Considerations in the Use of Propensity Scores in Observational Studies

Lawrence Rasouliyan, Estel Plana, Jaume Aguado
RTI Health Solutions, Barcelona, Spain

PhUSE 2016 / Real World Evidence Stream
Background

• Randomized Controlled Trials = Gold standard for treatment comparison
• Observational studies are increasingly being used to estimate the effects of treatments, exposures, and interventions on outcomes.
• Treated and untreated patients often have systematic differences in their distributions of underlying characteristics.
• Traditionally, multivariable models have been used.
• Propensity scores can be used to minimize the effects of bias in the estimate of the treatment effect.
  – Serves as a “balancing” score for measured confounders
  – Outcomes are compared taking into account this balancing to reduce potential bias of confounders
Propensity Score: Definition

- Probability of receiving Treatment A (versus Treatment B) given the patient’s underlying characteristics.
- Usually determined by logistic regression:
Propensity Score: Definition

- Probability of receiving Treatment A (versus Treatment B) given the patient’s underlying characteristics.
- Usually determined by logistic regression:

\[ \ln \left( \frac{P_A}{1 - P_A} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots \]
Propensity Score: Definition

- Probability of receiving Treatment A (versus Treatment B) given the patient’s underlying characteristics.
- Usually determined by logistic regression:

\[ \ln \left( \frac{P_A}{1 - P_A} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots \]

- Solving for \( P_A \), we get the propensity score:

\[ P_A = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots}} \]
Propensity Score: Definition

- Probability of receiving Treatment A (versus Treatment B) given the patient’s underlying characteristics.
- Usually determined by logistic regression:

\[
\ln \left( \frac{P_A}{1 - P_A} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots
\]

- Solving for \( P_A \), we get the propensity score:

\[
P_A = \frac{\exp[\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots]}{1 + \exp[\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots]}
\]

- Then, we account for this propensity score to make treatment comparison.
Illustrative Example: Data Source

• Prospective 8-year observational study in oncology

• Outcomes of interest:
  – Overall survival (OS) = Time to death
  – Progression-free survival (PFS) = Time to progression or death
  – Overall response rate (ORR) = Response to treatment

• Data simulated to mimic general attributes of actual results

• Disease indication: “Cancer”

• Two possible therapies…
Two Possible Therapies

Blue Pill
N = 1,871

Red Pill
N = 1,384

Which pill is “better” for outcomes of interest?
## Patient Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Blue Pill</th>
<th>Red Pill</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N = 1871</strong></td>
<td><strong>N = 1384</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age group, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 50</td>
<td>59 (3.2%)</td>
<td>234 (16.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>51 to 60</td>
<td>576 (30.8%)</td>
<td>684 (49.4%)</td>
<td></td>
</tr>
<tr>
<td>61 to 70</td>
<td>925 (49.4%)</td>
<td>412 (29.8%)</td>
<td></td>
</tr>
<tr>
<td>&gt; 70</td>
<td>311 (16.6%)</td>
<td>54 (3.9%)</td>
<td></td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>834 (44.6%)</td>
<td>593 (42.8%)</td>
<td>0.326</td>
</tr>
<tr>
<td>Female</td>
<td>1037 (55.4%)</td>
<td>791 (57.2%)</td>
<td></td>
</tr>
<tr>
<td><strong>Geographic region, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>846 (45.2%)</td>
<td>553 (40.0%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Europe</td>
<td>705 (37.7%)</td>
<td>702 (50.7%)</td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>320 (17.1%)</td>
<td>129 (9.3%)</td>
<td></td>
</tr>
<tr>
<td><strong>Performance status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1060 (56.7%)</td>
<td>1001 (72.3%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1</td>
<td>688 (36.8%)</td>
<td>352 (25.4%)</td>
<td></td>
</tr>
<tr>
<td>≥ 2</td>
<td>123 (6.6%)</td>
<td>31 (2.2%)</td>
<td></td>
</tr>
</tbody>
</table>
## Patient Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Blue Pill</th>
<th>Red Pill</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = 1871 )</td>
<td></td>
<td>( N = 1384 )</td>
<td></td>
</tr>
<tr>
<td><strong>Age group, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \leq 50 )</td>
<td>59 (3.2%)</td>
<td>234 (16.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>51 to 60</td>
<td>576 (30.8%)</td>
<td>684 (49.4%)</td>
<td></td>
</tr>
<tr>
<td>61 to 70</td>
<td>925 (49.4%)</td>
<td>412 (29.8%)</td>
<td></td>
</tr>
<tr>
<td>&gt; 70</td>
<td>311 (16.6%)</td>
<td>54 (3.9%)</td>
<td></td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>834 (44.6%)</td>
<td>593 (42.8%)</td>
<td>0.326</td>
</tr>
<tr>
<td>Female</td>
<td>1037 (55.4%)</td>
<td>791 (57.2%)</td>
<td></td>
</tr>
<tr>
<td><strong>Geographic region, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>846 (45.2%)</td>
<td>553 (40.0%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Europe</td>
<td>705 (37.7%)</td>
<td>702 (50.7%)</td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>320 (17.1%)</td>
<td>129 (9.3%)</td>
<td></td>
</tr>
<tr>
<td><strong>Performance status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1060 (56.7%)</td>
<td>1001 (72.3%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1</td>
<td>688 (36.8%)</td>
<td>352 (25.4%)</td>
<td></td>
</tr>
<tr>
<td>( \geq 2 )</td>
<td>123 (6.6%)</td>
<td>31 (2.2%)</td>
<td></td>
</tr>
</tbody>
</table>

Red Pill Patients

- Younger
- From Europe
- Better Performance
### Patient Characteristics (Continued)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Blue Pill</th>
<th>Red Pill</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prognosis risk category, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>297 (15.9%)</td>
<td>228 (16.5%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Medium risk</td>
<td>684 (36.6%)</td>
<td>656 (47.4%)</td>
<td></td>
</tr>
<tr>
<td>High risk</td>
<td>890 (47.6%)</td>
<td>500 (36.1%)</td>
<td></td>
</tr>
<tr>
<td><strong>Disease stage, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>1210 (64.7%)</td>
<td>851 (61.5%)</td>
<td>0.063</td>
</tr>
<tr>
<td>IV</td>
<td>661 (35.3%)</td>
<td>533 (38.5%)</td>
<td></td>
</tr>
<tr>
<td><strong>Extranodal sites, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2</td>
<td>591 (31.6%)</td>
<td>417 (30.1%)</td>
<td>0.374</td>
</tr>
<tr>
<td>&lt; 2</td>
<td>1280 (68.4%)</td>
<td>967 (69.9%)</td>
<td></td>
</tr>
<tr>
<td><strong>LDH level, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevated</td>
<td>498 (26.6%)</td>
<td>290 (21.0%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Normal</td>
<td>1373 (73.4%)</td>
<td>1094 (79.0%)</td>
<td></td>
</tr>
<tr>
<td><strong>Bone marrow involvement, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1165 (62.3%)</td>
<td>900 (65.0%)</td>
<td>0.106</td>
</tr>
<tr>
<td>No</td>
<td>706 (37.7%)</td>
<td>484 (35.0%)</td>
<td></td>
</tr>
<tr>
<td><strong>Center type, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>400 (21.4%)</td>
<td>197 (14.2%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Community</td>
<td>1471 (78.6%)</td>
<td>1187 (85.8%)</td>
<td></td>
</tr>
</tbody>
</table>

*Notes:*
- Red Pill Patients
  - Better Prognosis Risk
  - Normal LDH Level
  - From Community Center
Specifying the Propensity Score Model

• Choosing main effects is an “art” and science
• Several factors to consider (e.g., strength of association, clinical plausibility, multicollinearity, effect modification)
• Remember the purpose: Control for confounding variables

- Simulation studies have shown:
  - Confounding variables should always be included
Specifying the Propensity Score Model

- Choosing main effects is an “art” and science
- Several factors to consider (e.g., strength of association, clinical plausibility, multicollinearity, effect modification)
- Remember the purpose: Control for confounding variables

Simulation studies have shown:
- Confounding variables should always be included
- Variables associated with only the treatment decrease precision
Choosing main effects is an “art” and science

Several factors to consider (e.g., strength of association, clinical plausibility, multicollinearity, effect modification)

Remember the purpose: Control for confounding variables

Simulation studies have shown:
  - Confounding variables should always be included
  - Variables associated with only the treatment decrease precision
  - Variables associated with only the outcome increase precision
Specifying the Propensity Score Model

- Choosing main effects is an “art” and science
- Several factors to consider (e.g., strength of association, clinical plausibility, multicollinearity, effect modification)
- Remember the purpose: Control for confounding variables

- Simulation studies have shown:
  - Confounding variables should always be included
  - Variables associated with only the treatment decrease precision
  - Variables associated with only the outcome increase precision

- Our strategy: Include all variables associated with outcome
Selection of Patient Characteristics for Model

- Look at variables one at a time (univariable models)
- Determine which variables are associated with any outcome

* Univariable Cox Regression Model Predicting OS;

```
proc phreg data=ad01;
  class region (ref="North America");
  model osmo*osevt(0) = region;
run;
```

* Univariable Cox Regression Model Predicting PFS;

```
proc phreg data=ad01;
  class region (ref="North America");
  model pfsmo*pfsevt(0) = region;
run;
```

* Univariable Logistic Regression Model Predicting ORR;

```
proc logistic data=ad01;
  class region (ref="North America");
  model orr (event = "1") = region;
run;
```
<table>
<thead>
<tr>
<th>Variable</th>
<th>OS</th>
<th></th>
<th>PFS</th>
<th></th>
<th>ORR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>P Value</td>
<td>β</td>
<td>P Value</td>
<td>β</td>
<td>P Value</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 50</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>51 to 60</td>
<td>0.291</td>
<td>0.141</td>
<td>0.038</td>
<td>0.728</td>
<td>0.411</td>
<td>0.001</td>
</tr>
<tr>
<td>61 to 70</td>
<td>0.922</td>
<td>&lt;0.001</td>
<td>0.511</td>
<td>&lt;0.001</td>
<td>-0.189</td>
<td>0.070</td>
</tr>
<tr>
<td>&gt; 70</td>
<td>1.057</td>
<td>&lt;0.001</td>
<td>0.393</td>
<td>0.001</td>
<td>-0.806</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.049</td>
<td>0.536</td>
<td>0.002</td>
<td>0.970</td>
<td>0.021</td>
<td>0.734</td>
</tr>
<tr>
<td>Female</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Geographic region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Europe</td>
<td>-0.041</td>
<td>0.628</td>
<td>-0.058</td>
<td>0.307</td>
<td>0.093</td>
<td>0.274</td>
</tr>
<tr>
<td>Asia</td>
<td>0.167</td>
<td>0.144</td>
<td>0.138</td>
<td>0.076</td>
<td>-0.298</td>
<td>0.004</td>
</tr>
<tr>
<td>Performance status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>1</td>
<td>0.059</td>
<td>0.491</td>
<td>0.092</td>
<td>0.104</td>
<td>0.019</td>
<td>0.859</td>
</tr>
<tr>
<td>≥ 2</td>
<td>0.679</td>
<td>&lt;0.001</td>
<td>0.423</td>
<td>&lt;0.001</td>
<td>-0.334</td>
<td>0.036</td>
</tr>
<tr>
<td>Disease stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Ref</td>
<td>&lt;0.001</td>
<td>Ref</td>
<td>&lt;0.001</td>
<td>Ref</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IV</td>
<td>0.349</td>
<td>&lt;0.001</td>
<td>0.335</td>
<td>&lt;0.001</td>
<td>-0.217</td>
<td>0.001</td>
</tr>
<tr>
<td>Extranodal sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2</td>
<td>0.209</td>
<td>0.011</td>
<td>0.152</td>
<td>0.006</td>
<td>-0.201</td>
<td>0.001</td>
</tr>
<tr>
<td>&lt; 2</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>LDH level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevated</td>
<td>0.673</td>
<td>&lt;0.001</td>
<td>0.335</td>
<td>&lt;0.001</td>
<td>-0.417</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Normal</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Bone marrow involvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.019</td>
<td>0.813</td>
<td>0.001</td>
<td>0.986</td>
<td>-0.018</td>
<td>0.780</td>
</tr>
<tr>
<td>No</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Center type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>0.187</td>
<td>0.051</td>
<td>0.129</td>
<td>0.047</td>
<td>0.068</td>
<td>0.405</td>
</tr>
<tr>
<td>Community</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
</tbody>
</table>
## Selection of Patient Characteristics for Model

### Selected Variables
- Age group
- Geographic region
- Performance status
- Disease stage
- Extranodal sites
- LDH level

<table>
<thead>
<tr>
<th>Variable</th>
<th>OS β</th>
<th>P Value</th>
<th>PFS β</th>
<th>P Value</th>
<th>ORR β</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 50</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>51 to 60</td>
<td>0.291</td>
<td>0.141</td>
<td>0.038</td>
<td>0.728</td>
<td>0.411</td>
<td>0.001</td>
</tr>
<tr>
<td>61 to 70</td>
<td>0.922</td>
<td>&lt;0.001</td>
<td>0.511</td>
<td>&lt;0.001</td>
<td>-0.189</td>
<td>0.070</td>
</tr>
<tr>
<td>&gt; 70</td>
<td>1.057</td>
<td>&lt;0.001</td>
<td>0.393</td>
<td>0.001</td>
<td>-0.806</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.049</td>
<td>0.536</td>
<td>0.002</td>
<td>0.970</td>
<td>0.021</td>
<td>0.734</td>
</tr>
<tr>
<td>Female</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td><strong>Geographic region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Europe</td>
<td>-0.041</td>
<td>0.628</td>
<td>-0.058</td>
<td>0.307</td>
<td>0.093</td>
<td>0.274</td>
</tr>
<tr>
<td>Asia</td>
<td>0.167</td>
<td>0.144</td>
<td>0.138</td>
<td>0.076</td>
<td>-0.298</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>Performance status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>1</td>
<td>0.059</td>
<td>0.491</td>
<td>0.092</td>
<td>0.104</td>
<td>0.019</td>
<td>0.859</td>
</tr>
<tr>
<td>≥ 2</td>
<td>0.679</td>
<td>&lt;0.001</td>
<td>0.423</td>
<td>&lt;0.001</td>
<td>-0.334</td>
<td>0.036</td>
</tr>
<tr>
<td><strong>Disease stage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>IV</td>
<td>0.349</td>
<td>&lt;0.001</td>
<td>0.335</td>
<td>&lt;0.001</td>
<td>-0.217</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Extranodal sites</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2</td>
<td>0.209</td>
<td>0.011</td>
<td>0.152</td>
<td>0.006</td>
<td>-0.201</td>
<td>0.001</td>
</tr>
<tr>
<td>&lt; 2</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td><strong>LDH level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevated</td>
<td>0.673</td>
<td>&lt;0.001</td>
<td>0.335</td>
<td>&lt;0.001</td>
<td>-0.417</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Normal</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td><strong>Bone marrow involvement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.019</td>
<td>0.813</td>
<td>0.001</td>
<td>0.986</td>
<td>-0.018</td>
<td>0.780</td>
</tr>
<tr>
<td>No</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td><strong>Center type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>0.187</td>
<td>0.051</td>
<td>0.129</td>
<td>0.047</td>
<td>0.068</td>
<td>0.405</td>
</tr>
<tr>
<td>Community</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
</tbody>
</table>
Propensity Score Model

* PS Model: Multivariable Logistic Regression Predicting Receipt of BP;

```r
proc logistic data=ad01;
   class agegrp (ref="<=50") region (ref="North America")
       perfstat (ref="0") stage (ref="III")
       extranod (ref=">=2") ldhlevel (ref="Elevated") / param=ref;
   model tx (event="BP") = agegrp region perfstat stage extranod ldhlevel;
   output out=ad02 pred=propensity;
run;
```
Propensity Score Model

* PS Model: Multivariable Logistic Regression Predicting Receipt of BP;

```r
proc logistic data=ad01;
   class agegrp (ref="<=50") region (ref="North America")
      perfstat (ref="0") stage (ref="III")
      extranod (ref=">=2") ldhlevel (ref="Elevated") / param=ref;
   model tx (event="BP") = agegrp region perfstat stage extranod ldhlevel;
   output out=ad02 pred=propensity;
run;
```

Model receipt of the Blue Pill
Propensity Score Model

* PS Model: Multivariable Logistic Regression Predicting Receipt of BP;

```
proc logistic data=ad01;
  class agegrp (ref="<=50") region (ref="North America")
  perfstat (ref="0") stage (ref="III")
  extranod (ref=">=2") ldhlevel (ref="Elevated") / param=ref;
  model tx (event="BP") = agegrp region perfstat stage extranod ldhlevel;
  output out=ad02 pred=propensity;
run;
```

Model receipt of the Blue Pill

As a function of the 6 selected variables
Propensity Score Model

* PS Model: Multivariable Logistic Regression Predicting Receipt of BP;

```sas
proc logistic data=ad01;
    class agegrp (ref="<=50") region (ref="North America")
        perfstat (ref="0") stage (ref="III")
        extranod (ref=">=2") ldhlevel (ref="Elevated") / param=ref;
    model tx (event="BP") = agegrp region perfstat stage extranod ldhlevel;
    output out=ad02 pred=propensity;
run;
```

Model receipt of the Blue Pill

Output the results into a data set named AD02

As a function of the 6 selected variables
Propensity Score Model

* PS Model: Multivariable Logistic Regression Predicting Receipt of BP;

```sas
proc logistic data=ad01;
  class agegrp (ref="<=50") region (ref="North America")
    perfstat (ref="0") stage (ref="III")
    extranod (ref=">=2") ldhlevel (ref="Elevated") / param=ref;
  model tx (event="BP") = agegrp region perfstat stage extranod ldhlevel;
  output out=ad02 pred=propensity;
run;
```

Model receipt of the Blue Pill

Output the results into a data set named AD02

With the variable PROPENSITY for each patient

As a function of the 6 selected variables
Propensity Score Model

* PS Model: Multivariable Logistic Regression Predicting Receipt of BP;

```
proc logistic data=ad01;
    class agegrp (ref="<=50") region (ref="North America")
        perfstat (ref="0") stage (ref="III")
        extranod (ref=">2") ldhlevel (ref="Elevated") / param=ref;
    model tx (event="BP") = agegrp region perfstat stage extranod ldhlevel;
    output out=ad02 pred=propensity;
run;
```

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>-1.1882</td>
<td>0.1860</td>
<td>40.7886</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>agegrp 51 to 60</td>
<td></td>
<td>1.1653</td>
<td>0.1600</td>
<td>53.0770</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>agegrp 61 to 70</td>
<td></td>
<td>2.1720</td>
<td>0.1610</td>
<td>181.9532</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>agegrp &gt;=70</td>
<td></td>
<td>3.1191</td>
<td>0.2114</td>
<td>217.6932</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>region Asia</td>
<td></td>
<td>0.5041</td>
<td>0.1274</td>
<td>15.6471</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>region Europe</td>
<td></td>
<td>-0.4297</td>
<td>0.0835</td>
<td>26.4601</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>perfstat 1</td>
<td></td>
<td>0.6027</td>
<td>0.0856</td>
<td>49.5625</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>perfstat &gt;=2</td>
<td></td>
<td>1.2876</td>
<td>0.2182</td>
<td>34.8199</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>stage IV</td>
<td></td>
<td>0.1179</td>
<td>0.0811</td>
<td>2.1121</td>
<td>0.1461</td>
</tr>
<tr>
<td>extranod &lt;2</td>
<td></td>
<td>-0.0882</td>
<td>0.0849</td>
<td>1.0814</td>
<td>0.2984</td>
</tr>
<tr>
<td>ldhlevel Normal</td>
<td></td>
<td>-0.3877</td>
<td>0.0925</td>
<td>17.5754</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Propensity Score Trimming

- Propensity score distribution through stacked histograms
Propensity Score Trimming

- Propensity score distribution through stacked histograms

- 317 patients excluded from further analysis
- PS-trimmed cohort = 2,945 patients
Application of the Propensity Score

• We will cover 4 methods of applying the propensity score to estimate the adjusted treatment effect:
  – Stratification
  – Matching
  – Inverse Probability Treatment Weighting (IPTW)
  – Covariate adjustment

• Note about covariate balance
Application of the Propensity Score

- We will cover 4 methods of applying the propensity score to estimate the adjusted treatment effect:
  - Stratification
  - Matching
  - Inverse Probability Treatment Weighting (IPTW)
  - Covariate adjustment

- Note about covariate balance

Continuous variables

\[ d = \frac{(X_{BP} - X_{RP})}{\sqrt{\frac{s^2_{BP}}{2} - \frac{s^2_{RP}}{2}}} \]

Categorical variables

\[ d = \frac{(p_{BP} - p_{RP})}{\sqrt{\frac{p_{BP}(1 - p_{BP}) + p_{RP}(1 - p_{RP})}{2}}} \]
Stratification

• Rank propensity scores into quintiles.
• Patients within any particular quintile are similar in characteristics.
• Treatment effect within a particular quintile should not be influenced by measured differences between BP and RP patients.

* Categorize PS-Trimmed Cohort into PS Quintiles;

```sas
proc rank data=ad02 out=strat01 groups=5;
  where (pstrimcohort eq 1);
  var propensity;
  ranks psrank;
run;

data strat02;
  set strat01;
  psquin = (psrank + 1);
  label psquin = "Propensity Score Quintile (1 to 5)";
run;
```
Stratification: Relative Treatment Effect

- Model each outcome as a function of treatment and propensity score quintile:

* PS Stratification;
* Relative Treatment Effect (Hazard Ratio): OS;

```sas
proc phreg data=strat02;
  class tx (ref="BP") psquin;
  model osmo*osevt(0) = tx psquin / rl;
run;
```

* Relative Treatment Effect (Hazard Ratio): PFS;

```sas
proc phreg data=strat02;
  class tx (ref="BP") psquin;
  model pfsmo*pfsevt(0) = tx psquin / rl;
run;
```

* Relative Treatment Effect (Odds Ratio): ORR;

```sas
proc logistic data=strat02;
  class tx (ref="BP") psquin / param=ref;
  model orr (event = "1") = tx psquin / rl;
run;
```
Stratification: Relative Treatment Effect

- Model each outcome as a function of treatment and propensity score quintile:

  * PS Stratification;
  * Relative Treatment Effect (Hazard Ratio): OS;

```plaintext
proc phreg data=strat02;
  class tx (ref="BP") psquin;
  model osmo*osevt(0) = tx psquin / rl;
run;
```

* Relative Treatment Effect (Hazard Ratio): PFS;

```plaintext
proc phreg data=strat02;
  class tx (ref="BP") psquin;
  model pfsmo*pfsevt(0) = tx psquin / rl;
run;
```

* Relative Treatment Effect (Odds Ratio): ORR;

```plaintext
proc logistic data=strat02;
  class tx (ref="BP") psquin / param=ref;
  model orr (event = "1") = tx psquin / rl;
run;
```
Matching

- Each RP patient is matched to a BP patient with similar propensity score
- Various matching algorithms (we used 1:1 “greedy” matching)
- In our example:
  - 960 RP patients matched to 960 BP patients = 1,920 patients total
  - Additional 1,025 patients excluded
- Account for matched pairs in treatment effect.
Matching: Relative Treatment Effect

• Model each outcome adjusting for matched pairs:

* PS Matching Accounting for Clustering Within Matched Pairs;
* Relative Treatment Effect (Hazard Ratio): OS;

```r
proc phreg data=ad_matching covs(aggregate);
  id matchid;
  class tx (ref="BP");
  model osmo*osevt(0) = tx / rl;
run;
```

* Relative Treatment Effect (Hazard Ratio): PFS;

```r
proc phreg data=ad_matching covs(aggregate);
  id matchid;
  class tx (ref="BP");
  model pfsmo*pfsevt(0) = tx / rl;
run;
```

* PS Matching Accounting for Clustering Within Matched Pairs;
* Relative Treatment Effect (Odds Ratio): ORR;

```r
proc genmod data=ad_matching desc;
  class matchid;
  model orr = tx / dist=bin link=logit;
  repeated subject=matchid / type=un corrw covb;
  estimate 'RP vs. BP' tx 1 /exp;
run;
```
Matching: Relative Treatment Effect

- Model each outcome adjusting for matched pairs:

* PS Matching Accounting for Clustering Within Matched Pairs;
* Relative Treatment Effect (Hazard Ratio): OS;

```sas
proc phreg data=ad_matching covs(aggregate);
    id matchid;
    class tx (ref="BP");
    model osmo*osevt(0) = tx / rl;
run;
```

* Relative Treatment Effect (Hazard Ratio): PFS;

```sas
proc phreg data=ad_matching covs(aggregate);
    id matchid;
    class tx (ref="BP");
    model pfsmo*pfsevt(0) = tx / rl;
run;
```

* PS Matching Accounting for Clustering Within Matched Pairs;
* Relative Treatment Effect (Odds Ratio): ORR;

```sas
proc genmod data=ad_matching desc;
    class matchid;
    model orr = tx / dist=bin link=logit;
    repeated subject=matchid / type=un corrw covb;
    estimate 'RP vs. BP' tx 1 /exp;
run;
```
Matching: Relative Treatment Effect

- Model each outcome adjusting for matched pairs:
  
  * PS Matching Accounting for Clustering Within Matched Pairs;
  * Relative Treatment Effect (Hazard Ratio): OS;

```
proc phreg data=ad_matching covs(aggregate);
  id matchid;
  class tx (ref="BP");
  model osmo*osevt(0) = tx / rl;
run;
```

* Relative Treatment Effect (Hazard Ratio): PFS;

```
proc phreg data=ad_matching covs(aggregate);
  id matchid;
  class tx (ref="BP");
  model pfsmo*pfsevt(0) = tx / rl;
run;
```

* PS Matching Accounting for Clustering Within Matched Pairs;
* Relative Treatment Effect (Odds Ratio): ORR;

```
proc genmod data=ad_matching desc;
  class matchid;
  model orr = tx / dist=bin link=logit;
  repeated subject=matchid / type=un corrw covb;
  estimate 'RP vs. BP' tx 1 /exp;
run;
```
IPTW

• Each patient weighted by the inverse of the probability of receiving the treatment that he or she actually received:
  – BP patients: \( \text{WEIGHT} = \frac{1}{\text{PS}} \)
  – RP patients: \( \text{WEIGHT} = \frac{1}{(1 - \text{PS})} \)

• Weights are often “stabilized” as follows:
  – BP Patients: \( \text{STWEIGHT} = \frac{\text{P}_{BP}}{\text{PS}} \)
  – RP Patients: \( \text{STWEIGHT} = \frac{(1 - \text{P}_{BP})}{(1 - \text{PS})} \)
IPTW: Relative Treatment Effect

- Perform the appropriate regression while adjusting for stabilized weights:

  * IPTW with Stabilized Weights;
  * Relative Treatment Effect (Hazard Ratio): OS;

  ```
  proc phreg data=ad_iptw;
    class tx (ref="BP");
    model osmo*osevt(0) = tx / rl;
    weight stweight / normalize;
  run;
  *
  * Relative Treatment Effect (Hazard Ratio): PFS;

  proc phreg data=ad_iptw;
    class tx (ref="BP");
    model pfsmo*pfsevt(0) = tx / rl;
    weight stweight / normalize;
  run;
  *
  * Relative Treatment Effect (Odds Ratio): ORR;

  proc logistic data=ad_iptw;
    class tx (ref="BP") / param=ref;
    model orr (event="1") = tx / rl;
    weight stweight / normalize;
  run;
  ```
IPTW: Relative Treatment Effect

- Perform the appropriate regression while adjusting for stabilized weights:

* IPTW with Stabilized Weights;
* Relative Treatment Effect (Hazard Ratio): OS;

```sas
proc phreg data=ad_iptw;
  class tx (ref="BP");
  model osmo*osevt(0) = tx / rl;
  weight stweight / normalize;
run;
```

* Relative Treatment Effect (Hazard Ratio): PFS;

```sas
proc phreg data=ad_iptw;
  class tx (ref="BP");
  model pfsmo*pfsevt(0) = tx / rl;
  weight stweight / normalize;
run;
```

* Relative Treatment Effect (Odds Ratio): ORR;

```sas
proc logistic data=ad_iptw;
  class tx (ref="BP") / param=ref;
  model orr (event="1") = tx / rl;
  weight stweight / normalize;
run;
```
Covariate adjustment

- Most elementary approach
- Simply include propensity score as a continuous covariate in the model.

* Covariate Adjustment of PS;
* Relative Treatment Effect: OS;
```
proc phreg data=ad02;
    class tx (ref="BP");
    model osmo*osevt(0) = tx propensity / rl;
run;
```

* Relative Treatment Effect: PFS;
```
proc phreg data=ad02;
    class tx (ref="BP");
    model pfsmo*pfsevt(0) = tx propensity / rl;
run;
```

* Relative Treatment Effect: ORR;
```
proc logistic data=ad_matching;
    class tx (ref="BP") / param=ref;
    model orr (event = "1") = tx propensity / rl;
run;
```
Covariate adjustment

- Most elementary approach
- Simply include propensity score as a continuous covariate in the model.

* Covariate Adjustement of PS;
* Relative Treatment Effect: OS;

```
proc phreg data=ad02;
    class tx (ref="BP");
    model osmo*osevt(0) = tx propensity / rl;
run;
```

* Relative Treatment Effect: PFS;

```
proc phreg data=ad02;
    class tx (ref="BP");
    model pfsmo*pfsevt(0) = tx propensity / rl;
run;
```

* Relative Treatment Effect: ORR;

```
proc logistic data=ad_matching;
    class tx (ref="BP") / param=ref;
    model orr (event = "1") = tx propensity / rl;
run;
```
Results: OS

Crude (no adjustment)

HR (95% CI)
0.565 (0.478-0.667)

Red Pill is Better
Blue Pill is Better

Hazard Ratio

0.125 0.25 0.5 1 2 4 8
Results: OS

<table>
<thead>
<tr>
<th>Adjustment</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude (no adjustment)</td>
<td>0.565 (0.478-0.667)</td>
</tr>
<tr>
<td>Stratification</td>
<td>0.771 (0.643-0.924)</td>
</tr>
<tr>
<td>Matching</td>
<td>0.799 (0.656-0.974)</td>
</tr>
<tr>
<td>IPTW</td>
<td>0.823 (0.697-0.972)</td>
</tr>
<tr>
<td>Covariate Adjustment</td>
<td>0.792 (0.661-0.950)</td>
</tr>
<tr>
<td>Multivariable Model</td>
<td>0.765 (0.638-0.916)</td>
</tr>
</tbody>
</table>

Red Pill is Better
Blue Pill is Better
Results: PFS

<table>
<thead>
<tr>
<th>Adjustment Type</th>
<th>Hazard Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude (no adjustment)</td>
<td>0.594 (0.533-0.663)</td>
</tr>
<tr>
<td>Stratification</td>
<td>0.688 (0.609-0.777)</td>
</tr>
<tr>
<td>Matching</td>
<td>0.659 (0.575-0.756)</td>
</tr>
<tr>
<td>IPTW</td>
<td>0.707 (0.632-0.791)</td>
</tr>
<tr>
<td>Covariate Adjustment</td>
<td>0.686 (0.607-0.775)</td>
</tr>
<tr>
<td>Multivariable Model</td>
<td>0.679 (0.603-0.766)</td>
</tr>
</tbody>
</table>

Red Pill is Better
Blue Pill is Better
Results: ORR

<table>
<thead>
<tr>
<th>Covariate Adjustment</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude (no adjustment)</td>
<td>3.84 (2.82-5.23)</td>
</tr>
<tr>
<td>Stratification</td>
<td>2.62 (1.88-3.66)</td>
</tr>
<tr>
<td>Matching</td>
<td>2.62 (1.81-3.75)</td>
</tr>
<tr>
<td>IPTW</td>
<td>2.64 (1.94-3.59)</td>
</tr>
<tr>
<td>Covariate Adjustment</td>
<td>2.61 (1.87-3.63)</td>
</tr>
<tr>
<td>Multivariable Model</td>
<td>2.93 (2.10-4.10)</td>
</tr>
</tbody>
</table>

Blue Pill is Better
Red Pill is Better
Conclusions

• Propensity scores are a powerful tool to deal with confounding when comparing treatments in observational studies, especially when there are few outcomes.

• Propensity score methods can adjust only for measured confounding variables.

• In our simulation:
  – Propensity score methods attenuated the crude observed treatment effect across all outcomes (all measures of association closer to the null).
  – Stratification, matching, and covariate adjustment yielded similar results for OS and PFS, while IPTW yielded point estimates closer to the null.
  – For ORR, similar odds ratios were obtained for all propensity score methods.
  – Multivariable model gave similar results to the propensity score results.
Conclusions

Choose the Red Pill! 😊
Contact Information

Lawrence Rasouliyan  
+34 933 624 297  
lrasouliyan@rti.org

Estel Plana  
+34 933 622 832  
eplana@rti.org

Jaume Aguado  
+34 933 624 251  
jaguado@rti.org