CV SAFETY ASSESSMENTS: SIGNAL DETECTION WITH HETEROGENEOUS DATA SOURCES

Sharon-Lise T. Normand

Harvard Medical School & Harvard School of Public Health
MDEpiNet Methodology Center
sharon@hcp.med.harvard.edu

February 19, 2015
Heterogeneous data sources:

- Contain different data **types**
  - Natural language processed, laboratory measurement, billing claims data, abstracted data

- Elements have different **resolutions**
  - Measurement error varies across data elements (e.g., imaging information versus claims data)

- Sources may cover **different time periods**
  - Registries with a longer versus shorter history

- **Completeness** of data varies
  - Partially or completely missing data elements

- **Design** differences: prospective, retrospective, blinding

- Potential to **grow**
POOLING DATA FOR SIGNAL DETECTION

- **Pool** data (across devices, across indications, across patient populations) when uncertainty exists

- **Appropriateness** of pooling depends on
  - data quality
  - data completeness
  - how reported
  - units of measurement
  - design (blinded, prospective, retrospective, etc)

- What **level** of pooling is acceptable?

- Operator effects can be substantial (**learning**)
### TAVR REGISTRIES

<table>
<thead>
<tr>
<th>Feature</th>
<th>(1) U.S.</th>
<th>(2) Western Europe</th>
<th>(3) Canada</th>
<th>(4) U.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Patients</td>
<td>7710</td>
<td>996</td>
<td>339</td>
<td>2732</td>
</tr>
<tr>
<td>Centers</td>
<td>224</td>
<td>44</td>
<td>6</td>
<td>31</td>
</tr>
<tr>
<td>Year</td>
<td>11/11-5/13</td>
<td>6/10-7/11</td>
<td>6/05-6/09</td>
<td>1/07-12/11</td>
</tr>
<tr>
<td>Manufact.</td>
<td>Edwards</td>
<td>CoreValve</td>
<td>Edwards</td>
<td>Both</td>
</tr>
<tr>
<td>Age</td>
<td>84 ± 10</td>
<td>81 ± 6</td>
<td>81.8 ± 8</td>
<td>?</td>
</tr>
<tr>
<td>Severity</td>
<td>7 ± 6</td>
<td>19.2* ± 12.4</td>
<td>9.8 ± 6.4</td>
<td>?</td>
</tr>
<tr>
<td>30-Day Mortality (%)</td>
<td>7.6</td>
<td>4.5</td>
<td>10.4</td>
<td>5.4 in 2011</td>
</tr>
<tr>
<td>Follow-up (%)</td>
<td><strong>46</strong></td>
<td>99</td>
<td><strong>100</strong></td>
<td>?</td>
</tr>
</tbody>
</table>

**Sources:**

1. JAMA, 2013;310(19):2069-2077
3. JACC, 2010;55(11):1080-1090
4. TCT 2012 slides
STARCLOSE (NEW) VS ANGIOSEAL (OLD)
(Source = Sarma, 2013 ACC)
DISTRIBUTED NETWORKS

- Depends on how computations are conducted
  - Combining data at the individual-level among networks
  - Combining aggregated data across networks
- When some elements are partially or completely missing, aggregated approaches are not sufficient

**Distributed network approaches require substantial overhaul**

- But, **substantial** opportunities to learn faster by exploiting variation
Using registries for **signal detection** requires:

- Sharing **more** than summary statistics
- Theoretical and empirical experience with **causal modeling** using heterogeneous data
- Development of inferential tools for **pooling related but distinct** data
  - Methodology to validate results
- Methodology to separate **cumulative learning** from patient severity
- Rules of engagement: **what** & **how much** to pool?