

Worked Example
NHANES data

- ▶ How can we fit a propensity score?
- ▶ How can we check balance?
- ▶ How can we fit an ATE or ATT?

1. Fitting a propensity score

- ▶ Logistic regression
- ▶ Checking balance
- ▶ Alternatives to logistic regression

2. Fitting an ATE

- ▶ Traditional regression (G-computation)
- ▶ PS stratification
- ▶ PS matching
- ▶ PS regression
- ▶ inverse probability of treatment weighting

3. Fitting an ATT

- ▶ PS matching
- ▶ inverse probability of treatment weighting

- ▶ In PS-based methods, the goal of the treatment model is to eliminate imbalance in the distribution of covariates between treated and untreated subjects.
 - ▶ Achieving balance on other covariates (particularly strong predictors of treatment) is unhelpful.
- ▶ The goal is *not* to build an excellent predictive model for the treatment.
- ▶ Some methods to avoid:
 - ▶ C-statistic (AUC),
 - ▶ significance tests.

Common measures of balance:

- ▶ Standardized mean difference or proportion:

$$\frac{\bar{x}^{1,w} - \bar{x}^{0,w}}{\sqrt{0.5(v^{1,w} + v^{0,w})}}$$

where

$$\bar{x}^{z,w} = \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{1}_z(Z_i) X_i}{f_{Z|X}^O(Z_i|X_i)},$$

i.e. the weighted sample mean of variable X among those with treatment value z , and similarly $v^{z,w}$ is the weighted variance estimate.

- ▶ For all methods of analysis other than IPW, the weights are taken to be 1 for all subjects.
- ▶ SMD of 0.1 or less typically considered reasonable.
- ▶ Visual examination of weighted empirical CDFs among the treated and untreated (for binary or categorical treatment).

Assessing balance: example

- ▶ In this example, we will explore propensity score based analyses using the publicly available (U.S.) National Health and Nutrition Examination Survey (NHANES). For this, may use `NHANES`, `tableone`, and `Matching` in R.
- ▶ We will focus our analysis on the question of whether current smoking affects average systolic blood pressure. The variables we will need are: `BPSysAve`, `SmokeNow`, `Gender`, `Age`, `Race3`, `Education`, `MaritalStatus`, and `Poverty` where the first two are the outcome and exposure of interest and the remaining are potential confounders.
- ▶ Additionally, we will restrict our attention to adults (> 17 years old) in the second wave of the survey.

Assessing balance: example

```
> library(NHANES)
> library(tableone)
> library(Matching)
>
> NHANES$SmokeNow <- as.numeric(NHANES$SmokeNow)-1
> small.nhanes <- na.omit(NHANES[NHANES$SurveyYr=="2011_12"
  & NHANES$Age > 17,c(3,4,8:11,13,25,61)])
> dim(small.nhanes) ## 1377
>
> vars <- c("Gender", "Age", "Race3", "Education",
  "MaritalStatus", "Poverty")
> tabUnmatched <- CreateTableOne(vars = vars,
  strata = "SmokeNow", data = small.nhanes,
  test = FALSE)
```

Assessing balance: example

```
> print(tabUnmatched, smd = TRUE)
```

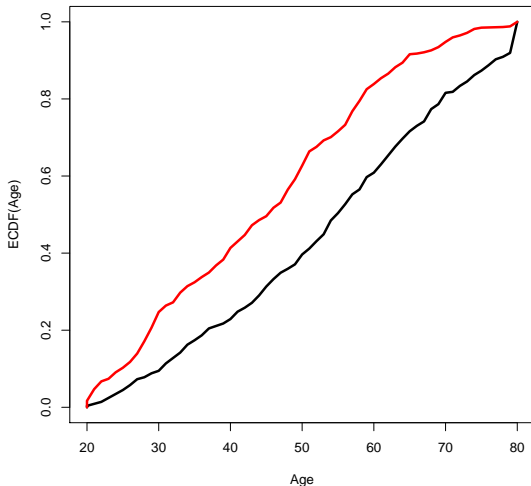
	Stratified by SmokeNow		
	0	1	SMD
n	782	595	
Gender = male (%)	432 (55.2)	369 (62.0)	0.138
Age (mean (sd))	54.33 (16.52)	44.96 (15.11)	0.592
Race3 (%)			0.315
Asian	25 (3.2)	15 (2.5)	
Black	43 (5.5)	64 (10.8)	
Hispanic	26 (3.3)	38 (6.4)	
Mexican	45 (5.8)	35 (5.9)	
White	630 (80.6)	416 (69.9)	
Other	13 (1.7)	27 (4.5)	

Assessing balance: example

Education (%)				0.512
8th Grade	59 (7.5)	33 (5.5)		
9 - 11th Grade	71 (9.1)	120 (20.2)		
High School	152 (19.4)	151 (25.4)		
Some College	256 (32.7)	210 (35.3)		
College Grad	244 (31.2)	81 (13.6)		
MaritalStatus (%)				0.488
Divorced	85 (10.9)	77 (12.9)		
LivePartner	61 (7.8)	96 (16.1)		
Married	453 (57.9)	240 (40.3)		
NeverMarried	108 (13.8)	142 (23.9)		
Separated	6 (0.8)	14 (2.4)		
Widowed	69 (8.8)	26 (4.4)		
Poverty (mean (sd))	3.11 (1.65)	2.38 (1.58)		0.453

Assessing balance: example

Assessing balance – original sample eCDFs in smokers and non-smokers for age:

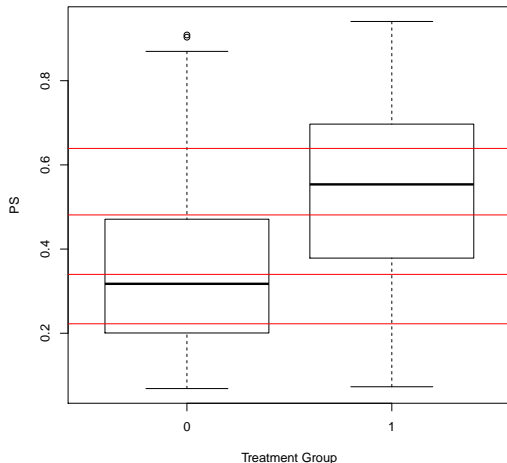


Assessing balance: example

```
> ps.mod <- glm(SmokeNow~Gender+Age+Race3+Education+
  MaritalStatus+HHIncome+Poverty,
  data=small.nhanes,family="binomial")
> ps.lr <- predict(ps.mod,type="response")
> summary(ps.lr)
>
> boxplot(ps.lr[small.nhanes$SmokeNow==0],
  ps.lr[small.nhanes$SmokeNow==1],
  ylab="PS",xlab="Treatment Group",names=c(0,1))
> quints <- c(0,quantile(ps.lr,seq(.2,1,.2)))
> abline(h=quints[2:5],col="red")
```

Assessing balance: example

Boxplots of propensity scores estimated via logistic regression (red lines indicate quintiles of the estimated PS distribution):



Assessing balance: example

The overlap is a bit better than the boxplots suggest:

```
> rbind(table(cut(ps.lnr[small.nhanes$SmokeNow==0],quints)),
        table(cut(ps.lnr[small.nhanes$SmokeNow==1],quints)))
```

(0,0.22]	(0.22,0.34]	(0.34,0.48]	(0.48,0.64]	(0.64,0.94]
231	194	167	121	69
47	82	105	157	204

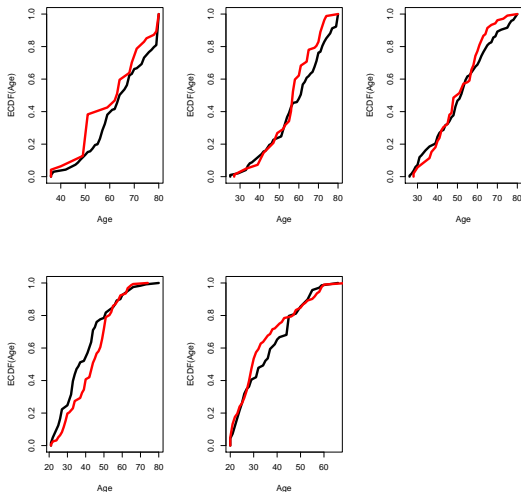
We can therefore proceed to check for balance knowing we have sufficient numbers of smokers and non-smokers in each quintile to ensure the stratum-specific estimates are not too unstable.

Table 1. Standardized mean differences: NHANES.

Var.	PS Quintiles					
	Orig.	Q1	Q2	Q3	Q4	Q5
Gender	0.138	0.102	0.104	0.029	0.200	0.031
Age	0.592	0.257	0.171	0.099	0.311	0.164
Race	0.315	0.317	0.112	0.344	0.415	0.287
Educ.	0.512	0.538	0.417	0.280	0.238	0.302
Marital	0.488	0.432	0.239	0.272	0.233	0.261
Poverty	0.453	0.087	0.126	0.114	0.004	0.146

Assessing balance: example

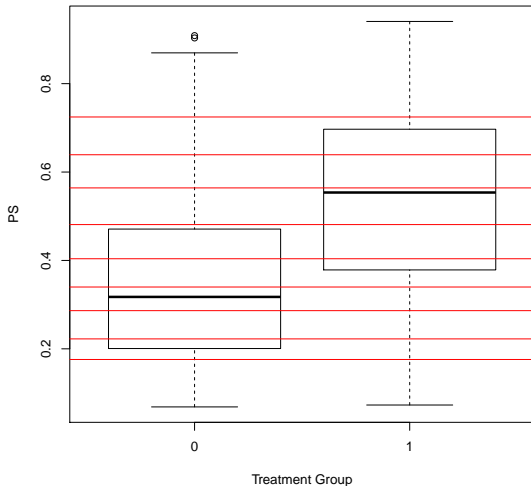
Assessing balance – eCDFs within quintiles of PS in smokers and non-smokers for age:



- ▶ Balance does not appear to have been achieved:
 - ▶ SMDs > 0.1 for at least three quintiles for all variables.
 - ▶ The empirical CDFs of age do not overlap in several quintiles.
- ▶ Should we try for finer strata?

Assessing balance: example

Boxplots of PS estimated via logistic regression (red lines indicate *deciles* of the estimated PS distribution):



Assessing balance: example

First decile has 20 smokers: still reasonable overlap.

Is balance improved?

```
> ps.lr.dec <- cut(ps.lr,dec,labels=1:10)
> SMD.10.table <- ExtractSmd(tabUnmatched)
> for(j in 1:10) {
  tabPSdec <- CreateTableOne(vars = vars, strata = "SmokeNow"
    data = small.nhanes[ps.lr.dec==j,], test = FALSE)
  SMD.10.table <- cbind(SMD.10.table,ExtractSmd(tabPSdec))
}
>
> summary(as.vector(SMD.10.table))
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.01194 0.12220 0.31360 0.31570 0.47980 0.78460
```

Most variables still showing significant imbalance. What if we were to use the PS in another way?

Assessing balance: example

Let's look at matching and IPW.

```
> ps.lr.match <- Match(Tr=small.nhanes$SmokeNow,  
  X=small.nhanes$ps.lr,estimand="ATE",ties=FALSE)  
> matched.samp <- small.nhanes[c(ps.lr.match$index.control,  
  ps.lr.match$index.treated),]  
> table(table(c(ps.lr.match$index.control,  
  ps.lr.match$index.treated)))  
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  
15  16  23  29  38  
851 246 131  47  42  18   7   5   2   4   5   4   3   3  
5   1   1   1   1
```

```
> tabMatched <- CreateTableOne(vars = vars, strata = "SmokeNow  
  data = matched.samp, test = FALSE)  
> MatchBalance(SmokeNow ~ Gender+Age+Race3+Education+  
  MaritalStatus+HHIncome+Poverty,data=small.nhanes,  
  match.out=ps.lr.match)
```

Assessing balance: example

The function `MatchBalance` from the `Matching` library provides many more details than `CreateTableOne`, including:

- ▶ mean, median, and maximum difference in empirical CDF plots,
- ▶ mean, median, and maximum difference in empirical QQ plots,
- ▶ Kolmogorov-Smirnov statistics,
- ▶ ratio of variances,
- ▶ p-value for t-test.
- ▶ Note that SMDs are $\times 100$.

```
> temp0 <- Ecdf(matched.samp$Age[matched.samp$SmokeNow==0])
> temp1 <- Ecdf(matched.samp$Age[matched.samp$SmokeNow==1])
> plot(temp0$x,temp0$y,ylab="ECDF (Age)",xlab="Age",main="",
      type="l",lwd=3)
> lines(temp1$x,temp1$y,col="red",lwd=3)
```

Assessing balance: example

```
> ps.lr.weight <- small.nhanes$SmokeNow/ps.lr +  
  (1-small.nhanes$SmokeNow)/(1-ps.lr)  
>  
> nhanes.IPW.lr <- svydesign(ids=~0, data=small.nhanes,  
  weights=ps.lr.weight)  
> tabIPW <- svyCreateTableOne(vars = vars, strata = "SmokeNow"  
  data = nhanes.IPW.lr, test = FALSE)  
> print(tabIPW, smd = TRUE)  
>  
> temp0 <- Ecdf(small.nhanes$Age[small.nhanes$SmokeNow==0],  
  weights=ps.lr.weight[small.nhanes$SmokeNow==0])  
> temp1 <- Ecdf(small.nhanes$Age[small.nhanes$SmokeNow==1],  
  weights=ps.lr.weight[small.nhanes$SmokeNow==1])  
> plot(temp0$x,temp0$y,ylab="ECDF (Age)",xlab="Age",main="",  
  type="l",lwd=3)  
> lines(temp1$x,temp1$y,col="red",lwd=3)
```

Table 1, con't. Standardized mean differences: NHANES.

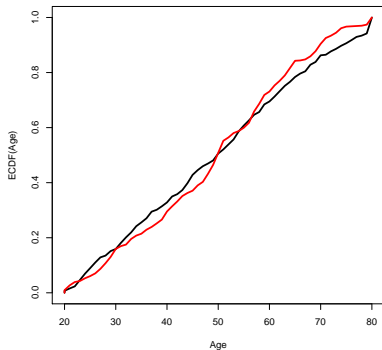
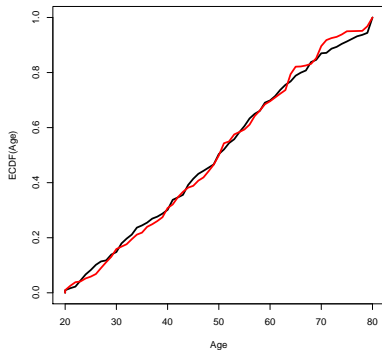
Var.	PS Quintiles						Match	IPW
	Orig.	Q1	Q2	Q3	Q4	Q5		
Gender	0.138	0.102	0.104	0.029	0.200	0.031	0.006	0.023
Age	0.592	0.257	0.171	0.099	0.311	0.164	0.002	0.014
Race	0.315	0.317	0.112	0.344	0.415	0.287	0.120	0.052
Educ.	0.512	0.538	0.417	0.280	0.238	0.302	0.133	0.029
Marital	0.488	0.432	0.239	0.272	0.233	0.261	0.094	0.023
Poverty	0.453	0.087	0.126	0.114	0.004	0.146	0.049	0.000

Table 1, con't. Standardized mean differences: NHANES.

Var.	PS Quintiles						Match	IPW
	Orig.	Q1	Q2	Q3	Q4	Q5		
Gender	0.138	0.102	0.104	0.029	0.200	0.031	0.006	0.023
Age	0.592	0.257	0.171	0.099	0.311	0.164	0.002	0.014
Race	0.315	0.317	0.112	0.344	0.415	0.287	0.120	0.052
Educ.	0.512	0.538	0.417	0.280	0.238	0.302	0.133	0.029
Marital	0.488	0.432	0.239	0.272	0.233	0.261	0.094	0.023
Poverty	0.453	0.087	0.126	0.114	0.004	0.146	0.049	0.000

Assessing balance: example

Assessing balance – eCDFs in smokers and non-smokers for age, matched and IPW:



Some authors have argued in favour of more complex or flexible methods of estimating the PS. Ridgeway and McCaffrey (2007), for example, recommend generalized boosted models (GBM):

```
> gbm.fit <- ps(SmokeNow ~ Gender+Age+Race3+Education+  
               MaritalStatus+HHIncome+Poverty, data=small.nhanes)  
> ps.gbm <- gbm.fit$ps$ks.mean.ATE
```

Let us check balance and overlap using this fit.

Assessing balance: example

Using GBM: boxplots of propensity scores estimated via logistic regression (red lines indicate quintiles of the estimated PS distribution, 1st quintile has 17 smokers):

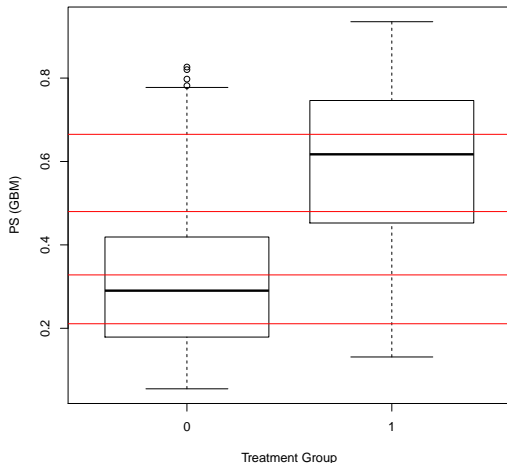


Table 2. Standardized mean differences using PS estimated by GBM.

Var.	PS Quintiles						Match	IPW
	Orig.	Q1	Q2	Q3	Q4	Q5		
Gender	0.138	0.040	0.168	0.055	0.262	0.109	0.108	0.065
Age	0.592	0.027	0.450	0.258	0.301	0.142	0.211	0.168
Race	0.315	0.512	0.434	0.166	0.308	0.693	0.304	0.116
Educ.	0.512	0.576	0.267	0.369	0.440	0.702	0.253	0.153
Marital	0.488	0.716	0.682	0.652	0.323	0.499	0.144	0.156
Poverty	0.453	0.372	0.118	0.193	0.556	0.099	0.237	0.096

Balance is noticeably worse than under a PS estimated by logistic regression.

We shall try one additional approach, Super Learning (SL):

```
> X.mat <- data.frame(cbind(small.nhanes$Gender,
  small.nhanes$Age, small.nhanes$Race3,
  small.nhanes$Education, small.nhanes$MaritalStatus,
  small.nhanes$HHIncome, small.nhanes$Poverty))
> my.library <- c("SL.knn", "SL.randomForest", "SL.glmnet",
  "SL.mean")
> SL.fit <- SuperLearner(Y = small.nhanes$SmokeNow,
  X = X.mat, SL.library = my.library, verbose = TRUE,
  method = "method.NNLS", family = binomial())
> ps.SL <- SL.fit$SL.predict
```

Let us again check balance and overlap using this fit.

Assessing balance: example

Using SL: boxplots of propensity scores estimated via logistic regression (red lines indicate quintiles of the estimated PS distribution):

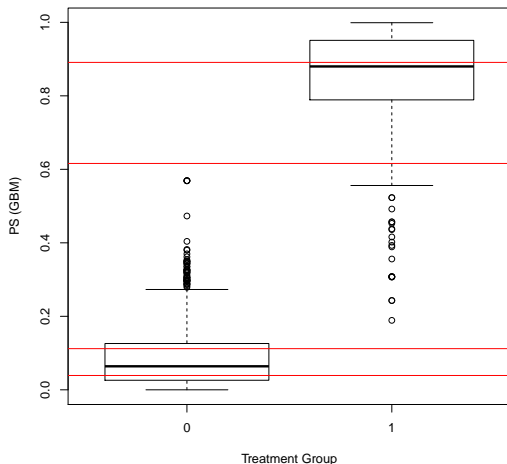


Table 3. Standardized mean differences using PS estimated by SL.

Var.	PS Quintiles						Match	IPW
	Orig.	Q1	Q2	Q3	Q4	Q5		
Gender	0.138	-	-	0.539	-	-	1.123	0.116
Age	0.592	-	-	0.507	-	-	0.174	0.488
Race	0.315	-	-	1.061	-	-	0.131	0.249
Educ.	0.512	-	-	1.142	-	-	1.551	0.408
Marital	0.488	-	-	0.789	-	-	0.218	0.382
Poverty	0.453	-	-	1.719	-	-	0.004	0.293

Key ideas:

- ▶ Creating or restoring confounder balance is essential to estimating a causal effect.
- ▶ It can be hard to assess overlap or achieve balance in high dimensions.
- ▶ The propensity score, a scalar summary of confounding variables, simplifies this task.
- ▶ However:
 - ▶ fitting a model for treatment does not guarantee balance,
 - ▶ fitting a model that predicts treatment with a high degree of precision can be unhelpful.

Let's proceed now to estimating the ATE, using:

- ▶ outcome regression,
- ▶ PS stratification,
- ▶ PS matching,
- ▶ PS regression,
- ▶ IPW.

We will use the PS estimated via logistic regression, as this provided the best balance.

Let us first look at regression coefficients:

```
> coef(lm(BPSysAve ~ SmokeNow, data=small.nhanes)) [2]
SmokeNow
-3.679357
```

```
> coef(lm(BPSysAve ~ SmokeNow+Gender+Age+Race3+
Education+MaritalStatus+HHIncome+Poverty,
data=small.nhanes)) [2]
SmokeNow
-1.097768
```

The naive conditional effect estimate is more than 3 times greater than its confounder-adjusted counterpart.

Now let's use the regression to obtain the ATE:

```
> nhanes.allsmoke <- small.nhanes
> nhanes.allsmoke$SmokeNow <- 1
> nhanes.nosmoke <- small.nhanes
> nhanes.nosmoke$SmokeNow <- 0

> mod1.lm <- lm(BPSysAve ~ SmokeNow+Gender+Age+Race3+
  Education+MaritalStatus+HHIncome+Poverty,
  data=small.nhanes)
> APO.lm.1 <- mean(predict(mod1.lm, nhanes.allsmoke))
> APO.lm.0 <- mean(predict(mod1.lm, nhanes.nosmoke))
> APO.lm.1 - APO.lm.0
[1] -1.097768
```

Conditional and marginal effect are the same in a linear model with no interaction!

With interactions:

```
> mod1.lmX <- lm(BPSysAve ~ SmokeNow+Gender+Age+Race3+Education+
  MaritalStatus+HHIncome+Poverty+SmokeNow:HHIncome+
  SmokeNow:Gender+SmokeNow:Age, data=small.nhanes)

> APO.lmX.1 <- mean(predict(mod1.lmX, nhanes.allsmoke))
> APO.lmX.0 <- mean(predict(mod1.lmX, nhanes.nosmoke))

> APO.lmX.1 - APO.lmX.0
[1] -1.402538
```

ATE: PS stratification

```
> ps.lr.quints <- cut(ps.lr,quints,labels=1:5)

> p.strat <- table(ps.lr.quints)/length(ps.lr.quints)
> p.strat
ps.lr.quints
           1           2           3           4           5
0.2018882 0.2004357 0.1975309 0.2018882 0.1982571

> ATE.strat <- rep(NA,5)
> for(j in 1:5) {
  ATE.strat[j] <-
    mean(BPSysAve[SmokeNow == 1 & ps.lr.quints==j]) -
    mean(BPSysAve[SmokeNow == 0 & ps.lr.quints==j])
}
> ATE.strat
[1] -8.1736207 -2.2701785 -0.2062732 -1.1820287  2.8633845
> sum(ATE.strat*p.strat)
[1] -1.816879
```

```
> ps.lr.match <- Match(Tr=small.nhanes$SmokeNow,  
  X=small.nhanes$ps.lr, estimand="ATE", ties=FALSE)  
> matched.samp <- small.nhanes[c(ps.lr.match$index.control,  
  ps.lr.match$index.treated),]  
  
> dim(matched.samp)  
[1] 2754  13  
  
> mean(matched.samp$BPSysAve[matched.samp$SmokeNow == 1]) -  
  mean(matched.samp$BPSysAve[matched.samp$SmokeNow == 0])  
[1] -0.4705882
```

ATE: PS regression

```
> mod1.PS1m1 <- lm(BPSysAve ~ SmokeNow+ps.lnr, data=small.nhanes)
> APO.PS1m1.1 <- mean(predict(mod1.PS1m1, nhanes.allsmoke))
> APO.PS1m1.0 <- mean(predict(mod1.PS1m1, nhanes.nosmoke))
> APO.PS1m1.1 - APO.PS1m1.0
[1] -1.10791

> mod1.PS1m2 <- lm(BPSysAve ~ SmokeNow+ps.lnr+I(ps.lnr^2),
  data=small.nhanes)
> APO.PS1m2.1 <- mean(predict(mod1.PS1m2, nhanes.allsmoke))
> APO.PS1m2.0 <- mean(predict(mod1.PS1m2, nhanes.nosmoke))
> APO.PS1m2.1 - APO.PS1m2.0
[1] -1.110337

> mod1.PS1m3 <- lm(BPSysAve ~ SmokeNow+bs(ps.lnr, df=4),
  data=small.nhanes)
> APO.PS1m3.1 <- mean(predict(mod1.PS1m3, nhanes.allsmoke))
> APO.PS1m3.0 <- mean(predict(mod1.PS1m3, nhanes.nosmoke))
> APO.PS1m3.1 - APO.PS1m3.0
[1] -1.133493
```

```
> ps.lr.weight <- SmokeNow/ps.lr + (1-SmokeNow)/(1-ps.lr)
> mean(SmokeNow*BPSysAve*ps.lr.weight) -
  mean((1-SmokeNow)*BPSysAve*ps.lr.weight)
[1] -1.928655

> coef(lm(BPSysAve ~ SmokeNow, weights = ps.lr.weight))
(Intercept)      SmokeNow
 124.237219     -1.991233
```

Table 4. Balance and ATE: a comparison across approaches.

	Outcome reg.	PS quints.	PS match	PS reg.	IPW
Max SMD	-	0.538	0.133	-	0.052
Mean SMD	-	0.221	0.067	-	0.024
Med. SMD	-	0.236	0.071	-	0.023
ATE	-1.098	-1.817	-0.471	-1.133	-1.929

For the ATT, we simply need to ensure that only the *exposed* are matched. This will reduce the sample size.

```
> matched.ATT <- Match(Y=small.nhanes$BPSysAve,
  Tr=small.nhanes$SmokeNow, X=ps.lf,
  estimand = "ATT", ties=FALSE)

> matched.samp.ATT <-
  small.nhanes[c(matched.ATT$index.control,
    matched.ATT$index.treated), ]
> dim(matched.samp.ATT)
[1] 1190   13

> mean(BPSysAve[SmokeNow == 1], data=matched.samp.ATT) -
  mean(BPSysAve[SmokeNow == 0], data=matched.samp.ATT)
[1] 0.6756303
```

Alternatively, we can estimate the ATT by re-weighting only the *unexposed* individuals by $\Pr(Z = 1|x)/\Pr(Z = 0|x)$:

```
> ATT.lr.weight <- small.nhanes$SmokeNow +  
  (1-small.nhanes$SmokeNow)*ps.lr/(1-ps.lr)  
  
> mean(SmokeNow*BPSysAve*ATT.lr.weight) -  
  mean((1-SmokeNow)*BPSysAve*ATT.lr.weight)  
[1] -0.3895692
```

Note that the estimated ATT under IPW is in the same direction as for the estimated ATE using all methods. The ATT has the opposite sign, and estimates vary considerably from one analysis to another (0.24-1.28).

Table 5. SMDs for estimation of the ATT.

Var.	Orig.	Match	IPW
Gender	0.138	0.110	0.016
Age	0.592	0.002	0.031
Race	0.315	0.158	0.110
Educ.	0.512	0.067	0.065
Marital	0.488	0.192	0.034
Poverty	0.453	0.064	0.006

Assessing balance: ATT example

Note that only the untreated are weighted:

```
> print(ATT.IPW, smd = TRUE)
```

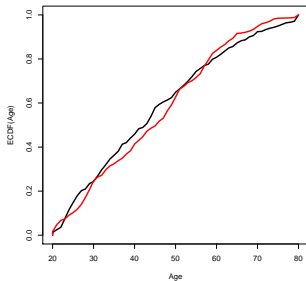
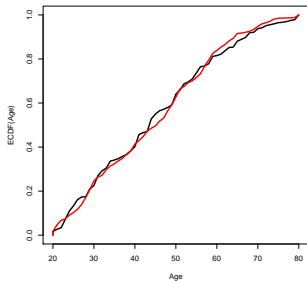
	Stratified by SmokeNow		
	0	1	SMD
n	597.0	595.0	
Gender = male (%)	365.4 (61.2)	369.0 (62.0)	0.016
Age (mean (sd))	44.46 (16.35)	44.96 (15.11)	0.031
Race3 (%)			0.110
Asian	15.4 (2.6)	15.0 (2.5)	
Black	76.0 (12.7)	64.0 (10.8)	
Hispanic	49.5 (8.3)	38.0 (6.4)	
Mexican	36.2 (6.1)	35.0 (5.9)	
White	389.8 (65.3)	416.0 (69.9)	
Other	30.0 (5.0)	27.0 (4.5)	

Assessing balance: ATT example

Education (%)				0.065
8th Grade	30.4 (5.1)	33.0 (5.5)		
9 - 11th Grade	113.6 (19.0)	120.0 (20.2)		
High School	141.9 (23.8)	151.0 (25.4)		
Some College	226.7 (38.0)	210.0 (35.3)		
College Grad	84.4 (14.1)	81.0 (13.6)		
MaritalStatus (%)				0.034
Divorced	76.0 (12.7)	77.0 (12.9)		
LivePartner	93.6 (15.7)	96.0 (16.1)		
Married	242.0 (40.5)	240.0 (40.3)		
NeverMarried	141.9 (23.8)	142.0 (23.9)		
Separated	17.1 (2.9)	14.0 (2.4)		
Widowed	26.4 (4.4)	26.0 (4.4)		
Poverty (mean (sd))	2.39 (1.60)	2.38 (1.58)		0.006

Assessing balance: ATT example

Assessing balance – eCDFs in smokers and non-smokers for age, matched and IPW:



Key ideas:

- ▶ From a coding perspective, all approaches to estimating the ATE are straightforward.
- ▶ All approaches are not, however, equally likely to satisfy the assumption of correct model specification.
- ▶ Using a PS, it is much easier to assess balance prior to computing the ATE.
- ▶ How the PS is then used in the analysis should be carefully considered and cannot be judged based on concordance between observed estimates, as all are subject to differing degrees of variability and bias.

Additional considerations: SEs and CIs

- ▶ All of the PS approaches considered rely on substitution estimators.
 - ▶ In PS regression, we plug in an estimated PS as a covariate.
 - ▶ In IPW, we plug in estimated weights.
- ▶ We need to account for this when estimating standard errors and/or confidence intervals.
- ▶ Analytically derived asymptotic variances can be used, but are not provided in many standard software packages.
- ▶ The easiest approach is to bootstrap.
- ▶ Note, however, that the bootstrap is *not* valid for matching.

- ▶ If data are missing, one can either impute or (if only missing the outcome but not covariates), “censor” the individual.
- ▶ Censored data can easily be handled by incorporating weights for censoring into estimator or the regression model for any of the approaches that we have considered.

Key points: Summary

- ▶ In a real-data setting, caution must be taken to ensure balance is achieved.
- ▶ Model choices should be based on subject-matter knowledge to the greatest extent possible.
- ▶ Many subtle and not-so-subtle issues remain, and must be accounted for carefully.