Accounting for Uncertainty in Observational Database Studies

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.”
Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort

Jane Green, clinical epidemiologist, Gabriela Czanner, statistician, Gillian Reeves, statistical epidemiologist, Joanna Watson, epidemiologist, Lesley Wise, manager, Pharmacoepidemiology Research and Intelligence Unit, Valerie Beral, professor of cancer epidemiology

Conclusions The risk of oesophageal cancer increased with 10 or more prescriptions for oral bisphosphonates and with prescriptions over about a five year period.
BMJ study design choices

- Data source: General Practice Research Database
- Study design: Nested case-control
- Inclusion criteria: Age > 40
- Case: cancer diagnosis between 1995-2005 with 12-months of follow-up pre-diagnosis
- 5 controls per case
- Matched on age at index date, sex, practice, observation period prior to index
- Exposure definition: >=1 prescription during observation period
- “RR” estimated with conditional logistic regression
- Covariates: smoking, alcohol, BMI before outcome index date
- Sensitivity analyses:
  - exposure = 2+ prescriptions
  - covariates not missing
  - time-at-risk = >1 yr post-exposure
Current Practice

- In the design of observational studies we rely heavily on “clinical judgment”

- We do so with very limited feedback

- Operating characteristics are unknown

- Like early days of lab testing – “trust me, I measured it myself”
Do these choices matter?
Range of estimates across propensity score inception cohort parameter settings

Parameter settings explored in OMOP:
- **Washout period** (1): 180d
- **Surveillance window** (3): 30 days from exposure start; exposure + 30d; all time from exposure start
- **Covariate eligibility window** (3): 30 days prior to exposure, 180, all-time pre-exposure
- **# of confounders** (2): 100, 500 covariates used to estimate propensity score
- **Propensity strata** (2): 5, 20 strata
- **Analysis strategy** (3): Mantel-Haenszel stratification (MH), propensity score adjusted (PS), propensity strata adjusted (PS2)
- **Comparator cohort** (2): drugs with same indication, not in same class; most prevalent drug with same indication, not in same class
Range of estimates across univariate self-controlled case series (USCCS) parameter settings

USCCS Parameter settings explored in OMOP:
Condition type (2): first occurrence or all occurrences of outcome
Defining exposure time-at-risk:
Days from exposure start (2): should we include the drug start index date in the period at risk?
Surveillance window (4): 30 d from exposure start
Duration of exposure (drug era start through drug era end)
Duration of exposure + 30 d
Duration of exposure + 60 d
Precision of Normal prior (4): 0.5, 0.8, 1, 2

For Bisphosphonates-GI Ulcer hospitalization, USCCS using incident events, excluding the first day of exposure, and using large prior of 2:
• When surveillance window = length of exposure, no association is observed
• Adding 30d of time-at-risk to the end of exposure increased to a significant RR=1.14
Fix everything *except* the database...
Cohort
OMOP 2010/2011 Research Experiment

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

Drug

<table>
<thead>
<tr>
<th>Outcome</th>
<th>ACE inhibitors</th>
<th>Amphotericin B</th>
<th>Antibiotics: erythromycin, sulfafoximes, 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>

Positives: 9
Negatives: 44
Adverse events associated with treatment of latent tuberculosis in the general population

Benjamin M. Smith MD, Kevin Schwartzman MD MPH, Gillian Bartlett PhD, Dick Menzies MD MSc

ABSTRACT

Background: Guidelines recommend treatment of latent tuberculosis in patients at increased risk for active tuberculosis. Studies investigating the association of therapy with serious adverse events have not included the entire treated population nor accounted for comorbidities or occurrence of similar events in the untreated general population. Our objective was to estimate the risk of adverse events requiring hospital admission that were associated with therapy for latent tuberculosis infection in the general population.

Methods: Using administrative health data from the province of Quebec, we created a historical cohort of all residents dispensed therapy for latent tuberculosis between 1998 and 2003. Each patient was matched on age, sex and postal region with two untreated residents. The observation period was 18 months (from 6 months before to 12 months after initiation of therapy). The primary outcome was hospital admission for therapy-associated adverse events.

Results: During the period of observation, therapy for latent tuberculosis was dispensed to 9145 residents, of whom 95% started isoniazid and 5% started rifampin. Pretreatment comorbid illness was significantly more common among patients receiving such therapy compared with the matched untreated cohort. Of all patients dispensed therapy, 45 (0.5%) were admitted to hospital for a hepatic event compared with 15 (0.1%) of the untreated patients. For people over age 65 years, the odds of hospital admission for a hepatic event among patients treated for latent tuberculosis infection was significantly greater than among matched untreated people after adjustment for comorbidities (odds ratio [OR] 6.4, 95% CI 2.2–18.3). Excluding patients with comorbid illness, there were two excess admissions to hospital for hepatic events per 100 patients initiating therapy compared with the rate among untreated people over 65 years (95% CI 0.1–3.87).

Interpretation: The risk of adverse events requiring hospital admission increased significantly among patients over 65 years receiving treatment for latent tuberculosis infection. The decision to treat latent tuberculosis infection in elderly patients should be made after careful consideration of risks and benefits.
Receiver Operating Characteristic (ROC) curve

- ROC plots sensitivity vs. false positive rate
- Rank-order all pairs by RR from largest to smallest
- Calculate sensitivity and specificity at all possible RR thresholds

- Area under ROC curve (AUC) provides probability that method will score a randomly chosen true positive drug-outcome pair higher than a random unrelated drug-outcome pair
  - AUC=1 is perfect predictive model
  - AUC=0.50 is random guessing (diagonal line)
  - Cohort method on MDCR: AUC = 0.64

**Isoniazid (RR=4.04):**
- Sensitivity = 4%
- Specificity = 98%
Performance after Customization

- AUC=0.92
- AUC=0.76

• Restricting to drugs with sufficient sample further increased AUC for all outcomes, but the degree of change varied by outcome
• Increased prediction comes as tradeoff with fewer drugs under surveillance
• Self-controlled cohort design continue to be optimal design, but specific settings changed in all outcomes
Performance across methods, by database

- All self-controlled designs (OS, ICTPD, SCCS) are consistently at or near the top of performance across all outcomes and sources.
- Cohort and case-control designs have comparable performance, consistently lower than all self-controlled designs.
- Substantial variability in performance across the optimal settings of each method.
Good performance?

- ...it all depends on your tolerance of false positives and false negatives...
- ...but we've created a tool to let you decide

http://elmo.omop.org
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>Antidepressants</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>Anticoagulant</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

Relative Risk (Log scale)

Density

0
1
2
Null distribution

CC: 2000314, CCAE, GI Bleed

clopidogrel
Negative controls & the null distribution

CC: 2000314, CCAE, GI Bleed

clopidogrel
Negative controls & the null distribution

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

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

CC: 2000314, CCAE, GI Bleed

- p < 0.05: 55%
- Calibrated p < 0.05: 6%

- clopidogrel:
  - RR: 1.9 (1.8 – 1.9)
  - p: <0.001
  - Calibrated p: 0.30
This analysis failed to reject the empirical null... but we know clopidogrel causes GI bleeding (it’s a positive control)
p-value calibration plot
Optimal method: SCCS:1931010, CCAE, GI Bleed

- **p < .05**: 33%
- **Calibrated p < .05**: 9%
- **clopidogrel**: RR 1.3 (1.2 – 1.3)
- **p**: <.001
- **Calibrated p**: .01
Applying case-control design to negative controls in real data

45% of the CIs of negative controls contain 1 (Expected: 95%)
Large-scale analytics can help reframe the patient-level prediction problem.

Given a patient’s clinical observations in the past,...

...can we predict outcomes for that patient in the future?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Age</th>
<th>Gender</th>
<th>Race</th>
<th>Location</th>
<th>Drug 1</th>
<th>Drug 2</th>
<th>Drug n</th>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition n</th>
<th>Procedure 1</th>
<th>Procedure 2</th>
<th>Procedure n</th>
<th>Lab 1</th>
<th>Lab 2</th>
<th>Lab n</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>76</td>
<td>M</td>
<td>B</td>
<td>441</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>77</td>
<td>F</td>
<td>W</td>
<td>521</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>96</td>
<td>F</td>
<td>B</td>
<td>215</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>76</td>
<td>F</td>
<td>B</td>
<td>646</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>64</td>
<td>M</td>
<td>B</td>
<td>379</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>74</td>
<td>M</td>
<td>W</td>
<td>627</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>68</td>
<td>M</td>
<td>B</td>
<td>348</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Demographics | All drugs | All conditions | All procedures | All lab values
---|------------|----------------|----------------|----------------|

Tools for Large-Scale Regression

BBR/BMR
- logistic, multinomial
- L1, L2 regularization
- sparse \(\rightarrow\) millions of predictors
- hierarchical, priors, autosearch
- stable

BXR
- bayesianregression.org
- cleaner

BOXER
- online logistic regression

CCD
- bscs.googlecode.com
- logistic, conditional logistic, multinomial, Poisson, Cox, ParamSurv, least squares
- L1, L2 regularization
- sparse \(\rightarrow\) millions of predictors
- imputation
- CPU, GPU
PLATO

Patient-Level Assessment of Treatment Outcomes

- Predictive models that assess the probability of a patient experiencing any outcome following initiation of any intervention, given his or her personal medical history

- Front-end prediction browser

- Challenge: \(~25K\) interventions \(\times\) \(~1K\) outcomes = \(25M\) models
Sparse Coding

Relational Random Forests
Shahn et al.

(\(\geq 30\), appendectomy, Y/N):
in the last 30 days, did the patient have an appendectomy?

(\(<0\), max(SBP), 140):
at any time in the past did the patient’s SBP exceed 140 mmHg?

(\(<90\), rofecoxib, Y/N):
in the time period up to 90 days ago, did the patient have a prescription for rofecoxib?

(\(\geq 7\), fever, Y/N):
in the last week, did the patient have a fever?

Bayesian Hierarchical Association Rule Mining
McCormick et al.

- Goal: Predict next event in current sequence given sequence database
- So far, successful application to RCT data

Bayesian List Machine
Rudin et al.

\[
\begin{align*}
\text{if hemiplegia and age} &> 60 \text{ then stroke risk } 58.9\% (53.8\% - 63.8\%) \\
\text{else if cerebrovascular disorder then stroke risk } &47.8\% (44.8\% - 50.7\%) \\
\text{else if transient ischaemic attack then stroke risk } &23.8\% (19.5\% - 28.4\%) \\
\text{else if occlusion and stenosis of carotid artery without infarction then stroke risk } &15.8\% (12.2\% - 19.6\%) \\
\text{else if altered state of consciousness and age} &> 60 \text{ then stroke risk } 16.0\% (12.2\% - 20.2\%) \\
\text{else if age} &\leq 70 \text{ then stroke risk } 4.6\% (3.9\% - 5.4\%) \\
\text{else stroke risk } &8.7\% (7.9\% - 9.6\%)
\end{align*}
\]
Conclusions

• Using an empirically driven, high throughput approach to study design, a risk identification system can perform at AUC>0.80

• Traditional p-values and confidence intervals require empirical calibration to account for bias in observational studies

• Advancing the science of observational research requires an empirical and reproducible approach to methodology and systematic application

• Predictive models face similar issues
Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

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

Read more about us, about our goals, and how you can help support the OHDSI community.

Join the Journey

ACHILLES Released

OHDSI released its first open-source software application, ACHILLES, at the 2014 EDM Forum in San Diego, CA. Congratulations to the ACHILLES

OHDSI on YouTube

Welcome to OHDSI

Latest News

- OHDSI paper published in Drug Safety

search here ...  Go
# Collaborators

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charles Bailey, MD, PhD</td>
<td>Assistant Professor of Pediatrics, Perelman School of Medicine at the University of Pennsylvania</td>
<td>Children’s Hospital of Philadelphia</td>
</tr>
<tr>
<td>Rimma Belinkaya, MS, MA</td>
<td>Knowledge Manager</td>
<td>Albert Einstein College of Medicine</td>
</tr>
<tr>
<td>Tomas Bengtsson, PhD</td>
<td>Research Engineer</td>
<td>WHO Uppsala Monitoring Centre</td>
</tr>
<tr>
<td>Jessie Berlin, ScD</td>
<td>Vice President and Global Head, Epidemiology</td>
<td>Johnson &amp; Johnson</td>
</tr>
<tr>
<td>Edward A. Bottrich, PhD</td>
<td>Global Head, Pharmacoepidemiology and Database Research Unit, Data Analytics and Observational Methods</td>
<td>Merck Research Laboratories</td>
</tr>
<tr>
<td>Richard D. Boyce, PhD</td>
<td>Assistant Professor of Biomedical Informatics</td>
<td>University of Pittsburgh School of Medicine</td>
</tr>
</tbody>
</table>
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
OHDSI’s vision

OHDSI collaborators access a network of 1 billion patients to generate evidence about all aspects of healthcare. Patients and clinicians and other decision-makers around the world use OHDSI tools and evidence every day.
Hill Criteria

“What aspects of that association should we especially consider before deciding that the most likely interpretation of it is causation?”

- Strength
- Consistency
- Temporality
- Plausibility
- Experiment
- Coherence
- Biological gradient
- Specificity
- Analogy

Introducing HOMER

- Health Outcomes and Medical Effectiveness Research (HOMER) system

- Live, interactive evidence exploration system with fully functional implementations of all of the components of Sir Austin Bradford Hill’s viewpoints for risk identification and assessment, plus some additional components designed by the OMOP team
HOMER implementation of Hill’s viewpoints

- Consistency
- Temporality
- Strength
- Plausibility
- Experiment
- Analogy
- Biological gradient
- Coherence
- Specificity
- Comparative effectiveness
- Predictive modeling
CCD

• CCD is a tool for performing massive regularized regressions (Poisson, linear, logistic, survival)
  – > 1+ million variables
  – > 10+ million rows

• Useful for
  – Data-driven propensity score estimation
  – Building predictive models
  – Large scale outcome models (e.g. SCCS)
Installing the OHDSI packages

• Install RTools:
  \[http://cran.r-project.org/bin/windows/Rtools/\]

• In R:

  install.packages(“devtools”)
  library(devtools)
  install_github(“ohdsi/DatabaseConnector”)
  install_github(“ohdsi/SqlRender”)
  install_github(“ohdsi/CCD”)
  install_github(“ohdsi/CohortMethod”)
Join us on the journey

• Observational data has tremendous potential value for healthcare, but systematic application of large-scale analysis is required to unlock that value
• OHDSI has established a collaborative led by people and expertise necessary to execute on this ambitious plan to develop and apply solutions to generate reliable evidence that can transform medical decision making
• With your support, together we can make this vision a reality

http://ohdsi.org