Adjusting for Bias in Observational Data

Inverse Probability of Treatment Weighting using the Propensity Score

Why Use Observational Data?

- Full spectrum of patients and providers.
- Some studies are not ethical to conduct as randomized trials.
- Much less expensive than randomized trials.
- Larger sample size, much longer follow-up for less common outcomes and long-term outcomes.

But

- Assignment bias
- Missing information (variables that are not available)

Dealing with Bias

- Regression
- Stratification
- Instrumental variable analysis
- Propensity score methods

Calculating the Propensity Score

- A way of summarizing the information in all of the prognostic variables
- PS = probability of one of the two treatments, given the observed covariates
- Logistic regression:

  P(treatment A rather than B) = f(age, sex, comorbidities, etc.)

Propensity score

Logistic regression estimates the propensity for patients to be treated with A rather than B, based on patient characteristics

```
proc logistic descending;
    model A = age sex diabetes COPD
    rurality …;
output out = propensity predicted = PS;
```

- PS ~ propensity of physicians to choose one treatment based on patient characteristics
- Patients *<u>predicted</u>* to be unlikely to be treated with A → low propensity score
- Patients <u>predicted</u> to be likely to be treated with A → high propensity score.

Using the Propensity Score

- Stratification
- Regression
- Matching
- Inverse probability of treatment weighting

Inverse Probability of Treatment Weighting Using the Propensity Score

Weights

$$\boldsymbol{W} = \frac{\boldsymbol{treatmentA}}{\boldsymbol{PS}} + \frac{\boldsymbol{1 - treatmentA}}{\boldsymbol{1 - PS}}$$

where treatmentA = 1 if the person received treatment A, and 0 if the person received treatment B

Weights

Treatment A*:* $W = \frac{1}{PS}$

Treatment B: $W = \frac{1}{1 - PS}$

Subjects weighted by the inverse of the probability of receiving the treatment that was actually received.

How it Works

- Create two datasets: one for each treatment group.
- Everyone contributes to both datasets

Our Data

ID	Treatment group	PS P(A)	Outcome
1	A	0.80	Y_1
2	A	0.77	Y_2
3	B	0.70	Y_3
4	A	0.53	Y_4
5	B	0.50	Y_5
6	B	0.45	Y_6
7	A	0.33	Y_7
8	B	0.25	Y_8

Data Set for the Effect of Treatment A

ID	Treatment group	PS P(A)	Outcome on Treatment A
1	A	0.80	Y_1
2	A	0.77	Y_2
3	B	0.70	**?**
4	A	0.53	Y_4
5	B	0.50	?
6	B	0.45	?
7	A	0.33	Y_7
8	B	0.25	?

Data Set for the Effect of Treatment A

ID	Treatment group	PS P(A)	Weight = 1 / PS	Outcome on Treatment A
1	A	0.80	1.25	Y_1
2	A	0.77	1.30	Y_2
3	B	0.70	0	**?**
4	A	0.53	1.89	Y_4
5	B	0.50	0	?
6	B	0.45	0	?
7	A	0.33	3.03	Y_7
8	B	0.25	0	?

In dataset 1 we are missing information about the effect of treatment A for people who received B.

- If Mr. X, who received A, had a low (e.g. 20%) probability of getting A, there must be 4 similar people who received B.
- Mr. X's weight is 1/0.2 = 5. He represents 5 people on treatment A (himself and 4 others).

- If they had received A, we expect their outcome would be the same as Mr. X's outcome. We impute the missing outcome for these people using Mr. X's outcome.

Data Set for the Effect of Treatment B

ID	Treatment group	PS P(A)	Weight = 1 / (1 – PS)	Outcome of Treatment B
1	A	0.80	0	?
2	A	0.77	0	?
3	B	0.70	3.33	Y_3
4	A	0.53	0	?
5	B	0.50	2.00	Y_5
6	B	0.45	1.82	Y_6
7	A	0.33	0	?
8	B	0.25	1.33	Y_8

Estimating the Effects of Treatment A and Treatment B

Estimated average effect of treatment A

$$= 1/N \sum_{i=1}^{N} \frac{treatmentA \times Yi}{PS}$$

Estimated average effect of treatment B

$$= 1/N \sum_{i=1}^{N} \frac{(1 - treatmentA) \times Yi}{1 - PS}$$

Treatment Difference

Estimated difference (treatment A – treatment B) =

1/N $\sum_{i=1}^{N} \frac{(treatment)\times Yi}{PS}$ -

1/N $\sum_{i=1}^{N} \frac{(1-treatment)\times Yi}{1-PS}$

However, the estimate of the variance is not as straightforward.

Interpretation

Result of an inversely weighted propensity score analysis is an aggregate estimate of the treatment effect, if it were <u>applied to the entire population</u>

- Unusual individuals (treated but don’t fit the description of those usually treated, $\therefore$ small PS) have high weights
- Unusual individuals (not treated, but look a lot like people who are usually treated, $\therefore$ low value for (1-PS)) have high weights

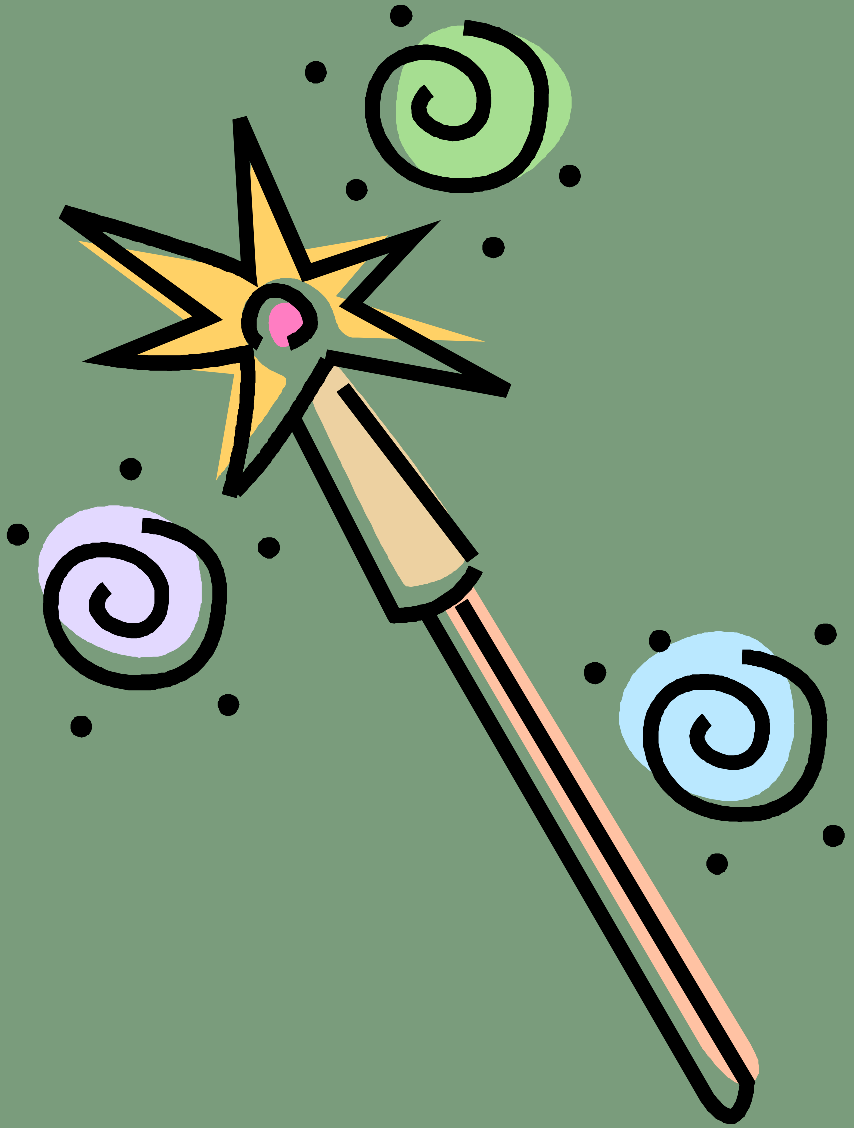
- May trim high weights.

IPW is Related to Survey Weights

In the CCHS, we oversample people from rural areas.

In order to obtain a population estimate, we upweight the responses from urban respondents and downweight the responses from rural people.

It's magic



Well, almost magic

Makes no claims to balance unmeasured covariates.

Remove hidden biases only to the extent that unmeasured variables are correlated with the measures used to compute the score.

What Questions are not Answered

- Does *not* predict the outcome for a person with a given set of characteristics
- Does *not* tell you the role of the other covariates in predicting the outcome (e.g. are older patients more likely to have a stroke) (this is what regression does).
- Does not tell you who will benefit most from a given treatment.

References

Peter C. Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research, 46:399–424, 2011

Peter C. Austin. A tutorial and case study in propensity score analysis: An application to estimating the effect of in-hospital smoking cessation counseling on mortality.

Multivariate Behavioral Research, 46:119–151, 2011

"OUR STATISTICIAN WILL DROP IN AND EXPLAIN WHY YOU HAVE NOTHING TO WORRY ABOUT."