Ec1127 Final Project: Job Training Program's Effect on Earnings
Meng Chen

Overview

The focus of this study is to infer the causal effect of a job training program on a subset of its participants. We will look at both the effect of assignment to the job training program and the effect of actually receiving the training; as the data shows, not all who are assigned to the program follow through. Such people are never-takers, assuming that the monotonicity assumption holds, i.e. that there are no defiers. In this subset of the population only 1-sided noncompliance is an issue, as none of those who are assigned to control received the training program.

First we go through the design phase, inducing balance in covariates by using propensity scores to minimize the bias of the point estimate for the average causal effect of the job training program on the outcome of interest (" weekly earnings of those in this sample population four years after assignment to treatment"). In the complete experiment population, the distribution of covariates within treatment groups should look similar, the differences due only to variance in randomizations. However, because this is a subset of the experiment data, we will need to induce balance on covariates rather than rely on randomized treatment assignment. We add interactions and transformations of existing terms to further increase the similarity in distributions between treated and control groups. In addition, we stratify on gender so that there is perfect balance in it for each of 10 subclasses, 5 for each gender stratum. Propensity scores are calculated separately for female and male subsets.

Upon establishing the design phase, we have a total of 10 subclasses and calculate the estimated average causal effect within each subclass, then weigh results by the proportion of the study population within the subclass to obtain a weighted point estimate. We use the same structure to find the weighted variance used to create 95% confidence intervals surrounding the point estimate.

We apply the design first to the analysis of the effect of assignment to the job training program, then repeat for the effect of actually receiving the training. The latter is assumed to be synonymous to the complier average causal effect, assuming exclusion restriction, where assignment to the job training program itself does not affect wages four years from assignment. We therefore can only infer the effect of receiving treatment for those who take it only when assigned to it. We cannot estimate the effect for noncompliers who would never actually receive treatment, which is why we use the $CACE$ estimate.

Section 1: Treatment Assigned

1. Design Stage

1.1 Prior to Subclassification using Propensity Scores

First we examine the covariate balance in the simplest design-comparison between the treatment assigned and the treatment not assigned (control) groups. We look at the distributions of the covariates in each group and see if they are balanced, or comparable to what would be observed if units had been randomly assigned to each group.

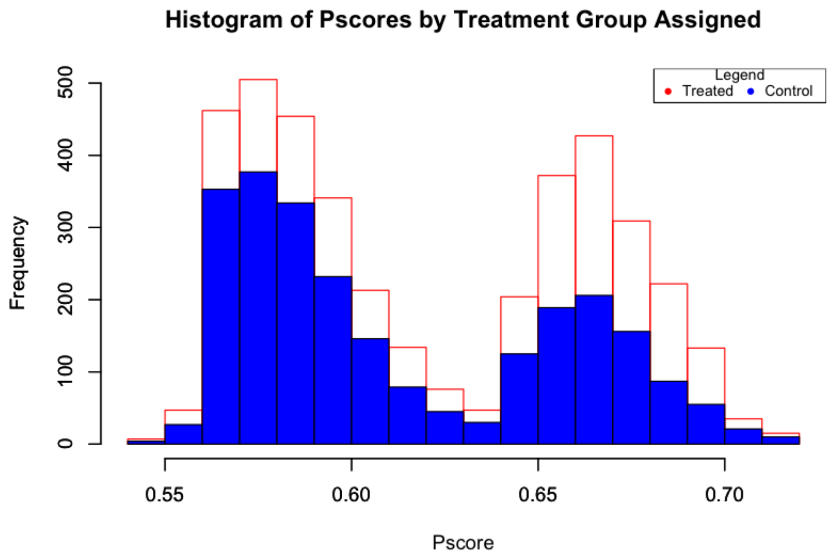
Table 1.1 Covariate means within each treatment assigned group

	Experiment	
	$\bar{X}_T$	$\bar{X}_C$
Female	0.44	0.35
Age	19.12	18.96
Has Child	0.20	0.18
Education	0.46	0.43
Employed	0.26	0.25
Months in Job	4.58	4.51
(Stand) Yrly Earn	0.12	0.11
White	0.3	0.31
Partnered	0.069	0.067

The balance is not too bad, although there is a higher proportion of females in the treated group than in the control group.

Gender could be an important covariate in this study since gender wage equality is often an issue. Mean age is also a bit higher in the treated group.

Graph 1.1 Histogram of propensity scores by assigned treatment



We generated propensity scores for all units in this population by using a logistic regression with all of the identified covariates above on the treatment assigned indicator, without additional interaction terms, then doing a t-test.
It is very apparent that there are two modes in the distribution of the propensity scores for both the treated and the control groups. This suggests that there might be two general types of individuals in the experiment coded for by an indicator covariate. However the balance seems decent between the treated and control groups since there is a similar distribution of propensity scores for both groups.

1.2 Subclassification of covariates using propensity scores and quintiles:

Now we use the propensity scores previously obtained for all units to generate quantile cut-offs, then subclassify all 6479 units into 5 subclasses.

Table 1.21: Units in Subclasses

Subclass	# Treated Units	# Control Units
1	734	562
2	752	545
3	787	508
4	846	450
5	885	411
Total	4003	2476

There is a higher proportion of treated units that control units, so it is likely that treatment was not assigned with a .5 probability.

Table 1.22 Covariate Means within Subclasses

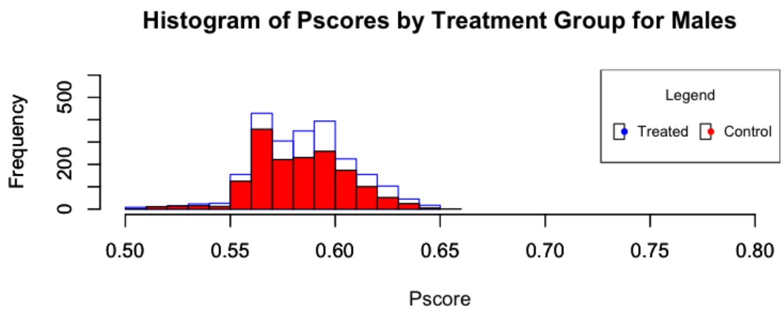
Subclass:	1		2		3		4		5	
	$\bar{X_T}$	$\bar{X_C}$	$\bar{X_T}$	$\bar{X_C}$	$\bar{X_T}$	$\bar{X_C}$	$\bar{X_T}$	$\bar{X_C}$	$\bar{X_T}$	$\bar{X_C}$
Female	0	0	0	0	0.23	0.2	1	1	1	1
Age	17	17	18.5	18.6	21.2	21.1	17.7	17.7	20.9	20.7
Has Child	0.11	0.12	0.1	0.086	0.13	0.14	0.3	0.3	0.3	0.3
Education	0.14	0.15	0.41	0.41	0.67	0.64	0.3	0.3	0.7	0.7
Employed	0.037	0.028	0.32	0.28	0.42	0.43	0.1	0.1	0.4	0.4
Months in Job	2.9	2.8	5.1	4.7	6.02	6.24	3.3	3.3	5.5	5.7
(Stand) Yrly Earn	-0.22	-0.18	0.24	0.14	0.57	0.57	-0.2	-0.2	0.2	0.3
White	0.32	0.32	0.36	0.36	0.35	0.39	0.2	0.2	0.25	0.23
Partnered	0.055	0.071	0.045	0.023	0.067	0.092	0.07	0.08	0.1	0.05

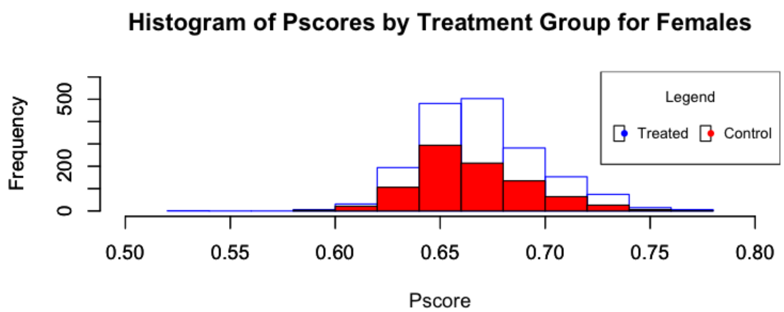
From Table 1.22, we can that there are no females in subclasses 1-2 and no males subclasses 4-5. This suggests that females have a higher probability of being assigned to take the job training program than males. It seems reasonable to stratify on gender since otherwise we would obtain estimates in all but subclass 3 that would account for only one gender. The resulting weighted average effect of treatment assigned would not be very robust when generalizing to the sample population regardless of gender. Therefore we will condition on gender to absolve this problem.

1.3 Stratifying/Conditioning on Gender

10 subclasses are created by inducing perfect balance in gender. Subsets of the sample population are made- one stratum for males and one stratum for females.

Graph 1.31: Assessing Balance between Treatment Groups, no additional covariate terms





Now the difference in the distribution of pscores by gender becomes more apparent.
We normally discard control units whose propensity score range falls outside of that of the range of pscore values for the treated units in each stratum. Such units would hypothetically never receive treatment since we assume that this set of covariates summarized by the pscore predict assignment to treatment. However, upon testing the range of pscores, we find that only 2 units in the male stratum would be discarded from the control group out of 1606 (and 0 in the female stratum) and conclude that the difference would be negligible if we did discard these units. Since our pscore estimates are imprecise, these small variations could be due to errors in calculation.

1.4 Creating Subclasses

Five subclasses were created for each gender stratum; the dataset was divided into male and female subgroups, then pscores calculated for each stratum separately. The subclasses 1-5 are for males and 6-10 correspond to the five subclasses for females.

Table 1.41: Number of Units in each Subclass

	Males		
Subclass	# Treated Units	# Control Units	Total
1	432	341	773
2	425	347	772
3	468	305	773
4	462	310	772
5	470	303	773
	Females		
	# Treated Units	# Control Units	
6	468	267	755
7	245	157	402
8	293	115	408
9	210	99	309
10	530	232	762

Note that the proportion of treated to untreated units varies little between subclasses, but proportion of population in subclass 6 and 10 are significantly higher than proportions in subclasses 7-9.

By substratifying on gender, we achieve perfect balance on it. We weigh the averages of the subclasses by the population in each subclass rather than weighting by the proportion of treated units in the subclass because we want to generalize this study to the sample population, including the units under control that could potentially receive treatment based on their covariates. We assume that all units in the sample population could potentially be assigned treatment; this makes sense because they were all in an experiment where treatment was randomized.

Table 1.42: Covariate Means for Male Stratum, Subclasses 1-5

Subclass:	1		2		3		4		5		Wtd	
	$\bar{X}_T$	$\bar{X}_C$	$\bar{X}_T$	$\bar{X}_C$	$\bar{X}_T$	$\bar{X}_C$	$\bar{X}_T$	$\bar{X}_C$	$\bar{X}_T$	$\bar{X}_C$	$\bar{X}_T$	$\bar{X}_C$
Age	17.5	17.5	17.6	17.8	18.9	19.1	19.7	19.6	20.1	20.8	19	18.9
Has Child	.1	.1	.05	.05	.08	.09	.08	.08	.2	.2	.1	.1
Education	.1	.1	.03	.04	.5	.5	.6	.6	.8	.8	.4	.4
Employed	.03	.04	.01	.03	.2	.1	.3	.4	.7	.7	.3	.3
Months in Job	2.4	2.7	2.9	2.6	4.4	4	5.5	5.9	8.1	8.1	4.6	4.7
(Stand) Yrly Earn	-.2	-.2	-.2	-.2	.07	.005	.3	.4	1.1	1	.2	.2
White	.6	.6	.2	.2	.45	.43	.3	.3	.2	.2	.3	.4
Partnered	.2	.2	.02	.03	.02	.03	.004	.01	0	0	.06	.05

From the covariate means within subclasses for each treatment assigned group, which we assume are good indicators of the covariate distributions due to the Central Limit Theorem, we infer that there is decent balance between treatment groups in most covariates in each subclass.

We observe some interesting trends: the proportion of partnered individuals decreases as the propensity score rises for males, suggesting that males without partners were more likely to be assigned to the job training program.
Those with higher standardized yearly earnings, age, education, were employed, and held their job for more months were also more likely to be assigned treatment. These observations seem to suggest that the assignment to this program favored those who were already better off.
There seems to be two modes for which there is a high proportion of white males- subclass 1 has means of 0.6, subclass 2 is significantly less with 0.2, then subclass 3 has ~.45, followed by a gradual decrease again.
It is difficult to identify trends relating to whether the male individual has children.

*Interaction and transformation terms were not added in this design but were later done, results shown further below.

Table 1.43: Covariate Means for Female Stratum, Subclasses 6-10

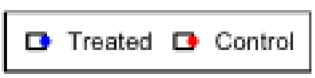
Subclass:	6		7		8		9		10		Wtd	
	X_T	X_C	X_T	X_C	X_T	X_C	X_T	X_C	X_T	X_C	X_T	X_C
Age	17.7	17.7	18.3	18.2	18.7	18.7	19.3	19.7	21.4	21.2	19	18.9
Has Child	.4	.4	.2	.2	.3	.2	.2	.3	.4	.4	.1	.3
Education	.5	.5	.5	.5	.5	.5	.5	.6	.6	.5	.4	.5
Employed	.4	.3	.2	.3	.2	.2	.2	.3	.2	.2	.3	.2
Months in Job	4.7	4.9	4.4	4.4	4	4	4	4.5	4.5	4.1	4.6	4.4
(Stand) Yrly Earn	0.1	0.2	-.04	-.03	-.07	-.05	-.1	-.04	.06	-.08	.2	.02
White	0.02	0.04	.1	.1	.3	.3	.4	.3	.4	.4	.3	.2
Partnered	0.002	0.01	.01	.03	.02	.03	.05	.08	.3	.2	.06	.08

From the covariate means, we can infer that the distributions between treatment groups for most covariates are decent in the subclasses.

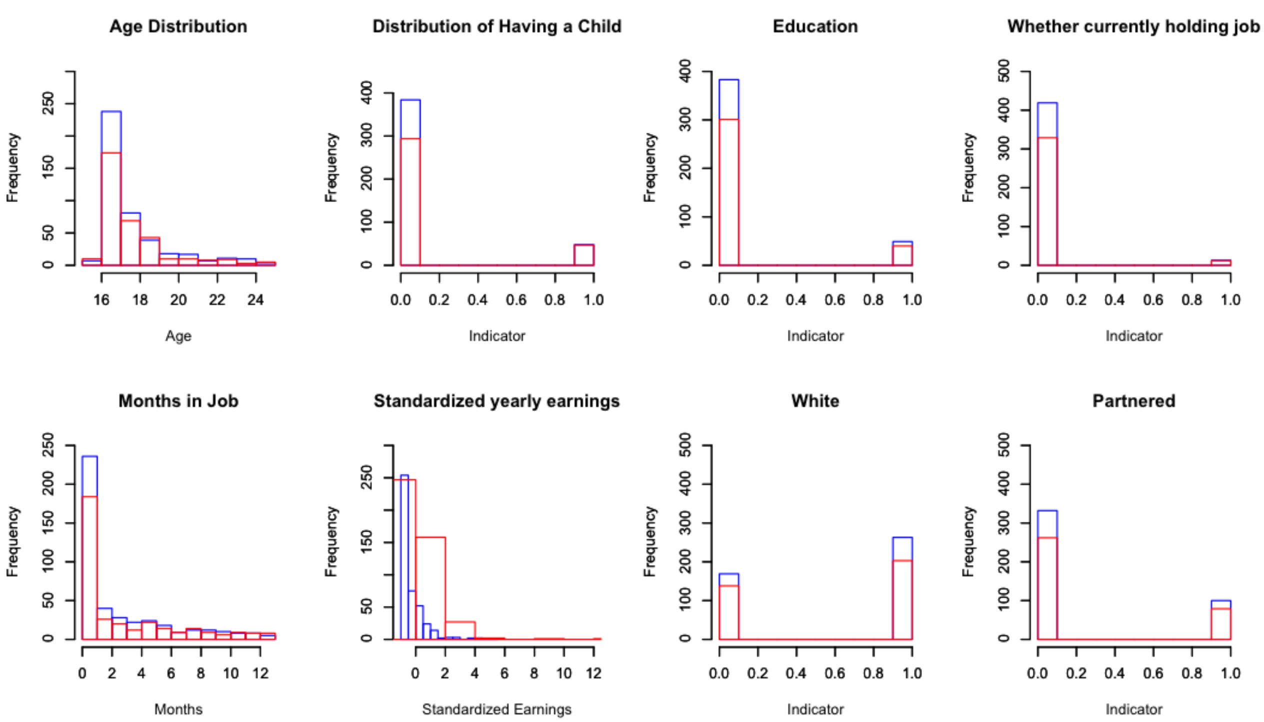
Some interesting trends: in contrast with the trend for males, the proportion of partnered individuals increases with the propensity score for females, suggesting that females with partners were more likely to be assigned to the job training program.
Those with higher age and proportion white were also more likely to be assigned treatment.
Proportion of females with children rises at the higher and lower ends of the propensity score range.
Mean months in job and proportion employed at the time of assignment to treatment are relatively consistent between subclasses.
Standardized yearly earnings could be better balanced within subclasses.

Table 1.44: Histograms comparing covariate distributions for subclasses 1-10

Legend

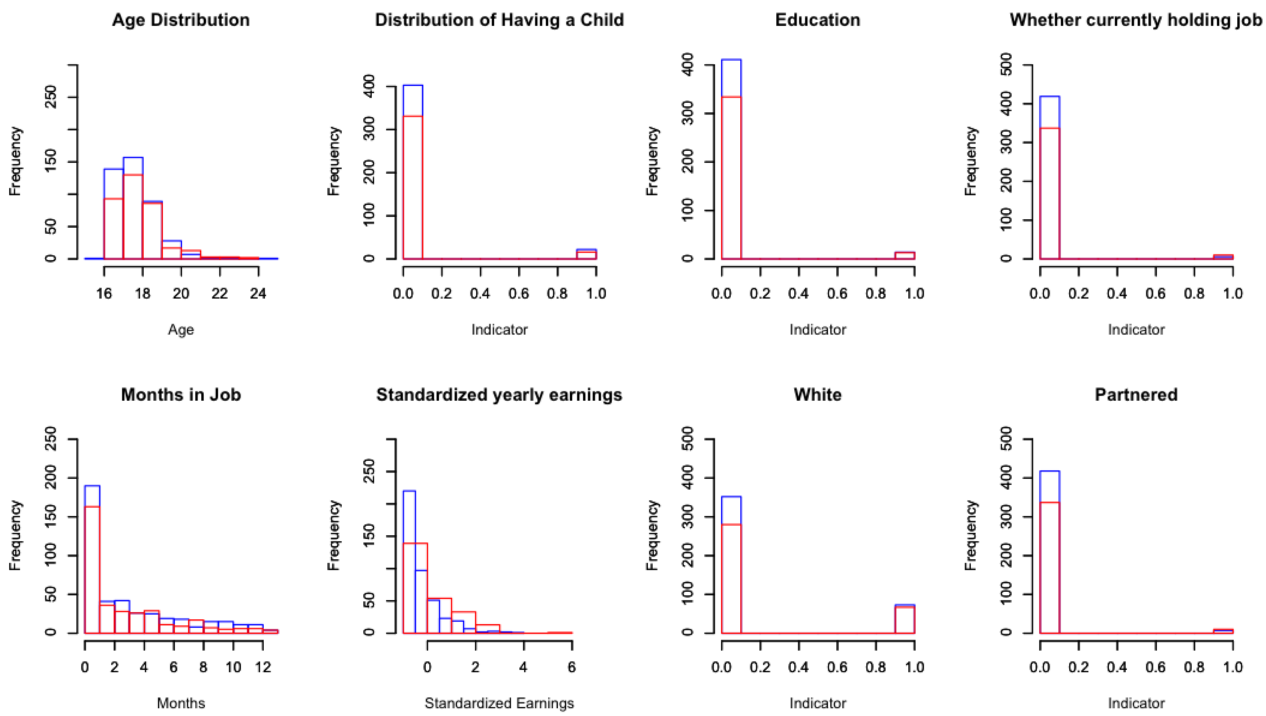


Subclass 1:

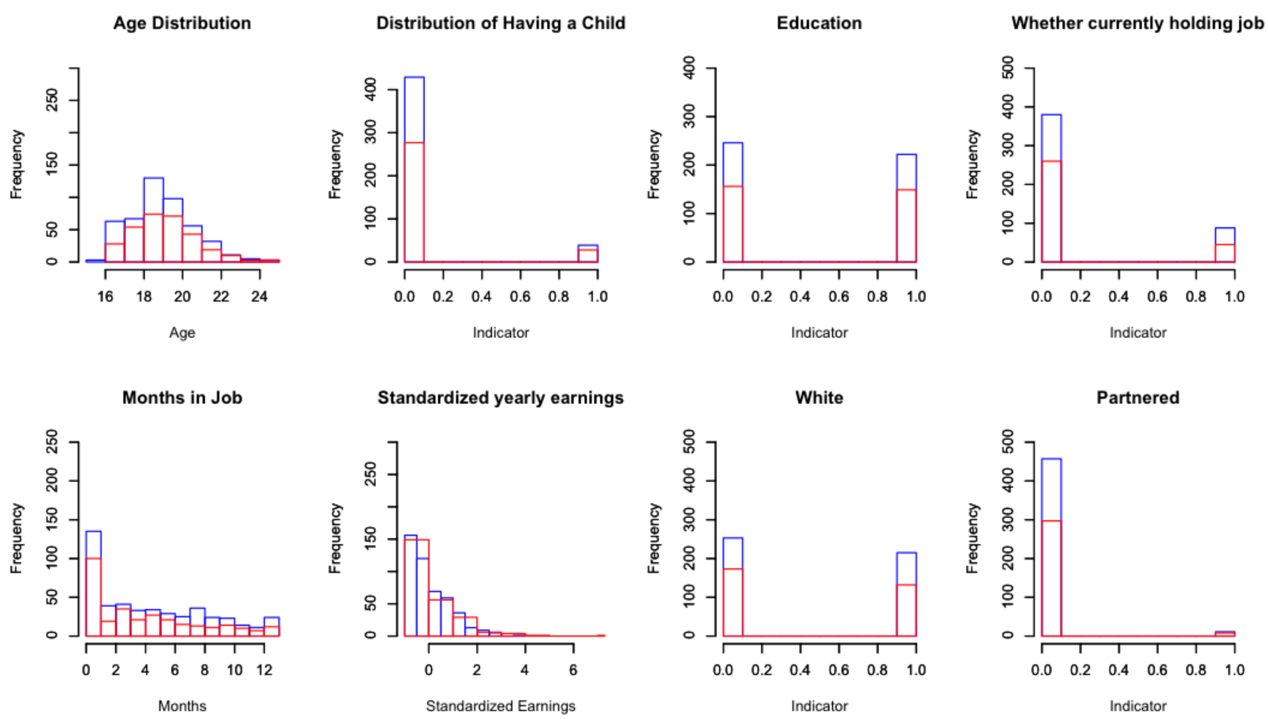


Balance between treated and control groups for yearly earnings is not very good- values for the latter

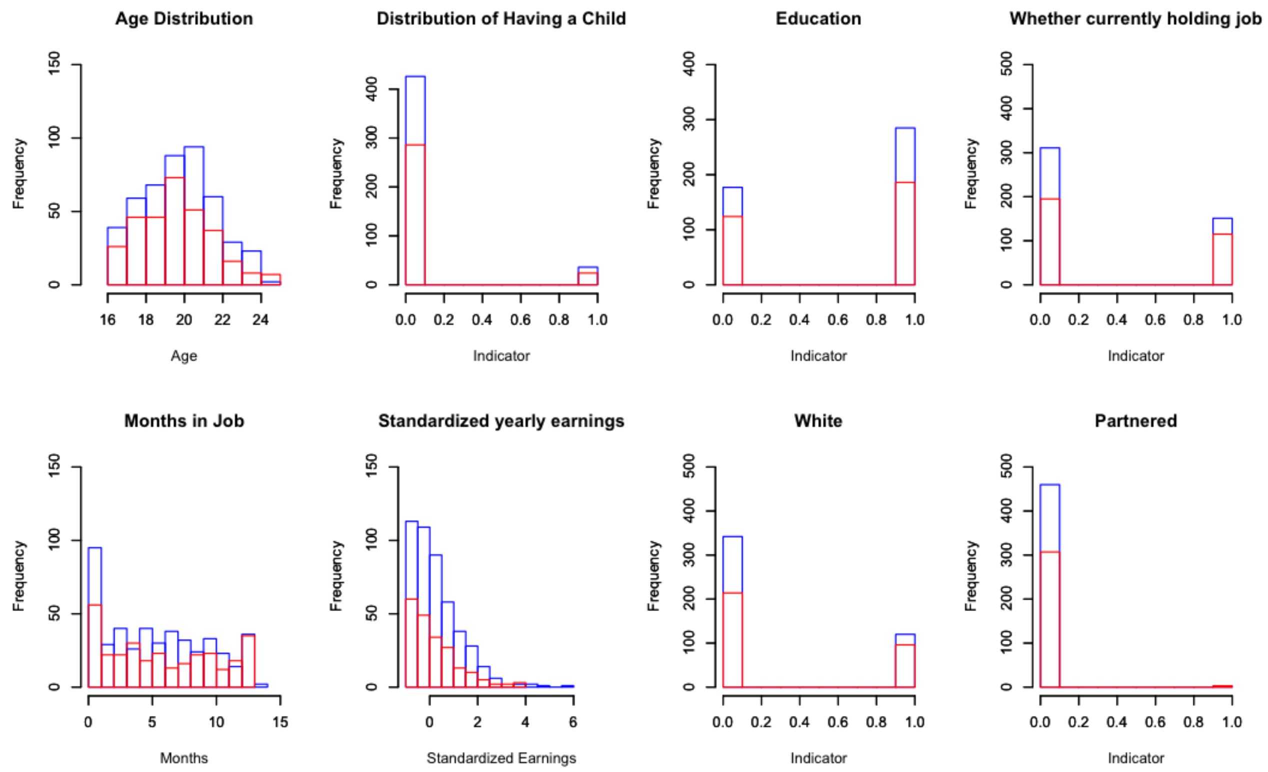
Subclass 2:



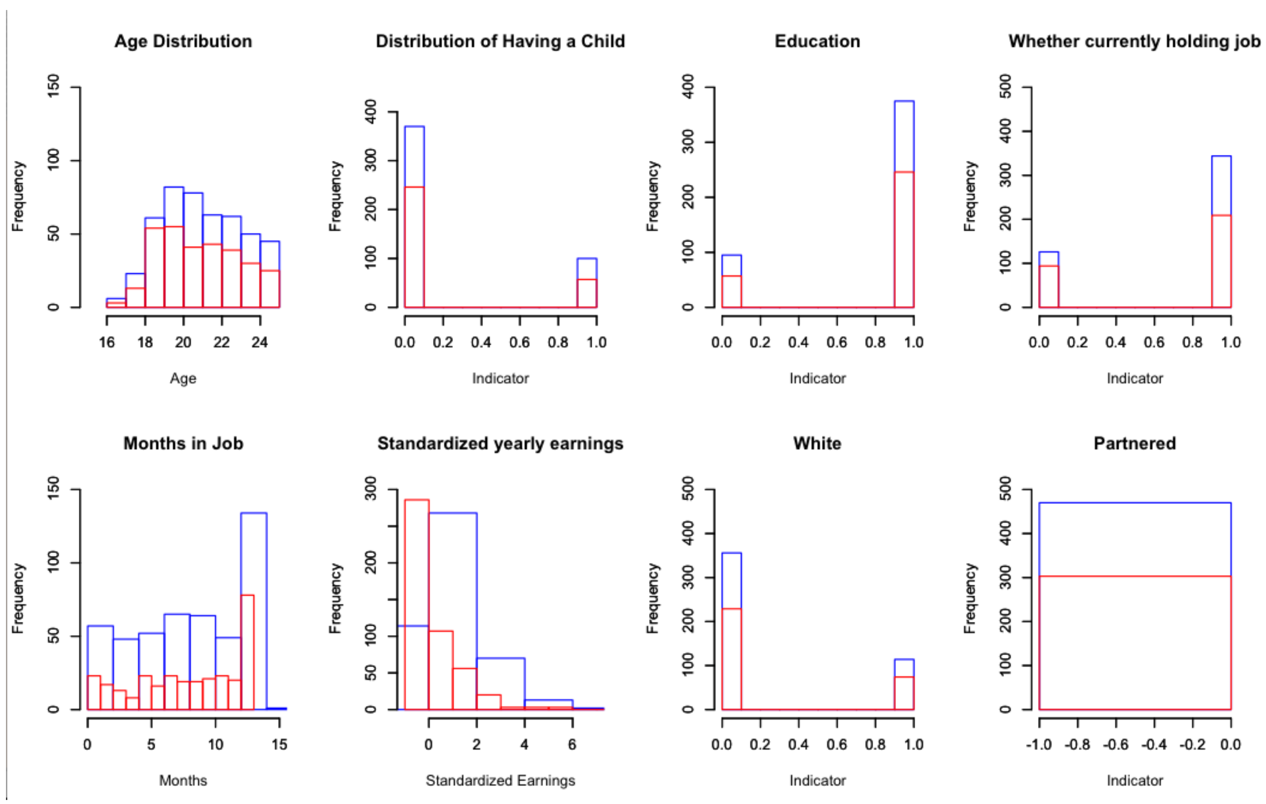
Subclass 3:



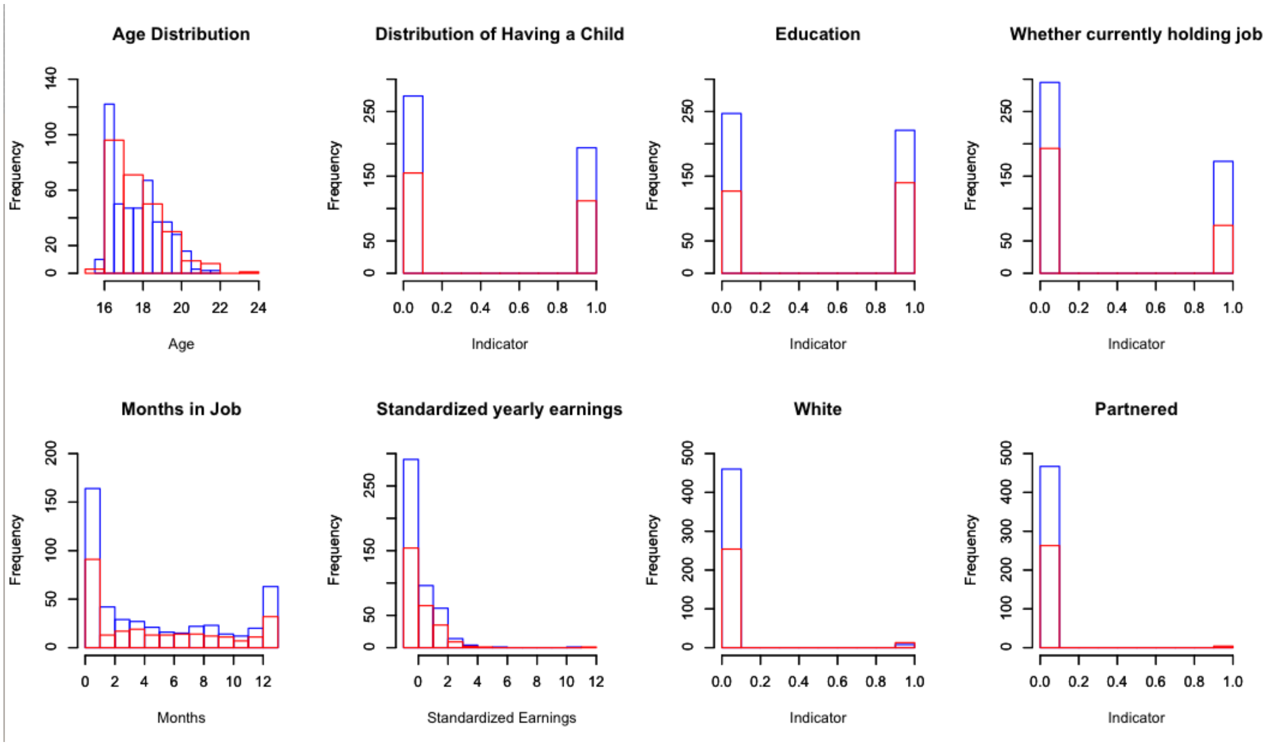
Subclass 4:



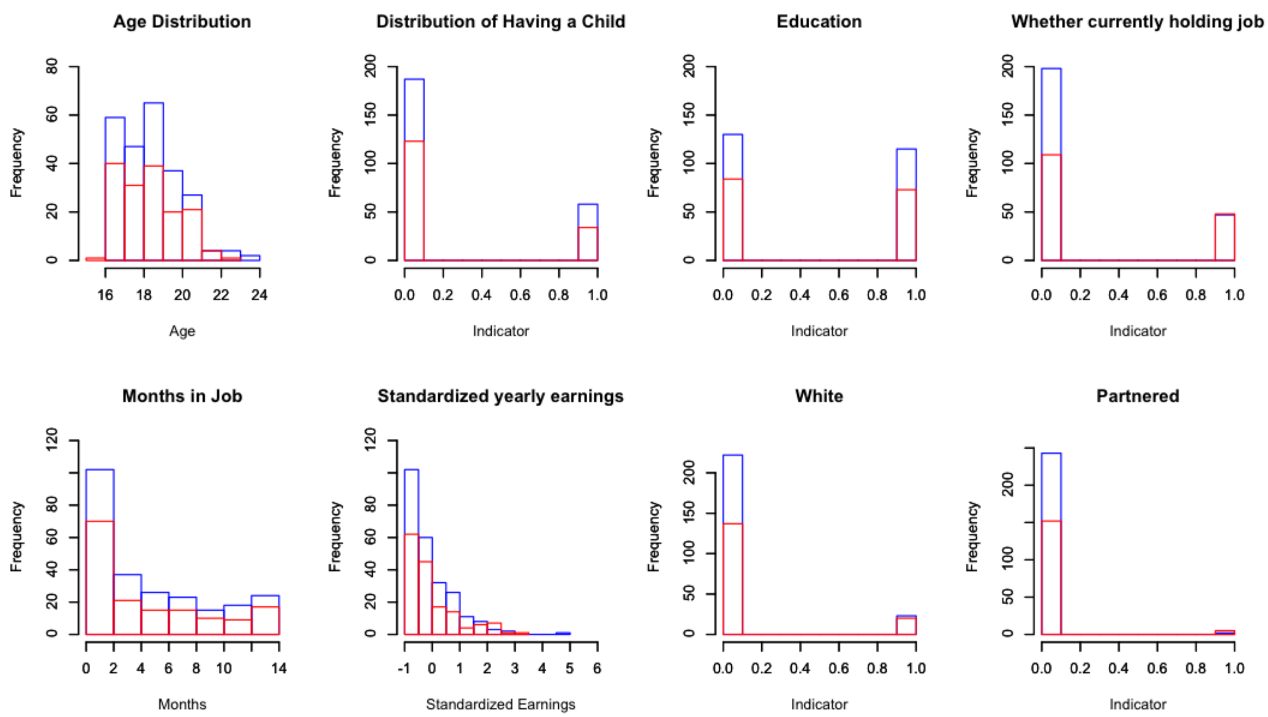
Subclass 5:



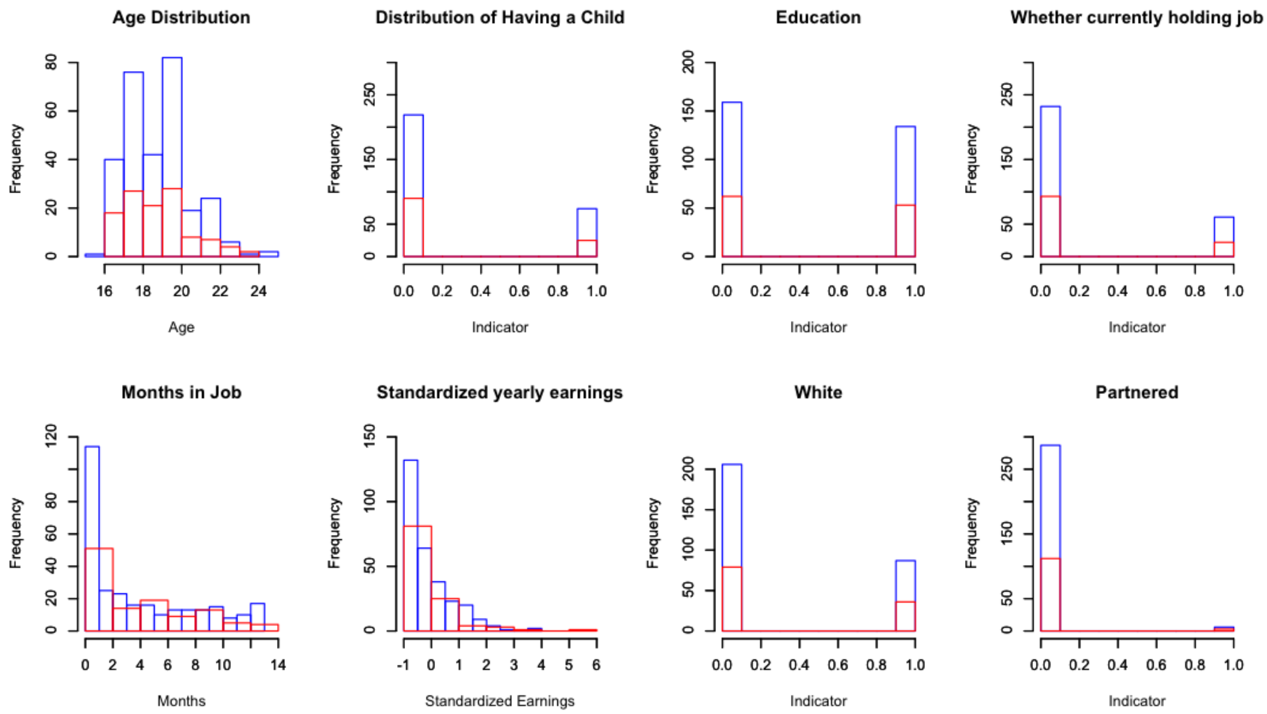
Subclass 6:



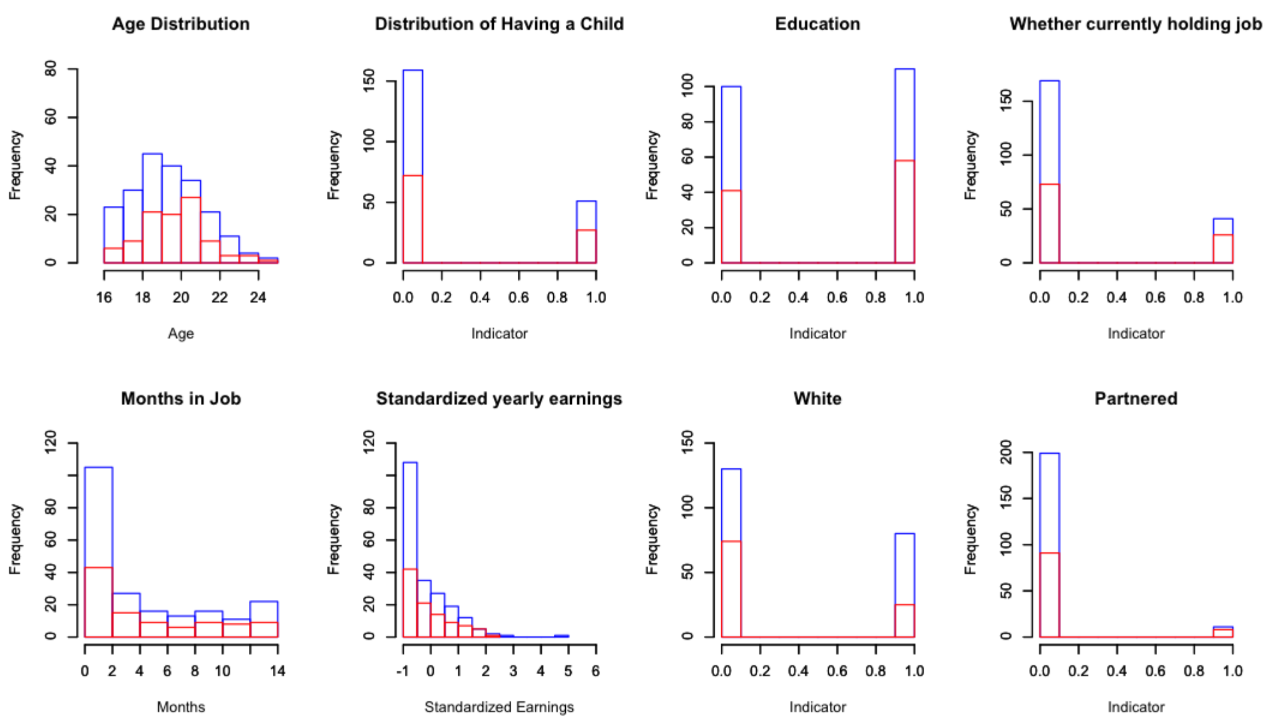
Subclass 7:



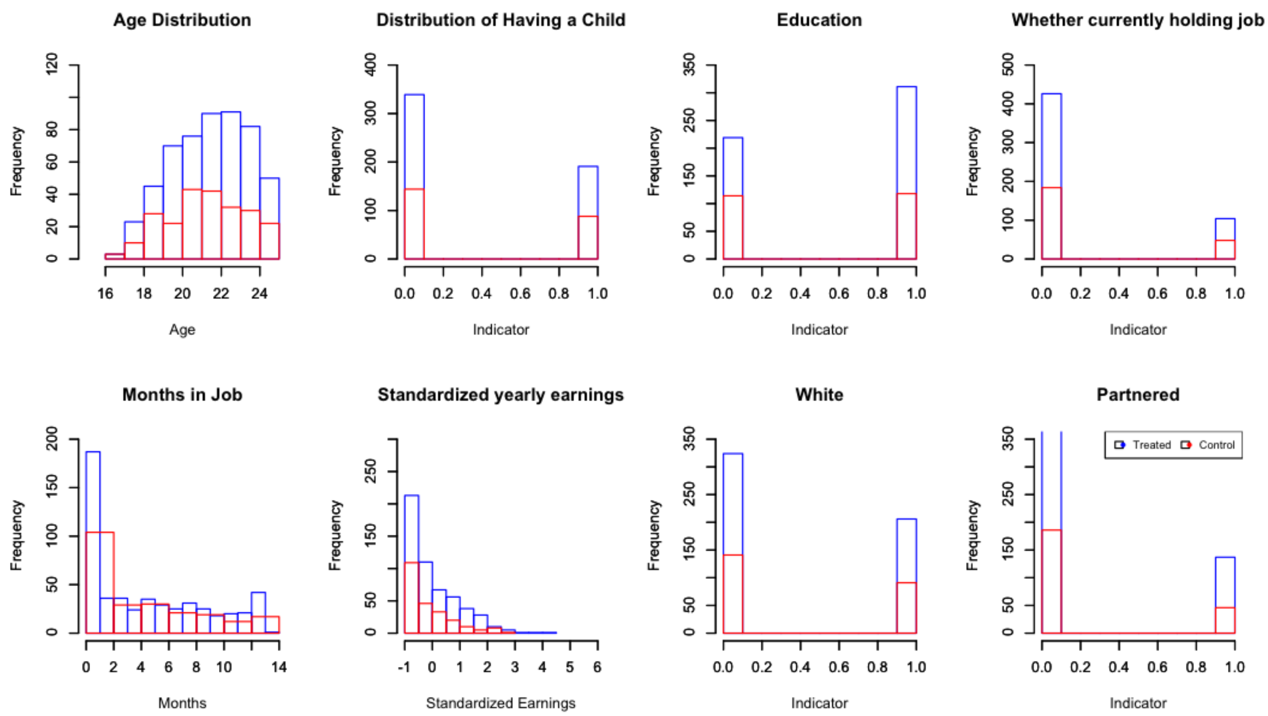
Subclass 8:



Subclass 9:



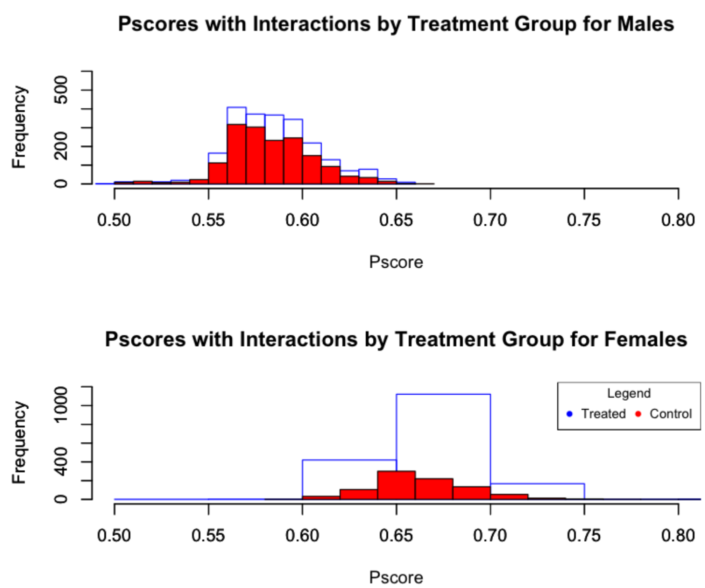
Subclass 10:



1.5 Further reducing bias due to covariates

We attempt to further improve the design by adding additional interactions.
We first added two interaction variables, including the second order interaction of the indicator of whether currently holding a job at time of assignment to the study with months in job since these two should be related. The other added term is the third order interaction of the indicator for whether the unit is partnered with the indicator for whether white and the education term.
However, balance actually decreased, so upon testing a few other interactions and transformations, we introduced only the square root of months in job and the square root of age for the best balance so far.

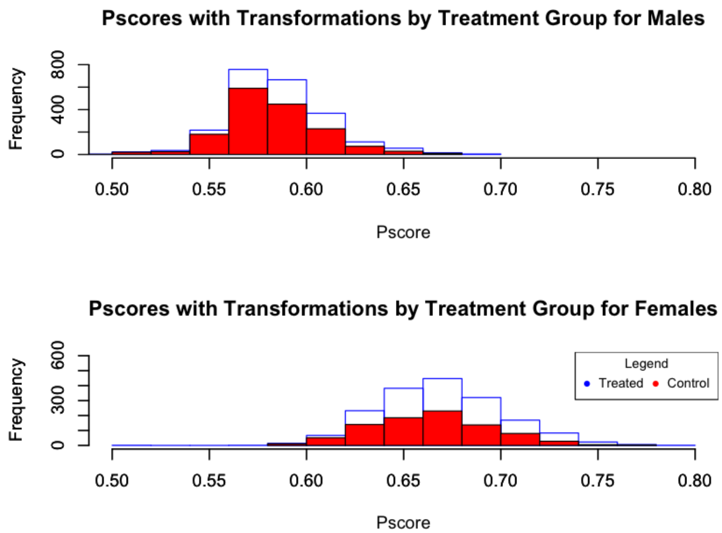
Graph 1.51: Observing Balance between Treatment Groups with Interaction Terms



Upon testing these interactions and comparing them with Graph 1.31 without interaction terms, the balance is worsened particularly in the female stratum, perhaps because there are too many 0's in the job interaction terms as well as the indicators that result in less useful covariate values with modes at 0. Instead we more conservatively add the terms sqrt(months in job) and sqrt(age).
We are careful not to do any transformations with the variable that would change the sign of the values or create

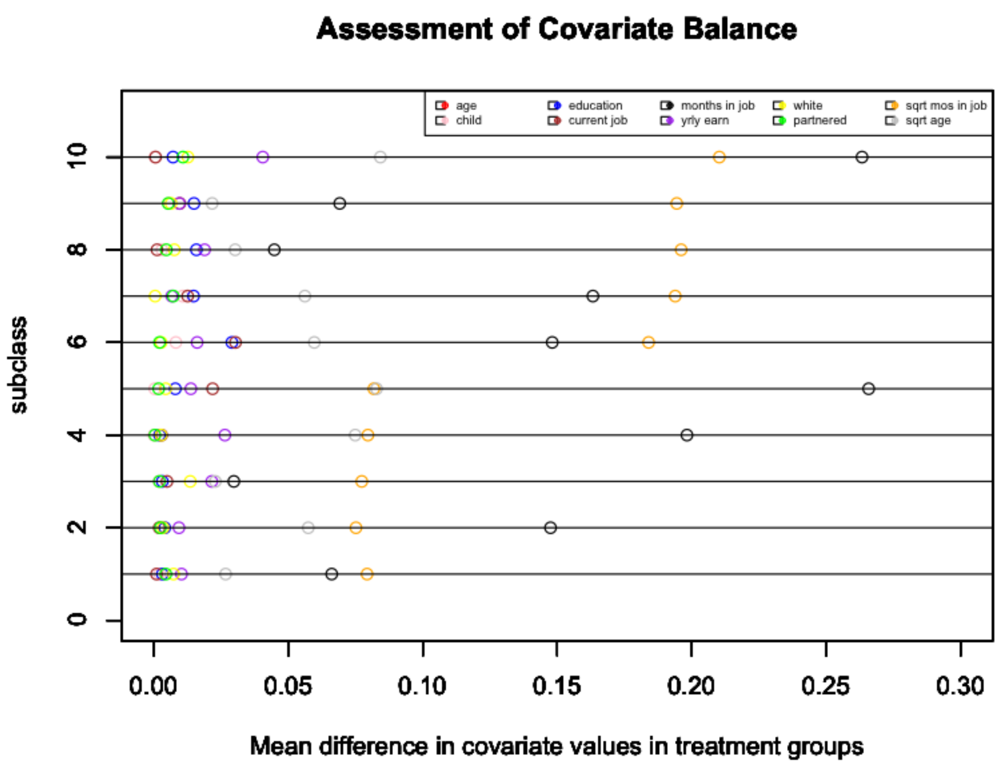
undefined terms due to 0's in the covariate values.

Graph 1.52



The distributions of pscores in the treated and control groups now appears more similar and Normal than before, suggesting better balance.

Graph 1.53: Covariate Balance with Sqrt Terms using Subclass Mean Differences



Balance is fairly good, with the absolute mean difference between treated and control groups clustered around zero, a few outliers but all within the 0-0.3 difference range. Values are not normalized, so it makes sense that the differences for the covariate months in job and square root of months in job would be farthest from 0. We also see that square root of months in job is closer to 0 for subclasses 1-5 for males than for subclasses 6-10 which consist of solely females.

1.6 Results

We use the 10 subclasses created in the design phase to estimate an unbiased answer for the causal effect of *assignment* *to treatment* on the earnings four years after assignment. Because we want to generalize our findings to all units in this study's population, we weight by the proportion of population (6479) in each subclass rather than by proportion treated.

We obtained $I\hat{T}T$ by calculating the Neyman average weighted by the proportion of the sample population in each subclass.
We obtain the variance for $I\hat{T}T$ by applying the same weights, but this time squared since $Var(\sum_s w_s(\bar{Y_{sT}} - \bar{Y_{sC}})) = \sum_s w_s^2 * Var((\bar{Y_{sT}} - \bar{Y_{sC}})) \approxeq \sum_s w_s^2 * (\frac{\sigma_{sT}^2}{N_{sT}} + \frac{\sigma_{sC}^2}{N_{sC}})$ so we will get a conservative estimate of the variance since we disregard the covariance term.

Table 1.61: Estimated ITT within each subclass k

Subclass k	$ITT_{Y,k}$
1	34.8
2	20.9
3	17.6
4	10.5
5	11.8
6	26.6
7	25
8	12.6
9	5.4
10	21.3

Note that the estimated mean difference in the observed outcome, wages, generally decreases as propensity score increases for males. There is less of a general trend for females.

Table 1.62: Estimated Weighted Effect of Intention to Treat on Wages

(w/ transformations)	Point Estimate	95% Interval
$I\hat{T}T_Y$	**18.7**	**(8.1, 29.4)**

We obtain our point estimate of the effect of assignment to the job training program. The confidence interval is comfortably in the positive range, not near 0, so one could infer that the ITT_Y is positive on weekly earnings four years from assignment to treatment. On average, assignment to treatment has this positive effect on the individuals in this study population and can be generalized to those with covariate values that fall within the range of those in the study population.

Section 2: Treatment Received

2.1 Implementing the design

2.11 Finding the Complier Average Causal Effect

We now use the same subclasses created Section 1: Treatment Assigned to estimate the complier average causal effect of receiving treatment. We use this estimate because we can only find the effect of receiving treatment for the units who have the possibility of both receiving and not receiving the treatment. It would be impossible to estimate this difference in wages for someone who would always receive training. Therefore, within each subclass, we calculate $\hat{p_c}$, the proportion of compliers, $\hat{p_N}$, the proportion of never-takers, and $\hat{p_A}$, the proportion of always-takers.
We find that of those individuals i in our study population who were not assigned treatment ($Z_i = 0$), none received the treatment ($W_i = 1$), so we only have one-sided noncompliance, i.e. $\hat{p_A} = 0$. The monotonicity assumption is therefore unnecessary because we can have no defiers.

For each subclass k, we know whether units i were assigned and/or received treatment. We know that $\hat{p_N}^k$ = (# units who were assigned treatment but didn't receive it)/(#units in subclass) and $\hat{p_C^k} = 1 - \hat{p_N}^k - \hat{p_A}^k = 1 - \hat{p_N}^k$.
Within each subclass, $C\hat{A}CE^k = \frac{I\hat{T}T_y^k}{\hat{P_c}^k}$. To obtain an overall weighted point estimate, we use the weights previously used in Section 1.
We use the delta method to calculate the variance for $C\hat{A}CE^k = \frac{I\hat{T}T_y^k}{\hat{P_c}^k}$, and again we use the weights from Section 1 to get an overall weighted variance.
We Taylor approximate $C\hat{A}CE$ for each subclass:

$X = I\hat{T}T_Y$, $Y = \hat{P_C}$. We already have point estimates for all random variables $C\hat{A}CE$, $I\hat{T}T_Y$, $\hat{P_C}$, and name them constants c, a, b, respectively.
$f(x,y) = C\hat{A}CE = \frac{X}{Y} \approx c + \frac{df}{dx}|_{x=a,\,y=b} * (x-a) + \frac{df}{dy}|_{x=a,\,y=b} * (y-b)$. This simplifies to
$f(x,y) \simeq c + \frac{1}{b}(x-a) - \frac{a}{b^2}(y-b) = c + \frac{1}{b}x - \frac{a}{b^2}y$

$Var(C\hat{A}CE^k) \simeq Var(c + \frac{1}{b}x - \frac{a}{b^2}y) = \frac{1}{b^2}Var(x) + \frac{a^2}{b^4} * Var(y) - (\frac{2a}{b^3})Cov(x,y)$,
where $Cov(x,y) = Cov(\bar{Y}_1^{obs} - \bar{Y}_0^{obs}, \bar{W}_t - \bar{W}_c) = Cov(\bar{Y}_1^{obs}, \bar{W}_t) + Cov(\bar{Y}_0^{obs}, \bar{W}_c) =$
$\frac{1}{N_t(N_t-1)}\sum_{i:Z_i=1}(Y_i^{obs} - \bar{Y}_1^{obs}) * (W_i^{obs} - \bar{W}_t^{obs}) + \frac{1}{N_c(N_c-1)}\sum_{i:Z_i=0}(Y_i^{obs} - \bar{Y}_c^{obs}) * (W_i^{obs} - \bar{W}_c^{obs})$

and $Var(x) = \frac{s_c^2}{N_C} + \frac{s_T^2}{N_T}$, $Var(y) = \frac{\sigma_c^2}{N_C} + \frac{\sigma_T^2}{N_T}$, where s^2 is the sample variance when the assigned treatment is the column corresponding to the earnings outcome and σ^2 for when this is true for received treatment.

2.5 Results

2.51 Estimates of CACE in every subclass

Subclass k	$\hat{P}_{Ck}$	$C\hat{A}CE_k$
1	.88	39.6
2	.87	24
3	.86	20.5
4	.87	12.1
5	.84	14.02
6	.89	29.9
7	.87	28.6
8	.86	14.6
9	.84	6.4
10	.83	25.6

As expected, the $C\hat{A}CE$ within each subclass is slightly higher than its $I\hat{T}T_Y$. The proportion of compliers in each subclass is also relatively consistent and does not increase with the likelihood of assignment to treatment, reflected by the propensity score. This makes sense because compliance type is latent and generally unobservable by those assigning people to the job training program.

2.52 Estimate of causal effect of Treatment Received:

Table 2.52

	Point Estimate	95% Interval
$C\hat{A}CE$	**21.6**	**(-17.2, 60.4)**

2.53 Analysis of Results

The estimated effect of actually receiving job training on the weekly wage four years after the start of the program is $21.6 - 18.7 \simeq 3.9$ higher than the $I\hat{T}T_Y$ point estimate. This slightly higher $CACE$ estimate makes sense, since $C\hat{A}CE = \frac{I\hat{T}T_Y}{P_C}$ with our previous assumption of exclusion restriction, and p_c within subclasses is generally quite high around ~.85 so that the estimate is not significantly higher than the ITT_Y estimate.
However, the variance of the $CACE$ estimate is considerably higher than that of the ITT_y estimate and the former's confidence interval spans zero. The high weighted $var(C\hat{A}CE)$ of *391.6* is partly due to the range of values [0, 60307.9] taken by the outcome of weekly earnings four years after assignment to treatment. The method-of-moments estimate is also unstable and sensitive to the estimated proportion of compliers.
Therefore if we heavily consider our obtained interval, we cannot fully accept that the effect of treatment received is positive. However, given our observed proportion of compliers within each subclass, this $CACE$ estimate should be higher than the ITT estimate, which had a low variance and a small confidence interval above 0. We should be able to infer the effect of treatment received from this $CACE$ point estimate for these reasons.

Conclusion

We saw various trends for how covariates are related the likelihood of being assigned treatment from to their relationship

with the propensity scores. These observations were often different for those of different genders. The program seemed to target young females more than young males, according to the farther right-centered propensity scores of the female stratum.
Our findings can be generalized to populations with similar covariates as those in our study: we would do so by comparing their propensity score with the range of propensity scores in the study population, conditioning on gender.

If we accept that the effect of assigning an individual in this study population to the job training program is noticeably positive on weekly earnings four years after being assigned to do the program, as our estimates suggest, we should similarly be able to infer that the effect of actually receiving the training program is positive and slightly higher than the effect of being assigned to take the program. This is assuming that we believe that exclusion restriction holds, where being assigned to take the program has no effect in itself. Due to the large sample size and small range of propensity scores, our conclusion should be relatively robust.