Open Science for Observational Research: Lessons from the OHDSI Collaborative

Patrick Ryan, PhD
Vice President, Observational Health Data Analytics, Janssen Research and Development
Assistant Professor, Adjunct, Department of Biomedical Informatics, Columbia University Medical Center

on behalf of the OHDSI community
The journey to real-world evidence

Patient-level data in source system/schema

Reliable evidence
The journey to real-world evidence

Different types of observational data:

- **Populations**
  - Pediatric vs. elderly
  - Socioeconomic disparities

- **Care setting**
  - Inpatient vs. outpatient
  - Primary vs. secondary care

- **Data capture process**
  - Administrative claims
  - Electronic health records
  - Clinical registries

- **Health system**
  - Insured vs. uninsured
  - Country policies
The journey to real-world evidence

Types of evidence desired:
• Clinical characterization
  • Clinical trial feasibility
  • Treatment utilization
  • Disease natural history
  • Quality improvement

• Population-level effect estimation
  • Safety surveillance
  • Comparative effectiveness

• Patient-level prediction
  • Precision medicine
  • Disease interception
# Desired attributes for reliable evidence

<table>
<thead>
<tr>
<th>Desired attribute</th>
<th>Question</th>
<th>Researcher</th>
<th>Data</th>
<th>Analysis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatable</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
</tr>
<tr>
<td>Reproducible</td>
<td>Identical</td>
<td>Different</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
</tr>
<tr>
<td>Replicable</td>
<td>Identical</td>
<td>Same or different</td>
<td>Similar</td>
<td>Identical</td>
<td>Similar</td>
</tr>
<tr>
<td>Generalizable</td>
<td>Identical</td>
<td>Same or different</td>
<td>Different</td>
<td>Identical</td>
<td>Similar</td>
</tr>
<tr>
<td>Robust</td>
<td>Identical</td>
<td>Same or different</td>
<td>Same or different</td>
<td>Different</td>
<td>Similar</td>
</tr>
<tr>
<td>Calibrated</td>
<td>Similar (controls)</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Statistically consistent</td>
</tr>
</tbody>
</table>
OHDSI’s mission

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

We’re all in this journey together…
OHDSI’s community engagement

• Active community online discussion: [forums.ohdsi.org](http://forums.ohdsi.org)
  – >2,770 distinct users have made >18,700 posts on >3,250 topics
  – Implementers, Developers, Researchers, CDM Builders, Vocabulary users, OHDSI in Korea, OHDSI in China, OHDSI in Europe

• Weekly community web conferences for all collaborators to share their research ideas and progress

• >25 workgroups for solving shared problems of interest
  – ex: Common Data Model, Population-level Estimation, Patient-level Prediction, Phenotype, NLP, GIS, Oncology, Women of OHDSI

• Quarterly tutorials in OHDSI tools and best practices, taught by OHDSI collaborators for OHDSI collaborators

• OHDSI Symposia held annually in North America, Europe and Asia to provide the community face-to-face opportunities to showcase research collaborations

• Follow us on Twitter @OHDSI and LinkedIn
OHDSI is an international data network
Data across the OHDSI community

• 152 entries on 2019 OHDSI data network inventory
• 133 different databases with patient-level data from various perspectives:
  – Electronic health records, administrative claims, hospital systems, clinical registries, health surveys, biobanks
• Data in 18 different countries, with >369 million patient records from outside US

All using one open community data standard: OMOP Common Data Model
Open community data standard: OMOP CDM v6

Standardized clinical data:
- Person
  - Observation_period
  - Visit_occurrence
  - Visit_detail
  - Condition_occurrence
  - Drug_exposure
  - Procedure_occurrence
  - Device_exposure
  - Measurement
  - Note
  - Note_NLP
  - Survey_conduct
  - Observation
  - Specimen
  - Fact_relationship

Standardized health system data:
- Location
- Location_history
- Care_site
- Provider

Standardized clinical data:
- Person

Standardized derived elements:
- Condition_era
- Drug_era
- Dose_era

Results Schema:
- Cohort
- Cohort_definition

Standardized health economics:
- Cost
- Payer_plan_period

Standardized metadata:
- CDM_source
- Metadata

Standardized vocabularies:
- Concept
- Vocabulary
- Domain
- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor
- Source_to_concept_map
- Drug_strength
OHDSI’s standardized vocabularies

- >130 Vocabularies across 40 domains
  - MU3 standards: SNOMED, RxNorm, LOINC
  - Disparate sources: ICD9CM, ICD10(CM), Read, NDC, Gemscript, CPT4, HCPCS...
- >7.4 million concepts
  - >3.0 million standard concepts
  - >3.8 million source codes
  - >511,000 classification concepts
- >45 million concept relationships
- >74 million ancestral relationships

Publicly available for download at: http://athena.ohdsi.org/
OHDSI is advancing science
What is OHDSI’s strategy to deliver reliable evidence?

• **Methodological research**
  – Develop new approaches to observational data analysis
  – Evaluate the performance of new and existing methods
  – Establish empirically-based scientific best practices

• **Open-source analytics development**
  – Design tools for data transformation and standardization
  – Implement statistical methods for large-scale analytics
  – Build interactive visualization for evidence exploration

• **Clinical evidence generation**
  – Identify clinically-relevant questions that require real-world evidence
  – Execute research studies by applying scientific best practices through open-source tools across the OHDSI international data network
  – Promote open-science strategies for transparent study design and evidence dissemination
Complementary evidence to inform the patient journey

Clinical characterization: What happened to them?

Patient-level prediction: What will happen to me?

Population-level effect estimation: What are the causal effects?

observation

inference

causal inference
<table>
<thead>
<tr>
<th>Analytic use case</th>
<th>Type</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical characterization</strong></td>
<td>Disease Natural History</td>
<td>Amongst patients who are diagnosed with <em>&lt;insert your favorite disease&gt;</em>, what are the patient’s characteristics from their medical history?</td>
</tr>
<tr>
<td></td>
<td>Treatment utilization</td>
<td>Amongst patients who have <em>&lt;insert your favorite disease&gt;</em> , which treatments were patients exposed to amongst <em>&lt;list of treatments for disease&gt;</em> and in which sequence?</td>
</tr>
<tr>
<td></td>
<td>Outcome incidence</td>
<td>Amongst patients who are new users of <em>&lt;insert your favorite drug&gt;</em> , how many patients experienced <em>&lt;insert your favorite known adverse event from the drug profile&gt;</em> within <em>&lt;time horizon following exposure start&gt;</em>?</td>
</tr>
<tr>
<td><strong>Population-level effect estimation</strong></td>
<td>Safety surveillance</td>
<td>Does exposure to <em>&lt;insert your favorite drug&gt;</em> increase the risk of experiencing <em>&lt;insert an adverse event&gt;</em> within <em>&lt;time horizon following exposure start&gt;</em>?</td>
</tr>
<tr>
<td></td>
<td>Comparative effectiveness</td>
<td>Does exposure to <em>&lt;insert your favorite drug&gt;</em> have a different risk of experiencing <em>&lt;insert any outcome (safety or benefit)&gt;</em> within <em>&lt;time horizon following exposure start&gt;</em> , relative to <em>&lt;insert your comparator treatment&gt;</em>?</td>
</tr>
<tr>
<td><strong>Patient level prediction</strong></td>
<td>Disease onset and progression</td>
<td>Amongst patients who are diagnosed with <em>&lt;insert your favorite disease&gt;</em> , which patients will go on to have <em>&lt;another disease or related complication&gt;</em> within <em>&lt;time horizon from diagnosis&gt;</em>?</td>
</tr>
<tr>
<td></td>
<td>Treatment response</td>
<td>Amongst patients who are new users of <em>&lt;insert your favorite chronically-used drug&gt;</em> , which patients will <em>&lt;insert desired effect&gt;</em> in <em>&lt;time window&gt;</em>?</td>
</tr>
<tr>
<td></td>
<td>Treatment safety</td>
<td>Amongst patients who are new users of <em>&lt;insert your favorite drug&gt;</em> , which patients will experience <em>&lt;insert your favorite known adverse event from the drug profile&gt;</em> within <em>&lt;time horizon following exposure start&gt;</em>?</td>
</tr>
</tbody>
</table>
OHDSI is a community
Case Western Reserve University: OHDSI face-to-face documentation-a-thon
The Journey From Data to Evidence OHDSI Europe 2019

- A platform to stimulate community building: 250 participants from 27 countries
- OHDSI Europe in action: 35 posters, 8 software demos
- Educate and train the community: 5 full day tutorials

www.ohdsi-europe.org
Fudan University – OHDSI tutorials
OHDSI Korea – Study design datathon
Building the LHC of observational data science?
ICMJE guidelines

The ICMJE recommends that authorship be based on the following 4 criteria:

• Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
• Drafting the work or revising it critically for important intellectual content; AND
• Final approval of the version to be published; AND
• Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

OHDSI in action

MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)
© 2015 IMIA and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms
of the Creative Commons Attribution Non-Commercial License.
doi:10.3233/978-1-61499-564-7-574

Observational Health Data Sciences and Informatics (OHDSI): Opportunities for
Observational Researchers

George Hripcsak\textsuperscript{a}, Jon D. Duke\textsuperscript{b}, Nigam H. Shah\textsuperscript{c}, Christian G. Reich\textsuperscript{d}, Vojtech Huser\textsuperscript{e}, Martijn J. Schuemie\textsuperscript{fg}, Marc A. Suchard\textsuperscript{h}, Rae Woong Park\textsuperscript{i}, Ian Chi Kei Wong\textsuperscript{j}, Peter R. Rijnbeek\textsuperscript{k}, Johan van der Lei\textsuperscript{l}, Nicole Pratt\textsuperscript{k}, G. Niklas Norèn\textsuperscript{m}, Yu-Chuan Li\textsuperscript{n}, Paul E. Stang\textsuperscript{a}, David Madigan\textsuperscript{a}, Patrick B. Ryan\textsuperscript{a}

\textsuperscript{a} Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA
\textsuperscript{b} Regenstrief Institute, Indianapolis, IN, USA
\textsuperscript{c} Center for Biomedical Informatics Research, Stanford University, CA, USA
\textsuperscript{d} AstraZeneca PLC, Waltham, MA, USA
\textsuperscript{e} NIH Clinical Center, Bethesda, MD, USA
\textsuperscript{f} Centre for Safe Medication Practice and Research, Dept. of Pharmacology and Pharmacy, University of Hong Kong, Hong Kong
\textsuperscript{g} Janssen Research & Development, LLC, Titusville, NJ, USA
\textsuperscript{h} Dept. of Biostatistics & Dept. of Human Genetics, David Geffen School of Medicine, Uni. of California, Los Angeles, CA, USA
\textsuperscript{i} Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea
\textsuperscript{j} Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands
\textsuperscript{k} School of Pharmacy and Medical Sciences, University of South Australia, Australia
\textsuperscript{l} Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala, Sweden
\textsuperscript{m} College of Medical Science and Technology (CoMST), Taipei Medical University, Taipei, Taiwan
\textsuperscript{n} Department of Statistics, Columbia University, New York, NY, USA
Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak\textsuperscript{a,b,c,1}, Patrick B. Ryan\textsuperscript{c,d}, Jon D. Duke\textsuperscript{e}, Nigam H. Shah\textsuperscript{c,f}, Rae Woong Park\textsuperscript{c,g}, Vojtech Huser\textsuperscript{c,h}, Marc A. Suchard\textsuperscript{c,i,j,k}, Martijn J. Schuemie\textsuperscript{c,d}, Frank J. DeFalco\textsuperscript{c,d}, Adler Perotte\textsuperscript{a,l}, Juan M. Banda\textsuperscript{c,i}, Christian G. Reich\textsuperscript{c,l}, Lisa M. Schilling\textsuperscript{c,m}, Michael E. Matheny\textsuperscript{c,n,o}, Daniella Meeker\textsuperscript{c,p,q}, Nicole Pratt\textsuperscript{c,l}, and David Madigan\textsuperscript{c,s}

\textsuperscript{a}Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; \textsuperscript{b}Medical Informatics Services, New York-Presbyterian Hospital, New York, NY 10032; \textsuperscript{c}Observational Health Data Sciences and Informatics, New York, NY 10032; \textsuperscript{d}Epidemiology Analytics, Janssen Research and Development, Titusville, NJ 08560; \textsuperscript{e}Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN 46205; \textsuperscript{f}Center for Biomedical Informatics Research, Stanford University, CA 94305; \textsuperscript{g}Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea, 443-380; \textsuperscript{h}Lister Hill National Center for Biomedical Communications (National Library of Medicine), National Institutes of Health, Bethesda, MD 20894; \textsuperscript{i}Department of Biostatistics, University of California, Los Angeles, CA 90095; \textsuperscript{j}Department of Human Genetics, University of California, Los Angeles, CA 90095; \textsuperscript{k}Real World Evidence Solutions, IMS Health, Burlington, MA 01809; \textsuperscript{l}Department of Medicine, University of Colorado School of Medicine, Aurora, CO 80045; \textsuperscript{m}Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37212; \textsuperscript{n}Geriatric Research, Education and Clinical Center, VA Tennessee Valley Healthcare System, Nashville, TN 37212; \textsuperscript{o}Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089; \textsuperscript{p}Department of Pediatrics, University of Southern California, Los Angeles, CA 90089; \textsuperscript{q}Division of Health Sciences, University of South Australia, Adelaide, SA, Australia 5001; and \textsuperscript{s}Department of Statistics, Columbia University, New York, NY 10027

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 5, 2016 (received for review June 14, 2015)

2016: 17 authors
11 data sources
Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network

*Jon D. Duke, **Patrick B. Ryan, ***Marc A. Suchard, **George Hripcsak, **Peng Jin, ***Christian Reich, ***Marie-Sophie Schwalm, ****Yuriy Khoma, ***Yonghui Wu, ***Hua Xu, **Nigam H. Shah, **Juan M. Banda, and ***Martijn J. Schuemie

Epilepsia, 58(8):e101–e106, 2017
doi: 10.1111/epi.13828

Summary

Recent adverse event reports have raised the question of increased angioedema risk associated with exposure to levetiracetam. To help address this question, the Observational Health Data Sciences and Informatics research network conducted a retrospective observational new-user cohort study of seizure patients exposed to levetiracetam (n = 274,445) across 10 databases. With phenytoin users (n = 74,482) as a comparator group, propensity score-matching was conducted and hazard ratios computed for angioedema events by per-protocol and intent-to-treat analysis. Angioedema events were rare in both the levetiracetam and phenytoin groups (0.6% in per-protocol and 2.0% in intent-to-treat). No significant increase in angioedema risk with levetiracetam was seen in any individual database (hazard ratios ranging from 0.63 to 2.31). Meta-analysis showed a summary hazard ratio of 0.72 (95% confidence interval [CI] 0.39–1.31) and 0.64 (95% CI 0.32–0.78) for the per-protocol and intent-to-treat analyses, respectively. The results suggest that levetiracetam is the same or lower risk for angioedema than phenytoin, which does not currently carry a labeled warning for angioedema. Further studies are warranted to evaluate angioedema risk across all antiepileptic drugs.

Key Words: Angioedema, Levetiracetam, Anticonvulsant hypersensitivity syndrome, Pharmacovigilance, Observational research, Adverse drug reactions.
Association of Hemoglobin A1c Levels With Use of Sulfonylureas, Dipeptidyl Peptidase 4 Inhibitors, and Thiazolidinediones in Patients With Type 2 Diabetes Treated With Metformin Analysis From the Observational Health Data Sciences and Informatics Initiative

Rohit Vashisht, PhD; Kenneth Jung, PhD; Alejandro Schuler, MS; Juan M. Banda, PhD; Rae Woong Park, MD, PhD; Sanghyung Jin, MS; Kipp W. Johnson, MD, PhD; Mark M. Shervey, PhD; Hua Xu, PhD; Yonghui Wu, PhD; Karthik Natraj, PhD; George Hripcsak, MD, MS; Anthony Reckard, BS; Christian G. Reich, MD; James Weaver, MPH, MS; Martijn J. Schuemie, PhD; Patrick B. Ryan, PhD; Alison Callahan
Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, Ruijun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripko, Patrick B Ryan

Summary

Background: Uncertainty remains about the optimal monotherapy for hypertension, with current guidelines recommending any primary agent among the first-line drug classes thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydropyridine calcium channel blockers, in the absence of comorbid indications. Randomised trials have not further refined this choice.

Methods: We developed a comprehensive framework for real-world evidence that enables comparative effectiveness and safety evaluation across many drugs and outcomes from observational data encompassing millions of patients, while minimising inherent bias. Using this framework, we did a systematic, large-scale study under a new-user cohort design to estimate the relative risks of three primary (acute myocardial infarction, hospitalisation for heart failure, and stroke) and six secondary effectiveness and 46 safety outcomes comparing all first-line classes across a global network of six administrative claims and three electronic health record databases. The framework addressed residual confounding, publication bias, and p-hacking using large-scale propensity adjustment, a large set of control outcomes, and full disclosure of hypotheses tested.

Findings: Using 4.9 million patients, we generated 22,000 calibrated, propensity-score-adjusted hazard ratios (HRs) comparing all classes and outcomes across databases. Most estimates revealed no effectiveness differences between classes; however, thiazide or thiazide-like diuretics showed better primary effectiveness than angiotensin-converting enzyme inhibitors: acute myocardial infarction (HR 0.84, 95% CI 0.75–0.95), hospitalisation for heart failure (0.83, 0.77–0.90), and stroke (0.82, 0.71–0.94).

OHDSI in action: LEGEND-HTN

Oct2018→2019:
11 authors
9 sources
1.4m TCOs
From question to preliminary results in 1 day.

Oct 2018 → 2019:
23 authors
11 data sources
OHDSI in action: Oxford study-a-thon

"To compare the risk of post-operative complications and mortality between unicompartmental vs total knee replace

The Lancet
Rheumatology

Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study


Summary
Background There is uncertainty around whether to use unicompartmental knee replacement (UKR) or total knee replacement (TKR) for individuals with osteoarthritis confined to a single compartment of the knee. We aimed to emulate the design of the Total or Partial Knee Arthroplasty Trial (TOPKAT) using routinely collected data to assess whether the efficacy results reported in the trial translate into effectiveness in routine practice, and to assess comparative safety.

Dec2018→2019: 26 authors
5 sources
### Contributors

Each chapter lists one or more chapter leads. These are the people who wrote the chapter. However, there are many others that have contributed to the Book of OHDSI, which we would like to acknowledge here:

<table>
<thead>
<tr>
<th>Name</th>
<th>Name</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamed Abdiash</td>
<td>Mustafa Ascha</td>
<td>Mark Beno</td>
</tr>
<tr>
<td>Clair Blacketer</td>
<td>David Blatt</td>
<td>Brian Christian</td>
</tr>
<tr>
<td>Gino Cloft</td>
<td>Frank DeFalco</td>
<td>Sara Dempster</td>
</tr>
<tr>
<td>Jon Duke</td>
<td>Sergio Eslava</td>
<td>Clark Evans</td>
</tr>
<tr>
<td>Thomas Falconer</td>
<td>George Hripcsak</td>
<td>Vojtech Huser</td>
</tr>
<tr>
<td>Mark Khayter</td>
<td>Greg Klebanov</td>
<td>Kristin Kostka</td>
</tr>
<tr>
<td>Bob Laneese</td>
<td>Wanda Lattimore</td>
<td>Chun Li</td>
</tr>
<tr>
<td>David Madigan</td>
<td>Sindoosha Malay</td>
<td>Harry Menegay</td>
</tr>
<tr>
<td>Akihiko Nishimura</td>
<td>Ellen Palmer</td>
<td>Nirav Patel</td>
</tr>
<tr>
<td>Jose P. Pacheco</td>
<td>Emily Pajer</td>
<td>John Pajer</td>
</tr>
<tr>
<td>Christina Parker</td>
<td>Patrick Parker</td>
<td>Paola Silvestri</td>
</tr>
<tr>
<td>Anthony Soos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matt Spoussz</td>
<td>Marc Stenard</td>
<td>Joes Sweerdien</td>
</tr>
<tr>
<td>Devin Tian</td>
<td>Don Torok</td>
<td>Kees van Bochoove</td>
</tr>
<tr>
<td>Mai Van Zandt</td>
<td>Erica Voss</td>
<td>Kristin Waite</td>
</tr>
<tr>
<td>Mike Warle</td>
<td>Jamie Weaver</td>
<td>James Wiggins</td>
</tr>
<tr>
<td>Andrew Williams</td>
<td>Chan You Seng</td>
<td></td>
</tr>
</tbody>
</table>

**2019: 56 contributors!**
Why do we need more collaboration?

• We want to learn from as many data sources as the world as possible (replicability, generalizability, heterogeneity)
  – Each data partner contributes source data understanding and shares in interpreting their results in the context of the entire network

• Large scale evidence generation requires large scale collaboration for interpretation
  – LEGEND : One causal evidence system → Many clinical insights to inform different health decisions
Building the LHC of observational data science?
What will be the research we do together that generates >1000 co-author papers?

• Methods research:
  “Examining data heterogeneity across a global health network”
  “Development and evaluation of methods for integrating causal inference design and machine learning algorithms for patient-level estimation”

• Open-source development:
  “Implementation of a large-scale analytics ecosystem to enable evidence generation within health systems and across a global health network”
  “Validation of an international phenotype library to define and identify disease across electronic health record systems”

• Clinical applications:
  “Characterization of disease incidence and treatment utilization patterns across the world”
  “Comprehensive comparative safety and effectiveness of treatments for <every disease>: an OHDSI LEGEND study”
OHDSI’s areas of focus:
Continuing our journey in 2020…

- **Evidence generation and dissemination:** Evolve from promising proof-of-concept to impactful production
- Expand engagement of the OHDSI data network in the evidence generation process, starting with characterization of concept prevalence
- Increase adoption of existing community data standards through improved documentation on shared ETL conventions and user guides
- Improve community connections between methods research, open-source development, and clinical applications to promote greater adoption of community best practices

…Or so we thought….

…and then a pandemic happened….
OHDSI COVID-19 Study-a-thon kickoff
26Mar2020 3amEST

https://www.ohdsi.org/covid-19-updates/
### OHDSI COVID-19 Study-a-thon Study Tracker

#### Analytic use case

<table>
<thead>
<tr>
<th>Study</th>
<th>Lit Review and protocol development</th>
<th>Phenotype development and evaluation</th>
<th>Study package development</th>
<th>Study execution across network</th>
<th>Clinical review and dissemination</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characterization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19 positive patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19 +ve subgroup analyses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influenza, symptoms, and complications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invasive treatments for respiratory distress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other questions?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prediction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Who presenting with flu, symptoms, or complications will be admitted to hospital?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Who sent home with symptoms will progress to require hospitalization?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Who admitted to hospital will require intensive care services or die?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other questions?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Estimation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of hydroxychloroquine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of IL6 and JAK inhibitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of HIV protease inhibitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of HepC protease inhibitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of ACE inhibitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other questions?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Tracking our collaboration**

26Mar2020 3amET
## Where did we end up by 29Mar2020 7pmET?

### OHDSI COVID-19 Study-a-thon Study Tracker

<table>
<thead>
<tr>
<th>Analytic use case</th>
<th>Study</th>
<th>Lit Review and protocol development</th>
<th>Phenotype development and evaluation</th>
<th>Study package development</th>
<th>Study execution across network</th>
<th>Clinical review and dissemination</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characterization</strong></td>
<td>COVID-19 positive patients</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>COVID-19 IVE subgroup analyses</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>Influenza, symptoms, and complications</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>Invasive treatments for respiratory distress</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>other questions?</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td><strong>Prediction</strong></td>
<td>1) Who presenting with flu, symptoms, or complications will be admitted to hospital?</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>2) Who sent home with symptoms will progress to require hospitalization?</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>3) Who admitted to hospital will require intensive care services or die?</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>other questions?</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td><strong>Estimation</strong></td>
<td>Effects of hydroxychloroquine</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>Effects of IL6 and JAK inhibitors</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>Effects of HIV protease inhibitors</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>Effects of HepC protease inhibitors</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>Effects of ACE inhibitors</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>other questions?</td>
<td>Green</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Green</td>
</tr>
</tbody>
</table>

*To be done*  
*In progress*  
*Results in, more to come*  
*Completed*
What have we done?

In only 88 hours, we have:

- Convened 351 participants brought together from 30 countries
- Held 12 Global Huddles, >100 collaborator calls, >13,000 chat messages
- Engaged 15 concurrent channels
- Reviewed >10,000 publications
- Drafted 9 protocols
- Released 13 study packages
- Designed 355 cohort definitions
- Assembled a distributed data network with 37 partners signed on to execute studies
3 things that we did in 4 days together that nobody had ever done before

• First large-scale international characterization of COVID hospitalized patients
• First prediction model externally validated on COVID patients to support triage to ‘flatten the curve’
• Largest study ever conducted on the safety of hydroxychloroquine
Open collaboration requires FULL transparency in every step of the research process

- Study registered in ENCEPP with full protocol posted: http://www.encepp.eu/encepp/viewResource.htm?id=34498
- Phenotype definitions and analysis specifications are both human-readable and computer-executable using ATLAS against any OMOP CDM: https://atlas.ohdsi.org/#/estimation/cca/6
- Analysis source code freely available and directly downloadable: https://github.com/ohdsi-studies/Covid19EstimationHydroxychloroquine
- Manuscript posted on Medrxiv while awaiting peer-review: https://www.medrxiv.org/content/10.1101/2020.04.08.20054551v1
- All analysis results available for public exploration through interactive R shiny application: http://evidence.ohdsi.org/Covid19EstimationHydroxychloroquine
Safety of hydroxychloroquine, alone and in combination with azithromycin, in light of rapid wide-spread use for COVID-19: a multinational, network cohort and self-controlled case series study


doi: https://doi.org/10.1101/2020.04.08.20054551

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.
Methods

- New user cohort studies were conducted including 16 severe adverse events (SAEs).
- Rheumatoid arthritis patients aged 18+ and initiating hydroxychloroquine were compared to those initiating sulfasalazine and followed up over 30 days.
- Self-controlled case series (SCCS) were conducted to further establish safety in wider populations.
- Separately, SAEs associated with hydroxychloroquine-azithromycin (compared to hydroxychloroquine-amoxicillin) were studied.
- Data comprised 14 sources of claims data or electronic medical records from Germany, Japan, Netherlands, Spain, UK, and USA.
- Propensity score stratification and calibration using negative control outcomes were used to address confounding. Cox models were fitted to estimate calibrated hazard ratios (CalHRs) according to drug use.
- Estimates were pooled where I²<40%.

https://www.medrxiv.org/content/10.1101/2020.04.08.20054551v1
**Key findings**

- **HCQ** appears safe in short term in RA, but long-term use may be associated with increased CV mortality
- **HCQ+azithromycin** increases 30-day risk of heart failure and cardiovascular mortality

---

**Figure 1.** Source-specific and meta-analytic cardiovascular risk estimates for hydroxychloroquine vs sulfasalazine and azithromycin vs amoxicillin new users during 30-day follow-up

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Database</th>
<th>CalHR (95% CI)</th>
<th>I²</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV-related mortality</td>
<td>Optum</td>
<td>1.04 (0.31-3.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VA</td>
<td>2.47 (0.45-13.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Meta-analysis</td>
<td>1.21 (0.51-3.03)</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Chest pain or angina</td>
<td>AmbEMR</td>
<td>1.09 (0.71-1.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCAE</td>
<td>0.91 (0.72-1.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CPRD</td>
<td>0.90 (0.36-2.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DAGEst Patients</td>
<td>0.41 (0.08-2.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IMRD</td>
<td>1.11 (0.39-3.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MDCD</td>
<td>0.96 (0.56-1.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MDCR</td>
<td>0.93 (0.56-1.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OpenClaims</td>
<td>0.92 (0.50-1.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optum</td>
<td>0.82 (0.60-1.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PanTher</td>
<td>1.15 (0.51-1.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VA</td>
<td>1.04 (0.73-1.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Meta-analysis</td>
<td>0.96 (0.83-1.11)</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Heart failure</td>
<td>AmbEMR</td>
<td>1.17 (0.56-2.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCAE</td>
<td>1.40 (0.55-3.90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CPRD</td>
<td>1.32 (0.89-2.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DAGEst Patients</td>
<td>0.65 (0.10-4.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IMRD</td>
<td>5.14 (0.29-89.87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MDCD</td>
<td>1.83 (0.80-4.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MDCR</td>
<td>0.75 (0.42-1.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OpenClaims</td>
<td>0.94 (0.40-2.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optum</td>
<td>1.20 (0.74-1.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PanTher</td>
<td>1.29 (0.81-1.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VA</td>
<td>1.28 (0.72-2.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Meta-analysis</td>
<td>0.95 (0.88-1.02)</td>
<td></td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

HCQ=hydroxychloroquine; SSZ=sulfasalazine; AZM=azithromycin (plus concurrent hydroxychloroquine exposure); AMX=amoxicillin (plus concurrent hydroxychloroquine exposure); CalHR=calibrated hazard ratio; CI=confidence interval; I²=estimate heterogeneity statistic. Meta-analytic estimates reported where I²<0.4. All database-specific estimates are reported in Appendix Table S7. AmbEMR=IQVIA Ambulatory EMR; CCAE=IBM Commercial Database; CPRD=Clinical Practice Research Datalink, DAGEst Patients=IQVIA Disease Analyzer Germany; IMRD=IQVIA UK Integrated Medical Record Data; MDCD=IBM IBM Multi-state Medicaid; MDCR=IBM Medicare Supplemental Database; OpenClaims=IQVIA Open Claims; Optum=Optum Clininformatics Datamart; PanTher=Optum PanTherapeutic Electronic Health Record; VA=Veteran’s Administration Database

https://www.medrxiv.org/content/10.1101/2020.04.08.20054551v1

*under peer review*
Open collaboration requires FULL transparency in every step of the research process

- Protocol and analysis source code freely available and directly downloadable: https://github.com/ohdsi-studies/Covid19HospitalizationCharacterization

- Phenotype definitions are both human-readable and computer-executable using ATLAS against any OMOP CDM: https://atlas.ohdsi.org/

- Manuscript posted on Medrxiv while awaiting peer-review: https://www.medrxiv.org/content/10.1101/2020.04.22.20074336v1

- All analysis results available for public exploration through interactive R shiny application: http://evidence.ohdsi.org/Covid19CharacterizationHospitalization/

- The study is a living evidence repository: any data partners can execute analysis and share aggregate results at any point, including updates as data accumulate
Key findings*

- Rates of comorbidities and medication use are high among individuals hospitalized with COVID-19.

- COVID-19 patients are more likely to be male and appear to be younger and, in the US, generally healthier than those typically admitted with influenza.

---

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and should not be used to guide clinical practice.
### OHDSI COVID and Influenza Hospitalization

#### Cohort Counts

<table>
<thead>
<tr>
<th>Covariate Name</th>
<th>CUIMC</th>
<th>DCMC</th>
<th>HRA</th>
<th>STARR-OMOP</th>
<th>VA-OMOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>age group: 00-04</td>
<td>&lt;1.0%</td>
<td>2.1%</td>
<td>0.15</td>
<td>&lt;1.7%</td>
<td>&lt;1.7%</td>
</tr>
<tr>
<td>age group: 05-09</td>
<td>&lt;1.0%</td>
<td>1.1%</td>
<td>0.54</td>
<td>&lt;3.2%</td>
<td>&lt;1.7%</td>
</tr>
<tr>
<td>age group: 10-19</td>
<td>1.2%</td>
<td>11.5%</td>
<td>0.30</td>
<td>&lt;7.1%</td>
<td>1.1%</td>
</tr>
<tr>
<td>age group: 20-24</td>
<td>3.3%</td>
<td>1.0%</td>
<td>0.61</td>
<td>9.1%</td>
<td>0.28</td>
</tr>
<tr>
<td>age group: 25-29</td>
<td>5.4%</td>
<td>2.3%</td>
<td>0.21</td>
<td>&lt;7.1%</td>
<td>2.4%</td>
</tr>
<tr>
<td>age group: 30-34</td>
<td>5.5%</td>
<td>2.3%</td>
<td>0.19</td>
<td>9.1%</td>
<td>0.31</td>
</tr>
<tr>
<td>age group: 35-39</td>
<td>5.6%</td>
<td>2.2%</td>
<td>0.26</td>
<td>8.4%</td>
<td>0.20</td>
</tr>
<tr>
<td>age group: 40-44</td>
<td>40.1%</td>
<td>3.3%</td>
<td>0.19</td>
<td>6.2%</td>
<td>0.29</td>
</tr>
<tr>
<td>age group: 45-49</td>
<td>8.4%</td>
<td>2.9%</td>
<td>0.19</td>
<td>8.1%</td>
<td>0.29</td>
</tr>
<tr>
<td>age group: 50-54</td>
<td>0.6%</td>
<td>3.1%</td>
<td>0.32</td>
<td>7.2%</td>
<td>0.27</td>
</tr>
<tr>
<td>age group: 55-59</td>
<td>9.3%</td>
<td>3.3%</td>
<td>0.49</td>
<td>6.5%</td>
<td>0.25</td>
</tr>
<tr>
<td>age group: 60-64</td>
<td>10.2%</td>
<td>3.2%</td>
<td>0.19</td>
<td>4.5%</td>
<td>0.21</td>
</tr>
<tr>
<td>age group: 65-69</td>
<td>9.5%</td>
<td>3.2%</td>
<td>0.19</td>
<td>3.6%</td>
<td>0.19</td>
</tr>
<tr>
<td>age group: 70-74</td>
<td>1.1%</td>
<td>3.3%</td>
<td>0.33</td>
<td>2.6%</td>
<td>0.15</td>
</tr>
<tr>
<td>age group: 75-79</td>
<td>6.4%</td>
<td>2.6%</td>
<td>0.37</td>
<td>2.3%</td>
<td>0.15</td>
</tr>
<tr>
<td>age group: 80-84</td>
<td>5.2%</td>
<td>2.3%</td>
<td>0.19</td>
<td>1.2%</td>
<td>0.11</td>
</tr>
<tr>
<td>age group: 85-89</td>
<td>5.2%</td>
<td>2.2%</td>
<td>0.4%</td>
<td>4.7%</td>
<td>0.07</td>
</tr>
<tr>
<td>age group: 90-94</td>
<td>&lt;1.0%</td>
<td>&lt;0.1%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
</tr>
<tr>
<td>condition_era group during day -30 through day 0 days relative to index: Abdominal abscess</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
</tr>
<tr>
<td>condition_era group during day -30 through day 0 days relative to index: Abdominal aortic aneurysm</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
</tr>
<tr>
<td>condition_era group during day -30 through day 0 days relative to index: Abdominal aortic aneurysm without rupture</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
</tr>
<tr>
<td>condition_era group during day -30 through day 0 days relative to index: Abdominal aortic ectasia</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
</tr>
<tr>
<td>condition_era group during day -30 through day 0 days relative to index: Abdominal distraction, gaseous</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
<td>&lt;1.0%</td>
</tr>
</tbody>
</table>

[Visit OHDSI COVID and Influenza Hospitalization](http://evidence.ohdsi.org/Covid19CharacterizationHospitalization/)
COVID-19 Study-A-Thon

ohdsi.org/covid-19-updates
Characterizing Health Associated Risks, and Your Baseline Disease In SARS-COV-2 (CHARYBDIS)

Objectives: 1) Describe the baseline demographic, clinical characteristics, treatments and outcomes of interest among individuals tested for SARS-CoV-2 and/or diagnosed with COVID-19 overall and stratified by sex, age and specific comorbidities; 2) Describe characteristics and outcomes of hospitalized influenza patients between September 2017 and April 2018 compared to the COVID-19 population.

Data Sources: OHDSI studies around COVID-19 currently leverage datasets from Asia, Europe and North America. We are actively looking for data partners to join the journey!

1 IRB / 1 Study Package = Dozens of studies & papers that will be submitted for peer review!

Study Topics: COVID-19 and ...

- Asthma
- Cancer
- Cardiac Outcomes
- Chronic Kidney Disease
- COPD
- Elderly
- End-Stage Renal Disease
- Gender Differences
- Heart Disease
- Hepatitis C
- HIV infection
- Hypertension
- Immune Disorders
- Obesity
- Pediatrics
- Pregnant Women
- Testing
- Tuberculosis
- Type 2 Diabetes

... And more!

OHDSI research into COVID-19 has generated many studies, including two currently in the peer-review process. Our baseline characterization of hospitalized COVID-19 patients is available on MedRxiv.

www.ohdsi.org/Covid-19-Updates
Objective: The aim of this study is to assess the comparative safety and effectiveness of all emerging drug therapies used in COVID-19 treatments ...

... administered during hospitalization and prior to intensive services.

... administered during hospitalization after initiating intensive services.

... administered after COVID-19 positive testing and prior to hospitalization.

Methods: Apply a comparative cohort design, comparing alternative treatments amongst 1) drugs with in vitro activity on SARS-COV-2 virus, and 2) drugs used as adjuvant therapy for COVID-19 disease. Building off **OHDSI LEGEND** framework, all treatment comparisons will be systematically evaluated for a large range of safety and effectiveness outcomes.

Data Setting: Analyses will be run across data partners in the OHDSI network across Asia, Europe and North America, using EHR and claims data from primary and secondary care.

Any researchers with data in OMOP CDM are encouraged to collaborate.
Join the journey!

[Logo]

ohdsi.org
Email: contact@ohdsi.org
Twitter: @OHDSI