The Journey of OHDSI: Where have we been?

George Hripcsak, MD, MS
Columbia University, New York, USA
NewYork-Presbyterian Hospital, New York, USA
OHDSI Europe
Odyssey (*noun*): \oh-d-si\n
1. A long journey full of adventures
2014: OHDSI’s 1st F2F meeting, Columbia University, NY, USA
2015: OHDSI’s 2nd F2F, Stanford University, CA, USA

2015: OHDSI’s 3rd F2F, National Library of Medicine, MD, USA

2015: 1st Annual OHDSI Symposium, Washington DC, USA
2016: 2nd Annual OHDSI Symposium, Washington DC, USA
2017: OHDSI Hadoop hack-a-thon, QuintilesIMS, PA, USA
2018: 1st OHDSI Europe Symposium, Rotterdam, NL

YOU ARE HERE
European OHDSI Symposium

Bridging Europe

23-24th March 2018, Rotterdam, The Netherlands

http://www.ohdsi-europe.org/
2014: OHDSI’s 1st F2F meeting, Columbia University, NY, USA

2015: OHDSI’s 2nd F2F, Stanford University, CA, USA

2015: OHDSI’s 3rd F2F, National Library of Medicine, MD, USA

2017: OHDSI Korea Symposium, Ajou University, Suwon, South Korea

2017: OHDSI Hadoop hack-a-thon, QuintilesIMS, PA, USA

2015: 1st Annual OHDSI Symposium, Washington DC, USA

2016: 2nd Annual OHDSI Symposium, Washington DC, USA

2017: 3rd Annual OHDSI Symposium, Bethesda, MD, USA

2018: 1st OHDSI Europe Symposium, Rotterdam, NL

You are here.
Welcome!
OHDSI is an open science community
OHDSI’s mission

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

We’re all in this journey together...
OHDSI’ community engagement

• Weekly community web conferences for all collaborators to share their research ideas and progress
• 15 workgroups for solving shared problems of interest
  – Common Data Model, Population-level Estimation, Patient-level Prediction, Architecture, Phenotype, NLP, GIS, Oncology, …
• Active community online discussion: forums.ohdsi.org
• 594 distinct users have made 8,894 posts on 1,631 topics:
  – Implementers, Developers, Researchers, CDM Builders, Vocabulary users, OHDSI in Korea, OHDSI in China, OHDSI in Europe
OHDSI is
an international data network
Data across the OHDSI community

• 84 organizations have access to 64 different databases
• Patient-level data from various perspectives:
  – Electronic health records, administrative claims, hospital systems, clinical registries, health surveys, biobanks
• Collectively, totaling 1.26 billion patient records
• Data in 17 different countries, with 115 million patient records from outside US

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

Journey of an open community data standard
Journey of an open community data standard

**OMOP CDM v1**
- **Nov2009** Strawman
- **Focus on drug safety surveillance, methods research**

**OMOP CDM v2**
- **June2012** Expanded to support comparative effectiveness research

**OMOP CDM v4**
- **2015-2017** Improvements to support additional analytical use cases of the community

**OMOP CDM v5**
- **Nov2014** Expanded to support medical device research, health economics, biobanks, freetext clinical notes; vocabulary-driven domains

**OMOP CDM v5.0.1**
- **OMOP CDM v5.1**
- **OMOP CDM v5.2**

**OMOP CDM v5.0.1**

**https://github.com/OHDSI/CommonDataModel**
Thanks to Rimma Belenkaya, Christian Reich, and Clair Blacketer for leading our community data model stewardship!
OHDSI’s standardized vocabularies

• 78 Vocabularies across 32 domains
  – MU3 standards: SNOMED, RxNorm, LOINC
  – Disparate sources: ICD9CM, ICD10(CM), Read, NDC, Gemscript, CPT4, HCPCS…

• 5,720,848 concepts
  – 2,361,965 standard concepts
  – 3,022,623 source codes
  – 336,260 classification concepts

Thank you Christian and the Odysseus team for continue to steward, maintain, and improve this invaluable resource for the entire community!

• 32,612,650 concept relationships

Publicly available for download at: http://athena.ohdsi.org/
OHDSI is collaborating to generate reliable evidence
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
### Estimation methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>New-user cohort</td>
<td>New-user cohort studies using large-scale regression for propensity and outcome models</td>
</tr>
<tr>
<td>Self-Controlled Case Series</td>
<td>Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.</td>
</tr>
<tr>
<td>Self-Controlled Cohort</td>
<td>A self-controlled cohort design, where time preceding exposure is used as control.</td>
</tr>
<tr>
<td>IC Temporal Pattern Disc.</td>
<td>A self-controlled design, but using temporal patterns around other exposures and outcomes to correct for time-varying confounding.</td>
</tr>
<tr>
<td>Case-control</td>
<td>Case-control studies, matching controls on age, gender, provider, and visit date. Allows nesting of the study in another cohort.</td>
</tr>
<tr>
<td>Case-crossover</td>
<td>Case-crossover design including the option to adjust for time-trends in exposures (so-called case-time-control).</td>
</tr>
</tbody>
</table>

---

Thank you Martijn Schuemie, Marc Suchard, Peter Rijnbeek, and Jenna Reps for leading methods research and development efforts!

---

### Supporting packages

<table>
<thead>
<tr>
<th>Package</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database Connector</td>
<td>Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.</td>
</tr>
<tr>
<td>Sql Render</td>
<td>Generate SQL on the fly for the various SQL dialects.</td>
</tr>
<tr>
<td>Cyclops</td>
<td>Highly efficient implementation of regularized logistic, Poisson and Cox regression.</td>
</tr>
<tr>
<td>Ohdsi R Tools</td>
<td>Support tools that didn’t fit other categories, including tools for maintaining R libraries.</td>
</tr>
</tbody>
</table>
Journey toward open-source analytics development

• 88 developers on 101 OHDSI GitHub repositories

• Applications released for:
  – CDM ETL design and implementation
  – Clinical characterization (ACHILLES, ATLAS)
  – Population-level effect estimation
  – Patient-level prediction
  – OHDSI network studies (protocol + source code, ARACHNE)
Journey toward open-source analytics development

ATLAS/WebAPI – a single community platform for:
- vocabulary browsing
- database characterization
- cohort definition
- cohort characterization
- incidence rate estimation
- patient profile exploration
- population-level effect estimation design
- patient-level prediction design

Demo at http://ohdsi.org/web/ATLAS

Thank you teams from Columbia, Google, Cloudera, Erasmus MC, Odysseus, BlueCrossBlueShield-South Carolina, Regenstrief, Janssen for contributing to the ATLAS 2.3 release!
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
Open Science

Open science

Data + Analytics + Domain expertise

Generate evidence

Open source software

Enable users to do something

Standardized, transparent workflows

Database summary
Cohort definition
Cohort summary
Compare cohorts
Exposure-outcome summary
Effect estimation & calibration
Compare databases
How OHDSI works

Source data warehouse, with identifiable patient-level data

Standardized, de-identified patient-level database (OMOP CDM v5)

Standardized large-scale analytics

Analysis results

OHDSI Data Partners

OHDSI Coordinating Center

Data network support
Analytics development and testing
Research and education

Summary statistics results repository

OHDSI.org
OHDSI in Action
Treatment Pathways

Global stakeholders
- Public
- Academics
- Industry
- Regulator

Evidence
- RCT, Obs

Conduits
- Social media
- Lay press
- Literature
- Guidelines
- Advertising
- Formulary
- Labels

Local stakeholders
- Family
- Patient
- Clinician
- Consultant

Inputs
- Indication
- Feasibility
- Cost
- Preference
## OHDSI partners for this query (250M)

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name</th>
<th>Description</th>
<th>Population, millions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUSOM</td>
<td>Ajou University School of Medicine</td>
<td>South Korea; inpatient hospital EHR</td>
<td>2</td>
</tr>
<tr>
<td>CCAE</td>
<td>MarketScan Commercial Claims and Encounters</td>
<td>US private-payer claims</td>
<td>119</td>
</tr>
<tr>
<td>CPRD</td>
<td>UK Clinical Practice Research Datalink</td>
<td>UK; EHR from general practice</td>
<td>11</td>
</tr>
<tr>
<td>CUMC</td>
<td>Columbia University Medical Center</td>
<td>US; inpatient EHR</td>
<td>4</td>
</tr>
<tr>
<td>GE</td>
<td>GE Centricity</td>
<td>US; outpatient EHR</td>
<td>33</td>
</tr>
<tr>
<td>INPC</td>
<td>Regenstrief Institute, Indiana Network for Patient Care</td>
<td>US; integrated health exchange</td>
<td>15</td>
</tr>
<tr>
<td>JMDC</td>
<td>Japan Medical Data Center</td>
<td>Japan; private-payer claims</td>
<td>3</td>
</tr>
<tr>
<td>MDCD</td>
<td>MarketScan Medicaid Multi-State</td>
<td>US; public-payer claims</td>
<td>17</td>
</tr>
<tr>
<td>MDCR</td>
<td>MarketScan Medicare Supplemental and Coordination of Benefits</td>
<td>US; private and public-payer claims</td>
<td>9</td>
</tr>
<tr>
<td>OPTUM</td>
<td>Optum ClinFormatics</td>
<td>US; private-payer claims</td>
<td>40</td>
</tr>
<tr>
<td>STRIDE</td>
<td>Stanford Translational Research Integrated Database Environment</td>
<td>US; inpatient EHR</td>
<td>2</td>
</tr>
<tr>
<td>HKU</td>
<td>Hong Kong University</td>
<td>Hong Kong; EHR</td>
<td>1</td>
</tr>
</tbody>
</table>
T2DM: All databases

**Treatment pathways for diabetes**

- **Only drug**
- **First drug**
- **Second drug**

<table>
<thead>
<tr>
<th>Drug</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metformin</td>
<td>29.42%</td>
</tr>
<tr>
<td>Pioglitazone</td>
<td></td>
</tr>
<tr>
<td>Sitagliptin</td>
<td></td>
</tr>
<tr>
<td>Glipizide</td>
<td></td>
</tr>
<tr>
<td>Glimepiride</td>
<td></td>
</tr>
<tr>
<td>Gliclazide</td>
<td></td>
</tr>
<tr>
<td>Glyburide</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone</td>
<td></td>
</tr>
<tr>
<td>Insulin, Glargine, Human</td>
<td></td>
</tr>
<tr>
<td>Exenatide</td>
<td></td>
</tr>
<tr>
<td>Insulin, Aspart, Human</td>
<td></td>
</tr>
<tr>
<td>Liraglutide</td>
<td></td>
</tr>
<tr>
<td>Saxagliptin</td>
<td></td>
</tr>
<tr>
<td>Insulin, Lispro, Human</td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td></td>
</tr>
<tr>
<td>Insulin, Isophane, Human</td>
<td></td>
</tr>
</tbody>
</table>
OHDSI’s first large scale study
August 2010: “Among patients in the UK General Practice Research Database, the use of oral bisphosphonates was not significantly associated with incident esophageal or gastric cancer.”

Sept 2010: “In this large nested case-control study within a UK cohort [General Practice Research Database], we found a significantly increased risk of oesophageal cancer in people with previous prescriptions for oral bisphosphonates.”
Distribution of possible results for one hypothesis

Stat signif > 1

OR

Databases

Methods
Distribution of possible results for one hypothesis

OR

Stat signif > 1

Databases

Methods
Distribution of possible results for one hypothesis
Distribution of possible results for one hypothesis
Distribution of possible results for one hypothesis

OR

Databases

Methods
Reproducible research

1. Address confounding that is measured
   • Propensity stratification
   • *Systematic* (not manual) variable selection
     • Balance 58,285 variables (“Table 1”)

After stratification on the propensity score, all 58,285 covariates have standardized difference of mean < 0.1
Reproducible research

2. Unmeasured (residual) confounding
   • Confidence interval calibration
     • Adjust for all uncertainty, not just sampling
   • Many negative controls
     • Unique to OHDSI (PNAS in press)

After calibration, 4% have $p < 0.05$ (was 16%)
Reproducible research

3. Multiple databases, locations, practice types
   • Exploit international OHDSI network
Reproducible research

4. Open: publish all
   • Hypotheses
   • Code
   • Parameters
   • Runs

[URL]
Potential Signals of Serious Risks/New Safety Information Identified by the FDA Adverse Event Reporting System (FAERS) between October - December 2015

| Keppra (levetiracetam) tablet, oral solution, injection | Angioedema | FDA is evaluating the need for regulatory action. |

- Protocol completed, code tested, study announced
- 50 viewed protocol, 25 viewed the code, and 7 sites ran the code on 10 databases (5 claims / 5 EHR), 59,367 levetiracetam patients matched with 74,550 phenytoin patients
No evidence of increased angioedema risk with levetiracetam use compared with phenytoin use

“The study is focused, appears well designed, and provides new insight that should be of interest to clinicians and regulators... This is an important contribution to improved pharmacovigilance.”

Add word to title, move diagram from supplement to body
How can we improve the literature
Drawing reproducible conclusions

Effect size (1 = no effect)

Standard error

P = 0.05
Drawing reproducible conclusions

- Not significant
- Protect
- Harmful
- P = 0.05
Drawing reproducible conclusions

85% of exposure-outcome pairs significant

29,982 estimates
11,758 papers
Drawing reproducible conclusions

29,982 estimates
11,758 papers

Publication bias
P hacking
5. Carry out on aligned hypotheses at scale

Duloxetine vs. Sertraline for these 22 outcomes:

<table>
<thead>
<tr>
<th>Acute liver injury</th>
<th>Hypotension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute myocardial infarction</td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>Alopecia</td>
<td>Insomnia</td>
</tr>
<tr>
<td>Constipation</td>
<td>Nausea</td>
</tr>
<tr>
<td>Decreased libido</td>
<td>Open-angle glaucoma</td>
</tr>
<tr>
<td>Delirium</td>
<td>Seizure</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>Stroke</td>
</tr>
<tr>
<td>Fracture</td>
<td>Suicide and suicidal ideation</td>
</tr>
<tr>
<td>Gastrointestinal hemorrhage</td>
<td>Tinnitus</td>
</tr>
<tr>
<td>Hyperprolactinemia</td>
<td>Ventricular arrhythmia and sudden cardiac death</td>
</tr>
<tr>
<td>Hyponatremia</td>
<td>Vertigo</td>
</tr>
</tbody>
</table>
## Many treatments at once

<table>
<thead>
<tr>
<th>Type</th>
<th>Class</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug</td>
<td>Atypical</td>
<td>Bupropion</td>
</tr>
<tr>
<td>Drug</td>
<td>Atypical</td>
<td>Mirtazapine</td>
</tr>
<tr>
<td>Procedure</td>
<td>ECT</td>
<td>Electroconvulsive therapy</td>
</tr>
<tr>
<td>Procedure</td>
<td>Psychotherapy</td>
<td>Psychotherapy</td>
</tr>
<tr>
<td>Drug</td>
<td>SARI</td>
<td>Trazodone</td>
</tr>
<tr>
<td>Drug</td>
<td>SNRI</td>
<td>Desvenlafaxine</td>
</tr>
<tr>
<td>Drug</td>
<td>SNRI</td>
<td>duloxetine</td>
</tr>
<tr>
<td>Drug</td>
<td>SNRI</td>
<td>venlafaxine</td>
</tr>
<tr>
<td>Drug</td>
<td>SSRI</td>
<td>Citalopram</td>
</tr>
<tr>
<td>Drug</td>
<td>SSRI</td>
<td>Escitalopram</td>
</tr>
<tr>
<td>Drug</td>
<td>SSRI</td>
<td>Fluoxetine</td>
</tr>
<tr>
<td>Drug</td>
<td>SSRI</td>
<td>Paroxetine</td>
</tr>
<tr>
<td>Drug</td>
<td>SSRI</td>
<td>Sertraline</td>
</tr>
<tr>
<td>Drug</td>
<td>SSRI</td>
<td>vilazodone</td>
</tr>
<tr>
<td>Drug</td>
<td>TCA</td>
<td>Amitriptyline</td>
</tr>
<tr>
<td>Drug</td>
<td>TCA</td>
<td>Doxepin</td>
</tr>
<tr>
<td>Drug</td>
<td>TCA</td>
<td>Nortriptyline</td>
</tr>
</tbody>
</table>
Drawing reproducible conclusions

11% of exposure-outcome pairs are significant once calibrated
Large-scale estimation for depression

• How do we use it? Troll for effects?

• Professor what should I study this year?
  – Simple, go to Pubmed and find the smallest p-values in the literature; surely those must be the most significant things to study

• Which is safer?
  • Seizure in 0.0000000001 to 0.0000000002 (p=0.00001)
  • Seizure in 0 to 0.2 (p=.45)

• Large-scale studies become the literature
  • Come with hypothesis and ask a question
Drawing reproducible conclusions

- Current literature is a data dredging machine
  - Dredging is not about how many things you test; it is about what you secretly throw away
- Open science approach effectively replaces the observational literature
- Collaborative open effort
• Incidence of side effects
  • Develop condition for first time after getting drug
  • Within time at risk
• Any drug on the world market
• Any condition
• Absolute risk
  • Not causal (Characterization)
• On the Internet