

Design Trumps Analysis: Propensity Score Weighting Analysis

Scott Cunningham

Recall Don and Bernie and Brigham all saying the importance of the "design" stage. Recall Rubin's article from 2008, "design trumps analysis". And it may seem abstract what that means, we will illustrate some of it here because none of what we will do now will look at the outcomes at all. We will only be focused on reconstructing the hypothetical experiment through reweighting the variables. The outcome will be mortality and the treatment will be health insurance treatment status. The context are near elderly individuals in a widely used retirement survey. The author is our leader Bernie Black with Espin-Sanchez, French, and Litvak in the 2017 *Journal of Health Economics*, "The Long-term Effect of Health Insurance on Near-Elderly Health and Mortality".

They wanted to estimate the average effect of health insurance on near elderly (50-61yo at start of panel) mortality using propensity score methods. For this design stage, we are going to focus on inverse probability weighting (IPW) because it's an easy way to reweight. Remember, propensity scores can be used for estimating treatment effects using matching or reweighting, but that's the analysis stage. We are in the design stage, so we will reweight since we don't have the outcome data in this stage.

First, let's look at the fact that the treatment and control group is imbalanced in Table 1. Note races are non-hispanic white, non-hispanic black, hispanic and non-hispanic other for race 1-4. Others are baseline variables that should be self-explanatory. And there is imbalance because the means are different. We could do t-tests on the difference in means, but I don't do that here. But in Bernie's do file, he does do that, and I leave it to you to update the `Day1_Cunningham.do` file according to Bernie's if you want to see how to do that. But for this document, I am just showing a simple difference in means for the treatment and control group. Age is around the same, but that's expected as it's a panel of people the same age. But beyond that we see differences for many things, especially whether they are employed. Note that our treatment group is that they don't have insurance, but the control is they do, which is why we see the treatment group having lower employment.

Next we estimate the propensity score using logit. This is a two step procedure. First estimate the propensity score:

$$Pr(D_i = 1|X_i) = \frac{e^{X_i'\beta}}{1 + e^{X_i'\beta}} \quad (1)$$

The using the fitted values from the logit, predict the conditional probability of treat which we will call $\widehat{\rho(x)}$. And notice that the interpretation of this, first of all, is a conditional probability of treatment. Given values of X , what percent of the population is in the treatment group *of units with those values*. It approximates this value, but in the end we have collapsed the information in our X variables into a single dimension scalar, the propensity score. Note that we estimated the propensity score; we do not know the true propensity score except in a randomized experiment, which is not happening here.

We can then in Figure 1 look at the distribution of the propensity score by treatment category. That's customarily done in a histogram. It smooths over all this using the kernel density, though, which I find can hide a little. But we that there.

Now we are going to reweight the variables using the inverse probability weights. You weight the treated units by $IPW_i = \frac{1}{\widehat{P}(X_i)}$ and you weight j control units by $IPW_j = \frac{1}{1-\widehat{P}(X_j)}$. I show that in Table 2 and then have a love plot in Figure 2. Note that this contains both table 1 and table 2 which is nice.

1 Summary Statistics Before Weighting

Table 1: Summary Statistics Before Weighting

Variables	Treatment		Control	
	Mean	SD	Mean	SD
age	55.52	3.14	55.85	3.16
gender	0.44	0.50	0.48	0.50
log_income_wave_1	9.17	2.34	10.40	1.27
diabetes_wave_1	0.11	0.31	0.09	0.29
cancer_wave_1	0.04	0.19	0.05	0.22
employed_ft_wave_1	0.37	0.48	0.58	0.49
veteran_wave_1	0.18	0.38	0.27	0.44
race_1	0.52	0.50	0.75	0.43
race_2	0.22	0.41	0.16	0.37
race_3	0.23	0.42	0.07	0.25
race_4	0.03	0.18	0.02	0.14

2 Propensity Score Distribution

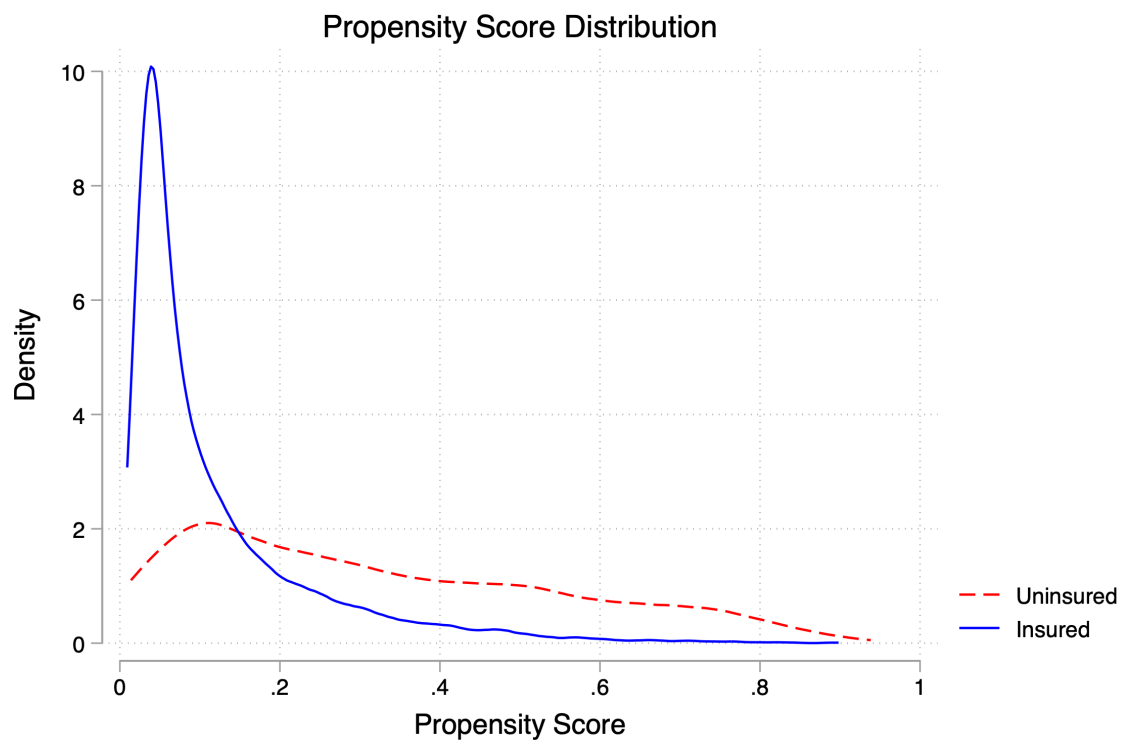


Figure 1: Distribution of Propensity Scores

3 Summary Statistics After Weighting

Table 2: Summary Statistics After Weighting

Variables	Treatment		Control	
	Mean	SD	Mean	SD
age	55.85	3.15	55.81	3.15
gender	0.47	0.50	0.47	0.50
log_income_wave_1	10.18	1.73	10.26	1.42
diabetes_wave_1	0.09	0.29	0.10	0.29
cancer_wave_1	0.05	0.22	0.05	0.21
employed_ft_wave_1	0.53	0.50	0.55	0.50
veteran_wave_1	0.26	0.44	0.26	0.44
race_1	0.72	0.45	0.72	0.45
race_2	0.16	0.37	0.17	0.38
race_3	0.09	0.29	0.09	0.29
race_4	0.03	0.16	0.02	0.14

4 Balance Before and After Weighting

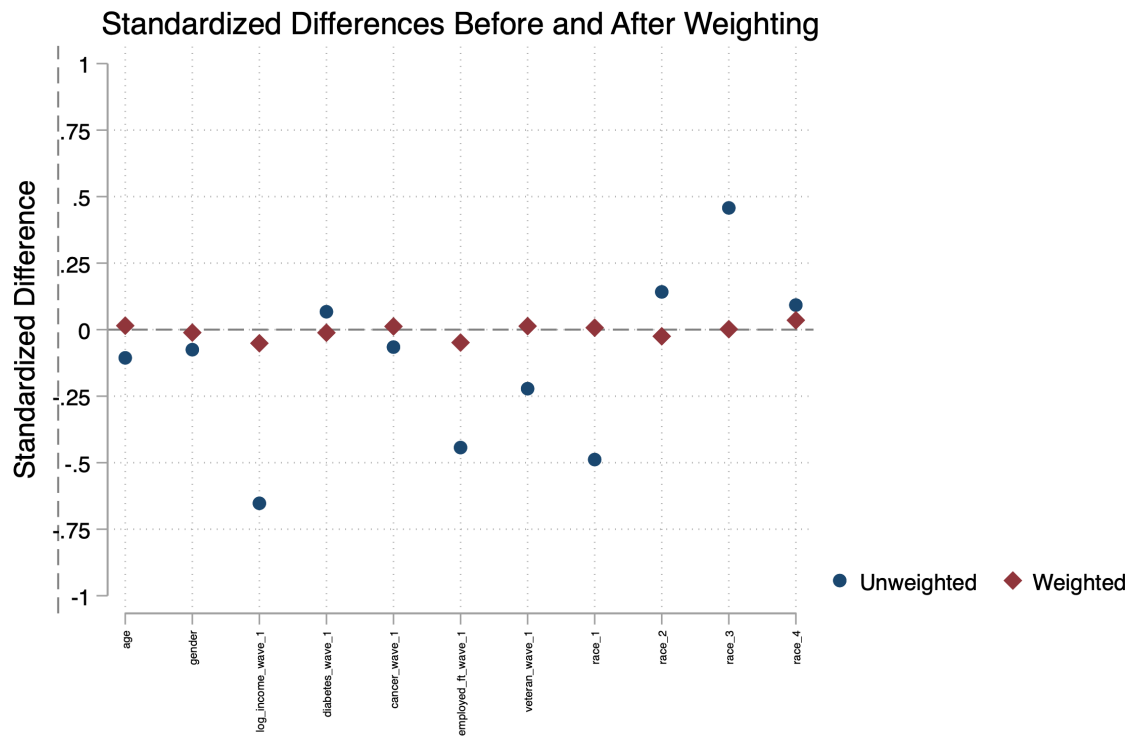


Figure 2: Standardized Differences Before and After Weighting