OHDSI’s Journey: Where we’ve been, where we can go

Patrick Ryan, PhD
Janssen Research and Development
Columbia University Medical Center
10 January 2017
Odyssey (noun): \oh-d-si\n
1. A long journey full of adventures

http://www.merriam-webster.com/dictionary/odyssey
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’s areas of focus

- Methodological research
- Open-source analytics development
- Clinical applications

Observational data management
Clinical characterization
Population-level estimation
Patient-level prediction
Journey toward best practices in data management
Journey toward best practices in data quality assessment

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

Objective: Harmonized data quality (DQ) assessment can establish a common understanding of the strength of data quality for operational analytics, quality improvement, and research. We propose a comprehensive unified terminology with a framework to support a common approach to DQ assessment.

Materials and Methods: DQ publications, informatics programs, and operational manuals from seven institutions were reviewed to identify potential DQ terms and categories. An initial set of DQ terms and definitions was created. Feedback and consultation were used to define terms and categories. Multiple rounds of iterative development of a framework consisting of DQ categories, subcategories, and definitions provided a basis for a more inclusive and unified framework.

Multisite Evaluation of a Data Quality Tool for Patient-Level Clinical Data Sets

Introduction: Data quality and fitness for analysis are crucial if outputs of analyses of electronic health record data or administrative claims data should be trusted by the public and the research community.

Methods: We describe a data quality analysis tool (called Achilles Heel) developed by the Observational Health Data Sciences and Informatics Collaborative (OHDSI) and compare outputs from this tool as it was applied to 24 large healthcare datasets across seven different organizations.
Journey toward open data standardization

Data Descriptor: A curated and standardized adverse drug event resource to accelerate drug safety research

Juan M. Banda, Lee Evans, Rami S. Vanguri, Nicholas P. Tatonetti, Patrick B. Ryan & Nigam H. Shah

Identification of adverse drug reactions (ADRs) during the post-marketing phase is one of the most important goals of drug safety surveillance. Spontaneous reporting systems (SRS) data, which are the mainstay of traditional drug safety surveillance, are used for hypothesis generation and to validate the newer approaches. The publicly available US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) data requires substantial curation before they can be used appropriately, and applying different strategies for data cleaning and normalization can have material impact on analysis results. We provide a curated and standardized version of FAERS removing duplicate case records, applying standardized vocabularies with drug names mapped to RxNorm concepts and outcomes mapped to SNOMED-CT concepts, and pre-computed summary statistics about drug-outcome relationships for general consumption. This publicly available resource, along with the source code, will accelerate drug safety research by reducing the amount of time spent performing data management on the source FAERS reports, improving the quality of the data and increasing efficiency.
Journey toward leveraging open data to inform new analyses
Journey toward large-scale open science

Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak*a,b,c,d, Patrick B. Ryanc,e, Jon D. Dukec,e, Nigam H. Shahc,f, Rae Woong Parkc,g, Vojtech Huserc,h, Marc A. Suchardc,i,j,k, Martijn J. Schuemiec,d, Frank J. DeFalcoj,d, Adler Perottiec, Juan M. Bandac,f, Christian G. Reichc,i, Lisa M. Schillingc,m, Michael E. Mathenyc,n,o, Daniella Meekerc,p,q, Nicole Prattr,s, and David Madiganc,s

aDepartment of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; bMedical Informatics Services, NewYork-Presbyterian Hospital, New York, NY 10032; cObservational Health Data Sciences and Informatics, New York, NY 10032; dEpidemiology Analytics, Janssen Research and Development, Titusville, NJ 08560; eCenter for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN 46205; fCenter for Biomedical Informatics Research, Stanford University, CA 94305; gDepartment of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea, 443-380; hLister Hill National Center for Biomedical Communications (National Library of Medicine), National Institutes of Health, Bethesda, MD 20894; iDepartment of Biomatics, University of California, Los Angeles, CA 90095; jDepartment of Biostatistics, University of California, Los Angeles, CA 90095; kDepartment of Human Genetics, University of California, Los Angeles, CA 90095; lReal World Evidence Solutions, IMS Health, Burlington, MA 01809; mDepartment of Medicine, University of Colorado School of Medicine, Aurora, CO 80045; nDepartment of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37212; oGeriatic Research, Education and Clinical Center, VA Tennessee Valley Healthcare System, Nashville, TN 37212; pDepartment of Preventive Medicine, University of Southern California, Los Angeles, CA 90089; qDepartment of Pediatrics, University of Southern California, Los Angeles, CA 90089; rDivision of Health Sciences, University of South Australia, Adelaide, SA, Australia 5001; and sDepartment of Statistics, Columbia University, New York, NY 10027
Journey to clinical characterization insights required network research

Hripcsak et al, PNAS, 2016
Journey of an open community data standard

- Evolving standard (In 2016: CDMv5 → CDMv5.0.1 → CDMv5.1) based on analytical use cases of the community
- Increased adoption of OMOP CDM standard across claims, EHR, registries around the world
- Increasing platform support: MS SQL Server, Oracle, PostgreSQL, MS APS, AWS Redshift, Impala
- Thanks to Rimma Belenkaya and Christian Reich for leading our community data model stewardship!
OHDSI’s standardized vocabularies

• 69 Vocabularies across 25 domains
  - Thank you Christian and the Odysseus team for continue to steward, maintain, and improve this invaluable resource for the entire community!

• 3,526,752 concepts
  - 1,174,628 standard concepts
  - 1,418,696 source codes
  - 323,159 classification concepts

• 24,064,798 concept relationships

Publicly available for download at: http://athena.ohdsi.org/
Our journey as a community of collaborators

OHDSI Collaborators:
• >140 researchers in academia, industry, government, health systems
• >20 countries
• Multi-disciplinary expertise: epidemiology, statistics, medical informatics, computer science, machine learning, clinical sciences

Databases converted to OMOP CDM within OHDSI Community:
• >50 databases
• >660 million patients
Forums.ohdsi.org

- 361 distinct users who have posted
- 5,823 posts on 1,100 topics
- Active discussions across all categories:
  - Implementers, Developers, Researchers, CDM Builders, Vocabulary users

Thank you Christian Reich, Lee Evans, and Chris Knoll for being our most diligent community responders!
Journey toward open-source analytics development

• 61 developers on OHDSI GitHub repositories

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

ATLAS – a single community platform for:
- vocabulary browsing
- database characterization
- cohort definition
- incidence rate
- patient profiles
- population-level effect estimation study design

Thank you Frank DeFalco for leading the OHDSI architecture workgroup!
Journey toward open-source analytics development

Thank you Martijn Schuemie and Marc Suchard for leading methods development efforts!
OHDSI Symposium 2016

• 410 registrants from 11 countries, 27 US states

• 48 poster presentations in observational data management, methodological research, analytics technology and infrastructure, and clinical applications

Congratulations to Christophe Lambert, Sigfried Gold, and Rupa Makadia on winning the OHDSI Best Poster awards!
Thank you OHDSI tutorial faculty!

CDM/ETL: Rimma Belenkaya, Karthik Natarajan, Mark Velez, Erica Voss

Technology stack: Taha Abdul-Basser, Lee Evans, Karthik Natarajan, Mark Velez

Cohort definition: Juan Banda, Jon Duke, Chris Knoll, Nigam Shah

Vocabulary:
Frank DeFalco, George Hripcsak, Christian Reich

Population-level estimation:
David Madigan, Martijn Schuemie, Marc Suchard
Observational research results in literature

85% of exposure-outcome pairs have p < 0.05

What’s going wrong?
- Observational study bias
- Publication bias
- P-hacking

29,982 estimates
11,758 papers

Schuemie, OHDSI Symposium, 2016
Large-scale analysis can start to produce reliable evidence to enable an honest learning healthcare system.

11% of exposure-outcome pairs have calibrated $p < 0.05$

In literature, 85% have $p < 0.05$
OHDSI’s recommended best practices for population-level effect estimation

**Evidence Generation**
- Write and share protocol
- Open source study code
- Use validated software
- Replicate across databases

**Evidence Evaluation**
- Produce standard diagnostics
- Include negative controls
- Create positive controls
- Calibrate confidence interval and p-value

**Evidence Dissemination**
- Don’t provide only the effect estimate
- Also share protocol, study code, diagnostics and evaluation
- Produce evidence at scale
Populations can be used to accurately predict outcomes for individuals.
OHDSI’s areas of focus:
A look back on progress from 2016...

<table>
<thead>
<tr>
<th>Methodological research</th>
<th>Open-source analytics development</th>
<th>Clinical applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational data management</td>
<td>Population-level estimation</td>
<td>Patient-level prediction</td>
</tr>
</tbody>
</table>

- Observational data management
- Clinical characterization
- Population-level estimation
- Patient-level prediction
NOT THE END
Join the journey
At the start of 2016...
Doctor X: “This paper says there’s side effects, but I’ve never seen them happen”
SYNAGIS- palivizumab injection, solution
MedImmune, LLC

-------------------------------

HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use SYNAGIS safely and effectively. See full prescribing information for SYNAGIS.
SYNAGIS® (palivizumab) injection, for intramuscular use
Initial U.S. Approval: 1998

------------------------------ INDICATIONS AND USAGE -------------------------------
Synagis is a respiratory syncytial virus (RSV) F protein inhibitor monoclonal antibody indicated for the prevention of serious lower respiratory tract disease caused by RSV in children at high risk of RSV disease.
- Safety and efficacy were established in children with bronchopulmonary dysplasia (BPD), infants with a history of premature birth (less than or equal to 35 weeks gestational age), and children with hemodynamically significant congenital heart disease (CHD).
- The safety and efficacy of Synagis have not been established for treatment of RSV disease. (1)

------------------------------ DOSAGE AND ADMINISTRATION ------------------------------
15 mg per kg of body weight, administered intramuscularly prior to commencement of the RSV season and remaining doses administered monthly throughout the RSV season. (2.1)
Children undergoing cardio-pulmonary bypass should receive an additional dose of Synagis as soon as possible after the cardio-pulmonary bypass procedure (even if sooner than a month from the previous dose). Thereafter, doses should be administered monthly as scheduled. (2.1, 12.3)

------------------------------ DOSAGE FORMS AND STRENGTHS -----------------------------
Single-dose liquid solution vials: 50 mg per 0.5 mL and 100 mg per 1 mL. (3)

------------------------------ CONTRAINDICATIONS ------------------------------------
Previous significant hypersensitivity reaction to Synagis. (4)

------------------------------ WARNINGS AND PRECAUTIONS -----------------------------
- Anaphylaxis and anaphylactic shock (including fatal cases), and other severe acute hypersensitivity reactions have been reported. Permanently discontinue Synagis and administer appropriate medications if such reactions occur. (5.1)
- As with any intramuscular injection, Synagis should be given with caution to children with thrombocytopenia or any coagulation disorder. (5.2)
- Palivizumab may interfere with immunological-based RSV diagnostic tests such as some antigen detection-based assays. (5.3, 12.4)

------------------------------ ADVERSE REACTIONS -----------------------------------
Adverse reactions occurring greater than or equal to 10% and at least 1% more frequently than placebo are fever and rash. (6.1)
In the middle of 2016...
At the end of 2016...
Why our journey isn’t finished yet

• No patient should have to wonder “what’s the chance that this event might happen to me?”

• All patients deserve to have reliable evidence about the safety and comparative effectiveness of alternative treatments when making medical decisions

• The potential for a ‘learning healthcare system’ can only be realized if we agree to learn together and collaborate to build a system
Building the LHC of observational research?
OHDSI’s areas of focus:
A look forward at 2017...

• Generate and disseminate more clinical evidence
• Maintain and evolve open community data and vocabulary standards
• Develop and improve tools to enable large-scale analysis
• Establish and promote community best practices
• Strengthen and expand collaborations across OHDSI research network
• Advance scholarship in observational data science through publication, presentations, and education
• Generate and disseminate more clinical evidence

Observational data management
Clinical characterization
Population-level estimation
Patient-level prediction

Methodological research
Open-source analytics development
Clinical applications
Join the journey

• Discussion / questions / comments

ryan@ohdsi.org