OHDSI Methods for Causal Effect Estimation

George Hripcsak, David Madigan, Patrick Ryan, Martijn Schuemie, Marc Suchard

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
When comparisons were made between NOACs, matched rivaroxaban patients had a significantly higher risk of major bleeding (HR: 1.82; 95 % CI: 1.36–2.43) compared to apixaban patients.
Reliability: Analysis Ignores...

- Selection bias
- Measurement error
- Model misspecification
- Multiple modeling
- Unmeasured confounding

“Grave errors are commonplace, perhaps typical. It does no good to append a claim that you have included in the regression all relevant covariates, a claim that there are no unmeasured confounders and that you could not be mistaken in making this claim. Who are you that you could not be mistaken?”

- Paul Rosenbaum
Observational research results in literature

85% of exposure-outcome pairs have $p < 0.05$

29,982 estimates
11,758 papers
A New Approach

• Reproducible, systematized, open source approach at scale

• Negative controls
  – Drugs and outcomes “known” to have no causal association
  – Literature, product labels, spontaneous reports
  – Empirical p-values

• Positive Controls
  – Inject signals onto negative controls with known effect size
  – Calibrated confidence intervals
LEGEND results

1,321,696 estimates
83.4% of CIs includes 1
Observational research results in literature

29,982 estimates
11,758 papers
Calibration Assumptions

\[ \theta_0 \quad \text{true effect size of interest} \]
\[ \theta_i \quad \text{true effect size for the controls, } i = 1, \ldots, m \]
\[ \hat{\theta}_i \quad \text{estimated effect sizes, } i = 0, \ldots, m \]
\[ \hat{\beta}_i = \hat{\theta}_i - \theta_i \quad \text{"estimated bias," } i = 0, \ldots, m \]
\[ E[\hat{\beta}_i] = \beta_i \quad i = 0, \ldots, m \]
Many models, many databases

\[ p(\theta_0|\mathcal{D}) = \sum_k p(\theta_0|M_k, \mathcal{D}) p(M_k|\mathcal{D}) \]

where the data, \( \mathcal{D} \), comprise:

\[ \hat{\theta}^k_{0j}, k = 1, \ldots, M, j = 1, \ldots, D \]
\[ \hat{\theta}^k_{ij}, k = 1, \ldots, M, j = 1, \ldots, D, i = 1, \ldots, Q \]
\[ \theta_i, i = 1, \ldots, Q \]

can show that:  \[ p(M_k|\mathcal{D}) \propto \prod_{i=1}^{Q} \prod_{j=1}^{D} p(\hat{\theta}^k_{ij}|\theta_i, M_k) \]

Combining calibration with random effects meta-analysis and BMA
Method: Study design (LEGEND)

Treatment strategies:
• Atenolol
• Nebivolol

Causal contrasts of interest:
• On-treatment effect
• Intent-to-treat effect

Eligibility criteria:
• Diagnosis of hypertension during previous 1 year
• No prior antihypertensive drug
• No prior cardiovascular outcome

Outcome (Major Adverse Cardio-Cerebrovascular Event):
• Hospitalized myocardial infarction, heart failure, stroke and sudden cardiac death

https://github.com/OHDSI/LEGEND
Cohort Methods

Subject 1
- Covariate capture
- Target

Subject 2
- Covariate capture
- Target

Subject 3
- Covariate capture
- Comparator

Subject 4
- Covariate capture
- Comparator

Outcome

Adjustment strategy

Time
Other Methods: Case Control
Other Methods: SCCS
# Gold Standard Performance: CCAE

<table>
<thead>
<tr>
<th>Analysis choices</th>
<th>AUC</th>
<th>95% CI coverage</th>
<th>Mean precision</th>
<th>MSE</th>
<th>Type 1 error</th>
<th>Type 2 error</th>
<th>Non-estimatable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohort method</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No PS, simple outcome model</td>
<td>0.78</td>
<td>0.92</td>
<td>3.35</td>
<td>0.51</td>
<td>0.08</td>
<td>0.60</td>
<td>0.19</td>
</tr>
<tr>
<td>1-on-1 matching, unstratified outcome model</td>
<td>0.78</td>
<td>0.89</td>
<td>9.35</td>
<td>0.33</td>
<td>0.08</td>
<td>0.37</td>
<td>0.31</td>
</tr>
<tr>
<td>Variable ratio matching, stratified outcome model</td>
<td>0.78</td>
<td>0.92</td>
<td>7.45</td>
<td>0.34</td>
<td>0.06</td>
<td>0.41</td>
<td>0.30</td>
</tr>
<tr>
<td>Stratification</td>
<td>0.79</td>
<td>0.90</td>
<td>8.68</td>
<td>0.41</td>
<td>0.10</td>
<td>0.35</td>
<td>0.26</td>
</tr>
<tr>
<td>IPTW</td>
<td>0.78</td>
<td>0.93</td>
<td>4.14</td>
<td>0.44</td>
<td>0.08</td>
<td>0.55</td>
<td>0.24</td>
</tr>
<tr>
<td>Var ratio matching + full outcome model</td>
<td>0.78</td>
<td>0.93</td>
<td>7.76</td>
<td>0.31</td>
<td>0.04</td>
<td>0.43</td>
<td>0.45</td>
</tr>
<tr>
<td><strong>Self-controlled cohort (SCC)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time exposed, inc. exp. start date</td>
<td>0.87</td>
<td>0.92</td>
<td>12.28</td>
<td>0.49</td>
<td>0.09</td>
<td>0.29</td>
<td>0.07</td>
</tr>
<tr>
<td>30 days, incl. exp. start date</td>
<td>0.87</td>
<td>0.92</td>
<td>12.25</td>
<td>0.19</td>
<td>0.09</td>
<td>0.34</td>
<td>0.04</td>
</tr>
<tr>
<td>Time exposed, inc. exp. start date, require full obs.</td>
<td>0.88</td>
<td>0.92</td>
<td>13.91</td>
<td>0.47</td>
<td>0.09</td>
<td>0.27</td>
<td>0.08</td>
</tr>
<tr>
<td>Time exposed, ex. exp. start date</td>
<td>0.87</td>
<td>0.94</td>
<td>11.75</td>
<td>0.22</td>
<td>0.08</td>
<td>0.18</td>
<td>0.23</td>
</tr>
<tr>
<td>30 days, ex. exp. start date</td>
<td>0.89</td>
<td>0.93</td>
<td>14.46</td>
<td>0.14</td>
<td>0.08</td>
<td>0.29</td>
<td>0.05</td>
</tr>
<tr>
<td>Time exposed, ex. exp. start date, require full obs.</td>
<td>0.87</td>
<td>0.94</td>
<td>13.04</td>
<td>0.20</td>
<td>0.07</td>
<td>0.17</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Case-control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 controls per case</td>
<td>0.84</td>
<td>0.92</td>
<td>7.35</td>
<td>0.59</td>
<td>0.08</td>
<td>0.48</td>
<td>0.04</td>
</tr>
<tr>
<td>10 controls per case</td>
<td>0.84</td>
<td>0.92</td>
<td>7.37</td>
<td>0.62</td>
<td>0.08</td>
<td>0.48</td>
<td>0.01</td>
</tr>
<tr>
<td>Nesting in indication, 2 controls per case</td>
<td>0.87</td>
<td>0.91</td>
<td>12.90</td>
<td>0.54</td>
<td>0.10</td>
<td>0.34</td>
<td>0.01</td>
</tr>
<tr>
<td>Nesting in indication, 10 controls per case</td>
<td>0.86</td>
<td>0.92</td>
<td>12.28</td>
<td>0.55</td>
<td>0.10</td>
<td>0.35</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Case-crossover</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple case-crossover, -30 days</td>
<td>0.87</td>
<td>0.92</td>
<td>10.17</td>
<td>0.46</td>
<td>0.08</td>
<td>0.43</td>
<td>0.04</td>
</tr>
<tr>
<td>Simple case-crossover, -180 days</td>
<td>0.85</td>
<td>0.93</td>
<td>10.79</td>
<td>0.61</td>
<td>0.07</td>
<td>0.44</td>
<td>0.02</td>
</tr>
<tr>
<td>Nested case-crossover, -30 days</td>
<td>0.87</td>
<td>0.92</td>
<td>11.78</td>
<td>0.41</td>
<td>0.08</td>
<td>0.34</td>
<td>0.06</td>
</tr>
<tr>
<td>Nested case-crossover, -180 days</td>
<td>0.86</td>
<td>0.93</td>
<td>12.26</td>
<td>0.55</td>
<td>0.06</td>
<td>0.36</td>
<td>0.03</td>
</tr>
<tr>
<td>Nested case-time-control, -30 days</td>
<td>0.87</td>
<td>0.92</td>
<td>10.76</td>
<td>0.38</td>
<td>0.06</td>
<td>0.38</td>
<td>0.05</td>
</tr>
<tr>
<td>Nested case-time-control, -180 days</td>
<td>0.87</td>
<td>0.94</td>
<td>10.80</td>
<td>0.21</td>
<td>0.07</td>
<td>0.35</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>Self-controlled case series (SCCS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple SCCS</td>
<td>0.90</td>
<td>0.95</td>
<td>15.78</td>
<td>0.17</td>
<td>0.07</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Including day 0</td>
<td>0.87</td>
<td>0.93</td>
<td>11.67</td>
<td>0.53</td>
<td>0.08</td>
<td>0.39</td>
<td>0.09</td>
</tr>
<tr>
<td>Using pre-exposure window</td>
<td>0.90</td>
<td>0.95</td>
<td>12.98</td>
<td>0.19</td>
<td>0.08</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>Using age and season</td>
<td>0.91</td>
<td>0.94</td>
<td>22.15</td>
<td>0.16</td>
<td>0.08</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>Using event-dependent observation</td>
<td>0.88</td>
<td>0.95</td>
<td>12.12</td>
<td>0.18</td>
<td>0.08</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>Using all other exposures</td>
<td>0.91</td>
<td>0.95</td>
<td>21.98</td>
<td>0.16</td>
<td>0.05</td>
<td>0.11</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Gold Standard Performance: CCAE
Concluding thoughts

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

• Common data model critically important

• Much work remains on establishing (and improving) actual operating characteristics of current approaches to causal inference