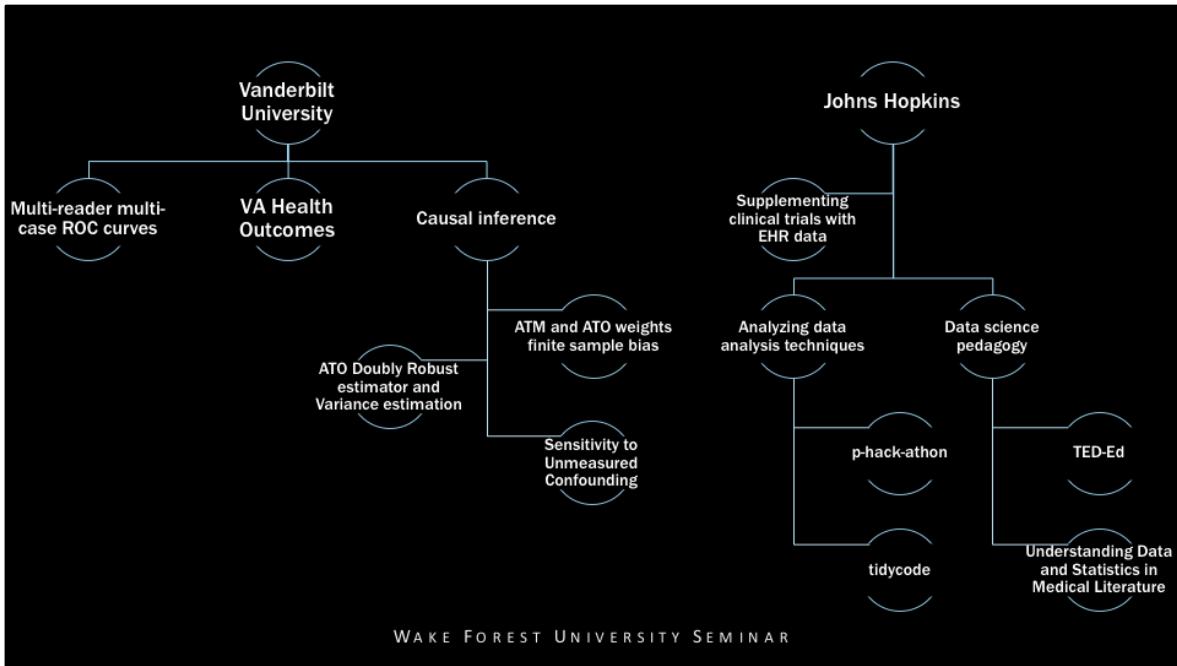


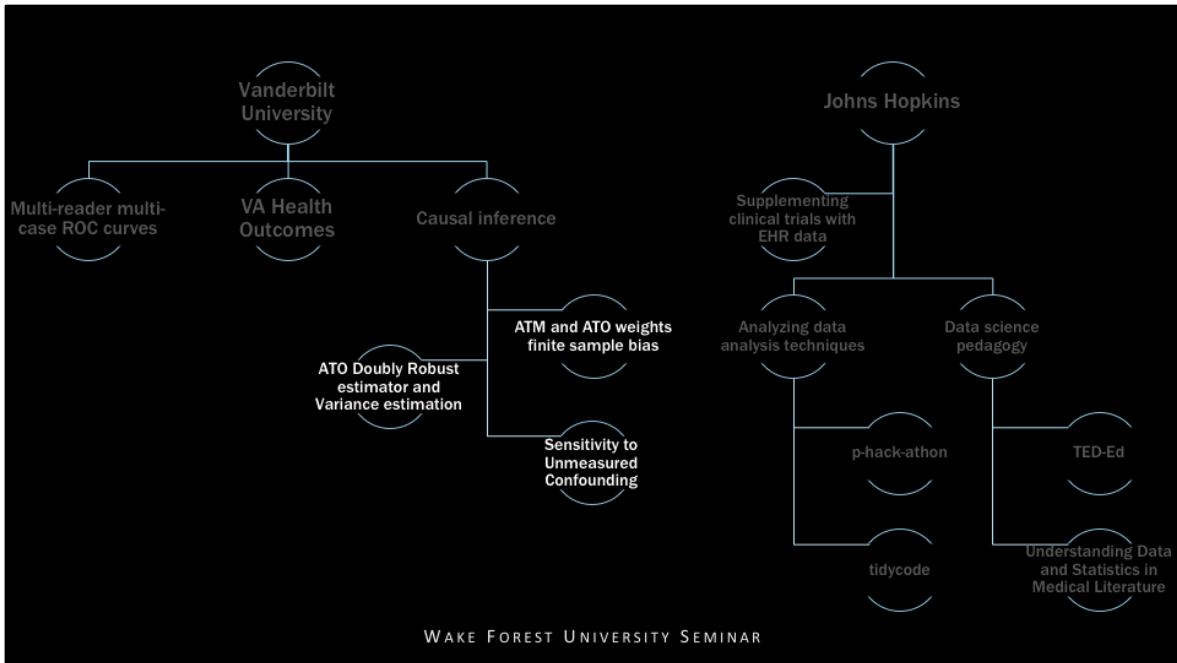
Improving Modern Techniques of Causal Inference

Lucy D'Agostino McGowan

WAKE FOREST UNIVERSITY SEMINAR



Broadly, I get excited about identifying a large problem in the scientific community and finding a way statistics can help.



Background

How do we compensate for non-random treatment assignment?
How do we incorporate the propensity scores?

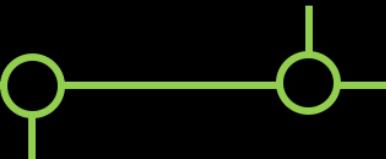


Method

How do we calculate these weights?
Which weights should we choose?

Sensitivity Analysis

What if our assumptions are violated?



Estimation

How do we estimate the treatment effect?
How do we estimate the variability?

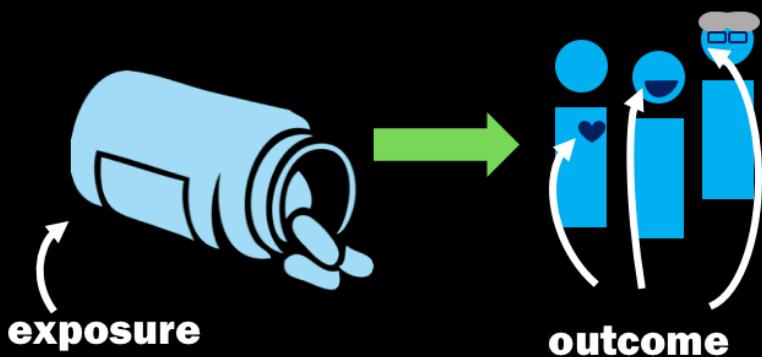
WAKE FOREST UNIVERSITY SEMINAR

- Dr. Raynor mentioned that it'd be best for this talk not to be too technical since there would be a broad audience. Going to have some equations for those of you that get excited about understanding precisely what I've done, but they will always be paired with a high level explanation along with some fun animations. SO if the equation doesn't quite make sense, hang on tight for the high level explanation, and if that still doesn't make sense, please interrupt me, I'm happy to clarify at any point!

Background

WAKE FOREST UNIVERSITY SEMINAR

Observational Studies

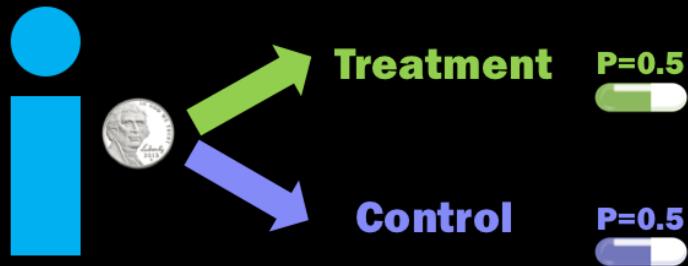


WAKE FOREST UNIVERSITY SEMINAR

A lot of my research so far has been in the observational study space. Like many statistical questions, our goal here is to assess the association between some [click] exposure, in my case often a treatment, and some [click] outcome.

Observational Studies

Randomized Controlled Trials



WAKE FOREST UNIVERSITY SEMINAR

An observational study is distinct from a [click] randomized controlled trial. A randomized study works something like this [click] a participant is assigned to either the [click] treatment or control group [click] this treatment assignment is via some random process, in the simplest case something like a coin flip. We know the probability of treatment assignment and often patients are just as likely to have been assigned to the treatment group as the control group

Estimate treatment effect

$$\frac{\sum_{i=1}^n Y_i Z_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n Y_i (1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

In this randomized setting, we can just take the the difference between the mean outcome among the treated and the mean outcome among the controls. I'm showing this specific notation because we are going to build on it in the subsequent slides.

Estimate treatment effect

$$\frac{\sum_{i=1}^n Y_i Z_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n Y_i (1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

Y_i is the observed outcome for participant i

Estimate treatment effect

$$\frac{\sum_{i=1}^n Y_i Z_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n Y_i (1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

Z_i is the indicator for treatment for participant i (it is 0 or 1)

Estimate treatment effect

$$\frac{\sum_{i=1}^n Y_i Z_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n Y_i (1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

the denominator of this first portion, the sum of Zi, is just your total number of treated people

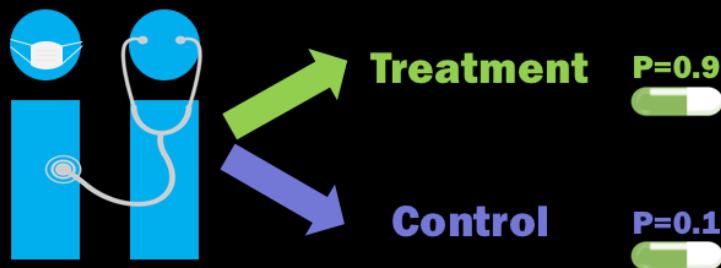
Estimate treatment effect

$$\frac{\sum_{i=1}^n Y_i Z_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n Y_i (1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

sum of (1-Zi) is your total number of controls

Observational Studies



WAKE FOREST UNIVERSITY SEMINAR

An observational study differs from a randomized study in that now instead of a random process such as a coin flip determining your treatment assignment, the treatment assignment [click] is determined by something else, for example your doctor. We may still be able to calculate [click] the probability that you were assigned to the treatment group, but it is no longer random. For example, one study I worked on looked at two different diabetes drugs and their association with heart disease. Instead of conducting a randomized trial, we looked at EHR data to see which treatment patients were given and then examined their subsequent heart disease. It turns out one of the drugs was more likely to be prescribed to sicker patients. If we didn't account for this, we would end up with a biased result, it could look like that drug led to more heart disease when in reality this might have been driven by the patients pre-treatment characteristics.



How do we compensate for non-random treatment assignment?

WAKE FOREST UNIVERSITY SEMINAR

This brings me to our first question. How do we compensate for non-random treatment assignment?

Propensity score

$$e_i = P(Z_i = 1 | \mathbf{X})$$

WAKE FOREST UNIVERSITY SEMINAR

One way to do this is via propensity scores. The propensity score is denoted here as e_i . It is the probability that you received treatment given your observed pre-treatment covariates, here denoted as \mathbf{x} . A lot of times we fit this with a logistic regression

Propensity scores

Rosenbaum and Rubin showed in observational studies, conditioning on **propensity scores** can lead to unbiased estimates of the treatment effect

1. There are no unmeasured confounders
2. Every subject has a nonzero probability of receiving either treatment

WAKE FOREST UNIVERSITY SEMINAR

Propensity scores can give unbiased estimates of the treatment effect as long as [click] there are no unmeasured confounders - that is you account for all of the pre-treatment characteristics that are associated with both the exposure and the outcome and [click] Every subject has a non-zero probability of receiving either treatment. In other words, you don't want the propensity score to be exactly 0 or 1.

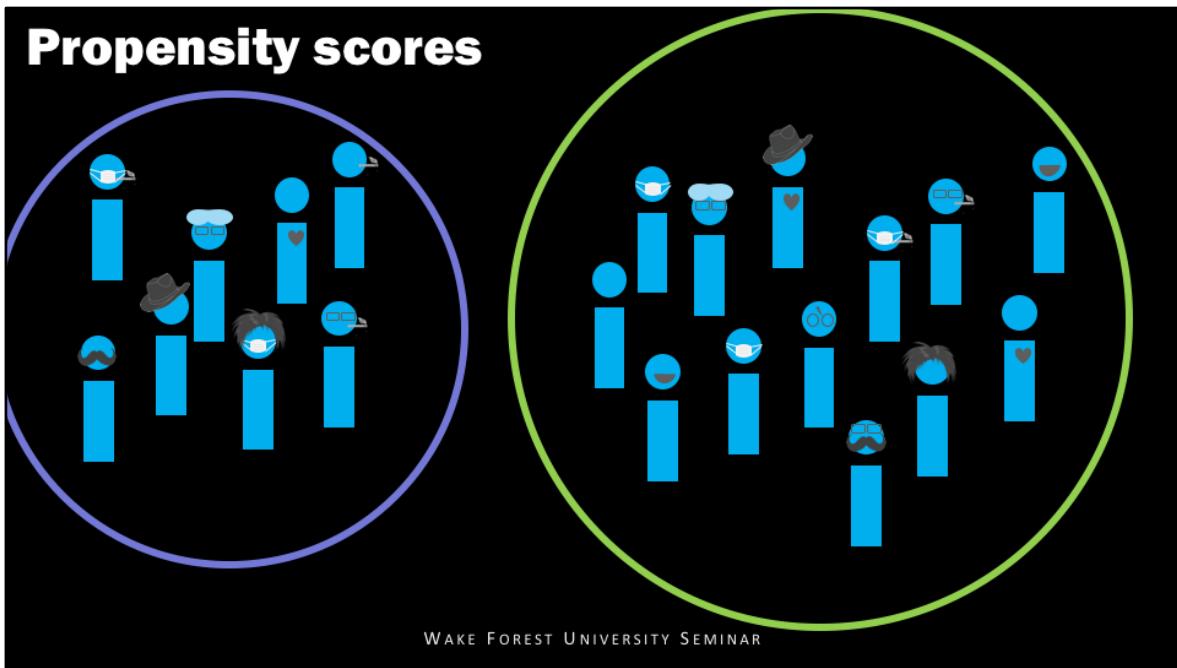
Propensity scores

Rosenbaum and Rubin showed in observational studies, conditioning on **propensity scores** can lead to unbiased estimates of the treatment effect

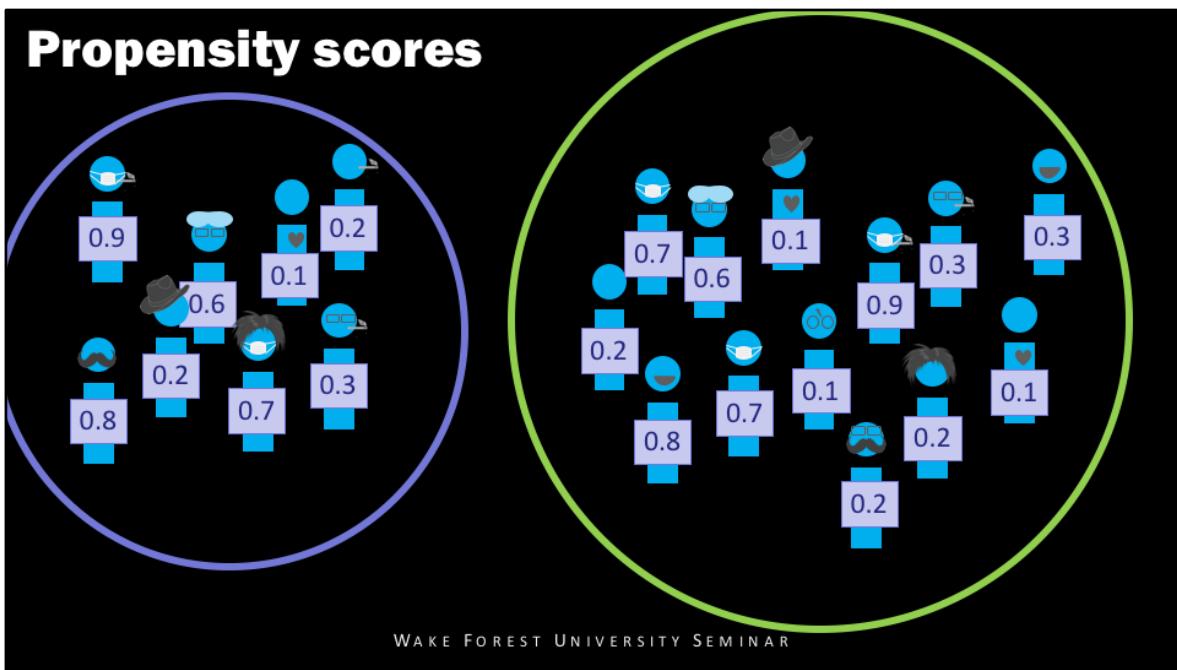
1. There are **no unmeasured confounders**
2. Every subject has a nonzero probability of receiving either treatment

WAKE FOREST UNIVERSITY SEMINAR

This first condition is essentially impossible to check, however sensitivity analysis can be conducted to measure the impact of an unmeasured confounder. The final section of this talk will address some research I have done in this area.



Alright so I've shown how that works mathematically, let's take a second to see it visually. Here I have two sets of participants. On the left, in my purple group I have participants that were given the treatment. On the right, my green group, I have participants given the control. Notice these participants all have various combinations of pre-treatment characteristics. For example, some are wearing cowboy hats, some smokers, some have healthy hearts, and some are mustachioed.



Using these pre-treatment characteristics, I can calculate the probability that these participants would have received treatment [click], and assign each their propensity score



How do we compensate for non-random treatment assignment?

propensity scores

WAKE FOREST UNIVERSITY SEMINAR

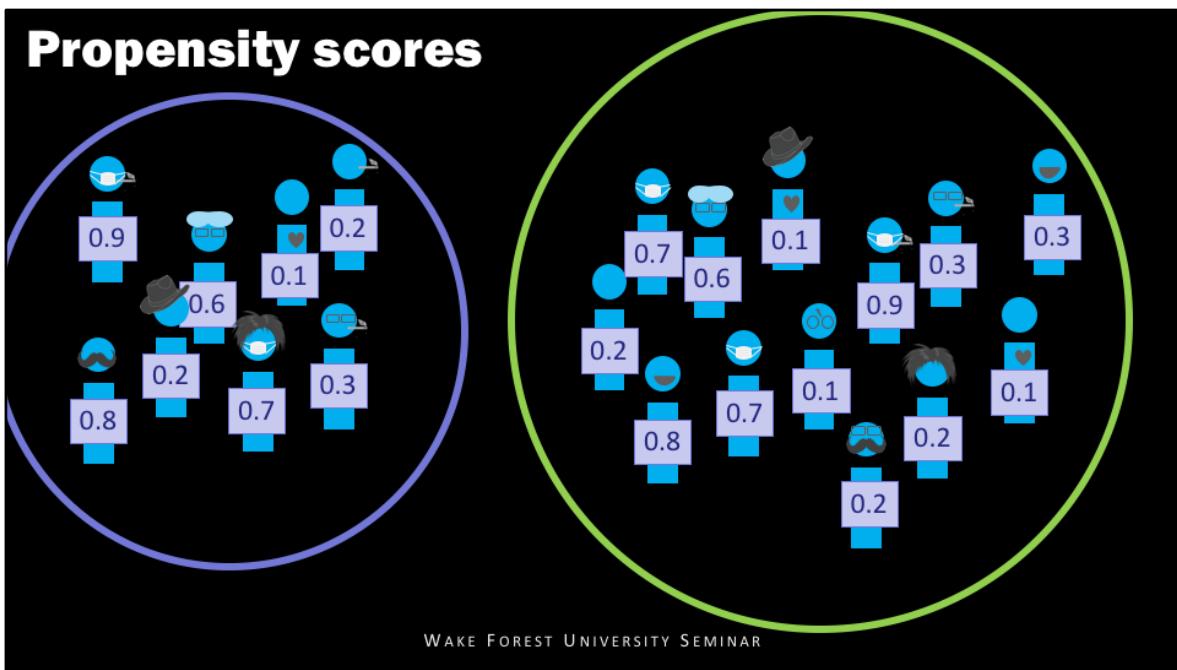
So that answers our first question, we compensate for non-random treatment assignment with propensity scores.

**How do we compensate for non-random treatment assignment?
*propensity scores***

How do we incorporate the propensity scores?

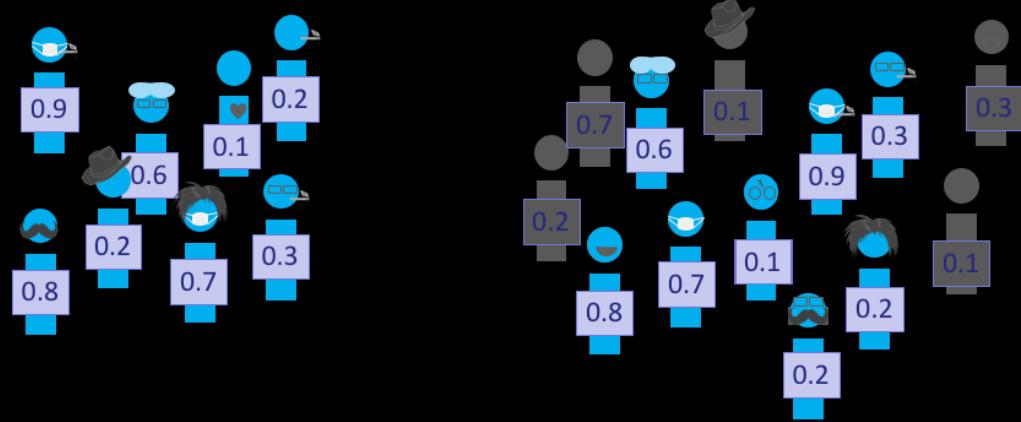
WAKE FOREST UNIVERSITY SEMINAR

Now that we've decided to use the propensity score, how do we incorporate it in our analysis?



There are lots of ways to incorporate propensity scores, but two that I prefer are matching and weighting. This is because both of these methods make it very straightforward to describe the characteristics of the study population

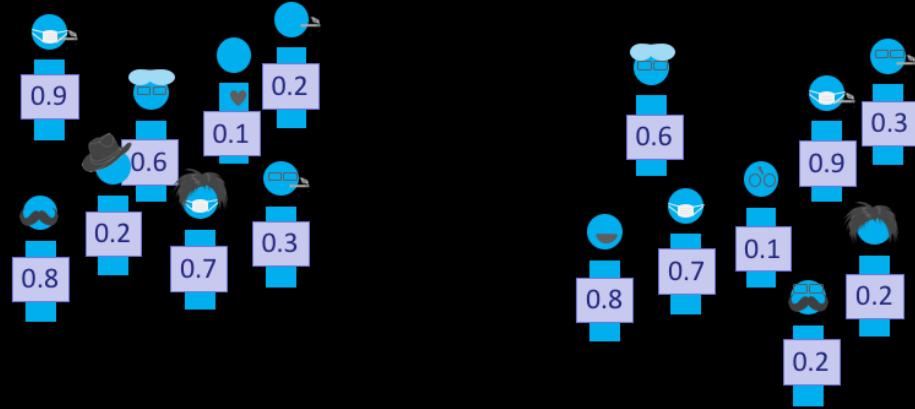
Propensity scores



WAKE FOREST UNIVERSITY SEMINAR

In matching, we often drop some patients that don't have a good match in the opposite group

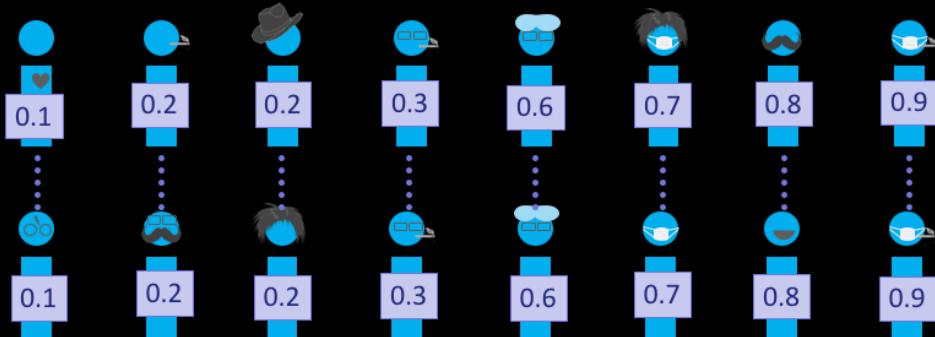
Propensity scores



WAKE FOREST UNIVERSITY SEMINAR

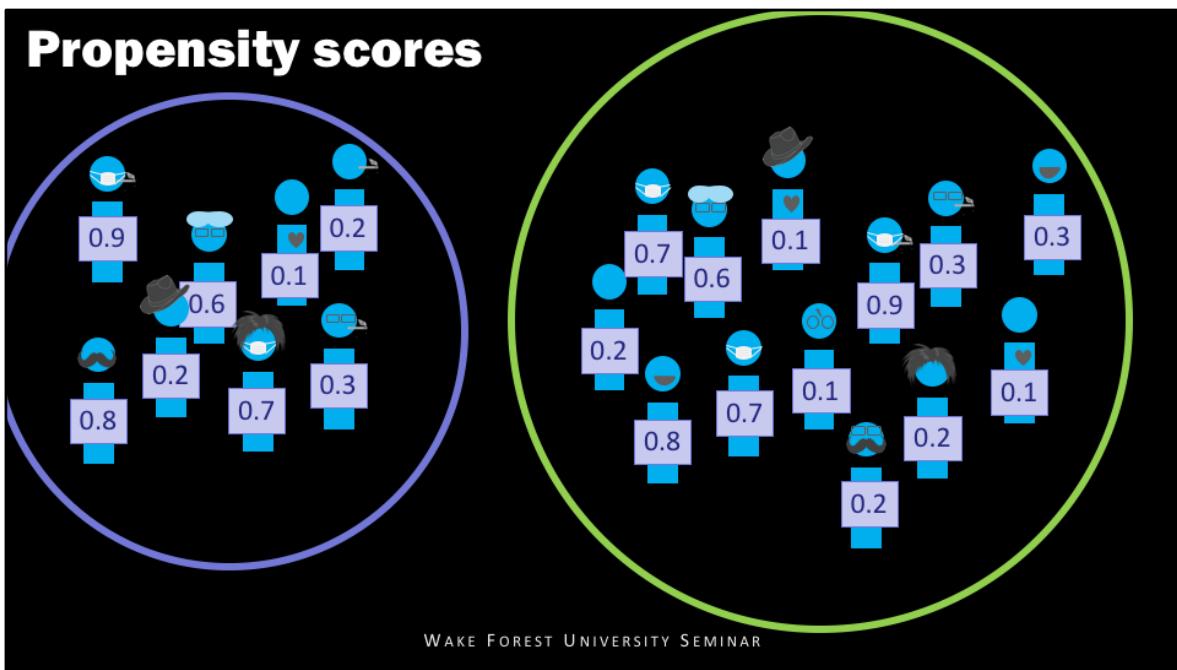
And then line up the remaining patients, here the top patients were assigned treatment, and the bottom control

Propensity scores



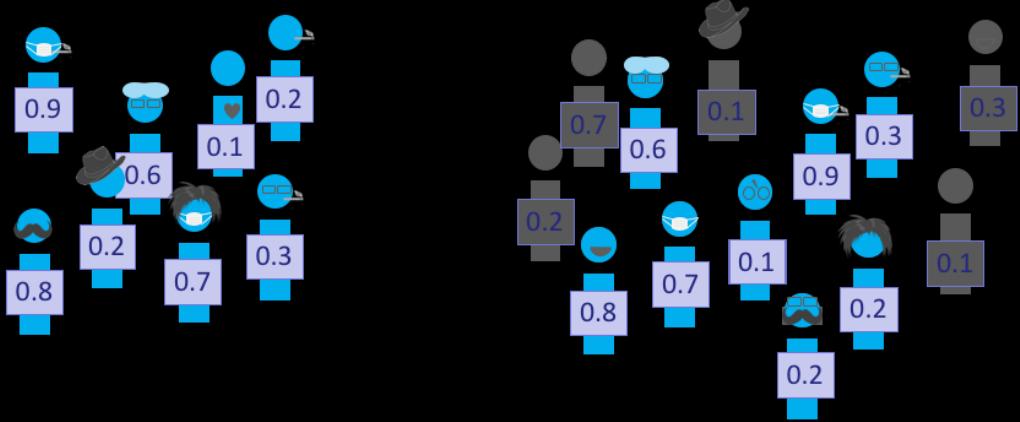
WAKE FOREST UNIVERSITY SEMINAR

We can now compare the outcome between the treatment group and the control group in the same way we would in a randomized trial.



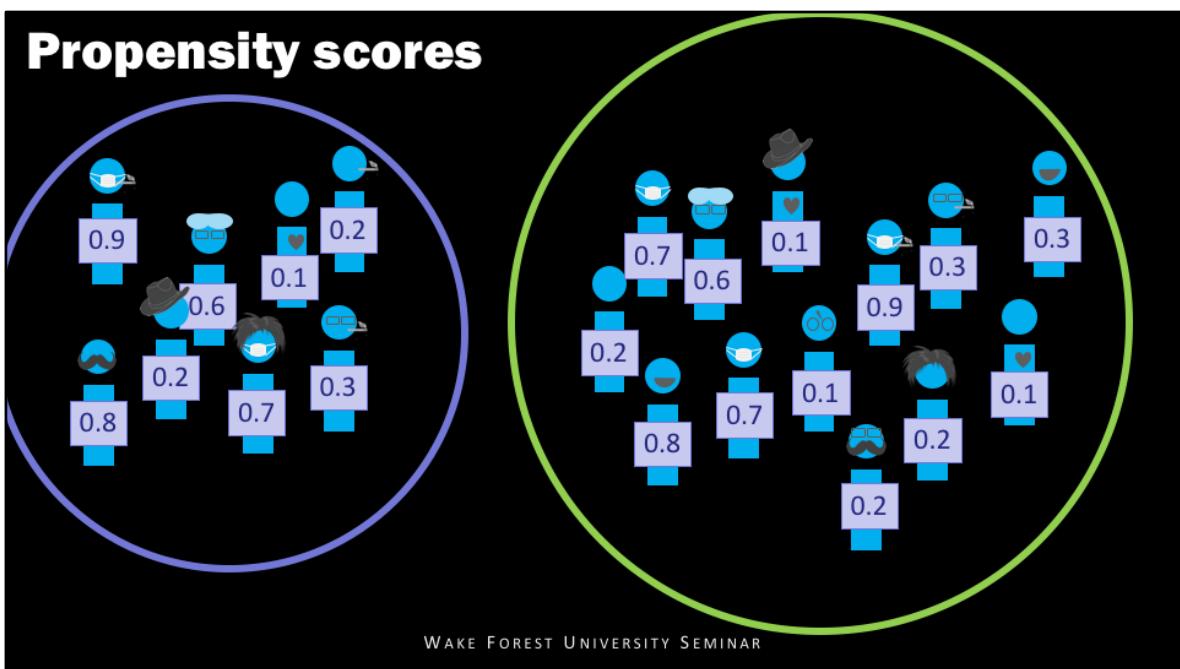
Instead of matching, We can create weights that are some function of the propensity score and the indicator of treatment.

Propensity scores



WAKE FOREST UNIVERSITY SEMINAR

Now instead of dropping participants that don't have a good match



We can [click] upweight and down weight participants so the two groups are comparable.

Propensity score

$$e_i = P(Z_i = 1 | \mathbf{X})$$

$$w_i = f(e_i, Z_i)$$

WAKE FOREST UNIVERSITY SEMINAR

Here is what that looks like mathematically. The weight here, w_i , is a function of the propensity score, e_i and the treatment, Z_i

Estimate treatment effect

$$\frac{\sum_{i=1}^n Y_i Z_i w_i}{\sum_{i=1}^n Z_i w_i} - \frac{\sum_{i=1}^n Y_i (1 - Z_i) w_i}{\sum_{i=1}^n (1 - Z_i) w_i}$$

WAKE FOREST UNIVERSITY SEMINAR

We can incorporate these weights to estimate the treatment effect.

Estimate treatment effect

$$\frac{\sum_{i=1}^n Y_i Z_i w_i}{\sum_{i=1}^n Z_i w_i} - \frac{\sum_{i=1}^n Y_i (1 - Z_i) w_i}{\sum_{i=1}^n (1 - Z_i) w_i}$$

WAKE FOREST UNIVERSITY SEMINAR

This is a consistent estimator if the propensity score is correct.

- How do we compensate for non-random treatment assignment?**
propensity scores
- How do we incorporate the propensity scores?**
weighting

WAKE FOREST UNIVERSITY SEMINAR

Alright, so we are using propensity scores and we've decided on weighting

10 min

- How do we compensate for non-random treatment assignment?
*propensity scores***
- How do we incorporate the propensity scores?
*weighting***
- How do we calculate these weights?**

WAKE FOREST UNIVERSITY SEMINAR

Now how do we actually calculate these weights?

Method

WAKE FOREST UNIVERSITY SEMINAR

This is where my research comes in. In particular, I've examined the finite sample bias and variance properties of these three weighting schemes.

Weighting

- **ATE** Average treatment effect
- **ATM** Average treatment effect among the evenly matchable
- **ATO** Average treatment effect among the overlap population

WAKE FOREST UNIVERSITY SEMINAR

There are many different potential weighting schemes based on the target population, or causal estimand, you are interested in. For this talk, we are going to talk about three, a classic weighting scheme for the Average treatment effect and two newer weighting schemes the Average treatment effect among the evenly matchable and the average treatment effect among the overlap population. I've put these along with their definitions on your handout, so you can refer to that if you forget, or feel free to stop me if an acronym is unclear.

Average treatment effect

$$w_{ATE} = \frac{Z_i}{e_i} + \frac{1 - Z_i}{1 - e_i}$$

WAKE FOREST UNIVERSITY SEMINAR

The first weight we are going to discuss allows you to estimate the average treatment effect. The target population here is the whole sample population, both treated and controlled. If we were to look at a table of pre-treatment characteristics of the patients before and after weighting, post weighting the characteristics would be balanced across treatment groups and would match that of the overall population. For example, if the average age overall in the cohort was 50, with the average age pre-weighting as 40 in the control group and 60 in the treatment group, post weighting the average age in the treatment group and control group would both match the overall unweighted average of 50.

While this is often declared as the population of interest, it is not always the medically or scientifically appropriate population. This is because estimating the ATE assumes that every participant can be switched from their current treatment to the opposite, which doesn't always make sense. For example, it may not be medically appropriate for every participant who didn't receive a treatment to receive it. I'm going to briefly step through this equation to give some intuition about what this weight is doing, and then show the pseudo population created when using this weight

Average treatment effect

$$w_{ATE} = \frac{Z_i}{e_i} + \frac{1 - Z_i}{1 - e_i}$$

WAKE FOREST UNIVERSITY SEMINAR

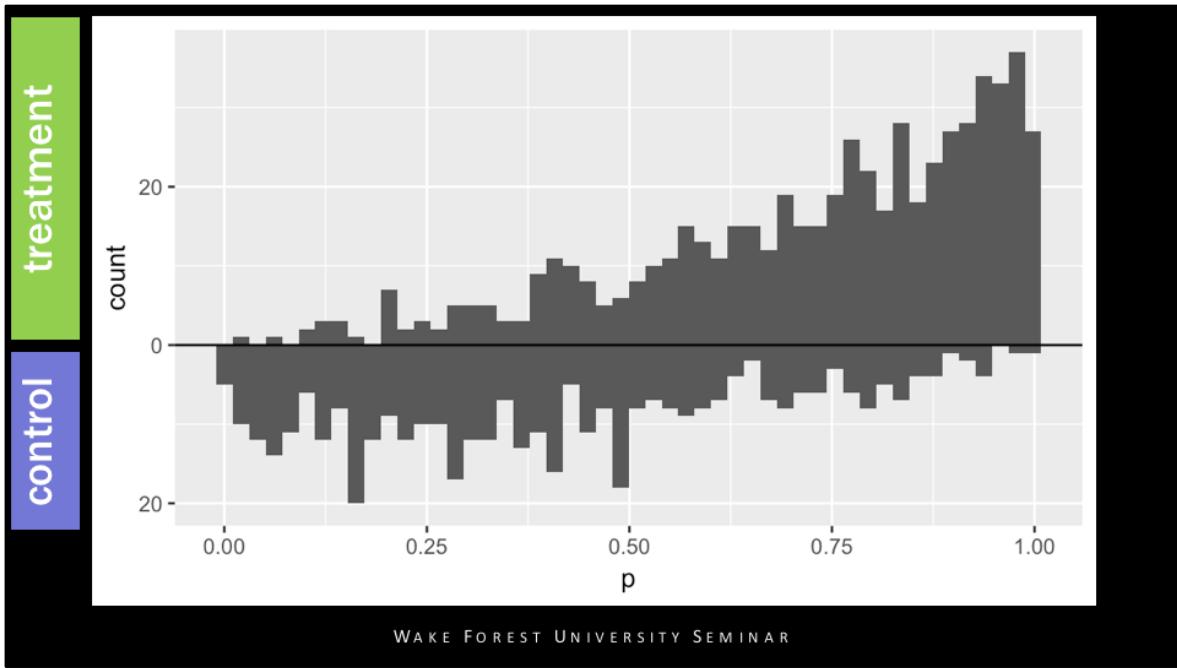
participants who receive treatment will just have this part of the equation, since Zi is 1, making the second part 0. For these participants there weight is just 1 over their propensity score

Average treatment effect

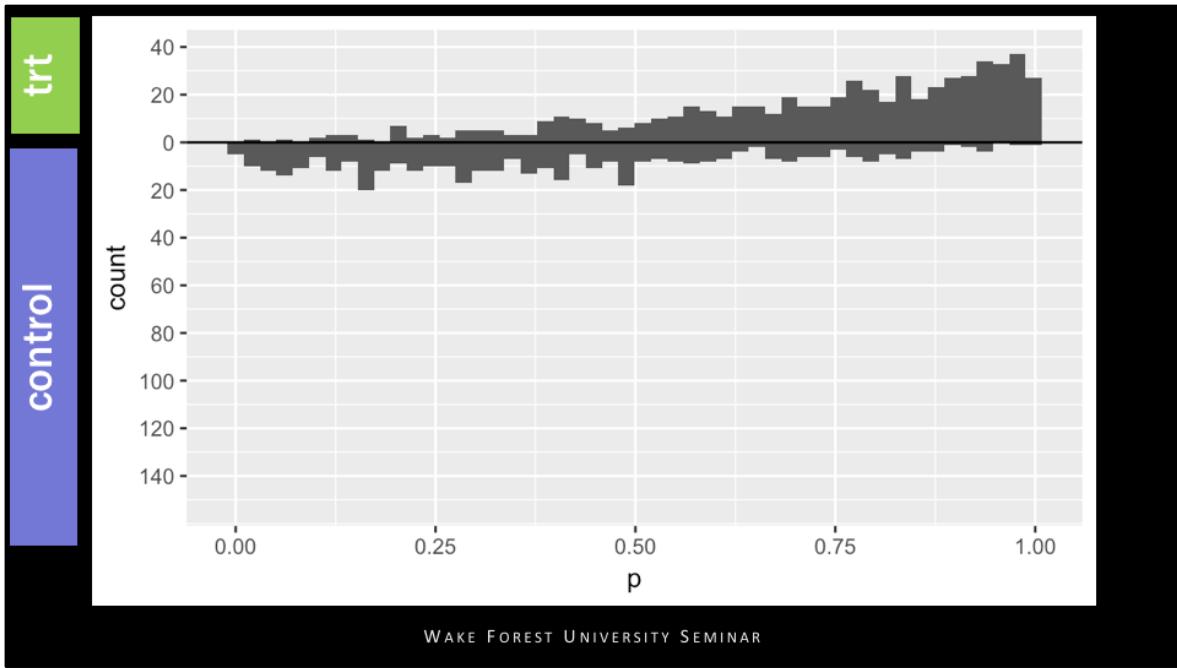
$$w_{ATE} = \frac{Z_i}{e_i} + \frac{1 - Z_i}{1 - e_i}$$

WAKE FOREST UNIVERSITY SEMINAR

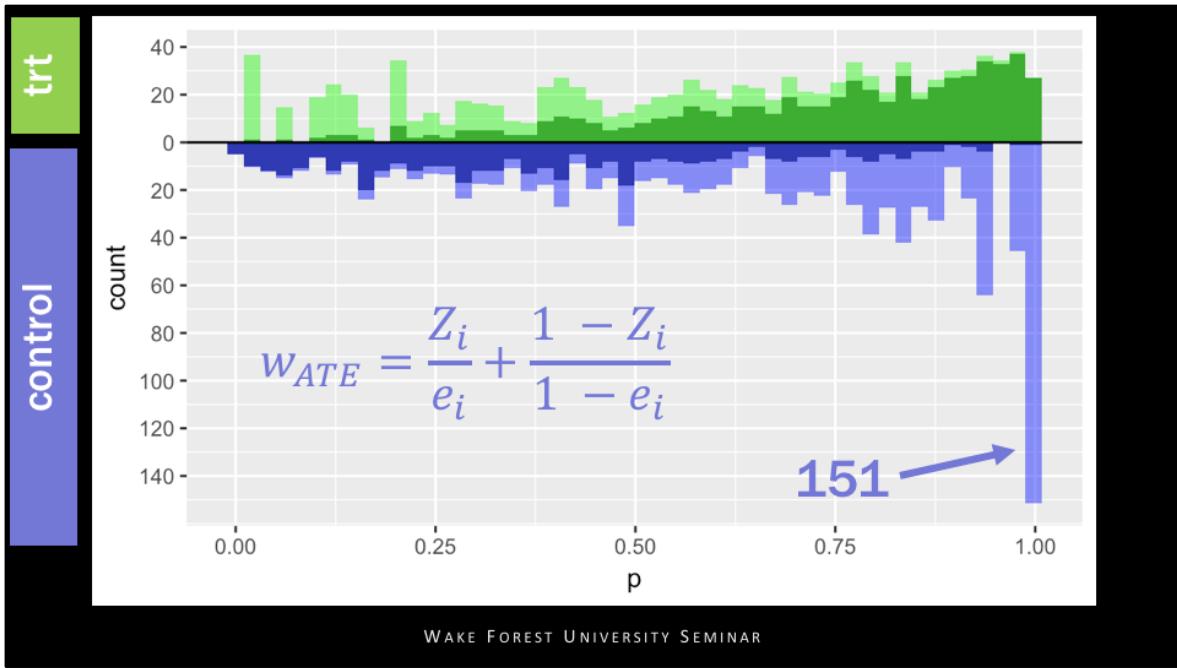
the participants in the control group, will just have this part of the equation. Their weight will just be 1 over 1 minus the propensity score. So these ATE weights overall are just 1 over the probability of receiving the treatment that you received. An important thing to notice here is these weights range from 0 to infinity, so in theory they can be very very large



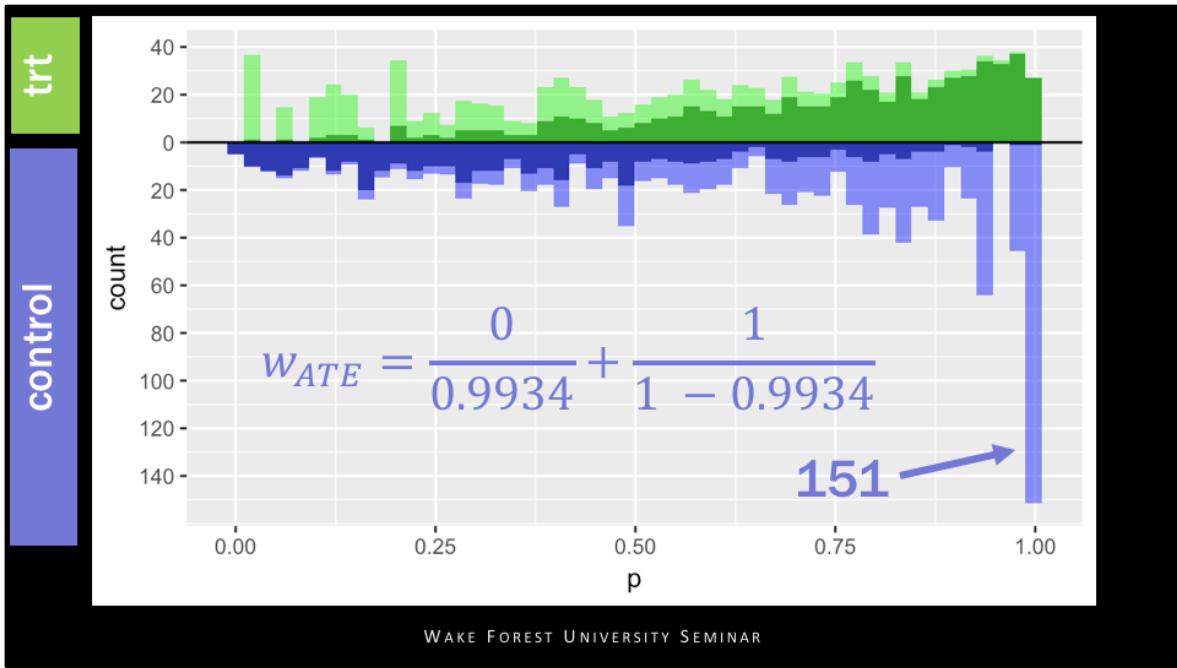
Here, I have a mirrored histogram of the actual observed sample. On the top, above 0, I have the histogram of propensity scores among the treated. Notice here that the graph is skewed such that most of the mass is on the higher end. This makes sense since the propensity score is the probability of receiving treatment, we'd expect those who DID receive treatment to have often been more likely to do so. Now let's look at the other side, the mirrored histogram below 0. This is the histogram of propensity scores for the control group. Here, there is more mass towards 0, since these participants were more likely to NOT receive the treatment. Also notice that it looks like we have more participants in the treatment group than the control group. The ATE weights are trying to make these two populations look similar. We would expect that we will need to upweight some of these control patients on the higher side of the propensity scores, and upweight some of these treated participants on the low side of the propensity scores.



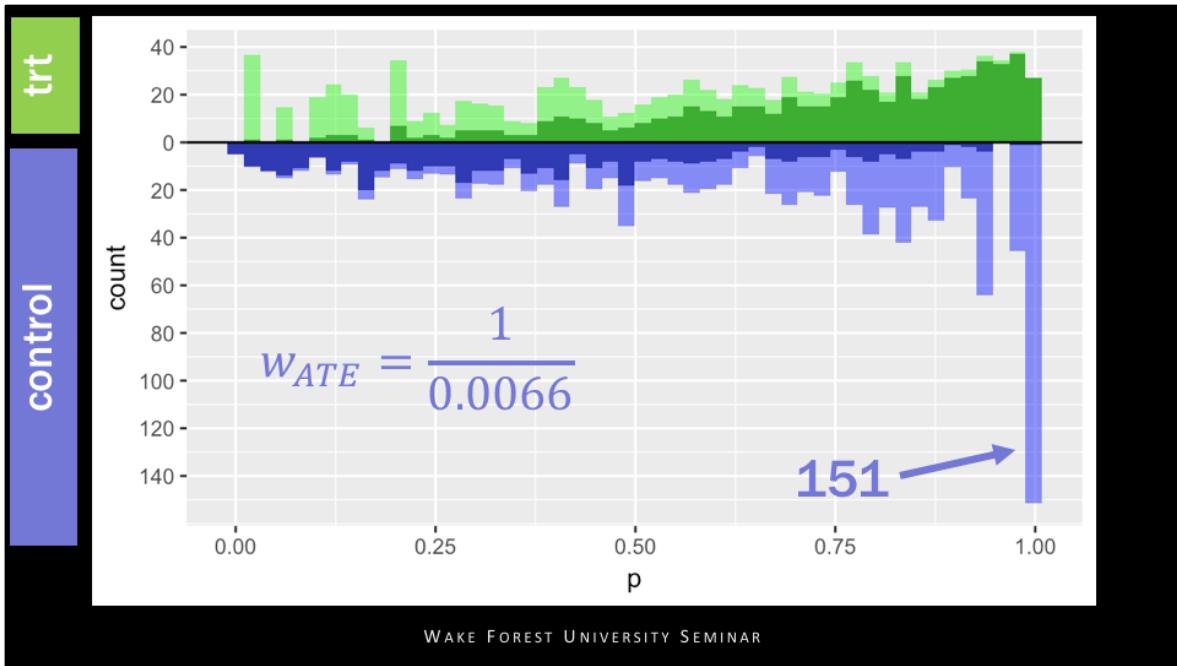
This is the same histogram as the previous slide, I've just shrunk it a bit to extend the y-axis so we can see the pseudo-population created from the ATE weights



The dark portion, the grey, are the distribution of propensity scores and again the top part of the mirrored histogram is the treated population and the bottom, the upside-down histogram is the control population. Then the green represents the weighted population for the treated and the blue is the weighted population for the controls. In other words, the green and blue represent the pseudo population created by using the weights (Frequency counts for a pseudo population based on the weights) [click]. This guy at the far end has a propensity score greater than .99, and therefore ends up accounting for 151 people in the pseudo cohort.



If we plug in 0 for Z_i , since this participant is in the control group, and 0.9934, their propensity score, for e_i



We end up with 1 over 0.0066 or 151. Intuitively, it can probably cause some problems if we are allowing single participants to represent a large number of participants in a pseudo cohort.

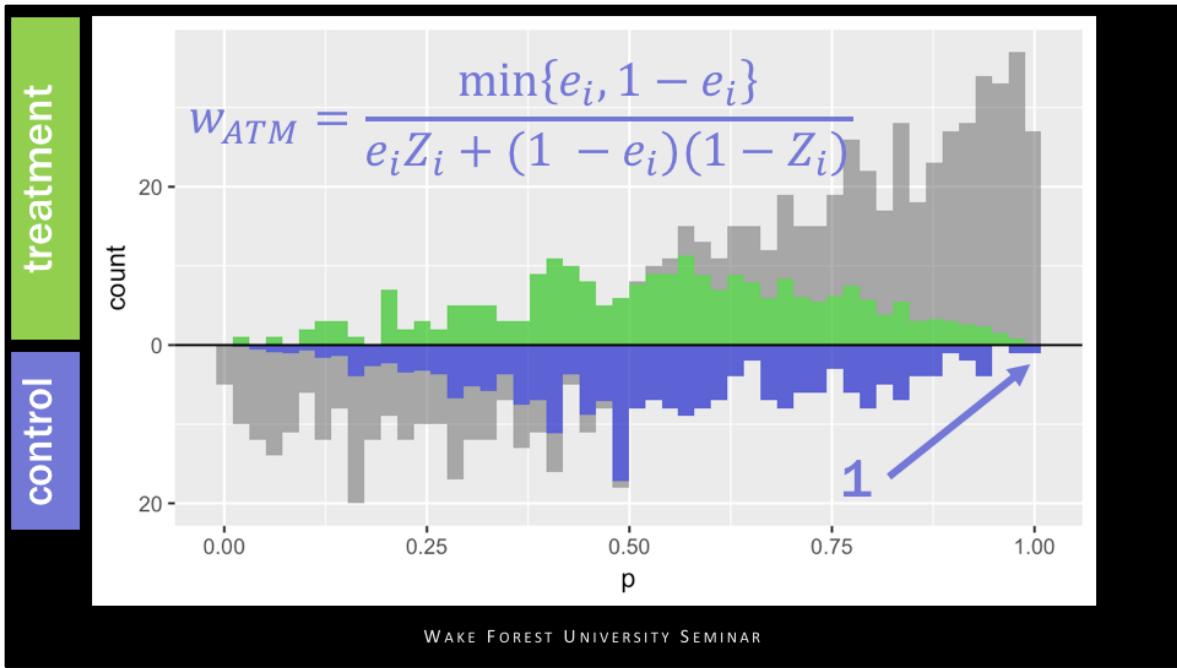
Average treatment effect among the evenly matchable

$$W_{ATM} = \frac{\min\{e_i, 1 - e_i\}}{e_i Z_i + (1 - e_i)(1 - Z_i)}$$

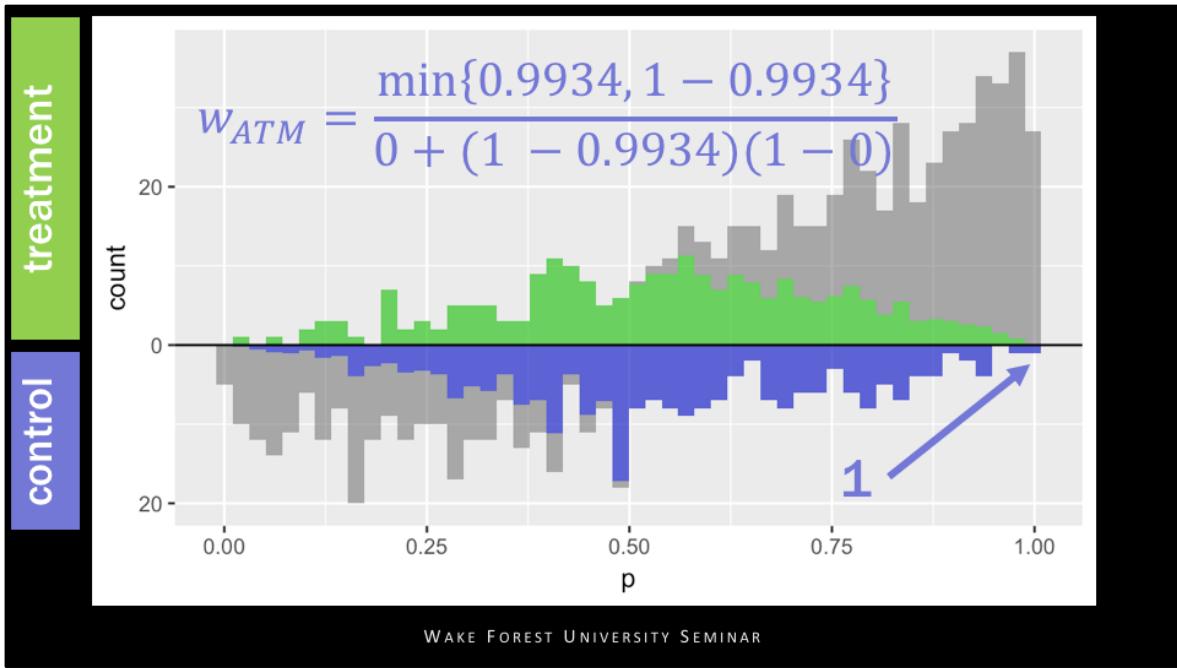
Li & Greene (2013)

WAKE FOREST UNIVERSITY SEMINAR

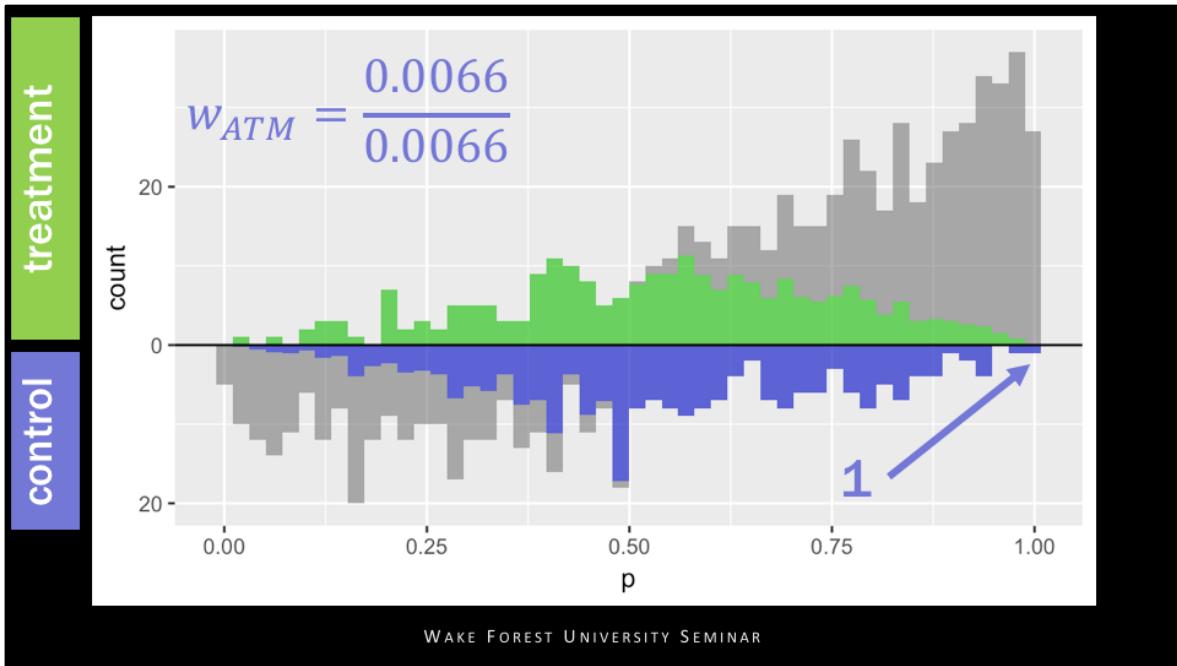
A newer weighting scheme introduced by Li and Greene in 2013 are matching weights. These estimate the average treatment effect among the evenly matchable. The pseudo-cohort they create is akin to the cohort that would be created via 1:1 pair matching. Here the weight is the minimum of the propensity score and 1 minus the propensity score over the propensity score for the treated and 1 minus the propensity score for the controls. Notably, this weight is more stable than the ATE weights in that it is bounded by 0 and 1.



[click] notice now that rather than representing 151, this participant now just has a weight of 1, and the 20 treated participants above it each have a weight of 1/20



plugging in that propensity score and 0 for Z_i



we end up with 0.0066 over 0.0066, or 1.

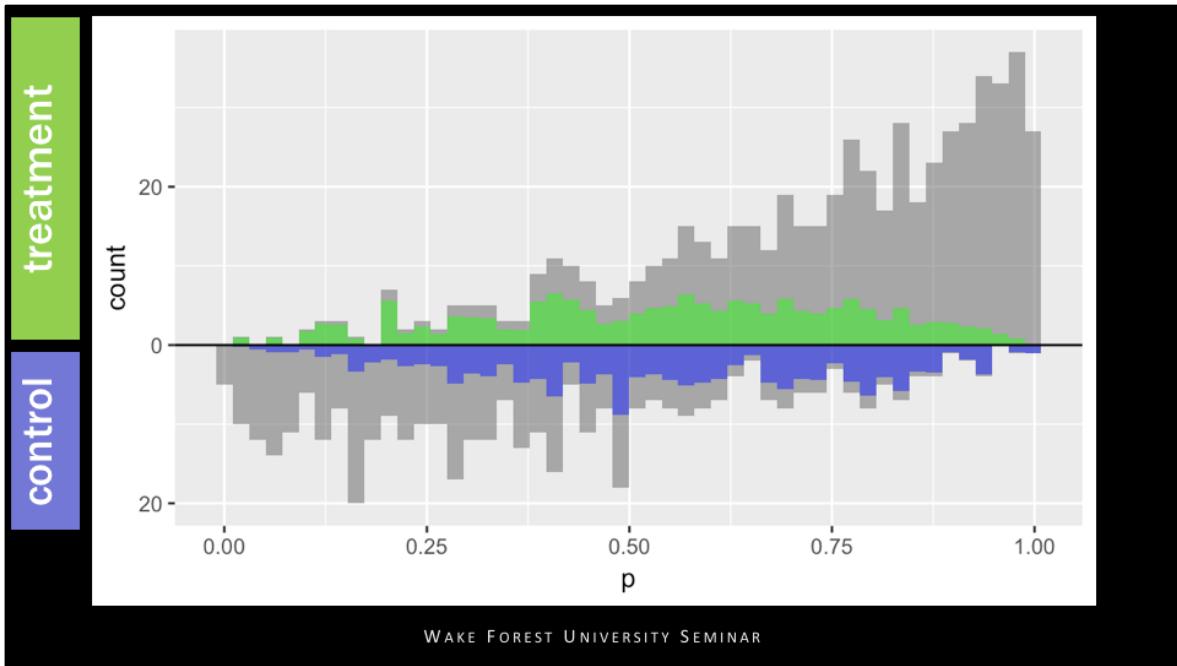
Average treatment effect among the overlap population

$$w_{ATO} = (1 - e_i)Z_i + e_i(1 - Z_i)$$

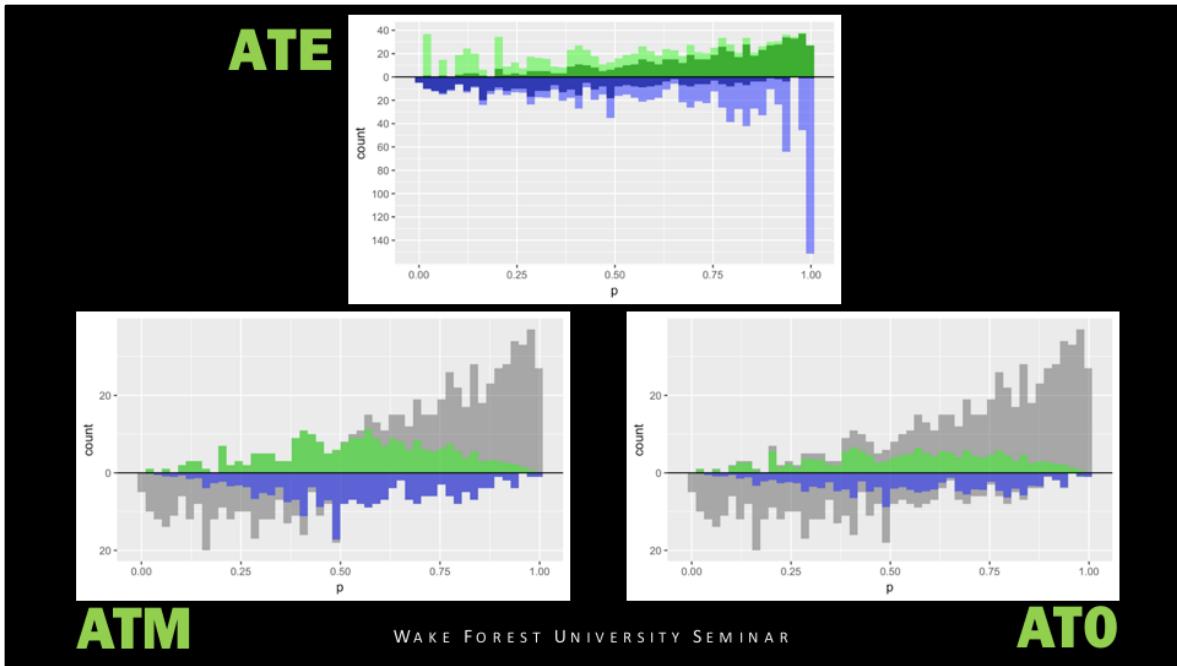
Li et al (2016)

WAKE FOREST UNIVERSITY SEMINAR

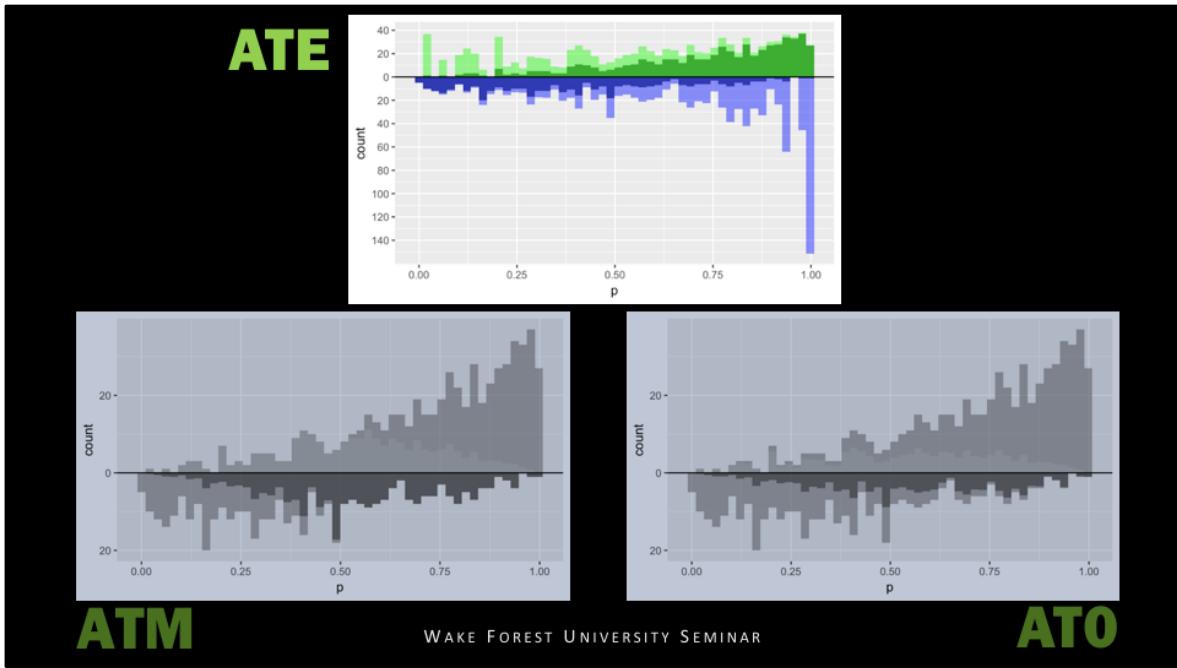
Finally, the overlap weights derived by Li et al in 2016 estimate the average treatment effect among the overlap population. Here, we are constructing a pseudo-population that will have the most overlap in the covariates between treatment groups. In practice, this creates a cohort often very similar to the matching weights, with some improved variance properties. Here the weight is 1 minus the propensity score for the treated and the propensity score for the controls. So again, this is going to be bounded by 0 and 1



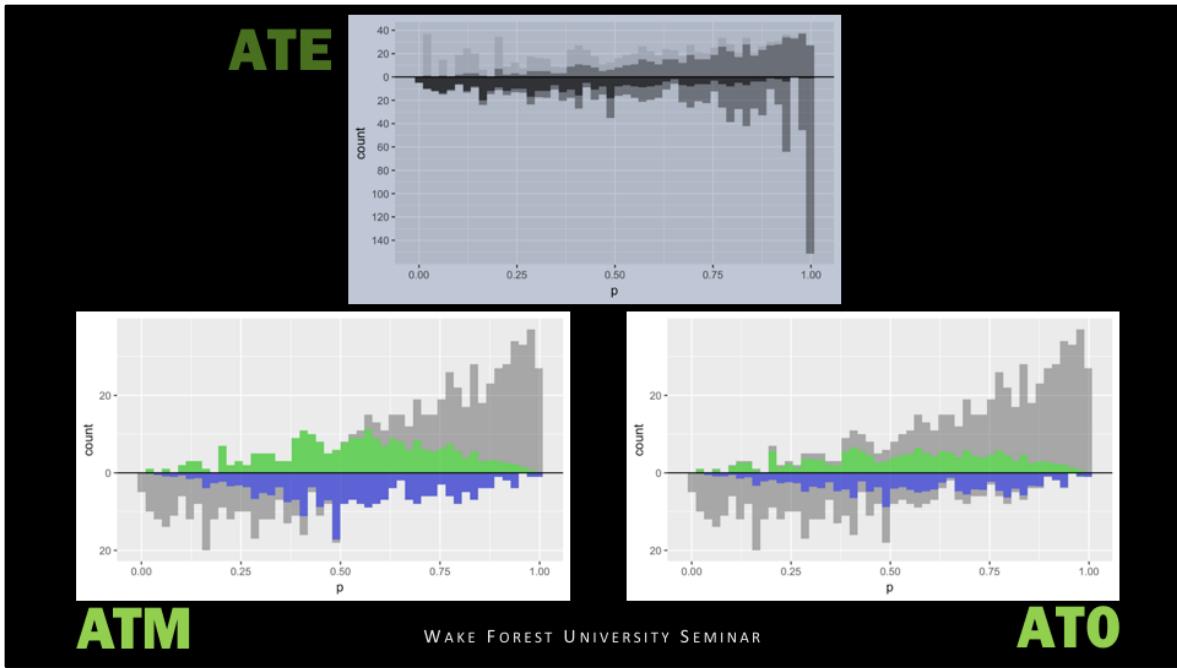
Here is what that pseudo population looks like. The population will look a lot like a Table 1 from a 1:1 matched trial, but shifted a bit to improve the variance.



Here are the three pseudo-populations side by side so you can get a better idea for the difference



With the weights for estimating the average treatment effect ranging from 0 to infinity



And these newer, more stable weights, ranging from 0 to 1.

- How do we compensate for non-random treatment assignment?**
propensity scores
- How do we incorporate the propensity scores?**
weighting
- How do we calculate these weights?**

WAKE FOREST UNIVERSITY SEMINAR

Alright so now we have a bit of a better understanding for how to calculate these weights,

- How do we compensate for non-random treatment assignment?**
propensity scores
- How do we incorporate the propensity scores?**
weighting
- How do we calculate these weights?**
- Which weights should we choose?**

WAKE FOREST UNIVERSITY SEMINAR

How do we pick which ones to use?

Weighting Regressions by Propensity Scores

David A. Freedman
University of California, Berkeley
Richard A. Berk
University of Pennsylvania

Evaluation
Volume 32 Number 4
August 2008 392-409
© 2008 Sage Publications
10.1177/0193841X08317586
<http://ers.sagepub.com>
hosted at
<http://online.sagepub.com>

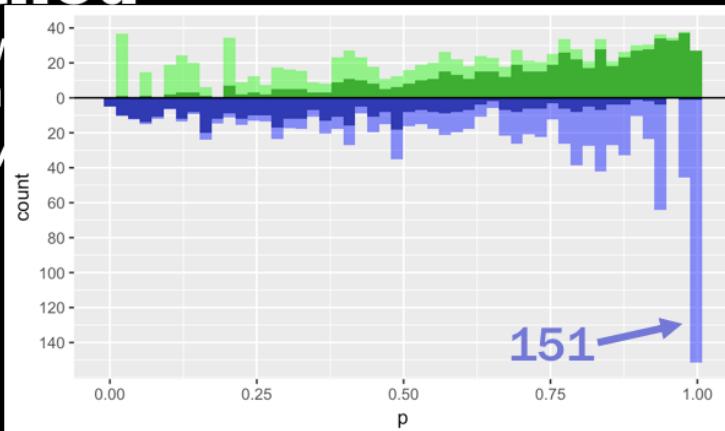
Regressions can be weighted by propensity scores in order to reduce bias. However, weighting is likely to increase random error in the estimates, and to bias the estimated standard errors downward, even when selection mechanisms are well understood. Moreover, in some cases, weighting will increase the bias in estimated causal parameters. If investigators have a good causal model, it seems better just to fit the model without weights. If the causal model is improperly specified, there can be significant problems in retrieving the situation by weighting, although weighting may help under some circumstances.

Keywords: causation; selection; models; experiments; observational studies;
regression; propensity scores

Estimating causal effects is often the key to evaluating social programs, but the interventions of interest are seldom assigned at random. Observational data are therefore frequently encountered. In order to

Propensity score weighting method

- The weights are estimated
- ATE weights



WAKE FOREST UNIVERSITY SEMINAR

These are known to be unstable because the weights can range from 0 to infinity

Propensity score weighting method

- The weights used in Freedman & Berk (2008) are estimating ATE
- **ATE weights can explode**
- **Hypothesize** that ATO and ATM weights will perform better

WAKE FOREST UNIVERSITY SEMINAR

These are known to be unstable because the weights can range from 0 to infinity

Fit a misspecified model

- Allow the propensity score to be **correct**
- The outcome model is **missing a covariate**

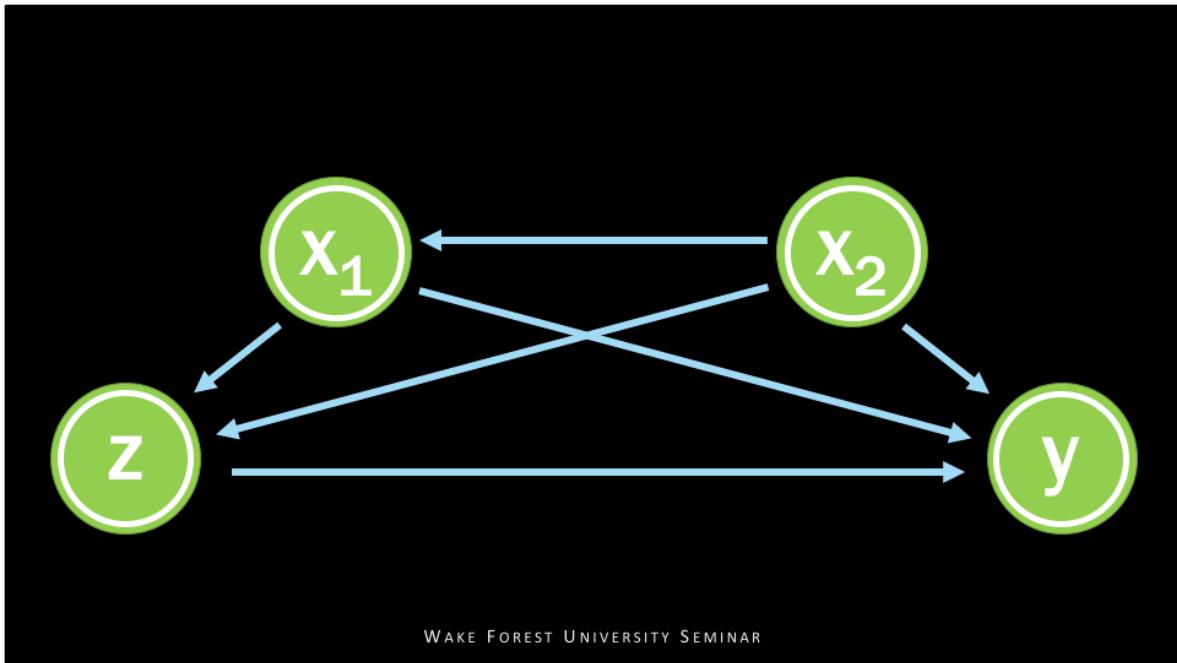
WAKE FOREST UNIVERSITY SEMINAR

They are intentionally creating a setup that is seemingly friendly to weighting to show that it fails even when it should have been working great

Freedman & Berk Simulation

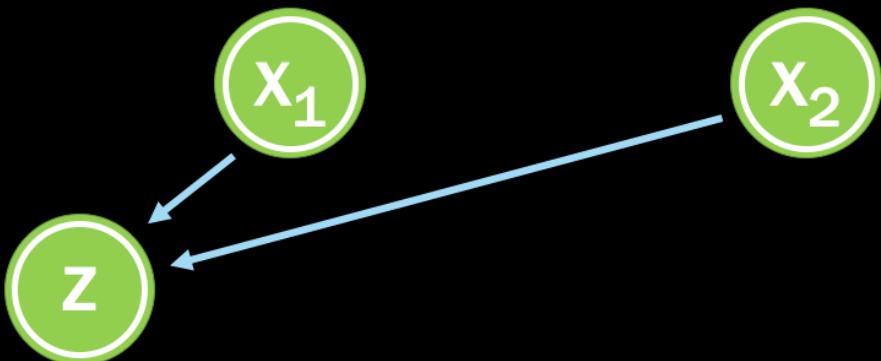
WAKE FOREST UNIVERSITY SEMINAR

Freedman and Berk demonstrated that propensity score weighting introduced bias and large standard errors. In their simulation, they fixed their sample size at 1000. To get a handle on this bias, we examined sample sizes from 100 to 10,000. We know that the propensity score was modeled correctly, so we would expect the estimate to be consistent, and therefore the bias to reduce as the sample size increases, but the rate at which it is decreasing is definitely of interest. We also introduced the two new weighting schemes to see how they fared under these simulated conditions.



In this simulation, we have two pre-treatment covariates, x_1 and x_2 . Both are confounders, in other words associated with the exposure, z as well as the outcome y . z , the exposure or treatment variable, is associated with y the outcome variable

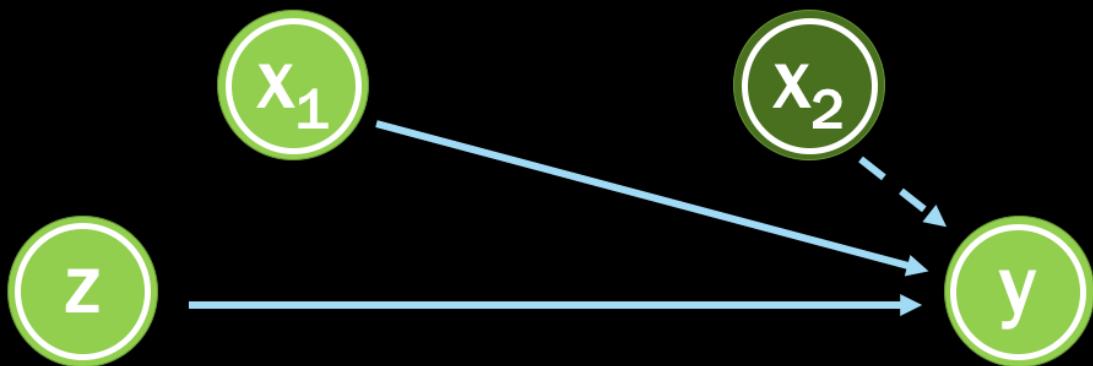
Propensity score model



WAKE FOREST UNIVERSITY SEMINAR

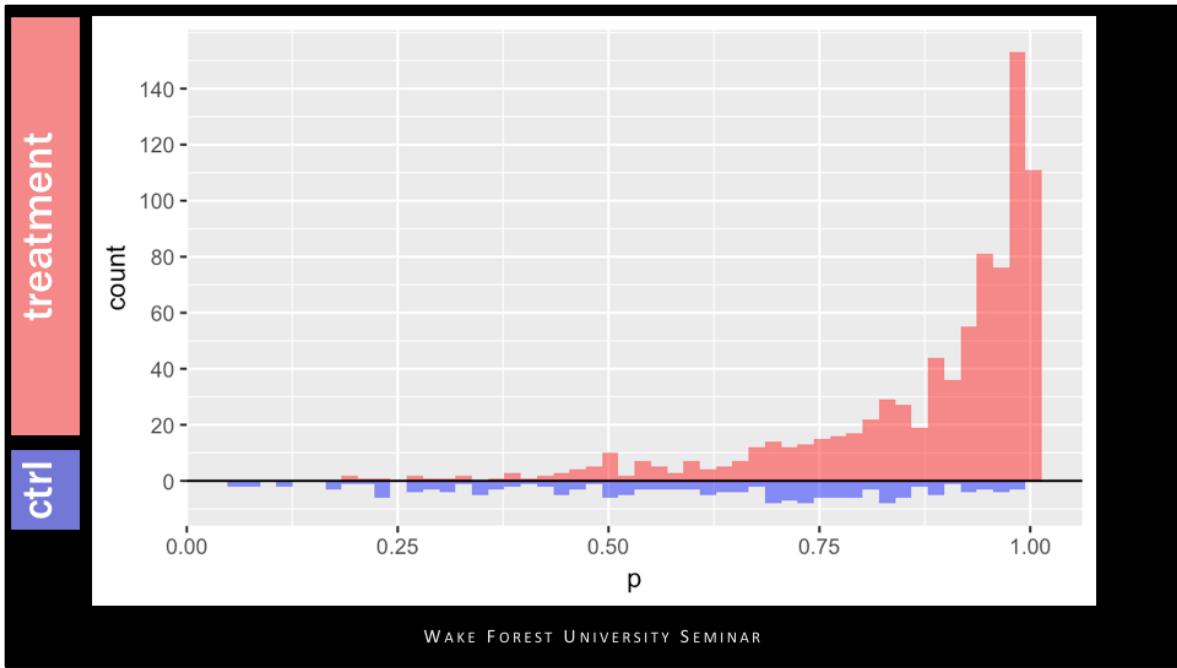
So we fit a propensity score model that is correctly specified, that is has the correct pre-treatment covariates. Because it is correctly specified, we know our treatment effect will be consistent, but it may be biased in finite samples.

Outcome model

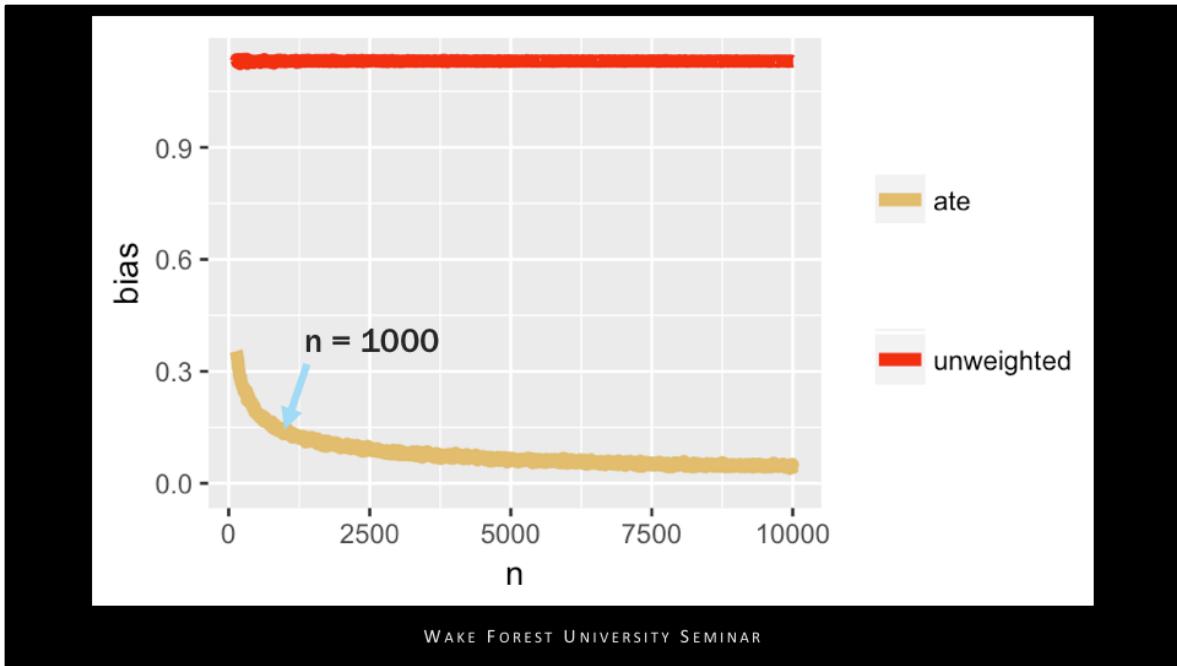


WAKE FOREST UNIVERSITY SEMINAR

The outcome model is missing one covariate, x2.

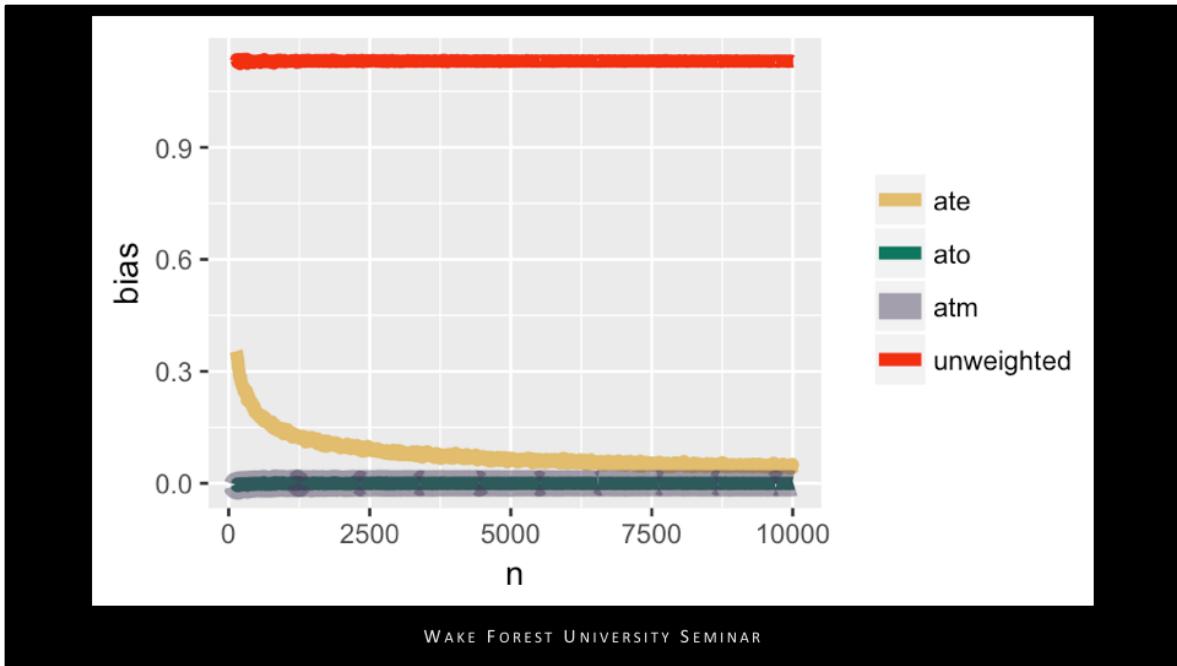


Here is a mirrored histogram of this population. Just looking at this, we can see that the ATE weights are going to have some trouble since there are so many more treated than control patients, especially towards the higher end.
TODO weighted hist

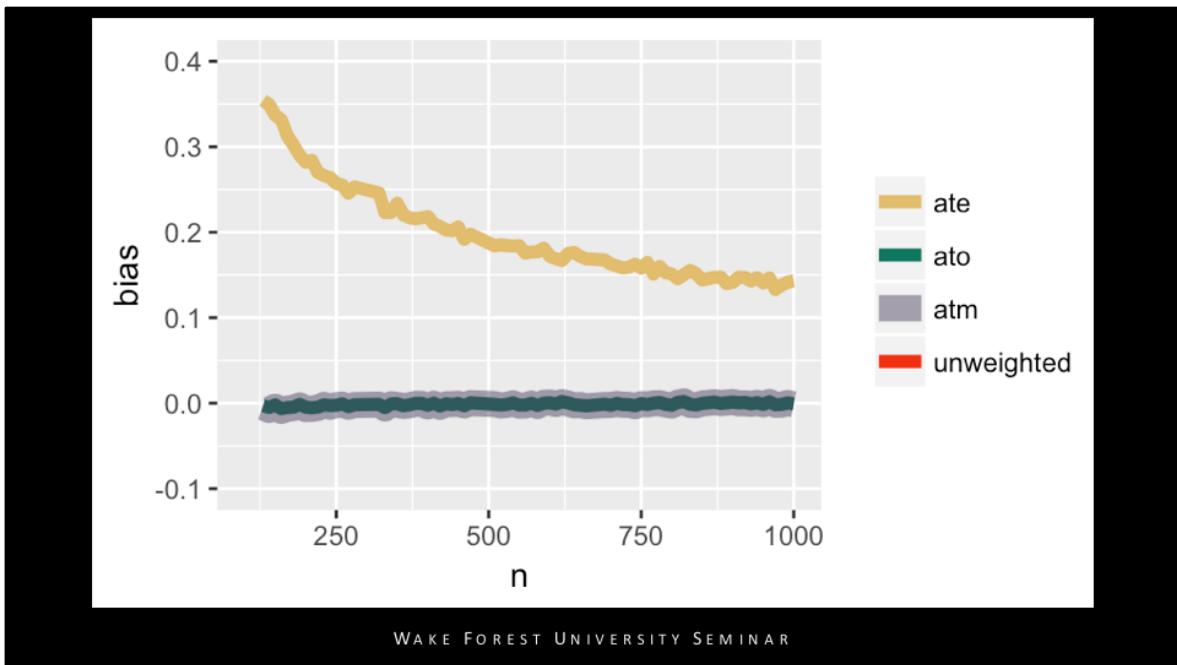


Let me take a few minutes to explain this graph.

At $n = 1000$, Freedman and Berk report a bias of 0.13 and we replicate this. Even when n is 10,000, the result is still biased. So they of course were correct, these weights do lead to bias in finite samples.



However it turns out to be less of a story about weighting in general, and more of a story about these particular weights.

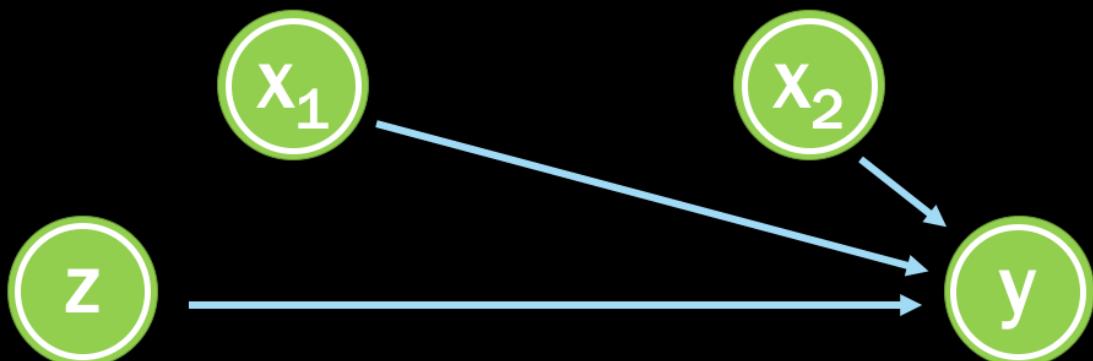


delete

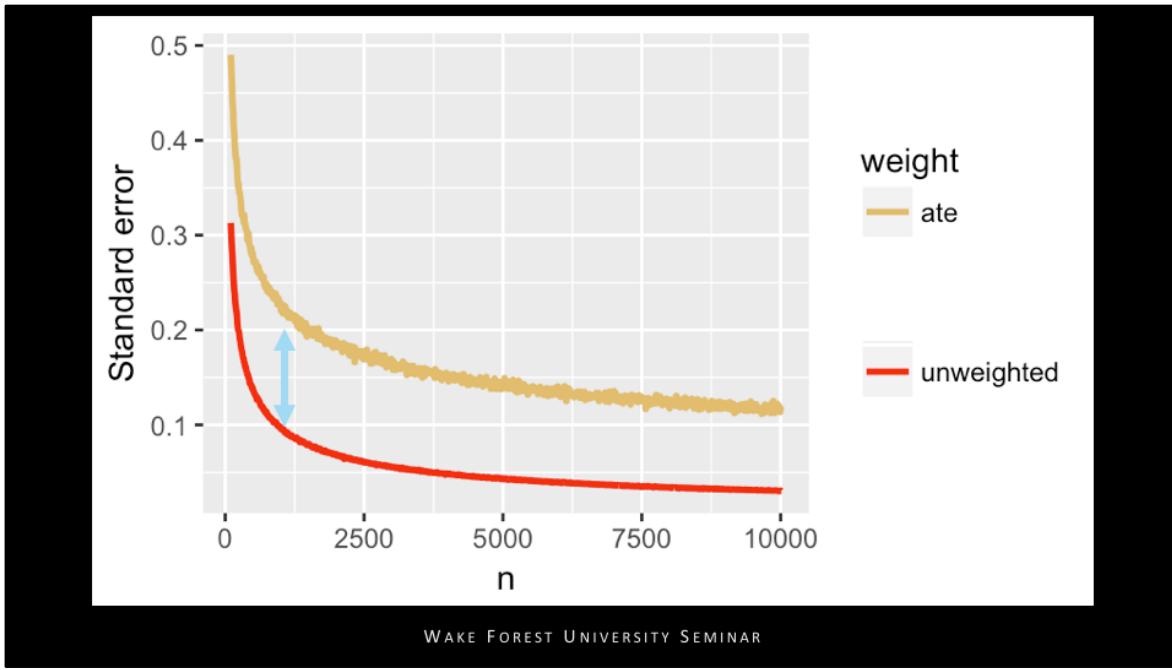
Standard errors

WAKE FOREST UNIVERSITY SEMINAR

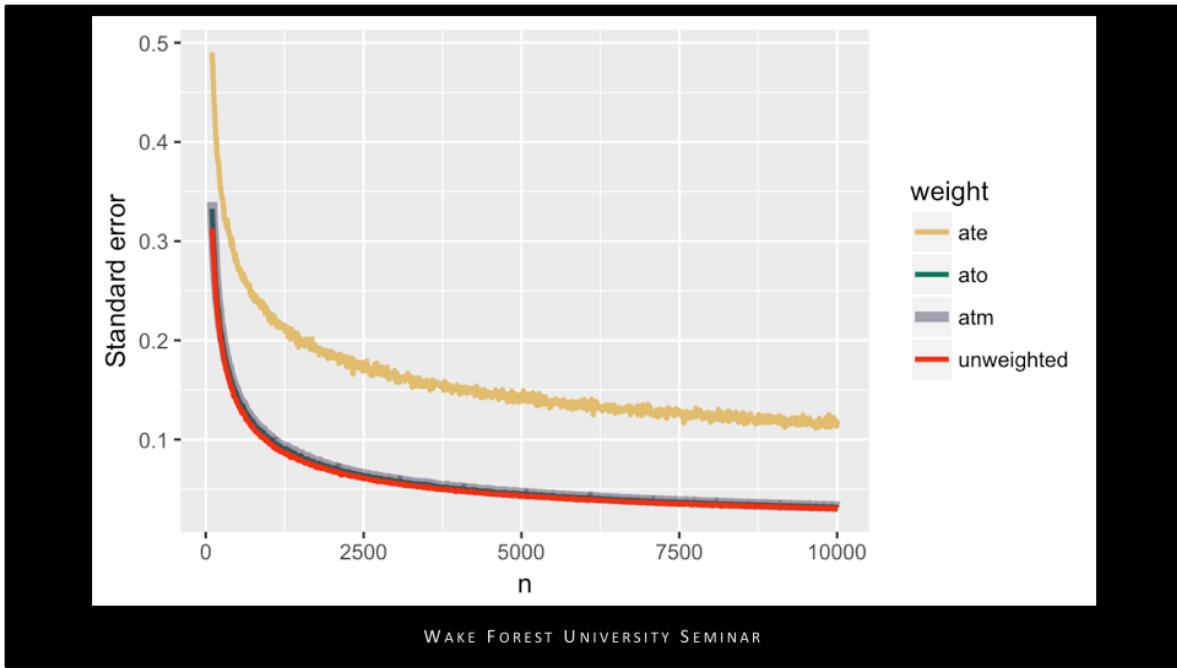
Outcome model



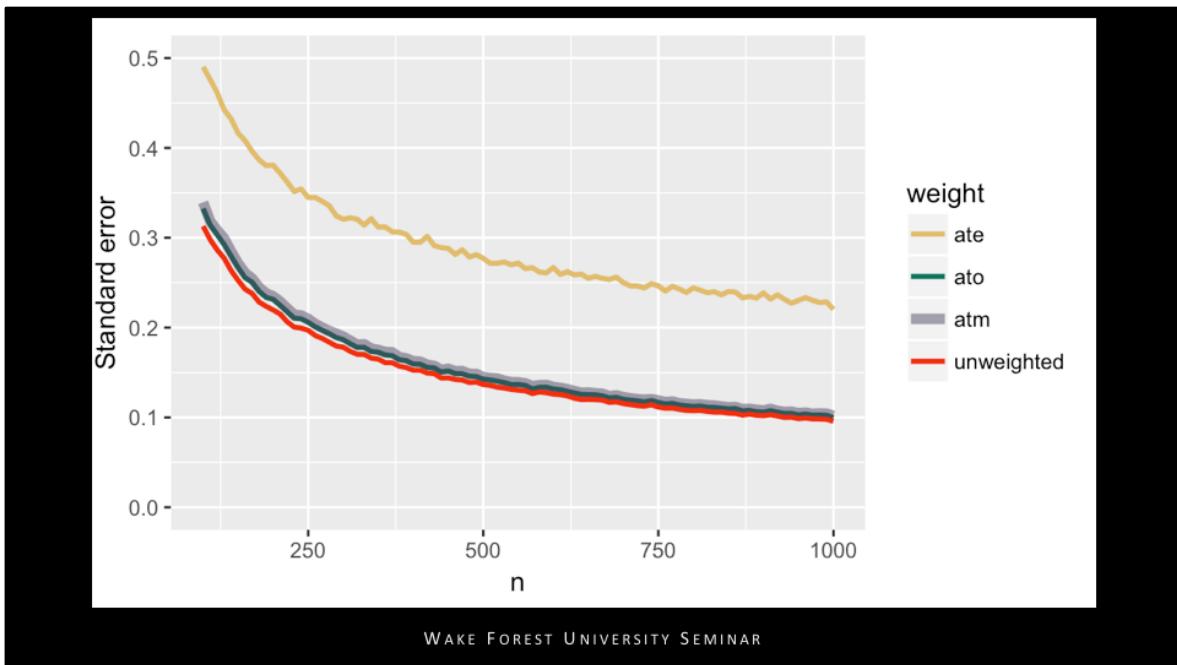
WAKE FOREST UNIVERSITY SEMINAR



We replicate what Freedman and Berk show, that this difference in the true standard error of these estimates is 0.12.



Correct-correct
Minimum variance unbiased estimator



- ✓ How do we compensate for non-random treatment assignment?
propensity scores
- ✓ How do we incorporate the propensity scores?
weighting
- ✓ How do we calculate these weights?
- ✓ Which weights should we choose?
ATM and ATO weights had improved finite-sample properties

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?**
propensity scores
- How do we incorporate the propensity scores?**
weighting
- How do we calculate these weights?**
- Which weights should we choose?**
The variance for the ATO and ATM is preferable to that of the ATE

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
*propensity scores***
- How do we incorporate the propensity scores?
*weighting***
- How do we calculate these weights?**
- Which weights should we choose?
*The variance for the ATO and ATM is preferable to that of the ATE***

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
propensity scores
- How do we incorporate the propensity scores?
weighting
- How do we calculate these weights?
- Which weights should we choose?
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?**

WAKE FOREST UNIVERSITY SEMINAR

Estimation

WAKE FOREST UNIVERSITY SEMINAR

28 min

**We want to use a propensity
score model and *also* adjust
the outcome model**

WAKE FOREST UNIVERSITY SEMINAR

Doubly robust estimators

- Fit a **propensity score model**
- Fit two **outcome models**
- **Use these** to estimate an average treatment effect
- As long as **one is correctly specified**, your result will be **consistent**

WAKE FOREST UNIVERSITY SEMINAR

We're going to focus on the example where what they do is the doubly robust estimator

Doubly robust estimators

- Lunceford and Davidian (2004) derived a **consistent estimator for the ATE**
- Li and Greene (2013) derived a consistent estimator for the **ATM**
- We derive a consistent estimator for the **ATO**

WAKE FOREST UNIVERSITY SEMINAR

Doubly robust estimator

$$\widehat{\Delta}_{DR,w} = \frac{\sum_{i=1}^n w_i (m_1(\mathbf{X}_i, \widehat{\alpha}_1) - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i} +$$
$$\frac{\sum_{i=1}^n w_i Z_i (Y_i - m_1(\mathbf{X}_i, \widehat{\alpha}_1))}{\sum_{i=1}^n w_i Z_i} -$$
$$\frac{\sum_{i=1}^n w_i (1 - Z_i) (Y_i - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

Consistent if either the propensity score model is correct or outcome

Doubly robust estimator

$$\widehat{\Delta}_{DR,w} = \frac{\sum_{i=1}^n w_i (m_1(\mathbf{X}_i, \widehat{\alpha}_1) - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i} +$$
$$\frac{\sum_{i=1}^n w_i Z_i (Y_i - m_1(\mathbf{X}_i, \widehat{\alpha}_1))}{\sum_{i=1}^n w_i Z_i} -$$
$$\frac{\sum_{i=1}^n w_i (1 - Z_i) (Y_i - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

Consistent if either the propensity score model is correct or outcome

Doubly robust estimator

$$\widehat{\Delta}_{DR,w} = \frac{\sum_{i=1}^n w_i (m_1(\mathbf{X}_i, \widehat{\alpha}_1) - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i} +$$
$$\frac{\sum_{i=1}^n w_i Z_i (Y_i - m_1(\mathbf{X}_i, \widehat{\alpha}_1))}{\sum_{i=1}^n w_i (1 - Z_i) (Y_i - m_0(\mathbf{X}_i, \widehat{\alpha}_0))} - \frac{\widehat{y}_{treated} - \widehat{y}_{controls}}{\sum_{i=1}^n w_i (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

Consistent if either the propensity score model is correct or outcome

Doubly robust estimator

$$\widehat{\Delta}_{DR,w} = \frac{\sum_{i=1}^n w_i (m_1(\mathbf{X}_i, \widehat{\alpha}_1) - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i} +$$
$$\frac{\sum_{i=1}^n w_i Z_i (Y_i - m_1(\mathbf{X}_i, \widehat{\alpha}_1))}{\sum_{i=1}^n w_i Z_i} -$$
$$\frac{\sum_{i=1}^n w_i (1 - Z_i) (Y_i - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

residuals

Doubly robust estimator

$$\widehat{\Delta}_{DR,w} = \frac{\sum_{i=1}^n w_i (m_1(\mathbf{X}_i, \widehat{\alpha}_1) - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i} +$$
$$\frac{\sum_{i=1}^n w_i Z_i (Y_i - m_1(\mathbf{X}_i, \widehat{\alpha}_1))}{\sum_{i=1}^n w_i Z_i} -$$
$$\frac{\sum_{i=1}^n w_i (1 - Z_i) (Y_i - m_0(\mathbf{X}_i, \widehat{\alpha}_0))}{\sum_{i=1}^n w_i (1 - Z_i)}$$

WAKE FOREST UNIVERSITY SEMINAR

Consistent if either the propensity score model is correct or outcome

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

WAKE FOREST UNIVERSITY SEMINAR

- We can write the doubly robust estimator for the ATO in 3 equations, delta 1, delta2 and delta 3

ATO DR estimator

$$0 = \sum_{i=1}^n u_i(\boldsymbol{\theta})$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

we can estimate the propensity score and the treatment effect simultaneously in one step by solving the following estimating equations with respect to theta (we don't have to solve them jointly for the point estimate, but having this form will make the variance estimation more straightforward)

ATO DR estimator

$$\mathbf{u}(\boldsymbol{\theta}) = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1))Z\mathbf{X} \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0))(1 - Z)\mathbf{X} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

This can be solved by simultaneously solving these estimating equations.

ATO DR estimator

$$\mathbf{u}(\boldsymbol{\theta}) = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1))Z\mathbf{X} \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0))(1 - Z)\mathbf{X} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

ATO DR estimator

$$\boldsymbol{u}(\boldsymbol{\theta}) = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \textbf{\textit{outcome model (treatment)}} \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0))(1 - Z)\mathbf{X} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

ATO DR estimator

$$\boldsymbol{u}(\boldsymbol{\theta}) = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \textit{outcome model (treatment)} \\ \textit{outcome model (control)} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

ATO DR estimator

$$u(\boldsymbol{\theta}) = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \textit{outcome model (treatment)} \\ \textit{outcome model (control)} \\ \textit{propensity score model} \end{pmatrix}$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
propensity scores
- How do we incorporate the propensity scores?
weighting
- How do we calculate these weights?
- Which weights should we choose?
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?**
Doubly robust estimator

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
propensity scores
- How do we incorporate the propensity scores?
weighting
- How do we calculate these weights?
- Which weights should we choose?
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?**
Doubly robust estimator
- How do we estimate the variability?**

WAKE FOREST UNIVERSITY SEMINAR

ATO DR estimator

$$0 = \sum_{i=1}^n u_i(\boldsymbol{\theta})$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

M-Estimation

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-T})$$

WAKE FOREST UNIVERSITY SEMINAR

Since these estimating equations are unbiased, U of theta is an example of an M-estimator. The asymptotic distribution of an M estimator is the following.

the negative T is just inverse transpose

Sandwich estimator

$$\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-T}$$

WAKE FOREST UNIVERSITY SEMINAR

Note here that this technically will only be correct if both the outcome model AND the propensity score model are correctly specified.

Sandwich estimator

$$\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-T}$$

$$\mathbf{A} = -E \left[\frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]$$

WAKE FOREST UNIVERSITY SEMINAR

A is the model-based estimate, B is the “robust” estimate of the information (the variance of the score function)

Sandwich estimator

$$\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-T}$$

$$\mathbf{B} = E[\mathbf{u}(\boldsymbol{\theta})\mathbf{u}(\boldsymbol{\theta})^T]$$

WAKE FOREST UNIVERSITY SEMINAR

, B is the “robust” estimate of the information (the variance of the score function)

Sandwich estimator

$$\widehat{\text{var}}(\widehat{\theta}) = \frac{1}{n} \widehat{\mathbf{A}}_n^{-1} \widehat{\mathbf{B}}_n \widehat{\mathbf{A}}_n^{-T}$$

WAKE FOREST UNIVERSITY SEMINAR

We can replace the expectations by the sample average to get an approximate large sample variance

- How do we compensate for non-random treatment assignment?
propensity scores
- How do we incorporate the propensity scores?
weighting
- How do we calculate these weights?
- Which weights should we choose?
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?**
Doubly robust estimator
- How do we estimate the variability?**
Sandwich estimator

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
propensity scores
- How do we incorporate the propensity scores?
weighting
- How do we calculate these weights?
- Which weights should we choose?
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?
Doubly robust estimator
- How do we estimate the variability?
Sandwich estimator

- How do we compensate for non-random treatment assignment?
*propensity scores***
- How do we incorporate the propensity scores?
*weighting***
- How do we calculate these weights?**
- Which weights should we choose?**
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?**
Doubly robust estimator
- How do we estimate the variability?**
Sandwich estimator

What if our assumptions are violated?

WAKE FOREST UNIVERSITY SEMINAR

Sensitivity Analysis

WAKE FOREST UNIVERSITY SEMINAR

Motivation

- Review of **90 observational studies** JAMA, NEJM, and AJE
- **41** mentioned the issue of unmeasured confounding as a limitation
- **4** performed a quantitative sensitivity analysis

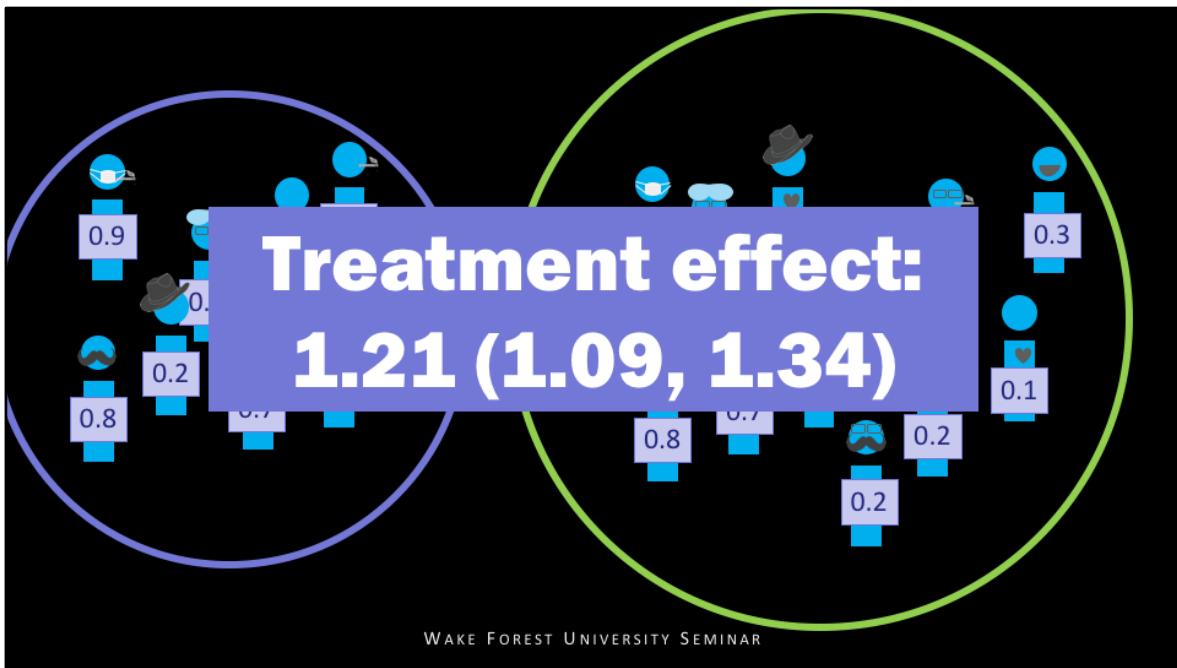
WAKE FOREST UNIVERSITY SEMINAR

We conducted a review of **90 observational studies** with statistically significant findings published in 2015 in the Journal of the American Medical Association, the New England Journal of Medicine, and the American Journal of Epidemiology. 41 (45.6%) mentioning the issue of unmeasured confounding as a limitation and only 4 (4.4%) performing a quantitative sensitivity analysis.

👉 Contributions

1. Examine the **measured confounders** via an **observed bias plot**
2. Use this to inform how we think about **unmeasured confounders** via **observed covariate E-values**

WAKE FOREST UNIVERSITY SEMINAR

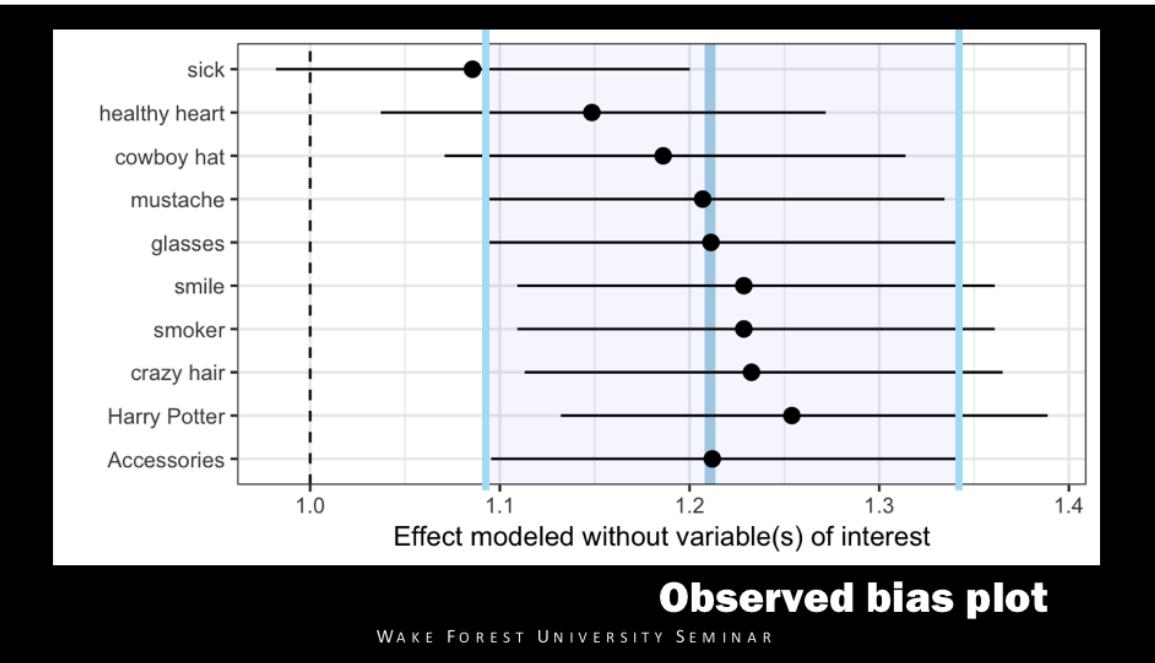


Let's bring back our goofy example. We do our propensity score weighting [click] to make the two groups comparable and then estimate the treatment effect using the methods we've discussed. [click] we end up with an observed treatment effect of 1.21 with a confidence interval of 1.09 to 1.34. Awesome, but what if we're missing something?

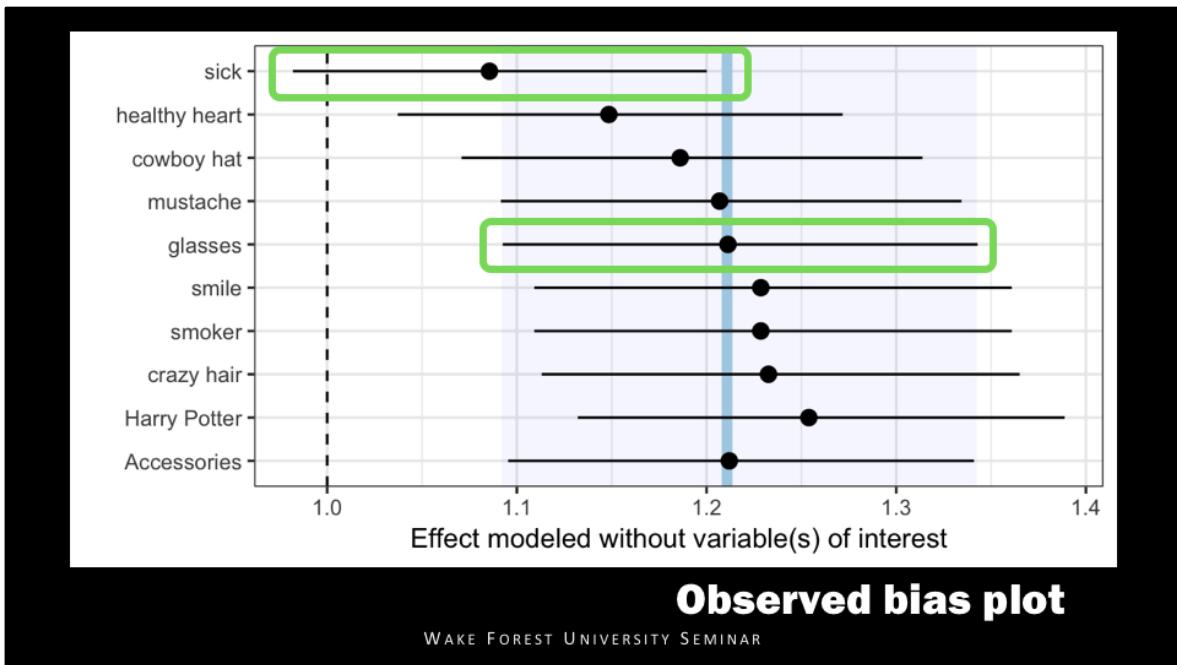
Meaningful confounders

1. How **imbalanced** is the unmeasured confounder between the exposure groups?
2. How **predictive** is the unmeasured confounder of the outcome?
3. How **Independent** is the unmeasured confounder from the other covariates?

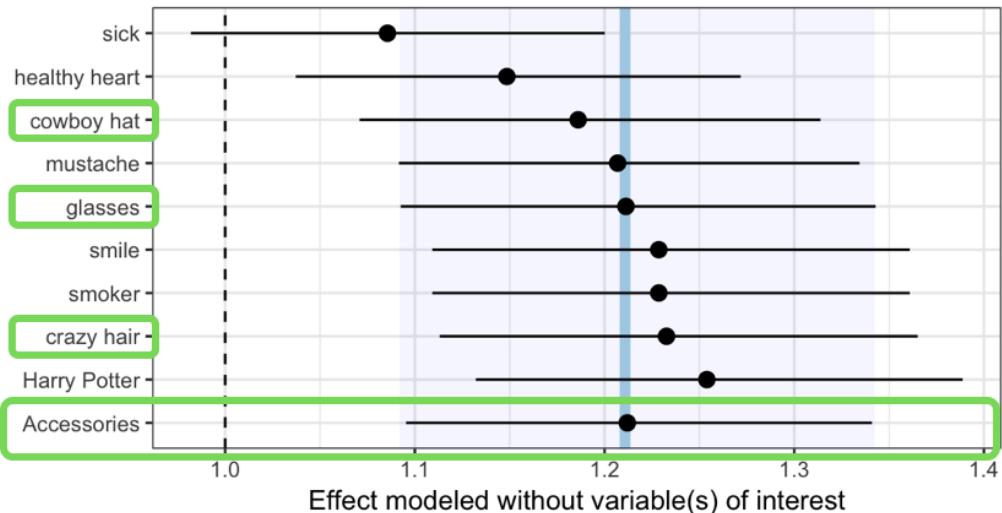
WAKE FOREST UNIVERSITY SEMINAR



Let me take a minute to explain this plot.



*observed bias factor



Observed bias plot

WAKE FOREST UNIVERSITY SEMINAR

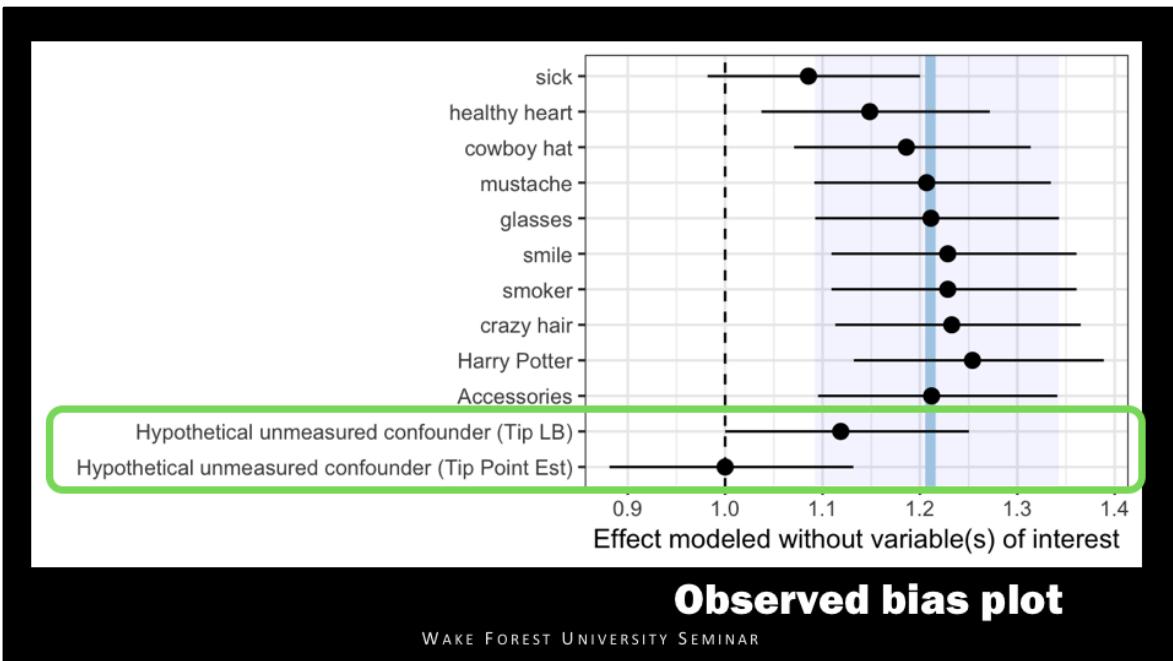


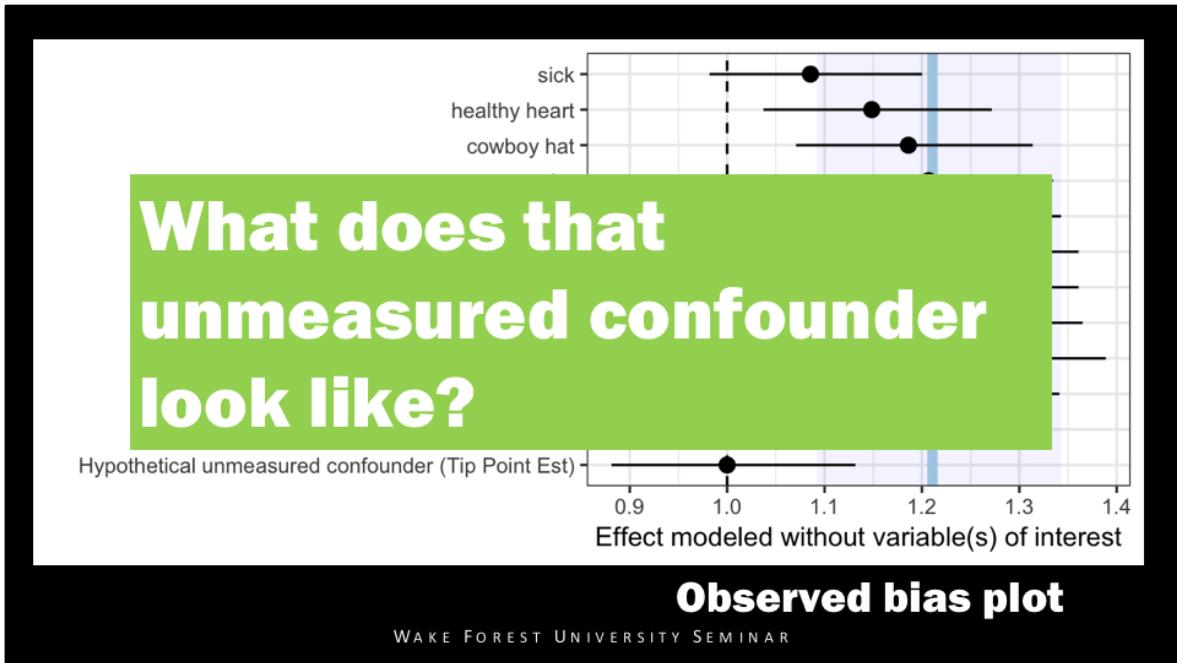
Tipping point analyses

what will tip our confidence bound to cross 1

WAKE FOREST UNIVERSITY SEMINAR

We've formally defined these tipping point analyses for binary and continuous unmeasured confounders





Is it plausible?

You need

- ✓ **Exposure-outcome effect**
- ✓ **Exposure-unmeasured confounder effect**
- ✓ **Outcome-unmeasured confounder effect**

WAKE FOREST UNIVERSITY SEMINAR
HEALTH POLICY TAKEOUT THURSDAY

You need

- ✓ **Exposure-outcome effect** generally estimated from a model, for example an odds ratio, hazard ratio, or risk ratio
- ✓ **Exposure-unmeasured confounder effect**
- ✓ **Outcome-unmeasured confounder effect**

WAKE FOREST UNIVERSITY SEMINAR

You need

- ✓ **Exposure-outcome effect** generally estimated from a model, for example an odds ratio, hazard ratio, or risk ratio
- ✓ **Exposure-unmeasured confounder effect**
- ✓ **Outcome-unmeasured confounder effect**

WAKE FOREST UNIVERSITY SEMINAR

You need

- ✓ **Exposure-outcome effect** generally estimated from a model, for example an odds ratio, hazard ratio, or risk ratio
- ✓ **Exposure-unmeasured confounder effect**
- ✓ **Outcome-unmeasured confounder effect**

WAKE FOREST UNIVERSITY SEMINAR

Instead of specifying these, we can use a new concept known as the E-value

explain LB

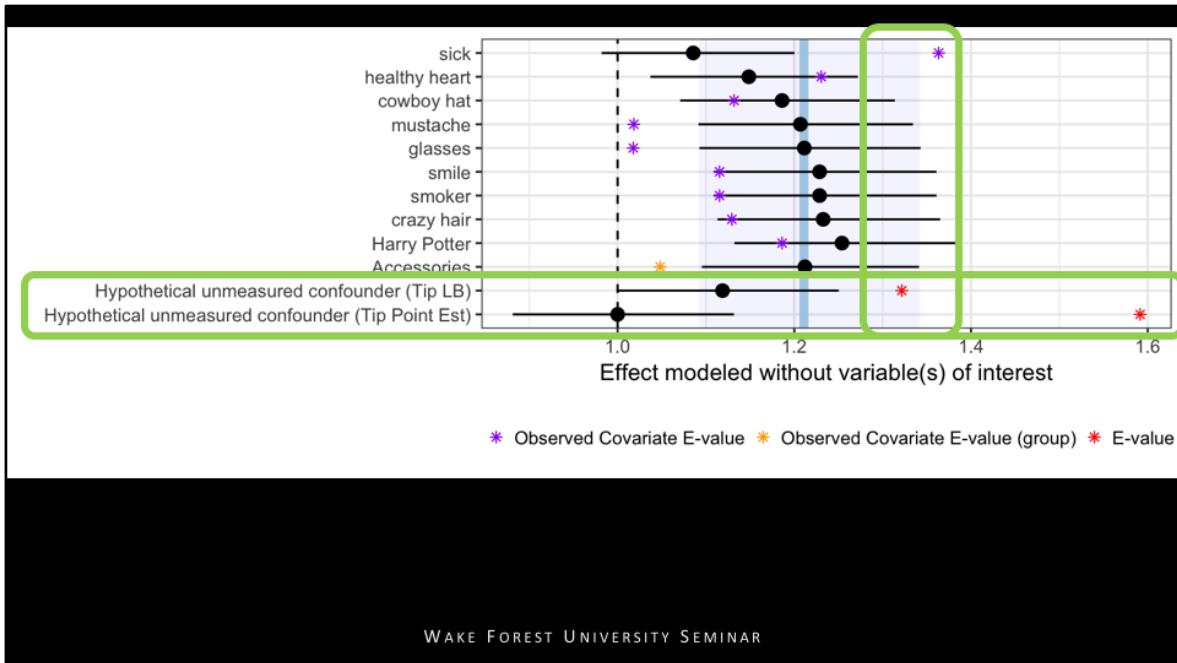
Vanderweele and Ding's, it is the point that minimizes the strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the exposure and outcome, conditional on the measured covariates, to explain away an observed exposure-outcome association. Again this is certainly one way to look at the analysis, but can be difficult contextualize without more information.

Observed covariate E-value

$$E\text{-value}_{adj} = \frac{LB_{obs}}{LB_{adj}} + \sqrt{\frac{LB_{obs}}{LB_{adj}} \times \left(\frac{LB_{obs}}{LB_{adj}} - 1 \right)}$$

WAKE FOREST UNIVERSITY SEMINAR

we've used similar methodology to create an observed covariate e-value, where we allow you to input the value you are trying to adjust to, rather than always setting the adjusted limiting bound to one. Replacing LB adj with the lower bound of the observed bias factors, we can calculate a quantity that would demonstrate what the E-value would be if the only unmeasured confounder were the one we dropped to calculate the observed bias factor



R package

tipr



WAKE FOREST UNIVERSITY SEMINAR

And we've created an R-package to calculate them

- How do we compensate for non-random treatment assignment?
propensity scores
- How do we incorporate the propensity scores?
weighting
- How do we calculate these weights?
- Which weights should we choose?
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?
Doubly robust estimator
- How do we estimate the variability?
Sandwich estimator
- What if our assumptions are violated?**
Perform a quantitative sensitivity analysis

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
propensity scores
- How do we incorporate the propensity scores?
weighting
- How do we calculate these weights?
- Which weights should we choose?
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?
Doubly robust estimator
- How do we estimate the variability?
Sandwich estimator
- What if our assumptions are violated?
Perform a quantitative sensitivity analysis

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
*propensity scores***
- How do we incorporate the propensity scores?
*weighting***
- How do we calculate these weights?**
- Which weights should we choose?
*The variance for the ATO and ATM is preferable to that of the ATE***
- How do we estimate the treatment effect?
*Doubly robust estimator***
- How do we estimate the variability?
*Sandwich estimator***
- What if our assumptions are violated?
*Perform a quantitative sensitivity analysis***

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
*propensity scores***
- How do we incorporate the propensity scores?
*weighting***
- How do we calculate these weights?**
- Which weights should we choose?**
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?**
Doubly robust estimator
- How do we estimate the variability?**
Sandwich estimator
- What if our assumptions are violated?**
Perform a quantitative sensitivity analysis

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
*propensity scores***
- How do we incorporate the propensity scores?
*weighting***
- How do we calculate these weights?**
- Which weights should we choose?
*The variance for the ATO and ATM is preferable to that of the ATE***
- How do we estimate the treatment effect?
*Doubly robust estimator***
- How do we estimate the variability?
*Sandwich estimator***
- What if our assumptions are violated?
*Perform a quantitative sensitivity analysis***

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
*propensity scores***
- How do we incorporate the propensity scores?
*weighting***
- How do we calculate these weights?**
- Which weights should we choose?
*The variance for the ATO and ATM is preferable to that of the ATE***
- How do we estimate the treatment effect?
*Doubly robust estimator***
- How do we estimate the variability?
*Sandwich estimator***
- What if our assumptions are violated?
*Perform a quantitative sensitivity analysis***

WAKE FOREST UNIVERSITY SEMINAR

- How do we compensate for non-random treatment assignment?
propensity scores
- How do we incorporate the propensity scores?
weighting
- How do we calculate these weights?
- Which weights should we choose?
The variance for the ATO and ATM is preferable to that of the ATE
- How do we estimate the treatment effect?
Doubly robust estimator
- How do we estimate the variability?
Sandwich estimator
- What if our assumptions are violated?
Perform a quantitative sensitivity analysis

WAKE FOREST UNIVERSITY SEMINAR

Q&A

WAKE FOREST UNIVERSITY SEMINAR

13 minutes

References

- Austin, Peter C. 2009. "Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples." *Statistics in Medicine* 28 (25): 3083–3107.
- Austin, Peter C, and Elizabeth A Stuart. 2015. "Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies." *Statistics in Medicine* 34 (28): 3661–79.
- Bang, Heejung, and James M Robins. 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics* 61 (4): 962–73.
- Belle, G. van. 2011. *Statistical Rules of Thumb*. Wiley Series in Probability and Statistics. Wiley. <https://books.google.com/books?id=UaEL-HI6v2YC>.
- Bross, IDJ. 1966. "Spurious effects from an extraneous variable." *Journal of Chronic Diseases*.
- Brumback, Babette A, Miguel A Hernán, Sébastien J P A Haneuse, and James M Robins. 2004. "Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures." *Statistics in Medicine* 23 (5): 749–67.
- Busso, Matias, John DiNardo, and Justin McCrary. 2009. "Finite sample properties of semiparametric estimators of average treatment effects." *Forthcoming in the Journal of Business and Economic Statistics*.

WAKE FOREST UNIVERSITY SEMINAR

References

- . 2014. "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators." *Review of Economics and Statistics* 96 (5): 885–97.
- Connors, A F, T Speroff, N V Dawson, and C Thomas. 1996. "The effectiveness of right heart catheterization in the initial care of critically ill patients." *Jama*.
- Cornfield, J, W Haenszel, E C Hammond, A M Lilienfeld, M B Shimkin, and E L Wynder. 1959. "Smoking and lung cancer: recent evidence and a discussion of some questions." *Journal of the National Cancer Institute* 22 (1): 173–203.
- Crump, R K, V J Hotz, G W Imbens, and O A Mitnik. 2009. "Dealing with limited overlap in estimation of average treatment effects." *Biometrika* 96 (1): 187–99.
- Ding, Peng, and Tyler J VanderWeele. 2016. "Sensitivity Analysis Without Assumptions." *Epidemiology (Cambridge, Mass.)* 27 (3): 368–77.
- D'Agostino, R B. 1998. "Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group." *Statistics in Medicine* 17 (19): 2265–81.
- Freedman, David A, and Richard A Berk. 2008. "Weighting regressions by propensity scores." *Evaluation Review* 32 (4): 392–409.

WAKE FOREST UNIVERSITY SEMINAR

References

- Funk, Michele Jonsson, Daniel Westreich, Chris Wiesen, Til Stürmer, M Alan Brookhart, and Marie Davidian. 2011. "Doubly Robust Estimation of Causal Effects." *American Journal of Epidemiology* 173 (7): 761–67.
- Greenland, Sander. 1998. "The sensitivity of a sensitivity analysis." 1997 *Proceedings of the Biometrics Section American Statistical Association*, 19–21.
- . 2001. "Sensitivity Analysis, Monte Carlo Risk Analysis, and Bayesian Uncertainty Assessment." *Risk Analysis* 21 (4): 579–84.
- . 2003. "The Impact of Prior Distributions for Uncontrolled Confounding and Response Bias." *Journal of the American Statistical Association* 98 (461): 47–54.
- . 2005. "Multiple-bias modelling for analysis of observational data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168 (2): 267–306.
- Groenwold, Rolf H H, Anna M M Van Deursen, Arno W Hoes, and Elko Hak. 2008. "Poor Quality of Reporting Confounding Bias in Observational Intervention Studies: A Systematic Review." *Annals of Epidemiology* 18 (10): 746–51.
- Groenwold, Rolf H H, Frank Vries, Anthonius Boer, Wiebe R Pestman, Frans H Rutten, Arno W Hoes, and Olaf H Klungel. 2011. "Balance measures for propensity score methods: a clinical example on beta-agonist use and the risk of myocardial infarction." *Pharmacoepidemiology and Drug Safety* 20 (11): 1130–7.

WAKE FOREST UNIVERSITY SEMINAR

References

- Hansen, Ben B, and Mark M Fredrickson. 2014. "Omitted Variable Sensitivity Analysis with the Annotated Love Plot." *Society for Research on Educational Effectiveness*.
- Hosman, C A, B B Hansen, and P W Holland. 2010. "The Sensitivity of Linear Regression Coefficients' Confidence Limits to the Omission of a Confounder." *The Annals of Applied Statistics*.
- Hsu, Jesse Y, and Dylan S Small. 2013. "Calibrating Sensitivity Analyses to Observed Covariates in Observational Studies." *Biometrics* 69 (4): 803-11.
- Imai, K, G King, and E A Stuart. 2008. "Misunderstandings between experimentalists and observationalists about causal inference." *Journal of the Royal Statistical Society*
- Imbens, Guido W, and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.
- Imbens, Guido W, and Jeffrey M Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5-86.
- Joffe, Marshall M, Thomas R Ten Have, Harold I Feldman, and Stephen E Kimmel. 2004. "Model Selection, Confounder Control, and Marginal Structural Models." *The American Statistician* 58 (4): 272-79.

WAKE FOREST UNIVERSITY SEMINAR

References

- Li, Liang, and Tom Greene. 2013. "A Weighting Analogue to Pair Matching in Propensity Score Analysis." *The International Journal of Biostatistics*.
- Lin, D Y, B M Psaty, and R A Kronmal. 1998. "Assessing the sensitivity of regression results to unmeasured confounders in observational studies." *Biometrics* 54 (3): 948–63.
- Lipsitz, Stuart R, Joseph G Ibrahim, and Lue Ping Zhao. 1999. "A Weighted Estimating Equation for Missing Covariate Data with Properties Similar to Maximum Likelihood." *Journal of the American Statistical Association* 94 (448): 1147–60.
- Love, T E. 2002. "Displaying covariate balance after adjustment for selection bias." Joint Statistical Meeting: Section on Health Policy Statistics.
- Lumley, Thomas. 2011. *Complex Surveys: A Guide to Analysis Using R*. Vol. 565. John Wiley & Sons.
- Lunceford, Jared K, and Marie Davidian. 2004. "Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study." *Statistics in Medicine* 23 (19): 2937–60.
- Mao, Huzhang, and Liang Li. 2018. *PSW: Propensity Score Weighting Methods for Dichotomous Treatments*. <https://CRAN.R-project.org/package=PSW>.
- McCandless, Lawrence C, Paul Gustafson, and Adrian Levy. 2007. "Bayesian sensitivity analysis for unmeasured confounding in observational studies." *Statistics in Medicine* 26 (11): 2331–47.
- McCandless, Lawrence C, Paul Gustafson, and Adrian R Levy. 2008. "A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding." *Journal of Clinical Epidemiology* 61 (3): 247–55.

References

- Neugebauer, Romain, and Mark van der Laan. 2005. "Why prefer double robust estimators in causal inference?" *Journal of Statistical Planning and Inference* 129 (1-2): 405–26.
- Neyman, Jersey. 1923. "Sur les applications de la théorie des probabilités aux expériences agricoles: Essai des principes." *Roczniki Nauk Rolniczych* 10: 1–51.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robins, James M. 2000. "Robust estimation in sequentially ignorable missing data and causal inference models." In *Proceedings of the American Statistical Association*, 6–10.
- Robins, James M, Andrea Rotnitzky, and Mark van der Laan. 2000. "On Profile Likelihood: Comment." *Journal of the American Statistical Association* 95 (450): 477.
- Robins, James M, Andrea Rotnitzky, and Daniel O Scharfstein. 2000. "Sensitivity Analysis for Selection bias and unmeasured Confounding in missing Data and Causal inference models." In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, 1–94. New York, NY: Springer New York.
- Robins, James M, Andrea Rotnitzky, and Lue Ping Zhao. 2012. "Estimation of Regression Coefficients When Some Regressors are not Always Observed." *Journal of the American Statistical Association* 89 (427): 846–66.

WAKE FOREST UNIVERSITY SEMINAR

References

- Robins, James, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky. 2007. "Comment: Performance of Double-Robust Estimators When Inverse Probability Weights Are Highly Variable." *Statistical Science* 22 (4): 544–59.
- Rosenbaum, P R, and D B Rubin. 1983a. "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome." *Journal of the Royal Statistical Society Series B*.
- . 1983b. "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome." *Journal of the Royal Statistical Society Series B*
- Rosenbaum, Paul R. 2012. "Optimal Matching of an Optimally Chosen Subset in Observational Studies." *Journal of Computational and Graphical Statistics* 21 (1): 57–71.
- Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66 (5): 688.

WAKE FOREST UNIVERSITY SEMINAR

References

- _____. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology* 2 (3-4): 169–88.
- Samuels, Lauren Ruth. 2017. "Aspects of Causal Inference within the Evenly Matchable Population: The Average Treatment Effect on the Evenly Matchable Units, Visually Guided Cohort Selection, and Bagged One-to-One Matching." PhD thesis, Vanderbilt University.
- Scharfstein, Daniel O, Andrea Rotnitzky, and James M Robins. 1999. "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models." *Journal of the American Statistical Association* 94 (448): 1096.
- Schlesselman, J J. 1978. "Assessing effects of confounding variables." *American Journal of Epidemiology* 108 (1): 3–8.
- Schneeweiss, Sebastian. 2006. "Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics." *Pharmacoepidemiology and Drug Safety* 15 (5): 291–303.
- Snow, Greg. 2013. *ObsSens: Sensitivity Analysis for Observational Studies*. <https://CRAN.R-project.org/package=obsSens>.
- Stuart, Elizabeth A, Brian K Lee, and Finbarr P Leacy. 2013. "Prognostic scorebased balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research." *Journal of Clinical Epidemiology* 66 (8): S84–S90.e1.

WAKE FOREST UNIVERSITY SEMINAR

References

- Stürmer, Til, Sebastian Schneeweiss, Jerry Avorn, and Robert J Glynn. 2005. "Adjusting Effect Estimates for Unmeasured Confounding with Validation Data using Propensity Score Calibration." *American Journal of Epidemiology* 162 (3): 279–89.
- VanderWeele, T J, and M A Hernán. 2008. "Causal directed acyclic graphs and the direction of unmeasured confounding bias." *Epidemiology (Cambridge)*
- VanderWeele, Tyler. 2015. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.
- VanderWeele, Tyler J. 2008a. "Sensitivity analysis: distributional assumptions and confounding assumptions." *Biometrics*.
- . 2008b. "The Sign of the Bias of Unmeasured Confounding." *Biometrics* 64 (3): 702–6.
- . 2013. "Unmeasured confounding and hazard scales: sensitivity analysis for total, direct, and indirect effects." *European Journal of Epidemiology* 28 (2): 113–17.

WAKE FOREST UNIVERSITY SEMINAR

References

- VanderWeele, Tyler J, and Onyebuchi A Arah. 2011. "Bias Formulas for Sensitivity Analysis of Unmeasured Confounding for General Outcomes, Treatments, and Confounders." *Epidemiology* 22 (1): 42–52.
- VanderWeele, Tyler J, and Peng Ding. 2017. "Sensitivity Analysis in Observational Research: Introducing the E-Value." *Ann Intern Med*, July.
- VanderWeele, Tyler J, Bhramar Mukherjee, and Jinbo Chen. 2012. "Sensitivity analysis for interactions under unmeasured confounding." *Statistics in Medicine* 31 (22): 2552–64.
- Venables, W N, and B D Ripley. 2002. *Modern Applied Statistics with S*. 4th edition. Statistics and Computing. New York, NY: Springer New York.
- Williamson, Elizabeth J, Andrew Forbes, and Ian R White. 2013. "Variance reduction in randomised trials by inverse probability weighting using the propensity score." *Statistics in Medicine* 33 (5): 721–37.
- Zeileis, Achim. 2004. "Econometric Computing with HC and HAC Covariance Matrix Estimators." *Journal of Statistical Software* 11 (10).
- . 2006. "Object-Oriented Computation of Sandwich Estimators." *Journal of Statistical Software* 16 (9).

WAKE FOREST UNIVERSITY SEMINAR

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

WAKE FOREST UNIVERSITY SEMINAR

- We can write the doubly robust estimator for the ATO in 3 equations, delta 1, delta2 and delta 3

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_1 = \frac{\sum_{i=1}^n \left(Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) + (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) (m_1(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_1) - m_0(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_0))}{\sum_{i=1}^n \left(Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) + (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right)}$$

WAKE FOREST UNIVERSITY SEMINAR

- Here is delta 1

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_1 = \frac{\sum_{i=1}^n \left(Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) + (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) (m_1(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_1) - m_0(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_0))}{\sum_{i=1}^n \left(Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) + (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right)}$$

WAKE FOREST UNIVERSITY SEMINAR

- This will be weighted by 1 minus the propensity score for the treated

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_1 = \frac{\sum_{i=1}^n \left(Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) + (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) (m_1(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_1) - m_0(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_0))}{\sum_{i=1}^n \left(Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) + (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right)}$$

WAKE FOREST UNIVERSITY SEMINAR

- And the propensity score for the controls

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_1 = \frac{\sum_{i=1}^n \left(Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) + (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) (m_1(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_1) - m_0(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_0))}{\sum_{i=1}^n \left(Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right) + (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) \right)}$$

WAKE FOREST UNIVERSITY SEMINAR

- Here is the difference between the Y predicted among the treated and the Y predicted among the controls

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_2 = \frac{\sum_{i=1}^n Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})\right) (Y_i - m_1(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_1))}{\sum_{i=1}^n Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})\right)}$$

WAKE FOREST UNIVERSITY SEMINAR

- Here is delta 2

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_2 = \frac{\sum_{i=1}^n Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})\right) (Y_i - m_1(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_1))}{\sum_{i=1}^n Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})\right)}$$

WAKE FOREST UNIVERSITY SEMINAR

- Notice this is just among the treated, so if the subject i is in the control group, this will be zero, since Zi will be 0

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_2 = \frac{\sum_{i=1}^n Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})\right) (Y_i - m_1(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_1))}{\sum_{i=1}^n Z_i \left(1 - e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})\right)}$$

WAKE FOREST UNIVERSITY SEMINAR

- This piece is essentially the residual between the observed Y among the treated and the predicted Y among the treated

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_3 = \frac{\sum_{i=1}^n (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) (Y_i - m_0(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_0))}{\sum_{i=1}^n (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})}$$

WAKE FOREST UNIVERSITY SEMINAR

- Here is delta 3

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_3 = \frac{\sum_{i=1}^n (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) (Y_i - m_0(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_0))}{\sum_{i=1}^n (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})}$$

WAKE FOREST UNIVERSITY SEMINAR

- This summation is just among the controls, if the participant is in the treated group, $1 - Z_i$ will be 0

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\delta}_3 = \frac{\sum_{i=1}^n (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}) (Y_i - m_0(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_0))}{\sum_{i=1}^n (1 - Z_i) e(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})}$$

WAKE FOREST UNIVERSITY SEMINAR

- And this is the residual of the observed y among the controls and the predicted y , fit among the controls

ATO DR estimator

$$\mathbf{u}(\boldsymbol{\theta}) = \begin{pmatrix} (m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_1)(Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) + (1 - Z)e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - \delta_2)Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_3)(1 - Z)e(\mathbf{X}, \boldsymbol{\beta}) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1))Z\mathbf{X} \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0))(1 - Z)\mathbf{X} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$
$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

Again this can be solved by simultaneously solving these estimating equations. Note here that this technically will only be correct if both the outcome model AND the propensity score model are correctly specified.

ATO DR estimator

$$u(\boldsymbol{\theta}) = \begin{pmatrix} (m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_1)(Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) + (1 - Z)e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - \delta_2)Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_3)(1 - Z)e(\mathbf{X}, \boldsymbol{\beta}) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1))Z\mathbf{X} \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0))(1 - Z)\mathbf{X} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$
$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

This is more complicated than the large-sample variance that we derived for the ATO IPW estimator because now we have more parameters to solve for, rather than just mu1, mu0 and beta, we now have delta 1 2 and 3, and alpha 1 and 0 from the outcome models

ATO DR estimator

$$\mathbf{u}(\boldsymbol{\theta}) = \begin{pmatrix} (m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_1)(Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) + (1 - Z)e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - \delta_2)Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_3)(1 - Z)e(\mathbf{X}, \boldsymbol{\beta}) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1))Z\mathbf{X} \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0))(1 - Z)\mathbf{X} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

These first 3 lines are the estimating equations that make up the 3 parts of the doubly robust equation we just went through

ATO DR estimator

$$\mathbf{u}(\boldsymbol{\theta}) = \begin{pmatrix} (m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_1)(Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) + (1 - Z)e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - \delta_2)Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_3)(1 - Z)e(\mathbf{X}, \boldsymbol{\beta}) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1))Z\mathbf{X} \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0))(1 - Z)\mathbf{X} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$
$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

These two are the estimating equations for fitting the outcome models, the top is for the outcome model among the treated, the bottom is the outcome model among the controls

ATO DR estimator

$$\mathbf{u}(\boldsymbol{\theta}) = \begin{pmatrix} (m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_1)(Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) + (1 - Z)e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1) - \delta_2)Z(1 - e(\mathbf{X}, \boldsymbol{\beta})) \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0) - \delta_3)(1 - Z)e(\mathbf{X}, \boldsymbol{\beta}) \\ (Y - m_1(\mathbf{X}, \boldsymbol{\alpha}_1))Z\mathbf{X} \\ (Y - m_0(\mathbf{X}, \boldsymbol{\alpha}_0))(1 - Z)\mathbf{X} \\ \mathbf{X}(Z - e(\mathbf{X}, \boldsymbol{\beta})) \end{pmatrix}$$
$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

This final equation accounts for the propensity score model

ATO estimator

$$0 = \sum_{i=1}^n u_i(\boldsymbol{\theta})$$

$$\boldsymbol{\theta} = (\delta_1, \delta_2, \delta_3, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_0^T, \boldsymbol{\beta}^T)^T$$

WAKE FOREST UNIVERSITY SEMINAR

we can estimate the propensity score and the treatment effect simultaneously in one step by solving the following estimating equations with respect to theta

M-Estimation

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-T})$$

WAKE FOREST UNIVERSITY SEMINAR

Since these estimating equations are unbiased, U of theta is an example of an M-estimator. The asymptotic distribution of an M estimator is the following.

the negative T is just inverse transpose

Sandwich estimator

$$\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-T}$$

WAKE FOREST UNIVERSITY SEMINAR

Sandwich estimator

$$\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-T}$$

$$\mathbf{A} = -E \left[\frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]$$

WAKE FOREST UNIVERSITY SEMINAR

A is the model-based estimate, B is the “robust” estimate of the information (the variance of the score function)

Sandwich estimator

$$\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-T}$$

$$\mathbf{B} = E[\mathbf{u}(\boldsymbol{\theta})\mathbf{u}(\boldsymbol{\theta})^T]$$

WAKE FOREST UNIVERSITY SEMINAR

Sandwich estimator

$$\widehat{\text{var}}(\widehat{\theta}) = \frac{1}{n} \widehat{\mathbf{A}}_n^{-1} \widehat{\mathbf{B}}_n \widehat{\mathbf{A}}_n^{-T}$$

WAKE FOREST UNIVERSITY SEMINAR

We can replace the expectations by the sample average to get an approximate large sample variance

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\text{var}}(\widehat{\Delta}_{DR,ATO}) = (1 \ 1 \ -1 \ 0 \ 0 \ 0) \frac{1}{n} \widehat{\mathbf{A}}_n^{-1} \widehat{\mathbf{B}}_n \widehat{\mathbf{A}}_n^{-T} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

WAKE FOREST UNIVERSITY SEMINAR

Because we are estimating delta 1 plus delta 2 minus delta 3, after we estimate our sandwich estimator

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\text{var}}(\widehat{\Delta}_{DR,ATO}) = (1 \ 1 \ -1 \ 0 \ 0 \ 0) \frac{1}{n} \widehat{\mathbf{A}}_n^{-1} \widehat{\mathbf{B}}_n \widehat{\mathbf{A}}_n^{-T} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

WAKE FOREST UNIVERSITY SEMINAR

This will be our variance estimator where

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\text{var}}(\widehat{\Delta}_{DR,ATO}) = (1 \ 1 \ -1 \ 0 \ 0 \ 0) \frac{1}{n} \widehat{\mathbf{A}}_n^{-1} \widehat{\mathbf{B}}_n \widehat{\mathbf{A}}_n^{-T} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

WAKE FOREST UNIVERSITY SEMINAR

We will multiply by the vector 1 1 -1 with the remaining entries 0 on both sides

ATO DR estimator

$$\widehat{\Delta}_{DR,ATO} = \widehat{\delta}_1 + \widehat{\delta}_2 - \widehat{\delta}_3$$

$$\widehat{\text{var}}(\widehat{\Delta}_{DR,ATO}) = (1 \ 1 \ -1 \ 0 \ 0 \ 0) \frac{1}{n} \widehat{\mathbf{A}}_n^{-1} \widehat{\mathbf{B}}_n \widehat{\mathbf{A}}_n^{-T} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

WAKE FOREST UNIVERSITY SEMINAR

We multiply by our estimated sandwich estimator where we've plugged in our sample averages for the expectations