Observational Health Data Sciences and Informatics (OHDSI): An International Network for Open Science and Data Analytics in Healthcare

Patrick Ryan, PhD
Janssen Research and Development
Columbia University Medical Center
9 May 2016
Odyssey (noun): \oh-d-si\ 

1. A long journey full of adventures
2. A series of experiences that give knowledge or understanding to someone

http://www.merriam-webster.com/dictionary/odyssey
What’s the core problem?

We have lots of DATA we’d like to learn from...

....and very little EVIDENCE we can actually trust
Why large-scale analysis is needed in healthcare

All health outcomes of interest
Introducing OHDSI

• The Observational Health Data Sciences and Informatics (OHDSI) program is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics

• OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University

http://ohdsi.org
OHDSI’s mission

To improve health, by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.
What evidence does OHDSI seek to generate from observational data?

• Clinical characterization
  – **Natural history:** Who are the patients who have diabetes? Among those patients, who takes metformin?
  – **Quality improvement:** what proportion of patients with diabetes experience disease-related complications?

• Population-level estimation
  – **Safety surveillance:** Does metformin cause lactic acidosis?
  – **Comparative effectiveness:** Does metformin cause lactic acidosis more than glyburide?

• Patient-level prediction
  – **Precision medicine:** Given everything you know about me and my medical history, if I start taking metformin, what is the chance that I am going to have lactic acidosis in the next year?
  – **Disease interception:** Given everything you know about me, what is the chance I will develop diabetes?
What is OHDSI’s strategy to generate 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 applications
  – 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
OHDSI’s approach to open science

- Open science is about sharing the journey to evidence generation
- Open-source software can be part of the journey, but it’s not a final destination
- Open processes can enhance the journey through improved reproducibility of research and expanded adoption of scientific best practices
Standardizing workflows to enable reproducible research

Open science

Population-level estimation for comparative effectiveness research:

Is \textit{<intervention X>} better than \textit{<intervention Y>} in reducing the risk of \textit{<condition Z>}?

**Defined inputs:**
- Target exposure
- Comparator group
- Outcome
- Time-at-risk
- Model specification

**Consistent outputs:**
- Analysis specifications for transparency and reproducibility (protocol + source code)
- Only aggregate summary statistics (no patient-level data)
- Model diagnostics to evaluate accuracy
- Results as evidence to be disseminated
  - Static for reporting (e.g. via publication)
  - Interactive for exploration (e.g. via app)
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

Standardized process for network analyses:
- Ask clinical question
- Design protocol
- Develop standardized analytics
- Generate and disseminate evidence
ADA T2DM Guidelines, 2015
## OHDSI participating data partners

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Description</th>
<th>Size (M)</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>

Hripcsak et al, PNAS, in press
Treatment pathways for diabetes

T2DM: All databases

First drug

Second drug

Only drug

Hripcsak et al, PNAS, in press
Population-level heterogeneity

Hripcsak et al, PNAS, in press
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)
## Pediatric patients across US observational databases

<table>
<thead>
<tr>
<th>Source</th>
<th>Age Group (at database entry)</th>
<th>Persons</th>
<th>Avg Years of Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAE</td>
<td>1. newborn 0 to 27d</td>
<td>3,360,896</td>
<td>2.23</td>
</tr>
<tr>
<td>MDCD</td>
<td>1. newborn 0 to 27d</td>
<td>1,862,651</td>
<td>1.74</td>
</tr>
<tr>
<td>Optum</td>
<td>1. newborn 0 to 27d</td>
<td>1,473,940</td>
<td>1.82</td>
</tr>
<tr>
<td>CCAE</td>
<td>2. infant and toddler 28d to 23mo</td>
<td>3,275,604</td>
<td>2.34</td>
</tr>
<tr>
<td>MDCD</td>
<td>2. infant and toddler 28d to 23mo</td>
<td>1,379,760</td>
<td>1.70</td>
</tr>
<tr>
<td>Optum</td>
<td>2. infant and toddler 28d to 23mo</td>
<td>963,770</td>
<td>1.97</td>
</tr>
<tr>
<td>CCAE</td>
<td>3. children 2 to 11 yo</td>
<td>14,904,293</td>
<td>2.46</td>
</tr>
<tr>
<td>MDCD</td>
<td>3. children 2 to 11 yo</td>
<td>4,037,836</td>
<td>1.77</td>
</tr>
<tr>
<td>Optum</td>
<td>3. children 2 to 11 yo</td>
<td>4,951,888</td>
<td>2.18</td>
</tr>
<tr>
<td>CCAE</td>
<td>4. adolescents 12 to 18 yo</td>
<td>12,218,224</td>
<td>2.41</td>
</tr>
<tr>
<td>MDCD</td>
<td>4. adolescents 12 to 18 yo</td>
<td>2,565,515</td>
<td>1.57</td>
</tr>
<tr>
<td>Optum</td>
<td>4. adolescents 12 to 18 yo</td>
<td>3,805,609</td>
<td>2.23</td>
</tr>
</tbody>
</table>
Exploring palivizumab exposure and hypersensitivity in observational data

<table>
<thead>
<tr>
<th>data source</th>
<th>age group (at time of exposure)</th>
<th>persons exposed</th>
<th>follow-up time (yrs)</th>
<th>persons with outcome</th>
<th>risk (events / 1000 persons); 95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAE</td>
<td>1. newborn 0 to 27d</td>
<td>1839</td>
<td>4829</td>
<td>0</td>
<td>0 (0 - 2.59)</td>
</tr>
<tr>
<td>MDCD</td>
<td>1. newborn 0 to 27d</td>
<td>381</td>
<td>760</td>
<td>0</td>
<td>0 (0 - 12.41)</td>
</tr>
<tr>
<td>OPTUM</td>
<td>1. newborn 0 to 27d</td>
<td>2610</td>
<td>5916</td>
<td>1</td>
<td>0.38 (0 - 2.45)</td>
</tr>
<tr>
<td>CCAE</td>
<td>2. infant and toddler 28d to 23mo</td>
<td>42843</td>
<td>106320</td>
<td>27</td>
<td>0.63 (0.43 - 0.92)</td>
</tr>
<tr>
<td>MDCD</td>
<td>2. infant and toddler 28d to 23mo</td>
<td>19910</td>
<td>41196</td>
<td>17</td>
<td>0.85 (0.53 - 1.38)</td>
</tr>
<tr>
<td>OPTUM</td>
<td>2. infant and toddler 28d to 23mo</td>
<td>22365</td>
<td>48632</td>
<td>11</td>
<td>0.49 (0.27 - 0.9)</td>
</tr>
<tr>
<td>CCAE</td>
<td>3. children 2 to 11 yo</td>
<td>544</td>
<td>1525</td>
<td>1</td>
<td>1.84 (0 - 11.68)</td>
</tr>
<tr>
<td>MDCD</td>
<td>3. children 2 to 11 yo</td>
<td>265</td>
<td>706</td>
<td>0</td>
<td>0 (0 - 17.78)</td>
</tr>
<tr>
<td>OPTUM</td>
<td>3. children 2 to 11 yo</td>
<td>204</td>
<td>574</td>
<td>0</td>
<td>0 (0 - 23.01)</td>
</tr>
<tr>
<td>CCAE</td>
<td>4. adolescents 12 to 18 yo</td>
<td>38</td>
<td>93</td>
<td>0</td>
<td>0 (0 - 115.33)</td>
</tr>
<tr>
<td>MDCD</td>
<td>4. adolescents 12 to 18 yo</td>
<td>33</td>
<td>28</td>
<td>0</td>
<td>0 (0 - 131.21)</td>
</tr>
<tr>
<td>OPTUM</td>
<td>4. adolescents 12 to 18 yo</td>
<td>11</td>
<td>44</td>
<td>0</td>
<td>0 (0 - 334.22)</td>
</tr>
</tbody>
</table>

Back of the envelope:
Assuming CCAE+MDCD+OPTUM represents 10% of US and exposures are evenly distributed across ~1000 NICUs, doctor would have seen ~50 newborns with exposure... even if the true event rate was 1%, there’s >60% chance they’d never see one case
OHDSI: what does it mean to me?

Where is there reliable data about the health of children?

Who are the children who are exposed to palivizumab?

Does palivizumab cause anaphylaxis in newborns?

Will my daughter be the one to develop anaphylaxis?
Concluding thoughts

• Observational databases can be a useful tool for generating evidence to important clinical questions in...
  – Clinical characterization
  – Population-level estimation
  – Patient-level prediction
• ...but ensuring that evidence is reliable requires developing scientific best practices, and transparent and reproducible processes to conduct analyses across the research enterprise
• An open science community allows all stakeholders to contribute to and benefit from a shared solution...anyone can get involved...that mean’s YOU!
• Every patient, caregiver, parent and child deserves to know what is known (and what remains uncertain) from the real-world experience of others in order to inform their medical decision-making
Join the journey

Interested in OHDSI?
Questions or comments?

ryan@ohdsi.org