Learning from observational databases: Lessons from OMOP and OHDSI

Patrick Ryan
Janssen Research and Development
David Madigan
Columbia University

http://www.omop.org
http://www.ohdsi.org

“The sole cause and root of almost every defect in the sciences is this: that whilst we falsely admire and extol the powers of the human mind, we do not search for its real helps.”

— Novum Organum: Aphorisms [Book One], 1620, Sir Francis Bacon
141 patients exposed in pivotal randomized clinical trial for metformin
>1,000,000 new users of metformin in one administrative claims database
Patient profiles from observational data
What is the quality of the current evidence from observational analyses?

**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.”
What is the quality of the current evidence from observational analyses?

April 2012: “Patients taking oral fluoroquinolones were at a higher risk of developing a retinal detachment”

Dec 2013: “Oral fluoroquinolone use was not associated with increased risk of retinal detachment”
What is the quality of the current evidence from observational analyses?

BMJ May 2012: “The use of pioglitazone is associated with an increased risk of incident bladder cancer among people with type 2 diabetes.”

BJCP May 2012: “In this study population, pioglitazone does not appear to be significantly associated with an increased risk of bladder cancer in patients with type 2 diabetes.”
• Unknown operating characteristics

• Type 1 error rate? “95%” confidence interval?

• Like early days of lab testing – “trust me, I measured it myself”
2010-2013 OMOP Research Experiment

- 10 data sources
- Claims and EHRs
- 200M+ lives

**Common Data Model**

- Open-source
- Standards-based

**OMOP Methods Library**

- 14 methods
- Epidemiology designs
- Statistical approaches adapted for longitudinal data

**Drug**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>ACE Inhibitors</th>
<th>Amphotericin B</th>
<th>Antibiotics: aminoglycosides, streptococci</th>
<th>Antihypertensives: calcium channel blockers, beta-blockers</th>
<th>Antihyperplasia: alpha-adrenergic</th>
<th>Bisphosphonates: alendronate</th>
<th>Triacyl antidepressants</th>
<th>Typical antipsychotics</th>
<th>Warfarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angioedema</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aplastic Anemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute Liver Injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bleeding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip Fracture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial Infarction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality after MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal Failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI Ulcer Hospitalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lesson 1: Database heterogeneity: Holding analysis constant, different data may yield different estimates

When applying a propensity score adjusted new user cohort design to 10 databases for 53 drug-outcome pairs:

- 43% had substantial heterogeneity ($I^2 > 75\%$) where pooling would not be advisable
- 21% of pairs had at least 1 source with significant positive effect and at least 1 source with significant negative effect

“Evaluating the Impact of Database Heterogeneity on Observational Study Results”
Lesson 2: Parameter sensitivity:

Holding data constant, different analytic design choices may yield different estimates

- Relative risk

Test cases from OMOP 2011/2012 experiment

Holding all parameters constant, except:
- Matching on age, sex and visit (within 30d)
  (CC: 2000205)

yields a RR = 0.73 (0.65 – 0.81)

Sertaline-GI Bleed: RR = 2.45 (2.06 – 2.92)

- Controls per case: up to 10 controls per case
- Required observation time prior to outcome: 180d
- Time-at-risk: 30d from exposure start
- Include index date in time-at-risk: No
- **Case-control matching strategy: Age and sex**
- Nesting within indicated population: No
- Exposures to include: First occurrence
- Metric: Odds ratio with Mantel Haenszel adjustment by age and gender
  (CC: 2000195)

Madigan D, Ryan PB, Scheumie MJ, Therapeutic Advances in Drug Safety, 2013: “Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies”
Lesson 3: Empirical performance:
Most observational methods do not have nominal statistical operating characteristics

Applying the cohort design to MDCR against 34 negative controls for acute liver injury:
- If 95% confidence interval was properly calibrated, then $95\% \times 34 = 32$ of the estimates should cover $RR = 1$
- We observed 17 of negative controls did cover $RR=1$
- Estimated coverage probability = $\frac{17}{34} = 50\%$
- Estimates on both sides of null suggest high variability in the bias

Ryan PB, Stang PE, Overhage JM et al, Drug Safety, 2013:
“A Comparison of the Empirical Performance of Methods for a Risk Identification System”
Lesson 4: Empirical calibration can help restore interpretation of study findings

- Negative controls can be used to estimate empirical null distribution: how much bias and variance exists when no effect should be observed.
- Empirical null can replace theoretical null to estimate calibrated p-value to test for statistical significance.

Negative controls & the null distribution

CC: 2000314, CCAE, G1 Bleed

55% of these negative controls have p < .05 (Expected: 5%)
Negative controls & the null distribution

CC: 2000314, CCAE, GI Bleed
Negative controls & the null distribution

CC: 2000314, CCAE, GI Bleed
p-value calibration plot

CC: 2000314, CCAE, GI Bleed
Clear path forward: systematic evaluation and calibration
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
Why large-scale analysis is needed in healthcare

All health outcomes of interest

<table>
<thead>
<tr>
<th>Drug</th>
<th>Outcome</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug 1</td>
<td>Outcome A</td>
<td>Result A</td>
</tr>
<tr>
<td>Drug 2</td>
<td>Outcome B</td>
<td>Result B</td>
</tr>
<tr>
<td>Drug 3</td>
<td>Outcome C</td>
<td>Result C</td>
</tr>
</tbody>
</table>

All drugs

20
What is large-scale?

- **Millions of observations**
  Need for performance in handling relational structure with millions of patients and billions of clinical observations, focus on optimization to analytical use cases.

- **Millions of covariates**
  No analytics software in the world can fit a regression with >1m observations and >1m covariates on typical hardware... but CYCLOPS can!

- **Millions of questions**
  Systematic solutions with massive parallelization should be designed to run efficiently for one-at-a-time AND all-by-all
Questions OHDSI seeks to answer 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
  – 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?
Community: a social unit of any size that shares common values
--http://en.wikipedia.org/wiki/Community

OHDSI’s communities:
• Research
• Open-source software development
• Data network
OHDSI’s global research community

- >120 collaborators from 11 different countries
- Experts in informatics, statistics, epidemiology, clinical sciences
- Active participation from academia, government, industry, providers

http://ohdsi.org/who-we-are/collaborators/
Data network accomplishments, 2014

• Databases in OMOP CDM
  – 58 databases reported in progress or completed
  – Types: Administrative claims, electronic health records, health information exchanges, hospital billing data, clinical registries, national surveys
  – 9 countries: US, UK, Italy, Germany, Netherlands, Korea, Taiwan, Hong Kong, Japan
  – >682 million patients covered across sources
The odyssey to evidence generation

Patient-level data in source system/schema
Preparing your data for analysis

1. **ETL design**
   - **WhiteRabbit**: profile your source data
   - **RabbitInAHat**: map your source structure to CDM tables and fields

2. **ETL implement**
   - **Usagi**: map your source codes to CDM vocabulary

3. **Patient-level data in OMOP CDM**
   - **CDM**: DDL, index, constraints for Oracle, SQL Server, PostgresQL; Vocabulary tables with loading scripts

4. **ETL test**
   - **ACHILLES**: profile your CDM data; review data quality assessment; explore population-level summaries

**OHDSI Tools built to help**
- WhiteRabbit: profile your source data
- RabbitInAHat: map your source structure to CDM tables and fields
- Usagi: map your source codes to CDM vocabulary
- CDM: DDL, index, constraints for Oracle, SQL Server, PostgresQL; Vocabulary tables with loading scripts
- ACHILLES: profile your CDM data; review data quality assessment; explore population-level summaries

**OHDSI Forums**
Public discussions for OMOP CDM Implementers/developers

[http://github.com/OHDSI](http://github.com/OHDSI)
Data Evidence sharing paradigms

Single study
- Write Protocol
- Develop code
- Execute analysis
- Compile result

Real-time query
- Develop app
- Design query
- Submit job
- Review result

Large-scale analytics
- Develop app
- Execute script
- Explore results

Patient-level data in OMOP CDM

One-time | Repeated
Standardized large-scale analytics tools under development within OHDSI

- **ACHILLES**: Database profiling
- **CIRCE**: Cohort definition
- **HERMES**: Vocabulary exploration
- **HERACLES**: Cohort characterization
- **OHDSI Methods Library**: CYCLOPS Cohort Method
- **LAERTES**: Drug-AE evidence base
- **PLATO**: Patient-level predictive modeling
- **HOMER**: Population-level causality assessment

Patient-level data in OMOP CDM

[http://github.com/OHDSI](http://github.com/OHDSI)
Large-scale analytics example: ACHILLES

http://ohdsi.org/web/ACHILLES

- >12 databases from 5 countries across 3 different platforms:
  - Janssen (Truven, Optum, Premier, CPRD, NHANES, HCUP)
  - Columbia University
  - Regenstrief Institute
  - Ajou University
  - IMEDS Lab (Truven, GE)
  - UPMC Nursing Home
  - Erasmus MC
  - Cegedim
Single study example: Treatment pathways

Open-source process:
- Program analysis: https://github.com/ohdsi
- Execute code on CDM and centrally share results
- Collaboratively explore statistics and jointly publish findings

Treatment pathway example:
- Conceived at AMIA 15Nov2014
- Protocol written, code written and tested at 2 sites 30Nov2014
- Analysis submitted to OHDSI network 2Dec2014
- Results submitted for 7 databases by 5Dec2014, other databases awaiting IRB approval
- Preview of findings now...
Treatment pathway protocol

INDEX: First exposure

>365 day of prior observation

≥1 exposure 121d-240d after index

≥1 exposure 241d-360d after index

≥1 exposure 361d-480d after index

≥1 exposure 481d-600d after index

≥1 exposure 601d-720d after index

≥1 exposure 721d-840d after index

≥1 exposure 841d-960d after index

≥1 exposure 961d-1080d after index

>1095 days of observation post-exposure

≥1 condition occurrence of disease of interest between all time prior to index and all time after index

≤0 condition occurrence of any excluded diseases between all time prior to index and all time after index

Treatment pathway results
Concluding thoughts

• An international community and global data network can be used to generate real-world evidence in a secure, reliable and efficient manner

• Multiple evidence sharing paradigms can and should be used, but all require systematic approaches enabled by a common data model

• Statisticians can and should play a leading role throughout the journey from data to evidence
One model, multiple use cases
Revisiting clopidogrel & GI bleed
(Opatrny, 2008)

<table>
<thead>
<tr>
<th>Agent</th>
<th>Cases (n = 4028)</th>
<th>Controls (n = 40171)</th>
<th>Crude rate ratio</th>
<th>Adjusted rate ratio*</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antidepressants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSRI</td>
<td>335 (8.3%)</td>
<td>1780 (4.4%)</td>
<td>1.97</td>
<td>1.33</td>
<td>1.09, 1.62</td>
</tr>
<tr>
<td>TCA</td>
<td>262 (6.5%)</td>
<td>1764 (4.4%)</td>
<td>1.52</td>
<td>1.04</td>
<td>0.83, 1.30</td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>56 (1.4%)</td>
<td>229 (0.6%)</td>
<td>2.48</td>
<td>1.85</td>
<td>1.34, 2.55</td>
</tr>
<tr>
<td><strong>Anticoagulant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warfarin</td>
<td>281 (7.0%)</td>
<td>1130 (2.8%)</td>
<td>2.64</td>
<td>2.17</td>
<td>1.82, 2.59</td>
</tr>
<tr>
<td>Clopidogrel</td>
<td>160 (4.0%)</td>
<td>532 (1.3%)</td>
<td>3.16</td>
<td>2.07</td>
<td>1.66, 2.58</td>
</tr>
</tbody>
</table>

OMOP, 2012 (CC: 2000314, CCAE, GI Bleed)

Relative risk: 1.86, 95% CI: 1.79 – 1.93
Standard error: 0.02, p-value: <.001
Null distribution
CC: 2000314, CCAE, GI Bleed
Null distribution

CC: 2000314, CCAE, GI Bleed

clopidogrel
Evaluating the null distribution?

- Current p-value calculation assumes that you have an unbiased estimator (which means confounding either doesn’t exist or has been fully corrected for).

- Traditionally, we reject the null hypothesis at \( p < .05 \) and we assume this threshold will incorrectly reject the null hypothesis 5% of time. Does this hold true in observational studies?

- We can test this using our negative controls.
Criteria for negative controls:
- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as ‘causative agent’ in Tisdale et al, 2010: “Drug-Induced Diseases”
- Literature review identified no evidence of potential positive association

<table>
<thead>
<tr>
<th>Event</th>
<th>Positive controls</th>
<th>Negative controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Liver Injury</td>
<td>81</td>
<td>37</td>
<td>118</td>
</tr>
<tr>
<td>Acute Myocardial Infarction</td>
<td>36</td>
<td>66</td>
<td>102</td>
</tr>
<tr>
<td>Acute Renal Failure</td>
<td>24</td>
<td>64</td>
<td>88</td>
</tr>
<tr>
<td>Upper Gastrointestinal Bleeding</td>
<td>24</td>
<td>67</td>
<td>91</td>
</tr>
<tr>
<td>Total</td>
<td>165</td>
<td>234</td>
<td>399</td>
</tr>
</tbody>
</table>
Negative controls & the null distribution

CC: 2000314, CCAE, Gi Bleed

clopidogrel