Propensity Score Methods to Adjust for Bias in Observational Data

SAS HEALTH USERS GROUP

APRIL 6, 2018
Overview

1. What is observational data?

2. What is the propensity score?

3. Statistical adjustment using the propensity score
   a) Matching on the propensity score
   b) Inverse probability of treatment weighting
Randomized Controlled Trials (the “gold standard”)

Inclusion/exclusion criteria

Population

Study population

Baseline

Treatment Group

Control Group

Follow-up

Outcome
Characteristics of RCTs

• Randomization ensures subjects in both treatment groups are equally matched on all factors

• Allow causal inference
But

- High cost

- Often short duration and/or underpowered.

- Problems with generalizability:
  - Treatment is “ideal” (high compliance, careful follow-up means that any problems may be caught early).
  - Many people who are given the treatments in “real life” are excluded from the trials

- Some situations cannot be randomized.
What is Observational Data?

The choice of treatment is not under the control of the researcher - the researcher can only ‘observe’ what treatment was given.

Examples:
- Data obtained using chart review
- Electronic medical records
- Survey data or health study data
- Administrative data.
The Study

Two medications used to treat chronic obstructive lung disease (COPD)

- Long-acting anticholinergic (LAAC)
- Long-acting beta-agonist (LABA)

Compare overall mortality and risk of hospital admission related to COPD

Gershon et al. Annals of Internal Medicine, 2011
Ontario Drug Benefit Plan (which drug is the person taking)

Hospital Discharge Database (diagnoses)

Physician Billing Database (diagnoses)

Registered Persons Database (age, sex, SES)

Hospital Discharge Database (for outcomes).
Analysis of Our Study

Exposure variable is choice of drug (LAAC vs. LABA)

Outcome is time to hospitalization or death. ∴ survival analysis will be used

Bias is a concern
## Bias

### Bias in confounders we can measure

<table>
<thead>
<tr>
<th>Covariate</th>
<th>LAAC</th>
<th>LABA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialist care in last year (%)</td>
<td>44.7</td>
<td>50.5</td>
</tr>
<tr>
<td>Prior lung function testing (%)</td>
<td>69.7</td>
<td>74.3</td>
</tr>
</tbody>
</table>

And bias in confounders we can’t measure, e.g., smoking, fitness
Statistical Adjustment for Observational Data

- Propensity score methods
- Instrumental variable analysis
- And others
Propensity Score

Rosenbaum and Rubin (1983) realized the bias from covariates can be eliminated by controlling for a scalar-valued function (a “balancing score”) calculated from the baseline covariates, i.e., the propensity score.

The propensity score is a way of summarizing the information in all the prognostic variables.
What is the Propensity Score?

PS = probability that a person received one treatment (rather than the other), given that person’s observed covariates

Calculated using logistic regression to estimate the propensity for a person to be prescribed a LAAC (rather than a LABA)

```plaintext
proc logistic descending;
    model LAAC = age sex diabetes hypertension rural_res incquint ...;
    output out = score predicted = ps;
run;
```
Calculating the Propensity Score

```latex
proc logistic descending;
  model LAAC = age sex diabetes hypertension rural_res incquint ....;
```

Patients predicted, based on their characteristics, to be likely to be prescribed a LAAC will have a high propensity score.

Patients predicted to be unlikely to be prescribed a LAAC (likely to be prescribed a LABA instead) will have a low propensity score.
Variable Selection

1. All measured baseline covariates
2. Baseline covariates associated with treatment choice
3. Baseline covariates associated with the outcome
4. Baseline covariates associated with both treatment assignment and outcome
Propensity Score Methods

1. Covariate adjustment using the Propensity Score

2. Stratification on the PS

3. Matching on the PS.

4. Inverse probability weighting
Matching
Matching

1. Create a matched sample based on logit(PS)
2. Assess balance between treated and untreated subjects in the matched sample.
   – The test of a good propensity score model is how well it balances the measured variables between treated and untreated subjects.
3. For unbalanced variables, add interactions or higher order terms to the propensity score logistic regression, recalculate the propensity score and repeat the process.
<table>
<thead>
<tr>
<th>Baseline Covariate</th>
<th>Before Matching</th>
<th>After Matching</th>
<th>Standard difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LAAC N=28,563</td>
<td>LABA N=17,840</td>
<td></td>
</tr>
<tr>
<td>Lung function testing (%)</td>
<td>69.7</td>
<td>74.3</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>72.4</td>
<td>73.0</td>
<td>1.3</td>
</tr>
<tr>
<td>Specialist care previous year (%)</td>
<td>44.7</td>
<td>50.5</td>
<td>11.6</td>
</tr>
<tr>
<td></td>
<td>49.0</td>
<td>49.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Also using inhaled corticosteroid</td>
<td>48.3</td>
<td>52.1</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>51.1</td>
<td>51.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Co-diagnosis of CHF</td>
<td>40.2</td>
<td>38.2</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>39.0</td>
<td>39.2</td>
<td>0.4</td>
</tr>
<tr>
<td>Hospitalized for COPD in previous 6 months</td>
<td>8.0</td>
<td>7.3</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>7.8</td>
<td>7.8</td>
<td>0.1</td>
</tr>
</tbody>
</table>
## Analysis of Matched Data

**Analysis of Matched Data Must Incorporate the Matching**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Means</strong></td>
<td>Paired t-test</td>
</tr>
<tr>
<td><strong>Proportions</strong></td>
<td>McNemar’s test</td>
</tr>
<tr>
<td><strong>Survival models</strong></td>
<td>Stratify on matched pairs</td>
</tr>
<tr>
<td><strong>Logistic regression</strong></td>
<td>GEE estimation to account for matched pairs</td>
</tr>
</tbody>
</table>
Matched Analyses ...

- Compares patients who are all potential candidates for both treatments.

- Matching pairs patients who are similar with respect to their propensity score matches on many confounders simultaneously.

- Unmatched individuals are discarded.

- The resulting matched sample may not be representative of all patients receiving treatment.
Interpretation of a Matched Analysis

Estimates the Average Treatment Effect for the Treated (ATT) – the average treatment effect for those who ultimately received the treatment
Inverse Probability of Treatment Weighting Using the Propensity Score
The Weights

\[ W = \frac{Z}{PS} + \frac{1 - Z}{1 - PS} \]

where \( Z = 1 \) for the treatment group and 0 for the control group
The Weights

Recall that our PS is the probability of receiving a LAAC (rather than a LABA)

\[ W = \frac{LAAC}{PS} + \frac{1 - LAAC}{1 - PS} \]

where LAAC is a 0/1 variable.
The Weights

\[ W = \frac{LAAC}{PS} + \frac{1 - LAAC}{1 - PS} \]

For those who received LAAC (LAAC = 1), weight = \( \frac{1}{PS} \) (probability of receiving LAAC):

\[ W = \frac{1}{PS} \]

For those who received LABA (LAAC = 0), weight = \( \frac{1}{1 - PS} \) (probability of receiving LABA):

\[ W = \frac{1}{1 - PS} \]
The Weights

Similar to survey weights

Respondents from oversampled groups are assigned low weights

- Selection probability = 1% \( \Rightarrow \) weight = \( \frac{1}{0.01} \) = 100

Respondents from undersampled groups are assigned high weights

- Selection probability = 0.2% \( \Rightarrow \) weight = \( \frac{1}{0.002} \) = 500
**Data Set to Estimate the Outcome of Treatment**

\[ W = \frac{Z}{PS} + \frac{1 - Z}{1 - PS} \]

<table>
<thead>
<tr>
<th>ID</th>
<th>Z treatment = 1</th>
<th>PS</th>
<th>Weight 1/PS</th>
<th>Outcome under treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>treatment</td>
<td>0.33</td>
<td>1 / 0.33 = 3</td>
<td>Y₁</td>
</tr>
<tr>
<td>2</td>
<td>control</td>
<td>0.33</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>control</td>
<td>0.33</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>treatment</td>
<td>0.67</td>
<td>1 / 0.67 = 1.5</td>
<td>Y₄</td>
</tr>
<tr>
<td>5</td>
<td>treatment</td>
<td>0.67</td>
<td>1 / 0.67 = 1.5</td>
<td>Y₅</td>
</tr>
<tr>
<td>6</td>
<td>control</td>
<td>0.67</td>
<td>0</td>
<td>?</td>
</tr>
</tbody>
</table>
Data Set to Estimate the Outcome of Treatment

Estimated average outcome of treatment  = \( \frac{1}{N} \sum_{i=1}^{N} w_i \times Y_i \)  

where \( w_i = \frac{1}{PS} \) for treated people and 0 for controls.

<table>
<thead>
<tr>
<th>ID</th>
<th>Z treatment = 1 control = 0</th>
<th>PS</th>
<th>Weight 1/PS</th>
<th>Outcome under treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>treatment</td>
<td>0.33</td>
<td>1 / 0.33 = 3</td>
<td>( Y_1 )</td>
</tr>
<tr>
<td>2</td>
<td>control</td>
<td>0.33</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>control</td>
<td>0.33</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>treatment</td>
<td>0.67</td>
<td>1 / 0.67 = 1.5</td>
<td>( Y_4 )</td>
</tr>
<tr>
<td>5</td>
<td>treatment</td>
<td>0.67</td>
<td>1 / 0.67 = 1.5</td>
<td>( Y_5 )</td>
</tr>
<tr>
<td>6</td>
<td>control</td>
<td>0.67</td>
<td>0</td>
<td>?</td>
</tr>
</tbody>
</table>
## Data Set to Estimate the Outcome for Controls

\[ W = \frac{Z}{PS} + \frac{1 - Z}{1 - PS} \]

<table>
<thead>
<tr>
<th>ID</th>
<th>Z</th>
<th>PS</th>
<th>Weight</th>
<th>Outcome under control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>treatment</td>
<td>0.33</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>control</td>
<td>0.33</td>
<td>( \frac{1}{1 - 0.33} = 1.5 )</td>
<td>( Y_2 )</td>
</tr>
<tr>
<td>3</td>
<td>control</td>
<td>0.33</td>
<td>( \frac{1}{1 - 0.33} = 1.5 )</td>
<td>( Y_3 )</td>
</tr>
<tr>
<td>4</td>
<td>treatment</td>
<td>0.67</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>treatment</td>
<td>0.67</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>control</td>
<td>0.67</td>
<td>( \frac{1}{1 - 0.67} = 3 )</td>
<td>( Y_6 )</td>
</tr>
</tbody>
</table>
Data Set to Estimate the Outcome for Controls

Estimated average effect for controls = \( \frac{1}{N} \sum_{i=1}^{N} w_i \times Y_i \) where \( w_i = \frac{1}{1 - PS} \) for people in the control group and 0 for people in the treated group

<table>
<thead>
<tr>
<th>ID</th>
<th>Z Treatment = 1; Control = 0</th>
<th>PS</th>
<th>Weight</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treatment</td>
<td>0.33</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>control</td>
<td>0.33</td>
<td>( \frac{1}{1 - 0.33} = 1.5 )</td>
<td>( Y_2 )</td>
</tr>
<tr>
<td>3</td>
<td>control</td>
<td>0.33</td>
<td>( \frac{1}{1 - 0.33} = 1.5 )</td>
<td>( Y_3 )</td>
</tr>
<tr>
<td>4</td>
<td>treatment</td>
<td>0.67</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>treatment</td>
<td>0.67</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>control</td>
<td>0.67</td>
<td>( \frac{1}{1 - 0.67} = 3 )</td>
<td>( Y_6 )</td>
</tr>
</tbody>
</table>
Estimating the Treatment Difference

Estimated difference (treatment A – treatment B) =

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{Z_i \times Y_i}{PS} - \frac{1}{N} \sum_{i=1}^{N} \frac{(1 - Z_i) \times Y_i}{1 - PS}
\]

Estimate of the variance

– Robust sandwich type variance estimators
– Bootstrapping

May trim very large weights (propensity score < 1\textsuperscript{st} percentile or > 99\textsuperscript{th} percentile)
Interpretation of an Inversely Weighted Analysis

Estimates the Average Treatment Effect (ATE): an estimate of the treatment effect, if it were applied to the entire population
It’s Magic
Well, Not Quite

The analyses make no claims to balance unmeasured covariates.

The analyses remove hidden biases only to the extent that the unmeasured variables are correlated with the available covariates.

Sensitivity analyses can help quantify the possible effects of unmeasured confounders.
Drawbacks to the Propensity Score

Available data is probably missing key covariates (e.g., living arrangements, smoking history)

Definition of the baseline time may be difficult (it should be the time at which the decision about treatment was made).

Does not eliminate the need to think about patient identification and selection
Advantages of the Propensity Score

Reduced dimensionality of covariates (important for rare outcomes)
Can demonstrate that the two groups are similar on all measured covariates

Like an RCT, does not predict the outcome for a person with a given set of characteristics
Like an RCT, does not tell you the role of the covariates in predicting the outcome
Like an RCT, can build planned sub-analyses into the design
Advantages of Observational Studies

Useful when it is not feasible to use an RCT
  – Unethical to withhold treatment
  – Exposure believed to be harmful
  – Patients will not agree to be randomized
  – RCT too expensive

Generalizable (all patients, all providers)
Allows studies of rare events, and studies with long follow-up times
Disadvantages of Observational Studies

Researcher has no control over assignment of subjects to treatments

Researcher often has no control over what covariates are available, their definitions, or the quality of their measurement
References


Anything else written by Peter Austin

Introducing the PSMATCH procedure for propensity score analysis: https://www.youtube.com/watch?v=JM2uu39zEAs (a very good introduction to both propensity scores and matching as well as the PSMATCH procedure)
Thank You