Large scale analytics for electronic health records: Lessons from Observational Health Data Science and Informatics (OHDSI)

Patrick Ryan, PhD
on behalf of OHDSI team
15 November 2016
Odyssey (noun): \oh-d-si\n
1. A long journey full of adventures
A caricature of the patient journey

- Conditions
- Drugs
- Procedures
- Measurements

Person time

Baseline time 0 Follow-up time
Each observational database is just an (incomplete) compilation of patient journeys.
Questions asked across the patient journey

- Which treatment did patients choose after diagnosis?
- Which patients chose which treatments?
- How many patients experienced the outcome after treatment?
- Does one treatment cause the outcome more than an alternative?
- Does treatment cause outcome?
- What is the probability I will develop the disease?
- What is the probability I will experience the outcome?
Classifying questions across the patient journey

- **Clinical characterization:** What happened to them?
  - What treatment did they choose after diagnosis?
  - Which patients chose which treatments?
  - How many patients experienced the outcome after treatment?

- **Patient-level prediction:** What will happen to me?
  - What is the probability that I will develop the disease?
  - What is the probability that I will experience the outcome?

- **Population-level effect estimation:** What are the causal effects?
  - Does treatment cause outcome?
  - Does one treatment cause the outcome more than an alternative?
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
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.

http://ohdsi.org
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
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
One common data model to support multiple use cases

Standardized clinical data

- Person
  - Observation_period
  - Specimen
  - Death
  - Visit_occurrence
  - Procedure_occurrence
  - Drug_exposure
  - Device_exposure
  - Condition_occurrence
  - Measurement
  - Note
  - Observation
  - Fact_relationship

Standardized health system data

- Location
- Care_site
- Provider
- Payer_plan_period
- Cost
- Cohort
- Cohort_attribute
- Condition_era
- Drug_era
- Dose_era

Standardized health economics

- Concept
- Vocabulary
- Domain
- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor

Standardized derived elements

- Source_to_concept_map
- Drug_strength
- Cohort_definition
- Attribute_definition

Standardized meta-data

- CDM_source
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?
How *should* patients with major depressive disorder be treated?

**Pharmacotherapy**

- The effectiveness of antidepressant medications is generally comparable between and within classes of medications, including selective serotonin reuptake inhibitors (SSRIs), serotonin norepinephrine reuptake inhibitors (SNRIs), bupropion, tricyclic antidepressants (TCAs), and monoamine oxidase inhibitors (MAOIs). Therefore, choose a medication largely based on the following:
  - Patient preference
  - Nature of prior response to medication
  - Safety, tolerability, and anticipated side effects
  - Co-occurring psychiatric or general medical conditions
  - Pharmacological properties of the medication (e.g., half-life, actions on cytochrome P450 enzymes, other drug interactions; consult the full guideline or a current drug database)
  - Cost
  - For most patients, a SSRI, a SNRI, mirtazapine, or bupropion is optimal.
  - In general, the use of MAOIs should be restricted to patients who do not respond to other treatments.
How are patients with major depressive disorder ACTUALLY treated?

Hripcsak et al, PNAS, 2016
### 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, 2016
Treatment pathway study design

- >250,000,000 patient records used across OHDSI network
- >=4 years continuous observation
- >=3 years continuous treatment from first treatment
- N=264,841 qualifying patients with depression
How are patients with major depressive disorder ACTUALLY treated?

• Substantial variation in treatment practice across data sources, health systems, geographies, and over time

• Consistent heterogeneity in treatment choice as no source showed one preferred first-line treatment

• 11% of depressed patients followed a treatment pathway that was shared with no one else in any of the databases

Hripcsak et al, PNAS, 2016
One standardized approach can be applied to multiple clinical areas

**Type 2 Diabetes Mellitus**

- CCAE
- CPRD
- JMDC

**Hypertension**

- CUMC
- INPC
- MDCR

**Depression**

- MDCD
- GE
- OPTUM

- Metformin
- Pioglitazone
- Sitagliptin
- Glimepiride
- Rosiglitazone
- Glyburide
- Insulin, Glargine, Human
- Exenatide
- Liraglutide
- Insulin, Aspart, Human
- Saxagliptin
- Hydrochlorothiazide
- Lisinopril
- Metoprolol
- Amlodipine
- Furosemide
- Losartan
- Atenolol
- Valsartan
- Carvedilol
- Triamterene
- Diltiazem
- Ramipril
- Benazepril
- Olmesartan
- Spironolactone
- Clonidine
- Citalopram
- Bupropion
- Sertraline
- Escitalopram
- Fluoxetine
- Trazodone
- Venlafaxine
- Duloxetine
- Paroxetine
- Amitriptyline
- Mirtazapine
- Desvenlafaxine
- Nortriptyline
- Doxepin

Hripcsak et al, PNAS, 2016
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
Comparison of the Effects of Serotonin-Norepinephrine Reuptake Inhibitors Versus Selective Serotonin Reuptake Inhibitors on Cerebrovascular Events

Yen-Chieh Lee, MDa,‡; Chin-Hsien Lin, MD, PhDb,‡; Min-Shung Lin, MDa; Yun Lu, MScb; Chia-Hsun Chang, MD, ScDb,c,d,*; and Jou-Wei Lin, MD, Phde

ABSTRACT

Background: Use of selective serotonin reuptake inhibitors (SSRIs) has been associated with an increased risk of intracranial hemorrhage. However, little is known about cerebrovascular risk in users of serotonin-norepinephrine reuptake inhibitors (SNRIs). Our aim was to determine the differential risk of cerebrovascular events between SSRIs and SNRIs.

Method: A nationwide population-based cohort study was conducted in adult patients who started taking SSRIs or SNRIs during the time period 2005 through 2009. The outcome of interest was defined by the first hospitalization diagnosis for ischemic stroke (ICD-9-CM codes 433, 434, 436) or intracranial hemorrhage (ICD-9-CM codes 430, 431, 432). We used a Cox regression model with time-varying medication use and adjusted for stroke risk factors to estimate the hazard ratios (HRs) of ischemic stroke and intracranial hemorrhage associated with SNRI use, using SSRI use as a reference.

Results: Among 582,650 SSRI and 76,920 SNRI initiators with an average follow-up period of 3.2 years, there was a nonsignificantly increased trend toward intracranial hemorrhage (adjusted HR = 1.24 [95% CI, 0.97–1.58]) in SNRI users compared to SSRI users. The risk of ischemic stroke was comparable between the 2 treatment groups (adjusted HR = 1.01 [0.90–1.12]). Similar results were obtained in sensitivity analyses, considering a dose-response relation, allowance of a 7-day grace period between study drug discontinuation and outcome occurrence, and restriction to exclusive users, who remained on the initial treatment. In the subgroup analysis, there was an increased incidence of intracranial hemorrhages in SNRI users compared to SSRI users in patients without prior depression (adjusted HR = 1.63 [1.14–2.32]).

Conclusions: Use of SNRIs is not associated with an increased risk of either ischemic stroke or intracranial hemorrhage as compared to use of SSRIs in adult patients with depression or anxiety. However, SNRIs should be used cautiously in patients without depression.
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
Observational research in depression

1,935 estimates
What if we considered all outcomes?

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>
What if we consider all treatments?

<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>
Large-scale estimation for depression

- 17 treatments
- 17 * 16 = 272 comparisons
- 22 outcomes
- 272 * 22 = 5,984 effect size estimates
- 4 databases so far (Truven CCAE, Truven MDCD, Truven MDCR, Optum)
- 4 * 5,984 = 23,936 estimates

NOT DATA MINING - Each analysis following best practice in causal inference
Estimates are in line with expectations

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
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
Populations can be used to accurately predict outcomes for individuals.
Building the LHC of observational research?
Join the journey

• Discussion / questions / comments

ryan@ohdsi.org