Fighting a battle up Hill’s: Discovering new roles for observational healthcare data in causality assessment

Jesse Berlin (VP, Epidemiology, Janssen Research & Development), leaning heavily on Patrick Ryan (on behalf of OMOP Research Team)

20 March 2013: LSHTM Seminar
Observational Medical Outcomes Partnership

*Public-Private Research Partnership established to inform the appropriate use of observational healthcare databases for studying the effects of medical products:*

- Conducting empirical methodological research to evaluate the performance of alternative methods with respect to their ability to identify true associations
- Developing tools and capabilities for transforming, characterizing, and analyzing disparate data sources across the health care delivery spectrum
- Establishing a shared resource so that the broader research community can collaboratively advance the science
OMOP Data Community – First Two Years

**OMOP Extended Consortium**

**OMOP Research Core**

**Research Lab & Coordinating Center**

Centralized data

**Distributed Network**

**178 million** persons with patient-level data

5.4 billion drug exposures, 5.8 billion procedures, 2.3 billion clinical observations
Risk identification and analysis system: One additional piece of evidence to inform medical decision-making

Risk Identification and Analysis System: a systematic and reproducible process to efficiently generate evidence to support the characterization of the potential effects of medical products from across a network of disparate observational healthcare data sources.

Pre-clinical toxicology
Pharmacology
Clinical trials
Spontaneous case reports
Perspectives in literature from medical experts
Pharmacoepidemiology evaluation studies

Risk identification and analysis system

Evidence to support safety assessment
Evidence about the benefits of the product
Evidence about alternative treatments
Decision-making about appropriate use
Is it “Evidence Synthesis?”

• Medical (regulatory) decision-making involves:
  – Summarizing results from RCTs (possibly using meta-analytic techniques)
  – Evaluating multiple epidemiologic studies (possibly using meta-analytic techniques)
  – Evaluating spontaneous adverse event reports
  – “Weighing” the other streams of evidence (e.g., pharmacology, preclinical toxicology)

• Now add “risk identification and analysis system”
  – Is it “active surveillance?”
  – In the context of multiple data sources, is it (can it be, should it be) meta-analysis?
OMOP Research Experiment

- Open-source
- Standards-based
- Systematic data characterization and quality assurance

10 data sources
- Claims and EHRs
- 170M+ lives
- Simulated data (OSIM)

OMOP Methods Library
- Inception cohort
- Case control
- Logistic regression

OMOP Extended Consortium
- OMOP Research Core
- Distributed partners
- Research Lab

Common Data Model

<table>
<thead>
<tr>
<th>Drug</th>
<th>ACE Inhibitors</th>
<th>Amphotericin B</th>
<th>Antibiotics: sulphonamides, tetracyclines</th>
<th>Antiepileptics: carbamazepine, phenytoin</th>
<th>Benzodiazepines</th>
<th>Beta Blockers</th>
<th>Bisphosphonates: alendronate</th>
<th>Tricyclic 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>
<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>
<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>
<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>
<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>
<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>
<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>
<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>
<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>
<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>
<td></td>
</tr>
</tbody>
</table>

Legend:
- Total
- True positive’ benefit
- True positive’ risk
- Negative control’
## Observational Medical Outcomes Partnership (OMOP) 2011/2012 Experiments

### Positive vs. Negative Controls

<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><strong>Total</strong></td>
<td><strong>165</strong></td>
<td><strong>234</strong></td>
<td><strong>399</strong></td>
</tr>
</tbody>
</table>

### Criteria for Positive Controls:
- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label
- Drug listed as ‘causative agent’ in Tisdale et al, 2010: “Drug-Induced Diseases”
- Literature review identified no powered studies with refuting evidence of effect

### 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 powered studies with evidence of potential positive association
Hill’s causality considerations
(OK – they are not criteria)

- Strength of association
- Consistency
- Specificity
- Temporality
- Biological gradient
- Plausibility
- Coherence
- Experimental evidence
- Analogy

Vision for a risk identification and analysis system ‘causal dashboard’

**Drug**: ACE inhibitors  
**Outcome**: Angioedema

**Strength of association**
- by data source
- by method and parameters
- by outcome definition

**Consistency**
- by data source
- by method and parameters
- by outcome definition

**Temporal**

**Specificity**
- Interactive patient profiles

**Plausibility**
- Interactive patient profiles

**Biological gradient**

**Analogy**
- Explore related conditions and treatments

**Experimental evidence**
- Dechallenge/Rechallenge

**Coherence**
- Understand data and cohort to assess potential confounding
Observational analyses to support each causal consideration

• **Strength of association**
  – Current focus: methods produce effect estimates (RR) of association between exposure and outcome

• Consistency
• Specificity
• Temporality
• Biological gradient
• Plausibility
• Coherence
• Experimental evidence
• Analogy
Exploring strength of association: Ex 1: ACE inhibitors - Angioedema

Current capability:
- Display strength of association (as relative risk) for any drug-outcome pair
- Sampling variability in effect estimate shown as 95% confidence intervals
- Results shown across methods and databases
- Composite estimates from meta-analysis
**Strength of association:**

**Ex 2: Antibiotics – Acute Renal Failure**

<table>
<thead>
<tr>
<th>CCO</th>
<th>DP</th>
<th>HDPS</th>
<th>ICTPD</th>
<th>USCCS</th>
</tr>
</thead>
</table>

**What have we learned?**
- Feasibility of establishing a data network with either a distributed network or centralized environment or both
- Multiple alternative perspectives, from epidemiology, statistics, informatics, are considered and can be implemented as methods to estimate effects
- Strength of association from standardized analysis is moderately predictive of true causal effects, poses risk of both false negatives and false positives

**What are existing needs for research?**
- Standardized procedures for data characterization, quality assurance, and software validation
- Better estimates of performance characteristics (e.g., sensitivity, specificity, positive predictive value)

**In some cases, the relative risks are consistent across methods and databases, but inconsistent with ground truth.**

**The strength of association alone is insufficient to understand why**
Observational analyses to support each causal consideration

- Strength of association
- **Consistency**
  - We currently consider four types of consistency:
    1. Consistency across different databases (including measures of heterogeneity)
    2. Consistency across different methods
    3. Consistency across parameters within method
    4. Consistency across different definitions of the health outcome of interest (HOI)
- Specificity
- Temporality
- Biological gradient
- Plausibility
- Coherence
- Experimental evidence
- Analogy
Range of estimates across databases when using high-dimensional propensity score inception cohort (HDPS)

**What have we learned?**
- Effect estimates are highly sensitive to choice of database
- In 21% of test cases, estimates ranged from statistically significant decreased risk to statistically significant increased risk
- No single database accounts for heterogeneity
- Network analyses should report source-specific estimates in addition to any meta-analyses

**What are existing needs for research?**
- Strategies to characterize data source populations and performance characteristics
- Standardized approaches to evaluate sources of heterogeneity
- Methods for pooling results across sources

**ACE Inhibitors-Angioedema** has strong database consistency: all database estimates are statistically significant with large effects (RR>2)

**Antibiotics-Acute renal failure** has weak database consistency: one database shows statistically significant decreased risk; others show statistically significant increased risks, but all with RR<2
Range of estimates across high-dimensional propensity score inception cohort (HDPS) parameter settings

- Each row represents a drug-outcome pair.
- The horizontal span reflects the range of point estimates observed across the parameter settings.
- ACE Inhibitors-Angioedema has strong parameter consistency toward a positive effect: the majority of design decisions result in positive, statistically significant estimates.

Parameter settings explored in OMOP:
- Washout period (1): 180d
- Surveillance window (3): 30 days from exposure start; exposure + 30d; all time

What have we learned?
- Effect estimates are highly sensitive to study design decisions.
- Comparable estimates across alternative standardized vocabularies (ICD9, SNOMED, MedDRA).
- Differential performance by alternative outcome definitions.

What are existing needs for research?
- Systematic process for defining and evaluating HOI definitions.
- Explicit rules to map decisions that would be made during custom evaluations into standardized systematic process.

Antibiotics-Acute renal failure has weak parameter consistency: study design decisions result in wide variation in estimates, with several of choices yielding insignificant.
Important Message

• Show the results by data source
• Show the sensitivity analyses (or at least report them)
• Showing ONLY a combined result is not enough (and may be misleading)
Observational analyses to support each causal consideration

• Strength of association
• Consistency
• Specificity
• **Temporality**
  – Evaluate time-to-event relationship between exposure and outcome
  – High incidence of events prior to exposure may suggest co-occurrence correlation without causal relationship
• Biological gradient
• Plausibility
• Coherence
• Experimental evidence
• Analogy
ACE Inhibitors-Angioedema has strong temporality:
- few incident outcomes prior to exposure
- largest fraction of events within 20 days of incident exposure

Antibiotics-Acute renal failure has weak temporality:
- high co-occurrence of outcome pre- and post-exposure

What have we learned?
• Other aspects of causal framework, beyond strength of association, can be operationalized and do contribute to better understanding of medical product effects

What are existing needs for research?
• Determine what customized analyses need to be implemented within systematic solution
• Standardize quantitative measures for each causal component to minimize subjectivity in assessment
Observational analyses to support each causal consideration

- Strength of association
- Consistency
- Specificity
- Temporality
- Biological gradient
- Coherence
- **Plausibility**
  - Explore interactive patient profiles to identify clinically relevant patterns or alternative explanations
  - Extend beyond population-level treatment effects to study patient-centered outcomes
- **Experimental evidence**
  - Use observational data to approximate natural experiments at the patient level
  - Summarizing dechallenge/rechallenge attempts, successes, and failures can provide supplemental evidence about provider suspicions and patient events
- Analogy
Patient profiles to explore plausibility and experimental evidence

<table>
<thead>
<tr>
<th>PERSON_ID</th>
<th>AGE_AT_INDEX</th>
<th>GENDER</th>
<th>COHORT_NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>20015229241</td>
<td>44</td>
<td>MALE</td>
<td>ACE inhibitor-A</td>
</tr>
</tbody>
</table>

Patient had two 30d pharmacy dispensings of benazepril

During assumed exposure period, angioedema diagnosis recorded and prednisolone prescribed

No further dispensings and no further events suggest potential positive dechallenge

Almost 2 years later, patient prescribed benazepril again...

...and within 9 days has another angioedema diagnosis recorded ...suggesting a potential positive rechallenge
Every patient with exposure and outcome can be considered a natural experiment.

Patients have event within first 30d of prescription, most with positive dechallenge.

Patients with angioedema diagnosis before first ACE inhibitor exposure.

Subsequent occurrence of angioedema diagnoses.

Exposure rechallenge.
Exploratory framework for studying effects

Urticaria

Angioedema

Anaphylactic reactions

What have we learned?
• Feasibility to establish standardized tools for risk identification and analysis system
• Exploratory process requires systematic solution for efficient data analysis

What are existing needs for research?
• Evaluation to determine which causal components provide most information within Bayesian framework
• Integrating observational analyses with other evidence to support safety assessment
Quantitative framework for studying effects

What has been learned?
- Bayesian framework can answer: ‘in light of the data, what is our revised belief of a true causal effect?’
- Here, \( p(\text{true} \mid \text{RR, SE}) \)
  - Logistic regression with 2 predictors
- RR<2 are largely uninformative

What are existing needs for research?
- Using Hill: \( p(\text{true} \mid \text{RR, SE, temporality, coherence, consistency, plausibility, biological gradient, specificity, etc.}) \)
  - Logistic regression with many predictors
- Framework rests on confidence in model, based on empirical evidence of how observational analyses correspond to true causal status

### Graph

**HDPS (Medium SE)**

- Probability
- Relative risk

---

\( p=0.1 \)
Opportunities for a coordinated system that leverages a network of observational healthcare databases to enhance our understanding of the effects of medical products

- Pre-clinical toxicology
- Pharmacology
- Clinical trials
- Spontaneous case reports
- Perspectives in literature from medical experts
- Pharmacoepidemiology evaluation studies
- Analysis system for observational healthcare databases

Evidence to support safety assessment

Evidence about the benefits of the product

Evidence about alternative treatments

Decision-making about appropriate use

p(true unfavorable effect | pre-clinical, pharmacology, clinical trials, spontaneous reports, observational data, ...)

p(true favorable effect | pre-clinical, pharmacology, clinical trials, spontaneous reports, observational data, ...)

Active surveillance

Comparative effectiveness
Concluding thoughts

- A standards-based common clinical information model is feasible and can accommodate disparate data sources
- Multiple analytical use cases can be satisfied within one framework, but scope of data needs may vary
- Standardized analytics enable efficient exploration of a large set of research/public health questions in a consistent, transparent, and reproducible process
- Large-scale analytics and interactive visualization can maximize value of EHR data resources by generating clinically meaningful knowledge for all stakeholders
Concluding thoughts from Sir Bradford Hill

“Yet too often I suspect we waste a deal of time, we grasp the shadow and lose the substance, we weaken our capacity to interpret data and to take reasonable decisions whatever the value of P. And far too often we deduce ‘no difference’ from ‘no statistical difference’. Like fire, the $\chi^2$ test is an excellent servant and a bad master.”

For more information

- [http://omop.fnih.org](http://omop.fnih.org)

- Everything is there