra Almond

Extending the OMOP CDM to store the output of natural language processing pipelines
Presenter: Monica Arrue

Validation of the Genomic Variant Vocabulary against TCGA
Presenter: Denys Kaduk

Evaluating the performance of Austin’s standardized difference heuristic in observational cohort studies with varying sample size
Presenter: Mitchell Conover

Leveraging APHRODITE to identify bias in statistical phenotyping algorithms
Presenter: Juan Banda

Assessing the impact of race on glomerular filtration rate prediction
Presenter: Linying Zhang

A Prediction Model Library
Presenter: Ross Williams

Detection of prone positioning in hospitalized COVID patients using NLP
Presenter: Patrick Alba

From metrics to intelligence using the OMOP CDM and Patient-Level-Prediction package as a foundation decision support tool
Presenter: Ismail Gögenur

Detecting PTSD and self-harm among US Veterans using positive unlabeled Learning
Presenter: Christophe Lambert

Revealing unknown benefits of existing medications to aid the discovery of new treatments for post-traumatic stress disorder
Presenter: David Kern

The European Health Data & Evidence Network (EHDEN) Sharing the OHDSI Journey and a Vision of Evidence Today, Not in Several tomorrows
Presenter: Nigel Hughes

Long term outcomes of prostate cancer patients managed by watchful waiting: results from the PIONEER/EHDEN/OHDSI study-a-thon
Presenter: Kees van Bochove, Asieh Golozar

Covid-19 Pandemic impacts on mental health Related conditions Via multi-database network: a Longitudinal Observational study (CERVELLO)
Presenter: Hao Luo

Large Scale Dissemination of OHDSI Methods and Tools: Introducing a New Community Resource, the OHDSI Center at the Roux Institute at Northeastern University
Presenter: Kristin Kostka

2021 Collaborator Showcase
Community Contribution Awards

Alberto Labarga: Extending the OMOP CDM to store the output of natural language processing pipelines

Anna Ostropolets: The concept of anchoring in observational study design and its influence

Kimberley Dickinson: Gold or Lead? Adjudicating Differences between CDM Data and Chart Reviews

Kelli Li: Competing risk regression models in cohort studies with the R package Cohort Method

Christophe Lambert: Detecting PTSD and self-harm among US Veterans using positive unlabeled Learning

#JoinTheJourney
The EHDEN Academy

The EHDEN Academy (academy.ehden.eu) serves as a free, publicly available online educational resource for anyone working in the domain of real-world data and real-world evidence.

Originating in the European Health Data & Evidence Network (EHDEN) IMI2 project, its goal is to build upon the foundations of that project and its collaboration with the OHDSI community. It is currently used across ~70 countries and has ~2900 enrollees.

The EHDEN Academy is a 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 analytical tools.

Current Courses in the EHDEN Academy

- Getting Started
- EHDEN Foundation
- Introduction to Real World Data & Real World Evidence (non-expert)
- Open Science & FAIR Principles
- Introduction to Data Quality
- Phenotype Definition, Characterisation and Evaluation
- Patient-Level Prediction R for Patient-Level Prediction
- Population-Level Effect Estimation
- Introduction to Usagi & Code Mappings for an ETL
- Infrastructure
- OHDSI in a Box
- OMOP CDM and Standardised Vocabularies
- Extract, Transform & Load
- ETL Learning Pathway: Data Partner & SME Real World Use Cases
- ATLAS
- Health Technology Assessment

Courses In Development

- Regulatory Learning Pathway
- Risk Minimization Management
- Data Quality
- HADES
- Outcome Standards
- OHDSI and Low/Middle Income 2 Countries (LMICs)

The European Health Data & Evidence Network (EHDEN.eu) aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care.

Its mission is to provide a new paradigm for the discovery and analysis of health data in Europe by building a large-scale, federated network of data sources standardized to the OMOP common data model and collaborating with OHDSI internationally.

As of the summer of 2022, EHDEN has built a federated network, so far, of 166 data partners from across 27 European nations, and has trained 65 small-to-medium enterprises to support mapping of ~650 million records to OMOP. The EHDEN study workflow portal was made public at the OHDSI Europe Symposium 2022 (portal.ehden.eu) with free access for the Data Partner Catalogue. A startup not-for-profit is being launched for sustainability.
The European Medicines Agency (EMA) announced Feb. 9, 2022 that Erasmus University Medical Center Rotterdam has been contracted to establish the DARWIN EU® (Data Analysis and Real World Interrogation Network) Coordination Centre.

The role of the Coordination Centre is to develop and manage a network of real-world healthcare data sources across the EU and to conduct scientific studies requested by medicines regulators and, at a later stage, requested by other stakeholders.

The vision of DARWIN EU® is to give EMA and national competent authorities in EU Member States access to valid and trust-worthy real-world evidence, for example on diseases, patient populations, and the use, safety and effectiveness of medicines, including vaccines, throughout the lifecycle of a medicinal product.

By supporting decision-making on the development, authorisation and surveillance of medicines, a wide range of stakeholders will benefit, from patients and healthcare professionals to health technology assessment bodies and the pharmaceutical industry. Additionally, DARWIN EU® will provide an invaluable resource to prepare for and respond to future healthcare crises and pandemics.

For example, the availability of timely and reliable real-world evidence can lead to innovative medicines becoming more quickly available to patients. Better evidence also supports more informed regulatory decision-making on the safe and effective use by patients of medicines on the market.

Peter Rijnbeek is the Executive Director of the DARWIN EU® Coordinating Center. He is a veteran OHDSI collaborator and Titan Award honoree, and he serves as Chair of the Department of Medical Informatics of the Erasmus MC.
The OHDSI community and SNOMED International formalized their long-time relationship with a five-year collaborative agreement that will benefit both of their user communities.

SNOMED CT is a core terminology within OHDSI’s common data model: Observational Medical Outcomes Partnership (OMOP), allowing the use of other terminologies and classifications through computable linkages.

The collaboration, which officially began in April 2022, provides OHDSI and its user community with comprehensive ontologies on specific healthcare domains and content such as devices, social determinants of health, disease severity scores and modifiers of cancers, as well as better concept definitions and resolutions of composite concepts in large-scale observational research.

In return, OHDSI and its user community can provide SNOMED International with information and feedback on clinical validation, frequency of use data, and validation of SNOMED CT content modeling. Ultimately, continuous feedback shared regarding identified content gaps will benefit both user communities as we move forward together.

The collaboration supports SNOMED International’s Member and stakeholder-driven five-year strategy, which includes genomic content collaboration as well as engagement with the research community. The work of OHDSI looks to showcase the use of SNOMED CT for the purposes of data analytics, supporting both healthcare research and audit, with the aim of enhancing healthcare globally.

"Both OHDSI and SNOMED International are working towards creating a healthcare environment that provides both patients and clinicians the real-world evidence needed to make informed and important decisions. SNOMED plays a critical role by delivering the comprehensive health terminologies necessary for OHDSI to generate reliable and reproducible evidence. Our community is thrilled to formalize this partnership so we can continue this important work."

- George Hripcsak
Chair of the OHDSI Coordinating Center at Columbia University.
Health Level Seven International (HL7®) and the OHDSI network agreed to a collaboration in 2021 to address the sharing and tracking of data in the healthcare and research industries by creating a single common data model. The organizations will integrate HL7 Fast Healthcare Interoperability Resources (FHIR®) and OHDSI’s Observational Medical Outcomes Partnership (OMOP) common data model to achieve this goal.

HL7 International CEO Dr. Charles Jaffe, M.D., Ph.D., underscored the significance of this partnership. “The Covid-19 pandemic has emphasized the need to share global health and research data. Collaboration with OHDSI is critical to solving this challenge and will help our mutual vision of a world in which everyone can securely access and use the right data when and where they need it.”

The organizations will align their standards to capture data in a clearly defined way into a single common data model. This will allow clinicians as well as researchers to pull data from multiple sources and compile it in the same structure without degradation of the information. This endeavor has global implications with the potential to permit the clinical community to define the elements they need, package and share them in a consistent single structure.

“We are excited to have the OHDSI community join this partnership with HL7 to evolve community standards around observational research and clinical care,” said George Hripcsak, MD, MS, OHDSI’s coordinating center director. “These standards set the foundation for our mission of global, open-science research, and this partnership will accelerate the development of effective and safe treatments for diseases facing today’s global population.”

Health Level Seven International is the global authority for healthcare information interoperability and standards with affiliates established in more than 30 countries. HL7 is a non-profit, ANSI accredited standards development organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. HL7’s members represent approximately 500 corporate members, which include more than 90 percent of the information systems vendors serving healthcare.
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.

Some of these initiatives have been mentioned in previous pages (EHDEN, HL7, SNOMED), while other organizations are highlighted below. If your organization would like to collaborate with OHDSI, please reach out on our forums!

In 2020, OHDSI was awarded a $10 million contract from the U.S. Food and Drug Administration (FDA) to provide support to the Biologics Effectiveness and Safety (BEST) program, which was launched by the FDA Center for Biologics Evaluation and Research (CBER) in 2017.

The lead research team, primarily comprised of OHDSI personnel from Columbia University, UCLA, Northeastern University and Johns Hopkins University provides support to the BEST system in its mission to conduct safety and effectiveness surveillance of biologic products (vaccines, blood and blood products, tissues and advanced therapeutics).

The All of Us Research Program is inviting one million people across the U.S. to help build one of the most diverse health databases in history.

Researchers will use the data, which is mapped to the OMOP CDM, to learn how our biology, lifestyle, and environment affect health. This may one day help them find ways to treat and prevent disease.

PIONEER is part of the Innovative Medicine Initiative’s (IMI’s) “Big Data for Better Outcomes” (BD4BO) umbrella program. The BD4BO mission is to improve health outcomes and healthcare systems in Europe by maximizing the potential of Big Data.

OHDSI collaborated with PIONEER in early 2021 on a five-day study-a-thon that investigated the natural history and outcomes of prostate cancer patients managed with watchful waiting.

The N3C is a partnership among the NCATS-supported Clinical and Translational Science Awards (CTSA) Program hubs, the National Center for Data to Health (CD2H), and NIGMS-supported Institutional Development Award Networks for Clinical and Translational Research (IDeA-CTR), with overall stewardship by NCATS.

Collaborators are contributing and using COVID-19 clinical data, mapped to the OMOP CDM, to answer critical research questions to address the pandemic.

The Federated E-Health Big Data for Evidence Renovation Network (FEEDER-NET) project was initiated in 2018 with a $10 million budget from the Ministry of Trade, Industry & Energy of Korea.

The main goal is to build a bio-health Big Data ecosystem, centered around an OMOP CDM-based data network. As of August 2021, the FEEDER-NET network included more than 54 million patients.
V. Data Standards
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.

“The OMOP Common Data Model 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
OHDSI Data Partners

What does it take to be an OHDSI data partner? Anyone with access to observational data can standardize their database in the OMOP Common Data Model, apply OHDSI’s open-source tools, and participate in collaborative research.

Who has already joined the journey and adopted the OMOP CDM? There are currently 453 databases, including 374 electronic health records, 34 registries and 30 administrative claims sources, that come from 41 different countries. Together, these databases represent more than 928 million unique patient records, approximately 12% of the world’s population.
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.

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

- 10,218,572 concepts
  - 3,549,524 standard concepts
  - 780,207 classification concepts
- 135 vocabularies
- 42 domains

1 Shared Resource to Enable Data Standards

This network diagram shows the relationships between vocabularies. Nodes are vocabularies, sized by the number of concepts. Edges show connections between concepts within vocabularies.

Want to learn more about the OHDSI vocabularies?
Read: book.ohdsi.org
Download: athena.ohdsi.org
Learn: academy.ehden.edu

“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 for Data Standards recipient

#JoinTheJourney
The open-source tools that empower OHDSI research are not only available to the community, but they are DEVELOPED by the community. Leaders within our global network, including 2018 Titan Award recipient Martijn Schuemie (pictured), have developed the foundation for OHDSI collaborators to engage in robust, reliable and reproducible observational health research.
VI.

Open-Source Software

COHORT METHOD
New-user cohort studies using large-scale regression for propensity and outcome models.
Learn more...

SELF-CONTROLLED CASE SERIES
Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.
Learn more...

CYCLOPS
Highly efficient implementation of regularized logistic, Poisson and Cox regression.
Learn more...

DATABASE CONNECTOR
Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.
Learn more...

SQL RENDERER
Generate SQL on the fly for the various SQL dialects.
Learn more...

SELF-CONTROLLED COHORT
A self-controlled cohort design, where time preceding exposure is used as control.
Learn more...

EVIDENCE SYNTHESIS
Routines for combining causal effect estimates and study diagnostics across multiple data sites in a distributed study.
Learn more...

PARALLEL LOGGER
Support for parallel computation with logging to console, disk, or e-mail.
Learn more...

FEATURE EXTRACTION
Automatically extract large sets of features for user-specified cohorts using data in the CDM.
Learn more...

ANDROMEDA
Storing very large data objects on a local drive, while still making it possible to manipulate the data in an efficient manner.
Learn more...

PATIENT LEVEL PREDICTION
Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.
Learn more...

EMPIRICAL CALIBRATION
Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.
Learn more...

BIG KON
A large scale k-nearest neighbor classifier using the Lucene search engine.
Learn more...

ROHTS Web API
Interact with OHDSI WebAPI web services.
Learn more...

OHDSI SHARING
Securely sharing (large) files between OHDSI collaborators.
Learn more...
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 observation study from data to estimates and supporting statistics, figures, and tables. The packages interact directly with observational data in the OMOP Common Data Model, and are designed to support both large datasets and large numbers of analyses.

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
This 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
This 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

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

EnsemblePatientLevelPrediction
This 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.
Cohort Construction

CAPR
This is an R package to develop and manipulate OHDSI cohort definitions. This package assists in creating a cohort definition that can be compiled by circe-be using CirceR. Cohort definitions developed in CapR are compatible with OHDSI ATLAS. Additionally, the package allows for development of cohort design components, sub-items of a cohort design that are meant to be reusable and mutable to assist in creating cohorts in study development.

CirceR
A R-wrapper 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 (CriteriaFeature). This package provides convenient wrappers for Circe functions, and includes the necessary Java dependencies.

Evidence Quality

CohortDiagnostics
This 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 comparison of multiple candidate cohort definitions across one or more data sources, facilitating reproducible research.

MethodEvaluation
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.

Empirical Calibration
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.

Kheiron Contributor Cohort
Paul Nagy and Adam Black, leads of the Open-Source Community workgroup, founded the Kheiron Contributor Cohort in 2022 as a way to welcome and mentor new developers in the OHDSI community. The first cohort included 25 individuals who committed 10% of their time for a year to join the journey with the open-source community, and the cohort has already made positive impacts on OHDSI tools. The cohort has hosted workshops, learned from developers around the community and made strong connections. Paul and Adam plan to lead a new cohort next year, and more information about that will be shared when available.
Supporting Packages

Andromeda
AsynchroNous Disk-based Representation of MassivE DAtA (ANDROMEDA) is 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.

BigKNN
This is an R package implementing a large scale k-nearest neighbor (KNN) classifier using the Lucene search engine.

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.

Eunomia
Eunomia is a standard dataset in the OMOP (Observational Medical Outcomes Partnership) Common Data Model (CDM) for testing and demonstration purposes. Eunomia is used for many of the exercises in the Book of OHDSI. For functions that require schema name, use ‘main’.

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

Hydra
This is 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.

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).

ROhdsiWebApi
ROhdsiWebApi is a 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 to JSON and vice versa. To ensure r-objects are analysis ready, the objects are type converted where possible, e.g. date/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, programatically 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”.

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.

OhdsiSharing
This is an R package for sharing data between OHDSI partners.

The eight HADES packages shown above have been released on CRAN and have been downloaded more than 400,000 times.
Certain factors for the success of an open-science community like OHDSI are more obvious than others. When hundreds of people come together to research a common cause, or studies are run against millions of patient records in a global database, it becomes clear that something impactful is happening.

One critical factor in OHDSI’s ability to perform rigorous, ground-breaking analyses lies under the surface, but it holds an equally important role in the overall community mission.

A core foundation for OHDSI is open-source software development, and a small group of community collaborators, led by Martin Schuemie, has generated a collection of analytics tools that enable research both in and out of the OHDSI community.

HADES — the Health Analytics Data-to-Evidence Suite — is a set of (currently) 25 open-source R packages for large scale analytics, including population characterization, population-level causal effect estimation, and patient-level prediction, as well as supporting packages that are critical throughout the journey of observational research. The packages offer a robust set of functions that together can be used to perform all the steps required to conduct a network study, from connecting to a database, translating queries into the appropriate SQL dialect, generating cohorts and extracting features, fitting large-scale statistical models, compiling results for meta-analysis and empirical calibration, and enabling exploration through interactive visualization dashboards.

The packages interact directly with any observational data in the OMOP Common Data Model, and are designed to support network research across large datasets with millions of patients and billions of observations, as well as smaller populations. HADES scales to enable large numbers of analyses so that researchers can systematically explore populations and hypotheses across a range of outcomes.

These packages, available on the HADES home page (ohdsi.github.io/Hades), have empowered at least 34 network studies. These include the OHDSI LEGEND study on hypertension, CHARYBDIS, hydroxychloroquine safety, the ongoing work with COVID AESI characterization, and many more. All packages are developed and released as open-source tools at github.com/OHDSI. Amongst the HADES ecosystem, all packages are made open source and publicly accessible through the OHDSI GitHub repository. Additionally eight packages are available on CRAN (The Comprehensive R Archive Network, a public repository for all R users). These eight HADES packages are available on CRAN and have been downloaded more than 400,000 times (see graphic, opposite page).

“Our community, and observational researchers in general, owe an enormous debt of gratitude to Martijn and the HADES team for leading this effort,” said Provost and Senior Vice President for Academic Affairs at Northeastern University David Madigan, who is leading efforts around the new OHDSI Center at the Roux Institute. “Open-source development within the OHDSI community is the quiet force that is impacting important evidence that can save lives, and it shouldn’t be taken for granted.”

Beyond network studies, HADES allows researchers to conduct analyses locally. It supports best practices for use of observational data as learned from previous and ongoing research; for example, the population-level estimation methods have been extensively evaluated using the OHDSI Methods Benchmark, as published in the Harvard Data Science Review.

Researchers can learn how to use HADES through documentation found in the Book of OHDSI (ohdsi.github.io/TheBookOfOhdsi/).

“We are very proud of the impact that HADES continues to make on real-world evidence generation,” said Schuemie, who leads the HADES group. “Our team develops, tests and continuously monitors a set of tools that empowers global research using best practices developed within our community.”

OHDSI’s reach has expanded recently, including its role supporting the FDA BEST program in vaccine surveillance, as well as informing best practices in the recent EMA revision of its guidelines. Researchers continue to join the community, and the breadth of work has expanded as collaboration efforts have matured. But for success to follow these positive developments, the HADES foundation and team continues to need greater support.

A small portion of the community maintains the set of packages, and one consistent HADES objective is to diversify the leadership within the ecosystem. There are several ways that OHDSI collaborators can support this critical piece of the puzzle. Developers can contribute by helping develop and test code. Users of the tools can help with testing, user documentation and other training resources. Those with the means can provide financial support to help pay for developers specifically focused on open-source development. Anybody can contribute ideas as part of the HADES workgroup.

Just as every piece of the HADES toolset has aided the growth of OHDSI, every small contribution from the community can aid the advancement of HADES.

“Open-source development within the HADES ecosystem has been critical to our growth and success as a community,” said George Hripcsak, Chair and Vivian Beaumont Allen Professor of Biomedical Informatics at Columbia, the coordinating center for OHDSI.

“Martijn and the HADES team have done extraordinary work to put us in position to run observational health studies that make a difference to patients around the world, but we can’t overlook the burden on this small core of our community who have enabled this growth. I believe we have people who are generous with both their time and talents to help take HADES to a sustainable level as we continue to mature as a community.”

#JoinTheJourney OHDSI.org

Open-Source Software

HADES
Health Analytics Data-to-Evidence Suite

Package Maintainers

Martijn Schuemie
BigKnn, CohortMethod, DatabaseConnector, EmpiricalCalibration, EvidenceSynthesis, MethodEvaluation, ParallelLogger, SelfControlledCaseSeries, SqlReader

Anthony Sena
CohortGenerator, FeatureExtraction, HYDRA

Jamie Gilbert
CohortDiagnostics, SelfControlledCohort

Gowtham Rao
PhenotypeLibrary, ROhdsiWebApi

Jenna Reps
EnsemblePatientLevelPrediction, PatientLevelPrediction

Marc Suchard
Cyclops, IterativeHardThresholding

Adam Black
Andromeda

Frank DeFalco
Eunomia

Lee Evans
OhdsiSharing

Christopher Knoll
CirceR

Martin Lavalee
CAPR

Peter Rijnbeek
PatientLevelPrediction

## Join the Journey

#JoinTheJourney OHDSI.org
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 59.
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.
1. **LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (e.g., the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (e.g., the effects of many treatments for a disease on many outcomes). **Aim:** 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:** 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:** 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 explicited and applied systematically. **Aim:** 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 advanced 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:** 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:** 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:** 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:** 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:** 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:** Privacy.
**Methods 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.
Thiazide diuretics demonstrate better effectiveness and cause fewer side effects than ACE inhibitors as first-line antihypertensive drugs, 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 across nine observational databases, making it 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.

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 reproducibly generate evidence by applying advanced analytics across a network of disparate databases for a wide array of exposures and outcomes.”

“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.”

- Marc Suchard
2018 Titan Award recipient for Methodological Research
The Journey To Reliable Evidence

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

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 for Methodological Research recipient
With Patient-Level Prediction

**Model Development**
- **Learning across datasets**
  - Models can be learned across datasets while maintaining privacy.
- **Test/Train split**
  - The design used to pick hyper-parameters and estimate internal validation matters, even with big p and big n data.
- **Sample size**
  - Learning curves provide a way for model developers to determine whether they have sufficiently sized data.

**Model Evaluation**
- **Model usability**
  - Simple score-based models are easier to apply and can be benchmarked against large-scale models.
- **Visualizing performance**
  - A simple plot with the operating characteristics for all cut-offs informs model usefulness.
- **Network validation**
  - The OHDSI network enables large-scale external validation and improves our understanding of models.

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

*Join The PLP Journey*

**PLP GitHub**: [github.com/OHDSI/PatientLevelPrediction](https://github.com/OHDSI/PatientLevelPrediction)

#JoinTheJourney
Members of the OHDSI community have published many papers together. Often, these studies are first showcased at our annual OHDSI Symposia, like the 2019 event pictured here. These events also provide opportunities for networking, which leads to new collaborations, and new collaborations lead to new evidence generation that impacts patients around the world.
VIII. OHDSI Publications


Martijn J. Schuemie,1,2,4 Faziah Arshad,1,2,4 Nicole Pratt,3 Fredrik Nyberg2,6 Thamir M Aishahmari,1 George Hilpca2,6 Patrick Ryan2,4,7 Daniel Pinto-Alhambra6,6,9 Lana Y. H. Lai10,11,12 Xintong Li11,12, Stephen Fortin,1 Evan Minty10,11 and Marc A. Suchard1,3,4,7,12

Phenotype Algorithms for the Identification and Characterization of Vaccine-Induced Thrombotic Thrombocytopenia in Real World Data: A Multinational Network Cohort Study

Azza Shaoul1,8,2,3,4 Goutham A. Rao1,2,4,5, Anna Ostropolets1,2,4,5 Miguel Angel Mayer1,8,2,4,5, Juan Manuel Ramirez-Anguita1,8,2,4,5, Filippo Maljkovic1,8,2,4,5, Biljana Carevski1,8,2,4,5, Scott Horban1,2,4,5, Daniel R. Morales1,2,4,5, Talita Duarte-Salles1,8,2,4,5, Clement Fraboulet1,8,2,4,5, Tanguy Le Carrou11,8,2,4,5, Spiros Denaxas11,8,2,4,5, Quinlivan Pappas11,8,2,4,5, Luis H. John12,6, Peter R. Rijnberk12,6, Evan Minty10,11, Thamir M. Aishahmari14, Rupa Makadia1,2,4,5, Claire Blacker12,4,5, Frank DeFalco12,4,5, Anthony G. Sena12,4,5, Marc A. Suchard2,4,5, Daniel Pinto-Alhambra6,6,9, Patrick B. Ryan12,6,7

Seek COVER: using a disease proxy to rapidly develop and validate a personalized risk calculator for COVID-19 outcomes in an international network

Ross D. Williams1, Aneek F. Markus2, Cynthia Yang3, Talita Duarte-Salles4, Scott L. DuVal5, Thomas Falconer5, Andrea Jonsson-Cole6, Chunguo Kim7, Yeungsook Rho7, Andrew E. Williams2, Amanda Abegaz Machado2, Min Ho An2, Maria Arangio1, Carlos-Amelia1, Edward Burns1,5, Young Hwa Choi2,4, Johann Drakos4, Maria Teresa Fernandez-Abascal1, Sergio Fernandez-Bertolin1, George Hilpca2,6, Benjamin Skov Kaas-Hansen1,2,4,5, Pia Paloma Kandasamy1,5, Jan A. Kors4, Katarzyna Kosta9, Saw-Tang Law11, Kristine E. Lynch1, Gerardo Machinick1, Michael E. Matheny2,22, Daniel Morales2,23, Fredrik Nyberg2,6, Rae Woong Park2,5, Albert Prato-Uribes2,4, Nicole Pratt3,4, David R. Rote2,5, Christian G. Reich2,6, Monica Rivera7,8, Tom Sarin11, Azza Shaoul1,8,2,3,4, Matthew E. Spottke1,2, Ewout W. Steyerberg2,22,7,22, Marc A. Suchard1,2, Seng Chen You1,2, Lin Zhang2,22, Lili Zhou1,5, Patrick B. Ryan4,6, Daniel Pinto-Alhambra6,6,9, Jenna M. Reps1,7,9 and Peter R. Rijnberk1,8

#JoinTheJourney

OHDSI.org
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.
Our OHDSI Community

- Each dot is an OHDSI collaborator with at least 2 OHDSI papers, which include studies involving OMOP.
- Size of the dot indicates the number of OHDSI/OMOP papers.
- The color indicates the first year someone wrote an OHDSI paper (see legend below).
- A line means two authors were on the same paper. The darker the color of the line, the more papers they co-authored.
- The layout is based on co-authorships, so people who collaborated more end up close together in the graph.

Legend:
- Node Fill Color: Year of first OHDSI publication
- Node Size: Number of OHDSI papers
  - 2
  - 34
  - 66
  - 99
  - 131
Community Dashboard

The OHDSI Community Dashboard is a tool to highlight the progress we are making toward our mission and the collective accomplishments and impact of our community. A goal of the dashboard is help our community identify how members can see the OHDSI ecosystem as an interconnected system to make a larger impact.

PubMed Publication Tracking highlights scholarship generated 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.

There are also dashboards monitoring YouTube video tracking and EHDEN course tracking.

Thank you to the team of Paul Nagy, Star Liu, Jody-Ann McLeggon, Asieh Golozar, and Adam Black for their leadership in developing this dashboard.


Zhu RL, Bissell M, Qureshi N, Saltz J, Haendel MA, Chute CG, Lehmann HP, Moffitt RA. Harmonizing units and values of quantitative data elements in a very large nationally

Network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety

Heider PM, Pipaliya RM, Meystre SM. A Natural Language Processing Tool Offering Data Extraction for COVID-19 Related Information (DECOVRI). Stud Health Technol

network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety


Zhu RL, Bissell M, Qureshi N, Saltz J, Haendel MA, Chute CG, Lehmann HP, Moffitt RA. Harmonizing units and values of quantitative data elements in a very large nationally


Tong J, Luo C, Islam MN, Shells NE, Buresh J, Edmondson M, Merkel PA, Lautenbach E, Duan R, Chen Y. Distributed learning for heterogeneous clinical data with application
PMCID: PMCPMC9198031.

PR, Development and external validation of prediction models for adverse health outcomes in rheumatoid arthritis: A multinational real-world cohort analysis. Semin Arthritis

Ostropolets A, Ryan PB, Schuemie MJ, Hirpsck G. Characterizing Anchoring Bias in Vaccine Comparator Selection Due to Health Care Utilization With COVID-19 and Influenza:
PMCID: PMCPMC9250064.


John LH, Kors JA, Reps JM, Ryan PB, Rijnbeek PR. Logistic regression models for patient-level prediction based on massive observational data: Do we need all data?

35705689.

s2589-7500(22)00048-6. PubMed PMID: 35859849; PubMed Central PMCID: PMCPMC9110014.


X. Join The Journey

Artwork by OHDSI collaborator Sarah Seager

#JoinTheJourney
Cheers, From The OHDSI Community!

2020 threatened to pull people apart, but the OHDSI community came closer together. Volunteer researchers from around the globe joined forces to study COVID-19 and other critical healthcare concerns. Collaboration in the spirit of open science drove us to do far more together than anybody could have done alone.

We also had a lot of fun in the process. To close our 2020 Global Symposium, we created a virtual “cheers” to celebrate our shared successes. To all of you who have done so much for the community, and to those of you who will join our future endeavors, CHEERS!
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!

Since as long as I can remember, I have loved sports. I love that sport is a meritocracy: at the start of any match, everyone has an equal chance to win. I love that there are objective criteria to determine success, clear rules that everyone must follow to play the game, clear goals to accomplish (like actually scoring goals!). At the end of any match, everyone knows who has won and lost.

Growing up, I especially loved playing team sports. Be it soccer, baseball or basketball, if there was a game to be played, I wanted to be in the action. I loved the competition, but I also enjoyed the comradery. I enjoyed the practices where we worked to improve our individual skills and also tried to perform better as a team. We would run drills one-on-one, two-on-two, five-on-five, each trying to focus on specific interactions between teammates. We’d rehearse set plays over and over to make sure everybody knew their role and were ready to execute. By gametime, we’d try our best to apply what we learned and see if we were good enough to earn the win. Together as a team, we shared the ‘thrill of victory or the agony of defeat.’

While I loved playing on teams, I was acutely aware that I was never one of the best players. On the soccer pitch, I wasn’t fast, didn’t have the best drilling skills, couldn’t kick it the furthest. On the basketball court, I wasn’t tall, still wasn’t fast, couldn’t jump, wasn’t much of a shooter. My dad often coached me while I was young, and he was also acutely aware of
my lack of ability. He taught me two things: 1. You can’t control how naturally talented you are, but you can control how much work you put in to try to get better; 2. You shouldn’t focus just on yourself, but think about how you can help make your team better.

Fast forward many years later, Craig Sachson and I coached the Killer Dragons (they came up with their own name!), a U-10 soccer team in Ithaca, N.Y. Our team was comprised of all the ‘leftover’ kids who weren’t selected by any of the other teams in the league. We knew they weren’t the fastest, didn’t have the best dribbling skills, couldn’t kick it the farthest. We wanted to instill that same love of the game that we had, to get them to enjoy both the competition and the practices together. We wanted them to understand the value of team and the importance of teamwork. I don’t remember all the practice drills or game scores that season, but I’ll never forget when the Dragons scored their first (and only) goal during the final game, which came after a couple successful passes up-field between teammates. Watching those kids come together to celebrate their shared accomplishment is a memory I’ll never forget.

For me, there’s nothing like the rush of team achievement, that feeling when you declare “we did it” and you celebrate and recognize how everyone came together to make it happen.

Herb Brooks, head coach of the ‘Miracle on Ice’ 1980 US hockey team that won Olympic gold, said, “When you pull on that jersey, you represent yourself and your teammates. The name on the front is a hell of a lot more important than the one on the back.” When a team has a shared purpose, it can have a powerful effect. Teams typically strive for common goals which are more ambitious than those that can be accomplished by individuals alone. Setting sights on a bigger objective can be motivating, but it can also feel intimidating on your own. With the support of a team, however, that shared objective can be broken into smaller pieces, and the specific tasks and contributions can seem more manageable and far less intimidating. When individual activities serve as building blocks, and teammates orchestrate how those blocks fit together, the team itself can build something great together.

**Teamwork scales accomplishment.** One person can only do so much, no matter how talented or productive she is. You can’t play all positions on the field at once. Collaboration is hard, and it takes practice to figure out how to work together most effectively. Once those relationships are made and trust in teammates is formed, it becomes more comfortable to divide-and-conquer. The World Cup takes place later this year, and it represents one of the great team competitions in the world. Players like Messi, Ronaldo and Neymar are recognized around the globe, but none can win for their home countries on their own. You don’t win one on eleven, no matter how special the one might be. A winning team needs a great striker, versatile midfielders, strong defenders and a dependable goalkeeper. Despite having different individuals having different responsibilities and different skills within different
position groups, the team works as one and builds on each other’s successes. There is no ‘me’ or ‘you’, there is only ‘we.’ Championships are often determined by the togetherness of the team.

I was quite struck and inspired when Martijn Schuemie introduced the analogy of the CERN Large Hadron Collider to our community (see image, right). More than 10,000 scientists across hundreds of universities and labs came together to build the world’s largest and highest-energy particle collider, and they have collaboratively conducted a wide range of large-scale experiments to answer some of the most fundamental questions in physics, such as detecting the Higgs boson. CERN is a marvel of modern engineering in terms of scale and precision, but it also required tremendous innovations in computing. Leaders in their respective disciplines could have chosen to focus on their own areas of interest, continuing to make incremental progress and producing quality scholarship in their local domains. Instead, they chose a path of multi-disciplinary collaboration, contributing their expertise — often a small part of the bigger whole — to collectively enable transformational advances in science that no one individual or lab could possibly have the resources or capabilities to achieve alone. Together, they have published thousands of papers, often with thousands of co-authors, providing compelling evidence toward explaining phenomena across the universe. Part of the CERN mission is to "unite people from all over the world to push the frontiers of science and technology, for the benefit of all." Imagine the thrill of being part of this team and playing your part, no matter how small, in something so big.

What is our sport? Who is our team? How do we win?

I believe wholeheartedly in the potential for real-world evidence to fundamentally transform medicine by putting the collective learnings from past patient experience into practice for future patient care. I am confident that using observational data to characterize disease natural history and treatment utilization patterns from around the world, estimate the safety and comparative effectiveness of medical interventions, and predict health outcomes in
populations of interest can empower patients and providers to make more informed, evidence-based and personalized treatment decisions. I also recognize that transformative change won’t happen overnight and can’t be completed by one person alone. I remain passionate about OHDSI’s mission, “to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.”

**I believe the only way that our open-science community can succeed is through effective teamwork.** If our sport is generating reliable evidence and our objective is to improve the lives of patients around the world, then that is a team I want to be on. That is a game I want to win.

Our team — the OHDSI community — already has a solid foundation. We have tremendous depth, with 3,266 collaborators on the field. We have incredible talent at all the key positions, with world-leading experts in open community data standards, open-source development, methodological research, and clinical applications guiding the way. A key strength of our team is diversity: we have strong representation across geographies, across disciplines, and across stakeholder communities. Our team has already demonstrated some of what we’re capable of:

- Together as a team, **we have established an open community data standard** that is now used by more than 450 organizations in more than 40 countries around the world.
- Together as a team, **we have developed a rich suite of open-source analytics tools**, with a subset of the HADES packages having been downloaded >400,000 times by the broader research community.
- Together as a team, **the EHDEN community has certified 47 small-to-medium size enterprises in best practices for data standards and OHDSI tool configuration**, amassing a core group capable of supporting health care institutions around the world on their journey to evidence using the OMOP CDM and OHDSI analytics.
- Together as a team, **we have empirically evaluated the performance of alternative causal inference methods for comparative effectiveness and safety surveillance**, establishing the LEGEND principles for conducting large-scale analyses within disease areas of interest.
- Together as a team, **we have established best practices for patient-level prediction and built advanced machine learning tools** to train and validate models across our network.
- Together as a team, **we have created a textbook, 15 courses and >400 hours of educational content** to train the broader research community on best practices in data standards and analysis.
- Together as a team, **we have published 475 scientific articles and generated evidence that has directly impacted clinical guidelines and regulatory policy.**
JOIN THE JOURNEY

Yet, I would contend that our team is still just learning how to play together.

We have many members in our community watching from the sideline, wondering if they should start to play and if the juice will be worth the squeeze. We have many data partners who have joined the journey by converting their data to the OMOP CDM, but relatively few who have contributed to an open network study. We have many users of OHDSI tools, but a much smaller number of contributors willing to communicate issues, recommend enhancements, develop solutions, and support their fellow colleagues in their appropriate use. We have innovated in the area of phenotype development and evaluation, but have only begun to build a shared library of cohort definitions for expanded use across our community. We’ve completed an impressive number of OHDSI network studies, but many more studies remain isolated to a specific data source or institution. We have many workgroups that have made progress on their local ambitions, but haven’t yet coordinated all the activities to align with an over-arching objective.

I’m excited for what’s possible if we all work as one team with a shared sense of purpose, each of us focusing our efforts to make our own contributions toward a common goal. Brick by brick, when we put it all together, we can build something profound. We can build a healthier world together.

I can’t imagine a sweeter victory.

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?

**Step One: Join The OHDSI Forums (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.

**Step Two: Join Our Workgroups & MS Teams Environment (ohdsi.org/ohdsi-workgroups)**

OHDSI has 27 active workgroups that always seek new collaborators. Our workgroups present opportunities for all community members to find a home for their talents and passions, and a place to make meaningful contributions. Our workgroups collaborate inside the OHDSI MS Teams environment; a form to join our Teams environment is available here: bit.ly/Join-OHDSI-Teams.

**Step Three: Join Our 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, debates and more. These calls are recorded, and you can access them (as well as the meeting link) at our Community Calls page.

**Step Four: Continue To Learn About OHDSI**

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

- 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

- 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

- Our OHDSI News page keeps you informed of recent news, publications, upcoming studies and more, while also profiling collaborators and providing other updates: ohdsi.org/ohdsi-news-updates

- Check out the OHDSI YouTube page (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), or reach out to us over email (contact@ohdsi.org).

Thank you for Joining The Journey with OHDSI!